HABILITATION À DIRIGER DES RECHERCHES

École Doctorale de Sciences Mathématiques de Paris Centre

présentée par

Pierre YOUSSEF

Matrices to random taste: phenomena in high dimension

Rapporteurs :	Charles Bordenave	Université Aix Marseille
	Stéphane Boucheron	Université Paris Diderot
	Roman Vershynin	University of California Irvine

Soutenue le **17 Décembre 2019** devant le jury composé de

Franck Barthe	Université Paul Sabatier, Toulouse
Charles Bordenave	Université Aix Marseille
Stéphane Boucheron	Université Paris Diderot
Djalil Chafaï	Université Paris Dauphine
Olivier Guédon	Université Paris-Est Marne-La-Vallée
Alice Guionnet	Ecole Normale Supérieure de Lyon
Sandrine Péché	Université Paris Diderot
Assaf Naor	Princeton University

Remerciements

J'aimerai commencer par remercier Charles, Roman et Stéphane qui ont accepté de lire et rapporter ce manuscrit. Je remercie également Alice, Assaf, Djalil, Franck, Olivier et Sandrine d'avoir accepté de faire partie du jury. Je suis vraiment honoré et reconnaissant.

Je remercie spécialement Olivier qui a joué le rôle d'un grand frère, et ce depuis mes années de thèse. Djalil, merci pour les nombreux conseils et pour ton inspirante vision des choses. Je remercie Assaf pour tout son support et pour avoir supporté, malgré son planning chargé, mes demandes excessives.

Je remercie Dario Cordero-Erausquin qui avait accepté de m'encadrer en Master. Je dois beaucoup à Gilles Godefroy qui a été celui qui m'a mis sur le chemin amenant aux maths que je fais maintenant en m'orientant vers Dario, et surtout celui qui a su avoir les mots déterminants lorsque j'ai été proche d'abandonner. Je remercie également Nicole Tomczak–Jaegermann, qui m'a tant inspiré par sa détermination et a surtout bienveillé à ce que je ne baisse pas le rythme de travail. J'ai une pensée à Alain Pajor qui m'a tant impressionné par son élégance mathématique et ses qualités personnelles.

J'aimerai remercier le LPSM qui a pris le risque de me recruter, moi qui était un peu étranger à la communauté probabiliste. Je remercie tous les membres du labo pour leur accueil (surtout Marie pour mes premiers pas) et pour tout ce que j'ai eu la chance d'apprendre pendant ces 4 ans. Je tiens à remercier Florence, Nathalie, Valérie et enfin Amina, pour leur patience avec mon incompétence administrative. Jean-François, merci pour tes conseils et le partage de tes expériences. Je remercie Mathieu qui a du supporter mes demandes, mais aussi pour les moments partagés avec Justin autour de bières. Enfin, merci Justin pour ton amitié, les innombrables rigolades, les moments de doute et d'espoir mathématiques ou pas.

Ce mémoire doit beaucoup au séjour que j'ai effectué au MSRI à Berkeley. J'aimerai remercier les organisateurs (Franck en faisant partie) pour ce semestre très réussi. Ce fut un régal de faire des maths, discuter avec les nombreux collègues/amis, jouer au basket (et se briser le genou), et profiter du magnifique cadre de vie.

Ce manuscrit est le fruit de nombreuses collaborations. J'ai eu la chance de collaborer avec d'excellents chercheurs de qui j'ai beaucoup appris. Je remercie: Anna, Assaf, Florence, Kostya, Marwa, Nicole, Omer, Rafal, Ramon, Richard, Sasha. Je remercie Marwa et Richard pour leur chère amitié, Ramon pour son support et le partage de ses connaissances. Je remercie Kostya pour son amitié et pour les randonnées mathématiques qu'on a pu faire.

Les amis à Paris, et vous vous connaissez, je vous remercie tous d'avoir supporté mes moments d'obsession. Mes parents, mes soeurs, vos sacrifices et votre confiance en moi, est ce qui a aidé à transformer les conjectures en théorèmes dans ce manuscrit.

A mes moments complexes, réels ou imaginaires, et à mes rêves irrationnels...

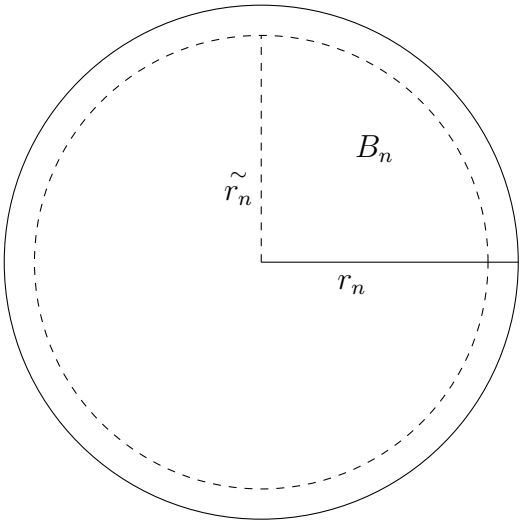
Overview

The present manuscript summarizes my research work which resulted in the publication list given in the sequel. It is meant to give an overview of my contributions without entering into technical details. It should be noted that the manuscript is far from being an extensive survey of the research directions which will be discussed, and focuses instead on my personal work.

I am mainly interested in high dimensional phenomena where one ideally aim at the following three stages:

1. Capture a phenomenon arising as the "dimension" grows.
2. Quantify the phenomenon.
3. Provide, when possible, a qualitative explanation as to why such a phenomenon occurs.

Let me illustrate this by a very basic observation. Let B_2^n be the Euclidean ball of radius 1 in $\mathbb{R}^n$ and B_n be the Euclidean ball of volume one. It can be checked that the radius r_n of B_n satisfies $r_n \sim \sqrt{\frac{n}{2\pi e}}$ as $n \to \infty$. Now $(B_n, \mathcal{B}(B_n), \mathrm{Vol})$ defines a probability space and by multi-linearity of the volume, one has that the volume of $(1-\varepsilon)B_n$ vanishes as $n \to \infty$.



- $B_n = r_n B_2^n$ of volume 1.
- $(B_n, \mathcal{B}, \mathrm{Vol})$ probability space.
- $\widetilde{r_n} = (1-\varepsilon) r_n$.
- $\mathrm{Vol}(\widetilde{r_n} B_2^n) = (1-\varepsilon)^n \underset{n\to\infty}{\longrightarrow} 0$.

This trivial observation is very enlightening as it asserts that the ball becomes "empty" as the dimension grows and the whole mass is pushed out to the surface. This phenomenon which is captured in the high dimension regime is known as the concentration of measure and

can be quantified as to control the decay of the mass outside a set of non-negligible volume. It is known to be implied by isoperimetric inequalities and is further established on other probability spaces. We refer to [88] and references therein for more on the concentration of measure phenomenon and to the excellent recent book [146] for an introduction to high dimensional probability.

The high dimensional componant is a common feature to all my research work combining random matrices, spectrum of random graphs, geometric functional analysis, concentration inequalities and data selection problems. At the intersection of these subjects is the word "Matrix", a very simple object which will be looked at from different angles. In Chapter 1, we will look at it from a data scientist point of view and study problems of extracting submatrices with certain properties from a large array of data. This will cover some results present in [Y2,Y4,Y9,Y10,Y19]. In Chapter 2, we add a geometric flavor by interpreting a matrix as a linear transformation, studying distances between high dimensional convex bodies and establishing some structural properties. The geometry is also mixed with probability as we study convex hulls of random walks in high dimensions [Y1,Y5,Y10,Y19]. In Chapter 3, we officially enter the random matrix world as we focus on the study of the norm of random matrices with independent entries and the outliers of their limiting spectral distribution [Y15,Y18]. In Chapter 4 , we further dive in random matrix theory by dropping the independence feature and investigating the universality phenomenon through the study of random regular graphs. Our journey there leads us to visit interesting problems as the rank of their adjacency matrices, their "well" invertibility captured by the smallest singular value, delocalization properties of their eigenvectors and finally their limiting spectral distribution [Y7,Y8,Y13,Y14,Y16,Y17]. In Chapter 5, we move from the macroscopic level and investigate the edge of the spectrum of random graphs by studying the spectral gap of the two famous random graph models: random regular graphs and the Erdős-Renyi graph [Y11,Y12,Y15,Y18]. Finally, in Chapter 6, we discuss classical concentration inequalities extensions in two different directions: implementing dependence and providing non-commutative analogues [Y3,Y6,Y12,Y15].

Publications

The articles [Y1,Y2,Y3] below were part of my PhD thesis. The papers [Y4,Y5,Y6] were completed while in postdoc at the University of Alberta, and the remaining works were done since I moved to Paris Diderot in September 2015 as a maître de conférences.

[Y1] **Restricted invertibility and the banach-mazur distance to the cube.**
Mathematika 60 (2014), no. 1, 201-218.

[Y2] **A note on column subset selection.**
International Mathematics Research Notices, 2014(23):6431–6447, 2014.

[Y3] **Estimating the covariance of random matrices.**
Electronic journal of Probability, Vol 18, no.107, 1-26.

[Y4] **Extracting a basis with fixed block inside a matrix.**
Linear Algebra and Its Applications, 469(Complete):28–38, 2015.

[Y5] **When does a discrete-time random walk in $\mathbb{R}^n$ absorb the origin into its convex hull?**
Ann. Probab. 45 (2017), no. 2, 965-1002.
(Joint with K. Tikhomirov)

[Y6] **Bernstein type inequality for a class of dependent random matrices.**
Random Matrices Theory Appl. 5 (2016), no. 2, 1650006, 28 pp.
(Joint with M. Banna and F. Merlevède)

[Y7] **Anti-concentration property for random digraphs and invertibility of their adjacency matrices.**
C.R. Math. Acad. Sci. Paris, 354 (2016), 121–124.
(Joint with A. Lytova, A. Litvak, K. Tikhomirov and N. Tomczak-Jaegermann)

[Y8] **Adjacency matrices of random digraphs: singularity and anti-concentration.**
J. of Math. Analysis and Appl., 445 (2017), 1447–1491. (Winner of AMES 2017 award)
(Joint with A. Lytova, A. Litvak, K. Tikhomirov and N. Tomczak-Jaegermann)

[Y9] **Restricted invertibility revisited.**
A Journey Through Discrete Mathematics. A Tribute to Jirí Matousek., Springer International Publishing, 657–691, 2017.
(Joint with A. Naor)

[Y10] **Approximating matrices and convex bodies.**
International Mathematics Research Notices, Vol. 2017, No. 00, pp. 1–19.
(Joint with O. Friedland)

[Y11] **On the norm of a random jointly exchangeable matrix.**
J. Theor. Probab (2018). https://doi.org/10.1007/s10959-018-0844-y
(Joint with K. Tikhomirov)

[Y12] **The spectral gap for dense random regular graphs.**
Ann. Probab. 47 (2019), no. 1, 362–419.
(Joint with K. Tikhomirov)

[Y13] **The smallest singular value of a shifted d-regular random square matrix.**
Prob. Th. Rel. Fields, 173 (2019), 1301–1347.
(Joint with A. Lytova, A. Litvak, K. Tikhomirov and N. Tomczak-Jaegermann)

[Y14] **The rank of random regular digraphs of constant degree.**
J. of Complexity, 48 (2018), 103–110.
(Joint with A. Lytova, A. Litvak, K. Tikhomirov and N. Tomczak-Jaegermann)

[Y15] **The dimension-free structure of nonhomogeneous random matrices.**
Inventiones mathematicae (2018), Volume 214, No. 3, pp. 1031–1080.
(Joint with R. Latala et R. van Handel)

[Y16] **Structure of eigenvectors of random d-regular digraphs.**
Trans. Amer. Math. Soc., 371 (2019), 8097–8172.
(Joint with A. Lytova, A. Litvak, K. Tikhomirov et N. Tomczak-Jaegermann)

[Y17] **Circular law for sparse random regular digraphs.**
To appear in Journal of the European Mathematical Society.
(Joint with A. Lytova, A. Litvak, K. Tikhomirov and N. Tomczak-Jaegermann)

[Y18] **Outliers in spetrum of sparse Wigner matrices.**
Available at arXiv:1904.07985
(Joint with K. Tikhomirov)

[Y19] **Courant-Fisher bound for coordinate subspaces.**
preprint.
(Joint with A. Naor)

Contents

Chapter 1

Data selection problems

In the current era of data science, determining among millions of features which are the informative ones is an important problem. The ability to select such features from high-dimensional data is therefore crucial. Column selection problems consist of searching inside a big data set for some specific subset. More precisely, we are given a flat matrix A of size $n \times m$ (the column vectors play the role of the input data) and we would like to select inside A a block of columns satisfying special properties. We investigate in this chapter several problems of this kind.

1.1 Approximating matrices [Y10]

A very natural problem is to look for a sub-matrix with much fewer columns, approximating the original one. One can ask for such an approximation using different notions of matrix norms. Perhaps the most interesting one is the operator norm as it provides a tight control on the singular values.

Let A be an $n \times m$ matrix. We denote by $s_i(A) = \sqrt{\lambda_i(A^t A)}$ the singular values of A, where A^t denotes the transpose matrix of A, and $\lambda_i(A^t A)$ denotes the i^{th}-eigenvalue of $A^t A$ when rearranged in the non-increasing order. The operator norm of A is given by

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

where $\|\cdot\|_2$ denotes the Euclidean norm on $\mathbb{R}^n$ and S^{n-1} its unit Euclidean sphere. It is easy to check that the operator norm is equal to the largest singular value of the matrix. The singular values play an important role in the numerical analysis of data. Indeed, several data mining and low rank approximation methods are based on the analysis of the singular value decomposition. For instance, writing the spectral decomposition of $AA^t = \sigma_1^2 v_1 v_1^t + \ldots + \sigma_n^2 v_n v_n^t$ where $v_1, \ldots, v_n$ are the left singular vectors of A, we can see that a good approximation of AA^t is obtained just by projecting A on the subspace spanned by singular vectors associated with "large" singular values. This observation indicates that the rank of the approximation should be somehow approximated by the number of "large" singular values. A quantity measuring this is what is usually called *numerical rank* or *stable rank* and

is given by

$$\operatorname{srank}(A) := \frac{\|A\|_{\mathrm{HS}}^2}{\|A\|^2}$$

where $\|A\|_{\mathrm{HS}}$ denotes the Hilbert-Schmidt norm (or Frobenius norm) of A, i.e. $\|A\|_{\mathrm{HS}} = \sqrt{\operatorname{Tr}(AA^t)} = \sqrt{\sum_{i\leq m} s_i(A)^2}$. It is easy to check that $\operatorname{srank}(A)$ is less or equal than $\operatorname{rank}(A)$. Since $\operatorname{srank}(A)$ is the ratio of the sum of all singular values squared to the squared largest one, it doesn't take into account the tiny singular values. Therefore, it indeed accounts to the number of "large" singular values. Our goal is to obtain a "canonical" approximation of AA^t by extracting a block of columns in A serving as the low rank approximation. More precisely, we aim to show the existence of a small subset $\sigma \subset [m] := \{1, \ldots, m\}$ so that A_σ, the matrix containing the columns indexed by σ, verifies that

$$\|A_\sigma A_\sigma^t - AA^t\|$$

is small. This ensures that the spectrum of $A_\sigma A_\sigma^t$ is close to that of AA^t implying the same for the sequence of singular values. Such statement would be much more informative than the approximation done using the singular value decomposition as one now reads the approximation inside the original data. We treat this problem while allowing a reweighting of the extracted columns. This corresponds to finding a multi-set $\sigma \subset [m]$ and looking at A_σ with the number of repetitions of an index in σ giving the value of the weight associated to the corresponding column. We denote by $\widetilde{A}$ the matrix obtained by normalizing the columns of A by their Euclidean norms. In [Y10], we obtain the following.

Theorem 1.1 ([Y10]). *Let A be an $n \times m$ matrix of unit norm. Then, for any $\varepsilon > 0$ there exists a multi-set $\sigma \subset [m]$ with $|\sigma| \leq \operatorname{srank}(A)/c\varepsilon^2$ so that*

$$\left\| c\varepsilon^2 \widetilde{A}_\sigma \widetilde{A}_\sigma^t - AA^t \right\| \leq \varepsilon$$

where $c > 0$ is a universal constant. Moreover if $A = \kappa\widetilde{A}$ for some $\kappa > 0$, then the above holds with σ being a set.

The conclusion of the above theorem can be also formulated as follows. There exists an $m \times m$ non-negative diagonal matrix with rank at most $\operatorname{srank}(A)/c\varepsilon^2$ so that

$$\|ADA^t - AA^t\| \leq \varepsilon.$$

When $A = \kappa\widetilde{A}$ for some $\kappa > 0$, the non-zero diagonal entries of D are given explicitly by $c\varepsilon^2/\kappa^2$.

Results of this kind, which falls under the name of "column subset selection", attract a lot of attention in the numerical analysis and the analysis of algorithms communities. Many papers are devoted to this problem, let us mention a few [39, 44, 14, 141] and refer to [73] and references therein for a detailed exposition on the topic. Theorem 1.1 produces the minimal size approximation compared to the results available in the literature. For instance, the best known bounds on the number of selected columns are of order $\operatorname{srank}(A)\log(\operatorname{srank}(A))$ (see

Table 2 and Theorem 4.1 in [76]). This is for example done in [121] using random sampling in which case the logarithmic factor is needed. Thus, we improve the existing results by removing the logarithmic factor and obtaining, up to a constant, the optimal sampling size for this approximation problem. It should be noted that our result is only existential and we do not know how to produce a polynomial time algorithm achieving the extraction promised in the theorem. The main reason for this is that our proof is based on the solution of the Kadison-Singer problem [93] which will be discussed in the next section.

We should mention that in [Y10], we developed as well a tensorization trick to treat constraint approximation problems. One may ask to achieve an approximation of the matrix while keeping some special properties it has. For example, given an $n \times m$ matrix A and a vector v in its kernel, one may be interested in approximating A by a submatrix with fewer columns while keeping v not far from the kernel of the restricted matrix. This is motivated by some geometric applications which will be discussed later. To give a brief idea, looking at the columns of the matrix A as vectors in $\mathbb{R}^n$, we see the vector v from the kernel as some weighted barycenter of the column vectors of A. Therefore, the constraint in this case can be seen as keeping v not far from being a weighted barycenter of the selected column vectors. We refer to Theorem 1.2 in [Y10] for a statement.

Finally, Theorem 1.1 also implies a sparsification result for quadratic forms improving on results of [22] by providing additional information on the sparsification weights.

Corollary 1.2. *Let $x_1, \dots, x_m \in \mathbb{R}^n$ with $\|x_i\|_2 = 1$ for any $i \leq m$ and $Id = \sum_{i \leq m} c_i x_i x_i^t$ for some positive scalars $(c_i)_{i \leq m}$. Then for any $\varepsilon \in (0,1)$, there exists a multi-set $\sigma \subset \{1, \dots, m\}$ with $|\sigma| \lesssim n/\varepsilon^2$ such that*

$$(1-\varepsilon) Id \preceq \frac{n}{|\sigma|} \sum_{i \in \sigma} x_i x_i^t \preceq (1+\varepsilon) Id,$$

where "$\preceq$" refers to the semi-definite order.

The notation $a \lesssim b$ above and in the remainder of the manuscript always refers to an inequality holding up to a universal constant i.e. there exists a universal constant c such that $a \leq cb$.

1.2 Well conditioned submatrices [Y2]

The condition number of an $n \times m$ matrix A is given by

$$\kappa(A) = \frac{s_{\max}(A)}{s_{\min(A)}},$$

where $s_{\max}$ and $s_{\min}$ denote the largest and smallest singular values of A. When the matrix is not of full rank, then its smallest singular value is equal to zero and its condition number explodes. Note that $\kappa(A) \geq 1$ and is equal to one whenever A is a multiple of an isometry. The condition number serves as a measure of precision of certain matrix algorithms [23, Chapter III], [128]. The problem of interest here is to extract inside any matrix A a "large" submatrix which is "well" conditioned. In this context, we proved the following.

Theorem 1.3 ([Y2]). *Let A be an $n \times m$ matrix and $\widetilde{A}$ the matrix obtained by normalizing the columns of A by their Euclidean norms. Then for any $\varepsilon \in (0,1)$, there exists $\sigma \subset \{1,\dots,m\}$ of size at least $c\varepsilon^2 \mathrm{srank}(A)$ such that $\kappa(\widetilde{A}_\sigma) \leq 1+\varepsilon$, where c is a universal constant.*

Theorem 1.3 improves on earlier results of Vershynin [144] and its proof which builds on the work of [22] provides a polynomial time algorithm to find the promised set σ. Theorem 1.3 answers a question raised by Naor as to obtain a constructive proof of a related result of Bourgain and Tzafriri [35, 36, 37]. The latter is related to the famous Kadison-Singer problem which originated in operator theory and was left open since 1959 until its resolution by Marcus, Spielman and Srivastava [93]. One of its many equivalent formulations is the paving problem which states that any $n \times n$ matrix with zero diagonal can be partitioned into a universal number of diagonal blocks of small norm. By iteratively using Theorem 1.3, one can obtain the following.

Theorem 1.4 ([Y2]). *Let A be an $n \times n$ symmetric matrix with zero diagonal. Then for any $\varepsilon \in (0,1)$, there exists a partition of $\{1,\dots,n\}$ into k sets $\sigma_1,\dots,\sigma_k$ such that $k \lesssim \varepsilon^{-2}\log n$ and*

$$\|P_{\sigma_i} A P_{\sigma_i}^t\| \leq \varepsilon \|A\|,$$

where $P_{\sigma_i} : \mathbb{R}^n \to \mathbb{R}^{\sigma_i}$ denotes the canonical projection on $\mathbb{R}^{\sigma_i}$.

This statement recovers results of Bourgain and Tzafriri [35], and Tropp [141], and has the advantage of producing a deterministic polynomial time algorithm to construct the above partition. The paving problem proved by Marcus, Spielman and Srivastava [93] significantly strengthens the above by showing the existence of a partition whose size is independent of the dimension of the matrices considered.

1.3 The restricted invertibility principle [Y4, Y9, Y19]

The restricted invertibility principle due to Bourgain and Tzafriri [35, 36, 37] asks to extract inside a matrix the largest number of columns possible such that the resulting sub-matrix is well invertible i.e. the norm of its inverse is well bounded. In addition to numerical analysis, this basic question finds applications in the local theory of Banach spaces as well as in harmonic analysis.

In this problem, we wish to find a large subset $\sigma \subset [m]$ such that A is injective on $\mathbb{R}^\sigma \subset \mathbb{R}^m$. Secondly, rather than being satisfied with mere invertibility we ask for A to be *quantitatively invertible* on $\mathbb{R}^\sigma$ in the sense that the operator norm of the inverse $A^{-1} : A(\mathbb{R}^\sigma) \to \mathbb{R}^\sigma$ is not too large. Following earlier results of Bourgain–Tzafriri [35] and Vershynin [144], Spielman–Srivastava [129] obtained the following improved restricted invertibility principle, relying nontrivially on a remarkable method for sparsifying quadratic forms that was developed by Batson–Spielman–Srivastava [22] (see also the survey [103]).

Theorem 1.5 (Spielman–Srivastava). *Suppose that $k, m, n \in \mathbb{N}$ and let A be an $n \times m$ matrix such that $k < \mathrm{srank}(A)$. Then there exists a subset $\sigma \subset [m]$ with $|\sigma| = k$ such that*

$$\|(AJ_\sigma)^{-1}\| \leq \frac{1}{1-\sqrt{k/\mathrm{srank}(A)}} \cdot \frac{\sqrt{m}}{\|A\|_{\mathrm{HS}}},$$

where J_σ denotes the formal identity from $\mathbb{R}^\sigma$ to $\mathbb{R}^m$.

Equivalently, the above theorem can be stated as a lower bound on the smallest singular value of the restricted matrix i.e.

$$s_{\min}(AJ_\sigma) \geq \big(1 - \sqrt{k/\mathrm{srank}(A)}\big) \frac{\|A\|_{\mathrm{HS}}}{\sqrt{m}}.$$

The quantity $\frac{\|A\|_{\mathrm{HS}}}{\sqrt{m}}$ appears naturally in this context. Indeed, if σ contains one element, then the corresponding sub-matrix is of rank one and its only non-zero singular value equals the Euclidean norm of the chosen columns. Now, $\frac{\|A\|_{\mathrm{HS}}}{\sqrt{m}}$ is the "ℓ_2-average" of the norms of the columns and provides therefore a sharp lower bound on the smallest singular value of the rank one matrix obtained. Due to Cauchy interlacing theorem, by selecting more columns inside the original matrix, the smallest singular value can only get smaller. Remarkably, Theorem 1.5 asserts that one can still lower bound the smallest singular value by $\frac{\|A\|_{\mathrm{HS}}}{\sqrt{m}}$ up to a loss by a multiplicative factor of $\big(1 - \sqrt{k/\mathrm{srank}(A)}\big)$.

In [Y9], we revisited this problem with the aim of increasing the size of the extracted sub-matrix and bringing back to life old methods from functional analysis essentially developed in the work of Bourgain and Tzafriri [35] as opposed to the modern linear algebraic method introduced in [22] and used in [129]. To this aim, we introduced for any $p > 2$ the notion of p-stable rank of an $n \times m$ matrix A as

$$\mathrm{srank}_p(A) := \left(\frac{\|A\|_{\mathsf{S}_2}}{\|A\|_{\mathsf{S}_p}}\right)^{\frac{2p}{p-2}}, \tag{1.1}$$

where

$$\|A\|_{\mathsf{S}_p} := \Big(\mathrm{Tr}(A^t A)^{\frac{p}{2}}\Big)^{\frac{1}{p}} = \left(\sum_{j=1}^m s_j(A)^p\right)^{\frac{1}{p}},$$

is the p-Schatten norm of A and the s_j's denote the singular values of A. With this definition, since $\|A\| = \|A\|_{\mathsf{S}_\infty}$ we have $\mathrm{srank}(A) = \mathrm{srank}_\infty(A)$. A direct application of Hölder's inequality implies that $\mathrm{srank}_p(A)$ is non-increasing in p so that $\mathrm{srank}(A)$ is the smallest stable rank. Taking the limit as $p \to 2^+$, we can also define the entropic stable rank (see [Y9] for more details). Let us state one of the many restricted invertibility statements obtained in [Y9].

Theorem 1.6 ([Y9])**.** *Suppose that $k, m, n \in \mathbb{N}$, $p > 2$, and let A be an $n \times m$ matrix such that $k < \mathrm{srank}_p(A)$. Then there exists a subset $\sigma \subset [m]$ with $|\sigma| = k$ such that*

$$\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \lesssim \psi_p\left(1 - \frac{k}{\mathrm{srank}_p(A)}\right) \frac{\sqrt{m}}{\|A\|_{\mathsf{S}_2}},$$

where $\psi_p : \mathbb{R} \to [0, \infty]$ is defined by $\psi_p(x) = \infty$ if $x \leq 0$, $\psi_p(x) = (\sqrt{p/(p-2)})/x$ if $0 < x < 1/2$, $\psi_p(x) = (\sqrt{p/(p-2)})/\log(1/(1-x))$ if $1/2 < x \leq 1 - e^{-p/(p-2)}$ and $\psi_p(x) = 1$ if $x > 1 - e^{-p/(p-2)}$.

In these results, one is interested in the dependence on $\varepsilon = 1 - \frac{k}{\mathrm{srank}_p(A)}$ in the regime where k is close to $\mathrm{srank}_p(A)$. Theorem 1.6 matches the optimal dependence on ε in Theorem 1.5. We should note that the case $p = 4$ was obtained in [95] using the method of interlacing polynomials. Theorem 1.6 can detect the well-invertibility of A on a much larger scale compared to what is captured by Theorem 1.5. For example suppose that the singular values of A are $s_1(A) \asymp \sqrt[3]{m}$ and $s_2(A) \asymp s_3(A) \asymp \ldots \asymp s_m(A) \asymp 1$. Then Theorem 1.5 yields a subset $\sigma \subset \{1, \ldots, m\}$ of size of order $m^{1/3}$ for which the operator norm of the inverse of AJ_σ is $O(1)$, while (the case $p = 3$ of) Theorem 1.6 yields such a subset whose size is proportional to m.

Through our study, we came to realize that the restricted invertibility problem is a two stage procedure: the first being a restricted invertibility on a global scale, and the second being a more refined one on a microscopic scale. More precisely, the first stage consists of finding a restriction for which the singular values, "on average", are large: this can be done by aiming to find $\omega \subset [m]$ of large size such that the quantity $\sum_{i=1}^{|\omega|} s_i^{-2}(AJ_\omega) = \|(AJ_\omega)^{-1}\|_{\mathsf{S}_2}^2$ is well bounded above, ensuring that most singular values of AJ_ω are well bounded below. This first stage, which can be seen as a restricted invertibility principle on a macroscopic scale, is much easier to achieve. For instance, we studied this problem [Y4] generalizing earlier results (see [Y4] for details and references). Once the first stage completed, one needs to refine the study by finding a further restriction on which all singular values are well bounded below. The difficulty and depth of the restricted invertibility principle lies in this second refined stage which can be seen as a restricted invertibility principle on a microscopic scale.

Theorem 1.7 ([Y19]). *Let $k < r \leq n \in \mathbb{N}$ and A be an $n \times r$ matrix of rank r. Then there exists $\sigma \subset [r]$ with $|\sigma| = k$ such that*

$$\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \lesssim \frac{\|A^{-1}\|_{\mathsf{S}_2}}{\sqrt{r-k}} \cdot \sqrt{\log\Big(\frac{r}{r-k}\Big)}.$$

The above theorem surprisingly asserts that any matrix is well invertible (up to removing few columns) as long as it is well invertible on average. The bound stated above is sharp up to the logarithmic term. Let us note that one always has $\frac{1}{\sqrt{r}}\|A^{-1}\|_{\mathsf{S}_2} \leq \|A^{-1}\|_{\mathsf{S}_\infty}$ and Theorem 1.7 asserts that this inequality can be reverted, up to a multiplicative factor, on a large sub-matrix. Interestingly, using the modern method of interlacing polynomials [94] produces a strictly weaker result as opposed to the classical methods developed by Bourgain and Tzafriri which we employ for the proof of the above theorem. Finally, before fully adding the geometric flavor in the next chapter, let us end this section by giving a geometric interpretation of the above result. Say A is an $n \times n$ invertible symmetric matrix and let $\mathcal{E} = AB_2^n$ be an ellipsoid in $\mathbb{R}^n$, where we denoted B_2^n the unit Euclidean ball in $\mathbb{R}^n$. The lengths of the axis of this ellipsoid are given by the singular values of A^{-1}. Theorem 1.7 states that there exists $\sigma \subset \{1, \ldots, n\}$ of size $n/2$ say such that the coordinate projection $P_\sigma \mathcal{E}$ contains a Euclidean ball of radius equal, up to a multiplicative constant, to the harmonic mean of the lengths of the axis of $\mathcal{E}$. Results of this form were previously studied by Giannopoulos [66], as well as Giannopoulos and Milman [67].

1.4 Perspectives

One of the major drawbacks of the modern techniques developed mainly in [22] and [94] is that they are very much spectral methods which therefore restrict their applicability to problems where matrices are seen as operators on ℓ_2. On the contrast, the old methods from Functional Analysis which we brought back to life in [Y9] and [Y19] are more flexible in this respect. Already in the work of Bourgain and Tzafriri [35], the restricted invertibility of operators on ℓ_p spaces was considered, and this gives us hope that with our current understanding, it would be possible to elaborate such invertibility results beyond the ℓ_2 setting. In a similar manner, extending the paving problem mentioned in Theorem 1.4 in terms of ℓ_p norms would be of great interest and requires the elaboration of new techniques.

Chapter 2

Matrices in convex geometry

In this chapter, matrices will appear naturally as linear transformations and embeddings between normed spaces. The object of interest is a convex body K in $\mathbb{R}^n$ i.e. a convex compact set of non-empty interior. When K is symmetric i.e. $K = -K$, one can see K as the unit ball of a norm defined as

$$\|x\|_K := \inf\{\lambda \in \mathbb{R}_+ : \, x \in \lambda K\},$$

so that the study of symmetric convex bodies is equivalent to that of n-dimensional normed spaces. Geometric Functional Analysis is concerned with the study of the structure of these convex bodies.

2.1 ℓ_1-analogue of Dvoretzky's theorem [Y1, Y19]

In order to capture structural properties of convex bodies, we start first by defining a distance which helps compare these objects. Given K, L two convex bodies in $\mathbb{R}^n$, the Banach-Mazur distance between K and L is given by

$$d(K, L) = \inf_{T,u} \big\{\alpha \geq 1 : \, L + u \subseteq T(K) \subseteq \alpha(L + u)\big\},$$

where the infimum is taken over all isomorphisms T on $\mathbb{R}^n$ and all vectors $u \in \mathbb{R}^n$. If K and L are symmetric, then $u = 0$. In words, two convex bodies are "close" if we can (up to an isomorphism) squeeze one into an another. In terms of the associated normed spaces, the distance measures the equivalence between the corresponding norms and we may as well write the distance between two normed spaces as the above distance between their unit balls. As an example, it is easy to check that $d(B_2^n, B_1^n) = \sqrt{n}$. Let us note that $\log d(\cdot, \cdot)$ is a distance so that $d(\cdot, \cdot)$ is multiplicative and two convex bodies are close whenever their distance is close to one. Moreover, $d(\cdot, \cdot)$ is invariant by polarity (or by duality of the norms).

One of the founding results in Geometric Function Analysis is Dvoretzky's theorem which looks for Euclidean structure inside high dimensional normed spaces.

Theorem 2.1 (Dvoretzky [51, 101])**.** *Let K be a symmetric convex body in $\mathbb{R}^n$. For any $\varepsilon \in (0, 1)$, there exists $k \gtrsim c(\varepsilon) \log n$ and a subspace E of dimension k such that $d(K \cap E, B_2^k) \leq 1 + \varepsilon$, where $c(\varepsilon)$ is a constant depending only on ε.*

The dimension of the corresponding subspace captured above is optimal as can be checked on the cube B_∞^n. There is a hidden fascinating aspect of Theorem 2.1 which asserts that if one takes a random (uniformly distributed according to the rotation-invariant measure) subspace E of dimension as above then it is most likely that $K \cap E$ is almost Euclidean. Such a statement can be made quantitative and the concentration of measure phenomenon plays an important role in its proof due to Milman [101] and in the development of the field (see [13, 102, 113, 146]).

A natural problem is to look for other structures such as the ℓ_1-ball for instance. The aim would be to find a large section which is close to $B_1^k := \{(x_1, \ldots, x_k) \in \mathbb{R}^k : |x_1| + \ldots + |x_k| \leq 1\}$. This is of course impossible as can be seen when looking at the Euclidean ball whose sections of dimension k are Euclidean and at distance $\sqrt{k}$ from the corresponding ℓ_1-ball. Instead, we will look to find a section with the largest possible dimension which is not far from the ℓ_1-ball i.e. matches the behavior of the Euclidean ball in this respect. This was studied by Bourgain and Szarek [34], Szarek and Talagrand [133], Giannopoulos [66]. Since B_1^k is the convex hull of the canonical basis which constitutes a perfectly conditioned matrix, finding an ℓ_1-structure inside K can be related to the restricted invertibility principle where one looks for a well invertible sub-matrix. This link was made explicit in [Y1] allowing to recover and improve the constants involved in the results of [66].

Theorem 2.2 ([Y1]). *Let K be a symmetric convex body in $\mathbb{R}^n$. For any $\varepsilon \in (0,1)$, there exists $k \geq (1-\varepsilon)n$ and a subspace E of dimension k such that $d(K \cap E, B_1^k) \lesssim \frac{\sqrt{k}}{\varepsilon(1-\varepsilon)}$.*

The above is related to a question concerning the structure of the Banach-Mazur compactum $\mathcal{M}_n$ defined as the collection of all centrally symmetric convex bodies in $\mathbb{R}^n$. It follows from John's theorem [79] that $d(K, B_2^n) \leq \sqrt{n}$ for any $K \in \mathcal{M}_n$ meaning that the diameter of $\mathcal{M}_n$ is less than n. On the other hand, it was shown by Gluskin [69] that this is sharp meaning that the diameter of $\mathcal{M}_n$ is of order n and that B_2^n is a center of $\mathcal{M}_n$ (in the sense that $\mathcal{M}_n$ is of diameter n and all convex bodies are at distance at most $\sqrt{n}$ from B_2^n). The question of (approximate) uniqueness of the center of $\mathcal{M}_n$ was raised by Pelczynski [109] i.e. if H is such that $d(H, K) \lesssim \sqrt{n}$ for all $K \in \mathcal{M}_n$, would that imply that H is somehow close to B_2^n? This was answered negatively by Szarek [132]. Thus, a natural problem is to look for other potential centers and to ask for instance if B_1^n is a center of $\mathcal{M}_n$. It was shown in [132] and further strengthened by Tikhomirov [138] that this is not the case. Nevertheless, the question which was left open is to estimate the distance between any convex body and B_1^n. Following the results of Bourgain and Szarek [34], Szarek and Talagrand [133], Giannopoulos [65], we obtained in [Y1] the estimate $d(K, B_1^n) \leq 2n^{5/6}$ for any $K \in \mathcal{M}_n$, recovering the dependence on the dimension from [65]. It was shown in [138] that, up to a logarithmic term, the quantity $\sup_{K \in \mathcal{M}_n} d(K, B_1^n)$ is lower bounded by $n^{5/9}$, leaving a substantial gap to fill as to the right order of magnitude of this quantity. In [Y19], with the aim to fill in this gap, we introduced the following geometric parameter for any symmetric convex body K in $\mathbb{R}^n$ and any $m \leq n$

$$\gamma_{K,m} := \inf_{\{x_1,\ldots,x_m\} \subseteq \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^m \frac{\|x_i\|_K}{\operatorname{dist}\big(x_i, \operatorname{span}\{x_j : j \neq i\}\big)},$$

where "dist" refers to the Euclidean one. This parameter captures in a sense the presence of vectors in K which are well separated mimicking the structure of the ℓ_1-ball.

Theorem 2.3 ([Y19]). *Let K be a symmetric convex body in $\mathbb{R}^n$ and $a, b > 0$ be such that $b\, B_2^n \subseteq K \subseteq a\, B_2^n$. Then*

$$d(K, B_1^n) \lesssim \frac{a}{b} \min_{m \in \{1,\dots,n\}} \sqrt{\max(n-m, bm\gamma_{K,m})} \cdot \sqrt{\log\Big(\frac{m}{\max(n-m, bm\gamma_{K,m})}\Big)}.$$

Although the statement is somehow complicated, one can apply a linear transformation to K so that the minimal volume ellipsoid containing it is B_2^n [79] and get that $a = 1, b = n^{-1/2}$ and

$$1 \leq \gamma_{K,m} \leq \sqrt{\frac{n}{n-m}}.$$

For bodies K where the right hand side inequality is attained, one recovers up to a logarithmic term the bound obtained in [65] and [Y1] while when the left hand side inequality is attained, the distance to B_1^n is bounded above by $n^{3/4}\sqrt{\log n}$. There are examples where this is indeed the case and Theorem 2.3 strictly improves on existing results.

2.2 Contact points of a convex body [Y10]

As was apparent in the previous section, some quantities of interest in convex geometry are invariant under linear transformations. Thus, it is useful to find suitable positions (i.e. transformations) of the convex body that make the corresponding problem tractable in a similar manner to what was discussed at the end of the previous section. One such important position is the one given by the ellipsoid of minimal volume containing the body. It was shown by John [79] that there exists a unique ellipsoid of minimal volume containing a given convex body K in $\mathbb{R}^n$. This ellipsoid is often referred to as John's ellipsoid and by applying a suitable linear transformation to the body K, one can suppose that John's ellipsoid is the standard Euclidean ball. This position plays a central role in convex geometry due to the following characterization due to John [79]. Given a convex body K in $\mathbb{R}^n$, B_2^n is it's John's ellipsoid if and only if there exists $m \leq n(n+3)/2$ points $x_1, \dots, x_m$ with $\|x_i\|_K = \|x_i\|_2 = 1$ and positive scalars $c_1, \dots, c_m$ such that

$$Id = \sum_{i=1}^m c_i x_i x_i^t \quad \text{and} \quad \sum_{i=1}^m c_i x_i = 0.$$

The points x_i's above are called *contact points* of K with its John's ellipsoid. Let us note that the above characterization is valid for symmetric and non-symmetric convex bodies. Moreover, the balancing condition $\sum_{i=1}^m c_i x_i$ can be trivially omitted in the symmetric case (by adding the points $-x_i$ which are still contact points). Contact points play an important role in convex geometry and have been successfully used to understand volume ratios [18, 19] and distances between convex bodies. Reducing the number of contact points is of interest as it allows for example to obtain non-trivial approximations of the convex body by polytopes

with few vertices. In view of John's characterization, the number of contact points cannot be smaller than the dimension. This motivates the problem of finding for any convex body K and any $\varepsilon \in (0,1)$, a convex body H such that $d(K,H) \leq 1+\varepsilon$ and H has at most $C(\varepsilon)n$ contact points.

This question was studied by Rudelson [117, 118] who partially answered the above question by finding an approximating body H with $C(\varepsilon)n \log n$ contact points. In view of John's decomposition, sparsification results such as Corollary 1.2 or the results of [22], provide an essential tool to reduce the number of contact points. Since the balancing condition plays no role in the symmetric case, this allowed Srivastava [130] to solve the problem for symmetric convex bodies. When K is a non-symmetric convex body, Srivastava [130] shows the existence of a convex body H with $O(n)$ contact points such that $d(K,H) \leq 2.24$. In [Y10], we developed a technique to deal with constraint sparsification problems allowing us to prove the following.

Theorem 2.4 ([Y10])**.** *Let K be a convex body in $\mathbb{R}^n$. For any $\varepsilon \in (0,1)$, there exists a convex body H with at most Cn/ε^2 contact points such that $d(K,H) \leq 1+\varepsilon$.*

2.3 Convex hull of high dimensional random walks [Y5]

In this section, we investigate convexity aspects of random walks in high dimensions. The convex hull of a random walk is a geometric characteristic of the walk which has been widely studied (see for instance [45, 116, 126] for surveys on the subject). While it has been extensively studied in the case of planar walks, the literature lacks results in high dimensions. Recently in [52, 81], some aspects of the convex hull of random walks in high dimensions were studied.

We were interested in the following question raised by Benjamini and first considered by Eldan [53].

Problem 2.3.1. *Given a discrete-time random walk $W(i)$ with values in $\mathbb{R}^n$, how many steps N are needed until the origin becomes an interior point of the convex hull of $\{W(i)\}_{i\leq N}$?*

The above can be seen as a multi-dimensional analogue of the persistence problem which consists of studying the probability that a one dimensional random walk stays positive. Different models of random walks can be studied, the standard random walk on $\mathbb{Z}^n$ being a natural one to consider. We can also study the standard Brownian motion BM_n in $\mathbb{R}^n$ captured at discrete times.

Theorem 2.5 ([Y5])**.** *There exists a universal constant C such that the following holds. Let $n \in \mathbb{N}$ and $N \geq e^{Cn}$. Then with probability at least $1-e^{-n}$, the convex hull of $\{\mathrm{BM}_n(i/N), i \leq N\}$ (resp. the standard random walk on $\mathbb{Z}^n$) contains the origin in its interior.*

The above result improves on an earlier result of Eldan [53] by removing an extra logarithmic factor in the exponent. Moreover, Theorem 2.5 answers Problem 2.3.1 for the two models considered (see [Y5] for another model of a random walk on the sphere). Indeed, we show that in the case of the Brownian motion (see [53] for the standard random walk on

$\mathbb{Z}^n$), Theorem 2.5 is optimal as if $N \leq e^{cn}$ for some small universal constant, then with high probability the origin won't belong to the interior of the convex hull of the walk. We should note that since then, some of our results were superseded by those of Kabluchko, Vysotsky and Zaporozhets [80] who completely characterize the probability that the origin belongs to the convex hull of a random walk with i.i.d symmetric increments.

Perhaps the main contribution hidden behind Theorem 2.5 is the interconnection between random walks, random matrix theory and high-dimensional convex geometry. Indeed, we connect Problem 2.3.1 to that of estimating the probability that the image of certain random matrices (whose rows are the increments of the random walk) does not intersect with a given subset of the unit sphere. The study of random sections of convex sets is a central theme in the area of geometric functional analysis and its importance has been highlighted in Milman's proof of Dvoretzky's theorem which we discussed at the beginning of this chapter.

Let us state a simplified version of the result we obtained (see [Y5] and Theorem D there for a more general statement).

Theorem 2.6 ([Y5]). *For any $K > 1$, $\varepsilon, B > 0$, there exist constants η, δ, L depending on K, ε, B such that the following holds. Let $N \geq Ln$ and A be an $N \times n$ random matrix whose rows R_i's are independent, centered, isotropic i.e. $\mathbb{E}\, R_i R_i^t = Id$ and satisfy the moment assumption*

$$\mathbb{E}\langle R_i, x\rangle^{2+\varepsilon} \leq B,$$

for any $i \leq N$. Then, for any $N \times N$ matrix F satisfying $\|F - Id\| \leq \eta$, we have

$$\mathbb{P}\Big(\mathrm{Im}(A) \cap F(\mathbb{R}_+^N) \neq \emptyset\Big) \leq e^{-\delta N} + \mathbb{P}\Big(\|A\| > K\sqrt{N}\Big).$$

The above statement reminds the so-called escape through a mesh theorem due to Gordon [71] which estimates the probability that the image of the standard Gaussian matrix intersects a subset S of the sphere in terms of the Gaussian width of S. On the contrast, the above theorem considers the image of a wide range of random matrices and study the intersection with a special type of convex cones given as some well conditioned image of $\mathbb{R}_+^N$. Questions of this type have attracted considerable attention in recent years due to its connection to the theory of compressed sensing [41].

2.4 Perspectives

Despite our modest advances, estimating the Banach-Mazur distance to the cross polytope remains open. We conjecture the following bound.

Conjecture 2.4.1. *For any symmetric convex body K in $\mathbb{R}^n$, we have $d(K, B_1^n) \lesssim n^{3/4}$.*

Our motivation for such a prediction comes from Theorem 2.2 where the dependence on ε, if improved to $\sqrt{\varepsilon}$, would lead to the above conjecture. Moreover, such a dependence, if true, would be optimal as was shown in [117]. Solving the above conjecture remains a very delicate task as estimating the distance to the cross polytope is dimension sensitive in the following sense. Given a symmetric convex body K in $\mathbb{R}^n$, the motivation is to find a

polytope with few vertices which is not far from K. If one aims to have a polytope with exactly $2n$ vertices, then finding the distance to it is still not known. However, if one is willing to give up on the exact number of vertices and allow $(2+\varepsilon)n$ ones, then it can be shown that there exists such a polytope with distance of order $\sqrt{n}$ to the body. In the language of embeddings, one can see the question of estimating the distance to the cross polytope as the problem of embedding an n-dimensional Banach space $(X, \|\cdot\|_X)$ into ℓ_1^n and estimating the embedding distortion. The difficulty of this problem lies into embedding in exactly the same dimension, while if one increase the dimension to $(1+\varepsilon)n$ then one can achieve the embedding with distortion of order $\sqrt{n}$ which is optimal as can be seen by taking X to be the Euclidean space. In a work in preparation, we are able to strengthen this and show that any n-dimensional Banach space X can be complemented by a "small" Euclidean space so that it embeds into ℓ_1 of the corresponding dimension with the best possible distortion. The proof also makes use of the sparsification results discussed in the previous chapter.

Theorem 2.7. *Let $(X, \|\cdot\|_X)$ be an n-dimensional Banach space. For any $\lambda > 1$, we have*

$$d\big(X \oplus_1 \ell_2^{(\lambda-1)n}, \ell_1^{\lambda n}\big) \leq 2\,\frac{\sqrt{\lambda}+1}{\sqrt{\lambda}-1}\,\sqrt{\lambda n}.$$

In particular, if $\varepsilon \in (0,1)$, then $d\big(X \oplus_1 \ell_2^{\varepsilon n}, \ell_1^{(1+\varepsilon)n}\big) \lesssim \sqrt{n}/\varepsilon$.

Chapter 3

Norm and outliers of random matrices

While we had a glance at Random Matrix Theory at the end of the previous chapter, here we officially enter this field. Spectral analysis of large random matrices is a very active area of research motivated by questions in statistics, mathematical physics, computer science. In this chapter, we restrict our attention to random matrices with independent entries before dropping this assumption in the next chapters. We refer to [99, 12, 34, 107, 55] for an introduction to random matrix theory.

3.1 Outliers in the sparse semi-circular law [Y18]

In its essence, random matrix theory aims at understanding the behavior of the eigenvalues of a large array of random data. This line of research has led to an enormous amount of works resulting in a very detailed analysis of the properties of the spectrum. A quantity of particular interest is the empirical spectral distribution which provides information on how the spectrum is distributed. Given an $n \times n$ symmetric matrix A, its *empirical spectral distribution* is a measure on $\mathbb{R}$ defined by

$$\mu_A := \frac{1}{n} \sum_{j=1}^{n} \delta_{\lambda_j(A)},$$

where $\lambda_j(A)$ denote the eigenvalues of A. When applied to an interval, this measure counts the average number of eigenvalues of A landing in this interval, providing therefore information on the location of the spectrum of A on the real line.

One of the celebrated results in this area is due to Wigner [150] and states that whenever $(\Xi_n)_{n\geq 1}$ is a sequence of $n\times n$ symmetric matrices whose entries on and above the diagonal are i.i.d centered random variables with unit variance, then the sequence of (random) measures $\mu_{\frac{1}{\sqrt{n}}\Xi_n}$ converges[1] almost surely to what is called the *Wigner semi-circular* distribution μ_{sc} with density $\frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{[-2,2]}(x)$. One of the amazing features of this result is its universality as the distribution of the matrix entries plays absolutely no role as long as they are centered

[1]By the convergence of measures, we mean *weak convergence* with continuous bounded test functions.

and of unit variance. Moreover, by considering the support of the limiting measure, it follows that asymptotically almost surely

$$\|\frac{1}{\sqrt{n}}\Xi\| \geq 2 - o(1),$$

where $\|\cdot\|$ stands for the spectral norm and $o(1)$ designates a term vanishing to 0 as $n \to \infty$. It turns out that whenever the entries of the matrix have finite fourth moment, then the extreme eigenvalues converge to the edge of the support of the limiting measure [16, 63, 64]. This determines the location of eigenvalues on a macroscopic scale and show that under the fourth moment assumption, one has asymptotically almost surely $\|\frac{1}{\sqrt{n}}\Xi\| \leq 2 + o(1)$ and there are no outlier eigenvalues detached from the support of the limiting measure.

While the global statistics of the eigenvalues are stable under finite rank additive perturbations (due to Weyl's inequalities), the behavior of the extreme eigenvalues can be severely affected. Understanding the appearance or not of outliers under the effect of perturbation is a natural problem as it is a way to study a "signal plus noise" model. Studying the spectrum of perturbed random matrices has been subject of an enormous number of articles. One is interested in studying the spectrum of $M_n + H_n$, where M_n is an $n \times n$ Wigner matrix while H_n is a fixed deterministic symmetric perturbation. When H_n is of finite rank, the limiting spectral distribution of $M_n + H_n$ remains the semi-circular distribution due to the interlacing property of eigenvalues. However, this perturbation can affect the extreme eigenvalues causing some of these to get detached from the rest of the spectrum. This has been first considered in [61] where the authors, motivated by estimating the largest eigenvalue of the adjacency matrix of an Erdős-Renyi graph, consider rank one deformations of a Wigner matrix. Later on, the interest in deformed random matrices took off with the work of [17] where the famous Baik, Ben Arous and Péché "BBP" phase transition was put forward. Later on, an enormous number of articles were devoted to investigate this phase transition in a variety of models as well as to the study of the fluctuations of eigenvalues separating or not from the bulk. We refer, among others, to [108] for a review on the subject and we state below an instance of the BBP phase transition.

Theorem 3.1. *[108, Theorem 2.1] Let M_n be an $n \times n$ Wigner matrix i.e. with entries on and above the diagonal i.i.d copies of $\frac{1}{\sqrt{n}}\xi$ where ξ is centered with unit variance. Suppose further that ξ has a finite fourth moment. Fix $r \in \mathbb{N}$ and an $n \times n$ deterministic symmetric matrix H_n of rank r with non-zero eigenvalues $\theta_1 \geq \ldots, \theta_r > 0$. Then for any $1 \leq i \leq r$,*

- *If $\theta_i \leq 1$, then $\lambda_i \xrightarrow[n\to\infty]{a.s.} 2$;*
- *If $\theta_i > 1$, then $\lambda_i \xrightarrow[n\to\infty]{a.s.} \theta_i + \frac{1}{\theta_i}$,*

where $\lambda_1 \geq \ldots \geq \lambda_n$ denote the eigenvalues of $M_n + H_n$.

In [Y18], we were interested in establishing a "sparsity counterpart" to the above phenomenon. More precisely, we interpret the sparsity as a perturbation and study its effect on the extreme eigenvalues of a Wigner matrix. Let $\Xi_n = (\xi_{ij})_{1\leq i,j\leq n}$ be an $n \times n$ symmetric

random matrix whose entries on and above the diagonal are i.i.d copies of ξ which is centered, of unit variance, and uniformly bounded (in particular, ξ has a finite fourth moment and thus Ξ_n has no outliers). Let $B_n = (b_{ij})_{1\leq i,j\leq n}$ be an $n \times n$ symmetric matrix whose entries on and above the diagonal are i.i.d Bernoulli variables with probability of success p_n (less than $1/2$ say) and suppose that B_n and Ξ_n are independent. We consider the random matrix W_n obtained as the entry-wise product of B_n and Ξ_n normalized by $\sqrt{np_n}$. In other words, W_n is obtained by deciding randomly and independently to zero-out entries of Ξ_n. One of the nice features of Wigner semi-circular law is its stability under sparsification as long as the average number of non-zero entries in each row is large. More precisely, as long as $np_n \to \infty$, we have $\mu_{W_n} \xrightarrow[n\to\infty]{a.s.} \mu_{sc}$. While the limiting law is stable under this operation, it is not clear at first glance if the extreme eigenvalues are rigid and would still stick to the rest of the spectrum. Checking the fourth moment, we have

$$\mathbb{E}\Big|\frac{b_{ij}\xi_{ij}}{\sqrt{p_n(1-p_n)}}\Big|^4 = \frac{\mathbb{E}\,|\xi_{ij}|^4}{p_n(1-p_n)^2},$$

which is clearly finite whenever p_n is fixed and diverges as $p_n \to 0$. Therefore, if one zeros out on average a fixed fraction of the entries, then there are still no outliers to the semi-circular law while when $p_n \to 0$, the fourth moment test is inconclusive. This leads us to several interesting questions regarding the above observations:

1. *Is there a phase transition (in terms of sparsity) in the appearance of outliers in the semi-circular law?*

2. *In the affirmative, can we obtain a quantitative answer as to capture the sparsity condition?*

3. *Would it be possible to obtain a qualitative explanation as to understand the reason for the appearance of outliers?*

4. *Finally, can we capture the exact asymptotic value of the outlier?*

In [Y18], we addressed these questions by proving the following.

Theorem 3.2 ([Y18])**.** *Let ξ be a real centered uniformly bounded random variable of unit variance. For each n, let W_n be $n \times n$ symmetric random matrix with i.i.d. entries, with each entry equidistributed with the product $\frac{1}{\sqrt{np_n}} b_n \xi$, where b_n is $0/1$ (Bernoulli) random variable independent of ξ, with probability of success equal to p_n. Assume further that $np_n \to \infty$ with n. For each n, define the quantity*

$$\theta_n := \sqrt{\max\big(\mathbb{E}\max_{i\leq n} \|\mathrm{row}_i(W_n)\|_2^2 - 1, 1\big)}.$$

Then the sequence $\big(\frac{\|W_n\|}{\theta_n + \theta_n^{-1}}\big)_{n\geq 1}$ converges to one in probability. In particular,

- *If* $\limsup_n \theta_n \leq 1$, *then* $\|W_n\| \xrightarrow[n\to\infty]{\mathbb{P}} 2$;
- *If* $\liminf_n \theta_n > 1$, *then* $\frac{\|W_n\|}{\theta_n+\theta_n^{-1}} \xrightarrow[n\to\infty]{\mathbb{P}} 1$.

We refer to [Y18] where a more general version is stated (treating the k-th largest eigenvalue) and a non-centered counterpart to the result is established. The condition on θ_n implicitly encapsulates a sparsity condition. Indeed, for example whenever $\frac{np_n}{\log n} \to \infty$, standard concentration inequalities show that $\limsup_n \theta_n \leq 1$ and there are no outliers in the sparse semi-circular law. In Chapter 5, we further illustrate this explicitly on an important particular case given by the Erdős-Renyi graphs. In addition to being quantitative, our result also provides a qualitative answer as to the exact reason of appearance of outliers, that is, when the matrix has a row whose Euclidean norm squared exceeds twice the average norm. In such a case, W_n has an outlier and its value is asymptotically given by $\theta_n + \theta_n^{-1}$.

Let us give the following non-rigorous explanation of the apparent resemblance between Theorem 3.2 and Theorem 3.1. Note that due to concentration inequalities, most of the rows of W_n have their norms squared concentrated around its mean equal to 1. Only a small fraction of these rows can have their norm far from the mean. For simplicity, suppose that only one row (say, the first one after suitable permutation) has its Euclidean norm signicantly larger than 1. Thus, this suggests that we may decompose our model as

$$W_n \approx M_n + H_n,$$

where M_n is obtained from W_n by a regularization procedure consisting of reducing the entries of the first row (the one with the largest norm) as to make its Euclidean norm equal to 1, and H_n is the $n \times n$ symmetric zero diagonal matrix whose first row/column's Euclidean norm is approximately given by $u_n = \sqrt{\max_{i\leq n} \|\mathrm{row}_i(W_n)\|_2^2 - 1}$. Conceptually, we interpret the excess Euclidean norm in the first row which was transferred to H_n as a deformation of the regularized matrix M_n. Clearly, H_n is of rank 2 with eigenvalues $\pm u_n$. On the other hand, all rows of M_n have their norms concentrated around the mean which could suggest that M_n has no outliers. Now, Theorem 3.2 states that if $u_n \leq 1$, then $M_n + H_n$ has all its eigenvalues asymptotically bounded by 2; while if $\liminf_n u_n > 1$, then $M_n + H_n$ has an outlier and its value is given by $u_n + \frac{1}{u_n}$. This parallels the suggested behavior in the BBP phase transition result. Let us emphasize that the above comparison is only on a conceptual level and we do not provide any rigorous link between the two. Nevertheless, it would be very interesting if such a relation could be elaborated. For instance, there has been recently several interesting works concerned with the regularization of random graphs/matrices i.e. procedures to reduce the norm of random matrices by changing few of its entries (see [56, 87, 121, 114]).

To end this section, let us record the following intriguing consequence of Theorem 3.2. For any $\varepsilon > 0$, we have

$$\max_{i\leq n} \|\mathrm{row}_i(W_n)\|_2 \leq \|W_n\| \leq (2+\varepsilon) \max_{i\leq n} \|\mathrm{row}_i(W_n)\|_2, \tag{3.1}$$

with probability going to 1 as n goes to infinity. The left hand side inequality is valid deterministically for any matrix since one always has $\|A\|_{\ell_1\to\ell_2} \leq \|A\|_{\ell_2\to\ell_2}$ for any linear

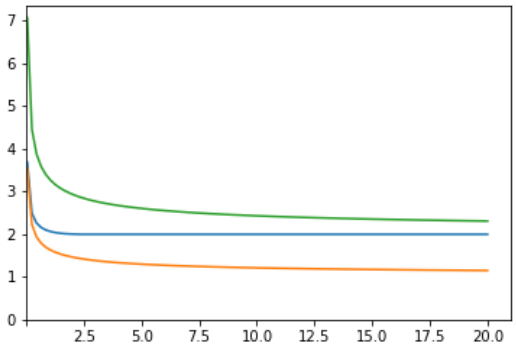


Figure 3.1: *This figure illustrates the asymptotic relation between the maximum row norm and the operator norm of W_n, a centered symmetric Bernoulli matrix of parameter p_n, with $\lim_{n\to\infty} np_n = \infty$ and $\lim_{n\to\infty} p_n = 0$. The horizontal axis is the value of the limit $\lim_n \frac{np_n}{\log n}$. The blue curve is the corresponding values of $\lim_n \|W_n\|$. The orange curve is the values of $\lim_n \max_{i\le n} \|\mathrm{row}_i(W_n)\|_2$. The green curve — the values of $\lim_n 2\max_{i\le n} \|\mathrm{row}_i(W_n)\|_2$. When $\frac{np_n}{\log n} \to 0$, the left side of inequality* (3.1) *is asymptotically sharp while for $\frac{np_n}{\log n} \to \infty$, the right side of* (3.1) *is sharp. This is an illustration of the effect of sparsity on the operator norm.*

operator A. While the reverse inequality is far from being true for a given matrix, the above consequence of Theorem 3.2 asserts that this is the case for our sparse Wigner matrix up to an asymptotically optimal constant equal to 2. We illustrate this inequality below in the case of symmetric centered Bernoulli matrices. One may wonder how general this phenomenon is and if it is some magical feature of random matrices. This will be investigated in the next section.

3.2 The operator norm of a random matrix [Y15]

We ended the last section wondering if the phenomenon captured in (3.1) is more general. Indeed, it was shown by Seginer [127] that random matrices X with i.i.d centered entries exhibit this phenomenon i.e.

$$\mathbb{E}\,\|X\| \approx \mathbb{E}\,\max_i \|\mathrm{row}_i(X)\|_2,$$

where the notation $a \approx b$ means $a \lesssim b$ and $b \lesssim a$. In the previous section, we recovered this observation for bounded entries as a byproduct of Theorem 3.2 while capturing the optimal constant 2 in the above comparison. It turns out that the distribution of the entries plays no role in Seginer's result but rather it is the invariance by permutation which is exploited by means of combinatorial methods.

In view of such a remarkably general probabilistic principle, it is natural to ask whether the same conclusion also extends to nonhomogeneous matrix models that do not possess

permutation symmetry. However, Seginer [127] already showed that this is not the case by considering the example of a block diagonal matrix with blocks of suitable size filled with independent Rademacher random variables. While it is customary in random matrix theory that some results for Gaussian matrices are valid for subgaussian ones, no counterexamples for the above phenomenon could be found for Gaussian matrices. This led to the following conjecture made by Latała 15 years ago.

Conjecture 3.2.1 (Latała [86]). *Let X be an $n \times n$ symmetric matrix whose entries on and above the diagonal are independent centered Gaussian random variables. Then*

$$\mathbb{E}\,\|X\| \approx \mathbb{E}\,\max_i \|\mathrm{row}_i(X)\|_2.$$

In [Y15], we solve this conjecture in a much more general way. Indeed, let us define the mixed norm

$$\|X\|_{\ell_p(\ell_2)} := \Big(\sum_i \Big(\sum_j X_{ij}^2\Big)^{p/2}\Big),$$

and analogously when $p = \infty$, we get $\|X\|_{\ell_\infty(\ell_2)} = \max_i \|\mathrm{row}_i(X)\|_2$.

Theorem 3.3 ([Y15]). *Let X be an $n \times n$ symmetric matrix whose entries on and above the diagonal are independent centered Gaussian random variables. Then for any $2 \le p \le \infty$*

$$\mathbb{E}\,\|X\|_{S_p} \approx \mathbb{E}\,\|X\|_{\ell_p(\ell_2)}.$$

The case $p = \infty$ of the above result solves Latała's conjecture. Theorem 3.3 is also true for a wider class of random matrices. For instance, we show its validity for a broad class of heavy-tailed entry distribution. Moreover, the phenomenon extends for rectangular matrices up to the obvious addition of the maximum Euclidean norm of the columns as well.

The striking observation made in [Y15] is that $\|X\|_{S_p}$ and $\|X\|_{\ell_p(\ell_2)}$ are even more intimately related as we show that the distributions of these random variables are comparable in a very strong sense.

Theorem 3.4 ([Y15]). *Let X be an $n \times n$ symmetric matrix whose entries on and above the diagonal are independent centered Gaussian random variables. Then*

$$\mathbb{P}\big(\|X\|_{\ell_p(\ell_2)} \ge t\big) \le \mathbb{P}\big(\|X\|_{S_p} \ge t\big) \le C\,\mathbb{P}\big(\|X\|_{\ell_p(\ell_2)} \ge t/C\big),$$

for any $t \ge 0$ and $2 \le p \le \infty$, where C is a universal constant.

It is rather fascinating to have such a strong connection between these two quantities. The case $p = \infty$ on its own is already very meaningful: it is trivial that the operator norm of a matrix must be large if the matrix possesses a row with large Euclidean norm; what we have shown is that for symmetric Gaussian matrices with independent centered entries, this is the *only* reason why the operator norm can be large, regardless of the variance pattern of the matrix entries.

The above observation provides a qualitative understanding of the operator norm of a Gaussian matrix. This is made possible thanks to matching quantitative estimates obtained for both quantities. In [Y15], we obtain the following explicit expression of the p-Schatten norm and operator norm in terms of the variances of the matrix entries.

Theorem 3.5 ([Y15]). *Let X be an $n \times n$ symmetric matrix with $X_{ij} = b_{ij} g_{ij}$, where $b_{ij} \geq 0$ and g_{ij} are i.i.d. standard Gaussian variables for $i \geq j$. Then*

$$\mathbb{E}\,\|X\|_{S_p} \asymp \left(\sum_i \left(\sum_j b_{ij}^2 \right)^{\frac{p}{2}} \right)^{\frac{1}{p}} + \max_{i \leq e^p} \max_j b_{ij}^* \sqrt{\log i} + \sqrt{p} \left(\sum_{i \geq e^p} \max_j b_{ij}^{*\,p} \right)^{\frac{1}{p}}$$

for all $2 \leq p < \infty$, and

$$\mathbb{E}\,\|X\|_{S_\infty} \asymp \max_i \sqrt{\sum_j b_{ij}^2} + \max_{ij} b_{ij}^* \sqrt{\log i}$$

for $p = \infty$. Here the matrix (b_{ij}^*) is obtained by permuting the rows and columns of the matrix (b_{ij}) such that $\max_j b_{1j}^* \geq \max_j b_{2j}^* \geq \cdots \geq \max_j b_{nj}^*$, and the constants in the estimates are universal (independent of $n, p, \{b_{ij}\}$).

It is remarkable that the above expressions hold with absolutely no restriction on the variance pattern. Moreover, the dimension of the matrix plays absolutely no role in this expression. This offers a lot of flexibility in applying the result and provide a way to tackle natural questions which were previously out of reach. For instance, one could ask under what conditions an infinite random matrix defines a bounded operator on ℓ_2. It is readily verified that such matrices could never have identically distributed entries; to obtain meaningful answers to such infinite-dimensional questions, it is therefore essential to consider non-homogeneous random matrix models. As a consequence of Theorem3.5, we obtain a characterization of all infinite matrices with independent Gaussian entries that define bounded operators on ℓ_2.

Corollary 3.6 ([Y15]). *Let $(X_{ij})_{i,j \in \mathbb{N}}$ be a symmetric infinite matrix with independent Gaussian entries $X_{ij} \sim N(a_{ij}, b_{ij}^2)$ for $i \geq j$. We have the following dichotomy:*

- *If*

$$\max_i \sum_j b_{ij}^2 < \infty, \qquad \max_{ij} b_{ij}^* \sqrt{\log i} < \infty, \qquad \|(a_{ij})\|_{S_\infty} < \infty,$$

  then X defines a bounded operator on $\ell_2(\mathbb{N})$ a.s.

- *Otherwise, X is unbounded as an operator on $\ell_2(\mathbb{N})$ a.s.*

Interestingly, the proof of these results provide considerable insight into what random matrices with bounded operator norm look like: they must be nearly block-diagonal. Indeed, we build on initial progress made by Bandeira and van Handel [20] as well as [143]. Using permutation of rows and columns, we put in place a special decomposition of the matrix into blocks on which we proceed using different techniques. More precisely, we show that the random matrix X can have a bounded operator norm only if it has a specific form (modulo a relabeling of the rows and columns): it consists of a nearly block-diagonal "core", which is made of blocks of controlled dimension in which the variance of the entries decay at a slow rate; and an off-diagonal remainder, in which the entry variances decay at a much faster

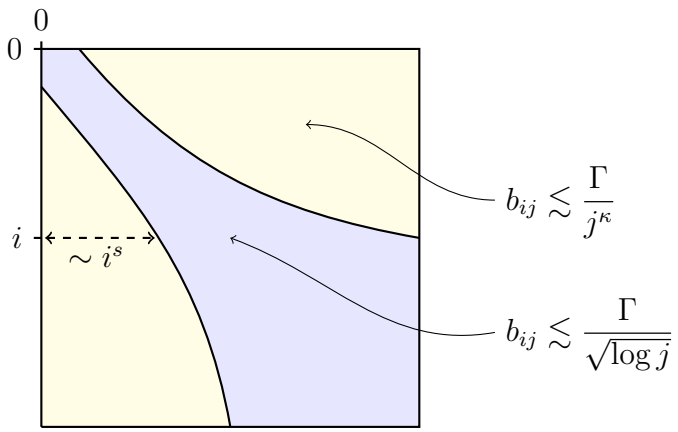


Figure 3.2: After permutation of rows/columns, this is the structure of bounded operators with independent Gaussian entries. A "core" where the variances decay slowly, while their decay is much faster outside. The width of the core increases at a polynomial speed. In [Y15], we take $s = \frac{1}{3}$, $\kappa = \frac{1}{4}$ and $\Gamma = \max_i \sqrt{\sum_j b_{ij}^2} + \max_{ij} b_{ij}^* \sqrt{\log i}$ in the above quantities.

rate (see Figure 3.2 for a rough illustration). On the part of the matrix with controlled dimension, we extend a dimension-compression argument developed in [20] and establish a comparison principle between the moments $\mathbb{E}\,\|X\|_{S_p}^p$ and that of a smaller Wigner matrix of appropriate dimension. To complement this analysis, a method based on the geometry of Gaussian processes [143] is used to deal with part of the matrix where the dimension is not controlled while the variances decay at a fast rate.

3.3 Perspectives

In the previous section, we discussed the solution to the characterization of the norm of a matrix with independent Gaussian entries. As we mentioned, the proof relies on a special decomposition and a combination of the moment method together with suprema of Gaussian processes. The former method is very much spectral as it relates the norm of the matrix to its trace, and is therefore limited to deal with spectral norms. On the other hand, dealing with suprema of Gaussian processes offers a lot of flexibility as such method readily extends to more general operator norms by considering suprema over unit balls of other norms than the Euclidean one. It would be very interesting to elaborate a proof of the previous results using only this method with the aim extending these to general norms. Suboptimal results going in this direction were obtained in [72].

The original BBP famous phase transition was concerned with sample covariance matrices. The limiting spectral distribution in this context is the Marcenko-Pastur law. What would be the effect of sparsity on this model? In a similar manner, the Marcenko-Pastur is stable under sparsification procedure which however may create outliers depending on its magnitude. As in the Wigner case, when the sparsity level is very high, this should create a

very much spiked population model and outliers should emerge in the limiting distribution. It would be interesting to study this model and eventually capture a similar phenomenon as in the information-plus-noise type model. We expect that a similar phenomenon to the Wigner case studied in this chapter can be elaborated for these models.

Chapter 4

Random regular matrices and digraphs

The universality phenomenon is one of the fascinating sides of random matrix theory. In its essence, it asserts that the occurrence of a particular phenomenon is independent in a sense of the distribution of the matrix entries. We started the previous chapter with one such instance of universality through the Wigner semi-circular distribution, and ended the chapter with a non-universal phenomenon valid only for some class of random matrices. In this chapter, we investigate the universality of the limiting spectral distribution in the non-symmetric case. While the spectrum of Hermitian matrices enjoy nice properties such as the interlacing phenomenon, its non-Hermitian counterpart lacks these making the study of the empirical spectral distribution supported in $\mathbb{C}$ much more delicate. Let us emphasize that our focus in this chapter will be on global statistics of eigenvalues and in the next chapter we investigate the extreme eigenvalues at the edge of the support of the limiting spectral distribution. The literature contains an enormous amount of results concerned with the universality of local statistics of eigenvalues which are not considered in this manuscript, we refer among others to [55] for references on this.

The universal object of interest in this chapter is the circular law i.e. the uniform measure on the disk in $\mathbb{C}$ which we denote by μ_{circ}. The circular law theorem asserts that whenever $(X_n)_{n\geq 1}$ is a sequence of $n \times n$ matrices whose entries on and above the diagonal are i.i.d centered random variables with unit variance, then the sequence of random measures $\mu_{\frac{1}{\sqrt{n}}X_n}$ converges almost surely to the circular law. In parallel to its symmetric counterpart encountered in the previous chapter, this result is a further illustration of universality as the distribution of the matrix entries play absolutely no role (beyond the first two moments). The circular law theorem is the culmination of efforts of several mathematicians producing partial results and introducing crucial ideas which led to the general case finally obtained by Tao and Vu [135]. We refer to the survey [31] for a detailed account of the history and references on this result.

One may wonder to what extend the circular law is universal. For instance, it is not known if the validity of the circular law for a sequence of random matrices with i.i.d entries imply the finiteness of their second moment (see [29] for heavy-tailed entries). In another

direction, can the independence assumption be dropped somehow? Is the circular law stable under sparsification? This line of research has attracted a lot of attention in the last ten years or so, resulting in several results aiming at extending the universality class of the Circular law. For instance, the circular law theorem was shown to hold under sparsification [136, 70, 151, 21, 120]. Models with dependent entries were considered, to name a few: random Markov matrices [30], random matrices with i.i.d log-concave rows [3], random sign matrices with i.i.d rows of given sum [106], random doubly stochastic matrices with the uniform distribution on the Birkhoff polytope [104], random matrices with exchangeable entries [4], etc. Our aim in this chapter is to further enlarge the universality class of the circular law by investigating a model encapsulating both dependence and sparsity. This will be provided by random regular graphs which we introduce next.

4.1 The limiting spectral distribution of random regular digraphs [Y17]

Let $d, n \in \mathbb{N}$. A directed graph G (*digraph*) on $[n]$ is called *d-regular* if every vertex has exactly d in-neighbors and d out-neighbors. A digraph can be encoded in a matrix $A = (a_{ij})_{1\leq i,j\leq n}$ called *the adjacency matrix* of the graph as follows

$$a_{ij} = 1 \text{ if there is an edge from } i \text{ to } j; \quad \text{and } a_{ij} = 0 \text{ otherwise.}$$

It turns out that expansion properties of the graph are intimately related to some properties of the spectrum of its adjacency matrix making the understanding of the latter interesting on its own. One trivial observation which follows from d-regularity is that $A\mathbf{1} = d\mathbf{1}$, where we denoted by $\mathbf{1}$ the n dimensional vector with all entries equal to 1. Thus, this is an eigenvector associated to the eigenvalue d. Using Perron-Frobenius theorem, one can further deduce that $\|A\| = d$.

We are interested in the "typical" behavior of the spectrum of a d-regular graph. To this aim, we denote $\mathcal{M}_{n,d}$ the set of adjacency matrices of d-regular digraphs on n vertices, where we allow loops but do not allow multiple edges. We condiser $A_{n,d}$ the random matrix uniformly distributed on $\mathcal{M}_{n,d}$. In other words, $A_{n,d}$ is the adjacency matrix of the random digraph with the uniform distribution on the set of all d-regular graphs on n vertices, an object very much studied in the graph theoretical literature. Our starting point is the following well-known conjecture:

Conjecture 4.1.1 ([31])**.** *Fix $d \geq 3$ and let $A_{n,d}$ be as above. Then, as $n \to \infty$, $\mu_{A_{n,d}}$ converges in probability to $\mu_{KM}^{(d)}$, the oriented Kesten-Mckay distribution on $\mathbb{C}$ whose density with respect to the Lebesgue measure is given by*

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

The above conjecture parallels the undirected case, where the Kesten-Mckay distribution (a symmetric version of the above) was shown to be the limiting spectral distribution

of the uniform random regular undirected graph [97]. When $d \to \infty$ and after a proper normalization, the circular law is recovered. This naturally leads to the following conjecture:

Conjecture 4.1.2. *Let $A_{n,d}$ be as above. Then, as $d \to \infty$, $\mu_{\frac{1}{\sqrt{d}}A_{n,d}}$ converges in probability to μ_{circ}.*

The above conjecture parallels the undirected case, where the semi-circular distribution was shown to be the limiting spectral distribution (as $d \to \infty$) of the uniform random regular undirected graph [140]. The above conjecture was shown to hold when $d \gtrsim \log^{96} n$ by Cook [47]. Our first contribution in this chapter is the resolution of the above conjecture in the complementary challenging sparse regime.

Theorem 4.1 ([Y17])**.** *Fix a constant $C \geq 1$ and for any $n > 1$ let $d = d(n)$ be a positive integer satisfying $d \leq \log^C n$. Then, as $d \to \infty$, $\mu_{\frac{1}{\sqrt{d}}A_{n,d}}$ converges in probability to μ_{circ}.*

The above theorem is another remarkable instance of the universality of the circular law, as the random matrices involved can be very sparse and exhibit dependencies among their entries.

As in all previous works, a key element in the proof of the circular law for d-regular digraphs is to transport the problem of the limiting spectral distribution to the singular values distribution, which is much easier to study. This method – called the Hermitization technique – goes back to Girko [68] and exploits a close relation between the log-potential functions of the spectral and singular values distributions. Following Girko, this idea was used in various papers dealing with non-Hermitian random matrices, and we refer to [31] for a detailed exposition.

The *singular values distribution* of an $n \times n$ random matrix B is the random probability measure on $\mathbb{R}$ given by

$$\nu_B := \frac{1}{n} \sum_{i=1}^{n} \delta_{s_i},$$

where $(s_i)_{i \leq n}$ denote the singular values of B. The *logarithmic potential* $U_\mu : \mathbb{C} \to (-\infty, \infty]$ of a probability measure μ on $\mathbb{C}$ is defined for any $z \in \mathbb{C}$ by

$$U_\mu(z) := -\int_{\mathbb{C}} \log|z - \lambda| \, d\mu(\lambda).$$

The logarithmic potential function uniquely determines the underlying measure, that is, if $U_\mu = U_{\mu'}$ Lebesgue almost everywhere then $\mu = \mu'$ (see, in particular, [31, Lemma 4.1]).

Given an $n \times n$ matrix B, it is easy to check that

$$U_{\mu_B}(z) = -\frac{1}{n} \log|\det(B - zId)| = -\int_0^\infty \log(t) \, d\nu_{B - z\mathrm{Id}}(t).$$

Therefore, knowing $\nu_{B-z\mathrm{Id}}$ for almost all $z \in \mathbb{C}$, we can determine U_{μ_B}, hence μ_B itself. This observation lies at the heart of the method. Studying the convergence of the singular value distribution is a much easier problem as one has access to standard methods such as the

Stieltjes transform method, and the moment method. In view of the above relation, one would hope to deduce the convergence of the empirical spectral distribution from that of the singular value distribution. However, the logarithm has singularities at 0 and ∞ which would require special treatment. The singularity at ∞ is easily dealt with thanks to the explicit knowledge of the largest singular value of a d-regular matrix. Showing the uniform integrability of log at 0 represents a serious challenge and opens the door to interesting problems such as the invertibility of the adjacency matrix of random regular graphs, as well as obtaining quantitative bounds on the smallest singular value and on the growth of the smallish ones.

4.2 Invertibility of adjacency matrices of random regular digraphs [Y7, Y8, Y13, Y14]

As became clear from the previous section, obtaining lower bounds on the smallest singular value of the adjacency matrix of a random regular graph is a key ingredient in the proof of the circular law stated above. Moreover, singularity of random discrete square matrices is a subject interesting on its own with a long history, many results and applications. Dealing with discrete distributions in singularity problems carries considerably more challenges than continuous ones. In a standard setting, when the entries of the $n \times n$ matrix are i.i.d. Bernoulli ± 1 random variables, the invertibility problem has been addressed by Komlós in [83, 84], and later considered in several papers [82, 134, 38]. A long-standing conjecture asserts that the probability that the Bernoulli matrix is singular is $\big(1/2 + o(1)\big)^n$. This was very recently proved in an outstanding breakthrough by Tikhomirov [139].

In the absence of independence between the matrix entries, the problem of singularity involves additional difficulties. Such a problem was considered for the (symmetric) adjacency matrix M_n of a random (with respect to the uniform probability) undirected d-regular graph on n vertices. The case $d = 1$ corresponds to a permutation matrix which is non-singular, and for $d = 2$ the graph is a union of cycles and the matrix is almost surely singular. Moreover, the invertibility of the adjacency matrix of the complementary graph is equivalent to that of the original one (in fact, the ranks of the adjacency matrices of a d-regular graph and of its complementary graph are the same). The following conjecture was raised by Costello and Vu [50, Section 10], and again reiterated in the survey [148, Problem 8.4] as well as in 2014 ICM talks by Frieze [60, Problem 7] and by Vu [149, Conjecture 5.8].

Conjecture 4.2.1. *For every $3 \le d \le n-3$, we have*

$$\mathbb{P}\Big(M_n \text{ non-singular}\Big) \underset{n\to\infty}{\longrightarrow} 1.$$

It is natural to consider the non-symmetric counterpart to the above conjecture i.e. for $A_{n,d}$ the adjacency matrix of the uniform random d-regular digraph (see, in particular, [46, Conjecture 1.5]). Cook [46] proved that $A_{n,d}$ is asymptotically almost surely non-singular for $\omega(\log^2 n) \le d \le n - \omega(\log^2 n)$, where $f = f(n) = \omega(a_n)$ means $f/a_n \to \infty$ as $n \to \infty$. We were able to complement this result as to cover all ranges of $d \to \infty$.

Theorem 4.2 ([Y7,Y8]). *There are universal constants c, C such that if $C \leq d \leq cn/\log^2 n$, we have*

$$\mathbb{P}\Big(A_{n,d} \text{ non-singular}\Big) \geq 1 - \frac{C\log^3 d}{\sqrt{d}}.$$

The above theorem shows the validity of Conjecture 4.2.1 for $d \to \infty$ at any rate. Beside extending the range of d compared to [46], the probability bound obtained above improves on that of [46]. Moreover, as opposed to [46], Theorem 4.2 sheds the light on a significant difference between the adjacency matrix of the uniform random d-regular digraph and a Bernoulli matrix with parameter d/n. Indeed, while the latter is singular with probability close to 1 when $d \ll \log n$, our result asserts the invertibility of the uniform d-regular model, thus highlighting a way in which random regular graphs behave very differently to Erdős-Renyi graphs.

To illustrate how the special structure of the random regular graph is used, note that the presence of many common in/out neighbors between vertices comes at the detriment of a well invertibility. Indeed, the corresponding adjacency matrix will have rows/columns which almost coincide making it possible to architect a defective vector close to the kernel. While in the independent case, such scenarios are easy to analyse, this requires a careful treatment in such a dependent setting and is an instance of the expansion properties enjoyed by random regular graphs. Let us state one of several graph theoretical results we had to establish.

Theorem 4.3 ([Y7]). *Let $8 \leq d \leq n$ and $\mathcal{D}_{n,d}$ be the uniform random d-regular digraph on n vertices. For any $2 \leq k \leq cn/d$, we have*

$$\mathbb{P}\Big(|\mathcal{N}_G^{in}(S)| \geq d|S|/2 \text{ for any } S \subseteq [n] \text{ of size } k\Big) \geq 1 - \exp\Big(-ckd\log\big(cn/(kd)\big)\Big),$$

where c is a universal constant and $\mathcal{N}_G^{in}$ denotes the set of in-neighbors to S in $\mathcal{G}$.

We refer to [Y7] for a much more refined statement. Let us note that by d-regularity, one always has $|S| \leq |\mathcal{N}_G^{in}(S)| \leq d|S|$ and the above captures the extend of expansion in our model as the neighborhood of sets of size n/d tends to contain almost all vertices. While such results were known in the undirected case (see e.g. [9] and references therein), the literature lacks analogue statements for its directed counterpart.

In view of the problem raised in the previous section, obtaining quantitative invertibility is crucial in this context. Providing lower bounds on the smallest singular value of a random matrix is a subject interesting on its own. Beside its connection to the Hermitization technique discussed previously, we already saw in the first two chapters its importance for problems in numerical linear algebra and convex geometry. The literature contains an enormous amount of work concerned with estimating the smallest singular value of a random matrix, ranging from square matrices to rectangular ones, dealing with independent entries or independent rows/columns. To avoid an extensive list of references, we refer to [123, 145, 119, 146] for more on this subject. In [Y13], we were able to significantly strengthen Theorem 4.2 as to obtain a lower bound on the smallest singular value.

Theorem 4.4 ([Y13]). *There are universal constants c, C such that if $C \leq d \leq cn/\log^2 n$, we have*

$$\mathbb{P}\Big(s_{\min}(A_{n,d}) \geq n^{-6}\Big) \geq 1 - \frac{C\log^2 d}{\sqrt{d}}.$$

Let us note that the above statement also holds for complex shifts of $A_{n,d}$ as to fit its required use for the Hermitization technique. Moreover, the polynomial lower bound is sufficient for this purpose as we were interested in the behavior of the logarithm of the smallest singular value. It is natural to wonder what would be the optimal lower bound, and to this aim one could draw comparisons with standard Wigner matrices leading to expect a lower bound of order $\sqrt{d}/n$. It is not clear if such a bound would hold and the current available tools are incapable of reaching such a level of precision. The above advances serve their need towards establishing the circular law theorem, however leave untreated the constant d case of conjecture 4.2.1. In an effort to prove the conjecture, we were able to show the following.

Theorem 4.5 ([Y14]). *There exists a universal constant C such that for any fixed integer $d \geq C$ the following holds. Let $n > d$ and $A_{n,d}$ be the adjacency matrix of the uniform random d-regular digraph. Then*

$$\mathbb{P}\Big(\operatorname{rank}(A_{n,d}) \geq n - 1\Big) \geq 1 - C\frac{\log^2 d}{\log n}.$$

After our result was published, Conjecture 4.2.1 was solved independently in [78] and [105] where the key idea is to study the singularity over finite fields, which has the drawback of not providing quantitative information on the smallest singular value.

A fundamental role in the context of invertibility is played by what is nowadays called *the Littlewood-Offord theory* which has been elaborated over the years, leading to a much better understanding of interrelationship between the singularity probability and arithmetic structure of potential null vectors (see [137, 123] for more details). Elaborating a related concept in our setting is essential and understanding the structure of potential null vectors or eigenvectors in general is of particular importance. Indeed, such structural properties can be efficiently combined with *the simple switching*, a standard tool applied first in the context of regular graph by Mckay [98], in order to estimate event probabilities. The simple switching is an operation performed on a d-regular graph as to obtain from it a new d-regular graph. As an illustration, let G be a d-regular digraph on n vertices and let $i_1 \neq i_2$ and $j_1 \neq j_2$ be vertices of G such that (i_1, j_1) and (i_2, j_2) are edges of G while (i_1, j_2) and (i_2, j_1) are not. The simple switching illustrated in the Figure 4.1 consists of replacing the edges (i_1, j_1), (i_2, j_2) with (i_1, j_2) and (i_2, j_1). Note that the operation does not destroy d-regularity of the graph. As this operation allows to form each time a new d-regular graph, estimating the probability of an event boils down to studying its stability under simple switching. Informally, an event in our context has a large probability if starting from one particular realization satisfying it, many simple switchings can be performed while staying inside this event; on the other hand, starting from a realization outside this event, very few simple switchings can keep this realization outside of this event of large probability. This observation urges the need to

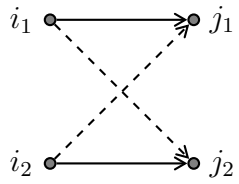


Figure 4.1: The simple switching.

understand where to operate switchings in the graph in such a way to capture singularity probabilities. This is where the structure of eigenvectors and of normal vectors to subspaces spanned by rows of $A_{n,d}$ enter into play. Indeed, given a localized vector (say $(1, 1, 0 \dots, 0)$), very few switchings on a realization of $A_{n,d}$ can change its action on this vector (namely, such switching should involve one of the first two vertices). On the contrast, when the vector is delocalized with some additional discrepancy among its coordinates, then many switching can be operated as to create a rich randomness in the action of the matrix on it. This naturally leads us to study localization/delocalization of eigenvectors of random regular graphs, which is the subject of the next section.

4.3 Delocalization of eigenvectors of random regular digraphs [Y16]

Delocalization properties of eigenvectors for various models have been a focus of active research, especially in the setting of Wigner (and generalized Wigner) matrices. The term *delocalization* usually refers to upper bounds on the ℓ_∞-norm of a random vector as to assert that no single coordinate of the eigenvector carries a significant mass. Another related concept called *no-gaps delocalization* was introduced in [125] and aim at capturing the uniformness of the eigenvectors in the sense that its ℓ_2 norm is more or less evenly spread over the coordinates. These notions are motivated by the previously mentioned universality phenomenon stipulating that the statistics of a random matrix do not depend on the distribution of its entries. Guided by this belief, one anticipate the eigenvectors of a broad class of random matrices to behave as those of a Gaussian random matrix for instance. The latter being invariant under orthogonal transformations, its normalized eigenvectors are uniformly distributed over the unit sphere and are therefore very much delocalized. The notions of delocalization considered in the literature aim mainly to mimic some characteristics of the eigenvectors of a Gaussian random matrix. We refer to [124, 125] for more on this concept and for an extensive list of references on the subject.

As for the eigenvalues distribution, understanding the eigenvectors in the non-Hermitian setting is significantly more challenging. Moreover, our model exhibit dependencies and high sparsity making it even more complicated. Nevertheless, we were able to prove the following in [Y16].

Theorem 4.6 ([Y16]). *There are universal constants $c, C > 0$ such that the following holds. Let $n \geq C$ and $C \leq d \leq \exp(c\sqrt{\log n})$ and let $A_{n,d}$ be the adjacency matrix of $\mathcal{D}_{n,d}$ the uniform random d-regular digraph on n vertices. Then with probability at least $1 - 1/n$ any eigenvector x of $A_{n,d}$, which is not parallel to $(1, 1, \ldots, 1)$, satisfies $x_i^* \leq d^3(n/i)^6 x_{\lfloor cn \rfloor}^*$ for all $1 \leq i \leq \lfloor cn \rfloor$, and, moreover,*

$$\forall \lambda \in \mathbb{C} \qquad \left| \left\{ i \leq n : \, |x_i - \lambda| \leq n^{-c} x_{\lfloor cn \rfloor}^* \right\} \right| \leq Cn \log^2 d / \log n,$$

where $(x_i^*)_i$ denote the non-increasing rearrangement of $(|x_i|)_i$.

The above Theorem is close in spirit to the concept of no-gap delocalization as it implies weak lower bounds on the Euclidean norm of restrictions of an eigenvector. However, in addition to that, it also measures cardinalities of sets of almost equal coordinates, thus giving an additional structural information very important in the context discussed in the previous section. This result, to our best knowledge, is the first statement which provides quantitative information on the delocalization for non-Hermitian random matrices with a *constant* number of non-zero elements in rows/columns.

Theorem 4.6 asserts for example that a vector in the kernel of $A_{n,d}$ necessarily have at least $\log n / \log^2 d$ levels of coordinates, where by a level we mean a subset of equal coordinates. Such a structural information guide our application of the switching technique as to show for instance that $A_{n,d}$ is of rank at least $n - 1$ with high probability. Informally speaking, knowing the level structure of a vector in the kernel of $A_{n,d}$, indicate the switchings to be operated on $A_{n,d}$ as to remove this vector from its kernel. We refer to [Y15] for a more precise and rigorous explanation of this idea.

In [Y16], we developed much more involved structural properties of normal vectors to rows of $A_{n,d}$. These properties turned out to be essential in understanding the smallest and intermediate singular values of $A_{n,d}$ and its complex shifts. An often employed approach to estimating the smallest singular value is to split the unit sphere and work separately with different types of vectors. In this context, the concept of *compressible* and *incompressible* vectors was introduced [89, 122] as the former are of low complexity (close to sparse) and efficient covering arguments could be used to deal with them, while the latter are spread and naturally provide access to concentration and anti-concentration inequalities. Building upon earlier works on the Littlewood–Offord theory, the concept of the least common denominator (LCD) of a vector was introduced [122], and the Euclidean sphere was partitioned into subsets according to the magnitude of the LCD. In our model, due to special structure of the matrices (in particular, due to the lack of independence and sparsity) these concepts are not sufficient and we developed a tailored splitting of the unit sphere into *steep*, *almost constant*, and *gradual vectors*. We refer to [Y16] for the exact definitions of these notions, and only give here an informal description. *Steep vectors* correspond to vectors containing jumps in the non-increasing rearrangement of their coordinates. In a similar manner, *very steep vectors* are those having a huge drop at the beginning of their coordinates' non-increasing rearrangement, thus consisting of "few" large coordinates while the remaining are much smaller. This notion is related to the notion of compressible vectors, however the distance to sparse vectors is measured in another metric. *Almost constant vectors* are those having a

large set of equal coordinates while all remaining vectors are *gradual vectors*. Let us state an informal version of the main result of [Y16] where a more detailed statement is presented.

Theorem 4.7 ([Y16])**.** *There are universal constants $c, C > 0$ such that the following holds. Let $n \geq C$ and $C \leq d \leq \exp(c\sqrt{\log n})$ and let $A_{n,d}$ be the adjacency matrix of $\mathcal{D}_{n,d}$ the uniform random d-regular digraph on n vertices. Let $I \subset [n]$ satisfying $0 \leq |I^c| \leq n/d^3$ and denote $A^I_{n,d}$ the restriction of $A_{n,d}$ to the rows indexed by I. Then with probability at meast $1 - 1/n$ every unit vector such that $\|M^I x\|_2 \leq |I^c|^3 n^{-6}$ is either very steep or gradual with many levels of coordinates.*

4.4 perspectives

Conjecture 4.1.1 remains open as of this writing. Solving it will require some additional work and new ideas need to be injected. For instance, the solution to Conjecture 4.2.1 is not fully satisfactory as it lacks quantitative estimates. On the other hand, with our quantitative approach we could only show that the adjacency matrix of the uniform random d-regular digraph is of rank at least $n-1$. This comes from a deficiency in our proof techniques which rely heavily on the structure of potential null vectors. As we explained previously, we use this structure to guide the application of the switching technique and show that given a singular adjacency matrix, many switchings make it invertible. Reversely, given a full rank adjacency matrix, one needs to show that few switchings would make it singular. Since in such a case, one has no null vector to start working with, we have limited understanding on how to operate switchings in an efficient way. At the same time, this seems to be the only obstacle to turning Theorem 4.5 into a proof of a potentially quantitative form of Conjecture 4.2.1.

All current proofs of the circular law in various models rely on the Hermitization technique. While the latter is clearly motivated by the desire to bridge back the problem to studying the singular values, it doesn't provide any qualitative understanding as to the universality of the circular law. It turns out that the latter minimizes the logarithmic energy under a second moment constraint (see [31, Remark 4.7]). It would be very interesting to elaborate a proof of the circular law theorem exploiting this characterization along the lines of the entropic proof of the central limit theorem.

Chapter 5

Spectral gap of random graphs

As we already mentioned previously, graph properties such as connectivity, bipartitedness, graph diameter, and the evolution of various random processes defined on the graph are closely related to the eigenvalues of its adjacency matrix (see for example [43]). In the previous chapter, we investigated the global behavior of the eigenvalues of the adjacency matrix of the uniform random regular graph. In this chapter, we move to a finer scale as to understand the edge of the limiting spectral distribution and the gap between the two largest eigenvalues/singular values.

5.1 The spectral gap of random regular graphs [Y11, Y12]

One of the important properties of a graph is its connectivity. In practice, one would need to build a network which is well connected for a signal to propagate fast in it. Such network quality can be measured through *the edge isoperimetric constant* of a graph G defined by

$$h(G) := \inf\Big\{\frac{|\partial S|}{|S|} : S \subset [n], |S| \leq n/2\Big\},$$

where ∂S denotes the set of edges connecting a vertex in S to one outside of it. Clearly, the bigger $h(G)$ is, the better is the connectivity of G. This naturally leads to the notion of *expanders*, a family of graphs $(G_n)_{n\in\mathbb{N}}$ whose edge expansion constant is bounded away from 0 for any n. Expander graphs play an important role in applications and we refer to [77] for a detailed treatment of these objects. Controlling the edge expansion constant is therefore of particular importance, and falls in the category of *isoperimetric inequalities* as it boils down to finding a relation between the volume of a set (the number of vertices it contains) and its surface (captured by the size of its boundary). This is the content of Cheeger's inequality, proved by Alon and Milman [8], which asserts that for any undirected d-regular graph G

$$\frac{d-\lambda_2}{2} \leq h(G) \leq \sqrt{2d(d-\lambda_2)},$$

where λ_2 stands for the second largest eigenvalue of the adjacency matrix of G. The above inequalities provide a remarkable connection between the expansion constant of the graph and its "spectral gap" given by $d-\lambda_2$. Indeed, recall that for any d-regular graph, the largest eigenvalue of its adjacency matrix is equal to d. The above asserts, that the bigger the gap is between the two largest eigenvalues, the better expander the graph is; moreover, the reverse is also true. Given an undirected d-regular graph G, we denote $\lambda = \lambda(G) = \max(|\lambda_2|, |\lambda_n|)$ its second largest (in absolute value) eigenvalue. In a similar manner, the expander mixing lemma [7] controls the pseudo-randomness of the graph in terms of λ. The smaller λ is, the bigger the spectral gap is and the better are the expansion properties of the graph. This naturally leads to wonder how small can λ be. The Alon-Boppana bound serves this purpose and asserts that

$$\lambda(G) \geq 2\sqrt{d-1}\left(1 - \frac{C\log^2 d}{\log^2 n}\right),$$

for some universal constant C. A graph satisfying $\mu \leq 2\sqrt{d-1}$ is called *Ramanujan* and in view of the above, a family of Ramanujan graphs can be considered as the optimal spectral expanders. Showing the existence and providing constructions of Ramanujan graphs was subject of several research investigations [90, 96, 92]. Attention naturally turns to random graphs, as one wonders when picking a d-regular undirected graph uniformly at random, whether it is almost Ramanujan or not. It was conjectured by Alon [6] and proved in a breakthrough by Friedman [58] (see [28] for an alternative proof), that whenever d is fixed, one has $\lambda(G) = 2\sqrt{d-1} + o_n(1)$ with probability going to one as n tends to infinity. One may wonder if such a phenomenon still holds as $d \to \infty$, and the proof of [58, 28] which relies on the configuration model only extends to values of d growing at a slow rate. This leads to the following conjecture made by Vu [148] (see also conjectures 7.3 and 7.4 in [149]).

Conjecture 5.1.1. *Let $\mathcal{G}_{n,d}$ be the uniform random undirected d-regular graph on n vertices. Then for any $d \leq n/2$ growing to infinity with n, we have*

$$\lambda(\mathcal{G}_{n,d}) \leq \big(2 + o(1)\big)\sqrt{d(1-d/n)},$$

with probability going to 1 with n.

In addition to its graph theoretical meaning described previously, the above conjecture asserts that, apart from the largest eigenvalue equal to d, there should be no other outliers to the limiting spectral distribution. Indeed, as we mentioned in the previous chapter, it was shown in [140] that the limiting spectral distribution of the properly normalized adjacency matrix of the uniform random undirected d-regular graph is the semi-circular distribution. Since the variance of each entry of this matrix is $\frac{d}{n}(1-d/n)$, then this suggests that most of the spectrum lies between $-2\sqrt{d(1-d/n)}$ and $2\sqrt{d(1-d/n)}$. The above conjecture predicts that all eigenvalues asymptotically live in this window, and there are no outliers as the second largest (in absolute value) eigenvalue should asymptotically stick to the boundary of the support of the limiting measure.

In [40], it was shown that $\lambda \lesssim \sqrt{d}$ for any $d = o(\sqrt{n})$ with probability going to one in n. Later on, this result was extended in [48] to cover the range $d \lesssim n^{2/3}$. In [Y12], we further extend this range to answer the isomorphic version of Vu's conjecture.

Theorem 5.1 ([Y12]). *Let $n^{1/4} \leq d \leq n/2$ and $\mathcal{G}_{n,d}$ be the uniform random undirected d-regular graph on n vertices. Then $\lambda(\mathcal{G}_{n,d}) \lesssim \sqrt{d}$, with probability at least $1 - 1/n$.*

Together with the previous results, this covers all ranges of degree d going to infinity and answers up to a constant the conjecture. The above statement can also be interpreted as a concentration result for the adjacency matrix around its expectation. Indeed, let A denote the adjacency matrix of $\mathcal{G}_{n,d}$ and note that $\mathbb{E}\, A = \frac{d}{n}\mathbf{1}\mathbf{1}^t$, where $\mathbf{1}$ denotes the n-dimensional vector with all entries equal to 1. Now recall that $\mathbf{1}$ is the eigenvector associated with the largest eigenvalue equal to d so that by the spectral theorem, we can write

$$A - \mathbb{E}\, A = \lambda_2 v_2 v_2^t + \ldots + \lambda_n v_n v_n^t,$$

where we denoted $\lambda_2 \geq \ldots \geq \lambda_n$ the eigenvalues of A and $v_2, \ldots, v_n$ the associated eigenvectors. Therefore, we see that $\lambda(\mathcal{G}_{n,d}) = \|A - \mathbb{E}\, A\|$ and the above result states that A concentrates well around its expectation, where the concentration is measured in terms of the operator norm.

Those results extend to the directed case. Indeed, a directed version of Friedman's theorem was elaborated in [49] following the approad introduced in [28]. In [Y12], we also proved a directed version of the above theorem.

Theorem 5.2 ([Y12]). *Let $n^{1/4} \leq d \leq n/2$, let $\widetilde{\mathcal{D}_{n,d}}$ be the uniform random directed d-regular graph on n vertices and $A_{n,d}$ its adjacency matrix. Then $s_2(A_{n,d}) \lesssim \sqrt{d}$, with probability at least $1 - 1/n$, were $s_2(\widetilde{A})$ stands for the second largest singular value of $A_{n,d}$.*

In fact, the above statements are proved in a more general setting as to cover random graphs with predefined degree sequences satisfying some assumptions. Interestingly, we show that it is possible to deduce Theorem 5.1 from Theorem 5.2. Indeed, in [Y11], we developed a sort of "de-symmetrization" allowing to relate the spectral gap of an undirected graph to that of the spectral gap of some directed sub-graph of it. The procedure which is quite general and interesting on its own, allows to estimate the norm of an $n \times n$ random matrix with jointly exchangeable entries in terms of the norm of its $n/2 \times n/2$ submatrix located in the top right corner (we refer to [Y11] for more details).

The proof of Theorem 5.2 uses standard ideas originated in Geometric Functional Analysis. Indeed, one starts using the Courant-Fisher formula and the singular value decomposition to write

$$s_2(A_{n,d}) = \sup_{x \in S^{n-1}, x \perp \mathbf{1}, y \in S^{n-1}} \langle A_{n,d} x, y \rangle,$$

and aim at showing that the supremum is controlled (up to a constant) by $\sqrt{d}$ with high probability. The standard approach consists in three steps which can be summarized as follows:

1. Fix $x, y \in S^{n-1}$ with $x \perp \mathbf{1}$. Show that $\langle A_{n,d} x, y \rangle \lesssim \sqrt{d}$ with high probability.

2. Discretize the sphere with a finite net, and elaborate an approximation argument to reduce the supremum over the sphere to a supremum over the net.

3. Combine the two steps through a union bound over the net.

While the last two steps are rather standard, the difficulty lies in establishing individual probability bounds strong enough to compensate the size of the net which is exponential in the dimension. Given $x, y \in S^{n-1}$ and denoting a_{ij} the entries of $A_{n,d}$, notice that $\langle A_{n,d}x, y\rangle = \sum_{1\leq i,j\leq n} a_{ij}x_j y_i$ is a sum of bounded random variables. When these are independent, standard concentration inequalities such as Bennett and Bernstein's inequality are available to deal with such quantity. It is therefore essential to develop matching concentration inequalities in dependent settings such as the one encountered here. This will be the subject of the next chapter. To finish this section, let us note that the above three step strategy is insufficient in this case, as it is not possible to obtain strong concentration inequalities for localized vectors x and y. Instead, such inequalities are used only for delocalized vectors, and one exploits the expansion properties of the graph when dealing with localized vectors. This approach was introduced by Kahn and Szemerédi [59] and used subsequently for similar problems.

5.2 Phase transition for the spectral gap of Erdős-Renyi graphs [Y15, Y18]

In this section, we move to the study of another famous random graph model. The Erdős-Renyi graph $\mathcal{G}(n, p_n)$, which we mentioned previously, is the random undirected graph on n vertices where each edge is drawn independently with probability p_n. Its adjacency matrix is a symmetric $n \times n$ matrix whose entries are independent (up to the symmetry) Bernoulli random variables with probability of success p_n. As for the uniform model, understanding the spectrum of the Erdős-Renyi graph is of interest from a graph theoretical point of view and from a random matrix one. If B_n denotes the adjacency matrix of $\mathcal{G}(n, p_n)$ then it is known that the limiting spectral distribution of $B_n/\sqrt{np_n}$ is the semi-circular distribution as long as $np_n \to \infty$. When np_n is constant, the behavior of the limiting measure is more involved and has been subject to several investigations (we refer to [54] and references therein for more on this regime which will not be investigated here). In the sequel, the focus will be on the regime where $np_n \to \infty$ and on trying to understand the spectral gap of the Erdős-Renyi graph in it.

It follows from the limiting spectral distribution that the leading eigenvalues of B_n are asymptotically larger than $2\sqrt{np_n}$. As previously, understanding the exact location of these eigenvalues is our aim. In [85], it is shown that the largest eigenvalue of B_n almost surely satisfies

$$\lambda_1(B_n) = \big(1 + o(1)\big) \max\Big(\sqrt{\max_{i\leq n} \deg(i)}, np_n\Big),$$

where $o(1)$ tends to 0 as $\max\big(\sqrt{\max_{i\leq n} \deg(i)}, np_n\big)$ tends to infinity, and $\deg(i)$ is the degree of the i-th vertex of $\mathcal{G}(n, p_n)$.

When $\frac{np_n}{\log n} \to \infty$, standard concentration inequalities for sums of independent random variables (or just binomial tail estimates) ensure that all degrees in the graph are around

their average np_n. This implies that in this regime, the Erdős-Renyi graph is almost np_n-regular. The above expression for λ_1 shows that it coincides asymptotically with np_n. One may naturally wonder if as for the uniform random regular graph, the Erdős-Renyi graph is almost Ramanujan. This was shown to be true by Fıuredi and J. Komlòs for $np_n \gtrsim \log^8 n$, by Vu [147] for $np_n \gtrsim \log^4 n$. For the general case, it was shown independently by Bordenave, Benaych-Georges and Knowles [25] and in our work [Y15]. More precisely, we proved the following:

Theorem 5.3 ([Y15]). *Let $\mathcal{G}(n, p_n)$ be the Erdős-Renyi random graph and B_n its adjacency matrix. Then*

$$\mathbb{P}\Big(\lambda(B_n) \leq 2\Big(1 + \sqrt[3]{C \log n/(np_n)}\Big)\sqrt{np_n}\Big) \underset{n\to\infty}{\longrightarrow} 1,$$

where C is a universal constant.

While the results of [25] improve on the error term given by the cubic root above, our result extends to inhomogeneous graphs without any restriction on the sequence of connection probabilities (we refer to [Y15] for more details). It is natural at this stage to question the restriction $\frac{np_n}{\log n} \to \infty$ and its necessity for the graph to be almost Ramanujan. Let us note that as in the previous section, one can interpret the above as a concentration of the adjacency matrix around its expectation.

It is not difficult to see that if $np_n \to \infty$ very slowly, vertices of degree higher than the average start emerging in the graph creating hubs and preventing expansion. This suggests that as $\frac{np_n}{\log n} \to 0$, the above Ramanujan property breaks down. This was for instance studied in [56, 87] where procedures were developed as to clean up the graph from these anomalies and recover the concentration of the adjacency matrix around its expectation. It was observed that reducing the degrees of the vertices as to make them all bounded by twice the average reinstaure the concentration in the graph. As we will see in the sequel, we will show that such a bound on the degrees is in fact necessary for optimal concentration. In [24], the exact value of $\lambda(B_n)$ was captured in this regime. More precisely, it is shown that if $\frac{np_n}{\log n} \to 0$, then $\lambda(B_n)$ concentrates around $\sqrt{\max_{i\leq n} \deg(i)}$ with probability going to 1 with n. In view of the estimate on λ_1, this means that the first two eigenvalues are asymptotically glued together and the graph is no longer an "optimal" expander.

It is known that $\log n$ is the threshold of connectivity in the sense that the graph is connected when np_n exceeds $\log n$ and disconnects when it drops below this threshold. We are lead to wonder if in a similar manner, there is a phase transition happening in the window around $\log n$ and if there is a particular connectivity probability above which the graph is an optimal expander and below which it no longer is. We answer this in the following theorem.

Theorem 5.4 ([Y18]). *Let $\mathcal{G}(n, p_n)$ be the Erdős-Renyi random graph and B_n its adjacency matrix. Assume that $p_n \to 0$ and $np_n \to \infty$. Then, we have the following dichotomy:*

- *(optimal expander) If $\liminf \frac{np_n}{\log n} \geq \frac{1}{\log(4/e)}$ then $\frac{\lambda(B_n)}{2\sqrt{np_n}}$ converges to 1 in probability.*
- *(non-optimal expander) If $\limsup \frac{np_n}{\log n} < \frac{1}{\log(4/e)}$ then there is $\varepsilon > 0$ such that*

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{\lambda(B_n)}{2\sqrt{np_n}} > 1 + \varepsilon\Big) = 1.$$

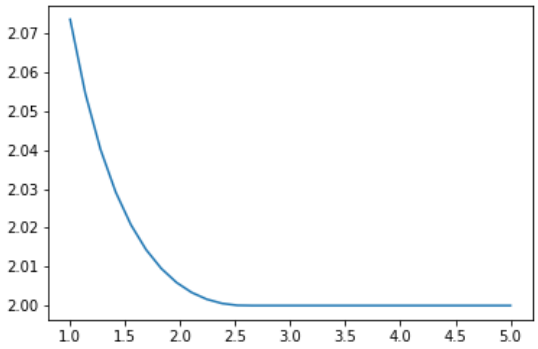


Figure 5.1: *The value of* $\lim_n(\rho_n/\sqrt{np_n})$ *(vertical axis) for the adjacency matrix of the Erdős–Renyi graph, viewed as a function of* $\lim_n(np_n/\log n)$ *(horizontal axis). The phase transition happens at* $\frac{1}{\log(4/e)} \approx 2.59$.

The threshold $\frac{1}{\log(4/e)}$ has a particular significance, it is also the connectivity probability below which vertices with degree larger than twice the average start emerging in the graph. This nicely complement the work of [56, 87], and rigorously relates the concentration of the graph to the presence or not of vertices with degree higher than twice the average degree. We mention that shortly after our result was posted on arxiv, this phase transition was independently (and with different techniques) captured in [2]. Moreover, the work of [2] further strengthens the above by showing that the emergence of each vertex of degree larger than twice the average, result in a new eigenvalue getting above $2\sqrt{np_n}$ asymptotically.

The above result is a consequence of the characterization of the norm of sparse random matrices which were discussed in Chapter 3. More precisely, we are able to show a completely explicit asymptotic expression for $\lambda(B_n)$. Indeed, setting

$$\rho_n := \theta_n + \frac{np_n}{\theta_n}, \quad \theta_n := \sqrt{\max\left(enp_n \exp\left[\mathcal{W}_0\Big(\frac{\log n - np_n}{enp_n}\Big)\right] - np_n, np_n\right)},$$

we have that $\frac{\lambda(B_n)}{\rho_n}$ converges to one in probability. The function $\mathcal{W}_0$ denotes the main branch of the Lambert function defined by $z = \mathcal{W}_0(z)e^{\mathcal{W}_0(z)}$. The presence of this function comes from the expression of the maximum degree in the Erdős-Renyi graph as it can be shown that

$$\frac{\max_{i\leq n}\deg(i)}{enp_n \exp\left[\mathcal{W}_0\big(\frac{\log n - np_n}{enp_n}\big)\right]} \xrightarrow[n\to\infty]{\mathbb{P}} 1.$$

We refer to [Y18] for more details, and note here that the above formula for $\lambda(B_n)$ can be formulated in terms of the maximum degree in the graph englobing for instance the previous results obtained when $\frac{np_n}{\log n} \to 0$ or ∞. Thanks to these expressions, we illustrate in Figure 5.1 the phase transition captured in Theorem 5.4.

5.3 Perspectives

Conjecture 5.1.1 remains open as of this writing. The methods we use originating in Geometric Functional Analysis can never lead to a solution to this conjecture as they naturally lead to sub-optimal constants. Reaching such a level of precision is usually done through the moment method which was used in the proof of results in Chapter 3. This method relies heavily on independence and finding a way to go beyond such a setting remains a challenge. While for small values of d, the uniform model can be studied through the configuration model which exhibits an independence structure, this is no longer the case when d grows relatively fast to infinity. From this perspective, it would be very interesting to bring a new insight into the moment method through the study of this problem.

In the previous section, we were able to establish a phase transition for the appearance of outliers in the spectrum of the Erdős-Renyi random graph. Moreover, we were able to relate these outliers to the emergence of high degree vertices in the graph. Naturally, these emerging hubs make the corresponding eigenvectors localized on them. Therefore, one expects a phase transition in the localization/delocalization of eigenvectors in accordance to the phenomenon captured in this chapter. This problem was recently studied in [74] where it was observed that the transition from delocalization to localization takes place in the window of connectivity around $\log n/n$. It would be interesting to capture the exact transition similarly to what is done in this chapter.

Chapter 6

Matrix concentration inequalities

The laws of large number assert that a sum of random variables is asymptotically close to its expectation. Concentration inequalities aim to quantify this standard fact and provide a probabilistic illustration of a high dimensional phenomena: a random variable measurable with respect to a large number of independent random variables and depending little on each individually, is almost constant with high probability. Concentration inequalities are versatile tools which found use in areas across Mathematics, and deriving such inequalities can be of great use to several pure and applied problems. We refer to [32] for more on concentration inequalities.

In the previous chapter, we encountered the need to elaborate concentration inequalities in a dependent setting. In this chapter, we will investigate some generalized concentration inequalities as to implement special types of dependence on one hand, and to consider non-commutative random variables on the other hand.

6.1 Bennett-Bernstein inequalities in a dependent setting [Y12]

Among all concentration inequalities, perhaps the most famous ones are Hoeffding, Bernstein and Bennett's inequality [32]. While Hoeffding's inequality deals with sub-Gaussian random variables, Bernstein's with sub-exponential random variables, these inequalities are commonly used to deal with a sum of independent bounded random variables. In such a setting, Bennett's inequality offers the most satisfactory results as it takes full account of the law of the variables in addition to their boundness assumption.

Bennett's inequality asserts that given $\xi_1, \dots, \xi_N$ independent random variables satisfying $|\xi_i - \mathbb{E}\,\xi_i| \leq K$ almost surely for every $i \leq N$, then for any $t > 0$ we have

$$\mathbb{P}\Big(\sum_{i=1}^{N}(\xi_i - \mathbb{E}\,\xi_i) \geq t\Big) \leq \exp\Big(-\frac{\sigma^2}{K^2}\,H\Big(\frac{Kt}{\sigma^2}\Big)\Big),$$

where $\sigma^2 = \sum_{i=1}^{N} \mathrm{Var}(\xi_i)$ and $H(x) = (1+x)\log(1+x) - x$. The above inequality combines a Gaussian tail behavior in the small deviation regime and a Poisson-like tail behavior in the

large deviation one. The function H in the above deviation bound is quite natural in this context. It implicitly appears in the classical inequality of Bennett for sums of independent variables (see, [26, Formula 8b]), and later in the well known paper of Freedman [57] where he extends BennettÕs inequality to martingales. The latter work offers a lot of flexibility as to derive concentration inequalities in a dependent setting as long as one can construct a martingale adapted to the problem. One of the main challenges lies in understanding the quadratic variations of the corresponding martingale as this determines the quality of the concentration inequality to be obtained.

Motivated by estimating the spectral gap of the uniform random regular graph discussed in the previous chapter, we developed a concentration inequality in this context. Let $\mathcal{D}_{n,d}$ be the random uniform directed d-regular graph on n vertices and denote $E_{\mathcal{D}}$ the (random) set of directed edges of $\mathcal{D}_{n,d}$. Given an $n \times n$ matrix $Q = (q_{ij})_{1 \leq i,j \leq n}$, our aim is to study the random variable

$$\xi = \sum_{(i,j) \in E_{\mathcal{D}}} q_{ij} = \sum_{i,j=1}^{N} q_{ij} \mathbf{1}_{(i,j) \in E_{\mathcal{D}}},$$

which is a sum of dependent random variables, where the dependence is dictated by the d-regularity of the graph. Each of these random variables is bounded above by $\max_{i,j} |q_{ij}| := \|Q\|_\infty$ and we have

$$\mathbb{E}\,\xi = \frac{d}{n} \sum_{i,j=1}^{N} q_{ij} \quad \text{and} \quad \sum_{i,j=1}^{N} \mathbb{E}\,[q_{ij}^2 \mathbf{1}_{(i,j) \in E_{\mathcal{D}}}] = \frac{d}{n} \|Q\|_{\mathrm{HS}}^2.$$

With these observations in mind, let us state a soft incomplete version of the Bennett's inequality we obtained in this context.

Theorem 6.1 ([Y12])**.** *With the above notations we have that for any $t > 0$*

$$\mathbb{P}\Big(\Big| \sum_{(i,j) \in E_{\mathcal{D}}} q_{ij} - \frac{d}{n} \sum_{i,j=1}^{N} q_{ij} \Big| \geq t \mid \mathcal{E} \Big) \leq 2 \exp\Big(- \frac{d\|Q\|_{\mathrm{HS}}^2}{n\|Q\|_\infty^2} \, H\Big(\frac{tn\|Q\|_\infty}{d\|Q\|_{\mathrm{HS}}^2} \Big) \Big),$$

where $\mathcal{E}$ is an event holding with probability at least $1 - 1/n$.

In view of the calculations made above, the probability bound obtained is in accordance with Bennett's inequality. The previous theorem can be extended to cover random graphs with predefined degree sequences under some additional assumptions. Perhaps the only drawback of this result is the presence of the event $\mathcal{E}$ which we did not specify here. This comes from technical obstacles in implementing the martingale based proof and we refer to [Y12] for a clear definition of it and a detailed statement. Let us note that many arguments can be usually carried under conditioning before getting rid of it once the whole argument is completed. This is for instance the case of our application to the study of the spectral gap discussed in the previous chapter and the above theorem serves its need as to deal with the first of the three steps strategy described there. Let us end this section by emphasizing the importance of such concentration inequalities. Indeed, Theorem 6.1 provides for instance

bounds on edge count statistics in $\mathcal{D}_{n,d}$. Given two sets of vertices S and T in $[n]$, take Q to be the 0/1 matrix with entries in the block $S \times T$ equal to 1. Then applying Theorem 6.1 one obtains that the number of edges heading from S to T is concentrated around its expectation $\frac{d}{n}|S| \cdot |T|$ in accordance with the pseudo-randomness property of these graphs.

6.2 Non-commutative Khintchine inequality and matrix covariance estimation [Y3, Y15]

Concentration inequalities for a sum of independent random variables is intimately related to the study of their moments. Classical Khintchine inequalities study the moments of a sum of independent sub-Gaussian random variables. Extending these inequalities to cover series with matrix coefficients is of particular importance due to the numerous applications in applied mathematics (see for instance [142]). The non-commutative Khintchine inequality was first obtained by Lust-Piquard [91] and later extended and extensively studied (see for example [113] and references therein). In its simplified form, it states that given $A_1, \dots, A_N$ a sequence of $n \times n$ deterministic matrices and $\varepsilon_1, \dots, \varepsilon_N$ independent Rademacher random variables, one has for any $p \in \mathbb{N}^*$

$$\Big[\mathbb{E}\mathrm{Tr}\Big(\big(\sum_{i=1}^N \varepsilon_i A_i\big)^{2p}\Big)\Big]^{\frac{1}{2p}} \lesssim \sqrt{p}\Big[\mathrm{Tr}\big(\sum_{i=1}^N A_i^2\big)^p\Big]^{\frac{1}{2p}}.$$

Let us note that taking p large enough (say of order $\log n$), one obtains bounds on the operator norm of X since the latter is approximated by the p-Schatten norm when p is large. The above inequality can be strengthened as to deal with Gaussian series. Indeed, using that $\mathbb{E}|g| = \sqrt{2/\pi}$, we can write

$$\Big[\mathbb{E}\mathrm{Tr}\Big(\sum_{i=1}^m \varepsilon_i A_i\Big)^{2p}\Big]^{\frac{1}{2p}} = \sqrt{\frac{\pi}{2}}\Big[\mathbb{E}\mathrm{Tr}\Big(\sum_{i=1}^m \varepsilon_i \mathbb{E}|g_i|\, A_i\Big)^{2p}\Big]^{\frac{1}{2p}},$$

where $(g_i)_{i\le m}$ are independent standard Gaussians, independent of the sequence of Rademachers $(\varepsilon_i)_{i\le n}$. Now combining this with Jensen's inequality, we deduce

$$\Big[\mathbb{E}\mathrm{Tr}\Big(\sum_{i=1}^m \varepsilon_i A_i\Big)^{2p}\Big]^{\frac{1}{2p}} \le \sqrt{\frac{\pi}{2}}\Big[\mathbb{E}\mathrm{Tr}\Big(\sum_{i=1}^m \varepsilon_i |g_i|\, A_i\Big)^{2p}\Big]^{\frac{1}{2p}}.$$

Finally, using that $\varepsilon_i|g_i|$ has the same distribution as g_i, we conclude that having a non-commutative Khintchine inequality for Gaussian series imply its Rademacher counterpart. Denoting $X = \sum_{i=1}^N g_i A_i$, the non-commutative Khintchine then states

$$\Big[\mathbb{E}\mathrm{Tr}(X^{2p})\Big]^{\frac{1}{2p}} \lesssim \sqrt{p}\Big[\mathrm{Tr}\big(\sum_{i=1}^N A_i^2\big)^p\Big]^{\frac{1}{2p}},$$

and since $\mathbb{E}\, X^2 = \sum_{i=1}^N A_i^2$, the above inequality can be formulated entirely in terms of X. Of particular interest is the study of the optimality of the above inequality. Taking A_i to be

the diagonal matrix with its i-th entry equal to 1, we see that the above inequality describes precisely the moments of X which is in this case just a diagonal matrix formed by Gaussian random variables. However, taking A_i to be $n \times n$ matrix having only the $(1,i)$, and $(i,1)$-th entries equal to 1 and the remaining 0, one can check that the non-commutative Khintchine inaccurately describes the reality.

This optimality question seems intractable in its full generality. In [Y15], we investigate an important special case given by $X = \sum_{1 \leq i \leq j \leq n} g_{ij} A_{ij}$ where A_{ij} is the $n \times n$ matrix with all its entries equal to zero except the (i,j) and (j,i)-th ones which are equal to some non-negative scalar a_{ij}. In other terms, X is an $n \times n$ inhomogeneous Gaussian matrix represented as a Gaussian series. In this context, we show that the non-commutative Khintchine inequality can be sharpened and obtain the following.

Theorem 6.2 ([Y15]). *Let $X = \sum_{1 \leq i \leq j \leq n} g_{ij} A_{ij}$ be as above. Then for any $p \geq 1$*

$$\Big[\mathbb{E}\mathrm{Tr}(X^{2p})\Big]^{\frac{1}{2p}} \asymeq \Big[\mathrm{Tr}\big(\sum_{1 \leq i \leq j \leq N} A_{ij}^2\big)^p\Big]^{\frac{1}{2p}} + \sqrt{p}\Big[\mathrm{Tr}\big(\sum_{1 \leq i \leq j \leq N} A_{ij}^{2p}\big)\Big]^{\frac{1}{2p}}.$$

A nice application of the non-commutative Khintchine inequality is to the problem of covariance estimation. In its classical form, we are given a centered random vector Y in $\mathbb{R}^n$ and we aim at approximating its covariance matrix by the empirical covariance estimator given by $\frac{1}{N}\sum_{i=1}^N Y_i Y_i^t$, where the Y_i's are independent copies of Y. The goal being to provide such an approximation with the least possible number of copies. The link to the non-commutative Khintchine inequality was made by Rudelson [118] who used symmetrization and Jensen's inequality to write

$$\mathbb{E}\,\Big\|\frac{1}{N}\sum_{i=1}^N Y_i Y_i^t - \mathbb{E}\,YY^t\Big\| \leq 2\mathbb{E}\,\Big\|\frac{1}{N}\sum_{i=1}^N \varepsilon_i Y_i Y_i^t\Big\|,$$

where the ε_i's are independent Rademacher random variables (independent from the Y_i's). One then conditions on the realization of the Y_i's and uses the non-commutative Khintchine inequality to estimate the right-hand side. Such an approach implies that if $\|Y\|_2 = O(\sqrt{n})$ almost surely, then taking $N = O(n \log n)$ copies suffices to accurately approximate the covariance matrix of Y. This result is very general as it requires no assumption on the actual distribution of Y. Nevertheless, it is also known that for sufficiently regular distributions the logarithmic oversampling factor is not needed. Covariance estimation problems are extensively studied in high-dimensional statistics, we refer to [146, Chapter 4] to avoid an extensive list of references.

In [Y2], we were interested in the above problem when additional regularity assumptions are made.

Theorem 6.3 ([Y2]). *Let B be an $n \times n$ positive semi-definite random matrix satisfying $\mathbb{E}\,B = \mathrm{Id}$ and*

$$\mathbb{P}\Big(\|PBP\| \geq t\Big) \leq \frac{c}{t^{1+\eta}}, \quad \forall t \geq c\mathrm{rank}(P) \text{ and for any orthogonal projection } P,$$

for some c and $\eta > 0$. Then for every $\varepsilon \in (0,1)$, taking $N \gtrsim n/\varepsilon^{2+2\eta^{-1}}$ we have

$$\mathbb{E}\Big\|\frac{1}{N}\sum_{i=1}^{N} B_i - \mathrm{Id}\Big\| \leq \varepsilon,$$

where $B_1, \dots, B_N$ are independent copies of B.

As opposed to applying the non-commutative Khintchine inequality, the above result saves a logarithmic term in its estimate on the number of copies N. The assumption made on B is not too restrictive as there are several examples of random matrices satisfying it. For instance if $B = UDU^t$ is the spectral decomposition of B with U and D being independent and the diagonal entries of D have a finite $(2+\varepsilon)$ moment, then the above theorem applies. Moreover, taking $B = YY^t$, one recovers the results of Srivastava and Vershynin [131] who studied the classical covariance estimation problem of a random vector with few regularity assumption. We should note that the above theorem and its proof are very much inspired by the work [131].

6.3 Bernstein inequality in a matrix/dependent setting [Y6]

In the first section of this chapter, we investigated concentration inequalities in a dependent setting while in the previous section we studied non-commutative extensions. In this final section, we look at concentration inequalities in a matrix dependent setting. In recent years, there has been an increased interest into matrix concentration inequalities. Extending classical inequalities such as Hoeffding, Bernstein and Bennett to a matrix setting was subject to several investigations. We refer to the excellent monograph of Tropp [142] who greatly contributed to this field.

One of the consequences of the scalar Bernstein's inequality states that if $\xi_1, \dots, \xi_N$ are independent centered random variables uniformly bounded by K, then for any $t > 0$

$$\mathbb{P}\Big(\sum_{i=1}^{N} \xi_i \geq t\Big) \leq \exp\Big(-\frac{t^2}{2(\sigma^2 + Kt/3)}\Big),$$

where $\sigma^2 = \mathrm{Var}\big(\sum_{i=1}^{N}\xi_i\big) = \sum_{i=1}^{N}\mathbb{E}\,\xi_i^2$. There have been several approaches as to extend such inequality to a matrix setting. The aim being to control the largest eigenvalue of a sum of bounded self adjoint matrices. Perhaps the most standard approach is the Laplace transform method which was generalized by Ahlswede and Winder [5]. The independence structure is heavily used as to exploit tensorization properties of the Laplace transform. Going beyond the independent setting remains a challenging and technical task. In [Y6], we aimed at such extension and considered the sum of self-adjoint, centered, geometrically absolutely regular random matrices.

We let $(X_i)_{i\geqslant 1}$ be a family of $n \times n$ self-adjoint random matrices whose entries are defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and that are geometrically absolutely regular in the following

sense. Let

$$\beta_0 = 1 \quad \text{and} \quad \beta_k = \sup_{j \geqslant 1} \beta(\sigma(X_i\,,\, i \leq j), \sigma(X_i\,,\, i \geqslant j+k))\,, \text{ for any } k \geq 1,$$

where

$$\beta(\mathcal{A}, \mathcal{B}) = \frac{1}{2} \sup \big\{ \sum_{i \in I} \sum_{j \in J} |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)| \big\}\,,$$

the maximum being taken over all finite partitions $(A_i)_{i \in I}$ and $(B_i)_{i \in J}$ of Ω respectively with elements in $\mathcal{A}$ and $\mathcal{B}$. The $(\beta_k)_{k \geqslant 0}$ are usually called the coefficients of absolute regularity and we shall assume that they decrease geometrically in the sense that there exists $c > 0$ such that for any integer $k \geqslant 1$,

$$\beta_k = \sup_{j \geqslant 1} \beta(\sigma(X_i\,,\, i \leq j), \sigma(X_i\,,\, i \geqslant j+k)) \leq \mathrm{e}^{-c(k-1)}\,.$$

We refer to [Y6] and references therein for more on these coefficients. Informally, the above assumption relaxes the independence hypothesis usually made and stipulates that matrices far apart in time (indexation) become more and more independent. The degree of independence is captured by these coefficients which decrease exponentially in time.

Theorem 6.4 ([Y6])**.** *Let $(X_i)_{i \geqslant 1}$ be a family of self-adjoint random matrices of size n satisfying the above assumption on the coefficient of absolute regularity. Moreover, assume that*

$$\mathbb{E}\, X_i = 0 \quad \textit{and} \quad \lambda_{\max}(X_i) \leq 1 \quad \textit{almost surely.}$$

Then there exists a universal positive constant C such that for any $t > 0$,

$$\mathbb{P}\Big(\lambda_{\max}\big(\sum_{i=1}^{n} X_i\big) \geq t\Big) \leq n \exp\Big(-\frac{Ct^2}{v^2 + t\log^2 n}\Big)\,,$$

where

$$v^2 = \sup_{K \subseteq \{1,\dots,n\}} \frac{n}{\mathrm{Card} K} \lambda_{\max}\Big(\mathbb{E}\big(\sum_{i \in K} X_i\big)^2\Big).$$

When the X_i's are independent, then the term v^2 corresponds to the variance term usually present in Bernstein's inequality. Let us note that the logarithmic term present above was shown to be necessary in [1]. This result extends to the matrix setting its scalar version obtained in [100] and we refer to [Y6] for examples of applications of it.

6.4 Perspectives

The optimality of the non-commutative Khintchine inequality is a fascinating question. Our solution to a particular case of it (Theorem 6.2) is already highly non-trivial. Moreover, it is not completely clear what would be a suitable conjecture. In view of what we were able to prove, it is tempting to conjecture the validity of Theorem 6.2 beyond the example considered

there. At this stage, it would be interesting to confirm this conjecture on other examples having some exploitable structure in them. For instance, even the case of a sequence of rank one matrices is not understood, band matrices, etc..

Establishing concentration inequalities in a matrix dependent setting remains a subject to be explored. A particularly interesting problem is to study matrix weighted sums of negatively associated random variables. Strong Rayleigh distributions emerged as a suitable candidate towards a theory of negative dependence (see [110, 27]). Following [111] where concentration inequalities for Lipschitz functions of strong Rayleigh distributions were established, these results were strengthened in [75] as to establish corresponding modified log-sobolev inequalities, known to imply such concentration bounds. It would be very interesting to extend these to a matrix setting as beyond the result itself, this would lead to a better understanding of matrix concentration inequalities. Indeed, as we already mentioned, there was an enormous amount of work devoted to extend all approaches of concentration in the real case to the matrix setting. The very elegant entropy based approach present challenges in its extension and only partial results were obtained in this direction [42]. It would be of great interest to efficiently elaborate an entropy approach to matrix concentration inequalities which would extend in a friendly way to dependent settings such as strong Rayleigh distributions.

Bibliography

[1] R. Adamczak, A tail inequality for suprema of unbounded empirical processes with applications to Markov chains, Electron. Journal Probab. (2008) 13, pp1000–1034.

[2] J. Alt, R. Ducatez and A. Knowles, Extremal eigenvalues of critical Erdős–Rényi graphs, arXiv:1905.03243.

[3] R. Adamczak, On the Marchenko-Pastur and circular laws for some classes of random matrices with dependent entries, Electron. J. Probab. 16(2011), 1065–1095.

[4] R. Adamczak, D. Chafaï and P. Wolff, Circular law for random matrices with exchangeable entries, Random Struct. Algorithms 48(3), 454–479 (2016).

[5] R. Ahlswede and A. Winter, Strong converse for identication via quantum channels, IEEE Trans. Inform. Theory, (2002) 48, no. 3, 569–579.

[6] N. Alon, Eigenvalues and expanders, Combinatorica **6** (1986), no. 2, 83–96. MR0875835

[7] N. Alon and F. R. K. Chung, Explicit construction of linear sized tolerant networks, Discrete Math., 72:15–19, 1989.

[8] N. Alon and V. D. Milman, λ_1, isoperimetric inequalities for graphs, and superconcentrators, J. Combin. Theory Ser. B **38** (1985), no. 1, 73–88. MR0782626

[9] N. Alon and J.H. Spencer, The probabilistic method, Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, third edition, 2008. With an appendix on the life and work of Paul Erdős.

[10] N. Anantharaman and E. Le Masson, Quantum ergodicity on large regular graphs, Duke Math. J. **164** (2015), no. 4, 723–765. MR3322309

[11] Á. Backhausz and B. Szegedy, On the almost eigenvectors of random regular graphs, arXiv:1607.04785.

[12] G. W. Anderson, A. Guionnet and O. Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, 118, Cambridge University Press, Cambridge, 2010.

[13] S. Artstein-Avidan, A. Giannopoulos and V. D. Milman, Asymptotic geometric analysis, Part I, Mathematical Surveys and Monographs, 202, American Mathematical Society, Providence, RI(2015).

[14] H. Avron and C. Boutsidis, Faster subset selection for matrices and applications, SIAM J. Matrix Anal. Appl. 34 (2013), no. 4, 1464–1499.

[15] Z. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*, second edition, Springer Series in Statistics, Springer, New York, 2010.

[16] Z. D. Bai and Y. Q. Yin, Limiting behavior of the norm of products of random matrices and two problems of Geman-Hwang, Probab. Theory Related Fields (1986) 73, 555–569.

[17] J. Baik, G. Ben Arous and S. Péché, Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices, Ann. Probab. 33 (2005), no. 5, 1643–1697.

[18] K. Ball, Volumes of sections of cubes and related problems, Lecture Notes in Mathematics 1376, Springer, Berlin, 1989, pp. 251–260.

[19] K. Ball, Volume ratios and a reverse isoperimetric inequality, Journal of the London Mathematical Society 44 (1991), no. 2, 351–359.

[20] A. Bandeira and R. van Handel, Sharp nonasymptotic bounds on the norm of random matrices with independent entries, Ann. Probab., 44(4): 2479–2506, 2016.

[21] A. Basak and M. Rudelson, The circular law for sparse non-Hermitian matrices, to appear in Ann. Probab.

[22] J. D. Batson, D. A. Spielman, and N. Srivastava, Twice-Ramanujan Sparsifiers, In STOC'09–Proceedings of the 2009 ACM International Symposium on Theory of Comput- ing, 255–62. ACM: New York, 2009.

[23] D. Bau and L. Trefethen, Numerical linear algebra. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

[24] F. Benaych-Georges, C. Bordenave, A. Knowles, Largest eigenvalues of sparse inhomogeneous Erdős–Renyi graphs, Ann. Probab., to appear, arXiv:1704.02953

[25] F. Benaych-Georges, C. Bordenave, A. Knowles, Spectral radii of sparse random matrices, arXiv:1704.02945

[26] G. Bennett, Probability Inequalities for the Sum of Independent Random Variables, Journal of the American Statistical Association **297** (1962), 33–45, doi:10.2307/2282438.

[27] J. Borcéa, P. Brändén and T. M. Liggett, Negative dependence and the geometry of polynomials, J. Amer. Math. Soc., 22(2): 521–567, 2009.

[28] C. Bordenave, A new proof of Friedman's second eigenvalue Theorem and its extension to random lifts, to appear in Annales de l'ENS.

[29] C. Bordenave, P. Caputo, and D. Chafaï, Spectrum of non-Hermitian heavy tailed random matrices, Comm. Math. Phys.307(2011), no. 2, 513–560.

[30] C. Bordenave, P. Caputo, and D. Chafaï, Circular law theorem for random Markov matrices, Probab. Theory Related Fields 152(2012), no. 3-4, 751–779.

[31] C. Bordenave and D. Chafaï, Around the circular law, Probab. Surv. **9** (2012), 1–89.

[32] S. Boucheron, G. Lugosi and P. Massart, *Concentration inequalities*, Oxford University Press, Oxford, 2013. MR3185193

[33] P. Bourgade, J. Huang and H.-T. Yau, Eigenvector statistics of sparse random matrices, Electron. J. Probab. **22** (2017), Paper No. 64, 38 pp. MR3690289

[34] J. Bourgain and S. J. Szarek, The Banach-Mazur distance to the cube and the Dvoretzky-Rogers factorization, Israel J. Math., 62(2):169–180, 1988.

[35] J. Bourgain and L. Tzafriri, Invertibility of large submatrices with applications to the geometry of Banach spaces and harmonic analysis, Israel Journal of Mathematics 57, no. 2 (1987): 137–224.

[36] J. Bourgain and L. Tzafriri, Restricted invertibility of matrices and applications, In Analysis at Urbana, Vol. II (Urbana, IL, 1986–1987), volume 138 of London Math. Soc. Lecture Note Ser., pages 61–107. Cambridge Univ. Press, Cambridge, 1989

[37] J. Bourgain and L. Tzafriri, On a problem of Kadison and Singer. J. Reine Angew. Math., 420:1–43, 1991.

[38] J. Bourgain, V.H. Vu, and P.M. Wood, *On the singularity probability of discrete random matrices,* J. Funct. Anal., 258(2) (2010), 559–603.

[39] C. Boutsidis, M.W. Mahoney, and P. Drineas, An improved approximation algorithm for the column subset selection problem, Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, PA, 2009, pp. 968–977.

[40] A. Z. Broder, A. M. Frieze, S. Suen, E. Upfal, Optimal construction of edge-disjoint paths in random graphs, SIAM J. Comput. **28** (1999), no. 2, 541–573 (electronic). MR1634360

[41] E. Candès, Mathematics of sparsity (and a few other things), Proceedings of the International Congress of Mathematicians, Seoul, South Korea (2014).

[42] R. Y. Chen and J. A. Tropp., Subadditivity of matrix ϕ-entropy and concentration of random matrices, Electron. J. Probab., 19(27): 1–30, 2014.

[43] F. R. K. Chung, Spectral graph theory, volume 92. American Mathematical Soc., 1997.

[44] A. Çivril and M. Magdon-Ismail, Column subset selection via sparse approximation of SVD, Theoret. Comput. Sci. 421 (2012), 1–14.

[45] A. Comtet, S. N. Majumdar and J. Randon-Furling, Random Convex Hulls and Extreme Value Statistics, Journal of Statistical Physics, Volume 138, Issue 6, pp 955–1009 (2010).

[46] N.A. Cook, *On the singularity of adjacency matrices for random regular digraphs,* Probab. Theory Related Fields, 167(1) (2017), 143–200.

[47] N. Cook, The circular law for random regular digraphs, to appear in Ann. Inst. Henri Poincaré Probab. Stat.

[48] N. Cook, L. Goldstein, T. Johnson, Size biased couplings and the spectral gap for random regular graphs, Ann. Probab.46 (1), 72–125 (2018).

[49] S. Coste, The spectral gap of sparse random digraphs, arXiv:1708.00530.

[50] K.P. Costello and V. Vu, The rank of random graphs, *Random Structures Algorithms* 33 (2008), no. 3, 269–285.

[51] A. Dvoretzky, Some results on convex bodies and Banach spaces, in Proc. Internat. Sympos. Linear Spaces, Jerusalem, 1960, Jerusalem Academic Press, Jerusalem, Pergamon, Oxford, 1961, pp. 123–160.

[52] R. Eldan, Volumetric properties of the convex hull of an n-dimensional Brownian motion, Electron. J. Probab. 19 no. 45, 34.

[53] R. Eldan, Extremal points of high dimensional random walks and mixing times of a Brownian motion on the sphere, Ann. Inst. Henri Poincaré Probab. Stat. 50 (2014) 95–110.

[54] N. Enriquez and L. Ménard, Spectra of large diluted but bushy random graphs, Random Structures and Algorithms, Volume 49, Issue 1 (2016), pp 160–184.

[55] L. Erdős and H.-T, Yau, A dynamical approach to random matrix theory, In: Courant Lecture Notes in Mathematics, Courant Institute of Mathematical Sciences, vol. 28. American Mathematical Society, New York, Providence, RI (2017).

[56] U. Feige and E. Ofek, Spectral techniques applied to sparse random graphs, Random Structures and Algorithms, 27(2):251–275, 2005.

[57] D. A. Freedman, On tail probabilities for martingales, Ann. Probability **3** (1975), 100–118. MR0380971

[58] J. Friedman, A proof of Alon's second eigenvalue conjecture and related problems, Mem. Amer. Math. Soc. **195** (2008), no. 910, viii+100 pp. MR2437174

[59] J. Friedman, J. Kahn, E. Szemerédi, On the second eigenvalue of random regular graphs, Proceedings of the twenty-first annual ACM symposium on Theory of computing (1989), 587–598.

[60] A. Frieze, *Random structures and algorithms*, Proceedings ICM, Vol. 1, 2014, 311–340.

[61] Z. Fıuredi and J. Komlòs, The eigenvalues of random symmetric matrices, Combinatorica, 1(3):233–241, 1981.

[62] L. Geisinger, Convergence of the density of states and delocalization of eigenvectors on random regular graphs. J. Spectr. Theory **5** (2015), no. 4, 783–827.

[63] S. Geman, The spectral radius of large random matrices, Ann. Probab. 14 (1986), 1318–1328.

[64] S. Geman and C.-R. Hwang, A chaos hypothesis for some large systems of random equations, Z. Wahrsch. Verw. Gebiete 60 (1982), 291–314.

[65] A. A. Giannopoulos, A note on the Banach-Mazur distance to the cube, In Geometric aspects of functional analysis (Israel, 1992–1994), volume 77 of Oper. Theory Adv. Appl., pages 67–73. Birkhäuser, Basel, 1995.

[66] A. A. Giannopoulos, A proportional Dvoretzky-Rogers factorization result, Proc. Amer. Math. Soc., 124(1): 233–241, 1996.

[67] A. Giannopoulos and V. Milman, Low M^*-estimates on coordinate subspaces, J. Funct. Anal. 147 (1997), 457–484.

[68] V. L. Girko, The circular law, Teor. Veroyatnost. i Primenen. **29** (1984), no. 4, 669–679. MR0773436

[69] E. D. Gluskin, The diameter of the Minkowski compactum is roughly equal to n, Funktsional. Anal. i Prilozhen. 15 (1981), no. 1, 72–73.

[70] F. Götze and A. Tikhomirov, The circular law for random matrices, Ann. Probab.38 (2010), no. 4, p. 1444–1491.

[71] Y. Gordon, On MilmanÕs inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$, In Geometric Aspects of Functional Analysis (1986/87), Lecture Notes in Math. 1317, 84–106. Springer, Berlin.

[72] O. Guédon, A. Hinrichs, A. E. Litvak, J. Prochno, On the expectation of operator norms of random matrices, GAFA, Lecture Notes in Math., 2169 (2017), 151–162.

[73] N. Halko, P.G. Martinsson, and J.A. Tropp, Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions, SIAM Rev. 53 (2011), no. 2, 217–288.

[74] Y. He, A. Knowles and M. Marcozzi, Local law and complete eigenvector delocalization for supercritical Erdős-Renyi graphs, to appear in Ann. Prob.

[75] J. Hermon and J. Salez, Modified log-Sobolev inequalities for strong-Rayleigh measures, arXiv:1902.02775.

[76] J.T. Holodnak and I.C.F. Ipsen, Randomized approximation of the Gram matrix: exact computation and probabilistic bounds, SIAM J. Matrix Anal. Appl. 36 (2015), no. 1, 110–137.

[77] S. Hoory, N. Linial, and A. Wigderson, Expander graphs and their applications, Bull. Amer.Math. Soc.43(2006), no. 4, 439–561.

[78] J. Huang, Invertibility of adjacency matrices for random d-regular graphs, arXiv:1807.06465.

[79] F. John, Extremum problems with inequalities as subsidiary conditions. In Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948, pages 187–204. Interscience Publishers, Inc., New York, N. Y., 1948

[80] Z. Kabluchko, V. Vysotsky, and D. Zaporozhets, Convex hulls of random walks, hyperplane arrangements, and Weyl chambers, Geom. Funct. Anal. 27 (2017) 880–918.

[81] Z. Kabluchko, V. Vysotsky, and D. Zaporozhets, Convex hulls of random walks: Expected number of faces and face probabilities, Adv. Math. 320 (2017) 595–629.

[82] J. Kahn, J. Komlós and E. Szemerédi, *On the probability that a random* ± 1*-matrix is singular,* J. Amer. Math. Soc. 8 (1995), no. 1, 223–240.

[83] J. Komlós, *On the determinant of* $(0, 1)$ *matrices,* Studia Sci. Math. Hungar 2 (1967), 7–21.

[84] J. Komlós, Circulated Manuscript, Available online at `http://math.rutgers.edu/~komlos/01short.pdf`.

[85] M. Krivelevich, B. Sudakov, The largest eigenvalue of sparse random graphs, Combin Probab Comput,12:61–72, 2003.

[86] R. Latała, Some estimates of norms of random matrices, Proc. Amer. Math. Soc., 133(5):1273–1282 (electronic), 2005.

[87] C. Le, E. Levina, and R. Vershynin, Concentration and regularization of random graphs, Random Structures and Algorithms, 51(3):538–561, 2017.

[88] M. Ledoux, The Concentration of Measure Phenomenon, American Mathematical Society, Providence, RI

[89] A.E. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes,* Adv. Math., **195** (2005), 491–523.

[90] A. Lubotzky, R. Phillips, and P. Sarnak, Ramanujan graphs, Combinatorica, 8(3): 261–277,1988.

[91] F. Lust-Piquard, Inégalités de Khintchine dans C_p $(1 < p < \infty)$, C. R. Acad. Sci. Paris Sér. I Math. 303 (7) (1986) 289–292 (in French).

[92] A. W. Marcus, D. A. Spielman, and N. Srivastava. Interlacing families I: Bipartite Ramanujangraphs of all degrees.Ann. of Math. (2), 182(1): 307–325, 2015.

[93] A.W. Marcus, D.A. Spielman, and N. Srivastava, Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem, Ann. of Math. (2) 182 (2015), no. 1, 327–350.

[94] A. W. Marcus, D. A. Spielman, and N. Srivastava, Ramanujan graphs and the solution of the Kadison–Singer problem. In Proceedings of the 2014 International Congress of Mathematicians. Volume III, pages 363–386. 2014.

[95] A.W. Marcus, D.A. Spielman, and N. Srivastava, Interlacing Families III: Sharper Restricted Invertibility Estimates, to appear in Israel Journal of Mathematics.

[96] G. A. Margulis, Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators, Problemy Peredachi Informatsii, 24(1): 51–60, 1988.

[97] B. D. McKay, The expected eigenvalue distribution of a large regular graph, Linear Algebra Appl. 40 (1981), p. 203–216.

[98] B. D. McKay, Subgraphs of random graphs with specied degrees, Congr. Numer. 33 (1981), 213–223.

[99] M. L. Mehta, *Random matrices*, third edition, Pure and Applied Mathematics (Amsterdam), 142, Elsevier/Academic Press, Amsterdam, 2004.

[100] F. Merlevède, M. Peligrad and E. Rio, Bernstein inequality and moderate deviations under strong mixing conditions, High dimensional probability V: the Luminy volume, 273–292, Inst. Math. Stat. Collect., 5, Inst. Math. Statist., Beachwood, OH.

[101] V. D. Milman, A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies, Funktsional. Anal. i Prilozhen. 5 (4) (1971) 28–37.

[102] V. D. Milman and G. Schechtman, Asymptotic theory of finite-dimensional normed spaces, Lecture Notes in Mathematics, 1200, Springer, Berlin, 1986.

[103] A. Naor, Sparse quadratic forms and their geometric applications (after Batson, Spielman and Srivastava), Asterisque volume 348 (2012), pages 189–217.

[104] H.H. Nguyen, Random doubly stochastic matrices: the circular law, Ann. Probab. 42(3), 1161–1196 (2014).

[105] H. H. Nguyen and W. M. Matchett, Cokernels of adjacency matrices of random r-regular graphs, arXiv:1806.10068.

[106] H.H. Nguyen and V.H. Vu, Circular law for random discrete matrices of given row sum, J. Comb. 4(1), 1–30 (2013).

[107] L. Pastur and M. Shcherbina, *Eigenvalue distribution of large random matrices*, Mathematical Surveys and Monographs, 171, American Mathematical Society, Providence, RI, 2011.

[108] S. Péché, Deformed ensembles of random matrices, Proceedings of the ICM-Seoul 2014, Vol. III, 1159–1174.

[109] A. Pelczynski, Structural theory of Banach spaces and its interplay with analysis and probability, Proceedings of the International Congress of Mathematics 1983, PWN-North Holland (1984), 237–269.

[110] R. Pemantle, Towards a theory of negative dependence, J. Math. Phys., 41(3): 1371–1390, 2000. Probabilistic techniques in equilibrium and nonequilibrium statistical physics.

[111] R. Pemantle and Y. Peres, Concentration of Lipschitz functionals of determinantal and other strong Rayleigh measures, Combin. Probab. Comput., 23(1): 140–160, 2014.

[112] G. Pisier, The volume of convex bodies and Banach space geometry, Cambridge Tracts in Mathematics, 94, Cambridge Univ. Press, Cambridge, 1989.

[113] G. Pisier, Remarks on the non-commutative Khintchine inequalities for $0 < p < 2$, J. Func. Anal. 256 (2009) 4128–4161.

[114] L. Rebrova, Constructive regularization of the random matrix norm, preprint, available at arXiv:1809.03926.

[115] L. Rebrova and R. Vershynin, Norms of random matrices: local and global problems, Adv. Math. (2018), no. 324, pp 40–83.

[116] M. Reitzner, Random polytopes, Survey in Molchanov, I., and Kendall, W. (eds.): New perspectives in stochastic geometry. p. 45–76, Oxford University Press, Oxford (2010).

[117] M. Rudelson, Contact points of convex bodies, Israel J. Math. 101 (1997), 93–124.

[118] M. Rudelson, Random vectors in the isotropic position, J. Funct. Anal. 164, no. 1 (1999): 60–72.

[119] M. Rudelson, Recent developments in non-asymptotic theory of random matrices, Modern aspects of random matrix theory. In: Proc. Sympos. Appl. Math., Vol. 72. Amer. Math. Soc., Providence, (2014), pp. 83–120.

[120] M. Rudelson and K. Tikhomirov, The sparse circular law under minimal assumptions, to appear in Geom. Func. Anal.

[121] M. Rudelson and R. Vershynin, Sampling from large matrices: an approach through geometric functional analysis, J. ACM 54 (2007), no. 4, Art. 21, 19 pp.

[122] M. Rudelson, R. Vershynin, *The Littlewood-Offord problem and invertibility of random matrices,* Adv. Math. **218** (2008), 600–633.

[123] M. Rudelson and R. Vershynin, Non-asymptotic theory of random matrices: extreme singular values, In: Proceedings of the International Congress of Mathematicians, Vol. III. Hindustan Book Agency, New Delhi, (2010), 1576–1602.

[124] M. Rudelson and R. Vershynin, Delocalization of eigenvectors of random matrices with independent entries, Duke Math. J. **164** (2015), 2507–2538. MR3405592

[125] M. Rudelson and R. Vershynin, No-gaps delocalization for general random matrices, Geom. Funct. Anal. **26** (2016), no. 6, 1716–1776. MR3579707

[126] R. Schneider, Recent results on random polytopes, Boll. Unione Mat. Ital. (9) 1, No. 1, 17–39 (2008).

[127] Y. Seginer. The expected norm of random matrices. Combin. Probab. Comput., 9(2):149–166, 2000.

[128] S. Smale, On the efficiency of algorithms of analysis, Bull. Amer. Math. Soc. (N.S.) 13 (1985), 87–121.

[129] D. A. Spielman and N. Srivastava, An elementary proof of the restricted invertibility theorem. Isr. J. Math. 190, 83–91 (2012).

[130] N. Srivastava, On contact points of convex bodies, in Geometric Aspects of Functional Analysis, Springer, Heidelberg, 2012, pp. 393–412.

[131] N. Srivastava and R. Vershynin, Covariance estimation for distributions with $2 + \varepsilon$ moments, Ann. Probab., 41(5): 3081–3111, 2013.

[132] S. J. Szarek, Spaces with large distance to ℓ_∞^n and random matrices, Amer. J. Math., 112(6):899–942, 1990.

[133] S. J. Szarek and M. Talagrand, An "isomorphic" version of the Sauer-Shelah lemma and the Banach-Mazur distance to the cube, In Geometric aspects of functional analysis (1987–88), volume 1376 of Lecture Notes in Math., pages 105–112. Springer, Berlin, 1989.

[134] T. Tao and V.H. Vu, *On the singularity probability of random Bernoulli matrices*, J. Amer. Math. Soc., 20(3):603–628, 2007.

[135] T. Tao and V. Vu, Random matrices: universality of ESDs and the circular law, Ann. Probab.38 (2010),no. 5, p. 2023–2065, With an appendix by Manjunath Krishnapur.

[136] T. Tao and V. Vu, Random matrices: the circular law, Commun. Contemp. Math.10(2008), no. 2, p. 261–307.

[137] T. Tao and V. Vu, From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices, Bull. Amer. Math. Soc. (N.S.) 46 (2009), 377–396.

[138] K. Tikhomirov, On the Banach-Mazur distance to cross-polytope, to appear in Adv. Math., arXiv:1804.08212.

[139] K. Tikhomirov, Singularity of random Bernoulli matrices, arXiv:1812.09016.

[140] L.V. Tran, V.H. Vu and K. Wang, Sparse random graphs: eigenvalues and eigenvectors, Random Structures Algorithms 42 (2013), 110–134.

[141] J. A. Tropp, Column subset selection, matrix factorization, and eigenvalue optimization, Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, PA, 2009, pp. 978–986.

[142] J. A. Tropp, An introduction to matrix concentration inequalities. Foundations and Trends in Machine Learning, 8(1-2): 1–230, 2015.

[143] R. van Handel, On the spectral norm of Gaussian random matrices, Trans. Amer. Math. Soc., 369(11): 8161–8178, 2017.

[144] R. Vershynin, JohnÕs decompositions: selecting a large part, Israel Journal of Mathematics 122 (2001): 253–77.

[145] R. Vershynin, Introduction to the non-asymptotic analysis of random matrices, Compressed sensing. Cambridge Univ. Press, Cambridge (2012), pp. 210–268.

[146] R. Vershynin, High-Dimensional Probability: An Introduction with Applications in Data Sciences, Camb. Ser. Stat. Probab. Math. 47, Cambridge University Press, Cambridge, 2018.

[147] V. Vu, Spectral norm of random matrices, Combinatorica, 27(6):721–736, 2007.

[148] V. Vu, *Random discrete matrices, Horizons of combinatorics,* Bolyai Soc. Math. Stud., 17, 257–280, Springer, Berlin, 2008.

[149] V. Vu, *Combinatorial problems in random matrix theory*, Proceedings ICM, Vol. 4, 2014, 489–508.

[150] E. Wigner, On the distribution of the roots of certain symmetric matrices, Ann. of Math. 67 (1958), no. 2, 325–327.

[151] P. M. Wood, Universality and the circular law for sparse random matrices, The Annals of Applied Probability 22 (2012), 1266–1300.