On the moment method for inhomogeneous matrices : From regular graphs to quantum channels

Patrick Oliveira Santos

Université Gustave Eiffel

Thèse de doctorat de l'Université Gustave Eiffel

École Doctorale n°532, Mathématiques et STIC (MSTIC)

Spécialité de doctorat: Mathématiques

Unité de recherche : Laboratoire d'analyse et de mathématiques appliquées (LAMA)

Thèse présentée et soutenue à l'Université Gustave Eiffel, le 27/06/2024, par :

Patrick OLIVEIRA SANTOS

On the moment method for inhomogeneous matrices: From regular graphs to quantum channels

Composition du Jury

Olivier GUÉDON Université Gustave Eiffel	Co-directeur de thèse
Pierre YOUSSEF New York Abu Dhabi University	Co-directeur de thèse
Cécilia LANCIEN Université Grenoble Alpes	Examinatrice
Florence MERLEVÈDE Université Gustave Eiffel	Examinatrice
Nathanael ENRIQUEZ Université Paris-Saclay	Examinateur
Mylene MAIDA Université de Lille	Examinatrice
Benoît COLLINS Kyoto University	Rapporteur
Mireille CAPITAINE Université de Toulouse	Rapportrice

Encadrement de la thèse

Olivier GUÉDON Université Gustave Eiffel	Co-directeur de thèse
Pierre YOUSSEF New York Abu Dhabi University	Co-directeur de thèse

Agradecimentos

The content of this part is written in Portuguese, my native language.

Gostaria de começar agradecendo a minha família. Aos meus pais Ulisses Souza dos Santos e Janaina dos Santos Oliveira Santos. Vocês são fonte de amor a todo instante, e sempre serei eternamente grato a isso. Estendo esses agradecimentos aos meus avós Ulisses Barnabé dos Santos e Neyde Souza dos Santos que me acompanharam durante minha trajetória no Rio de Janeiro. Obrigado por sempre estarem dispostos a me ajudar e pelo imensurável carinho. Quero também agradecer aos meus primos Danielle Menezes e Milleny Menezes por sempre estarem presentes. Por fim, como sempre, Bruna, maninha... ainda não é sua hora.

Para além da minha família, gostaria de agradecer aos amigos que o Rio de Janeiro me deu. Conversar com vocês sempre foi motivo de muita felicidade. Cito, em especial, a Beatriz Fraga e a Milenna Mesquita, que presente é ter a amizade de vocês! Amo vocês! Também cito a Nathanaely Abner Lima de Araujo. Obrigado pelas infinitas trocas e pela maneira carinhosa de como nos falamos. Nosso constante contato é um dos meus pequenos prazeres!

Aqui, também quero agradecer aos amigos que atravessaram a flecha do tempo e estão comigo desde muito tempo. A liga... Da Justica! Matheus Ribeiro e Julio Rama. Palavras não descreveriam o que vocês são pra mim e tudo que eu queria agradecer. Amo vocês e espero que nossa amizade continue até o fim, até porque não há como continuar depois do fim e só acabará no fim mesmo...

Mas também não posso deixar de citar meus amigos da UFRJ, aqueles que fizeram o Bob. Menciono, com muito carinho, os nomes Iago Leal, Pedro Aragão, Alexandre Moreira, Gabriela Lewenfus e Ivani Ivanova. Obrigado por terem me acompanho e pelas inúmeras conversas ao longo desses anos, seja sobre matemática, música, trilhas e várias outras histórias hilárias que compartilhamos. Obrigado por tudo!

Para muito além do Rio de Janeiro, quero agradecer aos seres que comigo estão desde Boa Vista. Vocês são parte da minha infância, mas muito mais do que isso, são pessoas que me fizeram ser quem sou. Aos companheiros de jogos: Bruno Zardo e Gabriel Fin. Que haja muito jogos entre a gente ainda, passando ou não raiva. Entretanto, tenho que enfatizar um nome: Arthur Barbosa. Você será eternamente meu melhor amigo, fonte de motivação. Já está na hora de você vir me visitar aqui na França, não acha?

Agradeço aos amigos e colegas de departamento, aqueles que me suportaram diante de infinitas piadas (ruins?). Cito, em especial, Nikiforos Biehler e Gayane Taturyan.

Obrigado pelos risos sinceros às minhas piadas!

Quero agradecer também as instituições que investiram em mim e na minha jornada. Anteriormente a universidade, cito: Instituto Batista de Roraima, Acadêmico News, Sistema Elite de Ensino. Em nível acadêmico, meu agradecimento é às UFRJ e Universidade Gustave Eiffel, que proporcionou muitos conhecimentos e experiências. Em parapelo, agradeço a CAPES, FAPERJ, Bézout Labex e escola de doutorado MSTIC por investirem em mim por meio de bolsas de estudo.

Além disso, agradeço também aos meus professores que marcaram minha jornada e me proporcionaram inúmeros conhecimentos. Aqui menciono em especial meu ex-orientador César Niche.

Agradeço aos meus orientadores Olivier Guédon e Pierre Youssef por sempre estarem dispostos a me ajudar e corrigir meus inúmeros erros. Esses últimos três anos (e até mais) têm sido uma grande fonte de experiência, sabedoria e conhecimento em geral. Agradeço todo o esforço que vocês dedicaram a mim.

Agradeço também a banca Cécilia Lancien, Florence Merlevède, Mylene Maida e Nathanael Enriquez. Um agradecimento especial aos relatores Benoît Collins e Mireille Capitaine por disporem do seu tempo à minha tese. Obrigado a todos pela leitura e valiosa contribuição ao manuscrito e a tese.

Por fim, cito minha irmã Bruna Oliveira Santos. Não cabe em palavras o quão importante você é para mim, embora nessas poucas linhas eu seja obrigado a tentar. Sempre lembro e lembrarei o fato de nossas personalidades sempre Yin e Yang, o que quer dizer que você precisa me tolerar diversas vezes, além de sermos perfeitos juntos. Mana, você é o ser humano mais incrível que conheço e não consigo imaginar minha vida sem você. Amo você.

Acknowledgments

I want to start by thanking my family. My parents Ulisses Souza dos Santos and Janaina dos Santos Oliveira Santos. You are a source of love at all times, and I will be forever grateful for that. I extend this thanks to my grandparents, Ulisses Barnabé and Neyde Souza, who followed my journey in Rio de Janeiro. Thank you for being willing to help me and for your immeasurable affection. I also thank my cousins Danielle Menezes and Milleny Menezes for being around. Last, as always, Bruna, my little sister... it is not your time yet.

In addition to my family, I would like to thank the friends that Rio de Janeiro has given me. Talking to you has always been a source of great happiness. I would especially like to mention Beatriz Fraga and Milenna Mesquita. What a gift it is to have your friendship! I love you! I'd also like to mention Nathanaely Abner Lima de Araujo. Thank you for the endless exchanges and for the affectionate way in which we speak to each other. Our constant contact is one of my little pleasures!

Here, I also want to thank the friends who have been with me for a long time. The Justice League! Matheus Ribeiro and Julio Rama. Words wouldn't describe what you are for me and everything I wanted to thank. I love you, and I hope our friendship continues to the end, even because there is no way to continue after the future, and it will only end at the end...

But I also have to mention my friends at UFRJ, those who made Bob. I mention, with great affection, the names Iago Leal, Pedro Aragão, Alexandre Moreira, Gabriela Lewenfus, and Ivani Ivanova. Thank you for following me and for the countless conversations over the years about math, music, hiking, and several other hilarious stories we share. Thanks for all!

I would like to thank my friends and colleagues in the department, those who put up with my endless (bad?) jokes. I would especially like to mention Nikiforos Biehler and Gayane Taturyan. Thanks for the hearty laughs at my jokes!

I also want to thank the institutions that invested in me and my journey. Previously, to university, I quoted Instituto Batista de Roraima, Acadêmico News, and Sistema Elite de Ensino. On an academic level, my sincere thanks to UFRJ and Gustave Eiffel University, where I learned a lot and gained experience. Moreover, I want to thank CAPES, FAPERJ, Bézout Labex, and Doctorate School MSTIC for investing in me through scholarships.

I would also like to thank my teachers, who have marked my journey and given me a wealth of knowledge. I would especially like to mention my former advisor César Niche.

My sincere thanks to the committee professors Matthieu Fradelizi and Cyril Nicaud. Thank you for taking the time to read and comment on this memoir.

I thank my advisors, Olivier Guédon and Pierre Youssef, for always being willing to help me and correct my numerous mistakes. These past 3 years (and even more) have been a great source of experience, wisdom, and knowledge overall. I appreciate all the effort you have put into me.

I would also like to thank Cécilia Lancien, Florence Merlevède, Mylene Maida, and Nathanael Enriquez. Special thanks to the rapporteurs, Benoît Collins and Mireille Capitaine, for giving their time to this thesis. Thank you all for reading and making a valuable contribution to the manuscript and the thesis.

Finally, I quote and thank my sister Bruna Oliveira Santos. How important you are to me cannot be put into words, though in those few lines, I'm forced to try. I always remember and will remember the fact that our yin and yang personalities mean you have to put up with me several times, in addition to being perfect together. Sis, you are the most amazing human being I know, and I can't imagine my life without you. I love you.

This work is dedicated to my sister.

Abstract

This thesis studies the spectral statistics of inhomogeneous random matrices, such as weighted random graphs, covariance matrices, and quantum channels. One of the main proof techniques is adapting the moment method to these models. Our objectives are twofold. First, we investigate the limiting spectral distribution of regular directed graphs, quantum channels, and tensor products of non-commutative random variables. Second, we establish precise asymptotic and non-asymptotic bounds on the norm of such matrices and their quadratic forms. To achieve our objectives, in the first part, we show the convergence of large directed d-regular graphs G_n in n vertices, analyze the combinatorics of its moments, and explore the connection between random uniformly chosen regular digraphs, the infinite regular directed tree, and the oriented Kesten-McKay conjecture. We also work on its quantum counterpart, which is known as quantum channels. We will derive a free central limit theorem with the semi-circle law as the limit. Additionally, we extend the notion of quantum channels to non-commutative probability spaces and algebras and prove a central limit theorem for these variables. We show that the limit is the semi-circle law if and only if the variables are centered; otherwise, the limit can be written as a free convolution of the semi-circle law and an explicit probability measure. In the second part, we examine regular weighted graphs whose adjacency matrices X_n are formed by taking the Hadamard product of the adjacency matrix of the graph and a weighted matrix. We prove that when the weights are subgaussian random variables, the norm of the inhomogeneous random matrix X_n shows a sharp transition around $d \sim \log n$, indicating the presence of outliers. Additionally, we investigate the centered quadratic form $X_n X_n^t - \mathbb{E}[X_n X_n^t]$ and provide precise upper bounds on its norm, which is known as the covariance estimation problem. We present examples that improve upon previous works and also lower bounds.

Résumé

Cette thèse étudie les statistiques spectrales de matrices aléatoires inhomogènes, telles que les graphes aléatoires pondérés, les matrices de covariance et les canaux quantiques. L'une des principales techniques de preuve est l'adaptation de la méthode des moments à ces modèles. Nos objectifs sont doubles. Premièrement, nous étudions la distribution spectrale limite des graphes dirigés réguliers, des canaux quantiques et des produits tensoriels de variables aléatoires non commutatives. Deuxièmement, nous établissons des limites asymptotiques et non asymptotiques précises sur la norme de ces matrices et de leurs formes quadratiques. Pour atteindre nos objectifs, dans la première partie, nous montrons la convergence des grands graphes aléatoires dirigés d-réguliers G_n en n sommets, analysons la combinatoire de leurs moments, et explorons la connexion entre les digraphes réguliers aléatoires uniformément choisis, l'arbre régulier dirigé infini, et la conjecture orientée de Kesten-McKay. Nous travaillons également sur sa contrepartie quantique, connue

sous le nom de canaux quantiques. Nous déduisons un théorème central limite libre avec la loi du demi-cercle comme la limite. En outre, nous étendons la notion de canaux quantiques aux espaces et algèbres de probabilité non commutatifs et prouvons un théorème de limite centrale pour ces variables. Nous montrons que la limite est la loi du demi-cercle si et seulement si les variables sont centrées ; sinon, la limite peut être écrite comme une convolution libre de la loi du demi-cercle et des lois gaussiennes. Dans la deuxième partie, nous examinons les graphes réguliers pondérés dont les matrices d'adjacence X_n sont formées en prenant le produit d'Hadamard de la matrice d'adjacence du graphe et d'une matrice pondérée. Nous prouvons que lorsque les poids sont des variables aléatoires sousgaussiennes, la norme de la matrice aléatoire inhomogène X_n présente une transition abrupte autour de $d \sim \log n$, indiquant la présence de valeurs aberrantes. En outre, nous étudions la forme quadratique centrée $X_n X_n^t - \mathbb{E}[X_n X_n^t]$ et fournissons des estimations précises de sa norme, ce qui est connu comme le problème d'estimation de la covariance. Nous présentons des exemples qui améliorent les travaux précédents ainsi que des limites inférieures.

Contents

Introduction

Introduction (English version)

This thesis focuses on inhomogeneous random matrices in various fields, such as free probability, quantum information theory, high-dimensional probability, and statistics. A fundamental proof technique consists of applying the moment method to these different scenarios. The manuscript is divided into two main parts: the spectral distribution and the operator norm of inhomogeneous matrices. The first part consists of three chapters. In Chapter 1, we examine a combinatorial variation of the Oriented Kesten-McKay conjecture, along with the spectra of directed random regular graphs. In Chapter 2, we analyze the Empirical Spectral Distribution (ESD) of quantum channels. Finally, in Chapter 3, we prove a new central limit theorem for tensor products of free random variables in a non-commutative probability space. The second part has two chapters. In Chapter 4, we focus on the outliers of the ESD of inhomogeneous symmetric random matrices with subgaussian entries. In Chapter 5, we discuss the problem of covariance estimation for inhomogeneous random covariance matrices.

$\star\star\star$

Let's begin by emphasizing the importance of the moment method and explaining how it works. For a comprehensive analysis, see [8, 103]. Let M_n be an $n \times n$ symmetric matrix and define its ESD as the measure

$$\mu_{M_n} = \frac{1}{n} \sum_{k \in [n]} \delta_{\lambda_k},$$

where $\lambda_1 \geq \cdots \geq \lambda_n$ are its n eigenvalues. The p-th moment of μ_{M_n} is then

$$\int x^p \, \mathrm{d}\mu_{M_n} = \frac{1}{n} \sum_{k \in [n]} \lambda_k^p = \frac{1}{n} \operatorname{tr}(M_n^p).$$

In order to study the weak convergence of the measure μ_{M_n}, it is sufficient to analyze the convergence of the traces $\frac{1}{n} \operatorname{tr}(M_n^p)$ for all $p \geq 1$ under certain conditions. The combinatorial nature of the trace of the p-th power of an $n \times n$ matrix M_n often simplifies the analysis. To illustrate, Wigner [111] proved the following using the moment method. A symmetric matrix $X_n = (\xi_{ij})_{i,j \in [n]}$ whose entries on and above the diagonal are i.i.d copies of a random variable ξ and $\xi_{ij} = \xi_{ji}$ for $i > j$ is called the Wigner matrix associated with ξ.

Theorem 0.1 (Wigner's Theorem)**.** *Let X_n be a Wigner matrix associated with a centered random variable ξ with variance $1/n$. Then the ESD μ_{X_n} converges weakly almost surely to the semi-circle law. Moreover, its (even) moments converge to the Catalan numbers*

$$\frac{1}{n}\operatorname{tr}(X_n^{2p}) \to \frac{1}{p+1}\binom{2p}{p} =: C_p \text{ a. s.}$$

For the sparse case, consider a graph G and its adjacency matrix A_G. Then

$$\frac{1}{n}\operatorname{tr}(A_G^p) = \frac{1}{n}\sum_{v_1,\dots,v_p \in G}\prod_{i\in[p]} A_G(v_i, v_{i+1})$$

is the average number of closed excursions of length p in G, where $v_{p+1} := v_1$ by convention. In particular, one can prove the convergence of ESD μ_{A_G} by counting closed excursions on graph G. For an example, see [78].

Let us now consider the independent problem of controlling the largest eigenvalue λ_1. The limiting law of μ_{M_n} provides no information on λ_1 as its weight is negligible for μ_{M_n}. In the case of a Wigner matrix, λ_1 does not always converge to 2, which is the extreme of the spectrum of the semi-circle law. The following is a theorem by Bai and Yin [12].

Theorem 0.2. *Let X_n be an $n \times n$ Wigner matrix associated with a centered random variable ξ with variance $1/n$. Then $\lambda_1(X_n)$ converges almost surely if and only if $\mathbb{E}\,\xi^4 < \infty$. In this case, $\lambda_1(X_n) \to 2$ almost surely.*

Bai-Yin's theorem again uses the moment method. We define $\|\cdot\|$, the operator norm, as

$$\|M_n\| := \sup_{x\in S^{n-1}} \|M_n x\|_2,$$

where $\|\cdot\|_2$ is the Euclidean norm in $\mathbb{R}^n$ and S^{n-1} is the Euclidean sphere. First note that

$$\liminf_{n\to\infty} \|X_n\| \geq 2.$$

Indeed, this follows by Wigner's Theorem 0.1. We then compare

$$\|M_n\|^{2p} \leq \operatorname{tr}(M_n^{2p}) \leq n\|M_n\|^{2p},$$

for any symmetric $n\times n$ matrix M_n. This is the $l_{2p}-l_\infty$ norm inequality for the eigenvalues of M_n. Jensen's Inequality implies

$$\limsup_{n\to\infty} \mathbb{E}\|X_n\| \leq \limsup_{n\to\infty} \left(\mathbb{E}\operatorname{tr}(X_n^{2p_n})\right)^{\frac{1}{2p_n}}$$

The trace of $X_n^{2p_n}$ has a combinatorial nature, making it easier to work with compared to the operator norm's geometric nature. To prove their result, Bai and Yin then showed

that there exists a sequence $(p_n)_n$ such that $p_n/\log n \to \infty$ and

$$\left(\mathbb{E}\,\mathrm{tr}(X_n^{2p_n})\right)^{\frac{1}{2p_n}} \leq n^{\frac{1}{2p_n}}(2+\Delta_n),$$

where $\Delta_n \to 0$. The proof of the almost sure convergence follows by a concentration argument; see [12].

The thesis is divided into two parts, corresponding to different relaxations of limit theorems (such as Wigner's Theorem). In Part I, we relax the assumption of independence and study random matrices with dependent entries. We analyze the limiting behavior of their ESD. In Part II, we relax the assumption of identical distribution and consider random matrices with independent entries and a variance profile. We derive limits and bounds for their norms. Part I consists of three chapters. In Chapter 1, we will analyze the ESD of regular directed graphs. In Chapter 2, we will examine the limit spectral distribution of quantum channels. Concluding this part, in Chapter 3, we will investigate the limit spectral distribution of tensor products of non-commutative random variables. Part II includes two chapters. In Chapter 4, we will focus on the presence or absence of outliers of subgaussian matrices. Finally, in Chapter 5, we will evaluate the operator norm of covariance matrices.

Part I. Spectral Distribution of Inhomogeneous Matrices

Recall that Wigner's Theorem 0.1 states that the ESD of an $n \times n$ random matrix W_n with i.i.d centered entries with variance $1/n$ converges weakly almost surely to the semi-circle law. In this part, we relax the independence assumption and consider more generally random matrices W_n with dependent structures. We consider two important examples: adjacency matrices of random uniformly chosen regular graphs and tensor products $W_n \otimes W_n$.

⋆ ⋆ ⋆

In Chapter 1, we present the main results from the paper "A combinatorial view on star moments of regular directed graphs and trees" (https://arxiv.org/abs/2309.02225) [41], joint work with Benjamin Dadoun.

Let $G_{n,d}$ be a d-regular graph in n vertices with adjacency matrix $A_{n,d}$. As $G_{n,d}$ is regular, the matrix $A_{n,d}$ has the largest eigenvalue (by the Perron-Frobenius theorem) isolated and equal to d, with associated eigenvector 1. Computational algorithms often rely on the concept of spectral gap, so it is crucial to have a good understanding of

$$\Gamma(G_{n,d}) = d - \lambda_2(G_{n,d}) > 0,$$

and the larger Γ is, the better expansion properties $G_{n,d}$ shares; see [63]. Equivalently, we

want to control

$$\mathbb{E}\|A_{n,d} - \mathbb{E}\,A_{n,d}\| = \mathbb{E}\sup_{x\in S^{n-1}}\|(A_{n,d} - \mathbb{E}\,A_{n,d})x\|_2.$$

The ESD of $G_{n,d}$ was analyzed in [78, 67], and it is known as the Kesten-McKay distribution.

Theorem 0.3. *Let d be fixed and $G_{n,d}$ be a random uniformly chosen d-regular graph on n vertices. Then the ESD $\mu_{A_{n,d}}$ of the adjacency matrix of $G_{n,d}$ converges almost surely to the Kesten-McKay distribution with density*

$$f_{KM}(x) = \frac{d}{2\pi}\frac{\sqrt{4(d-1)-x^2}}{d^2-x^2}\mathbf{1}_{|x|\le 2\sqrt{d-1}}.$$

Moreover, the moments

$$m_{2k} = \int x^{2k} f_{KM}(x)\,dx$$

satisfy the following recursion

$$m_{2k} = d\sum_{l=0}^{k-1} C_l (d-1)^l m_{2(k-1-l)}, \tag{1}$$

where C_l is the l-th Catalan number and $m_0 = 1$.

In this case, it is known that λ_2 remains in the bulk of the spectrum, namely,

$$\lambda_2(G_{n,d}) \to 2\sqrt{d-1}.$$

This was known as the Alon's Conjecture [3] and solved by Friedman [49]; see also [50]. Such graphs are called weakly Ramanujan's or almost Ramanujan's.

Consider the d-regular infinite undirected tree UT_d rooted at a distinguished vertex $o \in UT_d$. Let $\tilde{m}_{2k}$ be the number of closed excursions from o to o of length $2k$. Then, a routine computation shows that $\tilde{m}$ satisfies (1) and both $m_0 = \tilde{m}_0 = 1$. In this case, the spectral distribution of UT_d with respect to the root o is the Kesten-McKay distribution because $\tilde{m} = m$. A connection between UT_d and $G_{n,d}$ can be established by considering the local topology on graphs [24]. A randomly and uniformly drawn d-regular graph $G_{n,d}$ locally converges to a tree as the number of small cycles is sublinear in the dimension. The convergence of the ESD of $G_{n,d}$ can be recovered through Bordenave-Lelarge's criterion [29]. This provides a comprehensive understanding of undirected regular graphs.

Consider the case of directed d-regular graphs $D_{n,d}$, where every vertex has d incoming and d outgoing edges. As its adjacency matrix $A_{n,d}$ is now nonsymmetric, the previous equivalence between the weak convergence of $\mu_{A_{n,d}}$ and the convergence of the traces $\mathrm{tr}(A_{n,d}^p)$ no longer holds [14, Chapter 11]. The following is an analog of the Kesten-McKay theorem [28].

Conjecture 0.4 (Oriented Kesten-McKay Conjecture)**.** *Let $D_{n,d}$ be a directed uniformly chosen d-regular graph on n vertices and $\mu_{A_{n,d}}$ be the ESD of its adjacency matrix. Then $\mu_{A_{n,d}}$ converges weakly, almost surely, to a radial distribution on the unit disc of the complex plane with density*

$$f_{OKM}(z) = \frac{1}{\pi}\frac{d^2(d-1)}{(d^2-|z|^2)^2}\mathbf{1}_{|z|\leq\sqrt{d}}.$$

Let us introduce some notation. For $w \in \{1,*\}^p$ and an $n \times n$ matrix A, let

$$A^w = A^{w_1}\cdots A^{w_p}.$$

In the case $w \equiv 1$, we have $A^w = A^p$, for instance. The moments $\frac{1}{n}\operatorname{tr}(A^w)$ are the star moments of the matrix A, and their convergence is defined as the star convergence. Using a generalization of the Benjamini-Schram topology [24] for directed graphs, we prove that the random uniformly chosen directed graphs $D_{n,d}$ locally converge to the directed d-regular infinite tree T_d, and therefore its moments converge.

Theorem 0.5. *For every $p \geq 0$ and $w \in \{1,*\}^p$, we have*

$$\frac{1}{n}\mathbb{E}\operatorname{tr}(A_{n,d}^w) \to M_d(w),$$

where $A_{n,d}$ is the adjacency matrix of a random uniformly chosen d-regular digraph on n vertices.

Here, $M_d(w)$ is the number of closed w-excursions on the infinite tree T_d starting on a distinguished vertex o, where a closed w-excursion on a digraph G is a sequence of vertices $v_1,\ldots,v_p$ such that $(v_i,v_{i+1}) \in E(G)$ if $w_i = 1$ or $(v_{i+1},v_i) \in E(G)$ otherwise, where we recall $v_{p+1} := v_1$.

Our second main theorem provides a direct and combinatorial formula for calculating $M_d(w)$. For this purpose, we introduce some notation. Let $\pi \in P(k)$ be a partition of $[k]$, and we denote $i \sim_\pi j$ if i and j belong to the same block of π. We say that π is noncrossing if there exist no four-tuple (i_1,i_2,j_1,j_2) with $i_1 < i_2 < j_1 < j_2$ such that $i_1 \sim_\pi j_1$, $i_2 \sim_\pi j_2$, and $i_1 \nsim_\pi i_2$. We denote $NC(k)$, the set of all noncrossing partitions of $[k]$. The cardinal $|\mathrm{NC}(k)|$ is equal to the Catalan number $C_k := \frac{1}{k+1}\binom{2k}{k}$. We further say that π is an *alternating non-crossing partition* of w ($\pi \in \mathrm{ANC}(w)$) if for every block $V := \{i_1 < \cdots < i_m\} \in \pi$, the subword $w|_V := w_{i_1}\cdots w_{i_m}$ of w is *alternating*, that is either of the form $w|_V = 1*\cdots 1*$ or $w|_V = *1\cdots *1$.

Theorem 0.6 (Combinatorial formula for $M_d(w)$)**.** *For every $k \geq 0$ and every $w \in \{1,*\}^k$,*

$$M_d(w) \quad = \sum_{\pi\in\mathrm{ANC}(w)}\left(\prod_{V\in\pi}(-1)^{\frac{|V|}{2}-1}C_{\frac{|V|}{2}-1}\right)d^{|\pi|}. \tag{2}$$

Theorem 0.6 indicates a positive answer to the Kesten-McKay conjecture. Indeed, the w-moments of the measure defined in Conjecture 0.4 are precisely the right-hand side (2); see Section 1.3.

⋆ ⋆ ⋆

In Chapter 2, we present the results from the paper "Limiting spectral distribution of random self-adjoint quantum channels" (https://arxiv.org/abs/2311.12368) [73], joint work with Cécilia Lancien, and Pierre Youssef.

Quantum settings extend classical notions with equivalent questions explored in the literature. In the previous chapter, we examined regular graphs and directed graphs. These mathematical objects have a quantum equivalent: a quantum channel.

In quantum physics, a quantum channel is a linear map Φ from the space of complex-valued $n \times n$ matrices to itself. The Kraus representation can be used to describe the action of Φ ([10, Section 2.3.2], or [114, Chapter 2])

$$\Phi(X) = \frac{1}{d} \sum_{i \in [d]} K_i X K_i^*, \tag{3}$$

where K_i are known as the Kraus operators. The Kraus dimension of the quantum channel is represented by the smallest value of d in equation (3). The quantum channel is trace-preserving if $\mathrm{tr}(\Phi(X)) = \mathrm{tr}(X)$ or equivalently, if

$$\frac{1}{d} \sum_{i \in [d]} K_i K_i^* = \mathrm{Id}\,.$$

Moreover, a trace-preserving quantum channel sends quantum systems (i.e., positive matrices with trace equal to one) to quantum systems. In particular, the analog of the Perron-Frobenius theorem [114, Chapter 6] states that a trace-preserving quantum channel has an isolated largest eigenvalue equal to 1. The spectral gap is defined as

$$\Gamma(\Phi) = 1 - \lambda_2(\Phi),$$

where $\lambda_2(\Phi)$ is the second largest eigenvalue of Φ. As in the classical case mentioned before for regular graphs, the larger the spectral gap is, the better the expansion properties of Φ are, and the faster the mixing time is.

By the canonical identification $M_n(\mathbb{C}) = \mathbb{C}^n \otimes \mathbb{C}^n$, a quantum channel can be identified as the matrix

$$M_\Phi = \frac{1}{d} \sum_{i \in [d]} K_i \otimes \overline{K}_i \in M_{n^2}(\mathbb{C}).$$

Such an equivalence preserves the spectrum of Φ. Therefore, in order to study the spectral properties of Φ, it suffices to study the spectral properties of summations of tensor

products of matrices. Several results [54, 62, 72, 88] have proved that for a large class of independent random Kraus operators K_i, the associated quantum channel has a large spectral gap

$$\mathbb{E}\,\lambda_2(\Phi) = \mathbb{E}\left\|\frac{1}{d}\sum_{i\in[d]} K_i\otimes\overline{K}_i - \mathbb{E}\,K_i\otimes\overline{K}_i\right\| \leq \frac{C}{\sqrt{d}}, \tag{4}$$

which is the equivalent of Alon's conjecture [3] for quantum channels. In particular, for those families of random matrices, their quantum channel is fast mixing.

In the past, research on random quantum channels has primarily focused on the spectral gap and extreme eigenvalues. However, our recent study, as presented in [73], explores the global spectral distribution of Φ, as well as its ESD μ_{M_Φ}. This provides a complementary approach to understanding both the bulk and extreme eigenvalues of the spectrum of Φ.

Let us now state our main result informally.

Theorem 0.7. *Let $W_1,\dots,W_d \in M_n(\mathbb{C})$ be independent Hermitian (i.e., self-adjoint) random matrices such that, for each $i\in[d]$, $\mathbb{E}(W_i)=0$ and $\mathbb{E}(W_i^2)=\mathrm{Id}$. Suppose additionally that, for each $i\in[d]$, the ESD of W_i converges to some distribution μ_i as $n\to\infty$ and that the joint asymptotic distribution of the family $(W_i)_{i\in[d]}$ is determined by the family $(\mu_i)_{i\in[d]}$ (see Section 2.2). Define the quantum channel*

$$\Phi : X\in M_n(\mathbb{C}) \mapsto \frac{1}{\sqrt{d}}\sum_{i\in[d]} W_i X W_i^* \in M_n(\mathbb{C}).$$

Then, the following holds.

- *If d is fixed, then the spectral distribution of $\Phi-\mathbb{E}(\Phi)$ converges to a specific distribution, depending only on $\mu_1,\dots,\mu_d$, as $n\to\infty$.*
- *If $d=d(n)\to\infty$ as $n\to\infty$ and W_i are i.i.d, then the spectral distribution of $\Phi-\mathbb{E}(\Phi)$ converges to the semicircular distribution as $n\to\infty$.*

Remark 0.8. It is important to note that the normalization factor of $1/\sqrt{d}$ in quantum channels comes naturally from a central limit theorem-type result. On the other hand, the previous normalization factor of $1/d$ in the operator norm of (4) is similar to the one used in the law of large numbers.

⋆ ⋆ ⋆

In Chapter 3, we will discuss the results from a joint work in progress with Cécilia Lancien and Pierre Youssef.

In the previous chapter, we studied the quantum channels

$$M_\Phi - \mathbb{E}\,M_\Phi = \frac{1}{\sqrt{d}}\sum_{i\in[d]}\left(W_i\otimes\overline{W}_i - \mathbb{E}\,W_i\otimes\overline{W}_i\right),$$

and we proved that, as d approaches infinity and W_i are i.i.d centered, the ESD of $M_\Phi - \mathbb{E}\, M_\Phi$ converges to the semi-circle law. In this chapter, we remove the centering condition and study more generally the problem of finding the limit law of

$$m - \tau(m) = \frac{1}{\sqrt{d}} \sum_{i \in [d]} (a_i \otimes a_i - \tau \otimes \tau(a_i \otimes a_i)1),$$

where a_i are non-commutative random variables. We begin its description now.

Let $(\mathcal{A}, \tau)$ be a unital non-commutative algebra equipped with a faithful tracial linear state $\tau(ab) = \tau(ba)$ for all $a, b \in \mathcal{A}$ and an involution $*$, that is, $(a^*)^* = a$, for all $a \in \mathcal{A}$. Faithfulness means that $\tau(a^*a) \geq 0$ and equality only holds if $a = 0$. We say that subalgebras $\mathcal{A}_1, \ldots, \mathcal{A}_n \subset \mathcal{A}$ are free if

$$\tau(a_1 \cdots a_k) = 0,$$

whenever the following conditions are satisfied.

1. $k \geq 0$;
2. $\tau(a_i) = 0$ for any $i \in [k]$;
3. $a_i \in \mathcal{A}_{j_i}$, where $j_i \in [n]$ for any $i \in [k]$;
4. Consecutive a_i's are not in the same algebra, $j_1 \neq j_2, \ldots, j_{k-1} \neq j_k$.

We say that random variables $a_1, \ldots, a_n$ are free whenever their algebras are. Freeness is the analog of independence in non-commutative spaces, and limit theorems for independent random variables extend naturally to free random variables. To illustrate, let us recall the free central limit theorem [110], [85, Lecture 8]. We denote

$$\mathrm{var}(a) = \tau((a - \tau(a)1)^2),$$

the variance of a random variable with mean $\tau(a)$, and $1 \in \mathcal{A}$ is the unit in $\mathcal{A}$.

Theorem 0.9. *Let $a_1, \ldots, a_n \in \mathcal{A}$ be self-adjoint (i.e., $a^* = a$) free i.i.d centered random variables with variance one. Let*

$$S_n := \frac{1}{\sqrt{n}} \sum_{i \in [n]} a_i.$$

Then S_n converges in distribution to the semi-circle law, namely, for any $p \geq 1$, we have

$$\tau(S_n^p) \to \tau(s^p),$$

where s is a semi-circle random variable.

Recall that a semi-circle random variable has semi-circle distribution, namely, its odd moments vanish, and for any $p \geq 1$, we have

$$\tau(s^{2p}) = C_p.$$

Its law μ_{sc} has density

$$f(x) = \frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{|x|\leq 2}.$$

The authors of [37] investigated the tensor product of free random variables $b_k = a_k \otimes a_k$, which was inspired by the tensor product of random matrices for quantum channels. They discovered that in the product space $(\mathcal{A}\otimes\mathcal{A}, \tau\otimes\tau)$, such tensors are generally not free. In particular, Theorem 0.9 does not need to hold for the variables b_k. Our goal is then to identify this limit distribution.

Let $a, a_1, \ldots, a_n \in (\mathcal{A}, \tau)$ be self-adjoint free i.i.d random variables. To avoid degenerated random variables, assume $\mathrm{var}(a) > 0$. Otherwise, $a_i = \tau(a)1$, and the result is trivial. Consider

$$b_i = a_i \otimes a_i - \tau\otimes\tau(a_i\otimes a_i)1 = a_i\otimes a_i - \tau^2(a_i)1, \tag{5}$$

and

$$S_n := \frac{1}{\sqrt{n\,\mathrm{var}(b)}}\sum_{i\in[n]} b_i. \tag{6}$$

Note first that

$$\mathrm{var}(b) = \mathrm{var}(a)(\mathrm{var}(a) + 2\tau^2(a)),$$

hence,

$$q := \frac{2\tau^2(a)\,\mathrm{var}(a)}{\mathrm{var}(b)} \in [0,1]. \tag{7}$$

Our main theorem characterizes the limiting distribution of S_n as a function of q. We require some notation. We denote $\boxplus$ the free convolution and for a measure μ, we denote

$$(t\mu)(A) := \mu(t^{-1}A),$$

its dilation by $t > 0$ and Borel sets $A \subset \mathbb{R}$. For $p \in \mathbb{N}$, we denote $P(p)$ the set of all partitions of $[p]$ and $P_2(p)$ the set of all partitions $\pi \in P(p)$ such that for every block $V \in \pi$, $|V| = 2$, and in this case π is called a pair partition. We say that two distinct blocks $V_1, V_2 \in \pi$ cross if there exist $i < j < k < l$ such that $\{i,k\} \subseteq V_1$ and $\{j,l\} \subseteq V_2$. We say that a block V is crossing if there exists another block V' such that V and V' cross. Finally, we define the intersection graph $G(\pi)$ as follows. The vertices of $G(\pi)$ are

the blocks of π, and there exists an edge between blocks V_1, V_2 if they cross. We denote $P^{\text{bi}}(p)$ (resp. $P_2^{\text{bi}}(p)$) the set of all partitions (resp. pair partitions) whose intersection graph is bipartite. We define a measure ν whose odd moments are zero and even moments are equal to

$$\int_{\mathbb{R}} x^{2p} \, \mathrm{d}\nu = |P_2^{\text{bi}}(2p)| \leq \mathbb{E}\, g^{2p},$$

where $g \sim N(0,1)$ is a standard Gaussian random variable whose moments are precisely the cardinal of pair partitions. The existence of the measure ν is discussed in Remark 3.4, and it is the Moment Problem for the sequence $(|P_2^{\text{bi}}(2n)|)_{n\geq 1}$ [2, 92].

The following is our main theorem.

Theorem 0.10. *Let $a, a_1, \ldots, a_n \in (\mathcal{A}, \tau)$ be self-adjoint free i.i.d random variables with variance $\operatorname{var}(a) > 0$. Let b_i and S_n as in (5), (6), respectively, and*

$$q = \frac{2\tau^2(a)\operatorname{var}(a)}{\operatorname{var}(b)} \in [0,1].$$

Then S_n converges in distribution to the weighted free convolution of ν and the semi-circle law

$$S_n \Rightarrow \sqrt{\frac{q}{2}}\nu \boxplus \sqrt{\frac{q}{2}}\nu \boxplus \sqrt{1-q}\,\mu_{sc}.$$

Part II. Norm of Inhomogeneous Matrices

In this section, we relax the assumption of identical distribution to prove limit theorems and consider inhomogeneous random matrices $W_n = (w_{ij})$ with independent entries and a variance profile where $\mathbb{E}(w_{ij}^2) = \sigma_{ij}^2$. We will focus on two examples. Firstly, we will compute and study the convergence of the norm of a symmetric inhomogeneous matrix W_n. Secondly, we consider asymmetric $d \times n$ matrices W_n and study bounds on the operator norm of its quadratic form.

⋆ ⋆ ⋆

In Chapter 4, we present the main results from the paper "On spectral outliers of inhomogeneous symmetric random matrices" (https://arxiv.org/abs/2401.07852) [5], joint work with Dylan J. Altschuler, Konstantin Tikhomirov, and Pierre Youssef.

In Chapter 1, we explored the ESD of d-regular graphs G_n. However, we can also consider a weighted regular graph $\tilde{G}_n$, where we assign independent and identically distributed (i.i.d) random variables ξ_e to each edge $e \in E(G_n)$. In other words, the adjacency matrix of $\tilde{G}_n$ is given by $X_n = A_{G_n} \circ W_n$, where A_{G_n} is the adjacency matrix of G_n, W_n is a symmetric random matrix with i.i.d entries following the law ξ, and $\circ$ denotes the

Hadamard (entrywise) product. This model is known as a sparse Wigner matrix. A natural question that arises is the following. "What is the limit ESD of X_n when $d = d(n)$ is a function of n?"

The celebrated Wigner's Theorem 0.1 has a universal characteristic similar to the central limit theorem. It means that the limit is not affected by the distribution of entries but solely by its variance. Wigner's ensemble corresponds to the complete graph G_n. According to Bai-Yin's Theorem 0.2, for the extreme eigenvalue λ_1, such universality breaks down in a weak sense. Namely, they prove that if the fourth moment is finite, then there are no outliers

$$\|W_n\| \to 2 \text{ a. s.}$$

Otherwise, the operator norm does not converge. The universality of both of these results has gained significant focus for other models, in particular for those without the i.i.d condition [25, 81, 45, 47, 48]. One important model is the sparse case, which is used in theoretical analysis, numerical simulations, statistical inference, graph theory, and random matrices [18, 19, 31]. The model we will study in this chapter is as follows.

Model. Let W_n be an $n \times n$ (symmetric) Wigner matrix whose on and above the diagonal entries are i.i.d copies of a centered random variable ξ having unit variance. Let $\Sigma_n = (\sigma_{ij})$ be an $n \times n$ symmetric matrix with positive entries such that

$$\sum_{i \in [n]} \sigma_{ij}^2 = 1,$$

for all $j \in [n]$. We consider the dilation $X_n = \Sigma_n \circ W_n$, where $\circ$ denotes the Hadamard (i.e., entrywise) product. We call X_n a dilation of the Wigner matrix W_n.

It is natural to ask whether the universality of the spectrum and the extreme eigenvalues holds in the inhomogeneous case. Specifically, we can question whether the limit spectral distribution of X_n and its norm limit are dependent on ξ and Σ_n. The first has been successfully answered in a more general setting in [55].

Theorem 0.11. *Let $X_n = \Sigma_n \circ W_n$ be a dilation of an $n \times n$ Wigner matrix associated with a centered random variable ξ with unit variance. Then, the following holds.*

1. *If*

$$\sigma_n^* = \max_{ij} \sigma_{ij} \to 0,$$

   the limit law of μ_{X_n} is the semi-circle law, no matter the distribution of ξ and the structure of Σ_n.

2. *Otherwise, if σ_n^* does not converge to zero, the limit law of the ESD of X_n (if it exists) depends both on ξ and the structure of Σ_n.*

Regarding the extremity of the spectrum, the results from [74] already demonstrate that such universality does not hold. In order to determine the influence of the structure of Σ_n, we will restrict our case to subgaussian random variables. Recall that a subgaussian random variable ξ satisfies

$$\mathbb{E}\, e^{\lambda(\xi-\mathbb{E}\,\xi)} \leq e^{\sigma^2\lambda^2/2},$$

for all $\lambda \in \mathbb{R}$ and some $\sigma > 0$. In the Gaussian case, the following result is shown [19].

Theorem 0.12. *Suppose $\xi \sim N(0,1)$ and let $X_n = \Sigma_n \circ W_n$ be the dilation of its $n \times n$ Wigner matrix. Then*

$$\mathbb{E}\|X_n\| \leq (1+\varepsilon)\Big(2 + C(\varepsilon)\sigma_n^*\sqrt{\log n}\Big),$$

for any $\varepsilon \in (0, 1/2)$.

The main idea to prove Theorem 0.12 is a compression argument that compares the matrix X_n to a standard symmetric $r \times r$ random matrix G_r with iid Gaussian entries on and above the diagonal such that

$$\mathbb{E}\,\mathrm{tr}(X_n^{2p}) \leq \kappa\,\mathbb{E}\,\mathrm{tr}(G_r^{2p}).$$

Here, κ and the dimension r are tuned appropriately according to the variance profile; see [19]. The latter is then computed via sharp bounds on Gaussian processes and Gaussian concentration as

$$\mathbb{E}\,\mathrm{tr}(G_r^{2p}) \leq n\,\mathbb{E}\|G_r\|^{2p}.$$

Note that the convergence of the spectral distribution of X_n implies that

$$\liminf \mathbb{E}\|X_n\| \geq 2.$$

Therefore, Theorem 0.12 implies that whenever $\sigma_n^*\sqrt{\log n}$ goes to zero, there is no outliers

$$\mathbb{E}\|X_n\| \to 2.$$

Such a result was extended to bounded random variables ξ in [74], but it remained open whether one could extend it to all subgaussian random variables. This is indeed the case, as our first main theorem shows.

Theorem 0.13. *Let $X_n = \Sigma_n \circ W_n$ be a dilation of an $n \times n$ Wigner matrix W_n associated with a centered subgaussian random variable ξ having variance one. If $\sigma_n^*\sqrt{\log n}$ goes to zero, X_n has no outliers, namely, $\|X_n\| \to 2$ almost surely.*

Restricting to the matrices $\Sigma_n = \frac{1}{\sqrt{d}}A_{G_n}$, that is, normalized adjacency matrices of

d-regular graphs G_n, the previous result states that whenever $d \gg \log n$, there are no outliers. Seginer [93] partially answered the converse.

Theorem 0.14. *Let ξ be a Rademacher random variable and W_n be its $n \times n$ Wigner matrix. Then, there exists a regular graph G_n with degree $d = \lceil \sqrt{\log n} \rceil$ such that*

$$\mathbb{E}\left\| \frac{1}{\sqrt{d}} A_{G_n} \circ W_n \right\| \geq c \log^{1/4} n.$$

Our second main theorem lifts the suboptimal condition $d = \lceil \sqrt{\log n} \rceil$ all the way up to the sharp condition $d \sim \log n$.

Theorem 0.15. *Let ξ be a centered random variable with variance one and bounded fourth moment. Let W_n be its $n \times n$ Wigner matrix. Fix any sequence $d_n = O(\log n)$ such that $d_n \to \infty$ and nd_n is even. Then, there exists a sequence of d_n regular graphs G_n such that the matrices $\frac{1}{\sqrt{d_n}} A_{G_n} \circ W_n$ have outliers almost surely. Namely,*

$$\mathbb{P}\left(\liminf_{n\to\infty} \left\| \frac{1}{\sqrt{d_n}} A_{G_n} \circ W_n \right\| > 2 \right) = 1.$$

$\star \star \star$

In Chapter 5, we present the results from the paper "Almost sharp covariance and Wishart-type matrix estimation" (https://arxiv.org/abs/2307.09190) [91].

Recall that Theorem 0.12 yields

$$\mathbb{E}\|X\| \leq (1+\varepsilon)(2 + C(\varepsilon)\sigma^* \sqrt{\log n}),$$

where $X = X_n$ and $\sigma^* = \sigma_n^*$. Here, we drop the index n as in this chapter we will consider the dimensions to be fixed. We will usually denote by $X_1, \ldots, X_n$ the *columns* of the matrix X.

A natural extension is to consider asymmetric $d \times n$ matrices X and polynomials on X. Let $P \in \mathbb{R}[X, Y]$ be a polynomial. We aim to bound

$$c(P) := \mathbb{E}\left\| P(X, X^t) \right\|, \tag{8}$$

here X^t is the transpose of X. Such a problem is related to the concept of Strong Asymptotic Freeness and outliers of the ESD of random matrices; see, for instance, [59, 38, 12, 14, 35]. In this chapter, we will consider a centered quadratic form

$$P(X, X^t) := XX^t - \mathbb{E}\, XX^t = \sum_{k \in [n]} P(X_k, X_k^t), \tag{9}$$

where X_k are the vector-columns of X. This is known as the Covariance Estimation

Problem (CEP). Notice that, if instead we let $P_1(X, X^t) = X \otimes X^t$ and

$$M := \sum_{k\in[n]} P_1(M_k, M_k^t),$$

for matrices M_k, then M would be a quantum channel studied in Chapter 2.

The CEP is a well-known problem in statistics theory and has been studied extensively in literature [79, 100, 116, 89, 69, 34, 33]. It also has connections with convex geometry, as explained in Chapter 9 of [108], and outliers of the spectrum of random matrices. The problem can be described as follows. Given a collection of independent and identically distributed random vectors $X_1, \dots, X_n \in \mathbb{R}^d$, we define its empirical covariance matrix as

$$\Sigma_n = \frac{1}{n} \sum_{k\in[n]} X_k X_k^t.$$

By the Law of Large Numbers, it is known that Σ_n converges entrywise to the true covariance matrix $\Sigma = \mathbb{E}\, X_1 X_1^t$. The objective is to control the speed of convergence. Specifically, we aim to bound

$$\mathbb{E}\|\Sigma_n - \Sigma\| = \mathbb{E} \sup_{x\in S^{d-1}} \|(\Sigma_n - \Sigma)x\|_2.$$

In the homogeneous case, it is known [108, Theorem 4.7.1].

Theorem 0.16. *Let $G_1, \dots, G_n \in \mathbb{R}^d$ be i.i.d Gaussian random variables $N(0, \mathrm{Id})$. Then*

$$\mathbb{E}\|\Sigma_n - \Sigma\| \leq C \max\left\{\frac{d}{n}, \sqrt{\frac{d}{n}}\right\},$$

where C is a universal constant.

Independent vectors $X_1, \dots, X_n \in \mathbb{R}^d$, but not necessarily identically distributed, require good concentration bounds in the absence of the iid hypothesis. Recent works [69, 33, 18, 31] have lifted this hypothesis for more general settings.

By Bai-Yin's theorem, in order to have convergence for the norm, we require at least the existence of the fourth moment of the entrywise distribution X_{ij} of each vector $X_j \in \mathbb{R}^d$. For concreteness, we will consider the case of Gaussian random vectors with independent entries. Let $X = (X_{ij}) = (b_{ij} g_{ij})$ be the $d \times n$ matrix with columns $X_j = (b_{ij} g_{ij})$, where $b_{ij} \geq 0$ are deterministic coefficients and g_{ij} are i.i.d standard Gaussian random variables. Then our goal is to evaluate

$$n\,\mathbb{E}\|\Sigma_n - \Sigma\| = \mathbb{E}\left\|XX^t - \mathbb{E}\, XX^t\right\|.$$

In order to introduce some parameters, consider bounding $\mathbb{E}\|X\|$, as we did in Chapter

4. Since the operator norm is bigger than the norm of each column X_j, we have

$$\mathbb{E}\|X\| \geq \max_{j\in[n]} \mathbb{E}\|X_j\|_2 \geq c \max_{j\in[n]} (\mathbb{E}\|X_j\|^2)^{1/2},$$

where Gaussian Poincaré Inequality implies the second inequality. In particular, we have

$$\mathbb{E}\|X\| \geq c \max_{j\in[n]} \left(\sum_{i\in[d]} b_{ij}^2 \right)^{1/2} =: \sigma_C,$$

where σ_C denotes the maximum Euclidean norm of the columns. Similarly,

$$\mathbb{E}\|X\| \geq c\sigma_R,$$

and σ_R is the maximum Euclidean norm of the rows. On the other hand, the operator norm is bigger than the maximum of the entries; hence

$$\mathbb{E}\|X\| \geq \mathbb{E} \max_{ij} |b_{ij} g_{ij}|.$$

If $B = (b_{ij})$ is sufficiently homogeneous, meaning that a polynomial proportion of the coefficients b_{ij} is of the same order as

$$\sigma_* = \max_{i,j} b_{ij},$$

then we can immediately lower bound

$$\mathbb{E}\|X\| \geq c\sigma_* \sqrt{\log(n \wedge d)},$$

by using the maximum of Gaussian restricted to those large coordinates; see [19]. In this case, the parameters $\sigma_*, \sigma_C, \sigma_R$ are the only ones that matter. Therefore, no structural parameter is required. Define the following parameters:

- $\tilde{\sigma}_\infty^2 = \max_{i,l:i\neq l} \sum_{j\in[n]} b_{ij}^2 b_{lj}^2$;
- $\bar{\sigma}_\infty^2 = \max_{i\in[d]} \sum_{j\in[n]} b_{ij}^4$;
- $\sigma_\infty^2 = \max_{i\in[d]} \sum_{j\in[n]} \sum_{l:l\neq i} b_{ij}^2 b_{lj}^2$;
- $\beta_\infty = \dfrac{\tilde{\sigma}_\infty \sigma_C}{\sigma_\infty \sigma_*}$.

The following is our main theorem, which is going to be shown in Chapter 5.

Theorem 0.17. *Let X be a $d \times n$ Gaussian matrix with independent entries such that $X_{ij} = b_{ij} g_{ij}$ where $\{g_{ij} : (i,j) \in [d] \times [n]\}$ are i.i.d standard Gaussian r.v. Let $\varepsilon \in (0, 1/2)$. Then, the following holds.*

1. *If* $\beta_\infty \le 1$, *we have*

$$\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| = \mathbb{E}\left\|\sum_{j\in[n]} X_jX_j^T - \mathbb{E}\,X_jX_j^T\right\|$$
$$\le (1+\varepsilon)\left\{2\sigma_\infty + \sigma_C^2 + C(\varepsilon)\sigma_*\left(\sigma_C + \frac{\sigma_\infty}{\sigma_C}\right)\sqrt{\log(n\wedge d)} + C^2(\varepsilon)\sigma_*^2\log(n\wedge d)\right\}.$$

2. *Otherwise,* $\beta_\infty > 1$ *and we have*

$$\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\|$$
$$\le (1+\varepsilon)\left\{\frac{2\tilde\sigma_\infty\sigma_C}{\sigma_*} + \sigma_C^2 + C(\varepsilon)(\sigma_C\sigma_* + \bar\sigma_\infty)\sqrt{\log(n\wedge d)} + C^2(\varepsilon)\sigma_*^2\log(n\wedge d)\right\}.$$

The constant $C(\varepsilon)$ *is*

$$C(\varepsilon) \le \frac{C}{\sqrt{\varepsilon}},$$

where C *is a universal constant.*

Theorem 0.17 improves upon the recent works [33, Theorem 2.1] and [31, Theorem 3.17], [18, Theorem 3.12] when $\beta_\infty \le 1$. The condition on β_∞ is a byproduct of the proof, and it is unclear whether it is necessary. In particular, the leading term $2\sigma_\infty + \sigma_C^2$ seems to be the correct bound on all cases, as it is indicated by some very sparse examples, the lower bound on the norm and [18].

Introduction (Version française)

Cette thèse se concentre sur les matrices aléatoires inhomogènes dans divers domaines, tels que les probabilités libres, la théorie de l'information quantique, les probabilités en grande dimension et les statistiques. Une technique de preuve fondamentale consiste à appliquer la méthode des moments à ces différents scénarios. Le manuscrit est divisé en deux parties principales : la distribution spectrale et la norme d'opérateur des matrices inhomogènes. La première partie se décompose en trois chapitres. Dans le chapitre 1, nous examinons une variante combinatoire de la conjecture de Kesten-McKay orientée, ainsi que les spectres de graphes réguliers aléatoires dirigés. Dans le chapitre 2, nous analysons la distribution spectrale empirique (DSE) des canaux quantiques. Enfin, dans le chapitre 3, nous prouvons un nouveau théorème central limite pour les produits tensoriels de variables aléatoires libres dans un espace de probabilité non-commutatif. La deuxième partie comporte deux chapitres. Dans le chapitre 4, nous nous concentrons sur les valeurs aberrantes de la DSE de matrices aléatoires symétriques inhomogènes avec des entrées sous-gaussiennes. Dans le chapitre 5, nous discutons du problème de l'estimation de la covariance pour les matrices de covariance aléatoires inhomogènes.

⋆ ⋆ ⋆

Commençons par souligner l'importance de la méthode des moments et par expliquer comment elle fonctionne. Pour une analyse complète, voir [8, 103]. Soit M_n une matrice symétrique $n \times n$ et définissons sa DSE comme la mesure

$$\mu_{M_n} = \frac{1}{n} \sum_{k \in [n]} \delta_{\lambda_k},$$

où $\lambda_1 \geq \cdots \geq \lambda_n$ sont ses n valeurs propres. Le p-ième moment de μ_{M_n} est alors

$$\int x^p \, \mathrm{d}\mu_{M_n} = \frac{1}{n} \sum_{k \in [n]} \lambda_k^p = \frac{1}{n} \operatorname{tr}(M_n^p).$$

Pour étudier la convergence faible de la mesure μ_{M_n}, il suffit d'analyser la convergence des traces $\operatorname{tr}(M_n^p)$, pour tout $p \geq 1$, sous certaines conditions. La nature combinatoire de la trace de la puissance p d'une matrice $n \times n$ M_n simplifie souvent l'analyse. Par exemple, Wigner [111] a prouvé ce qui suit en utilisant la méthode des moments. Une matrice symétrique $X_n = (\xi_{ij})_{i,j \in [n]}$ dont les entrées sur et au-dessus de la diagonale sont

des copies i.i.d d'une variable aléatoire ξ et $\xi_{ij} = \xi_{ji}$ pour $i > j$ est appelée la matrice de Wigner associée à ξ.

Théorème 0.18. *Soit X_n une matrice de Wigner associée à une variable aléatoire centrée ξ de variance $1/n$. Alors la DSE μ_{X_n} converge faiblement presque sûrement vers la loi du demi-cercle. De plus, ses moments (pairs) convergent vers les nombres de Catalan*

$$\frac{1}{n}\operatorname{tr}(X_n^{2p}) \to \frac{1}{p+1}\binom{2p}{p} =: C_p \text{ a. s.}$$

Pour le cas parcimonieux, considérons un graphe G et sa matrice d'adjacence A_G. Dans ce cas

$$\frac{1}{n}\operatorname{tr}(A_G^p) = \frac{1}{n}\sum_{v_1,\dots,v_p \in G}\prod_{l\in[p]} A_G(v_l, v_{l+1})$$

est le nombre moyen d'excursions fermées de longueur p dans G, où $v_{p+1} := v_1$, par convention. En particulier, on peut prouver la convergence de la DSE μ_{A_G} en comptant les excursions fermées sur le graphe G. Pour un exemple, voir [78].

Considérons maintenant le problème indépendant du contrôle de la plus grande valeur propre λ_1. La loi limite de μ_{M_n} ne fournit aucune information sur λ_1 parce que son poids est négligeable pour μ_{M_n}. Dans le cas d'une matrice de Wigner, λ_1 ne converge pas toujours vers 2, qui est l'extrême du spectre de la loi du demi-cercle. Voici un théorème de Bai et Yin [12].

Théorème 0.19. *Soit X_n une matrice de Wigner $n \times n$ associée à une variable aléatoire centrée ξ de variance $1/n$. Alors $\lambda_1(X_n)$ converge presque sûrement si et seulement si $\mathbb{E}\,\xi^4 < \infty$. Dans ce cas, $\lambda_1(X_n)$ converge vers 2 presque sûrement.*

Le théorème de Bai-Yin utilise à nouveau la méthode des moments. Nous définissons $\|\cdot\|$, la norme d'opérateur, comme suit

$$\|M_n\| := \sup_{x\in S^{n-1}} \|M_n x\|_2,$$

où $\|\cdot\|_2$ est la norme euclidienne dans $\mathbb{R}^n$ et S^{n-1} est la sphère euclidienne. Notons tout d'abord que

$$\liminf_{n\to\infty} \|X_n\| \geq 2.$$

En effet, cela découle du théorème de Wigner 0.18. Nous comparons ensuite

$$\|M_n\|^{2p} \leq \operatorname{tr}(M_n^{2p}) \leq n\|M_n\|^{2p},$$

pour toute matrice symétrique $n \times n$ M_n. Il s'agit de l'inégalité de norme $l_{2p} - l_\infty$ pour

les valeurs propres de M_n. L'inégalité de Jensen implique

$$\limsup_{n\to\infty} \mathbb{E}\|X_n\| \leq \limsup_{n\to\infty} \Big(\mathbb{E}\,\mathrm{tr}(X_n^{2p_n})\Big)^{\frac{1}{2p_n}}.$$

La trace de $X_n^{2p_n}$ est de nature combinatoire, ce qui la rend plus facile à utiliser que la norme d'opérateur, de nature géométrique. Pour prouver leur résultat, Bai et Yin ont ensuite montré qu'il existe une suite $(p_n)_n$ telle que $p_n/\log n \to \infty$ et que

$$\Big(\mathbb{E}\,\mathrm{tr}(X_n^{2p_n})\Big)^{\frac{1}{2p_n}} \leq n^{\frac{1}{2p_n}}(2+\Delta_n),$$

où $\Delta_n \to 0$. La preuve de la convergence presque sûre découle d'un argument de concentration ; voir [12].

La thèse est divisée en deux parties, correspondant à différentes relaxations des théorèmes limites (tels que le théorème de Wigner). Dans la partie I, nous relâchons l'hypothèse d'indépendance et étudions les matrices aléatoires avec des entrées dépendantes. Nous analysons le comportement limite de leur ESD. Dans la partie II, nous relâchons l'hypothèse de distribution identique et considérons des matrices aléatoires avec des entrées indépendantes et un profil de variance. Nous dérivons des limites et des bornes pour leurs normes. La partie I se compose de trois chapitres. Au chapitre 1, nous analyserons la DSE des graphes dirigés réguliers. Au chapitre 2, nous examinerons la distribution spectrale limite des canaux quantiques. Pour conclure cette partie, au chapitre 3, nous étudierons la distribution spectrale limite des produits tensoriels de variables aléatoires non-commutatives. La partie II comprend deux chapitres. Dans le chapitre 4, nous nous concentrerons sur la présence ou l'absence de valeurs aberrantes de matrices sous-gaussiennes. Enfin, dans le chapitre 5, nous évaluerons la norme d'opérateur des matrices de covariance.

Partie I. Distribution spectrale des matrices inhomogènes

Rappelons que le Théorème de Wigner 0.18 stipule que la DSE d'une matrice aléatoire $n \times n$ W_n avec des entrées centrées i.i.d. de variance $1/n$ converge faiblement et presque sûrement vers le demi-cercle. Dans cette partie, nous relâchons l'hypothèse d'indépendance et considérons plus généralement des matrices aléatoires W_n avec des structures dépendantes. Nous considérons deux exemples importants : les matrices d'adjacence de graphes réguliers aléatoires uniformément choisis, et les produits tensoriels $W_n \otimes W_n$.

⋆ ⋆ ⋆

Dans le Chapitre 1, nous présentons les principaux résultats de l'article "A combinatorial view on star des graphes et arbres dirigés réguliers" (https://arxiv.org/abs/2309.02225) [41], travail en commun avec Benjamin Dadoun.

Soit $G_{n,d}$ un graphe d-régulier à n sommets avec une matrice d'adjacence $A_{n,d}$. Comme $G_{n,d}$ est régulière, la matrice $A_{n,d}$ a la plus grande valeur propre (par le théorème de

Perron-Frobenius) isolée et égale à d, avec le vecteur propre 1 associé. Les algorithmes de calcul s'appuient souvent sur le concept de trou spectral, il est donc crucial de bien comprendre

$$\Gamma(G_{n,d}) = d - \lambda_2(G_{n,d}) > 0,$$

et plus Γ est grand, meilleures sont les propriétés d'expansion de $G_{n,d}$; voir [63]. De manière équivalente, nous voulons contrôler

$$\mathbb{E}\|A_{n,d} - \mathbb{E}\,A_{n,d}\| = \mathbb{E}\sup_{x\in S^{n-1}}\|(A_{n,d} - \mathbb{E}\,A_{n,d})x\|_2.$$

La DSE de $G_{n,d}$ a été analysée dans [78, 67], et elle est connue sous le nom de distribution de Kesten-McKay.

Théorème 0.20. *Soit d fixé et $G_{n,d}$ un graphe d-régulier aléatoire uniformément choisi sur n sommets. Alors la DSE $\mu_{A_{n,d}}$ de la matrice d'adjacence de $G_{n,d}$ converge presque sûrement vers la distribution de Kesten-McKay avec la densité*

$$f_{KM}(x) = \frac{d}{2\pi}\frac{\sqrt{4(d-1)-x^2}}{d^2-x^2}\mathbf{1}_{|x|\le 2\sqrt{d-1}}.$$

De plus, les moments

$$m_{2k} = \int x^{2k} f_{KM}(x)\,dx$$

satisfant à la récursivité suivante

$$m_{2k} = d\sum_{l=0}^{k-1} C_l (d-1)^l m_{2(k-1-l)}, \tag{10}$$

où C_l est le l-ème nombre de Catalan.

Dans ce cas, on sait aussi que λ_2 reste dans la majeure partie du spectre, à savoir,

$$\lambda_2(G_{n,d}) \to 2\sqrt{d-1}.$$

Ce résultat, connu sous le nom de Conjecture d'Alon [3], a été prouvé par Friedman [49] ; voir également [50]. De tels graphes sont appelés faiblement de Ramanujan ou presque de Ramanujan.

Considérons l'arbre non-dirigé infini d-régulier UT_d enraciné à un sommet distingué $o \in UT_d$. Soit $\tilde{m}_{2k}$ le nombre d'excursions fermées de o à o de longueur $2k$. Un calcul de routine montre alors que $\tilde{m}$ satisfait (10) et que $m_0 = \tilde{m}_0 = 1$. Dans ce cas, la distribution spectrale de UT_d par rapport à la racine o est la distribution de Kesten-McKay puisque $\tilde{m} = m$. Un lien entre UT_d et $G_{n,d}$ peut être établi en considérant la topologie locale sur les graphes [24]. Un graphe d-régulier aléatoirement et uniformément choisi $G_{n,d}$

converge localement vers un arbre car le nombre de petits cycles est sous-linéaire dans la dimension. La convergence de la DSE de $G_{n,d}$ peut être retrouvée grâce au critère de Bordenave-Lelarge [29]. Ceci permet une compréhension complète des graphes réguliers non dirigés.

Considérons le cas des graphes réguliers dirigés $D_{n,d}$, où chaque sommet a d arêtes entrantes et d arêtes sortantes. Comme la matrice d'adjacence $A_{n,d}$ est maintenant non symétrique, l'équivalence précédente entre la convergence faible de $\mu_{A_{n,d}}$ et la convergence des traces $\mathrm{tr}(A_{n,d}^p)$ ne tient plus [14, Chapitre 11]. Ce qui suit est un analogue du théorème de Kesten-McKay [28].

Conjecture 0.21 (Conjecture de Kesten-McKay orientée)**.** *Soit $D_{n,d}$ un graphe d-régulier dirigé uniformément choisi sur n sommets et $\mu_{A_{n,d}}$ la DSE de sa matrice d'adjacence. Alors $\mu_{A_{n,d}}$ converge faiblement presque sûrement vers une distribution radiale sur le disque unitaire du plan complexe avec la densité*

$$f_{OKM}(z) = \frac{1}{\pi}\frac{d^2(d-1)}{(d^2-|z|^2)^2}\mathbf{1}_{|z|\le d}.$$

Introduisons quelques notations. Pour $w \in \{1,*\}^p$ et une matrice $n \times n$ A, notons

$$A^w = A^{w_1}\cdots A^{w_p}.$$

Dans le cas $w \equiv 1$, nous avons $A^w = A^p$, par exemple. Les moments $\frac{1}{n}\mathrm{tr}(D^w)$ sont les moments en étoile de la matrice A, et leur convergence est définie comme la convergence en étoile. En utilisant une généralisation de la topologie de Benjamini-Schram [24] pour les graphes dirigés, nous prouvons que les graphes dirigés aléatoires $D_{n,d}$ uniformément choisis convergent localement vers l'arbre infini d-régulier dirigé T_d, et donc que ses moments convergent.

Théorème 0.22. *Pour tout $p \ge 0$ et $w \in \{1,*\}^p$, nous avons*

$$\frac{1}{n}\mathbb{E}\,\mathrm{tr}(A_{n,d}^w) \to M_d(w),$$

où $A_{n,d}$ est la matrice d'adjacence d'un digraphe aléatoire uniformément choisi d-régulier sur n sommets.

Ici, $M_d(w)$ est le nombre de w-excursions fermées sur l'arbre infini T_d à partir d'un sommet distingué o, où une w-excursion fermée sur un digraphe G est une suite de sommets $v_1,\dots,v_p$ telle que $(v_i,v_{i+1}) \in E(G)$ si $w_i = 1$ ou $(v_{i+1},v_i) \in E(G)$ sinon, où nous rappelons que $v_{p+1} := v_1$.

Notre deuxième théorème principal fournit une formule directe et combinatoire pour calculer $M_d(w)$. Pour cela, nous introduisons quelques notations. Soit $\pi \in P(k)$ une partition de $[k]$, et nous notons $i \sim_\pi j$ si i et j appartiennent au même bloc de π. Nous disons que π est non croisée s'il n'existe aucun quadruplet (i_1,i_2,j_1,j_2) tel que

$i_1 < i_2 < j_1 < j_2$, $i_1 \sim_\pi j_1$, $i_2 \sim_\pi j_2$, et $i_1 \nsim_\pi i_2$. Nous notons $NC(k)$, l'ensemble de toutes les partitions non croisées de $[k]$. Le cardinal $|\mathrm{NC}(k)|$ est égal au nombre de Catalan $C_k := \frac{1}{k+1}\binom{2k}{k}$. Nous disons également que π est une *partition non croisée alternée* de w ($\pi \in \mathrm{ANC}(w)$) si pour chaque bloc $V := \{i_1 < \cdots < i_m\} \in \pi$, le sous-mot $w|_V := w_{i_1} \cdots w_{i_m}$ de w est *alternant*, c'est-à-dire de la forme $w|_V = 1*\cdots 1*$ ou $w|_V = *1\cdots *1$.

Théorème 0.23 (Formule combinatoire pour $M_d(w)$)**.** *Pour tout $k \geq 0$ et tout $w \in \{1, *\}^k$,*

$$M_d(w) \quad = \sum_{\pi \in \mathrm{ANC}(w)} \left(\prod_{V \in \pi} (-1)^{\frac{|V|}{2}-1} C_{\frac{|V|}{2}-1} \right) d^{|\pi|}. \tag{11}$$

Le théorème 0.23 indique une réponse positive à la conjecture de Kesten-McKay. En effet, les w-moments de la mesure définie dans la conjecture 0.21 sont précisément le côté droit de l'égalité (11) ; voir la section 1.3.

⋆ ⋆ ⋆

Dans le chapitre 2, nous présentons les résultats de l'article "Limiting spectral distribution of random self-adjoint quantum channels" (https://arxiv.org/abs/2311.12368) [73], travail en commun avec Cécilia Lancien, et Pierre Youssef.

Les paramètres quantiques étendent les notions classiques avec des questions équivalentes explorées dans la littérature. Dans le chapitre précédent, nous avons examiné les graphes réguliers et les graphes dirigés. Ces objets mathématiques ont un équivalent quantique appelé un canal quantique.

En physique quantique, un canal quantique est un endomorphisme Φ de l'espace des matrices $n \times n$ à valeurs complexes La représentation de Kraus peut être utilisée pour décrire l'action de Φ ([10, Section 2.3.2], ou [114, Chapitre 2])

$$\Phi(X) = \frac{1}{d} \sum_{i \in [d]} K_i X K_i^*, \tag{12}$$

où les K_i sont définis comme les opérateurs de Kraus. La dimension de Kraus du canal quantique est représentée par la plus petite valeur de d dans l'équation (12). Le canal quantique préserve la trace si $\mathrm{tr}(\Phi(X)) = \mathrm{tr}(X)$ ou, de manière équivalente, si

$$\frac{1}{d} \sum_{i \in [d]} K_i K_i^* = \mathrm{Id}\,.$$

De plus, un canal quantique à trace préservée envoie des systèmes quantiques (c'est-à-dire des matrices positives dont la trace est égale à un) vers des systèmes quantiques. En particulier, l'analogue du théorème de Perron-Frobenius [114, Chapitre 6] stipule qu'un

canal quantique préservant la trace a une plus grande valeur propre isolée égale à 1. Le trou spectral est défini comme suit

$$\Gamma(\Phi) = 1 - \lambda_2(\Phi),$$

où $\lambda_2(\Phi)$ est la deuxième plus grande valeur propre de Φ. Comme dans le cas classique mentionné précédemment pour les graphes réguliers, plus le trou spectral est grand, plus les propriétés d'expansion de Φ sont bonnes, et plus le temps de mélange est rapide.

Par l'identification canonique $M_n(\mathbb{C}) = \mathbb{C}^n \otimes \mathbb{C}^n$, un canal quantique peut être identifié a la matrice

$$M_\Phi = \frac{1}{d} \sum_{i \in [d]} K_i \otimes \overline{K}_i \in M_{n^2}(\mathbb{C}).$$

Une telle équivalence préserve le spectre de Φ. Par conséquent, pour étudier les propriétés spectrales de Φ, il suffit d'étudier les propriétés spectrales des sommations de produits tensoriels de matrices. Plusieurs résultats [54, 62, 72, 88] ont prouvé que pour une grande classe d'opérateurs de Kraus aléatoires indépendants K_i, le canal quantique associé a un grand trou spectral

$$\mathbb{E}\,\lambda_2(\Phi) = \mathbb{E}\left\| \frac{1}{d} \sum_{i \in [d]} K_i \otimes \overline{K}_i - \mathbb{E}\, K_i \otimes \overline{K}_i \right\| \leq \frac{C}{\sqrt{d}}, \tag{13}$$

ce qui est l'équivalent de la conjecture d'Alon [3] pour les canaux quantiques. En particulier, pour ces familles de matrices aléatoires, leur canal quantique est à mélange rapide.

Dans le passé, la recherche sur les canaux quantiques aléatoires s'est principalement concentrée sur le trou spectral et les valeurs propres extrêmes. Cependant, notre étude récente, présentée dans [73], explore la distribution spectrale globale de Φ, ainsi que sa DSE μ_{M_Φ}. Ceci fournit une approche complémentaire pour comprendre à la fois la masse et les valeurs propres extrêmes du spectre de Φ.

Théorème 0.24. *Soit $W_1, \ldots, W_d \in M_n(\mathbb{C})$ des matrices aléatoires hermitiennes (c'est-à-dire auto-adjointes) indépendantes telles que, pour chaque $i \in [d]$, $\mathbb{E}(W_i) = 0$ et $\mathbb{E}(W_i^2) =$ Id. Supposons en outre que, pour chaque $i \in [d]$, la DSE de W_i converge vers une certaine distribution μ_i lorsque $n \to \infty$ et que la distribution asymptotique conjointe de la famille $(W_i)_{i \in [d]}$ est déterminée par la famille $(\mu_i)_{i \in [d]}$ (voir la section 2.2). Définissons le canal quantique*

$$\Phi : X \in M_n(\mathbb{C}) \mapsto \frac{1}{\sqrt{d}} \sum_{i \in [d]} W_i X W_i^* \in M_n(\mathbb{C}).$$

. Alors, on a les résultats:

- *Si d est fixé, alors la distribution spectrale de $\Phi - \mathbb{E}(\Phi)$ converge vers une distribution spécifique, dépendant uniquement de $\mu_1, \ldots, \mu_d$, lorsque $n \to \infty$.*
- *Si $d = d(n) \to \infty$ lorsque $n \to \infty$ et W_i sont i.i.d, alors la distribution spectrale de*

$\Phi - \mathbb{E}(\Phi)$ *converge vers la loi du demi-cercle lorsque* $n \to \infty$.

Remarque 0.25. Il est important de noter que le facteur de normalisation de $1/\sqrt{d}$ dans les canaux quantiques provient naturellement d'un résultat de type théorème central limite. D'autre part, le facteur de normalisation précédent de $1/d$ dans la norme d'opérateur de (13) est similaire à celui utilisé dans la loi des grands nombres.

⋆ ⋆ ⋆

Dans le chapitre 3, nous discuterons des résultats d'un travail en commun en cours avec Cécilia Lancien et Pierre Youssef.

Dans le chapitre précédent, nous avons étudié les canaux quantiques

$$M_\Phi - \mathbb{E}\, M_\Phi = \frac{1}{\sqrt{d}} \sum_{i \in [d]} \Big(W_i \otimes \overline{W}_i - \mathbb{E}\, W_i \otimes \overline{W}_i \Big),$$

et nous avons prouvé que, lorsque d tend vers l'infini et que W_i est centré i.i.d, la DSE de $M_\Phi - \mathbb{E}\, M_\Phi$ converge vers la loi du demi-cercle. Dans ce chapitre, nous supprimons la condition de centrage et étudions plus généralement le problème de la recherche de la loi limite de

$$m - \tau(m) = \frac{1}{\sqrt{d}} \sum_{i \in [d]} (a_i \otimes a_i - \tau \otimes \tau(a_i \otimes a_i) 1),$$

où a_i sont des variables aléatoires non-commutatives. Nous commençons par quelques notations.

Soit $(\mathcal{A}, \tau)$ une algèbre non-commutative unitaire équipée d'un état linéaire tracial fidèle $\tau(ab) = \tau(ba)$ pour tout $a, b \in \mathcal{A}$ et d'une involution $*$, c'est-à-dire $(a^*)^* = a$, pour tout $a \in \mathcal{A}$. La fidélité signifie que $\tau(a^* a) \geq 0$ et l'égalité ne tient que si $a = 0$. Nous disons que les sous-algèbres $\mathcal{A}_1, \ldots, \mathcal{A}_n \subset \mathcal{A}$ sont libres si

$$\tau(a_1 \cdots a_k) = 0,$$

lorsque les conditions suivantes sont remplies.

1. $k \geq 0$;
2. $\tau(a_i) = 0$ pour tout $i \in [k]$;
3. $a_i \in \mathcal{A}_{j_i}$, où $j_i \in [n]$ pour tout $i \in [k]$;
4. Aucun a_i consécutif n'est dans la même algèbre, c'est-à-dire $j_1 \neq j_2, \ldots, j_{k-1} \neq j_k$.

Nous disons que les variables aléatoires $a_1, \ldots, a_n$ sont libres lorsque leurs algèbres le sont. La liberté est l'analogue de l'indépendance dans les espaces non-commutatifs, et les théorèmes limites pour les variables aléatoires indépendantes s'étendent naturellement

aux variables aléatoires libres. Pour l'illustrer, rappelons le théorème limite central libre [110], [85, Lecture 8]. Nous dénotons

$$\mathrm{var}(a) = \tau((a - \tau(a)1)^2),$$

la variance d'une variable aléatoire de moyenne $\tau(a)$, et $1 \in \mathcal{A}$ est l'unité dans $\mathcal{A}$.

Théorème 0.26. *Soit $a_1, \ldots, a_n \in \mathcal{A}$ des variables aléatoires centrées i.i.d. libres auto-adjointes ($a^* = a$) avec une variance unitaire. Soit*

$$S_n := \frac{1}{\sqrt{n}} \sum_{i \in [n]} a_i.$$

Alors S_n converge en distribution vers la loi du demi-cercle, c'est-à-dire que pour tout $p \geq 1$, on a

$$\tau(S_n^p) \to \tau(s^p),$$

où s est une variable aléatoire en demi-cercle.

Rappelons qu'une variable aléatoire est semi-circulaire lorsque ses moments impairs s'annulent, et pour tout $p \geq 1$, nous avons

$$\tau(s^{2p}) = C_p.$$

Sa loi μ_{sc} a pour densité

$$f(x) = \frac{1}{2\pi}\sqrt{4 - x^2}\mathbf{1}_{|x| \leq 2}.$$

Les auteurs de [37] ont étudié le produit tensoriel de variables aléatoires libres $b_k = a_k \otimes a_k$, qui a été inspiré par le produit tensoriel de matrices aléatoires pour les canaux quantiques. Ils ont découvert que dans l'espace produit $(\mathcal{A} \otimes \mathcal{A}, \tau \otimes \tau)$, de tels tenseurs ne sont généralement pas libres. En particulier, le théorème 0.26 ne s'applique pas aux variables b_k. Notre but est donc d'identifier cette distribution limite.

Soit $a, a_1, \ldots, a_n \in (\mathcal{A}, \tau)$ des variables aléatoires libres i.i.d. auto-adjointes. Pour éviter les variables aléatoires dégénérées, supposons $\mathrm{var}(a) > 0$. Sinon, $a_i = \tau(a)1$, et le résultat est trivial. Considérons

$$b_i = a_i \otimes a_i - \tau \otimes \tau(a_i \otimes a_i)1 = a_i \otimes a_i - \tau^2(a_i)1, \tag{14}$$

et

$$S_n := \frac{1}{\sqrt{n\,\mathrm{var}(b)}} \sum_{i \in [n]} b_i. \tag{15}$$

Notons tout d'abord que

$$\mathrm{var}(b) = \mathrm{var}(a)(\mathrm{var}(a) + 2\tau^2(a)),$$

donc,

$$q := \frac{2\tau^2(a)\,\mathrm{var}(a)}{\mathrm{var}(b)} \in [0,1]. \tag{16}$$

Notre théorème principal caractérise la distribution limite de S_n en fonction de q. Nous avons besoin de quelques notations. Nous notons $\boxplus$ la convolution libre et pour une mesure μ, nous notons $(t\mu)$ sa dilatation par $t > 0$ définie par

$$(t\mu)(A) := \mu(t^{-1}A),$$

pour tout ensemble de Borel $A \subset \mathbb{R}$. Pour $p \in \mathbb{N}$, nous notons $P(p)$ l'ensemble de toutes les partitions de $[p]$ et $P_2(p)$ l'ensemble de toutes les partitions $\pi \in P(p)$ telles que pour chaque bloc $V \in \pi$, $|V| = 2$, et dans ce cas π est appelé une partition pair. On dit que deux blocs distincts $V_1, V_2 \in \pi$ se croisent s'il existe $i < j < k < l$ tel que $\{i,k\} \subseteq V_1$ et $\{j,l\} \subseteq V_2$. Nous disons qu'un bloc V est croisé s'il existe un autre bloc V' tel que V et V' se croisent. Enfin, nous définissons le graphe d'intersection $G(\pi)$ comme suit. Les sommets de $G(\pi)$ sont les blocs de π, et il existe une arête entre les blocs V_1, V_2 s'ils se croisent. Nous notons $P^{\mathrm{bi}}(p)$ (resp. $P_2^{\mathrm{bi}}(p)$) l'ensemble de toutes les partitions (resp. partitions pairs) dont le graphe d'intersection est biparti. Nous définissons une mesure ν dont les moments impairs sont nuls et les moments pairs sont égaux à

$$\int_{\mathbb{R}} x^{2p}\,\mathrm{d}\nu = |P_2^{\mathrm{bi}}(2p)| \leq \mathbb{E}\,g^{2p},$$

où $g \sim N(0,1)$ est une variable aléatoire gaussienne standard dont les moments sont précisément le cardinal des partitions de paires. L'existence de la mesure ν est un problème des moments pour la suite $(P_2^{\mathrm{bi}}(2n))_{n\geq 1}$ [2, 92], et sera discutée dans le remarque 3.4.

Notre principal théorème est le suivant.

Théorème 0.27. *Soit $a, a_1, \ldots, a_n \in (\mathcal{A}, \tau)$ des variables aléatoires libres auto-adjointes avec une variance $\mathrm{var}(a) > 0$. Soit b_i et S_n comme dans (5), (6), respectivement, et*

$$q = \frac{2\tau^2(a)\,\mathrm{var}(a)}{\mathrm{var}(b)} \in [0,1].$$

Alors S_n converge en distribution vers la convolution libre pondérée de ν et la loi du demi-cercle

$$S_n \Rightarrow \sqrt{\frac{q}{2}}\nu \boxplus \sqrt{\frac{q}{2}}\nu \boxplus \sqrt{1-q}\,\mu_{sc}.$$

Partie II. Norme des matrices inhomogènes

Dans cette section, nous relâchons l'hypothèse d'une distribution identique pour prouver les théorèmes de limite et nous considérons des matrices aléatoires inhomogènes $W_n = (w_{ij})$ avec des entrées indépendantes et un profil de variance où $\mathbb{E}(w_{ij}^2) = \sigma_{ij}^2$. Nous nous concentrerons sur deux exemples. Premièrement, nous calculons et étudions la convergence de la norme d'une matrice symétrique inhomogène W_n. Deuxièmement, nous considérons des matrices asymétriques $d \times n$ W_n et étudions les bornes sur la norme d'opérateur de sa forme quadratique.

⋆ ⋆ ⋆

Dans le chapitre 4, nous présentons les principaux résultats de l'article "On spectral outliers of inhomogeneous symmetric random matrices" (https://arxiv.org/abs/2401.07852) [5], travail en commun avec Dylan J. Altschuler, Konstantin Tikhomirov, et Pierre Youssef.

Dans le chapitre 1, nous avons exploré la DSE des graphes d-réguliers G_n. Cependant, nous pouvons également considérer un graphe régulier pondéré $\tilde{G}_n$, dans lequel nous attribuons des variables aléatoires indépendantes et identiquement distribuées (i.i.d) ξ_e à chaque arête $e \in E(G_n)$. En d'autres termes, la matrice d'adjacence de $\tilde{G}_n$ est donnée par $X_n = A_{G_n} \circ W_n$, où A_{G_n} est la matrice d'adjacence de G_n, W_n est une matrice aléatoire symétrique avec des entrées i.i.d suivant la loi ξ, et $\circ$ désigne le produit de Hadamard (entrée par entrée). Ce modèle est connu sous le nom de matrice de Wigner parcimonieuse. Une question naturelle qui se pose est la suivante. "Quelle est la limite DSE de X_n lorsque $d = d(n)$ est une fonction de n ?"

Le célèbre théorème de Wigner 0.18 possède une caractéristique universelle similaire au théorème de la limite centrale. Cela signifie que la limite n'est pas affectée par la distribution des entrées mais uniquement par sa variance. L'ensemble de Wigner correspond au graphe complet G_n. Selon le théorème de Bai-Yin 0.19, pour la valeur propre extrême λ_1, cette universalité s'effondre dans un sens faible. En particulier, ils montrent que si le quatrième moment est fini, il n'y a pas de valeurs hors du spectre $[-2, 2]$ puisque

$$\|W_n\| \to 2 \text{ a. s.}$$

Sinon, la norme d'opérateur ne converge pas. L'universalité de ces deux résultats a fait l'objet d'une attention particulière pour d'autres modèles, en particulier pour ceux qui ne sont pas soumis à la condition i.i.d. [25, 81, 45, 47, 48]. Un modèle important est le cas parcimonieux, qui est utilisé dans l'analyse théorique, les simulations numériques, l'inférence statistique, la théorie des graphes et les matrices aléatoires [18, 19, 31]. Le modèle que nous allons étudier dans ce chapitre est le suivant.

Modèle. Soit W_n une matrice $n \times n$ de Wigner (symétrique) dont les entrées sur et au-dessus de la diagonale sont des copies i.i.d. d'une variable aléatoire centrée ξ ayant une

variance unitaire. Soit $\Sigma_n = (\sigma_{ij})$ une matrice symétrique $n \times n$ à entrées positives telle que

$$\sum_{i \in [n]} \sigma_{ij}^2 = 1,$$

pour tout $j \in [n]$. Nous considérons la dilatation $X_n = \Sigma_n \circ W_n$, où $\circ$ désigne le produit de Hadamard (c'est-à-dire entrée par entrée). Nous appelons X_n une dilatation de la matrice de Wigner W_n.

Il est naturel de se demander si l'universalité du spectre et des valeurs propres extrêmes se vérifie dans le cas inhomogène. Plus précisément, nous pouvons nous demander si la distribution spectrale limite de X_n et sa norme limite dépendent de ξ et Σ_n. La première question a été résolue avec succès dans un cadre plus général dans [55].

Théorème 0.28. *Soit $X_n = \Sigma_n \circ W_n$ une dilatation d'une matrice de Wigner $n \times n$ associée à une variable aléatoire centrée ξ de variance unitaire. Alors, on a:*

1. *Si*

$$\sigma_n^* = \max_{ij} \sigma_{ij} \to 0,$$

   la loi limite de μ_{X_n} est la loi du demi-cercle, quelles que soient la distribution de ξ et la structure de Σ_n.

2. *Sinon, σ_n^* ne converge pas vers zéro, la loi limite de la DSE de X_n (si elle existe) dépend à la fois de ξ et de la structure de Σ_n.*

En ce qui concerne l'extrémité du spectre, les résultats de [74] démontrent déjà que cette universalité n'existe pas. Afin de déterminer l'influence de la structure de Σ_n, nous allons restreindre notre cas aux variables aléatoires sous-gaussiennes. Rappelons qu'une variable aléatoire ξ est sous-gaussienne lorsque il existe $\sigma > 0$ telle que

$$\mathbb{E}\, e^{\lambda(\xi - \mathbb{E}\,\xi)} \leq e^{\sigma^2 \lambda^2 / 2},$$

pour tout $\lambda \in \mathbb{R}$. Dans le cas gaussien, le résultat suivant est démontré dans [19].

Théorème 0.29. *Supposons que $\xi \sim N(0,1)$ et que $X_n = \Sigma_n \circ W_n$ soit la dilatation de sa matrice de Wigner $n \times n$. Alors*

$$\mathbb{E}\|X_n\| \leq (1+\varepsilon)\Big(2 + C(\varepsilon)\sigma_n^* \sqrt{\log n}\Big),$$

pour tout $\varepsilon \in (0, 1/2)$.

L'idée principale pour prouver le théorème 0.29 est un argument de compression qui compare la matrice X_n à une matrice aléatoire symétrique standard $r \times r$ G_r avec des

entrées gaussiennes iid sur et au-dessus de la diagonale telle que

$$\mathbb{E}\,\mathrm{tr}(X_n^{2p}) \le \kappa\,\mathbb{E}\,\mathrm{tr}(G_r^{2p}).$$

Ici, κ et la dimension r sont ajustés de manière appropriée en fonction du profil de variance, voir [19]. Cette dernière est alors calculée par des estimations précises sur les processus gaussiens et la concentration gaussienne comme suit

$$\mathbb{E}\,\mathrm{tr}(G_r^{2p}) \le n\,\mathbb{E}\|G_r\|^{2p}.$$

Notons que la convergence de la distribution spectrale de X_n implique que

$$\liminf \mathbb{E}\|X_n\| \ge 2.$$

Par conséquent, le théorème 0.29 implique que lorsque $\sigma_n^*\sqrt{\log n}$ converge vers zéro, il n'y a pas de valeurs hors du spectre $[-2,2]$, c'est-à-dire que

$$\mathbb{E}\|X_n\| \to 2.$$

Un tel résultat a été étendu aux variables aléatoires bornées ξ en dimension [74], mais il restait à savoir si l'on pouvait l'étendre à toutes les variables aléatoires sous-gaussiennes. C'est effectivement le cas, comme le montre notre premier théorème principal.

Théorème 0.30. *Soit $X_n = \Sigma_n \circ W_n$ une dilatation d'une matrice de Wigner $n \times n$ W_n associée à une variable aléatoire sous-gaussienne centrée ξ de variance 1. Si $\sigma_n^*\sqrt{\log n}$ converge vers zéro, X_n n'a pas de valeurs hors du spectre $[-2,2]$, c'est-à-dire que $\|X_n\| \to 2$ presque sûrement.*

En se limitant aux matrices $\Sigma_n = \frac{1}{\sqrt{d}} A_{G_n}$, c'est-à-dire aux matrices d'adjacence normalisées des graphes d-réguliers G_n, le résultat précédent indique que lorsque $d \gg \log n$, il n'y a pas de valeurs hors du spectre $[-2,2]$. Seginer [93] a partiellement répondu à la question inverse.

Théorème 0.31. *Soit ξ une variable aléatoire de Rademacher et W_n sa matrice de Wigner. Il existe alors un graphe régulier G_n de degré $d = \lceil\sqrt{\log n}\rceil$ tel que*

$$\mathbb{E}\left\|\frac{1}{\sqrt{d}} A_{G_n} \circ W_n\right\| \ge c\log^{1/4} n.$$

Notre deuxième théorème principal lève la condition sous-optimale $d = \lceil\sqrt{\log n}\rceil$ jusqu'à la condition précise attendue $d \sim \log n$.

Théorème 0.32. *Soit ξ une variable aléatoire centrée de variance 1 et de quatrième moment borné. Soit W_n sa matrice de Wigner. Fixons une suite $(d_n)_{n\ge1}$ telle que $d_n = O(\log n)$, $d_n \to \infty$ et nd_n est pair. Alors, il existe une suite de graphes d_n-réguliers*

G_n telle que les matrices $\frac{1}{\sqrt{d_n}} A_{G_n} \circ W_n$ ont des valeurs hors du spectre $[-2, 2]$ presque sûrement. A savoir,

$$\mathbb{P}\left(\liminf_{n\to\infty} \left\| \frac{1}{\sqrt{d_n}} A_{G_n} \circ W_n \right\| > 2 \right) = 1.$$

⋆ ⋆ ⋆

Dans le chapitre 5, nous présentons les résultats de l'article "Almost sharp covariance and Wishart-type matrix estimation" (https://arxiv.org/abs/2307.09190) [91].

Rappelons que le théorème 0.29 donne

$$\mathbb{E}\|X\| \leq (1+\varepsilon)\left(2 + C(\varepsilon)\sigma^* \sqrt{\log n}\right),$$

où $X = X_n$ et $\sigma^* = \sigma_n^*$. Ici, nous n'indiquerons plus l'indice n puisque, dans ce chapitre, nous considérerons que les dimensions sont fixes. Nous désignerons cette fois par $X_1, \ldots, X_n$ les *colonnes* de la matrice X.

Une extension naturelle consiste à considérer des matrices asymétriques $d \times n$ X et des polynômes sur X. Soit $P \in \mathbb{R}[X, Y]$ un polynôme. Nous cherchons à étudier

$$c(P) := \mathbb{E}\left\| P(X, X^t) \right\|, \tag{17}$$

où X^t est la transposée de X. Un tel problème est lié au concept de liberté asymptotique forte et aux valeurs aberrantes de la DSE des matrices aléatoires ; voir, par exemple, [59, 38, 12, 14, 35]. Dans ce chapitre, nous examinerons une forme quadratique centrée

$$P(X, X^t) := XX^t - \mathbb{E}\, XX^t = \sum_{k\in[n]} P(X_k, X_k^t), \tag{18}$$

où X_k sont les colonnes vectorielles de X. C'est ce qu'on appelle le Problème d'Estimation de la Covariance (PEC). Remarquons que, si nous laissons $P_1(X, X^t) = X \otimes X^t$ et

$$M := \sum_{k\in[n]} P_1(M_k, M_k^t),$$

pour les matrices M_k, alors M serait un canal quantique étudié au chapitre 2.

Le PEC est un problème bien connu en théorie statistique et a été étudié de manière approfondie dans la littérature [79, 100, 116, 89, 69, 34, 33]. Il a également des liens avec la géométrie convexe, comme l'explique le chapitre 9 de [108], et les valeurs aberrantes du spectre des matrices aléatoires. Le problème peut être décrit comme suit. Étant donné un ensemble de vecteurs aléatoires indépendants et identiquement distribués $X_1, \ldots, X_n \in \mathbb{R}^d$, nous définissons sa matrice de covariance empirique comme étant

$$\Sigma_n = \frac{1}{n} \sum_{k\in[n]} X_k X_k^t.$$

Par la loi des grands nombres, on sait que Σ_n converge vers la vraie matrice de covariance $\Sigma = \mathbb{E}\, X_1 X_1^t$. L'objectif est de contrôler la vitesse de convergence. Plus précisément, nous cherchons à borner

$$\mathbb{E}\|\Sigma_n - \Sigma\| = \mathbb{E} \sup_{x \in S^{d-1}} \|(\Sigma_n - \Sigma)x\|_2.$$

Dans le cas homogène, on connaît le théorème suivant [108, Théorème 4.7.1].

Théorème 0.33. *Soit $G_1, \dots, G_n \in \mathbb{R}^d$ des variables aléatoires gaussiennes i.i.d $N(0, \mathrm{Id})$. Alors*

$$\mathbb{E}\|\Sigma_n - \Sigma\| \leq C \max\left\{\frac{d}{n}, \sqrt{\frac{d}{n}}\right\},$$

où C est une constante universelle.

Dans notre cas, les vecteurs $X_1, \dots, X_n \in \mathbb{R}^d$ sont indépendants, mais pas nécessairement identiquement distribués, et cela nécessite de bonnes bornes de concentration en l'absence de l'hypothèse iid. Des travaux récents [69, 33, 18, 31] ont levé cette hypothèse dans des contextes plus généraux.

En vertu du théorème de Bai-Yin, pour qu'il y ait convergence de la norme, il faut au moins l'existence du quatrième moment de la distribution X_{ij} de chaque vecteur $X_j \in \mathbb{R}^d$. Pour être concret, nous allons considérer le cas de vecteurs aléatoires gaussiens avec des entrées indépendantes. Soit $X = (X_{ij}) = (b_{ij} g_{ij})$ la matrice $d \times n$ dont les colonnes sont $X_j = (b_{ij} g_{ij})$, où $b_{ij} \geq 0$ sont des coefficients déterministes et g_{ij} des variables aléatoires gaussiennes standard i.i.d.. Notre objectif est alors d'évaluer

$$n\, \mathbb{E}\|\Sigma_n - \Sigma\| = \mathbb{E}\left\|XX^t - \mathbb{E}\, XX^t\right\|.$$

Afin d'introduire certains paramètres, considérons les bornes de $\mathbb{E}\|X\|$, comme nous l'avons fait au chapitre 4. Puisque la norme d'opérateur est plus grande que la norme de chaque colonne X_j, nous avons

$$\mathbb{E}\|X\| \geq \max_{j \in [n]} \mathbb{E}\|X_j\|_2 \geq c \max_{j \in [n]} (\mathbb{E}\|X_j\|^2)^{1/2},$$

où l'inégalité de Poincaré gaussienne implique la deuxième inégalité. En particulier, nous avons

$$\mathbb{E}\|X\| \geq c \max_{j \in [n]} \left(\sum_{i \in [d]} b_{ij}^2\right)^{1/2} =: \sigma_C,$$

où σ_C représente la norme euclidienne maximale des colonnes. De même,

$$\mathbb{E}\|X\| \geq c\sigma_R,$$

et σ_R est la norme euclidienne maximale des lignes. D'autre part, la norme d'opérateur est plus grande que le maximum des entrées ; par conséquent

$$\mathbb{E}\|X\| \geq \mathbb{E}\max_{ij}|b_{ij}g_{ij}|.$$

Si $B = (b_{ij})$ est suffisamment homogène, c'est-à-dire qu'une proportion polynomiale des coefficients b_{ij} est du même ordre que

$$\sigma_* = \max_{i,j} b_{ij},$$

alors nous pouvons immédiatement abaisser la borne inférieure de

$$\mathbb{E}\|X\| \geq c\sigma_*\sqrt{\log(n \wedge d)},$$

en utilisant le maximum de la gaussienne restreint à ces grandes coordonnées ; voir [19]. Dans ce cas, les paramètres $\sigma_*, \sigma_C, \sigma_R$ sont les seuls qui comptent. Par conséquent, aucun paramètre structurel n'est nécessaire. Définissons:

- $\tilde{\sigma}_\infty^2 = \max_{i,l:i\neq l}\sum_{j\in[n]} b_{ij}^2 b_{lj}^2$;
- $\bar{\sigma}_\infty^2 = \max_{i\in[d]}\sum_{j\in[n]} b_{ij}^4$;
- $\sigma_\infty^2 = \max_{i\in[d]}\sum_{j\in[n]}\sum_{l:l\neq i} b_{ij}^2 b_{lj}^2$;
- $\beta_\infty = \dfrac{\tilde{\sigma}_\infty\sigma_C}{\sigma_\infty\sigma_*}$.

Voici notre théorème principal, qui sera démontré au chapitre 5.

Théorème 0.34. *Soit X une matrice gaussienne $d \times n$ à entrées indépendantes telle que $X_{ij} = b_{ij}g_{ij}$ où $\{g_{ij} : (i,j) \in [d] \times [n]\}$ sont des v.a. gaussiens standard i.i.d. Soit $\varepsilon \in (0, 1/2)$. Dans ce cas, la règle suivante s'applique.*

1. *Si $\beta_\infty \leq 1$, nous avons*

$$\begin{aligned}\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| &= \mathbb{E}\left\|\sum_{j\in[n]} X_jX_j^T - \mathbb{E}\,X_jX_j^T\right\| \\ &\leq (1+\varepsilon)\left\{2\sigma_\infty + \sigma_C^2 + C(\varepsilon)\sigma_*\left(\sigma_C + \frac{\sigma_\infty}{\sigma_C}\right)\sqrt{\log(n\wedge d)} + C^2(\varepsilon)\sigma_*^2\log(n\wedge d)\right\}.\end{aligned}$$

2. *Sinon, $\beta_\infty > 1$ et nous avons*

$$\begin{aligned}&\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \\ &\leq (1+\varepsilon)\left\{\frac{2\tilde{\sigma}_\infty\sigma_C}{\sigma_*} + \sigma_C^2 + C(\varepsilon)(\sigma_C\sigma_* + \bar{\sigma}_\infty)\sqrt{\log(n\wedge d)} + C^2(\varepsilon)\sigma_*^2\log(n\wedge d)\right\}.\end{aligned}$$

La constante $C(\varepsilon)$ est

$$C(\varepsilon) \leq \frac{C}{\sqrt{\varepsilon}},$$

où C est une constante universelle.

Le théorème 0.34 améliore les travaux récents [33, Théorème 2.1] et [31, Théorème 3.17], [18, Théorème 3.12] lorsque $\beta_\infty \leq 1$. La condition sur β_∞ est un sous-produit de la preuve, et il n'est pas clair si elle est nécessaire. En particulier, le terme principal $2\sigma_\infty + \sigma_C^2$ semble être la borne correcte dans tous les cas, comme l'indiquent certains exemples très rares, la borne inférieure sur la norme et [18].

Part I

Spectral distribution of inhomogeneous matrices

Chapter 1

A combinatorial view on star moments of regular directed graphs and trees [41]

Did you hear about the guy who tells everyone what the colors on the graph mean?
That guy's a legend!

I am not this guy

We investigate the method of moments for d-regular digraphs and the limiting d-regular directed tree T_d as the number of vertices tends to infinity, in the same spirit as McKay [78] for the undirected setting. In particular, we provide a combinatorial derivation of the formula for the star moments (from a root vertex $o \in T_d$)

$$M_d(w) := \sum_{\substack{v_0,v_1\ldots,v_{k-1},v_k \in T_d \\ v_0=v_k=o}} A_{T_d}(v_0,v_1)\cdots A_{T_d}(v_{k-1},v_k).$$

with A_{T_d}, the adjacency matrix of T_d, where $w := w_1 \cdots w_k$ is any word on the alphabet $\{1, *\}$ and $A^*_{T_d}$ is the adjoint matrix of A_{T_d}. Our analysis highlights a connection between the non-zero summands of $M_d(w)$ and the non-crossing partitions of $\{1, \ldots, k\}$ which are in some sense compatible with w.

Contents

1.1 Introduction

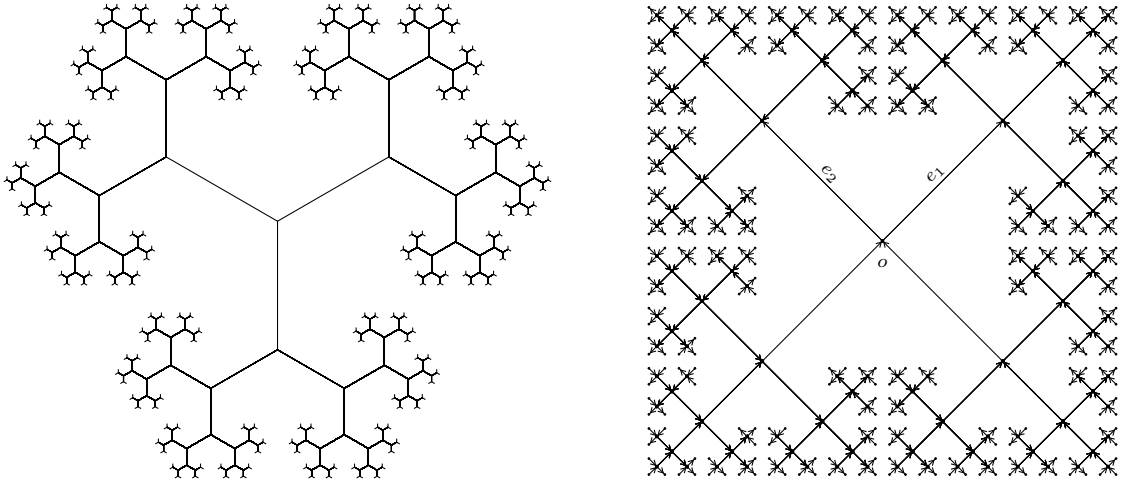
Counting paths in graphs and other discrete structures is a standard question with applications to many areas of mathematics, see [67, 23, 115, 113] or the book [26]. In random matrix theory, this question is typically raised when studying the convergence of empirical spectral distributions (ESDs) through the method of moments, of which Wigner's original proof of the semicircular law [111] is a renowned example. Essentially, for random Hermitian matrices $W_n := (W_n(i,j))_{1\leq i,j\leq n}$ whose coefficients on and above the diagonal are i.i.d. with mean 0 and variance 1 (so-called Wigner matrices), the different summands $\mathbb{E}W_n(i_1,i_2)\cdots W_n(i_k,i_1)$ occurring in the expansion of the states $\mathbb{E}\mathrm{Tr}\,W_n^k$, $k\geq 1$, can be related to certain cycles $i_1\to\cdots\to i_k\to i_1$ in a graph with vertex set $[n]:=\{1,\dots,n\}$, and understanding the combinatorics of these cycles helps to determine how each of those summands contributes to the k-th moment of the limiting spectral distribution (the semicircle distribution).

In contrast, when $A_{n,d}\in\{0,1\}^{n\times n}$ is the adjacency matrix of a uniformly sampled graph $G_{n,d}$ with n vertices and constant degree $d\geq 2$ (that is, $G_{n,d}$ is a uniform d-regular graph on $[n]$ and $A_{n,d}(i,j)=1$ if and only if $\{i,j\}$ is an edge in $G_{n,d}$), McKay [78] showed using the same method that the mean ESD $\mathbb{E}\frac{1}{n}\sum_{i=1}^n\delta_{\lambda_i(A_{n,d})}$ associated with $A_{n,d}$'s eigenvalues $\lambda_1(A_{n,d}),\dots,\lambda_n(A_{n,d})$ converges weakly (and in moments) towards a certain probability measure μ_{KM} which is now known as the Kesten–McKay distribution, in the sense that

$$\frac{1}{n}\mathbb{E}\sum_{i=1}^n f\big(\lambda_i(A_{n,d})\big)\xrightarrow[n\to\infty]{}\int f(x)\,\mu_{KM}(\mathrm{d}x) \tag{1.1}$$

holds for any polynomial or continuous bounded function $f:\mathbb{R}\to\mathbb{R}$. When $f(x):=x^k$ for some positive integer k, the left-hand side of (1.1), which can also be written $\frac{1}{n}\mathbb{E}\mathrm{Tr}\,A_{n,d}^k$, coincides with the expected total number of excursions of length k from a uniformly chosen vertex in $G_{n,d}$, while the right-hand side of (1.1) counts the number of such excursions (from a fixed vertex) in the (infinite) undirected d-regular tree UT_d, which is the Cayley graph of the free group with presentation $\langle e_1,\dots,e_d\mid e_i^2=1\rangle$, see Figure 1.1a. In fact, since the graphs $G_{n,d}$ converge locally to the tree UT_d as $n\to\infty$ (i.e., with respect to the Benjamini–Schramm topology [24]), the convergence (1.1) of their mean ESDs can be recovered from Bordenave–Lelarge's criterion [29].

In the present note, we adapt McKay's approach to the asymmetric (i.e., oriented) case. Although the local convergence as $n\to\infty$ of uniform d-regular digraphs $G_{d,n}$ towards the d-regular directed tree T_d does hold in a similar fashion (w.r.t. the "oriented" Benjamini–Schramm topology), it does not imply the convergence of ESDs anymore, and the analogue of (1.1) for oriented regular graphs is still an open question, known as the *oriented Kesten–McKay conjecture* [28]: the ESD of $G_{d,n}$ should converge towards a probability distribution on $\mathbb{C}$ corresponding in some sense to the spectral measure of T_d.



(a) (b)

Figure 1.1: (A) The undirected tree T_3 and (B) the directed tree T_2.

One difficulty for this conjecture is that because the adjacency matrix $A_{n,d}$ of $G_{n,d}$ is no longer Hermitian, the tracial moments $\mathbb{E}\mathrm{Tr}\, A_{n,d}^k$, $k \geq 0$, do not continuously determine the (now complex-valued) mean ESD of $A_{n,d}$. In fact, as $A_{n,d}$ is not even normal, neither do the star moments $\mathbb{E}\mathrm{Tr}\, A_{n,d}^w$ defined for any bit string $w := w_1 \cdots w_k$ on the alphabet $\Sigma := \{1, *\}$, where $A_{n,d}^w := A_{n,d}^{w_1} \cdots A_{n,d}^{w_k}$, and $A_{n,d}^*$ stands for the adjoint matrix of $A_{n,d} =: A_{n,d}^1$ (said differently, $A_{n,d}^*$ is the adjacency matrix of the graph $G_{d,n}^*$ obtained from $G_{d,n}$ by reversing each of its arcs). Nonetheless, investigating the star moments of regular digraphs remains interesting from a combinatorial perspective, and their convergence towards the corresponding star moments of the regular directed tree suggests that the conjecture holds.

By definition, the d-regular directed tree T_d is the unique infinite, connected, and acyclic digraph in which every vertex has constant in- and out-degree $d \geq 2$. In other words, T_d is the Cayley graph of the free group $F_d := \langle e_1, \ldots, e_d \rangle$ where unlike its symmetric version, the generators have no relations (see Figure 1.1b). We identify the vertex set of T_d with F_d, the root vertex $o \in T_d$ corresponding to the identity element, and we let A_{T_d} denote the adjacency matrix.

Theorem 1.1 (Convergence of star moments for uniform d-regular digraphs)**.** *For every $k \geq 0$ and every $w \in \Sigma^k$,*

$$\frac{1}{n}\mathbb{E}\mathrm{Tr}\, A_{n,d}^w \xrightarrow[n\to\infty]{} M_d(w) := A_{T_d}^w(o, o),$$

where $A_{T_d}^w(v, v')$ is defined for any pair of vertices $v, v' \in T_d$ by

$$A_{T_d}^w(v, v') := \sum_{\substack{v_0, v_1 \ldots, v_{k-1}, v_k \in T_d \\ v_0 = v,\ v_k = v'}} A_{T_d}^{w_1}(v_0, v_1) A_{T_d}^{w_2}(v_1, v_2) \cdots A_{T_d}^{w_k}(v_{k-1}, v_k).$$

Note that $A_{T_d}^w(v, v') = \mathbb{1}_{\{v=v'\}}$ if $w = \varnothing \in \Sigma^0$ is the empty word. In any case, all summands of $A_{T_d}(v, v')$ are either 0 or 1, and each non-zero summand corresponds to a

solution $(i_1, \dots, i_k) \in [d]^k$ to the word problem

$$v \cdot e_{i_1}^{w_1} \cdots e_{i_k}^{w_k} = v'$$

in the free group F_d, where e_i^* denotes the inverse of $e_i =: e_i^1$. We call such a solution $(i_1, \dots, i_k)$ a w-path from v to v', which we can also picture as

$$v =: v_0 \xrightarrow{w_1} v_1 \xrightarrow{w_2} \cdots \xrightarrow{w_{k-1}} v_{k-1} \xrightarrow{w_k} v_k := v',$$

where $v_j := v \cdot e_{i_1}^{w_1} \cdots e_{i_j}^{w_j}$ for all $0 \le j \le k$. In plain words, $M_d(w)$ is the cardinal of the set $P(w)$ of all w-paths from o to o (or from any other vertex to itself, by transitivity of the Cayley graph T_d). We stress that we do not consider any randomness on T_d: in this respect, our purpose is different from Kesten [67], who studied spectral properties of random walks on the undirected regular tree T_d.

In fact, Theorem 1.1 is a consequence of the following general criterion, similar to [78, Theorem 1.1]: under a growth assumption on the number of short cycles, we show that the star moments of deterministic d-regular digraphs converge to the star moments of the d-regular directed tree T_d.

Theorem 1.2 (Convergence of star moments for deterministic d-regular digraphs)**.** *Let G_n, $n \ge 1$, be a d-regular digraph with adjacency matrix A_n on a vertex set V_n. Let $k \ge 1$ and suppose that for every $j \in [k]$, the number $c_j(G_n)$ of cycles with length j in G_n (see (1.4)) fulfills*

$$\frac{c_j(G_n)}{|V_n|} \xrightarrow[n\to\infty]{} 0. \tag{1.2}$$

Then for every word $w \in \Sigma^k$,

$$\frac{1}{|V_n|} \operatorname{Tr} A_n^w \xrightarrow[n\to\infty]{} M_d(w). \tag{1.3}$$

Our last result is a combinatorial derivation of a formula for $M_d(w)$, which requires some notation. Recall that a partition π of $[k]$ can also be seen as the equivalence relation $\sim_\pi$ on $[k]$ such that $i \sim_\pi j \iff \exists V \in \pi,\ \{i, j\} \subseteq V$ for all $i, j \in [k]$. We say that π is *non-crossing*, written $\pi \in \mathrm{NC}(k)$, if $i_1 \sim_\pi j_1,\ i_2 \sim_\pi j_2 \implies j_1 \sim_\pi i_2$ for all $1 \le i_1 < i_2 < j_1 < j_2 \le k$. The cardinal $|\mathrm{NC}(k)|$ is equal to the ubiquitous Catalan number $C_k := \frac{1}{k+1}\binom{2k}{k}$, which is also [99] the cardinal $|\mathrm{NC}_2(2k)|$ of the non-crossing *pair* partitions of $[2k]$ (where each block has size 2). We further say that π is an *alternating non-crossing partition* of w ($\pi \in \mathrm{ANC}(w)$) if for every block $V := \{i_1 < \ldots < i_m\} \in \pi$, the subword $w|_V := w_{i_1} \cdots w_{i_m}$ of w is *alternating*, that is either of the form $w|_V = 1* \cdots 1*$ or $w|_V = *1 \cdots *1$ (so w and all blocks of π must have even size).

Theorem 1.3 (Combinatorial formula for $M_d(w)$)**.** *For every $k \geq 0$ and every $w \in \Sigma^k$,*

$$M_d(w) \quad = \sum_{\pi \in \mathrm{ANC}(w)} \left(\prod_{V \in \pi} (-1)^{\frac{|V|}{2}-1} C_{\frac{|V|}{2}-1} \right) d^{|\pi|}.$$

Notably, our proof shows that the w-paths may be counted by an inclusion-exclusion principle involving non-crossing pair partitions, thus explaining the presence of signs and Catalan numbers.

We mention that Theorems 1.1 and 1.3 may be recovered by taking a detour to free probability from a theorem of Nica [82], see Section 1.3, where we also put the oriented Kesten–McKay conjecture in more context. Our main motivation for this work is to provide direct combinatorial proofs, which we do in Section 1.2.

Acknowledgments. We would like to thank Pierre Youssef for suggesting to work on this problem and Charles Bordenave for informing us about [82].

1.2 Direct combinatorial proofs

1.2.1 Convergence of star moments.

In this section, we prove Theorem 1.2 and its corollary, Theorem 1.1. Let G be a multigraph with adjacency matrix A and vertex set V. We call a sequence of $j \geq 1$ distinct arcs $\varepsilon_1, \ldots, \varepsilon_j$ (read in any cyclic order) a *plain cycle of length j in G* if each of the pairs $\{\varepsilon_1, \varepsilon_2\}, \ldots, \{\varepsilon_{j-1}, \varepsilon_j\}, \{\varepsilon_j, \varepsilon_1\}$ has a common vertex (we disregard the arc orientations). Discounting the cyclic orderings, the number of plain cycles with length j in G is then given by

$$c_j(G) := \frac{1}{2j} \sum_{w \in \Sigma^j} \sum_{\mathbf{v}} A^{w_1}(v_0, v_1) \cdots A^{w_j}(v_{j-1}, v_j), \tag{1.4}$$

the second summation ranges over every $\mathbf{v} := (v_0, \ldots, v_{j-1}, v_j = v_0) \in V^{j+1}$ such that the sequence $\left((v_{i-1}, v_i)^{w_i}\right)_{1 \leq i \leq j}$ is injective, where $(v, v')^1 := (v, v')$ and $(v, v')^* := (v', v)$ for all $v, v' \in V$.

Proof of Theorem 1.2. Write $v \in \mathcal{C}_{n,k}$ if there exists a vertex $v' \in V_n$ at distance at most k from v (i.e., $A_n^{w'}(v, v') > 0$ for some $w' \in \Sigma^j$, with $j \leq k$) and belonging to a cycle of length at most k in G_n. Note that, by union bound,

$$|\mathcal{C}_{n,k}| \leq \sum_{i=1}^{k} (2d)^i \sum_{j=1}^{k} j\, c_j(G_n) \leq k^2 (2d)^k \sum_{j=1}^{k} c_j(G_n).$$

Thus, on the one hand,

$$\frac{1}{|V_n|} \sum_{v \in \mathcal{C}_{n,k}} A_n^w(v, v) \leq k^2 (2d^2)^k \, \frac{\sum_{j=1}^{k} c_j(G_n)}{|V_n|} \xrightarrow[n \to \infty]{} 0, \tag{1.5}$$

using the trivial upper bound $A_n^w(v,v) \leq d^k$ and (1.2). On the other hand, for $v \notin \mathcal{C}_{n,k}$, no vertex accessible in at most k steps from v belongs to a cycle of length at most k. Since G_n is d-regular, the ball $B_{G_n}(v,k)$ of radius k around v must then look exactly like the ball $B_{T_d}(o,k)$. In particular, $A_n^w(v,v) = A_{T_d}^w(o,o)$ for all $v \in V_n \setminus \mathcal{C}_{n,k}$, and thus

$$\frac{1}{|V_n|} \sum_{v \notin \mathcal{C}_{n,k}} A_n^w(v,v) = \frac{|V_n| - |\mathcal{C}_{n,k}|}{|V_n|} A_{T_d}^w(o,o) \xrightarrow[n\to\infty]{} M_d(w). \tag{1.6}$$

Adding (1.5) and (1.6) then shows as stated that

$$\frac{1}{|V_n|} \operatorname{Tr} A_n^w = \frac{1}{|V_n|} \sum_{v \in \mathcal{C}_{n,k}} A_n^w(v,v) + \frac{1}{|V_n|} \sum_{v \notin \mathcal{C}_{n,k}} A_n^w(v,v) \xrightarrow[n\to\infty]{} M_d(w). \qquad \square$$

Remark 1.4. As we can see from the proof, the condition (1.2) implies more generally that $G_n \to T_d$ with respect to the "oriented" Benjamini–Schramm topology: for every $k \geq 1$, the balls of radius k in G_n are eventually isomorphic to the ball of radius k in T_d. As such, Theorem 1.2 constitutes the non-symmetric version of [1, Proposition 14].

Next, we show that the growth condition (1.2) of Theorem 1.2 holds in expectation for the uniform d-regular digraph $G_{d,n}$.

Lemma 1.5 ($G_{d,n}$ has few short cycles on average)**.** *For every $k \geq 1$,*

$$\frac{1}{n} \mathbb{E} c_k(G_{d,n}) \xrightarrow[n\to\infty]{} 0.$$

Proof. To estimate this expectation, it is convenient to work with the so-called *configuration model* $\mathbf{CM}_{d,n}$, whose construction we briefly recall. First, to each vertex $i \in [n]$ we attach d unique incoming half-arcs $\varepsilon^+_{i+(p-1)n}$, $p \in [d]$, and another d unique outgoing half-arcs $\varepsilon^-_{i+(q-1)n}$, $q \in [d]$. Second, we choose uniformly at random a bijection $f_{d,n}$ joining each of the nd outgoing half-arcs to one of the nd incoming half-arcs (so there are $(nd)!$ possible choices for the bijection $f_{d,n}$). This gives rise to a random multigraph $\mathbf{CM}_{d,n}$ on $[n]$ in which the number of arcs $\mathbf{A}_{d,n}(i,j)$ from i to j equals the number of pairs $(p,q) \in [d]^2$ such that $f\big(\varepsilon^+_{i+(p-1)n}\big) = \varepsilon^-_{j+(q-1)n}$. Also, the distribution of $\mathbf{CM}_{d,n}$ conditional on the event

$$\mathcal{S}_{n,d} : \text{“}\mathbf{CM}_{d,n} \text{ is simple”} = \Big\{ \mathbf{A}_{d,n}(i,i) = 0 \text{ and } \mathbf{A}_{d,n}(i,j) \leq 1 \text{ for all } i \neq j \in [n] \Big\}$$

coincides with the law of $G_{n,d}$, that i,s $\mathcal{L}(G_{n,d}) = \mathcal{L}(\mathbf{CM}_{d,n} | \mathcal{S}_{n,d})$ Now, the expected number $\mathbb{E} c_k\big(\mathbf{CM}_{d,n}\big)$ of cycles with length k is easy to estimate from (1.4):

$$\begin{aligned}
\mathbb{E} c_k(\mathbf{CM}_{d,n}) &\leq \frac{1}{2k} \cdot 2^k \cdot n^k \cdot d^k \mathbb{P}\Big(f_{d,n}(\varepsilon_1^+) = \varepsilon_2^-, \dots, f_{d,n}(\varepsilon_{k-1}^+) = \varepsilon_k^-, f_{d,n}(\varepsilon_k^+) = \varepsilon_1^- \Big) \\
&= \frac{(2nd)^k \, (nd-k)!}{2k \, (nd)!} \\
&\sim \frac{2^{k-1}}{k},
\end{aligned}$$

by Stirling's formula. Furthermore, the probability $\mathbb{P}(\mathcal{S}_{n,d})$ of $\mathbf{CM}_{d,n}$ being simple was computed in [64] and is known [40] to be bounded away from zero as $n \to \infty$. Hence

$$\frac{1}{n}\mathbb{E}c_k(G_{d,n}) = \frac{1}{n}\mathbb{E}[c_k(\mathbf{CM}_{d,n}) \mid \mathcal{S}_{n,d}] \leq \frac{\mathbb{E}c_k(\mathbf{CM}_{d,n})}{n\mathbb{P}(\mathcal{S}_{n,d})} \xrightarrow[n\to\infty]{} 0,$$

which concludes the proof. □

We are now ready to prove Theorem 1.1.

Proof of Theorem 1.1. Let $(n_p)_{p\geq 1}$ be an increasing sequence of integers tending to ∞. By Lemma 1.5, the convergence

$$\frac{1}{n_p}\, c_k\big(G_{d,n_p}\big) \xrightarrow[p\to\infty]{} 0$$

holds in expectation, and thus also in probability. A classical application of the Borel–Cantelli lemma shows that it further holds almost surely along a subsequence: there exists $(n'_p)_{p\geq 1} \subseteq (n_p)_{p\geq 1}$ such that

$$\frac{1}{n'_p}\, c_k\big(G_{d,n'_p}\big) \xrightarrow[p\to\infty]{} 0$$

almost surely. Then Theorem 1.2 entails that

$$\frac{1}{n'_p}\operatorname{Tr} A^w_{n'_p,d} \xrightarrow[p\to\infty]{} M_d(w)$$

holds almost surely. Since $\operatorname{Tr} A^w_{n'_p,d} \leq n'_p\, d^k$, the dominated convergence theorem then yields

$$\frac{1}{n'_p}\mathbb{E}\operatorname{Tr} A^w_{n'_p,d} \xrightarrow[p\to\infty]{} M_d(w).$$

We have just shown that every subsequence of $\frac{1}{n}\mathbb{E}\operatorname{Tr} A^w_{n,d}$, $n \geq 1$, admits a further subsequence converging to $M_d(w)$, so Theorem 1.1 is proved. □

1.2.2 Combinatorial formula for $M_d(w)$.

Before establishing Theorem 1.3, let us warm up with a simple necessary condition for the set $P(w)$ of w-paths from o to o to be non-empty.

Lemma 1.6. *If $w := w_1 \cdots w_k$ is a word on Σ such that $P(w) \neq \emptyset$, then w is balanced:*

$$|\{i : w_i = 1\}| = |\{i : w_i = *\}|.$$

Proof. We proceed by induction on k. The lemma holds trivially if $k = 0$. Suppose $k \neq 0$.

By assumption, there exists a w-path $\mathbf{p} := (i_1, \dots, i_k)$ from $v_0 := o$ to itself:

$$\mathbf{p} \;=\; v_0 \xrightarrow{w_1} v_1 \xrightarrow{w_2} \cdots \xrightarrow{w_k} v_k = v_0,$$

with $v_j := e_{i_1}^{w_1} \cdots e_{i_j}^{w_j}$, $0 \le j \le k$, where we recall that the generators $e_1, \dots, e_d$ have no relations. In particular $v_1 \neq v_0$ (because $e_{i_1} \neq o$), and thus $k \ge 2$. Let $2 \le r \le k$ be the smallest index for which $v_r = v_0$, so

$$o = v_0 = v_r := e_{i_1}^{w_1}\big(e_{i_2}^{w_2} \cdots e_{i_{r-1}}^{w_{r-1}}\big)e_{i_r}^{w_r}, \qquad \text{that is,} \quad e_{i_r}^{w_r} e_{i_1}^{w_1} = \big(e_{i_2}^{w_2} \cdots e_{i_{r-1}}^{w_{r-1}}\big)^{-1}.$$

Since the generators have no relations, this forces $i_1 = i_r$ and $w_1 \neq w_r$ (in other words, T_d has no cycle, so the arc taken in $v_0 \xrightarrow{w_1} v_1$ must match the one in $v_{r-1} \xrightarrow{w_r} v_r = v_0$). Thus w is of the form $w = 1u{*}v$ or $w = {*}u1v$ where $u = w_2 \cdots w_{r-1}$ and $v := w_{r+1} \cdots v_k$, and

$$\mathbf{p} \;=\; v_0 \xrightarrow{w_1} \underbrace{v_1 \xrightarrow{w_2} \cdots \xrightarrow{w_{r-1}} v_{r-1}}_{u\text{-path}} \xrightarrow{w_r} v_r = \underbrace{v_0 \xrightarrow{w_{r+1}} v_{r+1} \xrightarrow{w_{r+2}} \cdots \xrightarrow{w_k} v_k}_{v\text{-path}} = v_0. \tag{1.7}$$

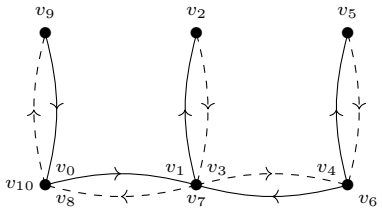
By induction, the smaller words u and v are balanced, and thus w is also balanced. □

A consequence of Lemma 1.6 is that $M_d(w) = 0$ if w is not balanced, so Theorem 1.3 is proved for such a word since then $\mathrm{ANC}(w) = \emptyset$. Note also that Theorem 1.3 holds if w is the empty word $\varnothing$, because $\mathrm{ANC}(\varnothing) := \{\emptyset\}$ is reduced to the empty partition and $P(\varnothing) := \{\emptyset\}$ is reduced to the empty path. We henceforth assume $w := w_1 \cdots w_{2p}$ non-empty and balanced. The decomposition (1.7) of a w-path (from o to o) with respect to its first return time to the origin is clearly unambiguous. Putting aside the choice of vertices along the path, this gives rise to a "skeleton" which, as we now claim, can be encoded as a certain partition $\pi \in \mathrm{ANC}(w)$ whose every block $V \in \pi$ has cardinal $|V| = 2$; we write $\pi \in \mathrm{ANC}_2(w)$ and call it an *alternating non-crossing pair partition* of w. Specifically, let $\mathbf{p} := (i_1, \dots, i_{2p}) \in P(w)$ and denote by $r \in \{2, \dots, 2p\}$ its first return time to o that is $v_j := e_{i_1}^{w_1} \cdots e_{i_j}^{w_j} \neq o$ for every $1 \le j < r$, and $v_r = o$. Then the skeleton of $\mathbf{p}$ is defined inductively as

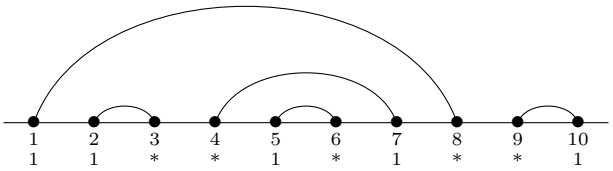
$$\sigma_w(\mathbf{p}) := \Big\{\{1, r\}\Big\} \cup \Big\{V + 1 : V \in \sigma_u(i_2, \dots, i_{r-1})\Big\} \cup \Big\{V + r : V \in \sigma_v(i_{r+1}, \dots, i_{2p})\Big\},$$

where $u := w_2 \cdots w_{r-1}$ and $v := w_{r+1} \cdots w_{2p}$, with the base case $\sigma_\varnothing(\emptyset) := \emptyset$ for the unique $\varnothing$-path $\emptyset$. Essentially, a block $V := \{j < k\}$ in $\sigma_w(\mathbf{p})$ means that $v_j, \dots, v_{k-1} \neq v_{j-1} = v_k$, i.e., k is the first return time to the vertex visited at time $j - 1$.

Conversely, given an alternating non-crossing *pair* partition $\pi \in \mathrm{ANC}_2(w)$ of w, what is $\sigma_w^{-1}\{\pi\}$, the subset of w-paths $\mathbf{p} := (i_1, \dots, i_{2p}) \in P(w)$ with skeleton $\sigma_w(\mathbf{p}) = \pi$? Clearly, since each block $V := \{j < k\} \in \pi$ indicates a segment of the path where it exits and first returns to the vertex v_{j-1}, we must have $i_j = i_k$, i.e., the arc taken at time j to exit v_{j-1} must be taken again at time k (but "backwards", since $w_j \neq w_k$) in order to



(a) $\mathbf{p} = (i, j, j, k, \ell, \ell, k, i, m, m)$, $k \notin \{i, \ell\}$.
(Arcs in T_d^* are drawn with a dashed line.)



(b) $\pi = \{\{1,8\},\{2,3\},\{4,7\},\{5,6\},\{9,10\}\}$, with $\lhd_\pi = \{(1,4),(4,5)\}$ and $B(\pi) = \{\{4,7\},\{5,6\}\}$.

Figure 1.2: (A) A w-path $\mathbf{p}$ and (B) its skeleton $\pi := \sigma_w(\mathbf{p})$, for $w := 11{*}{*}1{*}1{*}{*}1$.

return to v_{j-1}. This condition alone does not prevent a premature return $v_{j'} = v_{j-1}$ for some $j' \in \{j+1, \dots, k-1\}$. A premature return at time j' can, however, only happen if

(i) $w_{j'} \neq w_j$ (the arc at time j' must be taken in the opposite direction as when exiting v_{j-1}), and

(ii) j' is the lower element in its block, $U := \{j' < k'\} \in \pi$, which is directly surrounded by V: $j < j' < k' < k$ and there is no other block $\{j'' < k''\} \in \pi$ with $j < j'' < j' < k' < k'' < k$.

In case (i) and (ii) hold, we write $j \lhd_\pi j'$ as well as $U \in B(\pi)$, and say that j, j' form a *bad pair* and that U is a *bad block*. See Figure 1.2 for an illustration. Summarizing, for a w-path $(i_1, \dots, i_{2p})$ to have skeleton π, we must have $i_j \neq i_{j'}$ if j, j' form a bad pair (i.e., $j \lhd_\pi j'$), and $i_j = i_{j'}$ if j, j' belong to the same block ($j \sim_\pi j'$). It should be clear that these requirements are also sufficient:

Lemma 1.7. *The map $\sigma_w : P(w) \to \mathrm{ANC}_2(w)$ is surjective: for every $\pi \in \mathrm{ANC}_2(w)$,*

$$\sigma_w^{-1}\{\pi\} = \left\{ (i_1, \dots, i_{2p}) \in [d]^{2p} \;\middle|\; \forall (j, j') \in [2p]^2,\ \begin{cases} j \sim_\pi j' & \implies i_j = i_{j'} \\ j \lhd_\pi j' & \implies i_j \neq i_{j'} \end{cases} \right\}. \tag{1.8}$$

Furthermore,

$$\left|\sigma_w^{-1}\{\pi\}\right| = \prod_{V \in \pi} \left(d - \mathbb{1}_{\{V \in B(\pi)\}}\right). \tag{1.9}$$

Proof. First, the expression given for the cardinal (1.9) is always positive because $d \geq 2$, and is easily derived from (1.8): for each block $V := \{j < k\} \in \pi$, there are d degrees of freedom for the choice of $i_j = i_k \in [d]$, except if V is a bad block, in which case there is one degree of freedom less (because $i_j = i_k$ must be different from $i_{j'}$, where $j \lhd_\pi j'$). It remains to prove (1.8), which we do by induction on the balanced word $w := w_1 \cdots w_{2p}$. There is nothing to prove if $p = 0$. Suppose $p \geq 1$ and consider the decomposition of π with respect to the block containing 1,

$$\pi := \Big\{\{1, r\}\Big\} \cup \Big\{V + 1 : V \in \pi^{(u)}\Big\} \cup \Big\{V + r : V \in \pi^{(v)}\Big\},$$

where $u := w_2 \cdots w_{r-1}$, $v := w_{r+1} \cdots w_{2p}$, and

$$\pi^{(u)} := \Big\{V - 1 : V \in \pi,\ V \subseteq \{2, \dots, r-1\}\Big\} \ \in \mathrm{ANC}_2(u),$$
$$\pi^{(v)} := \Big\{V - r : V \in \pi,\ V \subseteq \{r+1, \dots, 2p\}\Big\} \in \mathrm{ANC}_2(v).$$

For $j \in [r-2]$, write $j \in J$ if $w_{j+1} \neq w_1$ and the block $\{j < k\} \in \pi^{(u)}$ containing j in $\pi^{(u)}$ is not surrounded by any other block (i.e., there is no $\{j' < k'\} \in \pi^{(u)}$ with $j' < j < k < k'$). Because of the previous decomposition and the definition of $\lhd_\pi$, we then have

$$\lhd_\pi = \Big\{(1, j+1) : j \in J\Big\} \cup \Big\{(j+1, j'+1) : j \lhd_{\pi^{(u)}} j'\Big\} \cup \Big\{(j+r, j'+r) : j \lhd_{\pi^{(v)}} j'\Big\}.$$

Recall also that $\{1, r\} \in \pi$ indicates that the w-paths with skeleton π first return to o at time r. Thus

$$(i_1, \dots, i_{2p}) \in \sigma_w^{-1}\{\pi\} \iff \begin{cases} i_1 = i_r,\ \forall j \in J,\ i_{j+1} \neq i_1, \\ (i_2, \dots, i_{r-1}) \in \sigma_u^{-1}\Big(\pi^{(u)}\Big),\ (i_{r+1}, \dots, i_{2p}) \in \sigma_v^{-1}\Big(\pi^{(v)}\Big), \end{cases}$$

and, by the induction hypothesis,

$$(i_1, \dots, i_{2p}) \in \sigma_w^{-1}\{\pi\} \iff \begin{cases} i_1 = i_r,\ \forall j \in J,\ i_{j+1} \neq i_1, \\ \forall (j, j') \in \{2, \dots, r-1\}^2,\ \begin{cases} (j-1) \sim_{\pi^{(u)}} (j'-1) \implies i_j = i_{j'}, \\ (j-1) \lhd_{\pi^{(u)}} (j'-1) \implies i_j \neq i_{j'}, \end{cases} \\ \forall (j, j') \in \{r+1, \dots, 2p\}^2,\ \begin{cases} (j-r) \sim_{\pi^{(v)}} (j'-r) \implies i_j = i_{j'}, \\ (j-r) \lhd_{\pi^{(v)}} (j'-r) \implies i_j \neq i_{j'}, \end{cases} \end{cases}$$
$$\iff \forall (j, j') \in [2p]^2,\ \begin{cases} j \sim_\pi j' \implies i_j = i_{j'}, \\ j \lhd_\pi j' \implies i_j \neq i_{j'}. \end{cases} \qquad \square$$

We can now complete the proof of Theorem 1.3.

Proof of Theorem 1.3. Let $w := w_1 \cdots w_{2p}$ be a balanced word on Σ. It follows from Lemma 1.7 that the set $P(w)$ of w-paths from o to o may be partitioned with respect to their skeleton as

$$P(w) \quad = \bigsqcup_{\pi \in \mathrm{ANC}_2(w)} \sigma_w^{-1}\{\pi\},$$

and passing to the cardinal, we get

$$M_d(w) \quad = \sum_{\pi \in \mathrm{ANC}_2(w)} \prod_{V \in \pi} \Big(d - \mathbb{1}_{\{V \in B(\pi)\}}\Big) \quad = \sum_{\pi \in \mathrm{ANC}_2(w)} \sum_{A \subseteq B(\pi)} (-1)^{|A|}\, d^{|\pi| - |A|},$$

by expanding out the product[1]. Now, given $\pi \in \mathrm{ANC}_2(w)$ and $A \subseteq B(\pi)$, we construct a coarser partition $\pi' := \gamma(\pi, A)$ from π by merging, for each bad pair $(j, j') \in A$, the block containing j' into its surrounding block (the one containing j). In other words, $\sim_{\pi'}$ is the smallest equivalence relation on $[2p]$ containing $\sim_\pi \cup A$. For instance, if π is the partition of Figure 1.2b and $A := \{(4,5)\}$, then $\pi' := \gamma(\pi, A) = \big\{\{1,8\},\{2,3\},\{4,5,6,7\},\{9,10\}\big\}$. It is clear that the conditions (i) and (ii) of forming a bad pair guarantee that π' remains non-crossing and alternating w.r.t. w: $\pi' \in \mathrm{ANC}(w)$. Further, π' has exactly $|A|$ fewer blocks than π, which has p blocks, so $(-1)^{|A|}\, d^{|\pi|-|A|} = (-1)^{p-|\pi'|}\, d^{|\pi'|}$. Conversely, given $\pi' \in \mathrm{ANC}(w)$, any pair partition π which is *finer* than π' (i.e., $\sim_\pi \subseteq \sim_{\pi'}$) automatically leads to an alternating non-crossing pair partition $\pi \in \mathrm{ANC}_2(w)$ of w having a certain set of bad pairs. Therefore,

$$M_d(w) \quad = \sum_{\pi' \in \mathrm{ANC}(w)} (-1)^{p-|\pi'|}\, d^{|\pi'|} \sum_{\substack{\pi \in \mathrm{ANC}_2(w) \\ \pi \preceq \pi'}} \sum_{\substack{A \subseteq B(\pi) \\ \gamma(\pi, A) = \pi'}} 1,$$

where we wrote $\pi \preceq \pi'$ for $\sim_\pi \subseteq \sim_{\pi'}$. Since

$$(-1)^{p-|\pi'|} = \prod_{V \in \pi'} (-1)^{\frac{|V|}{2}-1},$$

it remains to observe that

$$\sum_{\substack{\pi \in \mathrm{ANC}_2(w) \\ \pi \preceq \pi'}} \sum_{\substack{A \subseteq B(\pi) \\ \gamma(\pi, A) = \pi'}} 1 \quad = \quad \prod_{V \in \pi'} C_{\frac{|V|}{2}-1} \tag{1.10}$$

to conclude. But constructing $\pi \in \mathrm{ANC}_2(w)$ such that $\pi \preceq \pi'$ and $\gamma(\pi, A) = \pi'$ for some $A \subseteq B(\pi)$ is equivalent to partitioning each block $V := \{i_1, \ldots, i_{2m}\} \in \pi'$ using an alternating pair partition of $w|_V$ containing the block $\{1, 2m\}$. Since $w|_V$ is already alternating (because $\pi' \in \mathrm{ANC}(w)$), this amounts to choosing a non-crossing pair partition of $\{2, \ldots, 2m-1\}$, i.e., an element of $\mathrm{NC}_2(2m-2)$. Then (1.10) follows from the well-known fact $|\mathrm{NC}_2(2m-2)| = C_{m-1}$, see [99, Exercise 61]. □

[1] At the level of sets, this amounts to writing

$$\sigma_w^{-1}\{\pi\} = \Big\{(i_1, \ldots, i_{2p}) \in [d]^{2p} : \forall (j, j') \in [d]^{2p},\ j \sim_\pi j' \implies i_j = i_{j'}\Big\} \setminus \bigcup_{j \lhd_\pi j'} \Big\{(i_1, \ldots, i_{2p}) \in [d]^{2p} : i_j = i_j'\Big\}$$

and using the inclusion-exclusion formula.

1.3 Free probability and the oriented Kesten–McKay conjecture

1.3.1 Free probability.

Let us start this concluding section by showing how Theorems 1.1 and 1.3 can be recovered from Nica's work [82]. Free probability is a vast field initiated by Voiculescu; we only introduce the bare minimum, and refer to [85] for detail. The general framework is that of non-commutative variables $x, y, \ldots$ in some unital algebra $\mathcal{A}$ endowed with an adjoint operator * and a linear form $\varphi : \mathcal{A} \to \mathbb{C}$ such that $\varphi(1) = 1$ and $\varphi(x^*x) \geq 0$ for all $x \in \mathcal{A}$. The pair $(\mathcal{A}, \varphi)$ is called a *non-commutative probability space*, where the *state* φ plays the rôle of an expectation. The *distribution* of x or, more generally, the (joint) distribution of $(x_1, \ldots, x_k)$ is given by all mixed moments $\varphi(x_{i_1} \cdots x_{i_\ell})$ for $\ell \geq 1$ and $(i_1, \ldots, i_\ell) \in [k]^\ell$, which themselves may be expressed through the moment-cumulant formula [85, Lecture 11]:

$$\varphi(x_1 \cdots x_k) \;=\; \sum_{\pi \in \mathrm{NC}(k)} \prod_{\substack{V \in \pi \\ V := \{i_1 < \cdots < i_\ell\}}} \kappa_\ell(x_{i_1}, \ldots, x_{i_\ell}), \tag{1.11}$$

where the *free cumulants* $\kappa_\ell : \mathcal{A}^\ell \to \mathbb{C}$, $\ell \geq 1$, are defined inductively so that (1.11) holds for any $k \geq 1$ and any non-commutative variables $x_1, \ldots, x_k \in \mathcal{A}$. Similar to the log-Laplace transform of classical random variables, the free cumulants of $(x_1, \ldots, x_k)$ may be gathered into the so-called *R-transform* [85, Lecture 16]:

$$R_{(x_1,\ldots,x_k)}(z_1, \ldots, z_k) := \sum_{\ell=1}^{\infty} \sum_{i_1,\ldots,i_\ell \in [k]} \kappa_\ell(x_{i_1}, \ldots, x_{i_\ell})\, z_{i_1} \cdots z_{i_\ell}, \tag{1.12}$$

which is a formal series in non-commutative indeterminates $z_1, \ldots, z_k$. Analogously to independence for classical random variables, $x_1, \ldots, x_k$ are *free* if $R_{(x_1,\ldots,x_k)}(z_1, \ldots, z_k) = R_{x_1}(z_1) + \cdots + R_{x_k}(z_k)$, which is commonly phrased by the sentence "mixed cumulants vanish" (i.e., $\kappa_\ell(x_{i_1}, \ldots, x_{i_\ell}) = 0$ for every $\ell \geq 1$ and every non-constant sequence $(i_1, \ldots, i_\ell) \in [k]^\ell$).

Nica [82] showed that for $P_{1,n}, \ldots, P_{d,n} \in \{0,1\}^{n \times n}$ uniform, independently chosen permutation matrices, there exists a non-commutative probability space $(\mathcal{A}, \varphi)$ and free variables $u_1, \ldots, u_d \in \mathcal{A}$ such that:

(a) There is the convergence of mixed moments

$$\frac{1}{n} \mathbb{E}\mathrm{Tr}\, P_{i_1,n}^{w_1} \cdots P_{i_k,n}^{w_k} \xrightarrow[n \to \infty]{} \varphi\big(u_{i_1}^{w_1} \cdots u_{i_k}^{w_k}\big),$$

for every $k \geq 1$, every $(i_1, \ldots, i_k) \in [d]^k$, and every word $w \in \Sigma^k$.

(b) The u_i's are *Haar unitaries*, in the sense that $u_i^* u_i = u_i u_i^* = 1$ and $\varphi(u_i^k) = 0$ for every $k \geq 1$.

It follows from (a) and linearity that the star moments of $\mathbf{A}_{d,n} := P_{1,n}+\cdots+P_{d,n}$ converge to those of $a_d := u_1+\cdots+u_d$, and it is easy to see that $\mathbf{A}_{d,n}$ is distributed like the adjacency matrix of the configuration model $\mathbf{CM}_{d,n}$ introduced in the proof of Theorem 1.1: then, we may again condition on $\mathbf{CM}_{d,n}$ being simple ($\mathbf{A}_{d,n}(i,i) = 0$ and $\mathbf{A}_{d,n}(i,j) \le 1$ for all $i \neq j \in [n]$) to deduce the star-moment convergence, for every word w on Σ,

$$\frac{1}{n}\mathbb{E}\mathrm{Tr}\, A_{n,d}^w = \frac{1}{n}\mathbb{E}\Big[\mathrm{Tr}\, \mathbf{A}_{d,n}^w \mid \mathbf{CM}_{d,n} \text{ is simple}\Big] \xrightarrow[n\to\infty]{} \varphi(a_d^w),$$

of the uniform d-regular digraph $G_{d,n}$ with adjacency matrix $A_{n,d}$.

Finally, we check that the star moments $\varphi(a_d^w)$ coincide with the number $M_d(w)$ of w-paths in T_d. Using (b), it was derived in [86] that (for every $i \in [d]$)

$$R_{u_i,u_i^*}(z_1,z_2) = \sum_{k=1}^{\infty} (-1)^{k-1}\, C_{k-1}\Big[(z_1z_2)^k + (z_2z_1)^k\Big].$$

By freeness, $R_{a,a^*}(z_1,z_2) = R_{u_1,u_1^*}(z_1,z_2) + \cdots + R_{u_d,u_d^*}(z_1,z_2) = d\, R_{u_1,u_1^*}(z_1,z_2)$, and the structure of this R-transform shows that the free cumulants $\kappa_\ell(a_d^{w_1},\dots,a_d^{w_\ell})$ (which we recover from (1.12)) vanish if $w := w_1\cdots w_\ell$ is not alternating:

$$\kappa_\ell(a_d^{w_1},\dots,a_d^{w_\ell}) = \begin{cases} d\,(-1)^{p-1}\, C_{p-1}, & \text{if } w \text{ is alternating: } w = (1*)^p \text{ or } w = (*1)^p, \\ 0, & \text{otherwise.} \end{cases}$$

Recalling the definition of $\mathrm{ANC}(w)$, the moment-cumulant formula (1.11) then easily yields

$$\varphi(a_d^w) \;\; = \sum_{\pi\in\mathrm{ANC}(w)} \Bigg(\prod_{V\in\pi} (-1)^{\frac{|V|}{2}-1}\, C_{\frac{|V|}{2}-1}\Bigg) d^{|\pi|},$$

as in Theorem 1.3.

1.3.2 The oriented Kesten–McKay conjecture.

Theorem 1.1 states that the uniform d-regular digraph $G_{d,n}$ converges in star moments to the d-regular directed tree T_d. As we saw in the previous section, the star moments of T_d agree with those of the sum $a_d := u_1 + \cdots + u_d$ of d free Haar unitary elements in some non-commutative probability space $(\mathcal{A},\varphi)$. This implies the convergence of mean empirical *singular value* distributions: for every $z \in \mathbb{C}$ and every continuous bounded function f,

$$\frac{1}{n}\mathbb{E}\mathrm{Tr}\, f\Big(\sqrt{\big(A_{n,d}-z\mathrm{I}_n\big)^*\big(A_{n,d}-z\mathrm{I}_n\big)}\Big) \xrightarrow[n\to\infty]{} \varphi\Big(f\Big(\sqrt{(a_d-z1)^*(a_d-z1)}\Big)\Big),$$

that is,

$$\int f(t)\,\mu_{|A_{n,d}-z|}(\mathrm{d}t) \xrightarrow[n\to\infty]{} \int f(t)\,\mu_{|a_d-z|}(\mathrm{d}t), \tag{1.13}$$

where $X - z$ means that we subtract z times the identity element to X, and $\mu_{|X|}$ is the spectral measure of the positive operator $|X| := \sqrt{XX^*}$ (i.e., $\mu_{|X|}$ is the unique real probability measure having the same moments as $|X|$, as given by the Riesz–Markov–Kakutani theorem).

Although $X \in \{A_{n,d}, a_d\}$ is not a normal element, there still exists [58] a unique probability measure μ_X (on $\mathbb{C}$), known as the Brown measure of X, such that

$$\int \log|z-\lambda|\,\mu_X(\mathrm{d}\lambda) = \int \log(t)\,\mu_{|X-z|}(\mathrm{d}t)$$

for every $z \in \mathbb{C}$. When $X = A_{n,d}$, μ_X is nothing but the ESD $\frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i(A_{n,d})}$ of $G_{d,n}$. Since the star moments of X determine $(\mu_{|X-z|})_{z\in\mathbb{C}}$ and thus μ_X, and the star moments of a_d and T_d coincide, we can also view μ_{a_d} as the spectral measure of T_d. However, we cannot directly use (1.13) to show

$$\frac{1}{n}\mathbb{E}\sum_{i=1}^n f\big(\lambda_i(A_{n,d})\big) \xrightarrow[n\to\infty]{} \int_{\mathbb{C}} f(z)\,\mu_{a_d}(\mathrm{d}z) \tag{1.14}$$

because the logarithm is not a bounded function. There still lacks a uniform control on the smallest singular value of $A_{n,d}-z$ to validate the oriented Kesten–McKay conjecture (1.14), see [28, Lemma 4.3].

Chapter 2

Limiting spectral distribution of random self-adjoint quantum channels [73]

What does a quantum channel drink?
Uncertain-tea.

I need a coffee

We study the limiting spectral distribution of quantum channels whose Kraus operators are sampled as $n \times n$ random Hermitian matrices satisfying certain assumptions. We show that when the Kraus rank goes to infinity with n, the limiting spectral distribution (suitably rescaled) of the corresponding quantum channel coincides with the semi-circle distribution. When the Kraus rank is fixed, the limiting spectral distribution is no longer the semi-circle distribution. It corresponds to an explicit law, which can also be described using tools from free probability.

Contents

2.1 Introduction

In quantum physics, the state of an n-dimensional system is described by a density operator on $\mathbb{C}^n$, i.e. $\rho \in M_n(\mathbb{C})$ a Hermitian positive semidefinite matrix with trace 1: $\rho \succeq 0$ and $\mathrm{tr}(\rho) = 1$. A transformation of such a quantum system is described by a quantum channel on $M_n(\mathbb{C})$, i.e. $\Phi : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ a completely positive and trace-preserving linear map. We recall that a linear map $\Phi : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ is said to be

- positive if it preserves the fact of being Hermitian and positive semidefinite: for all $X \in M_n(\mathbb{C})$, $X \succeq 0$ implies that $\Phi(X) \succeq 0$;
- completely positive if $\Phi \otimes \mathrm{Id} : M_{n^2}(\mathbb{C}) \to M_{n^2}(\mathbb{C})$ is positive (where $\mathrm{Id} : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ denotes the identity map);
- trace-preserving if it preserves the trace: for all $X \in M_n(\mathbb{C})$, $\mathrm{tr}(\Phi(X)) = \mathrm{tr}(X)$.

A quantum channel Φ thus maps quantum states to quantum states (and so does $\Phi \otimes \mathrm{id}$).

The action of a completely positive map Φ on $M_n(\mathbb{C})$ can always be described in the following (non-unique) way, called a Kraus representation of Φ (see e.g. [10, Section 2.3.2] or [114, Chapter 2]): There exist $d \in \mathbb{N}$ and $K_1, \ldots, K_d \in M_n(\mathbb{C})$, called Kraus operators of Φ, such that

$$\Phi : X \in M_n(\mathbb{C}) \mapsto \sum_{i \in [d]} K_i X K_i^* \in M_n(\mathbb{C}), \tag{2.1}$$

where K_i^* stands for the adjoint of K_i (and where we have used the shorthand notation $[d]$ for the set of integers $\{1, \ldots, d\}$). The fact that Φ is trace-preserving is equivalent to the following constraint on the Kraus operators $K_1, \ldots, K_d$:

$$\sum_{i \in [d]} K_i^* K_i = \mathrm{Id}\,.$$

The smallest d such that an expression of the form of equation (2.1) for Φ exists is called the Kraus rank of Φ. It is always at most n^2 for a completely positive map on $M_n(\mathbb{C})$.

Given a completely positive map Φ on $M_n(\mathbb{C})$, its adjoint (or dual) is defined by duality with respect to the Hilbert-Schmidt inner product, i.e. as the map Φ^* on $M_n(\mathbb{C})$ (which is completely positive as well) such that, for all $X, Y \in M_n(\mathbb{C})$,

$$\mathrm{tr}(X \Phi^*(Y)) = \mathrm{tr}(\Phi(X) Y). \tag{2.2}$$

Φ being trace-preserving is equivalent to Φ^* being unital, i.e. such that $\Phi^*(\mathrm{Id}) = \mathrm{Id}$.

Note that, identifying $M_n(\mathbb{C})$ with $\mathbb{C}^n \otimes \mathbb{C}^n$, a linear map $\Phi : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ can equivalently be seen as a linear map $M_\Phi : \mathbb{C}^n \otimes \mathbb{C}^n \to \mathbb{C}^n \otimes \mathbb{C}^n$, i.e. an element of $M_{n^2}(\mathbb{C})$. Concretely, a completely positive linear map

$$\Phi : X \in M_n(\mathbb{C}) \mapsto \sum_{i \in [d]} K_i X K_i^* \in M_n(\mathbb{C})$$

can be identified with

$$M_\Phi = \sum_{i\in[d]} K_i \otimes \overline{K}_i \in M_{n^2}(\mathbb{C}), \tag{2.3}$$

where $\overline{K}_i$ stands for the entry-wise conjugate of K_i, in the canonical basis of $\mathbb{C}^n$. This identification preserves the spectrum, i.e. $\mathrm{spec}(\Phi) = \mathrm{spec}(M_\Phi)$. Moreover, the matrix version of the adjoint Φ^* of Φ is simply the adjoint of M_Φ, i.e. $M_{\Phi^*} = M_\Phi^*$.

In this paper, we will consider quantum channels Φ whose Kraus operators are Hermitian matrices, which ensures that Φ itself is Hermitian (in the sense that it is equal to its adjoint Φ^*, as defined by equation (2.2)), or equivalently that its matrix version M_Φ is Hermitian. This may look like a restrictive setting, but it, in fact, encompasses all Hermitian quantum channels. Indeed, it is not hard to see that any Hermitian completely positive map Φ on $M_n(\mathbb{C})$ admits a Kraus representation with Hermitian Kraus operators, where we can additionally guarantee that the number of such operators is at most twice the Kraus rank of Φ (i.e. in general at most $2n^2$). Concretely, if $K_1, \ldots, K_d \in M_n(\mathbb{C})$ are d Kraus operators for Φ, then $K_{1,R}, K_{1,I}, \ldots, K_{d,R}, K_{d,I} \in M_n(\mathbb{C})$ are $2d$ Hermitian Kraus operators for Φ, where given $K \in M_n(\mathbb{C})$, we set $K_R = (K + K^*)/2$, $K_I = -i(K - K^*)/2$, so that $K = K_R + iK_I$ and $K_R^* = K_R$, $K_I^* = K_I$. The latter claim follows from the observation that,if $M_\Phi^* = M_\Phi$, then we can re-write $M_\Phi = (M_\Phi + M_\Phi^*)/2$, and thus

$$M_\Phi = \frac{1}{2}\sum_{j\in[d]} \left(K_j \otimes \overline{K}_j + K_j^* \otimes \overline{K}_j^*\right) = \sum_{j\in[d]} \left(K_{j,R} \otimes \overline{K}_{j,R} + K_{j,I} \otimes \overline{K}_{j,I}\right).$$

Now, many interesting and well-studied classes of quantum channels are Hermitian: depolarizing channels, dephasing channels, and Pauli channels, to name just a few. Note that such channels are, in particular, unital.

This work aims to study the spectrum of a randomly generated quantum channel as the underlying dimension n goes to infinity. More precisely, given a quantum channel Φ on $M_n(\mathbb{C})$ whose Kraus operators $K_1, \ldots, K_d \in M_n(\mathbb{C})$ have been randomly sampled, we aim to characterize its asymptotic spectrum (i.e. its spectrum in the limit where n goes to infinity). Our study will involve two regimes, one where the Kraus rank d is fixed and only n grows, and the other one where $d = d(n)$ is a growing function of n.

Previous related works were mostly concerned with identifying the spectral gap of a random quantum channel, i.e. the difference between its largest and second largest eigenvalues. Indeed, it is known that a quantum channel Φ on $M_n(\mathbb{C})$ always has its largest (in modulus) eigenvalue $\lambda_1(\Phi)$ equal to 1 (with an associated eigenvector which is a positive semidefinite matrix), implying that Φ always has a fixed state. It was established that random quantum channels generically have their largest eigenvalue 1 isolated from the rest of the spectrum. All other eigenvalues are of order at most $1/\sqrt{d}$ (in modulus). This was shown first for specific models, where Kraus operators were sampled either as independent Haar unitaries [62, 88], or as blocks of a Haar isometry [54], or as independent Ginibre matrices (i.e. matrices having i.i.d. Gaussian entries) [71], and recently in greater generality [72].

On the other hand, much less is known concerning the asymptotic distribution of the bulk of the spectrum. To understand this, one studies the empirical spectral distribution, which, given a matrix $M \in M_n(\mathbb{C})$, is defined as

$$\mu_M := \frac{1}{n} \sum_{k \in [n]} \delta_{\lambda_k(M)},$$

where $\lambda_k(M)$ is the kth eigenvalue of M. In [32], a model where Kraus operators are sampled as blocks of a Haar isometry was studied. It was conjectured, from heuristic arguments and numerical simulations, that in the regime where $d = d(n) = n^2$, the empirical spectral distribution (rescaled by a factor $\sqrt{d}$) of such random quantum channel converges towards a circular distribution as n grows. It was later shown in [9] that, in the same regime $d = d(n)$ of order n^2, the empirical singular value distribution (again rescaled by a factor $\sqrt{d}$) of a random quantum channel whose Kraus operators are sampled as independent Ginibre matrices converges towards a quarter-circular distribution as n grows. We believe that for both models (and for more general ones) and in any regime $d = d(n) \to \infty$, the limiting spectral distribution coincides with the circular law.

In this work, we embark on the investigation of the limiting spectral distribution of quantum channels in the Hermitian setting. This serves as a precursor to our broader research goals involving the non-Hermitian case and the conjecture presented above. Classical problems in Random Matrix Theory have traditionally prioritized the exploration of Hermitian matrices before delving into their non-Hermitian counterparts, partly due to the inherent technical Complexities introduced by non-Hermitian systems, notably the instability of the spectrum under perturbations (see [14, Chapter 11]). For instance, while the limiting spectral distribution of an $n \times n$ Hermitian matrix with i.i.d. (up to symmetry) centered entries of variance $1/n$ was shown to be Wigner's semi-circle distribution, it is almost half a century later that the analogous result in the non-Hermitian case, Girko's circular law theorem, was established in full generality [104]. Understanding the limiting spectral distribution in the non-Hermitian case follows the Hermitization technique invented by Girko [51, 52]. The latter requires a quantitative control on the smallest singular value of the corresponding random matrix model (see [28] for an introduction to the method), making the problem significantly more involved. In the context of quantum channels, the associated random matrix model exhibits dependencies among its entries, which adds a layer of difficulty compared to classical random matrix models, even in the Hermitian setting.

Now, let us shift our focus back to the specific context of this paper. Here, we will consider the case where the Kraus operators of the quantum channel Φ are chosen to be random Hermitian operators. This ensures that the resulting completely positive map Φ, or equivalently its matrix version M_Φ, is Hermitian. Conversely, as explained earlier, any Hermitian completely positive map Φ can be written with Hermitian Kraus operators. Moreover, we aim to keep our assumptions regarding the distribution of these random Kraus operators as minimal as possible. Our objective is to gain a comprehensive

understanding of the spectrum of M_Φ as the dimension n increases without imposing specific constraints on the scaling relationship between d and n.

To keep the introduction light, we state an informal version of our main result and differ from the rigorous statement, requiring notions from free probability, to Section 2.2 (see Theorem 2.3).

Theorem 2.1 (Informal Statement)**.** *Let $W_1, \dots, W_d \in M_n(\mathbb{C})$ be independent Hermitian random matrices such that, for each $i \in [d]$, $\mathbb{E}(W_i) = 0$ and $\mathbb{E}(W_i^2) = \mathrm{Id}$. Suppose additionally that, for each $i \in [d]$, the spectral distribution of W_i converges to some distribution μ_i as $n \to \infty$ and that the joint asymptotic distribution of the family $(W_i)_{i\in[d]}$ is determined by the family $(\mu_i)_{i\in[d]}$. Next, set $K_i = W_i/\sqrt{d}$ for each $i \in [d]$ and define Φ as the random completely positive map on $M_n(\mathbb{C})$ having $K_1, \dots, K_d$ as Kraus operators, i.e.*

$$\Phi : X \in M_n(\mathbb{C}) \mapsto \sum_{i\in[d]} K_i X K_i \in M_n(\mathbb{C}).$$

First, Φ is, on average, trace-preserving and unital. And second,

- *If d is fixed, then the spectral distribution of $\Phi - \mathbb{E}(\Phi)$ converges to a specific distribution, depending only on $\mu_1, \dots, \mu_d$, as $n \to \infty$.*
- *If $d = d(n) \to \infty$ as $n \to \infty$ and W_i are i.i.d, then the spectral distribution of $\sqrt{d}(\Phi - \mathbb{E}(\Phi))$ converges to the semicircular distribution as $n \to \infty$.*

The specific assumption underlying the family of independent matrices $(W_i)_{i\in[d]}$ is precisely asymptotic freeness, a concept introduced by Voiculescu [109] within the framework of Free Probability. Asymptotic freeness proves instrumental in a systematical examination of the asymptotic behavior of random matrices. Importantly, this assumption is not restrictive as many classical independent random matrices are known to be asymptotically free. However, this property is no longer valid in the context of tensors (see [37]), preventing a direct derivation of the asymptotic joint spectrum of these random tensors, which is the setting of interest in this paper. To overcome this, we analyze the moment method directly on our model in order to derive the limiting distribution.

The required notions from Free Probability are presented in Section 2.2, where the main rigorous theorem is stated. The proofs are carried out in Section 2.3. Section 2.4 regroups a variety of examples to which our result applies.

Acknowledgments

Part of this work was completed during a stay of the second named author at New York University in Abu Dhabi, partly funded by a doctoral mobility grant delivered by Université Gustave Eiffel; he would like to thank both institutions for their support and the excellent working assumptions. The first named author was supported by the ANR projects ESQuisses (grant number ANR-20-CE47-0014-01), STARS (grant number ANR-20-CE40-0008), and QTraj (grant number ANR-20-CE40-0024-01).

2.2 Preliminaries and rigorous statement of the main result

2.2.1 Preliminaries

We begin by recalling some notation from Free Probability; see [85]. A noncommutative probability space is a pair $(\mathcal{A}, \tau)$, where $\mathcal{A}$ is a unital algebra equipped with a tracial state τ, that is, τ is linear, $\tau(1) = 1$ and $\tau(ab) = \tau(ba)$. We say that subalgebras $\mathcal{A}_1, \ldots, \mathcal{A}_d$ are free if

$$\tau(a_1 \ldots a_p) = 0 \tag{2.4}$$

whenever $a_i \in \mathcal{A}_{j_i}$, $\tau(a_i) = 0$ for all $i \in [p]$ and $j_1 \neq j_2 \neq \cdots \neq j_p$. We say that random variables $a_1, \ldots, a_d \in \mathcal{A}$ are free if their algebras are free. The distribution of a self-adjoint variable $a = a^*$ is defined as the collection of moments

$$\{\tau(a^p) : p \in \mathbb{N}\}.$$

In particular, there always exists a measure μ on $\mathbb{R}$ such that

$$\tau(a^p) = \int x^p \, \mathrm{d}\mu,$$

for every $p \in \mathbb{N}$. Such a measure is also called the distribution of a. A particular example of free variables is the free semicircular system $(s_1, \ldots, s_d)$, whose moments satisfy

$$\tau(s_{i_1} \cdots s_{i_p}) = \sum_{\pi \in NC_2(p)} \prod_{(l,k) \in \pi} \delta_{i_l i_k}.$$

Here, $NC_2(p)$ is the set of all noncrossing pair partitions of $[p]$, namely, each block of $\pi \in NC_2(p)$ has cardinality 2 and π does not contain two blocks $\{i, j\}, \{k, l\}$ such that $i < k < j < l$. In particular, the moments of the semicircular variable are

$$\tau(s^p) = |NC_2(p)| = \int x^p f_{sc}(x) \, \mathrm{d}x,$$

where f_{sc} denotes the density of the semi-circle distribution μ_{sc}

$$f_{sc}(x) = \frac{1}{2\pi} \sqrt{4 - x^2} \mathbf{1}_{|x| \leq 2}. \tag{2.5}$$

Given a matrix $M \in M_n(\mathbb{C})$, we define its normalized trace as

$$\tau^{(n)}(M) := \frac{1}{n} \operatorname{tr}(M).$$

In particular, for two matrices $M, N \in M_n(\mathbb{C})$, we have

$$\tau^{(n^2)}(M \otimes N) = \tau^{(n)} \otimes \tau^{(n)}(M \otimes N) = \tau^{(n)}(M)\tau^{(n)}(N).$$

We say that a random Hermitian matrix $M \in M_n(\mathbb{C})$ is *normalized* if

$$\mathbb{E}\Big(\tau^{(n)}(M)\Big) = 0;$$
$$\mathbb{E}\Big(\tau^{(n)}(M^2)\Big) = 1.$$

We say that a sequence $(M_n)_{n\in\mathbb{N}}$ of random Hermitian matrices $M_n \in M_n(\mathbb{C})$ such that $\mathrm{tr}(M_n^p)$ is integrable for all $p \geq 1$ converges weakly in probability (resp. in expectation) to μ, if $(\mu_{M_n})_{n\in\mathbb{N}}$ converges weakly to μ in probability (resp. in expectation). Equivalently, for any $p \geq 1$, we have

$$\tau^{(n)}(M_n^p) = \frac{1}{n}\mathrm{tr}(M_n^p) \underset{n\to\infty}{\to} \int x^p \,\mathrm{d}\mu$$

in probability and

$$\mathbb{E}\Big(\tau^{(n)}(M_n^p)\Big) = \mathbb{E}\Big(\frac{1}{n}\mathrm{tr}(M_n^p)\Big) \underset{n\to\infty}{\to} \int x^p \,\mathrm{d}\mu$$

for the convergence in expectation. Whenever $(\mu_{M_n})_{n\in\mathbb{N}}$ converges weakly to μ, we denote it by $\mu_{M_n} \Rightarrow \mu$, or simply $M_n \Rightarrow \mu$ as $n \to \infty$. In particular, we can find a noncommutative random variable a such that a has distribution μ and M_n converges to a.

Finally, we say that d random Hermitian matrices $M_1, \ldots, M_d \in M_n(\mathbb{C})$ are asymptotically free in probability if

$$\lim_{n\to\infty} \tau^{(n)}\Big(\Big(M_{i_1}^{p_1} - \tau^{(n)}(M_{i_1}^{p_1})\Big)\cdots\Big(M_{i_m}^{p_m} - \tau^{(n)}(M_{i_m}^{p_m})\Big)\Big) = 0$$

in probability, for all $m \geq 1$, $i_1 \neq i_2 \neq \cdots \neq i_m \in [d]$ and $p_1, \ldots, p_m \geq 1$. They are asymptotically free in expectation if

$$\lim_{n\to\infty} \mathbb{E}\left[\tau^{(n)}\Big(\Big(M_{i_1}^{p_1} - \mathbb{E}\Big(\tau^{(n)}(M_{i_1}^{p_1})\Big)\Big)\cdots\Big(M_{i_m}^{p_m} - \mathbb{E}\Big(\tau^{(n)}(M_{i_m}^{p_m})\Big)\Big)\Big)\right] = 0,$$

for all $m \geq 1$, $i_1 \neq i_2 \neq \cdots \neq i_m \in [d]$ and $p_1, \ldots, p_m \geq 1$. This is equivalent to (2.4) in the limit for the algebras $\mathcal{A}(M_i)$ generated by each M_i. As usual, we write $M + \lambda := M + \lambda\,\mathrm{Id}$, where $M \in M_n(\mathbb{C})$, $\lambda \in \mathbb{C}$.

Remark 2.2. We have defined weak convergence as the convergence of the moments of all order. In particular, this requires that $\mathrm{tr}(M_n^p)$ is integrable for all $p \geq 1$. In several cases, however, such strong integrability is not needed, and similar results can be proved for a larger class of random matrices via truncation techniques [7, Theorem 2.1.21].

2.2.2 Rigorous results

Let $\mu_1, \ldots, \mu_d$ be probability measures on and $a_1, \ldots, a_d$ be free variables with distribution μ_i, respectively. We define the tensor measure $\mu_i \star \mu_i$ as the measure associated with $a_i \otimes a_i$ and the tensor convolution $\mu_1 \circledast \cdots \circledast \mu_d$ as the distribution of

$$\sum_{i \in [d]} a_i \otimes a_i = a_1 \otimes a_1 + \cdots + a_d \otimes a_d.$$

In other words, we have

$$\int x^k \, \mathrm{d}\mu_i \star \mu_i = \tau \otimes \tau(a_i^k \otimes a_i^k) = \tau^2(a_i^k),$$

and

$$\int x^k \, \mathrm{d}\mu_1 \circledast \cdots \circledast \mu_d = \tau \otimes \tau \left(\sum_{i \in [d]} a_i \otimes a_i \right)^k = \sum_{i \in [d]^k} \tau^2(a_{i_1} \cdots a_{i_k}).$$

Note that $\mu_1 \circledast \cdots \circledast \mu_d$ is not necessarily the free convolution of $\mu_i \star \mu_i$. Indeed, it was proved in [37] that freeness for tensor products $a_i \otimes a_i$ does not follow from freeness of a_i.

In the sequel, we will be considering random Hermitian matrices $W_1, \ldots, W_d \in M_n(\mathbb{C})$ satisfying the following assumptions:

(**A.1**) For each $i \in [d]$, $W_i \in M_n(\mathbb{C})$ converges weakly in probability and in expectation to μ_i;

(**A.2**) The family $(W_i)_{i \in [d]}$ is in probability and in expectation asymptotically free, for each d fixed.

(**A.3**) For each $i \in [d]$, $\mathbb{E}(W_i \otimes \overline{W_i}) \in M_{n^2}(\mathbb{C})$ converges weakly to 0.

(**A.4**) The matrices $(W_i)_{i \in [d]}$ are independent and identically distributed.

As we will see in Section 2.4, a variety of classical random matrix models satisfy those assumptions. We are ready to state our main result.

Theorem 2.3. *Let $W_1, \ldots, W_d \in M_n(\mathbb{C})$ be centered Hermitian random matrices satisfying Assumptions (**A.1**),(**A.2**), (**A.3**) and $d = d(n)$. Then, the following holds.*

(2.3.i) If d is fixed, then

$$\Delta := \frac{1}{\sqrt{d}} \sum_{i \in [d]} \left(W_i \otimes \overline{W}_i - \mathbb{E}(W_i \otimes \overline{W}_i) \right)$$

converges weakly in probability and in expectation to the tensor convolution of $\tilde{\mu}_1, \ldots, \tilde{\mu}_d$, *where* $\tilde{\mu}$ *denotes the dilation of* μ, *i.e.*

$$\int x^p \, d\tilde{\mu} := \frac{1}{d^{p/2}} \int x^p \, d\mu.$$

This means that

$$\mu_\Delta \underset{n\to\infty}{\Rightarrow} \tilde{\mu}_1 \circledast \cdots \circledast \tilde{\mu}_d,$$

in probability and in expectation.

(2.3.ii) If $d = d(n)$ *diverges,* $(W_i)_{i\in[d]}$ *are also normalized and satisfy Assumption (**A.4**), then* Δ *converges weakly in probability and in expectation to the semicircular distribution* μ_{sc}. *This means that*

$$\mu_\Delta \underset{n\to\infty}{\Rightarrow} \mu_{sc},$$

in probability and in expectation.

2.3 Proofs

We start with the following observation.

Lemma 2.4. *Suppose that Assumptions (**A.1**) and (**A.3**) hold. For each* $i \in [d]$, $B_i = W_i \otimes \overline{W}_i - \mathbb{E}\big(W_i \otimes \overline{W}_i\big)$ *converges weakly in probability and in expectation to* $a_i \otimes a_i$, *where* a_i *has distribution* μ_i.

Proof. For Hermitian matrices $M_1, M_2 \in M_n(\mathbb{C})$, such that $M_1 \Rightarrow \mu_1$ and $M_2 \Rightarrow 0$ as $n \to \infty$, we have

$$M_1 + M_2 \Rightarrow \mu_1.$$

Indeed, using Holder's Inequality, we can bound the trace of a product of the matrices M_1 and M_2 by

$$\left| \tau^{(n)} \left(\prod_{j\in[m]} M_{i_j} \right) \right| \leq \prod_{j\in[m]} \left(\tau^{(n)} |M_{i_j}|^{p_j} \right)^{1/p_j}, \tag{2.6}$$

where $i \in \{1,2\}^m$ and $(p_j)_{j\in[m]}$ are conjugate exponents. If there is at least one $j \in [m]$ such that $i_j = 2$, then the right-hand side of (2.6) goes to 0. In particular, since

$$\tau^{(n)}((M_1 + M_2)^p) = \sum_{i\in\{1,2\}^p} \tau^{(n)} \left(\prod_{j\in[p]} M_{i_j} \right),$$

any term associated with $i \in \{1,2\}^p$ having at least one index $i_j = 2$ will asymptotically vanish. Therefore,

$$\lim_{n\to\infty} \tau^{(n)}((M_1 + M_2)^p) = \lim_{n\to\infty} \tau^{(n)}(M_1^p).$$

The result of the lemma follows by taking $M_1 = W_i \otimes \overline{W_i}$ and $M_2 = -\mathbb{E}(W_i \otimes \overline{W_i})$, for each $i \in [d]$. Indeed,

$$\tau^{(n)} \otimes \tau^{(n)}\Big(\big(W_i \otimes \overline{W}_i\big)^p\Big) = \Big(\tau^{(n)}(W_i^p)\Big)^2,$$

since the eigenvalues of W_i are real. Hence, the weak convergence in probability and in expectation of W_i to μ_i implies that

$$\tau^{(n)} \otimes \tau^{(n)}\Big(\big(W_i \otimes \overline{W}_i\big)^p\Big) \underset{n\to\infty}{\to} \left(\int x^p \,\mathrm{d}\mu_i\right)^2$$

in probability and in expectation. This finishes the proof. □

As an immediate consequence, we have the following.

Corollary 2.5. *Suppose that Assumptions (**A.1**),(**A.2**), and (**A.3**) hold. Let $B_i = W_i \otimes \overline{W_i} - \mathbb{E}(W_i \otimes \overline{W_i})$. Then, for each d fixed, $(B_i)_{i\in[d]}$ converges weakly in probability and in expectation to $(a_i \otimes a_i)_{i\in[d]}$, where a_i are free variables with distribution μ_i.*

Proof. By Assumption (**A.2**), we can assume $(W_i)_{i\in[d]}$ converges weakly in probability and in expectation to $(a_i)_{i\in[d]}$, where a_i are free. The result follows by Lemma 2.4. □

We are ready to prove Theorem (2.3.i).

Proof of Theorem (2.3.i). By Corollary 2.5, we have

$$\sum_{i\in[d]} B_i \underset{n\to\infty}{\Rightarrow} \sum_{i\in[d]} a_i \otimes a_i,$$

where a_i are free with distribution μ_i. To conclude, it suffices to note that $\tilde{\mu}_i$ is the distribution of $a_i \otimes a_i/\sqrt{d}$. □

Before proving Theorem (2.3.ii), let us give a sufficient condition for Assumption (**A.3**) to hold.

Lemma 2.6. *Let $W_1, \ldots, W_d$ satisfying Assumption (**A.1**) and assume that*

$$\mathbb{E}(\tau^{(n)}(W_i)) \underset{n\to\infty}{\to} 0, \tag{2.7}$$

*for all $i \in [d]$. Let $(W_i^{(l)})_{l\in[L]}$ be independent copies of W_i. If for each $i \in [d]$ and $L \geq 1$, $(W_i^{(l)})_{l\in[L]}$ are asymptotically free in probability and in expectation, then Assumption (**A.3**) holds.*

Proof. Fix $i \in [d]$. Then, we can easily see that

$$\frac{1}{n^2}\operatorname{tr}\Big(\Big(\mathbb{E}\Big(W_i \otimes \overline{W_i}\Big)\Big)^p\Big) = \frac{1}{n^2}\,\mathbb{E}\Big(\operatorname{tr}^2\Big(W_i^{(1)} \cdots W_i^{(p)}\Big)\Big).$$

Since $(W_i^{(l)})_{l\in[p]}$ are asymptotically free in probability and in expectation, we get that

$$\frac{1}{n^2}\operatorname{tr}\Big(\Big(\mathbb{E}\Big(W_i \otimes \overline{W_i}\Big)\Big)^p\Big) \underset{n\to\infty}{\to} \tau^2(a_i^{(1)} \cdots a_i^{(p)}),$$

where $(a_i^{(l)})_{l\in[p]}$ are free i.i.d with distribution μ_i. The centered Assumption (2.7) implies that μ_i is centered, hence $\tau(a_i^{(1)} \cdots a_i^{(p)}) = 0$ by freeness and the result follows. □

2.3.1 The asymptotic free Central Limit Theorem

In order to prove Theorem (2.3.ii), we begin with a simple lemma.

Lemma 2.7. *Suppose that Assumption (**A.4**) holds. For each $i \in [d]$, set $B_i := W_i \otimes \overline{W}_i - \mathbb{E}(W_i \otimes \overline{W}_i)$. Let $(i_1, \ldots, i_p) \in [d]^p$ and suppose that there exists $k \in [p]$ such that, for all $l \in [p]$ with $l \neq k$, $i_l \neq i_k$. Then, we have*

$$\mathbb{E}\Big(\operatorname{tr}(B_{i_1} \cdots B_{i_p})\Big) = 0.$$

Proof. We can assume without loss of generality that $k = 1$, i.e. $i_1 \notin \{i_2, \ldots, i_p\}$. In this case, notice that B_{i_1} is independent of $B := B_{i_2} \cdots B_{i_p}$ by Assumption (**A.4**), hence

$$\mathbb{E}(\operatorname{tr}(B_{i_1} B)) = \sum_{k_1,k_2\in[n]^2} \mathbb{E}(B_{i_1}(k_1,k_2))\,\mathbb{E}(B(k_2,k_1)).$$

Since B_{i_1} is centered, we have $\mathbb{E}(B_{i_1}(k_1,k_2)) = 0$ and the result follows. □

Although Lemma 2.7 is a simple consequence of independence and centering, it is a powerful property that will allow us to obtain a limit distribution for Δ. We are now ready to prove Theorem (2.3.ii).

Proof of Theorem (2.3.ii). We begin by writing Δ as

$$\Delta = \frac{1}{\sqrt{d}} \sum_{i\in[d]} B_i,$$

where $B_i = W_i \otimes \overline{W}_i - \mathbb{E}(W_i \otimes \overline{W}_i)$. We recall that, given $X \in M_{n^2}(\mathbb{C})$ a random matrix, we have

$$\mathbb{E}\Big(\tau^{(n^2)}(X)\Big) = \mathbb{E}\Big(\frac{1}{n^2}\operatorname{tr}(X)\Big).$$

Then, we can compute the moments of Δ by

$$\mathbb{E}\Big(\tau^{(n^2)}(\Delta^p)\Big) = d^{-p/2} \sum_{i\in[d]^p} \mathbb{E}\Big(\tau^{(n^2)}\Big(B_{i_1}\cdots B_{i_p}\Big)\Big).$$

For every $i \in [d]^p$, we associate a partition $\pi(i)$ of $[p]$ defined by placing any $k, \ell \in [p]$ in the same block of $\pi(i)$ whenever $i_k = i_l$. We denote by $P(p)$ the set of all partitions of $[p]$. Notice that for every $i \in [d]^p$, $\mathbb{E}\Big(\tau^{(n^2)}(B_{i_1}\cdots B_{i_p})\Big)$ only depends on the partition $\pi(i)$, since B_i are identically distributed by Assumption (**A.4**). In particular, let $\tau^{(n^2)}(\pi)$ be the common value of $\mathbb{E}\Big(\tau^{(n^2)}(B_{i_1}\cdots B_{i_p})\Big)$ for $\pi(i) = \pi$. Hence,

$$\mathbb{E}\Big(\tau^{(n^2)}(\Delta^p)\Big) = d^{-p/2} \sum_{\pi\in P(p)} \tau^{(n^2)}(\pi)|\{i \in [d]^p : \pi(i) = \pi\}|.$$

The cardinality can be computed by simply choosing an index for each block of π, namely,

$$|\{i \in [d]^p : \pi(i) = \pi\}| = d(d-1)\cdots(d-|\pi|+1) = d^{|\pi|}\Big(1 + O_p\Big(\frac{1}{d}\Big)\Big),$$

where $f = O_p(1/d)$ means that $f \le C_p/d$, for some constant C_p that depends only on p. We then have

$$\mathbb{E}\Big(\tau^{(n^2)}(\Delta^p)\Big) = \sum_{\pi\in P(p)} \tau^{(n^2)}(\pi) d^{|\pi|-p/2}\Big(1 + O_p\Big(\frac{1}{d}\Big)\Big).$$

By Lemma 2.7, if $\pi \in P(p)$ has a block V of size 1, then we would have $\tau^{(n^2)}(\pi) = 0$. Hence, we can restrict to partitions without single blocks, yielding

$$\mathbb{E}\Big(\tau^{(n^2)}(\Delta^p)\Big) = \sum_{\substack{\pi\in P(p)\\ \forall V\in\pi,\, |V|\ge 2}} \tau^{(n^2)}(\pi) d^{|\pi|-p/2}\Big(1 + O_p\Big(\frac{1}{d}\Big)\Big).$$

Using Assumption (**A.1**) and Holder's Inequality, we have that

$$\tau^{(n^2)}(\pi) \le \max_{i\in[d]}\Big[\mathbb{E}\Big(\tau^{(n^2)}(|B_i|^p)\Big)\Big]^{1/p} \le C_p, \tag{2.8}$$

where $C_p < \infty$ is a constant independent of n. Hence, whenever π has a block of size $|V| \ge 3$, we have $|\pi| < p/2$ and

$$\tau^{(n^2)}(\pi) d^{|\pi|-p/2} = O_p\Big(\frac{1}{\sqrt{d}}\Big).$$

In particular, the only partitions that contribute to the dominating term are the pair partitions; that is, every block has cardinality two. Let $P_2(p)$ be the set of all pair partitions, then

$$\mathbb{E}\Big(\tau^{(n^2)}(\Delta^p)\Big) = \sum_{\pi\in P_2(p)} \tau^{(n^2)}(\pi) + O_p\Big(\frac{1}{\sqrt{d}}\Big).$$

Note that $\tau^{(n^2)}(\pi)$ depends only on $B_1, \ldots, B_p$. Corollary 2.5 implies that $(B_i)_{i\in[d]}$ converges to $(a_i \otimes a_i)_{i\in[d]}$ in probability and in expectation, where a_i are free variables and the limit of W_i. Hence,

$$\tau^{(n^2)}(\pi) \underset{n\to\infty}{\to} \tau\otimes\tau(\pi) := \tau^2(a_{i_1}\cdots a_{i_p}),$$

In particular, freeness implies that $\tau\otimes\tau(\pi) = 0$ if and only if $\pi \in P_2(p)\setminus NC_2(p)$, where $NC_2(p)$ is the set of noncrossing pair partitions. If $\pi \in NC_2(p)$, then $\tau\otimes\tau(\pi) = 1$ because the matrices are normalized. We deduce that

$$\mathbb{E}\Big(\tau^{(n^2)}(\Delta^p)\Big) = |NC_2(p)| + o_p(1) + O_p\left(\frac{1}{\sqrt{d}}\right),$$

where $f = o_p(1)$ denotes a function that depends on p and $f \to 0$ as $n \to \infty$. This proves the convergence in expectation. To get convergence in probability, set

$$\tau^{(n^2)}(j) = \tau^{(n^2)}(B_{j_1}\cdots B_{j_p}),$$

for each $j \in [d]^p$. Then,

$$\mathrm{var}\Big(\tau^{(n^2)}(\Delta^p)\Big) = \frac{1}{d^p}\sum_{i,j\in[d]^p}\Big\{\mathbb{E}\Big(\tau^{(n^2)}(i)\tau^{(n^2)}(j)\Big) - \mathbb{E}\Big(\tau^{(n^2)}(i)\Big)\,\mathbb{E}\Big(\tau^{(n^2)}(j)\Big)\Big\}.$$

Now, each index i_l, j_l must appear at least twice in $i\cup j$, as the matrices are centered and independent. In particular, the summation has at most d^p indices. We thus have

$$\mathrm{var}\Big(\tau^{(n^2)}(\Delta^p)\Big) \le \max_{i,j\in[d]^p}\Big\{\mathbb{E}\Big(\tau^{(n^2)}(i)\tau^{(n^2)}(j)\Big) - \mathbb{E}\Big(\tau^{(n^2)}(i)\Big)\,\mathbb{E}\Big(\tau^{(n^2)}(j)\Big)\Big\}.$$

Assumption (**A.1**) implies that $(B_i)_{i\in[p]}$ are uniformly integrable, then the Dominated Convergence Theorem [43, Theorem 1.5.8] and Corollary 2.5 imply that

$$\mathbb{E}\Big(\tau^{(n^2)}(i)\tau^{(n^2)}(j)\Big) \underset{n\to\infty}{\to} \tau\otimes\tau(\pi(i))\tau\otimes\tau(\pi(j)),$$

The same limit holds for $\mathbb{E}\Big(\tau^{(n^2)}(i)\Big)\,\mathbb{E}\Big(\tau^{(n^2)}(j)\Big)$, and therefore

$$\mathrm{var}\Big(\tau^{(n^2)}(\Delta^p)\Big) = o_p(1),$$

from which convergence in probability follows. □

Remark 2.8. As we saw in the proof, the only assumption for the existence of a limit of Δ is the centering assumption in Lemma 2.7. In particular, the proof of Theorem (2.3.ii) works verbatim for noncommutative variables. If $a_1, \ldots, a_d$ are centered exchangeable variables such that Lemma 2.7 holds, that is, $\tau(a_{i_1}\cdots a_{i_p}) = 0$ whenever there exists an

index different than the others, we have

$$\lim_{d\to\infty} \tau\left(\frac{1}{\sqrt{d}}\sum_{i\in[d]} a_i\right)^p = \sum_{\pi\in P_2(p)} \tau(\pi).$$

If in addition $(a_i)_{i\in[d]} = (a_i^{(n)})_{i\in[d]}$ depends on some parameter n and are asymptotically free, as n grows, we get an asymptotic free Central Limit Theorem

$$\lim_{n,d\to\infty} \tau^{(n)}\left(\frac{1}{\sqrt{d}}\sum_{i\in[d]} a_i^{(n)}\right)^p = |NC_2(p)|.$$

2.4 Examples

We begin with an almost surely unital and trace-preserving Hermitian quantum channel.

Example 2.9 (Rademacher diagonal matrices)**.** Let $R = \mathrm{diag}(\varepsilon_k)_{k\in[n]} \in M_n(\mathbb{C})$ be a Rademacher diagonal matrix, i.e. $\varepsilon_1, \ldots, \varepsilon_n$ are independent Rademacher (or symmetric Bernoulli) random variables with parameter $1/2$. Let $U_1, \ldots, U_d \in M_n(\mathbb{C})$ be independent Haar unitary matrices independent of R as well and set $W_i = U_i R U_i^*$ for each $i \in [d]$. It follows from [109] that $(W_i)_{i\in[d]}$ are asymptotically free and Lemma 2.6 yields Assumption (**A.3**). Theorem (2.3.i) thus implies that

$$\Delta \underset{n\to\infty}{\Rightarrow} \frac{1}{\sqrt{d}}\sum_{i\in[d]} r_i \otimes r_i,$$

in probability and in expectation, where $(r_i)_{i\in[d]}$ are free with Rademacher distribution. In particular, for a polynomial

$$p(x) := b_0 + \sum_{l\in[m]} b_l x^l,$$

we immediately compute

$$p(r_i) - \tau(p(r_i)) = \left(\sum_{l\in[\lceil m/2\rceil]} b_{2l-1}\right) r_i;$$
$$p(r_i \otimes r_i) - \tau\otimes\tau(p(r_i\otimes r_i)) = \left(\sum_{l\in[\lceil m/2\rceil]} b_{2l-1}\right) r_i \otimes r_i.$$

Since $\tau(r_i) = 0$, we can readily see that the freeness of $(r_i)_{i\in[d]}$ is equivalent to the freeness of $(r_i \otimes r_i)_{i\in[d]}$ in this case. Hence $(r_i \otimes r_i)_{i\in[d]} \overset{d}{=} (r_i)_{i\in[d]}$ where the equality holds in distribution. If we denote the limit of Δ by z, we then have

$$z \overset{d}{=} \frac{1}{\sqrt{d}}\sum_{i\in[d]} r_i.$$

For d even, we can precisely compute this limit. Indeed, [85, Example 12.8.1] shows that

the free convolution $(r_1 + r_2)$ has arcsine distribution. Moreover, such a distribution was proved to be the law of $u + u^*$ in [85, Example 1.14], where u is a Haar unitary. We then deduce that

$$z \overset{d}{=} \frac{1}{\sqrt{d}} \sum_{i \in [d/2]} (u_i + u_i^*).$$

The distribution of the free sum of $u_i + u_i^*$ is called the Kesten-McKay distribution with parameter d; see [67],[85, Exercise 12.21]. Its density is given by

$$f_{KM(d)}(x) = \frac{1}{2\pi} \frac{d}{d^2 - x^2} \sqrt{4(d-1) - x^2}\, \mathbf{1}_{|x| \leq 2\sqrt{d-1}}.$$

It is also the limit spectral distribution of random d-regular graphs [78]. Therefore, the density of b is given by

$$\tilde{f}_{KM(d)}(x) = \frac{1}{2\pi} \frac{d}{d - x^2} \sqrt{4\Big(1 - \frac{1}{d}\Big) - x^2}\, \mathbf{1}_{|x| \leq 2\sqrt{1 - \frac{1}{d}}}.$$

A direct computation shows that, for all $x \in (-2, 2)$, $\tilde{f}_{KM(d)}(x) \to f_{sc}(x)$ as $d \to \infty$. This is a local version of Theorem (2.3.ii).

Let us interpret Example 2.9 above in terms of the corresponding random quantum channel. Given $W = URU^* \in M_n(\mathbb{C})$ a uniformly rotated Rademacher matrix, we have $W^2 = R^2 = \mathrm{Id}$. Hence, sampling $W_1, \ldots, W_d \in M_n(\mathbb{C})$ independently and uniformly rotated Rademacher matrices and setting $K_i = W_i/\sqrt{d}$ for each $i \in [d]$, the random completely positive map

$$\Phi : X \in M_n(\mathbb{C}) \mapsto \sum_{i \in [d]} K_i X K_i \in M_n(\mathbb{C})$$

is exactly trace-preserving and unital (not just on average), i.e. it is a random unital quantum channel. More precisely, it is a mixture of unitary conjugations. What is more, by what precedes, the spectral distribution of $\sqrt{d}(\Phi - \mathbb{E}(\Phi))$ converges weakly almost surely to $\tilde{\mu}_{KM(d)}$ (for d even) as $n \to \infty$. And the spectral distribution of $\sqrt{d}(\Phi - \mathbb{E}(\Phi))$ converges weakly in probability and in expectation to μ_{sc} as $n, d \to \infty$.

Remark 2.10. There is a straightforward generalization of the previous example to deterministic matrices. Let $M_1, \ldots, M_d \in M_n(\mathbb{C})$ be deterministic Hermitian matrices such that, for each $i \in [d]$, $M_i \Rightarrow \mu_i$ as $n \to \infty$ and μ is centered. A natural way to make them asymptotically free is by conjugating them with independent Haar unitaries. In particular, if we set, for each $i \in [d]$, $W_i = U_i M_i U_i^*$, where $U_1, \ldots, U_d \in M_n(\mathbb{C})$ are independent uniformly chosen unitaries, then $(W_i)_{i \in [d]}$ satisfies Assumptions (**A.1**),(**A.2**),(**A.3**) and (**A.4**). Indeed, for instance, Assumption (**A.3**) follows by Lemma 2.6 so that we can apply Theorem 2.3 to this kind of matrices.

Example 2.11 (Wigner matrices)**.** A random Hermitian matrix $W = (W_{kl})_{k,l\in[n]} \in M_n(\mathbb{C})$ is a Wigner matrix if W_{kl} are i.i.d. centered with variance $1/n$ for $k, l \in [n]$ such that $k \leq l$ and $W_{kl} = \overline{W_{lk}}$ for $k, l \in [n]$ such that $l < k$. It is well-known [7, 44, 109] that independent Wigner matrices $W_1, \ldots, W_d$ satisfy Assumptions (**A.1**), (**A.2**), (**A.3**) (by Lemma 2.6) and (**A.4**). Their joint law $(W_i)_{i\in[d]}$ converges weakly in probability to a free semicircular family $(s_i)_{i\in[d]}$. Therefore, Theorem 2.3 holds and the limit in Theorem (2.3.i) is expressed as

$$\Delta \underset{n\to\infty}{\Rightarrow} \frac{1}{\sqrt{d}} \sum_{i\in[d]} s_i \otimes s_i,$$

in probability and in expectation.

Let us interpret Example 2.11 above in terms of the corresponding random quantum channel. Given $W \in M_n(\mathbb{C})$ a Wigner matrix, we have $\mathbb{E}(W_{kl}) = 0$ and $\mathbb{E}\,|W_{kl}|^2 = 1/n$ for all $k, l \in [n]$ and $(W_{kl})_{k\leq l\in[n]}$ independent. Therefore,

$$\mathbb{E}\Big(W^2\Big) = \sum_{k\in[n]} E_{kk} = \mathrm{Id}\,.$$

So we can conclude that, sampling $W_1, \ldots, W_d \in M_n(\mathbb{C})$ independent Wigner matrices and setting $K_i = W_i/\sqrt{d}$ for each $i \in [d]$, the random completely positive map

$$\Phi : X \in M_n(\mathbb{C}) \mapsto \sum_{i\in[d]} K_i X K_i \in M_n(\mathbb{C})$$

is, on average, trace-preserving and unital. Moreover, the spectral distribution of $d(\Phi - \mathbb{E}(\Phi))$ converges weakly almost surely to $\mu_{sc}^{\boxtimes d}$ as $n \to \infty$, and the spectral distribution of $\sqrt{d}(\Phi - \mathbb{E}(\Phi))$ converges weakly in probability and in expectation to μ_{sc} as $n, d \to \infty$. In this case, we can actually explicitly compute $\mathbb{E}(\Phi)$. Indeed, denoting by $\{e_1, \ldots, e_n\}$ the canonical basis of $\mathbb{C}^n$ and setting $E_{kl} = e_k e_l^*$ for each $k, l \in [n]$, we have

$$\mathbb{E}\Big(W \otimes \overline{W}\Big) = \frac{1}{n} \sum_{k,l\in[n]} E_{kl} \otimes E_{kl} + \frac{1}{n} \sum_{k\neq l\in[n]} E_{kl} \otimes E_{lk} = \psi\psi^* + \frac{1}{n}(F - \mathrm{diag}(F)),$$

where $\psi = \sum_{k\in[n]} (e_k \otimes e_k)/\sqrt{n}$ is a maximally entangled unit vector, $F = \sum_{k,l\in[n]} E_{kl} \otimes E_{lk}$ is the flip operator and $\mathrm{diag}(F) = \sum_{k\in[n]} E_{kk} \otimes E_{kk}$ is its diagonal part (with respect to the canonical product basis of $\mathbb{C}^n \otimes \mathbb{C}^n$). $F - \mathrm{diag}(F)$ has spectrum $\{1, -1, 0\}$, where 1 and -1 have multiplicities $n(n-1)/2$ (with associated eigenvectors $\{(e_k \otimes e_l + e_l \otimes e_k)/\sqrt{2},\ k < l \in [n]\}$ and $\{(e_k \otimes e_l - e_l \otimes e_k)/\sqrt{2},\ k < l \in [n]\}$ respectively) and 0 has multiplicity n. Hence,

$$\mathbb{E}(M_\Phi) = \psi\psi^* + \frac{1}{n}(F - \mathrm{diag}(F)).$$

This can be re-written at the level of Φ as

$$\mathbb{E}(\Phi) : X \in M_n(\mathbb{C}) \mapsto \mathrm{tr}(X)\frac{\mathrm{Id}}{n} + \frac{1}{n}\Big(X^T - \mathrm{diag}(X)\Big) \in M_n(\mathbb{C}),$$

where X^T denotes the transposition of X and $\mathrm{diag}(X)$ its diagonal part, both with respect to the canonical basis of $\mathbb{C}^n$. This means that, up to a correction that vanishes as n grows, $\mathbb{E}(\Phi)$ is the so-called fully randomizing channel $\Pi : X \mapsto \mathrm{tr}(X)\,\mathrm{Id}\,/n$, which has one eigenvalue equal to 1, with associated eigenvector the maximally mixed state $\mathrm{Id}\,/n$, and all the other eigenvalues equal to 0. Our result thus gives a precise understanding of how the asymptotic spectrum of Φ deviates from the flat one of Π, for d either fixed or growing.

We can also consider non-homogeneous matrices as in [72, 17].

Example 2.12 (Non-homogeneous matrices)**.** Let $W_1, \ldots, W_d \in M_n(\mathbb{C})$ be independent Gaussian matrices such that, for each $i \in [d]$,

$$\|\mathbb{E}(W_i)\| \underset{n\to\infty}{\to} 0;$$
$$\left\|\mathbb{E}(W_i^2) - \mathrm{Id}\right\| \underset{n\to\infty}{\to} 0.$$

Then it was proven in [17, Theorem 2.10] that $(W_i)_{i\in[d]}$ are asymptotically free in expectation and almost surely as long as the covariance structure of each one goes to 0 sufficiently fast. Concretely, we have to impose that, for each $i \in [d]$,

$$v(W_i) := \|\mathrm{cov}(W_i)\| = o\Big(\log^{-3/2} n\Big).$$

The limit is a free semicircular family $(s_i)_{i\in[d]}$. So Theorem (2.3.i) implies that

$$\Delta \underset{n\to\infty}{\Rightarrow} \frac{1}{\sqrt{d}} \sum_{i\in[d]} s_i \otimes s_i,$$

in probability and in expectation. As a matter of fact, similar results hold for Hermitian random matrices with bounded entries (see [31, Theorem 3.25]).

Another classical example is Wishart-type matrices.

Example 2.13 (Wishart matrices)**.** A random Hermitian matrix $W \in M_n(\mathbb{C})$ is a (centered) Wishart matrix if

$$W = XX^* - \mathbb{E}(XX^*) = XX^* - \mathrm{Id},$$

where $X \in M_n(\mathbb{C})$ has i.i.d. centered entries with variance $1/n$. Independent Whishart matrices $W_1, \ldots, W_d$ also satisfy Assumptions (**A.1**), (**A.2**), (**A.3**) (by Lemma 2.6) and (**A.4**); see [35]. The limit of a Wishart matrix XX^* is equal to cc^*, where c is a circular element, i.e.

$$c = \frac{s + is'}{\sqrt{2}},$$

with s, s' free semicircular variables, and it is called the quarter circular law or Marchenko-

Pastur law. The limit law in (2.3.i) is given by

$$\Delta \underset{n\to\infty}{\Rightarrow} \frac{1}{\sqrt{d}} \sum_{j\in[d]} (c_j c_j^* - 1) \otimes (c_j c_j^* - 1),$$

in probability and in expectation, where $(c_j)_{j\in[d]}$ are free copies of c.

Chapter 3

Central Limit Theorem for tensor products of free variables

I want to break free!

Some random tensor product

We establish a central limit theorem for tensor product random variables $c_k := a_k \otimes a_k$, where $(a_k)_{k\in\mathbb{N}}$ is a free family of variables. We show that if the variables a_k are centered, the limiting law is the semi-circle. Otherwise, the limiting law depends on their mean and variance and corresponds to the free convolution of an explicit probability measure and the semi-circle distribution.

Contents

3.1 Introduction

The Free Central Limit Theorem serves as a foundational principle in free probability [110], [85, Lecture 8]. It asserts that as the number of freely independent operators summed

together approaches infinity, the distribution of the normalized sum tends towards an asymptotically semi-circular shape. This mirrors the classical Central Limit Theorem but with independence conditions replaced by free independence (also known as freeness) and the Gaussian limit substituted with a semi-circular limit. More precisely, let $(\mathcal{A}, \tau)$ be a unital noncommutative probability space equipped with a faithful tracial state τ [85, Lecture 1]. We say that subalgebras $\mathcal{A}_1, \ldots, \mathcal{A}_d \subset \mathcal{A}$ are free if

$$\tau(a_1 \ldots a_p) = 0,$$

whenever $a_i \in \mathcal{A}_{j_i}$, $\tau(a_i) = 0$ for all $i \in [p]$ and $j_1 \neq j_2 \neq \cdots \neq j_p$. We say that random variables $a_1, \ldots, a_d \in \mathcal{A}$ are free if their generated algebras are free. We say that a sequence of (self-adjoint) variables $a_n \in (\mathcal{A}_n, \tau_n)$ converges in distribution to a variable $a \in (\mathcal{A}, \tau)$ if

$$\tau_n(a_n^p) \to \tau(a^p),$$

for all integers $p \geq 0$ and we denote it $a_n \Rightarrow a$. We denote $a - \lambda := a - \lambda \mathbf{1}$, where $\mathbf{1} \in \mathcal{A}$ is the unit in the algebra. As usual, $\tau(a)$ is the mean of a and the variance is given by

$$\mathrm{var}(a) = \tau((a - \tau(a))^2).$$

The Free Central Limit Theorem states that if $a_1, \ldots, a_n \in (\mathcal{A}, \tau)$ are free iid random variables with mean λ and variance σ^2, then

$$\frac{1}{\sigma\sqrt{n}} \sum_{k \in [n]} (a_k - \lambda) \Rightarrow \mu_{sc},$$

where μ_{sc} denotes the semi-circle distribution whose density is given by

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

The goal of this paper is to establish a central limit theorem for the tensor product of free random variables. Concretely, given $a_1, \ldots, a_n \in (\mathcal{A}, \tau)$ free i.i.d random variables, we aim at studying the convergence of the normalized sequence

$$\frac{1}{\sqrt{n}} \sum_{k \in [n]} (a_k \otimes a_k - \tau \otimes \tau(a_k \otimes a_k)), \tag{3.1}$$

in the product space $(\mathcal{A} \otimes \mathcal{A}, \tau \otimes \tau)$.

Just as Free Probability captures the limiting behavior of random matrices, the above expression appears naturally as the limiting object corresponding to several models of random Quantum Channels [73]. Indeed, given $M_1, \ldots, M_n \in \mathcal{M}_d(\mathbb{C})$ independent random self-adjoint matrices, it was shown in [73] that the empirical spectral distribution

(ESD) of the quantum channel

$$\Delta_{d,n} := \frac{1}{\sqrt{n}} \sum_{k\in[n]} (M_k \otimes M_k - \mathbb{E}[M_k \otimes M_k]),$$

having the M_k's as random Kraus operators and with fixed Kraus rank n, converges as $d \to \infty$ to the expression in (3.1) with the a_k's being the corresponding limits of the ESD of the M_k's. Moreover, it was in particular shown that if the random matrices M_k are centered, then the ESD of $\Delta_{d,n}$ converges as $n, d \to \infty$ to the semi-circle distribution. These two statements combined suggest that, in the case where the a_k's are centered, an analogue of the free central limit theorem should hold for the $a_k \otimes a_k$'s. Whereas these heuristics indicate that the semi-circle distribution should appear as the limit of (3.1) when the a_k's are centered, the convergence and the explicit limit are not clear in the general case. The goal of this paper is to address this by establishing the convergence of the expression in (3.1) and identifying the limiting object. The latter, as we show, depends on the mean and variance of the variables a_k's and represents a free interpolation between an explicit measure and the semi-circle distribution.

Random matrix models of the form

$$M = \sum_{k\in[n]} M_k \otimes M_k, \tag{3.2}$$

for $M_1, \ldots, M_n \in \mathcal{M}_d(\mathbb{C})$ independent random self-adjoint matrices, are in fact useful in other areas of Quantum Information Theory. When the M_k's are positive semi-definite matrices, normalizing M by its trace produces a model for a random separable quantum state. Little is known about the typical asymptotic spectrum of separable states, contrary to that of entangled ones [6]. Moreover, a random matrix M of the form (3.2) appears naturally when performing a so-called realignment operation on a quantum state. Understanding the spectrum of the realignment of a state is important as it gives information on the entanglement of the state. In [9], this was done in the particular case where the M_k's are Gaussian matrices (corresponding to the case where the state is a normalized Wishart matrix). The results and techniques we develop here (combined with those in [73]) could be useful in addressing the questions mentioned above.

Given a measure μ, we denote

$$(t\mu)(A) := \mu(t^{-1}A),$$

its dilation by $t \neq 0$, where A is any Borel set in $\mathbb{R}$. Let $P_2^{\mathrm{bi}}(2p)$ be the set bipartite pair partitions, namely, those partitions whose intersection graph is bipartite (see definitions in Section 3.2). Define the symmetric measure ν such that its odds moments vanish and its even moments are given by

$$\int_{\mathbb{R}} x^{2p} \, \mathrm{d}\nu := |P_2^{\mathrm{bi}}(2p)|. \tag{3.3}$$

For the existence of such measure ν, see discussion in Remark 3.4.

The following is our main theorem.

Theorem 3.1. *Let $a \in (\mathcal{A}, \tau)$ be a self-adjoint random variable with mean $\tau(a) = \lambda$ and variance $\mathrm{var}(a) = \sigma^2 \neq 0$. Denote*

$$\delta^2 := \mathrm{var}(a \otimes a) = \sigma^2(\sigma^2 + 2\lambda^2),$$

and

$$q := \frac{2\lambda^2}{\sigma^2 + 2\lambda^2} \in [0, 1].$$

Given $(a_k)_{k\in\mathbb{N}}$ a sequence of free copies of a, the normalized sum

$$S_n := \frac{1}{\delta\sqrt{n}} \sum_{k\in[n]} (a_k \otimes a_k - \lambda^2)$$

converges as $n \to \infty$ to

$$\mu_q := \sqrt{\frac{q}{2}}\,\nu \boxplus \sqrt{\frac{q}{2}}\,\nu \boxplus \sqrt{1-q}\,\mu_{sc}, \tag{3.4}$$

where $\boxplus$ denotes the free convolution.

The difficulty in analyzing S_n stems from the complicated dependence structure exhibited by tensors, combining classical independence (between the two legs of the tensor) and freeness (between the variables across tensors). In the centered case, similar computations were made for semi-circle random variables [83, 42]. It would be of interest to design a general notion of independence corresponding to the tensor case, analyze its properties, derive the corresponding limit theorems, and characterize the corresponding universal objects. One particular generalization is by replacing the tensor product with the product of ε-independent random variables [98, 80, 97]. A direct consequence of Theorem 3.1 is that such a notion cannot, in general, reduce to freeness.

Corollary 3.2. *Let $a_1, \ldots, a_n \in (\mathcal{A}, \tau)$ be self-adjoint free i.i.d noncentered random variables. Then the $\{a_k \otimes a_k : k \in [n]\}$ are not free.*

The above corollary trivially follows from Theorem 3.1, since if the $a_k \otimes a_k$'s were free, the limit of their normalized sum would be the semi-circle distribution contradicting the conclusion of Theorem 3.1 (when $\lambda \neq 0$). This fact was originally proved in [37] where, more generally, the freeness of tensors of free variables was characterized.

This paper is organized as follows. In Section 3.2, we recall some definitions and notations. In Section 3.3, we provide some properties of the limiting measure appearing in Theorem 3.1. Section 3.4 establishes the existence of the limit, while Section 3.5 is dedicated to the proof of Theorem 3.1.

Acknowledgments

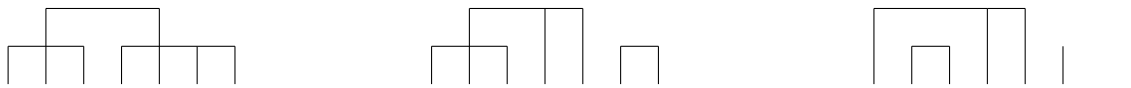
The last named author would like to thank Guillaume Cébron and Roland Speicher for helpful discussions. Part of this work was initiated during a stay of the second named author at New York University in Abu Dhabi, partly funded by a doctoral mobility grant delivered by Université Gustave Eiffel; he would like to thank both institutions for their support and the excellent working assumptions. The first named author was supported by the ANR projects ESQuisses (grant number ANR-20-CE47-0014-01), STARS (grant number ANR-20-CE40-0008) and QTraj (grant number ANR-20-CE40-0024-01).

3.2 Preliminaries and notations

Given $p \in \mathbb{N}$, a partition $\pi = \{V_1, \ldots, V_k\}$ of $[p]$ is a collection of disjoint sets $V_1, \ldots, V_k$ called blocks such that

$$V_1 \cup \cdots \cup V_k = [p].$$

We denote by $P(p)$ the set of partitions of $[p]$. We say that a partition $\pi \in P(p)$ is *connected* (also referred to as a linked diagram in [87]) if no proper subinterval of $[p]$ can be written as the union of blocks of π. A partition $\pi \in P(p)$ has a crossing $i < k < j < l$ if there exist two disjoint blocks $V_1, V_2 \in \pi$ such that $\{i, j\} \subset V_1$ and $\{k, l\} \subset V_2$. A block $V \in \pi$ is crossing if there exists another $V' \in \pi$ such that V' crosses V. We say that a partition $\pi \in P(p)$ is a *noncrossing* partition if all its blocks are noncrossing. We denote by $P^{\mathrm{con}}(p)$ (resp. $NC(p)$) the set of all connected (resp. noncrossing) partitions of $[p]$; see Figure 3.1.



(a) Connected partition (b) General partition (c) Noncrossing partition

Figure 3.1: Examples of partitions

Finally, for a partition $\pi \in P(p)$, we denote by $G(\pi)$ its *intersection graph*. It is the graph over the blocks of π such that two blocks are connected if they cross, under some arbitrary labeling. We say that a partition π is a bipartite partition if its intersection graph is bipartite and denote it $\pi \in P^{\mathrm{bi}}(p)$; see Figure 3.2 for the partitions in Figure 3.1.

(a) Connected partition (b) General partition (c) Noncrossing partition

Figure 3.2: Examples of graphs of crossings.

Given $\pi \in P(p)$, we denote $|\pi|$ its number of blocks and $\mathbf{cr}(\pi)$ its number of crossing

blocks. Therefore, the number of noncrossing blocks of π is

$$\mathbf{ncr}(\pi) := |\pi| - \mathbf{cr}(\pi).$$

Given $\pi \in P(p)$, we denote $\mathbf{cc}(\pi)$ its number of connected components. We denote $P_2(p)$, $P_2^{\mathrm{con}}(p)$, $P_2^{\mathrm{bi}}(p)$ and $NC_2(p)$ the set of pair partitions, connected pair partitions, bipartite pair partitions, and noncrossing pair partitions, respectively, that is, those partitions such that all of their blocks have cardinality two. Sometimes it will be useful to define similarly $P_2^{\mathrm{bicon}}(2p) := P_2^{\mathrm{bi}}(2p) \cap P_2^{\mathrm{con}}(2p)$ the set of bipartite connected pair partitions.

A pair partition $\pi \in P_2(2p)$ can be decomposed into its crossing connected components, namely, let $\hat{\pi} \in P(2p)$ be the choice of connected components and, for each block $T \in \hat{\pi}$, draw a connected partition $\pi_T \in P_2^{\mathrm{con}}(T)$. By definition, $\hat{\pi} \in NC(2p)$ as otherwise two disjoint components would meet ($\hat{\pi}$ is called the noncrossing closure of π in [77]). The mapping

$$\Phi : \pi \mapsto (\hat{\pi}, (\pi_T)_{T \in \hat{\pi}}) \tag{3.5}$$

is a bijection that will be used throughout the proof of Theorem 3.1; see Figure 3.3.

(a) π (b) $\hat{\pi}$

(c) $\pi_{\{1,2,3,4\}}$ (d) $\pi_{\{5,8\}} \cong \pi_{\{6,7\}}$

Figure 3.3: A partition π and its image $\Phi(\pi)$.

We denote

$$\mathrm{Proj}(\hat{\pi}) := \{(\pi_T)_{T \in \hat{\pi}}) : \pi_T \in P_2^{\mathrm{con}}(T),\ \forall T \in \hat{\pi}\}.$$

Note that $|\hat{\pi}| = \mathbf{cc}(\pi)$. At the level of cardinals, the above bijection implies that

$$|P_2(2p)| = \sum_{\hat{\pi} \in NC(2p)} \prod_{T \in \hat{\pi}} |P_2^{\mathrm{con}}(T)| = \sum_{\hat{\pi} \in NC(2p)} \sum_{(\pi_T)_{T \in \hat{\pi}}} 1, \tag{3.6}$$

where the second summation is over $(\pi_T)_{T \in \hat{\pi}} \in \mathrm{Proj}(\hat{\pi})$.

Given $a_1, \ldots, a_n \in \mathcal{A}$, we denote by $\kappa_n^{\mathrm{free}}(a_1, \ldots, a_n)$ their free cumulants, namely, for

any $i \in [n]^k$, we have

$$\tau(a_{i_1} \cdots a_{i_k}) = \sum_{\pi \in NC(k)} \kappa_\pi^{\text{free}}(a_{i_1}, \ldots, a_{i_k}),$$
$$\kappa_\pi^{\text{free}}(a_{i_1}, \ldots, a_{i_k}) = \prod_{V=\{v_1,\ldots,v_l\} \in \pi} \kappa_{|V|}^{\text{free}}(a_{i_{v_1}}, \ldots, a_{i_{v_l}}).$$

This is known as the moment-cumulant formula [85, Notation 11.5]. Note also that if the variables $a_1, \ldots, a_n$ are free, the free mixed cumulants vanish [85, Proposition 11.15]. Similarly, their classical cumulants $\kappa_n^{\text{class}}(a_1, \ldots, a_n)$ are defined by

$$\tau(a_{i_1} \cdots a_{i_k}) = \sum_{\pi \in P(k)} \kappa_\pi^{\text{class}}(a_{i_1}, \ldots, a_{i_k}),$$
$$\kappa_\pi^{\text{class}}(a_{i_1}, \ldots, a_{i_k}) = \prod_{V=\{v_1,\ldots,v_l\} \in \pi} \kappa_{|V|}^{\text{class}}(a_{i_{v_1}}, \ldots, a_{i_{v_l}}).$$

Here and throughout the paper, we denote $V = \{v_1, \ldots, v_l\}$ respecting the ordering, that is, $v_1 < \cdots < v_l$; see [85, Lecture 11]. We denote $\kappa_n^{\text{free}}(a)$ (resp. $\kappa_n^{\text{class}}(a)$) the free (resp. classical) cumulants of a random variable a. It follows from (3.6) that if $g \sim N(0,1)$, we have

$$\kappa_n^{\text{free}}(g) = |P_2^{\text{con}}(n)|.$$

This can also be deduced from a general relation between classical and free cumulants [77]

$$\kappa_n^{\text{free}}(a) = \sum_{\pi \in P^{\text{con}}(n)} \kappa_\pi^{\text{class}}(a).$$

We will equivalently denote $\kappa_n^{\text{free}}(\mu)$ (resp. $\kappa_n^{\text{class}}(\mu)$) the free (resp. classical) cumulants of a random variable a with distribution μ. Given two measures μ_a and μ_b, the free convolution $\mu_a \boxplus \mu_b$ denotes the distribution of $a + b$, where a and b are free random variables with distribution μ_a and μ_b, respectively.

3.3 Properties of the limiting measure

In this section, we summarize some of the properties of the measure μ_q appearing in (3.4).

We start by calculating the moments and free cumulants of μ_q.

Proposition 3.3. *Given $q \in [0,1]$, let*

$$\mu_q := \sqrt{\frac{q}{2}}\,\nu \boxplus \sqrt{\frac{q}{2}}\,\nu \boxplus \sqrt{1-q}\,\mu_{sc},$$

where ν is defined via Equation (3.3). Then the following hold:

1. *The free cumulants $\kappa_n^{\text{free}}(\mu_q)$ of μ_q vanish if n is odd, $\kappa_2^{\text{free}}(\mu_q) = 1$ and for any even*

$n \geq 4$, we have

$$\kappa_n^{\mathrm{free}}(\mu_q) = 2\left(\frac{q}{2}\right)^{n/2} |P_2^{\mathrm{bicon}}(n)|. \tag{3.7}$$

2. *The odd moments of μ_q vanish and, for every $p \in \mathbb{N}$, its $2p$-th moment is given by*

$$\sum_{\pi \in P_2^{\mathrm{bi}}(2p)} 2^{\mathbf{cc}(\pi)-p} q^{\mathbf{cr}(\pi)}. \tag{3.8}$$

Remark 3.4 (The moment problem [92])**.** It is unclear if the measure ν is well-defined, namely, if the sequence (a_n) defined via $a_{2n+1} = 0$ and $a_{2n} = |P_2^{\mathrm{bi}}(2n)|$ for all $n \geq 0$ is the moment sequence of some measure ν. Nonetheless, there exists a unique measure μ_q whose cumulants and moments are given by (3.7) and (3.8), respectively. Indeed, let $m_{2n+1} = 0$ and

$$m_{2n} = \sum_{\pi \in P_2^{\mathrm{bi}}(2p)} 2^{\mathbf{cc}(\pi)-p} q^{\mathbf{cr}(\pi)}$$

be the sequence of its moments. Hamburger's theorem [60] states that the existence of such measure μ_q is equivalent to the Hankel matrix $H_n = (m_{i+j})_{1 \leq i,j \leq n}$ being positive semi-definite. As we use the convergence of the moments to prove Theorem 3.1, we show that

$$\tau(S_n^p) \to m_p,$$

for all $p \geq 1$. In particular, by continuity, the sequence $(m_n)_n$ satisfies Hamburger's criterium, and then it is the moment sequence of some measure. The uniqueness holds by Carleman's criterium [2, 92], as its moments are controlled by the Gaussian moments.

Proof. We will use the notion of R-transform; see [85, Lecture 16]. For a random variable $a \in \mathcal{A}$, let

$$R_a(z) = \sum_{n \geq 1} \kappa_n^{\mathrm{free}}(a) z^n,$$

be its R-transform, defined as a formal series. The R-transform of a standard semi-circle law is given by

$$R_{\mu_{sc}}(z) = z^2,$$

whereas the R-transform of ν can easily be computed similarly to the Gaussian as

$$R_\gamma(z) = \sum_{n \geq 1} |P_2^{\mathrm{bicon}}(2n)| z^{2n}.$$

Since the free cumulants linearize the free convolution, we deduce that

$$\begin{aligned} R_{\mu_q}(z) &= 2R_{\sqrt{\frac{q}{2}}\gamma}(z) + R_{\sqrt{1-q}\mu_{sc}}(z) \\ &= 2R_{\gamma}\Big(\sqrt{\frac{q}{2}}z\Big) + R_{\mu_{sc}}\big(\sqrt{1-q}z\big) \\ &= z^2 + 2\sum_{n\geq 2}\left(\frac{qz^2}{2}\right)^n |P_2^{\text{bicon}}(2n)|. \end{aligned}$$

This proves the first part of the proposition. To prove the second part, we use the moment-cumulant formula to deduce that the odd moments vanish, while the $2p$-th moment can be expressed as

$$\begin{aligned} \sum_{\hat{\pi}\in NC(2p)}\prod_{T\in\hat{\pi}}\kappa_{|T|}^{\text{free}}(\mu_q) &= \sum_{\hat{\pi}\in NC(2p)}\prod_{\substack{T\in\hat{\pi}\\|T|\geq 4}} 2\Big(\frac{q}{2}\Big)^{|T|/2}|P_2^{\text{bicon}}(T)| \\ &= \sum_{\hat{\pi}\in NC(2p)}\sum_{(\pi_T)_{T\in\hat{\pi}}}\prod_{\substack{T\in\hat{\pi}\\|T|\geq 4}} 2\Big(\frac{q}{2}\Big)^{|T|/2}, \end{aligned}$$

where the second summation is over $(\pi_T)_{T\in\hat{\pi}} \in \text{Proj}(\hat{\pi})$, as in (3.6), such that $\pi_T \in P_2^{\text{bicon}}(T)$ for every $T \in \hat{\pi}$. Now noting that

$$|\{T \in \hat{\pi} : |T| \geq 4\}| = \mathbf{cc}(\pi) - \mathbf{ncr}(\pi),$$

and

$$\sum_{\substack{T\in\hat{\pi}\\|T|\geq 4}} \frac{|T|}{2} = \mathbf{cr}(\pi),$$

we finish the proof after using the bijection Φ from (3.5) to rewrite the above expression. □

3.4 Existence of the limit

The goal of this section is to show that the expression in (3.1) admits a limit that depends only on the first and second moments of the variables at hand. Let us first prove the following centering lemma.

Lemma 3.5. *Let $(\mathcal{A},\tau)$ be a unital faithful tracial noncommutative probability space. Let $a_1,\ldots,a_n \in (\mathcal{A},\tau)$ be free i.i.d random variables and let*

$$c_k := a_k \otimes a_k - \tau\otimes\tau(a_k\otimes a_k),$$

for every $k \in [n]$. Then for any $m \geq 1$ and $i_1, \ldots, i_m \in [n]$, we have

$$\tau \otimes \tau(c_{i_1} \cdots c_{i_m}) = 0,$$

whenever there exists an index i_l that is different from the others.

Proof. By cyclicity of the trace, we can assume that i_m is the index that is different from the others. In this case, we write

$$\tau \otimes \tau(c_{i_1} \cdots c_{i_m}) = \sum_{I \subseteq [m]} (-1)^{|I|} \tau \otimes \tau \left(\overrightarrow{\prod_{l \in I^c}} a_{i_l} \otimes a_{i_l} \right) \prod_{l \in I} \tau \otimes \tau(a_{i_l} \otimes a_{i_l}).$$

Here, $\overrightarrow{\prod}$ denotes the product respecting the ordering. Then,

$$\tau \otimes \tau(c_{i_1} \cdots c_{i_m}) = \sum_{I \subseteq [m]} (-1)^{|I|} \lambda^{2|I|} \tau^2 \left(\overrightarrow{\prod_{l \in I^c}} a_{i_l} \right),$$

where $\lambda := \tau(a_1)$. Splitting the summation depending on whether $m \in I$ or not, we can write

$$\tau \otimes \tau(c_{i_1} \cdots c_{i_m}) = \sum_{\substack{I \subseteq [m] \\ m \in I}} (-1)^{|I|} \lambda^{2|I|} \tau^2 \left(\overrightarrow{\prod_{l \in I^c}} a_{i_l} \right) + \sum_{\substack{I \subseteq [m] \\ m \in I^c}} (-1)^{|I|} \lambda^{2|I|} \tau^2 \left(\overrightarrow{\prod_{l \in I^c}} a_{i_l} \right).$$

For the first term, we set $I = \{m\} \cup I'$ with $I' \subseteq [m-1]$ to get

$$\sum_{\substack{I \subseteq [m] \\ m \in I}} (-1)^{|I|} \lambda^{2|I|} \tau^2 \left(\overrightarrow{\prod_{l \in I^c}} a_{i_l} \right) = \sum_{I' \subseteq [m-1]} (-1)^{|I'|+1} \lambda^{2|I'|+2} \tau^2 \left(\overrightarrow{\prod_{l \in (I')^c}} a_{i_l} \right).$$

For the second term, we use freeness of a_{i_m} from the rest of the a_i's so that

$$\sum_{\substack{I \subseteq [m] \\ m \in I^c}} (-1)^{|I|} \lambda^{2|I|} \tau^2 \left(\overrightarrow{\prod_{l \in I^c}} a_{i_l} \right) = \sum_{I' \subseteq [m-1]} (-1)^{|I'|} \lambda^{2|I'|} \tau^2 \left(\overrightarrow{\prod_{l \in (I')^c}} a_{i_l} \right) \tau^2(a_{i_m}).$$

Putting the above together, we deduce

$$\begin{aligned} \tau \otimes \tau(c_{i_1} \cdots c_{i_m}) = & \sum_{I' \subseteq [m-1]} (-1)^{|I'|+1} \lambda^{2|I'|+2} \tau^2 \left(\overrightarrow{\prod_{l \in (I')^c}} a_{i_l} \right) \\ & + \sum_{I' \subseteq [m-1]} (-1)^{|I'|} \lambda^{2|I'|+2} \tau^2 \left(\overrightarrow{\prod_{l \in (I')^c}} a_{i_l} \right). \end{aligned}$$

This finishes the proof. □

We are now ready to prove the existence of the limit and that it only depends on the first and second moments of the a_k's.

Proposition 3.6 (Existence)**.** *Let $(a_n)_{n\in\mathbb{N}} \in (\mathcal{A},\tau)$ be free iid random variables with mean λ, variance σ^2, and denote $\delta^2 := \operatorname{var}(a_1 \otimes a_1) = \sigma^2(\sigma^2 + 2\lambda^2)$. For every $k \in \mathbb{N}$, denote*

$$b_k := \frac{1}{\delta}\Big(a_k \otimes a_k - \tau \otimes \tau(a_k \otimes a_k)\Big),$$

and

$$S_n := \frac{1}{\sqrt{n}} \sum_{k\in[n]} b_k.$$

Then there exists a random variable $\mathbf{S} \in (\mathcal{A}',\tau')$ such that $S_n \Rightarrow \mathbf{S}$. Moreover, the law of $\mathbf{S}$ depends only on λ and σ, its odd moments vanish and

$$\tau'(\mathbf{S}^{2p}) = \sum_{\pi\in P_2(2p)} \tau \otimes \tau(b_{i_1} \cdots b_{i_{2p}}),$$

where $i \in [n]^{2p}$ is any sequence such that $i_j = i_k$ if and only $\{j,k\} \in \pi$.

Proof. We begin by writing

$$\tau \otimes \tau(S_n^p) = \frac{1}{n^{p/2}} \sum_{i\in[n]^p} \tau \otimes \tau(b_{i_1} \cdots b_{i_p}).$$

Since $b_1,\dots,b_n$ are identically distributed, the expression

$$\tau \otimes \tau(b_{i_1} \cdots b_{i_p}) \tag{3.9}$$

depends only on the partition $\pi = \pi(i) \in P(p)$ given by $l \sim_\pi k$ (that is, l,k belong to the same block of π) if and only if $i_l = i_k$. Denote the common value of (3.9) by $\tau \otimes \tau(\pi)$. Then we have

$$\tau \otimes \tau(S_n^p) = \frac{1}{n^{p/2}} \sum_{\pi\in P(p)} \tau \otimes \tau(\pi) \#\{i \in [n]^p : \pi(i) = \pi\}.$$

To count the cardinality, we choose an index for each block. Therefore, we have

$$\#\{i \in [n]^p : \pi(i) = \pi\} = n(n-1)\cdots(n-|\pi|+1) \sim n^{|\pi|}.$$

By Lemma 3.5, if π has a block of size 1, $\tau \otimes \tau(\pi) = 0$. Thus, we have

$$\tau \otimes \tau(S_n^p) = \sum_{\substack{\pi\in P(p)\\ |V|\geq 2;\forall V\in\pi}} \tau \otimes \tau(\pi) \frac{n(n-1)\cdots(n-|\pi|+1)}{n^{p/2}}.$$

Since $|V| \geq 2$ for all blocks $V \in \pi$, we have $|\pi| \leq p/2$. If there exists a block $V \in \pi$ such that $|V| \geq 3$, we immediately have $|\pi| < p/2$, and its contribution is negligible. In particular, this implies that the odd moments of S_n are asymptotically vanishing. We

deduce that

$$\lim_{n\to\infty} \tau\otimes\tau(S_n^p) = \lim_{n\to\infty} \sum_{\substack{\pi\in P(p)\\ |V|=2;\forall V\in\pi}} \tau\otimes\tau(\pi)\frac{n(n-1)\cdots(n-|\pi|+1)}{n^{p/2}}.$$

In this case, π is a pair partition and $|\pi| = p/2$, hence we deduce the formula

$$\lim_{n\to\infty} \tau\otimes\tau(S_n^p) = \sum_{\pi\in P_2(p)} \tau\otimes\tau(\pi),$$

which shows that S_n converges. To prove that the limit depends only on λ and σ, we write

$$\tau\otimes\tau(\pi) = \tau\otimes\tau(b_{i_1}\cdots b_{i_p}) = \frac{1}{\delta^p}\sum_{I\subseteq[p]}(-1)^{|I|}\lambda^{2|I|}\tau^2\left(\prod_{l\in I^c}^{\rightarrow} a_{i_l}\right),$$

where $\pi(i) = \pi$. By the moment-cumulant formula, we have

$$\tau\left(\prod_{l\in I^c}^{\rightarrow} a_{i_l}\right) = \sum_{\sigma\in NC(I^c)} \kappa_\sigma((a_{i_l})_{l\in I^c}).$$

Since the mixed cumulants of free variables vanish, the only partitions $\sigma\in NC(I^c)$ that contribute are those such that every block $V\in\sigma$ has cardinality at most two. We then have

$$\kappa_\sigma((a_{i_l})_{l\in I^c}) = \prod_{\substack{V\in\sigma\\ |V|=2}} \kappa_{|V|}(a,a)\prod_{\substack{V\in\sigma\\ |V|=1}}\kappa_{|V|}(a) = \sigma^{2\#\{V\in\sigma:|V|=2\}}\lambda^{\#\{V\in\sigma:|V|=1\}}.$$

This concludes the proof. □

3.5 Proof of Theorem 3.1

After proving the existence of the limit in the previous section, the goal here is to identify this limit as stated in Theorem 3.1. In all this section, $(a_n)_{n\in\mathbb{N}}$ denote free copies of a random variable a with mean λ and variance σ^2. Moreover, the common law of the normalized tensors will be denoted by

$$b = \frac{1}{\delta}\big(a\otimes a - \lambda^2\big), \tag{3.10}$$

where $\delta^2 = \mathrm{var}(a\otimes a) = \sigma^2(\sigma^2+2\lambda^2)$.

Throughout the proof, we will assume p is an even integer. Following Proposition 3.6, we denote $\mathbf{S}$ the limit of S_n and note that

$$\tau'(\mathbf{S}^p) = \sum_{\pi\in P_2(p)} \tau\otimes\tau(\pi),$$

where $\tau \otimes \tau(\pi) = \tau \otimes \tau(b_{i_1} \cdots b_{i_p})$ with $\pi(i) = \pi$.

The following lemma will be used throughout the proof.

Lemma 3.7. *Let $b_1, \ldots, b_n$ be identically distributed tensors of free variables with common distribution given in* (3.10) *and let d_1, d_2 be free from the variables $a_1, \ldots, a_n$. Then, for any $i \in [n]^p$, we have*

$$\tau \otimes \tau(d_1 \otimes d_2 b_{i_1} \cdots b_{i_p}) = \tau(d_1)\tau(d_2)\tau \otimes \tau(b_{i_1} \cdots b_{i_p}).$$

Proof. We begin by writing again

$$\tau \otimes \tau(d_1 \otimes d_2 b_{i_1} \cdots b_{i_p}) = \frac{1}{\delta^p} \sum_{I \subseteq [p]} (-1)^{|I|} \lambda^{2|I|} \tau \otimes \tau\left(d_1 \otimes d_2 \prod_{l \in I^c}^{\rightarrow} a_{i_l} \otimes a_{i_l}\right).$$

By freeness, we can remove $\tau(d_1)$ and $\tau(d_2)$ from the trace, hence

$$\tau \otimes \tau(d_1 \otimes d_2 b_{i_1} \cdots b_{i_p}) = \frac{\tau(d_1)\tau(d_2)}{\delta^p} \sum_{I \subseteq [p]} (-1)^{|I|} \lambda^{2|I|} \tau \otimes \tau\left(\prod_{l \in I^c}^{\rightarrow} a_{i_l} \otimes a_{i_l}\right).$$

Reducing the summation back to $\tau \otimes \tau(b_{i_1} \cdots b_{i_p})$, the conclusion follows. □

3.5.1 Contribution of noncrossing blocks

We begin by removing interval blocks.

Lemma 3.8. *Let $\pi \in P_2(p)$ and suppose that there exists $l \in [p]$ such that $\{l, l+1\} \in \pi$ (with the convention that $p+1 := 1$). Then*

$$\tau \otimes \tau(\pi) = \tau \otimes \tau(\pi \setminus \{l, l+1\}).$$

Proof. By cyclicity, we can assume $l = p - 1$. We then have

$$\begin{aligned}
\delta^2 \tau \otimes \tau(\pi) &= \tau \otimes \tau\Big(b_{i_1} \cdots b_{i_{p-2}} \cdot \Big(a_{i_p}^2 \otimes a_{i_p}^2 + \lambda^4 1 \otimes 1 - 2\lambda^2 a_{i_p} \otimes a_{i_p}\Big)\Big) \\
&= \tau \otimes \tau\Big(b_{i_1} \cdots b_{i_{p-2}} \cdot a_{i_p}^2 \otimes a_{i_p}^2\Big) + \lambda^4 \tau \otimes \tau\Big(b_{i_1} \cdots b_{i_{p-2}}\Big) \\
&\quad - 2\lambda^2 \tau \otimes \tau\Big(b_{i_1} \cdots b_{i_{p-2}} a_{i_p} \otimes a_{i_p}\Big) \\
&=: I + II + III.
\end{aligned}$$

We immediately recognize

$$II = \lambda^4 \tau \otimes \tau(\pi \setminus \{p-1, p\}).$$

By Lemma 3.7, we have

$$I = \tau^2(a^2)\tau \otimes \tau(\pi \setminus \{p-1, p\});$$
$$III = -2\lambda^4 \tau \otimes \tau(\pi \setminus \{p-1, p\}).$$

Hence

$$\delta^2 \tau \otimes \tau(\pi) = \left(\tau^2(a^2) - \lambda^4\right)\tau \otimes \tau(\pi \setminus \{p-1, p\}).$$

To conclude, we note that $\tau^2(a^2) - \lambda^4 = \delta^2$ and finish the proof. □

Recall that a noncrossing pair partition $\pi \in NC_2(p)$ always has an interval block $V = \{l, l+1\} \in \pi$ such that $\pi \setminus V$ is a noncrossing pair partition. In particular, by induction, Lemma 3.8 implies the following.

Corollary 3.9. *For any $\pi \in NC_2(p)$, we have $\tau \otimes \tau(\pi) = 1$.*

3.5.2 Decomposition of pair partitions

In order to capture the contribution of crossing partitions, we need to decompose a partition $\pi \in P_2(p) \setminus NC_2(p)$ using smaller partitions. We denote $\pi = \pi_1 \oplus \cdots \oplus \pi_k$ if $[p]$ can be decomposed into k intervals $I_1, \ldots, I_k$ such that $\pi_k \in P_2(I_k)$ and $V \in \pi$ if $V \in \pi_l$ for some $1 \leq l \leq k$. For $I \subseteq [p]$, let

$$a_I := \prod_{l \in I}^{\rightarrow} a_{i_l}.$$

Lemma 3.10. *Let $\pi \in P_2(p)$. Then, the following holds.*

(3.10.i) If $\pi = \pi_1 \oplus \cdots \oplus \pi_l$, then

$$\tau \otimes \tau(\pi) = \tau \otimes \tau(\pi_1) \cdots \tau \otimes \tau(\pi_l).$$

(3.10.ii) If $\pi = \{1, p\} \cup \pi_1$, where $\pi_1 \in P_2(\{2, \ldots, p-1\})$, then

$$\tau \otimes \tau(\pi) = \tau \otimes \tau(\pi_1).$$

Moreover, if there exists an interval $I \subseteq [p]$ such that $\pi|_I$ is a pair partition, then

$$\tau \otimes \tau(\pi) = \tau \otimes \tau(\pi|_I)\tau \otimes \tau(\pi|_{I^c}).$$

(3.10.iii) If π has a block $V = \{r, s\}$ such that for any block $U = \{l, k\} \in \pi$ with $r < l < s$, we have $r < k < s$ (i.e., every point inside V matches another one inside V), we

have

$$\tau \otimes \tau(\pi) = \tau \otimes \tau(\pi|_{V_-})\tau \otimes \tau(\pi|_{V_+}),$$

where $\pi|_{V_-}$ *is the restriction of* π *to inside of* V *and* $\pi|_{V_+}$ *is the restriction of* π *to outside of* V*.*

Proof. (3.10.i) By induction, it suffices to prove the case $\pi = \pi_1 \oplus \pi_2$. Let I_1, I_2 be the disjoint decomposition of $[p]$ given by π_1 and π_2. Then, we can write

$$\tau \otimes \tau(\pi) = \frac{1}{\delta^p} \sum_{\substack{J_1 \subseteq I_1 \\ J_2 \subseteq I_2}} (-1)^{|J_1|+|J_2|} \lambda^{2(|J_1|+|J_2|)} \tau^2 \left(\prod_{l \in J_1^c} a_{i_l} \prod_{l \in J_2^c} a_{i_l} \right).$$

Since π is the direct sum of π_1, π_2, the variables $a_{J_1^c}, a_{J_2^c}$ are free and we can write

$$\tau \otimes \tau(\pi) = \frac{1}{\delta^p} \sum_{\substack{J_1 \subseteq I_1 \\ J_2 \subseteq I_2}} (-1)^{|J_1|+|J_2|} \lambda^{2(|J_1|+|J_2|)} \tau^2 \left(\prod_{l \in J_1^c} a_{i_l} \right) \tau^2 \left(\prod_{l \in J_2^c} a_{i_l} \right).$$

It is immediate to check that the right-hand-side is equal to $\tau \otimes \tau(\pi_1)\tau \otimes \tau(\pi_2)$.

(3.10.ii) By cyclicity, we have

$$\tau \otimes \tau(\pi) = \tau \otimes \tau(b_{i_2} \cdots b_{i_{p-1}} b_{i_p} b_{i_1}).$$

Since $\{1, p\} \in \pi$, Lemma 3.8 implies that

$$\tau \otimes \tau(\pi) = \tau \otimes \tau(b_{i_2} \cdots b_{i_{p-1}}) = \tau \otimes \tau(\pi_1).$$

The second part follows again by cyclicity as we can assume $I = \{1, \ldots, k\}$ for some $k \in [p]$, and the variables are free.

(3.10.iii) By cyclicity, we can assume that $V = \{1, k\}$ for some $k \in [p]$. Note that V creates a direct sum $\pi = (V \cup \pi|_{V_-}) \oplus \pi|_{V_+}$. The result follows by (3.10.i) and (3.10.ii). □

Note that any block $V \in \pi$ that does not cross any other block of π is either an interval block or a block that satisfies (3.10.iii). In particular, its removal does not affect the value of $\tau \otimes \tau(\pi)$. It is clear then that $\tau \otimes \tau(\pi)$ is a multiplicative function [30] over the connected components of π. Using the mapping Φ defined in (3.5), we can write

$$\tau \otimes \tau(\pi) = \prod_{T \in \hat{\pi}} \tau \otimes \tau(\pi_T), \tag{3.11}$$

where $(\hat{\pi}, (\pi_T)_{T \in \hat{\pi}}) = \Phi(\pi)$. In view of this, we will now focus on the case where $\pi \in P_2^{\mathrm{con}}(p)$, for $p \geq 4$, as the case $p = 2$ corresponds to noncrossing blocks.

3.5.3 Contribution of connected partitions

Given $\pi \in P_2^{\text{con}}(p)$, for an even integer $p \geq 4$, we recall its intersection graph $G(\pi)$, where its vertices are the blocks (under some arbitrary labeling) of π, and there is an edge between two vertices if they cross. We recall that $P_2^{\text{bicon}}(p)$ is the set of connected pair partitions π whose intersection graph is bipartite.

The following is the main proposition of this subsection.

Proposition 3.11. *Let $p \geq 4$ be an even integer and $\pi \in P_2^{\text{con}}(p)$. Then, the following hold.*

1. *If $\pi \notin P_2^{\text{bicon}}(p)$, $\tau \otimes \tau(\pi) = 0$;*

2. *If $\pi \in P_2^{\text{bicon}}(p)$, we have*

$$\tau \otimes \tau(\pi) = 2\left(\frac{q}{2}\right)^{\frac{p}{2}},$$

where $q = \frac{2\lambda^2}{\sigma^2+2\lambda^2}$.

To prove Proposition 3.11, we require some notation. We uniquely order the blocks of π, $V_1, \ldots, V_{p/2}$, as follows. V_1 is the block associated with $1 \in [p]$. Now, given $V_1, \ldots, V_k$, we define V_{k+1} as the block with the lowest element among all remainder blocks such that V_{k+1} crosses at least one V_j, for $j < k$. This can always be accomplished as π is a connected pair partition. We call such ordering the canonical ordering. We denote N_j, the set of all block neighbors of $V_j(\pi)$ in the intersection graph of π.

Given integer $k \geq 0$, we define

$$\tau \otimes \tau_k(\pi) := \tau \otimes \tau(\pi, V_1, \ldots, V_k) = \tau \otimes \tau\left(\prod_{l \in [p]} \overline{b}_{i_l}\right)$$

recursively as follows. Let $(\tilde{a}_l)_{l \in \mathbb{N}}$ be a free independent family of copies of a, free from $(a_l)_{l \in \mathbb{N}}$. Given the variables

$$a_j^{(l)}, a_j^{(r)} \in \{a_j, \tilde{a_j}\},$$

and $\tilde{b}_j = a_j^{(l)} \otimes a_j^{(r)} - \lambda^2$ defined via

$$\tau \otimes \tau_k(\pi) := \tau \otimes \tau(\pi, V_1, \ldots, V_k) = \tau \otimes \tau\left(\prod_{j \in [p]} \tilde{b}_{i_j}\right),$$

we define $\tau \otimes \tau_{k+1}(\pi)$ for $V_{k+1} = \{r, s\}$ by the following procedure.

1. If $k = 0$, for all blocks $V_l = \{r^*, s^*\} \in N_1(\pi)$ with unique element $r < s^* < s$, we

define

$$\tilde{b}_{i_{r^*}} = a_{i_{r^*}} \otimes a_{i_{r^*}} - \lambda^2;$$
$$\tilde{b}_{i_{s^*}} = \tilde{a}_{i_{s^*}} \otimes a_{i_{s^*}} - \lambda^2,$$

that is, we replace the left leg of the tensor for the element inside V_1, and the left legs of the tensors are free independent. The rest of the variables are kept unchanged.

2. If we can write the joint law of (b_{i_r}, b_{i_s}) as

$$\tilde{b}_{i_r} = a_{i_r} \otimes \tilde{a}_{i_r} - \lambda^2;$$
$$\tilde{b}_{i_s} = a_{i_s} \otimes a_{i_s} - \lambda^2,$$

that is, the right legs of the tensors are free independent, for all $V_l = \{r^*, s^*\} \in N_{k+1}(\pi)$ with unique element $r < s^* < s$, we define

$$\overline{b}_{i_{r^*}} = a_{i_{r^*}} \otimes a^{(r)}_{i_{r^*}} - \lambda^2;$$
$$\overline{b}_{i_{s^*}} = \tilde{a}_{i_{s^*}} \otimes a^{(r)}_{i_{s^*}} - \lambda^2,$$

that is, we replace the left leg of the tensor for the element inside V_{k+1}, and the left legs of the tensors are free independent. The rest of the variables are kept unchanged.

3. If we can write the joint law of (b_{i_r}, b_{i_s}) as

$$\tilde{b}_{i_r} = \tilde{a}_{i_r} \otimes a_{i_r} - \lambda^2;$$
$$\tilde{b}_{i_s} = a_{i_s} \otimes a_{i_s} - \lambda^2,$$

that is, the left legs of the tensors are free independent, for all $V_l = \{r^*, s^*\} \in N_{k+1}(\pi)$ with unique element $r < s^* < s$, we define

$$\overline{b}_{i_{r^*}} = a^{(l)}_{i_{r^*}} \otimes a_{i_{r^*}} - \lambda^2;$$
$$\overline{b}_{i_{s^*}} = a^{(l)}_{i_{s^*}} \otimes \tilde{a}_{i_{s^*}} - \lambda^2,$$

that is, we replace the right leg of the tensor for the element inside V_{k+1}, and the right legs of the tensors are free independent. The rest of the variables are kept unchanged.

4. Otherwise, we can write the joint law of (b_{i_r}, b_{i_s}) as

$$\tilde{b}_{i_r} = \tilde{a}_{i_r} \otimes a_{i_r} - \lambda^2;$$
$$\tilde{b}_{i_s} = a_{i_s} \otimes \tilde{a}_{i_s} - \lambda^2,$$

that is, both legs of the tensors are free independent. Then, for all $V_l = \{r^*, s^*\} \in$

$N_{k+1}(\pi)$ with unique element $r < s^* < s$, we define

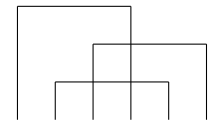
$$(\overline{b}_{i_{r^*}}, \overline{b}_{i_{s^*}}) = (\tilde{b}_{i_{r^*}}, \tilde{b}_{i_{s^*}}).$$

The rest of the variables are kept unchanged.

We denote $\tau \otimes \tau_0(\pi) := \tau \otimes \tau(\pi)$. Finally, we define

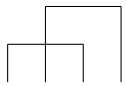
$$\begin{aligned}\tau \otimes \tau_{k_1,k_2}(\pi) :&= \tau \otimes \tau(\pi \setminus \{V_1, \ldots, V_{k_1}\}, V_1, \ldots, V_{k_2}) \\ &= \tau \otimes \tau \left(\overset{\rightarrow}{\prod_{\substack{j \in [p] \\ l \notin V_1 \cup \cdots \cup V_{k_1}}}} \overline{b}_{i_j} \right),\end{aligned}$$

where the variables $\overline{b}_j$ are defined as in $\tau \otimes \tau_{k_2}(\pi)$; see Figure 3.4 for the partition $\pi = \{\{1,4\},\{2,5\},\{3,6\}\}$, $V_1 = \{1,4\}, V_2 = \{2,5\}$, and where the top index indicates which legs of the tensors are replaced.



$b_1\ b_2\ b_3\ b_1\ b_2\ b_3$

(a) $\tau(\pi)$



$b_2^l\ b_3^l\ b_2\ b_3$

(b) $\tau(\pi \setminus V_1, V_1)$

$b_3^{lr}\ b_3$

(c) $\tau(\pi \setminus \{V_1, V_2\}, V_1, V_2)$

Figure 3.4: The terms in $\tau(\pi), \tau(\pi \setminus V_1, V_1), \tau(\pi \setminus \{V_1, V_2\}, V_1, V_2)$, respectively.

Summarily, fix the variables $\tilde{b}_j$ defined in $\tau \otimes \tau_k(\pi)$. If the pair $(\overline{b}_{i_r}, \overline{b}_{i_s})$ associated with V_{k+1} has a free independent right leg, we change the left leg of all crossing blocks of V_{k+1} on the element that lies inside V_{k+1}. If it has a free independent left leg, we change the right leg of all its crossing blocks on the element that lies inside V_{k+1}. If both legs are free independent, we do not change the variables. The rest of the blocks remain unchanged.

We begin with the following moment-cumulant formula; see [85, Lecture 5, Equation 5.6].

Lemma 3.12. *Let a, c_1, c_2 be variables such that a is free from $\{c_1, c_2\}$. Then*

$$\tau(ac_1ac_2) = \mathrm{var}(a)\tau(c_1)\tau(c_2) + \tau^2(a)\tau(c_1c_2).$$

The following proposition relates $\tau \otimes \tau(\pi)$ to $\tau \otimes \tau(\pi \setminus V, V)$ for some block $V \in \pi$.

Lemma 3.13. *Let $p \geq 4$ be an even integer, $\pi \in P_2^{\mathrm{con}}(p)$ and $V = \{r, s\} \in \pi$. Then*

$$\tau \otimes \tau(\pi) = q\, \tau \otimes \tau(\pi \setminus V, V),$$

where $q = \frac{2\lambda^2}{\sigma^2 + 2\lambda^2}$.

Proof. By cyclicity, we can assume that $r = 1$. Setting $b = b_{i_1}$, we write

$$\tau \otimes \tau(\pi) = \tau \otimes \tau\left(b \prod_{1<l<s}^{\rightarrow} b_{i_l} b \prod_{s<l\leq p}^{\rightarrow} b_{i_l}\right).$$

Expanding the variable b, we get

$$\tau \otimes \tau(\pi) = \frac{1}{\delta^2}\tau \otimes \tau\left(a \otimes a \prod_{1<l<s}^{\rightarrow} b_{i_l} a \otimes a \prod_{s<l\leq p}^{\rightarrow} b_{i_l}\right) - \frac{\lambda^4}{\delta^2}\tau \otimes \tau(\pi \setminus V), \tag{3.12}$$

where again we used Lemma 3.7 (see the proof of Lemma 3.8). To simplify the notation, let $I_1 = \{2, \ldots, s-1\}$ and $I_2 = \{s+1, \ldots, p\}$. Then,

$$\frac{1}{\delta^2}\tau \otimes \tau\left(a \otimes a \prod_{1<l<s}^{\rightarrow} b_{i_l} a \otimes a \prod_{s<l\leq p}^{\rightarrow} b_{i_l}\right) = \frac{1}{\delta^p} \sum_{\substack{J_1 \subseteq I_1 \\ J_2 \subseteq I_2}} (-1)^{|J_1|+|J_2|} \lambda^{2(|J_1|+|J_2|)} \tau^2\Big(a a_{J_1^c} a a_{J_2^c}\Big).$$

Using Lemma 3.12, we get

$$\begin{aligned}
&\frac{1}{\delta^2}\tau \otimes \tau\left(a \otimes a \prod_{1<l<s}^{\rightarrow} b_{i_l} a \otimes a \prod_{s<l\leq p}^{\rightarrow} b_{i_l}\right) \\
&= \frac{1}{\delta^p} \sum_{\substack{J_1 \subseteq I_1 \\ J_2 \subseteq I_2}} (-1)^{|J_1|+|J_2|} \lambda^{2(|J_1|+|J_2|)} \Big(\sigma^2 \tau\Big(a_{J_1^c}\Big)\tau\Big(a_{J_2^c}\Big) + \lambda^2 \tau\Big(a_{J_1^c} a_{J_2^c}\Big)\Big)^2.
\end{aligned}$$

We expand the square and note that

$$\tau \otimes \tau(\pi \setminus V) = \tau \otimes \tau(b_{I_1} b_{I_2}) = \frac{1}{\delta^{p-2}} \sum_{\substack{J_1 \subseteq I_1 \\ J_2 \subseteq I_2}} (-1)^{|J_1|+|J_2|} \lambda^{2(|J_1|+|J_2|)} \tau^2\Big(a_{J_1^c} a_{J_2^c}\Big),$$

and

$$\tau \otimes \tau(b_{I_1})\tau \otimes \tau(b_{I_2}) = \frac{1}{\delta^{p-2}} \sum_{\substack{J_1 \subseteq I_1 \\ J_2 \subseteq I_2}} (-1)^{|J_1|+|J_2|} \lambda^{2(|J_1|+|J_2|)} \tau^2\Big(a_{J_1^c}\Big)\tau^2\Big(a_{J_2^c}\Big).$$

Moreover, since π is connected and V is a crossing block of π, at least one index $l \in I_1$ is isolated in I_1, i.e., its matching symbol is in I_2. By Lemma 3.5, we therefore have

$$\tau \otimes \tau(b_{I_1})\tau \otimes \tau(b_{I_2}) = 0.$$

Combining the above identities in (3.12), we get

$$\tau \otimes \tau(\pi) = \frac{2\lambda^2\sigma^2}{\delta^p} \sum_{\substack{J_1 \subseteq I_1 \\ J_2 \subseteq I_2}} (-1)^{|J_1|+|J_2|} \lambda^{2(|J_1|+|J_2|)} \tau\Big(a_{J_1^c}\Big)\tau\Big(a_{J_2^c}\Big)\tau\Big(a_{J_1^c} a_{J_2^c}\Big).$$

It remains to check that the summation is precisely the one from $\tau \otimes \tau(\pi \setminus V, V)$. Indeed,

we have

$$\tau \otimes \tau(\pi \setminus V, V) = \frac{1}{\delta^{p-2}} \sum_{\substack{J_1 \subseteq I_1 \\ J_2 \subseteq I_2}} (-1)^{|J_1|+|J_2|} \lambda^{2(|J_1|+|J_2|)} \tau\Big(a_{J_1^c} a_{J_2^c}\Big) \tau\Big(\tilde{a}_{J_1^c} a_{J_2^c}\Big).$$

By freeness, the second product factorizes, and the conclusion holds. □

The following is the main induction step, and it is an analogue of Lemma 3.13 for $\tau \otimes \tau(\pi \setminus V, V)$.

Lemma 3.14. *Let $p \geq 4$ be an even integer, $\pi \in P_2^{\mathrm{con}}(p)$ and V_1, V_2 be two crossing blocks of π. Then*

$$\tau \otimes \tau(\pi \setminus V_1, V_1) = \frac{q}{2} \tau \otimes \tau(\pi \setminus \{V_1, V_2\}, V_1, V_2),$$

where $q = \frac{2\lambda^2}{\sigma^2 + 2\lambda^2}$.

Proof. By cyclicity of the trace, we assume $V_1 = \{1, k\}$ and $V_2 = \{r, s\}$ for $r < k < s$ and let $b = b_{i_r}$. We begin by writing

$$\tau \otimes \tau(\pi \setminus V_1, V_1) = \tau \otimes \tau\left(\prod_{1<l<r} \tilde{b}_{i_l} \tilde{b} \prod_{r<l<s} \bar{b}_{i_l} b \prod_{s<l\leq p} b_{i_l} \right),$$

where to simplify the notation, we used $\bar{b}_{i_l} = \tilde{b}_{i_l}$ for $l < k$, $\bar{b}_{i_l} = b_{i_l}$ for $l > k$ and $\bar{b}_{i_k} = 1$. By opening the expression for b, we get by Lemma 3.7 that

$$\begin{aligned} \tau \otimes \tau(\pi \setminus V_1, V_1) = \frac{1}{\delta^2} \tau \otimes \tau\left(\prod_{1<l<r} \tilde{b}_{i_l} a \otimes \tilde{a} \prod_{r<l<s} \bar{b}_{i_l} a \otimes a \prod_{s<l\leq p} b_{i_l} \right) \\ - \frac{\lambda^4}{\delta^2} \tau \otimes \tau(\pi \setminus \{V_1, V_2\}, V_1). \end{aligned}$$

Let $I_1 = \{2, \ldots, r-1\}, I_2 = \{r+1, \ldots, s-1\}, I_3 = \{s+1, \ldots, p\}$. Then, the first term can be computed as

$$\begin{aligned} &\frac{1}{\delta^2} \tau \otimes \tau\left(\prod_{1<l<r} \tilde{b}_{i_l} a \otimes \tilde{a} \prod_{r<l<s} \bar{b}_{i_l} a \otimes a \prod_{s<l\leq p} b_{i_l} \right) \\ &= \frac{1}{\delta^{p-2}} \sum_{\substack{J_1 \subseteq I_1 \\ J_2 \subseteq I_2 \\ J_3 \subseteq I_3}} (-1)^{|J_1|+|J_2|+|J_3|} \lambda^{2(|J_1|+|J_2|+|J_3|)} \tau\Big(a_{J_1^c} a a_{J_2^c} a a_{J_3^c}\Big) \tau\Big(\tilde{a}_{J_1^c} \tilde{a} \bar{a}_{J_2^c} a a_{J_3^c}\Big). \end{aligned}$$

By freeness and Lemma 3.12, we get

$$\begin{aligned}
&\frac{1}{\delta^2}\tau\otimes\tau\left(\prod_{1<l<r}\tilde{b}_{i_l}a\otimes\tilde{a}\prod_{r<l<k}\bar{b}_{i_l}a\otimes a\prod_{k<l\leq p}b_{i_l}\right)\\
&=\frac{1}{\delta^{p-2}}\sum_{\substack{J_1\subseteq I_1\\J_2\subseteq I_2\\J_3\subseteq I_3}}(-1)^{|J_1|+|J_2|+|J_3|}\lambda^{2(|J_1|+|J_2|+|J_3|)}\sigma^2\tau\Big(a_{J_1^c}a_{J_3^c}\Big)\tau\Big(a_{J_2^c}\Big)\lambda^2\tau\Big(\tilde{a}_{J_1^c}\overline{a}_{J_2^c}a_{J_3^c}\Big)\\
&\quad+\frac{1}{\delta^{p-2}}\sum_{\substack{J_1\subseteq I_1\\J_2\subseteq I_2\\J_3\subseteq I_3}}(-1)^{|J_1|+|J_2|+|J_3|}\lambda^{2(|J_1|+|J_2|+|J_3|)}\lambda^2\tau\Big(a_{J_1^c}a_{J_2^c}a_{J_3^c}\Big)\lambda^2\tau\Big(\tilde{a}_{J_1^c}\overline{a}_{J_2^c}a_{J_3^c}\Big).
\end{aligned}$$

The second term cancels out with $-\frac{\lambda^4}{\delta^2}\tau\otimes\tau(\pi\setminus\{V_1,V_2\},V_1)$. The first is equal to $\frac{\lambda^2\sigma^2}{\delta^2}\tau\otimes\tau(\pi\setminus\{V_1,V_2\},V_1,V_2)$, and the conclusion follows. □

It turns out, however, that V_3 cannot cross both V_1 and V_2.

Lemma 3.15. *Let $p\geq 4$ be an even integer, $\pi\in P_2^{\mathrm{con}}(p)$ and $V_1,\ldots,V_{p/2}$ be the canonical ordering of blocks of π. If V_3 crosses both V_1 and V_2, we have*

$$\tau\otimes\tau(\pi)=0.$$

Proof. We repeat the proof of Lemma 3.14. Let $V_j=\{r_j,s_j\}$ for $j=1,2,3$ and $a=a_{i_{r_3}}$. By cyclicity, we can assume $r_1<r_2<s_1<s_2$. There is only one possible configuration for V_3: V_3 starts inside V_1 and V_2 and ends outside both, as in Figure 3.5:

$$r_1<r_2<r_3<s_1<s_2<s_3;$$

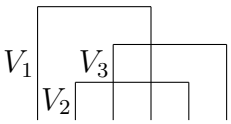


Figure 3.5: Three crossing structure.

We then have that $b_{i_{r_3}}$ gets both legs replaced, thus

$$\begin{aligned}
\tau\otimes\tau(\pi\setminus\{V_1,V_2\},V_1,V_2)=\frac{1}{\delta^2}\tau\otimes\tau\left(\prod_{1\leq l<r_3}\bar{b}_{i_l}\tilde{a}\otimes\tilde{a}\prod_{r_3<l<s_3}\bar{b}_{i_l}a\otimes a\prod_{s_3<l\leq p}\bar{b}_{i_l}\right)\\
-\frac{\lambda^4}{\delta^2}\tau\otimes\tau(\pi\setminus\{V_1,V_2,V_3\},V_1,V_2).
\end{aligned}$$

where the products and the definition of $\bar{b}$ follows the blocks of $\pi\setminus\{V_1,V_2\}$ and the replacements done by V_1 and V_2. By freeness and Lemma 3.13, we immediately get that the first term is equal to the second. □

We are ready to prove Proposition 3.11. To this end, denote $\omega:[p/2]\to\{0,1,2\}$ the coloring of vertices of $G(\pi)$ under the canonical ordering such that $\omega(l)=0$ if V_l replaces

the left leg of the tensor (Cases 1 and 2), $\omega(l) = 1$ if it replaces the right leg (Case 3) and $\omega(l) = 2$ otherwise (Case 4) in the definition of $\tau \otimes \tau_{k_1,k_2}(\pi)$, see Figure 3.4.

Proof of Proposition 3.11. Let $\pi \notin P_2^{\text{bicon}}(p)$ and $V_1, \ldots, V_{p/2}$ be the canonical ordering of blocks of π. Then, there exist $t_1 < t_3$ such that V_{t_1} crosses V_{t_3} and they would replace the same leg of the tensors by the coloring of replacements ω. Assume $\omega(t_1) = 1$. By the recursive procedure in Lemma 3.14, there must be a block V_{t_2} with $t_2 < t_3$ such that V_{t_2} crosses V_{t_3} and $\omega(t_2) = 0$, otherwise $\omega(t_3) = 0$ by Lemma 3.14. In particular, both legs of V_{t_3} get replaced. We have the following cases.

1. Either V_{t_2} crosses V_{t_1};
2. Or V_{t_2} does not crosses V_{t_1}.

Lemma 3.15 implies that the contribution of case (1) is zero. For case (2), if $V_{t_j} = \{r_j, s_j\}$, for $j = 1, 2, 3$, we have the following.

1. Either V_{t_2} and V_{t_1} are well-separated, that is,

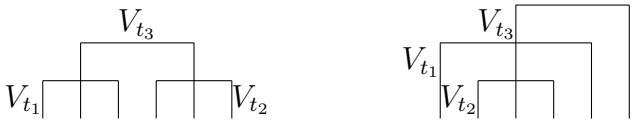
$$r_1 < r_3 < s_1 < r_2 < s_3 < s_2;$$

or

2. Or V_{t_2} are inside V_{t_1},

$$r_1 < r_2 < r_3 < s_2 < s_1 < s_3.$$

The block interactions are shown in Figure 3.6.



(a) Case (1) (b) Case (2)

Figure 3.6

For case (1), we have

$$\tau \otimes \tau(\pi \setminus \{V_1, \ldots, V_{t_3-1}\}, V_1, \ldots, V_{t_3-1}) = \frac{1}{\delta^2}\tau \otimes \tau\left(\prod_{1 \le l < r_3} \bar{b}_{i_l}\bar{a} \otimes \tilde{a} \prod_{r_3 < l < s_3} \bar{b}_{i_l}\tilde{a} \otimes \bar{a} \prod_{s_3 < l \le p} \bar{b}_{i_l}\right) - \frac{\lambda^4}{\delta^2}\tau \otimes \tau(\pi \setminus \{V_1, \ldots, V_l\}, V_1, \ldots, V_{l-1}),$$

where the definition of $\bar{b}$ follows the replacements done by $V_1, \ldots, V_{t_3-1}$. By Lemma 3.7, its contribution is zero, and we get the first part. Case (2) follows similarly.

For the second part, let π be a bipartite connected pair partition. Recall that this definition depends only on the structure of π rather than on the canonical ordering of blocks

$V_1, \dots, V_{p/2}$. In particular, let G_π be a representative graph isomorphic to the intersection graph of π under some arbitrary labeling. Let U_0, U_1 be its bipartite components of vertices and define the coloring $w : [p/2] \to \{0, 1\}$, where $w(l) = w$ if and only if $l \in U_w$. Since the graph is connected, there exists a sequence of blocks $V_{j_1}, \dots, V_{j_{p/2}}$ exhausting π such that V_{j_l} is connected to at least one V_{j_k}, for $k < l$. By Lemma 3.13, we have

$$\tau \otimes \tau(\pi) = q\tau \otimes \tau(\pi \setminus V_{j_1}, V_{j_1}).$$

Assume, by induction, that

$$\tau \otimes \tau(\pi) = 2\left(\frac{q}{2}\right)^l \tau \otimes \tau(\pi \setminus \{V_{j_1}, \dots, V_{j_l}\}, V_{j_1}, \dots, V_{j_l}), \tag{3.13}$$

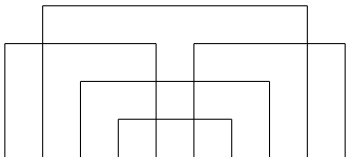
for some $1 \leq l < p/2$. Here, we define $\tau \otimes \tau(\pi \setminus \{V_{j_1}, \dots, V_{j_l}\}, V_{j_1}, \dots, V_{j_l})$ following the replacements done by the coloring w, namely, we replace the second leg of the tensors inside V_{j_l} if $w(j_l) = 1$ and the first if $w(j_l) = 0$, and $b_{i_l} \equiv 1$ if $l \in V_{j_1} \cup \cdots \cup V_{j_l}$. Let us prove that (3.13) holds for $l + 1 \leftarrow l$. Indeed, without loss of generality, let $w(j_{l+1}) = 1$ (the case $w(j_{l+1}) = 0$ is similar). Then, all crossing blocks V_{j_k} of $V_{j_{l+1}}$ have $w(j_k) = 0$. In particular, the second leg of the tensors associated with $V_{j_{l+1}}$ are still matched, namely, if $V_{j_{l+1}} = \{r, s\}$, the variables associated with $V_{j_{l+1}}$ in the interaction $\tau \otimes \tau(\pi \setminus \{V_{j_1}, \dots, V_{j_l}\}, V_{j_1}, \dots, V_{j_l})$ are given by

$$\begin{aligned} \overline{b}_{i_r} &= a' \otimes \tilde{a} - \lambda^2; \\ \overline{b}_{i_s} &= a'' \otimes \tilde{a} - \lambda^2, \end{aligned}$$

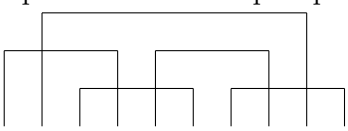
where $a', a'', \tilde{a}$ are free copies of a. We can then apply Lemma 3.14 verbatim and get

$$\tau \otimes \tau(\pi) = 2\left(\frac{q}{2}\right)^{l+1} \tau \otimes \tau(\pi \setminus \{V_{j_1}, \dots, V_{j_{l+1}}\}, V_{j_1}, \dots, V_{j_{l+1}}).$$

Induction is then proved. The result follows by the case $l = p/2$ and Lemma 3.14 applied for the connected pair partition in four elements $\pi = \{\{1, 3\}, \{2, 4\}\}$. □



(a) A bipartite connected pair partition

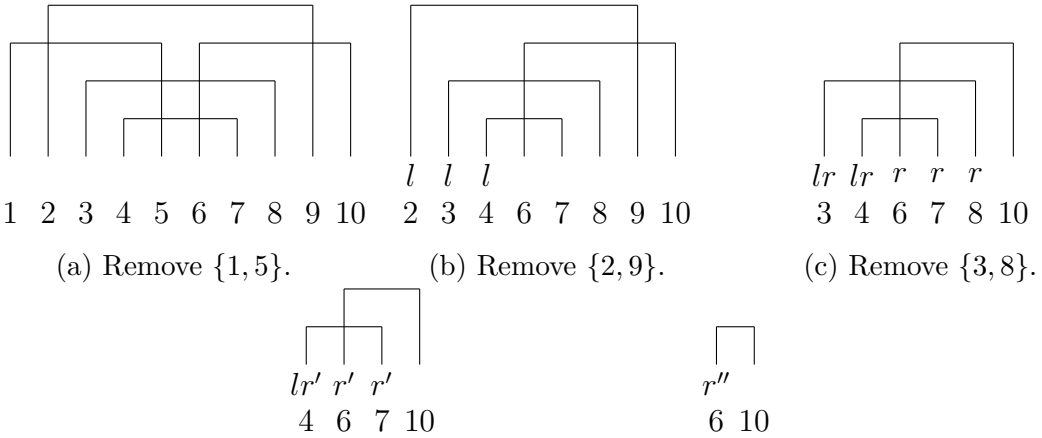


(b) A non-bipartite connected pair partition

Figure 3.7: Examples of connected pair partitions.

Figure 3.7 shows two examples of connected pair partitions, where in Figure 3.7a we

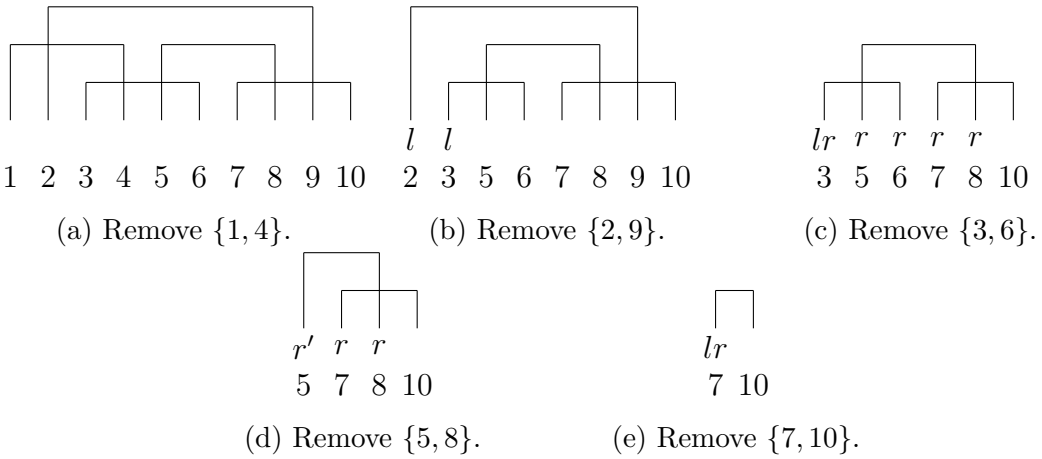
have a bipartite one, whose graph is the complete bipartite graph over 2 left vertices and 3 right vertices, and in Figure 3.7b we have a non-bipartite partition, whose graph is the cyclic graph C_5 (the pentagon). Figure 3.8 shows how we break down a bipartite partition, whereas Figure 3.9 shows how the breaking works for a non-bipartite partition.



(a) Remove $\{1, 5\}$. (b) Remove $\{2, 9\}$. (c) Remove $\{3, 8\}$.

(d) Remove $\{4, 7\}$. (e) Remove $\{6, 10\}$.

Figure 3.8: Induction over a bipartite partition. The letters $l, l', l'', ...$ indicate free independent replacements over the left leg, whereas $r, r', r'', ...$ indicate free independent replacement over the right leg. The first step has contribution q, whereas the rest have contribution $q/2$.



(a) Remove $\{1, 4\}$. (b) Remove $\{2, 9\}$. (c) Remove $\{3, 6\}$.

(d) Remove $\{5, 8\}$. (e) Remove $\{7, 10\}$.

Figure 3.9: Induction over a non-bipartite partition. The final result is zero, as both legs of b_{i_7} get replaced, by Lemma 3.7.

3.5.4 Proof of Theorem 3.1

We are now ready to prove the main theorem. Recall that

$$\tau'(\mathbf{S}^p) = \sum_{\pi \in P_2(p)} \tau \otimes \tau(\pi) = \sum_{\hat{\pi}, (\pi_T)_{T \in \hat{\pi}}} \prod_{T \in \hat{\pi}} \tau \otimes \tau(\pi_T),$$

where the second summation runs over $(\hat{\pi}, (\pi_T)_{T \in \hat{\pi}}) \in \Phi(P_2(p))$, using the bijection Φ defined in (3.5). Note that if $|T| = 2$, then by Corollary 3.9 we have $\tau \otimes \tau(\pi_T) = 1$.

Therefore, we deduce that

$$\tau'(\mathbf{S}^p) = \sum_{\hat{\pi},(\pi_T)_{T\in\hat{\pi}}} \prod_{\substack{T\in\hat{\pi}\\|T|\geq 4}} \tau\otimes\tau(\pi_T).$$

Using Proposition 3.11, we get

$$\tau'(\mathbf{S}^p) = \sum_{\hat{\pi}\in NC(p)} \sum_{(\pi_T)_{T\in\hat{\pi}}} \prod_{\substack{T\in\hat{\pi}\\|T|\geq 4}} 2\left(\frac{q}{2}\right)^{\frac{|T|}{2}},$$

where the second summation runs over bipartite connected pair partitions π_T, for $T \in \hat{\pi}$. Finally, note that the number of size-two blocks is precisely $\mathbf{ncr}(\pi)$ and thus

$$|\{T\in\hat{\pi} : |T|\geq 4\}| = \mathbf{cc}(\pi) - \mathbf{ncr}(\pi).$$

Since

$$\sum_{\substack{T\in\hat{\pi}\\|T|\geq 4}} \frac{|T|}{2} = \mathbf{cr}(\pi),$$

using again the bijection Φ, we deduce that

$$\tau'(\mathbf{S}^p) = \sum_{\pi\in P_2^{\mathrm{bi}}(p)} 2^{\mathbf{cc}(\pi)-\mathbf{ncr}(\pi)}\left(\frac{q}{2}\right)^{\mathbf{cr}(\pi)}.$$

This finishes the proof in view of Proposition 3.3 and of the fact that

$$\mathbf{ncr}(\pi) + \mathbf{cr}(\pi) = |\pi| = p/2.$$

Part II

Norm of inhomogeneous matrices

Chapter 4

On spectral outliers of inhomogeneous symmetric random matrices [5]

What do mathematicians sleep on?
Matrices.

Someone that needs resting

Sharp conditions for the presence of spectral outliers are well understood for Wigner random matrices with iid entries. In the setting of *inhomogeneous* symmetric random matrices (i.e., matrices with a non-trivial variance profile), the corresponding problem has been considered only recently. Of special interest is the setting of sparse inhomogeneous matrices since sparsity is both a key feature and a technical obstacle in various aspects of random matrix theory. For such matrices, the largest of the variances of the entries has been used in the literature as a natural proxy for sparsity. We contribute sharp conditions in terms of this parameter for an inhomogeneous symmetric matrix with sub-Gaussian entries to have outliers. Our result implies a "structural" universality principle: the presence of outliers is only determined by the level of sparsity rather than the detailed structure of the variance profile.

Contents

4.1 Introduction

Given an $n \times n$ random symmetric matrix M_n, define its *Empirical Spectral Distribution* (ESD) as the random probability measure

$$\mu_{M_n} := \frac{1}{n} \sum_{k \in [n]} \delta_{\lambda_k(M_n)},$$

where $\lambda_k(M_n)$ is the k-th largest eigenvalue of M_n. A central problem in random matrix theory is to establish necessary and sufficient conditions for the sequence μ_{M_n}, $n \geq 1$, to converge to a non-random measure (see, for example, [102, Section 2.4] for a discussion of different types of convergence: almost surely, in probability, in expectation).

A celebrated result of Wigner [112] asserts that the ESD of a normalized symmetric matrix with i.i.d entries converges weakly almost surely and in expectation to the *semi-circle distribution* μ_{sc} whose density f_{sc} is given by

$$f_{sc}(x) = \frac{1}{2\pi} \sqrt{4 - x^2}\, \mathbf{1}_{\{|x| \leq 2\}}. \tag{4.1}$$

The main feature of this result is its universality: the limiting ESD does not depend on the distribution of the matrix entries.

The universality phenomenon for the matrix spectrum has been actively studied for other models of randomness involving non-identically distributed entries. Regarding convergence to the semi-circle law, we refer, in particular, to [25, 81, 84] for random band matrices, [48, 47] for generalized Wigner matrices, and to the survey [Chapter 1][22] for further references. *Sparse* inhomogeneous random matrices have been a subject of much recent interest. The presence of sparsity often plays the role of both a key feature and a key challenge in statistical inference, graph theory, and random matrices [17, 19, 31].

We now introduce the matrix model to be studied in this paper.

Model. Let W_n denote an $n \times n$ Wigner matrix whose entries on and above the diagonal are iid copies of a centered random variable ξ having unit variance. Further, let $\Sigma_n = (\sigma_{ij})_{1 \leq i,j \leq n}$ be a symmetric $n \times n$ matrix with non-negative entries satisfying

$$\sum_{j=1}^{n} \sigma_{ij}^2 = 1, \text{ for all } 1 \leq i \leq n$$

(that is, the matrix $(\sigma_{ij}^2)_{1 \leq i,j \leq n}$ is *doubly stochastic*). The object of our study is the deformed matrix $X_n = \Sigma_n \circ W_n$, where "$\circ$" denotes the Hadamard (entry-wise) product. Note that the classical setting in Wigner's semi-circle theorem corresponds to $\Sigma_n = \frac{1}{\sqrt{n}} \mathbf{1}\mathbf{1}^t$, where $\mathbf{1}$ denotes the vector of all ones.

A natural question regarding the above model is: *does the limiting ESD of X_n depend on either Σ_n or the distribution of ξ?*

Denote by $\sigma_n^* = \max_{1\le i,j\le n} \sigma_{ij}$ the maximum of the standard deviations of the entries of X_n. The parameter σ_n^* can be viewed as a proxy of the matrix sparsity. The work [55] provides an essentially complete answer to the above question (we also refer to [17] for closely related statements):

Theorem 4.1 ([55]). *Let W_n be an $n \times n$ symmetric matrix whose entries on and above the diagonal are iid copies of a centered random variable ξ with unit variance. Let $\Sigma_n = (\sigma_{ij})_{1\le i,j\le n}$ be a symmetric matrix such that $(\sigma_{ij}^2)_{1\le i,j\le n}$ is doubly stochastic. Setting $X_n = \Sigma_n \circ W_n$ and assuming that $\sigma_n^* \to 0$, we have that μ_{X_n} converges weakly almost surely and in expectation to μ_{sc}.*

The main result of [55] can additionally handle W_n having entries that satisfy a Lindeberg-type condition rather than being identically distributed. The authors of [55] use the Stieltjes transform method to establish convergence in expectation. Additionally, it was noticed in [36] that this can be "upgraded" to almost sure convergence using a concentration inequality for the spectral measure of random matrices with independent entries [27, Lemma C.2], [57]. In this note, we provide an alternative proof of Theorem 4.1 based on the moment method (see appendix to this paper).

Take Σ_n to be the adjacency matrix of a d-regular graph (rescaled by $1/\sqrt{d}$ to have a doubly stochastic variance profile). Then X_n can be viewed as a weighted adjacency matrix with iid centered edge weights of variance $1/d$. Specialized to this setting, the above theorem asserts that the condition $d \to \infty$ is sufficient for almost sure convergence of the limiting ESD to the semi-circle law, regardless of the structure of the underlying graph. On the other hand, if $d \ge 2$ is fixed, then the distribution of μ_{X_n} is dependent both on the structure of Σ_n as well as the atomic distribution of W_n. For completeness, we provide a proof of this claim in the appendix (see Proposition 4.9 there).

4.1.1 Main results: spectral outliers of inhomogeneous matrices

Convergence of the ESD of X_n to the semi-circle law, which is supported on $[-2, 2]$, guarantees that $n - o(n)$ eigenvalues of X_n lie in the interval $[-2, 2]$. That does not rule out existence of *spectral outliers*, that is, eigenvalues near the spectral edges, which are at a non-vanishing distance to the support of the limiting ESD. In the classical setting of Wigner matrices with iid entries, it is known that the assumption of bounded fourth moment of the entries is necessary and sufficient to guarantee the absence of outliers [12].

Spectral outliers of Hermitian random matrices have been extensively studied [70, 15, 16, 20, 105, 4, 31], in part due to important applications in statistical inference and signal processing (we also refer, among others, to papers [106, 96, 65, 95, 94, 46, 76] dealing with the limiting distribution of the extreme eigenvalues of Wigner and sample covariance matrices). In particular, the seminal result of Baik–Ben Arous–Peche [15], and its generalizations unraveled a phase transition phenomenon in the appearance of outliers in the

spectrum of classical random matrix ensembles under small rank perturbations. In [105], a similar phase transition phenomenon was observed in the case of Wigner matrices under random sparsification, i.e., in the setting where each entry in the Wigner matrix is kept independently with some probability p_n (see also [4]). In the setting of the present paper, that corresponds to studying the norm of $B_n \circ W_n$, where W_n is a Wigner matrix and B_n is a symmetric matrix whose entries on and above the main diagonal are iid copies of a Bernoulli random variable with parameter p_n.

In this note, we give necessary and sufficient conditions for the presence of outliers for the matrix $X_n = \Sigma_n \circ W_n$ in terms of Σ_n and ξ. The main problem we want to address is whether structural universality phenomenon is observed for the spectral outliers, i.e., the presence/absence of outliers is characterized only by the level of the sparsity of Σ_n and does not depend on the fine structure of the variance profile.

Unlike for the bulk, it is readily checked that the presence or absence of outliers depends on the distribution of ξ regardless of the sparsity pattern. Indeed, this follows directly from the results of [74], where the norms of inhomogeneous matrices with Gaussian entries and with heavy-tailed entries were completely characterized. While distributional universality provably fails to hold, it remains a compelling question whether structural universality holds under some restrictions on ξ. We isolate the effect of the structure of Σ_n by restricting ξ to the class of sub-Gaussian distributions. Recall that a random variable ξ is said to be sub-Gaussian if there exists a constant C such that $\mathbb{E}e^{\lambda(\xi-\mathbb{E}\,\xi)} \leq e^{C\lambda^2}$ for every λ. Our question is as follows.

Question 1. *Let W_n be an $n \times n$ Wigner matrix whose entries are sub-Gaussian and consider $X_n = \Sigma_n \circ W_n$. Does the presence/absence of outliers for X_n depend on either the specific structure of Σ_n or the particular (sub-Gaussian) distribution of ξ? What conditions on Σ_n guarantee the presence/absence of outliers?*

We provide a sharp characterization for the appearance of outliers in terms of the sparsity proxy σ_n^*. Recall by Theorem 4.1 that whenever $\sigma_n^* \to 0$, the limiting ESD of X_n is the semi-circle distribution. Thus, if $\sigma_n^* \to 0$, the above question asks whether $\|X_n\| \to 2$, where $\|\cdot\|$ denotes the spectral norm.

Theorem 4.2 (Main result)**.** *Let W_n be an $n \times n$ symmetric matrix whose entries on and above the main diagonal are iid copies of a centered sub-Gaussian random variable ξ with unit variance. For any symmetric matrix $\Sigma_n = (\sigma_{ij})_{1\leq i,j\leq n}$ such that $(\sigma_{ij}^2)_{1\leq i,j\leq n}$ is doubly stochastic, if $\sigma_n^*\sqrt{\log n} \to 0$, then $\|\Sigma_n \circ W_n\| \to 2$ almost surely, implying the absence of outliers.*

Specialized to the setting of a weighted d-regular graph, the above theorem asserts that in the case of sub-Gaussian weights, if d grows faster than $\log n$, then there are no outliers regardless of the structure of the base graph. This result is sharp, as the next theorem shows.

Theorem 4.3 (A sufficient condition for existence of outliers)**.** *Let ξ be a centered random variable of unit variance and with bounded fourth moment, and for each n, let W_n be an $n \times n$ symmetric matrix whose entries above the main diagonal are iid copies of ξ. Fix any sequence $d := d(n) = O(\log n)$ with $d \to \infty$ such that nd is even for every n. Then there exists a sequence of non-random d-regular graphs with adjacency matrices A_n such that for all large n, the matrices $\frac{1}{\sqrt{d}} A_n \circ W_n$ have outliers almost surely. That is,*

$$\mathbb{P}\left(\liminf_{n\to\infty} \left\| \frac{1}{\sqrt{d}} A_{G_n} \circ W_n \right\| > 2 \right) = 1.$$

The sharp characterization provided by Theorem 4.2 combined with Theorem 4.3 was previously established in the special case of d-regular graphs with Gaussian edge weights [19]. Moreover, a succession of works [19, 21, 68, 74, 94] showed that for bounded weights, there are no outliers whenever $d/\log n \to \infty$. In the other direction, the existence of outliers for Rademacher weights was previously known when $d = O(\sqrt{\log n})$ [93], rather than up to the sharp rate of $d = O(\log n)$ that our results capture. It should be noted that the argument of [93] can be extended and adapted to the setting of Theorem 4.3 showing the existence of outliers in the regime $d/\log n \to 0$. However, that argument fails in the regime where d is of order $\log n$, while the proof we provide for Theorem 4.3 gives a unified treatment that works in all regimes. We refer to Remark 4.7 for more details.

Returning to the general setting of Theorem 4.2, to the best of our knowledge, the *absence* of outliers for the general sub-Gaussian distributions has only been treated in [21] for matrices with restricted bandwidth and graphs with suboptimal sparsity $d/\log^{9/2} n \to \infty$. However, there has been much recent progress in characterizing the spectral norm of inhomogeneous random matrices [17, 19, 21, 31, 74]. To the best of our knowledge, there are two existing upper bounds for the norm of inhomogeneous sub-Gaussian random matrices. The first is derived through a symmetrization procedure and comparison with the Gaussian setting [19], which leads to a suboptimal constant multiplicative factor and thus cannot capture sharp conditions for the presence/absence of outliers. The second result captures the correct constant 2 but at the expense of increasing the additive error term, which also results in a suboptimal regime for the absence of outliers (see [19, Corollary 3.6], [74, Remark 4.13]). Theorem 4.2 addresses these shortcomings and captures the correct asymptotic behavior for general sub-Gaussian matrices.

The proof of Theorem 4.2 avoids the use of symmetrization and instead directly estimates the trace of powers of $\Sigma_n \circ W_n$. It uses a truncation argument to split this matrix into the sum of a matrix with bounded weights and a remainder. The former is controlled using the results in [19, 21, 74].

The remainder is controlled by a new version of a compression argument, originally devised in [19, Proposition 2.1] and also used in [74], for comparing sparse matrices to dense matrices of smaller dimension. Implementing this argument in our setting requires tuning the dimension of the dense matrix as a function of the truncation level.

Matching upper and lower bounds on the norms of inhomogeneous Gaussian matrices were previously given in [74]. However, these results are only sharp up to a multiplicative constant, thus failing to capture the scale necessary for characterizing outliers. Upper bounds with the correct constant scale for the norm of inhomogeneous Gaussian and bounded random matrices were derived in [19, 74]. In the special case of Gaussian or bounded matrices, these results imply half of the dichotomy we prove, namely the absence of outliers. However, matching lower bounds would be needed to establish the presence of outliers. Theorem 4.3 addresses this shortcoming by showing that for *any* sparsity $d = \mathcal{O}(\log n)$, there is a sequence of graph adjacency matrices Σ_n such that $X_n := \Sigma_n \circ W_n$ has outliers for W_n having *any* non-atomic distribution with finite fourth moment. Thus, there is a sharp transition at sparsity $\mathcal{O}(\log n)$ below which structural universality is not observed. Below this level of sparsity, the presence/absence of outliers necessarily depends on the structure of Σ_n. The sequence of deterministic graphs in Theorem 4.3 is a union of cliques, and the proof of the result combines an anti-concentration argument with incompressibility properties of eigenvectors of Wigner matrices [90].

The paper is organized as follows: in Sections 4.2 and Section 4.3, we prove Theorems 4.2 and 4.3 respectively. Appendix 4.A contains a proof of Theorem 4.1 based on the moment method, together with a proof of Propositions 4.9. Although Theorem 4.1 is known, we prefer to include our argument here since it provides a uniform combinatorial treatment of both the bulk and the edges of the spectrum in an inhomogeneous setting.

Acknowledgments

The authors are grateful to Alexey Naumov for bringing to their attention the references [55] and [36]. We also thank Ramon van Handel for pointing out Remarks 4.4 and 4.7.

4.2 The regime with no outliers

The goal of this section is to prove Theorem 4.2. Let ξ be a centered sub-Gaussian random variable with unit variance and $W_n = (w_{ij})_{1\le i,j\le n}$ be an $n \times n$ Wigner matrix whose entries on and above the diagonal are iid copies of ξ. Fix an $n \times n$ symmetric matrix $\Sigma_n = (\sigma_{ij})_{1\le i,j\le n}$ satisfying

- $(\sigma_{ij}^2)_{1\le i,j\le n}$ is doubly stochastic;
- $\sigma_n^* \sqrt{\log n} \to 0$, where $\sigma_n^* = \max_{1\le i,j\le n} \sigma_{ij}$.

Given $L = L_n$ to be specified later, define $W_n^{\le L}$ (resp. $W_n^{>L}$) to be the matrix with entries $\left(w_{ij}\mathbf{1}_{|w_{ij}|\le L}\right)_{1\le i,j\le n}$ (resp. $\left(w_{ij}\mathbf{1}_{|w_{ij}|>L}\right)_{1\le i,j\le n}$). Correspondingly, define $X_n^{\le L} := \Sigma_n \circ W_n^{\le L}$ (resp. $X_n^{>L} := \Sigma_n \circ W_n^{>L}$) so that

$$X_n := \Sigma_n \circ W_n = X_n^{\le L} + X_n^{>L}.$$

Given an integer p, the triangle inequality combined with the fact that the entries of X_n are centered yields

$$\begin{aligned}\left(\mathbb{E}\,\mathrm{tr}(X_n^{2p})\right)^{1/(2p)} &\leq \left(\mathbb{E}\,\mathrm{tr}(X_n^{\leq L} - \mathbb{E}\,X_n^{\leq L})^{2p}\right)^{1/(2p)} + \left(\mathbb{E}\,\mathrm{tr}(X_n^{>L} - \mathbb{E}\,X_n^{>L})^{2p}\right)^{1/(2p)} \\ &=: \Gamma_1 + \Gamma_2.\end{aligned}$$

In order to bound Γ_1, we make use of [74, Theorem 4.8] as the entries of $X_n^{\leq L}$ are bounded, to obtain that

$$\Gamma_1 \leq 2\left(\sum_{j\in[n]}\left(\sum_{i\in[n]}\sigma_{ij}^2\,\mathbb{E}\left(w_{ij}\mathbf{1}_{|w_{ij}|\leq L} - \mathbb{E}\,w_{ij}\mathbf{1}_{|w_{ij}|\leq L}\right)^2\right)^p\right)^{1/(2p)} + Cn^{\frac{1}{p}}L\sigma_n^*\sqrt{p},$$

for some universal constant C. Using that $(\sigma_{ij}^2)_{1\leq i,j\leq n}$ is doubly stochastic, ξ has unit variance, and that the w_{ij}'s are iid, we deduce that

$$\Gamma_1 \leq 2n^{\frac{1}{2p}}\sqrt{\mathrm{var}(\xi\mathbf{1}_{|\xi|\leq L})} + Cn^{\frac{1}{p}}L\sigma_n^*\sqrt{p} \leq n^{\frac{1}{p}}\left(2 + CL\sigma_n^*\sqrt{p}\right). \tag{4.2}$$

To bound Γ_2, let $\tilde{X}_n^{>L}$ be an independent copy of $X_n^{>L}$. Since $X \mapsto \mathrm{tr}(X)^{2p}$ is convex, Jensen's Inequality implies that

$$\Gamma_2 = \left(\mathbb{E}_{X_n}\,\mathrm{tr}\left(\mathbb{E}_{\tilde{X}_n}\left(X_n^{>L} - \tilde{X}_n^{>L}\right)\right)^{2p}\right)^{1/(2p)} \leq \left(\mathbb{E}_{X_n,\tilde{X}_n}\,\mathrm{tr}(X_n^{>L} - \tilde{X}_n^{>L})^{2p}\right)^{1/(2p)}.$$

By symmetry, the distribution of $X_n^{>L} - \tilde{X}_n^{>L}$ coincides with that of $R\circ(X_n^{>L} - \tilde{X}_n^{>L})$, where $R = (r_{ij})_{1\leq i,j\leq n}$ is an $n\times n$ symmetric matrix with independent Rademacher variables independent of $X_n^{>L}$ and $\tilde{X}_n^{>L}$. By the triangle inequality, it follows that

$$\Gamma_2 \leq 2(\mathbb{E}\,\mathrm{tr}(R\circ X_n^{>L})^{2p})^{1/(2p)}.$$

Expanding the trace, we get

$$\mathbb{E}\,\mathrm{tr}(R\circ X_n^{>L})^{2p} = \sum_{u\in[n]^{2p}}\sigma_{u_1u_2}\cdots\sigma_{u_{2p}u_1}\,\mathbb{E}\,r_{u_1u_2}(X_n^{>L})_{u_1u_2}\cdots r_{u_{2p}u_1}(X_n^{>L})_{u_{2p}u_1}.$$

Let $u_1\to\cdots\to u_{2p}\to u_1$ be a closed path. We define its shape $s(u)$ as the relabelling of its vertices with labels $1,2,\ldots$ in order of appearance. Since the Rademacher variable is symmetric, for any path with non-zero contribution, all edges (u_i,u_{i+1}) must appear an even number of times on the path, and in this case, the corresponding contribution of the Rademacher variables is equal to 1. Let $\mathcal{S}_{even}$ be the set of all even shapes. Note that the quantity

$$T(s) := \mathbb{E}(X_n^{>L})_{u_1u_2}\cdots(X_n^{>L})_{u_{2p}u_1}$$

depends only on the shape $s(u) = s$ of the path u. In particular, we have

$$\mathbb{E}\,\mathrm{tr}(R \circ X_n^{>L})^{2p} = \sum_{s \in \mathcal{S}_{even}} T(s) \sum_{\substack{u \in [n]^{2p} \\ s(u)=s}} \sigma_{u_1 u_2} \cdots \sigma_{u_{2p} u_1}.$$

For a shape $s \in \mathcal{S}_{even}$, let $H_s = (V(s), E(s))$ be the graph generated by the vertices visited by the shape and edges given by $(u_1, u_2), \ldots, (u_{2p}, u_1)$. Denote $m(s) = |V(s)|$. Since s is a path, H_s is connected and we have $|E(s)| \geq m(s) - 1$. Applying [19, Lemma 2.5] (see also [74, Theorem 2.8]) and using that $(\sigma_{ij}^2)_{1 \leq i,j \leq n}$ is doubly stochastic, we have

$$\sum_{\substack{u \in [n]^{2p} \\ s(u)=s}} \sigma_{u_1 u_2} \cdots \sigma_{u_{2p} u_1} \leq n(\sigma_n^*)^{2p-2(m(s)-1)}.$$

Therefore, we deduce that

$$\mathbb{E}\,\mathrm{tr}(R \circ X_n^{>L})^{2p} \leq n(\sigma_n^*)^{2p} \sum_{s \in \mathcal{S}_{even}} T(s)(\sigma_n^*)^{-2(m(s)-1)}. \tag{4.3}$$

Now, let k_e be the number of times the edge $e \in E(s)$ is traversed by the shape s. Then

$$T(s) = \prod_{e \in E(s)} \mathbb{E}\,\xi^{k_e} \mathbf{1}_{|\xi|>L}.$$

Using Cauchy–Schwartz inequality and that ξ is sub-Gaussian (see [107, Section 2.5] for equivalent definitions of sub-Gaussian random variables), we can write

$$\mathbb{E}\,\xi^{k_e} \mathbf{1}_{|\xi|>L} \leq (\mathbb{E}\,\xi^{2k_e})^{1/2} \mathbb{P}(|\xi| > L)^{1/2} \leq \tilde{C}^{k_e} e^{-cL^2}\,\mathbb{E}\,g^{k_e},$$

where $g \sim N(0,1)$ and $\tilde{C}$, c are universal constants.

Since $\sum_{e \in E(s)} k_e = 2p$ and $|E(s)| \geq m(s) - 1$, putting together the above we deduce that

$$T(s) \leq \tilde{C}^{2p} e^{-c|E(s)|L^2} T_g(s) \leq \tilde{C}^{2p} e^{-c(m(s)-1)L^2} T_g(s),$$

where $T_g(s) = \prod_{e \in E(s)} \mathbb{E}\,g^{k_e}$. Replacing the above relations in (4.3), we deduce that

$$\mathbb{E}\,\mathrm{tr}(R \circ X_n^{>L})^{2p} \leq \tilde{C}^{2p} (\sigma_n^*)^{2p} n \sum_{s \in \mathcal{S}_{even}} T_g(s) b^{m(s)-1},$$

where we denoted $b := (\sigma_n^*)^{-2} e^{-cL^2}$. Setting $q = \lceil b \rceil + p$, using that $m(s) \leq p+1$ and that

$$\frac{q!}{(q - m(s))!} \geq q(q - m(s) + 1)^{m(s)-1} \geq q b^{m(s)-1},$$

we get

$$\mathbb{E}\,\mathrm{tr}(R \circ X_n^{>L})^{2p} \le \tilde{C}^{2p}(\sigma_n^*)^{2p}\frac{n}{q}\sum_{s\in\mathcal{S}_{even}} T_g(s)\frac{q!}{(q-m(s))!}.$$

Now note that for a standard $q \times q$ symmetric Gaussian matrix G_q, we readily have

$$\mathbb{E}\,\mathrm{tr}(G_q)^{2p} = \sum_{s\in\mathcal{S}_{even}} T_g(s)\frac{q!}{(q-m(s))!}.$$

Thus, we deduce that

$$\begin{aligned}\mathbb{E}\,\mathrm{tr}(R \circ X_n^{>L})^{2p} &\le \tilde{C}^{2p}(\sigma_n^*)^{2p}\frac{n}{q}\,\mathbb{E}\,\mathrm{tr}(G_q)^{2p}\\ &\le \tilde{C}^{2p}(\sigma_n^*)^{2p} n\,\mathbb{E}\|G_q\|^{2p}\\ &\le \tilde{C}^{2p}(\sigma_n^*)^{2p} n(2\sqrt{q}+8\sqrt{p})^{2p},\end{aligned}$$

where we have used [19, Lemma 2.2] to bound the moments of the norm of a standard Gaussian matrix. Replacing b in the above, we get that

$$\Gamma_2 \le C' n^{1/(2p)}\Big(e^{-cL^2/2} + \sigma_n^*\sqrt{p}\Big), \tag{4.4}$$

for some universal constant C'. Combining (4.2) and (4.4), and choosing $L^2 = -2c^{-1}\log(\sigma_n^*\sqrt{p})$, we get that

$$\Big(\mathbb{E}\,\mathrm{tr}(X_n^{2p})\Big)^{1/(2p)} \le n^{1/p}\left(2 + C''\sigma_n^*\sqrt{p\log\left(\frac{1}{\sigma_n^*\sqrt{p}}\right)}\right), \tag{4.5}$$

for some universal constant C''. Now choosing

$$p = \frac{\sqrt{\log n}}{\sigma_n^*}, \tag{4.6}$$

we deduce that

$$\mathbb{E}\,\|X_n\| \le \Big(\mathbb{E}\,\|X_n\|^{2p}\Big)^{\frac{1}{2p}} \le e^{\sigma_n^*\sqrt{\log n}}\left[2 + C''\sqrt{\sigma_n^*\sqrt{\log n}}\sqrt{\log\left(\frac{1}{\sqrt{\sigma_n^*\sqrt{\log n}}}\right)}\right].$$

Since $\sigma_n^*\sqrt{\log n} \to 0$, we have that

$$\limsup_{n\to\infty}\mathbb{E}\,\|X_n\| \le 2.$$

On the other hand, it follows from the almost sure convergence of the empirical spectral distribution (see Theorem 4.1) that almost surely,

$$\liminf_{n\to\infty}\|X_n\| \ge 2,$$

which, by Fatou's lemma, implies that

$$\liminf_{n\to\infty} \mathbb{E}\,\|X_n\| \geq 2.$$

This shows that $\mathbb{E}\,\|X_n\|$ converges to 2 whenever $\sigma_n^*\sqrt{\log n} \to 0$.

To prove the almost sure convergence, note that for any fixed $\eta > 2$, we have by Markov's inequality that

$$\mathbb{P}(\|X_n\| > \eta) \leq \frac{\mathbb{E}\,\mathrm{tr}(X_n^{2p})}{\eta^{2p}} \leq n^2 \left(\frac{2 + C''\sqrt{\sigma_n^*\sqrt{\log n}}\sqrt{\log\left(\frac{1}{\sqrt{\sigma_n^*\sqrt{\log n}}}\right)}}{\eta} \right)^{2p}.$$

Using the choice of p in (4.6) and that $\sigma_n^*\sqrt{\log n} \to 0$, it is easy to see that the above quantity is summable in n. A classical application of the Borel-Cantelli lemma finishes the proof.

Remark 4.4 (Communicated by Ramon van Handel)**.** A compression argument along the lines of [74, Theorem 4.8] can be used as the basis for an alternative proof of Theorem 4.2. Such a compression argument would yield a comparison to a matrix M with independent entries of the form $b|g|$, where b is a normalized Bernoulli variable (with properly chosen probability of success) and g is an independent standard Gaussian. By Talagrand's inequality [101] and Gaussian concentration, the largest eigenvalue of M is subgaussian. Combining this with the Bai–Yin theorem [12] rules out the presence of outliers in the asymptotic regime.

Remark 4.5 (Heavy-tailed distributions)**.** The above argument can be adapted to heavy-tailed distributions. Let ξ be a centered Weibull distribution with shape parameter at most 2, so that for some $\beta \geq 1/2$ and any $p \geq 1$, we have

$$\|\xi\|_p \leq Cp^\beta.$$

Let Σ_n be a doubly stochastic matrix and $X_n = \Sigma_n \circ W_n$. Then, similarly to the proof of [74, Theorem 4.4] and the above computations, using $\sqrt{p} \leq p^\beta$, we get

$$\left(\mathbb{E}\,\mathrm{tr}(X_n^{2p})\right)^{1/(2p)} \leq n^{1/p}\left[2 + C_\beta\left(e^{-c_\beta L^{1/\beta}} + \sigma_n^* L p^\beta\right)\right].$$

Choosing

$$L^{1/\beta} := \frac{1}{c_\beta}\log\left(\frac{1}{\beta\sigma_n^* p^\beta}\right); \quad p := \frac{\sqrt{\log n}}{(\sigma_n^*)^{1/(2\beta)}},$$

we get

$$\left(\mathbb{E}\,\mathrm{tr}(X_n^{2p})\right)^{1/(2p)} \leq e^{(\sigma_n^*\log^\beta n)^{1/(2\beta)}}\left(2 + C_\beta\sqrt{\sigma_n^*\log^\beta n}\log^\beta\left(\frac{1}{\sqrt{\sigma_n^*\log^\beta n}}\right)\right).$$

Therefore, whenever $\sigma_n^* \log^\beta n \to 0$, X_n has no outliers almost surely and in expectation. This improves upon [75, Theorem 4.4] since the cited theorem does not provide the correct leading constant 2 for the spectral norm estimate.

4.3 Outliers in the subcritical regime

The goal of this section is to prove Theorem 4.3. Note that since the random variable ξ is centered and non-constant, it satisfies

$$\sup_{u\in\mathbb{R}} \mathbb{P}(|\xi - u| \geq \rho) \geq \rho, \quad \mathbb{P}(\xi \geq \rho) \geq \rho, \quad \text{and} \quad \mathbb{P}(\xi \leq -\rho) \geq \rho, \tag{4.7}$$

for some $\rho \in (0,1)$ depending only on the distribution of ξ.

Let n be a large integer, and let $d = d(n) = O(\log n)$ and $d = \omega(1)$. To simplify the exposition, we will assume that $n/(d+1)$ is an integer; we note that our construction below can be easily adapted to cover all admissible choices of $d = d(n)$. Consider a d-regular graph G_n on n vertices whose adjacency matrix A_n is block diagonal with $n/(d+1)$ blocks of all ones (excluding the entries on the main diagonal). The graph G_n is thus a disjoint union of $n/(d+1)$ cliques of size $d+1$ each. For every integer k, we denote by $\tilde{W}_k$ the $k \times k$ symmetric matrix whose entries above the main diagonal are iid copies of ξ, and the main diagonal is zero. Finally, we define $X_n := \frac{1}{\sqrt{d}} A_n \circ \tilde{W}_n$, so that X_n is block diagonal with $n/(d+1)$ iid blocks, where each block is equidistributed with $\frac{1}{\sqrt{d}}\tilde{W}_{d+1}$. We denote these blocks by $W_{d+1}^{(i)}$, $i = 1, \dots, n/(d+1)$. To prove Proposition 4.3, we will show that $\liminf_{n\to\infty} \|X_n\|$ is almost surely bounded away from 2. We will need the following lemma:

Lemma 4.6. *For every $\varepsilon > 0$ there is $\delta > 0$ depending on ε and the distribution of ξ with the following property. Let k be a sufficiently large integer, and let $\tilde{W}_{k+1}$ be the matrix defined above. Then*

$$\mathbb{P}\Big(\Big\|\tilde{W}_{k+1}\Big\| \geq (2+\delta)\sqrt{k}\Big) \geq \exp(-\varepsilon k).$$

We provide the proof of the lemma at the end of the subsection. At this point, let $C > 0$ be the constant such that $d = d(n) \leq C \log n$ for all sufficiently large n. We choose $\varepsilon := \frac{1}{2C}$. Following Lemma 4.6, there is $\delta \in (0,1)$ depending on C and the distribution of

ξ such that

$$\begin{aligned}\mathbb{P}(\|X_n\| \geq 2+\delta) &= \mathbb{P}\Big(\max_{i=1,\dots,n/(d+1)} \|W_{d+1}^{(i)}\| \geq (2+\delta)\sqrt{d}\Big)\\ &= 1-\mathbb{P}\Big(\|\tilde{W}_{d+1}\| < (2+\delta)\sqrt{d}\Big)^{\frac{n}{d+1}}\\ &\geq 1-\Big(1-e^{-\varepsilon d}\Big)^{\frac{n}{d+1}}\\ &\geq 1-\left(1-\frac{1}{\sqrt{n}}\right)^{\frac{n}{d+1}}\\ &\geq 1-e^{-n^{1/4}},\end{aligned}$$

for all large n. A direct application of the Borel–Cantelli lemma implies the result.

Remark 4.7. A weaker version of Lemma 4.6 follows easily by an adaptation of an argument in [93] (see the proof of Theorem 3.2 there). Indeed, one can write

$$\mathbb{P}\Big(\Big\|\tilde{W}_{k+1}\Big\| \geq 3\sqrt{k}\Big) \geq \mathbb{P}(\xi \geq \rho)^{\frac{9}{\rho^2}k}.$$

This follows since the norm of $\tilde{W}_{k+1}$ is bounded below by the norm of its $(\frac{3}{\rho}\sqrt{k}) \times (\frac{3}{\rho}\sqrt{k})$ submatrix. The latter is larger than $3\sqrt{k}$ if all entries are larger than ρ. This trivial observation gives a version of Lemma 4.6 for some ε (rather than for any ε). One can easily check that using this, we have $\frac{\|X_n\|}{2} \to \infty$ almost surely whenever $d/\log n \to 0$. In this regime, the above argument indicates that the presence of outliers is caused by the emergence of a submatrix of size $O(\sqrt{\log n})$ with large entries of equal signs.

The extra quantification (in terms of ε) present in Lemma 4.6 is needed to treat the case of d of order $\log n$. On the other hand, we note that stronger and sharper statements than Lemma 4.6 follow from the large deviation principle for the largest eigenvalue of Wigner matrices [11, 56, 39]; however, those results only cover particular cases of sub-Gaussian random variables.

Proof of Lemma 4.6. We start the proof with the following observation. For every choice of parameters $\varepsilon, \beta > 0$ there is $\gamma > 0$ depending on β, ε and the distribution of ξ with the following property. Assuming k is sufficiently large, letting v be a non-random unit vector in $\mathbb{R}^k$ with

$$\Big|\Big\{i \leq k : \ |v_i| \geq \beta/\sqrt{k}\Big\}\Big| \geq \beta\, k,$$

and taking $\xi_1, \dots, \xi_k$ to be iid copies of ξ, we have

$$\mathbb{P}\left(\Big|\sum_{i=1}^k v_i\xi_i\Big| \geq \gamma\,\sqrt{k}\right) \geq \exp\Big(-\varepsilon k/2\Big).$$

Indeed, assuming k is large enough, we take ℓ to be the largest integer bounded above by $\beta\, k$ and such that $\rho^\ell \geq \exp\Big(-\varepsilon k/4\Big)$, where ρ is taken from (4.7) (observe that ℓ is of order k). Further, let $I \subset [k]$ be a non-random subset of cardinality ℓ such that

$|v_i| \geq \beta/\sqrt{k}$ for every $i \in I$. In view of (4.7) and the choice of ℓ, with probability at least $\exp\Big(-\varepsilon k/4\Big)$ we have

$$\sum_{i\in I} v_i\xi_i \geq \frac{\beta}{\sqrt{k}} \cdot \rho\ell.$$

On the other hand, since the variance of $\sum_{i\in[k]\setminus I} v_i\xi_i$ is less than one, and in view of Markov inequality, with a probability of at least $3/4$, we have

$$\Big|\sum_{i\in[k]\setminus I} v_i\xi_i\Big| \leq 2.$$

Combining the two estimates, we get that

$$\sum_{i=1}^{k} v_i\xi_i \geq \frac{\beta}{\sqrt{k}} \cdot \rho\ell - 2$$

with probability at least $\frac{3}{4}\exp\Big(-\varepsilon k/4\Big) \geq \exp\Big(-\varepsilon k/2\Big)$, and the claim follows.

For convenience, we will use the compact notation $M := \tilde{W}_{k+1}$, and we let M' be the top left $k \times k$ principal submatrix of M. Let $\alpha > 0$ be a small parameter (depending on ε and the distribution of ξ), which will be determined later. The standard covering arguments imply that, as long as k is sufficiently large, with a probability of at least 0.999 *every* eigenvector of M' is *incompressible* (see [90]). On the other hand, as a consequence of the Wigner semi-circle law (Theorem 4.1), with a probability of at least 0.999 there is an eigenvector v of M' with $\|M'v\|_2 \geq (2-\alpha)\sqrt{k}$. To summarize, we can define an M'–measurable random unit vector v satisfying

$$\mathbb{P}\Big(\|M'v\|_2 \geq (2-\alpha)\sqrt{k},\ \ \Big|\Big\{i \leq k:\ |v_i| \geq \beta/\sqrt{k}\Big\}\Big| \geq \beta\, k\Big) \geq 0.99,$$

where β depends on the distribution of ξ but not on α. Let z be a vector in $\mathbb{R}^{k+1}$ obtained from v by adding zero coordinate. Our goal is to estimate the probability that $\|Mz\|_2$ is bounded away from two. Denote the $(k+1)$-st row of M by row_{k+1}. We have

$$\|M\|^2 \geq \|Mz\|_2^2 = \|M'v\|_2^2 + \langle \mathrm{row}_{k+1}, z\rangle^2. \tag{4.8}$$

In view of the definition of v, we have

$$\|M'v\|_2^2 \geq (2-\alpha)^2\, k \tag{4.9}$$

with probability at least 0.99. On the other hand, conditioned on any realization of M' such that (4.9) holds and

$$\Big|\Big\{i \leq k:\ |v_i| \geq \beta/\sqrt{k}\Big\}\Big| \geq \beta\, k,$$

we obtain from the observation at the beginning of the proof that

$$\mathbb{P}_{\mathrm{row}_{k+1}}\Big(|\langle \mathrm{row}_{k+1}, z\rangle| \geq \gamma\sqrt{k}\Big) \geq \exp\Big(-\varepsilon k/2\Big),$$

for some γ depending on β, ε, and the distribution of ξ (but not on α). It remains to choose α so that $(2-\alpha)^2 + \gamma^2 > 4$, and the proof is complete.

□

Appendix

4.A The limiting Empirical Spectral Distribution: A moment method approach

The goal of this appendix is to provide a proof of Theorem 4.1 based on the moment method. We will assume that the variance profile Σ_n satisfies $\sigma_n^* \to 0$. By a standard truncation argument (see [102, Section 2.4.1]), it is sufficient to deal with matrices with bounded entries and zero diagonal. More precisely, we let $(L_n)_{n\geq 1}$ be a sequence of numbers defined by

$$L_n := \min\left\{\frac{1}{\sqrt{\sigma_n^*}}, \log^c n\right\} \tag{4.10}$$

(so that $L_n \to \infty$), and from now on, we will assume that the random matrices $W_n = (w_{ij})_{1\leq i,j\leq n}$ have i.i.d centered entries with unit variances and are *uniformly bounded by* L_n. Here, $c \geq 1$ is a fixed large constant.

The convergence of the ESD of X_n is characterized by the convergence of its moments. More precisely, we aim to show that for every integer $k \geq 1$, the sequence $\left(\frac{1}{n}\operatorname{tr}(X_n^k)\right)_{n\geq 1}$ converges almost surely and in expectation to the corresponding kth moment of the semi-circle distribution.

Given an integer vector $u \in [n]^k$, we will treat it as a closed path $u = u_1 \to \cdots \to u_k \to u_1$ in the complete graph over $[n]$. The path u generates a (connected) subgraph H_u whose vertices are $\{u_1, \ldots, u_k\}$ and the edges are $\{(u_l, u_{l+1}) : l \in [k]\}$, where $u_{k+1} = u_1$. With some abuse of notation, we will fix a labeling of the vertices of H_u, namely, $V(H_u) = \{1, \ldots, |V(H_u)|\}$. Let $q_u(e)$ be the number of times the edge $e \in E(H_u)$ is traversed by the path u, so that $\sum_e q_u(e) = k$. Finally, for every $u \in [n]^k$, we denote

$$\sigma_u := \sigma_{u_1u_2} \ldots \sigma_{u_ku_1} \quad \text{and} \quad w_u = w_{u_1u_2} \ldots w_{u_ku_1}.$$

4.A.1 The convergence in expectation

To prove the convergence in expectation, we write

$$\frac{1}{n}\,\mathbb{E}\operatorname{tr}(X_n^k) = \frac{1}{n}\sum_{u\in[n]^k} \sigma_u\,\mathbb{E}\,w_u.$$

Since W_n is centered, the case $k = 1$ is trivial. Thus, we suppose in the sequel that $k \geq 2$. Note that for a given $u \in [n]^k$, each edge $e \in E(H_u)$ must be traversed at least twice; otherwise, their contribution is zero as the variables are independent and centered. This implies that $q_u(e) \geq 2$ for every edge $e \in E(H_u)$ and therefore that $|E(H_u)| \leq k/2$. Since H_u is a connected graph, we have

$$|V(H_u)| \leq |E(H_u)| + 1 \leq k/2 + 1.$$

We will write $q_u \geq 2$ (resp. $q_u = 2$) to encode the fact that $q_u(e) \geq 2$ (resp. $q_u = 2$) for every edge $e \in E(H_u)$. Note that when $q_u \geq 2$, we have

$$\mathbb{E}\, w_u \leq L_n^{k-2|E(H_u)|} \leq L_n^{k-2(|V(H_u)|-1)}, \tag{4.11}$$

where we used that the entries of W_n are of unit variance and uniformly bounded by L_n.

Applying [19, Lemma 2.5] (see also [74, Theorem 2.8]) and using that $(\sigma_{ij}^2)_{1\leq i,j\leq n}$ is doubly stochastic, we can write for every $m \leq k/2+1$ that

$$\frac{1}{n} \sum_{\substack{u\in[n]^k \\ q_u\geq 2, |V(H_u)|=m}} \sigma_u\, \mathbb{E}\, w_u \leq (L_n\sigma_n^*)^{k-2(m-1)} = (\sqrt{\sigma_n^*})^{k-2(m-1)}, \tag{4.12}$$

by the choice of L_n in (4.10). Note that whenever $m \leq k/2$, we get that $k-2(m-1) \geq 2$ and the right-hand side of (4.12) goes to 0. Therefore, we can restrict to the case where $|V(H_u)| = k/2+1$, that is, H_u is a tree with $|E(H_u)| = k/2$ edges. Moreover, $q_u(e) = 2$ for all edges (otherwise $|E(H_u)| < k/2$). Hence, we get that

$$\lim_{n\to\infty} \frac{1}{n}\, \mathbb{E}\, \mathrm{tr}(X_n^k) = \lim_{n\to\infty} \frac{1}{n} \sum_{\substack{u\in[n]^k \\ q_u=2; H_u \text{ is a tree}}} \sigma_u\, \mathbb{E}\, w_u.$$

In this case, $\mathbb{E}\, w_u = 1$, so that

$$\lim_{n\to\infty} \frac{1}{n}\, \mathbb{E}\, \mathrm{tr}(X_n^k) = \lim_{n\to\infty} \frac{1}{n} \sum_{\substack{u\in[n]^k \\ q_u=2; H_u \text{ is a tree}}} \sigma_u.$$

Consider now the set $\mathcal{G}_k$ of all pairs (G,t) where G is a tree over $[k/2+1]$ with a closed walk $t = t_1 \to \cdots \to t_k \to t_1$ that traverses each edge exactly twice. Then

$$\lim_{n\to\infty} \frac{1}{n}\, \mathbb{E}\, \mathrm{tr}(X_n^k) = \lim_{n\to\infty} \sum_{(G,t)\in\mathcal{G}_k} \frac{1}{n} \sum_{\substack{u\in[n]^k \\ (H_u,u)=(G,t)}} \sigma_u.$$

As G is a tree and each edge is traversed twice, we can sum up over leaves. Using double stochasticity and induction over leaves, we have

$$\frac{1}{n} \sum_{\substack{u\in[n]^k \\ (H_u,u)=(G,t)}} \sigma_u = \frac{1}{n} \sum_{i,j\in[n]} \sigma_{ij}^2 = 1.$$

This also follows by a similar argument done in [19, Lemma 2.5]. By [66, Exercise 4.4.1], it is known that $|\mathcal{G}_k| = |NC_2(k)|$ which is precisely the kth moment of the semi-circle distribution. Hence, the result follows.

4.A.2 Tail bound and almost-sure convergence

By the Borel-Cantelli lemma, the following proposition immediately implies almost sure convergence of the ESD of X_n.

Proposition 4.8. *For all $k \geq 1$, $p \geq 2$, and $\varepsilon \in (0, 1/2)$, we have*

$$\mathbb{P}\Big[\frac{1}{n}\Big|\mathrm{tr}(X_n^k) - \mathbb{E}[\mathrm{tr}(X_n^k)]\Big| > n^{-1/2+\varepsilon}\Big] \leq \frac{2}{n^p},$$

for all n sufficiently large depending on k, p and ε.

Proof. Let $\mathcal{F}_t$ be the σ-algebra generated by revealing the first t of the entries on or above the diagonal of X under some arbitrary ordering. Denote $\mathbb{E}_t[\cdot] := \mathbb{E}[\cdot|\mathcal{F}_t]$. In this notation,

$$\begin{aligned}\mathrm{tr}(X_n^k) - \mathbb{E}\Big[\mathrm{tr}(X_n^k)\Big] &= \sum_{t=1}^{n(n+1)/2} \mathbb{E}_t\Big[\mathrm{tr}(X_n^k)\Big] - \mathbb{E}_{t-1}\Big[\mathrm{tr}(X_n^k)\Big] \\ &=: \sum_{t=1}^{n(n+1)/2} \Delta_t\,.\end{aligned}$$

The key estimate is the following "bounded difference" estimate for the martingale increments Δ_t. For all $t \in [n(n+1)/2]$,

$$\mathbb{P}\Big[|\Delta_t|^2 > n^{\varepsilon}\sigma_t^2\Big] \leq \frac{1}{n^{p+2}}\,, \tag{4.13}$$

Once established, this readily implies Proposition 4.8. By union bound:

$$\mathbb{P}\Big[\exists\, t \in [n(n+1)/2] \,:\, |\Delta_t| > n^{\varepsilon/2}\sigma_t\Big] \leq \frac{1}{n^p}$$

Thus, by the Azuma–Hoeffding inequality for martingales with bounded increments, since $\big(\sigma_{ij}^2\big)_{ij}$ is doubly stochastic:

$$\begin{aligned}\mathbb{P}\Big[\Big|\mathrm{tr}(X^k) - \mathbb{E}\Big[\mathrm{tr}(X^k)\Big]\Big| > y\Big] &\leq \frac{1}{n^p} + \mathbb{P}\Big[\Big|\mathrm{tr}(X^k) - \mathbb{E}\Big[\mathrm{tr}(X^k)\Big]\Big| > y,\ \Delta_t^2 \leq n^{\varepsilon}\sigma_t^2\ \forall t\Big] \\ &\leq \frac{1}{n^p} + \exp\Big\{-\frac{y^2}{2n^{1+\varepsilon}}\Big\}\,.\end{aligned}$$

Taking $y = n^{\frac{1}{2}+\varepsilon}$ concludes the proof. We turn towards proving the proposition.

Our strategy is to compute the p'th moment of Δ_t and then apply Markov's inequality, where p is a large even integer independent of n. Fix t to be the edge (i, j) where (i, j) is the t'th entry of A. Let p be an even positive integer. We will say that $t \in u$ if $(u_l, u_{l+1}) = t$ or $(u_{l+1}, u_l) = t$ for some $l \in [k]$. Noting that $\mathbb{E}_t[w_u] = \mathbb{E}_{t-1}[w_u]$ if $t \notin u$, we

obtain

$$\mathbb{E}[\Delta_t^p] = \mathbb{E}\left[\left(\sum_{u\in[n]^k} \sigma_u(\mathbb{E}_t[w_u] - \mathbb{E}_{t-1}[w_u])\right)^p\right] = \mathbb{E}\left[\left(\sum_{\substack{u\in[n]^k\\ t\in u}} \sigma_u(\mathbb{E}_t[w_u] - \mathbb{E}_{t-1}[w_u])\right)^p\right].$$

That is,

$$\mathbb{E}[\Delta_t^p] = \sum_{\substack{\mathbf{u}=(u^{(1)},\dots,u^{(p)})\in[n]^{kp}\\ t\in u^{(l)},\forall l\in[p]}} \sigma_{u^{(1)}}\cdots\sigma_{u^{(p)}}\,\mathbb{E}\left(\prod_{l\in[p]}\{\mathbb{E}_t[w_{u^{(l)}}] - \mathbb{E}_{t-1}[w_{u^{(l)}}]\}\right).$$

Denote $H = H_{\mathbf{u}}$ the subgraph generated by $\mathbf{u}$. Let $q(e) = q_{\mathbf{u}}(e)$ be the number of times an edge $e \in E(H)$ is traversed in $\mathbf{u}$. As before, we only consider $\mathbf{u}$ such that $q(e) \geq 2$ for all $e \in E(H)$. Since all paths $u^{(l)}$ contain the edge $t \in E(H)$, we have that H is connected and $q(t) \geq p$. We first bound the contribution of W_n.

$$\left|\mathbb{E}\left(\prod_{l\in[p]}\{\mathbb{E}_t[w_{u^{(l)}}] - \mathbb{E}_{t-1}[w_{u^{(l)}}]\}\right)\right| \leq C_p \max_{r\in\{t,t-1\}^p} \mathbb{E}\left(\prod_{l\in[p]}\mathbb{E}_{r_l}[w_{u^{(l)}}]\right).$$

Since $|w_{ij}| \leq L_n$ and w_{ij} has unit variance, recalling (4.11),

$$\left|\mathbb{E}\left(\prod_{l\in[p]}\{\mathbb{E}_t[w_{u^{(l)}}] - \mathbb{E}_{t-1}[w_{u^{(l)}}]\}\right)\right| \leq C_p L_n^{pk-2(|V(H)|-1)}. \tag{4.14}$$

Thus, it suffices to control

$$\sum_{\substack{\mathbf{u}\in[n]^{pk}:\\ (H_{\mathbf{u}},q_{\mathbf{u}})=(H,q),\\ t\in u^{(l)},\forall l\in[p]}} \sigma_{u^{(1)}}\cdots\sigma_{u^{(q)}} \leq \sum_{\substack{u\in[n]^{V(H)}\\ (u_i,u_j)=(i,j)}} \prod_{e=(a,b)\in E(H)} \sigma_{u_a u_b}^{q(e)}.$$

In order to use the results from [74], we will write this quantity in a better way. First, as $\sigma_{uv} \leq \sigma_n^*$ for all $u, v \in [n]$, for any spanning tree $T = (V(T), E(T))$ of H with $t \in E(T)$, we have

$$\sum_{\substack{u\in[n]^{V(H)}\\ (u_i,u_j)=(i,j)}} \prod_{e=(a,b)\in E(H)} \sigma_{u_a u_b}^{q(e)} \leq \prod_{e\in E(H)\setminus E(T)} (\sigma_n^*)^{q(e)} \sum_{\substack{u\in[n]^{V(T)}\\ (u_i,u_j)=(i,j)}} \prod_{e=(a,b)\in E(T)} \sigma_{u_a u_b}^{q(e)}.$$

Now define the following collection of (symmetric) matrices $(b^{(e)})_{e\in E(T)}$.

1. If $e = t$, then $b_{ij}^{(t)} = b_{ji}^{(t)} := \sigma_{ij}^{q(t)}$ and 0 otherwise.
2. If e is incident to i, set $b_{ui}^{(e)} = b_{iu}^{(e)} = \sigma_{ui}^{q(e)}$ for all $u \in [n]$ and zero otherwise.
3. If e is incident to j, set $b_{uj}^{(e)} = b_{ju}^{(e)} = \sigma_{uj}^{q(e)}$ for all $u \in [n]$ and zero otherwise.
4. If e is not incident to neither j nor i, set $b_{uv}^{(e)} = \sigma_{uv}^{q(e)}$ for all $u, v \in [n]$.

Then it is immediate to see that

$$\sum_{\substack{u\in[n]^{V(T)}\\(u_i,u_j)=(i,j)}}\prod_{e=(a,b)\in E(T)}\sigma_{u_au_b}^{q(e)}\le\sum_{u\in[n]^{V(T)}}\prod_{e=(a,b)\in E(T)}b_{u_au_b}^{(e)}$$
$$=:W^b(T).$$

We can readily use [74, Lemma 2.10] for $W^b(T)$ and deduce that

$$W^b(T)\le\prod_{e\in E(T)}\left\{\sum_{u\in[n]}\left(\sum_{v\in[n]}b_{uv}^{(e)}\right)^{p_e}\right\}^{1/p_e},$$

where $(p_e)_{e\in E(T)}$ is any collection of conjugate exponents (namely $p_e\ge1$ for all e, and $\sum_e 1/p_e=1$). We set $p_e=\infty$ for all $e\ne t$ and $p_t=1$, so

$$W^b(T)\le\sigma_t^{q(t)}\prod_{e\in E(T)\setminus\{t\}}\max_{u\in[n]}\sum_{v\in[n]}b_{uv}^{(e)}.$$

Since $q(e)\ge2$ for all $e\in E(H)$ and $\sigma_{uv}\le\sigma_n^*$, by double stochasticity, we get

$$W^b(T)\le\sigma_t^{q(t)}\prod_{e\in E(T)\setminus\{t\}}(\sigma_n^*)^{q(e)-2}.$$

We deduce that

$$\sum_{\substack{\mathbf{u}\in[n]^{pk}:\\(H_{\mathbf{u}},q_{\mathbf{u}})=(H,q),\\t\in u^{(l)}\forall l\in[p]}}\sigma_{u^{(1)}}\cdots\sigma_{u^{(q)}}\le\sigma_t^{q(t)}\prod_{e\in E(T)\setminus\{t\}}(\sigma_n^*)^{q(e)-2}\prod_{e\in E(H)\setminus E(T)}(\sigma_n^*)^{q(e)}.$$

As $\sum_{e\in E(H)}q(e)=pk$, $q(t)\ge p$ and $|E(T)|=|V(H)|-1$, we get

$$\sum_{\substack{\mathbf{u}\in[n]^{pk}:\\(H_{\mathbf{u}},q_{\mathbf{u}})=(H,q),\\t\in u^{(l)},\forall l\in[p]}}\sigma_{u^{(1)}}\cdots\sigma_{u^{(q)}}\le\sigma_t^{p}(\sigma_n^*)^{pk-p-2(|V(H)|-2)}.\tag{4.15}$$

Combining (4.14) and (4.15), we get

$$\mathbb{E}\,\Delta_t^p\le C_p(L_n\sigma_t)^p\sum_{(H,q)}(L_n\sigma_n^*)^{pk-p-2(|V(H)|-1)}(\sigma_n^*)^2.$$

To conclude the argument, since $q(t)\ge p$ and $q(e)\ge2$ for all $e\in E(H)$, we have

$$|E(H)|\le\frac{pk-p}{2}+1.$$

In particular,

$$|V(H)|\le|E(H)|+1\le\frac{pk-p}{2}+2.$$

By the choice of L_n in (4.10), we get $L_n\sigma_n^* \le 1$. The summation over the tuple (H, q) has a finite number of terms depending only on p and k. By replacing the value of L_n given in (4.10), we deduce that

$$\mathbb{E}\,\Delta_t^p \le C_{p,k}\sigma_t^p L_n^{p-2}$$

Markov's inequality implies that

$$\mathbb{P}(|\Delta_t| \ge s\sigma_t) \le C_{p,k}\frac{L_n^{p-2}}{s^p}. \tag{4.16}$$

Since $L_n \le \log^c n$, applying (4.16) for $s = n^\varepsilon$ and $p' = (p+3)/\varepsilon$, we get

$$\mathbb{P}(|\Delta_t| \ge n^\varepsilon\sigma_t) \le C_{p',k}\frac{\log^{(p+3)c/\varepsilon} n}{n^{p+3}} \le \frac{1}{n^{p+2}},$$

for all $n \ge n(\varepsilon, k, p)$, where $n(\varepsilon, k, p)$ is a large constant depending on ε, k, p. This implies (4.13), and Proposition 4.8 follows. □

4.A.3 The non-universal regime for the limiting ESD

The goal of this subsection is to show that the sparsity assumption $\sigma_n^* \to 0$ is necessary for Theorem 4.1 for convergence to the semi-circle law. The following proposition formalizes this, and the proof is provided for completeness.

Proposition 4.9. *Let $d \ge 2$ be constant with respect to n, and assume $n \to \infty$ with nd even.*

- *(Distributional Non-universality) Let W_n and W_n' be independent symmetric matrices, with entries drawn as independent copies of bounded random variables ξ and ξ' respectively, both of which are symmetrically distributed and have unit variance. If the laws of ξ and ξ' are not equal, then there is a sequence of d-regular graphs with adjacency matrices A_n, such that the matrices $X_n := \frac{1}{\sqrt{d}}A_n \circ W_n$ and $X_n' := \frac{1}{\sqrt{d}}A_n \circ W_n'$ have different limiting ESD almost surely.*
- *(Structural Non-universality) Let W_n be a symmetric matrix with independent Rademacher entries. There exist two sequences of d-regular graphs with adjacency matrices A_n and A_n' so that $X_n := \frac{1}{\sqrt{d}}A_n \circ W_n$ and $X_n' := \frac{1}{\sqrt{d}}A_n' \circ W_n$ have different limiting ESD almost surely.*

Remark 4.10. We note that the distributional non-universality in Proposition 4.9 was originally proved in [53] for graphs G_n that are locally tree-like in the Benjamini and Schramm topology [24]. The authors of [53], however, focus on the "eigendistributions" of the limit ESD of $A_{G_n} \circ W_n$ as a function of ξ.

Proof. Given a centered random variable ξ with unit variance, we denote by W_n^ξ the $n \times n$ Wigner matrix whose entries on and above the main diagonal are iid copies of ξ. We fix an

integer $d \geq 2$ independent of n, and for every d-regular graph G_n on n vertices, we denote by A_{G_n} its adjacency matrix. To prove distributional non-universality, we will construct a sequence of d-regular graphs G_n and show that the limiting spectral distribution of $\frac{1}{\sqrt{d}} A_{G_n} \circ W_n^\xi$ depends on ξ. To show the dependence on the graph structure, we will set ξ to be a Rademacher variable and construct two sequences of d-regular graphs G_n and G'_n such that the limiting spectral distributions of $\frac{1}{\sqrt{d}} A_{G_n} \circ W_n^\xi$ and $\frac{1}{\sqrt{d}} A_{G'_n} \circ W_n^\xi$ are different.

Regarding distributional non-universality, let C_d be the clique in $d+1$ vertices, i.e., a complete graph on $d+1$ vertices excluding self-loops. To simplify the exposition, we will suppose that n is a multiple of $d+1$ and let G_n be a d-regular graph on n vertices given by the union of $n/(d+1)$ cliques C_d. Denote the entries of A_{G_n} by $(a_{ij})_{1\leq i,j\leq n}$, the entries of W_n^ξ by $(w_{ij}^{(\xi)})_{1\leq i,j\leq n}$, and set $X_n = \frac{1}{\sqrt{d}} A_{G_n} \circ W_n^\xi$. We can write

$$\frac{1}{n}\mathbb{E}\,\mathrm{tr}(X_n^{2k}) = \frac{1}{d^k}\frac{1}{n}\sum_{u\in[n]^{2k}} a_{u_1u_2}\cdots a_{u_{2k}u_1}\,\mathbb{E}\,w_{u_1u_2}^{(\xi)}\cdots w_{u_{2k}u_1}^{(\xi)},$$

for every integer k. Fix a vertex $o \in C_d$. For a graph H and a vertex u_1 in H, let $\mathcal{P}(u_1, H, 2k)$ be the set of paths in H starting and ending at u_1 and of length $2k$. Then, from the above

$$\frac{1}{n}\mathbb{E}\,\mathrm{tr}(X_n^{2k}) = \frac{1}{nd^k}\sum_{u_1\in[n]}\sum_{u\in\mathcal{P}(u_1,C_d(u_1),2k)} \mathbb{E}\,w_{u_1u_2}^{(\xi)}\cdots w_{u_{2k}u_1}^{(\xi)},$$

where by $C_d(u_1)$ we denoted the $(d+1)$–clique in G_n containing u_1. Since G_n is a disjoint union of copies of C_d and the variables in each clique are iid, we deduce that

$$\frac{1}{n}\mathbb{E}\,\mathrm{tr}(X_n^{2k}) = \frac{1}{d^k}\sum_{u\in\mathcal{P}(o,C_d,2k)} \mathbb{E}\,w_{u_1u_2}^{(\xi)}\cdots w_{u_{2k}u_1}^{(\xi)} =: m((C_d,o),2k,\xi). \tag{4.17}$$

It is immediate to see that the sequence $(m((C_d,o),2k,\xi))_{k\geq 1}$ uniquely determines all the even moments of ξ (more specifically, if $k_{\min}$ is the smallest integer such that two distinct normalized symmetric distributions ξ and ξ' have different $2k_{\min}$–th moments then necessarily $m((C_d,o),2k_{\min},\xi) \neq m((C_d,o),2k_{\min},\xi'))$. This completes the proof of distributional non-universality. Let us remark here that the extra assumption $n \mod (d+1) = 0$ which we used in our construction, is not essential and can be easily removed by letting G_n to be a disjoint union of $\lfloor n/(d+1)\rfloor - 1$ cliques and a d–regular graph of size $n-(d+1)\lfloor n/(d+1)\rfloor + (d+1)$ having arbitrary topology.

To prove the structural non-universality, we now set ξ to be a ± 1 Rademacher random variable. We assume $d \geq 3$, as the case $d = 2$ can be verified by inspection. Again, to simplify the exposition, we will suppose that n is a multiple of $d+1$. Consider the sequence of graphs G_n introduced above, and let G'_n be a sequence of randomly uniformly chosen d-regular graphs on n vertices. Then it was proved in [78] that G'_n converges (in the Benjamini and Schramm sense [24]) to the infinite rooted d-regular tree (T_d, o). If $X' = \frac{1}{\sqrt{d}} A_{G'_n} \circ W_n^\xi$ where G'_n and W_n^ξ are independent, the above argument and the local

convergence of G'_n imply that

$$\frac{1}{n}\,\mathbb{E}\,\mathrm{tr}((X'_n)^{2k}) \to \frac{1}{d^k} \sum_{u\in\mathcal{P}(o,T_d,2k)} \mathbb{E}\, w^{\xi}_{u_1u_2}\cdots w^{\xi}_{u_{2p}u_1} = m((T_d,o),2k,\xi). \tag{4.18}$$

Let $k=3$, then, by counting the shapes, we can compute

$$|\mathcal{P}(o,T_d,6)| = 3d(d-1)^2 + 6d(d-1) + 2d(d-1)(d-2) + d,$$

so

$$\frac{d^3}{n}\,\mathbb{E}\,\mathrm{tr}((X'_n)^6) \to 3d(d-1)^2 + 6d(d-1) + 2d(d-1)(d-2) + d.$$

On the other hand, for the clique C_d, we have

$$\frac{d^3}{n}\,\mathbb{E}\,\mathrm{tr}(X_n^6) = 2d(d-1)^2 + 8d(d-1) + 3d(d-1)(d-2) + d.$$

Indeed, the shape $s = o \to v_1 \to v_2 \to v_3 \to v_2 \to v_1 \to o$ that originally appeared for T_d and its contribution $d(d-1)^2$ has to be decomposed on whether $v_3 = o$ for the clique. In this case, the shape is two laps on a triangle, and inverting the orientation of the second lap implies the contribution of $2d(d-1)$. If $v_3 \neq o$, then its contribution is $d(d-1)(d-2)$ and the result follows. Subtracting one from the other, we get

$$m((C_d,o),6,\xi) - m((T_d,o),6,\xi) = \frac{d(d-1)}{d^3}.$$

In particular, structural non-universality follows as $d \geq 2$. □

Chapter 5

Almost sharp covariance estimation for inhomogeneous random matrices [91]

What question does the Cauchy distribution not like?
"Got a moment?"

Unknown

Let $X_1, ..., X_n \in \mathbb{R}^d$ be independent Gaussian random vectors with independent entries and variance profile $(b_{ij})_{i\in[d],j\in[n]}$. A major question in the study of covariance estimation is to give precise control on the deviation of $\sum_{j\in[n]} X_j X_j^T - \mathbb{E}\, X_j X_j^T$. In this paper, we prove new bounds for

$$\mathbb{E}\left\| \sum_{j\in[n]} X_j X_j^T - \mathbb{E}\, X_j X_j^T \right\|.$$

The proofs are based on the moment method and a careful analysis of the structure of the shapes that matter. We also provide examples showing improvement over the past works and matching lower bounds.

Contents

5.1 Introduction

The study of the norm of random matrices has increased significantly over the years, and bounding the operator norm has been proved one central topic in the field [12, 61, 19, 74]. Particularly, several applications coming from statistics require a precise, sharp control on the deviations of the empirical covariance problem [69, 79, 117]. For instance, it is well-known [108, Theorem 4.7.1] that an i.i.d sample $X_1, ..., X_n \in \mathbb{R}^d$ of isotropic Gaussian random vectors satisfies the following deviation

$$\mathbb{E}\left\|\frac{1}{n}\sum_{j\in[n]} X_j X_j^T - \mathbb{E}\, X_1 X_1^T\right\| \lesssim \frac{d}{n} \vee \sqrt{\frac{d}{n}}. \tag{5.1}$$

Only very recently, the identically distributed condition was removed, and universal bounds were obtained for the sample covariance model [18, 31, 33]. Our contribution comes precisely in this direction. We improve their results and get sharper error factors.

We begin our results for the operator norm. Define the parameters:

- $\tilde{\sigma}_\infty^2 = \max_{i,l:i\neq l} \sum_{j\in[n]} b_{ij}^2 b_{lj}^2$;
- $\bar{\sigma}_\infty^2 = \max_{i\in[d]} \sum_{j\in[n]} b_{ij}^4$;
- $\sigma_\infty^2 = \max_{i\in[d]} \sum_{j\in[n]} \sum_{l:l\neq i} b_{ij}^2 b_{lj}^2$;
- $\sigma_C^2 = \max_{j\in[n]} \sum_{i\in[d]} b_{ij}^2$;
- $\sigma_* = \max_{(i,j)\in[d]\times[n]} |b_{ij}|$;
- $\beta_\infty = \dfrac{\tilde{\sigma}_\infty \sigma_C}{\sigma_\infty \sigma_*}$.

Notice in particular that $\tilde{\sigma}_\infty \leq \bar{\sigma}_\infty$, by Cauchy-Schwarz inequality. Moreover, we point out that the parameters σ_C and σ_* already appeared in previous works on the estimation of the operator norm [74, 19] and the covariance estimation [33], whereas $\sigma_\infty, \overline{\sigma}_\infty$ appeared in [31].

Theorem 5.1. *Let X be a $d \times n$ Gaussian matrix with independent entries such that $X_{ij} = b_{ij} g_{ij}$ where $\{g_{ij} : (i,j) \in [d]\times[n]\}$ are i.i.d standard Gaussian r.v. Let $\varepsilon \in (0, 1/2)$. Then, the following holds.*

1. *If $\beta_\infty \leq 1$, we have*

$$\mathbb{E}\left\|XX^T - \mathbb{E}\, XX^T\right\| = \mathbb{E}\left\|\sum_{j\in[n]} X_j X_j^T - \mathbb{E}\, X_j X_j^T\right\|$$
$$\leq (1+\varepsilon)\left\{2\sigma_\infty + \sigma_C^2 + C(\varepsilon)\sigma_*\left(\sigma_C + \frac{\sigma_\infty}{\sigma_C}\right)\sqrt{\log(n\wedge d)} + C^2(\varepsilon)\sigma_*^2 \log(n\wedge d)\right\}.$$

2. Otherwise, $\beta_\infty > 1$ and we have

$$\begin{aligned}&\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \\ &\leq (1+\varepsilon)\left\{\frac{2\tilde{\sigma}_\infty\sigma_C}{\sigma_*} + \sigma_C^2 + C(\varepsilon)(\sigma_C\sigma_* + \bar{\sigma}_\infty)\sqrt{\log(n\wedge d)} + C^2(\varepsilon)\sigma_*^2\log(n\wedge d)\right\}.\end{aligned}$$

The constant $C(\varepsilon)$ is

$$C(\varepsilon) = \frac{C(1+\varepsilon)}{\sqrt{\log(1+\varepsilon)}} \leq \frac{C}{\sqrt{\varepsilon}},$$

where C is a universal constant.

Theorem 5.1 improves Theorem 2.1 from [33] and in the case $\beta_\infty \leq 1$, it improves both [31, Theorem 3.17], [18, Theorem 3.12]. We do not know whether the two regimes over β_∞ are necessary. One would expect that the leading term $2\sigma_\infty + \sigma_C^2$ is always sharp, as it was previously obtained in [18, Thereom 3.12] with a larger error term (already dominant for the homogeneous case); see Theorem 5.18. Our present method allows us to capture better estimates and error factors. See discussion in Section 5.3. For the homogeneous case $b_{ij} = 1$ for all i, j, it was shown in [33, Proposition 2.3],[13] that

$$\liminf_{n,d\to\infty} \frac{\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\|}{2\sigma_\infty + \sigma_C^2} \geq 1,$$

hence, the numerical factors of 2 and 1 in the leading term are optimal.

Since the method of proof uses the moment method, we can extend Theorem 5.1 to estimate Schatten norms. Recall that the p-Schatten norm is defined by

$$\|A\|_{S_p}^p = \mathrm{Tr}(A)^p,$$

for a positive matrix A. It is also the same as the p-norm of the singular values of A. This time, we define the more involved parameters

- $\sigma_p = \left\{\sum_{i\in[d]}\left[\sum_{j\in[n]}\sum_{l\in[d]} b_{ij}^2 b_{lj}^2\right]^{p/2}\right\}^{1/p}$;
- $b_p = \left\{\sum_{i\in[d]}\max_{j\in[n]} b_{ij}^{2p}\right\}^{1/(2p)}$;
- $\overline{\sigma}_p = \left\{\sum_{i\in[d]}\left[\sum_{j\in[n]} b_{ij}^4\right]^{p/2}\right\}^{1/p}$;
- $\beta_p = \dfrac{\bar{\sigma}_p\sigma_C}{\sigma_p b_p}$.

Our second main theorem is the following.

Theorem 5.2. *Let $p \in \mathbb{N}$ and X be a $d\times n$ Gaussian matrix with independent entries such that $X_{ij} = b_{ij}g_{ij}$ where $\{g_{ij} : (i,j) \in [d]\times[n]\}$ are i.i.d standard Gaussian r.v. Then, the following holds.*

1. *If $\beta_p \le 1$, we have*

$$(\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\|_{S_p}^p)^{1/p} \le d^{1/p}\left\{2\sigma_p + \sigma_C^2 + C\sqrt{p}\left(\sigma_C\sigma_* + \frac{\sigma_p\sigma_*}{\sigma_C}\right) + C'pb_p^2\right\}.$$

2. *Otherwise, $\beta_p > 1$ and*

$$(\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\|_{S_p}^p)^{1/p} \le d^{1/p}\left\{\frac{2\overline{\sigma}_p\sigma_C}{\sigma_*} + \sigma_C^2 + C\sqrt{p}(\sigma_C\sigma_* + \overline{\sigma}_p) + C'pb_p^2\right\}.$$

For $p = \lceil \log d \rceil$ and $\beta_p \le 1$, Theorem 5.2 implies that

$$(\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\|^2)^{1/2} \lesssim \left(\sigma_\infty + \sigma_C^2 + \sqrt{\log d}\left(\sigma_C\sigma_* + \frac{\sigma_\infty\sigma_*}{\sigma_C}\right) + \sigma_*^2\log d\right).$$

We will show in Lemma 5.22 that, indeed, the leading term is a lower bound for the operator norm squared

$$(\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\|^2)^{1/2} \ge \max\{\sigma_\infty, 0.49\sigma_C^2\},$$

which proves that the leading term is optimal. We will also prove a lower bound for the Schatten norm in Proposition 5.23.

Remark 5.3. Since the proof uses the moment method and the Schatten norm, it is well-known (see, e.g., [74, Theorem 4.4], [33, Corollary 3.3] or [19, Corollary 3.2]), universal constants aside, that the conclusion of Theorems 5.1 and 5.2 holds for random matrices of the form $Y = (h_{ij})_{i\in[d],j\in[n]}$, where h_{ij} are independent symmetric subgaussian random variables with $\|h_{ij}\|_{\psi_2} \le b_{ij}$ (see definition in [108, Section 2.5]). This is why we focus exclusively on the Gaussian case in this paper.

5.1.1 Main ideas of the proof

The proof relies on the moment method and a careful analysis of paths. We will first remove the diagonal $\mathrm{Diag}(XX^T)$, so that

$$\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \le \mathbb{E}\left\|\Delta XX^T\right\| + \mathbb{E}\left\|\mathrm{Diag}(XX^T) - \mathbb{E}\,XX^T\right\|,$$

where ΔXX^T is the matrix of off-diagonal elements of XX^T. It turns out that the contribution of the diagonal is sufficiently small and can be added as an error factor (see Theorem 5.8). On the other hand, the combinatorics of ΔXX^T are much easier to deal with. In particular, all paths in the complete bipartite graph over $[d] \sqcup [n]$ have all right vertices with at least two neighbors. We then proceed with the moment method. Note that

$$\mathbb{E}\|Y\| \le (\mathbb{E}\|Y\|_{S_p}^p)^{1/p},$$

by Jensen's Inequality. We apply this for $Y = \Delta XX^T$, and our goal is to obtain a comparison lemma such as

$$\mathbb{E}\,\mathrm{Tr}(\Delta XX^T)^p \le \kappa\,\mathbb{E}\,\mathrm{Tr}(\Delta GG^T)^p,$$

where $\kappa > 0$ and G is a Gaussian matrix with reduced dimensions as in [19].

5.1.2 Outline of the paper

The paper is organized as follows. In Section 5.2, we will provide the proofs of Theorems 5.1 and 5.2. In Section 5.3, we will give examples to illustrate the improvement from the previous results. Finally, in Section 5.4, we will prove sharp matching lower bounds.

Notation. Let us clarify some notation used throughout the paper. We denote $a \lesssim b$ or $a = O(b)$ if there exists an absolute constant C such that $a \le Cb$. We also denote it as $b \gtrsim a$. If $a \lesssim b$ and $b \lesssim a$ hold, we denote $a \asymp b$. We write $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. We denote $[n] = \{1, \dots, n\}$ and $A \sqcup B$ is the disjoint union of two sets A and B. Finally, we use $C, c, C', \dots$ for universal numerical constants.

5.2 Proofs

5.2.1 Preliminaries

We begin by recalling the Gaussian integration by parts lemma.

Lemma 5.4. *Let $g \sim N(0,1)$ be a standard Gaussian r.v. and $f \in C^1(\mathbb{R})$, then*

$$\mathbb{E}\,gf(g) = \mathbb{E}\,f'(g).$$

The authors of [33, Lemma 5.2] deduced from this lemma a simple property of the joint moments of g and $g^2 - 1$.

Lemma 5.5. *Let $a_{n,m} = \mathbb{E}\,g^n(g^2-1)^m$, where $g \sim N(0,1)$. Then $a_{n,m} \ge 0$ and $a_{n,m} = 0$ if and only if n is odd or $(n,m) = (0,1)$. Moreover, if n is even, then*

$$(n+m-1)(n+2m-3)!! \le a_{n,m} \le (n+2m-1)!!,$$

where $(2n+1)!! = (2n+1)(2n-1)\dots 1$ is the double factorial. In particular, for any $p \ge 2$ or $p = 0$, we have

$$c^p p! \le \mathbb{E}(g^2-1)^p \le C^p p!.$$

We also recall the sharp bound on the operator norm for a standard Gaussian matrix shown in [33, Lemma 2.6].

Proposition 5.6. *Let G be a $d \times n$ Gaussian matrix with i.i.d standard Gaussian r.v. entries. Then, for any $p \geq 2$ we have*

$$(\mathbb{E}\left\|GG^T - \mathbb{E}\, GG^T\right\|^p)^{1/p} \leq 2\sqrt{dn} + d + 4\sqrt{p}(\sqrt{d} + \sqrt{n}) + 2p.$$

Note that

$$\mathbb{E}\left\|\mathrm{Diag}(GG^T) - \mathbb{E}\, GG^T\right\|^p = \mathbb{E} \max_{i\in[d]} \left(\sum_{j\in[n]} (g_{ij}^2 - 1)\right)^p.$$

Bernstein's Inequality [108, Theorem 2.8.1] implies then that

$$(\mathbb{E}\left\|\mathrm{Diag}(GG^T) - \mathbb{E}\, GG^T\right\|^p)^{1/p} \lesssim \sqrt{pn} + p.$$

Consequently, we end this subsection with a corollary for the off-diagonal part.

Corollary 5.7. *Let G be a $d \times n$ Gaussian matrix with i.i.d standard Gaussian entries. Then, for any $p \geq 2$ we have*

$$(\mathbb{E}\left\|\Delta(GG^T)\right\|^p)^{1/p} \leq 2\sqrt{dn} + d + C\sqrt{p}(\sqrt{d} + \sqrt{n}) + C'p.$$

5.2.2 The diagonal part

In this section, the main result is the following.

Theorem 5.8. *Let X be a $d \times n$ Gaussian matrix with independent entries such that $X_{ij} = b_{ij}g_{ij}$ and $\{g_{ij} : (i,j) \in [d] \times [n]\}$ are i.i.d Gaussian random variables. Then, for any even integer $p \geq 2$, we have*

$$\left(\mathbb{E}\left\|\mathrm{Diag}(XX^T) - \mathbb{E}\, XX^T\right\|_{S_p}^p\right)^{1/p} \asymp \sqrt{p}\overline{\sigma}_p + pb_p^2.$$

We first note that

$$\left(\mathbb{E}\left\|\mathrm{Diag}(XX^T) - \mathbb{E}\, XX^T\right\|_{S_p}^p\right) = \sum_{i\in[d]} \mathbb{E}\left(\sum_{j\in[n]} b_{ij}^2(g_{ij}^2 - 1)\right)^p,$$

and it suffices to control each summand.

Lemma 5.9. *Let $(b_k)_{k\in[n]}$ be positive numbers and $(g_k)_{k\in[n]}$ be independent standard Gaussians $N(0,1)$. Then, for any even integer $p \geq 2$, we have*

$$\left[\mathbb{E}\left(\sum_{j\in[n]} b_k^2(g_k^2 - 1)\right)^p\right]^{1/p} \asymp \sqrt{p}\left(\sum_{k\in[n]} b_k^4\right)^{1/2} + p\max_{k\in[n]} b_k^2.$$

Proof. Since g_k are Gaussian, $g_k^2 - 1$ are independent, centered, and subexponential ran-

dom variables. In particular, Bernstein's Inequality [108, Theorem 2.8.1] implies that

$$\mathbb{P}\left(\left|\sum_{k\in[n]} b_k^2(g_k^2-1)\right| \geq t\right) \leq 2\exp\left(-c\min\left\{\frac{t^2}{a_1^2}, \frac{t}{a_2}\right\}\right),$$

where

$$a_1 = \sum_{k\in[n]} b_k^4; \qquad a_2 = \max_{k\in[n]} b_k^2.$$

Therefore, we have

$$\left[\mathbb{E}\left(\sum_{j\in[n]} b_k^2(g_k^2-1)\right)^p\right]^{1/p} \lesssim \sqrt{p}\left(\sum_{k\in[n]} b_k^4\right)^{1/2} + p\max_{k\in[n]} b_k^2.$$

For the lower bound, we first note that

$$\mathbb{E}\left(\sum_{j\in[n]} b_k^2(g_k^2-1)\right)^p = \sum_{k_1+\cdots+k_n=p} \frac{p!}{k_1!\cdots k_n!}\prod_{l=1}^n b_l^{2k_l}\,\mathbb{E}(g_l^2-1)^{k_l}.$$

Here, whenever there exists $k_i = 1$, the summand is zero. Therefore, we can assume that the summation is restricted to $k_i \neq 1$ for all i. Lemma 5.5 implies then that

$$\mathbb{E}\left(\sum_{j\in[n]} b_k^2(g_k^2-1)\right)^p \geq c^p p! \sum_{\substack{k_1+\cdots+k_n=p\\ k_i\neq 1,\forall i}} \prod_{l=1}^n b_l^{2k_l}.$$

If we restrict it to $k_j = p$ for some j, we get

$$\mathbb{E}\left(\sum_{j\in[n]} b_k^2(g_k^2-1)\right)^p \geq c^p p! \sum_{j\in[n]} b_j^{2p},$$

that is,

$$\left(\mathbb{E}\left(\sum_{j\in[n]} b_k^2(g_k^2-1)\right)^p\right)^{1/p} \geq cp\left(\sum_{j\in[n]} b_j^{2p}\right)^{1/p}. \tag{5.2}$$

Moreover, since p is even, if we consider now $k_j = 2a_j$ and $\sum_{j\in[n]} a_j = p/2$, we get

$$\mathbb{E}\left(\sum_{j\in[n]} b_k^2(g_k^2-1)\right)^p \geq c^p p! \sum_{a_1+\cdots+a_n=p/2} \prod_{l=1}^n b_l^{4a_l}.$$

This implies that

$$\mathbb{E}\left(\sum_{j\in[n]} b_k^2(g_k^2-1)\right)^p \geq c^p \frac{p!}{(p/2)!} \sum_{a_1+\cdots+a_n=p/2} \frac{(p/2)!}{a_1!\cdots a_n!} \prod_{l=1}^n b_l^{4a_l}$$
$$\geq c^p \frac{p!}{(p/2)!} \left(\sum_{j\in[n]} b_j^4\right)^{p/2},$$

that is,

$$\left(\mathbb{E}\left(\sum_{j\in[n]} b_k^2(g_k^2-1)\right)^p\right)^{1/p} \geq c\sqrt{p}\left(\sum_{j\in[n]} b_j^4\right)^{1/2}. \tag{5.3}$$

Combining Inequalities (5.2) and (5.3), we get

$$\left(\mathbb{E}\left(\sum_{j\in[n]} b_k^2(g_k^2-1)\right)^p\right)^{1/p} \gtrsim \sqrt{p}\left(\sum_{k\in[n]} b_k^4\right)^{1/2} + p\left(\sum_{k\in[n]} b_k^{2p}\right)^{1/p}$$
$$\gtrsim \sqrt{p}\left(\sum_{k\in[n]} b_k^4\right)^{1/2} + p\max_{k\in[b]} b_k^2,$$

and this finishes the proof. □

Proof of Theorem 5.8. By Lemma 5.9, we get that

$$\left[\mathbb{E}\,\mathrm{Tr}(\mathrm{Diag}(XX^T) - \mathbb{E}\,XX^T)^p\right]^{1/p} \lesssim \left[\sum_{i\in[d]}\left(\sqrt{p}\left(\sum_{j\in[n]} b_{ij}^4\right)^{1/2} + p\max_{j\in[n]} b_{ij}^2.\right)^p\right]^{1/p}$$
$$\lesssim \sqrt{p}\overline{\sigma}_p + pb_p^2,$$

where the last inequality follows by the triangle inequality. The lower bound follows similarly considering Inequalities (5.2) and (5.3) separately. □

5.2.3 The off-diagonal part

The proof of the bounds for the off-diagonal part follows the moment method, but a precise analysis is needed to control the contribution of each path and summand. First, we open the trace so that

$$\mathbb{E}\,\mathrm{Tr}(\Delta XX^T)^p = \sum_{u\in[d]^p} \mathbb{E}\prod_{k=1}^p (XX^T)_{u_k u_{k+1}} \mathbf{1}_{u_k\neq u_{k+1}}$$
$$= \sum_{u\in[d]^p}\sum_{v\in[n]^p} \mathbb{E}\prod_{k=1}^p X_{u_k v_k} X_{u_{k+1} v_k} \mathbf{1}_{u_k\neq u_{k+1}},$$

where $u_{p+1} := u_1$ and $p \in \mathbb{N}$. We view the path $u_1 \to v_1 \to u_2 \to \cdots \to u_p \to v_p \to u_1$ as a cycle in the complete bipartite graph over $[d]^{(l)} \sqcup [n]^{(r)}$, where (l) and (r) indicate left

and right vertices (we will remove the indexes if the context is clear). For a path (u, v), we define its shape $s(u, v)$ as relabelling its vertices in order of appearance. For instance, the path

$$3 \to 2' \to 4 \to 1' \to 3 \to 1' \to 4 \to 5' \to 3$$

has shape

$$1 \to 1' \to 2 \to 2' \to 1 \to 2' \to 2 \to 3' \to 1.$$

Note that each edge $u_k v_k$ and $u_{k+1} v_k$ must appear at least twice in the path (u, v), by the independence of the Gaussian r.v. and symmetry. Call the shapes that satisfy this *even*. Let then $\mathcal{S}$ be the set of even shapes $s = (u, v)$ such that $u_k \neq u_{k+1}$ for all $k = 1, ..., p$. Moreover, the product

$$L(s) := \mathbb{E} \prod_{k=1}^{p} g_{u_k v_k} g_{u_{k+1} v_k} \tag{5.4}$$

only depends on the shape of (u, v), therefore we have

$$\mathbb{E} \operatorname{Tr}(\Delta X X^T)^p = \sum_{s \in \mathcal{S}} L(s) \sum_{\substack{(u,v) \in [d]^p \times [n]^p \\ s(u,v) = s}} \prod_{k=1}^{p} b_{u_k v_k} b_{u_{k+1} v_k}.$$

Let $(m_1, m_2) = (m_1(s), m_2(s))$ be the quantity of right and left vertices that appear in the shape s. The key proposition to prove Theorem 5.1 is to bound

$$W(s) := \sum_{\substack{(u,v) \in [d]^p \times [n]^p \\ s(u,v) = s}} \prod_{k=1}^{p} b_{u_k v_k} b_{u_{k+1} v_k} \tag{5.5}$$

according to the number of vertices visited by the path.

Proposition 5.10. *Assume $\sigma_* = 1$. If $\beta_\infty \leq 1$, we have*

$$W(s) \leq \left[d \Big(\frac{\sigma_\infty}{\sigma_C} \Big)^{2m_1} \sigma_C^{2(m_2 - 1)} \right] \wedge \left[n \Big(\frac{\sigma_\infty}{\sigma_C} \Big)^{2(m_1 - 1)} \sigma_C^{2m_2} \right].$$

Otherwise, $\beta_\infty > 1$ and we have

$$W(s) \leq \left[d \tilde{\sigma}_\infty^{2m_1} \sigma_C^{2(m_2 - 1)} \right] \wedge \left[n \tilde{\sigma}_\infty^{2(m_1 - 1)} \sigma_C^{2m_2} \right].$$

Let us prove Theorem 5.1 given Proposition 5.10.

Proof of Theorem 5.1. Assume $\sigma_* = 1$ (by homogeneity) and $\beta_\infty \le 1$. Let

$$a := \frac{\sigma_\infty}{\sigma_C};$$
$$b := \sigma_C.$$

Then, using the first bound on Proposition 5.10, we have

$$\mathbb{E}\,\mathrm{Tr}(\Delta X X^T)^p \le d \sum_{s\in\mathcal{S}} L(s) \Big(\frac{\sigma_\infty}{\sigma_C}\Big)^{2m_1} \sigma_C^{2(m_2-1)}.$$

On the other hand, for a standard Gaussian $r_2 \times r_1$ matrix G, we have

$$\mathbb{E}\,\mathrm{Tr}(\Delta G G^T)^p = \sum_{s\in\mathcal{S}} L(s) \frac{r_1!}{(r_1-m_1)!} \frac{r_2!}{(r_2-m_2)!},$$

for any $r_1, r_2 > p/2$ (see [33]). In particular, if $r_1 = \lceil a^2 \rceil + p/2$ and $r_2 = \lceil b^2 \rceil + p/2$, we have

$$\frac{r_1!}{(r_1-m_1)!} \ge r_1 \cdots (r_1 - m_1 + 1)^{m_1} \ge a^{2m_1},$$

and

$$\frac{r_2!}{(r_2-m_2)!} \ge r_2 b^{2(m_2-1)}.$$

Hence

$$\mathbb{E}\,\mathrm{Tr}(\Delta X X^T)^p \le \frac{d}{r_2}\,\mathbb{E}\,\mathrm{Tr}(\Delta G G^T)^p \le d\,\mathbb{E}\big\|\Delta G G^T\big\|^p.$$

Now we estimate the latter by Corollary 5.7 so that

$$(\mathbb{E}\big\|\Delta G G^T\big\|^p)^{1/p} \le 2\sqrt{r_1 r_2} + r_2 + C\sqrt{p}(\sqrt{r_1} + \sqrt{r_2}) + C' p.$$

Together with Theorem 5.8, we deduce that

$$\mathbb{E}\big\|X X^T - \mathbb{E}\,X X^T\big\| \le d^{\frac{1}{p}} \Big\{ 2\sigma_\infty + \sigma_C^2 + C\sqrt{p}\Big(\sigma_C + \frac{\sigma_\infty}{\sigma_C} + \bar{\sigma}_p\Big) + C p b_p^2 \Big\}.$$

Choose $p = \lceil \alpha \log d \rceil$. Since $\beta_\infty \le 1$, we have that

$$\bar{\sigma}_p \le d^{\frac{1}{p}} \bar{\sigma}_\infty \le d^{\frac{1}{p}} \frac{\sigma_\infty}{\sigma_C}.$$

Moreover, $b_p \le d^{\frac{1}{2p}} b_\infty$, thus

$$\mathbb{E}\big\|X X^T - \mathbb{E}\,X X^T\big\| \le e^{\frac{1}{\alpha}} \Big\{ 2\sigma_\infty + \sigma_C^2 + C e^{\frac{1}{\alpha}} \sqrt{\alpha \log d}\Big(\sigma_C + \frac{\sigma_\infty}{\sigma_C}\Big) + C\alpha e^{\frac{1}{2\alpha}} \log d \Big\}.$$

Finally, set $1+\varepsilon = e^{\frac{1}{\alpha}}$, hence

$$\alpha = \frac{1}{\log(1+\varepsilon)},$$

and we get

$$\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \le (1+\varepsilon)\left\{2\sigma_\infty + \sigma_C^2 + C(\varepsilon)\sqrt{\log d}\left(\sigma_C + \frac{\sigma_\infty}{\sigma_C}\right) + C^2(\varepsilon)\log d\right\}.$$

This gives the upper bound with $\log d$. The second bound in Proposition 5.10 yields the upper bound with $\log n$, and the general case for $\beta_\infty \le 1$ follows. The case $\beta_\infty > 1$ is proved similarly. Indeed, we now set (a, b) to be

$$a = \tilde{\sigma}_\infty;$$
$$b = \sigma_C,$$

and then the previous proof follows straightforwardly.

□

Now we prove Proposition 5.10.

Proof of Proposition 5.10. To simplify the notation, for a graph G, we will denote $e \in G$ if an edge e belongs to $E(G)$, $v \in G$ if $v \in V(G)$ and $G' = G \setminus \{v\}$ is the subgraph of G induced by the vertices $V(G) \setminus \{v\}$. We use a similar notation to $G \setminus \{e\}$ and an edge $e \in E(G)$.

Given a shape $s \in \mathcal{S}$, we define a bipartite graph G over $[m_2] \sqcup [m_1]$ so that $E(G) = \{(u_k v_k) : k \in [p]\}$. Here, $[m_2]$ denotes the left vertices and $[m_1]$ denotes the right vertices. Let k_e be the number of times each edge $e \in E(G)$ is traversed by the shape s, then $\sum_e k_e = 2p = |k|$. According to (5.5), we get an alternative expression for $W(s)$:

$$W(s) = \sum_{w_1 \ne \cdots \ne w_{m_2}} \sum_{t_1 \ne \cdots \ne t_{m_1}} \prod_{e=ij \in E(G)} b_{w_i t_j}^{k_e} =: W^k(G),$$

where the notation $w_1 \ne \cdots \ne w_{m_2}$ means that all w_k are different, similarly for t_k. Note that, by the assumption on $s \in \mathcal{S}$. Now, fix $u_1 = w_1 = z \in [d]$ and define the following first-time arrivals:

$$i_1(k) := \inf\{l : u_l = k\};\ k = 2, ..., m_2;$$
$$i_2(k) := \inf\{l : v_l = k\};\ k = 1, ..., m_1.$$

Let also $e_k^{(1)} = u_{i_1(k)} v_{i_1(k)-1}$ and $e_k^{(2)} = u_{i_2(k)} v_{i_2(k)}$. Then, all these $m_1 + m_2 - 1$ edges are distinct, and the subgraph H generated by them is a spanning tree of G.

The crucial distinction to [33, Theorem 2.1] is that we want to preserve the property that every right vertex has at least two neighbors. Call this property $\mathcal{P}$. Let us divide it into two cases.

Case I. Suppose the tree H satisfies property $\mathcal{P}$. Assume v, v' are extreme right vertices, that is,

$$d(v, v') = \max_{r,r' \in [m_1]} d(r, r').$$

(In case $m_1 = 1$, the result is trivial). Then v has exactly one neighbor $u \in [m_2]$ such that $|N(u)| \geq 2$ and it satisfies

$$d(v', u) = d(v', v) - 1,$$

that is, the unique path from v' to v passes through u. Indeed, if there are two of such vertices u, u' and u' is connected to both v and a different v'', we would have that

$$d(v', v'') = d(v', v) + d(v, v'') = d(v', v) + 2,$$

which contradicts the maximal distance of v and v'. Therefore, if $L(v) = \{u \in N(v) : |N(u)| = 1\} \cup \{v\}$ we have that the graph $H' = H \setminus L(v)$ is still a tree with the property $\mathcal{P}$. Without loss of generality, we can assume that $v = m_1$. Since $\sigma_* = 1$ and $k_e \geq 2$ for all $e \in G$, we have

$$\begin{aligned} W^k(G) &\leq d \sum_{w_2 \neq \cdots \neq w_{m_2}} \sum_{t_1 \neq \cdots \neq t_{m_1}} \prod_{e=ij \in E(H)} b^2_{w_i t_j} \\ &\leq d \left(\sum_{w_2 \neq \cdots \neq w_{m_2}} \sum_{t_1 \neq \cdots \neq t_{m_1 - 1}} \prod_{e=ij \in E(H')} b^2_{w_i t_j} \right) \max_{w \in [d]} \sum_{j \in [n]} b^2_{wj} \left(\sum_{l \in [d] : l \neq w} b^2_{lj} \right)^{|N(m_1)|-1}. \end{aligned}$$

For the second term, we further estimate

$$\max_{w \in [d]} \sum_{j \in [n]} b^2_{wj} \left(\sum_{l \in [d] : l \neq w} b^2_{lj} \right)^{|N(m_1)|-1} \leq \sigma_\infty^2 \sigma_C^{2(|N(m_1)|-2)}.$$

We then proceed by induction over the right vertices as we did for H. Here, induction is justified as H' is still in case I. In particular, that yields

$$W^k(G) \leq d \sigma_\infty^{2m_1} \sigma_C^{2 \sum_{v \in [m_1]} (|N(v)|-2)}.$$

Since $|E(H)| = \sum_{v \in [m_1]} |N(v)| = m_1 + m_2 - 1$, we get that

$$W^k(G) \leq d \left(\frac{\sigma_\infty}{\sigma_C} \right)^{2m_1} \sigma_C^{2(m_2-1)}.$$

Case II. In case the tree H does not satisfy property $\mathcal{P}$, we then add for each $v \in H \cap [m_1]$ with $|N(v)| = 1$ in H one extra edge $uv \in E(G)$ from G. This creates a graph H' that is not a tree, but it satisfies property $\mathcal{P}$.

Let

$$V = \{v \in H \cap [m_1] : |N(v)| = 1 \text{ in } H\}.$$

Then, for each $v \in V$ we have $|N(v)| = 2$ in H' and v belongs to a cycle in H'. In particular, we can remove v from H' and $H'' = H' \setminus \{v\}$ is still connected. Assume $v = m_1$, then we have

$$W^k(G) \leq d\left(\sum_{w_2 \neq \cdots \neq w_{m_2}} \sum_{t_1 \neq \cdots \neq t_{m_1-1}} \prod_{e=ij \in E(H'')} b^2_{w_i t_j}\right) \max_{i \neq l \in [d]} \sum_{j \in [d]} b^2_{ij} b^2_{lj}.$$

We deduce that

$$W^k(G) \leq d\tilde{\sigma}^2_\infty \sum_{w_2 \neq \cdots \neq w_{m_2}} \sum_{t_1 \neq \cdots \neq t_{m_1-1}} \prod_{e=ij \in E(H'')} b^2_{w_i t_j}.$$

By induction, we have

$$W^k(G) \leq d\tilde{\sigma}^{2|V|}_\infty \sum_{w_2 \neq \cdots \neq w_{m_2}} \sum_{t_1 \neq \cdots \neq t_{m_1-|V|}} \prod_{e=ij \in E(H \setminus V)} b^2_{w_i t_j}.$$

By assumption, $H \setminus V = H' \setminus V$ is a tree satisfying property $\mathcal{P}$. Therefore, case I implies that

$$W^k(G) \leq d\tilde{\sigma}^{2|V|}_\infty \left(\frac{\sigma_\infty}{\sigma_C}\right)^{2(m_1-|V|)} \sigma_C^{2(m_2-1)}.$$

By definition of β_∞, we have

$$W^k(G) \leq d\beta^{2|V|}_\infty \left(\frac{\sigma_\infty}{\sigma_C}\right)^{2m_1} \sigma_C^{2(m_2-1)}.$$

If $\beta_\infty \leq 1$, we choose $|V| = 0$, otherwise we choose $|V| = m_1$. A straightforward computation yields the bounds of Proposition 5.10 with factor d.

For the second bound, instead of fixing $u_1 = w_1 = z$, we fix $v_1 = t_1 = z$. Define the following first-time arrivals:

$$i_1(k) := \inf\{l : u_l = k\};\ k = 1, ..., m_2;$$
$$i_2(k) := \inf\{l : v_l = k\};\ k = 2, ..., m_1,$$

and let also $e^{(1)}_k = u_{i_1(k)} v_{i_1(k)-1}$ and $e^{(2)}_k = u_{i_2(k)} v_{i_2(k)}$. The same argument done before implies that these $m_1 + m_2 - 1$ edges are distinct, and the subgraph H generated by them is a spanning tree of G. We then repeat the proof as in the first bound, but now the first choice of vertex v_1 will contribute with a factor of n. □

5.2.4 Proof of Theorem 5.2

To get the correct parameters for the Schatten norm, we must improve Proposition 5.10 and the bound on $W(s)$. The main proposition of this subsection is the following.

Proposition 5.11. *For any shape $s \in \mathcal{S}$, if $\beta_p \leq 1$, we have*

$$W(s) \leq d\sigma_*^{2p}\left\{\frac{\sigma_p}{\sigma_*\sigma_C}\right\}^{2m_1(s)}\left\{\frac{\sigma_C}{\sigma_*}\right\}^{2(m_2(s)-1)}.$$

Otherwise $\beta_p > 1$ and

$$W(s) \leq d\sigma_*^{2p}\left\{\frac{\overline{\sigma}_p}{\sigma_*^2}\right\}^{2m_1(s)}\left\{\frac{\sigma_C}{\sigma_*}\right\}^{2(m_2(s)-1)}.$$

As soon as Proposition 5.11 is available, the proof of Theorem 5.2 follows similarly as the proof of Theorem 5.1 and the bound for the diagonal in Theorem 5.8.

Proposition 5.11 follows the same argument shown in [74]. On the other hand, we did not try to optimize the argument to our setting. Instead, we prefer to prove it directly.

We start with the reduction to tree argument done in [74, Lemma 2.9] for $W^k(G)$. In this case, however, we want to keep track of the exponents for each right leaf that appears in the final reduction. We hence present the proof for completeness.

Lemma 5.12. *Let G be a graph generated by a shape $s \in \mathcal{S}$ and $k_e \geq 2$ for each $e \in E(G)$. Then, there exist $k'_2, ..., k'_{m_1+m_2-1} \geq 2$ such that $\sum_i k'_i = \sum_e k_e$ and*

$$W^k(G) \leq \max_{T \in \mathrm{span}(G)} W^{k'}(T),$$

where $\mathrm{span}(G)$ is the set of spanning trees of G. Moreover, the maximum can be taken such that whenever T has a right leaf $v \in [m_1]$ with unique edge $e = uv \in T$ we have $k_e \geq 4$.

Proof. If G is a tree, the equality is rather trivial, so suppose G is not a tree. Let $r \in [m_1]$ be a right vertex in a cycle in G. In particular, there exist two distinct edges $e_1 = l_1 r$ and $e_2 = l_2 r$ such that $G_s = (V(G), E(G) \setminus \{e_s\})$ is still connected for $s = 1, 2$. Let $\bar{k} = k_{e_1} + k_{e_2}$. Then

$$W^k(G) = \sum_{w_1 \neq \cdots \neq w_{m_2}} \sum_{t_1 \neq \cdots \neq t_{m_1}} \prod_{s=1,2} \left(b^{\bar{k}}_{w_{l_s} t_r} \prod_{e=ij \neq e_1, e_2} b^{k_e}_{w_i t_j} \right)^{k_{e_s}/\bar{k}}.$$

Holder's Inequality implies that

$$W^k(G) \leq \max_{s=1,2} \sum_{w_1 \neq \cdots \neq w_{m_2}} \sum_{t_1 \neq \cdots \neq t_{m_1}} b^{\bar{k}}_{w_{l_s} t_r} \prod_{e=ij \neq e_1, e_2} b^{k_e}_{w_i t_j} = \max_{s=1,2} W^{k'_s}(G_s).$$

Notice that G_s runs over all vertices of G, $|E(G_s)| = |E(G)| - 1$ and G_i is still connected. Moreover, the neighborhood of $v \neq r$ is preserved and so are the weights for all $v \in [m_1]$,

namely,

$$(k'_s)_v := \sum_{u \in N(v,G_s)} (k'_s)_{(uv)} = \sum_{u \in N(v,G)} k_{(uv)} = k_v \geq 4,$$

where $N(v,G)$ denotes the neighborhood of v in G, and the last inequality follows as v has at least two neighbors in G. The result follows by induction (see [74, Lemma 2.9]). □

Let $\mathcal{T}_{m_1,m_2}$ be the set of bipartite trees over $[m_2] \sqcup [m_1]$. By Lemma 5.12, we can assume that $G \in \mathcal{T}_{m_1,m_2}$. In [74, Lemma 2.10], the authors developed a method to prune leaves of G iteratively. In our case, however, we will prune the right vertices. To keep the notation clean, let

$$W(G) = \sum_{w \in [d]_{\neq}^{m_2}} \sum_{t \in [n]_{\neq}^{m_1}} \prod_{e=ij \in E(G)} b^{(e)}_{w_i t_j},$$

where $(b^{(e)})_{e \in E(G)}$ is a family of $d \times n$ matrices and

$$[m]_{\neq}^{I} := \{w \in [m]^I : w_k \neq w_l \,, \forall k \neq l \in I\}.$$

We can easily recover $W^k(G)$ by setting $b^{(e)}_{wt} = b^{k_e}_{wt}$.

We have the analog of Lemma 2.10 in [74]. Let $\mathcal{L}(G)$ be the set of leaves of G, and for each $v \in \mathcal{L}(G) \cap [m_1]$, let u_v be its only neighbor.

Lemma 5.13. *For any $G \in \mathcal{T}_{m_1,m_2}$ and $p_v \geq 1$ such that*

$$\sum_{v \in [m_1]} \frac{1}{p_v} = 1,$$

we have

$$W(G) \leq \prod_{v \in \mathcal{L}(G) \cap [m_1]} \left\{ \sum_{i \in [d]} \left(\sum_{j \in [n]} b^{(u_v v)}_{ij} \right)^{p_v} \right\}^{\frac{1}{p_v}} \times \prod_{\substack{v \in \mathcal{L}(G)^c \cap [m_1] \\ u \in N(v) \cap \mathcal{L}(G)^c}} \left\{ \sum_{i \in [d]} \left[\sum_{j \in [n]} b^{(uv)}_{ij} \prod_{a \in N(v) \setminus \{u\}} \left(\sum_{l \neq i} b^{(av)}_{lj} \right) \right]^{p_v} \right\}^{\frac{1}{p_v} \frac{1}{\alpha_{uv}}},$$

where α_{uv} satisfies

$$\sum_{u \in N(v) \cap \mathcal{L}(G)^c} \frac{1}{\alpha_{uv}} = 1,$$

for all $v \in \mathcal{L}(G)^c \cap [m_1]$.

Before proving this result, we will use the following easier version. Let $u = u(v)$ be the choice $u \in N(v) \cap \mathcal{L}(G)^c$ that maximizes the second term in the bound, then the following holds.

Corollary 5.14. *For any $G \in \mathcal{T}_{m_1,m_2}$ and $p_v \geq 1$ such that*

$$\sum_{v\in[m_1]} \frac{1}{p_v} = 1,$$

we have

$$W(G) \leq \prod_{v\in\mathcal{L}(G)\cap[m_1]} \left\{ \sum_{i\in[d]} \left(\sum_{j\in[n]} b_{ij}^{(u_v v)} \right)^{p_v} \right\}^{\frac{1}{p_v}} \times \prod_{v\in\mathcal{L}(G)^c\cap[m_1]} \left\{ \sum_{i\in[d]} \left[\sum_{j\in[n]} b_{ij}^{(uv)} \prod_{a\in N(v)\setminus\{u\}} \left(\sum_{l\neq i} b_{lj}^{(av)} \right) \right]^{p_v} \right\}^{\frac{1}{p_v}}.$$

Proof of Lemma 5.13. The proof follows by induction. If $m_1 = 1$, then it is easy to check that $p_v = 1$ and

$$W(G) \leq \sum_{i\in[d]} \sum_{j\in[n]} \prod_{a\in N(1')\setminus\{1\}} b_{ij}^{(11')} \left(\sum_{l\neq i} b_{lj}^{(a1')} \right).$$

Therefore, if $|N(1')| > 1$, $W(G)$ has the second form on the bound shown in the lemma. Otherwise, $|N(v)| = 1$, and the bound has the first form. Hence, we can assume that $m_1 > 1$.

Let $\mathcal{L} = \mathcal{L}(G)$ and $v_1, v_2 \in [m_1]$ be such that

$$d(v_1, v_2) = \max_{r,r'\in[m_1]} d(r, r'),$$

where the distance is the graph distance. Therefore, both v_1 and v_2 have only one neighbor $u_1 \in N(v_1)\cap\mathcal{L}^c$ and $u_2 \in N(v_2)\cap\mathcal{L}^c$. This follows the argument shown in Proposition 5.10. Let then H be the subgraph generated by removing v_1, v_2 and all leaves $(N(v_1)\cup N(v_2))\cap\mathcal{L}$. Denote $H = (I \sqcup J, E(H))$. Then we have

$$W(G) \leq \sum_{w\in[d]^I_{\neq}} \sum_{t\in[n]^J_{\neq}} \left[\sum_{j\in[n]} b_{w_{u_1} j}^{(u_1 v_1)} \prod_{a\in N(v_1)\setminus\{u_1\}} \left(\sum_{l\neq w_{u_1}} b_{lj}^{(av_1)} \right) \right] \times \left[\sum_{j\in[n]} b_{w_{u_2} j}^{(u_2 v_2)} \prod_{a\in N(v_2)\setminus\{u_2\}} \left(\sum_{l\neq w_{u_2}} b_{lj}^{(av_2)} \right) \right] \prod_{e=ab\in E(H)} b_{w_a t_b}^{(ab)},$$

where we define

$$\prod_{a\in N(v)\setminus\{u\}} \left(\sum_{l\neq w_u} b_{lj}^{(av)} \right) = 1,$$

if $N(v) \setminus \{u\} = \varnothing$. Using Holder's Inequality, we can estimate

$$W(G) \leq \left\{ \sum_{w \in [d]^I_{\neq}} \sum_{t \in [n]^J_{\neq}} \left[\sum_{j \in [n]} b^{(u_1 v_1)}_{w_{u_1} j} \prod_{a \in N(v_1) \setminus \{u_1\}} \left(\sum_{l \neq w_{u_1}} b^{(a v_1)}_{lj} \right) \right]^{1 + \frac{p_{v_1}}{p_{v_2}}} \prod_{e = ab \in E(H)} b^{(ab)}_{w_a t_b} \right\}^{\frac{p_{v_2}}{p_{v_1} + p_{v_2}}} \times$$
$$\left\{ \sum_{w \in [d]^I_{\neq}} \sum_{t \in [n]^J_{\neq}} \left[\sum_{j \in [n]} b^{(u_2 v_2)}_{w_{u_2} j} \prod_{a \in N(v_2) \setminus \{u_2\}} \left(\sum_{l \neq w_{u_2}} b^{(a v_2)}_{lj} \right) \right]^{1 + \frac{p_{v_2}}{p_{v_1}}} \prod_{e = ab \in E(H)} b^{(ab)}_{w_a t_b} \right\}^{\frac{p_{v_1}}{p_{v_1} + p_{v_2}}}. \tag{5.6}$$

Note that this inequality preserves the number of summations of right and left vertices and homogeneity. Note also that if $v \in J$, the neighbors of v in H and G are the same.

The induction will be based on inequality (5.6). Suppose, for some $r > 1$ that

$$W(G) \leq \prod_{h=1}^{H} \left\{ \sum_{w \in [d]^{I_h}_{\neq}} \sum_{t \in [n]^{J_h}_{\neq}} \left[\sum_{j \in [n]} b^{(u_h v_h)}_{w_{u_h} j} \prod_{a \in N(v_h) \setminus \{u_h\}} \left(\sum_{l \neq w_{u_h}} b^{(a v_h)}_{lj} \right) \right]^{q_h} \prod_{e = ab \in E(G_h)} b^{(ab)}_{w_a t_b} \right\}^{\frac{1}{\alpha_h}},$$

where $H < \infty$, $N(v)$ is the neighbor of v in G,

1. For every h, $u_h \in I_h$, $v_h \notin J_h$ and $a v_h \notin E(G_h)$ for every $a \in N(v_h) \setminus \{u_h\}$;

2. For every h, G_h is a tree over $I_h \sqcup J_h$ and $|J_h| = r$;

3. The inequality is 1-homogeneous in all the variables $b^{(e)}$ and it preserves the number of left and right summations;

4. The exponents q_h satisfies

$$q_h = \sum_{v \in [m_1] \setminus J_h} \frac{p_{v_h}}{p_v},$$

and $\alpha_h \geq 1$.

We aim to show that if this holds for $r > 1$, so does it for $r - 1$. Indeed, fix one of the terms

$$T_h := \left\{ \sum_{w \in [d]^{I_h}_{\neq}} \sum_{t \in [n]^{J_h}_{\neq}} \left[\sum_{j \in [n]} b^{(u_h v_h)}_{w_{u_h} j} \prod_{a \in N(v_h) \setminus \{u_h\}} \left(\sum_{l \neq w_{u_h}} b^{(a v_h)}_{lj} \right) \right]^{q_h} \prod_{e = ab \in E(G_h)} b^{(ab)}_{w_a t_b} \right\}.$$

Since G_h is a tree and $r > 1$, there exists r_h such that u_h is not a leaf of r_h, and r_h has only one neighbor l_h such that $N(l_h) > 1$ in G_h. Let then H_h be the subgraph (a tree) of

G_h where we remove r_h and all of its leaves and let $H_h = (I_{h'} \sqcup J_{h'}, E(H_h))$, then

$$T_h \leq \sum_{w\in[d]_{\neq}^{I_{h'}}} \sum_{t\in[n]_{\neq}^{J_{h'}}} \left[\sum_{j\in[n]} b_{w_{u_h}j}^{(u_h v_h)} \prod_{a\in N(v_h)\setminus\{u_h\}} \left(\sum_{l\neq w_{u_h}} b_{lj}^{(av_h)}\right)\right]^{q_h} \times \left[\sum_{j\in[n]} b_{w_{l_h}j}^{(l_h r_h)} \prod_{a\in N(r_h)\setminus\{l_h\}} \left(\sum_{l\neq w_{l_h}} b_{lj}^{(ar_h)}\right)\right] \prod_{e=ab\in E(G_h)} b_{w_a t_b}^{(ab)}.$$

We can thus estimate by Holder's Inequality that

$$T_h \leq \left\{\sum_{w\in[d]_{\neq}^{I_{h'}}} \sum_{t\in[n]_{\neq}^{J_{h'}}} \left[\sum_{j\in[n]} b_{w_{u_h}j}^{(u_h v_h)} \prod_{a\in N(v_h)\setminus\{u_h\}} \left(\sum_{l\neq w_{u_h}} b_{lj}^{(av_h)}\right)\right]^{q_h'} \prod_{e=ab\in E(G_h)} b_{w_a t_b}^{(ab)}\right\}^{1/\alpha_h} \times \left\{\sum_{w\in[d]_{\neq}^{I_{h'}}} \sum_{t\in[n]_{\neq}^{J_{h'}}} \left[\sum_{j\in[n]} b_{w_{l_h}j}^{(l_h r_h)} \prod_{a\in N(r_h)\setminus\{l_h\}} \left(\sum_{l\neq w_{l_h}} b_{lj}^{(ar_h)}\right)\right]^{q} \prod_{e=ab\in E(G_h)} b_{w_a t_b}^{(ab)}\right\}^{1/q},$$

where q_h'/q_h and q are conjugate exponents. Again, the inequality is 1-homogeneous in all the variables it involves, and it preserves the number of summations. Moreover, we can set

$$q_h' = \sum_{v\in[m_1]\setminus J_{h'}} \frac{p_{v_h}}{p_v}$$
$$q = \sum_{v\in[m_1]\setminus J_{h'}} \frac{p_{l_h}}{p_v},$$

and it is easy to check that q_h'/q_h and q are conjugate exponents. Note that each new term has the same form as in the induction step with $|J_{h'}| = r - 1$. Therefore, the induction is proved.

The previous argument also shows that the induction holds for $r = 0$. Since the choice of $u \in \mathcal{L}^c \cap N(v)$ is arbitrary for each v, we deduce

$$W(G) \leq \prod_{h=1}^{H} \left\{\sum_{w\in[d]_{\neq}^{I_h}} \sum_{t\in[n]_{\neq}^{J_h}} \left[\sum_{j\in[n]} b_{w_{u_h}j}^{(u_h v_h)} \prod_{a\in N(v_h)\setminus\{u_h\}} \left(\sum_{l\neq w_{u_h}} b_{lj}^{(av_h)}\right)\right]^{q_h} \prod_{e=ab\in E(G_h)} b_{w_a t_b}^{(ab)}\right\}^{\frac{1}{\alpha_h}}$$
$$\leq \prod_{e=uv\in E(G):u\in\mathcal{L}^c} \left\{\sum_{i\in[d]} \left[\sum_{j\in[n]} b_{ij}^{(e)} \prod_{a\in N(v)\setminus\{u\}} \left(\sum_{l\neq i} b_{lj}^{(av)}\right)\right]^{p_v}\right\}^{\frac{1}{\alpha_e}}.$$

The conclusion of the lemma follows by the renormalization $\alpha_{uv} \leftarrow p_v \alpha_{uv}$ and splitting the product over $v \in \mathcal{L}$ and $v \notin \mathcal{L}$. □

Now we can prove Proposition 5.11.

Proof of Proposition 5.11. Let $|k| = \sum_v k_v = 2p$ and $\mathcal{L} = \mathcal{L}(G) \cap [m_1]$. By Lemma 5.12

and Corollary 5.14 with $p_v = |k|/k_v$, we get

$$W^k(G) \leq W^{k'}(T) \leq \prod_{v\in\mathcal{L}} \left\{ \sum_{i\in[d]} \left(\sum_{j\in[n]} b_{ij}^{k_v} \right)^{\frac{|k|}{k_v}} \right\}^{\frac{k_v}{|k|}} \times \prod_{v\in\mathcal{L}^c} \left\{ \sum_{i\in[d]} \left[\sum_{j\in[n]} b_{ij}^{k_{uv}} \prod_{a\in N(v)\setminus\{u\}} \left(\sum_{l\neq i} b_{lj}^{k_{av}} \right) \right]^{\frac{|k|}{k_v}} \right\}^{\frac{k_v}{|k|}},$$

where T is the spanning tree of G that maximizes $W^{k'}(T')$ in Lemma 5.12. Since $k_v \geq 4$ and $b_{ij}^{k_{uv}} \leq b_{ij}^2 \sigma_*^{k_{uv}-2}$, we get

$$W^k(G) \leq \sigma_*^{|k|-4|\mathcal{L}|-2\sum_{v\in\mathcal{L}^c}|N(v)|} \prod_{v\in\mathcal{L}} \left\{ \sum_{i\in[d]} \left(\sum_{j\in[n]} b_{ij}^4 \right)^{\frac{|k|}{k_v}} \right\}^{\frac{k_v}{|k|}} \times \prod_{v\in\mathcal{L}^c} \left\{ \sum_{i\in[d]} \left[\sum_{j\in[n]} b_{ij}^2 \prod_{a\in N(v)\setminus\{u\}} \left(\sum_{l\neq i} b_{lj}^2 \right) \right]^{\frac{|k|}{k_v}} \right\}^{\frac{k_v}{|k|}}.$$

As T is a spanning tree, we have

$$\sum_{v\in\mathcal{L}^c} |N(v)| + |\mathcal{L}| = m_2 + m_1 - 1;$$
$$\sum_{v\in\mathcal{L}^c} |N(v)| - 2|\mathcal{L}^c| = m_2 - m_1 - 1 + |\mathcal{L}|.$$

Moreover, we can remove σ_C from each term in the second product to get that

$$W^k(G) \leq \sigma_*^{|k|-2(m_1+m_2-1)-2|\mathcal{L}|} \sigma_C^{2(m_1+m_2-1)+2|\mathcal{L}|} \prod_{v\in\mathcal{L}} \left\{ \sum_{i\in[d]} \left(\sum_{j\in[n]} b_{ij}^4 \right)^{\frac{|k|}{k_v}} \right\}^{\frac{k_v}{|k|}} \times \prod_{v\in\mathcal{L}^c} \left\{ \sum_{i\in[d]} \left[\sum_{j\in[n]} b_{ij}^2 \left(\sum_{l\neq i} b_{lj}^2 \right) \right]^{\frac{|k|}{k_v}} \right\}^{\frac{k_v}{|k|}}.$$

Finally, the inequality of the norms in $\mathbb{R}^d$ implies that

$$\|\cdot\|_{\frac{|k|}{k_v}} \leq d^{\frac{k_v-4}{|k|}} \|\cdot\|_{\frac{|k|}{4}},$$

so we deduce

$$W^k(G) \leq d\sigma_*^{|k|-2(m_1+m_2-1)-2|\mathcal{L}|} \sigma_C^{2(m_1+m_2-1)+2|\mathcal{L}|} \sigma_p^{2|\mathcal{L}|} \bar{\sigma}_p^{2|\mathcal{L}^c|}.$$

The proof of Proposition 5.11 follows by a straightforward computation and the fact that $0 \leq |\mathcal{L}| \leq m_1$. □

Remark 5.15. Note that we rather proved Theorem 5.2 with a parameter σ'_p instead of σ_p, where σ'_p only takes $l \neq i$, that is,

$$\sigma'_p = \left\{ \sum_{i\in[d]} \left[\sum_{j\in[n]} \sum_{l\neq i} b_{ij}^2 b_{lj}^2 \right]^{p/2} \right\}^{1/p}.$$

This minor change is only essential for cases where the contribution of a column X_j appears only in the diagonal part, that is, when $X_j = b_{ij}e_i$ for some i.

5.3 Examples

To compare our estimates to the known bounds in the literature, we start by recalling the previous known results in [33, 18, 31].

Theorem 5.16 (Theorem 2.1 in [33]). *Let X be a $d\times n$ Gaussian matrix with independent entries such that $X_{ij} = b_{ij}g_{ij}$ where $\{g_{ij} : (i,j) \in [d]\times[n]\}$ are i.i.d standard Gaussian r.v. Then we have*

$$\begin{aligned}&\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \\ &\leq (1+\varepsilon)\Big\{2\sigma_R\sigma_C + \sigma_C^2 + C(\varepsilon)(\sigma_C\sigma_* + \sigma_R\sigma_*)\sqrt{\log(n\wedge d)} + C^2(\varepsilon)\sigma_*^2\log(n\wedge d)\Big\}.\end{aligned}$$

Here, σ_R denotes the maximum norm of the rows

$$\sigma_R^2 := \max_{i\in[d]} \sum_{j\in[n]} b_{ij}^2.$$

Notice that Theorem 5.1 strictly improves Theorem 5.16.

Theorem 5.17 (Theorem 3.17 in [31]). *Let X be a $d\times n$ Gaussian matrix with independent entries such that $X_{ij} = b_{ij}g_{ij}$ where $\{g_{ij} : (i,j) \in [d]\times[n]\}$ are i.i.d standard Gaussian r.v. Let $\varepsilon \in (0,1/2)$. Then*

$$\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \leq 2(1+\varepsilon)\sigma_\infty + \frac{C}{\varepsilon^3}\left(\overline{\sigma}_\infty + \sigma_C^2\right)\log^3(nd).$$

Theorem 5.18 (Theorem 3.12 in [18],Theorem 3.16 in [31]). *Let X be a $d\times n$ Gaussian matrix with independent entries such that $X_{ij} = b_{ij}g_{ij}$ where $\{g_{ij} : (i,j) \in [d]\times[n]\}$ are i.i.d standard Gaussian r.v. Then*

$$\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \leq 2\sigma_\infty + \sigma_C^2 + C\sigma_*^{1/2}\left(\sigma_C^{3/2} + \sigma_R^{3/2}\right)\log^{3/2}(nd).$$

In particular, let $\alpha \in [0,1]$ be such that

$$\alpha := \frac{\sigma_*\sigma_R}{\sigma_*^{1/2}\sigma_R^{3/2}} = \left(\frac{\sigma_*}{\sigma_R}\right)^{1/2}.$$

Since $\sigma_* = \sigma_R$ only when the rows R_i of X are of the form $R_i = b_{ij_i} e_{j_i}$, where $(e_j)_{j\in[n]}$ is the canonical basis and $j_i \in [n]$, we see that Theorem 5.1 strictly improves Theorem 5.18 when $\beta_\infty \le 1$. Moreover, even in the simplest case $b_{ij} = 1$ for all i, j, Theorems 5.17 and 5.18 cannot capture the right asymptotics of $XX^T - \mathbb{E}\,XX^T$. Indeed, Theorem 5.17 does not have the right leading term, and Theorem 5.18 yields the bound

$$\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \le 2\sqrt{nd} + d + C\left(n^{3/4} + d^{3/4}\right)\log^{3/2}(nd).$$

Here, we note that when $d \le \sqrt{n}$, the error factor becomes dominant, and the bound turns out to be

$$\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \lesssim n^{3/4}\log^{3/2}(nd),$$

which differs from the correct order $\sqrt{nd} + d$ from [13].

Now we discuss various examples and present how Theorem 5.1 improves upon Theorems 5.16, 5.17, and 5.18.

Proposition 5.19 (Case $b_{ij} = a_i b_j$)**.** *Let X be a $d \times n$ Gaussian matrix with independent entries such that $X_{ij} = a_i b_j g_{ij}$ where $\{g_{ij} : (i,j) \in [d]\times[n]\}$ are i.i.d standard Gaussian r.v. Then*

$$\begin{aligned}\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \lesssim &\|b\|_4^2\|a\|_2\|a\|_\infty + \|a\|_2^2\|b\|_\infty^2 + \\ &\left[\|a\|_2\|a\|_\infty\|b\|_\infty^2 + \|b\|_4^2\|a\|_\infty^2\right]\sqrt{\log(n\wedge d)} + \\ &\|a\|_\infty^2\|b\|_\infty^2\log(n\wedge d).\end{aligned}$$

Proof. Note that

- $\sigma_C = \|a\|_2\|b\|_\infty$;
- $\sigma_R = \|a\|_\infty\|b\|_2$;
- $\sigma_* = \|a\|_\infty\|b\|_\infty$;
- $\tilde\sigma_\infty \le \|b\|_4^2\|a\|_\infty^2$;
- $\bar\sigma_\infty = \|b\|_4^2\|a\|_\infty^2$;
- $\sigma_\infty \le \|b\|_4^2\|a\|_2\|a\|_\infty$.

We observe that

$$\frac{\tilde\sigma_\infty\sigma_C}{\sigma_*} \le \|b\|_4^2\|a\|_2\|a\|_\infty.$$

Therefore, Theorem 5.1 implies that

$$\begin{aligned}\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \lesssim &\|b\|_4^2\|a\|_2\|a\|_\infty + \|a\|_2^2\|b\|_\infty^2 + \\ &\left[\|a\|_2\|a\|_\infty\|b\|_\infty^2 + \|b\|_4^2\|a\|_\infty^2\right]\sqrt{\log(n\wedge d)} + \\ &\|a\|_\infty^2\|b\|_\infty^2\log(n\wedge d).\end{aligned}$$

□

The result in [33] yields

$$\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \lesssim \|b\|_2\|b\|_\infty\|a\|_2\|a\|_\infty + \|a\|_2^2\|b\|_\infty^2 + \left[\|a\|_2\|a\|_\infty\|b\|_\infty^2 + \|b\|_2\|b\|_\infty\|a\|_\infty^2\right]\sqrt{\log(n\wedge d)} + \|a\|_\infty^2\|b\|_\infty^2\log(n\wedge d).$$

Theorem 5.17 yields

$$\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \lesssim \|b\|_4^2\|a\|_2\|a\|_\infty + \left(\|b\|_4^2\|a\|_\infty^2 + \|a\|_2^2\|b\|_\infty^2\right)\log^3(nd)$$

Finally, Theorem 5.18 gives

$$\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \lesssim \|b\|_4^2\|a\|_2\|a\|_\infty + \|a\|_2^2\|b\|_\infty^2 + \|a\|_\infty^{1/2}\|b\|_\infty^{1/2}\left(\|a\|_2^{3/2}\|b\|_\infty^{3/2} + \|a\|_\infty^{3/2}\|b\|_2^{3/2}\right)\log^{3/2}(nd).$$

In this case, Proposition 5.19 strictly improves the previous results.

Another example is where all columns $B_j := Be_j$ have approximately the same norm, where $B = (b_{ij})$.

Proposition 5.20. *Let X be a $d\times n$ Gaussian matrix with independent entries such that $X_{ij} = b_{ij}g_{ij}$ where $\{g_{ij} : (i,j)\in[d]\times[n]\}$ are i.i.d standard Gaussian r.v and $B = (b_{ij})$ is a deterministic matrix. Let $\varepsilon\in(0,1/2)$ and $B_j = Be_j$. Assume that there exists $K\geq 1$ such that*

$$\frac{1}{K}\|B_k\|_2 \leq \|B_j\|_2 \leq K\|B_k\|_2,$$

for all $j,k\in[n]$. Then

$$\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \leq (1+\varepsilon)\left\{2K\sigma_\infty + \sigma_C^2 + C(\varepsilon)K(\sigma_C\sigma_* + \bar\sigma_\infty)\sqrt{\log(n\wedge d)} + C^2(\varepsilon)\sigma_*^2\log(n\wedge d)\right\}.$$

Proof. It is easy to compute

$$\beta_\infty \leq K.$$

Hence

$$\frac{\overline{\sigma}_\infty\sigma_C}{\sigma_*} \leq K\sigma_\infty.$$

By Theorem 5.1, we have

$$\begin{aligned}&\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \\ &\leq (1+\varepsilon)\Big\{2K\sigma_\infty + \sigma_C^2 + C(\varepsilon)K(\sigma_C\sigma_* + \bar{\sigma}_\infty)\sqrt{\log(n\wedge d)} + C^2(\varepsilon)\sigma_*^2\log(n\wedge d)\Big\}.\end{aligned}$$

□

Here, the previous known results only show the leading term with $\sigma_C\sigma_R$ in [33] and larger error factors in [18, 31].

To illustrate the unified approach from Theorem 5.1, we present another proof of the sharp result in [33, Theorem 3.13].

Example 5.21. Let X be a Gaussian matrix with i.i.d rows, that is, $b_{ij} = b_j$. In this case, we have

- $\sigma_C = \sqrt{d}\|b\|_\infty$;
- $\sigma_R = \|b\|_2$;
- $\sigma_* = \|b\|_\infty$;
- $\sigma_\infty = \sqrt{d-1}\|b\|_4^2$;
- $\tilde{\sigma}_\infty = \bar{\sigma}_\infty = \|b\|_4^2$.

In particular,

$$\frac{\tilde{\sigma}_\infty\sigma_C}{\sigma_*} = \sqrt{d}\|b\|_4^2.$$

Hence, Theorem 5.1 implies that

$$\begin{aligned}\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \leq& 2\sqrt{d}\|b\|_4^2 + d\|b\|_\infty^2 + \\ & C\left[\sqrt{d}\|b\|_\infty^2 + \|b\|_4^2\right]\sqrt{\log(n\wedge d)} + \\ & C\|b\|_\infty^2\log(n\wedge d).\end{aligned}$$

In this case, the error factor is smaller than the leading one, hence

$$\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\| \leq \left(2\sqrt{d}\|b\|_4^2 + d\|b\|_\infty^2\right)(1+o(1)),$$

where $o(1)$ is a quantity that goes to 0 when $n, d \to \infty$.

5.4 Lower bounds

For the sake of completeness, we present lower bounds on the operator norm and the Schatten norm. The former was already proven in a general version in [31, lemma 9.11].

We first begin the lower bounds for the operator norm.

Lemma 5.22. *Let X be a centered Gaussian matrix with independent entries such that $X_{ij} = b_{ij}g_{ij}$ where g_{ij} are i.i.d standard Gaussian random variables and $b_{ij} \geq 0$. Then*

$$\left(\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\|^2\right)^{1/2} \geq \max\{\sigma_\infty, 0.49\sigma_C^2\}.$$

Proof. Let $S_j = X_jX_j^T - \mathbb{E}\,X_jX_j^T$, then

$$\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\|^2 = \mathbb{E}\left\|\sum_{j\in[n]} S_j\right\|^2.$$

By submultiplicativity of the norm and Jensen's inequality, we have

$$\mathbb{E}\left\|\sum_{j\in[n]} S_j\right\|^2 \geq \mathbb{E}\left\|\left(\sum_{j\in[n]} S_j\right)^2\right\| \geq \left\|\mathbb{E}\left(\sum_{j\in[n]} S_j\right)^2\right\|.$$

The matrix on the right-hand side can be expanded as

$$\mathbb{E}\left(\sum_{j\in[n]} S_j\right)^2 = \sum_{j\in[n]} \mathbb{E}\,S_j^2 = \sum_{j\in[n]} \mathbb{E}\,X_jX_j^TX_jX_j^T - (\mathbb{E}\,X_jX_j^T)^2, \tag{5.7}$$

as S_j are centered and independent. Note that

$$\begin{aligned}(\sum_{j\in[n]} \mathbb{E}\,X_jX_j^TX_jX_j^T)_{il} &= \sum_{j\in[n]} \mathbb{E}\langle X_jX_j^TX_jX_j^Te_l, e_i\rangle \\ &= \sum_{j\in[n]} \mathbb{E}\langle X_j, X_j\rangle\langle X_j, e_l\rangle\langle X_j, e_i\rangle \\ &= \mathbf{1}_{i=l}\sum_{j\in[n]} \mathbb{E}\left(X_{ij}^4 + X_{ij}^2\sum_{k\neq i} X_{kj}^2\right),\end{aligned}$$

by the symmetry of the Gaussian and independence. Therefore, the right-hand side of (5.7) is a diagonal matrix D and

$$D_{ii} = \sum_{j\in[n]}\left(2b_{ij}^4 + b_{ij}^2\sum_{l\neq i} b_{lj}^2\right) \geq \sum_{j\in[n]}\sum_{l\in[d]} b_{il}^2b_{lj}^2.$$

Therefore

$$\left\|\sum_{j\in[n]} \mathbb{E}\,X_jX_j^TX_jX_j^T - (\mathbb{E}\,X_jX_j^T)^2\right\| \geq \sigma_\infty^2,$$

and

$$(\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\|^2)^{1/2} \geq \sigma_\infty. \tag{5.8}$$

On the other hand, Jensen's inequality implies that

$$(\mathbb{E}\left\|XX^T - \mathbb{E}\,XX^T\right\|^2)^{1/2} \geq \mathbb{E}\left\|\sum_{j\in[n]} S_j\right\|.$$

Let $k \in [n]$. Conditioning on the variables indexed by $J = [n] \setminus \{k\}$ and using Jensen's inequality, we get that

$$\mathbb{E}\left\|\sum_{j\in[n]} S_j\right\| \geq \mathbb{E}\|S_k\|.$$

By the inverse triangle inequality, we get

$$\mathbb{E}\|S_k\| \geq \mathbb{E}\|X_k\|^2 - \left\|\mathbb{E}\,X_k X_k^T\right\| = \|B_k\|_2^2 - \|B_k\|_\infty^2 \geq \|B_k\|_2^2 - \sigma_*^2,$$

where $B_k = (b_{ik})_{i\in[d]}$. Moreover, we have the trivial bound $\|S\| \geq |S_{ii}|$, hence

$$\mathbb{E}\|S_k\| \geq \mathbb{E}\,|X_{ik}^2 - \mathbb{E}\,X_{ik}^2| = b_{ik}^2\,\mathbb{E}\,|g^2 - 1|,$$

where $g \sim N(0,1)$. Since, $\mathbb{E}\,|g^2 - 1| = 2\sqrt{\frac{2}{\pi e}} := t$, we get that

$$\max_{k\in[n]} \mathbb{E}\|S_k\| \geq \sigma_C^2 - \sigma_*^2;$$
$$\max_{k\in[n]} \mathbb{E}\|S_k\| \geq t\sigma_*^2.$$

By summing both inequalities appropriately, we conclude that

$$\max_{k\in[n]} \mathbb{E}\|S_k\| \geq \frac{t}{1+t}\sigma_C^2 \geq 0.49\sigma_C^2. \tag{5.9}$$

We deduce the result with Inequality (5.8). □

We also prove a lower bound for the pth moment of the Schatten norm.

Proposition 5.23. *Let X be a centered Gaussian matrix with independent entries such that $X_{ij} = b_{ij}g_{ij}$ where g_{ij} are i.i.d standard Gaussian random variables and $b_{ij} \geq 0$. For any even integer $p \geq 2$, we have*

$$\left(\mathbb{E}\,\mathrm{Tr}(XX^T - \mathbb{E}\,XX^T)^p\right)^{1/p} \gtrsim \sigma_p + \sigma_C^2 + \sqrt{p}\bar{\sigma}_p + pb_p^2.$$

Proof. By Lemma 5.5, the joint moments of g and $g^2 - 1$ are always positive, thus it follows that

$$\left(\mathbb{E}\,\mathrm{Tr}(XX^T - \mathbb{E}\,XX^T)^p\right)^{1/p} \geq \left(\mathbb{E}\,\mathrm{Tr}(\mathrm{Diag}(XX^T) - \mathbb{E}\,XX^T)^p\right)^{1/p} \gtrsim \sqrt{p}\bar{\sigma}_p + pb_p^2,$$

where the last inequality follows by Theorem 5.8. For the leading factor, note that the Schatten norm is always lower bounded by the mixed $l_2(l_p)$ norm (see Lemma 2.12 in

[74]), then Jensen's Inequality implies that

$$\left(\mathbb{E}\,\mathrm{Tr}(XX^T - \mathbb{E}\,XX^T)^p\right)^{1/p} \geq \left[\sum_{i\in[d]}\left(\sum_{l\in[d]}\mathbb{E}(XX^T - \mathbb{E}\,XX^T)_{il}^2\right)^{p/2}\right]^{1/p}.$$

The latter can be estimated as

$$\begin{aligned}\mathbb{E}(XX^T - \mathbb{E}\,XX^T)_{il}^2 &= \sum_{j\in[n]}\mathbb{E}(X_jX_j^T - \mathbb{E}\,X_jX_j)_{il}^2\\ &= \sum_{j\in[n]}\mathbb{E}(b_{ij}b_{lj}(g_{ij}g_{lj} - \mathbf{1}_{i=l}))^2\\ &\geq \sum_{j\in[n]} b_{ij}^2b_{lj}^2.\end{aligned}$$

Hence

$$\begin{aligned}\left(\mathbb{E}\,\mathrm{Tr}(XX^T - \mathbb{E}\,XX^T)^p\right)^{1/p} &\geq \left[\sum_{i\in[d]}\left(\sum_{j\in[n]}\sum_{l\in[d]} b_{ij}^2b_{lj}^2\right)^{p/2}\right]^{1/p}\\ &\geq \sigma_p.\end{aligned}$$

Finally, let $k \in [n]$, then

$$\left(\mathbb{E}\,\mathrm{Tr}(XX^T - \mathbb{E}\,XX^T)^p\right)^{1/p} \geq \left(\mathbb{E}\,\mathrm{Tr}(X_kX_k^T - \mathbb{E}\,X_kX_k^T)^p\right)^{1/p}.$$

Since p is even, we get

$$\left(\mathbb{E}\,\mathrm{Tr}(X_kX_k^T - \mathbb{E}\,X_kX_k^T)^p\right)^{1/p} \geq \mathbb{E}\left\|X_kX_k^T - \mathbb{E}\,X_kX_k^T\right\|$$

Therefore, recalling that $S_k = X_jX_j^T - \mathbb{E}\,X_jX_j^T$, Inequality (5.9) implies that

$$\left(\mathbb{E}\,\mathrm{Tr}(XX^T - \mathbb{E}\,XX^T)^p\right)^{1/p} \geq \max_{k\in[n]}\mathbb{E}\|S_k\| \geq 0.49\sigma_C^2,$$

and this finishes the proof. □

Bibliography

[1] M. Abért, Y. Glasner, and B. Virág. "The measurable Kesten theorem". In: *Ann. Probab.* 44.3 (2016), pp. 1601–1646.

[2] N. I. Akhiezer. *The classical moment problem and some related questions in analysis*. Translated by N. Kemmer. Hafner Publishing Co., New York, 1965, pp. x+253.

[3] N. Alon. "Eigenvalues and expanders". In: vol. 6. 2. Theory of computing (Singer Island, Fla., 1984). 1986, pp. 83–96.

[4] J. Alt, R. Ducatez, and A. Knowles. "Extremal eigenvalues of critical Erdős-Rényi graphs". In: *Ann. Probab.* 49.3 (2021), pp. 1347–1401.

[5] D. J. Altschuler et al. *On spectral outliers of inhomogeneous symmetric random matrices*. 2024. arXiv: 2401.07852 [math.PR].

[6] A. Ambainis, A. W. Harrow, and M. B. Hastings. "Random Tensor Theory: Extending Random Matrix Theory to Mixtures of Random Product States". In: *Communications in Mathematical Physics* 310.1 (2012), pp. 25–74.

[7] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*. Cambridge Studies in Advanced Mathematics 118. Cambridge University Press, 2010.

[8] G. W. Anderson and O. Zeitouni. "A CLT for a band matrix model". In: *Probability Theory and Related Fields* 134.2 (2006), pp. 283–338.

[9] G. Aubrun and I. Nechita. "Realigning random states". In: *J. Math. Phys.* 53.10 (2012), p. 102210.

[10] G. Aubrun and S. Szarek. *Alice and Bob meet Banach: The interface of asymptotic geometric analysis and quantum information theory*. Vol. 223. Mathematical Surveys and Monographs. American Mathematical Society, 2017.

[11] F. Augeri, A. Guionnet, and J. Husson. "Large deviations for the largest eigenvalue of sub-Gaussian matrices". In: *Comm. Math. Phys.* 383.2 (2021), pp. 997–1050.

[12] Z. D. Bai and Y. Q. Yin. "Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix". In: *Ann. Probab.* 16.4 (1988), pp. 1729–1741.

[13] Z.-D. Bai and Y.-Q. Yin. "Limit of the smallest eigenvalue of a large dimensional sample covariance matrix". In: *Advances In Statistics*. World Scientific, 2008, pp. 108–127.

[14] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Vol. 20. Springer Series in Statistics. Springer, 2010.

[15] J. Baik, G. Ben Arous, and S. Péché. "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices". In: *Ann. Probab.* 33.5 (2005), pp. 1643–1697.

[16] J. Baik and J. W. Silverstein. "Eigenvalues of large sample covariance matrices of spiked population models". In: *J. Multivariate Anal.* 97.6 (2006), pp. 1382–1408.

[17] A. S. Bandeira, M. T. Boedihardjo, and R. van Handel. "Matrix concentration inequalities and free probability". In: *Inventiones mathematicae* (2023), pp. 1–69.

[18] A. S. Bandeira, M. T. Boedihardjo, and R. van Handel. "Matrix concentration inequalities and free probability". In: *Invent. Math.* 234.1 (2023), pp. 419–487.

[19] A. S. Bandeira and R. van Handel. "Sharp nonasymptotic bounds on the norm of random matrices with independent entries". In: *Ann. Probab.* 44.4 (2016), pp. 2479–2506.

[20] F. Benaych-Georges, C. Bordenave, and A. Knowles. "Largest eigenvalues of sparse inhomogeneous Erdős-Rényi graphs". In: *Ann. Probab.* 47.3 (2019), pp. 1653–1676.

[21] F. Benaych-Georges and S. Péché. "Largest eigenvalues and eigenvectors of band or sparse random matrices". In: *Electron. Commun. Probab.* 19 (2014), no. 4, 9.

[22] F. Benaych-Georges et al. *Advanced topics in random matrices*. Ed. by D. Chafaï, S. Péché, and B. de Tilière. Vol. 53. Panoramas et Synthèses [Panoramas and Syntheses]. Lecture notes based on the Conference "États de la Recherche en Matrices Aléatoires held at Institute Henri Poincaré, Paris, December 1–4, 2014. Société Mathématique de France, Paris, 2017, pp. ix+190.

[23] E. A. Bender and E. R. Canfield. "The asymptotic number of labeled graphs with given degree sequences". In: *J. Combinatorial Theory Ser. A* 24.3 (1978), pp. 296–307.

[24] I. Benjamini and O. Schramm. "Recurrence of distributional limits of finite planar graphs". In: *Selected Works of Oded Schramm* (2011), pp. 533–545.

[25] L Bogachev, S. Molchanov, and L. Pastur. "On the density of states of random band matrices.(Russian) Mat. Zametki 50 (1991), no. 6, 31–42; translation in Math". In: *Notes* 50 (1992), pp. 5–6.

[26] B. Bollobás. *Random graphs*. Springer, 1998.

[27] C. Bordenave, P. Caputo, and D. Chafaï. "Spectrum of non-Hermitian heavy tailed random matrices". In: *Comm. Math. Phys.* 307.2 (2011), pp. 513–560.

[28] C. Bordenave and D. Chafaï. "Around the circular law". In: *Probab. Surv.* 9 (2012), pp. 1–89.

[29] C. Bordenave and M. Lelarge. "Resolvent of large random graphs". In: *Random Structures & Algorithms* 37.3 (2010), pp. 332–352.

[30] M. Bożejko and R. Speicher. "Interpolations between bosonic and fermionic relations given by generalized Brownian motions". In: *Math. Z.* 222.1 (1996), pp. 135–159.

[31] T. Brailovskaya and R. van Handel. "Universality and sharp matrix concentration inequalities". In: *arXiv preprint arXiv:2201.05142* (2022).

[32] W. Bruzda et al. "Random quantum operations". In: *Phys. Lett. A* 373.3 (2009), pp. 320–324.

[33] T. T. Cai, R. Han, and A. R. Zhang. "On the non-asymptotic concentration of heteroskedastic Wishart-type matrix". In: *Electron. J. Probab.* 27 (2022), Paper No. 29, 40.

[34] T. T. Cai, Z. Ren, and H. H. Zhou. "Estimating structured high-dimensional covariance and precision matrices: optimal rates and adaptive estimation". In: *Electron. J. Stat.* 10.1 (2016), pp. 1–59.

[35] M. Capitaine and C. Donati-Martin. "Strong asymptotic freeness for Wigner and Wishart matrices". In: *Indiana University mathematics journal* 56.2 (2007), pp. 767–803.

[36] C. W. Chin. "Necessary and sufficient conditions for convergence to the semicircle distribution". In: *Random Matrices Theory Appl.* 12.1 (2023), Paper No. 2250045, 28.

[37] B. Collins and P. Y. G. Lamarre. "$*$-freeness in finite tensor products". In: *Advances in Applied Mathematics* 83 (2017), pp. 47–80.

[38] B. Collins and C. Male. "The strong asymptotic freeness of Haar and deterministic matrices". In: *Ann. Sci. Éc. Norm. Supér. (4)* 47.1 (2014), pp. 147–163.

[39] N. A. Cook, R. Ducatez, and A. Guionnet. "Full large deviation principles for the largest eigenvalue of sub-Gaussian Wigner matrices". In: *arXiv:2302.14823* (2023).

[40] C. Cooper and A. Frieze. "The size of the largest strongly connected component of a random digraph with a given degree sequence". In: *Combin. Probab. Comput.* 13.3 (2004), pp. 319–337.

[41] B. Dadoun and P. O. Santos. *A combinatorial view on star moments of regular directed graphs and trees*. 2023. arXiv: 2309.02225.

[42] P. Dehornoy and P. Biane. "Dual Garside Structure of Braids and Free Cumulants of Products". In: *Séminaire Lotharingien de Combinatoire* 72 (2014), p. 15.

[43] R. Durrett. *Probability: theory and examples*. Vol. 49. Cambridge university press, 2019.

[44] K. Dykema. "On certain free product factors via an extended matrix model". In: *Journal of functional analysis* 112.1 (1993), pp. 31–60.

[45] L. Erdős, B. Schlein, and H.-T. Yau. "Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices". In: *Ann. Probab.* 37.3 (2009), pp. 815–852.

[46] L. Erdős, H.-T. Yau, and J. Yin. "Rigidity of eigenvalues of generalized Wigner matrices". In: *Adv. Math.* 229.3 (2012), pp. 1435–1515.

[47] L. Erdős et al. "The local semicircle law for a general class of random matrices". In: *Electron. J. Probab.* 18 (2013), no. 59, 58.

[48] L. Erdős, H.-T. Yau, and J. Yin. "Bulk universality for generalized Wigner matrices". In: *Probability Theory and Related Fields* 154.1-2 (2012), pp. 341–407.

[49] J. Friedman. "A proof of Alon's second eigenvalue conjecture". In: *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*. 2003, pp. 720–724.

[50] J. Friedman. "Relative expanders or weakly relatively Ramanujan graphs". In: *Duke Math. J.* 118.1 (2003), pp. 19–35.

[51] V. L. Girko. "The circle law". In: *Teor. Veroyatnost. i Mat. Statist.* 28 (1983), pp. 15–21.

[52] V. L. Girko. "The circular law: ten years later". In: *Random Oper. Stochastic Equations* 2.3 (1994), pp. 235–276.

[53] L. Goldmakher et al. "On the spectral distribution of large weighted random regular graphs". In: *Random Matrices Theory Appl.* 3.4 (2014), pp. 1450015, 22.

[54] C. E. González-Guillén, M. Junge, and I. Nechita. "On the spectral gap of random quantum channels". In: *arXiv:1811.08847* (2018).

[55] F. Götze, A. A. Naumov, and A. N. Tikhomirov. "Limit theorems for two classes of random matrices with dependent entries". In: *Theory Probab. Appl.* 59.1 (2015), pp. 23–39.

[56] A. Guionnet and J. Husson. "Large deviations for the largest eigenvalue of Rademacher matrices". In: *Ann. Probab.* 48.3 (2020), pp. 1436–1465.

[57] A. Guntuboyina and H. Leeb. "Concentration of the spectral measure of large Wishart matrices with dependent entries". In: *Electron. Commun. Probab.* 14 (2009), pp. 334–342.

[58] U. Haagerup and F. Larsen. "Brown's spectral distribution measure for R-diagonal elements in finite von Neumann algebras". In: *J. Funct. Anal.* 176.2 (2000), pp. 331–367.

[59] U. Haagerup and S. Thorbjørnsen. "A new application of random matrices: is not a group". In: *Annals of Mathematics* (2005), pp. 711–775.

[60] H. Hamburger. "Über eine Erweiterung des Stieltjesschen Momentenproblems". In: *Mathematische Annalen* 81.2–4 (June 1920), 235–319.

[61] R. van Handel. "On the spectral norm of Gaussian random matrices". In: *Trans. Amer. Math. Soc.* 369.11 (2017), pp. 8161–8178.

[62] M. B. Hastings. "Random unitaries give quantum expanders". In: *Phys. Rev. A* 76 (3 2007), p. 032315.

[63] S. Hoory, N. Linial, and A. Wigderson. "Expander graphs and their applications". In: *Bull. Amer. Math. Soc. (N.S.)* 43.4 (2006), pp. 439–561.

[64] S. Janson. "The probability that a random multigraph is simple". In: *Combin. Probab. Comput.* 18.1-2 (2009), pp. 205–225.

[65] I. M. Johnstone. "On the distribution of the largest eigenvalue in principal components analysis". In: *Ann. Statist.* 29.2 (2001), pp. 295–327.

[66] T Kemp. "Introduction to Random Matrix Theory". In: *UCSD Lecture Notes for Math A* 247 (2013).

[67] H. Kesten. "Symmetric random walks on groups". In: *Transactions of the American Mathematical Society* 92.2 (1959), pp. 336–354.

[68] O. Khorunzhiy. "Estimates for moments of random matrices with Gaussian elements". In: *Séminaire de probabilités XLI* (2008), pp. 51–92.

[69] V. Koltchinskii and K. Lounici. "Concentration inequalities and moment bounds for sample covariance operators". In: *Bernoulli* 23.1 (2017), pp. 110–133.

[70] M. Krivelevich and B. Sudakov. "The largest eigenvalue of sparse random graphs". In: *Combin. Probab. Comput.* 12.1 (2003), pp. 61–72.

[71] C. Lancien and D. Pérez-García. "Correlation Length in Random MPS and PEPS". In: *Ann. Henri Poincaré* 23 (1 2022), pp. 141–222.

[72] C. Lancien and P. Youssef. "A note on quantum expanders". In: *arXiv preprint arXiv:2302.07772* (2023).

[73] C. Lancien, P. O. Santos, and P. Youssef. *Limiting spectral distribution of random self-adjoint quantum channels*. 2023. arXiv: 2311.12368 [quant-ph].

[74] R. Latała, R. van Handel, and P. Youssef. "The dimension-free structure of nonhomogeneous random matrices". In: *Invent. Math.* 214.3 (2018), pp. 1031–1080.

[75] R. Latała and W. Światkowski. "Norms of randomized circulant matrices". In: *Electronic Journal of Probability* 27 (2022), pp. 1–23.

[76] J. O. Lee and K. Schnelli. "Local law and Tracy-Widom limit for sparse random matrices". In: *Probab. Theory Related Fields* 171.1-2 (2018), pp. 543–616.

[77] F. Lehner. "Free cumulants and enumeration of connected partitions". In: *European J. Combin.* 23.8 (2002), pp. 1025–1031.

[78] B. D. McKay. “The expected eigenvalue distribution of a large regular graph”. In: *Linear Algebra and its applications* 40 (1981), pp. 203–216.

[79] A. Minasyan and N. Zhivotovskiy. “Statistically Optimal Robust Mean and Covariance Estimation for Anisotropic Gaussians”. In: *arXiv preprint arXiv:2301.09024* (2023).

[80] W. Młotkowski. “Λ-free probability”. In: *Infinite Dimensional Analysis, Quantum Probability and Related Topics* 07.01 (Mar. 2004), 27–41.

[81] S. A. Molchanov, L. A. Pastur, and A. Khorunzhii. “Limiting eigenvalue distribution for band random matrices”. In: *Theoretical and Mathematical Physics* 90.2 (1992), pp. 108–118.

[82] A. Nica. “Asymptotically free families of random unitaries in symmetric groups”. In: *Pacific J. Math.* 157.2 (1993), pp. 295–310.

[83] A. Nica. “Free probability aspect of irreducible meandric systems, and some related observations about meanders”. In: *Infinite Dimensional Analysis, Quantum Probability and Related Topics* 19.02 (June 2016), p. 1650011.

[84] A. Nica, D. Shlyakhtenko, and R. Speicher. “Operator-valued distributions. I. Characterizations of freeness”. In: *International Mathematics Research Notices* 2002.29 (2002), pp. 1509–1538.

[85] A. Nica and R. Speicher. *Lectures on the combinatorics of free probability*. Vol. 335. London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge, 2006, pp. xvi+417.

[86] A. Nica and R. Speicher. “R-diagonal pairs—a common approach to Haar unitaries and circular elements”. In: *Free probability theory (Waterloo, ON, 1995)*. Vol. 12. Fields Inst. Commun. Amer. Math. Soc., Providence, RI, 1997, pp. 149–188.

[87] A. Nijenhuis and H. S. Wilf. “The enumeration of connected graphs and linked diagrams”. In: *J. Combin. Theory Ser. A* 27.3 (1979), pp. 356–359.

[88] G. Pisier. “Quantum expanders and geometry of operator spaces”. In: *J. Eur. Math. Soc.* 16.6 (2014), pp. 1183–1219.

[89] P. Rigollet and J.-C. Hütter. “High-dimensional statistics”. In: *arXiv preprint arXiv:2310.19244* (2023).

[90] M. Rudelson and R. Vershynin. “No-gaps delocalization for general random matrices”. In: *Geometric and Functional Analysis* 26.6 (2016), pp. 1716–1776.

[91] P. O. Santos. *Almost sharp covariance and Wishart-type matrix estimation*. 2023. arXiv: 2307.09190 [math.ST].

[92] K. Schmüdgen. *The moment problem*. Vol. 277. Graduate Texts in Mathematics. Springer, Cham, 2017, pp. xii+535.

[93] Y. Seginer. “The expected norm of random matrices”. In: *Combinatorics, Probability and Computing* 9.2 (2000), pp. 149–166.

[94] S. Sodin. "The spectral edge of some random band matrices". In: *Annals of mathematics* (2010), pp. 2223–2251.

[95] S. Sodin. "The Tracy-Widom law for some sparse random matrices". In: *J. Stat. Phys.* 136.5 (2009), pp. 834–841.

[96] A. Soshnikov. "Universality at the edge of the spectrum in Wigner random matrices". In: *Comm. Math. Phys.* 207.3 (1999), pp. 697–733.

[97] R. Speicher and M. Weber. "Quantum groups with partial commutation relations". In: *Indiana University Mathematics Journal* 68.6 (2019), pp. 1849–1883.

[98] R. Speicher and J. Wysoczański. "Mixtures of classical and free independence". In: *Archiv der Mathematik* 107.4 (Sept. 2016), 445–453.

[99] R. P. Stanley. *Catalan numbers*. Cambridge University Press, New York, 2015, pp. viii+215.

[100] C. Stein. "Estimation of a covariance matrix". In: *39th Annual Meeting IMS, Atlanta, GA, 1975*. 1975.

[101] M. Talagrand. "Concentration of measure and isoperimetric inequalities in product spaces". In: *Inst. Hautes Études Sci. Publ. Math.* 81 (1995), pp. 73–205.

[102] T. Tao. *Topics in random matrix theory*. Vol. 132. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2012, pp. x+282.

[103] T. Tao. *Topics in random matrix theory*. Vol. 132. American Mathematical Society, 2023.

[104] T. Tao and V. Vu. "Random matrices: universality of ESDs and the circular law". In: *Ann. Probab.* 38.5 (2010). With an appendix by Manjunath Krishnapur, pp. 2023–2065.

[105] K. Tikhomirov and P. Youssef. "Outliers in spectrum of sparse Wigner matrices". In: *Random Structures & Algorithms* 58.3 (2021), pp. 517–605.

[106] C. A. Tracy and H. Widom. "On orthogonal and symplectic matrix ensembles". In: *Comm. Math. Phys.* 177.3 (1996), pp. 727–754.

[107] R. Vershynin. "High-dimensional probability". In: *University of California, Irvine* (2020).

[108] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*. Vol. 47. Cambridge university press, 2018.

[109] D. Voiculescu. "Limit laws for random matrices and free products". In: *Inventiones mathematicae* 104.1 (1991), pp. 201–220.

[110] D. Voiculescu. "Symmetries of some reduced free product C*-algebras". In: *Operator Algebras and their Connections with Topology and Ergodic Theory*. Springer Berlin Heidelberg, 1985, 556–588.

[111] E. P. Wigner. “On the distribution of the roots of certain symmetric matrices”. In: *Ann. of Math. (2)* 67 (1958), pp. 325–327.

[112] E. P. Wigner. “On the distribution of the roots of certain symmetric matrices”. In: *Annals of Mathematics* 67.2 (1958), pp. 325–327.

[113] W. Woess. *Random walks on infinite graphs and groups*. Vol. 138. Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, 2000, pp. xii+334.

[114] M. M. Wolf. “Quantum channels & operations: Guided tour”. Lecture notes. 2012.

[115] N. C. Wormald. “Some problems in the enumeration of labelled graphs”. In: *Bulletin of the Australian Mathematical Society* 21.1 (1980), pp. 159–160.

[116] R.-y. Yang and J. O. Berger. “Estimation of a covariance matrix using the reference prior”. In: *Ann. Statist.* 22.3 (1994), pp. 1195–1211.

[117] N. Zhivotovskiy. “Dimension-free bounds for sums of independent matrices and simple tensors via the variational principle”. In: *Electronic Journal of Probability* 29 (Jan. 2024).