Université Paris-Diderot (Paris 7)
Sorbonne Paris Cité

École doctorale de sciences mathématiques de Paris centre
Laboratoire de Probabilités, Statistique et Modélisation

Thèse de doctorat

Discipline : Mathématiques Appliquées

Présentée par
Lucas Benigni

Dynamique de vecteurs propres de matrices aléatoires et valeurs propres de modèles non-linéaires de matrices

Dynamics of eigenvectors of random matrices and eigenvalues of nonlinear matrix models

Sous la direction de **Paul Bourgade** et **Sandrine Péché**

Rapporteurs :

Jean-Philippe	**Bouchaud**	Capital Fund Management
László	**Erdős**	IST Austria

Présentée et soutenue à Paris le 20 juin 2019 devant le jury composé de

Giulio	Biroli	Directeur de recherche	École Normale Supérieure	Président du jury
Jean-Philippe	Bouchaud	Professeur	Capital Fund Management	Rapporteur
Paul	Bourgade	Professeur	New York University	Directeur
Sandrine	Péché	Professeure	Université Paris-Diderot	Directrice
Kevin	Schnelli	Professeur	KTH Royal Institute of Technology	Examinateur
Pierre	Youssef	Maître de conférences	Université Paris-Diderot	Examinateur

À César et Frédéric,

Remerciements

Mes plus vifs remerciements vont tout d'abord à Sandrine Péché et Paul Bourgade pour avoir dirigé cette thèse. Merci Paul pour avoir encadré mon année de recherche pré-doctorale puis co-encadré cette thèse et pour m'avoir énormément appris sur les mathématiques. Cette thèse n'aurait pas pu voir le jour sans ton expertise, tes réponses à mes questions incessantes et ta grande attention. Tu m'as aussi appris à faire partie intégrante de cette communauté par tes nombreux conseils et je t'en remercie. Sandrine, je voulais déjà te remercier pour avoir choisi de m'encadrer à l'aveugle et pour avoir ensuite partagé ton intuition et tes connaissances avec moi. Je te remercie aussi de m'avoir proposé un sujet très intéressant. Ce fut un réel plaisir de collaborer avec toi même si tu as eu le travail difficile de déchiffrer mes manuscrits de combinatoire incompréhensibles.

C'est un honneur pour moi que Jean-Philippe Bouchaud et Lázló Erdős aient accepté de rapporter cette thèse. Je les remercie d'avoir soumis leur expertise et leur attention à ce manuscrit. Je tiens aussi à remercier chaleureusement Giulio Biroli, Kevin Schnelli et Pierre Youssef. Je suis ravi d'avoir l'opportunité de présenter mes travaux devant vous.

I would like to thank Jun Yin for inviting me to speak at the UCLA Probability Seminar and Nick Cook for the stimulating discussion I had with him there. I am grateful for Lázló Erdős's invitation to speak at the Probability and Mathematical Physics Seminar at IST Austria. It was an enriching experience to converse with the random matrix team, especially with Giorgio Cipolloni and Dominik Schröder. I also wish to thank Maciej Nowak for bringing me to Krakow for a wonderful conference.

J'ai eu la chance de travailler dans ce laboratoire dans une excellente ambiance qui ne serait pas la même sans tous les doctorants et anciens-doctorants avec qui j'ai intéragi. Merci à tous les doctorants et postdocs de Paris 7 : Mina Abdel-Sayed, Cyril Bénézet, Clément Bonvoisin, Bohdan Bulanyi, S. C. et tous les théorèmes que l'on a renommés, Junchao Chen, Yann Chiffaudel, Guillaume Conchon--Kerjan, Fabio Coppini, Clément Cosco et nos moments dans l'ascenseur, Rémy Degenne, Barbara Dembin, Aurélia Deshayes, Mi-Song Dupuy et les instants où j'essayais de calmer sa rage, Sothea Has, Benjamin Havret et sa parfaite manière de lancer ses blagues, Hiroshi Horii, Côme Huré qui sait quand je repars pour New York, Chris Janjigian, Ziad Kobeissi, David Krief, William Lefebvre, Houzhi Li, Laure Marêché et ses chocolats sur son bureau, Enzo Miller, Marc Pegon, Luca Prezioso, Assaf Shapira et tout le temps écoulé dans nos discussions entre nos deux bureaux, Erik Slivken, Xiaoli Wei, Yiyang Yu et Arturo Zamorategui. Je tiens aussi à remercier les autres doctorants que j'ai croisés au Courant Institute, à Jussieu ou en conférence : Carlo Bellingeri, Jeanne Boursier et nos histoires palpitantes dans le bureau des visiteurs, Raphaël Butez, Guillaume Dubach et la très mignonne Louise, Henri Elad-Altman, Alejandro Fernandez Montero, Reza Gheissari, Jonathan Husson, Alexandre Krajenbrink, Thibaut Lemoine, Zhenyu Liao, Patrick Lopatto, Ben McKenna and his beautiful PCMI shirt, Paul Melotti, Krishnan Mody and his underappreciated New Yorker captions, Nathan Noiry, Michel Pain, Davide Parise, Othmane Safsafi et Yizhe Zhu. Je veux aussi exprimer toute ma gratitude aux personnes qui ont facilité ma vie administrative de chaque côté de l'Atlantique: Nathalie Bergame, Ross Edwards, Valérie Juvé et Julian Lopez.

D'un point de vue plus personnel, je voudrais remercier tous mes amis qui m'ont détendus pendant cette thèse. Michael et Julien, même si nos mathématiques sont éloignées, je n'ai aucun doute sur l'impact que vous avez eu sur ce travail et ma vie ces dernières années. Je pense bien sur à toutes les belles aventures, parisiennes ou non, que nous avons partagées mais aussi à toutes celles que nous vivrons. Je profite aussi de cette opportunité pour remercier toute la bande de Joffre : Alexandre, Antoine, Brice, Constantin, Farouk, Quentin et Thomas ainsi que Jonathan, Nicolas et Roxane pour toutes ces fêtes du Nouvel An.

Tout ce parcours dont cette thèse de doctorat est une des étapes finales n'aurait pas été possible sans le soutien incommensurable de toute ma famille, qu'elle soit de coeur ou de sang. Merci à ma mère qui m'a toujours rassuré de son oeil bienveillant, mon père qui m'a aidé en mathématiques jusqu'au lycée, mon beau-père qui a écouté mes comptes-rendus de colles chaque semaine, mon frère qui m'a appris les multiplications, mes soeurs et leur tribu que j'ai vue grandir ainsi que mes cousins, oncles et tantes et finalement ma grand-mère chez qui j'ai toujours ressenti une grande fierté.

I would like to thank the Rycz and Pfeiffer families for giving me a warm welcome and tremendous help in New York. Finally, Christina, thank you for all the happiness, love and support you brought in my life during these three years. Making every moment with or without you the best it could be.

Résumé

Cette thèse est constituée de deux parties indépendantes. La première partie concerne l'étude des vecteurs propres de matrices aléatoires de type Wigner. Dans un premier temps, nous étudions la distribution des vecteurs propres de matrices de Wigner déformées, elles consistent en une perturbation d'une matrice de Wigner par une matrice diagonale déterministe. Si les deux matrices sont du même ordre de grandeur, il a été prouvé que les vecteurs propres se délocalisent complètement et les valeurs propres rentrent dans la classe d'universalité de Wigner-Dyson-Mehta. Nous étudions ici une phase intermédiaire où la perturbation déterministe domine l'aléa: les vecteurs propres ne sont pas totalement délocalisés alors que les valeurs propres restent universelles. Les entrées des vecteurs propres sont asymptotiquement gaussiennes avec une variance qui les localise dans une partie explicite du spectre. De plus, leur masse est concentrée autour de cette variance dans le sens d'une unique ergodicité quantique. Ensuite, nous étudions des corrélations de différents vecteur propres. Pour se faire, une nouvelle observable sur les moments de vecteurs propres du mouvement brownien de Dyson est étudiée. Elle suit une équation parabolique close qui est un pendant fermionique du flot des moments de vecteurs propres de Bourgade-Yau. En combinant l'étude de ces deux observables, il est possible d'analyser certaines corrélations. La deuxième partie concerne l'étude de la distribution des valeurs propres de modèles non-linéaires de matrices aléatoires. Ces modèles apparaissent dans l'étude de réseaux de neurones aléatoires et correspondent à une version non-linéaire de matrice de covariance dans le sens où une fonction non-linéaire, appelée fonction d'activation, est appliquée entrée par entrée sur la matrice. La distribution des valeurs propres convergent vers une distribution déterministe caractérisée par une équation auto-consistante de degré 4 sur sa transformée de Stieltjes. La distribution ne dépend de la fonction que sur deux paramètres explicites et pour certains choix de paramètres nous retrouvons la distribution de Marchenko-Pastur qui reste stable après passage sous plusieurs couches du réseau de neurones.

Mots-clés : matrices aléatoires ; vecteurs propres ; universalité ; unique ergodicité quantique ; réseaux de neurones ; méthode des moments.

Abstract

This thesis consists in two independent parts. The first part pertains to the study of eigenvectors of random matrices of Wigner-type. Firstly, we analyze the distribution of eigenvectors of deformed Wigner matrices which consist in a perturbation of a Wigner matrix by a deterministic diagonal matrix. If the two matrices are of the same order of magnitude, it was proved that eigenvectors are completely delocalized and eigenvalues belongs to the Wigner-Dyson-Mehta universality class. We study here an intermediary phase where the deterministic perturbation dominates the randomness of the Wigner matrix : eigenvectors are not completely delocalized but eigenvalues are still universal. The eigenvector entries are asymptotically Gaussian with a variance which localize them onto an explicit part of the spectrum. Moreover, their mass is concentrated around their variance in a sense of a quantum unique ergodicity property. Then, we consider correlations of different eigenvectors. To do so, we exhibit a new observable on eigenvector moments of the Dyson Brownian motion. It follows a closed parabolic equation which is a fermionic counterpart of the Bourgade-Yau eigenvector moment flow. By combining the study of these two observables, it becomes possible to study some eigenvector correlations. The second part concerns the study of eigenvalue distribution of nonlinear models of random matrices. These models appear in the study of random neural networks and correspond to a nonlinear version of sample covariance matrices in the sense that a nonlinear function, called the activation function, is applied entrywise to the matrix. The empirical eigenvalue distribution converges to a deterministic distribution characterized by a self-consistent equation of degree 4 followed by its Stieltjes transform. The distribution depends on the function only through two explicit parameters. For a specific choice of these parameters, we recover the Marchenko-Pastur distribution which stays stable after going through several layers of the network.

Keywords: random matrices; eigenvectors; universality; quantum unique ergodicity; neural networks; moment method.

Contents

Part I

Dynamics of eigenvectors of random matrices

Chapter 1

Introduction

The study of eigenvalue statistics of large random matrices was spurred by the seminal work of Eugene Wigner [Wig55] in the 1950s. Wigner first observed that empirical data coming from the study of heavy nuclei seemed independent of the material and conjectured the universality of the gap statistics for energy levels. From the point of view of quantum mechanics, the energy levels are eigenvalues of a self-adjoint operator which, at the time, was neither explicitly known nor practically computable. Wigner's main idea was then to replace the unknown Hamiltonian by a large random matrix with independent entries while conserving the symmetry type of the system. Real symmetric random matrices were used to model systems with time reversal symmetry while Hermitian random matrices systems lacking this symmetry. This approximation was an outstanding success and unearthed an intricate universality principle.

While Wigner first studied the global statistics of these random matrices, the gap statistics of energy levels should be derived from the study of individual eigenvalues and not of the whole spectrum. In 1957, Wigner guessed at a conference the distribution of the spacings of energy levels and introduced what is now known as the *Wigner surmise*. It is only a few years later that Mehta showed [Meh60] that, while being a good approximation, it could not be the correct asymptotic distribution. Gaudin and Mehta [MG60] then gave explicitly the spacings distribution for Gaussian self-adjoint random matrices. These random matrices ensembles are integrable models that allows one to derive, by comparison, local spectral statistics for more general ensembles such as the Wigner ensemble. The Gaussian Orthogonal (respectively Unitary) Ensemble is described by a probability density on the space of symmetric (respectively Hermitian) matrices endowed with the Lebesgue measure given by

$$\mathbb{P}_{\mathrm{GOE}}(\mathrm{d}H) = \frac{1}{Z_N^1} e^{-\frac{N}{4}\mathrm{Tr}(H^2)}\mathrm{d}H \quad \text{on the space of } N\times N \text{ symmetric matrices,}$$
$$\mathbb{P}_{\mathrm{GUE}}(\mathrm{d}H) = \frac{1}{Z_N^2} e^{-\frac{N}{2}\mathrm{Tr}(H^2)}\mathrm{d}H \quad \text{on the space of } N\times N \text{ Hermitian matrices.}$$

Consider H a matrix from the GUE, with our normalization, the empirical spectral distribution converges almost surely to the semicircle distribution in the followginse sense, if we write $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)$ the ordered eigenvalues of H, we have the following convergence in probability [Wig58]

$$\mu := \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i} \xrightarrow[N\to\infty]{\mathbb{P}} \rho_{\mathrm{sc}} \quad \text{with} \quad \rho_{\mathrm{sc}}(\mathrm{d}x) = \frac{1}{2\pi}\sqrt{4-x^2}\mathbb{1}_{|x|\leqslant 2}\mathrm{d}x. \tag{1.0.1}$$

On a microscopic scale we have the following convergence in distribution: for any fixed and small enough κ and for any $E \in (-2+\kappa, 2-\kappa)$ (in other words in the bulk of the spectrum) we have

$$\sum_{k=1}^N \delta_{N\rho_{\mathrm{sc}}(E)(\lambda_k - E)} \xrightarrow[N\to\infty]{(d)} \chi_{\mathrm{GUE}} \tag{1.0.2}$$

where χ_{GUE} is a determinantal translation invariant point process given by the kernel

$$K(x,y) = \frac{\sin(\pi(x-y))}{\pi(x-y)} \quad \text{for } x\neq y \quad \text{and} \quad K(x,x) = \frac{1}{\pi}. \tag{1.0.3}$$

This analysis was then continued by Dyson and Mehta [Dys62b, Meh71] for correlation functions of eigenvalues and for all symmetry classes of Gaussian random matrices. Even though these estimates hold for eigenvalue in the bulk of the spectrum, similar results hold for eigenvalue at the edge. In the case of the integrable Gaussian ensembles, Tracy and Widom exhibited [TW94a, TW96] the law of the fluctuations of the largest eigenvalue. In other words, we have the following convergence in distribution,

$$N^{2/3}(\lambda_N - 2) \to \mathrm{TW}_{\mathrm{GUE}} \tag{1.0.4}$$

where the cumulative distribution function of $\mathrm{TW}_{\mathrm{GUE}}$ is a Fredholm determinant whose kernel is given in terms of Airy functions. Note that a similar convergence holds for symmetric ensembles.

1.1. Invariant Ensembles

Wigner's vision of universality is broader than matrices with independent entries and universality for *invariant ensembles* was also studied. They correspond to matrices given by a probability density on the space of Hermitian or symmetric matrices with respect to the Lebesgue measure of the form

$$\mathbb{P}_{\mathrm{Inv}}(\mathrm{d}H) = \frac{1}{Z_N} e^{-N\mathrm{Tr}V(H)}\mathrm{d}H. \tag{1.1.1}$$

This distribution is called invariant because it is invariant with respect to orthogonal or unitary conjugaison. Note also that one recovers the GOE/GUE for specific quadratic potential V, and is the only invariant matrix with independent entries.

Universality for eigenvalues in this context corresponds to convergence of the k-point correlation function through determinantal formulas. Let $p_N(\lambda_1, \dots \lambda_N)$ be the joint probability density of the unordered eigenvalues of a $N \times N$ random matrix, the k-point correlation functions are then defined by

$$p_N^{(k)}(\lambda_1, \dots, \lambda_k) = \int_{\mathbb{R}^{N-k}} p_N(\lambda_1, \dots, \lambda_k, \lambda_{k+1}, \dots, \lambda_N) \prod_{i=k+1}^{N} \mathrm{d}\lambda_i. \tag{1.1.2}$$

It has been shown by Dyson in [Dys70] that the correlations functions of a random matrix from the invariant ensembles can be written as a determinant in the following way,

$$p_N^{(k)}(\lambda_2 = 1, \dots, \lambda_k) = \frac{(N-k)!}{N!} N^{k/2} \det\left[K_N(\sqrt{N}\lambda_i, \sqrt{N}\lambda_j\right]_{i,j=1}^{k}$$

with K_N being given in terms of orthogonal polynomials with respect to weights depending on V. It now remains to show convergence of this kernel to the sine kernel (1.0.3) in the bulk of the spectrum or the Airy kernel to show convergence at the edge (1.0.4) which is given by the following theorem.

Theorem 1.1.1 (Pointwise convergence of K_N in the case of Hermitian matrices)**.** *Under technical assumptions on the potential V, let ρ be the asymptotic distribution of the eigenvalues ($\rho = \rho_{\mathrm{sc}}$ for $V(x) = x^2/2$), we have for bulk universality, uniformly in x and y*

$$\frac{1}{\rho(0)} \frac{1}{\sqrt{N-1}} K_N\left(\frac{x}{\rho(0)\sqrt{N}}, \frac{y}{\rho(0)\sqrt{N}}\right) \to K(x,y) := \frac{\sin(\pi(x-y))}{\pi(x-y)}.$$

As for edge universality we have, uniformly in x and y,

$$\frac{c_N}{\alpha_N N^{2/3}} K_N\left(c_N\left(1 + \frac{x}{\alpha_N N^{2/3}}\right), c_N\left(1 + \frac{y}{\alpha_N N^{2/3}}\right)\right) \to A(x,y) := \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\mathrm{Ai}(y)}{x-y}$$

where Ai *is the solution to the Painlevé equation $u''(x) - xu(x) = 0$ with vanishing boundary conditions at infinity and α_N and c_N are constant determining the position of the largest eigenvalue. In the case of a quadratic potential, we have $c_N = 2\sqrt{N}$ and $\alpha_N = 2$.*

These convergences were shown using Riemann-Hilbert problems to obtain asymptotics of the orthogonal polynomials which describe the kernels. This has been done in a series of work by different teams of people [BI99], [DKM+99, DG07b, DG07a], [PS97, PS03, PS08].

While this covers local eigenvalue statistics for invariant ensembles, universality for eigenvectors for this model is trivial. Indeed, the invariant ensembles are, by definition, invariant by orthogonal or unitary conjugaison which makes their eigenvectors Haar-distributed. Indeed, let $\mathbf{u} = (u_1, \dots, u_N)$ be an eigenbasis of a matrix from the symmetric invariant ensemble then for any $O \in O(N)$, $O\mathbf{u}$ has the same distribution as $\mathbf{u}$. As a consequence, any eigenvector u_k is distributed uniformly on the sphere $\mathbb{S}^{n-1}$.

We can also widen the class of invariant ensembles by constructing a probability measure on the simplex $\{\boldsymbol{\lambda} = (\lambda_1 \dots, \lambda_N),\ \lambda_1 \leqslant \lambda_N\}$ to generalize eigenvalues of invariant ensembles. Indeed, if we consider

$$\mathrm{d}\mu_\beta(\boldsymbol{\lambda}) = \frac{1}{Z_{N,\beta}} \exp\left(-\beta N V(\boldsymbol{\lambda})\right) \quad \text{with} \quad V(\boldsymbol{\lambda}) = V_0(\boldsymbol{\lambda}) - \frac{1}{N} \sum_{1 \leqslant i < j \leqslant N} \log(\lambda_i - \lambda_j)$$

then for $\beta = 1$, 2 or 4, we would obtain the distribution of eigenvalues of invariant ensembles. This point process distribution is called *β-ensembles* or *log-gases*. In the case of a quadratic potential, it has been shown in [DE02] that μ_β is the distribution of eigenvalues of certain tridiagonal matrix for any and is called the β-Gaussian ensembles. For a general potential, while not being eigenvalues of random matrices, the point process still follows a universal behavior linked to Gaussian ensembles.

Theorem 1.1.2 (Universality for β-ensembles)**.** *Let ρ be the equilibrium measure supported on an interval $[A, B]$. For bulk universality, let $E \in (A, B)$ be inside the support and let $E' \in (-2, 2)$. Then for some intervals I, and I' such that $|I| = |I'| = N^k$ for some $k \in (0, 1/2]$, we have for any smooth compacted supported function $O : \mathbb{R} \to \mathbb{R}$,*

$$\left| \mathbb{E}_{\mu_\beta} \frac{1}{N^k \rho(E)} \sum_{i \in I} O\left(\frac{N(\lambda_i - \lambda_{i+1})}{\rho(E)}\right) - \mathbb{E}_{G_\beta} \frac{1}{N^k \rho_{\mathrm{sc}}(E')} \sum_{i \in I'} O\left(\frac{N(\lambda_i - \lambda_{i+1})}{\rho_{\mathrm{sc}}(E')}\right) \right| = 0$$

where $\mathbb{E}_{G_\beta}$ corresponds to the β-Gaussian ensembles. For edge universality, we have the following convergence in distribution, for any fixed m,

$$\left(\frac{N}{2}\right)^{2/3} (\lambda_1 - A, \dots, \lambda_m - A) \to (\Lambda_1, \dots, \Lambda_m)$$

where $\Lambda_1, \dots, \Lambda_m$ are the smallest eigenvalues of the stochastic Airy operator on $\mathbb{R}_+$.

The dynamical method which we will present in the next section was applied to this model to show bulk universality in [BEY14b, BEY12] and to treat the edge case in [BEY14a]. Note that other methods were also used to treat these β-ensembles [Shc14, BFG15, KRV16, FG16].

1.2. Wigner ensembles

1.2.1. Presentation of the model and universality results

In the original work of Wigner, the entries of the random matrices considered were independent and identically distributed but not necessarily Gaussian. These ensembles were given the name of *Wigner ensembles* and can be defined in the following way,

Definition 1.2.1 (Wigner ensembles)**.** A real (respectively complex) $N \times N$ Wigner matrix is a symmetric (respectively Hermitian) matrix $W = (w_{ij})_{1 \leqslant i,j \leqslant N}$ whose entries are centered, have variance $1/N$, and are independent up to the symmetry constraint.

The first mathematical study of this ensemble was the convergence of the empirical spectral distribution and corresponds to a form of universality for global statistics. Indeed, Wigner showed that the convergence (1.0.1) holds for a more general type of entry distribution. While the global statistics are universal in the context of Wigner ensembles, the semicircle distribution is not the asymptotic empirical distribution to more general models which we can see in the next section.

The concept of universality for local eigenvalue statistics of random matrices taken from the Wigner ensembles, while being unearthed by Wigner, was originally stated by Mehta in his treatise [Meh67]. Define the k-point correlation functions as in (1.1.2), we can state bulk universality for symmetric Wigner matrices (and the same type of convergence holds for Hermitian Wigner matrices) as the following theorem.

Theorem 1.2.2 (Fixed energy universality for Wigner matrices)**.** *Under the assumptions* (1.2.7) *and* (1.2.8)*, for any continuous and compactly supported function* $O : \mathbb{R}^k \to \mathbb{R}$ *and for any* $\kappa > 0$ *we have uniformly in* $E \in [-2+\kappa, 2-\kappa]$,

$$\frac{1}{\rho_{\mathrm{sc}}(E)^k}\int O(\mathbf{v})p_N^{(k)}\left(E+\frac{\mathbf{v}}{N\rho_{\mathrm{sc}}(E)}\right)\mathrm{d}\mathbf{v} \xrightarrow[N\to\infty]{} \int O(\mathbf{v})p_{\mathrm{GOE}}^{(k)}(\mathbf{v})\mathrm{d}\mathbf{v} \tag{1.2.1}$$

with $p_{\mathrm{GOE}}^{(k)}$ *being the known limiting correlation functions for matrix from the Gaussian Orthogonal Ensemble.*

For Wigner matrices (and even a generalized model), a large series of work proved bulk universality for correlation functions. In [Joh01a], universality was proved for Gaussian divisible ensemble where a Hermitian Wigner matrix was perturbated by a large Gaussian component. The method used explicit Harish-Chandra-Itzykson-Zuber integral formulas from [BH96, BH97] available for Hermitian matrices. The perturbation of the original matrix by a Gaussian component was then refined in a series of work where the optimal relaxation time for the local equilibrium, first conjectured by Dyson in [Dys62a], was proved [ERSY10, EPR[+]10, ESY11]. In [TV11], universality was first proved for random matrices whose entry distribution matched the first four moments of Gaussian ensembles introducing a Green function comparison theorem. This method was then conjointly used with the relaxation method in [ERS[+]10] to prove universality for a large class or Wigner matrices. While this technique has been used for numerous models over the past few years, the mean-field Wigner matrices picture was completely understood with [EYY11, EYY12a, EYY12b] and finally in [BEYY16] where the present form (1.2.1) of bulk universality was shown. Note that universality can also be stated for the gap statistics and can be found in [EY15]. Fix n a positive integer, let $O : \mathbb{R}^n \to \mathbb{R}$ be any continuous and compactly supported test functions and j an index in the bulk in the spectrum,

$$(\mathbb{E}_W - \mathbb{E}_{\mathrm{GOE}})\left[O\left(N(\lambda_j-\lambda_{j+1}), N(\lambda_j-\lambda_{j+2}), \dots, N(\lambda_j-\lambda_{j+n})\right)\right] \xrightarrow[N\to\infty]{} 0. \tag{1.2.2}$$

As for the edge of the spectrum, one can state universality as the following theorem.

Theorem 1.2.3 (Edge universality for Wigner ensembles)**.** *For any* $k < 1/4$*, there exists a* $\chi > 0$*, such that for any fixed* $m \geqslant 1$ *and any smooth compactly supported function* O*, there exists* C *such that for any* N *and any* $I \subset [\![1, N^k]\!]$ *with* $|I| = m$,

$$\left|(\mathbb{E}_W - \mathbb{E}_{\mathrm{GOE}})O\left(\left(N^{2/3}i^{1/3}(\lambda_i-\gamma_i)\right)_{i\in I}\right)\right| \leqslant N^{-\chi}$$

where γ_i*'s are the quantiles of the semicircle distribution.*

Universality of the largest eigenvalue for Wigner matrices was first proved under some conditions on the moments of the entry distribution in [Sos99] studying large moments of matrix entries. This moment condition was improved in [EYY12b] and finally relaxed optimally in [LY14]. Note that the case of generalized Wigner matrix as in Definition 1.2.4 where variances can vary between entries was done in [BEY14a].

Universality for eigenvectors is however not as clear. Indeed comparison with the Gaussian ensembles, which are invariant by orthogonal or unitary conjugaison, would say that the strongest form of universality would be a form of asymptotic Haar distribution for the whole eigenbasis. However, the understanding of the behavior of eigenvectors is not yet as strong. We can instead state several different forms of universality for eigenvectors.

Gaussianity of projections: If we consider any deterministic sequence of indices $(k_N) \in [\![1, N]\!]$ and any deterministic sequence of normalized vector $\mathbf{q}_N$, we have the following convergence in distribution

$$\sqrt{N}\langle \mathbf{q}_N, u_{k_N}\rangle \xrightarrow[N\to\infty]{(d)} \mathcal{N}(0,1). \tag{1.2.3}$$

This form of convergence was first proved for Wigner matrices under moment conditions, namely, the entry distribution must have the same four moments has a normal random variable. The method consists in a Green function comparison theorem combined with a level repulsion estimate on the eigenvalues [TV12b, KY13b]. The moment condition was then relaxed in [BY17] where the flow of eigenvector moments under a certain dynamics on Wigner matrices was analyzed.

Complete delocalization: Eigenvectors of the Gaussian ensembles are completely delocalized in the sense that all their entries can not be greater than the typical size of $N^{-1/2}$, these bounds on the infinite norm hold with overwhelming probability in the following sense, for any large $D > 0$ and $\varepsilon > 0$,

$$\mathbb{P}\left(\|u_k\|_\infty \geqslant \frac{N^\varepsilon}{\sqrt{N}}\right) \leqslant N^{-D}. \tag{1.2.4}$$

In the case of Wigner matrices, this result has been proved using an optimal control of the Stieltjes transform on mesoscopic scales, called a *local law*, in [ESY09a, ESY09b]. Note that it has also been proved for numerous other models.

This form of delocalization has been improved in [VW15] under some concentration assumption on rows of the matrix. Indeed, they showed an optimal delocalization result for bulk and edge eigenvectors in the following sense, for any (D_1, D_2), there exist (C_1, C_2) such that,

$$\mathbb{P}\left(\|u_i\|_\infty \geqslant C_1\sqrt{\frac{\log N}{N}}\right) \leqslant N^{-D_1} \quad \text{for } i \text{ an index in the bulk of the spectrum,}$$
$$\mathbb{P}\left(\|u_i\|_\infty \geqslant C_2\frac{\log N}{\sqrt{N}}\right) \leqslant N^{-D_2} \quad \text{for } i \text{ an index at the edge of the spectrum.}$$

While complete delocalization of the form (1.2.4) controls peak of the eigenvector entries, it does not say anything about possible gaps in the entries. This type of delocalization, called *no-gaps delocalization*, was proved in [RV16] as the following statement, for any $\varepsilon \in (0,1)$, any $k \in [\![1, N]\!]$ and any set of indices $I \subset [\![1, N]\!]$ such that $|I| \geqslant \varepsilon N$, we have with high probability

$$\left(\sum_{\alpha\in I} |u_k(\alpha)|^2\right)^{1/2} \geqslant \phi(\varepsilon)\|u_k\|_2 \tag{1.2.5}$$

for some function $\phi : (0,1) \to (0,1)$.

Quantum unique ergodicity: If one considers the previous form of complete delocalization, eigenvectors could still be supported on a small fraction of the spectrum while verifying (1.2.4). Rudnick and Sarnack in [RS94] introduced another form of delocalization of eigenfunctions for the Laplace-Beltrami operator on negatively curved Riemannian manifolds. This quantum unique ergodicity has been proved for arithmetic surfaces [Hol10, HS10, Lin06] and a probabilistic version can be stated for eigenvectors of random matrices. In the case of Gaussian ensembles, since eigenvectors are uniformly distributed on the sphere, we have the following overwhelming probability bound for indices in the bulk of the spectrum. In other words, let α be a small positive constant and $k \in [\![\alpha N, (1-\alpha)N]\!]$ a $N-$dependent index and $I \subset [\![1, N]\!]$,

$$\mathbb{P}\left(\left|\sum_{\alpha\in I} u_k(\alpha)^2 - \frac{|I|}{N}\right| > N^\varepsilon\frac{\sqrt{|I|}}{N}\right) \leqslant N^{-D}. \tag{1.2.6}$$

This strong form of delocalization has not yet been proved with the optimal error $\sqrt{|I|}/N$ for Wigner matrices in full generality. However, this overwhelming probability bound has been proved for the Gaussian divisible ensemble: a Wigner matrix perturbated by a small Gaussian component in [BYY18] to study band matrices. Note also that a weaker form of quantum unique ergodicity has been proved for generalized Wigner matrices in [BY17].

1.2.2. Method of proof

In this subsection, we will give a sketch of the dynamical proof of universality for eigenvalue and eigenvectors introduced by Erdős-Shlein-Yau. The method consists in a short time relaxation by a variance-preserving dynamics on the space of symmetric or Hermitian random matrices. It can be dissected into three main steps [ESY11]: optimal estimates on the spectrum and eigenvectors by controlling the resolvent and its normalized trace, relaxation by the Dyson Brownian motion in order to reach local equilibrium at optimal speed and finally a form of invariance of local statistics either using the density of the Dyson Brownian motion and a Green function comparison theorem or using the continuity of the dynamics for the whole matrix structure. We will now give an idea of the proof and the interested reader shoud go to [EY17], a recent book on the subject.

We will describe those three main steps in the case of generalized Wigner matrices though it has been used for numerous other models. We will take the following definition for the model considered.

Definition 1.2.4 (Generalized Wigner matrices)**.** Let $W=(w_{ij})_{1\leqslant i,j\leqslant N}$ be a $N\times N$ symmetric (or Hermitian) centered random matrix with independent entries up to the symmetry constraint and such that there exists $c,C>0$ such that

$$\frac{c}{N}\leqslant \mathrm{Var}(w_{ij})\leqslant\frac{C}{N}\quad\text{and}\quad\sum_{j=1}^N\mathrm{Var}(w_{ij})=1\text{ for all i.}\tag{1.2.7}$$

Assume also that for every $p\in\mathbb{N}$ there exists μ_p independent of N such that

$$\mathbb{E}\left[|\sqrt{N}w_{ij}|^p\right]\leqslant\mu_p.\tag{1.2.8}$$

Remark 1.2.5. While the first proof of universality used a subexponential decay assumption for the matrix entries, we will use that all moments are finite. Note that this has been improved in [Agg18] where only moments of order $2+\varepsilon$ are needed.

With the normalization of the variance of matrix entries, the convergence (1.0.1) holds and the limiting spectral measure is given by the semicircle distribution. While this gives us the global behavior, we need a stronger estimate on the large N limit of the spectrum.

First step: local semicircle law. The global law (1.0.1) can also be stated as a convergence of Stieltjes transform in the following sense,

$$s(z):=\int\frac{\mathrm{d}\mu(x)}{x-z}\xrightarrow[N\to\infty]{(d)}m_{\mathrm{sc}}(z):=\int\frac{\mathrm{d}\rho_{\mathrm{sc}}(x)}{x-z}\tag{1.2.9}$$

for all $z\in\mathbb{C}_+$. The spectrum of our original matrix can be linked to the Stieltjes transform by the following identity and is the reason why the previous convergence gives (1.0.1),

$$\lim_{\eta\to0}\frac{1}{\pi}\mathrm{Im}\int_a^b m_{\mathrm{sc}}(E+\mathrm{i}\eta)\mathrm{d}E=\int_a^b\rho_{\mathrm{sc}}(E)\mathrm{d}E.\tag{1.2.10}$$

The global law gives us the asymptotic number of eigenvalues on a fixed interval independent of N. We would like a local estimate, in other words make the size of interval depend on N so that we can capture the behavior of only a few eigenvalues. If one writes the imaginary part of the Stieltjes transform as

$$\mathrm{Im}\,s(E+\mathrm{i}\eta)=\frac{1}{N}\sum_{i=1}^N\frac{\eta/\pi}{(\lambda_i-E)^2+\eta^2}=(\mu*\delta_\eta)(E)\tag{1.2.11}$$

where δ_η is an approximation of the identity as η goes to zero, one can see the Stieltjes transform as a smoothed version of the original measure on the scale η. But, with our normalization, the spectrum is supported in an interval of order 1, the typical eigenvalue spacings (in the bulk) is then of order N^{-1}. In order to have a similar convergence as (1.2.9) on smaller scales, we still need to average on a small part of the spectrum as we can not expect each eigenvalue to converge to a fixed point and we must take $\eta \gg N^{-1}$. We will then need to consider the following spectral domain for the bulk of the spectrum, corresponding to the *mesoscopic scales*, for any small κ, $\tau > 0$,

$$\mathcal{D}_\kappa^\tau = \{z = E + \mathrm{i}\eta, E \in [-2+\kappa, 2-\kappa], N^\tau/N \leqslant \eta \leqslant 10\}. \tag{1.2.12}$$

We can now state our (isotropic) local semicircle law in the following theorem

Theorem 1.2.6. *For any small κ, $\tau > 0$, we have*

$$\mathbb{P}\left(|s(z) - m_{\mathrm{sc}}(z)| \geqslant \frac{N^\varepsilon}{N\eta}\right) \leqslant N^{-D} \tag{1.2.13}$$

for any (small) $\varepsilon > 0$ and (large) $D > 0$ uniformly in $z = E + \mathrm{i}\eta \in \mathcal{D}_\kappa^\tau$. We have also, for any non random $\mathbf{x}$, $\mathbf{y} \in \mathbb{C}^N$, such that $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$,

$$\mathbb{P}\left(|\langle \mathbf{v}, G(z)\mathbf{w}\rangle - m_{\mathrm{sc}}(z)\langle \mathbf{v}, \mathbf{w}\rangle| \geqslant \frac{N^\varepsilon}{\sqrt{N\eta}}\right) \leqslant N^{-D} \quad \text{with} \quad G(z) = \sum_{k=1}^N \frac{|u_k\rangle\langle u_k|}{\lambda_k - z} \tag{1.2.14}$$

for any (small) $\varepsilon > 0$ and (large) $D > 0$ uniformly in $z = E + \mathrm{i}\eta \in \mathcal{D}_\kappa^\tau$.

This theorem has been improved over the last ten years or so, the first version of a local semicircle law was proved in a series of work [ESY09a, ESY09b, ESY10] with a weaker control of the error estimate which has been improved optimally in [EYY12b]. The isotropic local law as stated in (1.2.14) was first stated in [KY13a] and proved for generalized Wigner matrices in [BEK+14].

The proof is based on a diagrammatic analysis of the resolvent via the Schur complement formula and a stability analysis of a self-consistent equation followed by the Stieltjes transform. In order to get the optimal error of $(N\eta)^{-1}$, one has to average the resolvent fluctuation of order $(N\eta)^{-1/2}$. Note that recent methods used a cumulant expansion of the resolvent, inspired by [KKP96, LP09], and stability estimates to derive the isotropic local law [LS17, HKR17]. As we can not give a more detailed idea of the proof in this short review, we refer to [BGK16] for lecture notes on the subject. Note also that proofs of local laws are really model dependent and can differ widely from Wigner matrices to sparse graphs or correlated models for instance.

Theorem 1.2.6 gives us, among other things, two corollaries diretly linked to universality. The first is the complete delocalization of eigenvectors as in (1.2.4),

Corollary 1.2.7 ([ESY09b])**.** *Complete delocalization* (1.2.4) *holds for the generalized Wigner matrices.*

The second is rigidity of eigenvalues around their typical position with an optimal error estimate. While we stated the local law only in the bulk of the spectrum we can state a similar overwhelming bound at the edge of the spectrum, and we can state the following general corollary

Corollary 1.2.8 ([EYY12b])**.** *Define the typical positions γ_k implicitly by $\int_{-\infty}^{\gamma_k} \mathrm{d}\rho_{\mathrm{sc}} = \frac{k}{N}$ and let $\hat{k} = \min(k, N+1-k)$, then we have for any small $\varepsilon > 0$ and large $D > 0$,*

$$\mathbb{P}\left(|\lambda_k - \gamma_k| \geqslant \frac{N^\varepsilon}{N^{2/3}\hat{k}^{1/3}}\right) \leqslant N^{-D}. \tag{1.2.15}$$

Second step: Short-time relaxation. Now that we have strong estimates on our original model, we will interpolate to the GOE distribution with the Dyson Brownian motion, a dynamics on the set of symmetric (or Hermitian) random matrices whose equilibrium measure is given by the GOE (or GUE).

Definition 1.2.9 (Symmetric Ornstein-Uhlenbeck Dyson Brownian motion)**.**
Let $(B_{ij})_{i<j}$ and $(B_{ii}/\sqrt{2})$ be two families of independent standard Brownian motions, consider the following dynamics

$$\mathrm{d}H_t = \frac{\mathrm{d}B_t}{\sqrt{N}} - \frac{1}{2}H_t\mathrm{d}t. \tag{1.2.16}$$

Denote the ordered eigenvalues $\boldsymbol{\lambda}(t) = (\lambda_1(t) \leqslant \cdots \leqslant \lambda_N(t))$ of H_t and $\boldsymbol{u}(t) = (u_1(t), \dots u_N(t))$ the associated eigenvectors. Let $(\widetilde{B}_{k\ell})_{1\leqslant k\leqslant \ell\leqslant N}$ be independent standard Brownian motions, then $(\boldsymbol{\lambda}(t), \boldsymbol{u}(t))_t$ has the same distribution as the solution to the following coupled dynamics,

$$\mathrm{d}\tilde{\lambda}_k(t) = \frac{\mathrm{d}\widetilde{B}_{kk}(t)}{\sqrt{N}} + \left(\frac{1}{N}\sum_{\ell\neq k}\frac{1}{\tilde{\lambda}_k(t)-\tilde{\lambda}_\ell(t)} - \frac{1}{2}\tilde{\lambda}_k(t)\right)\mathrm{d}t, \tag{1.2.17}$$

$$\mathrm{d}\tilde{u}_k(t) = \frac{1}{\sqrt{N}}\sum_{\ell\neq k}\frac{\mathrm{d}\widetilde{B}_{k\ell}}{\tilde{\lambda}_k(t)-\tilde{\lambda}_\ell(t)}\tilde{u}_\ell(t) - \frac{1}{2N}\sum_{\ell\neq k}\frac{\mathrm{d}t}{(\tilde{\lambda}_k(t)-\tilde{\lambda}_\ell(t))^2}\tilde{u}_k(t). \tag{1.2.18}$$

Remark 1.2.10. The eigenvalue dynamics (1.2.17) was first given in [Dys62a] where the Dyson Brownian motion was first introduced. The eigenvector flow was computed in different settings such as the study of the Brownian motion on ellipsoids in [NRW86], for Wishart processes in [Bru89] or to study Gaussian ensembles in [AGZ10].

a. Relaxation for eigenvalues: The first use of the dynamics (1.2.17) was for the Hermitian Wigner case where explicit formulas for the eigenvalues density were available [Joh01a, EPR$^+$10, TV11] at any time t. For shorter-time relaxation, relative entropy was used in [ESY11, EYY12a] to show an averaged version of (1.2.1) over the energy E and an Helffer-Sjöstrand representation was used in [EY15] to show gap universality. We will give here a quick idea of the proof of fixed energy universality as in (1.2.1) from [BEYY16, LSY16] using a coupling argument.

This argument consists in taking two different trajectories of the same Dyson Brownian motion: the trajectory $\mathbf{x}$ with initial condition given by our generalized Wigner matrix (or another model of which we have strong estimates as in (1.2.15)) and $\mathbf{y}$ with initial condition given by a matrix taken from the GOE. Now, the observable $\delta_k(t) = e^{t/2}(x_k(t) - y_k(t))$ satisfies a parabolic equation

$$\partial_t \delta_k(t) = \sum_{\ell\neq k}\frac{\delta_k(t)-\delta_\ell(t)}{N(x_k(t)-x_\ell(t))(y_k(t)-y_\ell(t))}. \tag{1.2.19}$$

Since (1.2.15) holds along the dynamics, one can see that, for eigenvalues in the bulk of the spectrum, we have $(x_k(t)-x_\ell(t))(y_k(t)-y_\ell(t)) \simeq (k-\ell)^2$, so that by Hölder regularity of the dynamics for $t \gg N^{-1}$ we have $\delta_{k+1}(t) \simeq \delta_k(t)$ or in other words $x_{k+1}(t) - x_k(t) \simeq y_{k+1}(t) - y_k(t)$. But since the GOE measure is invariant along the dynamics (1.2.16), the gaps $y_{k+1}(t) - y_k(t)$ are given by the Mehta-Gaudin distribution and thus gives us universality for our original trajectory $\mathbf{x}$. In [LY17a, LSY16, LY17b], the short-time relaxation has been proved for a wide class of initial condition, effectively streamlining the whole second step for bulk universality [LY17a], fixed energy universality [LSY16] or edge universality [LY17b].

b. Relaxation for eigenvectors There is no coupling argument for eigenvectors and the study of the dynamics (1.2.18) is too complicated in high dimensions. The short-time relaxation was however used in [BY17] by looking at the joint moments of eigenvectors and reducing the complicated dynamics (1.2.18) to a random walk in random environment given by the eigenvalues. We will give here some details in the symmetric case but similar estimates and dynamics exist for the Hermitian case.

First consider $\boldsymbol{\xi} : [\![1, N]\!] \to \mathbb{N}$ a configuration of particles where ξ_k is the number of particles at the site k, we will denote $\boldsymbol{\xi}^{\boldsymbol{k},\boldsymbol{\ell}}$ the configuration $\boldsymbol{\xi}$ where we moved a single particle from the site k to the site ℓ (note that there is a direction to the particle movement). We will study the dynamics of the following observables, for $\mathbf{q} \in \mathbb{R}^N$ such that $\|\mathbf{q}\|_2 = 1$,

$$f_t(\boldsymbol{\xi}) = \frac{\mathbb{E}\left[\prod_{k=1}^N z_k^{2\xi_k}\,\middle|\,\boldsymbol{\lambda}\right]}{\prod_{k=1}^N \mathbb{E}\left[\mathcal{N}_k^{2\xi_k}\right]} \quad \text{with} \quad z_k = \sqrt{N}\langle \mathbf{q}, u_k(t)\rangle, \tag{1.2.20}$$

and $(\mathcal{N}_k)_{1\leqslant k\leqslant N}$ is an *i.i.d* collection of standard Gaussian random variables. If one wants to derive Gaussianity of projections of eigenvectors as in (1.2.3), one needs to see that for any configuration $\boldsymbol{\xi}$, $f_t(\boldsymbol{\xi})\to 1$ as N grows. This would indeed show convergence of moments. It was seen in [BY17] that f_t followed the parabolic differential equation

$$\partial_t f_t(\boldsymbol{\xi}) = \sum_{k\neq\ell}\frac{2\xi_k(1+2\xi_\ell)(f_t(\boldsymbol{\xi}^{\boldsymbol{k},\boldsymbol{\ell}})-f_t(\boldsymbol{\xi}))}{N(\lambda_k(t)-\lambda_\ell(t))^2} \tag{1.2.21}$$

which can be seen as a generator for a multi-particle random walk in random environment: each site can have a particle jumping to another site with a rate depending on the eigenvalues of the Dyson Brownian motion at time t. The dynamics for one particle was first obtained in the physics literature in [WW95]. The analysis of this equation is simplified by the fact that there exists an explicit reversible measure for the dynamics. With a maximum principle and a Gronwall argument it is then possible to show relaxation of f_t to 1 and thus the convergence (1.2.3) in the sense of moments.

A corollary from the Gaussianity of projections is a form of quantum unique ergodicity though weaker than the optimal (1.2.6), it states that for any $\delta>0$, and for any I a $N-$dependent set of indices, there exists $\varepsilon>0$ such that

$$\mathbb{P}\left(\left|\sum_{\alpha\in I}u_k(\alpha)^2-\frac{|I|}{N}\right|>\delta\right)\leqslant C\frac{N^{-\varepsilon}}{\delta^2}. \tag{1.2.22}$$

Third Step: invariance of local statistics The previous step gives us universality results for the addition of our original matrix with a small Gaussian component. One now needs to remove this small perturbation while conserving local statistics of eigenvalues or eigenvectors. The first method used to do so in [EPR+10] was the reverse heat flow by assuming smoothness of the entries on the original matrix. In [TV11], a Green function comparison theorem was used to show invariance of local eigenvalue statistics when the first four moments of the matrix entries are Gaussian moments, it was combined with the present dynamical method in [ERS+10] to show universality for a broader class of random matrices. They used the *density* of the Dyson Brownian motion in the following sense, one can find an initial condition $\widetilde{W}$, taken from the Wigner ensembles, for the Dyson Brownian motion $\widetilde{H}_t$ such that the first four moments approximate well enough our original matrix W. The result being proved for $\widetilde{H}_t$ by our previous relaxation argument, universality for W holds.

In [BY17], the authors proved invariance of local statistics along the dynamics H_t for a time $N^{-1}\ll t\ll N^{-1/2}$ in the case of generalized Wigner matrices. It is a consequence of an Itō lemma and can be stated as the following, for any smooth function F on symmetric matrices,

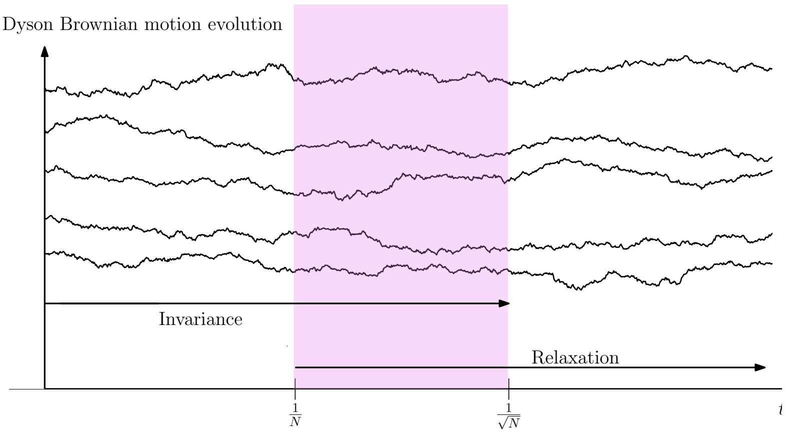
$$\mathbb{E}\left[F(H_t)\right]-\mathbb{E}\left[F(H_0=W)\right]=\mathcal{O}\left(tN^{1/2}\right)\mathbb{E}\left[\sup_{\substack{i\leqslant j,0\leqslant s\leqslant t\\ \boldsymbol{\theta}^{ij}}}\left(N^{3/2}|H_{ij}(s)|^3+\sqrt{N}|H_{ij}(s)|\right)|\partial_{ij}^3F(\boldsymbol{\theta}^{ij}H_s)|\right] \tag{1.2.23}$$

where $\boldsymbol{\theta}^{ij}H$ is a matrix given by $(\boldsymbol{\theta}^{ij}H)_{k\ell}=H_{k\ell}$ for $\{k,\ell\}\neq\{i,j\}$ and $(\theta^{ij}H)_{k\ell}=\theta^{ij}_{k\ell}H_{k\ell}$ for some $0\leqslant\theta^{ij}_{k\ell}\leqslant 1$ otherwise. The typical functionals F taken to show universality are products of resolvent entries. Since they are stable in the sense that $\partial^3_{ij}F=\mathcal{O}(N^\varepsilon)$ with overwhelming probability, we can show that local statistics essentially stays the same up to a time $t=N^{1/2-\varepsilon}$ for any ε.

The final picture is given by the following figure: invariance of local statistics holds until time $1/\sqrt{N}$ but the optimal relaxation time is $1/N$ so that we have a window where both holds.

1.3. Other mean-field models

This method has been applied to numerous mean-field models as only the first step is purely model dependent. We will try to give an extensive, though not exhaustive, list of models where this method has been applied to show either universality of eigenvalues or eigenvectors.



1.3.1. General mean and covariance

While we took centered entries with a stochastic variance matrix in Definition 1.2.4, it is possible to add an expectation matrix and change the covariance between entries. We first give the definition of the model in case of independent entries.

Definition 1.3.1. Let $W = (w_{ij})_{1\leqslant i,j\leqslant N}$ be a symmetric or Hermitian matrix with independent centered entries up to the symmetry such that

(i) For all $p \in \mathbb{N}$, there exists μ_p such that for all i, j, $\mathbb{E}\left[\sqrt{N}|w_{ij}|^p\right] \leqslant \mu_p$.

(ii) There exists c and C positive constants such that, for any positive sem-definite matrix T,

$$\frac{c}{N}\operatorname{Tr} T \leqslant \mathbb{E}[HTH] \leqslant \frac{C}{N}\operatorname{Tr} T.$$

(iii) There exists λ a positive constant such that, for any determinisic symmetric (or Hermitian) matrix B

$$\mathbb{E}\left[|\operatorname{Tr} BW|^2\right] \geqslant \frac{\lambda}{N}\operatorname{Tr} B^2.$$

Let A be a deterministic matrix such that there exists a $C > 0$ such that $\|A\| \leqslant C$ for all N. We will then consider the model $H = W + A$.

For such matrices, the global statistics is not given by the semicircle law anymore but the possible shape of the asymptotic spectrum has been characterized in [AEK18b]: the spectrum is supported on finitely many intervals, called bands, with an analytic density of states inside these, the behavior at the edge of such an interval is either a square-root growth which we call regular, or a cubic root cusp when the space between bands vanishes.

For the behavior in the bulk of each band and at the regular edge we have the following universality result. Define ρ to be the asymptotic density of states.

Theorem 1.3.2 (Bulk and regular edge universality)**.** *For all $\delta > 0$ and $n \in \mathbb{N}$ and any $O : \mathbb{R}^n \to \mathbb{R}$ smooth compactly supported function, there exists C and c positive constants such that for any $E \in \mathbb{R}$ such that $\rho(E) \geqslant \delta$*

$$\left|\mathbb{E}O\left(N\rho(\lambda_{i(E)})(\lambda_{i(E)} - \lambda_{i(E)+j})\right)_{j=1^n} - \mathbb{E}_{\text{Gauss}}O\left(N\rho_{\text{sc}}(0)(\lambda_{\lceil N/2\rceil} - \lambda_{\lceil N/2\rceil+j})\right)_{j=1}^n\right| \leqslant CN^{-c}$$

with $i(E)$ being the label of the closest eigenvalue to E.

Assume that $E \in \mathbb{R}$ is at the regular right (or left) edge of ρ such that there exists c a positive constant such that $\rho([E, E + c]) = 0$. There exists a γ such that if i_o is the label of the largest eigenvalue close to the band *edge with high probability, for any suitable function* $O:\mathbb{R}^{k+1}\to\mathbb{R}$ *we have*

$$\left|\mathbb{E}O\left(\gamma N^{2/3}(\lambda_{i_0}-E),\dots,\gamma N^{2/3}(\lambda_{i_0-k}-E)\right)-\mathbb{E}_{\mathrm{Gauss}}O\left(N^{2/3}(\lambda_N-2),\dots,N^{2/3}(\lambda_{N-k}-2)\right)\right| \leqslant N^{-\delta}$$

for some $\delta>0$.

Note that contrary to β-ensembles, the Tracy-Widom distribution holds at fixed label in the internal regular edges thanks to the remarkable band rigidity phenomenon exhibited in [AEKS18].

These results were proved using the three-step strategy in [AEK15, AEK17] for bulk universality and [AEKS18] for edge universality. Actually, these results hold under more general assumptions, with some summable decay of correlations between the matrix entries. The first case considered was Gaussian random matrix with correlations in [AEK16] where bulk universality and complete delocalization of eigenvectors were proved. It was generalized to any distribution with a fast correlation decay independently in [Che17, AEK18a]. The correlation decay has been improved and bulk universality was proved in [EKS17] and edge universality with a summable correlation decay was given in [AEKS18].

As we saw earlier, another singularity given by a cubic-root cusp (or almost-cusp) is possible in the support of the density of states. There, the local eigenvalue statistics is given by a Pearcey process. Universality for this process was proved for when A is diagonal and W has independent entries in [EKS18, CEKS18].

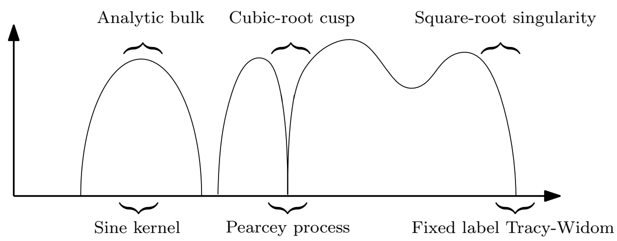


Figure 1.1: Possible shape of the density of states with the corresponding behavior of eigenvalues

1.3.2. Adjacency matrices of random graphs

We will focus in this subsection on two models of extensively studied random graphs: Erdős-Rényi graphs and d-regular graphs.

Erdős-Rényi graphs: Consider N vertices, each possible edge is chosen independently with probability p. If we consider the adjacency matrix of this graph, we can see that each row or column has on average pN nonzero entries. Thus, if we consider p depending on N such that $p\ll 1$, we obtain a sparse matrix. Note that the behavior of such a random matrix is different than a matrix from the Wigner ensemble with Bernoulli entry distribution since the moments of the entries of the Erdős-Rényi adjacency matrix have a way slowlier decay.

Results on universality began in [EKYY13, EKYY12] where both the bulk and edge universality was proved in the case of $Np\gg N^{2/3}$ and complete delocalization of eigenvectors for $Np\geqslant(\log N)^C$ for some C. Bulk universality was improved in [HLY15] for $Np\gg 1$ while edge universality was improved in [LS17] for $Np\gg N^{1/3}$. A phase transition has been discovered in [HLY17] where extreme eigenvalues have Gaussian fluctuations for $N^{2/9}\ll Np\ll N^{1/3}$ and have a combination of Gaussian and Tracy-Widom fluctuations for $Np=CN^{1/3}$. Asymptotic Gaussianity and quantum unique ergodicity has been proved in [BY17] for $Np\gg 1$.

d-regular graphs: Consider the adjacency matrix of a uniform random d-regular graph on N vertices, in other words it is a matrix chosen uniformly among matrices with $\{0,1\}$ entries such that each row sums to d and all diagonal entries are 0. The hard constraint on each row or column create a lot of dependencies between matrix entries which make some of the steps introduced in Section 1.2.2 trickier.

Bulk universality for regular graphs was first proved in [BHKY17] for $1\ll d\ll N^{2/3}$ where the third step is replaced by a switching dynamics approximation on the graphs. For the first step, an optimal local law was proved in [BKY17] for $d \geqslant (\log N)^4$ which implies eigenvector delocalization. Universality at the edge or at fixed d is currently an open question, however a local Kesten-McKay law for fixed d and a strong form of quantum unique ergodicity for eigenvectors have been proved in [BHY19] as well as a complete delocalization result.

Theorem 1.3.3 (Eigenvector delocalization for random d-regular graphs)**.** *Fix $\alpha > 4$ and $\omega \geqslant 8$, take d such that $\sqrt{d-1} \geqslant (\omega+1)2^{2\omega+45}$. Consider the rescaled adjacency matrix of a uniform d-regular graph with N vertices $H = A/\sqrt{d-1}$, then for u_i a L^2-normalized eigenvector of H whose corresponding eigenvalue λ_i follows $|\lambda_i \pm 2| \geqslant (\log N)^{1-\alpha/2}$, we have*

$$\mathbb{P}\left(\|u_i\|_\infty \geqslant \frac{\sqrt{2}(\log N)^{24\alpha+1/2}}{\sqrt{N}}\right) \leqslant N^{-\omega+8}.$$

1.3.3. Lévy matrices

In the previous subsections, the entries of each random matrix model had a second moment. Notably, in the case of Wigner ensembles, the moment assumption was relaxed to a $2+\varepsilon$ moment in [Agg18]. We will define here a model of random matrix with heavy-tailed entries given by α-stable laws.

Definition 1.3.4 (Lévy matrices)**.** Consider H a symmetric random matrix such that $(h_{ij})_{1\leqslant i\leqslant j\leqslant N}$ are independent and identically distributed random variables, distributed according to an α-stable law in the sense that

$$\mathbb{E}\left[\exp(\mathrm{i}th_{11})\right] = \exp(-N\sigma^\alpha|t|^\alpha) \quad \text{with} \quad \sigma = \left(\frac{\pi}{2\sin(\frac{\pi\alpha}{2})\Gamma(\alpha)}\right)^{1/\alpha}.$$

The global statistics of the spectrum for this model was first computed in [CB94, BAG08] where it has been shown that the empirical spectral distribution converges to a deterministic heavy-tailed distribution, in great contrast with the semicircle distribution. The first result concerning universality was on (de)localization of eigenvectors in [BG13, BG17]. They showed that for $1 < \alpha < 2$ almost all eigenvectors are delocalized in the following sense, for any $\delta > 0$,

$$\sup_{1\leqslant i\leqslant N} |u_k(i)| < N^{\delta-\rho} \quad \text{with high probability and with} \quad \rho = \frac{\alpha-1}{\max(2\alpha, 8-3\alpha)}.$$

For $0 < \alpha < 1$, they proved that there exists a E_α such that eigenvectors corresponding to eigenvalues on $[-E_\alpha, E_\alpha]$ are delocalized, for any $\delta > 0$

$$\sup_{1\leqslant i\leqslant N} |u_k(i)| < N^{\delta-\alpha/(4+2\alpha)}.$$

They also showed that eigenvectors are localized in some sense at large energy for $0 < \alpha < \frac{2}{3}$. These results of delocalization were improved in [ALY18] where bulk universality and complete delocalization was proved for $0 < \alpha < 2$ with similar distinctions. For $1 < \alpha < 2$, the results hold on any fixed compact away from 0 while for $0 < \alpha < 1$, there exists a E_α such that the results hold for all eigenvalues and corresponding eigenvector on a compact set $[-E_\alpha, E_\alpha]$.

Theorem 1.3.5 (Delocalization and universality for Lévy matrices)**.** *For $\alpha \in (1,2)$, fix a compact interval $K \subset \mathbb{R}\setminus\{0\}$, then*

(i) Eigenvectors of H are completely delocalized as in (1.2.4)*.*

(i) For $E \in K$, fixed energy universality for correlation function as in (1.2.1) *holds.*

Besides, there exists a countable set $\mathcal{A} \subset (0,2)$ with no accumulation points in $(0,2)$ such that for any $\alpha \in (0,2)$ outside of $\mathcal{A}$, there exists c (depending on α) such that

(i) Eigenvectors u_i corresponding to an eigenvalue in $[-c,c]$ are completely delocalized as in (1.2.4)*.*

(i) For $E \in [-c,c]$, fixed energy universality as in (1.2.1) *holds.*

1.3.4. Addition of random matrices

Another model one can consider is the following, consider two deterministic diagonal matrix D_1 and D_2 and draw independently two matrices U_1 and U_2 from the Haar-measure on the unitary group, define then

$$H = U_1^* D_1 U_1 + U_2^* D_2 U_2.$$

This model first appeared in free probability theory and the limiting empirical spectral distribution was first given by Voiculescu in [Voi91], it consists of the free convolution between the limiting empirical spectral distribution of D_1 and D_2. The first local law proved for the model was done in [Kar12] where the local law held down to the scale $(\log N)^{-1/2}$ which was then improved by the same author in [Kar15] to reach the scale $N^{-1/7}$. The optimal local law, down to the scale $N^{-1+\varepsilon}$, was finally obtained in a series of work [BES16, BES17a, BES17b] using a refined analysis of the Green function and fluctation averaging mechanism in order to obtain the optimal control parameter.

Using this local law as an input, a relaxation method was used in [CL17] to prove universality of local eigenvalue statistics. They used another diffusion to keep the structure of the matrix along the dynamics and showed the result by comparing it to the usual Dyson Brownian motion.

1.3.5. Deformed Wigner matrices

Another model of mean-field type consists in adding a random or deterministic potential to a Wigner matrix. This potential is given by a diagonal matrix. This model was first introduced by Rosenzweig and Porter in [RP60]. Let D be a deterministic (or random and independent from W) diagonal matrix and consider a parameter t which can depend on N, we then consider the matrix

$$W_t = D + \sqrt{t} W$$

The first result on this model was for $t \gtrsim 1$ and consisted in a local law and complete delocalization for eigenvectors in [LS13]. Bulk universality was proved for this phase in [LSSY16] and edge universality in [LS15].

Bulk universality was then proved in the case of $t \gg N^{-1}$ independently in [LY17a, ES17]. As for eigenvectors in this phase, they are not completely delocalized. However, it has been shown in [Ben17] that asymptotic Gaussianity of eigenvectors entries holds after renormalization by an explicit variance depending on t, D, the eigenvector and the entry considered. A strong form of quantum unique ergodicity is also proved for W a matrix from the Gaussian divisible ensemble while a weaker one holds for any Wigner matrix. This is the content of the next chapter.

Chapter 2

Eigenvector distribution and quantum unique ergodicity for deformed Wigner matrices

This chapter is based on the article [Ben17]

2.1. Introduction

In the study of large interacting quantum systems, Wigner conjectured that empirical results are well approximated by statistics of eigenvalues of large random matrices. The interested reader can go to [Meh04] for an overview and the mathematical formalization of the conjecture. This vision has not been shown for correlated quantum systems but is regarded to hold for numerous models. For instance, the Bohigas−Giannoni−Schmit conjecture in quantum chaos [BGS84] connects eigenvalues distributions in the semiclassical limit to the Gaudin distribution for GOE statistics. These statistics also conjecturally appear for random Schrödinger operators [And58] in the delocalized phase. Most of these hypotheses are unfortunately far from being proved with mathematical rigor. It is, however, possible to study systems given by large random matrices. One of the most important models of this type is the Wigner ensemble, random Hermitian or symmetric matrices whose elements are, up to the symmetry, independent and identically distributed zero-mean unit variance random variables. For this ensemble, local statistics of the spectrum only depend on the symmetry class and not on the laws of the elements (see [ESY11, TV11, EYY12a, EY15, BEYY16]). The Wigner−Dyson−Mehta conjecture was solved for numerous, more general mean-field models such as the generalized Wigner matrices, random matrices for which the laws of the matrix elements can have distinct variances (see [EY17] and references therein).

The statistics of eigenvectors were not used in Wigner's original study but localization, or delocalization, has been broadly studied in random matrix theory. For Wigner matrices, it has been shown in [ESY09b] that eigenvectors are completely delocalized in the following sense: denoting $u_1, \dots, u_N$ the L^2-normalized eigenvectors of an $N \times N$ Wigner matrix, we have with very high probability,

$$\sup_{\alpha} |u_i(\alpha)| \leqslant \frac{C(\log N)^{9/2}}{\sqrt{N}}.$$

Thus, eigenvectors cannot concentrate onto a set of size smaller than $N(\log N)^{-9/2}$. See also [VW15] for optimal bounds in some cases of Wigner matrices or [RV15] for an improved bound which also holds for non-Hermitian matrices and [EYY12b] for similar estimates for generalized Wigner matrices.

In the GOE and GUE cases, the distribution of the matrix is orthogonally invariant and eigenvectors are distributed according to the Haar measure on the orthogonal group. In particular, the entries of bulk eigenvectors are asymptotically normal:

$$\sqrt{N}u_i(\alpha) \xrightarrow[N\to\infty]{} \mathcal{N},$$

where $\mathcal{N}$ is a standard Gaussian random variable. Asymptotic normality was first proved for Wigner matrices in [KY13b, TV12b] under a matching condition on the first four moments of the entries using Green's function comparison theorems introduced in [TV11]. These conditions were later removed in [BY17] where asymptotic normality holds for generalized Wigner matrices. Beyond mean-field models, conjectures of interest, for example for band matrices, are still yet to be proved. A sharp transition is conjectured to occur when the band width W cross the critical value $\sqrt{N}$. For $W \ll \sqrt{N}$, eigenvectors are expected to be localized on $\mathcal{O}(W^2)$ sites and eigenvalue statistics are Poisson, while for $W \gg \sqrt{N}$ eigenvectors would be completely delocalized and one would get Wigner-Dyson-Mehta statistics for the eigenvalues. For the most recent works on this subject see [Sch09, PSSS19] for localization results, [BYY18] for delocalization results, [Sod10] for another transition occurring at the edge of the spectrum and [Bou18] for a recent review on the subject.

In this paper, we consider a generalized Rosensweig-Porter model, of mean-field type, which also interpolates between delocalized and localized (or partially delocalized) phases, but always with GOE/GUE statistics. It is defined as a perturbation of a potential, consisting of a deterministic diagonal matrix, by a mean field noise, given by a Wigner random matrix, scaled by a parameter t. This model follows two distinct phase transitions. When $t \ll 1/N$, eigenvalue statistics coincide with $t = 0$ and eigenvectors are localized on $\mathcal{O}(1)$ sites [vSW18a], while when $t \gtrsim 1$, local statistics fall in the Wigner-Dyson-Mehta universality class [LSSY16] with fully delocalized eigenvectors [LS13]. For $1/N \ll t \ll 1$, it has been shown in [LY17a, ES17] that eigenvalue statistics are in the Wigner-Dyson-Mehta universality class and in [vSW18b] that eigenvectors are not completely delocalized when the noise is Gaussian. In this intermediate phase, also called the *bad metal regime* (see [FVB16] or [TO16] for instance), eigenstates are partially delocalized over Nt sites, a diverging number as N grows but a vanishing fraction of the eigenvector coordinates. The existence of this regime for more intricate models is only conjectured or even debated in the physics literature though progress has been made recently, for instance for the Anderson model on the Bethe lattice and regular graph in [KAI17].

Our results give the asymptotic distribution of the eigenvectors for this model, giving a rather complete understanding of this regime for the Rosensweig-Porter model. We show that bulk eigenvectors are asymptotically Gaussian with a specific, explicit variance depending on the initial potential, the parameter t and the position in the spectrum. For a well-spread initial condition, this variance is heavy-tailed and follows a Cauchy distribution. This shape was first unearthed in a non-rigorous way in [ABB14, AB14] for W a matrix from the Gaussian ensembles, where the Gaussian distribution of eigenvectors (Corollary 2.1.4) was conjectured. Note that eigenvector dynamics was also considered in [AB12] and used for denoising matrices in [BABP16, BBP18]. In the case of Gaussian entries, the eigenvector distribution has been exhibited in the physics literature in [FVB16] using the resolvent flow and in [TO16] using supersymmetry techniques.

Another strong form of delocalization of eigenfunctions is quantum ergodicity. It has been proved for the Laplace-Beltrami operator on negative curved compact Riemannian manifold by Shnirel'man [Šni74], Colin de Verdière [CdV85] and Zelditch [Zel87] but also for regular graphs by Ananthamaran-Le Masson [ALM15]. In [RS94], Rudnick-Sarnack conjectured a stronger form of delocalization for eigenfunctions of the Laplacian called the quantum unique ergodicity. More precisely, denote $(\phi_k)_{k\geqslant 1}$ the eigenfunctions of the Laplace operator on any negatively curved compact Riemannian manifold $\mathcal{M}$, they then supposedly become equidistributed with respect to the volume measure μ in the following sense: for any open set $A \subset \mathcal{M}$

$$\int_A |\phi_k|^2 \mathrm{d}\mu \xrightarrow[k\to\infty]{} \int_A \mathrm{d}\mu.$$

This conjecture has not been proved for all negatively curved compact Riemannian manifold but has been rigorously shown for arithmetic surfaces (see [Hol10, HS10, Lin06]).

A probabilistic form of quantum unique ergodicity exists for eigenvectors of large random matrices. It first appeared in [BY17] for generalized Wigner matrices, large symmetric or Hermitian matrices whose entries are independent up to the symmetry and zero mean random variables but with varying variances. It is stated as a high-probability bound showing that eigenvectors are asymptotically flat in the following way: let $(u_k)_{1\leqslant k\leqslant N}$ be the eigenvectors of a $N\times N$ generalized Wigner matrix, then for any $k \in [\![1, N]\!]$, for any deterministic N-dependent set $I \in [\![1, N]\!]$ such that $|I| \to +\infty$ and any $\delta > 0$,

$$\mathbb{P}\left(\frac{N}{|I|}\left|\sum_{\alpha\in I}\left(u_k(\alpha)^2 - \frac{1}{N}\right)\right| > \delta\right) \leqslant \frac{N^{-\varepsilon}}{\delta^2}$$

for some $\varepsilon > 0$ using the Bienaymé–Chebyshev inequality. Similar high-probability bounds were proved for different models of random matrices such as d-regular random graphs in [BHY19], or band matrices in [BEYY17, BYY18]. In these last papers on band matrices, it was seen that quantum unique ergodicity is a useful property to study non mean-field models. In [BYY18], a stronger form of probabilistic quantum unique ergodicity has been found, showing that the eigenvectors mass is asymptotically flat with overwhelming probability (the probability decreases faster than any polynomial). Our result adapts the method introduced in [BYY18] to show a strong deformed quantum unique ergodicity for eigenvectors of a class of deformed Wigner matrices. Indeed, the probability mass is not flat but concentrates onto an explicit and deterministic profile with a quantitative error.

The key ingredient for this analysis is the Bourgade-Yau eigenvector moment flow [BY17], a multi-particle random walk in a random environment given by the trajectories of the eigenvalues. This method was used for generalized Wigner matrices [BY17] and sparse random graphs [BHY17], and both settings correspond to equilibrium or close to equilibrium situations. Our main contribution consists in treating the non-equilibrium case, which implies additional difficulties made explicit in the next section.

2.1.1. Main Results

Consider a deterministic diagonal matrix $D = \mathrm{diag}(D_1, \dots, D_N)$. The eigenvalues (or diagonal entries) need to be regular enough on a window of size r in the following way first defined in [LY17a].

Definition 2.1.1. Let $\eta_\star$ and r be two N-dependent parameters satisfying

$$N^{-1} \leqslant \eta_\star \leqslant N^{-\varepsilon'}, \quad N^{\varepsilon'}\eta_\star \leqslant r \leqslant N^{-\varepsilon'}$$

for some $\varepsilon' > 0$. A deterministic diagonal matrix D is said to be $(\eta_\star, r)$-regular at E_0 if there exists $c_D > 0$ and $C_D > 0$ independent of N such that for any $E \in [E_0 - r, E_0 + r]$ and $\eta_\star \leqslant \eta \leqslant 10$, we have

$$c_D \leqslant \operatorname{Im} m_D(E + \mathrm{i}\eta) \leqslant C_D,$$

where m_D is the Stieltjes transform of D:

$$m_D(z) = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{D_k - z}.$$

We want to study the perturbation of such a diagonal matrix, notably the eigenvectors, by a mesoscopic Wigner random matrix. We will now suppose that D is $(\eta_\star, r)$-regular at E_0, a fixed energy point. Letting $0 < \kappa < 1$, we will denote in the rest of the paper the spectral window as

$$\mathcal{I}_r^\kappa = [E_0 - (1-\kappa)r, E_0 + (1-\kappa)r].$$

We remove a certain window of energy to avoid any possible complications at the edge. We will also use the following domains. One of the domain will be used for the size of our deformation while the other will be the spectral domain in which we will perform our analysis. We will need the perturbation to be mesoscopic but smaller than the energy window size r, define then for any small positive ω,

$$\mathcal{T}_\omega = \left[N^\omega \eta_\star, N^{-\omega} r\right].$$

For the spectral domain, take first $t \in \mathcal{T}_\omega$ and note that we will consider only $\mathrm{Im}(z) := \eta$ smaller than t but most results such as local laws holds up to macroscopic η. Let $\vartheta > 0$ be an arbitrarily small constant and

$$\mathcal{D}_r^{\vartheta,\kappa} = \left\{ z = E + \mathrm{i}\eta : \, E \in \mathcal{I}_r^\kappa, \, N^\vartheta/N \leqslant \eta \leqslant N^{-\vartheta} t \right\}.$$

Hereafter is our assumptions on our Wigner matrix.

Definition 2.1.2. A Wigner matrix W is a $N \times N$ Hermitian/symmetric matrix satisfying the following conditions

(i) The entries $(W_{i,j})_{1 \leqslant i \leqslant j \leqslant N}$ are independent.

(ii) For all i, j, $\mathbb{E}[W_{i,j}] = 0$ and $\mathbb{E}[|W_{ij}|^2] = N^{-1}$.

(iii) For every $p \in \mathbb{N}$, there exists a constant C_p such that $\left\| \sqrt{N} W_{ij} \right\|_p \leqslant C_p$.

Let W be a Wigner matrix and define the following t-dependent matrix for $t \in \mathcal{T}_\omega$

$$W_t = D + \sqrt{t} W. \tag{2.1.1}$$

The eigenvectors of D are exactly the vectors of the canonical basis since the matrix is diagonal. However, if t were of order one instead of being in $\mathcal{T}_\omega$, the local statistics of W_t would become universal and would be given by local statistics from the Gaussian ensemble. In particular, the eigenvectors would be completely delocalized [LSSY16]. Our model consists in looking at the diffusion of the eigenvectors on the canonical basis after a mesoscopic perturbation. Our main result is that the coordinates of bulk eigenvectors are time and position dependent Gaussian random variables. Before stating our result, we first define the asymptotic distribution of the eigenvalues of the matrix W_t which is the free convolution of the semicircle law (coming from W) and the empirical distribution of D. We will define this distribution through its Stieltjes transform $m_t(z)$ as the solution to the following self-consistent equation

$$m_t(z) = \frac{1}{N} \sum_{\alpha=1}^N \frac{1}{D_\alpha - z - t m_t(z)}. \tag{2.1.2}$$

It is known that this equation has a unique solution with positive imaginary part and is the Stieltjes transform of a measure with density denoted by ρ_t (see [Bia97] for more details). Define the quantiles $(\gamma_{i,t})_{0 \leqslant i \leqslant N}$ of this measure by

$$\int_{-\infty}^{\gamma_{i,t}} \rho_t(x) \mathrm{d}x = \frac{i}{N}.$$

We can also now define the set of indices in the spectral window corresponding to the indices such that the corresponding classical locations lies in the energy window $\mathcal{I}_r^\kappa$,

$$\mathcal{A}_r^\kappa = \{i \in [\![1, N]\!], \gamma_{i,t} \in \mathcal{I}_r^\kappa\}.$$

We can now state our main results, denoting $u_1(t), \dots, u_N(t)$ the L^2-normalized eigenvectors of W_t (we will often omit the t-dependence for $\boldsymbol{u}$). We will define the following quantity, for $N^{-1} \ll \eta \ll t$ and $\eta_\star \ll t \ll r$:

$$\sigma_t^2(\mathbf{q}, k, \eta) = \sum_{\alpha=1}^{N} \frac{q_\alpha^2 t}{(D_\alpha - \gamma_{k_i,t} - t\mathrm{Re}\, m_t(\gamma_{k_i,t} + \mathrm{i}\eta))^2 + (t\mathrm{Im}\, m_t(\gamma_{k_i,t} + \mathrm{i}\eta))^2}. \tag{2.1.3}$$

It is the asymptotic deterministic variance of our eigenvector projections. By taking $\mathbf{q}$ to be a canonical basis vector $\mathbf{e}_\alpha$, we see that $u_k(\alpha)$ has a variance of the form

$$\frac{1}{N} \frac{t}{(D_\alpha - \gamma_{k_i,t} - t\mathrm{Re}\, m_t(\gamma_{k_i,t}))^2 + (t\mathrm{Im}\, m_t(\gamma_{k_i,t}))^2}.$$

For regularly spaced D_α's, this is heavy-tailed with Cauchy shape $\frac{t}{N(x^2+t^2)}$. It localizes the entries onto a subset of indices of size $Nt \ll N$: a fraction of the eigenvector coordinates vanishing as N grows. Such a partial localization appears in [vSW18b] for W GOE-distributed.

Theorem 2.1.3. *(Gaussianity of bulk eigenvectors) Fix $\kappa \in (0,1)$, $\omega \in (0, \varepsilon'/10)$ where ε' is as in Definition 2.1.1 and $m \in \mathbb{N}$. Let $t \in \mathcal{T}_\omega$ and $I \subset \mathcal{A}_r^\kappa$ be a deterministic (N-dependent) set of m elements. Let W as in Definition 2.1.2 and W_t as in (2.1.1). Write $I = \{k_1, \dots, k_m\}$, take a deterministic $\mathbf{q} \in \mathbb{R}^N$ such that $\|\mathbf{q}\|_2 = 1$, and define for $i \in [\![1, m]\!]$,*

$$\sigma_t^2(\mathbf{q}, k_i) := \lim_{\eta \downarrow 0} \sigma_t^2(\mathbf{q}, k_i, \eta) \tag{2.1.4}$$

Then we have

$$\left(\sqrt{\frac{N}{\sigma_t^2(\mathbf{q}, k_i)}} |\langle \mathbf{q}, u_{k_i} \rangle|\right)_{i=1}^{m} \xrightarrow[N \to \infty]{} (|\mathcal{N}_i|)_{i=1}^m \quad \textit{in the symmetric case,} \tag{2.1.5}$$

$$\left(\sqrt{\frac{2N}{\sigma_t^2(\mathbf{q}, k_i)}} |\langle \mathbf{q}, u_{k_i} \rangle|\right)_{i=1}^{m} \xrightarrow[N \to \infty]{} (|\mathcal{N}_i^{(1)} + i\mathcal{N}_i^{(2)}|)_{i=1}^m \quad \textit{in the Hermitian case} \tag{2.1.6}$$

in the sense of convergence of moments, where all $\mathcal{N}_i$, $\mathcal{N}_i^{(1)}$ and $\mathcal{N}_i^{(2)}$ are independent Gaussian random variables with variance 1. The convergence is uniform in over the choice of sets $I \subset [\![1, N]\!]$ of size m.

One can deduce joint weak convergence of eigenvector entries from the previous convergence of moments because $\mathbf{q}$ is arbitrary in $\mathbb{S}^{N-1}$ (see [BY17, Section 5.3]). However, since the eigenvectors are defined up to a phase, we first need to define the following equivalence relation: $u \sim v$ if and only if $u = \pm v$ in the symmetric case and $u = e^{\mathrm{i}\omega} v$ for some $\omega \in \mathbb{R}$ in the Hermitian case.

Corollary 2.1.4. *Let $\kappa \in (0,1)$ and $m \in \mathbb{N}$, let W as in Definition 2.1.2 and W_t as in Definition 2.1.1. Then for any deterministic $k \in \mathcal{A}_r^\kappa$ and $J \subset [\![1, N]\!]$ such that $|J| = m$ we have*

$$\left(\sqrt{\frac{N}{\sigma_t^2(\mathbf{e}_\alpha, k)}} u_k(\alpha)\right)_{\alpha \in J} \xrightarrow[N \to \infty]{} (\mathcal{N}_i)_{i=1}^m \quad \textit{in the symmetric case,} \tag{2.1.7}$$

$$\left(\sqrt{\frac{2N}{\sigma_t^2(\mathbf{e}_\alpha, k)^2}} u_k(\alpha)\right)_{\alpha \in J} \xrightarrow[N \to \infty]{} \left(\mathcal{N}_i^{(1)} + \mathrm{i}\mathcal{N}_i^{(2)}\right)_{i=1}^m \quad \textit{in the Hermitian case} \tag{2.1.8}$$

in the sense of convergence of moments modulo $\sim$, where all $\mathcal{N}_i$, $\mathcal{N}_i^{(1)}$ and $\mathcal{N}_i^{(2)}$ are independent Gaussian random variables with variance 1. In more precise terms, for any polynomial P in m variables there exists δ depending on P such that, for N large enough,

$$\sup_{\substack{J\subset[\![1,N]\!],\,|J|=m\\ k\in\mathcal{A}_r^\kappa}}\left|\mathbb{E}\left[P\left(\left(\epsilon\sqrt{\frac{N}{\sigma_t^2(\mathbf{e}_\alpha,k)}}u_k(\alpha)\right)_{\alpha\in J}\right)\right]-\mathbb{E}\left[P\left((\mathcal{N}_j)_{j=1}^m\right)\right]\right|\leqslant N^{-\delta}$$

where ϵ is taken uniformly at random in the set $\{-1,1\}$. The convergence in the Hermitian case is similar by taking ϵ uniform on the circle.

This result states that the entries of bulk eigenvectors are asymptotically independent Gaussian random variables with variance ${\sigma_t}^2$ which answers a conjecture from [ABB14, Section 3.2], stated in the more restrictive case where W is GOE. The asymptotic normality of the eigenvectors gives the following weak form of quantum unique ergodicity.

Corollary 2.1.5. *(Weak Quantum Unique Ergodicity) Let W as in 2.1.2 and W_t as in Definition (2.1.1). There exists $\vartheta>0$ such that for any $c>0$, there exists $C>0$ such that the following holds : for any $I\subset[\![1,N]\!]$ and $k\in\mathcal{A}_r^\kappa$, we have*

$$\mathbb{P}\left(\frac{Nt}{|I|}\left|\sum_{\alpha\in I}|u_k(\alpha)|^2-\frac{1}{N}\sum_{\alpha\in I}\sigma_t^2(\mathbf{e}_\alpha,k)\right|>c\right)\leqslant C(N^{-\vartheta}+|I|^{-1}). \tag{2.1.9}$$

This high probability bound is not the strongest form of quantum unique ergodicity one can obtain for random matrices. Indeed, if we consider the Gaussian ensembles for which the eigenbasis is Haar-distributed on the orthogonal group and each eigenvectors is uniformly distributed on the sphere, one can get that for any ε and D positive constants, for any $1\leqslant k\leqslant N$,

$$\mathbb{P}\left(\left|\sum_{\alpha\in I}|u_k(\alpha)|^2-\frac{|I|}{N}\right|\geqslant N^{\varepsilon}\frac{\sqrt{I}}{N}\right)\leqslant N^{-D}\quad\text{for any } N \text{ sufficiently large}$$

In this paper, we will obtain a similar overwhelming probability bound on the probability mass of a single eigenvector with an explicit error for a more restrictive model of matrices: deformed random matrix with smooth entries given by the following definition or from a Gaussian divisible ensemble.

Definition 2.1.6. A smooth Wigner matrix W is a $N\times N$ Hermitian/symmetric matrix with the following conditions

(i) The matrix entries $(W_{ij})_{1\leqslant i\leqslant j\leqslant N}$ are independent and identically distributed random variables following the distribution $N^{-1/2}\nu$ where ν has mean zero and variance 1.

(ii) The distribution ν has a positive density $\nu(x)=e^{-\Theta(x)}$ such that for any j, there are constants C_0 and C_1 such that
$$|\Theta^{(j)}(x)|\leqslant C_0(1+x^2)^{C_1} \tag{2.1.10}$$

(iii) The tail of the distribution ν has a subexponential decay. In other words, there exists C and q two positive constants such that
$$\int_{\mathbb{R}}\mathbb{1}_{|x|\geqslant y}\mathrm{d}\nu(x)\leqslant C\exp(-y^q) \tag{2.1.11}$$

We need the smoothness assumptions on W in order to use the reverse heat flow techniques from [EPR+10, ESY11]. Indeed, our result is an overwhelming probability bound on the eigenvectors of W_t. We think, however, that this property holds for a larger matrix ensembles and that the smoothness property is simply technical.

In the following, since the eigenvectors are concentrated on Nt sites, it is relevant to define the following notation for any set (which can be N-dependent) A, denote

$$\widehat{A} = \frac{|A|}{Nt} \wedge 1.$$

Indeed, having errors involving $\widehat{A}$ allows us to get bounds improving for $|A| \leqslant Nt$ but still holding for $|A| \gg Nt$.

Theorem 2.1.7. *Let $\kappa \in (0,1)$, ω a small positive constant, for instance $\omega < \varepsilon'/10$. Let $t \in \mathcal{T}_\omega$, $I \subset [\![1, N]\!]$ be a deterministic (N-dependent) set, W as in Definition 2.1.6 and W_t as in (2.1.1). Define now*

$$\Xi = \frac{\widehat{I}}{(Nt)^{1/3}} \quad \textit{and} \quad \sigma_t^2(\alpha, k) := \sigma_t^2(\alpha, k, \eta_0) \quad \textit{with} \quad N\eta_0 = \frac{\widehat{I}^{\,2}}{\Xi^2}. \tag{2.1.12}$$

Then we have, for any $\varepsilon > 0$ (small) and $D > 0$ (large) and for $k, \ell \in \mathcal{A}_r^\kappa$ with $k \neq \ell$, in the symmetric case

$$\mathbb{P}\left(\left|\sum_{\alpha \in I}\left(u_k(\alpha)^2 - \frac{1}{N}\sigma_t^2(\alpha, k)\right)\right| + \left|\sum_{\alpha \in I} u_k(\alpha)u_\ell(\alpha)\right| \geqslant N^\varepsilon \Xi\right) \leqslant N^{-D} \tag{2.1.13}$$

and in the Hermitian case,

$$\mathbb{P}\left(\left|\sum_{\alpha \in I}\left(|u_k(\alpha)|^2 - \frac{1}{2N}\sigma_t^2(\alpha, k)\right)\right| + \left|\sum_{\alpha \in I} u_k(\alpha)\bar{u}_\ell(\alpha)\right| \geqslant N^\varepsilon \Xi\right) \leqslant N^{-D}. \tag{2.1.14}$$

Remark 2.1.8. The choice of η_0 depends on our proof and is the one we should take to optimize our error Ξ. However, this error and the choice of η_0 do not seem optimal since we actually expect to have some form of Gausian fluctuations around this deterministic profile.

2.1.2. Method of Proof

Our proof is based on the three-step strategy from [EPR+10, ESY11] (see [EY17] for a recent book presenting this method). The first step is to have an optimal, local control of the spectral elements of the matrix ensemble given by a local law on the resolvent. The second step is to obtain the wanted result for a relaxation of the model by a small Gaussian perturbation. Finally, the third and last step consists of removing this Gaussian part. We will give the proof of Theorem 2.1.3, Corollary 2.1.5, Theorem 2.1.7 only in the symmetric case and refer the reader to [BY17, BYY18] for the tools needed in the Hermitian case.

First step: local laws for our model. In [LY17a], Landon-Yau showed a local law for the Dyson Brownian motion with a diagonal initial condition at all times. This result gives us an averaged local law on the Stieltjes transform but also an entrywise anisotropic local law for the resolvent. Since we want to consider any projection of the eigenvectors, we will also need a local law on the quadratic form $\langle \mathbf{q}, G(z)\mathbf{q}\rangle$. This control of the resolvent for mesoscopic perturbation has been showed in [BHY17]. Note that these results were done in the Gaussian case but can easily be generalized to the Wigner case with the right assumptions on moments.

Second step: short time relaxation. The second step consists of perturbing W_t by a small Gaussian component. We will obtain this perturbed model by making W_t undergo the Dyson Brownian motion given by the following definition.

Definition 2.1.9. Here is our choice of Dyson Brownian motion.
Let B be a $N \times N$ symmetric matrix such that B_{ij} for $i < j$ and $B_{ii}/\sqrt{2}$ are independent standard brownian motions. The $N \times N$ symmetric Dyson Brownian motion with initial condition H_0 is defined as

$$H_s = H_0 + \frac{1}{\sqrt{N}} B_s. \tag{2.1.15}$$

We also give the dynamics followed by the eigenvalues and the eigenvectors of such matrices.

Definition 2.1.10. Let $\boldsymbol{\lambda_0}$ be in the simplex $\Sigma_N = \{\lambda_1 < \cdots < \lambda_N\}$, $\boldsymbol{u_0}$ be an orthogonal $N \times N$ matrix, and $\tilde{B}$ as in (2.1.15). Consider the dynamics

$$\mathrm{d}\lambda_k = \frac{\mathrm{d}\tilde{B}_{kk}}{\sqrt{N}} + \frac{1}{N} \sum_{\ell \neq k} \frac{\mathrm{d}s}{\lambda_k - \lambda_\ell}, \tag{2.1.16}$$

$$\mathrm{d}u_k = \frac{1}{\sqrt{N}} \sum_{\ell \neq k} \frac{\mathrm{d}\tilde{B}_{k\ell}}{\lambda_k - \lambda_\ell} u_\ell - \frac{1}{2N} \sum_{\ell \neq k} \frac{\mathrm{d}s}{(\lambda_k - \lambda_\ell)^2} u_k \tag{2.1.17}$$

with initial condition $(\boldsymbol{\lambda_0}, \boldsymbol{u_0})$,

This eigenvector flow was first computed in different contexts such as [AGZ10] for GOE/GUE matrices, [Bru89] for real Wishart processes and [NRW86] for Brownian motion on ellipsoids.

Remark 2.1.11. If $\boldsymbol{\lambda_0}$ and $\boldsymbol{u_0}$ are the eigenvalues and eigenvectors of a fixed matrix H_0, then the solution to the dynamics from Definition 2.1.10 have the same distribution for any time s as the eigenvalues and eigenvectors of

$$H_s = H_0 + \sqrt{s}\mathrm{GOE}$$

with GOE being a matrix from the normalized Gaussian Orthogonal Ensemble (in the sense that its off-diagonal entries have variance $1/N$). In this paper, taking W to be such a matrix in (2.1.1), we study the eigenvectors of the Dyson Brownian motion with a diagonal initial condition after a mesoscopic time.

We will then need to study the eigenvectors of H_τ for a small $N^{-1} \ll \tau \ll t$. The convergence of joint moments of eigenvectors projections will be obtained by the maximum principle technique introduced in [BY17]. It is based on analyzing the dynamics followed by these moments. We will now recall notations and results on this eigenvector moment flow.

Take $\mathbf{q} \in \mathbb{R}^N$ such that $\|\mathbf{q}\|_2 = 1$ a fixed direction onto which we will project our eigenvectors. For $u_1^H, \ldots, u_N^H$ the eigenvectors of the matrix (2.1.15), define

$$z_k(s) = \sqrt{N} \langle \mathbf{q}, u_k^H(s) \rangle. \tag{2.1.18}$$

Now for $m \in [\![1, N]\!]$, denote by $j_1, \ldots, j_m$ positive integers and let $i_1, \ldots, i_m$ in $[\![1, N]\!]$ be distinct indices. We will consider the following normalized polynomials

$$Q_{i_1,\ldots,i_m}^{j_1,\ldots,j_m} = \prod_{l=1}^{m} z_{i_l}^{2j_l} a(2j_l)^{-1} \quad \text{where} \quad a(n) = \prod_{k \leq n, k \text{ odd}} k. \tag{2.1.19}$$

Note that $a(2n) = \mathbb{E}[\mathcal{N}^{2n}]$ with $\mathcal{N}$ a standard Gaussian random variables.

Consider a configuration of particles $\boldsymbol{\xi} : [\![1, N]\!] \to \mathbb{N}$ where $\xi_j := \boldsymbol{\xi}(j)$ is seen as the number of particles at the site j. We denote $\mathcal{N}(\boldsymbol{\xi}) = \sum_j \xi_j$ the total number of particles in the configuration $\boldsymbol{\xi}$.

Define $\boldsymbol{\xi}^{i,j}$ to be the configuration obtained by moving one particle from i to j. If there is no particle in i then $\boldsymbol{\xi}^{i,j} = \boldsymbol{\xi}$. It is clear that we can map $\{(i_1, j_1), \dots, (i_m, j_m)\}$ with distinct i_k's and positive j_k's summing to an $n > 0$ to a configuration $\boldsymbol{\xi}$ with $\xi_{i_k} = j_k$ and $\eta_l = 0$ if $l \notin \{i_1, \dots, i_m\}$. Define now, given this map,

$$f_{\boldsymbol{\lambda},s}(\boldsymbol{\xi}) := \mathbb{E}\left[Q_{i_1,\dots,i_m}^{j_1,\dots,j_m} | \boldsymbol{\lambda}\right], \tag{2.1.20}$$

a n-th joint moment of the coordinates of u_k. The conditioning here is on the full path of eigenvalues from 0 to ∞. The next theorem gives the eigenvector moment flow that $f_{\boldsymbol{\lambda},s}$ undergoes.

Theorem 2.1.12 ([BY17, Theorem 3.1])**.** *Suppose that $\boldsymbol{u}$ is the solution of the symmetric Dyson vector flow* (2.1.17) *and $f_{\boldsymbol{\lambda},s}(\boldsymbol{\xi})$ is given by* (2.1.20) *with the polynomials Q_s. Then it satisfies the equation*

$$\partial_s f_{\boldsymbol{\lambda},s}(\boldsymbol{\xi}) = \frac{1}{N} \sum_{i \neq j} \frac{2\xi_i(1 + 2\xi_j)\left(f_{\boldsymbol{\lambda},s}(\boldsymbol{\xi}^{i,j}) - f_{\boldsymbol{\lambda},s}(\boldsymbol{\xi})\right)}{(\lambda_i - \lambda_j)^2}. \tag{2.1.21}$$

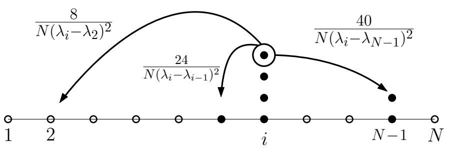


Figure 2.1: Example of the symmetric eigenvector moment flow with a configuration of 7 particles.

Note that this dynamics in the case of one particle was first obtained in [WW95]. Now that we have the expression of the eigenvector moment flow, we can give an heuristic for the apparition of a Cauchy profile in the variance (2.1.3). Indeed, the single particle case $m = 1$ gives us the variance of an entry of an eigenvector. To understand this result, consider the diagonal entries of the matrix D to be the quantiles of the semicircle law for instance, it is then interesting to consider the following continuous dynamics, define the operator K acting on smooth functions on $[-2, 2]$ as

$$(Kf)(x) = \int_{-2}^{2} \frac{f(x) - f(y)}{(x-y)^2} \mathrm{d}\rho(y). \tag{2.1.22}$$

The differential equation $\partial_t f = Kf$ can be seen as a deterministic and continuous equivalent of (2.1.21) because of the rigidity property of the Dyson Browian motion eigenvalues. We then get the following lemma from [BEYY16]

Lemma 2.1.13 ([BEYY16])**.** *Let f be smooth with all derivatives uniformly bounded. For any $x, y \in (-2, 2)$, denote $x = 2\cos\theta$, $y = 2\cos\phi$ with $\theta, \phi \in (0, \pi)$. Then*

$$\left(e^{-tK} f\right)(x) = \int p_t(x, y) f(y) \mathrm{d}\rho(y) \tag{2.1.23}$$

where the kernel is given by

$$p_t(x, y) := \frac{1 - e^{-t}}{\left|e^{i(\theta+\phi)} - e^{-t/2}\right|^2 \left|e^{i(\theta-\phi)} - e^{-t/2}\right|^2}. \tag{2.1.24}$$

Now at our small time-scale, we have

$$p_t(x,y) \sim \frac{t}{(x-y)^2+t^2}.$$

Hence (2.1.3) where $m=1$ can be considered as a result of stochastic homogenization in a non-equilibrium setting when we consider the dynamics (2.1.21) in the bulk.

For Theorem 2.1.7, we will study another observable which follows the same dynamics as in Theorem 2.1.12. This new observable has been analyzed in [BYY18] to obtain universality for a class of band matrices. Define now the centered eigenvectors overlaps for symmetric matrices,

$$p_{ij} = \sum_{\alpha\in I} u_i(\alpha)u_j(\alpha), \quad i\neq j\in I, \tag{2.1.25}$$

$$p_{ii} = \sum_{\alpha\in I} u_i(\alpha)^2 - C_0, \quad i\in I \tag{2.1.26}$$

where $\mathbf{u}$ are the eigenvectors of H_s and C_0 is any constant in the sense that it does not depend on i but can depend on N.

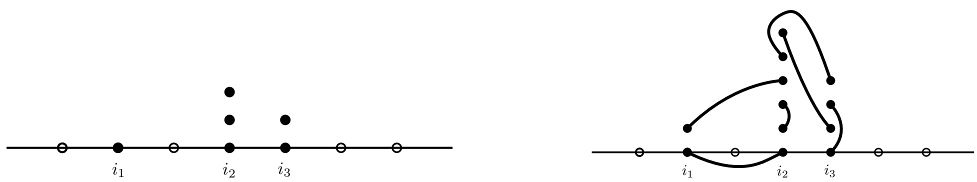
Now for $\boldsymbol{\eta}$ a configuration of n particles on N sites, define the following set

$$\mathcal{V}_{\boldsymbol{\eta}} = \{(i,a),\ 1\leqslant i\leqslant N,\ 1\leqslant a\leqslant 2\eta_i\}.$$

The set $\mathcal{V}$ will be a set of vertices. Consider now $\mathcal{G}_{\boldsymbol{\eta}}$ the set of perfect matchings on $\mathcal{V}_{\boldsymbol{\eta}}$. For any edge on G, $e=\{(i,a),(j,b)\}$, define $p(e)=p_{ij}$, $P(G)=\prod_{e\in\mathcal{E}(G)} p(e)$ and finally

$$F_{\boldsymbol{\lambda},s}(\boldsymbol{\eta}) = \frac{1}{\mathcal{M}(\boldsymbol{\eta})}\mathbb{E}\left[\sum_{G\in\mathcal{G}_{\boldsymbol{\eta}}} P(G)\middle|\boldsymbol{\lambda}\right] \tag{2.1.27}$$

where $\mathcal{M}(\boldsymbol{\eta})=\prod_{i=1}^N(2\eta_i)!!$, with $(2m)!!$ being the number of perfect matchings of the complete graph on $2m$ vertices. Note that this quantity depend on the eigenvalues trajectories $\boldsymbol{\lambda}$.



(a) A configuration $\boldsymbol{\eta}$ with $\mathcal{N}(\boldsymbol{\eta})=6$ particles

(b) An example of a perfect matching $G\in\mathcal{G}_{\boldsymbol{\eta}}^{(s)}$ with $P^{(s)}(G)=p_{i_1i_2}^2p_{i_2i_2}p_{i_2i_3}^2p_{i_3i_3}$

The previous quantity follows the same dynamics (2.1.21) as $f_{\boldsymbol{\lambda},s}(\boldsymbol{\xi})$.

Theorem 2.1.14 ([BYY18]). *Suppose that $\mathbf{u}$ is the solution of the symmetric Dyson vector flow (2.1.17) and $F_{\boldsymbol{\lambda},s}(\boldsymbol{\eta})$ is given by (2.1.27). Then it satisfies the equation*

$$\partial_s F_{\boldsymbol{\lambda},s}(\boldsymbol{\eta}) = \frac{1}{N}\sum_{i\neq j}\frac{2\eta_i(1+2\eta_j)\left(F_{\boldsymbol{\lambda},s}(\boldsymbol{\eta}^{i,j})-F_{\boldsymbol{\lambda},s}(\boldsymbol{\eta})\right)}{(\lambda_i-\lambda_j)^2}. \tag{2.1.28}$$

Third step: invariance of local statistics. The third and last step will be to obtain the result for the matrix W_t without any Gaussian part. We will do so using a variant of the dynamical method introduced in [BY17, Appendix A] which will show the continuity of resolvent statistics along the trajectory. This method was also used in [HLY15] to study sparse matrices. We will see that we need to take τ of order larger than N^{-1} but smaller than $\sqrt{t/N}$ to use the continuity argument. In order to remove the Gaussian component for Theorem 2.1.7, we will use the reverse heat flow and will need smoothness of the entries of our original matrix. Note that a moment matching scheme between the two matrix ensembles, which holds for any time $N^{-1} \ll \tau \ll 1$, could be used to obtain the invariance of local statistics. However it can be of later interest to obtain the continuity estimate up to time $\sqrt{t/N}$.

The next section will state the local laws proved in different papers ([LY17a] and [BHY17]). The third section is dedicated to prove Theorem 2.1.3, Corollary 2.1.5 and Theorem 2.1.7 for a short time relaxation of the matrix W_t. We use a maximum principle on f_s or F_s, a basic tool for the analysis of parabolic equations. However, since we want a local result, remember that the variance depends on the position of the spectrum, we need to localize the maximum principle. We finish with an induction on the number of particles in the multi-particle random walk or a mutli-scale argument. For the third step, we will need a continuity result for the Dyson Brownian motion which will be shown in Subsection 2.4.1 and we will give the reverse heat flow technique in Subsection 2.4.2. We will then conclude by combining the three steps in Section 2.5.

Acknowledgments. The author would like to kindly thank his advisor Paul Bourgade for many insightful and helpful discussions about this work.

2.2. Local laws

In this section, we focus on the different local laws result for W_t. These local laws are high probability bounds, for simplicity we will now introduce the following notation for stochastic domination. For

$$X = (X_N(u), N \in \mathbb{N}, u \in U_N), \quad Y = (Y_N(u), N \in \mathbb{N}, u \in U_N).$$

two families of nonnegative random variables depending on N (note that U_N can also depend on N), we will say that X is stochastically dominated by Y uniformly in ω, and write $X \prec Y$, if for all $\tau > 0$ and $D > 0$ we have

$$\sup_{u \in U_N} \mathbb{P}\left(X_N(u) > N^{\tau} Y_N(u)\right) \leqslant N^{-D}$$

for N large enough. If we have $|X| \prec Y$ for some family X, we will write $X = \mathcal{O}_{\prec}(Y)$.

Define now the resolvent of W_t and its normalized trace,

$$G(z) = (W_t - z)^{-1} = \sum_{k=1}^{N} \frac{|u_k\rangle\langle u_k|}{\lambda_k - z}, \quad \varsigma_t(z) = \frac{1}{N}\operatorname{Tr} G(z) = \frac{1}{N}\sum_{k=1}^{N} \frac{1}{\lambda_k - z} \tag{2.2.1}$$

and denote $G_{ij}(z)$ the (i,j) entry of the resolvent matrix. In the rest of the section we will omit the dependence in t of the resolvent since we are not looking at its dynamics.

2.2.1. Anisotropic local law for deformed Wigner matrices

An averaged local law was proved in [LY17a]. The proof relies on Schur's complement formula, large deviations bounds and interlacing formula in order to first state a weak local law on the resolvent entries and the Stieltjes transform. The result then follows from a fluctuation averaging lemma in order to go from the scale $(N\eta)^{-1/2}$ to $(N\eta)^{-1}$. We first give the definition of the limiting Stieltjes transform as the solution $m_t(z)$ such that $\operatorname{Im} m_t(z) > 0$ on the upper half plane of the following equation

$$m_t(z) = \frac{1}{N}\sum_{k=1}^{N} \frac{1}{D_i - z - tm_t(z)} = \frac{1}{N}\sum_{k=1}^{N} g_i(t,z) \tag{2.2.2}$$

where we defined

$$g_i(t,z) := \frac{1}{D_i - t - m_t(z)}.$$

We will also need the following lemma on the Stieltjes transform claiming that its imaginary part is of order one.

Lemma 2.2.1 ([LY17a, Lemma 7.2]). *Let $\vartheta, \omega > 0$ small constants and $\kappa \in (0,1)$. Take $z \in \mathcal{D}_r^{\vartheta,\kappa}$, for N large enough, the following bounds holds for $t \in \mathcal{T}_\omega$,*

$$c \leqslant \operatorname{Im} m_t(z) \leqslant C.$$

Moreover

$$ct \leqslant |D_i - z - tm_t(z)| \leqslant C.$$

Note that the constants above do not depend on any parameter.

Here is the averaged local law taken from [LY17a].

Theorem 2.2.2 ([LY17a, Theorem 3.3]). *Let W_t be as in Definition 2.1.1, $\vartheta > 0$ and $\kappa \in (0,1)$,*

$$|\varsigma_t(z) - m_t(z)| \prec \frac{1}{N\eta} \tag{2.2.3}$$

uniformly in $z = E + i\eta \in \mathcal{D}_r^{\vartheta,\kappa}$

The proof of Theorem 2.2.2 also gives the following entrywise local law from [LY17a] also properly stated in [BHY17].

Theorem 2.2.3 ([LY17a][BHY17, Theorem 2.4]). *Let W_t be as in Definition 2.1.1 and $\vartheta > 0$, $\kappa \in (0,1)$. Uniformly in $z = E + i\eta \in \mathcal{D}_r^{\vartheta,\kappa}$, we have for the diagonal entries*

$$|G_{ii}(z) - g_i(t,z)| \prec \frac{t}{\sqrt{N\eta}} |g_i(t,z)|^2, \tag{2.2.4}$$

and for the off-diagonal entries

$$|G_{ij}(z)| \prec \frac{1}{\sqrt{N\eta}} \min\{|g_i(t,z)|, |g_j(t,z)|\}. \tag{2.2.5}$$

In order to study $\langle \mathbf{q}, u_k \rangle$, we will need the following local law for $\langle \mathbf{q}, G(z)\mathbf{q} \rangle$ proved in [BHY17],

Theorem 2.2.4 ([BHY17]). *Let $\vartheta > 0$, $\kappa \in (0,1)$ and $\mathbf{q}$ a L^2-normalized vector of $\mathbb{R}^N$, we have*

$$\left|\langle \mathbf{q}, G(z)\mathbf{q}\rangle - \sum_{k=1}^N q_k^2 g_k(t,z)\right| \prec \frac{1}{\sqrt{N\eta}}\operatorname{Im}\left(\sum_{k=1}^N q_k^2 g_k(t,z)\right) \tag{2.2.6}$$

uniformly in $z = E + i\eta \in \mathcal{D}_r^{\vartheta,\kappa}$

This theorem also gives us control of the resolvent as a bilinear form by polarization. We will give the proof of this corollary for completeness.

Corollary 2.2.5. *Let $\vartheta > 0$, $\kappa \in (0,1)$, let $\mathbf{v}$ and $\mathbf{w}$ two L^2-normalized vectors of $\mathbb{R}^N$, we have*

$$\left|\langle \mathbf{v}, G(z)\mathbf{w}\rangle - \sum_{i=1}^N v_i w_i g_i(t,z)\right| \prec \frac{1}{\sqrt{N\eta}}\sqrt{\operatorname{Im}\left(\sum_{i=1}^N v_i^2 g_i(t,z)\right)\operatorname{Im}\left(\sum_{i=1}^N w_i^2 g_i(t,z)\right)} \tag{2.2.7}$$

uniformly in $z = E + i\eta \in \mathcal{D}_r^{\vartheta,\kappa}$

Proof. Let $\mu \in \mathbb{R}$, a parameter fixed later. Consider

$$\langle (\mathbf{v}+\mu\mathbf{w}), G(\mathbf{v}+\mu\mathbf{w})\rangle = \langle \mathbf{v}, G\mathbf{v}\rangle + \mu^2\langle \mathbf{w}, G\mathbf{w}\rangle + 2\mu\langle \mathbf{v}, G\mathbf{w}\rangle, \tag{2.2.8}$$

by linearity and symmetry of the resolvent G. On one hand, using Theorem 2.2.4 on the first two terms of the right hand side of (2.2.8), we get the equation

$$\begin{aligned}\langle (\mathbf{v}+\mu\mathbf{w}), G(\mathbf{v}+\mu\mathbf{w})\rangle &= 2\mu\langle \mathbf{v}, G\mathbf{w}\rangle + \sum_{i=1}^N v_i^2 g_i(t,z) + \mu^2\sum_{i=1}^N w_i^2 g_i(t,z)\\ &\quad + \mathcal{O}_\prec\left(\frac{1}{\sqrt{N\eta}}\left(\operatorname{Im}\left(\sum_{i=1}^N v_i^2 g_i(t,z) + \mu^2\sum_{i=1}^N w_i^2 g_i(t,z)\right)\right)\right).\end{aligned} \tag{2.2.9}$$

On the other hand, using Theorem 2.2.4 on the left hand side of (2.2.8), we obtain

$$\langle (\mathbf{v}+\mu\mathbf{w}), G(\mathbf{v}+\mu\mathbf{w})\rangle = \sum_{i=1}^N (v_i+\mu w_i)^2 g_i(t,z) + \mathcal{O}_\prec\left(\frac{1}{\sqrt{N\eta}}\operatorname{Im}\left(\sum_{i=1}^N (v_i+\mu w_i)^2 g_i(t,z)\right)\right). \tag{2.2.10}$$

Finally, combining (2.2.9) and (2.2.10) and choosing

$$\mu = \frac{\operatorname{Im}\left(\sum_{i=1}^N w_i v_i g_i(t,z)\right)}{\operatorname{Im}\left(\sum_{i=1}^N w_i^2 g_i(t,z)\right)},$$

we get the final result. □

We will also need the following rigidity result from [LY17a].

Theorem 2.2.6 ([LY17a, Theorem 3.5]). *Let $\omega > 0$ be a small constant and $\kappa \in (0,1)$. For any $t \in \mathcal{T}_\omega$, we have*

$$|\lambda_k - \gamma_{k,t}| \prec \frac{1}{N}$$

uniformly in $k \in \mathcal{A}_r^\kappa$

This control of the resolvent allows us to give an upper bound for the moments of the eigenvectors of W_t. Defining, for a fixed $\mathbf{q} \in \mathbb{S}^{N-1}$,

$$\varphi_t(\boldsymbol{\xi}) = \mathbb{E}\left[\prod_{k=1}^N \frac{(\sqrt{N}\langle \mathbf{q}, u_k\rangle)^{2\xi_k}}{a(2\eta_k)}\middle|\boldsymbol{\lambda}\right] \tag{2.2.11}$$

with $u_1, \dots, u_N$ the eigenvectors of W_t and $\boldsymbol{\lambda}$ its eigenvalues, we have the following corollary.

Corollary 2.2.7. *Let $\kappa \in (0,1)$, $\vartheta > 0$ and $\boldsymbol{\xi} : \mathcal{A}_r^\kappa \to \mathbb{N}$*

$$\varphi_t(\boldsymbol{\xi}) \prec \prod_{k=1}^N \sigma_t^{2\xi_k}(\mathbf{q}, k, \eta) =: \sigma_t^2(\mathbf{q}, \boldsymbol{\xi}, \eta). \tag{2.2.12}$$

uniformly in $N^{-1+\vartheta} \leqslant \eta \leqslant N^{-\vartheta} t$.

Proof. Fix any $\vartheta \in (0,1)$ such that $\eta = N^{-1+\vartheta} \ll t$ and $k \in \mathcal{A}_r^\kappa$, we have the following first high probability bound with $z_k = \lambda_k + \mathrm{i}\eta$

$$\begin{aligned}\frac{1}{\eta}\left(\sqrt{N}\langle \mathbf{q}, u_k\rangle\right)^2 &= \frac{\left(\sqrt{N}\langle \mathbf{q}, u_k\rangle\right)^2 \eta}{(\lambda_k - \mathrm{Re}(z_k))^2 + \eta^2} \leqslant N\mathrm{Im}\langle \mathbf{q}, G(z_k)\mathbf{q}\rangle \\ &= N\mathrm{Im}\left(\sum_{i=1}^N \frac{q_i^2}{D_i - z_k - tm_t(z_k)}\right) + \mathcal{O}_\prec\left(\frac{1}{\sqrt{N\eta}}\mathrm{Im}\left(\sum_{i=1}^N \frac{q_i^2}{D_i - z_k - tm_t(z_k)}\right)\right).\end{aligned}$$

We can then write,

$$\left(\sqrt{N}\langle \mathbf{q}, u_k\rangle\right)^2 \prec N\eta\mathrm{Im}\left(\sum_{i=1}^N \frac{q_i^2}{D_i - z_k - tm_t(z_k)}\right) \prec \sigma_t^2(\mathbf{q}, k, \eta).$$

where we used the definition of $\prec$ and that ϑ is as small as we want and the smoothness of $m_t(z)$. We finish the proof by definition of $\varphi_{\boldsymbol{\lambda},t}$. □

The bound from Theorem 2.2.4 is at the core of the proof of our main result and is the reason why σ_t has a Cauchy profile. It can be seen as an averaged version of Theorem 2.1.3. Indeed, let $z := E + \mathrm{i}\eta \in \mathcal{D}_r^{\vartheta,\kappa}$ then if we denote $\varphi_t(k)$ for $\varphi_t(\boldsymbol{\xi})$ where $\boldsymbol{\xi}$ is the configuration with a single particle in site k, we have

$$\begin{aligned}\mathrm{Im}\left(\frac{1}{N}\sum_{k=1}^N \frac{\varphi_t(k)}{\lambda_k - z}\right) &= \mathrm{Im}\left(\sum_{k=1}^N \frac{q_k^2}{D_k - z - tm_t(z)}\right) + \mathcal{O}_\prec\left(\frac{1}{\sqrt{N\eta}t}\right) \\ &= \sum_{k=1}^N \frac{q_k^2 t\mathrm{Im}\, m_t(z)}{(D_k - E)^2 + (t\mathrm{Im}\, m_t(z))^2} + \mathcal{O}_\prec\left(\frac{1}{\sqrt{N\eta}t}\right).\end{aligned} \tag{2.2.13}$$

The local law gives us a strong control on both the eigenvalues and eigenvectors of our matrix ensemble. As W_t will undergo the Dyson Brownian motion, these quantities will still be controlled through a local law up to a small error coming from the time of the relaxation. We will now define the event of good eigenvalue paths $(\boldsymbol{\lambda}(s))_{s\in(0,\tau)})$ in the sense that all the estimates and bound from the previous section such as local laws from Theorems 2.2.2, 2.2.4 and 2.2.6 hold. First denote the resolvent of H_s and its normalized trace by

$$G(s,z) = (H_s - z)^{-1} \quad \text{and} \quad \varsigma(s,z) = \frac{1}{N}\mathrm{Tr}\, H_s = \frac{1}{N}\sum_{k=1}^{N} \frac{1}{\lambda_k(t+s) - z}. \tag{2.2.14}$$

Definition 2.2.8. Let ε, $\vartheta > 0$ and $\kappa \in (0,1)$. An eigenvalue configuration $\boldsymbol{\lambda}$ is *good* is the following holds with overwhelming probability conditioning on $\boldsymbol{\lambda}(\tau_0) = \boldsymbol{\lambda}$ for N large enough,

1. $\sup_{0\leqslant s\leqslant\tau} |\varsigma(s,z) - m_{t+s}(z)| \leqslant N^{\varepsilon}(N\eta)^{-1}$ uniformly in $z \in \mathcal{D}_r^{\vartheta,\kappa}$.
2. $\sup_{0\leqslant s\leqslant\tau} |\langle \mathbf{q}, G(s,z)\mathbf{q}\rangle - \sum_{\alpha=1}^{N} q_\alpha^2 g_\alpha(t+s,z)| \leqslant N^{2\varepsilon}(N\eta)^{-1/2}\mathrm{Im}\sum_{\alpha=1}^{N} q_\alpha^2 g_\alpha(t+s,z)$ uniformly in $z \in \mathcal{D}_r^{\vartheta,\kappa}$.
3. $\sup_{0\leqslant s\leqslant\tau} |\lambda_i(t+s) - \gamma_{i,t+s}| \leqslant N^{\varepsilon}N^{-1}$ uniformly in $i \in \mathcal{A}_r^{\kappa}$

Note that by the considerations in this section and a continuity argument, we see that *good* eigenvalue paths occur with overwhelming probability so that we can condition on having such a path in the following section. Note that we will make W_t undergo the Dyson Brownian motion for a small time τ so that the classical location $\gamma_{i,t}$ and the deterministic counterpart to resolvent entries g_α have small variations. This is the statement of the next lemma.

Lemma 2.2.9. *Let ε, $\vartheta > 0$ and $\kappa \in (0,1)$. Conditionally on a* good *eigenvalue path as in Definition 2.2.8, we have the following control of the resolvent,*

$$\sup_{0\leqslant s\leqslant\tau} |\varsigma(s,z) - m_t(z)| \leqslant \frac{N^{\varepsilon}}{N\eta} + \frac{\tau}{t},$$

$$\sup_{0\leqslant s\leqslant\tau} \left| G(s,z)_{ii} - \frac{1}{D_i - z - tm_t(z)} \right| \leqslant \left(\frac{N^{2\varepsilon}}{\sqrt{N\eta}} + \frac{\tau}{t} \right) |g_\alpha(t,z)|$$

uniformly in $z \in \mathcal{D}_r^{\vartheta,\kappa}$. For the rigidity results for eigenvalues we have

$$\sup_{0\leqslant s\leqslant\tau} |\lambda_k(t+s) - \gamma_{k,t}| \leqslant N^{\varepsilon}\left(\frac{1}{N} + \tau\right)$$

uniformly in $k \in \mathcal{A}_r^{\kappa}$.

Proof. The first result comes from the fact that $|\partial_t m_t(z)| \leqslant N^{\varepsilon}/t$ which can be deduced from the time evolution of m_t from [LY17a, Lemma 7.6]

$$\partial_t m_t(z) = \partial_z \left(m_t(z)(m_t(z) + z) \right)$$

combined with the estimates $|\partial_z m_t(z)| \leqslant C/t$ and $|m_t(z)| \leqslant \log N$. The other error term comes from the local law holding for a *good* eigenvalue path. The proof of the second bound comes from the following simple identity

$$\partial_t g_\alpha(t,z) = \frac{m_t(z) + t\partial_t m_t(z)}{(D_i - z - tm_t(z))^2} \quad \text{which gives} \quad |\partial_t g_\alpha(t,z)| \leqslant \frac{N^{\varepsilon}}{t}|g_\alpha(t,z)|.$$

using the fact that $|m_t(z)| \leqslant C\log N$ and $|\partial_t m_t(z)| \leqslant N^{\varepsilon}/t$. For the rigidity estimate, we combine the estimate $|\partial_t \gamma_{i,t}| \leqslant C\log N$ which can be found in [LY17a, Lemma 7.6] as well with the rigidity coming from the *good* eigenvalue path.

□

2.3. Short time relaxation

In this section, we are going to prove Theorems 2.1.3 and 2.1.7 for the Dyson Brownian motion starting from W_t using maximum principle. Note that in this section we will omit the subscript $\boldsymbol{\lambda}$ for simplicity.

Recall the dynamics of the eigenvector moment flow with n particles for H_τ with $H_0 = W_t$.

$$\begin{cases} \partial_\tau f_\tau(\boldsymbol{\eta}) = \dfrac{1}{N}\displaystyle\sum_{i\neq j} \frac{2\eta_i(1+2\eta_j)\left(f_\tau(\boldsymbol{\eta}^{i,j}) - f_\tau(\boldsymbol{\eta})\right)}{(\lambda_i(t+\tau)-\lambda_j(t+\tau))^2} =: (\mathscr{B}_\tau f_\tau)(\boldsymbol{\eta}), \\ f_0(\boldsymbol{\eta}) = \varphi_t(\boldsymbol{\eta}) \end{cases} \tag{2.3.1}$$

with $\varphi_t(\boldsymbol{\eta})$ is defined in (2.2.11). where we noted $\lambda_i(t+\tau)$ the eigenvalues of H_τ. Note that in the case of a single particle in k, we can write the dynamics

$$\begin{cases} \partial_\tau f_\tau(k) &= \dfrac{2}{N}\displaystyle\sum_{j=1}^{N} \frac{f_\tau(j) - f_\tau(k)}{(\lambda_j(t+\tau)-\lambda_k(t+\tau))^2}, \\ f_0(k) &= \varphi_t(k). \end{cases} \tag{2.3.2}$$

We cut the dynamics into two parts : the short range where most of the information will be and the long range. This decomposition in this context was first introduced in [EY15]. Letting $1 \ll \ell \ll N$ be a parameter that we will choose later, we then define

$$(\mathscr{S}(\tau)f_\tau)(\boldsymbol{\eta}) = \frac{1}{N}\sum_{|j-k|\leqslant \ell}^{N} \frac{2\eta_i(1+2\eta_j)\left(f_\tau(\boldsymbol{\eta}^{i,j}) - f(\boldsymbol{\eta})\right)}{(\lambda_i-\lambda_j)^2}, \tag{2.3.3}$$

$$(\mathscr{L}(\tau)f_\tau)(k) = \frac{1}{N}\sum_{|j-k|>\ell}^{N} \frac{2\eta_i(1+2\eta_j)\left(f_\tau(\boldsymbol{\eta}^{i,j}) - f_\tau(\boldsymbol{\eta})\right)}{(\lambda_i-\lambda_j)^2}. \tag{2.3.4}$$

Denote by $U_{\mathscr{S}}(s,\tau)$ the semigroup associated with $\mathscr{S}$ from time s to τ :

$$\partial_\tau U_{\mathscr{S}}(s,\tau) = \mathscr{S}(\tau) U_{\mathscr{S}}(s,\tau) \tag{2.3.5}$$

for any $s \leqslant \tau$. We will denote in the same way $U_{\mathscr{B}}$. It has been proved in [BY17, BHY17] that the parabolic short range dynamics has a finite speed of propagation in the following sense: define the following distance on the set of configurations with n particles

$$d(\boldsymbol{\eta},\boldsymbol{\xi}) = \sum_{k=1}^{n} |x_k - y_k| \tag{2.3.6}$$

where $(x_1,\dots,x_n)$ are the positions of the particles in nondecreasing order of $\boldsymbol{\eta}$ and y_α of $\boldsymbol{\xi}$. The following lemma then states that if two configurations are far from each other, the short-range dynamics started at one and evaluated at the other is exponentially small with high probability.

Lemma 2.3.1 ([BHY17, Corollary 3.3])**.** *Choose $\ell \geqslant N\tau$, let $\varepsilon > 0$ be a small constant and $\kappa \in (0,1)$. Conditioning on a good eigenvalue path $(\boldsymbol{\lambda}(s))_s$, uniformly, for any function h on configurations of n particles and a configuration $\boldsymbol{\xi}$ outside of the support of h in the sense that $d(\boldsymbol{\xi},\boldsymbol{\eta}) \geqslant N^{\varepsilon}\ell$ for any configuration $\boldsymbol{\eta}$ inside the support of h, we have*

$$\sup_{0\leqslant s\leqslant s'\leqslant t} U_{\mathscr{S}}(s,s')h(\boldsymbol{\xi}) \leqslant N^n \|h\|_\infty e^{-cN^{\varepsilon}} \tag{2.3.7}$$

for any $D > 0$.

In this section, we condition on an event occurring with overwhelming probability so that we can state the results deterministically. We state it here as the following lemma.

Lemma 2.3.2. *Let $\omega, \mathfrak{a}, \mathfrak{b}, \vartheta$ and ε be small positive constants and D as in Definition 2.1.1. Let $t \in \mathcal{T}_\omega$, and W_t as in (2.1.1). Let $\kappa \in (0,1)$, $\tau \in [N^{-1+\mathfrak{a}}, N^{-\mathfrak{a}}t]$ and $\ell \in [\tau N^{1+\mathfrak{b}}, N^{-\mathfrak{b}}t]$. The dynamics $(H_s)_{0\leqslant s\leqslant\tau}$ induces a measure on the space of eigenvalues and eigenvectors $(\boldsymbol{\lambda}(t+s), \mathbf{u}(t+s))_{0\leqslant s\leqslant\tau}$. The event A of trajectories defined by the following holds with overwhelming probability:*

1. $\sup_{0\leqslant s\leqslant\tau} |\varsigma(s,z) - m_t(z)| \leqslant N^\varepsilon (N\eta)^{-1} + \tau t^{-1}$ *uniformly in* $z \in \mathcal{D}_r^{\vartheta,\kappa}$.
2. $\sup_{0\leqslant s\leqslant\tau} |\langle \mathbf{q}, G(s,z)\mathbf{q}\rangle - \sum_{\alpha=1}^N q_\alpha^2 g_\alpha(t,z)| \leqslant (N^{2\varepsilon}(N\eta)^{-1/2} + \tau t^{-1})\mathrm{Im}\sum_{\alpha=1}^N q_\alpha^2 g_\alpha(t,z)$ *uniformly in* $z \in \mathcal{D}_r^{\vartheta,\kappa}$.
3. $\sup_{0\leqslant s\leqslant\tau} |\lambda_i(t+s) - \gamma_{i,t}| \leqslant N^\varepsilon(N^{-1}+\tau)$ *uniformly in* $i \in \mathcal{A}_r^\kappa$.
4. *For any function h on configurations of n particles and a configuration $\boldsymbol{\xi}$ supported outside of the support of h in the sense that $d(\boldsymbol{\xi}, \boldsymbol{\eta}) \leqslant N^\varepsilon \ell$ for any configuration $\boldsymbol{\eta}$ inside the support of h we have*
$$\sup_{0\leqslant s\leqslant s'\leqslant t} U_{\mathscr{S}}(s,s')h(\boldsymbol{\xi}) \leqslant N^n \|h\|_\infty e^{-cN^\varepsilon}.$$

The following lemma gives us a bound on the difference between the short-range and long-range dynamics basically stating that most of the information lies in the short-range dynamics.

Lemma 2.3.3. *Fix $\ell \ll Nt$ and consider $\boldsymbol{\xi_0}$ to be a configuration of n particles supported on $\mathcal{A}_r^\kappa$ then for any eigenvalue paths $(\boldsymbol{\lambda}(t+s), \boldsymbol{u}(t+s))_{0\leqslant s\leqslant\tau}$ in A, we have for any $\vartheta > 0$ and $N^{-1+\vartheta} \leqslant \eta \leqslant N^{-\vartheta}\ell/N$,*

$$|(U_{\mathscr{B}}(0,\tau) - U_{\mathscr{S}}(0,\tau))\varphi_t(\boldsymbol{\xi})| \leqslant N^{(n+4)\varepsilon}\frac{N\tau}{\ell}\sigma_t^2(\mathbf{q},\boldsymbol{\xi},\eta). \tag{2.3.8}$$

Proof. Let $\eta \in [N^{-1+\vartheta}, N^{-\vartheta}t]$. Using Duhamel's formula we can write

$$|(U_{\mathscr{B}}(0,\tau) - U_{\mathscr{S}}(0,\tau))\varphi_t(\boldsymbol{\xi})| = \left|\int_0^\tau U_{\mathscr{S}}(s,\tau)\mathscr{L}(s)f_s(\boldsymbol{\xi})\mathrm{d}s.\right|$$

Now, by definition of the operator $\mathscr{L}(s)$ we have that

$$\mathscr{L}(s)f_s(\boldsymbol{\xi}) = \sum_{j,k,|j-k|>\ell} 2\eta_j(1+2\eta_k)\frac{f_s(\boldsymbol{\xi}^{\boldsymbol{j},\boldsymbol{k}}) - f_s(\boldsymbol{\xi})}{N(\lambda_j-\lambda_k)^2}$$

so that we can bound, using Corollary 2.2.7 since $\boldsymbol{\xi}$ is supported on $\mathcal{A}_r^\kappa$,

$$|\mathscr{L}(s)f_s(\boldsymbol{\xi})| \leqslant 2n(1+2n)\sum_{j,k:|j-k|>\ell}\frac{N^{n\varepsilon}\sigma_t^2(\mathbf{q},\boldsymbol{\xi},\eta) + |f_s(\boldsymbol{\xi}^{\boldsymbol{k},\boldsymbol{j}})|}{N(\lambda_j-\lambda_k)^2}.$$

Now, say the configuration is supported on p sites denoted $(k_1,\dots,k_p)$, then one can write

$$|\mathscr{L}(s)f_s(\boldsymbol{\xi})| \leqslant C_n\sum_{i=1}^p\sum_{j:|j-k_i|>\ell}\frac{N^{n\varepsilon}\sigma_t^2(\mathbf{q},\boldsymbol{\xi},\eta) + |f_s(\boldsymbol{\xi}^{\boldsymbol{k_i},\boldsymbol{j}})|}{N(\lambda_j-\lambda_{k_i})^2}. \tag{2.3.9}$$

Now, since $\boldsymbol{\xi}$ is supported on $\mathcal{A}_r^\kappa$, we have that by Corollary 2.2.7 and denoting $\boldsymbol{\xi}\setminus k_i$ the configuration $\boldsymbol{\xi}$ where we removed a particule from the site k_i,

$$f_s(\boldsymbol{\xi}^{k_i,j}) \leqslant N^{(n-1)\varepsilon}\sigma_t^2(\mathbf{q},\boldsymbol{\xi}\setminus k_i,\eta)f_s(j). \tag{2.3.10}$$

Consider now $\eta_q = 2^q\ell/N$ for $q=[0,\lceil\log_2(N/\ell)\rceil]$, then we can bound

$$\sum_{j:|j-k_i|>\ell}\frac{f_s(j)}{N(\lambda_j-\lambda_{k_i})^2} \leqslant \sum_{j:|j-k_i|>\ell}\sum_{q=0}^{\lceil\log_2(N/\ell)\rceil}\frac{1}{\eta_q}\frac{\mathbb{E}\left[\langle\mathbf{q},\,u_j\rangle^2|\boldsymbol{\lambda}\right]\eta_q}{(\lambda_j-\lambda_{k_i})^{+}\eta_q^2}$$
$$\leqslant \sum_{q=0}^{\lceil\log_2(N/\ell)\rceil}\frac{N}{2^q\ell}\mathrm{Im}\mathbb{E}\left[\langle\mathbf{q},\,G(s,\lambda_{k_i}+\mathrm{i}\eta_q)\mathbf{q}\rangle|\boldsymbol{\lambda}\right]. \tag{2.3.11}$$

We can now use the anisotropic local law since λ_{k_i} lies in the spectral window and since we are on the event A,

$$\mathrm{Im}\langle\mathbf{q},G(s,\lambda_{k_i}+\mathrm{i}\eta_q)\mathbf{q}\rangle \leqslant \mathrm{Im}\langle\mathbf{q},G(s,\lambda_{k_i}+\mathrm{i}\eta)\mathbf{q}\rangle \leqslant N^\varepsilon\mathrm{Im}\,m_t(\lambda_{k_i}+\mathrm{i}\eta)\sigma_t^2(\mathbf{q},k_i,\eta) \leqslant N^{2\varepsilon}\sigma_t^2(\mathbf{q},k_i,\eta). \tag{2.3.12}$$

where we used the fact that $\eta \ll \ell/N \leqslant \eta_q$ and that $m_t(z)$ is bounded in the spectral window. Combining the estimates (2.3.11) and (2.3.12) we obtain that

$$\sum_{j:|j-k_i|>\ell}\frac{f_s(j)}{N(\lambda_j-\lambda_{k_i})^2} \leqslant N^{2\varepsilon}\frac{N}{\ell}\sigma_t^2(\mathbf{q},k_i,\eta).$$

Injecting this bound in (2.3.9) with (2.3.10), we obtain that

$$|\mathscr{L}(s)\varphi_t(\boldsymbol{\xi})| \leqslant N^{(n+2)\varepsilon}\frac{N}{\ell}\sigma_t^2(\mathbf{q},\boldsymbol{\xi},\eta)$$

where we used the fact that, by the same argument as in (2.3.11),

$$\sum_{|j-k_i|>\ell}\frac{1}{N(\lambda_j-\lambda_{k_i})^2} \lesssim \frac{N}{\ell}.$$

Now, we can use that $U_{\mathscr{S}}$ is a contraction combined with the finite speed of propagation from Lemma 2.3.1 so that we can write that

$$|U_{\mathscr{S}}(s,\tau)\mathscr{L}(s)f_s(\boldsymbol{\xi})| \leqslant N^\varepsilon\sup_{\boldsymbol{\eta}:d(\boldsymbol{\eta},\boldsymbol{\xi})\leqslant N^\varepsilon\ell}|\mathscr{L}(s)f_s(\boldsymbol{\eta})| \leqslant N^{(n+3)\varepsilon}\frac{N}{\ell}\sup_{\boldsymbol{\eta}:d(\boldsymbol{\eta},\boldsymbol{\xi})\leqslant N^\varepsilon\ell}\sigma_t^2(\mathbf{q},\boldsymbol{\eta},\eta). \tag{2.3.13}$$

However, since we have that $\boldsymbol{\eta}$ is close to $\boldsymbol{\xi}$ we can use regularity of $\sigma_t^2(\mathbf{q},\boldsymbol{\eta},\eta)$. Indeed, if one looks at the function

$$\psi(x) = \sum_{\alpha=1}^N\frac{q_\alpha^2 t}{(D_\alpha-x-t\mathrm{Re}\,m_t(x))^2+(t\mathrm{Im}\,m_t(x))^2}.$$

Then we have that

$$\partial_x\psi(x) = \sum_{\alpha=1}^N\frac{q_\alpha^2 t(2(1+t\partial_x\mathrm{Re}\,m_t(x))(D_\alpha-x-t\mathrm{Re}\,m_t(x))-2t^2\partial_x\mathrm{Im}\,m_t(x)\mathrm{Im}\,m_t(x)}{\left[(D_\alpha-x-t\mathrm{Re}\,m_t(x))^2+(t\mathrm{Im}\,m_t(x))^2\right]^2}$$

So that we can obtain the bound using the fact that $|\partial_x m_t(x)| \leqslant N^{\varepsilon}/t$,

$$|\partial_x \psi(x)| \leqslant \frac{N^{\varepsilon}}{t}\psi(x).$$

We can use this bound in order to obtain the following variation formula for $\sigma_t^2(\mathbf{q}, k, \eta)$,

$$\left|\sigma_t^2(\mathbf{q}, \boldsymbol{\xi}, \eta) - \sigma_t^2(\mathbf{q}, \boldsymbol{\eta}, \eta)\right| \leqslant N^{\varepsilon}\frac{d(\boldsymbol{\xi}, \boldsymbol{\eta})}{Nt}\sigma_t^2(\mathbf{q}, \boldsymbol{\xi}, \eta). \tag{2.3.14}$$

In (2.3.13), one can see that the supremum is only taken over configurations close to each other, namely such that $d(\boldsymbol{\xi}, \boldsymbol{\eta}) \leqslant N^{\varepsilon\ell}$. Since we have $\ell \ll Nt$, by (2.3.14), $\sigma_t^2(\mathbf{q}, \boldsymbol{\xi}, \eta)$ varies slowly and we can finally bound

$$|U_{\mathscr{S}}(s, \tau)\mathscr{L}(s)f_s(\boldsymbol{\xi})| \leqslant N^{(n+4)\varepsilon}\frac{N}{\ell}\sigma_t^2(\mathbf{q}, \boldsymbol{\xi}, \eta)$$

□

The two previous lemmas will be very useful tools to prove Theorem 1.4. Indeed, the finite speed of propagation in Lemma 2.3.1 allows us to localize our problem but is a property of the short range dynamics, Lemma 2.3.3 then tells us that most of the information of the global dynamics is in this short range part.

2.3.1. Analysis of the moment observable

To prove Theorem 2.1.3 we will prove the following intermediary proposition,

Proposition 2.3.4. *Conditionally on $(\boldsymbol{\lambda}, \boldsymbol{u}) \in A$, let $\kappa \in (0, 1)$, $\varepsilon > 0$ and n be an integer. If $\mathbf{q} \in \mathbb{S}^{N-1}$, for any $\boldsymbol{\xi} : \mathcal{A}_r^{\kappa} \to \mathbb{N}$ such that $\mathcal{N}(\boldsymbol{\xi}) = n$, there exists a p depending on n such that we have*

$$f_\tau(\boldsymbol{\xi}) = \sigma_t^2(\mathbf{q}, \boldsymbol{\xi}, N^{-\varepsilon}\tau) + \mathcal{O}\left(N^{p\varepsilon}\left(\frac{1}{\sqrt{N\tau}} + \left(\frac{\tau}{t}\right)^{1/3}\right)\sigma_t^2(\mathbf{q}, \boldsymbol{\eta}, N^{-\varepsilon}\tau)\right). \tag{2.3.15}$$

where $\sigma_t(\mathbf{q}, k, \tau)$ is given by (2.1.3).

The $1/3$ exponent that we give here in the error is not optimal. We are not able to reach an optimal error because of the strong dichotomy we do between the short range and the long range dynamics and also the localization technique. Using a multi-scale partition of the dynamics could improve the error term. Note also the choice of the parameter η corresponds to $N^{-\varepsilon}\tau$ which optimize our error term.

Let $\varepsilon > 0$ be a small constant. Recall that $t \in \mathcal{T}_\omega$ and $\tau \ll t$. First, the following lemma gives us a local law for f_τ, in the case of a single particle, deduced from the isotropic local law for W_t in Theorem 2.2.4.

Lemma 2.3.5. *For $z \in \mathcal{D}_r^{\vartheta,\kappa}$, we have*

$$\operatorname{Im}\left(\frac{1}{N}\sum_{k=1}^{N}\frac{f_\tau(k)}{\lambda_k(t+\tau) - z}\right) = \operatorname{Im}\left(\sum_{k=1}^{N}\frac{q_k^2}{D_k - z - tm_t(z)}\right)$$
$$+ \mathcal{O}\left(N^{\varepsilon}\left(\frac{1}{\sqrt{N\eta}} + \frac{\tau}{t}\right)\operatorname{Im}\left(\sum_{k=1}^{N}\frac{q_k^2}{D_k - z - tm_t(z)}\right)\right).$$

Proof. See first that, by definition of f_τ, we have

$$\frac{1}{N}\sum_{k=1}^{N}\frac{f_\tau(k)}{\lambda_k(t+\tau)-z}=\mathbb{E}\left[\langle \mathbf{q}, G^{H_\tau}(z)\mathbf{q}\rangle|\boldsymbol{\lambda}\right]$$

where $G^{H_\tau} := (H_\tau - z)^{-1}$ is the resolvent of H_τ. Now, the law of H_τ is $D+\sqrt{t}W+\sqrt{\tau}\mathrm{GOE}\overset{(d)}{=}D+\sqrt{t+\tau}W'$ for some W' a Wigner matrix. We can use Theorem 2.2.4 for this matrix and write

$$\mathrm{Im}\left(\langle \mathbf{q}, G^{H_\tau}\mathbf{q}\rangle\right)=\mathrm{Im}\left(\sum_{k=1}^{N}\frac{q_k^2}{D_k-z-(t+\tau)m_{t+\tau}(z)}\right) + \mathcal{O}\left(N^{\varepsilon}\frac{1}{\sqrt{N\eta}}\mathrm{Im}\left(\sum_{k=1}^{N}\frac{q_k^2}{D_k-z-(t+\tau)m_{t+\tau}(z)}\right)\right)$$

with $m_{t+\tau}(z)$ the solution with positive imaginary part of the following self-consistent equation,

$$m_{t+\tau}(z)=\frac{1}{N}\sum_{k=1}^{N}\frac{1}{D_k-z-(t+\tau)m_{t+\tau}(z)}.$$

Note that we have, for $z\in\mathcal{D}_r^{\vartheta,\kappa}$, $0<\mathrm{Im}\left(m_{t+\tau}(z)\right)\leqslant C$ for some constant C so that, by a Taylor expansion in $\tau \ll t$ (remember that $\eta\ll t$ for $z\in\mathcal{D}_r^{\vartheta,\kappa}$), we obtain

$$\mathrm{Im}\left(\sum_{k=1}^{N}\frac{q_k^2}{D_k-z-(t+\tau)m_{t+\tau}(z)}\right)=\mathrm{Im}\left(\sum_{k=1}^{N}\frac{q_k^2}{D_k-z-tm_{t+\tau}(z)}\right) + \mathcal{O}\left(N^{\varepsilon}\frac{\tau}{t}\mathrm{Im}\left(\sum_{k=1}^{N}\frac{q_k^2}{D_k-z-tm_{t}(z)}\right)\right).$$

We get the final result with the bound $|\partial_t m_t(z)|\leqslant C(\log N)/t$. □

Let $\boldsymbol{\xi}_0\subset\mathcal{A}_r^{\kappa}$ be a fixed configuration, we want to use a maximum principle on a window centered around $\boldsymbol{\xi}_0$ of size w according to the distance (2.3.6). Since we make a small perturbation $\tau\ll t$, in order to notice the dynamics in this window, we need to have $w\gg N\tau$. Furthermore, we want to look in the part of the spectrum where the eigenvector will be of typical size $1/\sqrt{Nt}$, to localize the dynamics in this small part of the spectrum. We then need to take $w\ll Nt$.

We define the following flattening and averaging operator, for $a>0$

$$\left(\mathrm{Flat}^a_{\boldsymbol{\xi}_0}f\right)(\boldsymbol{\eta})=\begin{cases} f(\boldsymbol{\eta}) & \text{if } d(\boldsymbol{\eta},\boldsymbol{\xi}_{\boldsymbol{w}})\leqslant a,\\ \sigma_t^2(\mathbf{q},\boldsymbol{\xi}_{\mathbf{0}}) & \text{if } d(\boldsymbol{\eta},\boldsymbol{\xi}_{\boldsymbol{w}})>a\end{cases} \tag{2.3.16}$$

and

$$(\mathrm{Av}_{\xi_0}f)(\boldsymbol{\eta})=\frac{2}{w}\int_{w/2}^{w}(\mathrm{Flat}^a_{\xi_0}f)(\boldsymbol{\eta})\mathrm{d}a. \tag{2.3.17}$$

Notice that for every $\boldsymbol{\eta}$, there exists $a_{\boldsymbol{\eta}}\in[0,1]$ such that

$$\mathrm{Av}_{\xi_w}f(\boldsymbol{\eta})=a_{\boldsymbol{\eta}}f(\boldsymbol{\eta})+(1-a_{\boldsymbol{\eta}})\sigma_t^2(\mathbf{q},\boldsymbol{\xi}_{\mathbf{0}}). \tag{2.3.18}$$

Proof of Proposition 2.3.4 : Case of a single particle

To show the result (2.3.15) by induction, we will first prove in the case of one particle with the dynamics (2.3.2). We first prove under the hypotheses as in Proposition 2.3.4 in the case where $n=1$,

$$f_\tau(k) = \sigma_t^2(\mathbf{q},k,\eta) + \mathcal{O}\left(N^\varepsilon\left(\frac{1}{\sqrt{N\tau}}+\left(\frac{\tau}{t}\right)^{1/3}\right)\sigma_t^2(\mathbf{q},k,\eta)\right). \tag{2.3.19}$$

To do so, we want to use a localized maximum principle centered around k_w which is the position of the particle for the configuration $\boldsymbol{\xi_0}$ in this case. However, we need to know that the maximum stays in that window, that is why we first flatten and average $f_t(k)$ and use it as an initial condition for the dynamics (2.1.21). We will then make the short range dynamics work on $f_t(k)$ during a time $\tau \ll t$. Since we use the short range dynamics for a time τ we will be able to use the finite speed of propagation (2.3.1) and we should choose $Nt \gg \ell \geqslant N\tau$ for the range cut-off. We will take an explicit value at the end of the proof. The different parameters and scaling is illustrated in Figure 2.3.

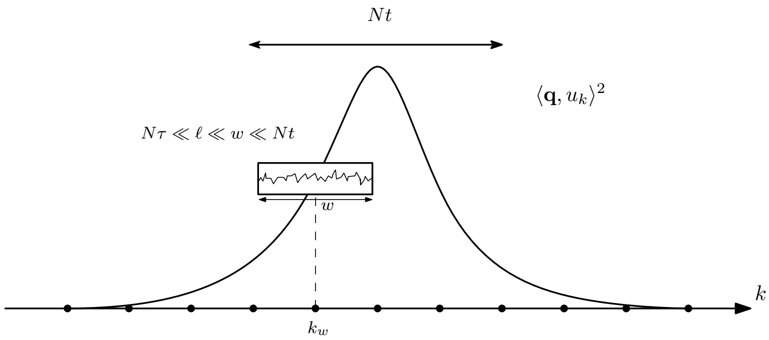


Figure 2.3: We see here a sketch of the variance profile plotted in the spectral dimension: the projection $\mathbf{q}$ is fixed and we plot the profile as a function of the eigenvector u_k. We then localize the dynamics onto the small window plotted here: the window is small enough so that the eigenvector can be seen as "flat" but large enough so that the short-range dynamics will not involve indices outside of this window.

Consider g_τ,

$$\partial_\tau g_\tau(k) = \frac{1}{N}\sum_{|j-k|\leqslant \ell}\frac{g_\tau(j)-g_\tau(k)}{(\lambda_j(t+\tau)-\lambda_k(t+\tau))^2} \quad \text{with} \quad g_0(k) = (\mathrm{Av}_{k_w}\varphi_t)(k).$$

First note that, in order to prove (2.3.19), it is enough to show that

$$g_\tau(k) = \sigma_t^2(\mathbf{q},k,\eta) + \mathcal{O}\left(\left(\frac{\ell}{w}+\frac{N\tau}{\ell}+\frac{1}{\sqrt{N\eta}}+\frac{\tau}{\eta}+\frac{w}{Nt}+\frac{N\eta}{\ell}\right)\sigma_t^2(\mathbf{q},k,\eta)\right) \tag{2.3.20}$$

where the parameter η, the spectral resolution, will be chosen so that $N^{-1} \ll \eta \ll t$. Indeed we have,

$$\begin{aligned} |f_\tau(k) - g_\tau(k)| &= |(U_{\mathscr{B}}(0,\tau)\varphi_t)(k) - (U_{\mathscr{S}}(0,\tau)\mathrm{Av}_{k_w}\varphi_t)(k)| && (2.3.21)\\ &= \left[(U_{\mathscr{B}}(0,\tau)-U_{\mathscr{S}}(0,\tau))\varphi_t\right](k) + [U_{\mathscr{S}}(0,\tau)(\mathrm{Id}-\mathrm{Av}_{k_w})\varphi_t](k). && (2.3.22)\end{aligned}$$

By Lemma 2.3.3, we get

$$[(U_{\mathscr{B}}(0,\tau)-U_{\mathscr{S}}(0,\tau))\varphi_t](k) \leqslant N^{(n+6)\varepsilon}\frac{N\tau}{\ell}\sigma_t^2(\mathbf{q},k,\eta). \tag{2.3.23}$$

Now, for the second term in (2.3.22). Since $(\varphi_t - \mathrm{Av}_{k_w}\varphi_t)(k) = 0$ for $k \in [\![k_w - w/2, k_w + w/2]\!]$, and taking $w \gg \ell N^{\varepsilon}$, looking at $k \in [\![k_w - w/3, k_w + w/3]\!]$ for instance, Lemma 2.3.1 tells us that the term is exponentially small. Thus, we obtain

$$f_\tau(k) = g_\tau(k) + \mathcal{O}\left(N^{(n+6)\varepsilon}\frac{N\tau}{\ell}\sigma_t^2(\mathbf{q}, k, \eta)\right).$$

We will first prove the following equation which can be seen as an averaged version of the result. We will now show the following lemma whic is analogous to [BY17, Lemma 7.3]

Lemma 2.3.6. *For $k_0 \in [\![k_w - u, k_w + u]\!]$, set $z^{(k_0)} = \gamma_{k_0,t} + \mathrm{i}\eta \in \mathcal{D}_r^{\vartheta,\kappa}$*

$$\left|\mathrm{Im}\left(\frac{1}{N}\sum_{k=1}^N \frac{g_\tau(k)}{\lambda_k(t) - z^{(k_0)}}\right) - \mathrm{Im}\left(m_t(z^{(k_0)})\right)\sigma_t^2(\mathbf{q}, k_w, \eta)\right|$$
$$\leqslant N^{(n+6)\varepsilon}\left(\frac{\ell}{w} + \frac{N\tau}{\ell} + \frac{1}{\sqrt{N\eta}} + \frac{\tau}{\eta} + \frac{w}{Nt} + \sqrt{\eta}\right)\sigma_t^2(\mathbf{q}, k_w, \eta). \quad (2.3.24)$$

Proof. First, decompose the left hand side term into three different terms :

$$\left|\mathrm{Im}\left(\frac{1}{N}\sum_{k=1}^N \frac{(U_{\mathscr{S}}(0,\tau)\mathrm{Av}_{k_w}\varphi_t)(k) - (\mathrm{Av}_{k_w}U_{\mathscr{S}}(0,\tau)\varphi_t)(k)}{\lambda_k - z^{(k_0)}}\right)\right| \quad (2.3.25)$$

$$+\left|\mathrm{Im}\left(\frac{1}{N}\sum_{k=1}^N \frac{(\mathrm{Av}_{k_w}U_{\mathscr{S}}(0,\tau)\varphi_t)(k) - \mathrm{Av}_{k_w}f_\tau(k)}{\lambda_k - z^{(k_0)}}\right)\right| \quad (2.3.26)$$

$$+\left|\mathrm{Im}\left(\frac{1}{N}\sum_{k=1}^N \frac{\mathrm{Av}_{k_w}f_\tau(k)}{\lambda_k - z^{(k_0)}}\right) - \mathrm{Im}\left(m_t(z^{(k_0)})\right)\sigma_t^2(\mathbf{q}, k_w, \eta)\right|. \quad (2.3.27)$$

To bound (2.3.25), we write

$$(U_{\mathscr{S}}(0,\tau)\mathrm{Av}_{k_w}\varphi_t)(k) - (\mathrm{Av}_{k_w}U_{\mathscr{S}}(0,\tau)\varphi_t)(k) = \frac{2}{w}\int_{w/2}^{w}\mathfrak{U}_{k_w}^a(\varphi_t)(k)\mathrm{d}a \quad (2.3.28)$$

with

$$\mathfrak{U}_{k_w}^a(\varphi_t)(k) = (U_{\mathscr{S}}(0,\tau)\mathrm{Flat}_{k_w}^a\varphi_t)(k) - (\mathrm{Flat}_{k_w}^a U_{\mathscr{S}}(0,\tau)\varphi_t)(k).$$

Look now at what happens around $k_w - a$, the other boundary of the window $k_w + a$ can be bounded exactly the same way. By finite speed of propagation from Lemma 2.3.1, for $k < k_w - a - \ell N^{\varepsilon}$, we easily get

$$\left(U_{\mathscr{S}}(0,\tau)\mathrm{Flat}_{k_w}^a\varphi_t\right)(k) = \mathrm{Flat}_{k_w}^a\left(U_{\mathscr{S}}(0,\tau)\varphi_t\right)(k) + \mathcal{O}\left(N^n e^{-N^{\varepsilon/2}}\right).$$

The same equality is true for $k > k_w - a + \ell N^{\varepsilon}$ using the same argument.

For $k_w - a - \ell N^{\varepsilon} \leqslant k \leqslant k_w - a + \ell N^{\varepsilon}$, since the operator $U_{\mathscr{S}}$ is bounded in ℓ^∞, we have

$$\left|\mathfrak{U}_{k_w}^a(\varphi_t)(k)\right| \leqslant 2\sup_{j:|j-k_w|\leqslant N^{\varepsilon}\ell}\varphi_t(j) \leqslant N^{\varepsilon}\sigma_t^2(\mathbf{q}, k_w, \eta) \quad (2.3.29)$$

where we used Corollary 2.2.7 combined with (2.3.14). Now, in the integrand in (2.3.28) there is a set of measure at most $N^{\varepsilon}\ell$ which is not exponentially small which gives, combined with Theorem 2.2.2, that we can bound (2.3.25) by

$$(2.3.25) \leqslant N^{\varepsilon}\frac{\ell}{w}\mathrm{Im}\, m_t(z^{(k_0)}))\sigma_t^2(\mathbf{q}, k_w) \leqslant N^{\varepsilon}\frac{\ell}{w}\sigma_t^2(\mathbf{q}, k_w, \eta), \quad (2.3.30)$$

where we used that $\operatorname{Im} m_t(z^{(k_0)})$ is bounded in the spectral window from Lemma 2.2.1.

To bound (2.3.26), noting that $f_\tau = U_{\mathscr{B}}(0,\tau)\varphi_t$,

$$|(\mathrm{Av}_{k_w} U_{\mathscr{S}}(0,\tau)\varphi_t)(k) - (\mathrm{Av}_{k_w} U_{\mathscr{B}}(0,\tau)\varphi_t)(k)| \leqslant |[(U_{\mathscr{S}}(0,\tau) - U_{\mathscr{B}}(0,\tau))\,\varphi_t]\,(k)| \leqslant N^{7\varepsilon}\frac{N\tau}{\ell}\sigma_t^2(\mathbf{q},k,\eta)$$

where we applied Lemma 2.3.3.
Thus we have

$$(2.3.26) \leqslant N^{7\varepsilon}\frac{CN\tau}{\ell}\sigma_t^2(\mathbf{q},k,\eta)\mathrm{Im}\left(\frac{1}{N}\sum_{k=1}^N \frac{1}{\lambda_k - z^{(k_0)}}\right) \leqslant N^{7\varepsilon}\frac{N\tau}{\ell}\sigma_t^2(\mathbf{q},k,\eta) \tag{2.3.31}$$

where we used the averaged local law from Theorem 2.2.2 and the fact that in the spectral window we have $\mathrm{Im} m_t(z^{(k_0)}) \leqslant C$.

To bound (2.3.27), we want to use (2.2.13) which comes from the local law. Recalling that $z^{(k_0)} = \gamma_{k_0} + \mathrm{i}\eta$, then

$$\mathrm{Im}\left(\frac{1}{N}\sum_{k=1}^N \frac{(\mathrm{Av}_{k_w} U_{\mathscr{B}}(0,\tau)\varphi_t)(k)}{\lambda_k - z^{(k_0)}}\right) = \mathrm{Im}\left(\frac{1}{N}\sum_{|k-k_0|\leqslant N\sqrt{\eta}} \frac{(\mathrm{Av}_{k_w} U_{\mathscr{B}}(0,\tau)\varphi_t)(k)}{\lambda_k - z^{(k_0)}}\right) + \mathcal{O}\left(\sqrt{\eta}\sigma_t^2(\mathbf{q},k_0)\right).$$

where we used the fact that similarly to the proof of Lemma 2.3.3, we have that, for any threshold l,

$$\mathrm{Im}\left(\frac{1}{N}\sum_{|k-k_0|>l} \frac{(\mathrm{Av}_{k_w} f_\tau)(k)}{\lambda_k - z^{(k_0)}}\right) = \mathcal{O}\left(N^{\varepsilon}\frac{N\eta}{l}\sigma_t^2(\mathbf{q},k_0)\right) \tag{2.3.32}$$

Taking $l = N\sqrt{\eta}$ gives here the bound. If we use the notation (2.3.18), we obtain

$$\mathrm{Im}\left(\frac{1}{N}\sum_{k=1}^N \frac{(\mathrm{Av}_{k_w} f_\tau)(k)}{\lambda_k - z^{(k_0)}}\right) = \tag{2.3.33}$$

$$= \mathrm{Im}\left(\frac{1}{N}\sum_{|k-k_0|\leqslant N\sqrt{\eta}} \frac{a_k f_\tau(k) + (1-a_k)\sigma_t^2(\mathbf{q},k_w,\eta)}{\lambda_k - z^{(k_0)}}\right) + \mathcal{O}\left(\sqrt{\eta}\sigma_t^2(\mathbf{q},k_0)\right) \tag{2.3.34}$$

$$= a_{k_0}\mathrm{Im}\left(\frac{1}{N}\sum_{k=1}^N \frac{f_\tau(k)}{\lambda_k - z^{(k_0)}}\right) + (1-a_{k_0})\sigma_t^2(\mathbf{q},k_w,\eta)\mathrm{Im}\left(\frac{1}{N}\sum_{k=1}^N \frac{1}{\lambda_k - z^{(k_0)}}\right) \tag{2.3.35}$$

$$+ \mathrm{Im}\left(\frac{1}{N}\sum_{|k-k_0|\leqslant N\sqrt{\eta}}^N \frac{(a_k - a_{k_0})f_\tau(k) + (a_{k_0} - a_k)\sigma_t^2(\mathbf{q},k_w,\eta)}{\lambda_k - z^{(k_0)}}\right) + \mathcal{O}\left(\sqrt{\eta}\sigma_t^2(\mathbf{q},k_0)\right). \tag{2.3.36}$$

Note now that

$$\mathrm{Im}\left(\sum_{k=1}^N \frac{q_k^2}{D_k - z^{(k_0)} - t m_t(z^{(k_0)})}\right) = \left(\mathrm{Im}(m_t(z^{(k_0)})) + \frac{\eta}{t}\right)\sigma_t^2(\mathbf{q},k_0,\eta),$$

so that, by Lemma 2.3.5 and Theorem 2.2.2, we obtain

$$(2.3.35) = a_{k_0}\mathrm{Im}(m_t(z^{(k_0)}))\sigma_t^2(\mathbf{q},k_0) + (1-a_{k_0})\sigma_t^2(\mathbf{q},k_w,\eta)\mathrm{Im}(m_t(z^{(k_0)})) + \mathcal{O}\left(N^{\varepsilon}\left(\frac{1}{\sqrt{N\eta}}+\frac{1}{N\eta}+\frac{\tau}{t}+\sqrt{\eta}\right)\sigma_t^2(\mathbf{q},k_w,\eta)\right)$$

Note that we do not keep the error η/t as we take $\eta \ll t$ so that τ/t is of larger order. Now using the deterministic bound (2.3.14), we obtain that since $k \in [\![k_w - w, k_w + w]\!]$,

$$(2.3.35) = \mathrm{Im}(m_t(z^{(k_0)}))\sigma_t^2(\mathbf{q},k_w,\eta) + \mathcal{O}\left(N^{\varepsilon}\left(\frac{1}{\sqrt{N\eta}}+\frac{\tau}{t}+\frac{w}{Nt}+\sqrt{\eta}\right)\sigma_t^2(\mathbf{q},k_w,\eta)\right)$$

Finally, with the elementary property $|a_i - a_k| \leqslant \frac{C|i-k|}{N}$, we get that $(2.3.36) \leqslant C\sqrt{\eta}\sigma_t^2(\mathbf{q},k_w,\eta)$. Putting these estimates together, we obtain

$$(2.3.27) = \mathcal{O}\left(N^{\varepsilon}\left(\frac{1}{\sqrt{N\eta}}+\frac{\tau}{t}+\frac{w}{Nt}+\sqrt{\eta}\right)\sigma_t^2(\mathbf{q},k_w,\eta)\right) \tag{2.3.37}$$

Combining (2.3.30),(2.3.31) and (2.3.37), we get the final result

$$\left|\mathrm{Im}\left(\frac{1}{N}\sum_{k=1}^{N}\frac{g_\tau(k)}{\lambda_k - z^{(k_0)}}\right) - \mathrm{Im}(m_t(z^{(k_0)}))\sigma_t^2(\mathbf{q},k_w,\eta)\right| = \mathcal{O}\left(N^{\varepsilon}\left(\frac{\ell}{w}+\frac{N\tau}{\ell}+\frac{1}{\sqrt{N\eta}}+\frac{\tau}{t}+\frac{w}{Nt}+\sqrt{\eta}\right)\sigma_t^2(\mathbf{q},k_w,\eta)\right).$$

□

Now, we just need to prove that (2.3.20). Let k_m be the index such that $g_\tau(k_m) = \max_k g_\tau(k)$ and $z = \lambda_{k_m} + i\eta$. If we have that

$$\left|g_\tau(k_m) - \sigma_t^2(\mathbf{q},k_m,\eta)\right| \leqslant N^{-10}, \tag{2.3.38}$$

there is nothing to prove. Now if the left hand side is greater than N^{-10}, by finite speed of propagation, k_m is in the interval $[\![k_w - u, k_w + u]\!]$. Indeed, if it is not then the difference in (2.3.38) would be exponentially small. We then have

$$\begin{aligned}
\partial_\tau g_\tau(k_m) &= \frac{1}{N}\sum_{\substack{|j-k_m|\leqslant \ell\\ j\neq k_m}}\frac{g_\tau(j)-g_\tau(k_m)}{(\lambda_j-\lambda_{k_m})^2}\\
&\leqslant \frac{1}{N\eta}\sum_{\substack{|j-k_m|\leqslant \ell\\ j\neq k_m}}\frac{\eta g_\tau(j)}{(\lambda_j-\lambda_{k_m})^2+\eta^2} - \frac{g_\tau(k_m)}{N\eta}\sum_{\substack{|j-k_m|\leqslant \ell\\ j\neq k_m}}\frac{\eta}{(\lambda_j-\lambda_{k_m})^2+\eta^2}\\
&\leqslant \frac{1}{\eta}\mathrm{Im}\left(\frac{1}{N}\sum_{k=1}^{N}\frac{g_\tau(k)}{\lambda_k - z}\right) - \frac{g_\tau(k_m)}{\eta}\mathrm{Im}(m_t(z)) + \mathcal{O}\left(\frac{N^{\varepsilon}}{\eta}\left(\frac{N\eta}{\ell}+\frac{1}{N\eta}\right)\sigma_t^2(\mathbf{q},k_m,\eta)\right)\\
&\leqslant \frac{C}{\eta}\left(\sigma_t^2(\mathbf{q},k_w,\eta) - g_\tau(k_m)\right) \qquad (2.3.39)\\
&+ \mathcal{O}\left(\frac{N^{\varepsilon}}{\eta}\left(\left(\frac{N\eta}{\ell}+\frac{1}{N\eta}\right)\sigma_t^2\mathbf{q},k_m,\eta) + \left(\frac{\ell}{w}+\frac{N\tau}{\ell}+\frac{1}{\sqrt{N\eta}}+\frac{\tau}{t}+\frac{w}{Nt}+\sqrt{\eta}\right)\sigma_t^2(\mathbf{q},k_w,\eta)\right)\right) \qquad (2.3.40)
\end{aligned}$$

where we used in the first inequality that $g_\tau(k_m)$ is the maximum, in the second inequality that extending the sum to all j adds an error $(N^{1+\varepsilon}\eta/\ell\sigma_t^2\mathbf{q},k_m,\eta))$ using (2.3.32) and Theorem 2.2.2 combined with the estimate from Lemma 2.2.1. Finally in the last inequality we used (2.3.24), $c \leqslant \mathrm{Im}(m_t(z)) \leqslant C$ in the spectral window and that the rigidity errors that appears from changing λ_{k_m} into $\gamma_{k_m,t}$ from Theorem 2.2.6 are smaller than the other terms. Injecting our variance $\sigma_t^2(\mathbf{q},k_w,\eta)$ in (2.3.39) which does not depend of τ,

$$\partial_\tau\left(g_\tau(k_m)-\sigma_t^2(\mathbf{q},k_w,\eta)\right) \leqslant -\frac{C}{\eta}\left(g_\tau(k_m)-\sigma_t^2(\mathbf{q},k_w,\eta)\right) + \mathcal{O}\left(\left(\frac{N\eta}{\ell}+\frac{\ell}{w}+\frac{N\tau}{\ell}+\frac{1}{\sqrt{N\eta}}+\frac{\tau}{t}+\frac{w}{Nt}+\sqrt{\eta}\right)\frac{N^\varepsilon}{\eta}\sigma_t^2(\mathbf{q},k_w,\eta)\right).$$

Thus, writing $S_\tau = g_\tau(k_m)-\sigma_t(\mathbf{q},k_w)^2$ we get

$$\partial_\tau S_\tau \leqslant -\frac{C}{\eta}S_\tau + \mathcal{O}\left(\left(\frac{N\eta}{\ell}+\frac{\ell}{w}+\frac{N\tau}{\ell}+\frac{1}{\sqrt{N\eta}}+\frac{\tau}{t}+\frac{w}{Nt}+\sqrt{\eta}\right)\frac{N^\varepsilon}{\eta}\sigma_t^2(\mathbf{q},k_w,\eta)\right).$$

Note that S_τ is not necessary differentiable as the maximum is not necesarily unique for instance, but one can get the result by instead considering

$$\partial_\tau S_\tau = \limsup_{s\to\tau}\frac{S_s-S_\tau}{s-\tau}.$$

Using Gronwall's lemma, we have if $\eta \ll \tau$,

$$S_\tau = \mathcal{O}\left(N^\varepsilon\left(\frac{N\eta}{\ell}+\frac{\ell}{w}+\frac{N\tau}{\ell}+\frac{1}{\sqrt{N\eta}}+\frac{\tau}{t}+\frac{w}{Nt}+\sqrt{\eta}\right)\sigma_t^2(\mathbf{q},k_w,\eta)+N^{-C}\right)$$

for any C. We can do the same reasoning with the infimum. Finally taking the following parameters

$$\eta = N^{-\varepsilon}\tau, \quad w = \left(N\tau(Nt)^2\right)^{1/3}, \quad \ell = \sqrt{N\tau u} \tag{2.3.41}$$

we get the result for a single particle. Note that with these choices of parameters we have the correct relations: $N^{-1} \ll \eta \ll \tau \ll \ell \ll w \ll Nt$.

Proof of Proposition 2.3.4: Case of n particles

In the previous part of the proof, we looked only at the second moment $\mathbb{E}\left[N\langle\mathbf{q},u_k(t)\rangle^2|\boldsymbol{\lambda}\right]$ which corresponds to a single particle in the site k. Now, we will do the proof of (2.3.15) by induction on the number of particles.

We can first define the same objects as the single particle case: we will consider the short range dynamics for a small time $\tau \ll t$ with initial condition an average of the eigenvectors moment of W_t localized onto a specific window. More precisely define, with $\boldsymbol{\xi}_0$ being the configuration with n particles that lies at the center of our window of size w in the sense of the distance (2.3.6),

$$\partial_\tau g_\tau(\boldsymbol{\xi}) = \frac{1}{N}\sum_{\substack{|i-j|\leqslant\ell\\ i\neq j}}\frac{g_\tau(\boldsymbol{\xi}^{ij})-g_\tau(\boldsymbol{\xi})}{(\lambda_i-\lambda_j)^2}, \tag{2.3.42}$$

$$g_0(\boldsymbol{\xi}) = \left(\mathrm{Av}_{\boldsymbol{\xi_0}}f_t\right)(\boldsymbol{\xi}). \tag{2.3.43}$$

By the same reasoning as for the one particle case, using Lemmas 2.3.3 and 2.3.1 with n particles, we get

$$|f_\tau(\boldsymbol{\xi}) - g_\tau(\boldsymbol{\xi})| \leqslant N^{(n+6)\varepsilon}\frac{CN\tau}{\ell}\sigma_t^2(\mathbf{q},\boldsymbol{\xi},\eta). \tag{2.3.44}$$

To reason by induction on the number of particles, we need to show the following equation, similar to (2.3.24) in the case of one particle. For $k_r \in \mathcal{A}_r^\kappa$, define $z^{(k_r)} = \gamma_{k_r} + i\eta$ and let $\boldsymbol{\xi}$ a configuration of n particles with at least one particle in k_r, we need to show

$$\begin{aligned}\operatorname{Im}\left(\frac{1}{N}\sum_{k=1}^N \frac{g_\tau\left(\boldsymbol{\xi}^{k_r,k}\right)}{\lambda_k(t+\tau)-z^{(k_r)}}\right) - \left(a_{\boldsymbol{\xi}} f_\tau(\boldsymbol{\xi}\setminus k_r)\sigma_t(\mathbf{q},k_r,\eta)^2 + (1-a_{\boldsymbol{\xi}})\sigma_t^2(\mathbf{q},\boldsymbol{\xi}_\mathbf{0})\right) \\ = \mathcal{O}\left(N^{(n+6)\varepsilon}\left(\frac{\ell}{w}+\frac{N\tau}{\ell}+\frac{1}{\sqrt{N\eta}}+\frac{\tau}{t}\right)\sigma_t^2(\mathbf{q},\boldsymbol{\xi}_\mathbf{0})\right).\end{aligned} \tag{2.3.45}$$

where $\boldsymbol{\xi}\setminus k_r$ denote the configuration where we removed one particle in k_r from $\boldsymbol{\xi}$.

We apply the same decomposition in three terms as in the single particle case, the first two terms can be bounded the same way and we can bound the left hand side of (2.3.45) by

$$\begin{aligned}\Bigg|\operatorname{Im}\left(\frac{1}{N}\sum_{k=1}^N \frac{\left(\operatorname{Av}_{\xi_w} f_\tau\right)\left(\boldsymbol{\xi}^{k_r,k}\right)}{\lambda_k - z^{(k_r)}}\right) \\ - \left(a_{\boldsymbol{\xi}} f_t(\boldsymbol{\xi}\setminus k_r)\operatorname{Im}(m_t(z^{(k_r)}))\sigma_t^2(\mathbf{q},k_r,\eta) + (1-a_{\boldsymbol{\xi}})\operatorname{Im}(m_t(z^{(k_r)}))\sigma_t^2(\mathbf{q},\boldsymbol{\xi}_\mathbf{0})\right)\Bigg| \\ + \mathcal{O}\left(N^{(n+6)\varepsilon}\left(\frac{\ell}{u}+\frac{N\tau}{\ell}\right)\sigma_t^2(\mathbf{q},\boldsymbol{\xi}_\mathbf{0})\right)\end{aligned} \tag{2.3.46}$$

Now we need to see that,

$$\begin{aligned}\operatorname{Av}_{\boldsymbol{\xi}_0} f_\tau(\boldsymbol{\xi}^{k_r,k}) &= a_{\boldsymbol{\xi}^{k_r,k}} f_\tau(\boldsymbol{\xi}^{k_r,k}) + (1-a_{\boldsymbol{\xi}^{k_r,k}})\sigma_t^2(\mathbf{q},\boldsymbol{\xi}_\mathbf{0}) \\ &= \left(a_{\boldsymbol{\xi}} f_\tau(\boldsymbol{\xi}^{k_r,k}) + (1-a_{\boldsymbol{\xi}})\sigma_t^2(\mathbf{q},\boldsymbol{\xi}_\mathbf{0})\right) \\ &+ \left((a_{\boldsymbol{\xi}^{k_r,k}} - a_{\boldsymbol{\xi}}) f_\tau(\boldsymbol{\xi}^{k_r,k}) + (a_{\boldsymbol{\xi}} - a_{\boldsymbol{\xi}^{k_r,k}})\sigma_t^2(\mathbf{q},\boldsymbol{\xi}_\mathbf{0})\right) \\ &= \left(a_{\boldsymbol{\xi}} f_\tau(\boldsymbol{\xi}^{k_r,k}) + (1-a_{\boldsymbol{\xi}})\sigma_t^2(\mathbf{q},\boldsymbol{\xi}_\mathbf{0})\right) + \mathcal{O}\left(\frac{d(\boldsymbol{\xi}^{k_r,k},\boldsymbol{\xi})\sigma_t^2(\mathbf{q},\boldsymbol{\xi}_\mathbf{0})}{N}\right).\end{aligned}$$

We can use the same decomposition into $|k-k_r| \leqslant N\sqrt{\eta}$ and the averaged local law from Theorem 2.2.2 to get

$$\begin{aligned}\operatorname{Im}\left(\frac{1}{N}\sum_{k=1}^N \frac{\operatorname{Av}_{\xi_w} f_\tau\left(\boldsymbol{\xi}^{k_r,k}\right)}{\lambda_k - z^{(k_r)}}\right) = a_{\boldsymbol{\xi}}\operatorname{Im}\left(\frac{1}{N}\sum_{k=1}^N \frac{f_\tau(\boldsymbol{\xi}^{k_r,k})}{\lambda_k - z^{(k_r)}}\right) \\ + (1-a_{\boldsymbol{\xi}})\operatorname{Im}(m_t(z^{(k_r)}))\sigma_t^2(\mathbf{q},\boldsymbol{\xi}_\mathbf{0}) + \mathcal{O}\left(\frac{N^\varepsilon}{N\eta}\sigma_t^2(\mathbf{q},\boldsymbol{\xi}_\mathbf{0})\right).\end{aligned} \tag{2.3.47}$$

Consider the sum in the right hand side, recall that there is at least one particle in k_r and denote $k_1,\dots,k_m$ with $m \leqslant n$, the sites where there is at least one particle in the configuration $\boldsymbol{\xi}$. Recall that $z^{(k_r)} = \gamma_{k_r,t} + i\eta$,

$$\frac{1}{N}\sum_{k=1}^N \frac{\eta f_\tau(\boldsymbol{\xi}^{k_r,k})}{(\gamma_{k_r,t}-\lambda_k)^2+\eta^2} = \frac{1}{N}\sum_{k\notin\{k_1,\dots k_m\}} \frac{\eta f_\tau(\boldsymbol{\xi}^{k_r,k})}{(\gamma_{k_r,t}-\lambda_k)^2+\eta^2} + \mathcal{O}\left(\frac{N^{n\varepsilon}}{N\eta}\sum_{i=1}^m \sigma_t^2(\mathbf{q},\boldsymbol{\xi}^{k_r,k_i},\eta)\right) \tag{2.3.48}$$

where we used Corollary 2.2.7 on the indices we removed from the sum. Now we have the following equality for the first sum by definition of f_τ,

$$\frac{1}{N}\sum_{k\notin\{k_1,\dots k_m\}}\frac{\eta f_\tau(\boldsymbol{\xi}^{k_r,k})}{(\gamma_{k_r,t}-\lambda_k)^2+\eta^2}=\mathbb{E}\left[\prod_{\substack{1\leqslant r'\leqslant m\\ r'\neq r}}\frac{z_{r'}^{2j_{r'}}}{a(2j_{r'})}\frac{z_r^{2(j_r-1)}}{a(2(j_r-1))}\sum_{j\notin\{k_1,\dots,k_m\}}\frac{\eta z_j^2}{N((\gamma_{k_r,t}-\lambda_j)^2+\eta^2)}\middle|\boldsymbol{\lambda}\right] \tag{2.3.49}$$

By Lemma 2.3.5, we have,

$$\frac{1}{N}\sum_{j\notin\{k_1,\dots,k_m\}}\frac{\eta z_j^2}{(\gamma_{k_r,t}-\lambda_j)^2+\eta^2}=\frac{1}{N}\sum_{k=1}^N\frac{\eta z_k^2}{(\gamma_{k_r,t}-\lambda_k)^2+\eta^2}+\mathcal{O}\left(\frac{N^\varepsilon}{N\eta}\sum_{j=1}^m\sigma_t^2(\mathbf{q},k_j,\eta)\right), \tag{2.3.50}$$

$$=\operatorname{Im}\left(\frac{1}{N}\sum_{k=1}^N\frac{z_k^2}{\lambda_k-z^{(k_r)}}\right)+\mathcal{O}\left(\frac{N^\varepsilon}{N\eta}\sum_{j=1}^m\sigma_t^2(\mathbf{q},k_j,\eta)\right), \tag{2.3.51}$$

$$=\operatorname{Im}(m_t(z^{(k_r)}))\sigma_t^2(\mathbf{q},k_r)+\mathcal{O}\left(N^\varepsilon\left(\frac{1}{\sqrt{N\eta}}+\frac{\tau}{t}\right)\sigma_t^2(\mathbf{q},k_r)+\frac{N^\varepsilon}{N\eta}\sum_{j=1}^m\sigma_t^2(\mathbf{q},k_j,\eta)\right). \tag{2.3.52}$$

Combining (2.3.52) and (2.3.49) and using the bounds on the variations of σ_t (2.3.14), we get

$$\frac{1}{N}\sum_{k=1}^N\frac{\eta f_\tau(\boldsymbol{\xi}^{k_r,k})}{(\gamma_{k_r,t}-\lambda_k)^2+\eta^2}=\operatorname{Im}(m_t(z^{(k_r)}))\sigma_t(\mathbf{q},k_r,\eta)^2 f_\tau(\boldsymbol{\xi}\setminus k_r)+\mathcal{O}\left(N^\varepsilon\left(\frac{1}{\sqrt{N\eta}}+\frac{\tau}{t}\right)\sigma_t^2(\mathbf{q},\boldsymbol{\xi},\eta)\right) \tag{2.3.53}$$

Finally, combining (2.3.53) and (2.3.47), we have

$$\operatorname{Im}\left(\frac{1}{N}\sum_{k=1}^N\frac{\operatorname{Av}_{\xi_0}f_\tau\left(\boldsymbol{\xi}^{k_r,k}\right)}{\lambda_k-z^{(k_r)}}\right)=a_{\boldsymbol{\xi}}f_t(\boldsymbol{\xi}\setminus k_r)\operatorname{Im}(m_t(z^{(k_r)}))\sigma_t^2(\mathbf{q},k_r,\eta)$$
$$+(1-a_{\boldsymbol{\xi}})\operatorname{Im}(m_t(z^{(k_r)}))\sigma_t^2(\mathbf{q},\boldsymbol{\xi_0})+\mathcal{O}\left(N^\varepsilon\left(\frac{1}{\sqrt{N\eta}}+\frac{\tau}{t}\right)\sigma_t^2(\mathbf{q},\boldsymbol{\xi},\eta)\right)$$

which, combined with (2.3.46), gives us (2.3.45).

We now follow the same proof as in the case of one particle : we state a maximum principle on the flattened and averaged moment. First define

$$\boldsymbol{\xi_m}=\max_{\substack{\boldsymbol{\xi}\\ \mathcal{N}(\boldsymbol{\xi})=n}}g_\tau(\boldsymbol{\xi}), \tag{2.3.54}$$

and let $k_1,\dots,k_m$ be the positions of the particles of the configuration $\boldsymbol{\xi_m}$ with $m\leqslant n$. We are going to use our induction hypothesis in the maximum principle inequalities by (2.3.45).

$$
\begin{aligned}
\partial_\tau g_\tau(\boldsymbol{\xi_m}) &\leqslant \frac{C}{N} \sum_{\substack{|i-j|\leqslant l \\ i\neq j}} \frac{g_\tau(\boldsymbol{\xi_m}^{i,j}) - g_\tau(\boldsymbol{\xi_m})}{(\lambda_i-\lambda_j)^2} \\
&\leqslant \frac{C}{N}\sum_{r=1}^m \left(\frac{1}{\eta} \sum_{\substack{|j-k_r|\leqslant l \\ j\neq k_r}} \frac{\eta g_\tau(\boldsymbol{\xi_m}^{k_r,j})}{(\lambda_j-\lambda_{k_r})^2+\eta^2} - \frac{g_\tau(\boldsymbol{\xi_m})}{\eta} \sum_{\substack{|j-k_r|\leqslant l \\ j\neq k_r}} \frac{\eta}{(\lambda_j-\lambda_{k_r})^2+\eta^2} \right) \\
&\leqslant \frac{C}{\eta}\sum_{r=1}^m \left[\operatorname{Im}\left(\frac{1}{N}\sum_{j=1}^N \frac{g_\tau(\boldsymbol{\xi_m}^{k_r,j})}{\lambda_j - z^{(k_r)}} \right) - g_\tau(\boldsymbol{\xi_m}) \operatorname{Im}\left(s_\tau(z^{(k_r)}) \right) \right] + \mathcal{O}\left(\frac{N\eta}{\ell} \frac{\sigma_t^2(\mathbf{q},\boldsymbol{\xi_m},\eta)}{\eta} \right) \\
&\leqslant \frac{C}{\eta}\sum_{r=1}^m \left(a_{\boldsymbol{\xi_m}} f_\tau(\boldsymbol{\xi_m}\setminus k_r) \operatorname{Im}(m_t(z^{(k_r)}))\sigma_t(\mathbf{q},k_r)^2 + (1-a_{\boldsymbol{\xi_m}})\operatorname{Im}(m_t(z^{(k_r)}))\sigma_t^2(\mathbf{q},\boldsymbol{\xi_0}) \right) \qquad (2.3.55) \\
&- \frac{C}{\eta}\sum_{r=1}^m g_\tau(\boldsymbol{\xi_m})\operatorname{Im}(m_t(z^{(k_r)})) + \mathcal{O}\left(\frac{N^{(n+6)\varepsilon}}{\eta} \left(\frac{1}{\sqrt{N\eta}} + \frac{\tau}{t} + \frac{\ell}{w} + \frac{N\tau}{\ell} + \frac{N\eta}{\ell} \right) \sigma_t^2(\mathbf{q},\boldsymbol{\xi_m},\eta) \right).
\end{aligned}
$$

Now, we use the induction assumption on $f_\tau(\boldsymbol{\xi_m}\setminus k_r)$ which is a $(n-1)$th moment and obtain

$$
f_\tau(\boldsymbol{\xi_m}\setminus k_r)\sigma_t(\mathbf{q},k_r,\eta)^2 = \sigma_t^2(\mathbf{q},\boldsymbol{\xi_m}) + \mathcal{O}\left(\left(\frac{1}{\sqrt{N\tau}} + \left(\frac{\tau}{t}\right)^{1/3} \right) \sigma_t^2(\mathbf{q},\boldsymbol{\xi_m},\eta) \right). \qquad (2.3.56)
$$

Besides, we can easily see that, since $d(\boldsymbol{\xi_w},\boldsymbol{\xi_m}) \leqslant 2w$, from (2.3.14),

$$
\left| \sigma_t^2(\mathbf{q},\boldsymbol{\xi_0})^2 - \sigma_t^2(\mathbf{q},\boldsymbol{\xi_m})^2 \right| \leqslant \frac{N^\varepsilon w}{Nt} \sigma_t^2(\mathbf{q},\boldsymbol{\xi_m},\eta). \qquad (2.3.57)
$$

Now, injecting (2.3.56) and (2.3.57) in (2.3.55), we get

$$
\begin{aligned}
\partial_\tau \left(g_\tau(\boldsymbol{\xi_m}) - \sigma_t^2(\mathbf{q},\boldsymbol{\xi_0},\eta) \right) &\leqslant -\frac{C}{\eta}\sum_{r=1}^m \left(g_\tau(\boldsymbol{\xi_m}) - \sigma_t^2(\mathbf{q},\boldsymbol{\xi_0},\eta) \right) \\
&+ \mathcal{O}\left(\left(\frac{N\tau}{\ell} + \frac{\ell}{w} + \frac{1}{\sqrt{N\eta}} + \frac{\tau}{t} + \frac{N\eta}{\ell} + \frac{w}{Nt} + \frac{1}{\sqrt{N\tau}} + \left(\frac{\tau}{t}\right)^{1/3} \right) \frac{N^{(n+6)\varepsilon}}{\eta} \sigma_t^2(\mathbf{q},\boldsymbol{\xi_m},\eta) \right).
\end{aligned}
$$

Doing the same reasoning as in the proof for one particle, we get, by applying Gronwall's lemma,

$$
\begin{aligned}
g_\tau(\boldsymbol{\xi_m}) &= \sigma_t^2(\mathbf{q},\boldsymbol{\xi_m},\eta) \\
&+ \mathcal{O}\left(\left(\frac{N\tau}{\ell} + \frac{\ell}{w} + \frac{1}{\sqrt{N\eta}} + \frac{\tau}{t} + \frac{N\eta}{\ell} + \frac{w}{Nt} + \left(\frac{1}{\sqrt{N\tau}} + \left(\frac{\tau}{t}\right)^{1/3} \right) \right) \sigma_t^2(\mathbf{q},\boldsymbol{\xi_m},\eta) \right)
\end{aligned}
$$

We can again do the same reasoning with the infimum and choosing the parameters as in (2.3.41) the claim from Proposition 2.3.4 follows.

2.3.2. Analysis of the perfect matching observable

In this subsection we will again condition on the event A of good eigenvalue paths where the local laws and the finite speed of propagation holds. Consider now a deterministic set of indices $I \subset [\![1,N]\!]$. Note that in the definition of the centered overlaps p_{ii}, we can only center by a constant not depending on i. However in Theorem 2.1.7, one can see that the expectation of the probability mass of the ith-eigenvector on I clearly depends on i. Thus, we will need to localize our perfect matching observables onto a window of size w chosen later and show that these p_{ii} are, up to an error depending on w, centered around the same constant. The size of the window w will be taken so that $N\tau \ll w \ll Nt$ similarly to the previous section. More precisely, we will fix an integer $i_0 \in \mathcal{A}_r^\kappa$ and consider the set of indices

$$\mathcal{A}_w^\kappa(i_0) = \{i \in [\![1, N]\!],\, \gamma_{t,i} \in [\gamma_{t,i_0} - (1-\kappa)w, \gamma_{t,i_0} + (1-\kappa)w]\}$$

so that we will take for our centered diagonal overlaps

$$p_{ii} = \sum_{\alpha \in I} u_i(\alpha)^2 - C_0 \quad \text{with} \quad C_0 = \frac{1}{N}\sum_{\alpha\in I}\sigma_t^2(\mathbf{e}_\alpha, i_0), \tag{2.3.58}$$

the overlaps for $i \neq j$ will not change. First consider these overlaps for the matrix W_t and define, similarly to the previous subsection

$$\Phi_t(\boldsymbol{\xi}) = \frac{1}{\mathcal{M}(\boldsymbol{\xi})}\mathbb{E}\left[\sum_{G\in\mathcal{G}_{\boldsymbol{\eta}}} P(G)\middle|\boldsymbol{\lambda}\right].$$

This quantity corresponds the perfect matching observable for our initial matrix W_t and we make it undergo the dynamics (2.1.28) so that we define

$$F_s(\boldsymbol{\xi}) = U_{\mathscr{B}}(0, s)\Phi_t(\boldsymbol{\xi})$$

where $\mathscr{B}$ is defined in (2.3.1). We now prove the result from Theorem 2.1.7 for a Gaussian divisible ensemble for $p_{i_0 i_0}$. We will need the following technical lemma allowing us to bound the p_{ij} by the perfect matching observables.

Lemma 2.3.7 ([BYY18, Lemma 3.6])**.** *Take an even integer n, there exists $C > 0$ depending on n such that for any $i < j$ and any time s we have*

$$\mathbb{E}\left[p_{ij}(s)^n|\boldsymbol{\lambda}\right] \leqslant C\left(F_s(\boldsymbol{\eta}^{(1)}) + F_s(\boldsymbol{\eta}^{(2)}) + F_s(\boldsymbol{\eta}^{(3)})\right) \tag{2.3.59}$$

where $\boldsymbol{\eta}^{(1)}$ is the configuration of n particles in the site i, $\boldsymbol{\eta}^{(2)}$ n particles in the site j, and $\boldsymbol{\eta}^{(3)}$ an equal number of particles between the site i and j.

We will also use repeatedly the following bound on the eigenvectors which comes from Corollary 2.2.7

$$\sum_{k\in I} u_k(\alpha)^2 \leqslant N^\varepsilon \widehat{I}.$$

The purpose of this section is to prove Theorem 2.1.7 for another matrix ensemble: a deformed Wigner matrix perturbed by a small Gaussian component. More precisely, we state it as the following theorem.

Theorem 2.3.8. *Consider $\mathfrak{a}$ and ω two small positive constants and $\kappa \in (0, 1)$, take $t \in \mathcal{T}_\omega$, D a deterministic diagonal matrix given by Definition 2.1.1 and W a Wigner matrix given by Definition 2.1.2. Let $\tau \in \mathcal{T}'_{\mathfrak{a}} := [N^{-1+\mathfrak{a}}, N^{-\mathfrak{a}}t]$, then if $u_1, \dots, u_N$ are the eigenvectors of the matrix*

$$H_\tau = D + \sqrt{t}W + \sqrt{\tau}\mathrm{GOE},$$

define the error

$$\Xi(\tau) = \frac{\widehat{I}}{\sqrt{N\tau}} + \widehat{I}\frac{\tau}{t},$$

we have, for any $k,\ell\in\mathcal{A}_r^\kappa$ with $k\neq\ell$ and any $\varepsilon>0$ and $D>0$,

$$\mathbb{P}\left(\left|\sum_{\alpha\in I}\left(u_k(\alpha)^2-\frac{1}{N}\sigma_t^2(\alpha,k,\tau)\right)\right|+\left|\sum_{\alpha\in I}u_k(\alpha)u_\ell(\alpha)\right|\geqslant N^\varepsilon\Xi(\tau)\right)\leqslant N^{-D}.$$

As one can see in the statement of Theorem 2.3.8, we will need the small time τ corresponding to the size of the Gaussian perturbation to be of order smaller than t so that the eigenvalues and eigenvectors barely changed during that time. We will later choose a specific τ and optimize all our different parameters when using the reverse heat flow technique to remove this small Gaussian component. One of these parameters will be a cut-off for the dynamics as in Subsection 2.3.1. Indeed, in order to use a maximum principle on the dynamics, we will split it in the same way: a short-range dynamics with generator $\mathscr{S}$ that will contain most of the information and a long-range part with generator $\mathscr{L}$ we need to control as in [EY15] where $\mathscr{S}$ and $\mathscr{L}$ are defined respectively in (2.3.3) and (2.3.4). Lemma 2.3.1 will help us localize the dynamics onto a small set of configurations. Now the following lemma says that most of the information of the dynamics is given by the short-range, bounding the difference between $\mathscr{B}$ and $\mathscr{S}$. It is analogous to Lemma 3.5 in [BYY18]. First define

$$S_I^{(u,v)}=\sup_{\boldsymbol{\eta}\subset I,u\leqslant s\leqslant v}F_s(\boldsymbol{\eta}).$$

Lemma 2.3.9. *For any intervals $J_{\rm in}\subset\mathcal{A}_w(i_0)$ and $J_{\rm out}=\{i,\,d(i,J_{\rm in})\leqslant N^\varepsilon\ell\}\subset\mathcal{A}_w^\kappa(i_0)$ since we will take $\ell\ll Nw$, any configuration $\boldsymbol{\xi}$ such that $\mathcal{N}(\boldsymbol{\xi})=n$ supported on $J_{\rm in}$ and any $N^{-1}\ll\tau\ll w$ we have*

$$|((U_{\mathscr{B}}(0,\tau)-U_{\mathscr{S}}(0,\tau))\Phi_t(\boldsymbol{\xi})|\leqslant N^\varepsilon\frac{N\tau}{\ell}\left(S_{J_{\rm out}}^{(0,\tau)}+\widehat{I}\frac{w}{t}\left(S_{J_{\rm out}}^{(0,\tau)}\right)^{\frac{n-1}{n}}+\frac{\widehat{I}}{\ell}\left(S_{J_{\rm out}}^{(0,\tau)}\right)^{\frac{n-2}{n}}\right)\tag{2.3.60}$$

where F_τ is the perfect matching observable defined in 2.1.27.

This bound is used in this form so we can obtain information on a box in space by extracting information from a larger box. Iterating this bound will give us Theorem 2.3.8.

Proof. We will follow the proof from [BYY18]. Define the following flattening operator. For f a function on configurations of n particles and $\boldsymbol{\eta}$ such a configuration,

$$(\mathrm{Flat}_a f)(\boldsymbol{\eta})=\begin{cases}f(\boldsymbol{\eta}) & \text{if}\quad \boldsymbol{\eta}\subset\{i,\,d(i,J_{\rm in})\leqslant a\},\\ 0 & \text{otherwise.}\end{cases}\tag{2.3.61}$$

We make the functions vanish outside of a certain interval. We use now Duhamel's formula and write

$$\left((U_{\mathscr{S}}(0,\tau)-U_{\mathscr{B}}(0,\tau)\Phi_t\right)(\boldsymbol{\xi})=\int_0^\tau U_{\mathscr{S}}(s,\tau)\mathscr{L}(s)F_s(\boldsymbol{\xi})\mathrm{d}s.$$

Now, see that, by definition of the flattening operator and the fact that $\boldsymbol{\xi}$ is supported on $J_{\rm in}$,

$$d(\mathrm{Supp}\left(\mathscr{L}(s)F_s-\mathrm{Flat}_{N^\varepsilon\ell}(\mathscr{L}(s)F_s)\right),\boldsymbol{\xi})\geqslant N^\varepsilon\ell.$$

With this bound, we can use the finite speed of propagation from Lemma 2.3.1 and obtain, using that $U_{\mathscr{S}}$ is a contraction in ℓ^∞,

$$|U_{\mathscr{S}}(s,\tau)\mathscr{L}(s)F_s(\boldsymbol{\xi})|\leqslant\max_{\tilde{\boldsymbol{\eta}}\subset J_{\rm out}}|\left(\mathrm{Flat}_{N^\varepsilon\ell}(\mathscr{L}(s)F_s)\right)(\tilde{\boldsymbol{\eta}})|+\mathcal{O}\left(e^{-cN^\varepsilon/2}\right).$$

Thus we need to control $|(\mathscr{L}(s)F_s)(\tilde{\boldsymbol{\eta}})|$ for $\tilde{\boldsymbol{\eta}} = \{(i_1, j_1), \dots (i_m, j_m)\}$ a configuration of $n \geqslant m$ particles supported in $J_{\rm out}$.

We have, by definition of $\mathscr{L}$,

$$|\mathscr{L}(s)F_s(\tilde{\boldsymbol{\eta}})| \leqslant C_n \left| \sum_{1\leqslant p\leqslant m} \sum_{|i_p-k|\geqslant \ell} \frac{F_s(\tilde{\boldsymbol{\eta}}^{i_p,k})}{N(\lambda_{i_p}-\lambda_k)^2} \right| + C_n\, |F_s(\tilde{\boldsymbol{\eta}})| \sum_{1\leqslant p\leqslant m} \sum_{|i_p-k|>\ell} \frac{1}{N(\lambda_{i_p}-\lambda_k)^2}.$$

For the second term in the previous inequality, we can use rigidity estimates from Theorem 2.2.6 and a dyadic decomposition and see that

$$\sum_{k,\,|i_p-k|>\ell} \frac{1}{N(\lambda_{i_p}-\lambda_k)^2} \leqslant N^{\varepsilon}\frac{N}{\ell}, \tag{2.3.62}$$

so that we have the bound

$$|\mathscr{L}(s)F_s(\tilde{\boldsymbol{\eta}})| \leqslant N^{\varepsilon} \left| \sum_{1\leqslant p\leqslant m} \sum_{k,|i_p-k|\geqslant \ell} \frac{F_s(\tilde{\boldsymbol{\eta}}^{i_p,k})}{N(\lambda_{i_p}-\lambda_k)^2} \right| + N^{\varepsilon}\frac{N}{\ell} S^{(0,\tau)}_{J_{\rm out}}. \tag{2.3.63}$$

For the first sum in (2.3.63), we will first restrict it to the sites k such that there are no particles in the configuration $\tilde{\boldsymbol{\eta}}$, so that we will have $\tilde{\eta}_k^{i_p,k} = 1$. Note that, if we have $\tilde{\eta}_k \neq 0$ then by definition of $\tilde{\boldsymbol{\eta}}$ supported on $J_{\rm out}$, $\tilde{\boldsymbol{\eta}}^{i_p,k}$ is also supported on $J_{\rm out}$. This gives us the bound

$$\begin{aligned}\sum_{k,\,|i_p-k|>\ell} \frac{F_s(\tilde{\boldsymbol{\eta}}^{i_p,k})}{N(\lambda_{i_p}-\lambda_k)^2} &\leqslant \sum_{\substack{k,\,|i_p-k|>\ell\\ \tilde{\eta}_k=0}} \frac{F_s(\tilde{\boldsymbol{\eta}}^{i_p,k})}{N(\lambda_{i_p-}\lambda_k)^2} + S^{(0,\tau)}_{J_{\rm out}} \sum_{\substack{k,\,|i_p-k|>\ell\\ \tilde{\eta}_k\neq 0}} \frac{1}{N(\lambda_{i_p}-\lambda_k)^2}.\\ &\leqslant \sum_{\substack{k,\,|i_p-k|>\ell\\ \tilde{\eta}_k=0}} \frac{F_s(\tilde{\boldsymbol{\eta}}^{i_p,k})}{N(\lambda_{i_p}-\lambda_k)^2} + CN^{\varepsilon}\frac{N}{\ell^2} S^{(0,\tau)}_{J_{\rm out}}\end{aligned}$$

where in the second inequality we used the rigidity of the eigenvalues, which gives us $(\lambda_{i_p}-\lambda_k)^2 \geqslant C(\ell/N)^2$ for $|i_p - k| > \ell$, and the fact that there is at most m sites k such that $\tilde{\eta}_k \neq 0$. By definition of the perfect matching observables, we can write

$$\sum_{\substack{k,\,|i_p-k|>\ell\\ \tilde{\eta}_k=0}} \frac{F_s(\tilde{\boldsymbol{\eta}}^{i_p,k})}{N(\lambda_{i_p}-\lambda_k)^2} = C(n)\mathbb{E}\left[\sum_{\substack{k,\,|i_p-k|>\ell\\ \tilde{\eta}_k=0}} \sum_{G\in\mathcal{G}_{\tilde{\boldsymbol{\eta}}^{i_p,k}}} \frac{\prod_{e\in\mathcal{E}(G)} p(e)}{N(\lambda_{i_p}-\lambda_k)^2} \middle| \boldsymbol{\lambda} \right].$$

In order to control this term, we will consider two different types of perfect matchings. Define the following partition of $\mathcal{G}_{\boldsymbol{\eta}}$ into two subsets

$$\mathcal{G}^{(1)}_{\boldsymbol{\eta}} = \{G\in\mathcal{G}_{\boldsymbol{\eta}},\ \{(k,1),(k,2)\}\in\mathcal{E}(G)\}\,, \tag{2.3.64}$$

$$\mathcal{G}^{(2)}_{\boldsymbol{\eta}} = \{G\in\mathcal{G}_{\boldsymbol{\eta}},\ \{(k,1),(k,2)\}\notin\mathcal{E}(G)\}\,. \tag{2.3.65}$$

(a) A perfect matching from $\mathcal{G}^{(1)}_{\boldsymbol{\eta}^{i_2,k}}$ (b) A perfect matching from $\mathcal{G}^{(2)}_{\boldsymbol{\eta}^{i_2,k}}$

We will begin by bounding the contribution from (2.3.64). Note first that, for $G \in \mathcal{G}_{\boldsymbol{\eta}}^{(1)}$, we have

$$\prod_{e\in\mathcal{E}(G)} p(e) = p_{kk} \times Q_1((p(e))_{e\in\mathcal{E}(G)}) \tag{2.3.66}$$

with

$$Q_1((p(e))_{e\in\mathcal{E}(G)}) = \prod_{e\in\mathcal{E}(G)\setminus\{(k,1),(k,2)\}} p(e).$$

See also that Q_1 is a monic monomial of degree $n-1$ so that we can use Lemma 2.3.7 and obtain

$$\mathbb{E}\left[\sum_{G\in\mathcal{G}_{\boldsymbol{\eta}\setminus i_p}} Q_1((p(e))_{e\in\mathcal{E}(G)})\middle|\boldsymbol{\lambda}\right] \leqslant C \sup_{\substack{0\leqslant s\leqslant\tau\\ \boldsymbol{\eta}\subset J_{\mathrm{out}},\mathcal{N}(\boldsymbol{\eta})=n-1}} |F_s(\boldsymbol{\eta})| \leqslant C\left(S_{J_{\mathrm{out}}}^{(0,\tau)}\right)^{\frac{n-1}{n}}. \tag{2.3.67}$$

Combining (2.3.66) and (2.3.67), we now only need to bound

$$\sum_{\substack{k,\,|k-i_p|>l,\\ \tilde{\eta}_k=0}} \frac{p_{kk}}{N(\lambda_{i_p}-\lambda_k)^2} = \sum_{k,\,|k-i_p|>\ell} \frac{p_{kk}}{N(\lambda_{i_p}-\lambda_k)^2} + \mathcal{O}\left(\frac{N}{\ell^2}\right).$$

In order to bound the sum from the right hand side of the previous equation, first define the following functions, for $|z-\lambda_{i_p}| \leqslant N^{-\varepsilon}\ell/N$,

$$f(z) = \sum_{k,\,\gamma_{t,k}\notin[E_1^-,E_2^+]} \frac{p_{kk}}{N(z-\lambda_k)},$$
$$g(z) = \sum_{k,\,\gamma_{t,k}\notin[E_1^-,E_1^+]\cup[E_2^-,E_2^+]} \frac{p_{kk}}{N(z-\lambda_k)}$$

where $E_1 = \gamma_{t,i_p-\ell}$, $E_1^- = \gamma_{t,i_p-\ell-N^\varepsilon}$, $E_1^+ = \gamma_{t,i_p-\ell+N^\varepsilon}$, $E_2 = \gamma_{t,i_p+\ell}$, $E_2^- = \gamma_{t,i_p+\ell-N^\varepsilon}$, $E_2^+ = \gamma_{t,i_p+\ell+N^\varepsilon}$. Let also Γ be the rectangle with vertices $E_1 \pm \mathrm{i}\ell/N$ and $E_2 \pm \mathrm{i}\ell/N$. We therefore want to bound, up to a N^ε term,

$$\sum_{k,\,|k-i_p|>\ell} \frac{p_{kk}}{N(\lambda_{i_p}-\lambda_k)^2} = \partial_z f(z)\Big|_{z=\lambda_{i_p}}$$

Now consider $\mathcal{C}_{i_p}$ the circle centered in λ_{i_p} with radius $N^{-\varepsilon}\frac{\ell}{N}$, then by Cauchy's formula, we can write,

$$\partial_z f(\lambda_{i_p}) = \frac{1}{2\mathrm{i}\pi}\int_{\mathcal{C}_{i_p}} \frac{f(z)}{(z-\lambda_{i_p})^2}\mathrm{d}\xi.$$

By using another Cauchy integral formula on the contour Γ for f and seeing that for λ_{int}, z inside the contour and λ_{ext} outside of the contour we have, by a residue calculus,

$$\int_\Gamma \frac{\mathrm{d}\xi}{(\xi-\lambda_{\mathrm{int}})(\xi-\lambda_{i_p})} = \int_\Gamma \frac{\mathrm{d}\xi}{(\overline{\xi}-\lambda_{\mathrm{ext}})(\xi-\lambda_{i_p})} = 0$$

we can write

$$\left|\partial_z f(\lambda_{i_p})\right| = \left|\frac{1}{2\pi}\int_\Gamma \frac{g(\xi)}{(\xi-\lambda_{i_p})^2}\mathrm{d}\xi\right| = \mathcal{O}\left(\frac{N}{\ell}\int_\Gamma\left|\frac{\mathrm{Im} g(\xi)}{\xi-\lambda_{i_p}}\right|\mathrm{d}\xi\right). \tag{2.3.68}$$

We will first control the part of the contour closest to the real axis. Consider

$$\Gamma_1 = \{z = E+\mathrm{i}\eta \in \Gamma,\, |\eta| < N^\varepsilon/N\},$$

as in [BYY18] and bounding p_{kk} by $\widehat{I}$, we obtain

$$\left|\int_{\Gamma_1}\frac{\mathrm{Im} g(\xi)}{\xi-\lambda_{i_p}}\right| = \mathcal{O}\left(N^{\varepsilon}\frac{\widehat{I}}{\ell}\right).$$

Now for the rest of the contour, note that we can add the missing eigenvalues to the total sum in g up to adding an error of order $N^{\varepsilon}\widehat{I}/\ell$. Finally we just have to bound

$$\frac{N}{\ell}\int_{\Gamma\backslash\Gamma_1}\left|\mathrm{Im}\sum_{k=1}^{N}\frac{p_{kk}}{N(\xi-\lambda_k)}\right||\mathrm{d}\xi| = \frac{N}{\ell}\int_{\Gamma\backslash\Gamma_1}\left|\mathrm{Im}\frac{1}{N}\sum_{\alpha\in I}G_{\alpha\alpha}(\xi)-\frac{\mathrm{Im} m_t(\xi)}{\mathrm{Im} m_t(z_0)}\frac{1}{N}\sum_{\alpha\in I}\mathrm{Im} g_\alpha(t,z_0)\right||\mathrm{d}\xi| \tag{2.3.69}$$

where we used the definition of p_{kk} and of σ_t and defined $z_0=\gamma_{t,i_0}+\mathrm{i}\eta_0$, with $\eta_0\ll t$ is the center of our window of size $w\ll t$ with positive imaginary part. Now, using the entrywise local law from Theorem 2.2.3 and expanding between z_0 and ξ since $|\xi-z_0|\leqslant w$, we have

$$\begin{aligned}
&\left|\mathrm{Im}\frac{1}{N}\sum_{\alpha\in I}G_{\alpha\alpha}(\xi)-\frac{\mathrm{Im} m_t(\xi)}{\mathrm{Im} m_t(z_0)}\frac{1}{N}\sum_{\alpha\in I}\mathrm{Im} g_\alpha(t,z_0)\right| \\
&\qquad\leqslant\left|\mathrm{Im}\frac{1}{N}\sum_{\alpha\in I}(G_{\alpha\alpha}(\xi)-g_\alpha(t,\xi))\right|+\left|\frac{1}{N}\sum_{\alpha\in I}\left(\mathrm{Im} g_\alpha(t,\xi)-\frac{\mathrm{Im} m_t(\xi)}{\mathrm{Im} m_t(z_0)}\mathrm{Im} g_\alpha(t,z_0)\right)\right| \\
&\qquad\leqslant N^{\varepsilon}\frac{\widehat{I}}{\sqrt{N|\mathrm{Im}\xi|}}+N^{\varepsilon}\widehat{I}\frac{w}{t}.
\end{aligned}$$

Injecting this bound in the contour integral, we get the following bound.

$$(2.3.69)\leqslant N^{\varepsilon}\widehat{I}\frac{w}{t}+N^{\varepsilon}\frac{\widehat{I}}{\sqrt{N}}.$$

Finally, putting all the contributions together and coming back to (2.3.68), we obtain

$$|\partial_z f(\lambda_{i_p})|\leqslant N^{\varepsilon}\frac{N}{\ell}\left(\frac{\widehat{I}}{\ell}+\widehat{I}\frac{w}{t}\right). \tag{2.3.70}$$

Consider now the contribution from (2.3.65), first see that for $G\in\mathcal{G}^{(2)}_{\boldsymbol{\eta}}$, there exists q_1 and q_2 in $\{1,\dots m\}$, such that

$$\prod_{e\in\mathcal{E}(G)}p(e)=p_{ki_{q_1}}p_{ki_{q_2}}\times Q_2((p(e))_{e\in\mathcal{E}(G)})$$

with Q_2 a monic monomials of degree $n-2$. Then using Lemma 2.3.7, we can bound

$$\mathbb{E}\left[\sum_{G\in\mathcal{G}'}Q_2((p(e))_{e\in\mathcal{E}(G)})\middle|\boldsymbol{\lambda}\right]=\mathcal{O}\left(\left(S^{(0,\tau)}_{J_{out}}\right)^{\frac{(n-2)}{n}}\right).$$

Besides, we can bound the term with the cross-edges in the following way

$$\sum_{\substack{k,\,|k-i_p|>\ell\\ \tilde{\eta}_k=0}}\frac{p_{i_{q_1}k}p_{i_{q_2}}}{N(\lambda_{i_p}-\lambda_k)^2}\leqslant\frac{N}{\ell^2}\sum_{k=1}^{N}(p^2_{i_{q_1}k}+p^2_{i_{q_2}k})=\frac{N}{\ell^2}\sum_{\alpha\in I}\left(u_{q_1}(\alpha)^2+u_{q_2}(\alpha)^2\right)\leqslant\widehat{I}\frac{N}{\ell^2}.$$

Putting everything together, we get

$$|((U_{\mathscr{B}}(0,\tau)-U_{\mathscr{S}}(0,\tau))\Phi_t(\boldsymbol{\xi})| \leqslant N^{\varepsilon}\frac{N\tau}{\ell}\left(S_{J_{\mathrm{out}}}^{(0,\tau)}+\widehat{I}\frac{w}{t}\left(S_{J_{\mathrm{out}}}^{(0,\tau)}\right)^{\frac{n-1}{n}}+\frac{\widehat{I}}{\ell}\left(S_{J_{\mathrm{out}}}^{(0,\tau)}\right)^{\frac{n-2}{n}}\right)$$

which is exactly the result wanted. □

We will first prove the following proposition in order to deduce Theorem 2.1.7.

Proposition 2.3.10. *For any ε and N large enough the following holds. For any intervals $J_{\mathrm{in}} \subset \mathcal{A}_r^{\kappa}$ and $J_{\mathrm{out}} \subset \{i,\, d(i,J_{\mathrm{in}}) \leqslant N^{-\varepsilon}Nw\}$ we have*

$$S_{J_{\mathrm{in}}}^{(0,\tau)} \leqslant N^{\varepsilon}\left(\frac{\ell}{Nw}+\frac{N\tau}{\ell}+\frac{1}{N\tau}\right)S_{J_{\mathrm{out}}}^{(0,\tau)}+N^{\varepsilon}\left(\frac{\widehat{I}}{\sqrt{N\tau}}+\widehat{I}\frac{w}{t}\right)(S_{J_{\mathrm{out}}}^{(0,\tau)})^{\frac{n-1}{n}}$$
$$+N^{\varepsilon}\left(\frac{\widehat{I}}{N\tau}+\widehat{I}\frac{N\tau}{\ell^2}\right)(S_{J_{\mathrm{out}}}^{(0,\tau})^{\frac{n-2}{n}}. \quad (2.3.71)$$

Proof. We will use the short-range dynamics and its finite speed of propagation property in order to localize the maximum principle in J_{in}. We will then use the local laws and Lemma 2.3.9 in order to get a Gronwall type bound. Define the following averaging operator. For f a function on configurations and $\boldsymbol{\eta}$ a configuration,

$$\mathrm{Av}(f)=\frac{3N^{\varepsilon}}{Nw}\sum_{\frac{1}{3}\frac{Nw}{N^{\varepsilon}}<a<\frac{2}{3}\frac{Nw}{N^{\varepsilon}}}\mathrm{Flat}_a(f). \quad (2.3.72)$$

Note that, if $\boldsymbol{\eta}$ is not included in J_{out}, by definition of the flattening operator, $(\mathrm{Av}(f))(\boldsymbol{\eta})=0$. The purpose of these operators is to change the initial condition in order to remove the particles far from the initial interval J_{in}. Note also that we can write, for any $\boldsymbol{\eta}$,

$$\mathrm{Av}(f)(\boldsymbol{\eta})=a_{\boldsymbol{\eta}}f(\boldsymbol{\eta}),$$

with $a_{\boldsymbol{\eta}}\in[0,1]$. Note that we have the elementary bound $|a_{\boldsymbol{\eta}}-a_{\boldsymbol{\xi}}|\leqslant CN^{\varepsilon}/Nw$. Define now the following dynamics

$$\begin{cases}\partial_s\Gamma_s=\mathscr{S}(s)\Gamma_s, \quad 0\leqslant s\leqslant\tau\\ \Gamma_0(\boldsymbol{\eta})=(\mathrm{Av}\,\Phi_t)(\boldsymbol{\eta}).\end{cases} \quad (2.3.73)$$

Now, if one takes a configuration $\boldsymbol{\eta}$ supported on J_{in}, it suffices to show the bound in Proposition 2.3.10 for Γ. Indeed

$$|F_\tau(\boldsymbol{\eta})-\Gamma_\tau(\boldsymbol{\eta})| \leqslant |(U_{\mathscr{B}}(0,\tau)\Phi_t-U_{\mathscr{S}}(0,\tau)\Phi_t)(\boldsymbol{\eta})|+|U_{\mathscr{S}}(0,\tau)(\Phi_t-\mathrm{Av}\Phi_t)(\boldsymbol{\eta})|$$
$$\leqslant N^{\varepsilon}\frac{CN\tau}{\ell}\left(S_{J_{\mathrm{out}}}^{(0,\tau)}+\widehat{I}\frac{w}{t}\left(S_{J_{\mathrm{out}}}^{(0,\tau)}\right)^{\frac{n-1}{n}}+\frac{\widehat{I}}{\ell}\left(S_{J_{\mathrm{out}}}^{(0,\tau)}\right)^{\frac{n-2}{n}}\right)+\exp(-cN^{\varepsilon})$$

where we bounded the first term by using Lemma 2.3.9 and the second term using the finite speed of propagation. Indeed, since $\boldsymbol{\eta}$ is supported on J_{in}, $(\mathrm{Id}-\mathrm{Av})\Phi_t$ vanishes for any configuration supported on J_{out}. Note that we can use Lemma 2.3.9 since we will take $\ell/N \ll w$. In the rest of the proof, we will prove the bound from Proposition 2.3.10 for Γ_τ. If we already have for some $C>0$,

$$\Gamma_\tau(\boldsymbol{\eta_m}):=\sup_{\boldsymbol{\eta},\,\mathcal{N}(\boldsymbol{\eta})=n}\Gamma_\tau(\boldsymbol{\eta})\leqslant N^{-C}$$

then we have nothing to prove by the argument above and the definition of F_τ. However, if this supremum is greater than N^{-C}, then by the finite speed of propagation of $\mathscr{S}$, we know that $\boldsymbol{\eta_m}$ will be supported in, for instance, $\left\{i,\, d(i, J_{\rm in}) \leqslant \frac{3Nw}{4N^\varepsilon}\right\}$.

Consider now, a parameter η that we will choose later and denote also $m \leqslant n$ the number of sites with at least a particle and $j_1, \ldots, j_m$ those sites. Then, we can write

$$\partial_\tau \Gamma_\tau(\boldsymbol{\eta_m}) = \sum_{0<|j-k|\leqslant \ell} \frac{2\eta_{m,k}(1+2\eta_{m,j})\left(\Gamma_\tau\left(\boldsymbol{\eta_m^{k,j_p}}\right) - \Gamma_\tau\left(\boldsymbol{\eta_m}\right)\right)}{N(\lambda_k(t+\tau)-\lambda_j(t+\tau))^2} \tag{2.3.74}$$

$$\leqslant \frac{C}{N\eta} \sum_{\substack{1\leqslant p\leqslant m\\ k,\, 0<|j_p-k|\leqslant \ell}} \frac{\eta(\Gamma_\tau\left(\boldsymbol{\eta_m^{k,j_p}}\right) - \Gamma_\tau(\boldsymbol{\eta_m}))}{(\lambda_k-\lambda_{j_p})^2+\eta^2} \tag{2.3.75}$$

$$\leqslant -\frac{C}{N\eta} \sum_{\substack{1\leqslant p\leqslant m\\ k,\, 0<|j_p-k|\leqslant \ell}} \operatorname{Im}\left(\frac{\Gamma_\tau\left(\boldsymbol{\eta_m^{k,j_p}}\right)}{\lambda_k - z_{j_p}}\right) - \frac{1}{N\eta}\Gamma_\tau(\boldsymbol{\eta_m}) \sum_{\substack{1\leqslant p\leqslant m\\ k,\, 0<|j_p-k|<\ell}} \operatorname{Im}\left(\frac{1}{z_{j_p}-\lambda_k}\right) \tag{2.3.76}$$

with $z_{j_p} = \lambda_{j_p} + i\eta$. For the second term, see that for $p \in [\![1, m]\!]$, if we choose η to be smaller than ℓ/N,

$$\#\left\{k,\, 0<|j_p-k|\leqslant \ell\right\} \geqslant C\ell \geqslant CN\eta \geqslant C\#\left\{k,\, |\lambda_{j_p}-\lambda_k| \leqslant \eta\right\} \geqslant C'N\eta$$

where we used, in the two last inequalities, the rigidity of the eigenvalues. Now we can write

$$\sum_{\substack{1\leqslant p\leqslant m\\ k,\, 0<|j_p-k|\leqslant \ell}} \operatorname{Im}\left(\frac{1}{z_{j_p}-\lambda_k}\right) \geqslant \sum_{\substack{1\leqslant p\leqslant m\\ k,\, 0<|\lambda_{j_p}-\lambda_k|\leqslant \eta}} \frac{\eta}{(\lambda_{j_p}-\lambda_k)^2+\eta^2} \geqslant CN.$$

We now need to control the first term in (2.3.76). To do so, we will split it in the three following terms:

$$\operatorname{Im}\left(\sum_{k,\, 0<|j_p-k|\leqslant \ell} \frac{(U_{\mathscr{S}}(0,\tau)\mathrm{Av}\Phi_t)\,(\boldsymbol{\eta_m^{j_p,k}}) - (\mathrm{Av}U_{\mathscr{S}}(0,\tau)\Phi_t)\,(\boldsymbol{\eta_m^{j_p,k}})}{N(z_{j_p}-\lambda_k)}\right) \tag{2.3.77}$$

$$+\operatorname{Im}\left(\sum_{k,\, 0<|j_p-k|\leqslant \ell} \frac{(\mathrm{Av}U_{\mathscr{S}}(0,\tau)\Phi_t)\,(\boldsymbol{\eta_m^{j_p,k}}) - (\mathrm{Av}U_{\mathscr{B}}(0,\tau)\Phi_t)\,(\boldsymbol{\eta_m^{j_p,k}})}{N(z_{j_p}-\lambda_k)}\right) \tag{2.3.78}$$

$$+\operatorname{Im}\left(\sum_{k,\, 0<|j_p-k|\leqslant \ell} \frac{(\mathrm{Av}U_{\mathscr{B}}(0,\tau)\Phi_t)\,(\boldsymbol{\eta_m^{j_p,k}})}{N(z_{j_p}-\lambda_k)}\right) \tag{2.3.79}$$

To bound the first term, we will use the finite speed of propagation property of $\mathscr{S}$ from Lemma 2.3.1. Indeed, we can write

$$(U_{\mathscr{S}}(0,\tau)\mathrm{Av}\Phi_t)\,(\boldsymbol{\eta_m^{j_p,k}}) - (\mathrm{Av}U_{\mathscr{S}}(0,\tau)\Phi_t)\,(\boldsymbol{\eta_m^{j_p,k}}) = \frac{3N^\varepsilon}{Nw} \sum_{\frac{1}{3}\frac{Nw}{N^\varepsilon}<a<\frac{2}{3}\frac{Nw}{N^\varepsilon}} \mathscr{U}_a\left(\boldsymbol{\eta_m^{j_p,k}}\right)$$

with

$$\mathscr{U}_a = U_{\mathscr{S}}(0,\tau)\mathrm{Flat_a}\Phi_t - \mathrm{Flat}_a U_{\mathscr{S}}(0,\tau)\Phi_t.$$

Fix a and consider three cases, if $\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}$ is supported on $\{i,\, d(i, J_{\rm In}) > a + N^{\varepsilon}\ell\}$ then by definition of $\mathrm{Flat_a}$ we have

$$\mathrm{Flat}_a U_{\mathscr{S}}(0,\tau)\Phi_t = 0,$$

and by Lemma 2.3.1 we have

$$|U_{\mathscr{S}}(0,\tau)\mathrm{Flat}_a\Phi_t| \leqslant \exp\left(-\frac{cN^{\varepsilon}}{2}\right).$$

Now if $\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}$ is supported on $\{i,\, d(i, J_{\rm In}) \leqslant a - N^{\varepsilon}\ell\}$ then again by definition of $\mathrm{Flat_a}$,

$$\mathrm{Flat}_a\left(U_{\mathscr{S}}(0,\tau)\Phi_t\right)\left(\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}\right) = \left(U_{\mathscr{S}}(0,\tau)\Phi_t\right)\left(\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}\right).$$

Thus

$$\left|\mathscr{U}_a\left(\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}\right)\right| \leqslant \left|U_{\mathscr{S}}(0,\tau)\left(\Phi_t - \mathrm{Flat}_a\Phi_t\right)\left(\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}\right)\right| \leqslant \exp\left(-\frac{cN^{\varepsilon}}{2}\right).$$

Finally, if $\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}$ is supported on $\{i,\, d(i, J_{\rm In}) \leqslant a + \ell N^{\varepsilon}\}$, first note that there can only be $2n\ell N^{\varepsilon}$ such a, then one can see that we can use the finite speed of propagation if we remove particle away from a at distance $2\ell N^{\varepsilon}$ for instance, then

$$\begin{aligned}
\left|\mathscr{U}_a\left(\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}\right)\right| &\leqslant |\mathrm{Flat}_a U_{\mathscr{S}}(0,\tau)\Phi_t| + |\mathrm{Flat}_a U_{\mathscr{S}}(0,\tau)\mathrm{Flat}_{a+2\ell N^{\varepsilon}}\Phi_t| \\
&\quad + |\mathrm{Flat_a} U_{\mathscr{S}}(0,\tau)\left(\Phi_t - \mathrm{Flat}_{a+2\ell N^{\varepsilon}}\right)| \\
&\leqslant \|\mathrm{Flat}_a\Phi_t\|_{\infty} + \|\mathrm{Flat}_{a+2\ell N^{\varepsilon}}\|_{\infty} + \exp\left(-\frac{cN^{\varepsilon}}{2}\right) \\
&\leqslant 2S_{J_{\rm out}}^{(0,\tau)} + \exp\left(-\frac{cN^{\varepsilon}}{2}\right),
\end{aligned}$$

where we used that $U_{\mathscr{S}}$ is a contraction in $\|\|_{\infty}$. Finally we can bound (2.3.77),

$$(2.3.77) \leqslant N^{\varepsilon}\frac{\ell}{Nw}S_{J_{\rm out}}^{(0,\tau)} + \exp\left(-\frac{cN^{\varepsilon}}{2}\right) \tag{2.3.80}$$

where we used the fact that

$$\left|\frac{1}{N}\mathrm{Im}\left(\sum_{k,0<|j_p-k|<\ell}^{N}\frac{1}{z_{j_p}-\lambda_k}\right)\right| \leqslant \left|\varsigma\left(\tau, z_{j_p}\right)\right| \leqslant N^{\varepsilon} \tag{2.3.81}$$

where the Stieltjes transform ς is defined in (2.2.14). For (2.3.78), we will use Lemma 2.3.9. Indeed, first note that in the short-range regime, the set of k such that $|j_p - k| \leqslant \ell$ is included in $\mathcal{A}_r^{2\kappa}$. Then we can bound

$$(2.3.78) \leqslant N^{\varepsilon}\frac{N\tau}{\ell}\left(S_{J_{\rm out}}^{(0,\tau)} + \widehat{I}\frac{w}{t}\left(S_{J_{\rm out}}^{(0,\tau)}\right)^{\frac{n-1}{n}} + \frac{\widehat{I}}{\ell}\left(S_{J_{\rm out}}^{(0,\tau)}\right)^{\frac{n-2}{n}}\right) \tag{2.3.82}$$

where we used the fact that Av is a contraction and (2.3.81).

Finally, in order to bound the third term (2.3.79), we will use the local law for Φ_t. First write

$$\begin{aligned}
\operatorname{Im}\left(\sum_{k,\,0<|j_p-k|\leqslant\ell}\frac{(\mathrm{Av}U_{\mathscr{B}}(0,\tau)\Phi_t)\left(\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}\right)}{N(z_{j_p}-\lambda_k)}\right) &= \operatorname{Im}\left(\sum_{k,\,0<|j_p-k|\leqslant\ell}\frac{a_{\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}}F_\tau\left(\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}\right)}{N(z_{j_p}-\lambda_k)}\right)\\
= \operatorname{Im}\left(\sum_{k,\,0<|j_p-k|\leqslant\ell}\frac{a_{\boldsymbol{\eta}_{\boldsymbol{m}}}F_\tau\left(\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}\right)}{N(z_{j_p}-\lambda_k)}\right) &+ \operatorname{Im}\left(\sum_{k,\,0<|j_p-k|\leqslant\ell}\frac{\left(a_{\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}}-a_{\boldsymbol{\eta}_{\boldsymbol{m}}}\right)F_\tau\left(\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}\right)}{N(z_{j_p}-\lambda_k)}\right)\\
= \operatorname{Im}\left(\sum_{k,\,0<|j_p-k|\leqslant\ell}\frac{a_{\boldsymbol{\eta}_{\boldsymbol{m}}}F_\tau\left(\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}\right)}{N(z_{j_p}-\lambda_k)}\right) &+ \mathcal{O}\left(N^{\varepsilon}\mathbb{1}_{|j_p-k|\leqslant\ell}\left|a_{\boldsymbol{\eta}_{\boldsymbol{m}}}-a_{\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}}\right|S^{(0,\tau)}_{J_{\mathrm{out}}}\right). \qquad (2.3.83)
\end{aligned}$$

But we have the bound, for $|j_p-k|\leqslant\ell$,

$$\left|a_{\boldsymbol{\eta}_{\boldsymbol{m}}}-a_{\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}}\right|\leqslant\frac{N^{\varepsilon}d\left(\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}},\boldsymbol{\eta}_{\boldsymbol{m}}\right)}{Nw}\leqslant\frac{N^{\varepsilon}\ell}{Nw}.$$

In order to bound the first term, see first that we can remove the contributions of $k\in\{j_1,\dots,j_p\}$ writing

$$\operatorname{Im}\left(\sum_{k,\,0<|j_p-k|\leqslant\ell}\frac{a_{\boldsymbol{\eta}_{\boldsymbol{m}}}F_\tau\left(\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}\right)}{N(z_{j_p}-\lambda_k)}\right)=\operatorname{Im}\left(\sum_{\substack{k,\,0<|j_p-k|\leqslant\ell\\ k\notin\{j_1,\dots,j_p\}}}\frac{a_{\boldsymbol{\eta}_{\boldsymbol{m}}}F_\tau\left(\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}\right)}{N(z_{j_p}-\lambda_k)}\right)+\mathcal{O}\left(\frac{N^{\varepsilon}}{N\eta}S^{(0,\tau)}_{J_{\mathrm{out}}}\right).$$

Recall now the definition of Φ_t from (2.1.27) and write,

$$\sum_{\substack{k,\,0<|j_p-k|\leqslant\ell\\ k\notin\{j_1,\dots,j_p\}}}\frac{F_\tau\left(\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}\right)}{N(z_{j_p}-\lambda_k)}=\frac{1}{\mathcal{M}\left(\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}\right)}\mathbb{E}\left[\sum_{\substack{k,\,0<|j_p-k|\leqslant\ell\\ k\notin\{j_1,\dots,j_p\}}}\sum_{G\in\mathcal{G}_{\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}}}\frac{\prod_{e\in\mathcal{E}(G)}p(e)}{N(z_{j_p}-\lambda_k)}\middle|\boldsymbol{\lambda}\right]. \qquad (2.3.84)$$

First consider the contribution of (2.3.64) in the sum in (2.3.84), denote $e_k=\{(k,1),(k,2)\}$ and write

$$\begin{aligned}
\sum_{\substack{k,\,0<|j_p-k|\leqslant\ell\\ k\notin\{j_1,\dots,j_p\}}}\sum_{G\in\mathcal{G}^{(1)}_{\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}}}\frac{\prod_{e\in\mathcal{E}(G)}p(e)}{N(z_{j_p}-\lambda_k)} &= \sum_{\substack{k,\,0<|j_p-k|\leqslant\ell\\ k\notin\{j_1,\dots,j_p\}}}\sum_{G\in\mathcal{G}^{(1)}_{\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{j_p},\boldsymbol{k}}}}\left(\prod_{e\in\mathcal{E}(G)\setminus\{e_k\}}p(e)\right)\frac{p_{kk}}{N(z_{j_p}-\lambda_k)}\\
&= \left(\sum_{G\in\mathcal{G}_{\boldsymbol{\eta}_{\boldsymbol{m}}\setminus\boldsymbol{j_p}}}\prod_{e\in\mathcal{E}(G)}p(e)\right)\sum_{\substack{k,\,0<|j_p-k|\leqslant\ell\\ k\notin\{j_1,\dots,j_p\}}}\frac{p_{kk}}{N(z_{j_p}-\lambda_k)}. \qquad (2.3.85)
\end{aligned}$$

To control the last term in (2.3.85), we can use the local law. First write,

$$\sum_{\substack{k,\,0<|j_p-k|\leqslant\ell\\ k\notin\{j_1,\dots,j_p\}}}\frac{p_{kk}}{N(z_{j_p}-\lambda_k)}=\sum_{k=1}^{N}\frac{p_{kk}}{N(z_{j_p}-\lambda_k)}+\sum_{k,\,|j_p-k|>\ell}\frac{p_{kk}}{N(z_{j_p}-\lambda_k)}+\mathcal{O}\left(N^{\varepsilon}\frac{\widehat{I}}{N\eta}\right), \qquad (2.3.86)$$

where we used the bound $|p_{kk}| \leqslant N^{\varepsilon}\widehat{I}$. Now, recall the definition of p_{kk} from (2.3.58) so that we have

$$\begin{aligned}
\left|\operatorname{Im}\sum_{k=1}^{N}\frac{p_{kk}}{N(z_{j_p}-\lambda_k)}\right| &= \left|\operatorname{Im}\frac{1}{N}\sum_{\alpha\in I}G_{\alpha\alpha}(z_{j_p}) - \frac{\operatorname{Im}m_t(z_{j_p})}{\operatorname{Im}m_t(z_0)}\frac{1}{N}\sum_{\alpha\in I}\operatorname{Im}g_\alpha(t,z_0)\right|\\
&\leqslant \left|\operatorname{Im}\frac{1}{N}\sum_{\alpha\in I}\left(G_{\alpha\alpha}(z_{j_p})-g_\alpha(t,z_{j_p})\right)\right| + \left|\frac{1}{N}\sum_{\alpha\in I}\left(\operatorname{Im}g_\alpha(t,z_{j_p}) - \frac{\operatorname{Im}m_t(z_{j_p})}{\operatorname{Im}m_t(z_0)}\operatorname{Im}g_\alpha(t,z_0)\right)\right|\\
&\leqslant N^{\varepsilon}\left(\frac{\widehat{I}}{\sqrt{N\eta}}+\widehat{I}\frac{w}{t}\right)
\end{aligned}$$

where $z_0 := \gamma_{t,i_0} + \mathrm{i}\eta$. Note that we used the fact that $\boldsymbol{\eta_m}$ is supported in J_{out}, so that $|z_0 - z_{j_p}| = |\gamma_{t,i_0} - \lambda_{j_p}| \leqslant N^{\varepsilon}w$. We can then use Lemma 2.3.7 and bound

$$\sum_{G\in\mathcal{G}_{\boldsymbol{\eta_m}\backslash \boldsymbol{j_p}}}\prod_{e\in\mathcal{E}(\mathcal{G}}p(e) = \mathcal{O}\left(\sup_{\boldsymbol{\eta}\subset J_{\text{out}},\mathcal{N}(\boldsymbol{\eta})=n-1}|\Phi_t(\boldsymbol{\eta})|\right) = \mathcal{O}\left((S^{(0,\tau)}_{J_{\text{out}}})^{\frac{n-1}{n}}\right)$$

Now, consider the contribution of (2.3.65) in the sum from (2.3.84). Note that for any graph G in $\mathcal{G}^{(2)}_{\boldsymbol{\xi}}$, there exists q and q' in $\{1,\dots,m\}$, $a\in\{1\dots,\eta_{i_q}\}$ and $b\in\{1,\dots,\eta_{i_{q'}}\}$, such that $e_q := \{(k,1),(q,a)\}$ and $e_{q'} := \{(k,2),(q',b)\}$ are edges in G. We can then write

$$\begin{aligned}
\sum_{k,\,0<|j_p-k|\leqslant\ell}\sum_{G\in\mathcal{G}^{(2)}_{\boldsymbol{\eta_m}^{\boldsymbol{j_p},\boldsymbol{k}}}}\frac{\prod_{e\in\mathcal{E}(G)}p(e)}{N(z_{j_p}-\lambda_k)} &= \sum_{k,\,0<|j_p-k|\leqslant\ell}\sum_{G\in\mathcal{G}^{(2)}_{\boldsymbol{\eta_m}^{\boldsymbol{j_p},\boldsymbol{k}}}}\left(\prod_{e\in\mathcal{E}(G)\backslash\{e_q,e_{q'}\}}p(e)\right)\frac{p_{i_q,k}p_{i_{q'},k}}{N(z_{j_p}-\lambda_k)},\\
&= \sum_{q,q'=1}^{m}\left(\sum_{G\in\mathcal{H}^{q,q'}_{\boldsymbol{\eta_m}\backslash\boldsymbol{j_p}}}\prod_{e\in\mathcal{E}(G)}p(e)\right)\sum_{k,\,0<|j_p-k|\leqslant\ell}\frac{p_{i_q,k}p_{i_{q'},k}}{N(z_{j_p}-\lambda_k)}, \qquad (2.3.87)
\end{aligned}$$

where we defined the set of graphs $\mathcal{H}^{q,q'}_{\boldsymbol{\eta}}$ to be the set of perfect matching of the complete graph on the set of vertices $\mathcal{V}_{\boldsymbol{\eta}}$ where we removed a single particle at the site i_q and $i_{q'}$. Note that for any graph $G\in\mathcal{H}^{q,q'}_{\boldsymbol{\eta_m}\backslash\boldsymbol{j_p}}$, $\prod_{e\in\mathcal{E}(G)}p(e)$ is a monomial of degree $n-2$.
Now, we can bound the imaginary part of the second sum in (2.3.87),

$$\left|\operatorname{Im}\left(\sum_{\substack{k,\,0<|j_p-k|\leqslant\ell\\ k\notin\{j_1,\dots,j_p\}}}\frac{p_{i_q,k}p_{i_{q'},k}}{N(z_{j_p}-\lambda_k)}\right)\right| \leqslant \frac{C}{N\eta}\sum_{k=1}^{N}\left(p^2_{i_q,k}+p^2_{i_{q'},k}\right) = \mathcal{O}\left(\frac{N^{\varepsilon}\widehat{I}}{N\eta}\right). \qquad (2.3.88)$$

For the last inequality, we used the following identity on eigenvectors

$$\sum_{k=1}^{N}p^2_{i,k} = \sum_{\alpha\in I}u_i(\alpha)^2$$

and that for any $\varepsilon>0$,using the entrywise local law from Theorem 2.2.3 on a diagonal entry of the resolvent,

$$u_k(\alpha)^2 \leqslant N^{-1+\varepsilon}\operatorname{Im}\left(G(\tau,\lambda_k+\mathrm{i}N^{-1+\varepsilon})_{\alpha,\alpha}\right) \leqslant \frac{N^{\varepsilon}}{Nt}$$

Again, we can bound the other term from (2.3.87) using Lemma 2.3.7,

$$\sum_{G\in\mathcal{H}^{q,q'}_{\boldsymbol{\eta m}\setminus\boldsymbol{jp}}}\prod_{e\in\mathcal{E}(G)}p(e)=\mathcal{O}\left(N^{\varepsilon}\sup_{\boldsymbol{\eta}\subset J_{\mathrm{out}},\mathcal{N}(\boldsymbol{\eta})=n-2}|\Phi_t(\boldsymbol{\eta})|\right)=\mathcal{O}\left(\left(S^{(0,\tau)}_{J_{\mathrm{out}}}\right)^{\frac{n-2}{n}}\right)$$

Finally, putting all these estimates together, we get the Gronwall-type inequality,

$$\partial_\tau\Gamma_\tau(\boldsymbol{\eta m})\leqslant-\frac{1}{\eta}\Gamma_\tau(\boldsymbol{\eta m})+\mathcal{O}\left(\frac{N^{\varepsilon}}{\eta}\left(\left(\frac{\ell}{Nw}+\frac{N\tau}{\ell}+\frac{1}{N\eta}\right)S^{(0,\tau)}_{J_{\mathrm{out}}}\right.\right.$$
$$\left.\left.+\left(\frac{\widehat{I}}{\sqrt{N\eta}}+\widehat{I}\frac{w}{t}\right)(S^{(0,\tau)}_{J_{\mathrm{out}}})^{\frac{n-1}{n}}+\left(\frac{\widehat{I}}{N\eta}+\frac{N\tau}{\ell^2}\right)(S^{(0,\tau}_{J_{\mathrm{out}}})^{\frac{n-2}{n}}\right)\right)\quad(2.3.89)$$

In order to get a proper bound using Gronwall's lemma, we need to take $\eta\ll\tau$ but to get the best estimates possible, we also have to take η as large as possible. Hence, considering $\eta=N^{-\varepsilon}\tau$ we have the bound

$$S^{(0,\tau)}_{J_{\mathrm{in}}}\leqslant N^{\varepsilon}\left(\frac{\ell}{Nw}+\frac{N\tau}{\ell}+\frac{1}{N\tau}\right)S^{(0,\tau)}_{J_{\mathrm{out}}}+N^{2\varepsilon}\left(\frac{\widehat{I}}{\sqrt{N\tau}}+\widehat{I}\frac{w}{t}\right)(S^{(0,\tau)}_{J_{\mathrm{out}}})^{\frac{n-1}{n}}$$
$$+N^{\varepsilon}\left(\frac{\widehat{I}}{N\tau}+\widehat{I}\frac{N\tau}{\ell^2}\right)(S^{(0,\tau}_{J_{\mathrm{out}}})^{\frac{n-2}{n}}\quad(2.3.90)$$

which gives the Proposition 2.3.10. □

Now that we have the bound from Proposition 2.3.10, we are able to get a bound on the p_{ij} using Lemma 2.3.7. To do so, we can use a sequence of set of indices with decreasing size and apply recursively Proposition 2.3.10. We will also need to choose the right parameters ℓ, w and τ.

Proof of Theorem 2.3.8. Consider first any ε small enough, such that if we write $t=N^{-1+\delta}$ (recall that $t\in\mathcal{T}_\omega$ so that $t\gg N^{-1}$) we have $\varepsilon<4\delta/3$. and a large $D>0$. Then we can take the following parameters :

$$w=N^{\varepsilon}\tau\quad\text{and}\quad\ell=N\sqrt{\tau w},\qquad(2.3.91)$$

note that we have the right bounds between these parameters: $N^{-1}\ll\tau\ll\ell/N\ll w\ll t$, and define the following sequence of sets of indices, defined implicitely,

$$\begin{cases}J_0=\mathcal{A}^{\kappa}_{w}(i_0),\\ J_i=\{i:\ d(i,J_{i+1})\leqslant N^{-\varepsilon}Nw\}.\end{cases}$$

From Proposition 2.3.10 we have the following bound holding with overwhelming probability,

$$S^{(0,\tau)}_{J_{i+1}}\leqslant N^{-\varepsilon/2}S^{(0,\tau)}_{J_i}+\left(\frac{\widehat{I}}{\sqrt{N\tau}}+\widehat{I}\frac{N^{\varepsilon}\tau}{t}\right)(S^{(0,\tau)}_{J_i})^{\frac{n-1}{n}}+\widehat{I}\frac{1}{N\tau}(S^{(0,\tau)}_{J_i})^{\frac{n-2}{n}}.$$

Now see that as long as we have

$$(S^{(0,\tau)}_{J_i})^{1/n}\geqslant CN^{3\varepsilon/2}\Xi(\tau)$$

with Ξ given in (2.1.12), we obtain the recursive bound

$$S^{(0,\tau)}_{J_{i+1}}\leqslant N^{-\varepsilon/2}S^{(0,\tau)}_{J_i}.$$

But if we take a very large i so that the previous bound cannot hold, for instance $i = \lceil 3\varepsilon^{-1} \rceil$, then it means that for such a i we have the bound

$$(S_{J_i}^{(0,\tau)})^{1/n} \leqslant CN^{3\varepsilon/2}\Xi.$$

Now using the definition of p_{ii}, we have for $i \in \mathcal{A}_w^\kappa(i_0)$,

$$\left| \sum_{\alpha \in I} \left(u_i(\alpha)^2 - \frac{1}{N}\sigma_t^2(\alpha, i) \right) \right| \leqslant |p_{ii}| + \widehat{I}\frac{w}{t} \leqslant |p_{ii}| + \Xi(\tau). \tag{2.3.92}$$

Finally, using Markov's inequality, taking for instance $n = \lfloor 3D/\varepsilon \rfloor$, and using Lemma 2.3.7 to bound the p_{ij} by $S^{(0,\tau)}$ we have

$$\mathbb{P}\left(|p_{ii}| + |p_{ij}| \geqslant N^\varepsilon \Xi(\tau) \right) \leqslant N^{-D}. \tag{2.3.93}$$

The result then follows from combining (2.3.92) and (2.3.93). □

2.4. Approximation by a Gaussian divisible ensemble

2.4.1. Continuity of the Dyson Brownian motion

In Subsection 2.3.1 we showed that the moments of the eigenvectors of the matrix H_τ are asymptotically those of a Gaussian random variable with variance σ_t^2. If we would have taken the time $t - \tau$ from the start, the previous section gives us (2.3.15) for W a matrix from the Gaussian Orthogonal Ensemble. Now, since τ is a small time, recall that $\tau \ll t$, we can use the continuity of the Dyson Brownian motion to show that H_τ and $H_0 = W_t$ have the same local statistics. In order to state a proper continuity lemma we need to have a dynamics with constant second moments and vanishing expectation.

First see that the variance of the centered model is

$$\mathbb{E}\left[(W_{t,ij} - D_{ij})^2 \right] = \frac{t}{N}.$$

Consider, for $0 \leqslant s \leqslant \tau$, the following variance-preserving dynamics on symmetric matrices.

$$\mathrm{d}\left(\tilde{H}(s) - D \right) = \frac{\mathrm{d}B}{\sqrt{N}} - \frac{1}{2t}\left(\tilde{H}(s) - D \right) \mathrm{d}s, \tag{2.4.1}$$

$$\tilde{H}(0) = W_t = D + \sqrt{t}W. \tag{2.4.2}$$

The following lemma gives us a continuity argument between $\tilde{H}(\tau)$ and W_t. It is similar to Lemma A.1 in [BY17] or Lemma 4.3 in [HLY15]. We will later use this lemma on the resolvent entries.

Lemma 2.4.1. *Denote $\partial_{ij} = \partial_{\tilde{H}_{ij}}$. Take F a smooth function of the matrix entries satisfying*

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}, 0 \leqslant s \leqslant \tau} \frac{1}{N} \sum_{i \leqslant j} \left(\frac{N|\left(\tilde{H}(s) - D\right)_{ij}|^3}{t} + |\left(\tilde{H}(s) - D\right)_{ij}| \right) \left| \partial_{ij}^3 F(\boldsymbol{\theta}\tilde{H}_s) \right| \right] \leqslant M \tag{2.4.3}$$

where $(\boldsymbol{\theta}H)_{ij} = \theta_{ij}H_{ij}$ with $\theta_{kl} = 1$ for $\{k,l\} \neq \{i,j\}$ and $\theta_{ij} \in [0,1]$. Then

$$\mathbb{E}[F(\tilde{H}(\tau))] - \mathbb{E}[F(\tilde{H}(0))] = \mathcal{O}(\tau)\, M.$$

Proof. By Itô's formula we have

$$\partial_s \mathbb{E}[F(\tilde{H}(s))] = -\frac{1}{2N}\sum_{i\leqslant j}\frac{N}{t}\mathbb{E}[\Big(\tilde{H}(s)-D\Big)_{ij}\partial_{ij}F(\tilde{H}_s)] - \mathbb{E}[\partial_{ij}^2 F(\tilde{H}_s)].$$

Using Taylor expansions, we can write, forgetting the dependence in time for clarity,

$$\begin{aligned}\mathbb{E}[\Big(\tilde{H}-D\Big)_{ij}\partial_{ij}F(\tilde{H})] &= \mathbb{E}[\Big(\tilde{H}-D\Big)_{ij}\partial_{ij}F_{\tilde{H}_{ij}=D_{ij}}] + \mathbb{E}[\Big(\tilde{H}-D\Big)_{ij}^2\partial_{ij}^2F_{\tilde{H}_{ij}=D_{ij}}]\\ &\quad + \mathcal{O}\left(\mathbb{E}\left[\sup_{\boldsymbol{\theta}}\left|\Big(\tilde{H}-D\Big)_{ij}^3\partial_{ij}^3F(\boldsymbol{\theta}\tilde{H})\right|\right]\right)\\ &= \frac{t}{N}\partial_{ij}^2F_{\tilde{H}_{ij}=D_{ij}} + \mathcal{O}\left(\mathbb{E}\left[\sup_{\boldsymbol{\theta}}\left|\Big(\tilde{H}-D\Big)_{ij}^3\partial_{ij}^3F(\boldsymbol{\theta}\tilde{H})\right|\right]\right).\end{aligned}$$

and

$$\mathbb{E}[\partial_{ij}^2F(\tilde{H}_s)] = \mathbb{E}[\partial_{ij}^2F_{\tilde{H}_{ij}=D_{ij}}] + \mathcal{O}\left(\mathbb{E}\left[\sup_{\boldsymbol{\theta}}\left|\Big(\tilde{H}-D\Big)_{ij}\partial_{ij}^3F(\boldsymbol{\theta}\tilde{H}_s)\right|\right]\right).$$

Putting everything together the claim follows. □

This continuity property of the Dyson Brownian motion gives us a control over the eigenvalues and eigenvectors of $\tilde{H}(0)$ and $\tilde{H}_s$.

Corollary 2.4.2. *Let $\kappa\in(0,1)$ and $m\in\mathbb{N}$. Let $\Theta:\mathbb{R}^{2m}\to\mathbb{R}$ be a smooth function satisfying*

$$\sup_{k\in[\![0,5]\!],x\in\mathbb{R}^{2m}}|\Theta^{(k)}(x)(1+|x|)^{-C}<\infty,$$

for some $C>0$. Denote $\tilde{u}_1(s),\dots,\tilde{u}_N(s)$ the eigenvectors of $\tilde{H}(s)$ associated with the eigenvalues $\tilde{\lambda}_1(s),\dots,\tilde{\lambda}_N(s)$. Define, for a small $\mathfrak{a}$ the time domain

$$\mathcal{T}'_{\mathfrak{a}} = \left[\frac{N^{\mathfrak{a}}}{N}, N^{-\mathfrak{a}}\sqrt{\frac{t}{N}}\right],$$

then for any $\tau\in\mathcal{T}'_{\mathfrak{a}}$, there exists $\mathfrak{p}>0$ depending on Θ, $\mathfrak{a}$, κ and r such that

$$\sup_{I\subset\mathcal{I}_r^\kappa,\,|I|=m,\,\|\mathbf{q}\|=1}\left|\left(\mathbb{E}^{\tilde{H}_s}-\mathbb{E}^{\tilde{H}_0}\right)\Theta\left(\left(N(\tilde{\lambda}_k-\gamma_{k,t}),\frac{N}{\sigma_t^2(\mathbf{q},k,\eta)}\langle\mathbf{q},\tilde{u}_k\rangle^2\right)_{k\in I}\right)\right|\leqslant N^{-\mathfrak{p}}. \tag{2.4.4}$$

Proof. We prove this corollary using the continuity estimate from Lemma 2.4.1. To do so, we use the techniques introduced in [KY13b] in order to change estimates of the resolvent below microscopic scales, in other words control $G(E+\mathrm{i}\eta)$ for $\eta\ll N^{-1}$, into estimates on eigenvectors. Indeed, it has been shown that such a control of the resolvent combined with an estimate on the number of eigenvalues in a very small interval allows us via integrating the resolvent over such an interval to gain estimates on eigenvectors. We can then split the proof into two results we need to show:

(*i*) A level repulsion estimate on the eigenvalues for both matrix ensembles of the following form: for $E\in\mathcal{I}_r^\kappa$ and a small $\xi>0$ there exists $\mathfrak{d}>0$ such that

$$\mathbb{P}\left(\left|\left\{\lambda_i\in[E-N^{-1-\xi},E+N^{-1-\xi}]\right\}\right|\geqslant 2\right)\leqslant N^{-\xi-\mathfrak{d}}.$$

(*ii*) Comparison of the resolvent below microscopic scales: for any smooth function F of polynomial growth, there exists a $\mathfrak{c} > 0$ and a $\xi > 0$ such that for all $N^{-1-\xi} < \eta < t$,

$$\sup_{\substack{\|\mathbf{q}\|_2=1,\\ E_1,\dots,E_m\in\mathcal{I}_r^\kappa}} \left|\left(\mathbb{E}^{\tilde{H}_\tau} - \mathbb{E}^{W_t}\right) F\left(\left(\frac{1}{\operatorname{Im}\left(\sum_{i=1}^N q_i^2 g_i(t,z_k)\right)}\langle \mathbf{q}, G(z_k)\mathbf{q}\rangle\right)_{k=1}^m\right)\right| \leqslant N^{-\mathfrak{c}} \tag{2.4.5}$$

where $z_k = E_k + \mathrm{i}\eta$.

We first prove (i) for the eigenvalues of W_t. This property can be deduced from gap universality, note that gap universality for Gaussian perturbation of size $t \in \mathcal{T}_\omega$ has been shown in [LY17a] (a stronger level repulsion estimate can also be found in [LY17a, Section 5]). In [LY17a, Subsection 2.4], Landon−Yau explains that they can deduce universality for deformed Wigner ensembles. However, they state the result for an initial condition such that $r^2 \gg t$. It has been confirmed by the authors that it is a simple typographical error and should be read as $r \gg t$. We will nonetheless give an idea of the proof of (i) for the sake of completeness.

As said earlier, we will first apply Lemma 2.4.1 to

$$F(\tilde{H}_s) = \frac{1}{N}\mathrm{Tr}(\tilde{H}_s - z)^{-1} \quad \text{for } z \text{ in} \quad \left\{z = E + \mathrm{i}\eta,\, E \in \mathcal{I}_r^\kappa,\, N^{-1-\xi} < \eta < t\right\}$$

for $\xi > 0$ arbitrarily small. See that in Lemma 2.4.1, we need to bound a functional of the form $F(\theta\tilde{H}_s)$ for θ a perturbation of two entries of the matrix. Since we will only need bounds such as Theorem 2.2.4 or Corollary 2.2.5, such bounds still hold for the perturbated matrix. So we will explain the bound for the third derivative of F applied directl to $\tilde{H}_s$. Note that by definition of $\tilde{H}$, we have $|(\tilde{H} - D)_{ij}| \prec \sqrt{\frac{t}{N}}$ so that we can bound the left hand side of 2.4.3 by

$$\frac{1}{N}\sqrt{\frac{t}{N}}\sum_{i\leqslant j}|\partial_{ij}^3 F(\tilde{H}_s)|. \tag{2.4.6}$$

Taking the third derivative of F with respect to an entry, we obtain, writing $G = (\tilde{H}_s - z)^{-1}$ for simplicity

$$\partial_{ij}^3 F(\tilde{H}_s) = -\frac{1}{N}\sum_{k=1}^N\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} G_{k\alpha_1}G_{\beta_1\alpha_2}G_{\beta_2\alpha_3}G_{\beta_3 k}$$

where $\{\alpha_\ell, \beta_\ell\} = \{i, j\}$ for $\ell = 1, 2, 3$. To bound the sum in the previous equation, we will need the following high probability bound for the off-diagonal entries of the resolvent which follows directly from Theorem 2.2.3

$$|G_{ij}(z)| \prec \frac{1}{\sqrt{N\eta}}\sqrt{|g_i(t,z)g_j(t,z)|}. \tag{2.4.7}$$

Note that these bounds holds for $\eta \gg N^{-1}$, we will first consider such η.

Finally we can bound (2.4.6),

$$\frac{1}{N}\sqrt{\frac{t}{N}}\sum_{i\leqslant j}\left|\partial_{ij}^3 F(\tilde{H}_s)\right| \prec \frac{N^{2\xi}}{N^2}\sqrt{\frac{t}{N}}\sum_{i\leqslant j}\sum_{k=1}^N\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}}|g_k|\left(|g_{\alpha_1}g_{\beta_1}g_{\alpha_2}g_{\beta_2}g_{\alpha_3}g_{\beta_3}|\right)^{1/2}. \tag{2.4.8}$$

Now, from Lemma 7.5 of [LY17a], we have

$$\frac{1}{N}\sum_{k=1}^{N}|g_k(t,z)| \leqslant C\log N \tag{2.4.9}$$

where the constant C only depend on D our diagonal matrix. Besides, in the last product of (2.4.8), there are, by definition of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, three occurrences of g_i and three occurrences of g_j. Thus,

$$(2.4.8) \leqslant \frac{CN^{2\xi}\log N}{N}\sqrt{\frac{t}{N}}\sum_{i\leqslant j}|g_i(t,z)g_j(t,z)|^{3/2} \leqslant \frac{CN^{2\xi}\log N}{Nt}\sqrt{\frac{t}{N}}\left(\sum_{i=1}^{N}|g_i(t,z)|\right)^2 \tag{2.4.10}$$

$$\leqslant N^{2\xi}\log^3 N\sqrt{\frac{N}{t}} \tag{2.4.11}$$

where we used (2.4.9) in the first inequality, the fact that $|g_j| \leqslant Ct^{-1}$ in the second and (2.4.9) again in the final inequality. In order to go below microscopic scales, recall that we used local laws that holds down to mesoscopic scales, we can use the following identity, for $y \leqslant \eta$,

$$\operatorname{Im}\left(\frac{1}{N}\operatorname{Tr}G(E+\mathrm{i}y)\right) \leqslant \frac{\eta}{y}\operatorname{Im}\left(\frac{1}{N}\operatorname{Tr}G(E+\mathrm{i}\eta)\right).$$

Finally, using Lemma 2.4.1, we get,

$$\sup_{E\in\mathcal{I}_r^\kappa}\left|\left(\mathbb{E}^{\tilde{H}_\tau}-\mathbb{E}^{W_t}\right)\left[\frac{1}{N}\operatorname{Tr}G(z)\right]\right| \leqslant N^{5\xi}\tau\sqrt{\frac{N}{t}} \leqslant N^{-\mathfrak{c}} \tag{2.4.12}$$

for some $\mathfrak{c} > 0$ by taking $\tau \in \mathcal{T}'_{\mathfrak{a}}$ for $\mathfrak{a} > 5\xi$. We can easily generalize this result to a product of trace, indeed taking

$$F(\tilde{H}_s) = \prod_{k=1}^{m} F_k \quad \text{with} \quad F_k = \frac{1}{N}\operatorname{Tr}G(z_k),$$

we can take the third derivative and write

$$\partial_{ij}^3 F = \sum_{k_1=1}^{m}\partial_{ij}^3 F\prod_{k\neq k_1}F_k + 3\sum_{k_1=1}^{m}\sum_{k_2\neq k_1}\partial_{ij}^2 F_{k_1}\partial_{ij}F_{k_2}\prod_{k\neq k_1,k_2}F_k$$
$$+\sum_{k_1=1}^{m}\sum_{k_2\neq k_1}\sum_{k_3\neq k_1,k_2}\partial_{ij}F_{k_1}\partial_{ij}F_{k_2}\partial_{ij}F_{k_3}\prod_{k\neq k_1,k_2,k_3}F_k. \tag{2.4.13}$$

Then, using the first and second derivative of F_k,

$$\partial_{ij}F_k = \frac{1}{N}\sum_{k=1}^{N}\sum_{\substack{\alpha,\beta\\ \{\alpha,\beta\}=\{i,j\}}} G_{k,\alpha}G_{\beta,k},$$

$$\partial_{ij}^2 F_k = \frac{1}{N}\sum_{k=1}^{N}\sum_{\substack{\boldsymbol{\alpha},\boldsymbol{\beta}\\ \{\alpha_k,\beta_k\}=\{i,j\}}} G_{k\alpha_1}G_{\beta_1\alpha_2}G_{\beta_2 k},$$

we can bound (2.4.13) in a similar way and finishing the bound by Lemma 2.4.1. Again, now that we have any polynomial of fixed degree, we can also extend to any smooth function F with polynomial growth.

Now, a consequence of these uniform bounds in $\mathrm{Re}(z)$ between $\tilde{H}_0 = W_t$ and $\tilde{H}_\tau$ for $\tau \in \mathcal{T}'_{\mathfrak{a}}$ for some small $\mathfrak{a}$ gives us a comparison of the gap distribution between these two matrix ensembles (see [EYY12a] for instance). Namely, there exists $c_1 > 0$ such that for any O a smooth test function of n variables and any index i such that $\gamma_{i,t} \in \mathcal{I}_r^\kappa$, we have for N large enough and $i_1, \dots, i_n$ indices such that $i_k \leqslant N^{c_1}$,

$$\left| \left(\mathbb{E}^{W_t} - \mathbb{E}^{\tilde{H}_\tau} \right) \left[O\left(N\rho_t^{(N)}(\gamma_{i,t})(\lambda_i - \lambda_{i,i+i_1}), \dots, N\rho_t^{(N)}(\gamma_{i,t})(\lambda_i - \lambda_{i+i_n}) \right) \right] \right| \leqslant N^{-c_1}. \tag{2.4.14}$$

But $\tilde{H}_s$ is a matrix with a small Gaussian component following the conditions of [LY17a], so that we have, for this matrix ensemble, gap universality. Hence, combining this gap universality with the continuity of the Green's function, we obtain gap universality of the matrix ensemble $D + \sqrt{t}W$. In other words, there exists $c_2 > 0$ such that, taking the same assumptions as for (2.4.14), we can write,

$$\begin{aligned} \Big| \mathbb{E}^{W_t} & \left[O\left(N\rho_t^{(N)}(\gamma_{i,t})(\lambda_i - \lambda_{i,i+i_1}), \dots, N\rho_t^{(N)}(\gamma_{i,t})(\lambda_i - \lambda_{i+i_n}) \right) \right] \\ & - \mathbb{E}^{\mathrm{GOE}} \left[O\left(N\rho_{sc}^{(N)}(\mu_i)(\lambda_i - \lambda_{i,i+i_1}), \dots, N\rho_{sc}^{(N)}(\mu_i)(\lambda_i - \lambda_{i+i_n}) \right) \right] \Big| \leqslant N^{-c_2} \end{aligned} \tag{2.4.15}$$

where ρ_{sc} is the density of Wigner's semicircular law and μ_i its quantiles defined by

$$\rho_{sc}(x) = \sqrt{4 - x^2}\mathbb{1}_{[-2,2]}, \quad \int_{-\infty}^{\mu_i} \mathrm{d}\rho_{sc}(E) = \frac{i}{N}. \tag{2.4.16}$$

Finally, (2.4.15) combined with Theorem 2.2.6 gives us the level repulsion estimate (i) for the matrix W_t, indeed consider $E \in \mathcal{I}_r^\kappa$, and ℓ the index such that

$$|\gamma_{\ell,t} - E| \leqslant \min_{k \in \mathcal{I}_r^\kappa} |\gamma_{k,t} - E|,$$

then, for any $\tilde{\varepsilon} > 0$, we have

$$\begin{aligned} \mathbb{P}\left(\left| \left\{ i,\, \lambda_i \in [E - N^{-1-\xi}, E + N^{-1-\xi}] \right\} \right| \geqslant 2 \right) & \leqslant \sum_{|k-\ell| \leqslant N^{\tilde{\varepsilon}}} \mathbb{P}^{W_t}\left(|\lambda_k - \lambda_{k+1}| < N^{-1-\xi} \right) \\ & \leqslant \sum_{|k-\ell| \leqslant N^{\tilde{\varepsilon}}} \mathbb{P}^{\mathrm{GOE}}\left((|\lambda_k - \lambda_{k+1}| < N^{-1-\xi} \right) + N^{-c_\varepsilon + \tilde{\varepsilon}} \\ & \leqslant N^{-2\xi + \tilde{\varepsilon}} + N^{-c_\varepsilon + \tilde{\varepsilon}} \\ & \leqslant N^{-\xi - \delta} \end{aligned}$$

for some $\delta > 0$ by taking $\tilde{\varepsilon}$ and $\xi > 0$ small enough. Note that we used rigidity in the first inequality, gap universality in the second and a level repulsion estimate for GOE matrix which for instance can be found in [EY15] .

In order to get the resolvent estimate (ii), we will use Lemma 2.4.1. To do so, we will first explain how to get the bound M for

$$F(\tilde{H}_s) = \frac{1}{\mathrm{Im}\left(\sum_{i=1}^N q_i^2 g_i(t,z) \right)} \langle \mathbf{q}, (\tilde{H} - z)^{-1} \mathbf{q} \rangle$$

for $z \in \mathbb{C}$ down to below microscopic scales. To get the right bound, we will first need to use local laws which hold down to mesoscopic scales $\eta = N^{-1+\xi}$.

Now for the third derivative of F, first write

$$|\partial_{ij}^3 F(H_s)| = \left| \frac{1}{\operatorname{Im}\left(\sum_{i=1}^N q_i^2 g_i(t,z)\right)} \sum_{1\leqslant a,b\leqslant N} \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} q_a G_{a\alpha_1} G_{\beta_1\alpha_2} G_{\beta_2\alpha_3} G_{\beta_3 b} q_b \right| \tag{2.4.17}$$

where $\{\alpha_k, \beta_k\} = \{i, j\}$ for $k = 1, 2, 3$. In order to bound the four terms coming up in the previous equation we will need Corollary 2.2.5. Writing (2.2.6) for $\mathbf{v} = \mathbf{q}$ and $\mathbf{w} = \mathbf{e_i}$, we obtain

$$\langle \mathbf{q}, G\mathbf{e_i} \rangle = q_i g_i(t,z) + \mathcal{O}_\prec \left(\frac{1}{\sqrt{N\eta}} \sqrt{\operatorname{Im}\left(\sum_{k=1}^N q_k^2 g_k(t,z) \right) \operatorname{Im}\left(g_i(t,z)\right)} \right).$$

Note that since we want a bound holding down to microscopic scales, the error terms has to be taken into account. In particular, we will consider η sufficiently small that we can bound every $(N\eta)^{1/2}$ by $N^{\xi/2}$. In the following computations, we will not bound the errors coming cross terms for simplicity, they can be bounded in a similar way.

We can divide the sum in (2.4.17) in three parts. The first case consists in $\{\beta_1, \alpha_2\} = \{\beta_2, \alpha_3\} = \{i, j\}$. In this case, note that, necessarily, $\{\alpha_1, \beta_3\} = \{i, j\}$ and write

$$\begin{aligned}
\sum_{1\leqslant a,b\leqslant N} q_a G_{a\alpha_1} G_{\beta_1\alpha_2} G_{\beta_2\alpha_3} G_{\beta_3 b} q_b &= \langle \mathbf{q}, G\mathbf{e}_{\boldsymbol{\alpha_1}} \rangle G_{\beta_1\alpha_2} G_{\beta_2\alpha_3} \langle \mathbf{e}_{\boldsymbol{\beta_3}}, G\mathbf{q} \rangle \\
&\prec \frac{1}{N\eta} \min\left(|g_i(t,z)|, |g_j(t,z)|\right)^2 \langle \mathbf{q}, G\mathbf{e}_{\boldsymbol{\alpha_1}} \rangle \langle \mathbf{e}_{\boldsymbol{\beta_3}}, G\mathbf{q} \rangle \\
&\prec N^{2\xi} \Big(\min\left(|g_i(t,z)|, |g_j(t,z)|\right)^2 |q_i g_i(t,z) q_j g_j(t,z)| \qquad (2.4.18) \\
&+ \min\left(|g_i(t,z)|, |g_j(t,z)|\right)^2 \operatorname{Im}\left(\sum_{k=1}^N q_k^2 g_k(t,z) \right) \sqrt{|g_i(t,z) g_j(t,z)|} \Big). \qquad (2.4.19)
\end{aligned}$$

Putting the leading order (2.4.18) in the sum of (2.4.6), we have the bound

$$\begin{aligned}
&\sqrt{\frac{t}{N}} \frac{1}{N\operatorname{Im}\left(\sum_{k=1}^N q_k^2 g_k(t,z)\right)} \sum_{1\leqslant i\leqslant j\leqslant N} (2.4.18) \qquad (2.4.20) \\
&\leqslant \sqrt{\frac{t}{N}} \frac{N^{2\xi}}{N\operatorname{Im}\left(\sum_{k=1}^N q_k^2 g_k(t,z)\right)} \sum_{1\leqslant i\leqslant j\leqslant N} |q_i q_j| |g_i(t,z) g_j(t,z)|^2 \qquad (2.4.21) \\
&\leqslant \sqrt{\frac{t}{N}} \frac{CN^{2\xi}}{N\operatorname{Im}\left(\sum_{k=1}^N q_k^2 g_k(t,z)\right)} \sum_{1\leqslant i\leqslant j\leqslant N} \left(q_i^2 + q_j^2\right) |g_i(t,z) g_j(t,z)|^2 \qquad (2.4.22) \\
&\leqslant \sqrt{\frac{t}{N}} \frac{CN^{2\xi}}{N\operatorname{Im}\left(\sum_{i=1}^N q_k^2 g_k(t,z)\right)} \left(\sum_{i=1}^N q_i^2 |g_i(t,z)|^2 \right) \sum_{j=1}^N |g_j(t,z)|^2. \qquad (2.4.23)
\end{aligned}$$

Note that by definition of $g_i(t,z) = (D_i - z - tm_t(z))^{-1}$, the fact that $\operatorname{Im} m_t(z) \asymp 1$ and $\eta \leqslant t$, we can write

$$|g_i(t,z)|^2 \asymp \frac{1}{t} \operatorname{Im}\left(g_i(t,z)\right). \tag{2.4.24}$$

Besides we also have from (2.4.9),

$$\sum_{i=1}^{N}|g_i(t,z)|^2 \leqslant \frac{C}{t}\sum_{i=1}^{N}|g_i(t,z)| \leqslant \frac{CN}{t}\log N. \tag{2.4.25}$$

Injecting now (2.4.24) and (2.4.25) in (2.4.20), we get the bound

$$\sqrt{\frac{t}{N}}\frac{1}{N\mathrm{Im}\left(\sum_{i=1}^{N}q_k^2 g_k(t,z)\right)}\sum_{1\leqslant i<j\leqslant N}(2.4.18) \leqslant C\sqrt{\frac{t}{N}}\frac{N^{2\xi}}{N}\frac{N}{t^2}\log N \tag{2.4.26}$$

$$\leqslant \frac{N^{3\xi}}{Nt}\sqrt{\frac{N}{t}}. \tag{2.4.27}$$

Looking now at the error term (2.4.19) and injecting it in the sum (2.4.6), we obtain

$$\sqrt{\frac{t}{N}}\frac{1}{N\mathrm{Im}\left(\sum_{i=1}^{N}q_k^2 g_k(t,z)\right)}\sum_{1\leqslant i<j\leqslant N}(2.4.19) \leqslant \sqrt{\frac{t}{N}}\frac{N^{2\xi}}{N}\sum_{1\leqslant i<j\leqslant N}|g_i(t,z)g_j(t,z)|^{3/2}$$

$$\leqslant \sqrt{\frac{t}{N}}\frac{N^{2\xi}}{Nt}\left(\sum_{i=1}^{N}|g_i(t,z)|\right)^2$$

$$\leqslant N^{3\xi}\sqrt{\frac{N}{t}}. \tag{2.4.28}$$

The second case are the terms where one term is diagonal and the other is an off-diagonal term. More precisely the set $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ such that $\beta_1 = \alpha_2$ and $\beta_2 \neq \alpha_3$ or $\beta_1 \neq \alpha_2$ and $\beta_2 = \alpha_3$. Note that necessarily, in that case, $\alpha_1 = \beta_3$. For instance consider the term

$$\langle \mathbf{q}, G\mathbf{e}_{\boldsymbol{\alpha_1}}\rangle G_{\beta_1\alpha_2}G_{\beta_2\alpha_3}\langle \mathbf{e}_{\boldsymbol{\beta_3}}, G\mathbf{q}\rangle = \langle \mathbf{q}, G\mathbf{e}_i\rangle G_{jj}G_{ij}\langle \mathbf{e}_i, G\mathbf{q}\rangle. \tag{2.4.29}$$

Putting all the leading terms from Theorem 2.2.3, (2.4.7) and (2.2.7), we obtain the bound

$$\langle \mathbf{q}, G\mathbf{e}_i\rangle G_{jj}G_{ij}\langle \mathbf{e}_i, G\mathbf{q}\rangle \prec |q_i|^2|g_i(t,z)|^2|g_j(t,z)|\frac{1}{\sqrt{N\eta}}\min(|g_i(t,z)|,|g_j(t,z)|) \tag{2.4.30}$$

$$+\frac{1}{(N\eta)^{3/2}}\mathrm{Im}\left(\sum_{k=1}^{N}q_k^2 g_k(t,z)\right)\sqrt{|g_i(t,z)g_j(t,z)|}|g_j(t,z)|\min(|g_i(t,z)|,|g_j(t,z)|) \tag{2.4.31}$$

$$\leqslant N^{2\xi}\left(q_i^2|g_i(t,z)g_j(t,z)|^2 + \mathrm{Im}\left(\sum_{k=1}^{N}q_k^2 g_k(t,z)\right)|g_i(t,z)g_j(t,z)|^{3/2}\right). \tag{2.4.32}$$

Then injecting the bounds (2.4.24) and (2.4.25) in the sum of (2.4.6), one gets

$$\sqrt{\frac{t}{N}}\frac{N^{2\xi}}{N\mathrm{Im}\left(\sum_{k=1}^{N}q_k^2 g_k(t,z)\right)}\sum_{1\leqslant i<j\leqslant N}q_i^2|g_ig_j|^2 \leqslant \frac{CN^{3\xi}}{Nt}\sqrt{\frac{N}{t}} \tag{2.4.33}$$

and for the second term,

$$\sqrt{\frac{t}{N}}\frac{N^{2\xi}}{N}\sum_{1\leqslant i<j\leqslant N}|g_i(t,z)g_j(t,z)|^{3/2} \leqslant N^{3\xi}\sqrt{\frac{N}{t}}. \tag{2.4.34}$$

The final case consists of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ such that $\{\{\beta_1, \alpha_2\}, \{\beta_2, \alpha_3\}\} = \{\{i, i\}, \{j, j\}\}$. Note that, in this case, we necessarily have $\alpha_1 \neq \beta_3$. For instance, consider the term

$$\langle \mathbf{q}, G\mathbf{e}_{\boldsymbol{\alpha_1}}\rangle G_{\beta_1\alpha_2} G_{\beta_2\alpha_3} \langle \mathbf{e}_{\boldsymbol{\beta_3}}, G\mathbf{q}\rangle = \langle \mathbf{q}, G\mathbf{e}_i\rangle G_{jj} G_{ii} \langle \mathbf{e}_j, G\mathbf{q}\rangle. \tag{2.4.35}$$

Again, taking the leading terms from the local laws from Threom 2.2.3 and Corollary 2.2.5,

$$\langle \mathbf{q}, G\mathbf{e}_i\rangle G_{jj} G_{ii} \langle \mathbf{e}_j, G\mathbf{q}\rangle \prec |q_i q_j||g_i(t,z)g_j(t,z)|^2 + N^{\xi}|g_i(t,z)g_j(t,z)|^{3/2}\mathrm{Im}\left(\sum_{k=1}^N q_k^2 g_k(t,z)\right). \tag{2.4.36}$$

Then using similar bounds as the first case one gets

$$\sqrt{\frac{t}{N}} \frac{1}{N\mathrm{Im}\left(\sum_{k=1}^N q_k^2 g_k(t,z)\right)} \sum_{1\leqslant i<j\leqslant N} (2.4.36) \prec N^{2\xi}\left(\frac{1}{Nt}\sqrt{\frac{N}{t}} + \sqrt{\frac{N}{t}}\right). \tag{2.4.37}$$

Finally, putting together (2.4.27), (2.4.28), (2.4.33), (2.4.34) and (2.4.37), we get the bound, for $\eta = N^{-1+\xi}$.

$$\sqrt{\frac{t}{N}} \frac{1}{N} \sum_{1\leqslant i<j\leqslant N} \partial_{ij}^3 F(\tilde{H}_s) \prec N^{3\xi}\sqrt{\frac{N}{t}}. \tag{2.4.38}$$

In order to get a bound for η below microscopic scales, we can use the following inequality, for any $y \leqslant \eta$, which can be found in [EYY12a, Section 8],

$$|\langle \mathbf{v}, G(E + \mathrm{i}y)\mathbf{w}\rangle| \leqslant C \log N \frac{\eta}{y} \mathrm{Im}\langle \mathbf{v}, G(E + i\eta)\mathbf{w}\rangle.$$

This bounds allows us to get below microscopic scales for F and its derivatives since they only involves such quantity as $\langle \mathbf{v}, G(E + \mathrm{i}y)\mathbf{w}\rangle$. Thus, uniformly in $E \in \mathcal{I}_r^\kappa$ and $N^{-1-\xi} \leqslant \eta \leqslant t$, we have

$$M = \mathcal{O}\left(N^{5\xi}\sqrt{\frac{N}{t}}\right). \tag{2.4.39}$$

Using now Lemma 2.4.1, we can make W_t undergo the dynamics $\tilde{H}_s$ up to a time $\tau \ll N^{-5\xi}\sqrt{\frac{t}{N}}$ with ξ arbitrarily small in order to get the right bound.

For a product of resolvent entries, one can do similar computations and bounds. Indeed consider $m \geqslant 0$, and

$$F(\tilde{H}_s) = \prod_{k=1}^m F_k(\tilde{H}_s) \quad \text{with} \quad F_k(\tilde{H}_s) = \langle \mathbf{q}, G(z_k)\mathbf{q}\rangle,$$

then one can write the third derivative of F as (2.4.13) and using the fact that

$$\partial_{ij} F_k = -\sum_{\{\alpha,\beta\}=\{i,j\}} \langle \mathbf{q}, G\mathbf{e}_\alpha\rangle\langle \mathbf{e}_\beta, G\mathbf{q}\rangle, \tag{2.4.40}$$

$$\partial_{ij}^2 F_k = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} \langle \mathbf{q}, G\mathbf{e}_{\alpha_1}\rangle G_{\beta_1,\alpha_2}\langle \mathbf{e}_{\beta_2}, G\mathbf{q}\rangle \tag{2.4.41}$$

where $\{\alpha_i, \beta_i\} = \{i, j\}$ and using the same type of bounds as for (2.4.38), we obtain the result (2.4.5) since the extension to any smooth function with polynomial growth is also clear. □

2.4.2. Reverse heat flow

In Subsection 2.3.2, we showed Theorem 2.3.8, which corresponds to our main result for the matrix $H_\tau = W_t + \sqrt{\tau}\text{GOE}$ for $N^{-1} \ll \tau \ll t$ with a general Wigner matrix W in the definition of W_t. Thus, the overwhelming probability bound holds for the eigenvectors of this matrix H_τ giving us a strong form of quantum unique ergodicity for the deformed Gaussian divisible ensemble. In order to remove the small Gaussian component in the matrix, we will use the reverse heat flow technique from [EPR+10, ESY11] which allows us to obtain an error as small as we want in total variation between two matrix ensembles. In order to use this technique, we need the smoothness assumption on the matrix W given by Definition 2.1.6. We first introduce some notation for this section.

As before, we let ν denote the distribution of the entries of W, and φ denote the density of ν with respect to ρ, the Gaussian distribution with mean zero and variance one, that is, $\mathrm{d}\nu = \varphi\mathrm{d}\varrho$. The reverse heat flow technique gives the existence of a probability distribution $\tilde{\nu}_s$ for any s small enough such that making $\tilde{\nu}_s$ undergo the Ornstein-Uhlenbeck process of generator

$$A := \frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{x}{2}\frac{\partial}{\partial x}$$

approaches the distribution ν in total variation.

This process on all the matrix entries induces the Dyson Brownian motion process on the eigenvalues. Thus the following proposition tells us that there exists a distribution of a matrix from the Gaussian divisible process of the form

$$\widetilde{W}_s = \sqrt{1-s}\,\widetilde{W} + \sqrt{s}\text{GOE}$$

that approximates as close as polynomially possible a smooth Wigner matrix W. The precise statement is written in the following proposition.

Proposition 2.4.3 ([ESY11])**.** *Let K be a positive integer and $\nu = \varphi\varrho$ a distribution smooth in the sense that it follows the conditions (ii) and (iii) of Definition 2.1.6. Then there exists s_K a small positive constant depending on K such that for any $0 < s \leqslant s_K$, there exists a probability density ψ_s with mean zero and variance one such that we have the inequality*

$$\int \left|e^{sA}\psi_s - \varphi\right| \mathrm{d}\varrho \leqslant Cs^K \tag{2.4.42}$$

for some positive constant C depending only on K. Besides we also have the inequality for the joint probability of all matrix entries in the following sense,

$$\int \left| e^{sA^{\otimes N^2}} \psi_s^{\otimes N^2} - \varphi^{\otimes N^2} \right| \mathrm{d}\varrho \leqslant CN^2 s^K \tag{2.4.43}$$

Now, see that this proposition holds for any fixed K so that, taking $s = N^{-\varepsilon}$ for some small ε we can choose a large K only depending on ε (and not on N) so that we can obtain any polynomial bound between the two matrix ensembles. This property allows us to get overwhelming probability bounds on the eigenvectors since the total variation distance of the distribution of the eigenvector entries is smaller than the total variation distance between the joint probability of the matrix entries.

2.5. Proofs of main results

Now that we have the result for the Gaussian divisible ensemble H_τ with $N^{-1} \ll \tau \ll t$ by Section 2.3, combining it with the continuity argument from the last subsection, we are able to prove Theorem 2.1.3 and Corollary 2.1.5. These two results are a consequence of the following proposition showing the convergence of moments for the eigenvectors of W_t.

Proposition 2.5.1.
Let $\kappa \in (0,1)$ and m an integer, for a set of indices $I \subset \mathcal{A}_r^\kappa$, such that $|I| = m$, we have for any deterministic unit vector $\mathbf{q} = \mathbf{q}_N$,

$$\mathbb{E}\left[P\left(\left(\frac{N}{\sigma_t(\mathbf{q},k)^2}|\langle \mathbf{q}, u_k\rangle|^2\right)_{k\in I}\right)\right] \xrightarrow[N\longrightarrow\infty]{} \mathbb{E}\left[P\left((\mathcal{N}_k^2)_{k=1}^m\right)\right] \tag{2.5.1}$$

with $(\mathcal{N}_k)_k$ a family of independent normal random variables.

See now the proof of Theorem 2.1.3 and Corollary 2.1.5 given by Proposition 2.3.4.

Proof of Theorem 2.1.3. Proposition 2.5.1 exactly gives us that the joint moments of the renormalized eigenvectors converge to those of independent normal random variables which is the result of Theorem 2.1.3. □

Proof of Corollary 2.1.5. By Proposition 2.3.4 and Corollary 2.4.2, we have the following inequality, for some $\varepsilon > 0$,

$$\mathbb{E}\left[|u_k(\alpha)^2|\right] = \frac{1}{N}\sigma_t^2(\alpha,k) + \mathcal{O}\left(\frac{N^{-\varepsilon}}{Nt}\right). \tag{2.5.2}$$

By Markov's inequality, we can write

$$\begin{aligned}\mathbb{P}\left(\frac{Nt}{|A|}\left|\sum_{\alpha\in A}|u_k(\alpha)|^2 - \frac{1}{N}\sum_{\alpha\in A}\sigma_t^2(\alpha,k)\right| > c\right) &\leqslant \frac{N^2t^2}{c^2|A|^2}\mathbb{E}\left[\left|\sum_{\alpha\in A}|u_k(\alpha)|^2 - \frac{1}{N}\sum_{\alpha\in A}\sigma_t^2(\alpha,k)\right|^2\right],\\ &\leqslant \frac{N^2t^2}{c^2|A|^2}\left(\mathfrak{A} - 2\mathfrak{B} + \mathfrak{C}\right) \end{aligned}\tag{2.5.3}$$

We now need to evaluate the three terms in the last inequality using (2.5.2), first we have

$$\mathfrak{A} := \mathbb{E}\left[\left(\sum_{\alpha\in A}|u_k(\alpha)|^2\right)^2\right] = \frac{1}{N^2}\left(\sum_{\alpha\in A}\sigma_t^2(\alpha,k)\right)^2 + \frac{2}{N^2}\sum_{\alpha\in A}\sigma_t^4(\alpha,k) + \mathcal{O}\left(\frac{N^{-\varepsilon}|A|^2}{N^2t^2}\right).$$

Likewise,

$$\mathfrak{B} := \frac{1}{N}\sum_{\beta\in A}\sigma_t^2(\beta,N)\mathbb{E}\left[\sum_{\alpha\in A}|u_k(\alpha)|^2\right] = \frac{1}{N^2}\sum_{\alpha,\beta\in A}\sigma_t^2(\alpha,k)\sigma_t^2(\beta,k) + \mathcal{O}\left(\frac{N^{-\varepsilon}|A|^2}{N^2t^2}\right).$$

Finally, $\mathfrak{C}$ is just a deterministic term,

$$\mathfrak{C} := \left(\frac{1}{N}\sum_{\alpha\in A}\sigma_t^2(\alpha,k)\right)^2.$$

Putting all three terms together, we get

$$\mathfrak{A} - 2\mathfrak{B} + \mathfrak{C} = \frac{2}{N^2}\sum_{\alpha\in A}\sigma_t^4(\alpha,k) + \mathcal{O}\left(\frac{N^{-\varepsilon}|A|^2}{N^2t^2}\right) \leqslant C\left(\frac{|A|}{N^2t^2} + \frac{N^{-\varepsilon}|A|^2}{N^2t^2}\right) \tag{2.5.4}$$

The claim then follows from injecting the last inequality in (2.5.3). □

We finish now with the proof of Proposition 2.5.1.

Proof of Proposition 2.5.1. By Corollary 2.4.2, we know that for some $\tau \in \mathcal{T}'_{\mathfrak{a}}$ there exists $\varepsilon > 0$ such that

$$\left| \mathbb{E}\left[P\left(\left(\frac{N}{\sigma_t(\mathbf{q},k)^2} |\langle \mathbf{q}, u_k \rangle|^2 \right)_{k\in I} \right) \right] - \mathbb{E}\left[P\left(\left(\frac{N}{\sigma_t(\mathbf{q},k)^2} |\langle \mathbf{q}, u_k^{\tilde{H}_\tau} \rangle|^2 \right)_{k\in I} \right) \right] \right| \leqslant N^{-\varepsilon} \tag{2.5.5}$$

and by Proposition 2.3.4, we know that, recalling the definition of H_s, for some $\tau' \ll t$ there exists a $\varepsilon' > 0$ such that

$$\left| \mathbb{E}\left[P\left(\left(\frac{N}{\sigma_t(\mathbf{q},k)^2)} |\langle \mathbf{q}, u_k^{H_{\tau'}} \rangle|^2 \right)_{k\in I} \right) \right] - \mathbb{E}\left[P\left(\left(\mathcal{N}_k^2 \right)_{k=1}^m \right) \right] \right| \leqslant N^{-\varepsilon'}. \tag{2.5.6}$$

Now, we need to see that $\tilde{H}_\tau$ defined in (2.4.1) has the same law as $H_{\tau'}$ for some $\tau' \ll t$. Note that we can write the law of the entries of $\tilde{H}_\tau$ as

$$\tilde{H}_{ij}(\tau) \overset{d}{=} D_{ij} + e^{-\frac{\tau}{2t}} \sqrt{t} W_{ij} + \sqrt{t\left(1 - e^{-\frac{\tau}{t}}\right)} \frac{1}{\sqrt{N}} \mathcal{N}^{(ij)}, \tag{2.5.7}$$

where $\left(\mathcal{N}^{(ij)}\right)_{i\leqslant j}$ is a family of independent standard Gaussian random variables. Doing the scaling

$$\begin{aligned} t' &= t e^{-\tau/t} = \mathcal{O}(t), \\ \tau' &= \sqrt{t\left(1 - e^{-\frac{\tau}{t}}\right)} = \mathcal{O}(\tau) \end{aligned} \tag{2.5.8}$$

one can write

$$\tilde{H}_\tau = D + \sqrt{t'} W + \sqrt{\tau'} \mathrm{GOE}.$$

Finally, we can apply Proposition 2.3.4 to $\tilde{H}$ so that (2.5.6) applies and combining it with (2.5.5) we get the convergence of moments for the eigenvectors of W_t. □

Combining Theorem 2.3.8 and Proposition 2.3.4, we are now able to prove Theorem 2.1.7.

Proof of Theorem 2.1.7. Let ε and D two positive constants and consider $s = \tau/t,$. There exists then a large K, which does not depend on N, such that by Proposition 2.4.3 there exists a matrix $\widetilde{W}$ such that the total variation distance between the distribution of W and $\sqrt{1-s}\widetilde{W} + \sqrt{Ns}\mathrm{GOE}$ is smaller than N^{-D}.
Denote $u_1, \dots, u_N$ the L^2-normalized eigenvectors of $W_t = D + \sqrt{t}W$ and $\tilde{u}_1, \dots, \tilde{u}_N$ the normalized eigenvectors of $\widetilde{W}_t(s) = D + \sqrt{t(1-s)}\,\widetilde{W} + \sqrt{ts}\mathrm{GOE}$. Now, since we have in the overwhelming probability bound (2.1.13) the N^ε degree of liberty, we can do the scaling $t' = \sqrt{t(1-s)}$ as $s \ll 1$ and still get (2.1.13) for the deformed Gaussian divisible ensemble $\widetilde{W}_t(s)$, thus one can write

$$\begin{aligned} &\mathbb{P}\left(\left| \sum_{\alpha\in I} \left(u_k(\alpha)^2 - \frac{1}{N}\sigma_t^2(\alpha,k) \right) \right| \geqslant N^\varepsilon \Xi(\tau) \right) \\ &\quad\leqslant \mathbb{P}\left(\left| \sum_{\alpha\in I} \left(\tilde{u}_k(\alpha)^2 - \frac{1}{N}\sigma_t^2(\alpha,k) \right) \right| \geqslant N^{\varepsilon/2} \Xi(\tau) \right) + \mathbb{P}\left(\left| \sum_{\alpha\in I} \left(u_k(\alpha)^2 - \tilde{u}_k(\alpha) \right) \right| \geqslant N^{\varepsilon/2} \Xi(\tau) \right) \\ &\quad\leqslant N^{-D} \end{aligned}$$

where for the last inequality we used the quantum unique ergodicity proved in Theorem 2.3.8 for the deformed Gaussian divisible ensemble of which $\tilde{\mathbf{u}}$ are the eigenvectors and Proposition 2.4.3 in order.

Now, in order to get the error Ξ we now need to optimize the error

$$\Xi(\tau_0) = \frac{\widehat{I}}{\sqrt{N\tau_0}} + \widehat{I}\frac{\tau_0}{t} = \frac{\widehat{I}}{(Nt)^{1/3}} = \Xi \quad \text{with} \quad \tau_0 = \left(\frac{t^2}{N}\right)^{1/3}.$$

We can do the same thing for the quantity $\sum_{\alpha\in I} u_k(\alpha)u_l(\alpha)$ and get the final result. □

Chapter 3

Fermionic observable for the eigenvector moment flow and fluctuations of eigenvectors of random matrices

This chapter is based on the article [Ben19]

3.1. Introduction

The purpose of this chapter is to study the following quantity: if we denote W a symmetric random matrix with independent entries up to the symmetry and consider $\lambda_1 \leqslant \cdots \leqslant \lambda_N$ its ordered eigenvalues and $(u_1, \ldots, u_N)$ the associated eigenvectors, we then want to look at, for a fixed deterministic sequence of indices k and $I \subset [\![1, N]\!]$ a N-dependent set of indices,

$$\frac{1}{\sqrt{2|I|}} \sum_{\alpha \in I} \left(N u_k(\alpha)^2 - 1 \right) =: \tilde{p}_{kk}. \tag{3.1.1}$$

It has been shown that eigenvectors entries are normally distributed asymptotically, it was also shown that for a fixed eigenvector, its entries are asymptotically independent in the sense of moments, thus we would expect that this random variables converges in some sense to a Gaussian random variable in light of the central limit theorem.

While the earlier studies of moments of eigenvector give some information of such fluctuations, it is not yet possible to study joint moments between different entries of different eigenvectors such as $u_k(1)u_\ell(2)$ and thus the correlations between fluctuations for two fixed distinct eigenvectors were not computable. The key new ingredient in this paper is the exhibition of a new moment observable of fluctuations that follows the eigenvector moment flow, a dynamics introduced in [BY17]. In [BYY18], another observable involving fluctuations of eigenvectors such as (3.1.1) and eigenvectors overlaps, for $k \neq \ell$,

$$\frac{1}{\sqrt{|I|}} \sum_{\alpha \in I} N u_k(\alpha) u_\ell(\alpha) =: \tilde{p}_{k\ell} \tag{3.1.2}$$

was introduced. By gaining information through the observable from [BYY18] and the one from this chapter we are able to obtain the Gaussianity and decorrelation of fluctuations of type (3.1.1). Even if we expect asymptotic Gaussianity of the mixed overlap (3.1.2), we are able here to only obtain its asymptotic variance. Indeed, the mixed overlap contains information on two distinct eigenvectors and thus obtaining joint moments becomes harder. Note that while our main theorem gives the asymptotic decorrelation of fluctuations and the variance of the mixed overlap, we actually obtain more information given by different sum of joint moments.

The study of fluctuations of eigenvectors were first on the global scale, in the sense that they involved a macroscopic number of eigenvectors. The first result comes from the eigenvectors of large sample covariance matrices in [Sil90] where it was seen that some form of fluctuations involving all eigenvectors converges weakly to the Brownian bridge. Also, in the case of Gaussian matrices, say symmetric matrices, it was first shown in [DMR12] that the process

$$\left(\frac{1}{\sqrt{2}}\sum_{\substack{1\leqslant i\leqslant Ns\\ 1\leqslant j\leqslant Nt}}\left(|u_i(j)|^2-\frac{1}{N}\right)\right)_{(s,t)\in[0,1]^2}$$

converges to a bivariate Brownian bridge. This result was then generalized to more general model of matrices such as Wigner matrices in [BG12]. Another form of convergence to the Brownian bridge for Wigner matrices was also proved in [BPZ14].

It is however difficult to obtain information on a finite number of eigenvectors and this paper aims to reach a better understanding of correlations between a couple of eigenvectors.

3.1.1. Main results

We will study eigenvectors of generalized symmetric Wigner matrices of size N given by the following definition.

Definition 3.1.1. Let W be a $N\times N$ symmetric matrix such that its entries $(w_{ij})_{1\leqslant i\leqslant j\leqslant N}$ are centered independent random variables of variance s_{ij} such that there exists two positive constants c and C such that

$$\frac{c}{N}\leqslant s_{ij}\leqslant\frac{C}{N}\quad\text{for all } i,j\quad\text{and}\quad\sum_{i,j=1}^{N}s_{ij}=1\quad\text{for all } j.$$

We will also assume that the matrix entries have all finite moments in the following sense, for every $p\in\mathbb{N}$ there exists a constant μ_p independent of N such that

$$\mathbb{E}\left[\sqrt{N}w_{ij}^p\right]\leqslant\mu_p.$$

The global statistics of eigenvalues for this model are given by the semicircle law, namely if we consider $\lambda_1\leqslant\cdots\leqslant\lambda_N$ the eigenvalues of W, we have the following almost sure convergence

$$\frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i}\xrightarrow[N\to\infty]{}\rho_{sc}\quad\text{with}\quad\rho_{sc}(x)\mathrm{d}x=\sqrt{4-x^2}\mathrm{d}x. \tag{3.1.3}$$

For the local eigenvalues statistics bulk universality has been proved in [EYY12a] and edge universality [BEY14a]. While this concerns eigenvalue statistics, the local behavior of eigenvectors was first considered in the case of Wigner matrices in [TV12b] with a matching condition. For generalized Wigner matrices, it was shown in [KY13b] that if two matrix ensembles have the same four moments, the bulk and edge eigenvectors have asymptotically the same distribution. The moment condition was removed in [BY17] using a dynamical proof to show asymptotic Gaussianity of projections of eigenvectors as stated in the following theorem.

Theorem 3.1.2 ([BY17])**.** *Let $I \subset [\![1, N]\!]$ a set of indices such that $|I| = m$, then for any sequence of deterministic vector $\mathbf{q}_N$ such that $\|\mathbf{q}_N\|_2 = 1$*

$$\left(|\sqrt{N}\langle \mathbf{q}, u_k\rangle|\right)_{k\in I} \xrightarrow[N\to\infty]{} (|\mathcal{N}_i|)_{i=1}^m \tag{3.1.4}$$

with $(\mathcal{N}_i)$ is family of centered unit variance independent Gaussian random variables and the convergece holds in the sense of moments.

We also have the following convergence of moments, for $k \in [\![1, N]\!]$,

$$\left(\sqrt{N}u_k(\alpha)\right)_{\alpha\in I} \xrightarrow[N\to\infty]{} (\mathcal{N}_i)_{i=1}^m$$

where this convergence holds modulo the phase choice for the eigenvector u_k.

Actually, it is possible to use this theorem in order to study the fluctuations $\tilde{p}_{kk}$ as in (3.1.1). Indeed, as we can only hope using our technique to obtain a convergence of moments, Theorem 3.1.2 gives us all the information on moments for a single fixed eigenvector which gives us the following corollary.

Corollary 3.1.3. *Let $\varepsilon > 0$, let $k_N \in [\![1, N]\!]$ be a sequence of deterministic indices and $I \subset [\![1, N]\!]$ such that $N^{\varepsilon} \leqslant |I| \leqslant N^{1-\varepsilon}$, we have the following convergence in the sense of moments,*

$$\frac{1}{\sqrt{2|I|}}\sum_{\alpha\in I}\left(Nu_k(\alpha)^2 - 1\right) \xrightarrow[N\to\infty]{} \mathcal{N}(0,1)$$

Thus the main contribution of this paper does not concern the Gaussianity of fluctuations of eigenvectors but the correlations between fluctuations.

Theorem 3.1.4. *Let ε be a (small) positive constant. Consider $(\mathbf{q}_i)_{i\in I}$ a family of unit orthogonal vectors. Consider k_N and ℓ_N two distinct deterministic sequences of indices in $[\![1, N]\!]$, let I be a N-dependent set of indices such that $|I| \leqslant N^{1/2-\varepsilon}$ then there exists a $\delta > 0$ such that*

$$\mathbb{E}\left[\frac{N^2}{2|I|}\left(\sum_{\alpha\in I}\langle \mathbf{q}_\alpha, u_k\rangle^2 - \frac{|I|}{N}\right)\left(\sum_{\alpha\in I}\langle \mathbf{q}_\alpha, u_\ell\rangle^2 - \frac{|I|}{N}\right)\right] \leqslant N^{-\delta}. \tag{3.1.5}$$

Besides, we also have that

$$\left|\mathbb{E}\left[\left(\frac{N}{\sqrt{|I|}}\sum_{\alpha\in I}\langle \mathbf{q}_\alpha, u_k\rangle\langle \mathbf{q}_\alpha, u_\ell\rangle\right)^2\right] - 1\right| \leqslant N^{-\delta}. \tag{3.1.6}$$

Remark 3.1.5. The condition on the cardinality of the set I is technical and the result should hold for any growing set I. It comes from the difficulty to bound a technical term involving the eigenvectors and the resolvent. Indeed, the optimal local law and the complete delocalization are not strong enough here to obtain an optimal result, this difficulty more precisely comes from the dependence between the resolvent and eigenvectors.

Our result gives decorrelations of fluctuations of fixed eigenvectors, thus giving a bigger understanding of the whole transfer matrix. Indeed, in order to obtain Theorem 3.1.4, we need to study statistics involving different entries of different eigenvectors for which Theorem 3.1.2 only gives partial information. As for the mixed overlap $\tilde{p}_{k\ell}$, we can not obtain Gaussianity, either with Theorem 3.1.2 or with the additional information we obtain in this paper, but we can compute its variance which was not computable before. These results give us a deeper understanding on the convergence to the eigenvector matrix to a Haar-distributed on a local scale since we consider finitely many eigenvectors. As a corollary, we obtain a probability bound on the mixed overlap (3.1.2).

Corollary 3.1.6. *For any $\varepsilon > 0$ there exist a $\delta > 0$ such that*

$$\mathbb{P}\left(\left|\sum_{\alpha\in I} u_k(\alpha)u_\ell(\alpha)\right| \geqslant N^{\varepsilon}\frac{\sqrt{|I|}}{N}\right) \leqslant N^{-\delta}. \tag{3.1.7}$$

Proof. The proof is direct by combining 3.1.4 with the Bienaymé-Chebyshev inequality. □

3.1.2. Method of proof

The proof is based upon the three-step strategy used to prove universality of eigenvalues and eigenvectors of random matrices first introduced in[ERS+10, ESY11] (see [EY17, BB19] for recent writings on the subject).

The first step of the strategy is a local law, a local version of the convergence (3.1.3). While the convergence (3.1.3) gives information on the global statistics of eigenvalues, we will need a more local form of this convergence as the consideration of the fluctuations involve a single or two eigenvectors but not the whole spectrum. A local law consists of a high-probability bound on the resolvent of our generalized Wigner matrix controlling it down to the optimal scale $N^{-1+\varepsilon}$ for any $\varepsilon > 0$.

Define the resolvent G and the Stieltjes transform of the semicircle law m to be for $z \in \mathbb{C}$ with $\operatorname{Im} z > 0$

$$G(z) = \sum_{k=1}^{N} \frac{|u_k\rangle\langle u_k|}{\lambda_k - z} \quad \text{and} \quad m(z) = \int \frac{\mathrm{d}\rho_{sc}(x)}{x - z} = \frac{-z + \sqrt{z^2 - 4}}{2} \tag{3.1.8}$$

where the choice of the square root is given by m being holomorphic in the upper half plane and $m(z) \to 0$ as $z \to \infty$. We will need two forms of local law, one will be an averaged local law on the Stieltjes transform of the empirical spectral distribution of W, $s(z) = N^{-1}\operatorname{Tr} G(z)$, the other will be on the resolvent as a quadratic form, also called an isotropic local law.

Theorem 3.1.7 ([EYY12b, BEK+14])**.** *Consider the following spectral domain, for any (small) $\omega > 0$,*

$$\mathcal{D}_\omega = \left\{z = E + \mathrm{i}\eta,\, |E| \leqslant \omega^{-1},\, N^{-1+\omega} \leqslant \eta \leqslant \omega^{-1}\right\},$$

then we have for any positive ε and $D > 0$,

$$\sup_{z\in\mathcal{D}_\omega} \mathbb{P}\left(|s(z) - m(z)| \geqslant \frac{N^{\varepsilon}}{N\eta}\right) \leqslant N^{-D}, \tag{3.1.9}$$

and for any vector $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^N$, for any positive ε and D,

$$\sup_{z\in\mathcal{D}_\omega} \mathbb{P}\left(|\langle \mathbf{v}, G(z)\mathbf{w}\rangle - m(z)\langle \mathbf{v}, \mathbf{w}\rangle| \geqslant N^{\varepsilon}\langle \mathbf{v},]\rangle\mathbf{w}\left(\sqrt{\frac{\operatorname{Im} m(z)}{N\eta}} + \frac{1}{N\eta}\right)\right) \leqslant N^{-D} \tag{3.1.10}$$

As a corollary of this theorem, one obtains the complete delocalization of eigenvectors as an overwhelming probability bound. We will need this optimal estimate (up to logarithmic corrections) in order to control eigenvectors.

Corollary 3.1.8. *Let $k \in [\![1, N]\!]$ and $\mathbf{q} \in \mathbb{R}^N$ such that $\|\mathbf{q}\|_2 = 1$, we have, for any D and any ε positive*

$$\mathbb{P}\left(|\langle \mathbf{q}, u_k\rangle| \geqslant \frac{N^{\varepsilon}}{\sqrt{N}}\right) \leqslant N^{-D}. \tag{3.1.11}$$

The second step of the method consists of the relaxation of our original matrix through the Dyson Brownian motion. It consists of an Ornstein-Uhlenbeck process on the space of symmetric matrices. The main characteristics of this dynamics for our problem is the explicit dynamics of eigenvectors along the process and the short-time to relaxation to the equilibrium measure. For eigenvectors, this measure consists in the Haar measure on orthogonal matrices so that asymptotically we obtain Gaussianity and independence of eigenvectors entries. We will now give our definition for the Dyson Brownian motion.

Definition 3.1.9. Let B be a symmetric $N \times N$ matrix such that B_{ij} for $i < j$ and $B_{ii}/\sqrt{2}$ are standard independent brownian motions. The symmetric Dyson Brownian motion is given by the stochastic differential equation

$$\mathrm{d}H_s = \frac{1}{\sqrt{N}}\mathrm{d}B_s - \frac{1}{2}H_s \mathrm{d}t \tag{3.1.12}$$

Besides, it induces the following dynamics on eigenvalues and eigenvectors: the eigenvalues and eigenvectors of H_s have the same distribution as the solutions of this coupled dynamics,

$$\mathrm{d}\lambda_k = \frac{\mathrm{d}\widetilde{B}_{kk}}{\sqrt{N}} + \left(\frac{1}{N}\sum_{\ell \neq k}\frac{1}{\lambda_k - \lambda_\ell} - \frac{\lambda_k}{2}\right)\mathrm{d}s, \tag{3.1.13}$$

$$\mathrm{d}u_k = \frac{1}{\sqrt{N}}\sum_{\ell \neq k}\frac{\mathrm{d}\widetilde{B}_{k\ell}}{\lambda_k - \lambda_\ell}u_\ell - \frac{1}{2N}\sum_{\ell \neq k}\frac{\mathrm{d}s}{(\lambda_k - \lambda_\ell)^2}u_k \tag{3.1.14}$$

where $\widetilde{B}$ is an independent copy of B.

The explicit form of the dynamics of eigenvectors (3.1.14) is one of the key ingredient in our study. Indeed, while this dynamics by itself is too hard to analyze, one can look at some observables on eigenvector moments which will follow a simpler parabolic equation. This equation, the eigenvector moment flow, was first considered in [BY17] to study eigenvectors of generalized Wigner matrices, then in [BHY17] to study sparse matrices and in [Ben17] to look at the behavior of eigenvectors for deformed Wigner matrices. We will give here a new observable, called the Fermionic observable, which will follow a similar equation.

This Fermionic observable can be stated as an observable directly on the fluctuations (3.1.1) and (3.1.2). Consider $\mathbf{u}_s = (u_1^s, \dots, u_N^s)$ the L^2- normalized eigenvectors of H_s as in (3.1.12) associated to its ordered eigenvalues $\boldsymbol{\lambda}_s = (\lambda_1(s) \leqslant \dots \leqslant \lambda_N(s))$ and a family of deterministic fixed vectors $(\mathbf{q}_i)_{i\in I}$ non necesarily orthogonal. We will now work with the non-normalized fluctuations, denote for $k \neq \ell$ in $[\![1, N]\!]$,

$$p_{kk}(s) = \sum_{\alpha\in I}\langle \mathbf{q}_i, u_k^s\rangle^2 - \frac{|I|}{N} \quad \text{and} \quad p_{k\ell}(s) = \sum_{\alpha \in I}\langle \mathbf{q}_i, u_k^s\rangle\langle \mathbf{q}_i, u_\ell^s\rangle. \tag{3.1.15}$$

For $\mathbf{k} = (k_1, \dots, k_n)$ with k_i pairwise distinct indices in $[\![1, N]\!]$, we will define the following $n \times n$ (symmetric) matrix of fluctuations

$$P_s(\mathbf{k}) = \begin{pmatrix} p_{k_1k_1}(s) & p_{k_1k_2}(s) & \dots & p_{k_1k_n}(s) \\ \vdots & \vdots & \dots & \vdots \\ p_{k_nk_1}(s) & p_{k_nk_2}(s) & \dots & p_{k_nk_n}(s) \end{pmatrix}. \tag{3.1.16}$$

Then our Fermionic observable consists in the expectation of the determinant of our matrix of fluctuations,

$$f_s^{\mathrm{Fer}}(\mathbf{k}) = \mathbb{E}\left[\det P_s(\mathbf{k}) | \boldsymbol{\lambda}\right]. \tag{3.1.17}$$

We will use the following notation in order to describe the dynamics followed by f_{Fer}, it consists of replacing the i-th coordinated by another indices, for $\ell \notin \{k_1, \dots, k_n\}$ we will write

$$\mathbf{k}^i(\ell) = (k_1, \dots, k_{i-1}, \ell, k_{i+1}, \dots, k_n) \quad \text{and} \quad |\mathbf{k}| = |\mathbf{k}^i(\ell)| = n$$

and we now gives the flow that f_s^{Fer} undergoes.

Theorem 3.1.10. *Let $(\mathbf{u}, \boldsymbol{\lambda})$ be the solution to the coupled flows as in Definition 3.1.9 and let f_s^{Fer} be as in (3.1.17), it satisfies the following equation, for $\mathbf{k}$ a pairwise distinct set of indices such that $|\mathbf{k}| = n$,*

$$\partial_s f_s^{\mathrm{Fer}}(\mathbf{k}) = 2 \sum_{i=1}^{n} \sum_{\substack{\ell \in [\![1,N]\!] \\ \ell \notin \{k_1,\dots,k_n\}}} \frac{f_s^{\mathrm{Fer}}(\mathbf{k}^i(\ell)) - f_s^{\mathrm{Fer}}(\mathbf{k})}{N(\lambda_{k_i} - \lambda_\ell)^2} \tag{3.1.18}$$

Remark 3.1.11. We call this observable Fermionic by comparing to the observable in [BY17] or in [BYY18]. Indeed this dynamics is very similar, if we take the point of view of a multi-particle random walk in a random environment, the only difference is that we can only consider configurations of particles with at most one particle on each site, also see that the jump of the particles can only be on an empty site so that we have a form of an exclusion principle for the particles.

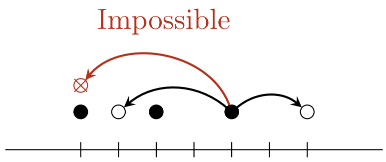


Figure 3.1: Multi-particle random walk representation of (3.1.18)

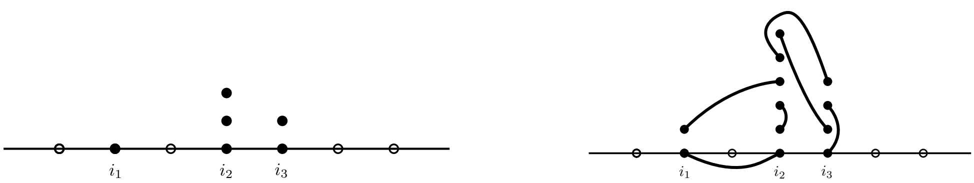
We will now develop more on the Bosonic observable introduced in [BYY18]. Indeed, since we want look at both edge and bulk eigenvectors, we will need to prove an a priori bound on the overlap p_{kk} and $p_{k\ell}$ at the edge. Now for $\boldsymbol{\eta} : [\![1, N]\!] \to \mathbb{N}$ a configuration of n particles, define the following set of vertices

$$\mathcal{V}_{\boldsymbol{\eta}} = \{(k, \alpha),\ 1 \leqslant k \leqslant N,\ 1 \leqslant \alpha \leqslant 2\eta_i\}.$$

Consider now $\mathcal{G}_{\boldsymbol{\eta}}$ the set of perfect matchings on $\mathcal{V}_{\boldsymbol{\eta}}$. We will denote such a graph $G = (\mathcal{V}_{\boldsymbol{\eta}}, \mathcal{E}(G))$. Now for $e \in \mathcal{E}(G)$, write $e = \{(k, \alpha), (\ell, \beta)\}$ and define $p(e) = p_{k\ell}$, $P(G) = \prod_{e \in \mathcal{E}(G)} p(e)$ and finally

$$f_s^{\mathrm{Bos}}(\boldsymbol{\eta}) = \frac{1}{\mathcal{M}(\boldsymbol{\eta})} \mathbb{E}\left[\sum_{G \in \mathcal{G}_{\boldsymbol{\eta}}} P(G) \middle| \boldsymbol{\lambda} \right] \tag{3.1.19}$$

where $\mathcal{M}(\boldsymbol{\eta}) = \prod_{i=1}^{N} (2\eta_i)!!$, with $(2m)!!$ being the number of perfect matchings of the complete graph on $2m$ vertices. Note that this quantity depend on the eigenvalues trajectories $\boldsymbol{\lambda}$.



(a) A configuration $\boldsymbol{\eta}$ of 6 particles

(b) An example of a perfect matching $G \in \mathcal{G}_{\boldsymbol{\eta}}$ with $P(G) = p_{i_1 i_2}^2 p_{i_2 i_2} p_{i_2 i_3}^2 p_{i_3 i_3}$

The previous quantity follows the usual eigenvector moment flow.

Theorem 3.1.12 ([BYY18])**.** *Suppose that $\boldsymbol{u}$ is the solution of the Dyson vector flow* (3.1.14) *and* $f_s^{\mathrm{Bos}}(\boldsymbol{\eta})$ *is given by* (3.1.19)*. Then it satisfies the equation*

$$\partial_s f_s^{\mathrm{Bos}}(\boldsymbol{\eta}) = \frac{1}{N}\sum_{i\neq j} \frac{2\eta_i(1+2\eta_j)\left(f_s^{\mathrm{Bos}}(\boldsymbol{\eta}^{i,j}) - f_s^{\mathrm{Bos}}(\boldsymbol{\eta})\right)}{(\lambda_i-\lambda_j)^2}. \tag{3.1.20}$$

where $\boldsymbol{\eta}^{i,j}$ is the configuration obtained by moving a particle from the site i to the site j.

Another possible representation comes from a Wick theorem on the original observable from [BY17]. Let $\mathbf{q} = \mathbf{q}^{(1)} + \mathrm{i}\sqrt{\frac{|I|}{N}}\mathbf{q}^{(2)}$ be a linear combination of two Gaussian vectors given by the following: consider $(\mathcal{N}_\alpha)_{\alpha\in I}$ and $(\mathcal{N}'_\alpha)_{\alpha=1}^N$ two independent families of independent centered Gaussian with unit variance and define for $\alpha \in [\![1,N]\!]$,

$$\mathbf{q}^{(1)}_\alpha = \mathcal{N}_\alpha \mathbb{1}_{\alpha\in I} \quad \text{and} \quad \mathbf{q}^{(2)}_\alpha = \mathcal{N}'_\alpha.$$

Then we have the following identity, where $\mathbb{E}_{\mathbf{q}}$ denotes the expectation with respect to the two families of Gaussian random variables

$$f^{\mathrm{Bos}}(\boldsymbol{\eta}) = \frac{1}{\mathcal{M}(\boldsymbol{\eta})}\mathbb{E}_{\mathbf{q}}\left[\mathbb{E}\left[\prod_{i=1}^N \langle \mathbf{q}, u_k\rangle^{2\eta_i}\middle|\boldsymbol{\lambda}\right]\right]. \tag{3.1.21}$$

This identity can be derived through a Wick theorem. The Fermionic observable will actually be derived in the same way using a Fermionic Wick theorem and Grassmann variables.

Remark 3.1.13. The Bosonic observable can also be given in a matricial way, define the matrices, for $i,j\in[\![1,n]\!]$,

$$E^{(ij)} = \left(\frac{\delta_{ik}\delta_{j\ell}+\delta_{i\ell}\delta_{jk}}{1+\delta_{ij}}\right)_{1\leqslant k,\ell\leqslant n} \quad \text{and} \quad Q_s^{(ij)} = \begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} p_{k_ik_j}(s)$$

Then one can define the following symmetric $2n\times 2n$ matrix involving fluctuations

$$Q_s(k_1,\dots,k_n) = \sum_{1\leqslant i\leqslant j\leqslant N} E^{(ij)}\otimes Q_s^{(ij)} \quad \text{for instance} \quad Q_s(k_1,k_2) = \begin{pmatrix} p_{k_1k_1} & p_{k_1k_1} & p_{k_1k_2} & p_{k_1k_2}\\ p_{k_1k_1} & p_{k_1k_1} & p_{k_1k_2} & p_{k_1k_2}\\ p_{k_2k_1} & p_{k_2k_1} & p_{k_2k_2} & p_{k_2k_2}\\ p_{k_2k_1} & p_{k_2k_1} & p_{k_2k_2} & p_{k_2k_2}\end{pmatrix}.$$

Then we can define our Bosonic observable as

$$f_s^{\mathrm{Bos}}(\boldsymbol{\eta}) = \frac{1}{\mathcal{M}(\boldsymbol{\eta})}\mathbb{E}\left[\mathrm{Haf}\, Q_s(\boldsymbol{\eta})|\boldsymbol{\lambda}\right]$$

where the Hafnian is given by the following definition, for A a $2n\times 2n$ matrix,

$$\mathrm{Haf}\, A = \frac{1}{n!2^n}\sum_{\sigma\in\mathfrak{S}_{2n}}\prod_{j=1}^n A_{\sigma(2j-1),\sigma(2j)}.$$

3.2. Proof of Theorem 3.1.10

3.2.1. Preliminaries

The proof of Theorem 3.1.10 involves a supersymmetric representations of our determinant (3.1.17). In order to develop the proof and the tools, we will recall in this subsection notions of Grassmann variables and Gaussians expectations with respect to these variables. Grassmann variables can be seen as anticommutative numbers, we will first consider four families of Grassmann variables $\{\eta_i, \xi_i, \varphi_i, \psi_i\}_{i=1}^N$, they follow the relations of commutation for i, j two indices in $[\![1, N]\!]$ given by

$$\eta_i\eta_j = -\eta_j\eta_i, \quad \xi_i\xi_j = -\xi_j\xi_i \text{ and } \quad \eta_i\xi_j = -\xi_j\eta_i$$

and all similar relations between the other families. In particular, see that $\eta_i^2 = \xi_i^2 = \varphi_i^2 = \psi_i^2 = 0$ and that the variables $\{\eta_i\xi_j\}$ and $\{\varphi_i\psi_j\}$ all commute.

Remark 3.2.1. A possible representation of such variables is given by matrices. See for instance the Clifford-Wigner-Jordan representation of these Grassmann variables.

Now that we have defined these Grassmann variables, we will define our generalized projections. Namely, we can define for a N-dimensional vector v the following quantity

$$\langle v\rangle_{\boldsymbol{\eta}} = \sum_{\alpha=1}^{N} v(\alpha)\eta_\alpha. \tag{3.2.1}$$

We can also define functions of these Grassmann variables, note that by Taylor expansion and the commutations rules, it is enough to define polynomials of such variables. Thus we will define a function

$$F(\boldsymbol{\eta},\boldsymbol{\xi},\boldsymbol{\varphi},\boldsymbol{\psi}) = \sum_{I,J,K,L\subset[\![1,N]\!]} a_{I,J,K,L} \prod_{i_1\in I\cap J}\eta_{i_1}\xi_{i_1}\prod_{j_1\in I\setminus J}\eta_{j_1}\prod_{k_1\in J\setminus I}\xi_{k_1}\prod_{i_2\in K\cap L}\varphi_{i_2}\psi_{i_2}\prod_{j_2\in K\setminus L}\varphi_{j_2}\prod_{k_2\in L\setminus K}\psi_{k_2}$$

where $a_{I,J,K,L}$ will be real numbers for our purpose. By the matricial representation, one can then see such a function as a matrix. From this definition of a function, we can define the integral of a function by,

$$\int F(\boldsymbol{\eta},\boldsymbol{\xi},\boldsymbol{\varphi},\boldsymbol{\psi})\prod_{i=1}^{N}\mathrm{d}\eta_i\mathrm{d}\xi_i\mathrm{d}\varphi_i\mathrm{d}\psi_i = a_{[N],[N],[N],[N]}$$

where we shortened $[N] := [\![1, N]\!]$. As explained earlier, we can define functions through a Taylor expansion. In order to construct a Gaussian expectation, we need to construct the exponential. It is straightforward to define it as

$$\exp\left(F(\boldsymbol{\eta},\boldsymbol{\xi},\boldsymbol{\varphi},\boldsymbol{\psi})\right) = \sum_{m=1}^{\infty}\frac{F(\boldsymbol{\eta},\boldsymbol{\xi},\boldsymbol{\varphi},\boldsymbol{\psi})^m}{m!} = \sum_{m=1}^{m_0}\frac{F(\boldsymbol{\eta},\boldsymbol{\xi},\boldsymbol{\varphi},\boldsymbol{\psi})^m}{m!}$$

for some m_0 via the commutation relations. We can define our Gaussian expectation as, for an invertible $N\times N$ matrix Δ,

$$\mathcal{E}^{\Delta}_{\boldsymbol{\eta},\boldsymbol{\xi},\boldsymbol{\varphi},\boldsymbol{\psi}}\left[F(\boldsymbol{\eta},\boldsymbol{\xi},\boldsymbol{\varphi},\boldsymbol{\psi})\right] = \int F(\boldsymbol{\eta},\boldsymbol{\xi},\boldsymbol{\varphi},\boldsymbol{\psi})\exp\left(\sum_{i,j=1}^{N}\eta_i\Delta^{-1}_{ij}\xi_j + \sum_{i=1}^{N}\varphi_i\psi_i\right)\prod_{i=1}^{2N}\mathrm{d}\eta_i\mathrm{d}\xi_i\mathrm{d}\varphi_i\mathrm{d}\psi_i \tag{3.2.2}$$

The Fermionic Wick theorem allows us to compute joint Gaussian moments with respect to this super-expectation. We give it here with respect to our Gaussian expectation and the moments we will need later.

Lemma 3.2.2 (Fermionic Wick theorem [ZJ89]). *Consider* $\{(i_k,j_k)\}_{k=1}^m \subset [\![1,N]\!]\times[\![1,N]\!]$, *and* $\{\eta_i,\xi_i,\varphi_i,\psi_i\}_{i=1}^N$ *a family of Grassmann variables, we have*

$$\mathcal{E}^{\Delta}_{\boldsymbol{\eta},\boldsymbol{\xi},\boldsymbol{\varphi},\boldsymbol{\psi}}\left[\prod_{k=1}^m\left(\eta_{i_k}+\mathrm{i}\sqrt{\frac{|I|}{N}}\varphi_{i_k}\right)\left(\xi_{j_k}+\mathrm{i}\sqrt{\frac{|I|}{N}}\psi_{i_k}\right)\right]=\det\left((\Delta-\frac{|I|}{N}\mathrm{Id}_m)_{i_k,j_\ell}\right)_{k,\ell=1}^m \tag{3.2.3}$$

3.2.2. Construction of the Fermionic observable

We will be able to construct an observable based on the families of Grassman variables which will follow (3.1.18). Then by taking the Gaussian expectation defined in (3.2.2) we will obtain the observable (3.1.17) by choosing the right covariance matrix Δ. In the following definitions we will fix a set of indices $I\subset[\![1,N]\!]$ and consider $(\mathbf{q}_i)_{i\in I}$ a family of vectors of $\mathbb{R}^N$ not necesarily orthogonal.

We will consider the observable, for $\boldsymbol{u}^s$ the solution to the Dyson vector flow (3.1.14)

$$g_s^{\mathrm{Fer}}(k_1,\dots,k_n)=\mathbb{E}\left[\prod_{i=1}^n\langle u_{k_i}^s\rangle_{\boldsymbol{\eta}+\mathrm{i}\sqrt{\frac{|I|}{N}}\boldsymbol{\varphi}}\langle u_{k_i}^s\rangle_{\boldsymbol{\xi}+\mathrm{i}\sqrt{\frac{|I|}{N}}\boldsymbol{\psi}}\middle|\boldsymbol{\lambda}\right]. \tag{3.2.4}$$

Remark 3.2.3. Note that in this definition, the product is commutative since we have quantities of order 2 in Grassmann variables. See also that this is a similar quantity as the moment observable from [BY17]. Indeed if one considers a configuration with a single particle at sites $k_1,\dots,k_n$ then the observable would be written as

$$g_s^{\mathrm{Bos}}(k_1,\dots,k_n)=\mathbb{E}\left[\prod_{i=1}^n\langle\mathbf{q},u_{k_i}\rangle^2\middle|\boldsymbol{\lambda}\right].$$

In order to see that g_s^{Fer} follows a form of the eigenvector moment flow (3.1.18), first see the following proposition from [BY17] which gives us the generator of the Dyson vector flow.

Proposition 3.2.4 ([BY17]). *The generator acting on smooth functions of the diffusion* (3.1.14) *is given by*

$$L_t=\sum_{1\leqslant k<\ell\leqslant N}\frac{1}{N(\lambda_k-\lambda_\ell)^2}X_{k\ell}^2 \tag{3.2.5}$$

with the operator $X_{k\ell}$ *defined by*

$$X_{k\ell}=X_{k\ell}^{(1)}-X_{k\ell}^{(2)}\quad\text{with}\quad X_{k\ell}^{(1)}=\sum_{\alpha=1}^N u_k(\alpha)\partial_{u_\ell(\alpha)}\quad\text{and}\quad X_{k\ell}^{(2)}=u_\ell(\alpha)\partial_{u_k(\alpha)} \tag{3.2.6}$$

We will thus need to prove the following lemma, showing that g_s^{Fer} follows the eigenvector moment flow

Lemma 3.2.5. *For* g_s^{Fer} *defined as in* (3.2.4) *and* $\mathbf{k}=(k_1,\dots,k_n)$ *with* $k_i\neq k_j$ *for* $i\neq j$, *we have*

$$\partial_s g_s^{\mathrm{Fer}}(\mathbf{k})=\sum_{i=1}^n\sum_{\substack{\ell=1\\ \ell\notin\{k_1,\dots,k_n\}}}^N\frac{g_s^{\mathrm{Fer}}(\mathbf{k}^i(\ell))-g_s^{\mathrm{Fer}}(\mathbf{k})}{N(\lambda_{k_i}-\lambda_\ell)^2}. \tag{3.2.7}$$

Proof. As the lemma does not depend on the family of Grassmann variables, we will develop the proof for any families $\boldsymbol{\eta},\boldsymbol{\xi}$. First see by definition of the operator that since the eigenvectors for $k\notin\{k_1,\dots,k_n\}$ are not considered in the observable $g_s^{\mathrm{Fer}}(\mathbf{k})$ we clearly have

$$X_{k\ell}^2 g_s^{\mathrm{Fer}}(\mathbf{k})=0\quad\text{for } k\notin\{k_1,\dots,k_n\}.$$

Now, we need to show that for fixed $i,j\in[\![1,n]\!]$ we also have $X^2_{k_ik_j}g_s^{\mathrm{Fer}}(\mathbf{k})=0$. This equality actually comes from the anticommutativity of the Grassmann variables. First see that we have the relations

$$X_{k\ell}\langle u_k\rangle_{\boldsymbol{\eta}}=-\langle u_\ell\rangle_{\boldsymbol{\eta}}\quad\text{and}\quad X_{k\ell}\langle u_\ell\rangle_{\boldsymbol{\eta}}=\langle u_k\rangle_{\boldsymbol{\eta}}.$$

Besides, by definition of the operator $X_{k\ell}$ we only need to look at the part of the observable involving the eigenvectors u_k and u_ℓ, hence computing the quantity

$$\begin{aligned}X^2_{k_ik_j}&\left(\langle u_{k_i}\rangle_{\boldsymbol{\eta}}\langle u_{k_i}\rangle_{\boldsymbol{\xi}}\langle u_{k_j}\rangle_{\boldsymbol{\eta}}\langle u_{k_j}\rangle_{\boldsymbol{\xi}}\right)=2(\langle u_{k_j}\rangle_{\boldsymbol{\eta}}\langle u_{k_j}\rangle_{\boldsymbol{\xi}}\langle u_{k_j}\rangle_{\boldsymbol{\eta}}\langle u_{k_j}\rangle_{\boldsymbol{\xi}}+\langle u_{k_i}\rangle_{\boldsymbol{\eta}}\langle u_{k_i}\rangle_{\boldsymbol{\xi}}\langle u_{k_i}\rangle_{\boldsymbol{\eta}}\langle u_{k_i}\rangle_{\boldsymbol{\xi}}\\&-2\langle u_{k_i}\rangle_{\boldsymbol{\eta}}\langle u_{k_i}\rangle_{\boldsymbol{\xi}}\langle u_{k_j}\rangle_{\boldsymbol{\eta}}\langle u_{k_j}\rangle_{\boldsymbol{\xi}}-\langle u_{k_j}\rangle_{\boldsymbol{\eta}}\langle u_{k_i}\rangle_{\boldsymbol{\xi}}\langle u_{k_i}\rangle_{\boldsymbol{\eta}}\langle u_{k_j}\rangle_{\boldsymbol{\xi}}-\langle u_{k_i}\rangle_{\boldsymbol{\eta}}\langle u_{k_j}\rangle_{\boldsymbol{\xi}}\langle u_{k_j}\rangle_{\boldsymbol{\eta}}\langle u_{k_i}\rangle_{\boldsymbol{\xi}}\\&-\langle u_{k_j}\rangle_{\boldsymbol{\eta}}\langle u_{k_i}\rangle_{\boldsymbol{\xi}}\langle u_{k_j}\rangle_{\boldsymbol{\eta}}\langle u_{k_i}\rangle_{\boldsymbol{\xi}}-\langle u_{k_i}\rangle_{\boldsymbol{\eta}}\langle u_{k_j}\rangle_{\boldsymbol{\xi}}\langle u_{k_i}\rangle_{\boldsymbol{\eta}}\langle u_{k_j}\rangle_{\boldsymbol{\xi}})\\&=0\end{aligned}$$

where we used the fact that $\langle u_{k_i}\rangle^2_{\boldsymbol{\eta}}=0$ and the anticommutativity relations. Finally, we need to compute $X^2_{k_i\ell}g_s^{\mathrm{Fer}}(\mathbf{k})$ for $i\in[\![1,n]\!]$ and $\ell\in[\![1,N]\!]\setminus\{k_1,\dots,k_N\}$, to do so we just need to compute

$$X^2_{k_i\ell}\langle u_{k_i}\rangle_{\boldsymbol{\eta}}\langle u_{k_i}\rangle_{\boldsymbol{\xi}}=2\left(\langle u_\ell\rangle_{\boldsymbol{\eta}}\langle u_\ell\rangle_{\boldsymbol{\xi}}-\langle u_{k_i}\rangle_{\boldsymbol{\eta}}\langle u_{k_i}\rangle_{\boldsymbol{\xi}}\right)$$

which means that we have

$$X^2_{k_i\ell}g_s^{\mathrm{Fer}}(\mathbf{k})=2\left(g_s^{\mathrm{Fer}}(\mathbf{k}^i(\ell))-g_s^{\mathrm{Fer}}(\mathbf{k})\right).$$

Combining all these equalities, we obtain Lemma 3.2.5. □

We now only need to show that we can obtain f_s^{Fer} using our observable g_s^{Fer}, this will involve the Fermionic Wick theorem given by Lemma 3.2.2.

Lemma 3.2.6. *There exists Δ such that*

$$\mathcal{E}^{\Delta}_{\boldsymbol{\eta},\boldsymbol{\xi},\boldsymbol{\varphi},\boldsymbol{\psi}}\left[g_s^{\mathrm{Fer}}(\mathbf{k})\right]=f_s^{\mathrm{Fer}}(\mathbf{k}).$$

Proof. By definition of g_s^{Fer}, we have the following, forgetting the dependence in s,

$$\begin{aligned}&\mathcal{E}^{\Delta}_{\boldsymbol{\eta},\boldsymbol{\xi},\boldsymbol{\varphi},\boldsymbol{\psi}}\left[\prod_{i=1}^n\langle u_{k_i}\rangle_{\boldsymbol{\eta}+\mathrm{i}\sqrt{\frac{|I|}{N}}\boldsymbol{\varphi}}\langle u_{k_i}\rangle_{\boldsymbol{\xi}+\mathrm{i}\sqrt{\frac{|I|}{N}}\boldsymbol{\psi}}\right]\\&\qquad=\sum_{\substack{i_1,\dots,i_n\\j_1,\dots,j_n}}^N\mathcal{E}^{\Delta}_{\boldsymbol{\eta},\boldsymbol{\xi},\boldsymbol{\varphi},\boldsymbol{\psi}}\left[\prod_{i=1}^n\left(\eta_{i_k}+\mathrm{i}\sqrt{\frac{|I|}{N}}\varphi_{i_k}\right)\left(\xi_{j_k}+\mathrm{i}\sqrt{\frac{|I|}{N}}\psi_{j_k}\right)\right]\prod_{m=1}^n u_{k_m}(i_m)u_{k_m}(j_m)\end{aligned}$$

Now we can use the Fermionic Wick theorem 3.2.2 in order to compute these Gaussian moments,

$$\mathcal{E}^{\Delta}_{\boldsymbol{\eta},\boldsymbol{\xi},\boldsymbol{\varphi},\boldsymbol{\psi}}\left[g_s^{\mathrm{Fer}}(\mathbf{k})\right]=\sum_{\substack{i_1,\dots,i_n\\j_1\dots j_n}}^N\det\left(\left(\Delta-\frac{|I|}{N}\mathrm{Id}\right)_{i_pj_q}u_{k_p}(i_p)u_{k_q}(j_q)\right)_{p,q=1}^n$$

Thus by multilinearity of the determinant we obtain that

$$\begin{aligned}\mathcal{E}^{\Delta}_{\boldsymbol{\eta},\boldsymbol{\xi},\boldsymbol{\varphi},\boldsymbol{\psi}}\left[g_s^{\mathrm{Fer}}(\mathbf{k})\right]&=\det\left(\sum_{i,j=1}^N\left(\Delta-\frac{|I|}{N}\mathrm{Id}\right)_{ij}u_{k_p}(i)u_{k_q}(j)\right)_{p,q=1}^n\\&=\det\left(\sum_{i,j=1}^N\Delta_{ij}u_{k_p}(i)u_{k_q}(j)-\frac{|I|}{N}\mathbb{1}_{k_p=k_q}\right)_{p,q=1}^n\end{aligned}$$

Now, we will consider the following covariance matrix

$$\Delta_{ij} = \sum_{\alpha \in I} q_\alpha(i) q_\alpha(j) \quad \text{for } i, j \in [\![1, N]\!].$$

Thus we can finally see that the entries of the matrix we take the determinant of are given by, for $\alpha, \beta \in [\![1, n]\!]$,

$$\sum_{i,j=1}^{N} \Delta_{ij} u_{k_\alpha}(i) u_{k_\beta}(j) - \frac{|I|}{N} \mathbb{1}_{k_\alpha = k_\beta} = \sum_{i \in I} \langle \mathbf{q}_i, u_{k_\alpha} \rangle \langle \mathbf{q}_i, u_{k_\beta} \rangle - \frac{|I|}{N} \mathbb{1}_{k_\alpha = k_\beta} = p_{k_\alpha k_\beta}(s).$$

□

Remark 3.2.7. We gave here a proof of Theorem 3.1.10 with supersymmetry and a link to the first observable following this equation from [BY17]. However, knowing Proposition 3.2.4, it is possible to give a combinatorial proof of the theorem with no consideration of Grassmann variables but simply of the properties of the determinant. We will give this proof in Appendix 3.5.

3.3. Relaxation by the Dyson Brownian motion

In this section, we will make our initial matrix W undergo the Dyson Brownian motion from Definition 3.1.9. The point being to obtain the asymptotic value of f_s^{Fer} after a short time s and see that it coincides with the family $(p_{k\ell})_{k\ell}$ being independent Gaussian random variables.

Before beginning the proof, we will state a priori results we need on the $p_{k\ell}(s)$. The overwhelming probability bound for $p_{k\ell}$ was studied for Gaussian divisible ensembles in [BYY18] in order to study band matrices but only consider bulk eigenvectors. While they only consider $|I| \geqslant cN$ for some constant $c > 0$, we will adapt the proof to any $|I|$ and we obtain the following result in the case of generalized Wigner matrices. In order to prove this estimate we need a priori bounds on eigenvalues and eigenvectors along the dynamics. This is given by the following lemma

Lemma 3.3.1 ([BY17, Lemma 4.2])**.** *Let $\delta, \xi, \omega > 0$ and $t \in [N^{-1+\delta}, N^{-\delta}]$. Consider W a generalized Wigner matrix and consider the dynamics 3.1.12 $(H_s)_{0 \leqslant s \leqslant t}$ with $H_0 = W$. Define the resolvent and its normalized trace for z in the upper plane,*

$$G_s(z) = (H_s - z)^{-1} \quad \textit{and} \quad m_s(z) = \frac{1}{N} \operatorname{Tr} G_s(z).$$

It induces a measure on the space of eigenvalues and eigenvectors $(\boldsymbol{\lambda}(s), \boldsymbol{u}(s))$ for $0 \leqslant s \leqslant t$ such that the following event A_1 holds with overwhelming probability,

- *We have rigidity of eigenvalues: $\forall s \in [0, t]$, $|\lambda_k(s) - \gamma_k| < N^{-2/3+\xi} (\hat{k})^{-1/3}$ uniformly in $k \in [\![1, N]\!]$.*
- *The averaged local law holds: for all $s \in [0, t]$, $|m_s(z) - m(z)| < N^\xi (N\eta)^{-1}$ and the anisotropic local law holds $|\langle \mathbf{v}, G_s(z) \mathbf{w} \rangle - m(z) \langle \mathbf{v}, \mathbf{w} \rangle| \leqslant N^\xi \langle \mathbf{v}, \mathbf{w} \rangle (\sqrt{\operatorname{Im} m(z) (N\eta)^{-1}} + (N\eta)^{-1})$ uniformly in $z \in \mathcal{D}_\omega$.*
- *Eigenvector delocalization holds: $\forall s \in [0, t]$, $\langle \mathbf{q}, u_k(s) \rangle^2 \leqslant N^{-1+\xi}$ uniformly in $k \in [\![1, N]\!]$.*

We will now give all our estimates conditionally on this event which occurs with overwhelming probability thus working using these a priori bounds and estimates deterministically.

Lemma 3.3.2. *For $k, \ell \in [\![1, N]\!]$, we have for any ε and D positive*

$$\mathbb{P}\left(|p_{kk}(s)| + |p_{k\ell}(s)| \geqslant N^{\varepsilon}\left(\frac{|I|}{N\sqrt{Ns^2}} + \sqrt{\frac{|I|}{N^2 s^{3/2}}}\right)\right) \leqslant N^{-D}.$$

Remark 3.3.3. The error term is the sum of two terms and it is not clear if one is larger than the other since it depends on the regime of $|I|$ or s. However, in the regime we will look at here, it will be the term $\sqrt{\frac{|I|}{N^2 s^{3/2}}}$ which will be dominating. Note that in the case of bulk eigenvectors, we have the following overwhelming probability bound from [BYY18] for a general class of initial condition.

Theorem 3.3.4 ([BYY18])**.** *Let $\alpha > 0$, for $k, \ell \in [\![\alpha N, (1-\alpha)N]\!]$ (an index of the bulk) we have, for any ε and D positive,*

$$\mathbb{P}\left(|p_{kk}(s)| + |p_{k\ell}(s)| \geqslant N^{\varepsilon}\sqrt{\frac{|I|}{N^2 s}}\right) \leqslant N^{-D}.$$

Before beginning the proof of Lemma 3.3.2, we will need the following lemma relating our fluctuations $p_{k\ell}$ to the Bosonic observable.

Lemma 3.3.5 ([BYY18])**.** *Take an even integer n, there exists $C > 0$ depending on n such that for any $i < j$ and any time s we have*

$$\mathbb{E}\left[p_{ij}(s)^n | \boldsymbol{\lambda}\right] \leqslant C\left(f_s^{\mathrm{Bos}}(\boldsymbol{\eta}^{(1)}) + f_s^{\mathrm{Bos}}(\boldsymbol{\eta}^{(2)}) + f_s^{\mathrm{Bos}}(\boldsymbol{\eta}^{(3)})\right) \tag{3.3.1}$$

where $\boldsymbol{\eta}^{(1)}$ is the configuration of n particles in the site i and no particle elsewhere, $\boldsymbol{\eta}^{(2)}$ n particles in the site j, and $\boldsymbol{\eta}^{(3)}$ an equal number of particles between the site i and the site j.

Using this lemma we can now adapt the proof of [BYY18, Theorem 2.5] to the edge case. Note that the proof is actually simpler since we do not need to localize the dynamics in the bulk of the spectrum.

Proof of Lemma 3.3.2. Let $\xi > 0$. Consider $f^{\mathrm{Bos}}(\boldsymbol{\eta})$ the Bosonic obervable for the eigenvector moment flow. Consider n fixed and look at the configuration $\boldsymbol{\eta_m}$ to be such that

$$f^{\mathrm{Bos}}(\boldsymbol{\eta_m}) = \sup_{\boldsymbol{\eta}, \mathcal{N}(\boldsymbol{\eta}) = n} f^{\mathrm{Bos}}(\boldsymbol{\eta}) \quad \text{and} \quad S_s = \sup_{\boldsymbol{\eta}} f_s^{\mathrm{Bos}}(\boldsymbol{\eta}).$$

Let η be a small parameter such that, if $\eta \in [N^{-1+\omega}, \omega^{-1}]$ for some small ω. We then have, forgetting about the superscript Bos,

$$\partial_s f_s(\boldsymbol{\eta_m}) = \sum_{k \neq \ell} 2\eta_k(1 + 2\eta_\ell)\frac{f_s(\boldsymbol{\eta_m^{k,\ell}}) - f_s(\boldsymbol{\eta_m})}{N(\lambda_k - \lambda_\ell)^2} \leqslant \frac{C}{N\eta}\sum_{i=1}^{p}\sum_{\ell \neq k_i}\frac{\eta(f_s(\boldsymbol{\eta_m^{k,\ell}}) - f_s(\boldsymbol{\eta_m}))}{(\lambda_{k_i} - \lambda_\ell)^2 + \eta^2}$$

where we denoted $(k_1, \dots, k_p)$ the sites k such that $\eta_k \neq 0$. In particular, $p \leqslant n$ and $\sum_{i=1}^{p}\eta_{k_i} = n$. Now, we have that

$$f_s(\boldsymbol{\eta_m})\frac{1}{N}\sum_{i=1}^{p}\sum_{\ell \neq k_i}\frac{\eta}{(\lambda_{k_i} - \lambda_\ell)^2 + \eta^2} = \left(\sum_{i=1}^{p}\operatorname{Im} m(z_{k_i})\right) f_s(\boldsymbol{\eta_m}) + \mathcal{O}\left(\frac{N^{\xi}}{N\eta}S_s\right)$$

For the other term, we will use an implicit bound using Hőlder inequalities. Indeed, if we denote $z_{k_i} = \lambda_{k_i} + \mathrm{i}\eta$,

$$\operatorname{Im}\sum_{\ell\neq k_i}\frac{f_s(\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{k},\boldsymbol{\ell}})}{N(\lambda_\ell - z)} = \operatorname{Im}\sum_{\ell\notin\{k_1,\dots,k_p\}}\frac{f_s(\boldsymbol{\eta}_{\boldsymbol{m}}^{\boldsymbol{k},\boldsymbol{\ell}})}{N(\lambda_\ell - z)} + \mathcal{O}\left(\frac{N^\xi}{N\eta}S_s\right)$$

Now, we can expand by the definition of $f_s(\boldsymbol{\eta})$ in terms of a sum over perfect matchings. Since we move one particle from k to ℓ, which is an empty site for the configuration $\boldsymbol{\eta}_{\boldsymbol{m}}$, we only have two particles in the graph on the site ℓ. Thus, there is two possibilities for the perfect matching, either there is an edge $\{(\ell,1),(\ell,2)\}$ or there is not. If there is such an edge, then we can write the contribution of such perfect matchings as

$$\mathbb{E}\left[Q_{n-1}(\boldsymbol{\eta}_{\boldsymbol{m}})\operatorname{Im}\sum_{\ell\notin\{k_1,\dots,k_p\}}\frac{p_{\ell\ell}}{N(\lambda_\ell - z)}\middle|\boldsymbol{\lambda}\right].$$

Now, we can use the isotropic local law to see that

$$\operatorname{Im}\sum_{\ell\notin\{k_1,\dots,k_p\}}\frac{p_{\ell\ell}}{N(\lambda_\ell - z)} = \operatorname{Im}\frac{1}{N}\sum_{\alpha\in I}\langle\mathbf{q}_\alpha, G_s(z)\mathbf{q}_\alpha\rangle - \frac{|I|}{N}\operatorname{Im} m(z) = \mathcal{O}\left(\frac{N^\xi|I|}{N\sqrt{N\eta}}.\right)$$

See that $Q_{n-1}(\boldsymbol{\eta})$ is a sum of monomial of degree $n-1$ involving the fluctuations $p_{k\ell}$. Thus by a Young inequality, using Lemma 3.3.5, we have that

$$Q_{n-1}(\boldsymbol{\eta}) = \mathcal{O}\left(S_s^{\frac{n-1}{n}}\right).$$

Now for perfect matchings where $\{(\ell,1),(\ell,2)\}$ is not an edge, we can write the contribution in the following way,

$$\mathbb{E}\left[Q_{n-2}(q_1,q_2,\boldsymbol{\eta}_{\boldsymbol{m}})\operatorname{Im}\sum_{\ell\notin\{k_1,\dots,k_p\}}\frac{p_{k_{q_1}\ell}p_{k_{q_2}\ell}}{N(\lambda_\ell - z)}\middle|\boldsymbol{\lambda}\right].$$

We can write, in order to control the sum

$$\operatorname{Im}\sum_{\ell\notin\{k_1,\dots,k_p\}}\frac{p_{k_{q_1}\ell}p_{k_{q_2}\ell}}{N(\lambda_\ell - z)} = \mathcal{O}\left(\frac{1}{N\eta}\sum_{\ell=1}^N(p_{k_{q_1}\ell}^2 + p_{k_{q_2}\ell}^2)\right) = \mathcal{O}\left(\frac{N^\xi|I|}{N^2\eta}\right)$$

where we used the complete delocalization property from Lemma 3.1.8 and the fact that

$$\sum_{\ell=1}^N p_{k\ell}^2 = \sum_{\alpha\in I}u_k(\alpha)^2 = \mathcal{O}\left(N^\xi\frac{|I|}{N}\right).$$

In the same way, using a Young inequality with Lemma 3.3.5, we control the polynomial of degree $n-2$ in terms of $p_{k\ell}$,

$$Q_{n-2}(q_1,q_2,\boldsymbol{\eta}_{\boldsymbol{m}}) = \mathcal{O}\left(S_s^{\frac{n-2}{n}}\right)$$

Thus, combining all these inequalities, we obtain the following Gronwall-type inequality,

$$\partial_s f_s(\boldsymbol{\eta}_{\boldsymbol{m}}) \leqslant -\frac{C}{\eta}\left(\sum_{i=1}^p\operatorname{Im} m(z_{k_i})\right)f_s(\boldsymbol{\eta}_{\boldsymbol{m}}) + \mathcal{O}\left(\frac{N^\xi}{\eta}\left(\frac{1}{N\eta}S_s + \frac{|I|}{N\sqrt{N\eta}}S_s^{\frac{n-1}{n}} + \frac{|I|}{N^2\eta}S_s^{\frac{n-2}{n}}\right)\right)$$

Now, using the fact that for $\eta \geqslant N^{-2/3+\varepsilon}$, we have $\operatorname{Im} m(E+\mathrm{i}\eta) \geqslant \sqrt{\eta}$, we obtain by Gronwall's lemma, by taking $\eta = sN^{-\omega}$, for some small ω,

$$f_s(\boldsymbol{\eta}_{\boldsymbol{m}}) = \mathcal{O}\left(\frac{N^{3\omega/2}}{Ns^{3/2}}S_s + \frac{|I|N^\omega}{N\sqrt{Ns^2}}S_s^{\frac{n-1}{n}} + \frac{|I|N^\omega}{N^2s^{3/2}}S_s^{\frac{n-2}{n}}\right)$$

Now, using the same machinery as in [BYY18, Theorem 2.5], we obtain that

$$|p_{kk}| + |p_{k\ell}| = \mathcal{O}\left(N^\xi \frac{|I|}{N\sqrt{Ns^2}} + N^\xi\sqrt{\frac{|I|}{N^2s^{3/2}}}\right)$$

which gives the lemma.

□

We will give the asymptotic value of f^{Fer} in the following lemma, while this is a simple computation, we will give a short proof in order to see a recursion relation. Indeed, it is this recursion relation which will occur in the later proof.

Lemma 3.3.6. *Consider $A_n = \mathbb{E}[\det \mathrm{G}]$ where G is a symmetric $n \times n$ matrix with independent entries (up to the symmetry) given by $\mathrm{G}_{ij} \sim \mathcal{N}(0,1)$ for $i \neq j$ and $\mathrm{G}_{ii} \sim \mathcal{N}(0,2)$. Then we have*

$$A_n = \begin{cases} (-1)^{n/2} n!! & \text{if } n \text{ is even,} \\ 0 & \text{otherwise.} \end{cases}$$

Proof. We will see a recursion relation of order 2 by developing according to some rows and columns. We will write in the following $M_{(i)}^{(j)}$ the matrix M where we removed the line i and the column j. We can then develop the determinant in the following way

$$A_n = \mathbb{E}[\det \mathrm{G}] = \mathbb{E}\left[\sum_{i=1}^n \mathrm{G}_{1,i}(-1)^{i+1}\det \mathrm{G}_{(1)}^{(i)}\right] = \sum_{i=1}^n\sum_{j=1}^{n-1}(-1)^{i+j}\mathbb{E}[\mathrm{G}_{1,i}\mathrm{G}_{j+1,1}]\,\mathbb{E}\left[\det \mathrm{G}_{(1,j+1)}^{(i,1)}\right]$$
$$= -\sum_{i=2}^n A_{n-2}.$$

And this recursion formula gives us the result knowing that

$$A_1 = 0 \quad \text{and} \quad A_2 = -1.$$

□

The following theorem will show that our determinant is asymptotically close to A_n and thus confirming the idea that, in the sense of moments, the family of $(p_{k\ell})$ are independent Gaussian. However, the knowledge of these moments is not enough to say that the whole family behaves that way. We will actually only use the case $n = 2$ to obtain Theorem 3.1.4 but we state here the theorem for any value of n. Note that we will now work on the overwhelming probability even A_2 which corresponds to the intersection of A_1 where the estimate from Proposition 3.3.2 holds.

Theorem 3.3.7. *Let $n \in \mathbb{N}$ and f_s^{Fer} as in (3.2.4), there exists a $\vartheta_n > 0$ such that*

$$\sup_{\mathbf{k},|\mathbf{k}|=n}\left|f_s^{\mathrm{Fer}}(\mathbf{k}) - \left(\frac{\sqrt{|I|}}{N}\right)^n A_n\right| = \mathcal{O}\left(\left(\frac{\sqrt{|I|}}{N}\right)^n N^{-\vartheta_n}\right). \tag{3.3.2}$$

Note that in the proof we will always do a maximum principle in order to obtain our leading order but the same estimates can be done on the infimum of our observable so that we get our result.

Proof of Theorem 3.3.7. Let $\xi > 0$. We will also use a recursion formula in order to obtain the value of our Fermionic observable. and thus need to obtain an estimate on the observable for small n, the size of the determinant. For $n = 1$, we have that $f_s^{\mathrm{Fer}}(k) = \mathbb{E}\left[p_{kk}|\boldsymbol{\lambda}\right]$. We can obtain an estimate by using a maximum principle on f_s^{Fer}. Consider k_m the index such that

$$f_s^{\mathrm{Fer}}(k_m) = \sup_{k\in[\![1,N]\!]} f_s^{\mathrm{Fer}}(k).$$

Then we have, since f_s^{Fer} follows the dynamics (3.1.18), we have for any $\eta > 0$,

$$\partial_s f_s^{\mathrm{Fer}}(k_m) = 2\sum_{\ell\neq k_m} \frac{f_s^{\mathrm{Fer}}(\ell) - f_s^{\mathrm{Fer}}(k_m)}{N(\lambda_\ell - \lambda_{k_m})^2} \leqslant \frac{2}{\eta}\mathbb{E}\left[\frac{1}{N}\sum_{\ell\neq k_m}\frac{(p_{\ell\ell} - p_{k_m k_m})\eta}{(\lambda_\ell - \lambda_{k_m})^2 + \eta^2}\middle|\boldsymbol{\lambda}\right].$$

Now, one can see that

$$p_{k_m k_m}\frac{1}{N}\sum_{\ell\neq k_m}\frac{\eta}{(\lambda_\ell - \lambda_{k_m})^2 + \eta^2} = p_{k_m k_m}\operatorname{Im} m(z_{k_m}) + \mathcal{O}\left(\frac{N^\xi}{N\eta}\sqrt{\frac{|I|}{N^2 s^{3/2}}}\right)$$

where we introduced the notation $z_{k_i} = \lambda_{k_i} + I\eta$. For the other term, we will use the isotropic local law from (3.1.10),

$$\begin{aligned}\frac{1}{N}\sum_{\ell\neq k_m}\frac{p_{\ell\ell}\eta}{(\lambda_\ell - \lambda_{k_m})^2 + \eta^2} &= \frac{1}{N}\sum_{i\in I}\langle \mathbf{q}_i, G(z_{k_m})\mathbf{q}_i\rangle - \frac{|I|}{N}\operatorname{Im} m(z_{k_m}) + \mathcal{O}\left(\frac{N^\xi}{N\eta}\sqrt{\frac{|I|}{N^2 s^{3/2}}}\right)\\ &= \mathcal{O}\left(\frac{N^\xi |I|}{N\sqrt{N\eta}} + \frac{N^\xi}{N\eta}\sqrt{\frac{|I|}{N^2 s^{3/2}}}\right).\end{aligned}$$

Thus, we obtain the following gronwall type inequality,

$$\partial_s f_s^{\mathrm{Fer}}(k_m) = -\frac{2\operatorname{Im} m(z_{k_m})}{\eta} f_s^{\mathrm{Fer}}(k_m) + \mathcal{O}\left(\frac{|I|N^\xi}{N\eta\sqrt{N\eta}} + \frac{N^\xi}{N\eta^2}\sqrt{\frac{|I|}{N^2 s^{3/2}}}\right)$$

which gives us that, as long as we consider $\eta \ll s$,

$$f_s^{\mathrm{Fer}}(k_m) = \mathcal{O}\left(\frac{|I|N^\xi}{N\sqrt{N\eta}} + \frac{N^\xi}{N\eta}\sqrt{\frac{|I|}{N^2 s^{3/2}}}\right).$$

Now, for any $\varepsilon > 0$ small enough, we can consider $\eta = N^{-\varepsilon}s$ and s be taken such that, $N^{-1} \ll s \ll 1$ and

$$\frac{|I|}{N\sqrt{Ns}} \vee \frac{1}{Ns}\sqrt{\frac{|I|}{N^2 s^{3/2}}} \ll \frac{\sqrt{|I|}}{N}.$$

Note that these choices of parameters are possible since we consider $|I| \ll \sqrt{N}$. Thus, the case $n = 1$ goes in the direction of Lemma 3.3.6.

We will now study the case $n = 2$, in this case we can write our Fermionic observable as

$$f_s^{\mathrm{Fer}}(k_1, k_2) = \mathbb{E}\left[p_{k_1k_1}p_{k_2k_2} - p_{k_1k_2}^2\middle|\boldsymbol{\lambda}\right].$$

We will use a maximum principle for this observable since it follows the parabolic equation (3.1.18) and obtain the result by a Gronwall argument. Consider $\mathbf{k^m} = (k_1^m, k_2^m)$ the multi-index corresponding to the maximum of the function f_s^{Fer} so that

$$f_s^{\mathrm{Fer}}(\mathbf{k^m}) = \sup_{\mathbf{k},\, |\mathbf{k}|=2} f_s^{\mathrm{Fer}}(\mathbf{k^m}).$$

Then we have, since f_s^{Fer} follows (3.1.18) and $\mathbf{k^m}$ is the index for which f_s^{Fer} is the maximum, for any positive η,

$$\partial_s f_s^{\mathrm{Fer}}(\mathbf{k^m}) = 2\sum_{i=1}^{2} \sum_{\ell \notin \{k_1^m, k_2^m\}} \frac{f_s^{\mathrm{Fer}}((\mathbf{k^m})^i(\ell)) - f_s^{\mathrm{Fer}}(\mathbf{k^m})}{N(\lambda_{k_i^m} - \lambda_\ell)^2}$$
$$\leqslant \frac{2}{\eta}\sum_{i=1}^{2} \frac{1}{N} \sum_{\ell \notin \{k_1^m, k_2^m\}} \frac{(f_s^{\mathrm{Fer}}((\mathbf{k^m})^i(\ell)) - f_s^{\mathrm{Fer}}(\mathbf{k^m}))\eta}{(\lambda_{k_i^m} - \lambda_\ell)^2 + \eta^2}.$$

Now, we will consider only the terms in the first sum of the right hand side for readability. First note that adding this parameter η made imaginary part arise. Namely, we have the formula

$$\frac{1}{N} \sum_{\ell \notin \{k_1^m, k_2^m\}} f_s^{\mathrm{Fer}}(\mathbf{k^m}) \frac{\eta}{(\lambda_{k_i^m} - \lambda_\ell)^2 + \eta^2} = f_s^{\mathrm{Fer}}(\mathbf{k^m}) \operatorname{Im} m(z_{k_i}) + \mathcal{O}\left(\frac{N^\xi}{N\eta} \frac{|I|}{N^2 s^{3/2}}\right)$$

where we used Theorem 3.3.4 for the error term. Now, we need to control the term involving $f_s^{\mathrm{Fer}}((\mathbf{k^m})^i(\ell)) = \mathbb{E}\left[p_{k_{3-i}k_{3-i}} p_{\ell\ell} - p_{k_{3-i}\ell} \middle| \boldsymbol{\lambda}\right]$. In the following we will look at the term $i = 1$, the term $i = 2$ can be bounded in exactly the same way. Thus we can write

$$\frac{1}{N} \sum_{\ell \notin \{k_1^m, k_2^m\}} \frac{(p_{\ell\ell} p_{k_2 k_2} - p_{\ell k_2}^2)\eta}{(\lambda_{k_1} - \lambda_\ell)^2 + \eta^2} = p_{k_2 k_2} \operatorname{Im} \sum_{\ell=1}^{N} \frac{p_{\ell\ell}}{N(\lambda_\ell - z_{k_1})} - \operatorname{Im} \sum_{\ell=1}^{N} \frac{p_{\ell k_2}^2}{N(\lambda_\ell - z_{k_1})} + \mathcal{O}\left(\frac{N^\xi}{N\eta} \frac{|I|}{N^2 s^{3/2}})\right). \tag{3.3.3}$$

Now we can write these two sums in terms of the resolvent defined in (3.1.8) and control them with the isotropic local law (3.1.10). Indeed, by definition of our overlaps we have

$$\operatorname{Im} \sum_{\ell=1}^{N} \frac{p_{\ell\ell}}{N(\lambda_\ell - z_{k_1})} = \operatorname{Im} \sum_{\alpha \in I} \sum_{\ell=1}^{N} \frac{\langle \mathbf{q}_\alpha, u_\ell \rangle^2}{N(\lambda_\ell - z_{k_1})} - \frac{|I|}{N} \operatorname{Im} m(z_{k_1})$$
$$= \frac{1}{N} \operatorname{Im}\left(\sum_{\alpha \in I} (\langle \mathbf{q}_\alpha, G(z_{k_1}) \mathbf{q}_\alpha \rangle - m(z_{k_1}))\right) = \mathcal{O}\left(\frac{N^\xi |I|}{N\sqrt{N\eta}}\right). \tag{3.3.4}$$

For the second term in (3.3.3), we can also write it in terms of the resolvent

$$\operatorname{Im} \sum_{\ell=1}^{N} \frac{p_{\ell k_2}^2}{N(\lambda_\ell - z_{k_1})} = \frac{1}{N} \sum_{\alpha,\beta \in I} \langle \mathbf{q}_\alpha, u_k \rangle \langle \mathbf{q}_\beta, u_k \rangle \operatorname{Im} \langle \mathbf{q}_\alpha, G(z_{k_1}) \mathbf{q}_\beta \rangle$$
$$= \frac{1}{N} \sum_{\alpha \in I} \langle \mathbf{q}_\alpha, u_k \rangle^2 \operatorname{Im} m(z_{k_1}) + \mathcal{O}\left(\frac{|I|^2}{N^2} \frac{N^{3\xi}}{\sqrt{N\eta}}\right) = \frac{|I|}{N^2} \operatorname{Im} m(z_{k_1}) + \mathcal{O}\left(\frac{N^\xi}{N} \sqrt{\frac{|I|}{N^2 s^{3/2}}} + \frac{|I|^2}{N^2} \frac{N^{3\xi}}{\sqrt{N\eta}}\right) \tag{3.3.5}$$

where we used in these inequalities the isotropic local law from (3.1.10), the delocalization property (3.1.11) and Theorem 3.3.4 and the fact that $\operatorname{Im} m(z)$ is of order smaller than one. Now, combining the estimates we obtain the inequality,

$$\partial_s f_s^{\mathrm{Fer}}(\mathbf{k}^{\mathbf{m}}) = -\frac{C}{\eta}\left(f_s^{\mathrm{Fer}}(\mathbf{k}^{\mathbf{m}}) + \frac{|I|}{N^2}\right) + \mathcal{O}\left(\frac{1}{\eta}\left(\frac{N^{3\xi}}{N\eta}\frac{|I|}{N^2 s^{3/2}} + \frac{|I|}{N\sqrt{N\eta}} + \frac{1}{N}\sqrt{\frac{|I|}{N^2 s^{3/2}}} + \frac{|I|^2}{N^2}\frac{1}{\sqrt{N\eta}}\right)\right).$$

Now, consider a positive $\varepsilon > 0$ small enough so that we can take $\eta = sN^{-\varepsilon}$ and s such that $N^{-2/3} \ll s \ll 1$ such that all the terms in the parenthesis are of order smaller that $\frac{|I|}{N^2}$. This bound is possible only since we took $|I| \ll \sqrt{N}$ for the term $\frac{|I|^2}{N^2}\frac{1}{\sqrt{N\eta}}$. And we finally obtain by Gronwall's lemma, since we have for $\eta \geqslant N^{-2/3+\varepsilon}$, $\operatorname{Im} m(E+\mathrm{i}\eta) \geqslant \sqrt{\eta}$,

$$f_s^{\mathrm{Fer}}(\mathbf{k}^{\mathbf{m}}) = -\frac{|I|}{N^2} + \mathcal{O}\left(\frac{|I|}{N^2}N^{-\vartheta_2}\right)$$

for some $\vartheta_2 > 0$. Thus, the case $n = 2$ has been proved and we obtain the same initial conditions as in Lemma 3.3.6. Consider now the case where n is an integer greater than 2. For the general case, we will develop our Fermionic observable via the Leibniz formula, for $\mathbf{k}$ such that $|\mathbf{k}| = n$ and $\mathfrak{S}_n$ the set of permutations of $[\![1, n]\!]$,

$$f_s^{\mathrm{Fer}}(\mathbf{k}) = \mathbb{E}\left[\det P_s(\mathbf{k}) | \boldsymbol{\lambda}\right] = \sum_{\sigma\in\mathfrak{S}_n} \epsilon(\sigma)\mathbb{E}\left[\prod_{i=1}^n p_{k_i k_{\sigma(i)}}(s)\middle|\boldsymbol{\lambda}\right].$$

As earlier, we will use a maximum principle technique in order to obtain the leading order for Theorem 3.3.7. Consider $\mathbf{k}^m$ maximizing $f_s^{\mathrm{Fer}}(\mathbf{k})$ and write

$$\partial_s f_s^{\mathrm{Fer}}(\mathbf{k}^{\mathbf{m}}) = 2\sum_{i=1}^n \sum_{\ell\notin\{k_1^m,\dots,k_n^m\}} \frac{f_s^{\mathrm{Fer}}((\mathbf{k}^{\mathbf{m}})^i(\ell)) - f_s^{\mathrm{Fer}}(\mathbf{k}^{\mathbf{m}})}{N(\lambda_{k_i^m} - \lambda_\ell)^2}$$
$$\leqslant \frac{2}{\eta}\sum_{i=1}^n \sum_{\ell\notin\{k_1^m,\dots,k_n^m\}} \frac{(f_s^{\mathrm{Fer}}((\mathbf{k}^{\mathbf{m}})^i(\ell)) - f_s^{\mathrm{Fer}}(\mathbf{k}^{\mathbf{m}}))\eta}{N((\lambda_{k_i^m} - \lambda_\ell)^2 + \eta^2)}. \quad (3.3.6)$$

Now, we can also write that since n is fixed independent of N,

$$\frac{1}{N}\sum_{\ell\notin\{k_1^m,k_2^m\}} f_s^{\mathrm{Fer}}(\mathbf{k}^{\mathbf{m}})\frac{\eta}{(\lambda_{k_i^m} - \lambda_\ell)^2 + \eta^2} = f_s^{\mathrm{Fer}}(\mathbf{k}^{\mathbf{m}})\operatorname{Im} m(z_{k_i}) + \mathcal{O}\left(\frac{N^{n\xi}}{N\eta}\left(\frac{|I|}{N^2 s^{3/2}}\right)^{n/2}\right).$$

In order to control $f_s^{\mathrm{Fer}}((\mathbf{k}^{\mathbf{m}})^i(\ell))$, we will partition $\mathfrak{S}_n$ into three sets which will give different contributions to the result, note that we will make the permutations on the set given by the indices in $(\mathbf{k}^{\mathbf{m}})^i(\ell)$ but since the number of indices stay constant and equals n, this dependence does not matter in our computations,

$$\mathfrak{S}_n^{(1)}(\ell) = \{\sigma \in \mathfrak{S}_n : \sigma(\ell) = \ell\},$$
$$\mathfrak{S}_n^{(2)}(\ell) = \left\{\sigma \in \mathfrak{S}_n : \sigma(\ell) = \sigma^{-1}(\ell) \quad \text{and} \quad \sigma(\ell) \neq \ell\right\},$$
$$\mathfrak{S}_n^{(3)}(\ell) = \mathfrak{S}_n \setminus (\mathfrak{S}_n^{(1)} \sqcup \mathfrak{S}_n^{(2)}).$$

Now see that we can write, for a fixed $i \in [\![1,n]\!]$,

$$\frac{1}{N}\sum_{\ell\notin\{k_1^m,\dots,k_n^m\}}\frac{\eta f_s^{\mathrm{Fer}}((\mathbf{k}^{\mathbf{m}})^i(\ell))}{(\lambda_{k_i^m}-\lambda_\ell)^2+\eta^2}=\frac{1}{N}\operatorname{Im}\sum_{\ell=1}^N\frac{f_s^{\mathrm{Fer}}((\mathbf{k}^{\mathbf{m}})^i(\ell))}{(\lambda_\ell-z_{k_i^m})}+\mathcal{O}\left(\frac{N^{n\xi}}{N\eta}\left(\frac{|I|}{N^2s^{3/2}}\right)^{n/2}\right).$$

Now, by developing $f_s^{\mathrm{Fer}}((\mathbf{k}^{\mathbf{m}})^i(\ell))$ according to the Leibiniz formula and separating this sum in three terms with respect to the prior partition of $\mathfrak{S}_n$, we now have to control these terms, the first one can be written as

$$(\mathrm{I}):=\sum_{\sigma\in\mathfrak{S}_n^{(1)}(i)}\epsilon(\sigma)\left(\prod_{\substack{j=1\\ j\neq i}}^n p_{k_jk_{\sigma(j)}}\right)\operatorname{Im}\sum_{\ell=1}^N\frac{p_{\ell\ell}}{N(\lambda_\ell-z_{k_i})}=\mathcal{O}\left(\frac{N^{n\xi}|I|}{N\sqrt{N\eta}}\left(\frac{|I|}{N^2s^{3/2}}\right)^{(n-1)/2}\right)$$

using the local law from (3.1.10). Now, for the contribution of $\mathfrak{S}_n^{(2)}$ we have to control

$$\begin{aligned}(\mathrm{II}):=&\sum_{\sigma\in\mathfrak{S}_n^{(2)}(i)}\epsilon(\sigma)\left(\prod_{\substack{j\neq i\\ j\neq\sigma(i)}}p_{k_jk_{\sigma(j)}}\right)\operatorname{Im}\sum_{\ell=1}^N\frac{p^2_{\ell k_{\sigma(i)}}}{N(\lambda_\ell-z_{k_i})}\\ =&\frac{|I|}{N^2}\operatorname{Im}m(z_{k_i})\sum_{\sigma\in\mathfrak{S}_n^{(2)}(i)}\epsilon(\sigma)\left(\prod_{\substack{j\neq i\\ j\neq\sigma(i)}}p_{k_jk_{\sigma(j)}}\right)+\mathcal{O}\left(\left(\frac{N^{n\xi}}{N}\sqrt{\frac{|I|}{N^2s^{3/2}}}+\frac{|I|^2N^{n\xi}}{N^2\sqrt{N\eta}}\times\right)\right.\\ &\left.\times\left(\frac{|I|}{N^2s^{3/2}}\right)^{(n-2)/2}\right)\end{aligned}$$

where we used the estimate (3.3.5) and Theorem 3.3.4. It is possible to see (II) as a sum over the possible $\sigma(i)$ in the product so that we can write it as a sum of determinant of size $n-2$ in order to conclude later by induction. Indeed, we have

$$(\mathrm{II})=-\frac{|I|}{N^2}\operatorname{Im}m(z_{k_i})\sum_{\substack{i_0=1\\ i_0\neq i}}^n\det P_{s(i,i_0)}^{(i,i_0)}+\mathcal{O}\left(\left(\frac{N^{n\xi}}{N}\sqrt{\frac{|I|}{N^2s^{3/2}}}+\frac{N^{n\xi}|I|^2}{N^2\sqrt{N\eta}}\right)\left(\frac{|I|}{N^2s^{3/2}}\right)^{(n-2)/2}\right).$$

Note that in the previous equation we obtain a minus sign from the signatures of the permutations. Indeed, as the estimate removed the cycle $(k_ik_{\sigma(i)})$, it removed two elements from the set so that if one writes the signature as $\epsilon(\sigma)=(-1)^{n-\mathscr{C}(\sigma)}$ with $\mathscr{C}(\sigma)$ the number of cycles of the permutation σ, the new signature becomes $(-1)^{n-2-\mathscr{C}(\sigma)+1}=-\epsilon(\sigma)$. It remains to bound the last term coming from $\mathfrak{S}_n^{(3)}$,

$$(\mathrm{III})=\sum_{\sigma\in\mathfrak{S}_n^{(3)}(i)}\epsilon(\sigma)\left(\prod_{\substack{j\neq i\\ j\neq\sigma^{-1}(j)}}p_{k_jk_{\sigma(j)}}\right)\operatorname{Im}\sum_{\ell=1}^N\frac{p_{\ell k_{\sigma(i)}}p_{k_{\sigma^{-1}(i)}\ell}}{N(\lambda_\ell-z_{k_i})}.$$

Now, we can write the last sum as,

$$\operatorname{Im}\sum_{\ell=1}^{N}\frac{p_{\ell k_{\sigma(i)}}p_{k_{\sigma^{-1}(i)}\ell}}{N(\lambda_\ell - z_{k_i})} = \frac{1}{N}\sum_{\alpha,\beta\in I}\langle \mathbf{q}_\alpha, u_{k_{\sigma(i)}}\rangle\langle \mathbf{q}_\beta, u_{k_{\sigma^{-1}(i)}}\rangle \operatorname{Im}\langle \mathbf{q}_\alpha, G(z_{k_i})\mathbf{q}_\beta\rangle$$
$$= \frac{1}{N}\operatorname{Im} m(z) p_{k_{\sigma(i)}}p_{k_{\sigma^{-1}(i)}} + \mathcal{O}\left(\frac{N^{\xi}|I|}{N^2\sqrt{N\eta}} + \frac{N^{3\xi}|I|^2}{N^2\sqrt{N\eta}}\right) = \mathcal{O}\left(\frac{N^{3\xi}}{N}\sqrt{\frac{|I|}{N^2 s^{3/2}}} + \frac{|I|^2}{N^2\sqrt{N\eta}}\right)$$

which gives us that

$$(\mathrm{III}) = \mathcal{O}\left(N^{(n+3)\xi}\left(\frac{1}{N}\sqrt{\frac{|I|}{N^2 s}} + \frac{|I|^2}{N^2\sqrt{N\eta}}\right)\left(\frac{|I|}{N^2 s^{3/2}}\right)^{(n-2)/2}\right).$$

Finally, putting all these estimates together in (3.3.6) , we obtain the following inequality

$$\partial_s f_s^{\mathrm{Fer}}(\mathbf{k}^{\mathbf{m}}) \leqslant -\sum_{i=1}^{n}\operatorname{Im} m(z_{k_i})\frac{C}{\eta}\left(f_s^{\mathrm{Fer}}(\mathbf{k}^{m}) + \frac{|I|}{N^2}\sum_{\substack{i_0=1\\ i_0\neq i}}^{n}\mathbb{E}\left[\det P_{s}{}_{(i,i_0)}^{(i,i_0)}\Big|\boldsymbol{\lambda}\right]\right)$$
$$+\mathcal{O}\left(\frac{1}{\eta}\left(\frac{N^{(n+3)\xi}}{N\eta} + \sqrt{\frac{|I|}{s^{3(n-1)/2}N\eta}} + \frac{1}{\sqrt{s^{3(n-1)/2}|I|}} + \frac{|I|}{\sqrt{s^{3(n-2)/2}N\eta}}\right)\left(\frac{\sqrt{|I|}}{N}\right)^{n}\right).$$

Now, we are going to use our induction hypothesis, since $\mathbb{E}\left[\det P_{s}{}_{(i,i_0)}^{(i,i_0)}\Big|\boldsymbol{\lambda}\right]$ corresponds to the Fermionic observable with a configuration of $n-2$ particles (we removed a particle in i and in i_0), thus we will suppose that there exists a ϑ_{n-2} such that for any i and i_0

$$\mathbb{E}\left[\det P_{s}{}_{(i,i_0)}^{(i,i_0)}\Big|\boldsymbol{\lambda}\right] = \left(\frac{\sqrt{|I|}}{N}\right)^{(n-2)} A_{n-2} + \mathcal{O}\left(\left(\frac{\sqrt{|I|}}{N}\right)^{n-2} N^{-\vartheta_{n-2}}\right).$$

So that, since we got the right initial conditions earlier, we obtain that

$$\partial_s f_s^{\mathrm{Fer}}(\mathbf{k}^{\mathbf{m}}) \leqslant -\sum_{i=1}^{n}\operatorname{Im} m(z_{k_i})\frac{C}{\eta}\left(f_s^{\mathrm{Fer}}(\mathbf{k}^{\mathbf{m}}) - \left(\frac{\sqrt{|I|}}{N}\right)^{n} A_n\right)$$
$$+\mathcal{O}\left(\frac{N^{(n+3)\xi}}{\eta}\left(\frac{1}{N\eta} + \sqrt{\frac{|I|}{s^{3(n-1)/2}N\eta}} + \frac{1}{\sqrt{s^{3(n-1)/2}|I|}} + \frac{|I|}{\sqrt{s^{3(n-2)/2}N\eta}} + N^{-\vartheta_{n-2}}\right)\left(\frac{\sqrt{|I|}}{N}\right)^{n}\right).$$

So that, by taking $\eta = sN^{-\varepsilon}$ for some small $\varepsilon > 0$, we have by Gronwall's lemma,

$$f_s^{\mathrm{Fer}}(\mathbf{k}^{\mathbf{m}}) = \left(\frac{\sqrt{|I|}}{N}\right)^{n}\Bigg(A_n$$
$$+\mathcal{O}\left(N^{(n+3)\xi}\left(\frac{N^{\varepsilon+(n+3)\xi}}{Ns} + \sqrt{\frac{|I|N^{\varepsilon}}{Ns^{(3n-1)/2}}} + \frac{1}{\sqrt{s^{3(n-1)/2}|I|}} + \frac{|I|N^{\varepsilon}}{\sqrt{Ns^{(3n-4)/2}}} + N^{-\vartheta_{n-2}}\right)\right)\Bigg).$$

Thus taking $\varepsilon > 0$ small enough and s such that

$$s \geqslant N^{\varepsilon}\max\left\{\left(N^{\varepsilon}\frac{|I|}{N}\right)^{\frac{2}{3n-1}}, \frac{1}{|I|^{\frac{2}{3(n-1)}}}, \left(\frac{|I|^2 N^{\varepsilon}}{N}\right)^{\frac{2}{3n-4}}, \frac{N^{\varepsilon}}{N}\right\} \quad \text{and} \quad s \leqslant N^{-\varepsilon}$$

we obtain the result. □

Now that we have the leading order for our Fermionic observable, we can obtain Theorem 3.1.4 for the class of matrices given by H_s for $s \in [N^{-1+\varepsilon}, N^{-\varepsilon}]$ for some $\varepsilon > 0$.

Proposition 3.3.8. *Let $\varepsilon > 0$, there exists $s \in [N^{-1+\varepsilon}, N^{-\varepsilon}]$ such that Theorem 3.1.4 holds for H_s.*

Proof. By the analysis of the Fermionic observable in Theorem 3.3.7, we know that there exists $\delta > 0$ such that

$$\mathbb{E}\left[p_{kk}p_\ell\right] - \mathbb{E}\left[p_{k\ell}^2\right] = -\frac{|I|}{N^2} + \mathcal{O}\left(\frac{|I|}{N^2}N^{-\delta}\right). \tag{3.3.7}$$

Now, while we studied our Fermionic observable it was also possible to study the Bosonic observable from [BYY18] which in the case of two particles consists of, for k and ℓ two distinct indices in $[\![1, N]\!]$,

$$f_s^{\text{Bos}}(k,\ell) = \mathbb{E}\left[p_{kk}p_{\ell\ell} + 2p_{k\ell}^2|\boldsymbol{\lambda}\right] \quad \text{and} \quad f_s^{\text{Bos}}(k,k) = \frac{1}{3}\mathbb{E}\left[p_{kk}^2|\boldsymbol{\lambda}\right]$$

and it follows the usual eigenvector moment flow so that by a similar analysis, we can obtain

$$\mathbb{E}\left[p_{kk}p_\ell\right] + 2\mathbb{E}\left[p_{k\ell}^2\right] = 2\frac{|I|}{N^2} + \mathcal{O}\left(\frac{|I|}{N^2}N^{-\delta'}\right) \tag{3.3.8}$$

for some positive δ'. So that, combining (3.3.7) and (3.3.8), we obtain our result for the eigenvector of the matrix H_s. □

Remark 3.3.9. While we used the Bosonic observable only in the case of two particles, one could do a similar analysis as for the Fermionic observable and see that the moments behaves as the $p_{k\ell}$ were independent Gaussian. Using the construction from Remark 3.1.13, if you consider G′ constructed in the same way but with independent centered unit variance Gaussian instead of the $p_{k_ik_j}$ and write $B_n(\mathbf{k}) = \mathcal{M}(\mathbf{k})^{-1}\mathbb{E}[\text{haf}\,\mathrm{G}']$ we have a similar convergence than Theorem 3.3.7

$$\left|f_s^{\text{Bos}}(\mathbf{k}) - \left(\frac{\sqrt{|I|}}{N}\right)^n B_n(\mathbf{k})\right| = \mathcal{O}\left(\left(\frac{\sqrt{|I|}}{N}\right)^n N^{-\vartheta_n'}\right).$$

3.4. Proof of Theorem 3.1.4

We have now our result for the Gaussian divisible ensemble

$$H_s(W) = e^{-s/2}W + \sqrt{1-e^{-s}}\text{GOE}$$

with s a small parameter (in particular $s \leqslant N^{-\varepsilon}$ for some ε) and any W being a generalized Wigner matrix. The point of this section is to remove the Gaussian term in order to obtain the result for our original matrix W. We will do so by using a moment matching scheme and the density of the Gaussian divisible ensemble. The main point being that we can find a generalized Wigner matrix W_0 such that $H_s(W_0)$ has the same first moments as W and finish the proof by a Green function comparison theorem. We will give this theorem now, a variant of [KY13b, Theorem 1.10] which can be found in [BY17, Theorem 5.2]. It needs as an assumption a level repulsion estimate. The following theorem states that the level repulsion estimate holds for generalized Wigner matrices, it can be found in [EY15, BEY14a].

Theorem 3.4.1 ([EY15, BEY14a])**.** *Consider W a generalized Wigner matrix and $\lambda_1 \leqslant \cdots \leqslant \lambda_N$ its ordered eigenvalues. There exists $\alpha_0 > 0$ such that for any $0 < \alpha < \alpha_0$, there exists $\delta > 0$ such that for any $E \in (-2, 2)$, see that we have $\gamma_k \leqslant E \leqslant \gamma_{k+1}$ for some $k \in [\![1, N]\!]$, we have*

$$\mathbb{P}\left(\left|\{i,\, \lambda_i \in [E - N^{-2/3}\hat{k}^{-1/3}, E + N^{-2/3}\hat{k}^{-1/3}]\}\right| \geqslant 2\right) \leqslant N^{-\alpha-\delta}$$

with $\hat{k} = \min(k, N-k+1)$.

Remark 3.4.2. Note that this result has only been technically proved in the regime where either $\hat{k} \leqslant N^{1/4}$ for the edge case or in the bulk of the spectrum. But as remarked in [BY17], this estimate can be proved to any regime of k with minor modifications in the proof.

This uniform level repulsion estimate for W allows us to use the generalization of the following Green function comparison theorem

Theorem 3.4.3 ([BY17]). *Consider W and W' two generalized wigner ensembles such that the first three moments of off-diagonal entries of W and W' are equal and that the first two moments of diagonal entries of W and W' are equal. Suppose also that there exists a positive a such that for any $i \neq j$,*

$$\left|\mathbb{E}[w_{ij}^4] - \mathbb{E}[{w'_{ij}}^4]\right| \leqslant N^{-2-a}.$$

Let $\alpha > 0$, then there exists $\varepsilon = \varepsilon(a) > 0$ such that for any $k \in \mathbb{N}$ and any $\mathbf{q}_1, \dots, \mathbf{q}_k$ and any indices $j_1, \dots, j_k \in [\![\alpha N, (1-\alpha)N]\!]$ we have

$$\left(\mathbb{E}^W - \mathbb{E}^{W'}\right) O\left(N\langle \mathbf{q}_1, u_{j_1}\rangle^2, \dots, N\langle \mathbf{q}_k, u_{j_k}\rangle^2\right) = \mathcal{O}(N^{-\varepsilon})$$

for any smooth function O with polynomial growth,

$$|\partial^m O(x)| \leqslant C(1+|x|)^C$$

for some C and for any $m \in \mathbb{N}^k$ such that $|m| \leqslant 5$.

We can now give the proof of our main result.

Proof of Theorem 3.1.4. We have proved that our result holds for any matrix from the Gaussian divisible ensemble $H_s(W_0)$ for any generalized Wigner matrix W_0 and $s \in [N^{-\delta}, N^{-\varepsilon}]$ for some positive $\delta < \varepsilon$. Now, in order to use the Green function comparison theorem for eigenvector Theorem 3.4.3, we simply need to be able to construct a matrix W_0 such that the assumption of Theorem 3.4.3 hold for the matrices $H_s(W_0)$ and W. Such a construction can be seen in [EYY11, Lemma 3.4] and the result has been proved. □

3.5. Combinatorial proof of Theorem 3.1.10

Proof of Theorem 3.1.10. First define

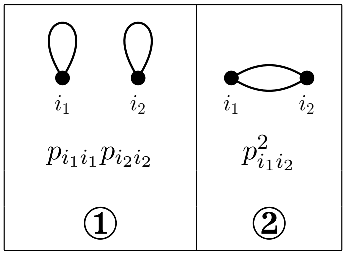
$$g(\boldsymbol{\eta}) = \sum_{\sigma \in \mathfrak{S}_n} \epsilon(\sigma) \prod_{i=1}^n p_{i_k i_{\sigma(k)}} \quad \text{so that} \quad f_s^{\mathrm{Fer}}(\boldsymbol{\eta}) = \mathbb{E}\left[g(\boldsymbol{\eta}) | \boldsymbol{\lambda}\right].$$

We therefore need to show the following two equality

$$X_{i_k j}^2 g(\boldsymbol{\eta}) = 2(g(\boldsymbol{\eta}^{i_k j}) - g(\boldsymbol{\eta})) \quad \text{for} \quad k \in \{1 \dots, n\},\, j \notin \{i_1, \dots, i_n\} \tag{3.5.1}$$

$$X_{i_k i_\ell}^2 g(\boldsymbol{\eta}) = 0 \quad \text{for} \quad k, \ell \in \{1, \dots, n\}. \tag{3.5.2}$$

Part of the reasoning will be done via induction. We will first describe the proof for two particles. For simplicity, we will describe one of the joint moments of the family $(p_{k\ell})$ as a graph corresponding to a permutation in the determinant. For instance, for two particles, we have two distinct graphs.



Thus we can write our fermionic observable for these two particles as

$$g_t(\boldsymbol{\eta}) = p_{i_1i_1}p_{i_2i_2} - p^2_{i_1i_2} = ① - ②.$$

Note that we have a sign difference between the terms because of the changing signature between these two permutations. Now, see that the generator X operates on our family of overlaps $(p_{i_ki_\ell})$ in the following way

$$\begin{gathered} X_{i_ki_\ell}p_{i_ki_k} = -2p_{i_ki_\ell} = -X_{i_ki_\ell}p_{i_\ell i_\ell},\ X_{i_ki_\ell}p_{i_ki_\ell} = p_{i_ki_k} - p_{i_\ell i_\ell} \\ X_{i_ki_\ell}p_{i_kj} = -p_{i_\ell j},\ X_{i_ki_\ell}p_{i_\ell j} = p_{kj}. \end{gathered} \tag{3.5.3}$$

With these algebraic relations, one can easily see that we have (3.5.1) for $g_t(\boldsymbol{\eta})$. One can also easily deduce (3.5.2) with these simple relations, however, in order to explain the more detailed approach of the case of n particles, we will disclose the proof in more details. We can first operate X on both of the terms in $g_t(\boldsymbol{\eta})$ and see that

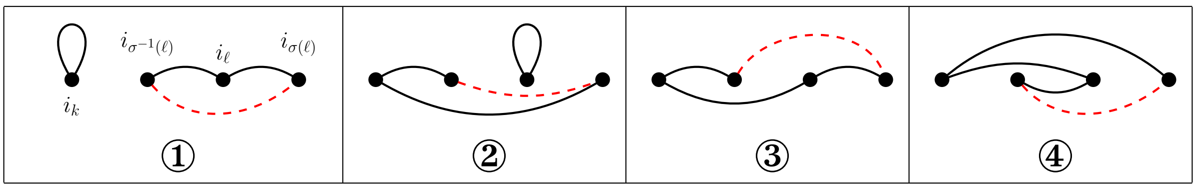
$$X^2_{i_1i_2}\left(p_{i_1i_1}p_{i_2i_2}\right) = 2\left(p^2_{i_1i_1} + p^2_{i_2i_2} - 2(p_{i_1i_1}p_{i_2i_2} + 2p^2_{i_1i_2})\right) = X^2_{i_1i_2}p^2_{i_1i_2}. \tag{3.5.4}$$

First see that (3.5.4) gives us that $X^2_{i_1i_2}g_t(\boldsymbol{\eta}) = 0$ and the proof for two particles is clear. However, to introduce the notations we will use in the case of n particles, we will write (3.5.4) as the following, using the graphical representation from the previous table,

$$X^2_{i_1i_2}① = 2\left(p^2_{i_1i_1} + p^2_{i_2i_2} - 2(① + 2②)\right) = X^2_{i_1i_2}②.$$

Consider now the case of n particles on $\{i_1, \dots, i_n\}$. By induction, we can only look at the permutations in the sum where either $\ell(i_k) + \ell(i_\ell) = n$ or $\ell(i_k) = \ell(i_\ell) = n$ where $\ell(j)$ is the length of the cycle containing j. Note that the second condition is there to take in account the fact that i_k and i_ℓ can be in the same cycle. Also see that by definition of $X_{i_ki_\ell}$, we will only be interested in the sites i_k, $i_{\sigma(k)}$, $i_{\sigma^{-1}(k)}$, i_ℓ, $i_{\sigma(\ell)}$ and $i_{\sigma^{-1}(\ell)}$.

First consider the permutations such that $\ell(i_k)$ or $\ell(i_\ell)$ is equal to 1, such a permutation wil be represented by ① or ② in the following table.



These four graphs will be the one involved when applying X^2 to ①. Note that while we display $i_{\sigma^{-1}(\ell)}$ and $i_{\sigma(\ell)}$ as distinct points, they could potentially be the same. The dashed red line represent the rest of the permutation, note also that we have two distinct cycles for the graphs ① and ② while there is a single cycle for the graphs ③ and ④ so that $\epsilon(①) = \epsilon(②) = -\epsilon(③) = -\epsilon(④)$.For simplicity, consider the notations

$$P_\ell^{(1)} = p_{i_\ell i_\ell} p_{i_{\sigma^{-1}(\ell)} i_\ell} p_{i_\ell i_{\sigma(\ell)}},$$
$$P_k^{(1)} = p_{i_k i_k} p_{i_{\sigma^{-1}(\ell)} i_k} p_{i_k i_{\sigma(\ell)}}.$$

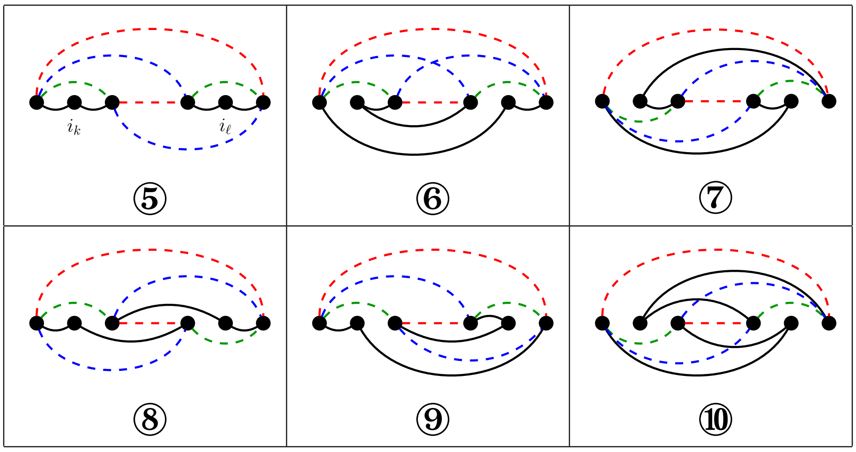
Now, using the relations (3.5.3) we obtain

$$X_{i_k i_\ell}^2 ① = 2\left(P_\ell^{(1)} + P_k^{(1)} - (2① + 2③ + 2④)\right), \qquad X_{i_k i_\ell}^2 ② = 2\left(P_\ell^{(1)} + P_k^{(1)} - (2② + 2③ + 2④)\right)$$
$$X_{i_k i_\ell}^2 ③ = 2\left(P_\ell^{(1)} + P_k^{(1)} - (① + ② + 3③ + ④)\right), \quad X_{i_k i_\ell}^2 ④ = 2\left(P_\ell^{(1)} + P_k^{(1)} - (① + ② + ③ + 3④)\right)$$

So that finally, taking in account the different signatures, we finally have

$$X_{i_k i_\ell}^2 (2① + 2② - 2③ - 2④) = 0.$$

Note that we have a coefficient of 2 in front of each graph because both σ and σ^{-1} follows the same graph. Now, we will consider permutations where $\ell(i_k)$ and $\ell(i_\ell)$ are greater than 1. Thus we will consider such a permutation as the graph ⑤ in the following table.



In this table, we represented all the permutations which will be relevant when applying $X_{i_k i_\ell}$ to a general permutation of type ⑤. The different colors explains the different behavior of the permutation on the rest of the sites that will not be seen by the operator X but will be relevant when counting the signatures on the different graphs. We first introduce the following notations as earlier

$$P_\ell^{(2)} = p_{i_{\sigma^{-1}(k)} i_\ell} p_{i_\ell i_{\sigma(k)}} p_{i_{\sigma^{-1}(\ell)} i_\ell} p_{i_\ell i_{\sigma(\ell)}},$$
$$P_k^{(2)} = p_{i_{\sigma^{-1}(k)} i_k} p_{i_k i_{\sigma(k)}} p_{i_{\sigma^{-1}(\ell)} i_k} p_{i_k i_{\sigma(\ell)}}.$$

Now, if we apply $X_{i_k i_\ell}$ to a permutation such that the cycle of i_k and of i_ℓ are greater than 1 we obtain the following set of equations:

$$X_{i_k i_\ell}^2 ⑤ = 2\left(P_k^{(2)} + P_\ell^{(2)} - (2⑤ + ⑥ + ⑦ + ⑧ + ⑨)\right),$$
$$X_{i_k i_\ell}^2 ⑥ = 2\left(P_k^{(2)} + P_\ell^{(2)} - (2⑥ + ⑤ + ⑦ + ⑧ + ⑩)\right),$$
$$X_{i_k i_\ell}^2 ⑦ = 2\left(P_k^{(2)} + P_\ell^{(2)} - (2⑦ + ⑤ + ⑥ + ⑨ + ⑩)\right),$$
$$X_{i_k i_\ell}^2 ⑧ = 2\left(P_k^{(2)} + P_\ell^{(2)} - (2⑧ + ⑤ + ⑥ + ⑨ + ⑩)\right),$$
$$X_{i_k i_\ell}^2 ⑨ = 2\left(P_k^{(2)} + P_\ell^{(2)} - (2⑨ + ⑤ + ⑦ + ⑧ + ⑨)\right),$$
$$X_{i_k i_\ell}^2 ⑩ = 2\left(P_k^{(2)} + P_\ell^{(2)} - (2⑩ + ⑥ + ⑦ + ⑧ + ⑩)\right),$$

Now, in order to put all these equations together, one needs to see the number of permutations following these graphs and their respective signature. Both of these values depend on the number of cycles, which is equal to 1 or 2 in these cases, of the permutations and thus depend on the corresponding color in the previous table. We finally have

$$\begin{aligned}
&\text{Green case:} \quad \epsilon(⑤)X_{i_k i_\ell}^2(2⑤ - ⑥ - ⑦ - ⑧ - ⑨ + 2⑩) = 0.\\
&\text{Red Case:} \quad \epsilon(⑥)X_{i_k i_\ell}^2(2⑥ - ⑤ - ⑦ - ⑧ + 2⑨ - ⑩) = 0.\\
&\text{Blue Case:} \quad \epsilon(⑦)X_{i_k i_\ell}^2(2⑦ - ⑤ - ⑥ + 2⑧ - ⑨ - ⑩) = 0.
\end{aligned}$$

Combining this result with the case where $\ell(i_k)$ or $\ell(i_\ell)$ is equal to 1 gives us the result for any permutation which finally gives

$$X_{i_k i_\ell}^2 g_t(\boldsymbol{\eta}) = 0.$$

□

3.6. Case of Hermitian matrices

In this paper, we focused and developed the proof for symmetric random matrices, but the proof holds for Hermitian matrices as well. While the maximum principle technique can clearly be directly applied to the Hermitian case, we will focused here in the definition of the Fermionic observable for the Hermitian Dyson Brownian motion. The Dyson vector flow in this case have a different generator and it is not necessarily clear that the determinant will still. Indeed, the Bosonic observable has a different form for Hermitian matrices [BYY18, Appendix] since we obtain the permanent of a matrix instead of a Hafnian. We will now give the Dyson flow of eigenvalues and eigenvectors for Hermitian matrices.

Definition 3.6.1. Let B be a Hermitian $N \times N$ matrix such that $\operatorname{Re} B_{ij}, \operatorname{Im} B_{ij}$ for $i < j$ and $B_{ii}/\sqrt{2}$ are standard independent brownian motions. The Hermitian Dyson Brownian motion is given by the stochastic differential equation

$$\mathrm{d}H_s = \frac{\mathrm{d}B_s}{\sqrt{2N}} - \frac{1}{2}H_s \mathrm{d}t. \tag{3.6.1}$$

Besides, it induces the following dynamics on eigenvalues and eigenvectors,

$$\mathrm{d}\lambda_k = \frac{\mathrm{d}\widetilde{B}_{kk}}{\sqrt{2N}} + \left(\frac{1}{N}\sum_{\ell \neq k} \frac{1}{\lambda_k - \lambda_\ell} - \frac{\lambda_k}{2}\right)\mathrm{d}s, \tag{3.6.2}$$

$$\mathrm{d}u_k = \frac{1}{\sqrt{2N}}\sum_{\ell \neq k} \frac{\mathrm{d}\widetilde{B}_{k\ell}}{\lambda_k - \lambda_\ell} u_\ell - \frac{1}{2N}\sum_{\ell \neq k} \frac{\mathrm{d}s}{(\lambda_k - \lambda_\ell)^2} u_k \tag{3.6.3}$$

where $\widetilde{B}$ is distributed as B.

The generator for the Hermitian Dyson vector flow is also known and given in the following proposition.

Proposition 3.6.2 ([BY17])**.** *The generator acting on smooth functions of the diffusion (3.6.3) is given by*

$$L_t = \frac{1}{2}\sum_{1 \leqslant k < \ell \leqslant N} \frac{1}{N(\lambda_k - \lambda_\ell)^2}\left(X_{k\ell}\overline{X}_{k\ell} + \overline{X}_{k\ell}X_{k\ell}\right) \tag{3.6.4}$$

with the operator $X_{k\ell}$ defined by

$$X_{k\ell}=\sum_{\alpha=1}^{N}\left(u_k(\alpha)\partial_{u_\ell(\alpha)}-\overline{u}_\ell(\alpha)\partial_{\overline{u}_\ell(\alpha)}\right)\quad \textit{and}\quad \overline{X}_{k\ell}=\sum_{\alpha=1}^{N}\left(\overline{u}_k(\alpha)\partial_{\overline{u}_\ell(\alpha)}-u_\ell(\alpha)\partial_{u_\ell(\alpha)}\right)$$

We will show that the determinant of fluctuations is again an observable which follows the Fermionic eigenvector moment flow. In the Hermitian case, if one considers $(u_1,\dots,u_N)$ the eigenvectors associated to the eigenvalues $\lambda_1\leqslant\cdots\leqslant\lambda_N$ of H_s given by (3.6.1), we will define the fluctuations and mixed overlap by, for a family $(\mathbf{q}_\alpha)_{\alpha\in I}\in(\mathbb{R}^N)^{|I|}$,

$$p_{kk}=\sum_{\alpha\in I}|\langle\mathbf{q}_\alpha,u_k\rangle|^2-\frac{|I|}{2N}\quad\text{and}\quad p_{k\ell}=\sum_{\alpha\in I}\langle\mathbf{q}_\alpha,u_k\rangle\langle\mathbf{q}_\alpha,\overline{u}_\ell\rangle\ \text{ for }\ k\neq\ell.$$

Note in particular that we have $p_{k\ell}\neq p_{\ell k}$ but $p_{k\ell}=\overline{p_{\ell k}}$. Now, we will define the same observable, for $\mathbf{k}=(k_1,\dots,k_n)$, with $k_i\neq k_j$,

$$f_s^{\mathrm{Fer}}(\mathbf{k})=\mathbb{E}\left[\det P_s(\mathbf{k})|\boldsymbol{\lambda}\right]\tag{3.6.5}$$

with $P_s(\mathbf{k})$ given by (3.1.16), note that it becomes a Hermitian matrix instead of a symmetric matrix in the symmetric case. We then have the same fact that f_s^{Fer} follows the eigenvector moment flow.

Theorem 3.6.3. *Let $(\mathbf{u},\boldsymbol{\lambda})$ be the solution to the coupled flows as in Definition 3.6.1 and let f_s^{Fer} be as in (3.6.5), it satisfies the following equation, for $\mathbf{k}$ a pairwise distinct set of indices such that $|\mathbf{k}|=n$,*

$$\partial_s f_s^{\mathrm{Fer}}(\mathbf{k})=\sum_{i=1}^{n}\sum_{\substack{\ell\in[\![1,N]\!]\\ \ell\notin\{k_1,\dots,k_n\}}}\frac{f_s^{\mathrm{Fer}}(\mathbf{k}^i(\ell))-f_s^{\mathrm{Fer}}(\mathbf{k})}{N(\lambda_{k_i}-\lambda_\ell)^2}\tag{3.6.6}$$

The proof of Theorem 3.6.3 can also be done using Grassmann variables and a Fermionic Wick theorem as in Section 3.2 or by carefully expanding the determinant and following the contribution of each permutation as in Appendix 3.5. We will not develop the proof here as it is very similar but it is interesting to note that the determinant and the Fermionic eigenvector moment flow is universal regarding the symmetry of the system contrary to the Bosonic observable. Indeed, we saw the definition of the Bosonic observable via (3.1.19) or a matricial representation in Remark 3.1.13 for the symmetric Dyson flow, but the Bosonic observable in the Hermitian case is different.

While we can also define it as a sum over (colored) graphs similarly to (3.1.19) another possible definition can be given in the following way: Let $\boldsymbol{\eta}$ be a configuration of n particles, denote the position of the sites where each particle is situated as $(k_1,\dots,k_n)$ (note that we can have $k_i=k_j$ for some i's and j's) then we can define

$$f^{\mathrm{Bos}}(\boldsymbol{\eta})=\frac{1}{\mathcal{M}(\boldsymbol{\eta})}\mathbb{E}\left[\operatorname{per}P_s(\boldsymbol{\eta})|\boldsymbol{\lambda}\right]\quad\text{with}\quad P_s(\boldsymbol{\eta})=\left(p_{k_ik_j}\right)_{1\leqslant i,j\leqslant n}\quad\text{and}\quad\mathcal{M}(\boldsymbol{\eta})=\prod_{i=1}^{N}\eta_i!$$

where per denote the permanent of the matrix,

$$\operatorname{per}A=\sum_{\sigma\in\mathfrak{S}_n}\prod_{i=1}^{n}A_{i,\sigma(i)}.$$

Part II

Eigenvalues of nonlinear matrix models

Chapter 1

Introduction

In the previous part of this thesis, we studied Wigner-type random matrices: symmetric matrices with independent entries. While these type of matrices have a lot of applications in mathematical physics, other models of random matrices occur in multivariate statistics. Sample covariance matrices were actually the first model of random matrices introduced in mathematics in the seminal work of Wishart [Wis28]. We give a description of this model in the next section as well as a presentation of the method of moments to compute the asymptotic empirical eigenvalue distribution.

1.1. Sample covariance matrices

1.1.1. Wishart distribution

We begin by introducing the corresponding integrable model called the Wishart distribution. Let Σ be a symmetric positive definite matrix and X be a $n \times p$ random matrix where each column X_i for $i \leqslant n$ are independently distributed according to a p-variate centered Gaussian distribution with covariance Σ. We then consider the following $p \times p$ matrix

$$M = \sum_{i=1}^{n} X_i X_i^* = XX^*$$

and say that M is distributed according to the Wishart distribution with n degrees of freedom and write $M \sim W_p(\Sigma, n)$. This construction actually gives the following joint distribution of the entries M_{ii} and M_{ij}.

Theorem 1.1.1. *Let Σ be a symmetric positive definite $p \times p$ matrix and $n \geqslant p$. If $M \sim W_p(\Sigma, n)$ then M has the following probability density function (according to the Lebesgue distribution)*

$$f_W(M) = \frac{1}{2^{np/2} \det(\Sigma)^{n/2} \Gamma_p\left(\frac{n}{2}\right)} \det(M)^{(n-p-1)/2} e^{-\frac{1}{2}\Sigma^{-1}M}$$

where the multivariate Gamma distribution Γ_p is defined by

$$\Gamma_p\left(\frac{n}{2}\right) = \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma\left(\frac{n+1-i}{2}\right).$$

Now, one can see that if we consider $\Sigma = \mathrm{Id}_p$ the probability density function becomes a symmetric function of the eigenvalues of the matrix M. If we denote $\lambda_1, \dots, \lambda_p$ the eigenvalues of the matrix M

they have the following joint eigenvalue distribution

$$\frac{1}{Z_{n,p}}\prod_{i=1}^{p}\lambda_i^{(n-p-1)/2}e^{-\frac{1}{2}\sum_{i=1}^{p}\lambda_i}\prod_{i<j}|\lambda_i-\lambda_j|.$$

From the joint eigenvalue distribution above, it is possible to compute the asymptotic empirical spectral distribution of the Wishart distribution. It was actually computed for general distribution of entries in [MP67, Wac78].

Theorem 1.1.2 ([MP67]). *Suppose that $M \sim W_p(\mathrm{Id}_p/n, n)$ and that $p/n \to \gamma \geqslant 1$. Denote $\lambda_1, \ldots, \lambda_p$ the eigenvalues of M and its empirical spectral distribution $\mu_{n,p} = \frac{1}{p}\sum_{i=1}^{p}\delta_{\lambda_i}$, we have the following weak convergence*

$$\mu_{n,p} \to \mu_\gamma := \frac{\sqrt{(\lambda_+ - x)_+(x-\lambda_-)_+}}{2\pi x\gamma}\mathrm{d}x \text{ almost surely} \tag{1.1.1}$$

with

$$\lambda_+ = (1+\sqrt{\gamma})^2 \quad \text{and} \quad \lambda_- = (1-\sqrt{\gamma})^2.$$

Note that if we have $\gamma < 1$, we have a mass at zero since the rank of the matrix M is n, there is $p-n$ zero eigenvalues which corresponds asymptotically to a mass of $(1-\gamma)$ at 0 for the limiting empirical spectral distribution, in other words the limiting eigenvalue distribution is given by

$$\mu_{n,p} \to (1-\gamma)\delta_0 + \mu_\gamma.$$

The convergence of the empirical eigenvalue distribution does not imply the convergence of the largest (or smallest) eigenvalue to the support. These convergences were proved in more generality than Gaussian entries in several papers: the case of the largest eigenvalue was first proved in [Gem80, YBK88], for the convergence of the smallest singular value of X, it was first proved in the Gaussian case in [Sil85]. To show these types of convergence, the method of moments is often used and is explained in the next subsection.

For local eigenvalue statistics, it has been proved that correlation functions as defined in (1.1.2) converges after rescaling to different determinantal processes for the Wishart distribution: in the bulk of the spectrum, via the study of classes of orthogonal polynomials [NW91, NW92a, NW92b] and for the largest eigenvalue [Joh00, Joh01b] , it has the same asymptotics as the Gaussian Orthogonal Ensemble. For the smallest singular value there are two distinct behaviors depending on the presence of eigenvalues at zeros, in other words if $\gamma = 1$ (called the hard edge case) or $\gamma > 1$ (the soft edge case). For $\gamma > 1$ there are no eigenvalues at zeros and the fluctuations of the smallest eigenvalue is still given by the Tracy-Widom distribution with the corresponding scaling [For93, TW94a]. For the hard edge case, the behavior is different as the scaling is now p^2, the limiting distribution of the smallest singular value is given by an exponential law if $p = n$ [Ede88] and the distribution of the m smallest singular values can be described using Bessel kernels when $n = p + \alpha$ for any fixed α [For93, TW94b]. We state this result here for complex Gaussian sample covariance matrices and for $n = p$ for simplicity

Theorem 1.1.3 ([Ede88, For93]). *Let $m \geqslant 1$ be a fixed integer and $\sigma_1 \leqslant \cdots \leqslant \sigma_r$ be the singular values of $\frac{1}{\sqrt{n}}X \in \mathbb{C}^{n\times n}$. For any bounded symmetric function $f : \mathbb{R}^m \to \mathbb{R}$ we have*

$$\mathbb{E}\left[\sum_{1\leqslant i_1\leqslant \ldots\leqslant i_m\leqslant n} f(4n^2\sigma_{i_1}^2, \ldots, 4n^2\sigma_{i_m}^2)\right] \longrightarrow \int_{\mathbb{R}^n} f(t_1,\ldots,t_m)\det(K_{\mathrm{Bess}}(t_i,t_j))_{i,j=1}^m \mathrm{d}t_1\ldots\mathrm{d}t_m$$

where K_{Bess} is given in terms of Bessel function by

$$K_{\mathrm{Bess}}(t,s) = \frac{J_0(\sqrt{t})\sqrt{s}J_1(\sqrt{s}) - J_1(\sqrt{s})\sqrt{t}J_0(\sqrt{t})}{2(t-s)} \quad \text{with} \quad J_k(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-\mathrm{i}(ks - t\sin s)}\mathrm{d}s$$

For the smallest singular value we have the exact formula,

$$\mathbb{P}\left(n^2\sigma_1^2 \leqslant t\right) = 1 - e^{-t}.$$

While convergence of the largest eigenvalue to the edge of the spectrum was proved for $\Sigma = \mathrm{Id}$ a remarkable phase transition occur when deviating from this model. This so-called Baik−Ben Arous−Péché phase transition discovered in [BBAP05] describes the possible separation of the largest eigenvalue from the bulk of the spectrum. We here describe the result in the simplest form but note that the result in [BBAP05] and subsequent results of this type are stronger. Since we state the result as in [BBAP05], one should consider the complex Wishart distribution instead of the real one,

Theorem 1.1.4 ([BBAP05])**.** *Let ℓ_1 be a real number and we take for our covariance matrix*

$$\Sigma = \mathrm{diag}(\ell_1, 1, \dots, 1).$$

We then have three distinct behaviors for the largest eigenvalue depending on the value of ℓ_1.

1. *If $\ell_1 < 1 + \sqrt{\gamma}$, then for any $x \in \mathbb{R}$,*

$$\mathbb{P}\left(\frac{p^{2/3}}{\sigma_1}\left(\lambda_n - (1+\sqrt{\gamma})^2\right) \leqslant x\right) \to \mathrm{TW}_2(x) \quad \textit{with} \quad \sigma_1 = \sqrt{\gamma}\left(1 + \frac{1}{\sqrt{\gamma}}\right)^{4/3}$$

where TW_2 is the asymptotic probability density function of the largest eigenvalue of the GUE.

2. *If $\ell_1 = 1 + \sqrt{\gamma}$, then for any $x \in \mathbb{R}$,*

$$\mathbb{P}\left(\frac{p^{2/3}}{\sigma_1}\left(\lambda_n - (1+\sqrt{\gamma})^2\right) \leqslant x\right) \to (\mathrm{TW}_1(x))^2$$

where TW_1 is the asymptotic probability density function of the largest eigenvalue of the GOE.

3. *If $\ell_1 > 1 + \sqrt{\gamma}$, then for any $x \in \mathbb{R}$,*

$$\mathbb{P}\left(\frac{\sqrt{p}}{\sigma_2}\left(\lambda_n - \left(\ell_1 + \frac{\gamma\ell_1}{(\ell_1 - 1)}\right)^2\right) \leqslant x\right) \to \mathbb{P}\left(\mathcal{N}(0,1) \leqslant x\right) \quad \textit{with} \quad \sigma_2 = \sqrt{\ell_1^2 - \frac{\gamma\ell_1^2}{(\ell_1-1)^2}}.$$

One can see that while the largest eigenvalue still converges to the edge of the Marčenko-Pastur distribution for $\ell_1 \leqslant 1 + \sqrt{\gamma}$ and has fluctuations of order $p^{2/3}$, the largest eigenvalue separate from the rest of the spectrum when $\ell_1 > 1 + \sqrt{\gamma}$ and have Gaussian fluctuations of order $p^{1/2}$.

There has been numerous generalizations of this phenomenon and we do not give a comprehensive list of these results. For real sample covariance matrices, the locations of the largest eigenvalue after a finite-rank deformation was studied in [BS06] regardless of the entry distribution and the fluctuations were given in [Pau07] for Gaussian matrices. Detailed asymptotics for finite n and p were given in [Nad08]. This phase transition also occurs for the Wigner ensembles: it was first proved in [Péc06] for Hermitian Gaussian Wigner matrices deformed by a rank one perturbation. In [FP07], both the symmetric and Hermitian case were considered for rank one deformation for entries with symmetric distribution. This result was generalized in [CDMF09] for any finite rank deformation and the non-universality of fluctuations of eigenvalues separated from the spectrum was exhibited. The distribution of fluctuations actually depends on the entry distribution beyond its second moment for some specific deformation. A general phase transition result which holds for numerous models of random matrices was later obtained in [BGN11].

1.1.2. Universality results

Universality with respect to matrix entries. The previous subsection focused on the Wishart distribution which corresponds to the integrable model for sample covariance matrices. The limiting eigenvalue distribution given in Theorem 1.1.2 was showed in [MP67] for more general random variables. Indeed, it holds for matrices with centered random variables with finite second moment [Wac78, Yin86] which is optimal (as the distribution depends on the second moment of the matrix entries). The convergence of the largest eigenvalue to the edge of the support was already proved for broader random variables than Gaussian distribution in [Gem80, YBK88]. However, the convergence of the smallest singular value was proved for general random variables in [BY93].

Universality results for local eigenvalue statistics in the bulk of the spectrum have been proved in the similar fashion than the Wigner ensembles explained in Part I. The method introduced in [Joh01a] using explicit formula has been used for complex sample covariance matrices in [BAP05] in order to show bulk universality for deformation of order one of sample covariance matrices by a Gaussian matrix. This deformation was then removed in [Péc12] using similar exact fomula in the complex case. Bulk universality was also proved for covariance matrices with a moment matching condition in [TV12a]. The dynamical method presented in the previous part was also generalized to both real and complex sample covariance matrices in [ESYY12, PY14].

When considering the largest eigenvalue, the scheme introduced by Soshnikov in [Sos99] for Wigner ensembles was applied to sample covariance matrices in [Sos02] to obtain universality for the fluctuations of the largest eigenvalue when $\gamma = 1 + \mathcal{O}(N^{-1/3})$ and for symmetric entry distribution. The condition on γ was then removed in [Péc09] where the Tracy-Widom fluctuations of the largest eigenvalue was proved for any $\gamma \in (0, \infty)$. The condition on the symmetry of the entry was removed in [PY14] using a Green function comparison theorem.

To study the fluctuations of the smallest eigenvalue, we saw that different behaviors occur for the Wishart distribution when γ is equal to 1 (hard edge) or not (soft edge). For the soft edge case, under symmetry of entry distribution, the Tracy-Widom fluctuations were proved in [FS10] for the smallest eigenvalue and this condition was also removed in [Wan12, PY14]. For the hard edge case, universality for the smallest singular value was proved in [TV10] using ideas coming from combinatorics and theoretical computer science showing that the distribution computed in [Ede88] is universal.

Universality with respect to Σ. We stated here all the results for $\Sigma = \mathrm{Id}$ corresponding to uncorrelated population entries. The seminal work of Marčenko and Pastur [MP67] gives the asymptotic eigenvalue distribution as the solution of an integral equation depending on the spectrum Σ.

Theorem 1.1.5. *Under some assumptions on the covariance matrix Σ, denote $(\sigma_1, \ldots, \sigma_p)$ its eigenvalues and its empirical eigenvalue distribution*

$$\pi_p = \frac{1}{p}\sum_{i=1}^{p} \sigma_i.$$

Let m be the number of distinct positive eigenvalues of Σ and denote these eigenvalues $\{s_1 > \cdots > s_m\}$. The asymptotic deterministic eigenvalue density ϱ_n, characterized by its Stieltjes transform m_n, is given by the following equation, for each $z \in \mathbb{C}_+$ the upper half-plane,

$$z = \gamma \sum_{i=1}^{m} \frac{\pi_p(\{s_i\})}{m_n(z) + \frac{1}{s_i}} - \frac{1}{m_n(z)} \quad \textit{with the condition} \quad \operatorname{Im} m(z) > 0. \tag{1.1.2}$$

Note that the asymptotic eigenvalue density here though deterministic depends on n as the equation itself depends on n. In the case where π_p converges to some distribution we can write the exact asymptotic global density. For instance, in the case where $\Sigma = \mathrm{Id}_p$ we saw that we can compute the solution of (1.1.2) and obtain (1.1.1). Thus, while the exact formula from Theorem 1.1.2 is not universal, we have a universal way with respect to the covariance structure to characterize the global density of eigenvalues.

The behavior at the soft edge of the spectrum of this model has been widely studied beyond the case where $\Sigma = \mathrm{Id}$. Note that we do not consider the possible eigenvalue which separates from the spectrum here. In the case of the Wishart distribution when entries are Gaussian, Tracy-Widom fluctuations were first proved for the largest eigenvalue for complex entries in [EK07, Ona08] using the explicit Harish-Chandra-Itzykson-Zuber formula. For complex Gaussian entries, the case of all soft and hard edges were studied in [HHN16]. For real Gaussian entries, Tracy-Widom fluctuations for the largest eigenvalue were proved in [LS16] using a comparison scheme with the uncorrelated model $\Sigma = \mathrm{Id}$. This paper also gives the result for general real random variables but diagonal Σ. For random variables which follows a moment matching condition and general Σ, a similar Green function comparison theorem as in [TV12a] was used to show Tracy-Widom fluctuations for the largest eigenvalue in [BPZ15]. Finally, for general random variables and for general Σ, universality at every soft edge was proved in [KY17].

1.2. The method of moments

In this section we give a brief idea on how to use the method of moments for sample covariance matrices to prove Theorem 1.1.2 for a rather general class of entry distribution. The next chapter is based around this method for a more intricate model of random matrices. The strategy of the method of moments is to show the convergence of the kth-moment of the empirical eigenvalue distribution to the kth moment of the Marčenko-Pastur distribution. The key observation to use this strategy is the following identity: first denote $\mu_n = n^{-1}\sum_{i=1}^n \delta_{\lambda_i/p}$ the empirical eigenvalue distribution and see that

$$m_k(\mu_n) := \int x^k \mathrm{d}\mu_n(x) = \frac{1}{np^k}\sum_{i=1}^n \lambda_i^k = \frac{1}{n}\mathrm{Tr}\left(\frac{1}{p}M\right)^k.$$

Thus we can compute asymptotic moments of the empirical eigenvalue distribution by considering tracial moments of our matrix M. The other key observation is the development of the tracial moment. Indeed we have

$$\frac{1}{n}\mathrm{Tr}\left(\frac{1}{p}M\right)^k = \frac{1}{np^k}\sum_{i_1,\dots,i_k=1}^n M_{i_1i_2}M_{i_2i_3}\dots M_{i_{k-1}i_k}M_{i_ki_1}.$$

Now, since the entries of M are given by $M_{ij} = \sum_{k=1}^p X_{ik}X_{jk}$,

$$m_k(\mu_n) = \frac{1}{np^k}\sum_{i_1,\dots,i_k=1}^n \sum_{j_1,\dots,j_k=1}^p X_{i_1j_1}X_{i_2j_1}X_{i_2j_2}X_{i_3j_2}\dots X_{i_pj_p}X_{i_1j_p}. \tag{1.2.1}$$

In order to show convergence of the (random) moment $m_k(\mu_n)$, we first prove convergence of its expectation and finish by showing that its variance is vanishing asymptotically. First, we need to know the moments of the Marčenko-Pastur distribution from Theorem 1.1.2 given by the following proposition.

Proposition 1.2.1. *Let k be an integer and μ_γ be the distribution defined by (1.1.1). We have*

$$m_k(\mu_\gamma) := \int x^k \mathrm{d}\mu_\gamma(x) = \sum_{r=0}^k \frac{\gamma^r}{r+1}\binom{k-1}{r}\binom{k}{r}.$$

Thus the first part of the proof consists in proving convergence of the expected moments of the empirical eigenvalue distribution: for any integer k,

$$\overline{m}_k(\mu_n) := \mathbb{E}\left[m_k(\mu_n)\right] \xrightarrow[n\to\infty]{} m_k(\mu_\gamma).$$

Now, we can see the product of X entries in (1.2.1) as some weight on closed paths on the complete bipartite graph $\mathscr{B}(n,p)$. The complete bipartite graph $\mathscr{B}(n,p)$ is the graph whose edges join any pairs $i \in [\![1,n]\!]$ and $j \in [\![1,p]\!]$. A closed path of length $2k$ starting (and ending) at i_0 on this graph which we denote ℓ is a sequence of indices

$$\ell = i_1, j_1, i_2, j_2, \ldots, i_k, j_k, i_1$$

so that if we define the weight

$$w(\ell) = \mathbb{E}\left[X_{i_1j_1}X_{i_2j_1}X_{i_2j_2}X_{i_3j_2}\ldots X_{i_pj_p}X_{i_1j_p}.\right]$$

and denote $\mathscr{L}$ the set of all closed path on $\mathscr{B}(n,p)$ we have

$$\overline{m}_k(\mu_n) = \frac{1}{np^k}\sum_{\ell\in\mathscr{L}} w(\ell).$$

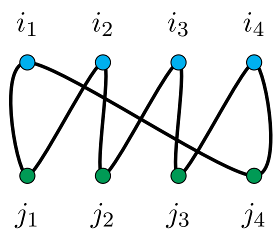


Figure 1.1: Example of a closed path of length 8 on the bipartite complete graph

Now, since the entries of X are centered and independent, in order to have a non zero weight $w(\ell)$ for a given path ℓ, each edge needs to appear at least twice in the path. Since the path is of length $2k$ and each edge appear twice, we have at most k distinct edges in the path and at most $k+1$ distinct vertices. However, if the number of vertices $m = m_1 + m_2 \leqslant k$ with m_1 indices i-labeled vertices and m_2 j-labeled vertices then the total number of choices for this indices is given by $\mathcal{O}(n^{m_1}p^{m_2})$ so that with the renormalization given by $n^{-1}p^{-k}$, the contribution of such closed paths vanishes. Finally, if we denote $\mathscr{L}_{k+1}$ the set of closed paths on the bipartite graph where each edge occurs at least twice and with $k+1$ vertices we have

$$\overline{m}_k = \frac{1}{np^k}\sum_{\ell\in\mathscr{L}_{k+1}} w(\ell) + o(1).$$

We can represent each closed path with $k+1$ vertices as an oriented double tree where one of every two vertices is either labeled by i's or by j's. This representation is given in Figure 1.2. Note that this double tree only depends on the order we read the indices and not on the actual values of indices, we call this order the *shape* of the tree. Thus, if we suppose that we have $r+1$ i-labeled vertices and

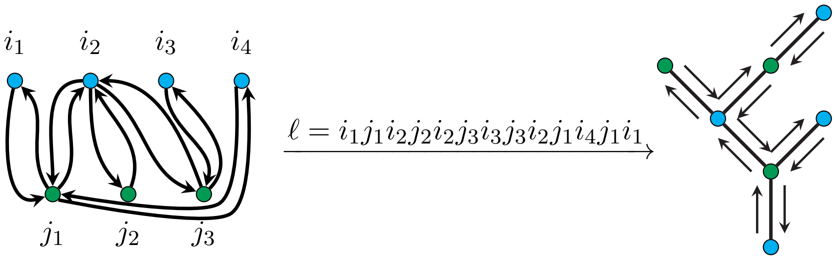


Figure 1.2: Double tree representation of a closed path on the bipartite complete graph.

$k-r$ j-labeled vertices, the total number of choices for these vertices is given by $n^{r+1}p^{k-r}(1+o(1))$ and we have

$$\begin{aligned}\overline{m}_k &= \frac{1+o(1)}{np^k}\sum_{r=0}^{k-1} n^{r+1}p^{k-r}\#\{\text{shapes of double tree with } r+1 \ i\text{-vertices and } k-r \ j\text{-vertices}\}\\ &= (1+o(1))\sum_{r=0}^{k-1}\gamma^r\#\{\text{shapes of double tree with } r+1 \ i\text{-vertices and } k-r \ j\text{-vertices}\}\end{aligned}$$

It finally only remains to show

$$\#\{\text{shapes of double tree with } r+1 \ i\text{-vertices and } k-r \ j\text{-vertices}\} = \frac{1}{r+1}\binom{k}{r}\binom{k-1}{r}.$$

In order to count these shapes of double trees we encode them bijectively with another set of paths denoted by a finite a sequence $(s_q)_{q=1}^{2k}$ with the following rules

- If q is even then $s_q \in \{0,1\}$ with $s_{2k}=0$ and if q is odd then $s_q \in \{-1,0\}$.
- $\sum_{q=1}^{m} s_q \geqslant 0$ for any $m \geqslant 1$ and $\sum_{q=1}^{2k} s_q = 0$.
- $\#\{q : s_q = 1\} = \#\{q : s_q = -1\} = r$.

The way to encode a double tree by such a sequence is done by doing the following: consider only the i-labeled vertices in the tree except for the root (the first index i_1 since its value is forced by the constraint $s_{2k}=0$); for each i-labeled vertex, mark the edge leading to it; now that the edges are marked, read the tree according to his shape and put $s_q=-1$ if the marked edge is going down and $s_q=1$ if the edge is going up in the tree. This is illustrated in Figure 1.3.

Now, the fact that we have a double tree gives the first condition on the path and since we have $r+1$ i-labeled vertices we obtain the third condition. The second condition is not as clear but it can be seen that a path not satisfying the second condition cannot give a double tree. In order to count these paths, one can see that if we do not consider the second condition (which states that the path does not go below zero) we only have to choose r steps going up among $k-1$ possible positions and r steps going down among k possible positions. Thus the total number of paths following the first and third rule is given by $\binom{k-1}{r}\binom{k}{r}$. By a reflection principle, one can count the sequences of steps which do not follow the second rule and see that they are equal to $\binom{k-1}{r-1}\binom{k}{r+1}$ so that we finally have the

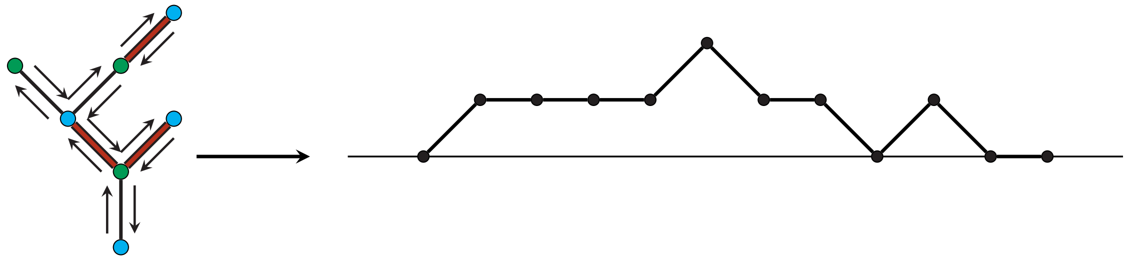


Figure 1.3: Bijection between a double tree with $r+1=4$ i-vertices and a path going from zero and back by going up $r=3$ times. We marked the edges which encode the up and down steps in the path.

identity:

$$\#\{\text{shapes of double tree with } r+1 \ i\text{-vertices and } k-r \ j\text{-vertices}\}$$
$$= \binom{k-1}{r}\binom{k}{r} - \binom{k-1}{r-1}\binom{k}{r+1} = \frac{1}{r+1}\binom{k}{r}\binom{k-1}{r}$$

and the convergence of the expected moments is done.

In order to prove convergence of the k-th moment of the empirical eigenvalue distribution we need to show a weak form of concentration of the moments toward its expectation and this can be done by simply showing that $\mathrm{Var}(m_k(\mu_n)) = o(1)$. From this, using Bienaymé-Chebyshev inequality, we can show that, for any $\varepsilon > 0$,

$$\mathbb{P}\left(|m_k(\mu_n) - m_k(\mu)| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0.$$

Note that this gives convergence in probability of the moments. To show an almost sure convergence, we can prove a stronger bound on the variance: $\mathrm{Var}(m_k(\mu_n)) = \mathcal{O}(n^{-2})$ and finish with the Borel-Cantelli lemma. The control of the variance of the moments can be done similarly by expanding the variance and counting the contributing graphs in the sum.

Chapter 2

Eigenvalue distribution of nonlinear matrix models

This chapter is based on a joint work with S. Péché [BP19].

2.1. Introduction

Deep learning has shined through a large list of succesful applications over the past five years or so (see for instance applications in image or speech recognition [KSH12, HDY+12] or translation [WSC+16]). However, the theoretical and mathematical understanding of deep learning has had a slow progress. The main difficulty comes from the complexity of studying highly non-convex functions of a very large number of parameters [CHM+15]. Thus, a possible idea to understand better such large complex systems is to approximate the elements of the system by random variables as it is done in statistical physics and thermodynamics. Indeed, even Wigner's original idea of introducing random symmetric or Hermitian matrices was to understand the energy levels of large nuclei whose study was too complicated.

In [BGC16, CBG16] for instance, a random matrix approach has been used to do a theoretical study of spectral clustering by looking at the Gram matrix WW^* where the columns of W are given by random vectors. They compute the asymptotic deterministic empirical distribution of this matrix which allows the analysis of the spectral clustering algorithm in large dimensions.

For random neural networks, there are two distinct parts of the system that are large, the number of samples m and the number of parameters n. We then need to consider rectangular matrices of size $n \times m$. Random matrix thoery usually applies to large matrices where the number of samples and that of parameters grow in the same way: one considers the regime where n/m goes to some constant ϕ. The study of random matrix models for random neural networks was done in [LLC18, PW17], it consists of nonlinear random matrix models of the form

$$M = \frac{1}{m} YY^* \in \mathbb{R}^{n_1 \times n_1} \quad \text{with} \quad Y_{ij} = f\left(\frac{1}{\sqrt{n_0}}(WX)_{ij}\right) \quad \text{for} \quad 1 \leqslant i \leqslant n_1, \quad 1 \leqslant j \leqslant m$$

where f is a nonlinear activation function, W is the $n_1 \times n_0$ matrix corresponding to the weights and X the $n_0 \times m$ matrix of the data. There are several possibilities to incorporate randomness in this model. In [LLC18], the authors considered random weights with *deterministic data*. The weights are given by functional of Gaussian and they studied eigenvalues through concentration inequalities for finite n_0, n_1 and m and the function f is Lipschitz continuous. We give the limiting asymptotic empirical eigenvalue distribution in terms of Stieltjes transform as the following theorem

Theorem 2.1.1 ([LLC18]). *Consider $(\lambda_1, \dots, \lambda_m)$ the eigenvalues of M. Its empirical eigenvalue distribution $\mu_m = \frac{1}{m}\sum_{i=1}^m \delta_{\lambda_i}$ then converges almost surely in distribution to $\bar{\mu}$ defined through its Stieltjes transform by*

$$m_{\bar{\mu}}(z) = \frac{1}{m}\mathrm{Tr}\left(\frac{n_1}{m}\frac{\overline{M}}{1+s(z)} - zI_m\right)^{-1} \quad \textit{with} \quad \overline{M} = \mathbb{E}\left[M\right]$$

and $s(z)$ is the solution such that $\mathrm{Im}\, s(z) > 0$ of

$$s(z) = \frac{1}{m}\mathrm{Tr}\left(\overline{M}\left(\frac{n_1}{m}\frac{\overline{M}}{1+s(z)} - zI_m\right)^{-1}\right)$$

Note that the dependence in f comes from the deterministic matrix $\overline{M}$. This type of eigenvalue distribution is not new to random matrix theory as it corresponds to sample covariance matrix with general population of type TXX^*T^* with T a deterministic matrix such that $TT^* = \overline{M}$ [SB95]. Thus the nonlinearity coming from applying the function f entrywise is not clearly seen in the eigenvalue distribution.

The main difference with general sample covariance matrices is the non universality of the eigenvalue distribution. Indeed, the deterministic matrix $\overline{M}$ depend on the distribution of W, beyond its first two moments. In [LLC18], a discussion is made on the effect of the fourth moments of the distribution for the efficiency of the neural networks.

On the other side, in [PW17], the randomness comes from both the matrices W and X as they are chosen to be independent random matrices with Gaussian entries. Interestingly, they obtain the asymptotic eigenvalue spectral distribution via a self-consistent equation for the Stieltjes transform of degree 4. This equation corresponds to a quartic equation. In some special cases of the parameters, one recovers the recursion for the Marčenko-Pastur with parameter ϕ/ψ:

$$zm(z)^2 + \left(\left(1 - \frac{\psi}{\phi}\right)z - 1\right)m(z) + \frac{\psi}{\phi} = 0.$$

This eigenvalue distribution is that of Wishart matrices. Thus, there exists a class of functions such that the nonlinear matrix model is simply given by a linear model with only one degree of randomness. It was then conjectured in [PW17] that choosing such an activation function could speed up training through the network. The fourth degree of the recursion relation can disappear for another class of functions and we then obtain a cubic equation. This cubic equation corresponds to the product Wishart matrix, indeed it can be seen that f is linear and the resulting matrix is simply ZZ^* with $Z = WX$. The eigenvalue distribution of such matrices has been computed in [AIK13, DC14]. However in all generality, a new type of eigenvalue distribution was given as the solution of a fourth order equation.

The aim of this paper is to study the asymptotic empirical eigenvalue distribution of such nonlinear functionals of random matrices $f(WX)$ where f is applied entrywise and to extend the result established by [PW17] to non Gaussian matrices. Such a study is also of interest in random matrix theory itself as it introduces a new class of ensembles of random matrices as well as a new class for universality. We also consider the case of several layers for a specific class of activation function.

2.2. Description of the model

Consider a random matrix $X \in \mathbb{R}^{n_0 \times m}$ a random matrix with *i.i.d* elements with distribution ν_1. Let also $W \in \mathbb{R}^{n_1 \times n_0}$ be a random matrix with *i.i.d* entris with distribution ν_2. W is called the weight matrix. Both distribution are centered and we denote the variance of each distribution by

$$\mathbb{E}\left[X_{ij}^2\right] = \sigma_x^2 \quad \text{and} \quad \mathbb{E}\left[W_{ij}^2\right] = \sigma_w^2. \tag{2.2.1}$$

We also need the following assumption on the tail of W and X: there exist constants ϑ_w, $\vartheta_x > 0$ and $\alpha > 1$ such that for any $t > 0$ we have

$$\mathbb{P}\left(|W_{11}| > t\right) \leqslant e^{-\vartheta_w t^\alpha} \quad \text{and} \quad \mathbb{P}\left(|X_{11}| > t\right) \leqslant e^{-\vartheta_x t^\alpha}. \tag{2.2.2}$$

Note that in light of the central limit theorem it gives us that there exists a constant $C > 0$ such that

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{n_0}}\sum_{k=1}^{n_0} W_{1k}X_{k1}\right| > t\right) \leqslant Ce^{-t^2/2}. \tag{2.2.3}$$

We now consider a smooth function $f : \mathbb{R} \to \mathbb{R}$ with zero Gaussian mean in the sense that

$$\int f(\sigma_w\sigma_x x)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x = 0. \tag{2.2.4}$$

As an additional assumption, we also suppose that there exists positive constants C_f and c_f and $A_0 > 0$ such that for any $A \geqslant A_0$ and any $n \in \mathbb{N}$ we have,

$$\sup_{x\in[-A,A]} |f^{(n)}(x)| \leqslant C_f A^{c_f n}. \tag{2.2.5}$$

We consider the following random matrix,

$$M = \frac{1}{m}YY^* \in \mathbb{R}^{n_1\times n_1} \quad \text{with} \quad Y = f\left(\frac{WX}{\sqrt{n_0}}\right) \tag{2.2.6}$$

where f is applied entrywise. We suppose that the dimensions of both the columns and the rows of each matrix grow together in the following sense: there exist positive constants ϕ and ψ such that

$$\frac{n_0}{m} \xrightarrow[m\to\infty]{} \phi, \quad \frac{n_0}{n_1} \xrightarrow[m\to\infty]{} \psi.$$

Now we can give the limiting empirical eigenvalue distribution of the matrix M. It consists in a deterministic compactly supported measure. Denote by $(\lambda_1, \dots, \lambda_{n_1})$ the eigenvalues of M given by (2.2.6) and

$$\mu_{n_1} = \frac{1}{n_1}\sum_{i=1}^{n_1}\delta_{\lambda_1} \tag{2.2.7}$$

its empirical eigenvalue distribution

Theorem 2.2.1. *There exists a deterministic compactly supported measure μ such that we have*

$$\mu_{n_1}^{(f)} \xrightarrow[n_1\to\infty]{} \mu \quad \textit{weakly almost surely.}$$

The moments of the asymptotic empirical eigenvalue distribution depend on the two following parameters of the function f

$$\theta_1(f) = \int f^2(\sigma_w\sigma_x x)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x \quad \text{and} \quad \theta_2(f) = \left(\sigma_w\sigma_x\int f'(\sigma_w\sigma_x x)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x\right)^2 \tag{2.2.8}$$

Theorem 2.2.2. *The measure μ is characterized through a self-consistent equation for its Stieljes transform G defined for $z \in \mathbb{C} \setminus \mathbb{R}$ by*

$$G(z) := \int \frac{\mathrm{d}\mu(x)}{x - z}, \quad \textit{denote also} \quad H(z) := \frac{\psi - 1}{\psi} + \frac{z}{\psi} G(z),$$
$$H_\phi(z) := 1 - \phi + \phi H(z) \quad \textit{and} \quad H_\psi(z) := 1 - \psi + \psi H(z)$$

We have the following fourth-order self-consistent equation,

$$H(z) = 1 + \frac{H_\phi(z) H_\psi(z)(\theta_1(f) - \theta_2(f))}{\psi z} + \frac{H_\phi(z) H_\psi(z) \theta_2(f)}{\psi z - H_\phi(z) H_\psi(z) \theta_2(f)},$$

with $\theta_1(f)$ and $\theta_2(f)$ are defined in (2.2.8).

Remark 2.2.3. The measure μ is characterized by its moments which is given in (2.3.18) below.

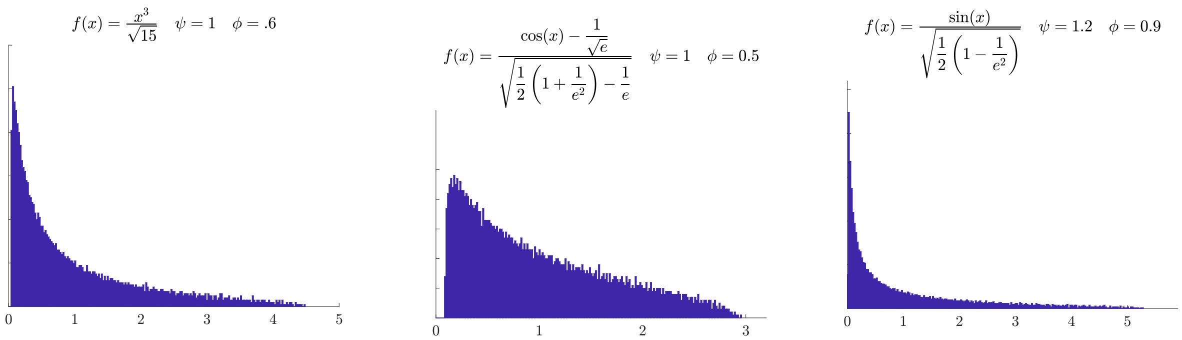


Figure 2.1: Examples of eigenvalue distribution for different activation function and parameters for Gaussian random variables. Note that the activation functions are centered and normalized so that $\theta_1(f) = 1$. In the second figure, we have $\theta_2(f) = 0$ and the asymptotic empirical eigenvalue distribution is given by the Marčenko-Pastur distribution of shape parameter $\phi/\psi/$.

The model given by (2.2.6) consists in passing the data covariance matrix through one layer of a neural network as we apply the function f a single time. However, we could put the matrix Y through the network again and obtain a new covariance matrix. It was conjectured in [PW17] that for activation functions such that $\theta_2(f) = 0$ the eigenvalue distribution stays invariant and we obtain the Marčenko-Pastur distribution at each layer. We give here a positive answer to this conjecture. We denote by L the number of layers and consider, for $p \in [\![0, L-1]\!]$ a family of independent matrices $W^{(p)} \in \mathbb{R}^{n_{p+1} \times n_p}$ where $(n_p)_p$ is a family of growing sequences of integers such that there exists $(\phi_p)_p$ and $(\psi_p)_p$ such that

$$\frac{n_0}{m} \xrightarrow[m \to \infty]{} \phi \quad \frac{n_p}{n_{p+1}} \xrightarrow[m \to \infty]{} \psi_p.$$

We suppose that all matrix entries $(W_{ij}^{(p)})_{ij}$ are $i.i.d$ with variance σ_w^2. Define the sequence of random matrices

$$Y^{(p+1)} = f\left(\frac{\sigma_x}{\sqrt{\theta_1(f)}} \frac{W^{(p)} Y^{(p)}}{\sqrt{n_p}}\right) \in \mathbb{R}^{n_{p+1} \times m} \quad \text{with} \quad Y^{(0)} = X. \tag{2.2.9}$$

The normalization is there in order to keep the same variance for Y entries along every layer. Indeed, we need to use the fact that f is centered with respect to the Gaussian distribution with the variance given by $\sigma_w^2\sigma_x^2$. Now, one can define

$$M^{(L)} = \frac{1}{m}Y^{(L)}Y^{(L)} \quad \text{and} \quad \mu_{n_L}^{(L)} = \frac{1}{n_L}\sum_{i=1}^{n_L}\delta_{\lambda_i^{(L)}}$$

where $(\lambda_k^{(L)})$ are the eigenvalues of $M^{(L)}$. We then have the following theorem under the additional assumption that the function f is bounded.

Theorem 2.2.4. *Suppose that f is a bounded smooth function such that (2.2.4) and (2.2.5) hold and $\theta_2(f) = 0$, then the asymptotic eigenvalue distribution of $\mu_{n_L}^{(L)}$ is given almost surely by the Marčenko-Pastur distribution of shape parameter $\frac{\phi}{\psi_0\psi_1\cdots\psi_{L-1}}$.*

The next section is dedicated to proving Theorem (2.2.1) for polynomials using the moment method. We first consider the case where f is a polynomial in order to compute the moments and will generalize to other functions using a polynomial approximation in Section 2.4. In Section 2.5, we show that the largest eigenvalue of our model for a single layer sticks to the edge of the support of μ by considering high moments of the matrix [FK81, Sos99, Sos02]. Finally, in Section 2.6 we first give a description of the expected moments after two layers for polynomials and then prove Theorem 2.2.4.

2.3. Moment method when f is a polynomial

The point of this section is to compute the moments of the empirical eigenvalue distribution of the matrix M when the activation is a polynomial. The following statement will compute the expected moment of the distribution in this case using a graph enumeration. We will extend the result to other functions f in Section 2.4

Theorem 2.3.1. *Let $f = \sum_{k=1}^{K}\frac{a_k}{k!}(x^k - k!!\mathbb{1}_{k\ \text{even}})$ be a polynomial such that (2.2.4) holds. The degree of f, K, can grow with n_1 but suppose that*

$$K = \mathcal{O}\left(\frac{\log n_1}{\log\log n_1}\right). \tag{2.3.1}$$

Let $\mu_{n_1}^{(f)}$ be defined in (2.2.7) and its expected moments

$$\overline{m}_q := \mathbb{E}\left[\langle\mu_{n_1}^{(f)}, x^q\rangle\right] = \mathbb{E}\left[\frac{1}{n_1}\sum_{i=1}^{n_1}\lambda_i^q\right].$$

We then have the following asymptotics

$$\overline{m}_q = \sum_{I_i,I_j=0}^{q}\sum_{b=0}^{I_i+I_j+1}\mathcal{A}(q,I_i,I_j,b)\theta_1(f)^b\theta_2(f)^{q-b}\psi^{I_i+1-q}\phi^{I_j}\left(1+o(1)\right). \tag{2.3.2}$$

where $\mathcal{A}(q,I_i,I_j,b)$ denote the number of admissible graphs with $2q$ edges, I_i i-identifications, I_j j-identifications and b cycles of size 1 defined in Definition 2.3.2 below and θ_1 and θ_2 are defined in (2.2.8).

Note that in this theorem we take the degree of polynomial to grow with n as in (2.3.1) but see that this theorem holds for any fixed q. It is possible that we can get a stronger assumption than (2.3.1) in the sense that K could grow faster with n_1. However, this bound is enough to do the polynomial approximation we will need later. Note that we will use a Taylor-Lagrange approximation on our final function f.

2.3.1. Case where f is a monomial of odd degree

We first consider the case where $f(x) = \frac{x^k}{k!}$ for k odd. In this section, we explain the combinatorics needed to compute the moments of the spectral measure of M in the case where f is an odd monomial. To that aim we need some definitions from graph theory. We will assume first that the entries of W and X are bounded in the following sense: there exists a $A > 0$ such that

$$\max_{ij} |W_{ij}| + |X_{ij}| \leqslant A \quad \text{almost surely.}$$

Basic definitions

For this activation function, the entries of $Y = f(WX/\sqrt{n_0})$ are of the form

$$Y_{ij} = \frac{1}{k!}\left(\frac{WX}{\sqrt{n_0}}\right)^k_{ij} = \frac{1}{n_0^{k/2}k!}\left(\sum_{\ell=1}^{n_0} W_{ik}X_{kj}\right)^k = \frac{1}{n_0^{k/2}k!}\sum_{\ell_1,\dots\ell_k=1}^{n_0}\prod_{p=1}^{k} W_{i\ell_p}X_{\ell_p j}. \tag{2.3.3}$$

We want to study the normalized tracial moments of the matrix M. Thus we want to consider, for a positive integer q,

$$\frac{1}{n_1}\mathbb{E}\left[\operatorname{Tr} M^q\right] = \frac{1}{n_1 m^q}\mathbb{E}\left[\operatorname{Tr}(YY^*)^q\right] = \frac{1}{n_1 m^q}\mathbb{E}\sum_{i_1,\dots,i_q=1}^{n_1}\sum_{j_1,\dots,j_q=1}^{m} Y_{i_1j_1}Y_{i_2j_1}Y_{i_2j_2}Y_{i_3j_2}\dots Y_{i_qj_q}Y_{i_1j_q} \tag{2.3.4}$$

Thus injecting (2.3.3) in the previous equation we obtain the following development of the tracial moment of M

$$\frac{1}{n_1}\mathbb{E}\left[\operatorname{Tr} M^q\right] = \frac{1}{n_1 m^q n_0^{kq}(k!)^{2q}}\mathbb{E}\sum_{i_1,\dots,i_q}^{n_1}\sum_{j_1,\dots,j_q}^{m}\sum_{\substack{\ell_1^1,\dots\ell_k^1\\ \dots\\ \ell_1^{2q}\dots\ell_k^{2q}}}^{n_0}\prod_{p=1}^{k} W_{i_1\ell_p^1}X_{\ell_p^1 j_1}\prod_{p=1}^{k} W_{i_2\ell_p^2}X_{\ell_p^2 j_1}\cdots\prod_{p=1}^{k} W_{i_1\ell_p^{2q}}X_{\ell_p^{2q} j_q} \tag{2.3.5}$$

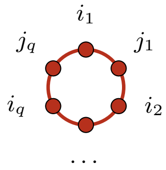
We encode each term in of the sums as a graph with $\#\{i_1,\dots,i_q,j_1,\dots,j_q\}$ red vertices and $2kq$ blue vertices. We can represent the vertices in a graph such as Figure 2.2a. Since the W_{ij} and X_{ij} are centered and independent, we need at least two of each of them in the summand in equation (2.3.5). Thus, the main contribution comes from those summands maximizing the number of pairwise distinct indices. We define these admissible graphs, corresponding to the leading order, as the following.

Definition 2.3.2. An admissible graph corresponding to a summand in (2.3.4) is a sequence of simple even cycles of *red* vertices labeled by the $\{i_1,\dots,i_q\}$ and $\{j_1,\dots,j_q\}$ such that each factor $Y_{i_{p_1}j_{p_2}}$ correspond to a *red* edge. The cycles are joined to another by a common vertex.

Remark 2.3.3. Such an admissible graph always has $2q$ *red* edges. It can also be seen as a tree of cycles also called a cactus graph. These graphs appeared also in random matrix theory in the theory of traffics when expanding injective traces [CDM16].

We call a red edge a *niche*. Each *niche* is decorated by k *blue* vertices from which leaves a blue edge corresponding to a term $W_{i\ell}X_{\ell j}$ in (2.3.5). Thus to compute the spectral moment one needs to match the blue edges so that each entry arises with multiplicity greater than 2. The matching of ℓ indices in (2.3.5) corresponds to a matching of the *blue* vertices.

As we can see each admissible graph as a tree of decorated *red* cycles, the basic figure is given by such a cycle:

Basic figure : 

For general admissible graphs, they may be coincidences among the i's or the j's.

The simplest admissible graph: a cycle of length $2q$

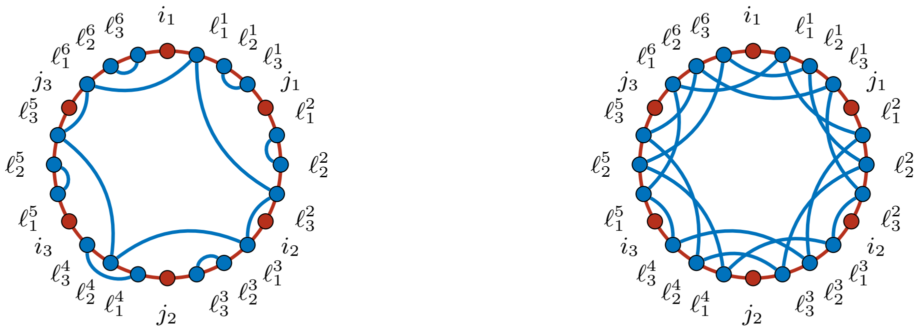
In this subsection, we assume that the i and j indices are pairwise distinct and consider the associated contribution to the spectral moment. In this case, we can really encode the products in the summand as in Figure 2.2a. Since we need at least two occurences of each matrix entry, say for instance W_{i_1,ℓ_1^1}, it needs to occur at least an other time in the product. There are then two different ways it can happen:

(i) There exists $p \in \{2, \dots, k\}$ such that $\ell_p^1 = \ell_1^1$.

(ii) There exists $p \in \{1, \dots, k\}$ such that $\ell_p^{2q} = \ell_1^1$. Applying the same reasoning for $X_{\ell_1^1, j_1}$, there exists $p' \in \{1, \dots, k\}$ such that $\ell_{p'}^2 = \ell_1^1$.

We will see that the dominant term corresponds to the maximization of the number of pairwise distinct indices. The way to have the most pairwise distinct indices without vanishing is to perform the most perfect matchings inside each niche. Note that since we supposed that k was odd, there is necessary an occurence of case (ii) which corresponds to the cycle of size $2q$ as in Figure 2.2a. Thus we can construct the graphs corresponding to the dominant term in the following way, which is illustrated in Figure 2.2a:

- Choose an index ℓ_p in each niche which will be in the only cycle of the graph.
- Do a perfect matching of the rest of the indices inside each niche.



(a) Leading order graph for $k = q = 3$

(b) Lower order graph for $k = q = 3$

Case $q > 1$. We have the following computation for the corresponding contribution on the moment, since every entry exactly occurs two times in the products, using (2.2.1), we obtain

$$E_q(k) = \frac{((\sigma_w \sigma_x)^k k(k-1)!!)^{2q} n_0}{n_1 m^q n_0^{kq} (k!)^{2q}} \frac{m!}{(m-q)!} \frac{n_1!}{(n_1-q)!} \frac{n_0!}{(n_0-(k-1)q)!}$$

To obtain this formula, note that we choose the i-labels over n_1 possible indices and the j-labels over m indices. Now, we also choose the ℓ-labels over n_0 indices but one has to be careful not to overmatch indices on adjacent niches. Finally, we have to determine the *blue* vertices forming the cycle parcouring each niche, there are k^{2q} possible ways to do so. The number of perfect matchings on the rest of the vertices in each niche is then equal to $((k-1)!!)^{2q}$.
We obtain that

$$E_q(k) = \left(\frac{(\sigma_w\sigma_x)^k k(k-1)!!}{k!}\right)^{2q} \psi^{1-q} + \mathcal{O}\left(\left(\frac{(\sigma_w\sigma_x)^k k(k-1)!!}{k!}\right)^{2q} \frac{q+k}{n_0}\right). \tag{2.3.6}$$

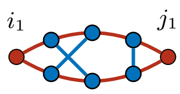
Note that we have in the case of f a monomial of odd degree, by (2.2.8), we can write

$$E_q(k) = \theta_2^q(f)\psi^{1-q} + \mathcal{O}\left(\theta_2^q(f)\frac{q+k}{n_0}\right) \quad \text{since} \quad \theta_2(f) = \left(\frac{(\sigma_w\sigma_x)^k k(k-1)!!}{k!}\right)^2.$$

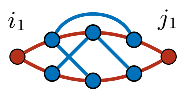
Case $q=1$. The behavior in the case where $k=1$ is slightly different and will be useful to study the general case later. Indeed in this case, we can do any perfect matching between the $2k$ *blue* vertices since there is no difference between any factor W or X in the summand in (2.3.5). The graph can bee seen in Figure 2.3a. Thus, the contribution of the moments in this case is the following

$$E_1(k) = \frac{(\sigma_w\sigma_x)^{2k}(2k)!!}{(k!)^2} + \mathcal{O}\left(\frac{k^2(2k-2)!!}{n_0(k!)^2}\right) = \theta_1(f) + \mathcal{O}\left(\frac{k^2(2k-2)!!}{n_0(k!)^2}\right)$$

where the error comes from the fact that the second order is performing a perfect matching on all sites except two and then identifying the two remaining *blue* vertices to already matched pairs.



(a) Contribution in the case where $q=1$



(b) Subleading term in the case $q=1$.

We now show that E_q is indeed the typical contribution from the basic cycle, that is all other matchings will lead to a negligible contribution with respect to E_q. There are four different phenomena that can give a lower order contribution. Firstly, there could be more than one cycle linking every niche such as Figure 2.2b. Also, in at least one niche there could be more identifications between ℓ-indices, which raises moments of entries of W and X. There could be an identification between the index of the cycle and an index from a perfect matching inside a niche. Finally, there could also be identifications between two distinct niches, note we can only get higher moments in the case where the two niches are adjacent. While these four behaviors can happen simultaneously, we will see the contribution separately since it would induce an even smaller order if counted together.

There is more than one cycle between niches. In this case, we can compute the contribution $E_q^{(1)}$ on the moments in the following way. Suppose there are c cycles. Note that necessarily c is odd since k is odd and entries are centered, then we can write, if we suppose that indices ℓ not in cycles are being perfectly matched,

$$E_q^{(1)} = \frac{(k^c(k-c)!!)^{2q}}{n_1 m^q n_o^{kq}(k!)^{2q}} \sum_{\substack{i_1,\dots,i_q \\ \text{pairwise} \\ \text{distinct}}}^{n_1} \sum_{\substack{j_1,\dots,j_q \\ \text{pairwise} \\ \text{distinct}}}^{m} \sum_{\ell_0,\dots,\ell_c}^{n_0} \sum_{\substack{\ell_1^1,\dots\ell^1_{\frac{k-c}{2}} \\ \dots \\ \ell_1^{2q}\dots\ell^{2q}_{\frac{k-c}{2}}}}^{n_0} (\sigma_w\sigma_x)^{2kq}$$

$$= \frac{((\sigma_w\sigma_x)^k k^c(k-c)!!)^{2q}}{n_1 m^q n_0^{kq-c}(k!)^{2q}} \frac{m!}{(m-q)!} \frac{n_1!}{(n_1-q)!} \frac{n_0!}{(n_0-(k-c)q)!}$$

In order to understand the very first term, note that we have to select in each niche c *blue* vertices in order to create the cycles and we then do a perfect matching in the rest of the vertices. Hence, we can see that

$$E_q^{(1)} = \frac{((\sigma_w\sigma_x)^k k(k-c)!!)^{2q}\psi^{1-q}}{n_0^{(c-1)(q-1)}(k!)^{2q}} (1+o(1)) \tag{2.3.7}$$

Thus this is of smaller order than (2.3.6) when the number of cycles is strictly greater than 1 as in Figure 2.2b for instance. Indeed, if one considers the ratio

$$\frac{E_q^{(1)}}{E_q} = \mathcal{O}\left(\frac{1}{n_0^{(c-1)(q-1)}}\left(\frac{(k-c)!!}{(k-1)!!}\right)^{2q}\right)$$

The graph in each niche is not a perfect matching. For this case, we said that the leading order is given by a perfect matching in each niche where we removed the vertex which belongs to the cycle. This graph gives only second moments of the matrix entries but note that we could have instead higher moments. Suppose that in one niche we have an identification between $a_1,\dots,a_b$ entries such that $a_1+\dots+a_b = k-1$. For ease we suppose that $a_1=\dots=a_{b_1}=2$ and $a_{b_1+1},\dots,a_b>2$ for some $b_1 \in [\![1,b-1]\!]$. We also suppose that this occurs in the niche $\{i_1,j_1\}$ as illustrated in Figure 2.4.

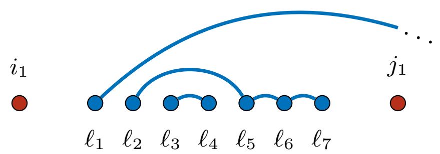


Figure 2.4: Niche where the induced graph is not a perfect matching which raises a fourth moment in the case where $k=7$

In this case, we can compare the contribution $E_q^{(2)}$ of all such matchings to the contribution of the perfect matching where $a_1=\dots=a_{(k-1)/2}=2$ that is E_q,

$$\frac{E_q^{(2)}}{E_q} = \sum_{b=1}^{\frac{k-1}{2}-1}\sum_{b_1=1}^{b-1}\sum_{\substack{a_{b_1+1}\dots a_b>2 \\ \sum a_j = k-1-2b_1}} \frac{(k-1)!}{(k-1)!!\prod_{i=1}^b a_i!}\frac{n_0!}{(n_0-(1+b+\frac{k-1}{2}(2q-1)))!}\times$$

$$\times\frac{(n_0-(1+(k-1)q)!}{n_0!}\frac{\prod_{b_1+1}^b \mathbb{E}|W_{11}|^{a_p}\mathbb{E}|X_{11}|^{a_p}}{(\sigma_w\sigma_x)^{2(\frac{k-1}{2}-b_1)}}.$$

For the first term in the summand, it corresponds to assigning the $k-1$ remaining *blue* vertices (after the choice of the cycle) into b classes of size $a_1 \dots a_b$ compared with simply doing a perfect matching between these vertices. We can bound it in the following way

$$\frac{(k-1)!}{(k-1)!!\prod_{i=1}^{b} a_i} \leqslant C\frac{2^{\frac{k-1}{2}-b_1}(\frac{k-1}{2})!}{\prod_{i=b_1+1}^{b} a_i} \leqslant C\left(\frac{k-1}{2}\right)^{\frac{k-1}{2}-b} \frac{2^{\frac{k-1}{2}-b_1}}{\prod_{i=b_1+1}^{b} a_i} \leqslant C\,(k-1)^{(\frac{k-1}{2}-b)}.$$

In the first equality we used the fact that $a_1 = \cdots = a_{b_1} = 2$ and the definition of the double factorial. In the next equality we expanded the factorial and in the last inequality we used the fact that $a_i \geqslant 3$ for $i > b_1$.
Now, for the second term, we compare the number of possible choices for ℓ indices.

$$\frac{(n_0-(1+(k-1)q)!}{(n_0-(1+b+\frac{k-1}{2}(2q-1)))!} \leqslant \frac{1}{n_0^{\frac{k-1}{2}-b}} e^{-\frac{C(kq)^2}{N}}.$$

Finally, the last term in the summand corresponds to the different moments we obtain since only variances intervene in the leading contribution while higher moments can appear inside the niche $\{i_1, j_1\}$,

$$\frac{\prod_{b_1+1}^{b} \mathbb{E}|W_{11}|^{a_p}\mathbb{E}|X_{11}|^{a_p}}{(\sigma_w\sigma_x)^{2(\frac{k-1}{2}-b_1)}} \leqslant \frac{A^{2\sum_{i\geqslant b_1+1} a_i}}{\prod_{i\geqslant b_1+1}\sigma_w^{a_i}\sigma_x^{a_i}} = \left(\frac{A^4}{\sigma_w^2\sigma_x^2}\right)^{\frac{k-1}{2}-b_1}. \tag{2.3.8}$$

Now that the terms in the summand are bounded, we need to bound the combinatorial factor coming from the sums, we can do that in the following way

$$\sum_{b_1=1}^{b-1} \sum_{\substack{a_{b_1+1},\dots,a_b\geqslant 3 \\ \sum a_j = k-1-2b_1}} \leqslant \sum_{b_1=1}^{b-1} \binom{k-1-3b-b_1+b-b_1-1}{b-b_1-1} \leqslant \sum_{b_1=1}^{b-1} (k-1)^{k-1-3b+b_1} \leqslant (k-1)^{2(\frac{k-1}{2}-b)}$$

where we used in the first inequality that $\sum_j (a_j-3) = k-1-2b_1-3(b-b_1)$. Finally, putting all these contributions together, we obtained the following comparison between $E_q^{(2)}$ and E_q,

$$\frac{E_q^{(2)}}{E_q} \leqslant \sum_{b=1}^{\frac{k-1}{2}-1} \left(\frac{CA^4}{\sigma_w^2\sigma_x^2}\frac{(k-1)^3}{n_0}\right)^{3(\frac{k-1}{2}-b)} = \mathcal{O}\left(\frac{Ck^3}{n_0}\right) \tag{2.3.9}$$

Since we have that $k^3 = o(n_0)$. Note that we supposed here that we have a perfect matching in all other niches and a single cycle to compute the contribution on the moments. This is not mandatory for the computation and we would just get a contribution of even smaller order in this case.

There are identifications between matchings from different niches Assume there are identifications between matchings of different niches. If the niches are not adjacent then such matchings would not increase the moments of the entries of W or X. However, matchings between adjacent niches may result into moments of higher order of the entries instead of the variance. We can then perform the same analysis as the previous one where we replace $k-1$ (the remaining indices after the choice of the cycle in one niche) to $2k-2$ corresponding to the number of vertices of two adjacent niches. And we recover the same order for the error as in (2.3.9).

There are identifications between the cycle and perfect matchings inside niches. We will now bound the contribution of possible identifications between the cycle parcouring every niche and the other ℓ-indices. Suppose that these identifications happen in d niches, and for $p \in \{1, \dots, d\}$, we identify the index from the cycle with $2b_p$ *blue* vertices from the niche. Note that we take an even number of identifications. Indeed if we would have an odd number, in order to obtain a non-vanishing term, we would need to either create another cycle or perform more identifications inside the niches. These possibilities are bounded by the two previous considerations. Thus, we can bound the contribution in the following way

$$\frac{E_q^{(3)}}{E_q} = \sum_{d=1}^{2q} \sum_{b_1,\dots,b_d=1}^{\frac{k-1}{2}} \binom{2q}{d} \left[\prod_{p=1}^{d} \binom{k-1}{b_p}\right] \prod_{i=1}^{d} \mathbb{E}|W_{11}|^{2+2b_p} \mathbb{E}|X_{11}|^{2+2b_p} \times$$
$$\times \frac{((k-1)!!)^{2q-d} \prod_{p=1}^{d}((k-2b_p-1)!!)}{n_0^{\sum_{p=1}^{d} b_p}((k-1)!!)^{2q}(\sigma_w\sigma_x)^{2d+\sum_{p=1}^{d} 2b_p}}.$$

This comes from the choices of the niches and the identifications we will do in each niche, the perfect matchings we perform in the other niches. Finally, we suppose that we perform perfect matchings in the rest of the d niches. Firstly, we can use the bounds

$$\prod_{p=1}^{d} \frac{1}{b_p!} \leqslant 1, \quad \prod_{i=1}^{d} \mathbb{E}|W_{11}|^{2+2b_p} \mathbb{E}|X_{11}|^{2+2b_p} \leqslant A^{4d+4\sum b_p} \quad \text{and} \quad \frac{(k-1)!!^{2q-d}\prod_{p=1}^{d}(k-1-2b_p)!!}{(k-1)!!^{2q}} \leqslant 1. \tag{2.3.10}$$

Using these bound we obtain that

$$\frac{E_q^{(3)}}{E_q} \leqslant \sum_{d=1}^{2q} \binom{2q}{d} \left(\frac{A^4}{\sigma_w^2\sigma_x^2}\right)^d \sum_{b_1,\dots,b_p=1}^{\frac{k-1}{2}} \left(\frac{A^4(k-1)}{2\sigma_w^2\sigma_x^2 n_0}\right)^{\sum b_p} = \sum_{d=1}^{2q} \binom{2q}{d} \left(\frac{A^4}{\sigma_w\sigma_x} \sum_{b=1}^{\frac{k-1}{2}} \left(\frac{A^4(k-1)}{2\sigma_w^2\sigma_x^2 n_0}\right)^b\right)^d$$

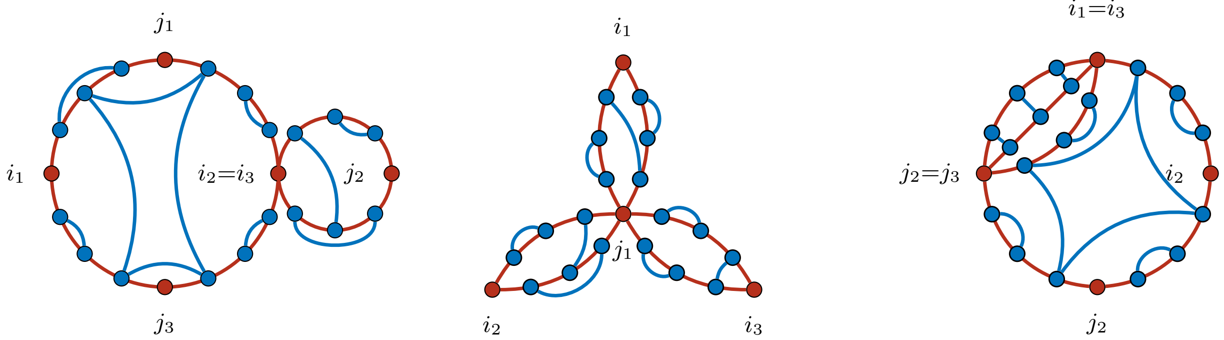
Now since we have that $k \ll n_0$, we obtain that

$$\frac{E_q^{(3)}}{E_q} = \mathcal{O}\left(\frac{Ck}{n_0}\right)$$

Contribution of general admissible graphs

We will now suppose that there are I_i identifications between the vertices indexed by i labels and I_j identifications between the vertices indexed by j labels. Note that by our definition, such a graph is admissible if and only if it consists of $I_i + I_j + 1$ cycles. See for example Figures 2.5a and 2.5b. As we saw earlier in the case of a simple cycle, the case of a cycle of size 1 has to be considered separately, thus we also suppose that the number of cycles of size 1 is given as a parameter b.

We can do a similar analysis in the case of general admissible graphs because we can realize *blue* identifications inside each cycle as they are well defined. Indeed, our *red* admissible graph being a cactus tree, in other words a tree of cycles, there is no ambiguity to define our *red* cycles. For instance, compare Figures 2.5a and 2.5b with Figure 2.5c where there are several possible choices of cycles.
Thus, if we denote $\mathcal{A}(q, I_i, I_j, b)$ the number of admissible graphs with $2q$ red edges, with I_i i-identifications and I_j j-identifications and with b cycles of size 1, we can write the contribution coming from all ad-



(a) Admissible graph with the i-identification $i_2 = i_3$ for $k = q = 3$.

(b) Admissible graph with the j-identification $j_1 = j_2 = j_3$ for $k = q = 3$.

(c) Non-admissible graph for $k = q = 3$.

Figure 2.5: Examples of admissible and non-admissible graphs

missible graphs as

$$E'_q(k) = \frac{1+o(1)}{n_1 m^q n_0^{kq}} \sum_{I_i,I_j=0}^{q} \sum_{b=0}^{I_i+I_j+1} \frac{n_1!}{(n_1-q+I_i)!} \frac{m!}{(m-q+I_j)!} \times$$
$$\times \mathcal{A}(k,q,I_i,I_j,b)\theta_1^b(f) n_0^{kb} \theta_2^{q-b}(f) n_0^{(k-1)(q-b)+I_i+I_j+1-b}$$

Looking at the first order in the equation we obtain the following

$$E'_q(k) = \sum_{I_i,I_j=0}^{q} \sum_{b=0}^{I_i+I_j+1} \mathcal{A}(q,I_i,I_j,b)\theta_1(f)^b \theta_2(f)^{q-b} \psi^{I_i+1-q} \phi^{I_j} \left(1+o(1)\right).$$

Note that the same error terms arise as before due to additional possible identifications: their contribution is then negligible as before as soon as matchings are still performed inside each cycle. Another contribution may arise actually due to an i-identification or a j-identification. One indeed has to check that performing cross-cycle *blue* identifications is subleading.
Suppose, we are around an i-identification as in Figure 2.6, these *blue* edges will match entries of W to get a non-vanishing moments. However, in order to match the corresponding X entries, some new identifications are needed. Either inside a niche, the matching is not a perfect matching and we have a lower order as in (2.3.9) or one has a cycle going along two cycles instead of two separate cycles, it consists of identifying two *blue* cycles and thus losing an order of n_0.

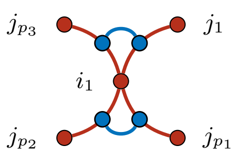


Figure 2.6: Subleading *blue* identifications around an i-identification

Contribution from non-admissible graphs

In this subsection we will control the contribution of non-admissible graphs which we denote $E_q^{(\mathrm{NA})}$. To explain how we proceed we first come back to admissible graphs. An admissible graph can be encoded into a rooted tree $T = (V, E)$ as follows. Consider such a tree: one replaces each edge with a cycle of length $2L$ with $L \geqslant 1$ though one may have to choose where cycles are identified along a cycle. These cycles will be called the *fundamental* cycles, they correspond to the cycles where we will perform a matching on the *blue* vertices. Note that in this case, we have the identity

$$I_i + I_j + 1 = \sum_{v \in V} (d(v) - 1) + 1 = 2|E| - |V| + 1 = |E|$$

where $d(v)$ denotes the degree of the vertex v in the tree and we used the fact that $|V| = |E| + 1$ and $\sum_{v \in V} d(v) = 2|E|$. In other words, this previous identity means that the total number of fundamental cycles in an admissible graph is given by $I_i + I_j + 1$.

A non-admissible graph can be encoded in a similar fashion but in a more complicated way. Indeed, we need to now consider a multigraph $G = (V, E_1, E_2)$ where E_1 denotes the set of single edges and E_2 the set of multiple edges and perform the same construction where each edge corresponds to a cycle of a certain length. We will first consider $E = \{E_1, E_2\}$ the set of all edges where we removed the multiplicity of each edge so that $G' = (V, E)$ is now a graph. Now, if we denote the *surplus* $s(G')$ to be the minimal number of edges we need to remove to G' in order to obtain a tree, we see that $s(G') = I_i + I_j + 1 - |E|$.

The problem with non-admissible graphs is the fact that there are multiple ways to determine fundamental cycles. Thus, we will count the number of non-admissible graphs labeled by its fundamental cycles. First note that if we know which are the fundamental cycles in our non-admissible graph, we can simply see it as an admissible graph with additional identifications, we can then consider such an admissible graph. It can then be obtained in the following way and see Figure 2.7 for an illustration:

Consider the tree T encoding the admissible graph, we can then choose two edges and *glue* them together in the sense of identifying one vertex of one edge to one of the other. This performed an additional identification in the initial graph and will encode a non-admissible graph. Now, while these two cycles (corresponding to these two edges) are identified at an additional vertex, there could be more identifications for the same two cycles by choosing additional vertices in each cycle to be again identified. So finally doing this step a single time, the number of possible ways to choose two cycles which are then identified at r pairs of vertices is at most:

$$C \binom{E}{2} (2q)^r \leqslant q^{r+2} \tag{2.3.11}$$

Since we need to choose two edges and then two vertices. However, see that we lose a power of n_1 (or m) for each additional identification as we lose a choice of index without gaining a cycle. This is a $o(1)$ for $q^3 \ll n_0$ and doing this step multiple times will just lower the order. Finally, the number of non-admissible graphs labeled by its fundamental cycles and weighted by n_1^a where a is the number of additional identifications is $\mathcal{O}(q^3/n_0)$ times the number of admissible graphs with the same fundamental cycles.

Now, once the fundamental cycles are identified, cross identifications between *blue* edges from distinct niches (or fundamental cycles) are subleading unless in the following case: there are multiple cycles of length 2. For these multiple cycles of length 2, we have pk *blue* vertices to match together where p is the number of these cycles. While the leading order is given by performing a perfect matching between these vertices such as in Figure 2.8, we can do any kind of matching and use the similar analysis we did for (2.3.9). Suppose that we have an identification between $a_1, \dots, a_b$ entries

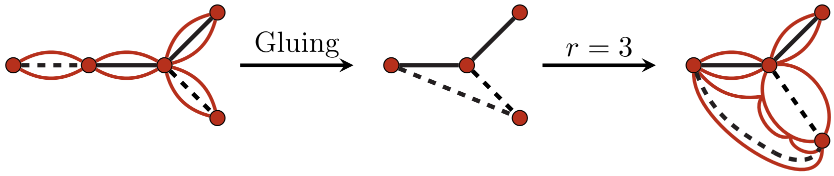


Figure 2.7: In the first picture we represent an admissible graph with its encoding tree. The two dashed lines correspond to the two edges we will glue together. The second graph correspond to the glued tree which is not a tree anymore since we created a cycle. Now the last step consists in choosing the number of identifications we want to make between the two cycles, here we have three total identifications. We keep the encoding graph to see what the choices of fundamental cycles are for this non-admissible graph. Note that we only represented the *red* vertices which join two cycles together but there are more *red* vertices inside each cycle.

such that $a_1 + \cdots + a_b = pk$. For ease we suppose that $a_1 = \cdots = a_{b_1} = 2$ and $a_{b_1+1}, \ldots, a_b > 2$ for some $b_1 \in [\![1, b-1]\!]$, then we can bound the contribution by

$$\sum_{p=2}^{q} \sum_{b=1}^{\frac{m(e)k}{2}} \frac{1}{n_0^{p-1}} \sum_{b_1=1}^{b} \sum_{\substack{a_{b_1+1},\ldots,a_b>2 \\ \sum a_i = pk-2b_i}} \frac{(pk)!}{((2k)!!)^{p/2} b! \prod_{i=b_1+1}^{b} a_i} \frac{n_0^b \prod_{i=b_1+1}^{b} \mathbb{E}|W_{11}|^{a_i} \mathbb{E}|X_{11}|^{a_i}}{n_0^{kp/2} (\sigma_w^2 \sigma_x^2)^{kp/2-b}} \tag{2.3.12}$$

The factor of n_0^{1-p} comes from the additional identifications between i's and j's in order to obtain a multiple edge. For instance in Figure 2.8 we have less identifications in the admissible graph than in the corresponding non-admissible graph. The first term in the summand consists in comparing the different possible matchings to performing a perfect matching in every single cycle and we can bound from above by

$$\frac{(pk)!}{((2k)!!)^{p/2} b! \prod_{i=b_1+1}^{b} a_i} \leqslant (Cp)^{kp}.$$

The second term now comes from the number of ℓ indices chosen and the ratio of moments and we bound it in the same way as in (2.3.9),

$$\frac{\prod_{b_1+1}^{b} \mathbb{E}|W_{11}|^{a_p} \mathbb{E}|X_{11}|^{a_p}}{n_0^{kp/2-b} (\sigma_w \sigma_x)^{2(kp/2-b)}} \leqslant \frac{A^{2\sum_{i \geqslant b_1+1} a_i}}{n_0^{kp/2-b} \prod_{i \geqslant b_1+1} \sigma_w^{a_i} \sigma_x^{a_i}} = \left(\frac{A^4}{n_0 \sigma_w^2 \sigma_x^2} \right)^{\frac{kp}{2}-b}.$$

Also in the same way as in (2.3.9), we can bound the combinatorial factor coming from the sums as

$$\sum_{b_1=1}^{b} \sum_{\substack{a_{b_1+1},\ldots,a_b>2 \\ \sum a_i = pk-2b_i}} \leqslant (m(e)k)^{2(\frac{pk}{2}-b)}.$$

Finally, putting all the contribution together we have

$$n_0 \sum_{p=2}^{q} \left(\frac{Cp^k}{n_0} \right)^p \sum_{b=1}^{pk/2} \left(\frac{A^4 p^2 k^2}{n_0 \sigma_w^2 \sigma_x^2} \right)^{\frac{kp}{2}-b} = \mathcal{O}\left(\frac{q^{2k}}{n_0} \right) \tag{2.3.13}$$

where we used the fact that the leading order comes from the fact that $b = \frac{km(e)}{2}$.

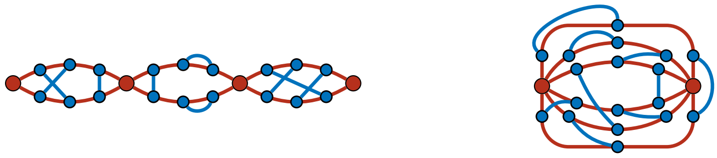


Figure 2.8: Different behavior between an admissible graph and a multiple edge.

Thus, the total contribution for non admissible graph, by combining the choices of fundamental cycles (2.3.11) and the combinatorial change coming from multiple single cycles (2.3.13) we obtain a total contribution of

$$\frac{E_q^{(\mathrm{NA})}}{E_q} = \mathcal{O}\left(\frac{q^3(1+q^{2k})}{n_0}\right). \tag{2.3.14}$$

2.3.2. Case where f is a monomial of even degree

In the case of an even monomial we center the function f, to do so we substract a constant given by the corresponding expectation. We will then consider centered monomial of the form

$$f(x) = \frac{x^k - k!!}{k!} \quad \text{so that} \quad \theta_1(f) = \frac{(\sigma_w\sigma_x)^{2k}}{(k!)^2}\left((2k)!! - (k!!)^2\right) \quad \text{and} \quad \theta_2(f) = 0.$$

Here, the fact that $\theta_2(f)$ vanishes means that all admissible graphs which have at least one cycle of size greater than 1 will be subleading so that we will see admissible graphs consisting only in cycles of size 1 such as Figure 2.5b for instance.

Note that we have seen earlier that we can write

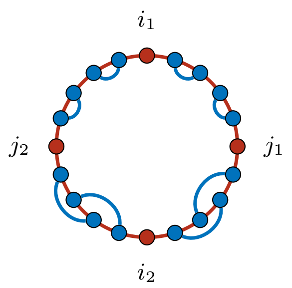
$$\mathbb{E}\left[\frac{1}{k!}\left(\frac{(WX)_{ij}}{\sqrt{n_0}}\right)^k\right] = \frac{1}{n_0^{k/2}k!}\mathbb{E}\sum_{\ell_1,\dots,\ell_k=1}^{n_0}\prod_{p=1}^{k} W_{i\ell_p}X_{\ell_p j} = \frac{k!!}{k!}(\sigma_w\sigma_x)^k\left(1+\mathcal{O}\left(\frac{1}{n_0}\right)\right).$$

Thus, by developing the tracial moments of M we obtain the following formula,

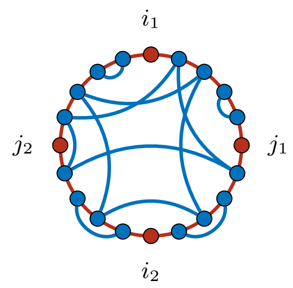
$$\frac{1}{n_1}\mathbb{E}\left[\operatorname{Tr} M^q\right] = \left(1+\mathcal{O}\left(\frac{1}{n_0}\right)\right)\frac{1}{n_1 m^q}\mathbb{E}\sum_{i_1,\dots,i_q}^{n_1}\sum_{j_1,\dots,j_q}^{m}\left[\frac{1}{n_0^{kq}(k!)^{2q}}\sum_{\substack{\ell_1^1,\dots\ell_k^1\\ \dots\\ \ell_1^{2q}\dots\ell_k^{2q}}}^{n_0}\prod_{p=1}^{k} W_{i_1\ell_p^1}X_{\ell_p^1 j_1}\times\right.$$
$$\left.\times\prod_{p=1}^{k} W_{i_2\ell_p^2}X_{\ell_p^2 j_1}\cdots\prod_{p=1}^{k} W_{i_1\ell_p^{2q}}X_{\ell_p^{2q} j_q} - c_0^{2q}\right] \tag{2.3.15}$$

Note that we can actually write c_0 as a sum over possible admissible graphs, where *blue* vertices form a perfect matching in each niche as in Figure 2.9a. Note that this is now possible since each niche contains an even number of vertices. Thus at the first order, we can write

$$c_0^{2q} = \left(1+\mathcal{O}\left(\frac{1}{n_0}\right)\right)\mathbb{E}\sum_{\substack{\ell_1^1,\dots\ell_k^1\\ \dots\\ \ell_1^{2q}\dots\ell_k^{2q}\\ \text{form a perfect}\\ \text{matching in each}\\ \text{niche}}}^{n_0}\prod_{p=1}^{k} W_{i_1\ell_p^1}X_{\ell_p^1 j_1}\prod_{p=1}^{k} W_{i_2\ell_p^2}X_{\ell_p^2 j_1}\cdots\prod_{p=1}^{k} W_{i_1\ell_p^{2q}}X_{\ell_p^{2q} j_q}$$



(a) Leading order graph before centering



(b) Leading order graph after centering.

We can then see that the graphs corresponding to the expectation are the admissible graphs where *blue* vertices make a perfect matching inside each niche. Thus the leading order graphs after centering will be those which have two cycles between the niches and a perfect matching with the rest of the vertices in each niche as in Figure 2.9b or additional identifications inside niches. Note that this comes from the fact that we want to maximize the number of distinct ℓ-indices and the previous subsection, since we have seen that adding more cycles or not doing perfect matchings in the niches are subleading. In the same way we will first see the contribution of one *red* cycle on the moments to then deduce the contribution of all admissible graphs. Thus we find that, in the case of a cycle with $2q$ vertices for $q > 1$,

$$E_q(k) = \frac{1}{n_0^{q-1}} \left((\sigma_w \sigma_x)^k k(k-1)(k-2)!! \right)^{2q} \psi^{1-q} + o\left(\frac{\hat{\theta}_2(f)}{n_0^{q-1}} \right) = \frac{1}{n_0^{q-1}} \hat{\theta}_2(f)^q \psi^{1-q} + \mathcal{O}\left(\frac{\hat{\theta}_2(f)}{n_0^{q-1}} \right)$$

$$\text{with } \hat{\theta}_2(f) = \left(\frac{(\sigma_w \sigma_x)^k k(k-1)(k-2)!!}{k!} \right)^2 = \left(\sigma_w \sigma_x \int f''(\sigma_x \sigma_x x) \frac{e^{-x^2/2}}{\sqrt{2\pi}} \mathrm{d}x \right)^2$$

In this case, we obtain $(k-2)q+2$ distinct ℓ-indices. See that in the leading order before centering we would obtain kq distinct indices as we perform a perfect matching in every niche. Since such identifications are forbidden by the centering, if we do not create a cycle between niches but identify *blue* vertices inside the same niche we can only obtain at most $(k-4)q + 2q$ distinct indices which is of lower order than Figure 2.9b. Now if we only create one cycle, we need to perform at least identifications between three vertices in each niche since we would have an odd number of *blue* vertices left and the number of distinct indices becomes at most $(k-4)q + 2q + 1$ which is also of lower order than Figure 2.9b. Thus we can summarize the distinct subleading behaviors as the following

Two cycles:	at most $k - 2q + 2$ distinct indices.
One cycle:	at most $k - 2q + 1$ distinct indices.
No cycle:	at most $k - 2q$ distinct indices.

See that the contribution of a *red* cycle of size greater than 2 for f being an even monomial is of smaller order than when it is an odd monomial. Now, in the same way, the case of a simple cycle with 2 vertices is slightly different because of the centering. Indeed, the centering prevents the graphs from performing a perfect matching inside each niche, thus at least one (thus two) vertices has to be connected to the other niche. Note also that any perfect matching where the two niches are connected is of the same order, thus we obtain for the leading order

$$E_1(k) = \frac{(\sigma_w \sigma_x)^{2k}}{(k!)^2} \left((2k)!! - (k!!)^2 \right) + \mathcal{O}\left(\frac{(2k-2)!!k^2}{(k!)^2 n_0} \right) = \theta_1(f) + \mathcal{O}\left(\frac{(2k-2)!!k^2}{(k!)^2 n_0} \right).$$

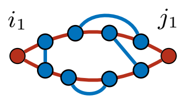


Figure 2.10: Contribution in the case $q = 1$ for an even monomial.

For the general case of admissible graphs with possible identifications, we can do similar computations as those in the previous subsection by seeing that the contribution is just a product over the different cycles. Thus, the leading order of a q-moment, corresponding to the admissible graphs with $2q$ edges can be written as

$$E_q(k) = \frac{1+p(1)}{n_1 m^q n_0^{kq}} \sum_{I_i,I_j=0}^{q} \sum_{b=0}^{I_i+I_j+1} \frac{n_1!}{(n-q+I_i)!} \frac{m!}{(m-q+I_j)!} \times$$
$$\times \mathcal{A}(q,k,I_i,I_j,b)\theta_1(f)^b n_0^{kb} \hat{\theta}_2(f)^{q-b} n_0^{2(I_i+I_j+1-b)+(k-2)(q-b)}$$

which gives asymptotically,

$$\begin{aligned} E_q(k) &= (1+o(1)) \sum_{I_i,I_j=0}^{q} \sum_{b=0}^{I_i+I_j+1} \frac{\mathcal{A}(q,I_i,I_j,b)\theta_1(f)^b \hat{\theta}_2(f)^{q-b} \phi^{I_j} \psi^{I_i+1-q}}{n_0^{q-(I_i+I_j+1)}} \\ &= (1+o(1)) \sum_{\substack{I_i,I_j=0 \\ I_i+I_j+1=q}} \mathcal{A}(q,I_i,I_j,I_i+I_j+1)\theta_1(f)^{I_i+I_j+1} \phi^{I_j} \psi^{I_i+1-q} \\ &= (1+o(1)) \sum_{I_i,I_j=0}^{q} \sum_{b=0}^{I_i+I_j+1} \mathcal{A}(q,I_i,I_j,b)\theta_1(f)^b \theta_2(f)^{q-b} \psi^{I_i+1-q} \phi^{I_j} \end{aligned}$$

where we used in the last equality the fact that $\theta_2(f) = 0$ in order to retrieve the expression (2.3.2). Note again that we did not give here all the errors since we have computed them in the previous subsection, the case of even monomials can be done similarly. Thus we can see that only the graphs which corresponds to a tree of simple cycles contribute to the moments.

Note that there is no difference for the study of non-admissible graphs as it only concerns *red* vertices while the polynomial involves only *blue* vertices.

2.3.3. Case where f is a polynomial

We will now suppose that we can write

$$f(x) = \sum_{k=1}^{K} a_k f_k(x) \quad \text{with} \quad f_k(x) = \frac{x^k - k!!\mathbb{1}_{k \text{ even}}}{k!} \quad \text{and} \quad \sup_{k\in[\![1,K]\!]} |a_k| \leqslant C^k \quad \text{for some } C.$$

In particular, our parameters are in this case

$$\theta_1(f) = \sum_{\substack{k_1,k_2=1 \\ k_1+k_2 \text{ even}}}^{K} \frac{a_{k_1} a_{k_2} (\sigma_w \sigma_x)^{k_1+k_2} \left((k_1+k_2)!! - k_1!!k_2!!\mathbb{1}_{k_1 \text{ even}}\right)}{k_1!k_2!},$$

$$\theta_2(f) = \left(\sum_{k=1}^{K} \frac{a_k (\sigma_w \sigma_x)^k k(k-1)!!\mathbb{1}_{k \text{odd}}}{k!} \right)^2.$$

Note that for any polynomial, by expanding the moment as in (2.3.5), we have to compute the following quantity, for any $k_1, \ldots, k_{2q}$ integers,

$$\frac{1}{n_1}\mathbb{E}\left[\operatorname{Tr} M^q\right] = \sum_{k_1,\ldots,k_{2q}=1}^{K} \frac{a_{k_1}\ldots a_{k_{2q}}}{n_1 m^q \prod_{i=1}^{2q} k_i!} \times$$
$$\times\, \mathbb{E} \sum_{i_1,\ldots,i_q}^{n_1} \sum_{j_1,\ldots,j_q}^{m} \sum_{\substack{\ell_1^1,\ldots\ell_k^1 \\ \cdots \\ \ell_1^{2q}\ldots\ell_k^{2q}}}^{n_0} f_{k_1}\left(\frac{WX}{\sqrt{n_0}}\right)_{i_1 j_1} f_{k_2}\left(\frac{WX}{\sqrt{n_0}}\right)_{i_2 j_1} \ldots f_{k_{2q}}\left(\frac{WX}{\sqrt{n_0}}\right)_{i_1 j_q} \quad (2.3.16)$$

In order to compute the dominant term of this moment, first note that the centering creates disparity between even and odd monomials. Indeed, if we consider one *red* cycle, we now have $2q$ niches of different sizes, namely $k_1, \ldots, k_{2q}$. We will first bound these moments in order to see that, in each cycle, the niches with an even number of vertices are subleading so that the dominant term in the asymptotic expansion of the moment corresponds to admissible graphs with only odd niches when expanding the polynomial.

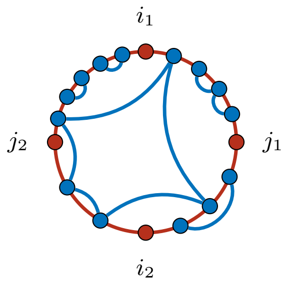


Figure 2.11: Admissible graph in the case of a polynomial with $(k_1, k_2, k_3, k_4) = (4, 3, 2, 5)$.

The behavior in a cycle can be understood as follows: there has to be at least one cycle between in each niche for the odd or the centered even niches. Now, in each odd niche, we saw that the dominant term is to realize a perfect matching in the $k_i - 1$ vertices remaining in the niche from (2.3.9). However, in the even niches, since there is already a cycle, there remains an odd number of vertices to be matched either with an existing cycle or a matching and the leading order is to perform a perfect matching in the $k_i - 2$ remaining vertices. Thus we obtain, if we consider the number of choices of indices for *red* and *blue* vertices in one configuration $k_1, \ldots, k_{2q}$ denoted $C(k_1, \ldots, k_{2q})$,

$$C(k_1, \ldots, k_{2q}) = \frac{n_0^{-\sum_{i=1}^{2q} \frac{k_i}{2}}}{n_1 m^q} n_1^q m^q n_0^{1+\sum_{k_i \text{ odd}} \frac{k_i - 1}{2} + \sum_{k_i \text{ even}} \frac{k_i - 2}{2}} = \frac{\psi^{1-q}}{n_0^{\frac{\#k_i \text{ even}}{2}}}.$$

This contribution can be understood in the following way, we have the normalization and then we have to choose the q i-indices, the q j-indices, the ℓ-index coming from the cycle, if the niche consists of an odd monomial we perform a perfect matching in the rest of the niche, if it is an even monomial we can only perform a perfect matching on $(k_i - 2)/2$ of the remaining vertices as we need to match one vertex elsewhere by the centering. Thus, if we consider the contribution of cycles of size $q > 1$ for the polynomial $P = \sum \frac{a_k}{k!}(X^k - k!!\mathbb{1}_{k \text{ even}})$, we get the following asymptotic expansion for the moments

$$(2.3.16) = \frac{1+\mathcal{O}\left(\frac{1}{\sqrt{n_0}}\right)}{n_1 m^q} \sum_{\substack{k_1,\dots,k_{2q} \\ k_i \text{ odd}}} \left[\prod_{i=1}^{2q} \frac{a_{k_i}}{k_i!}\right] \frac{1}{n_0^{\sum_{i,k_i \text{ odd}} \frac{k_i}{2}}} n_1^q m^q n_0^{1+\sum_{i,k_i \text{ odd}} \frac{k_i-1}{2}} \prod_{i,k_i \text{ odd}} (\sigma_w\sigma_x)^{k_i} k_i(k_i-1)!!$$

$$= \psi^{1-q}\left(\sum_{k \text{ odd}} a_k(\sigma_w\sigma_x)^k k(k-1)!!\right)^{2q}\left(1+\mathcal{O}\left(\frac{1}{\sqrt{n_0}}\right)\right) = \psi^{1-q}\theta_2^q(f) + \mathcal{O}\left(\frac{\theta_2^q(f)}{\sqrt{n_0}}\right),$$

as we now explain, in the case of a cycle consisting of two edges decorated by k_1 and k_2 *blue* vertices, there are three different possibilities:

(i) k_1 and k_2 are odd: $(\sigma_x\sigma_w)^{k_1+k_2}(k_1+k_2)!!$.

(ii) k_1 and k_2 are even: $(\sigma_w\sigma_x)^{k_1+k_2}((k_1+k_2)!! - k_1!!k_2!!)$.

(iii) k_1 is even and k_2 is odd: the leading term in the asymptotic expansion is of order $n_0^{-1/2}$.

Thus, the 1-moment for f a polynomial is

$$\sum_{\substack{k_1,k_2=1 \\ k_1+k_2 \text{ even}}}^{K} \left(\frac{a_{k_1}a_{k_2}}{k_1!k_2!}(\sigma_w\sigma_x)^{k_1+k_2}\left((k_1+k_2)!! - k_1!!k_2!!\mathbb{1}_{k_1 \text{ even}}\right) + \mathcal{O}\left(\frac{(k_1+k_2)(k_1+k_2-1)!!}{\sqrt{n_0}k_1!k_2!}\right)\right)$$

$$= \theta_1(f) + \mathcal{O}\left(\frac{K}{\sqrt{n_0}}\right)$$

where we used the fact that for any k_1 and k_2, $(k_1+k_2)!!/(k_1!k_2!)$ is bounded. While these analysis work in the case of a single cycle, we can do the same generalization to any admissible graphs as before. Thus we get the following q-moment in the case of a polynomial

$$m_q := \frac{1}{n_1}\mathbb{E}\left[\mathrm{Tr} M^q\right] = (1+o(1)) \sum_{I_i,I_j=0}^{q} \sum_{b=0}^{I_i+I_j+1} \mathcal{A}(q,I_i,I_j,b)\theta_1^b(f)\theta_2^{q-b}(f)\psi^{I_i+1-q}\phi^{I_j}.$$

2.3.4. Convergence of moments in probability

In the previous subsection, we proved convergence of the expected moments of the empirical eigenvalue distribution. In order to prove convergence of the actual moments of this distribution we will do as in the proof of Wigner's theorem.

Lemma 2.3.4. *Let $f(x) = \sum_k^K a_k x^k$ be a polynomial activation function and consider the associated matrix M with enpirical eigenvalue distribution μ_{n_1}. Denote by m_q th moments*

$$m_q = \frac{1}{n_1}\sum_{i=1}^{n_1} \lambda_i^q = \frac{1}{n_1}\mathrm{Tr}\, M^q \quad \text{and} \quad \overline{m}_q = \mathbb{E}\left[m_q\right]$$

we then have, for any $\varepsilon > 0$,

$$\mathbb{P}\left(|m_q - \overline{m}_q| > \varepsilon\right) \xrightarrow[n_1\to\infty]{} 0. \tag{2.3.17}$$

Actually, we even have that

$$\mathrm{Var}\, m_q = \mathcal{O}\left(\frac{(q^2K^2+q^4)C^q}{n_1^2}\right)$$

for some constant C.

Proof. We can write the variance of the moments in the following way

$$\operatorname{Var} m_q = \mathbb{E}\left[\left(\frac{1}{n_1}\operatorname{Tr} M^q\right)^2\right] - \overline{m}_q^2 = \frac{1}{n_1^2}\sum_{\mathcal{G}_1,\,\mathcal{G}_2}\sum_{\boldsymbol{\ell_1},\boldsymbol{\ell_2}} \mathbb{E}\left[M_{\mathcal{G}_1}(\boldsymbol{\ell_1})M_{\mathcal{G}_2}(\boldsymbol{\ell_2})\right] - \mathbb{E}\left[M_{\mathcal{G}_1}(\boldsymbol{\ell_1})\right]\mathbb{E}\left[M_{\mathcal{G}_2}(\boldsymbol{\ell_2})\right]$$

with $\mathcal{G}_p = (G_p, \mathbf{i_p}, \mathbf{j_p})$ are labeled graphs with the i-labels and j-labels given respectively by $\mathbf{i_p}$, $\mathbf{j_p}$. For a given labeled graph $\mathcal{G} = (G, \mathbf{i}, \mathbf{j})$ and a matching $\boldsymbol{\ell}$, the notation $M_{\mathcal{G}}(\boldsymbol{\ell})$ corresponds to the following product after expansion

$$M_{\mathcal{G}}(\boldsymbol{\ell}) = \sum_{k_1,\dots,k_{2q}=1}^{K} \frac{a_{k_1}\dots a_{k_{2q}}}{m^q n_0^{\sum k_i/2}} \prod_{p=1}^{k_1} W_{i_1\ell_p^1}X_{\ell_p^1 j_1}\prod_{p=1}^{k_2} W_{i_2\ell_p^2}X_{\ell_p^2 j_1}\cdots\prod_{p=1}^{k_{2q}} W_{i_1\ell_p^{2q}}X_{\ell_p^{2q} j_q}.$$

Now, note that the shape of the graph and the possible expansion of the polynomial f does not depend on n_0, n_1 or m. See also that the previous considerations still hold and the dominant term can only be given by admissible graphs where the polynomial expansion consist only of odd monomials.

By independence of the matrix entries W and X and by the definition of the variance as a sum on pairs of graphs, we only need to consider graphs $\mathcal{G}_1$, $\mathcal{G}_2$ which share a common edge $X_{\ell j}$ or $W_{i\ell}$ for some i, j, and ℓ. Suppose that $\mathcal{G}_1$ and $\mathcal{G}_2$ have $2q$ edges. For simplicity and to explain the computation we will suppose that they are both a cycle and f is an odd monomial x^k. Note that the generalization comes from the fact that admissible graph are a tree of cycles and non-admissible graphs are of lower order from (2.3.14). If we suppose that the coincidence between the two graphs comes from an i-label and a ℓ-label, in other words an entry $W_{i\ell}$, we have different possibilities.

The first case consists in taking the two *red* cycles and attaching them at a fixed vertex i_0. We then perform an identification cross-cycle as in Figure 2.6 in order to match two entries $W_{i_0\ell_0}$ together from $\mathcal{G}_1$ and $\mathcal{G}_2$. Now that we matched these W entries, note that the corresponding X entries have not been matched yet since we performed an identification between two distinct niches adjacent to i_0. We then need to identify this vertex with another vertex from an adjacent niche (and then creating a *blue* cycle going over the whole *red* cycle) or to another vertex in the same niche. Finally, it can be seen as simply performing the dominant matching into each graph, identifying two i indices and then identifying two *blue* edges from niches adjacent to i. Finally we can compute the contribution of these graphs in the covariance as

$$\sum_{\boldsymbol{\ell_1},\boldsymbol{\ell_2}} \operatorname{Cov}^{(1)}(M_{\mathcal{G}_1}(\boldsymbol{\ell_1}), M_{\mathcal{G}_2}(\boldsymbol{\ell_2})) = \mathcal{O}\left(q^2k^2\psi^{1-2q}\theta_2(f)^{2q}\left(\frac{\mathbb{E}W_{11}^4}{\sigma_w^4} - 1\right)\right)$$

Indeed, in each graph we perform the typical matching corresponding to a *blue* cycle going over every niche and perfect matchings between the remaining indices in each niche. Now the fact that we identify two $W_{i_0\ell_0}$ entries create a moment of order 4 when we compute $\mathbb{E}\left[M_{\mathcal{G}_1}M_{\mathcal{G}_2}\right]$. we then have to cpint the number of possible choices for indices: we have n_1^{2q-1} choices for the i indices as we identify two from $\mathcal{G}_1$ and $\mathcal{G}_2$, m^{2q} for the j indices, $n_0^{2+4q(k-1)/2-1}$ choices for the ℓ indices (2 cycles, $4q$ niches and an identification between the two graphs). Since we have a normalization of $m^{-2q}n_0^{-2kq}$, combining the whole yields a factor ψ^{1-2q} asymptotically. In the same way, for general polynomial and admissible graphs, for such an identification we would obtain that

$$\frac{1}{n_1^2}\sum_{\mathcal{G}_1,\mathcal{G}_2}\sum_{\boldsymbol{\ell_1},\boldsymbol{\ell_2}} \operatorname{Cov}^{(1)}(M_{\mathcal{G}_1}(\boldsymbol{\ell_1}), M_{\mathcal{G}_1}(\boldsymbol{\ell_1})) = \mathcal{O}\left(\frac{q^2k^2}{n_1^2\psi}\left(1-\frac{\phi}{\psi}\right)\mathfrak{m}_q^2\left(\frac{\mathbb{E}W_{11}^4}{\sigma_w^4} - 1\right)\right) = \mathcal{O}\left(\frac{q^2k^2C^q}{n_1^2}\right)$$

for some $C > 0$. Indeed, we get the q^2k^2 from the choices for the edge we want to identify between the two graphs, the constant factor in ϕ and ψ consists in the choice of choosing a $\{i,\ell\}$ edge or a $\{j,\ell\}$

edge. Then the previous computation in the case of a cycle can be generalized to all graphs as the construction only involves one cycle in each graph. For the second equality we used (2.3.19) proved in the next subsection.

The second case consists in identifying two *red* vertices in each graph. We need to choose two *red* vertices belonging to the same cycle in each graph and we identify the pair from one graph to the other pair. The whole graph $\mathcal{G}$ created by this construction is not admissible as we have two identifications and two fundamental cycles. We thus need to choose the fundamental cycles.

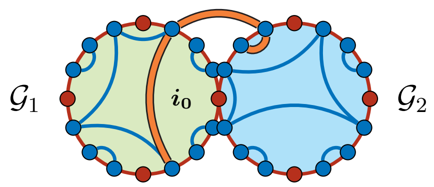
The fundamental cycles we choose for this *red* graph are given by the cycles between the two vertices with edges belonging to both graphs in each cycle. Since we need to choose a pair of vertices in each graph we have q^4 choices. In each fundamental cycles, we perform the typical *blue* matching and we have an edge between a niche from $\mathcal{G}_1$ and a niche from $\mathcal{G}_2$ (corresponding to the cycle going over every niche for instance). Thus we have a common W or X entry between the two graphs and the contribution in the covariance does not vanish. Considering the q^4 choices for the *red* vertices, we can see that we have

$$\frac{1}{n_1^2}\sum_{\mathcal{G}_1,\mathcal{G}_2}\sum_{\boldsymbol{\ell_1},\boldsymbol{\ell_2}}\mathrm{Cov}^{(1)}(M_{\mathcal{G}_1}(\boldsymbol{\ell_1}),M_{\mathcal{G}_2}(\boldsymbol{\ell_2}))=\mathcal{O}\left(\frac{q^4C^q}{n_1^2}\right).$$

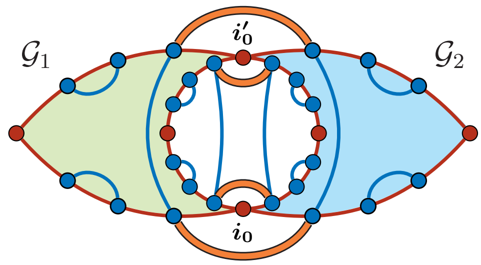
We have the same power of n_1 as the lose a choice of *red* vertex index (since we choose two in each graph) but we gain a choice of *blue* vertex (since we do not perform additional identifications). Finally, we obtain that

$$\mathrm{Var}\, m_q=\mathcal{O}\left(\frac{(q^2k^2+q^4)C^q}{n_1^2}\right).$$

Using Bienaymé-Chebyshev inequality, one easily deduces (2.3.17). □



(a) In this figure, the two highlighted cycles correspond to the graph $\mathcal{G}_1$ and $\mathcal{G}_2$ which are attached to a common vertex i_0. We perform a typical *blue* matching in each graph and then add an identification between the two graphs. The highlighted edges in orange corresponds the common edges between the two graphs which raise a moment of order 4.



(b) In this figure, the two highlighted cycles correspond to the graph $\mathcal{G}_1$ and $\mathcal{G}_2$ which are attached to two vertices i_0 and i_0'. The graph is non-admissible and we choose the fundamental cycles so that neither $\mathcal{G}_1$ or $\mathcal{G}_2$ are fundamental cycles. The typical matching in the chosen cycles create common edges between the two graphs highlighted in orange in the figure. Note that in this case we do not obtain moments of order 4.

2.3.5. Passage to sub-Gaussian random variables

We performed the computation of the limiting expected moments and of the covariance in the case of bounded random variables. However, note that while high moments of W or X can appear in the error terms, as in (2.3.8), (2.3.10) and (2.3.12) we have the following bound for our moments

$$\mathbb{E}\left[|X_{11}|^k\right] \leqslant C^k k^{k/\alpha}$$

for some constant C and the same boud holds for the entries of W. Thus it consists simply in taking all the errors and replacing A by $k^{1/\alpha}$. Since we always compare in the errors (2.3.8), (2.3.10) and (2.3.12) powers of A to powers of n_1, and k is of order $\frac{\log n_1}{\log\log n_1}$ all the errors are still a $o(1)$.

2.3.6. Weak convergence of the empirical spectral measure

In this subsection, we will see that the sequence given by the limiting moments of the empirical eigenvalue distribution μ_{n_1} characterize a compactly supported measure μ.

Lemma 2.3.5. *There exists a measure μ such that for any $q \in \mathbb{N}$, defining the sequence*

$$\mathfrak{m}_q := \sum_{I_i,I_j=0}^{q} \sum_{b=0}^{I_i+I_j+1} \mathcal{A}(q,I_i,I_j,b)\theta_1(f)^b\theta_2(f)^{q-b}\psi^{I_i+1-q}\phi^{I_j} \tag{2.3.18}$$

we have

$$\int x^q \mathrm{d}\mu(x) = \mathfrak{m}_q.$$

Besides, the measure μ is characterized by this sequence of moments and compactly supported.

Proof. This is simply a consequence of Carleman's condition. Indeed, We have a simple bound for the moment $\mathfrak{m}_q$, namely, there exists a constant C such that

$$\mathfrak{m}_q \leqslant C^q. \tag{2.3.19}$$

In particular, note that we have

$$\sum_{q=1}^{\infty} \frac{1}{\mathfrak{m}_{2q}^{1/2q}} = \infty$$

which guarantee this sequence of moments to be moments of a probability measure μ and the condition above is sufficient to say that the measure is determinate or characterized by its moments. Note that the bound (2.3.19) also gives that the measure has compact support. It thus remains to show (2.3.19), this comes from a bound on the total number of unlabeled cactus graphs. Indeed, it has been shown that regardless of the number of identifications or simple cycles, we have the following asymptotic for $\Theta(q)$ the number of unlabeled cactus graphs with q vertices from [FU56], there exists numerical constants $\delta > 0$ and $\xi > 1$ such that

$$\Theta(q) \sim \frac{3\delta}{4\sqrt{\pi}} \frac{\xi^{q+3/2}}{q^5}.$$

Now using this asymptotic, we know that for a large constant C we have the bound (2.3.19). □

Now, since we know that μ is a determinate measure, and we have the convergence of expected moments with the variance bound from Lemma 2.3.4. We have the corresponding almost sure weak convergence

Theorem 2.3.6. *For any continuous bounded function g we have*

$$\int g\mathrm{d}\mu_{n_1} \xrightarrow[n_1\to\infty]{} \int g\mathrm{d}\mu \quad \textit{almost surely.}$$

2.3.7. Recursion relation for the Stieltjes transform

Consider the Stieltjes transform of the limiting empirical eigenvalue distribution of M defined by Lemma 2.3.5,

$$G(z) = \int \frac{\mathrm{d}\mu(x)}{x-z}.$$

One can also write it as the following generating function of moments, since from the bound (2.3.19) the following equality makes sense at least on a neighborhood of infinity,

$$G(z) = \frac{1}{z} + \sum_{q=1}^{\infty} \frac{\mathfrak{m}_q}{z^{q+1}}.$$

Using that

$$\mathfrak{m}_q = \psi^{1-q} \sum_{I_i,I_j=0}^{q} \sum_{b=0}^{I_i+I_j+1} \mathcal{A}(q,I_i,I_j,b)\theta_1^b(f)\theta_2^{q-b}(f)\psi^{I_i}\phi^{I_j},$$

one can write the Stieltjes transform as

$$G(z) = \frac{1-\psi}{z} + \frac{\psi}{z}H(z) \quad \text{with} \quad H(z) = \sum_{q=0}^{\infty} \frac{1}{(\psi z)^q} \sum_{I_i,I_j=0}^{q} \sum_{b=0}^{I_i+I_j+1} \mathcal{A}(q,I_i,I_j,b)\theta_1^b(f)\theta_2^{q-b}(f)\psi^{I_i}\phi^{I_j}.$$

Fix a vertex v and denote q_0 the length of one of the fundamental cycles containing v. Suppose first that we have $q_0 > 1$, this cycle contains $2q_0$ edges with q_0 vertices labeled with i and q_0 vertices labeled with j. On each vertex labeled with i, either a graph is attached and we have a i-identification on this vertex, or nothing is attached. Thus, considering the formula above, we have that the contributions for identifications for each vertex is

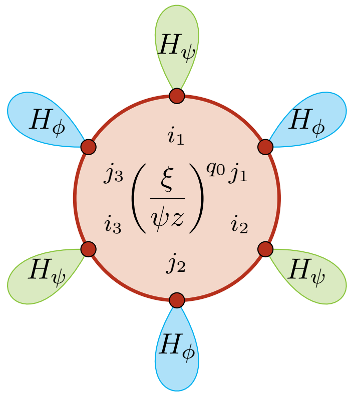
$$H_\psi(z) := 1 - \psi + \psi H(z) \quad \text{for } i\text{-labels and} \quad H_\phi(z) := 1 - \phi + \phi H(z) \quad \text{for } j\text{-labels.}$$

Also, one can see in the leading order of the moment that a cycle of length q_0 give a contribution of $\left(\frac{\xi}{\psi z}\right)^{q_0}$. Now, if the cycle is of length 1, in the same way, there is a single i-labeled vertex and a single j-labeled vertex which can give a contribution of H_ψ and H_ϕ but the contribution of a simple cycle is not given in terms of ξ but by $\frac{\eta}{\psi z}$. Thus, we have the following recursion relation for H,

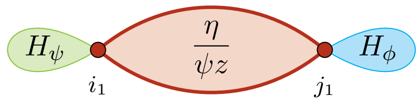
$$\begin{aligned} H(z) &= 1 + \frac{H_\phi(z)H_\psi(z)\theta_1}{\psi z} + \sum_{q_0=2}^{\infty} \left(\frac{H_\phi(z)H_\psi(z)\theta_2}{\psi z} \right)^{q_0} \\ &= 1 + \frac{H_\phi(z)H_\psi(z)(\theta_1-\theta_2)}{\psi z} + \frac{H_\phi(z)H_\psi(z)\theta_2}{\psi z - H_\phi(z)H_\psi(z)\theta_2}. \end{aligned}$$

2.4. Polynomial approximation for general activation function

In this section, we will now go from the activation function being a polynomial where we could use the moment method to a wider class of activation functions, thus proving Theorem 2.2.1. Note that for simplicity, we will consider $\sigma_w = \sigma_x = 1$ but the general case is true by simple renormalization.



(a) Contributions for the recursion formula in the case of a large cycle ($q_0 = 3$)



(b) Contribution for the recursion formula in the case of a simple cycle.

Proof of Theorem 2.2.1. We begin by defining the following polynomial which approximates f up to a constant, for $x \in \mathbb{R}$ we define

$$P_k(x) := \sum_{j=1}^{n} f^{(j)}(0)\frac{x^j - j!!}{j!} = \sum_{j=0}^{k} f^{(j)}(0)\frac{x^j}{j!} - a_n \quad \text{with} \quad a_n = \sum_{j=0}^{k} f^{(j)}(0)\frac{j!!}{j!} \tag{2.4.1}$$

with the convention that $j!! = 0$ for j odd and $0!! = 1$. This choice is made so that the polynomial is centered with respect to the Gaussian distribution. Thus, using a Taylor's theorem, we obtain the following approximation for any $A > 0$

$$\sup_{x\in[-A,A]} |(f(x) - a_{k-1})(x) - P_{k-1}(x)| \leqslant C_f \frac{A^{(1+c_f)k}}{k!} \tag{2.4.2}$$

Now, we will compare the Hermitized version of our matrix M, if we define

$$Y^{(a_k)} = f\left(\frac{WX}{\sqrt{n_0}}\right) - a_k, \quad \text{and} \quad Y_k = P_k\left(\frac{WX}{\sqrt{n_0}}\right), \tag{2.4.3}$$

we want to control the following $(m + n_1) \times (m + n_1)$ symmetric matrix

$$\mathcal{E} = \frac{1}{\sqrt{m}} \begin{pmatrix} 0 & Y^{(a_{k-1})} - Y_k \\ \left(Y^{(a_{k-1})} - Y_k\right)^* & 0 \end{pmatrix}. \tag{2.4.4}$$

If we consider the spectral radius of this matrix $\rho(\mathcal{E})$, we have the following bound

$$\rho(\mathcal{E}) \leqslant \frac{1}{\sqrt{m}} \max\left\{ \max_{1\leqslant i\leqslant n_1} \sum_{j=1}^{m} \left| \left(Y^{(a_{k-1})} - Y_k\right)_{ij} \right|, \max_{1\leqslant i\leqslant m} \sum_{j=1}^{n_1} \left| \left(Y^{(f-a_{k-1})} - Y^{(P_{k-1})}\right)_{ji} \right| \right\}. \tag{2.4.5}$$

Now consider the event, for $\delta_1 \in (0, \frac{1}{2})$,

$$\mathcal{A}_{n_1}(\delta_1) = \bigcap_{1\leqslant i\leqslant n_1} \bigcap_{1\leqslant j\leqslant m} \left\{ \left| \left(\frac{WX}{\sqrt{n_0}}\right)_{ij} \right| \leqslant (\log n_1)^{1/2+\delta_1} \right\}. \tag{2.4.6}$$

On this event, we have, considering the approximation (2.4.2),

$$\rho(\mathcal{E}) \leqslant C_f \sqrt{m} \frac{(\log n_1)^{k(1/2+\delta_1)(1+c_f)}}{k!}.$$

We thus need to consider k such that the right hand side of the previous inequality goes to zero. We can then take

$$k \geqslant c_0 \frac{\log n_1}{\log \log n_1} \quad \text{with} \quad c_0 > \frac{1}{2(1-(1+c_f)(\frac{1}{2}+\delta_1))}. \tag{2.4.7}$$

We obtain, by using Stirling formula, that there exists a $\delta_2 > 0$ such that for any $\varepsilon > 0$ we have

$$\rho(\mathcal{E}) = \mathcal{O}\left(\frac{n_1^{\varepsilon}}{n_1^{\delta_2}}\right).$$

By taking ε small enough we then see that, on the event $\mathcal{A}_{n_1}(\delta_1)$ and with k as in (2.4.7), $\rho(\mathcal{E}) \to 0$. It remains to see that the event $\mathcal{A}_{n_1}(\delta_1)$ occurs with high probability which comes from the additional assumption we put on the entries W_{ij} and X_{ij}. Indeed,

$$\begin{aligned}
\mathbb{P}\left(\mathcal{A}_{n_1}(\delta_1)^c\right) &= \mathbb{P}\left(\text{there exist } i,j \text{ such that } \left|\left(\frac{WX}{\sqrt{n_0}}\right)_{ij}\right| > (\log n_1)^{1/2+\delta_1}\right)\\
&\leqslant n_1 m \mathbb{P}\left(\left|\left(\frac{WX}{\sqrt{n_0}}\right)_{11}\right| > (\log n_1)^{1/2+\delta_1}\right)\\
&\leqslant C n_1 m e^{-(\log n_1)^{1+2\delta_1}/2}
\end{aligned}$$

which goes to zero faster than any polynomial in n_1. Thus, since we know the limiting empirical eigenvalue distribution of the matrix M constructed with the centered polynomial P_n as activation function, we know the limiting distribution for M constructed with $f - a_k$ instead. But if we denote $\mathbf{e}_p = {}^t(1,\dots,1) \in \mathbb{R}^p$ we have that

$$Y^{(a_k)} - Y = a_k \mathbf{e}_{n_1} {}^t\mathbf{e}_m. \tag{2.4.8}$$

Hence, $Y^{(a_k)}$ is just a rank one deformation of Y and we know by the rank inequalities (see [BS10] for instance) that they have the same limiting empirical eigenvalue distribution. □

2.5. Behavior of the largest eigenvalue

2.5.1. Convergence of the largest eigenvalue to the edge of the support

In this section we show that the largest eigenvalue of M sticks to the support of μ. We denote by u_+ the top edge of this support. This is the statement of the following theorem.

Theorem 2.5.1. *Let M be constructed as in (2.2.6) and denote λ_1 its largest eigenvalue. Then*

$$\lambda_1 \xrightarrow[n_1\to\infty]{} u_+ \quad \textit{in probability.}$$

The proof of Theorem 2.5.1 is again based on a preliminary polynomial approximation of f. Note that as in the previous section we will take $\sigma_w = \sigma_x = 1$ for simplicity of writing. One considers the centered Taylor-Lagrange approximation polynomial P_k defined in (2.4.1) and consider also $Y^{(a_k)}$ and Y_k defined in (2.4.3). Define then

$$R_k := \frac{1}{m} Y^{(a_k)} (Y^{(a_k)})^* - \frac{1}{m} Y_k Y_k^*. \tag{2.5.1}$$

The spectral radius of R_k can be bounded from above on the very high probability event $\mathcal{A}_{n_1}(\delta_1)$ defined in (2.4.6) by

$$\rho(R_{k-1}) \leqslant C_f\sqrt{m}\frac{(\log n_1)^{(1+c_f)(1/2+\delta_1)k}}{k!}. \tag{2.5.2}$$

The above goes to 0 as n_0 tends to infinity provided that $k \geqslant c_0 \frac{\log n_1}{\log\log n_1}$ for a constant $c_0 > 1$. From now on, we fix such a degree k for the approximation. Then, the largest eigenvalue of $Y^{(a_k)}(Y^{(a_k)})^*/m$ will be suitably approximated by that of $M_k = Y_k Y_k^*/m$.

Our next task is to show that the largest eigenvalue of M_k cannot exceed $u_+ + \delta$ for any $\delta > 0$ with probability arbitrarily close to 1. This is done in the next proposition using the method of high moments [FK81] and Markov's inequality :

$$\mathbb{P}(\lambda_1(M_k) > u_+ + \delta) \leqslant \frac{\mathbb{E}\mathrm{Tr}\, M_k^{2q}}{(u_+ + \delta)^{2q}}. \tag{2.5.3}$$

Proposition 2.5.2. *Let $0 < \alpha_1 < \alpha_2$ and $q = q(n_1)$ a sequence such that $q(n_1) \leqslant (\log n_1)^{1+\alpha_1}$. Assume that $k \leqslant k_0 := \frac{1}{1+\alpha_2}\frac{\log n_1}{\log\log n_1}$, then*

$$\mathbb{E}\left[\mathrm{Tr}\, M_k^{2q}\right] = n_1 \mathfrak{m}_{2q}(P_k)(1 + o(1)).$$

Assume that $k \geqslant k_0$, then there exists a constant $C > 0$ such that

$$\mathbb{E}\left[\mathrm{Tr}\, M_k^{2q}\right] \leqslant n_1 \mathfrak{m}_{2q}(P_{k_0})\left(1 + o(1)\right)$$

where $\mathfrak{m}_q$ is defined as (2.3.18) and $\mathfrak{m}_{2q}(P_{k_0})$ denotes the moment where the activation function is given by P_{k_0}.

Proof. We know that for q up to order $(\log n_1)^{1+\alpha_1}$ and $k < k_0$ we have that

$$\overline{m}_{2q}^{(P_k)} = \mathfrak{m}_{2q}(P_k)(1 + o(1))$$

which gives the first result of the proposition. This comes from Section 2.3 and in particular the bound (2.3.13) which gives that our convergence of moments is true up to q and k such that $q^k \ll n_0$ which holds for our choice of q and k.

However, in order to obtain the correct polynomial approximation we will need our degree k to be larger than $\frac{\log n_1}{\log\log n_1}$, this is not a problem because of our choice of polynomial. Indeed, for such high degrees, the $k!$ normalization will make the contribution of very high degrees vanish. This can be seen, for instance, by looking at the parameters θ_1 and θ_2 where the polynomial appears in the moment and in the errors. Consider $k > k_0$, then we can write, if we normalize so that the variances are equal to 1,

$$\theta_1(P_k) = \mathbb{E}\left[P_k(\mathcal{N})^2\right] = \mathbb{E}\left[(P_k(\mathcal{N}) - P_{k_0}(\mathcal{N}))^2\right] + \mathbb{E}\left[P_{k_0}(\mathcal{N})^2\right] + 2\mathbb{E}\left[P_{k_0}(\mathcal{N})(P_k(\mathcal{N}) - P_{k_0}(\mathcal{N})\right]$$

where $\mathcal{N}$ is a standard Gaussian random variable. By the Cauchy-Schwarz inequality, we now simply need to bound the first term,

$$\mathbb{E}\left[(P_k(\mathcal{N}) - P_{k_0}(\mathcal{N}))^2\right] = \sum_{\substack{i,j=k_0+1 \\ i+j \text{ even}}}^{k} a_i a_j \frac{(i+j)!!}{i!j!}$$

which goes to zeros exponentially by Stirling's formula. Thus we can see that we have for any $D > 0$,

$$\theta_1(P_k) = \theta_1(P_{k_0}) + \mathcal{O}\left(N^{-D}\right)$$

and the same thing holds for $\theta_2(P_k)$. In the leading order of the moment, this is the only part where the polynomial intervene since the admissible graphs do not depend on the activation function. However, since we choose k large, actually large enough for $q^k \gg n_0$, we need to check that the errors do not explode and actually vanish for $f_{k_0} = \sum_{i>k_0} \frac{a_i}{i!} x^k$. We saw from the previous analysis that the largest error comes from (2.3.14), for such a polynomial f_{k_0}, we can see that we thus need to bound the two quantities for $k_i, k_j > k_0$

$$\frac{k_i(k_i-1)!!}{k_i!}\frac{q^{k_i}}{n_0} \quad \text{and} \quad \frac{(k_i+k_j)!!}{k_i!k_j!}\frac{q^{k_i+k_j}}{n_0}.$$

We will bound the first one but the second one can be bounded in the same way. Note that these bound comes from the two different behaviors in the case of a cycle of length 2 and larger cycles. Now using Stirling's formula we can see that

$$\frac{k_i(k_i-1)!!}{k_i!}\frac{q^{k_i}}{n_0} = \mathcal{O}\left(\frac{\sqrt{k_i}}{n_0}\left(\sqrt{e}\frac{q}{\sqrt{k_i}}\right)^{k_i}\right)$$

This bound is decreasing in k_i so that we need to check its order for $k_i = k_0 = \frac{1}{1+\alpha_2}\frac{\log n_1}{\log\log n_1}$. And we obtain the following bound,

$$\frac{k_i(k_i-1)!!}{k_i!}\frac{q^{k_i}}{n_0} = \mathcal{O}\left(\frac{\psi(n_1)}{n_1^{-\frac{1+2(\alpha_2-\alpha_1)}{2(1+\alpha_2)}}}\right)$$

with the function ψ given by

$$\psi(n_1) = \sqrt{\frac{1}{1+\alpha_2}\frac{\log n_1}{\log\log n_1}}\left(\frac{e}{1+\alpha_2}\log\log n_1\right)^{\frac{1}{2(1+\alpha_2)}\frac{\log n_1}{\log\log n_1}} = \mathcal{O}(n_1^{\varepsilon})$$

for any $\varepsilon > 0$. Thus, recalling that $\alpha_2 > \alpha_1$ we have that for any $\varepsilon > 0$,

$$\frac{k_i(k_i-1)!!}{k_i!}\frac{q^{k_i}}{n_0} = \mathcal{O}\left(\frac{n_1^{\varepsilon}}{n_1^{-\frac{1+2(\alpha_2-\alpha_1)}{2(1+\alpha_2)}}}\right) = o(1)$$

by taking ε small enough. □

Proof of Theorem 2.5.1. First, see that we know from the convergence of the empirical eigenvalue distribution that for any $\delta > 0$,

$$\mathbb{P}\left(\lambda_1(M_k) < u_+ - \delta\right) \to 0$$

Now, for the other inequality we saw from (2.5.3) that we simply need to bound $\mathbb{E}\mathrm{Tr}\, M_k^{2q}$ using Proposition 2.5.2. We can see that even for $k > \frac{\log n_1}{\log\log n_1}$, we have the bound $\overline{m}_{2q}^{(P_k)} \leqslant 2u_+^{2q}$ from Proposition 2.5.2. Now, injecting this bound for the control of the largest eigenvalue we have that

$$\mathbb{P}\left(\lambda_{n_1}(M_k) > u_+ + \delta\right) \leqslant 2n_1\left(\frac{u_+}{u_+ + \delta}\right)^{2q} \xrightarrow[n_1\to\infty]{} 0.$$

Thus we have that convergence of the largest eigenvalue of M_k to the edge of the support which also gives convergence of the largest eigenvalue of $Y^{(a_k)}$ by the bound of the spectral radius from (2.5.2). Now we want to control the largest eigenvalue of $M = YY^*/m$ and we have seen that it is a rank one perturbation of Y by (2.4.8). Such a perturbation can possibly change the behavior of the largest eigenvalue but the perturbation here is small, indeed since our activation function f has a zero Gaussian mean we have that

$$\sum_{j=0}^{\infty} f^{(j)}(0)\frac{j!!}{j!} = \mathbb{E}\left[\sum_{j=0}^{\infty} f^{(j)}(0)\frac{\mathcal{N}^j}{j!}\right] = \mathbb{E}\left[f(\mathcal{N})\right] = 0$$

with $\mathcal{N}$ a standard Gaussian random variable. Thus, we have that there exists a $C > 0$ such that

$$|a_k| = \left|\sum_{j=0}^{k} f^{(j)}(0)\frac{j!!}{j!}\right| = \left|\sum_{j=k}^{\infty} f^{(j)}(0)\frac{j!!}{j!}\right| \leqslant \frac{C^{(k-1)}}{(k-1)^{(k-1)/2}}$$

We know that we can take k of order $c_0\frac{\log n_1}{\log\log n_1}$ for any $c_0 > 0$ by Proposition 2.5.2. In this case we obtain that for any $\varepsilon > 0$ we have $a_k = \mathcal{O}(n^{-c_0/2+\varepsilon})$. Now, we will use the Hoffman-Wielandt inequality for singular values to finish, indee we have

$$\left|\sqrt{\lambda_1(Y^{(a_k)})} - \sqrt{\lambda_1(Y)}\right| \leqslant \sqrt{\sum_{i=1}^{n_1}\left(\sqrt{\lambda_i(Y^{(a_k)})} - \sqrt{\lambda_i(Y)}\right)^2} \leqslant \left\|Y^{(a_k)} - Y\right\|$$

with $\|A\|^2 = \operatorname{Tr} AA^*$. But since we exactly know our rank one deformation we have that

$$\left\|Y^{(a_k)} - Y\right\| = \sqrt{a_k^2 m n_1} = \mathcal{O}\left(\frac{C}{n_1^{c_0/2-2-\varepsilon}}\right).$$

We finally obtain our result by taking $c_0 > 2 + 2\varepsilon$.

□

2.6. Propagation of eigenvalue distribution through multiple layers

In this section, we will study the eigenvalue distribution of the following nonlinear matrix model consisting of the covariance data after passing through several layers of the neural network. It has been conjectured in [PW17] that we have stability of eigenvalue distribution through the layers in the case where $\theta_2(f) = 0$. We give here a positive answer to this conjecture with the corresponding normalization to obtain stability and we also obtain a general formula for the moments in the case of going through two layers. The case of a single layer has been considered in Theorems 2.2.1 and 2.2.2 where we described the asymptotic empirical eigenvalue distribution in the one layer case. We explicit a combinatorial formula in the case of going through another layer and we see that if $\theta_1(f) \neq 0$ we obtain moments that differs from the single layer moments.

2.6.1. Eigenvalue distribution of $Y^{(2)}$

In the following theorem, we will give the moments of the empirical eigenvalue distribution of the matrix

$$M^{(2)} = \frac{1}{m}Y^{(2)}\left(Y^{(2)}\right)^* \in \mathbb{R}^{n_2\times n_2}, \quad \mu_{n_2}^{(2)} = \frac{1}{n_2}\sum_{i=1}^{n_2}\delta_{\lambda_i^{(2)}} \tag{2.6.1}$$

where $\lambda_{n_0}^{(2)} \leqslant \dots \leqslant \lambda_1^{(2)}$ are the eigenvalues of $M^{(2)}$. Define its expected moments

$$\overline{m}_q^{(2)} := \mathbb{E}\left[\langle\mu_{n_2}^{(2)}, x^q\rangle\right] = \mathbb{E}\left[\frac{1}{n_2}\sum_{i=1}^{n_2}\lambda_i^{(2)\,q}\right].$$

Theorem 2.6.1. *Let $f=\sum_{k=1}^K \frac{a_k}{k!}(x^k-k!!\mathbb{1}_{k\ even})$ be a polynomial such that (2.2.4) holds and q an integer. The degree of f, K, can grow with n_1 but suppose that*

$$K\leqslant \frac{\log n_1}{\log\log n_1}. \tag{2.6.2}$$

We then have the following asymptotics

$$\overline{m}_q^{(2)}=(1+o(1))\sum_{I_i,I_j=0}^{q}\sum_{b_0=0}^{I_i+I_j+1}(\psi_0\psi_1)^{I_i+1-q}\phi^{I_j}\theta_2^{q-b_0}(f)\theta_1^{b_0}(f)$$
$$\sum_{\mathbf{m}}\mathcal{A}(q,I_i,I_j,\mathbf{m},b_0)\prod_{i=2}^{q}\left(\sum_{I_\ell=0}^{i}\sum_{b_1=0}^{I_\ell+1}\mathcal{A}(i,I_\ell,0,b_1)\psi_0^{I_\ell}\left(\frac{\theta_2(f)}{\theta_1(f)}\right)^{i-b_1}\right)^{m_i}. \tag{2.6.3}$$

where $\mathcal{A}(q,I_i,I_j,b)$ denotes the number of admissible graphs with $2q$ edges, I_i i-identifications, I_j j-identifications and b cycles of size 1 as in Definition 2.3.2, the sum over $\mathbf{m}=(m_2,\dots,m_q)$ is over q-uplets such that $\sum im_i=q-b_0$ and $\sum m_i=I_i+I_j+1-b_0$, $\mathcal{A}(q,I_i,I_j,\mathbf{m},b_0)$ corresponds to the number of admissible graphs with the additional condition that there is m_i cycles of length i.

As in the previous section, we will first consider the case of an odd monomial of the form $f(x)=\frac{x^k}{k!}$. Note that the same argument as Subsection 2.3.2 gives that the even monomial are subleading so that the contribution of odd monomial gives the leading order in the asymptotic expansion. We can write the entries of $Y^{(2)}$ as

$$\begin{aligned}Y_{ij}^{(2)}=\frac{1}{k!}\left(\frac{\sigma_x}{\sqrt{\theta_1(f)}}\frac{W^{(1)}Y^{(1)}}{\sqrt{n_1}}\right)^k&=\frac{\sigma_x^k}{n_1^{k/2}k!\theta_1(f)^{k/2}}\left(\sum_{k=1}^{n_1}W_{ik}^{(1)}Y_{kj}^{(1)}\right)^k\\&=\frac{\sigma_x^k}{n_1^{k/2}k!\theta_1(f)^{k/2}}\sum_{\ell_1,\dots,\ell_k=1}^{n_1}\prod_{p=1}^{k}W_{i\ell_p}^{(1)}Y_{\ell_pj}^{(1)}.\end{aligned} \tag{2.6.4}$$

Then, developing the expected moment of the empirical eigenvalue distribution we obtain the following

$$\frac{1}{n_2}\mathbb{E}\left[\operatorname{Tr}\left(M^{(2)}\right)^q\right]=\frac{1}{n_0m^q}\mathbb{E}\sum_{i_1,\dots,i_q=1}^{n_2}\sum_{j_1,\dots,j_q=1}^{m}Y_{i_1j_1}^{(2)}Y_{i_2j_1}^{(2)}Y_{i_2j_2}^{(2)}Y_{i_3j_2}^{(2)}\dots Y_{i_qj_q}^{(2)}Y_{i_1j_q}^{(2)}$$

Thus injecting (2.6.4) in the previous equation we obtain the following development of the tracial moment of M

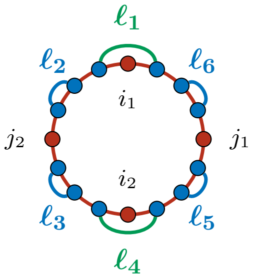
$$\begin{aligned}&\frac{1}{n_2}\mathbb{E}\left[\operatorname{Tr}\left(M^{(2)}\right)^q\right]=\\&=\frac{\sigma_x^{2kq}}{n_2m^qn_1^{kq}(k!)^{2q}\theta_1(f)^{kq}}\mathbb{E}\sum_{i_1,\dots,i_q}\sum_{j_1,\dots,j_q}\sum_{\substack{\ell_1^1,\dots\ell_k^1\\ \dots\\ \ell_1^{2q}\dots\ell_k^{2q}}}^{n_1}\prod_{p=1}^{k}W_{i_1\ell_p^1}^{(1)}Y_{\ell_p^1j_1}^{(1)}\prod_{p=1}^{k}W_{i_2\ell_p^2}^{(1)}Y_{\ell_p^2j_1}^{(1)}\cdots\prod_{p=1}^{k}W_{i_1\ell_p^{2q}}^{(1)}Y_{\ell_p^{2q}j_q}^{(1)}\end{aligned} \tag{2.6.5}$$

We call the terms contributing in a non negligible way *typical*. Now, we can give a graphical representation of these terms as in the previous sections.We will see that the contributing graphs are actually the same admissible graphs from Definition 2.3.2. However, there are less constraints in the choices of the *blue* edges. Indeed, the entries of the matrix Y are not independent and thus we do not need each entry to be matched with at least another. This constraint however holds for the entries of the matrix $W^{(1)}$.

We will first suppose that all the i-labeled vertices and j-labeled vertices are pairwise distinct and explain the combinatorics in this simpler case. We can see that the largest number of distinct ℓ indices we can get is kq by doing the following:

- Match at least two indices from different adjacent niches of an i-label index and perform a perfect matching between the $2k-2$ remaining indices. This type of matching gives kq different ℓ indices and match every $W^{(1)}$ entry with another. This is illustrated in the leftmost graph in Figure 2.14. Note that this type of matching gives the most distinct indices but is actually not necessarily of leading order (see Figure 2.15 for an illustration) and is not the sole leading order.
- Similarly to having possible identifications between i-labeled and j-labeled vertices in the single layer case, it will be possible to perform identifications between the *blue* edges and obtain something contributing in the dominant term in the asymptotic expansion (see Figure 2.16 for an illustration). This behavior will be explained in the second step when we develop the entries of $Y^{(1)}$.

A typical matching of ℓ indices corresponds to one such that all W_{ij} arises twice. This matching on the W entries induces one on the entries of $Y^{(1)}$. This corresponding joint moment thus induces another graph between j-labeled and ℓ-labeled vertices. The i-labeled vertices do not appear in the graph as they corresponded to the entries of W. This graph can be constructed from the initial graph by seeing which niches a *blue* edge links together. For instance, in Figure 2.14, we can see that ℓ_2 links the same niche adjacent to j_2 while ℓ_1 links the niches adjacent to j_1 and j_2.



Corresponding moment:
$Y_{\ell_1 j_1} Y^2_{\ell_2 j_1} Y^2_{\ell_3 j_i} Y_{\ell_4 j_1} Y_{\ell_4 j_2} Y^2_{\ell_5 j_2} Y^2_{\ell_6 j_2} Y_{\ell_1 j_2}$

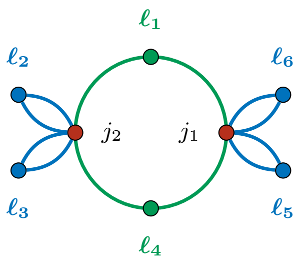


Figure 2.14: Graph obtained after a *blue* matching in the initial graph. The *green* edges, corresponding to bridges between niches, induce a cycle in the final graph. The remaining edges coming from matched pairs inside a niche will create simple cycles attached to j labeled indices. The basic graph obtained after another step is thus the following: one large cycle coming from bridges between niches and $k-1$ simple cycles attached to each j-labeled vertex.

In Figure 2.14, we can see that the matching on the initial graph induces another admissible graph. Note that it does not consist in one cycle but in a cycle (in green on the figure) where we attached to each j-labeled vertex $k-1$ cycles of length 2. However, in Figure 2.15, we can see that a *blue* matching on the initial cycle which maximizes the number of distinct indices gives a non-admissible graph. This comes from the fact that we have too many edges linking the same two distinct niches.

As we will see, the matching on the final graph will be subleading if we obtain a non-admissible graph. The following lemma gives us the typical matching on the initial graph. It states that the leading order is actually given by the matchings as in Figure 2.14.

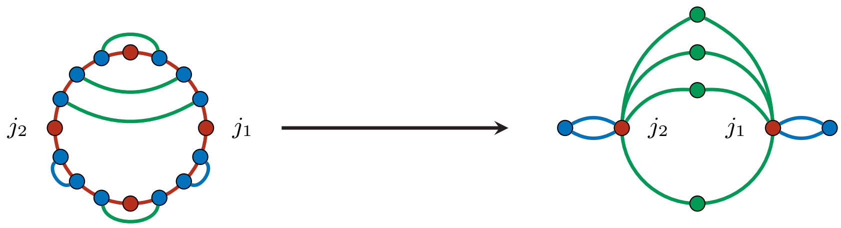


Figure 2.15: Non-admissible graph obtained after a *blue* matching which induces a maximum number of distinct indices in the initial cycle. We can see that several *green* bridges between the same niches create a non-admissible graph and is thus subleading via the analysis from the previous section.

Lemma 2.6.2. *Consider a cycle of length greater than 2, then the typical matchings on the* blue *vertices consist in the following:*

- *There is a single edge linking two niches adjacent to the same i-labeled vertex which we will call a* bridge.
- *Remaining edges inside a niche are matched according to a perfect matching.*
- *We can add identifications between only* bridges.

If the cycle is of length 2 *then we perform a perfect matching between the* $2k$ blue *vertices in the cycle.*

Proof. The proof is actually given by the construction of the second graph and the fact that admissible graphs are the dominant term in the asymptotic expansion. We will first see that any other matching gives a non-admissible second layer graph. Indeed, first see that if we have more than one *bridge* between two distinct niches, we will obtain too many paths from the two corresponding j-labeled indices which breaks the tree structure of an admissible graph. The same reasoning holds for possible identifications between bridges and a matched pairs inside a niche. If we identify two matched pairs inside a niche, we can see via the construction of the graph that it will create double edges as we would obtain an entry of $Y^{(1)}$ to the power of 4.

However, note that in the cycle of size q, we can add identifications between the q *bridges* and keep having an admissible graph. Indeed, one can see that such identifications do not increase the number of paths from a j-label vertex to another and the tree structure of the graph is conserved. In other words, every edge belongs to a unique cycle. This behavior is illustrated inf Figure 2.16 where we perform identifications between bridges and still obtain an admissible graph.

We need to see now that non-admissible graphs are subleading in the asymptotic expansion. In the case where the first *red* graph (on the i and j indices) is non-admissible the considerations from Subsection 2.3.1 hold and state that their contribution is subleading. Indeed, we first need to choose the fundamental cycles of the graph and then perform a matching. We saw that it involves at most an error of q^k/n_0 which is a $o(1)$.

Now, even if the initial *red* graph is admissible, we saw that some matchings on the *blue* vertices can create a non-admissible graph for the second layer as in Figure 2.15. We now need to see that the contribution of these matchings is also subleading. As in Subsection 2.3.1, we have additional identifications between the vertices and we need to choose the fundamental cycles. We saw also that the largest error comes from the possible multiple cycles of length 2 attached together as in Figure 2.8.

Fix a vertex j_0, if we match together $2p$ ℓ indices together in the niche adjacent to j_0, we saw in (2.3.13) that the corresponding error will be given by $\mathcal{O}(n_0(k(2p)^k/n_0)^p)$. However, we can now have up to $2k$ indices matched together so that the contribution of non admissible graphs in this case is given by

$$\sum_{p=2}^{k} n_0 \left(\frac{k(2p)^k}{n_0} \right)^p = o(1) \quad \text{for} \quad k \leqslant \frac{\log n}{\log \log n}.$$

It actually decays faster than any polynomial for such k. Note that this actually is the only case which differs from Subsection 2.3.1 as the multiplicity of the cycle of length 2 can be k and thus we can match up to $2k^2$ indices together in this multiple cycle. □

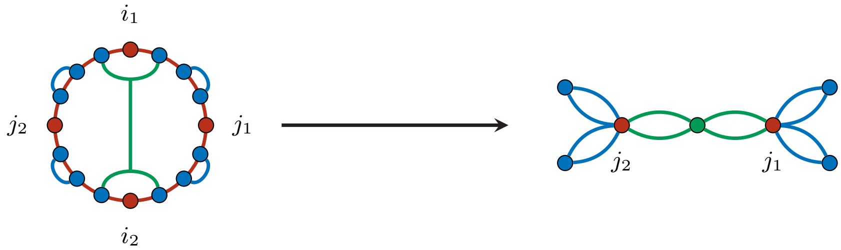


Figure 2.16: Admissible graph obtained by a matching in the initial graph with an identification between the two *bridges*. While we identified two ℓ vertices, and thus lost an order of n_1, this matching is still of leading order because we gained a cycle in the induced graph.

We are now able to give the contribution of one cycle through this two layers construction. It consists in first doing a typical matching as in Lemma 2.6.2 on the initial graph and then perform a typical matching on the induced admissible graph as in Subsection 2.3.1.

Lemma 2.6.3. *The contribution to the expected moment $\mathbb{E}\langle \mu_{n_2}, x^k \rangle$ of one cycle of size q is given by the following: if $q > 1$,*

$$E_q(f) = \theta_2^q(f) \sum_{I_\ell=0}^{q} \sum_{b_1=0}^{q} \mathcal{A}(q, I_\ell, 0, b_1) \psi_0^{I_\ell+1-q} \psi_1^{1-q} \left(\frac{\theta_2(f)}{\theta_1(f)} \right)^{q-b_1}$$

and if $q = 1$,

$$E_1(f) = \theta_1(f).$$

Proof. We will begin by the case of $q = 1$. We have a cycle of length 2 as in Figure 2.3a and we saw that the dominant term in the asymptotic expansion consists in performing a perfect matching between all edges. We can thus see that the contribution coming from this first construction is given by

$$\frac{\sigma_x^{2k}}{n_2 m n_1^k} n_2 m n_1^k \left(\sigma_w^{2k} (2k)!! \right) = \theta_1(f)(1 + o(1)).$$

Now, this construction on the initial graph will give us another graph as in Figure 2.17. Note that the orders of n in the previous contribution comes from the choices for the i index, the j index and the ℓ indices. We see that the graph obtained from a cycle of length 2 will be an admissible graph where all j's are identified to a single vertex and k cycles of length 2 are attached to it (corresponding to the k *blue* edges in the initial cycle).

Now that we have our second layer graph constructed, we can do the same reasoning as before and develop the entries $Y^{(1)}$ as a product of entries of $W^{(0)}$ and X. Since we have an admissible graph, we now that the dominant term in the asymptotic expansion will be to perform a perfect matching in all cycles of length 2 as in the Section 2.3 and illustrated in Figure 2.17. Thus we will add a contribution of

$$\frac{1}{n_0^{k^2}\theta_1(f)^k} n_0^{k^2}\left(\sigma_w^{2k}\sigma_x^{2k}(2k)!!\right)^k = 1+o(1).$$

Here, the normalization in ${n_0}^{-k^2}$ comes from the total number of entries in the joint moment coming from the initial graph: there are $2k$ entries with a normalization of ${n_0}^{-k/2}$. We then have to choose $n_0^{k^2}$ indices in the second graph. Finally, we obtain for the final contribution for a cycle of length 2

$$E_1(f) = \theta_1(f)(1+o(1)).$$

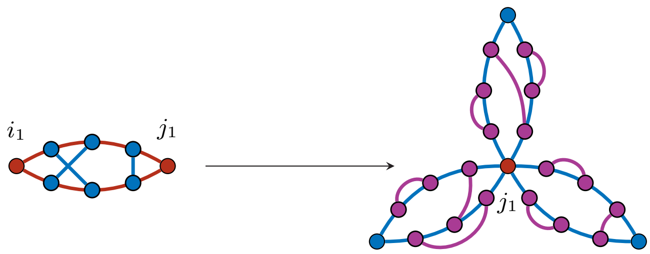


Figure 2.17: Construction and matching on the second layer graphs from a matching on the initial graph. One can see that the first graph gives a combinatorial factor of $(2k)!!$ while the second graph will give a factor of $(2k)!!^k$.

For the case of $q > 1$, we saw in Lemma 2.6.3 that the typical matchings consist in one bridge between niches, perfect matchings inside the niches and possible identifications between the bridges (as in Figure 2.16 for instance). Thus, we can sum over the number of identifications I_ℓ we do between the bridges and b_1 the number of cycles of length 2 we obtain after construction of the induced graph. We see that we can obtain any admissible graph with $2q$ edges, b simple cycles and I_ℓ identifications while there are no additional j-identifications.

First see the contribution if we do I_ℓ identifications coming from the initial graph:

$$\frac{\sigma_x^{2kq}}{n_2 m^q n_1^{kq}} n_2^q m^q n_1^{kq-I_\ell}\left(\sigma_w^k k(k-1)!!\right)^{2q} = \frac{1}{n_1^{I_\ell} n_2^{1-q}}\theta_2(f)^q. \tag{2.6.6}$$

Indeed, we have q choices for the i's and j's labels and $kq - I_\ell$ choices for the ℓ indices. The choices of the bridges between niches gives k^{2q} and the perfect matchings in the remaining vertices in each niche gives $(k-1)!!^{2q}$.

Now, we can have any admissible graph with I_ℓ identifications and b_1 cycles of length 2. Note that there are two types of cycles of length 2 in the second layer graph (which is illustrated by having distinct colors in Figure 2.16). The cycles of length 2 coming from possible identifications between the bridges (in green) and the cycles of length 2 coming from all the matched pairs inside each niche (in blue). These last cycles will always appear and does not depend on the shape of the admissible graph coming from the bridges. Thus, while we sum the number of cycles of length 2 in the second layer graph b_1 between 0 and q, the total number of cycles of length 2 will be $b_1 + q(k-1)$.
Thus for I_ℓ and b_1 fixed, and for an admissible graph with $2q$ edges, I_ℓ identifications and b_1 cycles of

length 2, the additional contribution will be given by

$$\frac{1}{n_0^{k^2 q}\theta_1(f)^{kq}} n_0^{I_\ell+1-b_1+(q-b_1)(k-1)} n_0^{kb_1} n_0^{qk(k-1)} \left(\sigma_w^k \sigma_x^k k(k-1)!!\right)^{2(q-b_1)} \times$$
$$\times \left(\sigma_w^{2k}\sigma_x^{2k}(2k)!!\right)^{b_1} \left(\sigma_w^{2k}\sigma_x^{2k}(2k)!!\right)^{(k-1)q} = n_0^{I_\ell} n_0^{1-q} \left(\frac{\theta_2(f)}{\theta_1(f)}\right)^{q-b_1}. \quad (2.6.7)$$

To understand this contribution, see the following: we have and admissible graph with $I_\ell + 1$ cycles including b_1 cycles of length 2. In the $I_\ell + 1 - b_1$ cycles of length greater than 2, we know that the dominant term is to perform a cycle parcouring each niche and perfect matchings between the remaining vertices in each niche. So if we look at a cycle of length q_i we obtain a contribution of

$$n_0^{1+(k-1)q_i} \left(\sigma_w^k \sigma_x^k k(k-1)!!\right)^{2q_i}.$$

Since it holds for any cycle of length greater than 2, and we have $I_\ell + 1$ of them, we can do the product over every such cycles and see that $\sum q_i = q - b_1$. Now, for each of the b_1 cycles of length 2, we perform a perfect matching inside each cycle and obtain

$$n_0^{kb_1} \left(\sigma_w^{2k}\sigma_x^{2k}(2k)!!\right)^{b_1}.$$

Finally it remains the $q(k-1)$ cycles of length 2 coming from the matched pairs in the initial graph. We also perform a perfect matching in each of these cycles which gives

$$n_0^{q(k-1)k} \left(\sigma_w^{2k}\sigma_x^{2k}(2k)!!\right)^{q(k-1)}. \quad (2.6.8)$$

Now, we can see that this last contribution only depends on I_ℓ and b_1 so that we can sum over admissible graph and obtain

$$\sum_{I_\ell=0}^{q} \sum_{b_1=0}^{q} \mathcal{A}(q, I_\ell, 0, b_1) \left(\frac{n_0}{n_2}\right)^{1-q} \left(\frac{n_0}{n_1}\right)^{I_\ell} \theta_2^q(f) \left(\frac{\theta_2(f)}{\theta_1(f)}\right)^{q-b_1}$$

and we obtain the final result by seeing that $\frac{n_0}{n_2} \to \psi_0\psi_1$ and $\frac{n_0}{n_1} \to \psi_0$. □

Now that we have the contribution for a given cycle of size q, it is easy to generalize to any admissible graph. Indeed, we will see that we can perform every matching and construct the induced graph for every cycle independently. Note that for any admissible graph, the induced graph is not necessarily connected but for a typical matching where we perform in each cycle a matching as in Lemma 2.6.2 we will obtain a forest of admissible graphs. This is illustrated in Figure 2.18. The fact that the induced graph is not connected is of no importance to compute the moment as the dominant contribution, as we have seen in the Section 2.3, consists of performing matchings independently in each cycle.

Proof of Theorem 2.6.1. We will begin to see that we can perform the typical matchings independent in each cycle. This statement is less clear than in Section 2.3 since we can perform cross-cycles *blue* edges without diminishing the number of ℓ indices in the initial graph. However, we lose a choice of index in the induced graph. Indeed, consider two cycles of length $2q_1$ and $2q_2$ that are attached on a vertex. If the vertex corresponds to a j index then the argument as in Figure 2.6 still holds since we need to match the W entries corresponding to i vertices by independence. We can then suppose that the vertex is a certain i-labeled vertex i_0.

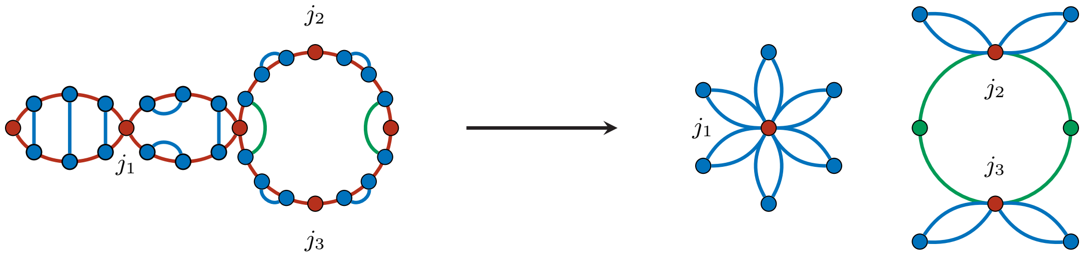


Figure 2.18: Admissible graph which induces a non connected second layer graph. However, note that the leading order matching in the initial graph induces each connected part of the second layer graph to be an admissible graph. We can thus do the analysis of admissible graphs by looking at each cycle separately.

If we perform the matchings independently in each cycle we have: $2(q_1+q_2)-1$ choices for the *red* vertices (remember that they are attached to a common vertex), in each cycle we have kq_i choices of *blue* vertices, in the induced graph for each cycle we have $1+(k-1)q_i+k(k-1)q_i$ choices since we have the choice for the cycle going over every niche, the choices for the matched pairs in each niche and the choices in the $k-1$ cycles of length 2 attached to each j vertex. Finally we have a total of $1+(q_1+q_2)(2+k^2+k-1)$ choices of indices.

If we perform cross-cycles edges between the cycles at the common vertex i_0 then we still have in the initial graph $(q_1+q_2)(2+k)-1$ choices. However, the induced graph consists now of one cycle of length $2(q_1+q_2)$. In this cycle we now have 1 choice for the cycle parcouring each niche, $(k-1)(q_1+q_2)$ for the matched pairs in each niche and $k(k-1)(q_1+q_2)$ choices for the cycles of length 2 attached to each j vertex. We then have a total of $(q_1+q_2)(2+k^2+k-1)$ choices to make. Thus, we can see that we lose a power of n_0 by performing cross-cycles edges.

As we saw from Lemma 2.6.3, the contribution of one cycle only depends on its length with a different behavior if it is of length 2 or not. In order to compute the contribution of an admissible graph, it thus depends on the length of its cycles. Thus we will denote, for an admissible graph $\mathcal{G}$ with $2q$ edges and b_0 cycles of length 2, $\mathbf{m}(\mathcal{G})=(m_2(\mathcal{G}),\dots,m_q(\mathcal{G}))$ where

$$m_i(\mathcal{G}) = \text{ number of cycles of length } 2i \text{ in } \mathcal{G} \quad \text{in particular} \quad \sum_{i=2}^{q} i m_i(\mathcal{G}) = q-b_0$$
$$\text{and} \quad \sum_{i=2}^{q} m_i(\mathcal{G}) = I_i+I_j+1-b_0.$$

Using the same reasoning as in Lemma 2.6.3, we are able to compute the limiting expected moment, we obtain

$$\overline{m}_q = \frac{1+o(1)}{n_2 m n_1^{kq} n_0^{k^2 q}} \sum_{I_i,I_j=0}^{q} \sum_{b_0=0}^{I_i+I_j+1} n_2^{q-I_i} m^{q-I_j} \sum_{\mathbf{m}} \mathcal{A}(q,I_i,I_j,\mathbf{m},b_0)\times$$
$$\times \left[\prod_{i=2}^{q}\left(\sum_{I_\ell=0}^{i}\sum_{b_1=0}^{I_\ell+1} \mathcal{A}(i,I_\ell,0,b_1) n_1^{ki-I_\ell} n_0^{I_\ell+1+(k-1)(i-b_1)-b_1+kb_1+ik(k-1)} \left(\frac{\theta_2(f)}{\theta_1(f)}\right)^{i-b_1}\theta_2^i(f)\right)^{m_i}\right]\times$$
$$\times\, n_1^{kb_0} n_0^{k^2 b_0}\theta_1^{b_0}(f)$$

This identity comes from applying Lemma 2.6.3 to each cycle independently. Now, using the fact that

$\sum_i im_i = q - b_0$ and that $\sum m_i = I_i + I_j + 1 - b_0$ and simplifying this expression, we obtain the following formula for the expected moment,

$$\overline{m}_q = (1+o(1)) \sum_{I_i, I_j = 0}^{q} \sum_{b=0}^{I_i+I_j+1} (\psi_0\psi_1)^{I_i+1-q} \phi_0^{I_j} \theta_2^{q-b_0}(f)\theta_1^{b_0}(f) \times$$
$$\times \sum_{\mathbf{m}} \mathcal{A}(q, I_i, I_j, \mathbf{m}, b_0) \prod_{i=2}^{q} \left(\sum_{I_\ell=0}^{i} \sum_{b_1=0}^{I_\ell+1} \mathcal{A}(i, I_\ell, 0, b_1) \psi_0^{I_\ell} \left(\frac{\theta_2(f)}{\theta_1(f)}\right)^{i-b_1} \right)^{m_i}$$

which gives the final result. □

2.6.2. Invariance of the distribution in the case when $\theta_2(f)$ vanishes

In the last subsection we computed the moments of the eigenvalue distribution of the covariance data matrix after two layers. While this formula does not seem to be constructive, it is interesting to look at the special case of $\theta_2(f) = 0$. Indeed, for the first-layer covariance data matrix, from Theorem 2.2.1, we obtain that the limiting eigenvalue distribution is given by the Marčenko-Pastur distribution with shape $\frac{\phi}{\psi}$. We can see this by looking at the limiting moments when $\theta_2(f)$ vanishes and this is given by the following lemma

Lemma 2.6.4. *Let q be a positive integer we have the following equality*

$$\sum_{\substack{I_j, I_i = 0 \\ I_i + I_j + 1 = q}}^{q-1} \mathcal{A}(q, I_i, I_j, q) \psi^{1-q+I_i} \phi^{I_j} \theta_1^q(f) = \theta_1^q(f) \sum_{k=0}^{q-1} \left(\frac{\phi}{\psi}\right)^k \frac{1}{k+1} \binom{q}{k} \binom{q-1}{k} = \theta_1^q(f) \langle x^q, \mu_{\phi/\psi} \rangle$$

where $\mu_{\phi/\psi}$ is the Marčenko-Pastur distribution with shape parameter $\frac{\phi}{\psi}$.

Proof. Firstly, see that we can slightly rewrite the left hand side,

$$\sum_{\substack{I_j, I_i = 0 \\ I_i + I_j + 1 = q}}^{q-1} \mathcal{A}(q, I_i, I_j, q) \psi^{1-q+I_i} \phi^{I_j} \theta_1^q(f) = \theta_1^q(f) \sum_{k=0}^{q-1} \left(\frac{\phi}{\psi}\right)^k \mathcal{A}(q, q-k-1, k, q).$$

Now it only remains to see that

$$\mathcal{A}(q, q-k-1, k, q) = \frac{1}{k+1} \binom{q}{k} \binom{q-1}{k}. \tag{2.6.9}$$

This fact comes from considering the right representation of admissible graphs. Here, we look at admissible graphs with $2q$ edges, q simple cycles, k j-identifications and $q-k-1$ i-identifications. The only admissible graphs following these conditions are graphs made of simple cycles with $q-k$ j-labeled vertices and $k+1$ i-labeled vertices. Thus we can count this at double trees, in the sense that one of every two vertices are i-labeled and the others are j-labeled, with the corresponding amount of each type of vertex. This number is known [CYY08] and is actually given by (2.6.9).

□

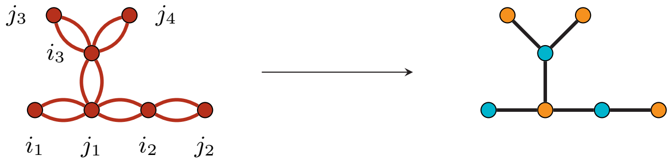


Figure 2.19: When an admissible graph is only given by simple cycles we can entirely encode it by a double tree where the two colors gives us the choice of which vertices are i-labeled and which one are j-labeled.

Note that the computation performed in Lemma 2.6.4 exactly corresponds to computing our limiting moment in the case of $\theta_2(f) = 0$. Indeed, in our formula for $\mathfrak{m}_q$, we see that if $\theta_2(f)$ vanishes, the only remaining terms consist in the graphs where $b = q$ which corresponds to graphs where all cycles have length 2.

This fact then means that if we consider a function f such that $\theta_2(f) = 0$, the distribution (up to a change in variance and shape) stays the same for our covariance data matrix after going through one layer of the network. Indeed, if one considers the matrix $\frac{1}{m\sigma_x^2}XX^*$, the asymptotic eigenvalue distribution is given by μ_ϕ the Marčenko-Pastur distribution with shape parameter ϕ. Now, after a layer of the network, we see that $\frac{1}{m\theta_1(f)}YY^*$ is given by $\mu_{\phi/\psi}$. This observation was made in [PW17] where it was conjectured that the distribution would stay invariant through several layers of the spectrum for this family of activation function. We can already answer the question for the second layer as we computed the expected moments in the previous subsection. Indeed, if one look at the formula for the expected moment (2.6.3) for the deterministic limiting moment, we obtain for $\theta_2(f) = 0$

$$\mathfrak{m}_q^{(2)} := \theta_1^q(f)\sum_{k=0}^{q-1}\left(\frac{\phi}{\psi_0\psi_1}\right)^k \mathcal{A}(q, q-k-1, k, q)$$

$$= \theta_1^q(f)\sum_{k=0}^{q-1}\left(\frac{\phi}{\psi_0\psi_1}\right)^k \frac{1}{k+1}\binom{q}{k}\binom{q-1}{k} = \theta_1^q(f)\langle x^q, \mu_{\phi/(\psi_0\psi_1)}\rangle$$

So we can see that we have the following behavior for the eigenvalue distribution depending on the activation function in the case of $\theta^{(2)}(f)$ vanishes for zeros to two layers of the neural network

$\theta_2(f)$	Data Covariance Matrix	Eigenvalue distribution
	$\frac{1}{m\sigma_x^2}XX^*$	μ_ϕ
$\theta_2^{(0)} = 0$	$\frac{1}{m\theta_1^{(0)}(f)}Y^{(1)}Y^{(1)*}$	μ_{ϕ/ψ_0}
$\theta_2^{(1)} = 0$	$\frac{1}{m\theta_1(f)}Y^{(2)}Y^{(2)*}$	$\mu_{\phi/(\psi_0\psi_1)}$

Thus we can conjecture the following pattern, for L the number of layers the data has gone through, we have the following limiting distribution if $\theta_2(f) = 0$ for the matrix

$$M^{(L+1)} = \frac{1}{m\theta_1(f)}Y^{(L+1)}Y^{(L+1)^*} \tag{2.6.10}$$

has for limiting eigenvalue distribution $\mu_{\phi/\prod_{i=0}^{\ell}\psi_i}$ the Marčenko-Pastur distribution with shape parameter $\frac{\phi}{\psi_0\psi_1\cdots\psi_\ell}$. This is the statement of the following theorem. Again we will first describe the moments with polynomial activation function and finish via a polynomial approximation.

Theorem 2.6.5. *Let $f = \sum_{k=1}^{K} \frac{a_k}{k!}(x^k - k!!\mathbb{1}_{k \text{ even}})$ be a polynomial such that (2.2.4) holds. The degree of f, K, can grow with n_1 but suppose that*

$$K = \leqslant \frac{1}{L-1}\frac{\log n_1}{\log\log n_1}. \tag{2.6.11}$$

Denote the empirical eigenvalue distribution of $M^{(L)}$ constructed as in (2.6.10), $\mu_{n_L}^{(L)} = \frac{1}{n_L}\sum_{i=1}^{n_L}\delta_{\lambda_i^{(\ell)}}$ and its expected moments

$$\overline{m}_q^{(L)} := \mathbb{E}\left[\langle \mu_{n_L}^{(L)}, x^q\rangle\right] = \mathbb{E}\left[\frac{1}{n_L}\sum_{i=1}^{n_L}\lambda_i^{(L)q}\right].$$

We then have the following asymptotics

$$\overline{m}_q^{(L)} = \left(\sum_{k=1}^{q-1}\left(\frac{\phi}{\prod_{i=0}^{L-1}\psi_i}\right)^k \frac{1}{k+1}\binom{q}{k}\binom{q-1}{k} + \theta_2(f)T(q,k,L)\right)(1+o(1)). \tag{2.6.12}$$

where $T(q,k,L)$ is a nonexplicit factor.

Proof. We will again first develop the arguments in the case of a monomial of odd degree since the case of an even monomial is completely similar (we only consider graphs with simpple cycles). The reasoning is actually similar to that of Theorem 2.6.1 as we will study and count the admissible graphs along each layer. However, if we look at the expression we want for our limiting formula, since we are actually only interested in the case where $\theta_2 = 0$, we want to exhibit the leading order where no θ_2 appears. It can easily be seen from the previous arguments that this consists in looking only at admissible graphs made of cycles of length 2 (and corresponding to double trees as in Figure 2.19).

The process is actually simpler than in the proof of 2.6.1 since we only look at graph with cycles of length 2. For the first step of the procedure see that by the construction explained above we will obtain a forest of *star admissible graph* where each graph is given by a unique j-labeled vertex attached to a certain number of cycles of length 2. Indeed, we saw that we should perform matching independently in each cycle to obtain the dominant term. To obtain a larger cycles in the induced graph, we need a bridge between cycles which is subleading as we perform a cross-cycle edge.

Consider now a connected component of the induced forest which corresponds to a unique j vertex. The number of cycles of length 2 attached to j consists in the total number of cycles adjacent to j in the previous steps multiplied by k (since we have k *blue* edges in each simple cycle). From this first process we then get the following contribution for this first two steps

$$\frac{\sigma_x^{2kq+2k^2q}(1+o(1))}{n_L m^q \theta_1(f)^{q+kq+k^2q}} \sum_{\substack{I_i, I_j \\ I_i+I_j+1=q}} \mathcal{A}(q, I_j, I_j, q) n_L^{q-I_i} \times$$

$$\times m^{q-I_j}\frac{1}{n_{L-1}^{kq}} n_{L-1}^{kq}\left(\sigma_w^{2k}(2k)!!\right)^q \frac{1}{n_{L-2}^{k^2q}} n_{L-2}^{k^2q}\left(\sigma_w^{2k}(2k)!!\right)^{kq} = (1+o(1))\sum_{k=0}^{q-1}\mathcal{A}(q, q-k-1, k, q)\left(\frac{n_L}{m}\right)^k.$$

To understand this contribution see that we have $n_L^{q-I_i}$ choices needed to label the i-labeled vertices and m^{q-I_j} for the j-labeled vertices, then for the powers of n_{L-1} we have the normalization and the corresponding number of ℓ indices to choose. Finally in each cycle of length 2 we perform a perfect matching between the two niches, we have q cycles of length 2 in the initial graph and kq such cycles in the forest obtained. See Figure 2.20 for an illustration.

Now, we can perform the third step of the procedure, we now have a forest of these *star admissible graphs* where each graph has only one j vertex. We saw that the j vertex is now attached to k times the number it was attached to in the previous step. Thus, the total number of cycles of length 2 in the forest is given by k^3q. We can perform this for each layer the data covariance matrix goes through as the only thing changing is the number of cycles of length 2 attached to each j vertex.

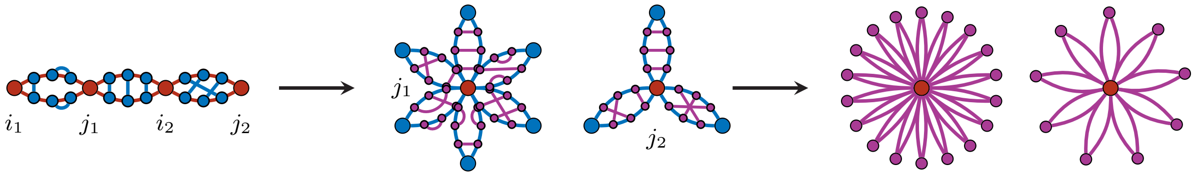


Figure 2.20: Effect on going through several layers for admissible graphs with only cycles of length 2. The first step consists of separating each j-labeled vertex into his own graph where he is attached to cycles of length 2. The number of these cycles is the number of matched pairs in every cycles adjacent to j. At each layer after the first one, we multiply by k (here 3) the number of cycles attached. For instance, in the initial graph, j_1 is attached to 2 simple cycles with $2 \times 3 = 6$ *blue* edges. Thus, in the next step, j_1 will be attached to 6 simple cycles and for the next step $6 \times 3 = 18$ simple cycles.

We can then see that at the layer L_0 we multiply by the term

$$\frac{1}{n_{L_0}^{k^{L-L_0}q}\theta_1^{k^{L-L_0}q}} n_{L_0}^{k^{L-L_0}q}\theta_1^{k^{L-L_0}q}(f).$$

Thus the whole contribution can be written in the following way

$$\sum_{k=0}^{q-1}\left(\frac{n_\ell}{m}\right)^k \mathcal{A}(q, q-k-1, k, q).$$

And we obtain our final result by seeing that

$$\left(\frac{n_\ell}{m}\right) \to \frac{\phi}{\psi_0\psi_1\dots\psi_{\ell-1}} \quad \text{and} \quad \mathcal{A}(q, q-k-1, k, q) = \frac{1}{k+1}\binom{r}{k}\binom{r-1}{k}.$$

Now, in the statement of the theorem we do not explicit the leading contribution of admissible graphs with at least one cycle of length greater than 2. We only need now to get an estimate on the other possible errors and show that they are a $o(1)$. The errors in the computation can only come from subleading matching on the graph at each possible step. Since we now know that the dominant term at each step is given by admissible graphs the whole analysis of errors from Section 2.3 stays true. However, the main difference comes from the number of vertices at each step which is $k^{L_0}q$ instead of just kq. Note that it still only consists of a power of k which grows slower than any power of n_1.

Again, the leading contribution of the errors comes from possible multiple edges arising in the graph. Say that a given j vertex is first connected to r cycles of length 2 in the initial graph. At the step L_0, it is now connected to $k^{L_0-1}r$ cycles of length 2. Thus if at this stage we connect *blue* indices together, say p of them we will obtain at the next step a multiple edge of multiplicity $2p$. We have a total of $2k^{L_0}r$ *blue* indices to match at this stage since we have $2k$ vertices per cycle of length 2. Thus, by comparing the contribution of such matchings with the typical matching we obtain, similarly to (2.3.13),

$$\sum_{p=2}^{k^{L_0}r} n_0 \left(\frac{Ckp^k}{n_0}\right)^p = o(1) \quad \text{for} \quad k \leqslant \frac{1}{L_0}\frac{\log n_1}{\log\log n_1}.$$

Now L_0 ranges from 1 to $L-1$ so that we obtain the bound that we need $k \leqslant \frac{1}{L-1}\frac{\log n_1}{\log\log n_1}$. □

This analysis of admissible graphs consisting in a tree of cycles of length 2 gives us that the Marčenko-Pastur distribution can be attained in any layer of the network by choosing the corresponding activation function.

Proof of Theorem 2.2.4. We showed that for a polynomial for up to degree $\frac{1}{L-1}\frac{\log n_1}{\log\log(n_1)}$, the expected moments of the eigenvalue distribution are given by the moments of the Marčenko-Pastur distribution with the correct shape parameter. We will first see that the variance of the moments is of order k^L/n_1^2 in order to show convergence of the actual moments. The principle is similar to that of Lemma 2.3.4 as we will count the corresponding graphs such that their covariance is non zero.

We can perform the same expansion as in Lemma 2.3.4 and see that we have for the first layer

$$\operatorname{Var} m_q^{(L)} = \frac{1}{n_1^2}\sum_{\mathcal{G}_1,\mathcal{G}_2}\sum_{\boldsymbol{\ell_1},\boldsymbol{\ell_2}} \mathbb{E}\left[M_{\mathcal{G}_1}^{(L)}(\boldsymbol{\ell_1})M_{\mathcal{G}_2}^{(L)}(\boldsymbol{\ell_2})\right] - \mathbb{E}\left[M_{\mathcal{G}_1}^{(L)}(\boldsymbol{\ell_1})\right]\mathbb{E}\left[M_{\mathcal{G}_2}^{(L)}(\boldsymbol{\ell_2})\right] \tag{2.6.13}$$

with

$$M_{\mathcal{G}}^{(L)}(\boldsymbol{\ell}) = \sum_{k_1,\dots,k_{2q}=1}^{K} \frac{a_{k_1}\dots a_{k_{2q}}}{m^q n_0^{\sum k_i/2}} \prod_{p=1}^{k_1} W_{i_1\ell_p^1}^{(L)} Y_{\ell_p^1 j_1}^{(L)} \prod_{p=1}^{k_2} W_{i_2\ell_p^2}^{(L)} Y_{\ell_p^2 j_1}^{(L)} \cdots \prod_{p=1}^{k_{2q}} W_{i_1\ell_p^{2q}}^{(L)} Y_{\ell_p^{2q} j_q}^{(L)}.$$

Now, in order to have a non vanishing contribution in the variance (2.6.13), we need to have additional identifications between the two graphs. Indeed, we need either at a given layer L_0 to have an entry of $W^{(L_0)}$ to be matched between $\mathcal{G}_1$ and $\mathcal{G}_2$ or at the last layer to have identifications between the X entries. It is possible to have identifications between $Y^{(L_0)}$ but after expansion into entries of $W^{(L_0-1)}$ and $Y^{(L_0-1)}$ we can obtain independent matchings. Thus, to have a non vanishing term, we would need to have other additional identifications in the layers beyond L_0. Since at each step we would lose an order $\mathcal{O}(q^2(k)^{2L_0}/n_0)$ (from the choice of which vertices to identify and the fact that we have one less choice of index), we see that it would be of higher order to simply identify X entries in the two last layers.

Thus, since our moments are still given by admissible graphs a similar analysis can be done as in Lemma 2.3.4: we can right at the first layer identify a i and j vertices to obtain an identification on the $W^{(L)}$ entries or choose two $W^{(L_0)}$ entries to identify at a given layer L_0 (or X entries at the last layers $L_0 = 1$) and thus we obtain

$$\operatorname{Var} m_q^{(L)} = \mathcal{O}\left(\frac{q^4 + q^2\sum_{L_0=1}^{L} k^{2L_0} + \sum_{L_0=1}^{L} k^{4L_0}}{n_0^2} C^q\right) = \mathcal{O}\left(\frac{k^{4L+4}}{n_0^2}\right)$$

since q is fixed here.

Let us now extend the result to a bounded function f. As in Section 2.4, we consider a polynomial P_k such that, for some $A > 0$,

$$\sup_{x\in[-A,A]} |(f(x) - a_k) - P_k(x)| \leqslant C_f \frac{A^{(1+c_f)k}}{(n+1)!}.$$

Now, we can consider $Y^{(L,a_k)}$ the matrix constructed as (2.2.9) with $f - a_k$ as an activation function and $Y^{(L,P_k)}$ the same matrix constructed with P_k. Note that we consider the same sampling of W and X for the construction of this model. We describe the case of $L = 2$ as we can recursively do the same reasoning for a higher number of layers, for simplicity we also forget the change of variance $\sigma_x/\sqrt{\theta_1(f)}$ at each layer. As we saw in Section 2.4, we simply need to bound

$$\frac{1}{\sqrt{m}}\max_{1\leqslant i\leqslant n_2}\sum_{j=1}^{m}\left|Y_{ij}^{(2,a_k)}-Y_{ij}^{(2,P_k)}\right| = \frac{1}{\sqrt{m}}\max_{1\leqslant i\leqslant n_2}\sum_{j=1}^{m}\left|f\left(\frac{W^{(1)}Y^{(1,a_k)}}{\sqrt{n_1}}\right)_{ij} - a_k - P_k\left(\frac{W^{(1)}Y^{(1,P_k)}}{\sqrt{n_1}}\right)_{ij}\right|.$$

We split the right hand side into two parts and write

$$\begin{aligned}&\left|Y_{ij}^{(2,a_k)}-Y_{ij}^{(2,P_k)}\right|\\ &\leqslant\left|f\left(\frac{W^{(1)}Y^{(1,a_k)}}{\sqrt{n_1}}\right)_{ij} - f\left(\frac{W^{(1)}Y^{(1,P_k)}}{\sqrt{n_1}}\right)_{ij}\right| + \left|f\left(\frac{W^{(1)}Y^{(1,P_k)}}{\sqrt{n_1}}\right)_{ij} - a_k - P_n\left(\frac{W^{(1)}Y^{(1,P_k)}}{\sqrt{n_1}}\right)_{ij}\right|.\end{aligned} \tag{2.6.14}$$

For the first term on the right hand side of the previous equation, we bound it from the polynomial approximation. Indeed, we consider the following event

$$\mathcal{A}_1(\delta_1) = \bigcap_{i=1}^{n_1}\bigcap_{j=1}^{m}\left\{\left|\left(\frac{W^{(0)}X}{\sqrt{n_0}}\right)_{ij}\right| \leqslant (\log n_1)^{1/2+\delta_1}\right\}\bigcap\left\{\left|W_{ij}^{(1)}\right| \leqslant (\log n)^{1/\alpha+\delta_1}\right\}.$$

This event occurs with overwhelming probability for any $\delta_1 > 0$ in the sense that its probability decays faster than any polynomial. Now, on this event we can bound

$$\left|\left(\frac{W^{(1)}Y^{(1,a_k)}}{\sqrt{n_1}}\right)_{ij} - \left(\frac{W^{(1)}Y^{(1,P_n)}}{\sqrt{n_1}}\right)_{ij}\right| \leqslant C^n\sqrt{n_1}(\log n_1)^{1/\alpha+\delta_1}\frac{(\log n_1)^{(1/2+\delta_1)n}}{n!},$$

where we expand the entries and use the polynomial approximation. If $n = c_0\frac{\log n_1}{\log\log n_1}$ for some contant $c_0 > 1$ this also decays faster than any polynomial. Even though we can only consider $n \leqslant \frac{1}{L-1}\frac{\log n_1}{\log\log n_1}$, note that this constraint on n is not a problem by the considerations in the proof of Proposition 2.5.2. Finally, using the fact that f has a bounded derivative on the event $\mathcal{A}_2(\delta_2)$ defined in (2.6.15), the first term in (2.6.14) goes to zero providing that $\mathcal{A}_2$ occurs with high probability.

For the second term in (2.6.14), by the previous analysis and as in Section 2.4 we only need to prove that the following event occurs with probability tending to one:

$$\mathcal{A}_2(\delta_2) = \bigcap_{i=1}^{n_2}\bigcap_{j=1}^{m}\left\{\frac{1}{\sqrt{n_1}}\sum_{\ell_1=1}^{n_1}W_{i\ell_1}^{(1)}P_n\left(\frac{1}{\sqrt{n_0}}\sum_{\ell_0=1}^{n_0}W_{\ell_1\ell_0}^{(0)}X_{\ell_0 j}\right) \leqslant (\log n_1)^{1/2+\delta_1}\right\}. \tag{2.6.15}$$

Since we suppose that f is bounded we know that on the event $\mathcal{A}_1(\delta_1)$ (which occurs with very high probability) we have that $\sup_{ij}|Y_{ij}^{(1,P_k)}| \leqslant C$. Besides, since $W_{i\ell_1}^{(1)}$ has zero expectation, has a sub-Gaussian tail and is independent of the entries of $W^{(0)}$ and X, the random variable $(W^{(1)}Y^{(1)})_{ij}$ is sub-Gaussian as well. So that we obtain that there exists a $C > 0$ such that

$$\mathbb{P}\left(\sum_{\ell_1=1}^{n_1}W_{i\ell_1}^{(1)}P_n\left(\frac{1}{\sqrt{n_0}}\sum_{\ell_0=1}^{n_0}W_{\ell_1\ell_0}^{(0)}X_{\ell_0 j}\right) > \sqrt{n_1}(\log n_1)^{1/2+\delta_1}\right) \leqslant Ce^{-c(\log n_1)^{1+2\delta_1}}.$$

And finally $\mathbb{P}(\mathcal{A}_2(\delta_2)) \geqslant 1 - n_1^{-D}$ for any $D > 0$. □

Bibliography

[AC17] A. Adhikari and Z. Che, *The Edge Universality of Correlated Matrices*, arXiv preprint (2017).

[Agg18] A. Aggarwal, *Bulk universality for generalized Wigner matrices with few moments*, Probab Theory Related Fields (2018).

[ALY18] A. Aggarwal, P. Lopatto, and H.-T. Yau, *GOE Statistics for Levy Matrices*, arXiv preprint (2018).

[AEK15] O. H. Ajanki, L. Erdős, and T. Krüger, *Quadratic vector equations on complex upper half-plane*, arXiv preprint (2015).

[AEK16] ______, *Local spectral statistics of Gaussian matrices with correlated entries*, Journal of statistical physics **163** (2016), no. 2, 280–302.

[AEK17] ______, *Universality for general Wigner-type matrices*, Probab. Theory Related Fields **169** (2017), no. 3-4, 667–727.

[AEK18a] ______, *Stability of the matrix Dyson equation and random matrices with correlations*, Probab. Theory Related Fields (2018).

[AIK13] G. Akemann, J. R. Ipsen, and M. Kieburg, *Products of rectangular random matrices: singular values and progressive scattering*, Physical Review E **88** (2013), no. 5, 052118.

[AB12] R. Allez and J.-P. Bouchaud, *Eigenvector dynamics: general theory and some applications*, Physical Review E **86** (2012), no. 4, 046202.

[AB14] ______, *Eigenvector dynamics under free addition*, Random Matrices: Theory and Applications **3** (2014), no. 03.

[ABB14] R. Allez, J. Bun, and J.-P. Bouchaud, *The eigenvectors of Gaussian matrices with an external source*, arXiv preprint arXiv:1412.7108 (2014).

[AEK18b] J. Alt, L. Erdős, and T. Krüger, *The Dyson equation with linear self-energy: spectral bands, edges and cusps*, arXiv preprint (2018).

[AEKS18] J. Alt, L. Erdős, T. Krüger, and D. Schröder, *Correlated Random Matrices: Band Rigidity and Edge Universality*, arXiv preprint (2018).

[And58] P. Anderson, *Absences of diffusion in certain random lattices*, Phys. Rev. (1958), 1492-1505.

[ALM15] N. Anantharaman and E. Le Masson, *Quantum ergodicity on large regular graphs*, Duke Math. J. **164** (2015), no. 4, 723–765.

[AGZ10] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010.

[BS10] Z. D. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*, 2nd ed., Springer Series in Statistics, Springer, New York, 2010.

[BY93] Z. D. Bai and Y. Q. Yin, *Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix*, Ann. Probab. **21** (1993), no. 3, 1275–1294.

[BBAP05] J. Baik, G. Ben Arous, and S. Péché, *Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices*, Ann. Probab. **33** (2005), no. 5, 1643–1697.

[BS06] J. Baik and J. W. Silverstein, *Eigenvalues of large sample covariance matrices of spiked population models*, J. Multivariate Anal. **97** (2006), no. 6, 1382–1408.

[BES16] Z. Bao, L. Erdős, and K. Schnelli, *Local stability of the free additive convolution*, J. Funct. Anal. **271** (2016), no. 3, 672–719.

[BES17a] ______, *Convergence rate for spectral distribution of addition of random matrices*, Adv. Math. **319** (2017), 251–291.

[BES17b] ______, *Local law of addition of random matrices on optimal scale*, Comm. Math. Phys. **349** (2017), no. 3, 947–990.

[BPZ14] Z. Bao, G. Pan, and W. Zhou, *Universality for a global property of the eigenvectors of Wigner matrices*, J. Math. Phys. **55** (2014).

[BPZ15] ———, *Universality for the largest eigenvalue of sample covariance matrices with general population*, Ann. Statist. **43** (2015), no. 1, 382–421.

[BHKY17] R. Bauerschmidt, J. Huang, A. Knowles, and H.-T. Yau, *Bulk eigenvalue statistics for random regular graphs*, Ann. Probab. **45** (2017), no. 6A, 3626–3663.

[BHY19] R. Bauerschmidt, J. Huang, and H.-T. Yau, *Local Kesten–Mckay law for random regular graphs*, Comm. Math. Phys. (2019), to appear.

[BKY17] R. Bauerschmidt, A. Knowles, and H.-T. Yau, *Local semicircle law for random regular graphs*, Comm. Pure Appl. Math. **70** (2017), no. 10, 1898–1960.

[BFG15] F. Bekerman, A. Figalli, and A. Guionnet, *Transport maps for β-matrix models and universality*, Comm. Math. Phys. **338** (2015), no. 2, 589–619.

[BAG08] G. Ben Arous and A. Guionnet, *The spectrum of heavy tailed random matrices*, Comm. Math. Phys. **278** (2008), no. 3, 715–751.

[BAP05] G. Ben Arous and S. Péché, *Universality of local eigenvalue statistics for some sample covariance matrices*, Comm. Pure Appl. Math. **58** (2005), no. 10, 1316–1357.

[BG12] F. Benaych-Georges, *A universality result for the global fluctuations of the eigenvectors of Wigner matrices*, Random Matrices Theory Appl. **1** (2012), no. 04.

[BGC16] F. Benaych-Georges and R. Couillet, *Spectral analysis of the Gram matrix of mixture models*, ESAIM Probab. Stat. **20** (2016), 217–237.

[BGK16] F. Benaych-Georges and A. Knowles, *Lectures on the local semicircle law for Wigner matrices*, arXiv preprint (2016).

[BGN11] F. Benaych-Georges and R. R. Nadakuditi, *The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices*, Adv. Math. **227** (2011), no. 1, 494–521.

[Ben17] L. Benigni, *Eigenvectors distribution and quantum unique ergodicity for deformed Wigner matrices*, arXiv preprint (2017).

[Ben19] ———, *Fermionic observable for the eigenvector moment flow and fluctuations of eigenvectors of random matrices*, Prepublication (2019).

[BB19] L. Benigni and P. Bourgade, *Universality of eigenvalues and eigenvectors of random matrices*, Prepublication (2019).

[BP19] L. Benigni and S. Péché, *Eigenvalue distribution of nonlinear matrix models*, Prepublication (2019).

[Bia97] P. Biane, *On the free convolution with a semi-circular distribution*, Indiana Univ. Math. J. **46** (1997), no. 3, 705–718.

[BI99] P. Bleher and A. Its, *Semiclassical asymptotics of orthogonal polynomials, Riemann-Hilbert problem, and universality in the matrix model*, Ann. of Math. (2) **150** (1999), no. 1, 185–266.

[BEK$^+$14] A. Bloemendal, L. Erdős, A. Knowles, H.-T. Yau, and Jun Yin, *Isotropic local laws for sample covariance and generalized Wigner matrices*, Electron. J. Probab. **19** (2014), no. 33, 53.

[BGS84] O. Bohigas, M.-J. Giannoni, and C. Schmit, *Characterization of chaotic quantum spectra and universality of level fluctuation laws*, Phys. Rev. Lett. **52** (1984), no. 1, 1–4.

[BG13] C. Bordenave and A. Guionnet, *Localization and delocalization of eigenvectors for heavy-tailed random matrices*, Probability Theory and Related Fields **157** (2013), no. 3-4, 885–953.

[BG17] ———, *Delocalization at small energy for heavy-tailed random matrices*, Communications in Mathematical Physics **354** (2017), no. 1, 115–159.

[Bou18] P. Bourgade, *Random band matrices*, Proceedings ICM-2018 (2018).

[BEY12] P. Bourgade, L. Erdős, and H.-T. Yau, *Bulk universality of general β-ensembles with non-convex potential*, J. Math. Phys. **53** (2012), no. 9, 095221, 19.

[BEY14a] ———, *Edge universality of beta ensembles*, Comm. Math. Phys. **332** (2014), no. 1, 261–353.

[BEY14b] ———, *Universality of general β-ensembles*, Duke Math. J. **163** (2014), no. 6, 1127–1190.

[BEYY16] P. Bourgade, L. Erdős, H.-T. Yau, and J. Yin, *Fixed energy universality for generalized Wigner matrices*, Comm. Pure Appl. Math. **69** (2016), no. 10, 1815–1881.

[BEYY17] ———, *Universality for a class of random band matrices*, Adv. Theo. Math. Phys. **21** (2017), 739-800.

[BHY17] P. Bourgade, J. Huang, and H.-T. Yau, *Eigenvector statistics of sparse random matrices*, Electron. J. Probab. **22** (2017), Paper No. 64, 38.

[BY17] P. Bourgade and H.-T. Yau, *The eigenvector moment flow and local quantum unique ergodicity*, Comm. Math. Phys. **350** (2017), no. 1, 231–278.

[BYY18] P. Bourgade, H.-T. Yau, and J. Yin, *Random band matrices in the delocalized phase: Quantum uniqe ergodicity and universality*, In preparation (2018).

[BH96] E. Brézin and S. Hikami, *Correlations of nearby levels induced by a random potential*, Nuclear Physics B **479** (1996), no. 3, 697–706.

[BH97] ———, *Spectral form factor in a random matrix theory*, Physical Review E **55** (1997), no. 4, 4067.

[Bru89] M.-F. Bru, *Diffusions of perturbed principal component analysis*, J. Multivariate Anal. **29** (1989), no. 1, 127–136.

[BABP16] J. Bun, R. Allez, J.-P. Bouchaud, and M. Potters, *Rotational Invariant Estimator for General Noisy Matrices*, IEEE Trans. Information Theory **62** (2016), no. 12, 7475–7490.

[BBP18] J. Bun, J.-P. Bouchaud, and M. Potters, *Overlaps between eigenvectors of correlated random matrices*, Physical Review E **98** (2018), no. 5.

[CDMF09] M. Capitaine, C. Donati-Martin, and D. Féral, *The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations*, Ann. Probab. **37** (2009), no. 1, 1–47.

[CDM16] G. Cébron, A. Dahlqvist, and C. Male, *Universal constructions for spaces of traffics*, arXiv preprint (2016).

[Che17] Z. Che, *Universality of random matrices with correlated entries*, Electro. J. of Probab. **22** (2017).

[CL17] Z. Che and B. Landon, *Local spectral statistics of the addition of random matrices*, arXiv preprint (2017).

[CYY08] W. Y. C. Chen, S. H. F. Yan, and L. L. M. Yang, *Identities from weighted Motzkin paths*, Adv. in Appl. Math. **41** (2008), no. 3, 329–334.

[CHM+15] A. Choromanska, M. Henaff, M. Mathieu, G. Ben Arous, and Y. LeCun, *The loss surfaces of multilayer networks*, Proceedings of the 18th International Conference on Artificial Intelligence and Statistics, AISTATS 2015 (2015).

[CEKS18] G. Cipolloni, L. Erdős, T. Krüger, and D. Schröder, *Cusp Universality for Random Matrices II: the real symmetric case*, arXiv preprint (2018).

[CB94] P. Cizeau and J. P. Bouchaud, *Theory of Lévy matrices*, Phys. Rev. E **50** (1994), no. 3, 1810–1822.

[CdV85] Y. Colin de Verdière, *Ergodicité et fonctions propres du laplacien*, Comm. Math. Phys. **102** (1985), no. 3, 497–502.

[CBG16] R. Couillet and F. Benaych-Georges, *Kernel spectral clustering of large dimensional data*, Electron. J. Stat. **10** (2016), no. 1, 1393–1454.

[DG07a] P. Deift and D. Gioev, *Universality in random matrix theory for orthogonal and symplectic ensembles*, Int. Math. Res. Pap. IMRP **2** (2007).

[DG07b] ———, *Universality at the edge of the spectrum for unitary, orthogonal, and symplectic ensembles of random matrices*, Comm. Pure Appl. Math. **60** (2007), no. 6, 867–910.

[DKM+99] P. Deift, T. Kriecherbauer, K. T.-R. McLaughlin, S. Venakides, and X. Zhou, *Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory*, Comm. Pure Appl. Math. **52** (1999), no. 11, 1335–1425.

[DMR12] C. Donati-Martin and A. Rouault, *Truncations of Haar distributed matrices, traces and bivariate Brownian bridges*, Random Matrices Theory Appl. **1** (2012), no. 1.

[DE02] I. Dumitriu and A. Edelman, *Matrix models for beta ensembles*, J. Math. Phys. **43** (2002), no. 11, 5830–5847.

[DC14] T. Dupic and I. P. Castillo, *Spectral density of products of Wishart dilute random matrices. Part I: the dense case*, arXiv preprint (2014).

[Dys62a] F. J. Dyson, *A Brownian-motion model for the eigenvalues of a random matrix*, J. Mathematical Phys. **3** (1962), 1191–1198.

[Dys62b] ———, *Statistical theory of the energy levels of complex systems. I, II, III*, J. Mathematical Phys. **3** (1962), 140–175.

[Dys70] ———, *Correlations between eigenvalues of a random matrix*, Comm. Math. Phys. **19** (1970), 235–250.

[Ede88] A. Edelman, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. Appl. **9** (1988), no. 4, 543–560.

[EK07] N. El Karoui, *Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices*, Ann. Probab. **35** (2007), no. 2, 663–714.

[EKYY12] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *Spectral statistics of Erdős-Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues*, Comm. Math. Phys. **314** (2012), no. 3, 587–640.

[EKYY13] ———, *Spectral statistics of Erdős-Rényi graphs I: Local semicircle law*, Ann. Probab. **41** (2013), no. 3B, 2279–2375.

[EKS17] L. Erdős, T. Krüger, and D. Schröder, *Random matrices with slow correlation decay*, arXiv preprint (2017).

[EKS18] ———, *Cusp Universality for Random Matrices I: Local Law and the Complex Hermitian Case*, arXiv preprint (2018).

[EPR+10] L. Erdős, S. Péché, J. A. Ramírez, B. Schlein, and H.-T. Yau, *Bulk universality for Wigner matrices*, Comm. Pure Appl. Math. **63** (2010), no. 7, 895–925.

[ERSY10] L. Erdős, J. A. Ramírez, B. Schlein, and H.-T. Yau, *Universality of sine-kernel for Wigner matrices with a small Gaussian perturbation*, Electron. J. Probab. **15** (2010), no. 18, 526–603.

[ERS+10] L. Erdős, J. A. Ramírez, B. Schlein, T. Tao, V. Vu, and H.-T. Yau, *Bulk universality for Wigner Hermitian matrices with subexponential decay*, Math. Res. Lett. **17** (2010), no. 4, 667–674.

[ESY09a] L. Erdős, B. Schlein, and H.-T. Yau, *Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices*, Ann. Probab. **37** (2009), no. 3, 815–852.

[ESY09b] ———, *Local semicircle law and complete delocalization for Wigner random matrices*, Comm. Math. Phys. **287** (2009), no. 2, 641–655.

[ESY10] ———, *Wegner estimate and level repulsion for Wigner random matrices*, Int. Math. Res. Not. IMRN **3** (2010), 436–479.

[ESY11] ———, *Universality of random matrices and local relaxation flow*, Invent. Math. **185** (2011), no. 1, 75–119.

[ESYY12] L. Erdős, B. Schlein, H.-T. Yau, and J. Yin, *The local relaxation flow approach to universality of the local statistics for random matrices*, Ann. Inst. Henri Poincaré Probab. Stat. **48** (2012), no. 1, 1–46 (English, with English and French summaries).

[ES17] L. Erdős and K. Schnelli, *Universality for random matrix flows with time-dependent density*, Ann. Inst. Henri Poincaré Probab. Stat. **53** (2017), no. 4, 1606–1656 (English, with English and French summaries).

[EY15] L. Erdős and H.-T. Yau, *Gap universality of generalized Wigner and β-ensembles*, J. Eur. Math. Soc. (JEMS) **17** (2015), no. 8, 1927–2036.

[EY17] ———, *A dynamical approach to random matrix theory*, Courant Lecture Notes in Mathematics, vol. 28, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2017.

[EYY11] L. Erdős, H.-T. Yau, and J. Yin, *Universality for generalized Wigner matrices with Bernoulli distribution*, J. Comb. **2** (2011), no. 1, 15–81.

[EYY12a] ———, *Bulk universality for generalized Wigner matrices*, Probab. Theory Related Fields **154** (2012), no. 1-2, 341–407.

[EYY12b] ———, *Rigidity of eigenvalues of generalized Wigner matrices*, Adv. Math. **229** (2012), no. 3, 1435–1515.

[FVB16] D. Facoetti, P. Vivo, and G. Biroli, *From non-ergodic eigenvectors to local resolvent statistics and back: A random matrix perspective*, EPL (Europhysics Letters) **115** (2016), no. 4, 47003.

[FS10] O. N. Feldheim and S. Sodin, *A universality result for the smallest eigenvalues of certain sample covariance matrices*, Geom. Funct. Anal. **20** (2010), no. 1, 88–123.

[FP07] D. Féral and S. Péché, *The largest eigenvalue of rank one deformation of large Wigner matrices*, Comm. Math. Phys. **272** (2007), no. 1, 185–228.

[FG16] A. Figalli and A. Guionnet, *Universality in several-matrix models via approximate transport maps*, Acta Math. **217** (2016), no. 1, 81–176.

[FU56] G. W. Ford and G. E. Uhlenbeck, *Combinatorial problems in the theory of graphs. III*, Proc. Nat. Acad. Sci. U.S.A. **42** (1956), 529–535.

[For93] P. J. Forrester, *The spectrum edge of random matrix ensembles*, Nuclear Phys. B **402** (1993), no. 3, 709–728.

[FK81] Z. Füredi and J. Komlós, *The eigenvalues of random symmetric matrices*, Combinatorica **1** (1981), no. 3, 233–241.

[Gem80] S. Geman, *A limit theorem for the norm of random matrices*, Ann. Probab. **8** (1980), no. 2, 252–261.

[HHN16] W. Hachem, A. Hardy, and J. Najim, *Large complex correlated Wishart matrices: fluctuations and asymptotic independence at the edges*, Ann. Probab. **44** (2016), no. 3, 2264–2348.

[HKR17] Y. He, A. Knowles, and R. Rosenthal, *Isotropic self-consistent equations for mean-field random matrices*, Probab. Theory and Related Fields (2017), 1–47.

[HDY+12] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and Others, *Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups*, IEEE Signal processing magazine **29** (2012), no. 6, 82–97.

[Hol10] R. Holowinsky, *Sieving for mass equidistribution*, Ann. of Math. (2) **172** (2010), no. 2, 1499–1516.

[HS10] R. Holowinsky and K. Soundararajan, *Mass equidistribution for Hecke eigenforms*, Ann. of Math. (2) **172** (2010), no. 2, 1517–1528.

[HLY15] J. Huang, B. Landon, and H.-T. Yau, *Bulk universality of sparse random matrices*, J. Math. Phys. **56** (2015), no. 12, 123301, 19.

[HLY17] ———, *Transition from Tracy-Widom to Gaussian fluctuations of extremal eigenvalues of sparse Erdős-Rényi graphs*, arXiv preprint (2017).

[Joh00] K. Johansson, *Shape fluctuations and random matrices*, Comm. Math. Phys. **209** (2000), no. 2, 437–476.

[Joh01a] ———, *Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices*, Comm. Math. Phys. **215** (2001), no. 3, 683–705.

[Joh01b] I. M. Johnstone, *On the distribution of the largest eigenvalue in principal components analysis*, Ann. Statist. **29** (2001), no. 2, 295–327.

[Kar12] V. Kargin, *A concentration inequality and a local law for the sum of two random matrices*, Probab. Theory Related Fields **154** (2012), no. 3-4, 677–702.

[Kar15] ———, *Subordination for the sum of two random matrices*, Ann. Probab. **43** (2015), no. 4, 2119–2150.

[KKP96] A. M. Khorunzhy, B. A. Khoruzhenko, and L. A. Pastur, *Asymptotic properties of large random matrices with independent entries*, J. Math. Phys. **37** (1996), no. 10, 5033–5060.

[KY13a] A. Knowles and J. Yin, *The isotropic semicircle law and deformation of Wigner matrices*, Comm. Pure Appl. Math. **66** (2013), no. 11, 1663–1750.

[KY13b] ———, *Eigenvector distribution of Wigner matrices*, Probab. Theory Related Fields **155** (2013), no. 3-4, 543–582.

[KY17] ———, *Anisotropic local laws for random matrices*, Probab. Theory Related Fields **169** (2017), no. 1-2, 257–352.

[KAI17] V. E. Kravtsov, B. L. Altshuler, and L. B. Ioffe, *Non-ergodic delocalized phase in Anderson model on Bethe lattice and regular graph*, Ann. Phys. **389** (2017).

[KRV16] M. Krishnapur, B. Rider, and B. Virág, *Universality of the stochastic Airy operator*, Comm. Pure Appl. Math. **69** (2016), no. 1, 145–199.

[KSH12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *ImageNet classification with deep convolutional neural networks*, Advances in Neural Information Processing Systems **25** (2012), 1097–1105.

[LSY16] B. Landon, P. Sosoe, and H.-T. Yau, *Fixed energy universality for Dyson Brownian motion*, arXiv preprint (2016).

[LY17a] B. Landon and H.-T. Yau, *Convergence of local statistics of Dyson Brownian motion*, Comm. Math. Phys. **355** (2017), no. 3, 949–1000.

[LY17b] ———, *Edge statistics of Dyson Brownian motion*, arXiv preprint (2017).

[LS13] J. O. Lee and K. Schnelli, *Local deformed semicircle law and complete delocalization for Wigner matrices with random potential*, J. Math. Phys. **54** (2013), no. 10, 103504, 62.

[LS15] ———, *Edge universality for deformed Wigner matrices*, Rev. Math. Phys. **27** (2015), no. 8, 1550018, 94.

[LS16] ———, *Tracy-Widom distribution for the largest eigenvalue of real sample covariance matrices with general population*, Ann. Appl. Probab. **26** (2016), no. 6, 3786–3839.

[LS17] ———, *Local law and Tracy–Widom limit for sparse random matrices*, Probab. Theory Related Fields (2017), 1–74.

[LSSY16] J. O. Lee, K. Schnelli, B. Stetler, and H.-T. Yau, *Bulk universality for deformed Wigner matrices*, Ann. Probab. **44** (2016), no. 3, 2349–2425.

[LY14] J. O. Lee and J. Yin, *A necessary and sufficient condition for edge universality of Wigner matrices*, Duke Math. J. **163** (2014), no. 1, 117–173.

[Lin06] E. Lindenstrauss, *Invariant measures and arithmetic quantum unique ergodicity*, Ann. of Math. (2) **163** (2006), no. 1, 165–219.

[LLC18] C. Louart, Z. Liao, and R. Couillet, *A random matrix approach to neural networks*, Ann. Appl. Probab. **28** (2018), no. 2, 1190–1248.

[LP09] A. Lytova and L. A. Pastur, *Central limit theorem for linear eigenvalue statistics of random matrices with independent entries*, Ann. Probab. **37** (2009), no. 5, 1778–1840.

[MP67] V. A. Marčenko and L. A. Pastur, *Distribution of eigenvalues in certain sets of random matrices*, Mat. Sb. (N.S.) **72 (114)** (1967), 507–536 (Russian).

[Meh60] M. L. Mehta, *On the statistical properties of the level-spacings in nuclear spectra*, Nuclear Phys. **18** (1960), 395–419.

[Meh67] ———, *Random matrices and the statistical theory of energy levels*, Academic Press, New York-London, 1967.

[Meh04] ———, *Random matrices*, 3rd ed., Pure and Applied Mathematics (Amsterdam), vol. 142, Elsevier/Academic Press, Amsterdam, 2004.

[Meh71] ———, *A note on correlations between eigenvalues of a random matrix*, Comm. Math. Phys. **20** (1971), 245–250.

[MG60] M. L. Mehta and M. Gaudin, *On the density of eigenvalues of a random matrix*, Nuclear Phys. **18** (1960), 420–427.

[Nad08] B. Nadler, *Finite sample approximation results for principal component analysis: a matrix perturbation approach*, Ann. Statist. **36** (2008), no. 6, 2791–2817.

[NW91] T. Nagao and M. Wadati, *Correlation functions of random matrix ensembles related to classical orthogonal polynomials*, J. Phys. Soc. Japan **60** (1991), no. 10, 3298–3322.

[NW92a] ———, *Correlation functions of random matrix ensembles related to classical orthogonal polynomials. II*, J. Phys. Soc. Japan **61** (1992), no. 1, 78–88.

[NW92b] ———, *Correlation functions of random matrix ensembles related to classical orthogonal polynomials. III*, J. Phys. Soc. Japan **61** (1992), no. 6, 1910–1918.

[NRW86] J. R. Norris, L. C. G. Rogers, and D. Williams, *Brownian motions of ellipsoids*, Trans. Amer. Math. Soc. **294** (1986), no. 2, 757–765.

[Ona08] A. Onatski, *The Tracy-Widom limit for the largest eigenvalues of singular complex Wishart matrices*, Ann. Appl. Probab. **18** (2008), no. 2, 470–490.

[TO16] K. Truong and A. Ossipov, *Eigenvectors under a generic perturbation: Non-perturbative results from the random matrix approach*, EPL (Europhysics Letters) **116** (2016), no. 3, 37002.

[PS97] L. Pastur and M. Shcherbina, *Universality of the local eigenvalue statistics for a class of unitary invariant random matrix ensembles*, J. Statist. Phys. **86** (1997), no. 1-2, 109–147.

[PS03] ———, *On the edge universality of the local eigenvalue statistics of matrix models*, Mat. Fiz. Anal. Geom. **10** (2003), no. 3, 335–365.

[PS08] ———, *Bulk universality and related properties of Hermitian matrix models*, J. Stat. Phys. **130** (2008), no. 2, 205–250.

[Pau07] D. Paul, *Asymptotics of sample eigenstructure for a large dimensional spiked covariance model*, Statist. Sinica **17** (2007), no. 4, 1617–1642.

[Péc06] S. Péché, *The largest eigenvalue of small rank perturbations of Hermitian random matrices*, Probab. Theory Related Fields **134** (2006), no. 1, 127–173.

[Péc09] ———, *Universality results for the largest eigenvalues of some sample covariance matrix ensembles*, Probab. Theory Related Fields **143** (2009), no. 3-4, 481–516.

[Péc12] ______, *Universality in the bulk of the spectrum for complex sample covariance matrices*, Ann. Inst. Henri Poincaré Probab. Stat. **48** (2012), no. 1, 80–106.

[PSSS19] R. Peled, J. Schenker, M. Shamis, and S. Sodin, *On the Wegner Orbital Model*, Int. Math. Res. Not. IMRN **4** (2019), 1030–1058.

[PW17] J. Pennington and P. Worah, *Nonlinear random matrix theory for deep learning*, Advances in Neural Information Processing Systems, 2017, pp. 2637–2646.

[PY14] N. S. Pillai and J. Yin, *Universality of covariance matrices*, Ann. Appl. Probab. **24** (2014), no. 3, 935–1001.

[RP60] N. Rosenzweig and C. E. Porter, *"Repulsion of Energy Levels" in Complex Atomic Spectra*, Phys. Rev. **120** (1960), no. 5, 1698–1714.

[RV15] M. Rudelson and R. Vershynin, *Delocalization of eigenvectors of random matrices with independent entries*, Duke Math. J. **164** (2015), no. 13, 2507–2538.

[RV16] ______, *No-gaps delocalization for general random matrices*, Geom. Funct. Anal. **26** (2016), no. 6, 1716–1776.

[RS94] Z. Rudnick and P. Sarnak, *The behaviour of eigenstates of arithmetic hyperbolic manifolds*, Comm. Math. Phys. **161** (1994), no. 1, 195–213.

[Sch09] J. Schenker, *Eigenvector localization for random band matrices with power law band width*, Comm. Math. Phys. **290** (2009), no. 3, 1065–1097.

[Shc14] M. Shcherbina, *Change of variables as a method to study general β-models: bulk universality*, J. Math. Phys. **55** (2014), no. 4, 043504, 23.

[Sil85] J. W. Silverstein, *The smallest eigenvalue of a large-dimensional Wishart matrix*, Ann. Probab. **13** (1985), no. 4, 1364–1368.

[Sil90] ______, *Weak convergence of random functions defined by the eigenvectors of sample covariance matrices*, The Annals of Probability (1990), 1174–1194.

[SB95] J. W. Silverstein and Z. D. Bai, *On the empirical distribution of eigenvalues of a class of large-dimensional random matrices*, J. Multivariate Anal. **54** (1995), no. 2, 175–192.

[Šni74] A. I. Šnirel'man, *Ergodic properties of eigenfunctions*, Uspehi Mat. Nauk **29** (1974), no. 6(180), 181–182.

[Sod10] S. Sodin, *The spectral edge of some random band matrices*, Ann. of Math. (2) **172** (2010), no. 3, 2223–2251.

[Sos99] A. Soshnikov, *Universality at the edge of the spectrum in Wigner random matrices*, Comm. Math. Phys. **207** (1999), no. 3, 697–733.

[Sos02] ______, *A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices*, J. Statist. Phys. **108** (2002), no. 5-6, 1033–1056. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays.

[TV10] T. Tao and V. Vu, *Random matrices: the distribution of the smallest singular values*, Geom. Funct. Anal. **20** (2010), no. 1, 260–297.

[TV11] ______, *Random matrices: universality of local eigenvalue statistics*, Acta Math. **206** (2011), no. 1, 127–204.

[TV12a] ______, *Random covariance matrices: universality of local statistics of eigenvalues*, Ann. Probab. **40** (2012), no. 3, 1285–1315.

[TV12b] ______, *Random matrices: universal properties of eigenvectors*, Random Matrices Theory Appl. **1** (2012), no. 1, 1150001, 27.

[TW94a] C. A. Tracy and H. Widom, *Level-spacing distributions and the Airy kernel*, Comm. Math. Phys. **159** (1994), no. 1, 151–174.

[TW94b] ______, *Level spacing distributions and the Bessel kernel*, Comm. Math. Phys. **161** (1994), no. 2.

[TW96] ______, *On orthogonal and symplectic matrix ensembles*, Comm. Math. Phys. **177** (1996), no. 3, 727–754.

[Voi91] D. Voiculescu, *Limit laws for random matrices and free products*, Invent. Math. **104** (1991), no. 1, 201–220.

[vSW18a] P. von Soosten and S. Warzel, *The phase transition in the ultrametric ensemble and local stability of Dyson Brownian motion*, Electron. J. Probab. **23** (2018), 24 pp.

[vSW18b] ______, *Non-ergodic delocalization in the Rosenzweig–Porter model*, Letters in Mathematical Physics (2018).

[VW15] V. Vu and K. Wang, *Random weighted projections, random quadratic forms and random eigenvectors*, Random Structures Algorithms **47** (2015), no. 4, 792–821.

[Wac78] K. W. Wachter, *The strong limits of random matrix spectra for sample matrices of independent elements*, Ann. Probability **6** (1978), no. 1, 1–18.

[Wan12] K. Wang, *Random covariance matrices: universality of local statistics of eigenvalues up to the edge*, Random Matrices Theory Appl. **1** (2012), no. 1, 1150005, 24.

[Wis28] J. Wishart, *The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population*, Biometrika **20A** (1928), no. 1/2, 32–52.

[Wig55] E. P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. of Math. (2) **62** (1955), 548–564.

[Wig58] ———, *On the distribution of the roots of certain symmetric matrices*, Ann. of Math. (2) **67** (1958), 325–327.

[WW95] M. Wilkinson and P. N. Walker, *A Brownian motion model for the parameter dependence of matrix elements*, Journal of Physics A: Mathematical and General **28** (1995), no. 21, 6143–6160.

[WSC+16] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, and Others, *Google's neural machine translation system: Bridging the gap between human and machine translation*, arXiv preprint (2016).

[Yin86] Y. Q. Yin, *Limiting spectral distribution for a class of random matrices*, J. Multivariate Anal. **20** (1986), no. 1, 50–68.

[YBK88] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah, *On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix*, Probab. Theory Related Fields **78** (1988), no. 4, 509–521.

[Zel87] S. Zelditch, *Uniform distribution of eigenfunctions on compact hyperbolic surfaces*, Duke Math. J. **55** (1987), no. 4, 919–941.

[ZJ89] J. Zinn-Justin, *Quantum field theory and critical phenomena*, International Series of Monographs on Physics, vol. 77, The Clarendon Press, Oxford University Press, New York, 1989. Oxford Science Publications.