Random quantum states

and

random matrix theory techniques

in

quantum information theory

Master's Thesis

Michail Louvaris

Advisor: Professor Dimitris Gatzouras

National and Kapodistrian University of Athens

Department of Mathematics

Athens — 2020

Abstract

In this thesis we present results and techniques from random matrix theory, the theory which studies matrices whose entries are random variables, and their connection to quantum information theory.

More precisely, first we define the concept of the empirical spectral distribution of a matrix, which is the uniform discrete measure which is induced by the eigenvalues of the matrix. Next we study the limiting behaviour, in the sense of weak convergence of random variables, of the empirical spectral distribution of random matrices as their size grows. The cases we study, under certain conditions such as i.i.d. entries, finite moments and more, are the following:

1. The case of square symmetric random matrices.

2. The case of the product of a random matrix (not necessary square) with its conjugate transpose matrix, when the dimensions of the matrix are proportional.

3. The case of the product of a random matrix (not necessary square) with its conjugate transpose matrix, when one dimension grows faster than the other.

In each of these cases we show that the limit is a probability measure which is absolutely continuous to the Lebesgue measure and has compact support. Moreover, in each of these cases, we prove that the extreme eigenvalues of the matrices converge to the extreme points of the support of the corresponding limit of the empirical spectral distribution of the matrices, when the entries follow the standard Gaussian distribution.

Next we present tools from random matrix theory that are useful in quantum information theory. We define several concepts such as the ∞–Wasserstein distance and the random (quantum) induced states, and prove some of their properties. It is proven that the concept of random quantum state is strongly related with matrices with standard Gaussian entries. Taking advantage of this connection we apply the results of the previous chapter to the study of random quantum states.

In the last chapter, using the results that we obtain for the random quantum states, we prove the existence of a threshold function which depends only on the dimension of the space and separates with high probability the states which are entangled from those that are not entangled, having as a criterion the dimension of the space from which the states have been induced.

Περίληψη

Στην παρούσα διπλωματική εργασία παρουσιάζονται αποτελέσματα και τεχνικές της θεωρίας τυχαίων πινάκων, δηλαδή της θεωρίας που μελετά πίνακες με στοιχεία τυχαίες μεταβλητές, και τη σχέση αυτών με αποτελέσματα της κβαντικής θεωρίας πληροφορίας.

Πιο συγκεκριμένα, αρχικά ορίζεται η έννοια της εμπειρικής φασματικής κατανομής ενός πίνακα, που είναι το ομοιόμορφο διακριτό μέτρο που επάγεται από τις ιδιοτιμές του. Στη συνέχεια μελετάται το όριο, με την έννοια της ασθενούς σύγκλισης τυχαίων μεταβλητών, της εμπειρικής φασματικής κατανομής τυχαίων πινάκων καθώς η διάστασή τους μεγαλώνει. Οι περιπτώσεις που εξετάζουμε, πάντα υπό κάποιες προϋποθέσεις, π.χ. ανεξαρτησία και ισονομία των στοιχείων του πίνακα, πεπερασμένες ροπές και άλλα, είναι οι ακόλουθες.

1. Η περίπτωση των τυχαίων τετραγωνικών συμμετρικών πινάκων.

2. Η περίπτωση του γινομένου ενός τυχαίου πίνακα (όχι κατ' ανάγκην τετραγωνικού) με τον ανάστροφό του, ή τον συζυγή ανάστροφό του αντίστοιχα, υποθέτοντας ότι οι διαστάσεις του είναι ανάλογες.

3. Η περίπτωση του γινομένου ενός τυχαίου πίνακα (όχι κατ' ανάγκην τετραγωνικού) με τον ανάστροφό του, όταν η μία διάσταση μεγαλώνει πολύ γρηγορότερα από την άλλη.

Σε κάθε μία περίπτωση αποδεικνύεται ότι το όριο είναι ένα μέτρο πιθανότητας, απόλυτα συνεχές ως προς το μέτρο Lebesgue, με φραγμένο φορέα. Παράλληλα, σε κάθε μία από αυτές τις περιπτώσεις αποδεικνύουμε και την σύγκλιση της μεγαλύτερης και της μικρότερης ιδιοτιμής των τυχαίων πινάκων

στα αντίστοιχα άκρα του φορέα του ορίου, με την υπόθεση ότι τα στοιχεία του πίνακα ακολουθούν την τυπική κανονική κατανομή.

Στη συνέχεια παρουσιάζονται εργαλεία της θεωρίας τυχαίων πινάκων που είναι χρήσιμα στην περιοχή της κβαντικής θεωρίας πληροφορίας. Ορίζονται οι έννοιες της ∞-απόστασης Wasserstein και της τυχαίας επαγόμενης (κβαντικής) κατάστασης, και αποδεικνύονται ιδιότητές τους. Αποδεικνύεται συγκεκριμένα πως η έννοια της τυχαίας κβαντικής κατάστασης συνδέεται ισχυρά με τυχαίους πίνακες με στοιχεία που ακολουθούν την κανονική κατανομή. Εκμεταλλευόμενοι αυτή τη σύνδεση εφαρμόζουμε τα αποτελέσματα των προηγούμενων κεφαλαίων στις τυχαίες κβαντικές καταστάσεις.

Στο τελευταίο κεφάλαιο χρησιμοποιώντας τα αποτελέσματα για τις τυχαίες κβαντικές καταστάσεις αποδεικνύουμε την ύπαρξη συνάρτησης (threshold) που εξαρτάται από την διάσταση του χώρου και χωρίζει με μεγάλη πιθανότητα τις καταστάσεις που είναι entangled από αυτές που δεν είναι, με βάση την διάσταση του περιβάλλοντος από το οποίο έχουν επαχθεί.

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε στα πλαίσια της απόκτησης του Διπλώματος Μεταπτυχιακών Σπουδών με ειδίκευση στην Κατεύθυνση των Θεωρητικών Μαθηματικών. Η τριμελής επιτροπή απαρτίζεται από τους κ.κ. Δημήτρη Γατζούρα, Αριστείδη Κατάβολο και Παντελή Δοδό.

Θα ήθελα να ευχαριστήσω όλα τα μέλη της επιτροπής για την συμμετοχή τους σε αυτήν. Επιπλέον θα ήθελα να ευχαριστήσω τους κ. Αριστείδη Κατάβολο και κ. Παντελή Δοδό για τα μαθήματα και τα σεμινάρια που διοργάνωσαν και παρέδωσαν, κατά τη διάρκεια του πρώτου και δεύτερου κύκλου σπουδών μου, τα οποία έπαιξαν σημαντικό ρόλο στην πορεία μου στο Μαθηματικό.

Ιδιαίτερα θα ήθελα να ευχαριστήσω τον κ. Δημήτρη Γατζούρα που με εισήγαγε στις περιοχές της Μαθηματικής Ανάλυσης και της Θεωρίας Πιθανοτήτων μέσα από τα προπτυχιακά μαθήματα «Πραγματική Ανάλυση» και «Θεωρία Μέτρου», και το μεταπτυχιακό μάθημα «Εργοδική Θεωρία», τα οποία δίδαξε σε πολύ υψηλό επίπεδο και με τρόπο τέτοιο που με ενέπνευσε να ασχοληθώ παραπάνω με αυτή την περιοχή των Μαθηματικών. Επίσης τον ευχαριστώ για τις ώρες που αφιέρωσε σε αυτή τη διπλωματική εργασία, την καθοδήγηση και τις ιδιαίτερα βοηθητικές παρατηρήσεις του, και συνολικά για το ειλικρινές ενδιαφέρον που έδειξε κατά τη διάρκεια των σπουδών μου.

Επίσης θα ήθελα να ευχαριστήσω τον κ. Γιαννόπουλο για την βοήθειά του και τις χρήσιμες παρατηρήσεις του στη διπλωματική αυτή, καθώς και για την υψηλής ποιότητας διδασκαλία του στα μαθήματα που παρακολούθησα μαζί του.

Τέλος θα ήθελα να ευχαριστήσω την οικογένεια και τους φίλους μου για την στήριξη τους. Ειδικότερα θα ήθελα να ευχαριστήσω τους φίλους μου Αλέξαν-

δρο και Μάνο για την συμπαράσταση και την συνεργασία που είχαμε, και ελπίζω να ξαναέχουμε, κατά τη διάρκεια των σπουδών μας, και τη Δήμητρα για την αναγκαία στήριξή της όλον αυτόν τον καιρό.

Μιχαήλ Λούβαρης
Αθήνα, 2020

Στην μνήμη του καθηγητή μου Δ.Γατζούρα

Contents

Introduction

In this thesis we are going to study the interface between random matrix theory and quantum information theory.

Historical background and general description

Classical information theory studies the transmission, processing, extraction, and utilization of information. Abstractly, information can be thought of as the resolution of uncertainty. As quantum mechanics progressed, several information theoretic concepts were introduced such as quantum information, which is the information of the state of a quantum system, and played important role in the area. This is hardly surprising, since quantum mechanics, as usually presented, is a probabilistic theory.

However, in the 1990s quantum information theory emerged as a distinct discipline. Moreover, as quantum information theory has been progressing it has been characterized as the mathematical framework necessary for the building of a quantum computer.

On the other hand, in 1955 the nuclear physicist E. Wigner [1] introduced the concept of random matrices (i.e. matrices whose entries are random variables) making the assumption that the spacings between the lines in the spectrum of a heavy atom nucleus should resemble the spacings between the eigenvalues of a random matrix, and should depend only on the symmetry class of the underlying evolution. By that, the mathematical field of random matrix theory was born and since then it has been connected with several research areas such as asymptotic geometric analysis (and more precisely high-dimensional probability), physics, numerical analysis, mathematical statistics, theoretical neuroscience, optimal con-

trol and more.

Over the last dozen or so years, it has become clear that quantum information theory is closely linked to geometric functional analysis (Banach space theory, operator spaces, high-dimensional probability) and random matrix theory. In this thesis we study the interface between quantum information theory and random matrix theory.

We have separated the thesis in three parts.

Part 1

Both quantum information theory and random matrix theory use tools from several research areas of mathematics such as geometric functional analysis, combinatorics, probability theory, operator theory and linear algebra. In the first part of this thesis we present all these tools and give a few proofs. We have avoided to give extended proofs in this part because if we did we would lose focus on the main goal of the thesis. More precisely this part contains the following:

· In both random matrix theory and in quantum information theory matrices play a crucial role. So it is only natural that we introduce the appropriate norms (the analogues of the ℓ_p norms in $\mathbb{R}^n$ or $\mathbb{C}^n$) on the matrix spaces.

· Next we present Dirac's Bra-Ket notation which is a well known way to denote elements in quantum mechanics.

· Probability theory is in the core of this thesis so several "classical" results such as the Borel-Cantelli lemma and Fubini-Tonelli theorem are presented since they are necessary.

· In the same section we introduce the concept of a probability metric space, i.e a metric space equipped with the Borel σ-algebra (the smallest σ-algebra that contains the open sets produced by the metric) and a probability measure defined on that σ-algebra.

So one may define the space $\mathbb{P}(X, d)$ of the (Borel) probability measures of a metric space (X,d). It has been proven (see [2]) that this space is a subspace of the dual of the space of continuous and bounded functions with the sup $-$norm. On this space one can construct a metric which can

metricize weak convergence, i.e the convergence with respect to the weak topology. Note that, when we are in $\mathbb{R}$, weak convergence is the convergence in distribution that we have seen in probability theory.

We also present Skorohod's theorem, a beautiful theorem which converts the convergence in distribution to almost sure convergence and Haar's theorem which states that every locally-compact group has a unique measure which is invariant under multiplication.

· As mentioned above, geometric functional analysis and high-dimensional probability are strongly related to both random matrix theory and quantum information theory. So it is natural to use several results from that area such as the *isoperimetric inequality on the sphere*: The n-dimensional sphere has a unique probability measure invariant under orthogonal transformations. There are various ways to define this probability measure but from a probabilistic point of view it can be defined as follows: Let $\{X_i\}_{i\in[n]}$ be i.i.d. random variables all following $N(0,1)$. Then

$$s^{n-1}(A) = \mathbb{P}\left(\frac{1}{(\sum_{i=1}^n X_i^2)^{1/2}}(X_1, X_2, \ldots, X_n) \in A\right).$$

The isoperimetric inequality on the sphere states that if C is a ball with respect to the geodesic metric on the sphere then

$$s^{n-1}(C_\epsilon) \leq s^{n-1}(A_\epsilon)\ , \forall A \subseteq S^{n-1} : s^{n-1}(A) = s^{n-1}(C),$$

where $A_\epsilon = \{x : \mathrm{dist}(x, A) < \epsilon\}$ is the ϵ-extension of A.

After the isoperimetric inequality we also present the well-known Krein-Milman theorem which states that for a convex and compact set K we have that

$$K = \overline{\mathrm{conv}(\mathrm{ext}(K))}.$$

Next we define several ways to "measure the size" of convex sets and mention some notions and inequalities about the volume radius, the mean width and the Gaussian mean width of sets.

The last result that we need from geometric functional analysis concerns the ℓ-norm, the ℓ–position and and the MM^* estimate for convex bodies. The ℓ-norm is defined for all matrices of $\mathbb{R}^n$ (or $\mathbb{C}^n$) as follows: if G is an

n-dimensional vector whose coordinates are i.i.d. random variables all following $N(0, 1)$ and K is a convex body containing 0 in its interior then

$$\ell_K(M) = \mathbb{E}\|T(G)\|_K.$$

The ℓ-position is a special position of a convex body: we say that a convex body in $\mathbb{R}^n$ is in ℓ-position iff the (unique) positive semi-definite matrix M of largest determinant among all matrices in the unit ball with respect to the ℓ_K-norm is a multiple of the identity matrix. A useful property of convex bodies which are in the $\ell-$position is that

$$1 \leq w(K)w(K^\circ) \leq C \log n$$

where $w(K)$ denotes the mean width of K.

· Next we present the tensor product of two Hilbert spaces. The tensor product of Hilbert spaces is the appropriate way to define the states used in quantum mechanics.

The physical phenomena that characterize quantum states such as entanglement and separability are all defined in this section as well.

Despite the way they are defined, in the rest of the thesis we will try to avoid tensor products, as much as we can, taking advantage of the following important property. If H_1, H_2 are two Hilbert spaces then

$$H_1 \otimes H_2 = B(H_1, H_2)$$

where $B(H_1, H_2)$ denotes the space of linear operators from H_1 to H_2.

So, since we will work on multi-dimensional complex spaces, we will translate states into complex matrices.

· In the last section of the first part we present tools from graph theory and combinatorics such as simple graphs, trees, bipartite graphs and Hall's theorem.

Part 2

In the second part of the thesis we present and prove several important results from random matrix theory. First we introduce the concept of the

empirical spectral distribution (E.S.D.) of an $n \times n$ matrix A which will be denoted by μ_A. More precisely, if $\{\lambda_i(A)\}_{i\in[n]}$ are the eigenvalues of A then

$$\mu_A = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(A)}$$

where δ denotes the Dirac measure. In other words, the empirical spectral distribution of a matrix A is exactly the discrete uniform measure on the set of the eigenvalues of A.

So one may note that if A is a random matrix then μ_A is also a random measure in $\mathbb{P}(\mathbb{R})$, the set of all Borel probability measures on $\mathbb{R}$.

The three results from random matrix theory that we are going to discuss are Wigner's semicircular law [1], the Marchenko-Pastur law [3] and Bai-Yin's theorem on convergence to the semicircular law [4]. The first two are "classical" results in random matrix theory. The third one is not so well-known but it is very useful in our case.

Wigner's semicircular law

In this section we prove the weak convergence of the empirical spectral distribution of a sequence of random symmetric matrices with i.i.d. entries which have finite moments (the theorem has been generalised for entries with are assumed to have only finite second moment, see [5]).

Let A_n be a sequence of matrices as above. Then $\mu_{\frac{A_n}{\sqrt{n}}}$ converges weakly in probability to the semicircular law, i.e the measure σ with density (with respect to the Lebesgue measure)

$$\sigma(x)dx = \frac{\mathbf{1}_{[-2,2]}(x)}{2\pi}\sqrt{4-x^2}.$$

There are mainly two known ways to prove such theorems: the Stieltjes method and the moment method. In this thesis we discuss the moment method.

As the name suggests, the moment method is based on establishing that the k-th moment of the sequence of the E.S.D. $\mu_{A_n/\sqrt{n}}$ converges weakly in probability to the k-th moment of the semicircular law. Taking into account:

(i) that since $\sigma(x)$ has bounded support it is uniquely determined by its moments,

(ii) the Weierstrass approximation theorem which states that there exist polynomials as close as we want to a continuous bounded function,

we prove that the deterministic measure $\mathbb{E}(\mu_{A_n/\sqrt{n}})$ is close (as the dimension grows) to the measure $\mu_{A_n/\sqrt{n}}$ and we reduce the proof of the weak convergence to the semicircular law to the following:

$$\mathbb{E}\left(\int x^k d\mu_{A_n/\sqrt{n}}\right) \to \int x^k d\sigma(x) \ \forall k \in \mathbb{N} \text{ in probability.}$$

But it easy to compute that the moments of the semicircular law are

$$\int x^k d\sigma = \begin{cases} 0 & \text{if } k \text{ is odd} \\ C_{\frac{k}{2}} & \text{if } k \text{ is even,} \end{cases}$$

where

$$C_n = \frac{1}{n+1}\frac{(2n)!}{(n!)^2}.$$

So, using combinatorial analysis we prove that the limit of the k-th moment (when k is even) of the E.S.D.'s is in fact the cardinality of the set of all the sequences $\{a_j\}_{j\in[k]}$ with k elements which are all either $+1$'s or -1's such that $\forall j \in [k-1] \sum_{m=1}^{j} a_j \geq 0, \sum_{j=1}^{k} a_j = 0$.

Lastly we prove that the cardinality of the set of the sequences mentioned above is given exactly by the sequence of Catalan numbers, which ends the proof of the semicircular law.

CONVERGENCE OF THE EXTREME EIGENVALUES

The next main result that we prove in the thesis is the convergence of the extreme eigenvalues of matrices seen in the Wigner's semicircular law (in the case where the entries are standard Gaussian random variables) to 2 and -2 respectively. In order to do that, we prove two well-known inequalities for the Gaussian measure:

(i) The Gaussian isoperimetric inequality, a result analogous to the spherical isoperimetric inequality, and a consequence of it, which

states that if

$$\gamma_n(A) = \gamma_1((-\infty, a])$$

then for any $\epsilon > 0$

$$\gamma_n(A_\epsilon) \geq \gamma_1(-\infty, a + \epsilon),$$

where $\forall m \in \mathbb{N}$, $\gamma_m(A) = \mathbb{P}((X_1, \ldots, X_m \in A)$, where $X_1, \ldots, X_m$ are standard Gaussian random variables.

(ii) Erhard's inequality which states that for any pair of Borel sets A, B in $\mathbb{R}^n$

$$\Phi^{-1}(\gamma_n(\lambda A + (1 - \lambda)B)) \geq \lambda\Phi^{-1}(\gamma_n(A)) + (1 - \lambda)\Phi^{-1}(\gamma_n(B)),$$

where $\Phi(x)$ is the distribution function of a standard Gaussian random variable.

Using the results above we prove that for any 1-Lipschitz function f, if M_f is its median then

$$M_f \leq E_{\gamma_n}(f),$$

and a concentration inequality for a 1-Lipschitz function and its median.

Applying these results for $f = \|\cdot\|_\infty$ we prove the convergence of the extreme eigenvalues.

MARCHENKO-PASTUR LAW

In this case we prove the weak convergence of the E.S.D. of a sequence of random matrices $X_{p\times n}X^*_{n\times p}/n$, where $p/n \to y \in (0, 1]$ and the entries of $X_{p\times n}$ are i.i.d. random variables with finite moments, to a deterministic measure μ that has density (with respect to the Lebesgue measure)

$$d\mu = \frac{1}{2\pi xy}\sqrt{(b - x)(x - a)}\mathbf{1}_{a\leq x\leq b},$$

where

$$a(y) = (1 - \sqrt{y})^2, b(y) = (1 + \sqrt{y})^2.$$

Note that a similar result is true when y $\in (1, \infty)$ under the weaker assumption that just the second moment of the entries of $X_{p\times n}$ is finite.

The methods that are used in order to prove the M-P law are very similar to those used for Wigner's semicircular law. Again we use the moment method and prove that every moment of the E.S.D. tends to the respective moment of μ.

Convergence of the extreme eigenvalues in M-P

As in the Wigner's law case we prove the convergence of the extreme eigenvalues of matrices in the M-P law, when the entries are Gaussian, to $a(y)$ and $b(y)$ respectively. The proof can be found in [6]. It is done by working with a more convenient matrix Y which has the same eigenvalues as X.

The tools that are used in the proof include results about the $\chi-$squared distribution and the Gershgorin circle theorem which states that, for any complex matrix A, every eigenvalue of it lies into a circle whose radius is the sum of the 2-norms of the elements of some of the rows of the matrix.

Bai-Yin's convergence to the semicircular law

In this section we prove another theorem concerning the weak convergence of the E.S.D. of a sequence of random matrices $A_p = \frac{1}{2\sqrt{np}}(X_pX_p^* - n(p)I_p)$ to the semicircular distribution.

Here X_p is a $p \times n(p)$ random matrix with i.i.d. entries with variance 1 and finite fourth moment. Also $n(p), p \to \infty$ and $p/n(p) \to 0$.

In order to prove the convergence we prove several lemmas. The most crucial one simplifies the random matrices we work with. The precise statement is as follows:

Let Y_p be a sequence $p \times n$ random matrices with i.i.d. entries such that

(i) $\mathbb{E}Y_{1,1} = 0$ and $\mathbb{E}Y_{1,1}^2 = 1 + a_p$, where $a_p \to 0$ as $p \to \infty$, and

(ii) $|Y_{1,1}| \leq \epsilon_p n^{1/4}$, where $\epsilon_p \to 0$ and $\epsilon_p n^{1/4} \to \infty$.

Then the matrix Z_p with

$$Z_{i,i} = 0$$

and

$$Z_{i,j} = \frac{1}{2\sqrt{np}} \sum_{l_p} Y_{i,l_p} Y_{j,l_p} \qquad \text{when } i \neq j$$

has E.S.D. that converges to the semicircular distribution. For the proof we use similar methods (the combinatorial approach, the moment method) as the ones we used for Wigner's semicircular law.

Next we prove several tools which we use in order to prove that the convergence of the simplified matrices Y_p to the semicircular law is sufficient for the convergence of the E.S.D. of X_p.

More specifically we prove that the E.S.D.'s of the truncated and centered matrices X_p', i.e. the matrices with entries $X_{i,j}1_{|X_{i,j}|<\epsilon_p n_p} - \mathbb{E}(X_{i,j}1_{|X_{i,j}|<\epsilon_p n_p})$ where $X_{i,j}$ is an entry of X_p, have the same limiting behaviour as the E.S.D. of X_p. The proof is completed by combining the simplified lemma and the previous fact. The complete proof can be found in [4].

Convergence of the extreme eigenvalues in Bai-Yin's case

Like in the Marchenko-Pastur case and Wigner's case we prove the convergence of the extreme eigenvalues of matrices in Bai-Yin's case, when the entries are standard Gaussian, to -2 and 2 respectively.

In order to do that, we use the same method we used in the Marchenko-Pastur case and work with more convenient matrices. This way we show that it is sufficient to prove the convergence of the extreme eigenvalues when the matrices have entries that are **real** standard Gaussian.

The proof is completed by the following very important lemmas from high-dimensional probability:

(i) (Slepian's inequality) Let $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ be two Gaussian processes such that for any $t, s \in T$

$$\mathbb{E}(Y_t - Y_s)^2 \leq \mathbb{E}(X_t - X_s)^2$$

and

$$\mathbb{E}X_t^2 = \mathbb{E}Y_t^2.$$

Then $\forall x \in \mathbb{R}$

$$\mathbb{P}(\sup_{t\in T} X_t \geq x) \leq \mathbb{P}(\sup_{t\in T} Y_t \geq x),$$

which implies that

$$\mathbb{E}\sup_{t\in T} X_t \leq \mathbb{E}\sup_{t\in T} Y_t.$$

(ii) (Gaussian interpolation) Consider two independent n-dimensional real random vectors $X \sim N(0, \Sigma^X)$ and $Y \sim N(0, \Sigma^Y)$. Then define the Gaussian vector

$$Z(u) = \sqrt{u}X + \sqrt{1-u}Y \qquad u \in [0,1]$$

For any $f : \mathbb{R}^n \to \mathbb{R}$ which is twice diffentiable, it is true that

$$\frac{d}{du}\mathbb{E}(f(Z(u)) = \frac{1}{2}\sum_{i,j}^{n}(\Sigma^X_{i,j} - \Sigma^Y_{i,j})\mathbb{E}\frac{d^2}{dx_i dx_j}(Z(u)).$$

(iii) (Chevet-Gordon inequalities) Let $B \in M_{p,n}$ be a random matrix with independent $N(0,1)$ entries. Let $K \subseteq \mathbb{R}^n$ and $L \subseteq S^{p-1}$ be compact sets and $r_k > 0$ such that $K \subseteq r_k B_2^n$. Then

$$\mathbb{E}\max_{u\in L}\max_{t\in K}\langle Bt, u\rangle \leq w_G(K) + r_k w_G(L),$$

where w_G denotes the Gaussian-mean width of a set.

Combining the previous facts we complete the proof.

Part 3

In this part we introduce and prove several important tools used in quantum information theory and then use the theorems from Part 2 in order to prove a threshold theorem.

RANDOM MATRICES IN QUANTUM INFORMATION THEORY

The first tool we introduce is the $\infty-$Wasserstein distance. It is defined as follows: for two probability measures μ_1, μ_2,

$$d_\infty(\mu_1, \mu_2) := \inf \|\mu_1 - \mu\|_{L_\infty},$$

where the infimum is over all couples (X_1, X_2) of random variables with (marginal) laws μ_1, μ_2 defined on a common probability space. It is shown that convergence with respect to d_∞ of a sequence of random variables to a random variable with compact support (say $[a, b]$) is equivalent to the weak convergence of these random variables and the convergence of the inf and sup of the random variables to a and b respectively.

Next we introduce two models of random states. The most important is the **random induced state**. Although this model is defined as the partial trace of a random variable uniformly distributed on the sphere of $\mathbb{C}^n \otimes \mathbb{C}^s$, we prove and use that the random induced state has distribution $\frac{W_{n,s}}{\mathrm{tr}(W_{n,s})}$ where $W_{n,s} = BB^*$ and B is an $n \times s$ random matrix with i.i.d. entries all following $N_{\mathbb{C}}(0,1)$.

We prove an important concentration result for $\chi-$squared distribution, which shows that the element $\mathrm{tr}(W_{n,s})$ can be virtually treated as a constant. So we conclude to the following result:

Let $\mu(A)$ denote the E.S.D. of a matrix A and let $\rho_{n,s}$ be the distribution of a random $n \times s$ induced state. If $s/n \to \lambda \in (0,\infty)$ then $\mu_{n,s}(s\rho_{n,s})$ converges with respect to the $\infty-$Wasserstein distance to the Marchenko-Pastur distribution (a consequence of the M-P theorem and the convergence of the extreme eigenvalues).

Likewise, if $s/n, s \to \infty$ then $\mu(\sqrt{ns}(\rho_{n,s} - \frac{I}{n}))$ converges with respect to the $\infty-$Wasserstein distance to the semicircular distribution (a consequence of Bai-Yin's theorem and the convergence of the extreme eigenvalues).

RANDOM QUANTUM STATE

In this section we prove a threshold theorem. For the proof we need the following tools:

1. For any $x, y \in \mathbb{R}^{n,0}$, i.e $\sum_i x_i = \sum_i y_i = 0$ with $y \neq 0$, and for every permutation invariant real convex function φ on $\mathbb{R}^n$ it is true that

$$\varphi(x) \leq \varphi\left(\frac{2n\|x\|_\infty}{\|y\|_1} y\right).$$

2. Let A, B be two unitary invariant random self-adjoint matrices with zero trace that satisfy

$$\mathbb{P}((d_\infty(\mu(A), \mu_{sc}) \leq \epsilon) \geq 1 - p$$

and

$$\mathbb{E}(d_\infty(\mu_{sp}(A), \mu_{sc}) \leq \epsilon,$$

and likewise for B. Here μ_{sc} denotes the distribution of the semicircular law. Then

$$\frac{1-p}{C\epsilon+1}\mathbb{E}\|A\|_K \leq \mathbb{E}\|B\|_K$$

for any convex subset of $M_n^{sa,0}$ (the set of all n-dimensional self-adjoint matrices with zero trace).

Combining the previous tools with the main theorem from the section of random matrices in quantum information theory, we prove that if $n/s, n \to \infty$ and if $\rho_{n,s}$ is the distribution of a random quantum state and A_n is an $n \times n$ random matrix with Gaussian entries and

$$G_n = A_n - \mathrm{tr}(A_n)I,$$

then

$$C_{n,s}^{-1}\mathbb{E}\left\|\frac{G_n}{n\sqrt{s}}\right\| \leq \mathbb{E}\left\|\rho_{n,s} - \frac{I}{n}\right\|_K \leq C_{n,s}\mathbb{E}\left\|\frac{G_n}{n\sqrt{s}}\right\|_K$$

for $C_{n,s} \to 1$.

Next we prove an appropriate form of the well-known concentration inequality, Lévy's inequality. More precisely, let K be a convex body which is a subset of the states of $\mathbb{C}^n$, with inradius r, and let $K_0 = K - \frac{I}{n}$ and $\rho_{n,s}$ be a random induced state. Then, if M is the median of $\|\rho_{n,s} - \frac{I}{n}\|_K$ (and likewise for any central value) we have that for any $\epsilon > 0$

$$\mathbb{P}\left(\left|\left\|\rho_{n,s} - \frac{I}{n}\right\|_{K_0} - M\right| \geq \epsilon\right) \leq \exp(-s) + 2\exp(-\epsilon^2 s r^2 n/72).$$

Combining the above with results from convex geometric analysis and asymptotic geometric analysis (mentioned in Part 1) we conclude the following very important threshold theorem:

Let $s_0(d) := w(\mathrm{Sep}(\mathbb{C}^d \otimes \mathbb{C}^d)^\circ)^2$, where $w(K)$ denotes the mean width of a convex set K and $\mathrm{Sep}(H)$ denotes the set of separable states of a Hilbert space. If ρ is a random induced state of $\mathbb{C}^d \otimes \mathbb{C}^d$, induced by the environment $\mathbb{C}^s$, then for any $\epsilon > 0$ we have that

(i) If $s \leq (1-\epsilon)s_0(d)$ then

$$\mathbb{P}(\rho \text{ is entangled}) \geq 1 - 2\exp(-c(\epsilon)d^3).$$

(ii) If $s \geq (1+\epsilon)s_0(d)$ then

$$\mathbb{P}(\rho \text{ is separable}) \geq 1 - 2\exp(-c(\epsilon)s).$$

The above threshold theorem can be translated as follows:

"Given N identical particles in a generic pure state, if we assign k of them to Alice and k of them to Bob, their shared state suddenly jumps from typically entangled to typically separable when k crosses a certain threshold value $k_N \sim \frac{N}{5}$."

Lastly we give a result of almost sure entanglement of low-dimensional environments which is a consequence of asymptotic geometric analysis. It states that if $s, d \in \mathbb{N}$ are such that $s \leq (d-1)^2$ and if ρ is a random $d^2 \times s$ induced state then

$$\mathbb{P}(\rho \text{ is separable}) = 0.$$

Part I

Preliminaries

Background

1.1 Matrix norms

In this section we present the concept of matrix norms; we give some examples and prove some of their properties. Note that an n-dimensional space of real (or complex) matrices is in fact a real (or complex) vector space of dimension n^2. But since the space of matrices is equipped with an additional operation, multiplication of matrices, one may use a slightly different method to estimate matrices.

Definition 1.1.1. Throughout this thesis we will use the following notations.

- $M_{n,m}$ for the class of $n \times m$, either real or complex, matrices and M_n for the class of $n \times n$ matrices.
- $M_n^{sa}(\mathbb{C})$ for the class of self-adjoint complex matrices. Note that $M_n^{sa}(\mathbb{C})$ is in fact an n^2-dimensional **real** vector space.
- A^T for the transpose of a matrix A and A^* for the conjugate transpose of A. Note that $A^* = A^T$ when A has real entries.
- Given a finite dimensional complex or real Hilbert space H, we will denote by $B(H_1, H_2)$ the space of linear maps (operators) from H_1 to H_2 and by $B(H_1)$ the space of linear operators from H_1 to H_1. When $H_1 = \mathbb{C}^n$ and $H_2 = \mathbb{C}^m$ then $B(H_1, H_2)$ can be identified with $M_{m,n}(\mathbb{C})$.

Next we present the ℓ_p-norms in $\mathbb{R}^n$ (equivalently $\mathbb{C}^n$) and then we will present the analogous norms for matrices.

Definition 1.1.2. We define the ℓ_p-norm, $p \in [1, \infty)$, on $\mathbb{R}^n$ by

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

for any $x \in \mathbb{R}^n$, while for $p = \infty$ we set

$$\|x\|_\infty := \max_{i \in [n]} |x_i|,$$

where $[n] = \{1, 2, \ldots, n\}$.

Next we give the definition and/or some properties of matrix norms.

Since M_n is itself a vector space of dimension n^2, one can measure the "size" of a matrix by using any norm on $\mathbb{C}^{n^2}$. However, M_n is not just a high-dimensional vector space; it has a natural multiplication operation, and, when we want to obtain estimates, it is common to relate the "size" of a product AB to the "sizes" of A and B.

Definition 1.1.3. A function $\|\cdot\|\colon M_n \to \mathbb{R}$ is a *matrix norm* if, for all $A, B \in M_n$, the following hold:

1. $\|A\| \geq 0$ and $\|A\| = 0$ if and only if $A = 0$,
2. $\|c \cdot A\| = |c| \cdot \|A\|$, for all $c \in \mathbb{C}$,
3. $\|A + B\| \leq \|A\| + \|B\|$,
4. $\|A \cdot B\| \leqslant \|A\| \cdot \|B\|$.

A matrix norm is sometimes called a *ring norm*. The first three properties of a matrix norm are identical to the axioms for a norm. A norm on matrices that does not satisfy property (4) for all A and B is a *vector norm* on matrices.

We are now ready to present the analogues of the ℓ_p norms for matrices.

Definition 1.1.4. Let $M \in M_{n,m}$ be a real or complex Euclidean space. We will denote $|M| := (M^* M)^{1/2}$. Then we define its Schatten p-norm, $p \in [1, \infty)$, as

$$\|M\|_p := (\mathrm{tr}|M|^p)^{1/p}.$$

Remark 1.1.5. The most commonly used norms in quantum information theory are the following.

- The Schatten 1-norm (the trace norm).
- The Hilbert–Schmidt norm (Frobenius norm) or Schatten 2-norm, which is the analogue of the ℓ_2-norm. In the rest of the thesis we will use this norm if not otherwise specified.
- The Schatten ∞-norm, which can be considered to be the limit of $||M||_p$ as p tends to infinity. This implies that Schatten ∞-norm is the operator norm, meaning

$$||M||_\infty = ||M||_{\text{op}} = \sup_{\{x:\, ||x||_2 \leq 1\}} ||Mx||_2.$$

An equivalent way to define the Schatten p-norms is via the singular values of a matrix M, meaning the eigenvalues of $|M|$. Denote $s(M)$ the singular values of M arranged in non-increasing order. Then

$$||M||_p = ||s(M)||_p \qquad p \geq 1,$$

where on the right-hand side of the equality the norm is the ℓ_p-norm of the vector $s(M)$. By this equivalent definition it easy to show that the Schatten p-norms are in fact matrix norms and that the matrices M and M^* have the same Schatten p-norm (obviously considered as elements of different matrix spaces) since MM^* and M^*M have the same non-zero eigenvalues.

We end this section with a useful tool from linear algebra.

Theorem 1.1.6 (Singular value decomposition)**.** *Let $M \in M_{n,m}$ be a real or complex Euclidean space. Assume $n \leq m$. Then*

$$M = U\Sigma V$$

where U is an $n \times n$ unitary matrix, V is an $m \times m$ unitary matrix and Σ is an $n \times m$ "diagonal" matrix, i.e. $\Sigma_{i,j} = 0$ when $i \neq j$, whose diagonal entries are the singular values of M.

Everything from this section and more about matrices and matrix norms can be found in the first chapters of [7].

1.2 Bra-ket notation

When working with objects related to Hilbert spaces, particularly the complex ones, we use throughout the thesis Dirac's bra-ket notation. This notation generalises the column-row vector convention for elements of real (or complex) spaces.

More precisely, if H is a Hilbert space, then a standard element $x \in H$ is written $|x\rangle$ (a ket-vector). The same element x can be considered as a linear mapping from H to $\mathbb{C}$ which acts on an element $y \in H$ via the scalar product $\langle y, x\rangle$ and then it is being denoted by $\langle x|$.

Moreover, let H_1, H_2 be two finite dimensional Hilbert spaces (real or complex) and let y_1, y_2 be elements of H_1, H_2, respectively. Then we use the notation $|y_1\rangle\langle y_2|$ for the operator $H_2 \to H_1$ which acts on a ket-vector $x \in H_2$ as

$$|x\rangle \to \langle y_2|x\rangle|y_1\rangle$$

or in the standard notation $x \to \langle y_2, x\rangle y_1$.

1.3 Tools from probability theory

In this section we are going to present some tools from probability theory needed in the rest of the thesis.

1.3.1 Weak convergence

Firstly we are going to present some properties of weak convergence of probability measures on metric spaces.

Definition 1.3.1. Let (X, d) be a metric space. We will use the notation $B(X)$ for the Borel σ-algebra (the smallest σ-algebra that contains all open sets of X with respect to the metric d). When (X, d) is separable, then equivalently $B(X)$ is the smallest σ-algebra that contains every open (or closed) ball of X (with respect to the metric d). The definition of Borel sets can be extended to arbitrary topological spaces in a similar way.

Definition 1.3.2. Let (X, d) be a metric space. A function $\mu\colon B(X) \to \mathbb{R}^+$ will be called a Borel probability measure if

$$\mu(\emptyset) = 0, \qquad \mu(X) = 1$$

and

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

whenever $\{A_n\}_{n\in\mathbb{N}} \subseteq B(X)$ is a sequence of pairwise disjoint Borel subsets of X. We will use the notation $\mathbb{P}(X, d)$ for the set of all Borel probability measures of (X, d). We will also use the notation $\mathbb{P}(X)$ when the underlying metric is clear from the context.

Lemma 1.3.3. *Any $\mu \in \mathbb{P}(X)$ has the following properties.*

1. *If $\{A_n\}_{n\in\mathbb{N}}$ is an increasing sequence of Borel sets then*

$$\lim_n \mu(A_n) = \mu\left(\bigcup_n A_n\right).$$

2. *If $\{A_n\}_{n\in\mathbb{N}}$ is a decreasing sequence of Borel sets then*

$$\lim_n \mu(A_n) = \mu\left(\bigcap_n A_n\right).$$

3. *μ is inner regular, meaning that for any $B \in B(X)$,*

$$\mu(B) = \sup\{\mu(C)\colon C \subseteq B, C \text{ closed}\}.$$

4. *μ is outer regular, meaning that for any $B \in B(X)$,*

$$\mu(B) = \inf\{\mu(C)\colon U \supseteq B, U \text{ open}\}.$$

5. *If (X, d) is a compact metric space, then*

$$\mu(A) = \sup\{\mu(K)\colon K \subseteq A, K \text{ compact}\}.$$

We are now ready to give the definition of weak convergence in any metric space.

Definition 1.3.4. Let (X, d) be a metric space. We will use the notation

$$C_b(X) = \{f : X \to \mathbb{R} \mid f \text{ continuous and bounded}\}.$$

Let $\{\mu_n\}_{n\in\mathbb{N}} \subseteq \mathbb{P}(X)$ and $\mu \in \mathbb{P}(X)$. We will say that μ_n *converges weakly* to μ iff

$$\int f d\mu_n \to \int f d\mu \qquad \forall f \in C_b(X).$$

Lemma 1.3.5. *Let (X, d) be a metric space and $\{\mu_n\}_{n\in\mathbb{N}} \subseteq \mathbb{P}(X)$ and $\mu \in \mathbb{P}(X)$. Then the following statements are equivalent.*

(i) *μ_n converges weakly to μ.*

(ii) *$\int g d\mu_n \to \int g d\mu$ for all real uniformly continuous and bounded functions g on X.*

(iii) *$\limsup_n \mu_n(C) \leq \mu(C)$ for every closed $C \subseteq X$.*

(iv) *$\liminf_n \mu_n(U) \geq \mu(U)$ for every open $U \subseteq X$.*

(v) *$\lim_n \mu_n(A) = \mu(A)$ for every Borel set A such that $\mu(\partial A) = 0$, where ∂A denotes the boundary of A.*

Note that weak convergence is in fact the convergence in distribution of real random variables as seen in probability theory.

Definition 1.3.6. Let $\mathbb{P}(X)$ be the set of Borel probability measures of the metric space (X, d). Then the function $d_P : \mathbb{P}(X) \times \mathbb{P}(X) \to [0, +\infty)$ defined by

$$d_P(\mu, \nu) = \inf\{a > 0 : \mu(A) \leq \nu(A_a) + a, \nu(A) \leq \mu(A_a) + a, \forall A \in B(X)\}$$

is called the *Prokhorov metric* on $\mathbb{P}(X, d)$. Here

$$A_a = \{x \in X : d(x, A) < a\}.$$

Obviously if $A \in B(X)$, then $A_a \in B(X)$.

Proposition 1.3.7. *The Prokhorov metric d_P is a metric on the space $\mathbb{P}(X, d)$. Also, if $d_P(\mu_n, \mu) \to 0$ then μ_n converges weakly to μ. Moreover, when (X, d) is separable, then the other direction is also true, i.e.,*

$$\mu_n \text{ converges weakly to } \mu \iff d_P(\mu_n, \mu) \to 0.$$

Note that when $X = \mathbb{R}$ then there exists an equivalent way to metrize weak convergence of real random variables (or equivalently probability measures on $\mathbb{R}$), called *Lévy's distance*, defined as follows

$$d_L(\mu, \nu) = \inf\{\epsilon > 0 \colon \mu(-\infty, t - \epsilon] - \epsilon \leq \nu(-\infty, t] \leq \mu(-\infty, t + \epsilon] + \epsilon \ \forall\, t \in \mathbb{R}\}.$$

One may find the proofs of all the results of this subsection, and more about probability measures on metric spaces, in [2].

1.3.2 Haar measure

Now we proceed with the definition of Haar measure. The following theorem proves that in locally compact topological groups there exists a measure (essentially unique) which is invariant under the group operation.

Definition 1.3.8. A topological group $(G, \cdot)$ is a group with a topology such that the functions

$$G \times G \to G \colon (x, y) \mapsto x \cdot y$$

and

$$G \to G \colon x \mapsto x^{-1}$$

are continuous.

Definition 1.3.9. Let (X, T) be a topological space. Then (X, T) will be called *locally compact* if for every $x \in X$ there exist $U \in T$ and a compact subset K of X such that $x \in U \subseteq K$.

Definition 1.3.10. Let (X, T) be a topological space. Then (X, T) will be called a *Hausdorff* space if for all $x, y \in X$ with $x \neq y$, there exist $U, V \in T$ such that $x \in U$, $y \in V$ and $U \cap V = \emptyset$.

Theorem 1.3.11 (Haar)**.** *Let $(G, \cdot)$ be a locally compact, Hausdorff topological group. Then there exists a Borel measure μ which is invariant under left (right) multiplication, meaning that, for all $A \in B(G)$ and all $g \in G$,*

$$\mu(gA) = \mu(A).$$

Moreover μ is unique in the following sense: if μ, ν are both invariant under left (respectively right) multiplication, then

$$\mu = c\nu$$

for some constant $c \in \mathbb{R}^+$. Finally, the Haar measure is finite if and only if the group G is compact.

A measure which is invariant with respect to left (respectively right) multiplication will be called a left (respectively right) *Haar measure*.

Note that in the case $G = \mathbb{R}^n$ (or equivalently $\mathbb{C}^n$), with addition $\cdot = +$ as the group operation and the usual topology, any Haar measure will be a multiple of the Lebesgue measure.

Proposition 1.3.12. *One may extend Haar's theorem to the space of left (respectively right) cosets of a locally compact Hausdorff topological group, on which the group acts on the left (respectively right) by multiplication.*

One may find a proof of Haar's theorem and more in [8].

1.3.3 Skorohod's theorem

In this subsection we are going to present a beautiful theorem which converts the convergence in distribution of a sequence of real random variables to almost sure convergence.

Theorem 1.3.13 (Skorohod)**.** *Let X_n be a sequence of real random variables that converges weakly to a random variable X, meaning that the Borel probability measures defined as distributions of the random variables X_n converge weakly to the distribution of X. Then there exists a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and random variables Y_n, $n \in \mathbb{N}$, such that*

(i) *the random variables* Y_n, Y *are all defined on the probability space* $(\Omega, \mathcal{A}, \mathbb{P})$;

(ii) $Y_n \to Y$ *almost surely*;

(iii) $Y_n \sim X_n$, $n \in \mathbb{N}$, *and* $Y \sim X$, $\sim$ *meaning that the corresponding random variables have the same distributions*.

An important tool used in the proof of Skorohod's theorem is the generalised inverse function of a random variable, which can be found in [9].

1.3.4 Several results from Probability theory

We end this section with several results from Probability theory. A random variable is a measurable function defined on a probability space.

First we introduce the $\chi-$squared distribution.

Definition 1.3.14. Let $\{X_i\}_{i\in[n]}$ be independent random variables all following the standard Gaussian distribution $N(0,1)$. Then we will say that $X := \sum_{i=1}^{n} X_i^2$ follows the $\chi-$squared distribution with n degrees of freedom. One can check that $X \sim \Gamma(\frac{n}{2}, \frac{1}{2})$, where Γ denotes the gamma distribution.

Lemma 1.3.15 (Borel–Canteli Lemma)**.** *Let* $\{A_n\}$ *be a sequence of events in a probability space* $\{\Omega, \mathcal{A}, \mathbb{P}\}$. *Then*

$$\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty \quad \Longrightarrow \quad \mathbb{P}(\limsup A_n) = 0.$$

Definition 1.3.16. Let $(\Omega_1, \mathcal{A}, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{B}, \mathbb{P}_2)$ be two probability spaces. The probability space $(\Omega, \mathcal{A}\times\mathcal{B}, \mathbb{P})$ will be called their *product probability space* if

1. $\Omega = \Omega_1 \times \Omega_2$
2. $\mathcal{A}\times\mathcal{B} = \sigma(A\times B : A \in \mathcal{A}\ , B \in \mathcal{B}\}$, where for a class of sets D we denote by $\sigma(D)$ the smallest σ-algebra containing D.
3. $\forall A \in \mathcal{A}$ and $\forall B \in \mathcal{B}$ it is true that

$$\mathbb{P}(A \times B) = \mathbb{P}_1(A)\,\mathbb{P}_2(B).$$

It is a standard fact that, given two probability spaces $(\Omega_1, \mathcal{A}, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{B}, \mathbb{P}_2)$, such a product probability space always exists.

Theorem 1.3.17 (Fubini–Tonelli theorem)**.** *Let $(\Omega_1, \mathcal{A}, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{B}, \mathbb{P}_2)$ be two probability spaces. Let $(\Omega, \mathcal{A} \times \mathcal{B}, \mathbb{P})$ be their product probability space. Let X be a real random variable defined on that space. Then if X is either integrable with respect to $\mathbb{P}$ or non-negative, we have that*

$$\int X d\mathbb{P} = \int_{\Omega_1} \int_{\Omega_2} X d\mathbb{P}_2 d\mathbb{P}_1 = \int_{\Omega_2} \int_{\Omega_1} X d\mathbb{P}_1 d\mathbb{P}_2.$$

Now we present a useful corollary of the Fubini–Tonelli theorem.

Lemma 1.3.18. *If X is a real random variable defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with $\mathbb{E}|X| < +\infty$, then*

$$\mathbb{E}X = \int_0^\infty \mathbb{P}(X > t) dt - \int_{-\infty}^0 \mathbb{P}(X \leq t) dt.$$

Definition 1.3.19. Let X, Y be two real random variables, not necessarily defined on a common probability space, with distribution functions F_X, F_Y respectively, that is $F_X(x) := \mathbb{P}(X \leq x)$ and similarly for F_Y. We will use the notation $X \leq_{st} Y$ and say that Y *stochastically dominates* X if $F_X(t) \geq F_Y(t)$ $\forall\, t \in \mathbb{R}$.

From the previous lemma one has the following.

Corollary 1.3.20. *If $X \leq_{st} Y$ and $\mathbb{E}|X| < +\infty$ and $\mathbb{E}|Y| < +\infty$, then*

$$\mathbb{E}X \leq \mathbb{E}Y.$$

Next we present some "classical" results from probability and measure theory.

Theorem 1.3.21. *Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of real random variables.*

(i) (*Fatou*) *If $X_n \geq 0$ for all $n \in \mathbb{N}$, then it is true that*

$$\mathbb{E} \liminf_n X_n \leq \liminf_n \mathbb{E}X_n.$$

(ii) (*Beppo-Levi*) *If* $X_n \geq 0$ *for all* $n \in \mathbb{N}$, *then it is true that*

$$\mathbb{E}\sum_{n=1}^{\infty} X_n = \sum_{n=1}^{\infty} \mathbb{E}X_n.$$

(iii) (*Monotone Convergence Theorem*) *If* $X_{n+1} \geq X_n \geq 0$ *for all* $n \in \mathbb{N}$, *then it is true that*

$$\lim_n \mathbb{E}X_n = \mathbb{E}\lim_n X_n$$

(iv) (*Dominated Convergence Theorem*) *If* $X_n \to X$ *a.e. and there exists a random variable* Y *such that* $|X_n| \leq Y$ *for all* $n \in \mathbb{N}$ *and* $\mathbb{E}|Y| < +\infty$, *then*

$$\mathbb{E}X = \lim_n \mathbb{E}X_n.$$

We end this subsection with the definition of a median of a random variable.

Definition 1.3.22. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let X be a real random variable defined on that probability space. Then a real number $M \in \mathbb{R}$ will be called median of X if

$$\min\{\mathbb{P}(X \geq M), \mathbb{P}(X \leq M)\} \geq \tfrac{1}{2}.$$

Remark 1.3.23. One may prove that every random variable has a median. Moreover, there are many concentration inequalities for the deviation of a random variable from its median.

All the results of this subsection, including Skorohod's theorem, can be found in any textbook on probability theory, for example in [9].

1.4 Tools from Convex Analysis

In this part we will gather some important results from convex analysis.

1.4.1 Isoperimetric inequality on the sphere

We denote by S^{n-1} the unit sphere of $\mathbb{R}^n$, $n \geq 2$.

Definition 1.4.1. We will say that a set $K \subseteq \mathbb{R}^n$ is a *convex body* if it is a convex compact set with non-empty interior.

Definition 1.4.2. Consider the function $g\colon S^{n-1} \times S^{n-1} \to [0, \infty)$ which assigns to a pair of points $x, y \in S^{n-1}$ the angle xOy in the plane defined by x, y and the origin. Note that $g(x, y) = 2 \arcsin\left(\frac{1}{2}\|x - y\|_2\right)$.

The function g is a metric and is equivalent with the restriction to the sphere of the metric induced by the 2-norm on $\mathbb{R}^n$.

Definition 1.4.3. Let $n \in \mathbb{N}$. Then we define a probability measure on the sphere S^{n-1}, called the *spherical measure* (the unique Haar probability measure on the sphere), as follows: For every Borel subset A of S^{n-1},

$$s^{n-1}(A) = \frac{1}{\lambda_n(B_2^n)}\lambda_n([0, 1]A)$$

where λ_n is the Lebesgue measure in $\mathbb{R}^n$, $B_2^n = \{x \in \mathbb{R}^n\colon \|x\|_2 \leq 1\}$ and $[0, 1]A = \{at\colon t \in [0, 1],\ a \in A\}$. Equivalently s^{n-1} can be expressed as

$$s^{n-1}(A) = \gamma_n((0, +\infty)A),$$

where γ_n denotes the standard Gaussian measure on $\mathbb{R}^n$, i.e., the measure on $\mathbb{R}^n$ with density $(2\pi)^{-n/2}\exp\left(-\frac{1}{2}\|x\|_2^2\right)$, $x \in \mathbb{R}^n$, with respect to λ_n.

By the rotational invariance of the Lebesgue measure λ_n, or equivalently by the rotational invariance of the Gaussian measure γ_n, s^{n-1} is rotationally invariant. Hence it is the unique Haar probability measure on S^{n-1}. Indeed, S^{n-1} can be identified with the set of cosets of the group of orthogonal transformations on $\mathbb{R}^n$ and hence its Haar measure is a measure invariant under orthogonal transformations; furthermore, this Haar measure is unique up to multiplication by a constant.

Remark 1.4.4. By the representation via the standard Gaussian measure on $\mathbb{R}^n$ above, s^{n-1} can also be expressed as follows. Let $X_1, X_2, \ldots, X_n$ be

i.i.d. random variables defined on a common probability space $(\Omega, \mathbb{A}, \mathbb{P})$, such that $X_1 \sim N(0,1)$, i.e., X_1 has a standard Gaussian distribution γ_1 in $\mathbb{R}$. Then for every Borel subset A of S^{n-1} it is true that

$$s^{n-1}(A) = \mathbb{P}\left(\left(\sum_{i=1}^{n} X_i^2\right)^{-1/2} (X_1, X_2, \ldots, X_n) \in A\right).$$

In fact, integrating in polar coordinates yields

$$\begin{aligned}
&\mathbb{P}\left(\left(\sum_{i=1}^{n} X_i^2\right)^{1/2} \in (a,b),\ \left(\sum_{i=1}^{n} X_i^2\right)^{-1/2} (X_1, X_2, \ldots, X_n) \in A\right) \\
&\qquad = (2\pi)^{-n/2} \int_0^{\infty} \int_{S^{n-1}} \mathbf{1}_{(a,b)\times A}(r,\partial) r^{n-1} e^{-r^2/2} d\partial dr \\
&\qquad = (2\pi)^{-n/2} \int_a^b r^{n-1} e^{-r^2/2} dr \cdot \int_A d\partial \\
&\qquad = \mathbb{P}\left(\left(\sum_{i=1}^{n} X_i^2\right)^{1/2} \in (a,b)\right) \cdot \mathbb{P}\left(\left(\sum_{i=1}^{n} X_i^2\right)^{-1/2} (X_1, X_2, \ldots, X_n) \in A\right)
\end{aligned}$$

for all Borel $A \subseteq S^{n-1}$ and $a, b \in [0, +\infty]$ with $a < b$, and this shows that, furthermore, the random variable $\left(\sum_{i=1}^{n} X_i^2\right)^{1/2}$ and the random vector $\left(\sum_{i=1}^{n} X_i^2\right)^{-1/2} (X_1, X_2, \ldots, X_n)$ are also independent.

Theorem 1.4.5 (Isoperimetric inequality on the sphere)**.** *Let $n \in \mathbb{N}$ with $n \geq 2$. Consider the probability metric space $(S^{n-1}, B(S^{n-1}), s^{n-1})$. Let C be an open ball of the sphere and $A \subseteq S^{n-1}$ measurable such that*

$$s^{n-1}(C) = s^{n-1}(A).$$

Then for every $\epsilon > 0$, it is true that

$$s^{n-1}(C_\epsilon) \leq s^{n-1}(A_\epsilon).$$

Corollary 1.4.6. *If $n > 2$ and if $s^{n-1}(A) \geq \frac{1}{2}$ for some $A \in B(S^{n-1})$, then*

$$s^{n-1}(A_\epsilon) \geq s^{n-1}\left(C\left(x, \tfrac{1}{2}\pi + \epsilon\right)\right) \geq 1 - e^{-n\epsilon^2/2}$$

for any $\epsilon > 0$.

One may find the isoperimetric inequality on the sphere in [7].

1.4.2 Krein–Milman theorem

Next we present an important result from functional analysis which implies that in Banach spaces (or more generally locally convex topological vector spaces) a convex and compact set is the convex hull of some of its elements (most of the times significantly fewer). In the thesis we use the Krein–Milman theorem to simplify several proofs.

Definition 1.4.7. Let X be a Banach space and let $K \subseteq X$ be a convex subset. Then a subset $F \subseteq K$ will be called *extreme* subset of K if

$$x, y \in K,\ \lambda \in (0,1),\ \lambda x + (1-\lambda)y \in F \implies x, y \in F.$$

If $F = \{f\}$, we will call f an *extreme point* of K. We will use the notation $\mathrm{ext}(K)$ for the set of the extreme points of K.

Theorem 1.4.8 (Krein–Milman theorem)**.** *Let K be a compact and convex subset of a Banach space X. Then* $\mathrm{ext}(K) \neq \emptyset$ *and*

$$\overline{\mathrm{conv}(\mathrm{ext}(K))} = K.$$

One may find the Krein–Milman theorem in [10].

1.4.3 Some facts about convex sets

At this point we are going to introduce some geometric parameters of convex sets (the most important one is volume, i.e., Lebesgue measure). First we give a useful inequality.

Definition 1.4.9. Let K be a convex body of $\mathbb{R}^n$. Then we define its *centroid*, $g(K)$, to be

$$g(K) := \frac{\int_K x dx}{\lambda_n(K)}$$

where λ_n denotes Lebesgue measure on $\mathbb{R}^n$.

Lemma 1.4.10. *If K is a convex body in $\mathbb{R}^n$ (or $\mathbb{C}^n$) with its centroid at the origin, then*

$$\lambda_n(K \cap (-K)) \geq 2^{-n}\lambda_n(K).$$

Proof. We will prove the lemma in several steps. First we mention the Brunn–Minkowski inequality, which is an important tool from convex geometry and will be used in the proof of this lemma.

Theorem 1.4.11 (Brunn–Minkowski inequality)**.** *Let K, L be two Borel subsets of $\mathbb{R}^n$. Then*

$$\lambda_n(K+L)^{1/n} \geq \lambda_n(K)^{1/n} + \lambda_n(L)^{1/n}.$$

An equivalent statement is as follows. For any two Borel sets $K, L \subseteq \mathbb{R}^n$ and any $\lambda \in [0,1]$,

$$\lambda_n(\lambda K + (1-\lambda)L) \geq \lambda_n(K)^{\lambda}\lambda_n(L)^{1-\lambda}.$$

Note. The first inequality implies the second in the theorem above; this is an immediate consequence of the inequality $\lambda x + (1-\lambda)y \geq x^{\lambda}y^{1-\lambda}$, valid for $x, y \geq 0$ and $\lambda \in [0,1]$:

$$\lambda_n(\lambda K + (1-\lambda)L) \geq [\lambda\lambda_n(K)^{1/n} + (1-\lambda)\lambda_n(L)^{1/n}]^n \geq \lambda_n(K)^{\lambda}\,\lambda_n(L)^{1-\lambda}.$$

If $H \subseteq \mathbb{R}^n$ is a linear or affine subspace of $\mathbb{R}^n$, we will use the notation λ_H for the $\dim(H)$-dimensional Lebesgue measure on H.

Lemma 1.4.12. *Let $K \subseteq \mathbb{R}^n$ be a convex body with its centroid at the origin. If E is a subspace of $\mathbb{R}^n$ and F is the orthogonal complement of E, then*

$$\lambda_n(K) \leq \lambda_E(K \cap E)\,\lambda_F(P_F K),$$

where P_F denotes projection onto the subspace F.

Proof. Define the function $D\colon P_F K \to \mathbb{R}^+$ as follows

$$D(x) = \lambda_{E+x}(K \cap E + x)^{1/k}.$$

Here $k = \dim(E)$. By convexity and by Theorem 1.4.11 we see that the function D is concave. Applying the Fubini–Tonelli theorem and Hölder's inequality we get that

$$\lambda_n(K) = \int_{P_F K} D(x)^k dx \leq \lambda_F(P_F K)^{1/(k+1)} \left(\int_{P_F K} D(x)^{k+1} dx \right)^{k/(k+1)}.$$

Since D is concave, there exists $y \in F$ such that for any $x \in P_F K$

$$D(x) \leq D(0) + \langle x, y \rangle.$$

It follows that

$$\begin{aligned}\int_{P_F K} D(x)^{k+1} dx &\leqslant \int_{P_F K} D(x)^k (D(0) + \langle x, y \rangle) dx \\ &= D(0) \int_{P_F K} D(x)^k dx = D(0) \lambda_n(K),\end{aligned}$$

because, since the centroid of K is at the origin we have $\int_{P_F(K)} D(x)^k \langle x, y \rangle dx = 0$. It follows that

$$\lambda_n(K) \leq \lambda_F(P_F K)^{1/(k+1)} D(0)^{k/(k+1)} \lambda_n(K)^{k/(k+1)}.$$

Since $D(0)^k = \lambda_E(K \cap E)$, the inequality follows. □

We can apply the previous lemma for the convex body $K \times -K \subseteq \mathbb{R}^n \times \mathbb{R}^n$ and the subspaces $E = \{(x, x)\, x \in \mathbb{R}^n\}$ and $F = \{(x, -x),\ x \in \mathbb{R}^n\}$. Note that

1. $\lambda_{2n}(K \times (-K)) = \lambda_n(K)\lambda_n(-K) = \lambda_n(K)^2$,
2. $\lambda_n(K \times (-K) \cap E) = 2^{n/2} \lambda(K \cap (-K))$,
3. $\lambda_n(P_F K \times (-K)) = 2^{-n/2} \lambda_n(K - (-K))$.

Using also the fact that $\lambda_n(K - (-K)) = \lambda_n(2K) = 2^n \lambda(K)$ we conclude the proof of Lemma 1.4.10. □

Definition 1.4.13. Let V be a (real or complex) Hilbert space. Let also K be a convex subset of V with the origin in its interior. We call the function $\|\cdot\|_K$ defined below the *gauge* of K (or *Minkowski functional* of K):

$$\|x\|_K := \inf\{t > 0 \colon x \in tK\}.$$

It is easy to prove that if K is an origin symmetric convex body then $\|\cdot\|_K$ is a norm. In the case where K is not symmetric, then the gauge is not a norm because there exists $x \in V$ such that

$$\|x\|_K \neq \|(-1)x\|_K.$$

Definition 1.4.14. Let K be a convex subset of $\mathbb{R}^n$ with the origin in its interior. Then we define the set

$$K^\circ := \{y \in \mathbb{C}^n : \langle y, x\rangle \leq 1 \ \forall\, x \in K\}$$

and call it the *polar* set of K.

Definition 1.4.15. Let $K \subseteq \mathbb{R}^n$ be a Borel set. The *volume radius* of K is defined as

$$\mathrm{vrad}(K) := \left(\frac{\lambda(K)}{\lambda(B_2^n)}\right)^{1/n},$$

where λ is the Lebesgue measure on $\mathbb{R}^n$ and B_2^n is the unit ball of $\mathbb{R}^n$ with respect to the 2-norm. In words, the volume radius of K is the radius of the Euclidean ball which has the same volume as K.

Equivalently, if K is a convex body, then

$$\mathrm{vrad}(K) = \int_{S^{n-1}} \|\theta\|_K^{-n} \, ds^{n-1}(\theta),$$

where S^{n-1} is the unit sphere of $\mathbb{R}^n$ and s^{n-1} is the spherical measure.

Definition 1.4.16. Let K be a convex body of $\mathbb{R}^n$. Then we define the *mean width* of K as follows:

$$w(K) := \int_{S^{n-1}} \sup_{x \in K} \langle u, x\rangle \, ds^{n-1}(u).$$

Alternatively,

$$w(K) = \int_{S^{n-1}} \|u\|_{K^\circ} \, ds^{n-1}(u).$$

Definition 1.4.17. Let Z_1, Z_2 be two i.i.d. random variables both following $N(0,1)$. Then we will use the notation $N_{\mathbb{C}}(0,1)$ for the distribution of the complex random variable $Z = \frac{1}{\sqrt{2}}(Z_1 + iZ_2)$

Definition 1.4.18. Let V be a real (resp. complex) finite-dimensional Hilbert space equipped with a Euclidean (resp. Hilbertian) norm. By definition, the standard Gaussian vector in V is a V-valued random variable whose coordinates with respect to any orthonormal basis of V are independent real (resp. complex) standard normal (Gaussian) random variables.

Remark 1.4.19. A standard Gaussian vector in $\mathbb{R}^n$ (or $\mathbb{C}^n$) is an n-dimensional random vector whose entries are independent $N(0,1)$ (or $N_{\mathbb{C}}(0,1)$)-variables.

Many results, such as the (multivariate) central limit theorem, can be generalised in all vector spaces (real or complex) using a standard Gaussian vector of the space.

Remark 1.4.20. Let G be a standard Gaussian vector in a Hilbert space V. Then $D := G/\|G\|$ is uniformly distributed, i.e., distributed according to the normalized Haar measure, on the unit sphere of V. Moreover, D is stochastically independent from $\|G\|$.

To see this, consider the case of a real vector space first. Fix an orthonormal basis $\{v_1, \dots, v_n\}$ of V, where $n = \dim(V)$, and consider the isomorphism $J\colon \mathbb{R}^n \to V$ defined by $J(e_i) = v_i$, $i \in \{1, \dots, n\}$, and then extended by linearity on $\mathbb{R}^n$, where $\{e_1, \dots, e_n\}$ is the standard (say) orthonormal basis of $\mathbb{R}^n$. To show that D is distributed according to the normalized Haar measure on the unit sphere of V, one has to show that the distribution of D is invariant under the unitary group of V, i.e. the group of linear transformations on V satisfying $T^*T = I_V$, where I_V the identity operator on V. The random variables $\langle v_1, G\rangle, \dots, \langle v_n, G\rangle$ are i.i.d. $N(0,1)$ random variables, hence one may invoke Remark 1.4.4. Hence, for any Borel subset A of the unit sphere S_V of V and any $a, b \in [0, +\infty]$ with $a < b$, one has that

$$
\begin{aligned}
&\mathbb{P}(\|G\| \in (a,b),\ \|G\|^{-1}G \in A) = \mathbb{P}(\|G\| \in (a,b),\ \|G\|^{-1}J^{-1}(G) \in J^{-1}(A)) \\
&= \mathbb{P}\left(\left(\sum_{i=1}^n |\langle v_i, G\rangle|^2\right)^{1/2} \in (a,b),\ \left(\sum_{i=1}^n |\langle v_i, G\rangle|^2\right)^{-1/2} \sum_{i=1}^n \langle v_i, G\rangle e_i \in J^{-1}(A)\right) \\
&= \mathbb{P}\left(\left(\sum_{i=1}^n |\langle v_i, G\rangle|^2\right)^{1/2} \in (a,b)\right) \mathbb{P}\left(\left(\sum_{i=1}^n |\langle v_i, G\rangle|^2\right)^{-1/2} \sum_{i=1}^n \langle v_i, G\rangle e_i \in J^{-1}(A)\right) \\
&= \mathbb{P}\left(\|G\| \in (a,b)\right) \mathbb{P}\left(\left(\sum_{i=1}^n |\langle v_i, G\rangle|^2\right)^{-1/2} \sum_{i=1}^n \langle v_i, G\rangle v_i \in A\right) \\
&= \mathbb{P}\left(\|G\| \in (a,b)\right) \mathbb{P}(D \in A),
\end{aligned}
$$

which shows the independence of $\|G\|$ and D. Furthermore, for any linear

transformation $T \in B(V)$ satisfying $T^*T = I_V$,

$$\begin{aligned}
\mathbb{P}(T(D) \in A) &= \mathbb{P}(\|G\|^{-1}T(G) \in A) \\
&= \mathbb{P}(\|G\|^{-1}J^{-1}TJ(J^{-1}(G)) \in J^{-1}(A)) \\
&= \mathbb{P}\left(\left(\sum_{i=1}^{n}|\langle v_i, G\rangle|^2\right)^{-1/2} J^{-1}TJ\left(\sum_{i=1}^{n}\langle v_i, G\rangle e_i\right) \in J^{-1}(A)\right) \\
&= \mathbb{P}\left(\left(\sum_{i=1}^{n}|\langle v_i, G\rangle|^2\right)^{-1/2} \sum_{i=1}^{n}\langle v_i, G\rangle e_i \in J^{-1}(A)\right) \\
&= \mathbb{P}\left(\left(\sum_{i=1}^{n}|\langle v_i, G\rangle|^2\right)^{-1/2} \sum_{i=1}^{n}\langle v_i, G\rangle v_i \in A\right) \\
&= \mathbb{P}(D \in A),
\end{aligned}$$

the fourth equality using the fact that JTJ^{-1} is an orthogonal transformation in $\mathbb{R}^n$ and that $\left(\sum_{i=1}^{n}|\langle v_i, G\rangle|^2\right)^{-1/2}\sum_{i=1}^{n}\langle v_i, G\rangle e_i$ is Haar (uniformly) distributed on the sphere S^{n-1}.

Definition 1.4.21. Let G be a standard Gaussian vector in $\mathbb{R}^n$. Then, for any non empty bounded set $K \subseteq \mathbb{R}^n$ we define the *Gaussian* mean width of K as

$$w_G(K) = \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{n/2}} \sup_{x\in K}\langle u, x\rangle e^{-\|u\|_2^2/2} du.$$

We next compare the mean width, the Gaussian mean width and the volume radius of convex bodies.

Lemma 1.4.22 (Urysohn inequality)**.** *Let K be a convex body in $\mathbb{R}^n$. Then*

$$\operatorname{vrad}(K) \leq w(K).$$

Moreover the above result is true for all bounded Borel sets.

Proof. We will give a sketch of the proof.

For a probability space (Ω, A, μ), where μ is a discrete probability measure or the limit of discrete probability measures, the following generalisation of the Brunn-Minkowski inequality holds:

$$\int_\Omega \lambda_n(K_t)^{1/n} d\mu(t) \leq \lambda_n^{1/n}\left(\int_\Omega K_t \, d\mu(t)\right). \tag{1.4.1}$$

When the measure μ is purely atomic with N atoms, the result can be proved by induction on N, the case $N = 2$ being exactly the Brunn{-Minkowski inequality. Moreover, the continuous case can then be derived by approximation. The inequality above makes sense when the function $t \to w(K_t, \partial)$ is measurable for every $\partial \in \mathbb{R}^n$.

If we equip the space $O(n)$ (where $O(n)$ is the class of $n \times n$ orthogonal matrices) with the Haar measure, and set $K_t = t(K), t \in O(n)$ then the convex body $L := \int_{O(n)} t(K)d\mu(t)$ is necessarily a Euclidean ball centered at the origin. Computing the width of L in an arbitrary direction we see that L is a Euclidean ball of radius $w(K)$. So, by applying (1.4.1) one may show that Urysohn's inequality holds.

For a more detailed proof see [7, Exercise 4.49]. □

Lemma 1.4.23. *Let $\gamma_n = \mathbb{E}\|G\|_2$ where G is a standard Gaussian vector in $\mathbb{R}^n$ and $\|\cdot\|_2$ is the 2-norm. Then one may compute that*

$$\sqrt{n-1} \leq \gamma_n \leq \sqrt{n}$$

Moreover, for any convex body K in $\mathbb{R}^n$ it is true that

$$w_G(K) = \gamma_n w(K).$$

Proof. For the first part we know that if $X \sim \chi^2(n)$ then

$$\mathbb{E}\|G\|_2 = \mathbb{E}\sqrt{X}.$$

So, if f_X is the probability density function of X and f_Z is the probability density function of a $\chi-$squared random variable with $n+1$ degrees of freedom, we get

$$\begin{aligned}\mathbb{E}\sqrt{X} &= \int_{\mathbb{R}} \sqrt{x} f_X(x)dx = \int_0^\infty \left(\frac{1}{2}\right)^{n/2} x^{\frac{n+1}{2}-1} \exp\left(-\frac{x}{2}\right)\frac{1}{\Gamma(\frac{n}{2})}dx \\ &= \frac{\sqrt{2}\Gamma((n+1)/2)}{\Gamma(n/2)} \int_0^\infty f_Z(x)dx = \sqrt{2}\left(\frac{\Gamma((n+1)/2)}{\Gamma(n/2)}\right).\end{aligned}$$

In order to prove the first assertion of the lemma we need several well known properties of the Γ-function.

1. $\forall x \in \mathbb{R}^+$ it is true that $\Gamma(x+1) = x\Gamma(x)$,

2. $\Gamma(0) = 1$ and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$,

3. The function $\log \Gamma$ is concave.

The first two properties imply that $\forall n \in \mathbb{N}$ it is true that $\Gamma(n) = (n-1)!$.

In order to get the lower bound for γ_n we use induction. The basic observation, which follows from the properties of the Γ-function, is that $\gamma_n \gamma_{n+1} = n$. For $n = 1$, the lower bound obviously holds, because $\sqrt{2/\pi} > 0$. Assuming that the lower bound is true for some $k \leq n$, we use the induction hypothesis and the recursion formula to write

$$\gamma_{n+1} = \frac{n}{\gamma_n} \geq \frac{n}{\sqrt{n}} = \sqrt{n}.$$

For the upper bound we note that, by the Cauchy{Schwarz inequality,

$$\mathbb{E}\sqrt{x} \leq (\mathbb{E}x)^{1/2} = \sqrt{n}$$

as needed.

The second part of the lemma is a simple consequence of Remark 1.4.20.

□

1.4.4 ℓ-norm, ℓ-position and the MM^*-estimate

Next we present a norm on the space of n-dimensional real (or complex) operators, the so-called ℓ-norm, and some useful properties of it.

Definition 1.4.24. Let $K \subseteq \mathbb{R}^n$ be a convex body containing 0 in its interior. Then, for any $T \in M_n$ we define the quantity

$$\ell_K(T) = \mathbb{E}\|T(G)\|_K$$

where G is a standard Gaussian vector in $\mathbb{R}^n$ (or $\mathbb{C}^n$).

The function $\ell_K : M_n \to \mathbb{R}^+$ is a norm and is called ℓ-norm.

Proposition 1.4.25. *If K is a convex body with the origin in its interior then*

(i) ℓ_K *obeys the ideal property: for any* $S, T \in M_n$,

$$\ell_K(TS) \leq \ell_K(T)\|S\|_{\text{op}}.$$

(ii) $\ell_K(\mathbb{I}) = w_G(K^\circ) = w(K^\circ)\mathbb{E}\|G\|_2$.

(iii) *If* $T \in M_n$ *is* 1-1 *then* $\ell_K(T) = \ell_{T^{-1}K}(\mathbb{I})$.

(iv) *If* P_E *denotes the orthogonal projection onto a subspace* $E \subseteq \mathbb{R}^n$ *then*

$$\ell_K(P_E) = w_G((K \cap E)^\circ) = w_G(P_E K)$$

where we denote by $(K \cap E)^\circ$ *the polar of* $K \cap E$ *inside* E.

Proof. The only part of the proposition which is not straightforward is (i).

Let $S, T \in M_n$. By homogeneity we may assume that $\|S\|_{\text{op}} = 1$ and since ℓ is a norm we may also assume that S is an extreme point of the unit ball of M_n with respect to the operator norm. One may show that T is an n-dimensional orthogonal matrix.

Since G is assumed to be a standard Gaussian vector in $\mathbb{R}^n$ (or $\mathbb{C}^n$), we know that under any orthogonal transformation G will remain a standard Gaussian vector. So, $\ell_K(T) = \ell_K(TS)$. □

We now introduce the ℓ-position.

Proposition 1.4.26. *For any convex body* $K \subseteq \mathbb{R}^n$ *containing* 0 *in its interior, there exists a unique positive semi-definite matrix* T_0 *that is a solution to the maximization problem*

$$\max\{\det(T) : T \text{ is a positive semi-definite matrix, } \ell_K(T) \leq 1\}.$$

If this unique solution is a multiple of the identity matrix (equivalently operator) then we say that K *is in the* $\ell-$*position.*

Proof. We will prove that the solution of the maximization problem in the statement of the proposition is unique.

Assume that there exist T_1, T_2 which both solve the maximization problem. Consider the matrix $T = (T_1 + T_2)/2$. Note that $\ell_K(T) \leq 1$.

We will show that the function $\log\det$ is strictly concave on the set of positive semi-definite matrices, which will lead to a contradiction.

We know that a function is concave iff its restriction to any line that intersects its domain is also concave.

So, we will prove that the function $g(t) = \log(\det(A + tB))$ is self-adjoint where $B \neq 0$ is self-adjoint and A is positive semi-definite. Since A is semi-definite the matrices $A^{1/2}$ and $A^{-1/2}$ are well defined and so we get:

$$\begin{aligned} g(t) &= \log\det(A + tB) = \log\det(A^{1/2}(I + tA^{1/2}BA^{-1/2})A^{1/2} \\ &= \log\det(A) + \sum_{i=1}^{n}(1 + t\lambda_i), \end{aligned}$$

where $\{\lambda_i\}_{i\in[n]}$ are the eigenvalues of $A^{-1/2}BA^{-1/2}$. But the function $\sum \log(1+t\lambda_i)$ is concave. So $g(t)$ is concave and as a result the function $\log\det$ is concave, which ends the proof. □

Now we present a crucial result from geometric analysis called "MM^*-estimate".

Theorem 1.4.27 (MM^*-estimate)**.** *For any convex body K which is in the ℓ − position we have that*

$$1 \leq w(K)w(K^0) \leq C\log n.$$

Proof. See [7]. □

Another crucial result about the ℓ−position is the following.

Lemma 1.4.28. *Let K be a symmetric convex body in $\mathbb{R}^n$ (or $\mathbb{C}^n$) and let Γ be the isometry group of K (i.e. the set of all orthogonal transformations U such that $UK = K$). Then there exists a linear map T such that TK is in the ℓ−position and*

$$T = \sum_i \lambda_i P_{E_i},$$

where $\lambda_i > 0$ and E_i are subspaces invariant under the action of Γ.

The proof of all the results from Convex Analysis that we presented can be found in [7].

In the next parts of the thesis we present and prove some additional tools from convex analysis, such as the Gaussian isoperimetric inequality 2.1.27 and Erhard's inequality 2.1.26.

1.5 Quantum information theory

In this section we give several definitions from quantum information theory.

Firstly we need to present the tensor product of Hilbert spaces. Throughout this thesis, all Hilbert spaces will be meant to be complex Hilbert spaces unless we specify differently.

Definition 1.5.1. Let A, B and C be finite dimensional Hilbert spaces. Then a mapping $f : A \times B \to C$ is called bilinear if

$$\begin{aligned} f(x_1 + x_2, y) &= f(x_1, y) + f(x_2, y) \\ f(x, y_1 + y_2) &= f(x, y_1) + f(x + y_2) \\ f(\lambda x, y) &= f(x, \lambda y) = \lambda f(x, y) \end{aligned}$$

for all vectors $x, y \in A \times B$ and $\lambda \in \mathbb{C}$.

Definition 1.5.2. Let A, B be finite dimensional Hilbert spaces. Then we say that a Hilbert space P is a tensor product of A, B with a bilinear mapping $f : A \times B \to P$ if f has the following properties:

- The closed linear hull of $f(A \times B)$ is P.
- $\langle f(x_1, y_1), f(x_2, y_2)\rangle_P = \langle x_1, x_2\rangle_A \langle y_1, y_2\rangle_B$.

Lemma 1.5.3. *Let A, B be two complex finite dimensional Hilbert spaces. Then their tensor product exists and is unique under isomorphism.*

Proof. For a proof see [11]. □

So, we will use the notation $A \otimes B$ for the space and $\otimes$ for the bilinear function f.

FACT: Let H_1, H_2 be two finite dimensional complex Hilbert spaces of dimension n and m respectively. Let $\{e_i\}_{i\in[n]}$ and $\{f_j\}_{j\in\mathbb{N}}$ be bases of H_1 and H_2 respectively. Then the set $\{e_i \otimes f_j\}_{i,j\in[n]\times[m]}$ is a basis for $H_1 \otimes H_2$.

Lemma 1.5.4. *Let H_1 and H_2 be two finite dimensional complex Hilbert Spaces. Then,*

$$B(H_1 \otimes H_2) = B(H_1) \otimes B(H_2).$$

Proof. Let $S \in B(H_1)$ and $T \in B(H_2)$. Then consider the function

$$S \otimes T : H_1 \otimes H_2 \to H_1 \otimes H_2$$

with $x \otimes y \mapsto S(x) \otimes T(y)$. This proves that $B(H_1) \otimes B(H_2) \subseteq B(H_1 \otimes H_2)$.

On the other hand, for any m-dimensional complex Hilbert space A it is true that $B(A) \cong M_m(\mathbb{C})$ since we can associate every linear map to its matrix. So,

$$\dim(B(A)) = m^2.$$

This implies that $\dim(B(H_1 \otimes H_2)) = n^2m^2$ and $\dim(B(H_1) \otimes B(H_2)) = n^2m^2$, which completes the proof. □

Lemma 1.5.5. *Let H_1, H_2 be two Hilbert spaces of dimension m and n respectively. Then it is true that*

$$H_1 \otimes H_2 \cong B(H_1, H_2).$$

Proof. Fix bases $\{e_j\}_{j\in[n]}$ and $\{f_i\}_{i\in[m]}$ of H_1 and H_2 respectively. Consider the function $\mathrm{vec} : H_1 \otimes H_2 \to B(H_1, H_2)$ with

$$\mathrm{vec}(e_i \otimes f_j) = |e_i\rangle\langle f_j|$$

and extend it linearly to all the elements of $H_1 \otimes H_2$ by $\mathbb{C}$–linearity. The function we obtain is a canonical identification between the two spaces. □

Corollary 1.5.6. *From the previous lemma, if $H_1 = \mathbb{C}^n$ and $H_2 = \mathbb{C}^m$ where $m, n \in \mathbb{N}$ we get*

$$\mathbb{C}^n \otimes \mathbb{C}^m \cong M_{n,m}(\mathbb{C}).$$

In general, in quantum information theory, tensors are more suitable to describe and model the problems; the previous proposition allows us to identify tensor products as spaces of matrices.

Next we give several definitions necessary in quantum information theory.

Definition 1.5.7. Let H be a finite dimensional Hilbert space. Then the set

$$D(H) = \{\rho \in B_{sa}(H) : \rho \geq 0,\ \mathrm{tr}(\rho) = 1\}$$

is called the set of states of H.

Definition 1.5.8. Let H be a finite dimensional Hilbert space. Then a state ρ of H is called pure if there exists a unitary vector y such that

$$\rho = |y\rangle\langle y|.$$

Definition 1.5.9. If H is the tensor product of a finite family of finite dimensional Hilbert spaces, i.e.

$$H = H_1 \otimes H_2 \otimes \cdots \otimes H_n,$$

then a pure state ρ of H is called pure separable if the unit vector $x \in H$ for which $\rho = |x\rangle\langle x|$ is a tensor product of unit vectors i.e

$$x = x_1 \otimes x_2 \otimes \cdots \otimes x_n,$$

where $x_i \in H_i$ are unit vectors.

In general, a state of H is called separable if it can be written as a convex combination of pure separable states.

We will use the notation $\mathrm{Sep}(H)$ for the set of all separable states of H.

Remark 1.5.10. The set $\mathrm{Sep}(H)$ is a convex body containing 0 in its interior.

Definition 1.5.11. The states of a tensorized Hilbert space which are not separable are called entangled.

1.6 Tools from Combinatorics and Graph theory

In this section we give several definitions and results from combinatorics that are used in the rest of the thesis.

Definition 1.6.1. We will call (simple) graph a couple (V, E) if V is a finite set (the vertex set) and E is a subset of $V \times V$: for all $e \in E$, $|e| = 2$ (the edge set).

Moreover, if $\mathcal{A}$ is the set of all graphs, we will work on the set $\mathcal{A}/\sim$. Here the relation $\sim$ is an equivalence relation defined as follows: for two graphs G, H we have $G \sim H$ iff there exists $f : V(G) \to V(H)$ such f is an isomorphism and

$$\forall x, y \in V(G), \{x, y\} \in E(G) \iff \{f(x), f(y)\} \in E(H).$$

Remark 1.6.2. In the previous definition, if we assume that $|e| \leq 2$ for all $e \in E$, we have allowed loops to exist in the graph.

Moreover if we assume that the set E is oriented, meaning that $(x, y) \in E$ does not imply that $(y, x) \in E$, then we get an oriented graph.

In the rest of the thesis we might come across to oriented graphs and/or graphs with loops but it is sufficient for us to present and use properties for simple graphs.

Definition 1.6.3. Let $G = (V, E)$ be a simple graph. We will say that G is connected iff

$$\forall x, y \in V \ \exists v_1, v_2, \ldots, v_k \in V : \{x, v_1\}, \{v_1, v_2\}, \ldots, \{v_k, y\} \in E,$$

meaning that we can "travel", via edges, from any vertex to any vertex.

Any set of consecutive edges via which we can travel from x to y is called a path that connects x and y.

Proposition 1.6.4. *If G is a connected graph, then*

$$|E(G)| \geq |V(G)| - 1.$$

Definition 1.6.5. A graph G will be called a tree iff it is connected and there exists a unique path from any vertex to any other.

One may also define a tree as a connected graph with no circles (meaning that we cannot travel from any vertex to any other in two different ways).

Proposition 1.6.6. *Let G be a connected graph. Then G is a tree iff $E(G) = V(G) - 1$.*

We end this section with a well-known theorem from combinatorics.

Definition 1.6.7. Let G be a graph. Then we will call G bipartite iff there exist two disjoint sets $X, Y \subseteq V(G)$ with $X \cup Y = V(G)$, such that

$$\{x, y\} \in E \Longrightarrow x \in X, y \in Y.$$

Definition 1.6.8. Let G be a simple graph. Let $M \subseteq E$. We will call M a matching of G iff

$$\forall e, e' \in M \Longrightarrow \ e \cap e' = \emptyset.$$

If M is a matching, then a vertex $x \in V$ such that $\exists e \in M : x \in e$ will be said to be covered by M.

Moreover we will say that G has a perfect matching iff $\exists M \subseteq E$ such that

$$(\forall x \in V \ \exists e_x \in M : x \in e_x).$$

Theorem 1.6.9 (Hall)**.** *Let G be a bipartite graph with parts X and Y. Then G has a matching that covers all the vertices in X iff for any subset $A \subseteq X$ it is true that*

$$|\{y \in Y : \exists x \in A : \{x, y\} \in E\}| \geq |A|.$$

The proof of the results presented in this section can be found in [12].

Part II

Random Matrix Theory

Convergence of the empirical spectral distribution

In this chapter we prove three classical results from random matrix theory. First we introduce the concept of the empirical spectral distribution (E.S.D.) of an $n \times n$ matrix A which will be denoted by μ_A. More precisely, if $\{\lambda_i(A)\}_{i\in[n]}$ are the eigenvalues of A then

$$\mu_A = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(A)}$$

where δ denotes the Dirac measure. In other words, the empirical spectral distribution of a matrix A is exactly the discrete uniform measure on the set of the eigenvalues of A.

So one may note that if A is a random matrix then μ_A is also a random measure in $\mathbb{P}(\mathbb{R})$, the set of all Borel probability measures on $\mathbb{R}$.

Then we present Wigner's semicircular law [1], the Marchenko-Pastur law [3] and Bai-Yin's theorem on convergence to the semicircular law [4].

2.1 Wigner's semicircular law

2.1.1 Convergence of the E.S.D.

In this subsection we present and prove a fundamental result from random matrix theory first proved in [1] by Wigner.

Definition 2.1.1. A matrix $[A]_{i,j} \in M_{n,m}[\mathbb{F}]$ will be called a *random matrix* if $\exists i_0, j_0 \in [n] \times [m]$ and a probability space $(\Omega, \mathbb{A}, \mathbb{P})$ such that A_{i_0,j_0} is a

random variable from that probability space to $\mathbb{F}$. Here $\mathbb{F}$ is the field we are working with. Normally it will be either $\mathbb{R}$ or $\mathbb{C}$. In the next chapters we will work on either $\mathbb{R}$ or $\mathbb{C}$ and it will be insignificant to further clarify $\mathbb{F}$.

Also, we assume that if there are more than one entries that are random, all of them are defined on the same probability space.

Definition 2.1.2. Suppose $A \in M_n$. Then the following measure is called the *empirical spectral distribution* (E.S.D.) of A:

$$\mu_A = \frac{1}{n} \sum_{i \in [n]} \delta_{\lambda_i},$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A in increasing order and δ_{λ_i} is the Dirac measure at the eigenvalue λ_i.

Remark 2.1.3. Note that if A is random then the E.S.D. will be a random probability measure.

Definition 2.1.4. The *semicircular distribution* is the probability measure with density function

$$\sigma(x) = \frac{1_{[-2,2]}(x)}{2\pi} \sqrt{4 - x^2}$$

with respect to the Lebesgue measure.

Remark 2.1.5. Note that the support of the semicircular distribution is the closed interval $[-2, 2]$.

We are ready now to present the main result of this subsection.

Theorem 2.1.6 (Semicircular law)**.** *Suppose A_n, $n \in \mathbb{N}$, is a sequence of random matrices such that*

1. *$A_n \in M_n[\mathbb{R}]$ and A_n is symmetric for all $n \in \mathbb{N}$, or $A_n \in M_n(\mathbb{C})$ and A_n is Hermitian for all $n \in \mathbb{N}$.*

2. *For every n, all the entries of A_n are independent random variables with zero mean. Moreover the diagonal entries of A_n are identically distributed. Likewise for the non-diagonal entries.*

3. *For every* $n \in \mathbb{N}$, $\mathbb{E}(A_n(1,2)^2) = 1$, *and for every* $k \in \mathbb{N}$,

$$\max\{\mathbb{E}\,|A_n(1,2)|^k, \mathbb{E}\,|A_n(1,1)|^k\} < \infty.$$

Here $A_n(1,2)$ *is the* $(1,2)$*-entry of the matrix* A_n *and* $A_n(1,1)$ *is the* $(1,1)$*-entry of* A_n.

4. *For every* $n \in \mathbb{N}$ *we have that* $A_n(1,1)$ *is i.i.d. with* $A_{n+1}(1,1)$ *and* $A_n(1,2)$ *is i.i.d. with* $A_{n+1}(1,2)$, *where, again,* $A_n(i,j)$ *is the* (i,j)*-entry of* A_n.

Let $X_n = (1/\sqrt{n})A_n$, $n \in \mathbb{N}$. *Then the empirical spectral distribution of* X_n *converges weakly in probability to the semicircular distribution.*

We will use the following notation: for any probability measure μ on $\mathbb{R}$ and every function f,

$$\langle \mu, f \rangle := \int_{\mathbb{R}} f d\mu,$$

and if $X_n = (1/\sqrt{n})A_n$ are the matrices above, then

$$\mu_n := \frac{1}{n} \sum_{i \in [n]} \delta_{\lambda_i}$$

and

$$\overline{\mu}_n := \mathbb{E}\Big(\frac{1}{n} \sum_{i \in [n]} \delta_{\lambda_i}\Big).$$

In order to prove the theorem we will need a number of lemmas and the next remark.

Remark 2.1.7. It is easy to compute the moments of the semicircular law. They are given by

$$\langle \sigma, x^k \rangle = \begin{cases} 0 & \text{if } k \text{ is odd} \\ C_{\frac{k}{2}} & \text{if } k \text{ is even,} \end{cases}$$

where

$$C_n = \frac{1}{n+1} \frac{(2n)!}{(n!)^2},$$

$n \in \mathbb{N}$, are the Catalan numbers.

Lemma 2.1.8. *For any positive integer k, $\langle \overline{\mu}_n, x^k \rangle$, converges to $\langle \sigma, x^k \rangle$ as n tends to infinity .*

Note. Note that $\overline{\mu}_n$ is not a random measure.

Proof. Our starting point is

$$\langle \mu_{X_n}, x^k \rangle = \int_{\mathbb{R}} x^k d\mu_{X_n} = \frac{1}{n} \operatorname{tr} X_n^k,$$

which holds true because both sides of the equality are equal to

$$(1/n)(\lambda_1^k + \cdots + \lambda_n^k).$$

Taking expectations and writing ζ_{ij} for the (i,j)-entry of X_n, we get that

$$\langle \overline{\mu}_n, x^k \rangle = \frac{1}{n} \sum_{i_1, i_2, \ldots, i_n = 1}^{n} \mathbb{E} \zeta_{i_1, i_2} \cdots \zeta_{i_{k-1}, i_k} \zeta_{i_k, i_1}. \tag{2.1.1}$$

To compute the sum in the right hand side of the equality we will use combinatorial analysis. Consider a sequence $I = (i_1, i_2, \ldots, i_k)$. This sequence can be thought of as a (multi)graph $G_I = (V_I, E_I)$ as follows:

1. It has as vertex set V_I the set of distinct points of I.

2. A vertex in this graph corresponding to an i_j in I is connected via an (undirected) edge with the vertex corresponding to i_{j+1}, for each $j \in \{1, \ldots, k\}$, with $i_{k+1} = i_1$.

Observe that there may be multiple edges between two given vertices and that the number of edges in the graph G_I (i.e., the cardinality of E_I) is always k.

For each sequence I we define the weight of I, denoted by t_I, as the cardinality of the vertex set V_I of the corresponding graph (or equivalently the cardinality of I). From the independence between entries and the fact that each entry has zero mean, it follows that it suffices to only consider those sequences I for which each edge in the corresponding graph G_I appears at least twice, as otherwise the expectation in (2.1.1) will be zero and the sequence will not contribute to the total sum. So we only need to compute

the sum over all sequences I for which $t_I \leq k/2+1$, because if a sequence I has weight $t_I > k/2+1$, then there are t_I indices $j_1 < j_2 < \cdots < j_{t_I}$ in $[k]$ for which the corresponding vertices in the graph are distinct and since there is an edge of the graph connecting each i_{j_l} with i_{j_l+1}, there would be at least $t_I - 1 > k/2$ edges in the graph between the distinct[1] pairs of vertices $\{i_{j_1}, i_{j_1+1}\}, \{i_{j_2}, i_{j_2+1}\}, \ldots, \{i_{j_{t_I-1}}, i_{j_{t_I-1}+1}\}$, and as there are exactly k edges in each graph corresponding to a sequence, there would not be enough edges left to satisfy the requirement that each edge $\{u, v\}$ in the graph appears at least twice.

Furthermore, we say that two sequences $I = (i_1, \ldots, i_k)$ and $J = (j_1, \ldots, j_k)$ are equivalent, if there exists a bijection in S_n, i.e., a permutation of $[n]$, which, for every $a \in [k]$, maps i_a to j_a. Obviously, if two sequences I and J are equivalent they have the same weight, as $i_a = i_b \iff j_a = j_b$, but more importantly, since the diagonal entries of X_n are i.i.d., and the same is true for all non diagonal entries, their corresponding terms in (2.1.1) are equal. Moreover, observe that the number of distinct equivalence classes depends on k but not on n, since each class has a representative where all $i_1, i_2, \ldots, i_k$ are in $\{1, \ldots, k\}$ (we can assume that $n > k$ since we will be concerned with the limit as n tends to infinity).

Given a sequence $I = (i_1, \ldots, i_k)$ with weight t, the number of sequences equivalent to it is

$$n(n-1)\cdots(n-t+1) \leq n^t, \tag{2.1.2}$$

because we obtain a sequence equivalent to I as follows. If $\{v_1, \ldots, v_t\}$ are the distinct elements of the set $\{i_1, \ldots, i_k\}$, i.e., the vertex set V_I of the graph G_I, we obtain a sequence $J = (\pi(i_1), \ldots, \pi(i_k))$ equivalent to I, where π is a permutation of $[n]$, simply by choosing the values $\pi(v_1), \ldots, \pi(v_t)$, and there are $n(n-1)\cdots(n-t+1)$ ways of doing this if these values are to be distinct and lie in the set $[n]$. This then completely determines J, because for each $a \in [k]$, $i_a = v_p$ for some $p \in [t]$ and hence $j_a = \pi(i_a) = \pi(v_p)$.

As a result, for a sequence $I = (i_1, \ldots, i_k)$ with weight $t_I < k/2+1$ we

[1] If $j_1 = 1$ and $j_{t_I} = k$, then the pair of vertices $\{i_{j_1}, i_{j_1+1}\} = \{i_1, i_2\}$ may coincide with $\{i_{t_I}, i_{t_I+1}\} = \{i_k, i_1\}$, i.e., we might have $i_k = i_2$.

have that

$$\frac{1}{n}\mathbb{E}\zeta_{i_1,i_2}\cdots\zeta_{i_{k-1},i_k}\zeta_{i_k,i_1} \leq C_{k,t}\cdot\frac{1}{n}\cdot\frac{1}{\sqrt{n}^k},$$

with $C_{k,t}$ being a constant depending only on k and t (since A_n has uniformly bounded moments for all n). So the total sum of all the sequences equivalent with I in the sum in (2.1.1) is $O(n^{t-k/2-1})$. So it is negligible as $n \longrightarrow \infty$.

Also, if k is odd then $t \neq k/2+1$, so the limit is zero (as one might suspect from the previous remark). Next we focus on the case where k is even and the equivalent classes with weight $t = k/2+1$ (and $k/2$ unique edges since we can distinct $k/2$ edges by assigning every vertex, except i_k, to an edge which has this vertex as first coordinate and appears for the first time in the sequence). For each such I we get that there are no loops in the graph (meaning there are no equal successive points in the sequence) since otherwise we could obtain the simple sub-graph, of that graph, with all the vertices and all the edges (each edge once), except the loops. Call that graph G. Then G is a connected (since there exists a path from any vertex to any other vertex) simple graph. But $|E(G)| < |V(G)| - 1$ which would be a contradiction. As a result we get that in every sequence with $t = k/2+1$ we must have every edge appearing exactly twice. So for each such sequence we have that

$$\frac{1}{n}\mathbb{E}\zeta_{i_1,i_2}\cdots\zeta_{i_{k-1},i_k}\zeta_{i_k,i_1} = \frac{1}{n}\cdot\frac{1}{\sqrt{n}^k}, \tag{2.1.3}$$

since all the non diagonal entries of A_n have variance one.

So by (2.1.2) and (2.1.3), we get that if k is even, and writing m for the number of equivalence classes of sequences with weight $k/2+1$ and length k,

$$\lim_{n\to\infty}\langle\overline{\mu}_n, x^k\rangle = m.$$

But since every class has a representative with vertices in $[k]$ and every type sequence with vertices in $[k]$ belongs to a class, it does not depend of n. For every sequence, with weight $t = k/2+1$, as above we define its *type*

sequence $(s_j)_{j=1}^{k}$ as

$$s_j = \sum_{i=1}^{j} a_i$$

where $a_j = 1$ if the edge $\{j, j+1\}$ appears for the first time in the sequence, and $a_j = -1$ otherwise. So every type sequence has the following properties:

(i) It starts with 1 ends with 0.

(ii) $|s_j - s_{j-1}| = 1 \ \forall j \in [k]$.

(iii) $s_j \geq 0 \ \forall j \in [k]$.

We will show that these properties characterize type sequences, i.e., every sequence with these properties is a type sequence for some equivalence class.

Proposition 2.1.9. *Two sequences in* $[k]$ *are equivalent iff they have the same type sequence.*

Proof. If two sequences (call them I and J) are equivalent then they have the same type sequence since an edge will appear for the first time in I if and only if it appears for the first time in J, which is true by what was done before.

For the other direction we will use induction: For $k = 2$ the assertion is true (obvious).

Suppose it is true for all $m \leq k - 1$. Then let I, J be two sequences with the same type sequence $(\{s\}_{i=1}^{k})$.

(i) We know that the corresponding graph of I and J is a tree. So there exist at least two leaves in I and in J respectively. The leaves in the sequences are the vertices that appear only once in the sequence (otherwise, there would be at least four edges in the sequence that would participate and since each distinct edge appears exactly twice we would have a contradiction).

(ii) Every leaf of the corresponding tree of I belongs to a symmetric sub-sequence of I with the leaf in the center of it (meaning a sub-sequence

such that every edge appears twice, a first time before the leaf and a second after, and every vertex appears also exactly twice, and in each appearance it has the same distance from the leaf). Pick the maximum such sequence for every leaf.

(iii) Note that a maximal sub-sequence of a leaf corresponds to a maximal sub-sequence of the type sequence of I (or J) with the property that it has the same number of $+1$'s and -1's with the $+1$'s preceding in the sequence.

(iv) Every maximum sub-sequence of the type sequence of I (or J) which has the same number of $+1$'s and -1's and the $+1$ appearing first, corresponds to a path, with every edge appearing twice, in the graph of I (or J) starting and ending at the same point. If not (let the length of the sub-sequence be h) this implies that there exists i such that the i-th ($i > h/2$) element of the sequence is -1 does not represent the edge that appears at the $i-h/2$ spot of the sub-sequence. Then there exists a circle in the corresponding graph of I starting and ending at the first vertex of the edge which appears at the i-th spot of the sub-sequence. This is a contradiction since the corresponding graph is a tree.

(v) The end-point of every path mentioned in (iv) (the center vertex in the corresponding sub-sequence of I or J) must be a leaf. For a given path call that point a. This is true, otherwise there would exist a cycle in the corresponding graph (from the vertex that connects with a in the path to the vertex that does not belong in the path but connects with a and to a) which would be a contradiction since the corresponding graph is a tree.

So we have proven that a vertex is a leaf in the sequence I (or J) if and only if there exists a maximal sub-sequence in the type sequence which has the same amount of $+1$'s and -1's and starts with $+1$. So a vertex is a leaf in I if and only if the corresponding vertex in J is a leaf. So by deleting

two leaves of I and the corresponding leaves of J and using the induction hypothesis we get that I and J are equivalent. □

Proposition 2.1.10. *Every sequence with the properties* (i)-(iii) *is the type sequence of an equivalence class.*

Proof. We will use induction. For $k = 2$ the statement is true.

Assume that the result holds true for all $m < k$. Suppose that $\{s_i\}_{i=1}^{k}$ is a sequence that satisfies (i)-(iii).

Using the same techniques we used in the previous proposition we get that we can delete a maximal sub-sequence of s_i that contains the same amount of $+1$'s and -1's in non-increasing order (since all terms of the sequence are non-negative there exists at least one). If the remaining sequence is empty then we associate the sequence to the sequence with weight $k/2 + 1$ and length k

$$\mathbf{i} = (i_1, \ldots, i_{k/2}, i_{k/2+1}, i_{k/2}, \ldots, i_1),$$

where $(i_1, i_2, \ldots, i_{k/2+1}) \subseteq [k]$. If the remaining sequence is not empty then we can apply the induction hypothesis and since the deleted sequence corresponds to a path as mentioned in the previous proposition we get that there exists an equivalent class with type sequence s_i. □

So we need to count the type sequences of length k. We start by choosing any subset of $[k]$ of length $k/2$. Then we assign the value $+1$ to every a_j with j in the chosen subset and -1 for every other j. But this way we have allowed s_j to be negative for some $j \in [k-1]$.

We will prove next that in order to count all the sequences with a negative term it suffices to count all the subsets of $[k]$ with cardinality $k/2 - 1$. We start with the next observation.

Fact: If X and Y are two finite sets for which there exists a surjective map $f : X \to Y$, then $|Y| \leq |X|$.

The proof of this fact is elementary since for every y in Y we can pick a point x in X so that $f(x) = y$. Let A be the subset of all these points x in X. Then, $|Y| = |A| \leq |X|$.

So let

$$X = \left\{ A \subseteq [k] : |A| = \frac{k}{2} - 1 \right\}$$

and

$$Y = \{[s]_{j=0}^{k} : \exists j_0 \in [k] : s_{j_0} < 0\} \cap \{|s_j - s_{j-1}| = 1\} \cap \{s_k = 0\} \cap \{s_0 = 0\}.$$

We will prove that X and Y have the same cardinality using the previous fact. For each sequence $[s]_{j=0}^{k} \in Y$, define $a_j = s_j - s_{j-1}$, $j \in \{1, \ldots, k\}$. Since $[s]_{j=0}^{k}$ determines $[a]_{j=0}^{k}$ and vice-versa, we can equivalently think of Y as the set of sequences of length k, with the same number of $+1$'s and -1's, for which there exists $j \in \{2, \ldots, k\}$ such that there are more -1's than $+1$'s in the set of coordinates up to j.

Let $f : X \to Y$ be defined as follows. For $A \in X$ define $f(A)$ by setting $a_j = +1$ for every $j \in A$ and also j is the largest integer with the property that

$$|[j] \cap [k] \setminus A| = \max\{|[i] \cap [k] \setminus A| \, , i \in [k]\}. \tag{2.1.4}$$

We will prove that f is well defined. Let $A, B \in X$ such that $f(A) \neq f(B)$.Then there exists $i \in [k]$ such that $f(A)_i \neq f(B)_i$. Without loss of generality suppose that $f(A)_i = 1$.

Then either $i \in A$ in which case (since $f(B)_i = -1 \implies i \in B^c$) we get that $B \neq A$ either i satisfies (2.1.4) for A and not for B (otherwise $f(B)_i = 1$) so $A \neq B$.

Also f is surjective since for every $y \in Y$ we can define A to be the set of all j with $s_j - s_{j-1} = +1$ except for the largest j for which $s_{j-1} = \min\{s_i \, , i \in [k]\}$. Then $f(A) = y$.

Conversely, let $g : Y \to X$ be defined as follows. Let $\{s_i\}_{i=1}^{k} \in Y$. Pick the smallest negative term. Note that the smallest negative term must be an odd number. So let $2l - 1$, where $l \in [k/2]$, be that number. We create a new sequence $\{d_i\}_{i=1}^{k}$ setting

$$d_i = \begin{cases} s_i & \text{if } i \leq 2l - 1 \\ -s_i & \text{otherwise} \end{cases}.$$

Define $g(\{s_i\}_{i=1}^k)$ to be the set of j with $d_j - d_{j-1} = +1$. Note that the cardinality of the set of j with $d_j - d_{j-1} = +1$ will be exactly $k/2 - 1$ (since $2l - 1$ was the first negative term).

The map described by the procedure above is well defined. Also, given a subset (call it A) of $[k]$ with cardinality $k/2 - 1$:

(i) We find the smallest coordinate that does not belong to A and there are more coordinates which do not belong in A than the coordinates that belong to A, until that point.

(ii) After that point we follow the reverse procedure to the one described above for all the coordinates (assume that every coordinate in A is assigned with $+1$ and every coordinate not in A with -1).

(iii) The result of that procedure will be a sequence in Y whose g-image will be A. So, g is onto.

(iv) From the previous remark we have that $|X| = |Y|$. As a result, since Y is exactly the set of all the sequences which are not type sequences, the cardinality of the set of all the type sequences is:

$$\binom{k}{\frac{k}{2}} - \binom{k}{\frac{k}{2} - 1} = C_{\frac{k}{2}},$$

which proves the lemma.

□

Lemma 2.1.11. *Fix $\epsilon > 0$ and $k \in \mathbb{N}$. Then:*

$$\lim_{n \longrightarrow \infty} \mathbb{P}(|\langle \mu_n, \chi^k \rangle - \langle \overline{\mu}_n, \chi^k \rangle| > \epsilon) = 0.$$

Proof. From Chebychev's inequality we get:

$$\mathbb{P}(|\langle \mu_n, \chi^k \rangle - \langle \overline{\mu}_n, \chi^k \rangle| > \epsilon) \leq \frac{1}{\epsilon^2} |\mathbb{E}(\langle \mu_n, \chi^k \rangle^2) - (\mathbb{E}\langle \mu_n, \chi^k \rangle)^2|.$$

Again, as in the previous lemma, we can rewrite moments in terms of matrix traces:

$$\begin{aligned} |\mathbb{E}(\langle \mu_n, \chi^k \rangle^2) - (\mathbb{E}\langle \mu_n, \chi^k \rangle)^2| &= \frac{1}{n^2} |\mathbb{E}(\mathrm{tr} X_n^k)^2 - (\mathbb{E}\mathrm{tr} X_n^k)^2| \\ &= \frac{1}{n^2} \sum_{I,J} |\mathbb{E}\zeta_I \zeta_J - \mathbb{E}\zeta_I \mathbb{E}\zeta_J|, \end{aligned} \tag{2.1.5}$$

where ζ_I stands for the product $\zeta_{i_1,i_2}\cdots\zeta_{i_{k-1},i_k}\zeta_{i_k,i_1}$, $I=(i_1,i_2,\ldots,i_k)\in[n]^k$ and $\zeta_{i,j}$ is the (i,j)-entry of X_n. Similarly for ζ_J.

As before, each pair (I,J) generates a (multi)graph with vertices

$$V_{I,J}=(i_1,i_2,\ldots,i_k)\cup(j_1,j_2,\ldots,j_k)$$

and edges

$$E=(i_1i_2,\ldots,i_ki_1)\cup(j_1j_2,\ldots,j_kj_1).$$

As before, the weight of (I,J) is defined as the cardinality of $V_{I,J}$. Also, as in the previous lemma, two pairs (I,J) and (W,D) are called equivalent if there exists a bijection on S_n mapping corresponding indices. As in the previous lemma, equivalent pairs of sequences have the same weight and contribute the same in the sum in (2.1.5).

Also if a term in (2.1.5) corresponding to (I,J) is non-zero then we necessarily have:

(i) Each edge in $E_{I,J}$ should appear at least twice since the entries of X_n have mean zero and are pairwise independent.

(ii) The graphs generated by I and J (as in the previous lemma) should have at least one edge in common, otherwise from independence, $\mathbb{E}\zeta_I\zeta_J=\mathbb{E}\zeta_I\mathbb{E}\zeta_J$.

So, as in the previous lemma, we get that for a pair (I,J) to be non-zero we must have $t\le k/2+1+k/2+1-1=k+1$. More precisely, $t\le k$. To see this, suppose that $t=k+1$. Then, since the graph produced by I, and equivalently for J, must contain each edge twice, there could not exist a common edge in I and J which is a contradiction.

Also, given a pair of sequences (I,J) there are $n(n-1)\cdots(n-t+1)\le n^{k+1}$ equivalent pairs (as in the previous lemma). Moreover, the contribution of each such sequence in the sum (2.1.5) is $\mathbb{O}(1/n^{k+2})$ since $X_n=A_n/\sqrt{n}$, the entries of X_n are independent, and the moments of A_n are uniformly bounded for every n. Thus, the equivalent classes with weight $t\le k+1$ contribute an asymptotically negligible ($\mathbb{O}(\frac{1}{n^2})$) amount to (2.1.5). Finally, since the number of equivalent classes depends on k and not on n (as in the previous lemma) the sum in (2.1.5) tends to zero as $n\to\infty$. □

We are now ready to prove the theorem.

Proof of Theorem 2.1.6. To conclude that $\mu_n \to \sigma$ in probability, in the weak sense, we need to prove that for any bounded continuous function $f : \mathbb{R} \longrightarrow \mathbb{R}$,

$$\langle \mu_n, f \rangle \longrightarrow \langle \sigma, f \rangle \qquad \text{in probability.}$$

The trick is to replace f by a polynomial (by the Weierstrass theorem) and thus, we can rewrite the integrals above as a linear combination of the moments (since integrals are linear). Because f needs to be compactly supported (and since σ has support $[-2, 2]$) by Markov's inequality we get that

$$\mathbb{P}(\langle \mu_n, |x|^k 1_{|x|\geq 5} \rangle \geq \epsilon) \leq \frac{1}{\epsilon}\mathbb{E}(\langle \mu_n, |x|^k 1_{|x|\geq 5} \rangle) \leq \frac{\langle \overline{\mu_n}, x^{2k} \rangle}{\epsilon 5^k}.$$

In the last inequality we used the fact that $\chi^k/5^k \geq 1$ inside the interval. Using the fact that $C_k \leq 4^k$ and the previous lemma, we let n tend to infinity and get

$$\limsup_{n\to\infty} \mathbb{P}(\langle \mu_n, |x|^k 1_{|x|\geq 5} \rangle \geq \epsilon) \leq \langle \sigma, x^{2k} \rangle \leq \frac{4^k}{5^k \epsilon},$$

which holds true for every $k \in \mathbb{N}$. Since the left hand side becomes larger as k grows (the sets become bigger as k becomes bigger), the right hand side must either be strictly increasing as k gets larger or zero. Since the right hand side is decreasing as k grows, we get that the left hand side is zero.

Next, consider $\delta > 0$ and let $f : \mathbb{R} \to \mathbb{R}$ be bounded and continuous (we can assume that f has compact support, say $[-5, 5]$, considering what was done previously). Let p_δ be a polynomial such that $|p_\delta(x) - f(x)| < \delta/4$ for all $x \in [-5, 5]$. Then, from the triangle inequality we get

$$|\langle \mu_n, f \rangle - \langle \sigma, f \rangle| \leq |\langle \mu_n, f - p_\delta \rangle 2 - \langle \sigma, f - p_\delta \rangle| + |\langle \mu_n, p_\delta \rangle - \langle \sigma, p_\delta \rangle|.$$

Splitting p_δ into the parts where it is smaller or bigger than five in the first inequality, and by the way it was chosen, we get:

$$\begin{aligned} |\langle \mu_n, f \rangle - \langle \sigma, f \rangle| \leq \frac{\delta}{2} + |\langle \mu_n, p_\delta 1_{|x|\geq 5} \rangle| + |\langle \mu_n, p_\delta \rangle - \langle \overline{\mu_n}, p_\delta \rangle| \\ + |\langle \overline{\mu_n}, p_\delta \rangle - \langle \sigma, p_\delta \rangle|. \end{aligned}$$

Applying this inequality we get

$$\begin{aligned}\mathbb{P}(|\langle \mu_n, f\rangle - \langle \sigma, f\rangle| \geq \delta) \leq \mathbb{P}\Big(|\langle \mu_n, p_\delta 1_{|x|\geq 5}\rangle| \geq \frac{\delta}{2}\Big) \\ + \mathbb{P}\Big(|\langle \mu_n, p_\delta\rangle - \langle \overline{\mu_n}, p_\delta\rangle| \geq \frac{\delta}{2}\Big) \\ + \mathbb{P}\Big(|\langle \overline{\mu_n}, p_\delta\rangle - \langle \sigma, p_\delta\rangle| \geq \frac{\delta}{2}\Big).\end{aligned}$$

The first term in this inequality tends to zero from what was done before, and the same is true for the second and the third by the previous lemmas.

□

We can also prove that the convergence can be stronger (meaning almost surely).

Remark 2.1.12. For any fixed k there exists a constant C_k not depending on n such that for sufficient large n:

$$|\mathbb{E}(\langle \mu_n, \chi^k\rangle)^2 - (\mathbb{E}\langle \mu_n, \chi^k\rangle)^2| \leq \frac{C_k}{n^2}$$

We have essentially proven in lemma 3.9 that the term above is $\mathbb{O}(1/n^2)$.

Corollary 2.1.13. *The convergence in the semicircular law is with probability* 1 *(almost surely).*

Proof. By Chebyshev's inequality,

$$\begin{aligned}\sum_{n=1}^{\infty} \mathbb{P}(|\langle \mu_n, \chi^k\rangle - \langle \overline{\mu}_n, \chi^k\rangle| > \epsilon) &\leq \sum_{n=1}^{\infty} \frac{1}{\epsilon^2} |\mathbb{E}(\langle \mu_n, \chi^k\rangle)^2 - (\mathbb{E}\langle \mu_n, \chi^k\rangle)^2| \\ &\leq c + \sum_{n=1}^{\infty} \frac{C_k}{n^2},\end{aligned}$$

where c is a constant, since the inequality in the previous corollary is true for large n.

So, the corollary follows from the Borel-Cantelli lemma. Using the same techniques as in the proof of the semicircular law we conclude the proof. □

Remark 2.1.14. It has been proven that for the convergence as presented in Corollary 2.1.13 we can assume that only the second moment of the entries is finite. The proof of this generalisation is done by approximation via matrices with entries which have all their moments finite. For a proof, see [13].

2.1.2 Gaussian isoperimetry

In this subsection we present the main isoperimetric inequalities for the Gaussian measure on $\mathbb{R}^n$ and concentration inequalities that are consequences of them and will be used in the next subsection. We start with the definition of the concentration function in the general setting of a metric probability space.

Definition 2.1.15. Let (X, d, μ) be a metric probability space. The concentration function of the space is defined on $(0, \infty)$ as follows:

$$a_\mu(t) = \sup\left\{1 - \mu(A_t) : \mu(A) \geq \frac{1}{2}\right\},$$

where $A_t = \{x \in X : d(x, A) \leq t\}$.

Proposition 2.1.16. *The concentration function satisfies*

$$\lim_{t\to\infty} a_\mu(t) = 0.$$

Proof. It is clear that the concentration function is decreasing. Now, let $0 < \epsilon < \frac{1}{2}$ and $x \in X$. Note that since

$$\lim_{n\to\infty} B(x, n) = \bigcup_{n=1}^{\infty} B(x, n) = X,$$

by the continuity of the measure we get that there exists $r \in \mathbb{N}$ such that

$$\mu(B(x, r)) \geq 1 - \epsilon.$$

Then, for any Borel subset A of X with $\mu(A) > \frac{1}{2}$ we get that

$$\mu(A \cap B(x, r)) > 0,$$

which implies that $B(x, r) \subseteq A_{2r}$, since there exists $a \in A$ such that $d(a, x) < r$, and hence for every $y \in B(x, r)$ we get $d(y, a) \leq d(x, y) + d(y, a) < 2r$. Then, for every $t \geq 2r$,

$$1 - \mu(A_t) \leq 1 - \mu(A_{2r}) \leq 1 - \mu(B(x, r)) < \epsilon.$$

□

Now we are going to present Erhard's inequality and, as a consequence, the Gaussian isoperimetric inequality. We are going to give a sketch of the proofs of both these fundamental results but we will not get into the details. For more detailed proofs see [14].

Definition 2.1.17. We use the notation γ_n for the probability measure on $\mathbb{R}^n$ with density function (with respect to the Lebesque measure on $\mathbb{R}^n$)

$$d\gamma_n(x) = (2\pi)^{-\frac{n}{2}} e^{-\|x\|_2^2/2} dx.$$

Note: The probability measure γ_n can be thought as follows. For every $A \in B(\mathbb{R}^n)$

$$\gamma_n(A) = \mathbb{P}(X \in A)$$

where X is an n-dimensional random vector whose entries are independent random variables such that $X_i \sim N(0,1)$ for all $i \in [n]$.

Definition 2.1.18. We use the notation $\Phi(x)$ for the distribution function of the standard normal random variable.

Definition 2.1.19. Let $n \in \mathbb{N}$ and $k \in [n]$. Let F be an $(n-k)$-dimensional subspace of $\mathbb{R}^n$ and let e be any unit vector orthogonal to F. For every $A \subseteq \mathbb{R}^n$ which is open or closed, we define $A' \subseteq \mathbb{R}^n$ (which will be called the Gaussian k-symmetrization of A with respect to F along e) as follows. For every $x \in F$

(i) If $\gamma_k(A \cap (x + F^\perp)) = 0$ then $A' \cap (x + F^\perp) = \emptyset$.

(ii) If $\gamma_k(A \cap (x + F^\perp)) = 1$ then $A' \cap (x + F^\perp) = x + F^\perp$.

(iii) If $0 < \gamma_k(A \cap (x + F^\perp)) < 1$ then: if A is open we set $A' \cap (x + F^\perp) = H(e, a) \cap (x + F^\perp)$, while if A is closed we set $A' \cap (x + F^\perp) = \overline{H(e, a)} \cap (x + F^\perp)$. Here a is defined so that $\gamma_k(A \cap (x + F^\perp)) = \gamma_k(H(e, a) \cap (x + F^\perp))$.

We may also use the notation $S(A)$ or $S_{F,e}(A)$ for the Gaussian symmetrization of A.

For every $x \in \mathbb{R}^n$ and any $r \in \mathbb{R}$ we use the notation

$$H(x, r) = \{y \in \mathbb{R}^n : \langle y, x \rangle > r\}.$$

Lemma 2.1.20. *For any $n \in \mathbb{N}$ and every $k \in [n]$, any k-Gaussian symmetrization S has the following properties:*

(i) *If $A \subseteq B$ then $S(A) \subseteq S(B)$.*

(ii) *If $\{A_j\}_{j\in\mathbb{N}}$ are open subsets of $\mathbb{R}^n$ then $S(\cup_j A_j) = \cup_j S(A_j)$.*

(iii) $S_{F,u}(A^c) = [S_{F,-u}(A)]^c$.

(iv) $S_{F,u}(A) + (F + \langle u\rangle)^\perp = S_{F,u}(A)$.

(v) *For any $z \in F$ we have $S(A + z) = S(A) + z$.*

(vi) *If $B \in B(\mathbb{R}^n)$ and $F^\perp + B = B$ then*

$$\gamma_n(B \cap A) = \gamma_n(B \cap S_{F,u}(A)).$$

Moreover,

$$\gamma_n(A) = \gamma_n(S(A)).$$

(vii) *If S is a 1 or 2-Gaussian symmetrization and A is a closed subset of $\mathbb{R}^n$ then $S(A)$ is also a closed subset of $\mathbb{R}^n$.*

Theorem 2.1.21. *If $C \in \mathbb{R}^n$ is closed and convex then every Gaussian symmetrization $S(C)$ of C is convex too.*

For the proof of Theorem 2.1.21 we need some lemmas.

Lemma 2.1.22. *Let F_1, F_2, F_3 be pairwise orthogonal subspaces of $\mathbb{R}^n$ and let $u \in S^{n-1}$ be orthogonal to F_i. Define $S_1 = S_{F_1+F_2,u}$ and $S_2 = S_{F_2+F_3,u}$. If A and $S_2(A)$ are closed subsets of $\mathbb{R}^n$ then*

$$S_1(S_2(A)) = S_{F_2,u}(A).$$

Proof. Set $F = (F_1 + F_2 + F_3 + \langle u\rangle)^\perp$. By Lemma 2.1.20 (claim (iv)) we have that, for every closed set A,

$$S_1(A) = S_1(A) + (F_1 + F_2 + \langle u\rangle)^\perp = S_1(A) + F_3 + F$$

and

$$S_2(A) = S_2(A) + F_1 + F.$$

Also, by the previous lemma, since $F_1 \subseteq F_1 + F_2$ we get that

$$S_1(S_2(A)) = S_1(S_2(A)) + F_1.$$

Using the previous equalities we get

$$S_1(S_2(A)) = S_1(S_2(A)) + (F_2 + \langle u \rangle)^{\perp}.$$

From Lemma 2.1.20, again, we have

$$S_{F_2,u}(A) = S_{F_2,u}(A) + (F_2 + \langle u \rangle)^{\perp}.$$

Since for the sets $S_1(S_2(A))$ and $S_{F_2,u}(A)$ we may apply claim (v) of Lemma 2.1.20, and since symmetrization preserves measure, if $k = \dim(F_2^{\perp})$ then for every $x \in F_2$ and for the set $R_x = x + F_2^{\perp}$ we have

$$\gamma_k(S_2(A) \cap R_x) = \gamma_k(A \cap R_x) = \gamma_k(S_1(A) \cap R_x) = \gamma_k(S_1(S_2(A) \cap R_x)).$$

It follows that $S_1(S_2(A)) = S_{F_2,u}(A)$. □

Lemma 2.1.23. *Let $m \in \mathbb{N}$ with $m > 2$ and $k \in [m]$ with $k > 1$. For any k-Gaussian symmetrization $S_{F,u}$ we can find 2-Gaussian symmetrizations $T_1, T_2, \ldots, T_{k-1}$ so that*

$$S_{F,u} = T_1 \circ T_2 \circ \cdots \circ T_{k-1}.$$

Proof. Let $v \in (F + \langle u \rangle)^{\perp}$. We define three subspaces as follows:

$$F_3 = (F + \operatorname{span}\{u, v\})^{\perp}, F_2 = F, F_1 = \langle v \rangle,$$

and apply the previous lemma for the symmetrizations

$$S_1 = S_{F_1+F_2,u}, S_2 = S_{F_2+F_3,u}.$$

Note that S_2 is a 2-symmetrization so it preserves closeness of sets. So, for every closed set A we get

$$S_1 \circ S_2(A) = S_{F_2,u}.$$

Setting $T_{k-1} = S_2$ and continuing inductively we can prove the lemma. □

Proof of Theorem 2.1.21. We will prove the assertion of the theorem for every 2-Gaussian symmetrization and by the previous lemma the theorem will be true for every $k \in \mathbb{N}$.

Firstly we will prove it for 1-Gaussian symmetrizations. Let $u \in S^{n-1}$, set $F = \langle u \rangle^{\perp}$ and consider the 1-symmetrization $S_{F,u}$. The symmetrization is done along the lines of the form $R_x = x + \langle u \rangle$ where $x \in F$. Let $A \subseteq \mathbb{R}^n$ be convex and closed. We clearly have

$$A \cap R_x = x + D\langle u \rangle$$

where $D \subseteq \mathbb{R}$. Then D is convex, since both A and R_x are convex. So let $a, b \in D$. Then for any c such that $a < c < b$ we have that

$$c = \Big(1 - \frac{b-c}{b-a}\Big)b + \frac{b-c}{b-a}a \in D.$$

So, D is an interval. Also, since A and R_x are closed, we get that D is also closed. So,

$$R_x \cap A = x + [a_x, b_x], \qquad a_x, b_x \in \mathbb{R} \cup \{-\infty, +\infty\}, \quad a_x < b_x.$$

As a result we get that since $S(A) \cap R_x = H(u, c_x) \cap R_x$ and hence $\gamma_1(A \cap R_x) = \gamma_1(H(u, c_x) \cap R_x)$, we must have

$$S(A) \cap R_x = [c_x, \infty]u + x,$$

where $c_x = -\Phi^{-1}(\Phi(b_x) - \Phi(a_x))$. So, for the convexity of $S(A)$ it is sufficient to show that, for every $x, y \in F$ and any $\lambda \in [0, 1]$,

$$S(A) \cap R_{\lambda x + (1-\lambda)y} \supseteq \lambda(S(A) \cap R_x) + (1 - \lambda)(S(A) \cap R_y),$$

which is equivalent to the following inequality:

$$[c_{\lambda x + (1-\lambda)y}, \infty] \supseteq \lambda[c_x, \infty] + (1 - \lambda)[c_y, \infty].$$

So it is sufficient to show that

$$c_{\lambda x + (1-\lambda)y} \leq \lambda c_x + (1 - \lambda)c_y,$$

which is equivalent to the following:

$$\Phi^{-1}(\Phi(b_{\lambda x+(1-\lambda)y}) - \Phi(a_{\lambda x+(1-\lambda)y})) \geq \lambda\Phi^{-1}(\Phi(b_x) - \Phi(a_x)) + (1-\lambda)\Phi^{-1}(\Phi(b_y) - \Phi(a_y)).$$

Let $d_1 \in R_x \cap A$ and $d_2 \in R_y \cap A$. Then, for any $\lambda \in [0,1]$, we have that $\lambda d_1 + (1-\lambda)d_2 \in [\lambda x + (1-\lambda)y + \langle u \rangle]$ and $\lambda d_1 + (1-\lambda)d_2 \in A$ since A is convex. So,

$$\lambda x+(1-\lambda)y+\lambda[a_x, b_x]+(1-\lambda)[a_y, b_y] \subseteq \lambda x+(1-\lambda)y+[a_{\lambda x+(1-\lambda)y)}, b_{\lambda x+(1-\lambda)y}],$$

which implies

$$a_{\lambda x+(1-\lambda)y} \leq \lambda a_x + (1-\lambda)a_y \leq \lambda b_x + (1-\lambda)b_y \leq b_{\lambda x+(1-\lambda)y}.$$

Then, we show the following, which will imply the desired inequality:

$$\Phi^{-1}(\Phi(\lambda b_x + (1-\lambda)b_y) - \Phi(\lambda a_x + (1-\lambda)a_y)) \geq \lambda\Phi^{-1}(\Phi(b_x) - \Phi(a_x)) + (1-\lambda)\Phi^{-1}(\Phi(b_y) - \Phi(a_y)).$$

So we need to prove that the function $g(a,b) = \Phi^{-1}(\Phi(b) - \Phi(a))$ on $\{(a,b) \in \mathbb{R}^2 : a < b\}$ is concave, which can be proven by computing the Hessian matrix of g.

In order to show the theorem for any 2-symmetrization, it is now enough to prove the following.

Lemma 2.1.24. *Any 2-symmetrization in $\mathbb{R}^n$ is the limit a sequence of compositions of 1-symmetrizations.*

Proof. We will show the lemma in $\mathbb{R}^2$. One can check that the following sequence $\{T_j\}_{j\in\mathbb{N}}$ of symmetrizations approaches any 2-symmetrization in $\mathbb{R}^2$:

$$T_j = S_j \circ S_{j-1} \circ \cdots \circ S_1 \circ S_0,$$

where

$$S_j = S_{e_{j+1}^{\perp}, e_j}$$

and

$$e_j := \left[\cos\left(\frac{3\pi}{2} + 2^{-j}\pi\right), \sin\left(\frac{3\pi}{2} + 2^{-j}\pi\right)\right].$$

Note that $e_j + e_0 \perp e_{j+1}$ and $e_j \to -e_0$. □

Proposition 2.1.25. *Note that for every $c, c' > 0$ and any closed $A \subseteq \mathbb{R}^n$, and $x \in T_j(A)$,*

$$x + ce_0 + c'e_j \in T_j(A).$$

Proof. For a proof see [14]. □

Combining the previous lemma with the fact that every 1-symmetrization preserves convexity we get that every 2-symmetrization preserves convexity. □

Theorem 2.1.26 (Erhard's inequality)**.** *For every pair of non-empty closed convex subsets $A, B \subseteq \mathbb{R}^n$ the following inequality is true for every $\lambda \in [0, 1]$:*

$$\Phi^{-1}(\gamma_n(\lambda A + (1-\lambda)B)) \geq \lambda\Phi^{-1}(\gamma_n(A)) + (1-\lambda)\Phi^{-1}(\gamma_n(B)).$$

Proof. A sketch of the proof is the following. Firstly suppose that A and B are also compact. Then consider the sets

$$A' = A \times \{1\}, B' = B \times \{0\}$$

and

$$C = \{y \in \mathbb{R}^{n+1} : y = \lambda a + (1-\lambda)b, a \in A', b \in B', \lambda \in [0, 1]\}.$$

Let $e = (0, 0, \ldots, 1) \in \mathbb{R}^{n+1}$ and $u = (1, 0, \ldots, 0) \in \mathbb{R}^{n+1}$. Obviously, u is a unit vector orthogonal to $\langle e \rangle$. Since C is convex, we have that $S_{\langle e \rangle, u}(C)$ is also convex. So it is true that

$$S(C) \cap (\mathbb{R}^n \times \{\lambda x + (1-\lambda)y)\}) \supseteq \lambda S(C) \cap [\mathbb{R}^n \times \{x\}] + (1-\lambda)S(C) \cap [\mathbb{R}^n \times \{y\}] \quad (2.1.6)$$

since $S(C)$ is convex. But for any $z \in \mathbb{R}$ we know that $S(C) \cap (R^n \times \{z\}) = H(u, r_z) \cap (\mathbb{R}^n \times \{z\})$, where r_z is defined by

$$\gamma_n(C \cap (\mathbb{R}^n \times \{z\})) = \gamma_n(H(u, r_z) \cap (\mathbb{R}^n \times \{z\})),$$

which implies that

$$\Phi(-r_z) = \gamma_n(C \cap (\mathbb{R}^n \times \{z\})).$$

So, $r_z = -\Phi^{-1}(\gamma_n((C \cap (\mathbb{R}^n \times \{z\}))$. But, by (2.1.6),

$$r_{\lambda x + (1-\lambda)y} \leq \lambda r_x + (1-\lambda)r_y.$$

Therefore, the function

$$Q(\lambda) = \Phi^{-1}(\gamma_n(C \cap (\mathbb{R}^n \times \{\lambda\}))) = \Phi^{-1}(\gamma_n(\lambda A + (1-\lambda)B))$$

is concave on $[0,1]$. This completes the proof of Erhard's inequality, since it is equivalent with the following inequality:

$$Q(\lambda) \geq (1-\lambda)Q(0) + \lambda Q(1).$$

In the case of non-compact sets we can approximate them from the "inside" by compact sets and thus prove the desired inequality in the general case. □

Next we present the isoperimetric inequality for the Gaussian measure.

Theorem 2.1.27 (Gaussian isoperimetric inequality)**.** *Let $A \subseteq \mathbb{R}^n$ be a Borel set and define $a \in \mathbb{R}$ by the equation*

$$\gamma_1((-\infty, a]) = \gamma_n(A).$$

Then, for any $\epsilon > 0$ we have

$$\gamma_n(A_\epsilon) \geq \gamma_1(-\infty, a + \epsilon),$$

or equivalently

$$\Phi^{-1}(\gamma_n(A_\epsilon)) \geq \Phi^{-1}(\gamma_n(A)) + \epsilon.$$

A first proof may be given via Erhard's inequality.

Proof of Theorem 2.1.27. Let $A \subseteq \mathbb{R}^n$ and $B_2^n = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$. Note that

$$A_\epsilon = A + \epsilon B_2^n = (1-\lambda)[(1-\lambda)^{-1}A] + \lambda(\lambda^{-1}\epsilon B_2^n).$$

From Erhard's inequality we get

$$\Phi^{-1}(\gamma_n(A_\epsilon)) \geq (1-\lambda)\Phi^{-1}(\gamma_n((1-\lambda)^{-1}A)) + \lambda\Phi^{-1}(\gamma_n(\lambda^{-1}\epsilon B_2^n)).$$

Now letting $\lambda \to 0^+$ it is clear by the continuity of Φ^{-1} and the continuity of γ_n that the first term of the right hand side of the previous inequality tends to $\Phi^{-1}(\gamma_n(A))$. For the second term we observe that

$$\lambda\Phi^{-1}(\gamma_n(\lambda^{-1}\epsilon B_2^n)) = \epsilon(\lambda^{-1}\epsilon)^{-1}\Phi^{-1}(\gamma_n(\lambda^{-1}\epsilon B_2^n)) \to \epsilon,$$

which proves the desired inequality. □

We shall give a second proof of the Gaussian isoperimetric inequality using the spherical isoperimetric inequality and the next lemma which is attributed to Poincaré although it seems it was known before Poincaré.

Lemma 2.1.28. *Let $n \in \mathbb{N}$. Then for any $N > n$ we denote by $P_{N+1,n}$ the projection from $\mathbb{R}^{N+1}$ onto $\mathbb{R}^n$. Let $\sigma^N = \sqrt{N}s^N$ be the Haar probability measure of the sphere $\sqrt{N}S^N$ of radius $\sqrt{N}$ in $\mathbb{R}^{N+1}$. Then, for every Borel subset A of $\mathbb{R}^n$,*

$$\lim_{N\to\infty} \sigma^N(P_{N+1,n}^{-1}(A) \cap \sqrt{N}S^N) = \gamma_n(A).$$

Proof. A sketch of the proof is as follows: Let $\{g_i\}_{i\in\mathbb{N}}$ be a sequence of independent random variables such that $g_i \sim N(0,1)$ for all $i \in \mathbb{N}$. Also, for every $k \geq 1$ let $R_k^2 = \sum_{i=1}^k g_i^2$. Then the distribution of the random vector $\frac{\sqrt{N}}{R_{N+1}}(g_1, g_2, \ldots, g_{N+1})$ is σ^N. So, the distribution of $\frac{\sqrt{N}}{R_{N+1}}(g_1, g_2, \ldots, g_n)$ is $\sigma^N(P_{N+1,n})$. Note that $R_n^2, R_{N+1}^2 - R_n^2$ and $\frac{1}{R_n}(g_1, g_2, \ldots, g_n)$ are pairwise independent random variables. So, R_n^2/R_{N+1}^2 and $\frac{1}{R_n}(g_1, g_2, \ldots, g_n)$ are independent. Also note that from the properties of the χ-squared distribution (see the previous subsection) $R_n^2 \sim \text{Gamma}(n/2, 2)$ and $R_{N+1}^2 - R_n^2 \sim \text{Gamma}((N+1-n)/2, 2)$. So $\frac{R_n^2}{R_n^2+R_{N+1}^2-R_n^2} \sim \text{Beta}(\frac{n}{2}, \frac{(N+1-n)}{2})$. It follows that

$$\begin{aligned}\sigma^N(P_{N+1,n}^{-1}(A) \cap \sqrt{N}S^N) &= \mathbb{P}\left(\frac{\sqrt{N}}{R_{N+1}}(g_1, g_2, \ldots, g_n) \in A\right) \\ &= \mathbb{P}\left(\frac{\sqrt{N}R_n}{R_{N+1}}\frac{1}{R_n}(g_1, \ldots, g_n) \in A\right),\end{aligned}$$

which implies that

$$\begin{aligned}&\sigma^N(P_{N+1,n}^{-1}(A) \cap \sqrt{N}S^N) \\ &= \int_{S^{n-1}} \int_0^1 1_{\{t\in[0,1]:x\sqrt{Nt}\in A\}} \frac{1}{B(n/2,(N+1-n)/2)} t^{\frac{n}{2}-1}(1-t)^{\frac{N+1-n}{2}-1} dt d\sigma^n(x).\end{aligned}$$

Setting $r = \sqrt{Nt}$ we get

$$\begin{aligned}&\sigma^N(P_{N+1,n}^{-1}(A) \cap \sqrt{N}S^N) \\ &= B(n/2,(N+1-n)/2)^{-1} \frac{2}{N^{n/2}} \int_{S^{n-1}} \int_0^{\sqrt{N}} 1_A(rx) r^{n-1}(1-\frac{r^2}{N})^{\frac{N+1-n}{2}-1} dr d\sigma^n(x).\end{aligned}$$

By the dominated convergence theorem we get

$$\lim_{N\to\infty} \sigma^N(P_{N+1,n}^{-1}(A) \cap \sqrt{N}S^N) = \frac{2}{\Gamma(n/2)2^{n/2}} \int_{S^{n-1}} \int_0^\infty 1_A(rx) r^{n-1} e^{-\frac{r^2}{2}} dr d\sigma^n(x)$$

which is exactly $\gamma_n(A)$ in polar coordinates. □

Now we can return to the proof of the Gaussian isoperimetric inequality.

Second proof of Theorem 2.1.27. Firstly note that for S^{n-1}, the unit sphere in $\mathbb{R}^n$ equipped with the geodesic metric, there exists a similar result to the Gaussian isoperimetric inequality, the spherical isoperimetric inequality, which states that:

Given $a \in (0,1)$ let $B_q(x,r)$ the ball with the geodesic metric $q(x,y)$, such that $s^n(B_q(x,r)) = a$. Then, for every $A \subseteq S^{n-1}$ such that $s^n(A) = a$ and for every $t > 0$ we have that

$$s^n(A_t) \geq s^n(B_q(x, r+t)).$$

So the previous lemma reduces the proof to the isoperimetric inequality on the sphere. Moreover, for a Borel set $A \subseteq \mathbb{R}^n$ and an $a \in \mathbb{R}$ such that $\Phi(a) = \gamma_n(A)$ it is true that, for sufficient large $N \in \mathbb{N}$ and any $b < a$,

$$\sigma^N(P_{N+1,n}^{-1}(A) \cap \sqrt{N}S^N) \geq \sigma^N(P_{N+1,1}^{-1}(-\infty, b) \cap \sqrt{N}S^N).$$

Also, for any $t > 0$ one can show that

$$P_{N+1,n}^{-1}(A_t) \cap \sqrt{N}S^N \supseteq [P_{N+1,n}^{-1}(A) \cap \sqrt{N}S^N]_t.$$

Here in the right hand side of the inclusion the t-extension is taken with respect to the geodesic metric. A very important observation is that the set $P_{N+1,1}^{-1}(-\infty, b) \cap \sqrt{N}S^N$ is a geodesic ball in $\sqrt{N}S^N$. So, by the isoperimetric inequality on the sphere, we have

$$\begin{aligned} s^N(P_{N+1,n}^{-1}(A_t) \cap \sqrt{N}S^N) &\geq s^N([P_{N+1,n}^{-1}(A) \cap \sqrt{N}S^N]_t) \\ &\geq \sigma^N([P_{N+1,1}^{-1}(-\infty, b) \cap \sqrt{N}S^N]_t). \end{aligned}$$

One can show that

$$[P^{-1}(-\infty, b) \cap \sqrt{N}S^N]_t = P^{-1}(-\infty, b + a_N) \cap \sqrt{N}S^N,$$

where

$$\lim_{N\to\infty} a_N = t.$$

So, by the previous lemma,

$$\gamma_n(A_t) \geq \Phi(b + t).$$

Since this last inequality is true for any $b < a$, we get that

$$\gamma_n(A_t) \geq \Phi(a + t).$$

□

A consequence of the isoperimetric inequality is the following lemma.

Lemma 2.1.29. *If $f : \mathbb{R}^n \to \mathbb{R}$ is 1-Lipschitz and M_f is the median of the function on the probability space $(\mathbb{R}^n, \gamma_n, B(\mathbb{R}^n))$ then*

$$\gamma_n(|f - M_f| \geq t) \leq 2\gamma_1(t, \infty).$$

Proof. Let $A = \{x \in \mathbb{R}^n : f(x) \geq M_f\}$ and $B = \{x \in \mathbb{R}^n : f(x) \leq M_f\}$. From the definition of the median we have $\gamma_n(A) \geq \frac{1}{2}$ and $\gamma_n(B) \geq \frac{1}{2}$.

So, there exist $a, b \geq 0$ such that $\gamma_1(-\infty, a) = \gamma_n(A)$ and $\gamma_1(-\infty, b) = \gamma_n(B)$. Also, since f is 1-Lipschitz we get that if $y \in A_t$ then there exists $x \in A$ such that $\|x - y\|_2 < t$ and hence

$$f(y) = f(x) - f(x) + f(y) \geq -d(y, x) + M_f \geq M_f - t.$$

Similarly, if $y \in B_t$ then there exists $x \in B$ such that $\|x - y\|_2 < t$. Therefore,

$$f(y) = f(x) - f(x) + f(y) \leq d(x, y) + M_f \leq t + M_f.$$

It follows that if $y \in A_t \cap B_t$ then

$$|f(y) - M_f| < t,$$

which implies that

$$\gamma_n(|f - M_f| \geq t) \leq \gamma_n(A_t^c \cup B_t^c) \leq \gamma_n(A_t^c) + \gamma_n(B_t^c).$$

On the other hand, by the isoperimetric inequality we have

$$\gamma_n(A_t^c) \leq \gamma_1(a + t, +\infty) \leq \gamma_1(t, +\infty)$$

and

$$\gamma_n(B_t^c) \leq \gamma_1(b + t, +\infty) \leq \gamma_1(t, +\infty)$$

since $a, b \geqslant 0$. □

Finally, we shall also need the next simple lemma.

Lemma 2.1.30. *If $Z \sim N(0,1)$ then $\mathbb{P}(Z > z) \leq e^{\frac{-z^2}{2}}$ for any $z > 0$.*

Proof. For any $\lambda > 0$ and any $z > 0$ we have that

$$\mathbb{P}(Z > z) = \mathbb{P}(e^{\lambda Z - \lambda z} \geq 1) \leq e^{-\lambda z}\mathbb{E}(e^{\lambda Z}) = e^{-\lambda z}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{\lambda x}e^{-\frac{x^2}{2}}\,dx.$$

But

$$\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{\lambda x}e^{-\frac{x^2}{2}}\,dx = e^{\frac{\lambda^2}{2}}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{(x-\lambda)^2}{2}}\,dx = e^{\frac{\lambda^2}{2}}$$

since the density function for a random variable $X \sim N(\lambda, 1)$ is

$$f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\lambda)^2}{2}}.$$

As a result we get

$$\mathbb{P}(Z > z) \geq e^{-\lambda z + \frac{\lambda^2}{2}}.$$

For $\lambda = z$ we get the desired inequality. □

2.1.3 Convergence of the extreme eigenvalues

In this subsection we seek the limit of the extreme eigenvalues of A_n when $A_{i,j} \sim N(0,1)$.

Lemma 2.1.31. *If A is a symmetric $n \times n$ matrix on $\mathbb{R}$ (or hermitian, respectively, on $\mathbb{C}$) then*

$$\|A\|_\infty = \max\{-\lambda_1, \lambda_n\},$$

where λ_1 is the smallest eigenvalue of A and λ_n is the largest one.

Proof. If A is symmetric then there exists an orthogonal matrix Q and a diagonal matrix D such that

$$A = QDQ^*.$$

Also, if $x \in \mathbb{R}^n$ is such that $\|x\|_2 = 1$ we get that $\|Q^*x\|_2 = 1$ and the function

$$x \mapsto Q^*x$$

is a bijection. So,

$$\begin{aligned}\|A\|_\infty &= \sup_{\{x\in\mathbb{R}^n:\|x\|_2=1\}} |x^*Ax| = \sup_{\{x\in\mathbb{R}^n:\|x\|_2=1\}} |x^*QDQ^*x| \\ &= \sup_{\{y\in\mathbb{R}^n:\|y\|_2=1\}} |y^*Dy| = \max\{-\lambda_1, \lambda_n\}.\end{aligned}$$

□

Lemma 2.1.32. *If A is an $n \times n$ symmetric or hermitian matrix then we have*

$$\|A\|_\infty \leq \|A\|_{\mathrm{HS}}.$$

Proof. Let $\{\lambda_i\}_{i\in[n]}$ be the eigenvalues of A. Then, for every $i \in [n]$,

$$|\lambda_i| \leq \Big(\sum_{j=1}^n \lambda_j^2\Big)^{\frac{1}{2}}.$$

So,

$$\max_{i\in[n]} |\lambda_i| \leq \Big(\sum_{j=1}^n \lambda_j^2\Big)^{\frac{1}{2}}.$$

□

Therefore, in the space M_n^{sa} of all symmetric (or Hermitian) $n \times n$ matrices with the Hilbert-Schmidt norm, all we get is that the function $\|\cdot\|_\infty$ is 1-Lipschitz.

Proposition 2.1.33. *Let A_n be a symmetric (or Hermitian) random matrix whose entries are independent random variables with standard normal distribution. Then,*

$$\mathbb{E}\|A_n\|_\infty < 2\sqrt{n}.$$

Proof. We will prove it for Hermitian matrices. The symmetric case is similar.

Let $A \in M_n(\mathbb{C})$ be a Hermitian matrix whose entries are independent random variables such that

$$\{a_{j,j}\}_{j\in\mathbb{N}}, \{\sqrt{2}\mathrm{Re}(a_{j,k})\}_{j<k}, \{\sqrt{2}\mathrm{Im}(a_{j,k})\}_{j<k} \sim N(0,\sigma^2).$$

So, the probability distribution μ of A (as a probability measure on the set $M_n^{sa}(\mathbb{C})$ of self-adjoint matrices) has density

$$d\mu(H) = c_1 \exp\left(-\frac{\mathrm{tr}(H^2)}{2\sigma^2}\right) dH,$$

where

$$c_1 = \frac{1}{(2\pi\sigma^2)^{n^2/2}}$$

and dH is the Lebesque measure on $M_n^{sa}(\mathbb{C})$.

If we consider the set

$$\Lambda_n = \{\{\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n : \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n\}$$

it is known (see [15]) that the function $h : M_n^{sa} \to \Lambda_n$ which sets the eigenvalues of a self-adjoint matrix in increasing order maps the probability measure μ, mentioned above, to

$$h(d\mu) = d\mu(h^{-1}) = c_2 \prod_{1\leq j<k\leq n} (\lambda_j - \lambda_k)^2 \exp\left(\frac{-1}{2\sigma^2}\sum_{k=1}^{n} \lambda_k^2\right) d\lambda_1 d\lambda_2 \cdots d\lambda_n,$$

where $c_2 > 0$ is another normalization constant:

$$c_2 = \left(\pi^{n(n-1)/2} \prod_{j=1}^{n-1} (j!)\right)^{-1}.$$

Hence, after averaging over all permutations of $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ we get that for any symmetric function $\phi : \mathbb{R}^n \to \mathbb{C}$ one has (when both integrals are defined)

$$\int_{M_n^{sa}} \phi(\lambda_1(H), \lambda_2(H), \ldots, \lambda_n(H))dH = \int_{\mathbb{R}^n} \phi(\lambda)g(\lambda)d\lambda_1 d\lambda_2 \cdots d\lambda_n,$$

where

$$g(\lambda) = \frac{c_2}{n!} \prod_{1 \le j < k \le n} (\lambda_j - \lambda_k)^2 \exp\left(\frac{-1}{2\sigma^2} \sum_{k=1}^{n} \lambda_k^2\right).$$

The marginal density

$$h(\lambda) = \int_{\mathbb{R}^{n-1}} g(\lambda, \lambda_2, \ldots, \lambda_n)d\lambda_2 \cdots d_{\lambda_n}$$

can be computed explicitly (see [15]). Moreover, for $\sigma^2 = \frac{1}{2}$ we have

$$h(\lambda) = \sum_{k=0}^{n-1} \phi_k^2(\lambda)$$

where

$$\phi_k(x) = \frac{1}{(\sqrt{\pi}k!2^k)^{1/2}} e^{-\frac{x^2}{2}} H_k(x).$$

Here $\{H_k\}_{k \in \mathbb{N}_0}$ are the Hermite polynomials defined as follows:

$$H_k(x) = (-1)^k e^{x^2} \left(\frac{d^k}{dx^k} e^{-x^2}\right).$$

So, in general one can show that for any σ^2

$$h(\lambda) = \frac{1}{n\sqrt{2\sigma^2}} \sum_{k=0}^{n-1} \phi_k\left(\frac{\lambda}{\sqrt{2\sigma^2}}\right).$$

In order to prove the proposition we need the following lemma.

Lemma 2.1.34. *Let $f : \mathbb{R} \to \mathbb{R}$ be a Borel function and consider the mapping $A \mapsto f(A)$, $a \in M_n^{sa}(\mathbb{C})$ obtained by the usual function calculus for self-adjoint operators on Hilbert space. Then,*

$$\mathbb{E}(\mathrm{tr}(f(A))) = n \int_{\mathbb{R}} f(\lambda)h(\lambda)d\lambda$$

given that the right hand side of the equality is well defined, i.e. either $f \ge 0$ or $\int_{\mathbb{R}} |fh|d\lambda < \infty$.

Proof. By what was done before, and since the function

$$\mathrm{tr}(f(A)) = \sum_{i=1}^{n} f(\lambda_i(A))$$

is a symmetric function over the eigenvalues of A, we have

$$\mathbb{E}(\mathrm{tr}(f(A)) = \int_{\mathbb{R}^n} \sum_{j=1}^{n} f(\lambda_j) g(\lambda_1, \lambda_2, \ldots, \lambda_n) d\lambda_1 d\lambda_2 \cdots d\lambda_n.$$

But one can easily verify that g is invariant under permutations of the λ_i's, so

$$\mathbb{E}(\mathrm{tr}(f(A))) = n \int_{\mathbb{R}} f(\lambda) h(\lambda) d\lambda.$$

For the general case we can consider f^+ and f^-. □

Corollary 2.1.35. *For any $s \in \mathbb{C}$,*

$$\begin{aligned}\mathbb{E}(\mathrm{tr}(\exp(sA))) &= \frac{1}{n\sigma\sqrt{2}} \int_{\mathbb{R}} e^{s\lambda} \sum_{k=0}^{n-1} \phi_k\left(\frac{\lambda}{\sigma\sqrt{2}}\right) d\lambda \\ &= n \exp\left(\frac{s^2\sigma^2}{2}\right) \Phi(1-k, 2, -s^2\sigma^2),\end{aligned}$$

where Φ is the hyper-geometric function defined for any $a, c, x \in \mathbb{C}$ by

$$\Phi(a, c, x) = \sum_{n=0}^{\infty} \frac{(a)_n x^n}{(c)_n n!}.$$

In order to proceed we need the following lemma.

Lemma 2.1.36. *Let $\sigma = 1$. Define $C(p, n) = \mathbb{E}(\mathrm{tr}(A^{2p}))$. Then, $C(0, 1) = n$, $C(1, n) = n^2$ and for every $n \in \mathbb{N}$ the numbers $C(p, n)$ satisfy the recursion formula*

$$C(p+1, n) = n\frac{C(p, n)(4p+2)}{p+2} + \frac{p(4p^2-1)}{p+2} C(p-1, n).$$

Proof. Let $a, c \in \mathbb{C}$ such that $c \in \mathbb{Z} \setminus \mathbb{N}$. Then the function

$$x \to \Phi(a, c, x)$$

is an entire function which satisfies the differential equation ($y = \Phi(a, c, x)$)

$$x\frac{d^2y}{dx^2} + (c - x)\frac{dy}{dx} - ay = 0$$

(see [16]). By what was done before we get

$$\mathbb{E}(\mathrm{tr}(\exp(sA)) = n\,\exp\left(\frac{s^2\sigma^2}{2}\right)\Phi(1 - n, 2, -s^2).$$

But since all the moments of sA are finite, by the properties of the moment generating function (note that A has the same distribution as $-A$ so $\mathbb{E}(\mathrm{tr}(A^{2q-1})) = 0$ for any $q \in \mathbb{N}$) we get

$$\mathbb{E}(\exp(sA)) = \sum_{p=0}^{\infty} \frac{s^{2p}}{(2p)!}\mathbb{E}(\mathrm{tr}(A^{2p})).$$

It follows thus, that $\frac{C(p,n)}{(2p)!}$ is the coefficient of x^p in the power series expansion of the function

$$\sigma_n(x) = n\exp\left(\frac{n}{2}\right)\Phi(1 - n, 2, -x).$$

Since $\Phi(a, b, x)$ satisfies the differential equation mentioned above, we get

$$x\sigma_n''(x) + 2\sigma_n'(x) - \left(\frac{x}{4} + n\right)\sigma_n(x) = 0.$$

Therefore, $\sigma_n(x) = \sum_{p=0}^{\infty} a_p x^p$ where $a_p = \frac{C(p,n)}{(2p)!}$. Going back to the differential equation and equating the power series we get

$$(p + 1)(p + 2)a_{p+1} - na_p - \frac{1}{4}a_{p-1} = 0$$

for $p \in \mathbb{N}$ and

$$2a_1 - na_0 = 0.$$

So, since $\mathrm{tr}(I_n) = \mathrm{tr}(A^0)) = n$ we get the recursion formulas. □

The above discussion shows that the quantity $d_p := \mathbb{E}(tr(A^{2p}))\frac{1}{n^p 2^{2p}}$ for $p \geq 1$ and $d_0 = 1$ satisfies the recursion formula

$$d_p = \frac{2p - 1}{2p + 2}\left(d_{p-1} + \frac{p(p - 1)(2p - 3)}{4n^2 2p}d_{p-2}\right)$$

and $d_1 = \frac{n}{4}$. Since we know that for $p \in \{0, 1\}$

$$d_p \leq n\binom{2p}{p}\frac{1}{2^{2p}(p+1)}\prod_{j=1}^{p}\left(1+\frac{j(j-1)}{4n^2}\right),$$

we may use induction and prove that for any $p \in \mathbb{N}_0$ the previous inequality is true.

Next, using successively Stirling's formula to majorize the binomial coefficient, the inequalities $\sum_{j=1}^{p} j(j-1) \leq \frac{p^3}{3}$ and $1 + x \leq e^x$ to estimate the product, and denoting $t = pn^{-2/3}$, we arrive at the following estimate:

$$d_p \leq n\frac{e^{\frac{p^3}{12n^2}}}{\sqrt{\pi}p^{\frac{3}{2}}} = \frac{e^{\frac{t^3}{12}}}{\sqrt{t^3\pi}}.$$

This is valid for $t > 0$, at least if the corresponding value of $p = tn^{2/3}$ is an integer. So, for $t = 1, 3$ and for sufficiently large p, by the continuity of $\frac{e^{\frac{t^3}{12}}}{\sqrt{t^3\pi}}$, after a calculation we get

$$d_p \leq e^{-0.3} < 1 \implies \mathbb{E}\mathrm{tr}(A^{2p})^{1/2p} < 2\sqrt{n}.$$

By Jensen inequality and since for any $p \in \mathbb{N}$ it is true that $\mathrm{tr}(A^{2p}) \geq \|A^{2p}\|_\infty$, we get

$$\mathbb{E}\|A\|_\infty \leq \mathbb{E}(\|A^{2p}\|_\infty)^{1/2p} \leq \mathbb{E}(\mathrm{tr}(A^{2p}))^{1/2p} < 2\sqrt{n}.$$

□

Lemma 2.1.37. *If $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function then the median M_f of f with respect to the standard Gaussian measure γ_n on $\mathbb{R}^n$ satisfies*

$$M_f \leq \mathbb{E}(f).$$

Proof. Firstly we are going to prove that the function $g : t \to \Phi^{-1}(\gamma_n\{f \leq t\})$ is concave. Note that, since f is convex, for any $t \in \mathbb{R}$ the set $A_t := \{x \in \mathbb{R}^n : f(x) \leq t\}$ is convex: for any $x, y \in A$ and $\lambda \in (0, 1)$ we have that

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \leq \lambda t + (1-\lambda)t.$$

Applying Erhard's inequality we have

$$\lambda\Phi^{-1}(\gamma_n(A_{t_1})) + (1-\lambda)\Phi^{-1}(\gamma_n(A_{t_2})) \leq \Phi^{-1}(\gamma_n(\lambda A_{t_1} + (1-\lambda)A_{t_2})),$$

which proves that g is concave. Note that $g(M_f) = 0$ so there exists a supporting line to the graph of g. More precisely, there exists $a > 0$ such that, for all $t \in \mathbb{R}$,

$$g(t) \leq a(t - M_f) + g(M_f) = a(t - M_f).$$

An equivalent way to express the previous inequality is

$$\gamma_n(\{f \leq t\}) \leq \mathbb{P}(Z \leq t),$$

where $Z \sim N(M_f, a^{-2})$. Since stochastic domination implies inequality for the means of random variables we have

$$M_f \leq \mathbb{E}(f).$$

□

Now we can state and prove the main result of this subsection.

Theorem 2.1.38. *Let A_n a sequence of $n \times n$ real symmetric (or complex Hermitian) random matrices whose entries are independent random variables all following the standard normal distribution. Then, for any epsilon > 0,*

$$\mathbb{P}\left(\lambda_1\left(\frac{1}{\sqrt{n}}A_n\right) \geq 2 + \epsilon\right) \to 0$$

and

$$\mathbb{P}\left(\lambda_n\left(\frac{1}{\sqrt{n}}A_n\right) \leq -2 - \epsilon\right) \to 0,$$

where $\{\lambda_i(A)\}_{i\in[n]}$ are the eigenvalues of an $n \times n$ matrix A in decreasing order.

Proof. Note that

$$\begin{aligned}\max\left\{\mathbb{P}\left(\lambda_1\left(\frac{1}{\sqrt{n}}A_n\right) \geq 2 + \epsilon\right), \mathbb{P}\left(\lambda_n\left(\frac{1}{\sqrt{n}}A_n\right) \leq -2 - \epsilon\right)\right\} \\ = \mathbb{P}\left(\left\|\frac{1}{\sqrt{n}}A_n\right\|_\infty \geq 2 + \epsilon\right)\end{aligned}$$

by Lemma 2.1.31. We have also proven in Lemma 2.1.32 that $\|\cdot\|_\infty$ is a convex 1-Lipschitz function on the space of the $n \times n$ symmetric (or Hermitian) matrices with the Hilbert-Schmidt norm. So,

$$\mathbb{P}\left(\left\|\frac{1}{\sqrt{n}}A_n\right\|_\infty \geq 2+\epsilon\right) = \gamma_n(\{\|x\|_\infty \geq 2\sqrt{n}+\epsilon\sqrt{n}\}.$$

Since the infinity norm is also convex, from Lemma 2.1.37 we have

$$M_{\|\cdot\|_\infty} \leq \mathbb{E}_{\gamma_n}(\|\cdot\|_\infty) \leq 2\sqrt{n}.$$

So by the previous lemmas we get that

$$\mathbb{P}(\|A_n\|_\infty \geq 2\sqrt{n}+\epsilon\sqrt{n}) \leq 2\gamma_1(\epsilon\sqrt{n},\infty) \leq 2e^{\frac{-\epsilon^2 n}{2}} \to 0.$$

Since it is obvious that $\mathbb{P}(\limsup_n \lambda_n \leq -2) = 0$ and $\mathbb{P}(\liminf_n \lambda_1 \geq 2) = 0$, by the semicircular law the proof is complete. □

Corollary 2.1.39. *By the Borel-Cantelli lemma we conclude that*

$$\lambda_1(A_n) \to -2 \ a.s.$$

and

$$\lambda_n(A_n) \to 2 \ a.s.$$

2.2 Marchenko-Pastur Law

2.2.1 Convergence of E.S.D.

We are going to present now another important theorem in random matrix theory which can be thought as a generalisation of the semicircular law and was first proven in [3] by Marchenko-Pastur. The techniques and the notation we are going to use are similar to the ones we used in the proof of the semicircular law.

Theorem 2.2.1. *Let $X_n \in M_{p\times n}$ be a sequence of random matrices such that all the entries are i.i.d., $\mathbb{E}(X_{1,1}) = 0$ and $\mathbb{E}(X_{1,1}^2) = 1$ and $X_{1,1}$ has finite moments, as in the previous theorem. Let*

$$S_n = \frac{1}{n}X\cdot X^* \in M_p.$$

Also let μ_n be the the E.S.D. of S_n. Assume that $p(n)/n \to y \in (0,1]$. Then μ_n converges (weakly) to μ almost surely, where μ is the deterministic measure with density (with respect to the Lebesque measure)

$$d\mu = \frac{1}{2\pi xy}\sqrt{(b-x)(x-a)}\mathbf{1}_{a\le x\le b}$$

and

$$a(y) = (1-\sqrt{y})^2, b(y) = (1+\sqrt{y})^2.$$

Remark 2.2.2. Observe that if $y = 1$ then μ is the semicircular distribution under the mapping $x \to x^2$.

In order to prove the theorem we will need the following.

Remark 2.2.3. It is easy to compute that:

$$\int x^k d\mu = \sum_{r=1}^{k-1} \frac{y^{r+1}}{r+1}\binom{k}{r}\binom{k-1}{r}.$$

Lemma 2.2.4. *It is true that*

$$\langle \mu_n, x^k\rangle \longrightarrow \langle \mu, x^k\rangle.$$

Proof. Some details of the proof will not be explained because the techniques that are used are the same as in the proof of the semicircular law. We have

$$\begin{aligned}\langle \mu_n, x^k\rangle &= \frac{1}{p}\mathbb{E}\left(\sum_{i=1}^{p}\lambda_i\right) = \frac{1}{p}\mathbb{E}\left(\frac{\mathrm{tr}(X\cdot X^*)^k}{n^k}\right) \\ &= \frac{1}{pn^k}\sum_{I,J}\mathbb{E}(X_{i_1,j_1}X_{i_2,j_1}X_{i_2,j_2}\cdots X_{i_k,j_k}X_{i_1,j_k}),\end{aligned} \tag{2.2.1}$$

where $I \in [p]^k$ and $J \in [n]^k$.

Note that each term in the sum (2.2.1) is associated with a bipartite (multi)graph with vertex set $V = I \cup J$ and edge set the coordinates on each term of the product in (2.2.1), meaning $E = \cup_{m\in[k]}(i_m, j_m)\cup_{m\in[k]}(i_{m+1 \pmod k}, j_m)$. We can imagine the edge set as a sequence $E = (i_1, j_1, i_2, j_2, \ldots, i_k, j_k, i_1)$ such that any two successive terms of the sequence are edges.

As in the previous theorem, in order for a term to be non-zero, each edge must appear twice meaning that we have at most k edges, and as a result at most $k+1$ vertices in each such graph.

Suppose that the vertex set of a term (I,J) has cardinality $\leq k$. Then let $V = A+B$ where A is the cardinality of I and B is the cardinality of J. Then the total number of ways of choosing A vertices for I and B vertices for J is bounded by $Cp^A n^B$, where C is a constant independent from n. So, the contribution of these terms in the expectation is

$$O(p^A n^B / pn^k) \longrightarrow_{n\to\infty} 0.$$

Thus we need to look at graphs with exactly $k+1$ vertices and k edges (meaning there are no loops and every edge appears exactly twice in the sequence and more importantly in reverse, meaning (i,j) and (j,i)).

Let (I,J) be a pair, where $\mathrm{card}(I) = r+1$ and $\mathrm{card}(J) = k-r$. The number of equivalence classes (there exists a bijection of $[n]\times[p]$ mapping each term of one graph to another) is the number of permutations of $r+1$ objects from p distinct objects and the same respectively for $k-r$ objects from n distinct objectd. So,

$$\binom{p}{r+1}\binom{n}{k-r} = np\left(\frac{p}{n}\right)^r\left(1+\mathbb{O}\left(\frac{1}{n}\right)\right)$$

Thus:

$$\langle \mu_n, x^k\rangle = \frac{1}{pn^k}\sum_{I,J}\mathbb{E}(X_{i_1,j_1}X_{i_2,j_1}X_{i_2,j_2}\cdots X_{i_k,j_k}X_{i_1,j_k}) = \sum_{r=1}^{k-1}\left(\frac{p}{n}\right)^r\left(1+\mathbb{O}\left(\frac{1}{n}\right)\right)\times D_r, \tag{2.2.2}$$

where D_r is the number of equivalence classes with $r+1$ I-vertices and $k-r$ J-vertices. Letting $n\to\infty$ in (2.2.2) we get that it is sufficient to show that:

$$D_r = \frac{1}{r+1}\binom{k}{r}\binom{k-1}{r}.$$

So we need to count D_r. In order to do that, for every equivalence class we define its *type sequence* $[s_i]_{i=1}^{2k} = 1$ if in the edge $(j,j+1)$, $j+1$ appears for the first time in the sequence and $j+1\in I$, and $=-1$ if in the edge $(j,j+1)$, j appears for the last time in the sequence and $j\in I$, Otherwise, we set $=0$. □

Proposition 2.2.5. *Every type sequence is well defined, meaning that each sequence in an equivalence class has the same type sequence.*

Proof. Given two pairs (I, J) and (W, V) with the same type sequence $\{s\}_{i=1}^{2k}$ we can use Proposition 2.1.9 for the sequence $\{d_i\}_{i=1}^{2r} = \{s_i : s_i \in \{-1, 1\}\}$ to find a subgraph of each pair which are equivalent. So we can prove the proposition by adding at the graph above all vertices of J and V respectively, join all edges between an element of I and an element of J that are successive with the same direction,and lastly delete all vertices amongst I and W respectively. Using the same technique for the a sequence with J and W to be non zero and Proposition 2.1.9, we get which points are equal and that J and W are equivalent. So after we merge the elements of J and W that are equal we get the same graph. As a result the two pairs are equivalent. □

Proposition 2.2.6. *Every sequence has the following properties which uniquely determine it:*

(i) *If i is odd then $s_i \in \{0, -1\}$.*

(ii) *If i is even then $s_i \in \{1, 0\}$.*

(iii) *For every $l \in [2k]$ we have that $\sum_{i=1}^{l} s_i \geq 0$.*

(iv) $\#\{i : s_i = 1\} = \#\{i : s_i = -1\} = r$ *(since $\forall i \in I$ we get that i appears exactly once for the first time and once for the last time and those times can not be the same).*

(v) *As a result of the previous properties, $\sum_{i=1}^{2k} s_i = 0$.*

Every type sequence has the properties above (obviously) and every sequence with the properties (i)-(v) *is a type sequence.*

Proof. Given a sequence $[s_i]_{i=1}^{2k}$ with the properties above firstly we distinct the non zero terms and construct a graph as in the equivalent case in Proposition 2.1.10. After that we use the same method as in Proposition 2.1.9 and construct the graph we want.

So we need to count all the sequences with the properties (i)-(v). If A is the set of all those sequences and $X_b = (j \subseteq [b] : |j| = r)$ then we define

$$f : A \longrightarrow X_k \times X_{k-1}$$

where f sends every sequence to the support of the even elements of the sequence (under the mapping $2m \rightarrow m$) times the support of the odd elements of the sequence (under the mapping $2m-1 \rightarrow m$). Note that $s_{2k} = 0$ so we just need a subset of $[k-1]$.

Obviously, f is well-defined and 1-1. Also, every subset of $X_k \times X_{k-1}$ under the mapping f^{-1} satisfies the properties (i), (ii), (iv) and (v). So we just need to count all the subsets which fail property (iii).

So let $\{s_i\}_{i=1}^{2k}$ be a sequence that fails property (iii). Then there exists $l \in [k]$ such that $\sum_{i=1}^{2l-1} s_i < 0$. We pick the smallest l with that property. Then we create the following sequence:

$$d_i = s_i \qquad \forall i \in [2l-1] \cup \{2k\}$$

$$(d_{2i}, d_{2i+1}) = (s_{2i}, s_{2i+1}) \quad \textit{if} \quad l \leq i \leq k-1 \quad \textit{and}$$

$$[(s_i, s_{i+1}) = (0,0) \quad \textit{or} \quad (s_i, s_{i+1}) = (1,-1)]$$

$$(d_i, d_{i+1}) = (s_{i+1}, s_i) \quad \textit{otherwise.}$$

Now the sequence d_i has $r+1$ odd elements assigned with -1 (and the rest 0) and $r-1$ even elements assigned with $+1$ (and the rest 0). The mapping above is clearly well-defined (since there will be one more $+1$ in $[2l, 2k]$ so in the "reflected" sequence d_i this will be reversed so there will be two -1's more than $+1$'s). Also the procedure above is reversible. Given a sequence with $r+1$ odd elements assigned with -1 and $r-1$ assigned with $+1$, and the rest 0, firstly find the first odd number with more -1's than $+1$'s until that point, and then follow the reversed procedure. So the set of sequences with a negative term has cardinality equal to the number of ways in which we can pick $r+1$ -1's from k of them and $r-1$ $+1$'s from $k-1$ of them. So:

$$D_r = \binom{k-1}{r}\binom{k}{r} - \binom{k-1}{r-1}\binom{k}{r+1} = \frac{1}{r+1}\binom{k-1}{r}\binom{k}{r}.$$

□

Lemma 2.2.7. It is true that:

$$\mathrm{Var}_{\mu_n}(x^k) \leq \frac{C_k}{n^2}.$$

Proof of Theorem 2.2.1. The proof of the lemma above and as a result the proof of Marchenko-Pastur law are exactly the same as the proof of the corresponding steps in the proof of the semicircular law. □

Remark 2.2.8. If $y > 1$ then one can show that a similar result is true. A sketch of the proof is the following: Since $\mathrm{rank}(S_n) = \min\{p, n\}$ we will have roughly $n(y-1)$ zero eigenvalues. Since $\mu_n = \frac{1}{p}\sum_{i=1}^{p} \delta_{\lambda_i}$ we see that there will be $(1-y^{-1})$ at 0 in the limiting measure. Since the non-zero eigenvalues of XX^* and X^*X are the same we get that the limiting distribution is

$$(1-y^{-1})\delta_{\{0\}} + \mu,$$

where μ is the same as in the Marchenko-Pastur law.

2.2.2 Convergence of the extreme eigenvalues

We will continue this section with the proof of some results about the limits of the extreme eigenvalues of $S_n = \frac{1}{n}X_nX_n^*$, where $X_{p,n}$ is a matrix as in the beginning of this section, with i.i.d. entries that have distribution $N(0,1)$. The results presented where first proved in [6].

Theorem 2.2.9. *Let $X_{p,n}$ be as above. Let $\lambda_1\left(\frac{1}{n}X_{p,n}X_{p,n}^*\right)$ be the smallest eigenvalue of the matrix $\frac{1}{n}X_{p,n}X_{p,n}^*$ and $p/n \to y$ where $y \in (0,1]$. Then*

$$\lim_{p,n\to\infty} \lambda_1\left(\frac{1}{n}X_{p,n}X_{p,n}^*\right) = (1-\sqrt{y})^2.$$

Proof. First note that the conditions of the Marchenko-Pastur law are satisfied. So we get

$$\mu_n \to \mu \qquad \text{weakly almost surely,}$$

where μ_n is the E.S.D. of $\frac{1}{n}X_{p,n}X_{p,n}^*$ and μ is the measure with density (with respect to the Lebesque measure)

$$d\mu = \frac{1}{2\pi xy}\sqrt{(b-x)(x-a)}\mathbf{1}_{a\leq x\leq b}$$

$$a(y) = (1 - \sqrt{y})^2, b(y) = (1 + \sqrt{y})^2.$$

Let $S_p = \frac{1}{n} X_{p,n} X_{p,n}^*$. The proof of the theorem is based on the following lemma and definition.

Lemma 2.2.10 (Gershgorin circle theorem)**.** *Let A be a complex $n \times n$ matrix with entries $a_{i,j}$. For each $i \in [n]$ let $R_i = \sum_{i \neq j} |a_{i,j}|$ be the sum of the absolute values of all the non-diagonal entries of the i-th row. Let $D_i(a_{i,i}, R_i)$ be the closed disc with center $a_{i,i}$ and radius R_i. Then every eigenvalue of A lies in at least one of those discs.*

Proof. Let λ be an eigenvalue of A. Choose a corresponding eigenvector $x = (x_j)$ so that one of its components x_i is equal to 1 and the others are of absolute value less than or equal to 1. We may always assume that such an x exists, simply by dividing any eigenvector by its component with largest modulus. Since $Ax = \lambda x$, in particular we have

$$\sum_j a_{i,j} x_j = \lambda x_i = \lambda.$$

So, splitting the sum and taking into account once again that $x_i = 1$, we get

$$\sum_{j \neq j} a_{i,j} x_j + a_{i,i} = \lambda.$$

Therefore, applying the triangle inequality,

$$|\lambda - a_{i,i}| = \left| \sum_{j \neq j} a_{i,j} x_j \right| \leq \sum_{j \neq i} |a_{i,j}||x_j| \leq \sum_{j \neq i} |a_{i,j}| = R_i.$$

□

Returning to the theorem, note that since μ is positive to the right of $(1 - y^{1/2})^2$ we immediately get

$$\limsup \lambda_1(S_p) \leq (1 - y^{1/2})^2.$$

Definition 2.2.11. Let $\{X_i\}$ be a sequence of independent standard Gaussian random variables. Then, for every $n \in \mathbb{N}$ we denote by $\chi^2(n)$ the random variable

$$\chi^2(n) = \sum_{i=1}^{n} X_i^2.$$

Now we can return to the proof of the theorem. Assume that $p < n$ (since $p/n \to y < 1$). We will show inductively that we can replace S_p by a matrix which is easier to work with. Let O_n^1 be the matrix whose first column is the first row of $X_{p\times n}$ normalized, by the Gram-Schmidt method, and the rest columns are non-random linearly independent n-dimensional vectors. Then, $X_{p\times n}^1 = X_{p\times n}O_n^1$. After calculation one can see that the first row of $X_{p\times n}^1$ is $(X_n, 0, \ldots, 0)$, where $X_n^2 \sim \chi^2(n)$ and $X_n \geq 0$, while the remaining rows are made up by independent $N(0, 1)$ random variables. Now let O_p^1 be a $p \times p$ orthogonal matrix of the form

$$O_p^1 = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & O_{p-1}^1 \end{bmatrix},$$

where O_{p-1}^1 is orthogonal, its first row is the normalization (by the Gram-Schmidt method) of the first column (without the first element) of $X_{p\times n}^1$, as a vector in $\mathbb{R}^{p-1}$, and the remaining columns are linearly independent. So, after calculations one can verify that

$$O_p^1 X_{p\times n}^1 = \begin{bmatrix} X_n & \mathbf{0} \\ Y_{p-1} & W_{(p-1)\times(n-1)} \\ \mathbf{0} & . \end{bmatrix},$$

where $Y_{p-1}^2 \sim \chi^2(p-1)$ and $Y_{p-1} \geq 0$, while $W_{(p-1)\times(n-1)}$ is a $(p-1)\times(n-1)$ random matrix made up by i.i.d. $N(0, 1)$ entries.

Following the same technique one can show inductively that there exist two orthogonal matrices $O_{p\times p}$ and $O_{n\times n}$ such that

$$O_{p\times p}X_{p\times n}O_{n\times n} = \begin{bmatrix} X_n & 0 & 0 & 0 & 0 & \ldots & 0 \\ Y_{p-1} & X_{n-1} & 0 & 0 & 0 & \ldots & 0 \\ 0 & Y_{p-2} & X_{n-2} & 0 & 0 & \ldots & 0 \\ 0 & 0 & . & . & . & \ldots & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & \\ 0 & 0 & 0 & 0 & Y_1 & X_{s-(n-1)} \ldots 0 & 0 \end{bmatrix}.$$

So, after calculations one can show that S_p is orthogonally similar (so with the same eigenvalues) to the matrix D_p with the following rows:

$$D_1 = \frac{1}{n}(X_n^2, X_n Y_{p-1}, 0, 0, \ldots, 0)$$
$$D_p = \frac{1}{n}(0, 0, \ldots, X_{n-p+2} Y_1, Y_1^2 + X_{n-p+1}^2)$$

and for every $j \in [p-2]$ the $(j+1)$-th row has non-zero diagonal element

$$D_{j+1,j+1} = \frac{1}{n} Y_{p-j}^2 + X_{n-j}^2$$

and non-diagonal elements

$$\frac{1}{n} X_{n-j+1} Y_{p-j}, \frac{1}{n} X_{n-j} Y_{p-j-1},$$

where $\{Y_i\}_{n\in\mathbb{N}}, \{X_i\}_{n\in\mathbb{N}}$ is an independent sequence of random variables such that $X_i^2, Y_i^2 \sim \chi^2(i)$ and $X_i, Y_i \geq 0$, for each i.

By Gershgorin's circle theorem we get

$$\lambda_1 \geq \min\Big\{\frac{1}{n}(X_n^2 - X_n Y_{p-1}), \frac{1}{n}(Y_1^2 + X_{n-p+1}^2 - X_{n-p+2} Y_1),$$
$$\min_j \frac{1}{n}(Y_{p-j}^2 + X_{n-j}^2 - X_{n-j+1} Y_{p-j} - X_{n-j} Y_{p-j-1})\Big\}.$$

Now notice that for a sequence of independent random variables $\{Z_n\}_{n\in\mathbb{N}}$ such that $Z_n \sim \chi^2(1)$ for every $n \in \mathbb{N}$ we have

$$\mathbb{P}(Z_1 = \infty) = 0 \Longrightarrow \lim_m \frac{Z_1}{m} = 0 \qquad a.s.$$

and by Kolmogorov's strong law of large numbers for i.i.d. random variables and the fact that $\mathbb{E}(Z_1) = 1$,

$$\frac{\chi^2(m)}{m} = \frac{\sum_{n=1}^m Z_m}{m} \to_{m\to\infty} 1 \qquad a.s.$$

From these results and the assumption that $p/n \to y \in (0, 1]$ we get

$$\frac{1}{n}(X_n^2 - X_n Y_{p-1}) = \frac{X_n^2}{n} - \frac{\sqrt{p-1}}{\sqrt{n}} \frac{X_n}{\sqrt{n}} \frac{Y_{p-1}}{\sqrt{p-1}} \to 1 - \sqrt{y} \cdot 1 \qquad a.s.$$

and likewise

$$\frac{Y_1^2 + X_{n-p+1}^2 - X_{n-p+2} Y_1}{n} \to 1 - y \qquad a.s.$$

Notice that

$$1 - y \geq 1 - \sqrt{y} \geq (1 - \sqrt{y})^2.$$

Also, for $0 < \epsilon < 1$ and $s, m \in \mathbb{N}$, by Markov's inequality we see that

$$\begin{aligned}\mathbb{P}(e^{t\chi^2(m)-tm} \geq e^{ts\epsilon}) &\leq e^{-ts\epsilon}\mathbb{E}(e^{t\chi^2(m)-tm}) = e^{-ts\epsilon-tm}\mathbb{E}\left(e^{\sum_{n=1}^{m} Z_n t}\right) \\ &= e^{-ts\epsilon}(\mathbb{E}(e^{(Z_1-1)t}))^m.\end{aligned}$$

Now we will show a concentration inequality for a random variable $Z \sim \chi^2(1)$. Let $t \in (0, 1/4)$. Then

$$\mathbb{E}(\exp(t(Z-1)) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(z^2-1)} e^{\frac{-z^2}{2}} dz.$$

Setting $y = \sqrt{1-2tz}$ we get

$$\mathbb{E}(\exp(t(Z-1)) = \frac{e^{-t}}{\sqrt{1-2t}} \int_{-\infty}^{\infty} e^{\frac{-y^2}{2}} dy = \frac{e^{-t}}{\sqrt{1-2t}}.$$

But $t \in (0, 1/4)$, and hence

$$\frac{e^{-t}}{\sqrt{1-2t}} \leq e^{2t^2}.$$

It follows that for $t_0 \in (0, \min\{\frac{1}{4}, \frac{\epsilon}{2}\})$ and $m \in [s]$,

$$\mathbb{P}\left(e^{t\chi^2(m)-tm} \geq e^{ts\epsilon}\right) \leq e^{-ts\epsilon}\left(\mathbb{E}(e^{(Z_1-1)t})\right)^m \leq e^{2t^2 m - ts\epsilon} \leq e^{(2t^2-t\epsilon)s} < 1.$$

Similarly one can show that for every $0 < \epsilon < 1$ and $s \in \mathbb{N}$ there exists a constant $c < 1$, depending only on ϵ, such that for all $m \in [s]$

$$\mathbb{P}\left(e^{-tx^2+tm} > e^{ts\epsilon}\right) \leq c^s.$$

So we can conclude that for every $\epsilon \in (0, 1)$, $s \in \mathbb{N}$ and $m \in [s]$ there exists a constant $a \in (0, 1)$, depending only on ϵ, such that

$$\mathbb{P}\left(\left|\frac{\chi^2(m)}{s} - \frac{m}{s}\right| > \epsilon\right) \leq 2a^s.$$

Returning to the proof of the theorem we have that, for $\epsilon \in (0, 1)$,

$$\begin{aligned}&\sum_{n=1}^{\infty} \mathbb{P}\left(\left\{\max_{n-(p-2)\leq m\leq n} \left|\frac{X^2(m)}{n} - \frac{m}{n}\right| > \epsilon\right\} \cup \left\{\max_{m\leq p-1} \left|\frac{X^2(m)}{n} - \frac{m}{n}\right| > \epsilon\right\}\right) \\ &\leq \sum_{n=1}^{\infty} (2p(n) - 1)4a^n) < \infty.\end{aligned}$$

Thus, by the Borel-Cantelli lemma, the following is valid almost surely:

$$\lim_{n\to\infty}\left(\max\left[\left\{\max_{n-(p-2)\le m\le n}\left|\frac{X^2(m)}{n}-\frac{m}{n}\right|\right\},\left\{\max_{m\le p-1}\left|\frac{X^2(m)}{n}-\frac{m}{n}\right|\right\}\right]\right)=0. \quad (2.2.3)$$

In order to continue we need the following inequality.

Lemma 2.2.12. *For any non-negative real numbers $a_1, a_2, b_1, b_2 \ge 0$ it is true that*

$$|a_1b_1 - a_2b_2| \le |a_2^2 - a_1^2|^{1/2}|b_2^2 - b_1^2|^{1/2} + |a_2^2 - a_1^2|^{1/2}|b_1| + |a_1||b_2^2 - b_1^2|^{1/2}.$$

Proof. We will use the following two facts.

Fact 1: For any two non-negative numbers,

$$a + b \le (\sqrt{a} + \sqrt{b})^2 \Longrightarrow \sqrt{a+b} \le \sqrt{a} + \sqrt{b}.$$

Fact 2: For any two non-negative numbers a, b we have that

$$|a - b| \le |a^2 - b^2|^{1/2}.$$

To see this, assume without loss of generality that $a \le b$. Then

$$2a^2 \le 2ab \Longrightarrow (a-b)^2 \le b^2 - a^2 \Longrightarrow |a-b| \le |b^2 - a^2|^{1/2}.$$

By the previous facts we have

$$\begin{aligned}|a_1b_1 - a_2b_2| &\le a_1|b_1 - b_2| + b_2|a_1 - a_2| \\ &\le a_1|b_1 - b_2| + (|b_2 - b_1| + b_1)|a_1 - a_2| \\ &\le a_1|b_1^2 - b_2^2|^{1/2} + (|b_2^2 - b_1|^{1/2} + b_1)|a_1^2 - a_2^2|^{1/2}.\end{aligned}$$

□

So,

$$\begin{aligned}A_j^n &:= \left|\frac{1}{n}(Y_{p-j}^2 + X_{n-j}^2 - X_{n-j+1}Y_{p-j} - X_{n-j}Y_{p-j-1})\right. \\ &\qquad \left. - \left(\frac{p-j}{n} + \frac{n-j}{n} - \frac{\sqrt{n-j+1}}{\sqrt{n}}\frac{\sqrt{p-j}}{\sqrt{n}} - \frac{\sqrt{n-j}}{\sqrt{n}}\frac{\sqrt{p-j-1}}{\sqrt{n}}\right)\right| \\ &\le \left|\frac{Y_{p-j}^2 - p + j}{n}\right| + \left|\frac{X_{n-j}^2 - n - j}{n}\right| + \left|\frac{X_{n-j+1}Y_{p-j} - \sqrt{n-j}\sqrt{p-j}}{n}\right| \\ &\qquad + \left|\frac{X_{n-j}Y_{p-j-1} - \sqrt{n-j}\sqrt{p-j-1}}{n}\right|.\end{aligned}$$

From the previous inequality and (2.2.3) we get

$$\max_{j\leq p-2} A_j^n \longrightarrow 0 \qquad a.s.$$

But the expression

$$\left|\frac{p-j}{n}+\frac{n-j}{n}-\frac{\sqrt{n-j+1}}{\sqrt{n}}\frac{\sqrt{p-j}}{\sqrt{n}}-\frac{\sqrt{n-j}}{\sqrt{n}}\frac{\sqrt{p-j-1}}{\sqrt{n}}\right|$$

achieves its smallest value for $j = 1$. So,

$$\frac{p-1}{n}+\frac{n-1}{n}-\frac{\sqrt{p-1}}{\sqrt{n}}-\sqrt{\frac{n-1}{n}}\sqrt{\frac{p-2}{n}} \to y+1-2\sqrt{y}.$$

□

For the maximal eigenvalue, the proof of $\lambda_{max} \to (1+y^{1/2})^2$ is similar. We use the same matrix D_p and the fact that for every eigenvalue λ of a matrix A we have $\lambda_i \leq \max\{\sum_{i=1}^{p} |a_{i,j}|\}$ from Gershgorin's circle theorem.

2.3 Bai-Yin's Convergence to the semicircular law

2.3.1 Convergence of the E.S.D.

In this subsection we prove another generalisation of the semicircular law, first proved in [4] by Bai-Yin. We are going to use the notation we used in the discussion of the semicircular law.

Theorem 2.3.1. *Let $X_n \in M_{p\times n}$ be a random matrix with i.i.d. entries for all p and n (also i.i.d. amongst different p and n). Assume that $n(p) \longrightarrow \infty$ and $p/n \longrightarrow 0$, also that $\mathbb{E}|X_{1,1}|^4 < \infty$ and $\mathrm{Var}(X_{1,1}) = 1$. Let*

$$A_p = \frac{1}{2\sqrt{np}}(XX^* - nI_p).$$

Then:

$$\mu_{A_p} \to \sigma \qquad a.s.$$

where σ is the semicircular distribution with density function $\mathbf{1}_{|x|\leq 1}\frac{2}{\pi}\sqrt{1-x^2}$ and μ_A is the E.S.D. of A.

Note that the k-th moment of σ is $\frac{1}{2^k}C_{k/2}$ when k is odd and zero otherwise.

In order to prove the theorem we need several lemmas.

Lemma 2.3.2. *For each p let $Y_p = [X_{i,j_p}]$ be a $p \times n$ random matrix with i.i.d. entries, where $n = n(p) \to \infty$ and $p/n \to 0$ as p tends to infinity, such that:*

1. $\mathbb{E}X_{1,1_p} = 0$ *and* $\mathbb{E}X^2_{1,1_p} = 1 + a_p$, *where* $a_p \to 0$.
2. $|X_{11_p}| \leq \epsilon_p n^{\frac{1}{4}}$, *where* $\epsilon_p \to 0$ *but* $\epsilon_p n^{\frac{1}{4}} \to +\infty$.

Let $B_p = [Z_{i,j}]$ be a $p \times p$ random matrix such that $Z_{i,i} = 0$ and

$$Z_{i,j} = \frac{1}{2\sqrt{np}} \sum_{l=1}^{n} X_{i,l}X_{j,l_p}, \quad i \neq j.$$

Then, μ_{B_p} converges to the semicircular distribution $\sigma(x)$ almost surely.

Proof. As in the semicircular law we will prove the following:

1. $\langle \mu_{B_p}, x^k \rangle \to \langle \sigma, x^k \rangle$.
2. $\sum_{p=1}^{\infty} \mathrm{Var}_{\mu_{B_p}}(x^k) < \infty$.

Firstly we prove (i). We write:

$$\begin{aligned}\langle \mu_{B_p}, x^k \rangle &= \mathbb{E}\left(\frac{1}{p}\operatorname{tr} B_P^k\right) \\ &= \frac{1}{p(2\sqrt{pn})^k} \sum_{I,J} \mathbb{E}\left(X_{i_1,j_1}X_{i_2,j_1}X_{i_2,j_2}X_{i_2,j_3}\cdots X_{i_k,j_k}X_{i_1,j_k}\right),\end{aligned}$$

where $I \subseteq [p]^k$ and $J \subseteq [n]^k$.

Notation: We shall use the following notation:

- $\psi(e_1, e_2, \ldots, e_m)$ is the number of distinct elements among $e_1, \ldots, e_m$.
- $I = (i_1, \ldots, i_k)$ and $J = (j_1, \ldots, j_k)$.
- $i_a \in [p]$ and $j_b = [n]$, where $a, b \in [k]$.
- $r = \psi(I)$ and $c = \psi(J)$.

· $\Gamma(I,J)$ denotes the multi-graph defined as follows: Let the I-line and the J-line be two parallel lines, plot $i_1,\dots,i_k$ on the I-line and $j_1\dots,j_k$ on the J-line. These are the vertices. The graph has $2k$ distinct edges joining the vertices as follows: $i_1,j_1,i_2,j_2,\dots,i_k,j_k,i_1$.

Let d_m denote the number of edges of multiplicity m (meaning the vertices that are connected with exactly m edges). Obviously, for a given multi-graph $\Gamma(I,J)$ we have

$$d_1 + 2d_2 + \cdots + 2kd_{2k} = 2k,$$

since each edge in $\Gamma(I,J)$ has multiplicity in $[2k]$.

Now define

$$A(r,c) = \{(I,J) : \psi(I) = r, \psi(J) = c\ , d_1 = 0, i_1 \neq i_2 \cdots \neq i_k \neq i_1\}.$$

By the above definition we get:

$$\begin{aligned}\langle \mu_{B_p}, x^k\rangle &= \frac{1}{p(2\sqrt{pn})^k}\sum_{r,c=1}^{k}\sum_{A(r,c)}\mathbb{E}(X_{i_1,j_1}X_{i_2,j_1}X_{i_2,j_2}X_{i_2,j_3}\cdots X_{i_k,j_k}X_{i_1,j_k})\\ &= \sum_{r,c=1}^{k} S_{r,c},\end{aligned}$$

where

$$S(r,c) = \frac{1}{p(2\sqrt{pn})^k}\sum_{A(r,c)}\mathbb{E}(X_{i_1,j_1}X_{i_2,j_1}X_{i_2,j_2}X_{i_2,j_3}\cdots X_{i_k,j_k}X_{i_1,j_k}).$$

We will prove that $S_{r,c} \to 0$ as $p \to \infty$ unless if $r = k/2+1$ and $c = k/2$. Note that

$$\begin{aligned}\mathbb{E}|X_{i_1,j_1}X_{i_2,j_1}X_{i_2,j_2}X_{i_2,j_3}\cdots X_{i_k,j_k}X_{i_1,j_k}| &\le |\mathbb{E}X_{1,1}|^{d_1}|\mathbb{E}X_{1,1}^2|^{d_2}\cdots|\mathbb{E}X_{1,1}^{2k}|^{d_{2k}}\\ &\le |1+a_p|^k(\epsilon_p n^{1/4})^{2k-2(d_2+\cdots+d_{2k})}.\end{aligned}$$

Since $r+c$ are the distinct vertices of the graph, we get that $r+c \le d_1 + d_2 + \cdots + d_{2k} + 1$, because $d_1+d_2+\cdots+d_{2k}$ is the cardinality of the distinct edges. Also,

$$|A(r,c)| \le \binom{p}{r} r^k c^k \binom{n}{c} \le p^r n^c r^k c^k.$$

It follows that

$$|S(r,c)| \leq \frac{1}{p(2\sqrt{pn})^k}|1+a_p|^k(\epsilon_p n^{1/4})^{2k-2(r+c-1)}p^r n^c r^k c^k. \tag{2.3.1}$$

We need one more inequality for $S(r,c)$. Let $l_1, l_2, \ldots, l_c$ be the different values of J. Then,

$$\begin{aligned} E := \mathbb{E}\left(X_{i_1,j_1}X_{i_2,j_1}X_{i_2,j_2}X_{i_2,j_3}\cdots X_{i_k,j_k}X_{i_1,j_k}\right) &= \prod_{b=1}^{c}\mathbb{E}\left(\prod_{j_a=l_b}(X_{i_a,j_a}X_{i_{a+1},j_a})\right) \\ &= \prod_{b=1}^{c}\mathbb{E}\left(X_{1,1}^{n_{b1}}\right)\mathbb{E}\left(X_{1,1}^{n_{b2}}\right)\cdots\mathbb{E}\left(X_{1,1}^{n_{bs}}\right), \end{aligned}$$

where $n_{b1}, n_{b2}, \ldots, n_{bs}$ are all ≥ 2 (or else the mean will be zero) and $s \geq 2$ depends on b (meaning how many vertices from I are connected with the b-th element of J). Then,

$$|E| \leq \prod_{b=1}^{c}(\epsilon_p n^{1/4})^{\sum_a n_{ba}-2s}(1+|a_p|)^2 \leq (\epsilon_p n^{1/4})^{2k-4s}(1+|a_p|)^k. \tag{2.3.2}$$

Now, suppose that $r \neq k/2+1$ and $c \neq k/2$. We distinguish three cases:

Case 1: $r > k/2+1$. Then since $c+r \leq k+1$ we have

$$\frac{1}{2}(r+c-1-k) \leq 0.$$

By (2.3.1) and the assumption that $p/n \to 0$, we get that $|S(r,c)| \to 0$.

Case 2: $c > k/2$. Any J-vertex cannot be connected via an edge with only one I-vertex since every two successive vertices are different, so there would be at least $4c$ edges, which is impossible

Case 3: $r < k/2+1, c < k/2$. In this case $S(r,c) \to 0$ by (2.3.2). So we just need to compute the case that $k = 2m$ is even and $r = m+1$, $c = m$. In this case, and since $d_1 = 0$, we have that

$$k+1 = r+c \leq d_2 + d_3 + \cdots + d_{2k} \leq 1/2(2d_2 + 3d_3 + \cdots + 2kd_{2k}) + 1 = k+1,$$

so

$$d_3 = \cdots = d_{2k} = 0.$$

So, each edge appears exactly twice. Define two pairs I, J and W, V to be equivalent if the following holds: two vertices are equal in I, J if and only if the equivalent vertices are equal in W, V. Now by assigning to each edge $+1$ if it appears for the first time and -1 otherwise we get, by the corresponding lemma in the proof of the semicircular law, that the cardinality of the equivalence classes is C_m (the Catalan number). So:

$$\langle \mu_{B_p}, x^{2m} \rangle = S(m, m+1) = \frac{1}{p(2\sqrt{np})^k} C_m (1 + a_p)^k = \mathbb{O}(1).$$

Letting $p \to \infty$ and using the same method as in the proof of the semicircular law we conclude the proof of (i).

The proof of (ii) is similar to the one of Corollary 2.1.13. □

Lemma 2.3.3. *Let X be a real random variable such that $E|X| < \infty$. Then*

$$\sum_{n=1}^{\infty} \mathbb{P}(|X| \geq n) < \infty.$$

Proof. Let $A_n = \{\omega \in \Omega : n \leq |X_n(\omega)| < n+1\}$. Then

$$\sum_{i=1}^{\infty} n \mathbf{1}_{A_n} \leq |X|.$$

So, by integration,

$$\mathbb{E} \sum_{i=1}^{\infty} n \mathbf{1}_{A_n} \leq \mathbb{E}|X|.$$

But, by the Beppo-Levi and Tonelli theorems,

$$\begin{aligned} \mathbb{E} \sum_{i=1}^{\infty} n \mathbf{1}_{A_n} &= \sum_{i=1}^{\infty} n \mathbb{P}(A_n) = \sum_{i=1}^{\infty} \mathbb{P}(A_n) \sum_{k=1}^{n} 1 \\ &= \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \mathbb{P}(A_n) = \sum_{k=1}^{\infty} \mathbb{P}(|X| \geq k). \end{aligned}$$

□

Lemma 2.3.4. *If $\mathbb{E}(X^4) < \infty$ then there exists $\{\epsilon_p\}_{p \in \mathbb{N}}$ such that*

1. *$\epsilon_p \to 0$, but $\epsilon_p p^{1/4} \to \infty$.*

2. $\mathbb{P}(|X| \geq \epsilon_p n^{1/4}) \leq \epsilon_p/n$.

Proof. Since $p/n \to 0$ we have that $p < n$ if n is sufficient large. So, we can assume that $p(n) < n$ for all $n \in N$.

Let $\{a_p\}$ be a decreasing sequence such that $a_p \to 0$ but $a_p p^{1/4} \to \infty$ increasingly. We denote by δ_p the sequence

$$\delta_p = \sqrt[5]{\mathbb{E}(X^4 \mathbf{1}_{|X|>a_p p^{1/4}})}.$$

Note that δ_p is a decreasing sequence which tends to zero by the dominated convergence theorem (for all $p \in \mathbb{N}$ we have that $X^4 \geq X^4 \mathbf{1}_{|X|>a_p p^{1/4}}$). Then, define

$$\epsilon_p := \max\{\delta_p, a_p\}.$$

Since $a_p \leq \epsilon_p$ we get that $\epsilon_p p^{1/4} \to \infty$. Note also that ϵ_p is non-increasing and tends to zero as it is the maximum of two non-increasing sequences which tend to zero. So,

$$\begin{aligned} n\mathbb{P}(|X| > \epsilon_p n^{1/4}) \leq n\mathbb{P}(|X| > \epsilon_n n^{1/4}) &\leq \frac{\mathbb{E}(X^4 1_{|X|>\epsilon_n p^{1/4}})}{\epsilon_n^4} \\ &= \frac{\delta_n^5}{\epsilon_n^4} \leq \delta_n \leq \epsilon_n \leq \epsilon_p. \end{aligned}$$

□

Lemma 2.3.5. *Let $Y_1, Y_2, Y_3, \ldots$ be i.i.d. random variables such that $\mathbb{P}(Y_1 = 1) = q = 1 - \mathbb{P}(Y_1 = 0)$. Then,*

$$\mathbb{P}\left(\sum_{i=1}^{n} Y_i - nq \geq n\epsilon\right) \leq e^{-nh(\epsilon - qh)}$$

for all $\epsilon > 0$, $n \in \mathbb{N}$ and $h \in [0, 1/2]$.

Proof. Let $\epsilon > 0$, $n \in \mathbb{N}$ and $h \in [0, \frac{1}{2}]$. Then, by Markov's inequality and since the Y_i's are i.i.d. we get

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{n} Y_i \geq n(q+\epsilon)\right) = \mathbb{P}\left(e^{\sum_{i=1}^n hY_i} \geq e^{hn(q+\epsilon)}\right) &\leq e^{-hn(q+\epsilon)} \mathbb{E} \prod_{i=1}^{n} e^{hY_i} \\ = e^{-hn(q+\epsilon)} \prod_{i=1}^{n} \mathbb{E} e^{hY_i} &= e^{-hn(q+\epsilon)} (\mathbb{E} e^{hY_1})^n. \end{aligned}$$

But, $\mathbb{E}e^{hY_1} = e^h q + 1 - q = (e^h - 1)q + 1$ and since for all $x \in (0, +\infty]$ we have $x + 1 \leq e^x$ and $e^h - 1 < (h+1)h$, we get that

$$\mathbb{E}\left(e^{hY_1}\right) \leq e^{(e^h-1)q} \leq e^{(h+1)hq}.$$

So,

$$\mathbb{P}\left(\sum_{i=1}^{n} Y_i \geq n(q+\epsilon)\right) \leq e^{-hn(q+\epsilon)} e^{(h+1)hnq},$$

which proves the lemma. □

Remark 2.3.6. Note that if F, G are the mass functions (with respect to the Lebesque measure) of two empirical spectral distributions of size n then

$$\int |F(x) - G(x)|\, dx = \frac{1}{n} \sum_{i=1}^{n} |\lambda_i - \mu_i|,$$

where λ_i are the eigenvalues of the matrix which corresponds to F in increasing order and similarly for μ_i and G.

Lemma 2.3.7. *Let* $\{(a_i, b_i), i \in \mathbb{N}\}$ *be the set of all intervals with rational endpoints and length less than* 1. *Let*

$$f_i(x) = \int_{-\infty}^{x} \mathbf{1}_{(a_i,b_i)}(t)\, dt$$

and

$$D(F, G) = \sum_i \frac{1}{2^i} \left| \int f_i d(F(x) - G(x)) \right|,$$

where F, G *are empirical spectral distributions. Then* $D(F_n, F) \to 0$ *implies that* $F_n \to F$ *weakly.*

Proof. In order to prove the lemma we are going to use the characterization of weak convergence from Lemma 1.3.5 which says that a sequence μ_n of Borel probability measures on a metric space (X, d) converges weakly to a probability measure μ if and only if, for all open subsets U of X,

$$\liminf_{n\to\infty} \mu_n(U) \geq \mu(U).$$

Let μ be the probability measure on $\mathbb{R}$ with distribution F, and likewise μ_n for F_n. It is easy to compute, using Fubini's theorem, that for every $i \in \mathbb{N}$,

$$\left| \int f_i d(F(x) - G(x)) \right| = |(\mu_n - \mu)((a_i, b_i))|,$$

which tends to zero as n tends to infinity. Also, for all $x \in \mathbb{R}$. Given an open subset U of $\mathbb{R}$, we write it as a infinite (countable) union of disjoint intervals with rational endpoints and length less than 1 plus a countable set. Let $U = \cup_{i=1}^{\infty} U_i$ be these intervals. Since $D(F_n, F) \to 0$ we get that $\mu_n(U_i) \to \mu(U_i)$. So, if $i, n \in \mathbb{N}$ then

$$\mu_n(U) \geq \sum_{k=1}^{i} \mu_n(U_k)$$

As a result, since this is true for every n,

$$\liminf_{n\to\infty} \mu_n(U) \geq \sum_{k=1}^{i} \liminf \mu_n(U_k) = \sum_{k=1}^{i} \mu(U)$$

and since this is true for every i we get

$$\liminf_{n\to\infty} \mu_n(U) \geq \sum_{k=1}^{\infty} \mu(U_k) = \mu(U).$$

□

Lemma 2.3.8. *Let A,B be two $p \times p$ symmetric matrices with eigenvalues $\{\lambda_1 \leq \lambda_2 \cdots \leq \lambda_p\}$ and $\{\mu_1 \leq \mu_2 \cdots \leq \mu_p\}$ respectively. Then,*

$$\sum_{i=1}^{p} (\lambda_i - \mu_i)^2 \leq tr(A - B)^2.$$

Proof. We begin by diagonalizing A and B. Since they are symmetric, there exist orthogonal matrices U, V and diagonal matrices Λ and M, which have as entries the eigenvalues of A and B respectively (in the order we have mentioned), such that

$$A = U\Lambda U^*$$

and

$$B = VMV^{T*}.$$

So,

$$\mathrm{tr}(AB) = \mathrm{tr}(U\Lambda U^* VMV^*) = \mathrm{tr}((VU^*)\Lambda(V^*U)M).$$

Setting $W = V^*U$ and after some calculations we get:

$$\mathrm{tr}(AB) = \sum_{1\le i,j\le p} \lambda_i \mu_j W_{i,j}^2.$$

In order to proceed we need some definitions.

Definition 2.3.9. An $n \times n$ matrix will be called *doubly stochastic* if all its entries are non-negative and the sum of the elements of each row and each column is equal to 1.

Definition 2.3.10. Let A be a $p \times p$ matrix. Then A will be called a *permutation matrix* if there exists $p \in S_n$ such that $A_{i,j} = 1 \iff p(i) = j$ and $A_{i,j} = 0 \iff p_i \neq j$.

Note: Since every $p \in S_n$ is a bijection from $[n]$ to $[n]$, every permutation matrix is doubly stochastic.

Now, since W is orthogonal we get $\sum_{i=1}^p W_{i,j}^2 = 1$ for each j and $\sum_{j=1}^p W_{i,j}^2 = 1$ for each i. So, setting $u_{i,j} = W_{i,j}^2$ we have that $\{u_{i,j}\}_{i,j=1}^p$ is doubly stochastic. Let D_p denote the set of all doubly stochastic $p \times p$ matrices. Then,

$$\mathrm{tr}(AB) = \sum_{1\le i,j\le p} \lambda_i \mu_j u_{i,j} \le \sup_{(a_{i,j})\in D_p} \sum_{1\le i,j\le p} \lambda_i \mu_j a_{i,j}$$

Note that D_p is convex (any convex combination of doubly stochastic $p \times p$ matrices will remain doubly stochastic) and the function

$$\{a_{i,j}\}_{i,j=1}^p \mapsto \sum_{1\le i,j\le p} \lambda_i \mu_j a_{i,j}$$

is linear. Therefore, the supremum on D_p is achieved at an extreme point of D_p.

We will prove that the extreme points of D_p are the permutation matrices (Birkhoff Theorem).

The proof that every permutation matrix is an extreme point is elementary since, if A, B are doubly stochastic matrices, P is a permutation matrix and $r \in (0, 1)$ such that

$$rA + (1 - r)B = P,$$

then for each $i,j \in [p]$ such that $p_{i,j} = 0$ we have that $ra_{i,j} = -(1-r)b_{i,j}$ and since every entry of a doubly stochastic matrix is non-negative we get

$$a_{i,j} = b_{i,j} = 0.$$

So, since the sum of the elements of each row and column of A and B must be equal to 1, we get that $A = B = P$.

In order to prove the opposite direction, we will prove that every doubly stochastic matrix is a convex combination of permutation matrices. To prove this, we will need another definition and a very important lemma:

Definition 2.3.11. For every doubly stochastic $p \times p$ matrix A we define its associated graph G with

$$V(G) = \{i_k : k \in [p]\} \cup \{j_k : k \in [p]\}$$

and

$$E(G) = \{(i_k, j_m) : A_{k,m} > 0\}.$$

Note: The graph of every doubly stochastic matrix is bipartite.

The idea behind this definition is that for a doubly stochastic matrix we create its graph by turning each row and each column into a vertex and we connect a row (call it i) and a column (call it j) via an edge if the element on the spot (i,j) is not zero.

Lemma 2.3.12. *The graph of every doubly stochastic matrix has a perfect matching.*

Proof. Assume, by way of contradiction, that there exists a doubly stochastic matrix such that its graph does not have a perfect matching. Call this matrix A, and call $R(A)$ the first part (the rows) and $C(A)$ the other part (the columns). By Hall's theorem, without loss of generality, we get that there exists $B \subseteq V(R(A))$ such that $N(B) < |B|$. Now, we see that

$$\sum_{i \in B, j \in N(B)} A_{i,j} = |B|.$$

This is true since for any vertex (column) in B every row connected to it belongs to $N(B)$, and since the matrix is doubly stochastic we get that the

sum above gives the cardinality of B (a term 1 for each vertex in B). But the fact that A is doubly stochastic gives that the same sum equals to the cardinality of $N(B)$ which is a contradiction. □

So in order to prove the opposite direction of Birkhoff's theorem, we will use induction on the number of non-zero entries of a matrix. For $k = 2$ the opposite direction is true.

Let the opposite direction be true for all $m < k$ and let A be a matrix with k non-zero elements. By the previous lemma, the associated graph of A has a perfect matching. Underline the entries associated to the edges of the perfect matching. Since the edges in the perfect matching are disjoint we get that there is exactly one element in each row and column. Let P be the permutation matrix with entries 1 exactly at the positions of the underlined elements. Let c be the minimum of those entries. If $c = 1$ then $A = P$. If not, then the matrix $M = A - cP/(1-c)$ is doubly stochastic with one non-zero entry less than the ones for A. So, since $A = (1-c)M + cP$, and by the induction hypothesis, the proof is complete.

Now we can return to the proof of Lemma 2.3.8. By what was done above we get that

$$\mathrm{tr}(AB) \leq \max_{\sigma \in S_p} \sum_{1 \leq i \leq p} \lambda_i \mu_{\sigma(i)}.$$

Using the fact that μ_i and λ_i are in non-decreasing order we will prove that the maximum is achieved by the identity permutation. If not, then for $i > j$

$$\lambda_i \mu_i + \lambda_j \mu_j - \lambda_j \mu_i - \lambda_i \mu_j = (\lambda_j - \lambda_i)(\mu_i - \mu_j) \geq 0.$$

Let $\sigma \in S_p$ be a permutation different than the identity. So, there exist $i, j \in [p]$ such that $j < i$ and $\sigma(j) < \sigma(i)$. Let σ' be a permutation with the property that $\sigma'(i) = i, \sigma'(j) = j$ and for all $d \in [p] \setminus \{i, j\}$ we have $\sigma'(d) = \sigma(d)$. The permutation σ has one more order reversal than σ'. Iterating this process we see that the sum is maximized for the permutation id. So,

$$\mathrm{tr}(AB) \leq \sum_{i=1}^{p} \lambda_i \mu_i$$

Finally since $\operatorname{tr} A^2 = \sum_{i=1}^{p} \lambda_i^2$ and $\operatorname{tr} B^2 = \sum_{i=1}^{p} \mu_i^2$ we get:

$$\begin{aligned}\sum_{i=1}^{p} (\lambda_i - \mu_i)^2 &= \sum_{i=1}^{p} \lambda_i^2 + \mu_i^2 - 2\lambda_i\mu_i \\ &\leq \sum_{i=1}^{p} \lambda_i^2 + \sum_{i=1}^{2} \mu_i^2 - 2\operatorname{tr} AB \\ &= \operatorname{tr} A^2 + \operatorname{tr} B^2 - 2\operatorname{tr} AB = \operatorname{tr}(A-B)^2.\end{aligned}$$

□

Definition 2.3.13. Given a matrix $A \in M_n$, the matrix obtained after deleting the i-th row and the i-th column of A for some $i \in [n]$ is called principal sub-matrix of A.

Definition 2.3.14. Let A be symmetric or Hermitian matrix in $M_n(\mathbb{R}^n)$ or $M_n(\mathbb{C}^n)$ respectively. Consider the standard inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^n$ or $\mathbb{C}^n$. The *Rayleigh-Rietz quotient* is the function

$$R_A(x) = \frac{\langle Ax, x \rangle}{\langle x, x \rangle}$$

defined on all the non-zero elements of $\mathbb{R}^n$ or $\mathbb{C}^n$.

Lemma 2.3.15 (min-max theorem)**.** *Let A be a symmetric matrix $A \in M_n$ and let $\{\lambda_i\}_{i=1}^n$ be the eigenvalues of A in non-decreasing order. For every $k \in [n]$ let*

$$\mathcal{A}_k = \{U \subseteq \mathbb{R}^n : dim(U) = k\}.$$

Then,

$$\lambda_k = \min_{U \in \mathcal{A}_k} \max_{x \in U \setminus \{0\}} R_A(x)$$

and

$$\lambda_k = \max_{U \in \mathcal{A}_{n-k+1}} \min_{x \in U \setminus \{0\}} R_A(x).$$

Proof. Since A is symmetric it is diagonalizable and we can chose an orthogonal basis of eigenvectors $\{u_1, u_2, \ldots, u_n\}$, where u_i is the eigenvector corresponding to λ_i for each $i \in [n]$. If U is a subspace of dimension k then its intersection with $\operatorname{span}\{u_k, u_{k+1}, \ldots, u_n\}$ is non-empty. Let

$u \in \operatorname{span}\{u_k, u_{k+1}, \ldots, u_n\} \cap U \setminus \{0\}$. Then we can write u as

$$u = \sum_{i=k}^{n} a_i u_i,$$

and its Rayleigh quotient is

$$R_A(u) = \frac{\sum_{i=k}^{n} a_i^2 \lambda_i}{\sum_{i=k}^{n} a_i^2} \geq \lambda_k \frac{\sum_{i=k}^{n} a_i^2}{\sum_{i=k}^{n} a_i^2} = \lambda_k.$$

Since this is true for every subspace U we get:

$$\lambda_k \leq \min_{U \in \mathcal{A}_k} \max_{x \in U \setminus \{0\}} R_A(x).$$

For the other direction note that for the subspace $V = \operatorname{span}\{u_1, u_2, \ldots, u_k\}$ and for every $u \in V$ we have

$$R_A(u) \leq \lambda_k,$$

since λ_k is the largest eigenvalue for U. So,

$$\lambda_k = \min_{U \in \mathcal{A}_k} \max_{x \in U \setminus \{0\}} R_A(x) = \max_{v \in V \setminus \{0\}} R_A(v).$$

The proof of the other equality is similar. In the case where U is a subspace of dimension $n - k + 1$, we proceed in a similar fashion: Consider the k-dimensional subspace $\operatorname{span}\{u_1, \ldots, u_k\}$. Its intersection with U is not $\{0\}$ (by simply checking dimensions) and hence there exists a non-zero vector v in this intersection, which we can write as

$$v = \sum_{i=1}^{k} a_i u_i.$$

So,

$$R_A(v) = \frac{\sum_{i=1}^{k} a_i^2 \lambda_i}{\sum_{i=1}^{k} a_i^2} \leq \lambda_k,$$

and since this is true for all U we have the first inequality.

To get the other inequality, note again that every eigenvector u of λ_k is contained in $V = \operatorname{span}\{u_k, \ldots, u_n\}$ so that we can conclude the equality. Also, as before, we get

$$\lambda_k = \max_{U \in \mathcal{A}_{n-k+1}} \min_{x \in U \setminus \{0\}} R_A(x) = \min_{x \in V \setminus \{0\}} R_A(v).$$

□

Lemma 2.3.16 (Cauchy interlacing theorem)**.** *Let A be a symmetric (or Hermitian in $\mathbb{C}$) matrix in M_n. Let B be a principal sub-matrix of A and let $\{\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n\}$ be the eigenvalues of A and $\{\mu_1 \leq \mu_2 \leq \cdots \leq \mu_{n-1}\}$ be the eigenvalues of B, both in non-decreasing order. Then:*

$$\lambda_1 \leq \mu_1 \leq \lambda_2 \leq \mu_2 \leq \lambda_3 \cdots \leq \lambda_{n-1} \leq \mu_{n-1} \leq \lambda_n.$$

Proof. Assume without loss of generality that we have deleted the n-th row, and so, let

$$A = \begin{bmatrix} B & x \\ x & z \end{bmatrix}.$$

Let $\{x_1, x_2, \ldots, x_n\}$ be the eigenvectors of A and let $\{y_1, \ldots, y_{n-1}\}$ be the eigenvectors of B. We define the following vector spaces:

$$V = \operatorname{span}\{x_k, \ldots, x_n\}$$
$$W = \operatorname{span}\{y_1, \ldots, y_k\}$$
$$W' = \left\{ \begin{pmatrix} w \\ 0 \end{pmatrix} : w \in W \right\}.$$

Since $\dim(V) = n - k + 1$ and $\dim(W') = \dim(W) = k$ we see that the intersection of W' and V is non-trivial, meaning that there exists $u \in W' \cap V \setminus \{0\}$. So,

$$u = \begin{pmatrix} w \\ 0 \end{pmatrix}$$

for some $w \in W$. Then

$$u^* A u = \begin{bmatrix} w^* & 0 \end{bmatrix} \begin{bmatrix} B & x \\ x & z \end{bmatrix} \begin{bmatrix} w \\ 0 \end{bmatrix} = w^* B w$$

But from the min-max theorem we get

$$\lambda_k = \min_{v \in V} R_A(v)$$

and

$$\mu_k = \max_{d \in W} R_B(d).$$

So, $\lambda_k \leq \mu_k$.

The proof of the other inequality is similar. We now define the vector spaces

$$V = \text{span}\{x_1, \ldots, x_{k+1}\}$$
$$W = \text{span}\{y_k, \ldots, y_{n-1}\}$$
$$W' = \left\{ \begin{pmatrix} w \\ 0 \end{pmatrix} : w \in W \right\}.$$

Since $\dim(V) = k+1$ and $\dim(W) = \dim(W') = n-k$, there exists $u \in W' \cap V \setminus \{0\}$. So, as before,

$$u = \begin{pmatrix} w \\ 0 \end{pmatrix}$$

for some $w \in W$. Then we have $u^* A u = w^* B w$. So, again from the min-max theorem,

$$\lambda_{k+1} = \max_{v \in V} R_A(v) \geq R_A(u) = R_B(w) \geq \min_{d \in W} R_B(d) = \mu_k.$$

□

Note: In the previous theorem we can replace $\mathbb{R}^n$ with $\mathbb{C}^n$ and the symmetric matrix with a Hermitian matrix.

We are now ready to prove the main theorem of this chapter.

Proof. By Lemma 2.3.4 we can choose ϵ_p such that $\epsilon_p \to 0$ and $\epsilon_p p^{1/4} \to \infty$ such that $\mathbb{P}(|X_{1,1}| \geq \epsilon_p n^{1/4}) \leq \epsilon_p / n$. Define

$$X'_p = [X'_{i,j} : i = 1, 2 \ldots, p; j = 1, 2 \ldots, n]$$

where

$$X'_{i,j} = X_{i,j} \mathbf{1}_{|X_{i,j}| < \epsilon_p n^{1/4}}$$

Let

$$A'_p = \frac{1}{2\sqrt{np}} (X'_p X'^*_p - n I_p)$$

and

$$h_{i,j} = \mathbf{1}_{|X_{i,j}| \geq \epsilon_p n^{1/4}}.$$

Proposition 2.3.17. *The following inequality holds true:*

$$\sup_x |F^{A_p}(x) - F^{A'_p}(x)| \leq \frac{1}{p} \sum_{i=1}^{p} \sum_{j=1}^{n} h_{i,j}.$$

Proof. Firstly note that

$$\sup_x |F^{A_p}(x) - F^{A'_p}(x)| = \sup_x |F^{X_p X_p^*}(x) - F^{X'_p X'^*_p}(x)|.$$

Since

$$\det(A_p - xI_p) = \frac{1}{(2\sqrt{np})^p} \det(X_p X_p^* - (n + 2x\sqrt{pn})I_p),$$

substituting $u := n + x2\sqrt{pn}$ we see that the roots with respect to u are the eigenvalues of $X_p X'_p$. So, solving for x we get that if λ_i is the i-th (in non-decreasing order) eigenvalue of A_p and d_i is the i-th eigenvalue of $X_p X_p^*$ then we have

$$\lambda_i = n + d_i 2x\sqrt{pn}.$$

Since the same is true for A'_p and since we are interested in the supremum over all x, we see that we can investigate the eigenvalues of XX^* and $X'X'^*$ instead.

Let $x \in \mathbb{R}$ and let

$$L_p = \{(i,j) \in [p] \times [n] : X_{i,j} \neq X'_{i,j}\}.$$

We will prove that, for every $p \in \mathbb{N}$, if $g \leq p$ and $X_p X_p^*$ is the matrix defined above for $p, n(p)$, and

$$g = p|F^{X_p X_p^*}(x) - F^{X'_p X'^*_p}(x)|,$$

then

$$g \leq |L_p|,$$

using induction on p.

For $p = 1$ the assertion is true; if not, we would have $X_1 = X'_1$ i.e. the two matrices would be equal but with different eigenvalues (the root of the polynomial $f = X_1 - 1 \cdot \lambda$) which would be a contradiction.

Suppose that for all $m \leq p-1$ the statement is true. Consider the matrices $X_pX_p^*$ and $X'_pX'^*_p$ as before. If

$$|F^{X_pX_p^*}(x) - F^{X'_pX'^*_p}(x)| = 0$$

then obviously

$$|F^{X_pX_p^*}(x) - F^{X'_pX'^*_p}(x)| \leq |L_p|.$$

If not, then there exist i,j such that $X_{i,j} \neq X'_{i,j}$. Without loss of generality let $i = p$. Let B be the principal sub-matrix of $X_pX_p^*$ deleting the p-th row and column, and similarly let B' be the corresponding sub-matrix for $X'_pX'_{p-1}$. Note that $B = X_{p-1}X_{p-1}^*$ and $B' = X'_{p-1}X'^*_{p-1}$. So, we can apply the induction hypothesis for B and B', which gives that

$$p|F^B(x) - F^B(x)| \leq |L_{p-1}|.$$

But, by Cauchy's interlacing theorem, we get that

$$p|F^{X_pX_p^*}(x) - F^{X'_pX'^*_p}(x)| \leq p|F^{X_{p-1}X_{p-1}^*}(x) - F^{X'_{p-1}X'^*_{p-1}}(x)| + 1.$$

Also, since there exists an element on the p-th row such that $X_{p,j} \neq X'_{p,j}$ for some j, we get that

$$|L_{p-1}| + 1 \leq |L_p|.$$

It follows that

$$p|F^{X_pX_p^*}(x) - F^{X'_pX'^*_p}(x)| \leq p|F^{X_{p-1}X_{p-1}^*}(x) - F^{X'_{p-1}X'^*_{p-1}}(x)| + 1 \leq |L_{p-1}| + 1 \leq |L_p|.$$

The inequality above is true for all $x \in \mathbb{R}$ and for all $p \in \mathbb{N}$. Also note that if $h_{i,j}$ denotes the event that the (i,j)-th element of X_p is different from the element in the same spot of X'_p, we get that

$$|L_p| \leq \sum_{i=1}^{p} \sum_{j=1}^{n} h_{i,j}.$$

So, the proof of the proposition is complete. □

Since $\mathbb{P}(h_{i,j} = 1) = \mathbb{P}(|X_{i,j}| \leq \epsilon_p n^{1/4}) = q_p$ (say), if $\delta > 0$, by what was done before we get

$$\begin{aligned}
\mathbb{P}\left(\sup_x |F^{A_p} - F_{A'_p}| \geq \delta\right) &\leq \mathbb{P}\left(\frac{1}{p}\sum_i \sum_j h_{i,j} \geq \delta\right) \\
&= \mathbb{P}\left(\sum_i \sum_j h_{i,j} - pnq_p \geq pn\frac{\delta}{n} - q_p\right) \\
&\leq \exp\left(-nph\left(\frac{\delta}{n} - q_p - q_p h\right)\right) \\
&\leq \exp\left(-nqh\left(\frac{\delta}{n} - (1+h)\frac{\epsilon_p}{n}\right)\right) \\
&\leq \exp\left(-p\frac{\delta h}{2}\right)
\end{aligned}$$

for $\epsilon_p < \delta/3$. We can choose $h = 1/2$. Thus, by the Borel-Cantelli lemma,

$$\sup_x |F^{A_p}(x) - F^{A'_p}(x)| \to 0 \qquad a.s.$$

Note that we can replace X'_p with $Y_p = X'_p - \mathbb{E}(X_p)$ and we would have the same result (see [17] p.81 and A-46)). But by Lemma 2.3.2 we know that

$$\sup_x |F^{A_p} - F_{B_p}| \to 0 \qquad a.s.$$

where B_p is defined as in Lemma 2.3.2 starting from the matrix Y_p defined above. Thus in order to prove the theorem it is sufficient to prove that

$$D(F^{A''_p}, F^{B_p}) = \sum_i \frac{1}{2^i}\left|\int f_i d(F^{A''_p}(x) - F^{B_p}(x))\right| \to 0$$

as in Lemma 2.3.7. Here, $A''_p = \frac{1}{2\sqrt{pn}}(Y_p Y_p^* - nI)$, also $X''_{i,j} = X'_{i,j} - \mathbb{E}(X'_{i,j})$, and λ_i are the eigenvalues of A''_p and μ_i the eigenvalues of B_p. So by integration by parts, Remark 2.3.6 and Lemma 2.3.8 we get that

$$\begin{aligned}
D^2(F^{A''_p}, F^{B_p}) \leq \left(\frac{1}{p}\sum_{i=1}^p |\lambda_i - \mu_i|\right)^2 &\leq \frac{1}{p}\sum_{i=1}^p (\lambda_i - \mu_i)^2 \leq \frac{1}{p}\operatorname{tr}(A''_p - B_p)^2 \\
&= \frac{1}{4np^2}\sum_{i=1}^p \left[\sum_{m=1}^n (X''_{i,m} - 1)\right]^2 \\
&\leq \frac{1}{2np^2}\sum_{i=1}^p \left[\sum_{m=1}^n (X''^2_{i,m} - \mathbb{E}(X''^2_{i,m})\right]^2 + \frac{n}{2p}(1 - \mathbb{E}X''^2_{i,m})^2.
\end{aligned}$$

For sufficiently large p we have:

$$\frac{n}{2p}(1-\mathbb{E}X''^2) \le \frac{n}{2p}\frac{4}{n\epsilon_p^4}\mathbb{E}^2X_{1,1}^4 \to 0.$$

So, for the first term on the right hand side of the inequality we have:

$$\begin{aligned}
&\frac{1}{4np^2}\sum_{i=1}^{p}\left[\sum_{m=1}^{n}(X''_{i,m}-1)\right]^2 \\
&\le \frac{1}{2np^2}\sum_{i=1}^{p}\left[\sum_{m=1}^{n}(X''^2_{i,m}-\mathbb{E}(X''^2_{i,m}))\right]^2 \\
&\le \frac{1}{2p^2n}\sum_{i=1}^{p}\sum_{m=1}^{n}[X''^2_{im}-\mathbb{E}(X''^2_{i,m})]^2 \\
&+ \frac{1}{2np^2}\sum_{i=1}^{p}\sum_{m_1\neq m_2}(X''^2_{i,m_1}-\mathbb{E}X''^2_{i,m_1})(X''^2_{i,m_2}-\mathbb{E}X''^2_{i,m_2}) \\
&= S_{1,p}+S_{2,p}.
\end{aligned}$$

Then, for any $\epsilon > 0$, by Markov's inequality we have:

$$\begin{aligned}
\sum_{p=1}^{\infty}\mathbb{P}(|S_{2,p}|>\epsilon) &= \sum_{p=1}^{\infty}\mathbb{P}(S_{2,p}^2>\epsilon^2) \\
&\le \frac{1}{\epsilon^2}\sum_{p=1}^{\infty}\mathbb{E}(S_{2,p}^2) \\
&= \sum_{p=1}^{\infty}\frac{1}{2np^2}\sum_{m=1}^{p}2n(n-1)\mathbb{E}^2(X''^2_{1,1}-\mathbb{E}X''^2_{1,1})^2 \\
&\le \sum_{p=1}^{\infty}\frac{1}{2p^3}\mathbb{E}^2X_{1,1}^4 < \infty.
\end{aligned}$$

Thus, by the Borel-Cantelli lemma we get

$$S_{2,p} \to 0 \qquad a.s.$$

For $S_{1,p}$ we have

$$\begin{aligned} S_{1,p} &= \frac{1}{2p^2n}\sum_{i=1}^{p}\sum_{m=1}^{n}[X''^2_{im} - \mathbb{E}(X''^2_{i,m})]^2 \\ &\leq \frac{1}{2p^2n}\sum_{i=1}^{p}\sum_{m=1}^{n} X''^4_{im} + \mathbb{E}^2(X''^2_{i,m}) \\ &\leq \frac{1}{2p^2n}\sum_{i=1}^{p}\sum_{m=1}^{n} X''^4_{im} - \mathbb{E}(X''^4_{i,m})] + \frac{K}{p}\mathbb{E}X^4_{1,1} = \Delta_p + \frac{K}{p}\mathbb{E}X^4_{1,1} \end{aligned}$$

for some $K > 0$. So, as before, we have

$$\begin{aligned} \sum_{p=1}^{\infty}\mathbb{E}\Delta_p^2 &= \sum_{p=1}^{\infty}\frac{1}{4p^4n^4}\sum_{i=1}^{p}\sum_{m=1}^{n}\mathbb{E}(X''^4_{im} - \mathbb{E}(X''^4_{i,m}))^2 \\ &\leq \sum_{p=1}^{\infty}\frac{K^2}{4n^2p^4}(n^{\frac{1}{4}}\epsilon_p)\mathbb{E}X^4_{1,1} < \infty. \end{aligned}$$

This completes the proof. □

2.3.2 Convergence of the extreme eigenvalues

In this subsection we are going to prove the convergence of the extreme eigenvalues of A_p to -2 and 2 respectively (using the same notation as in the previous subsection) under the assumption that all the entries of $X_p \in M_{p\times n}$ follow the standard normal distribution, and $n(p) \to \infty$ and $p/n \to 0$. By the convergence to the semicircular law we have that $\mu_{A_p} \to \sigma$ (here, again, we use the same notation as in the previous subsection). We will use similar arguments as in the previous similar cases. So we need the following:

Proposition 2.3.18. *Let B be a random $p \times n$ matrix whose entries are independent random variables with distribution $N_{\mathbb{C}}(0,1)$. Then, for any $t > 0$,*

$$\mathbb{P}(\|B\|_{\text{op}} \geq \sqrt{n} + \sqrt{p} + t) \leq 2e^{-t^2/2}.$$

Also, if $n > p$ then for any $t > \frac{4\sqrt{2\ln p}}{\sqrt{\frac{n}{p}}-1}$ it is true that

$$\mathbb{P}(s_n(B) \leq \sqrt{n} - \sqrt{p} - t) \leq 2e^{\frac{-t^2}{4}},$$

where $s_n(B)$ denotes the smaller singular value of B

Suppose that the previous proposition is proven. Then we get

$$\begin{aligned}
&\mathbb{P}(\|B\|_{\text{op}} \geq \sqrt{n} + \sqrt{p} + t) \\
&\quad = \mathbb{P}(\|B\|_{\text{op}}^2 - n \geq t^2 + p + 2t\sqrt{n} + 2\sqrt{np} + 2t\sqrt{p}) \\
&\quad = \mathbb{P}\left((\|B\|_{\text{op}}^2 - n)\frac{1}{\sqrt{np}} \geq \frac{t^2}{\sqrt{np}} + \sqrt{\frac{p}{n}} + \frac{2t}{\sqrt{p}} + 2 + \frac{2t}{\sqrt{n}}\right).
\end{aligned}$$

Letting $(p, n) \to \infty$ we get that

$$\lim_{(p,n)\to\infty} f(n, s) = 0$$

where

$$f(n, s) = \frac{t^2}{\sqrt{np}} + \sqrt{\frac{p}{n}} + \frac{2t}{\sqrt{p}} + \frac{2t}{\sqrt{n}},$$

and since this is true for any $t > 0$ we get

$$\mathbb{P}(\limsup \|A_{n,s}\|_{\text{op}} \geq 2) = 0.$$

Then, since the E.S.D of $A_{n,s}$ tends to the semicircular law and the operator norm is the largest eigenvalue of a matrix we get that

$$\mathbb{P}(\liminf \|A_{n,p}\|_{\text{op}} \leq 2) = 0.$$

Likewise one can show that the second part of the proposition implies the convergence of the smallest eigenvalue of $A_{n,p}$. So it is sufficient to prove the proposition.

Proof of Proposition 2.3.18. In order to prove the first part of the proposition we need the following lemmas

Lemma 2.3.19. *It is true that*

$$\mathbb{E}\|B\|_{\text{op}} \leq \sqrt{n} + \sqrt{p}.$$

Proof. Using the same method as the one we used in the corresponding subsection for the Marchenko-Pastur law we get that for any $p \leq n$ and any $p \times n$ random matrix A with independent entries which all follow $N(0, 1)$

there exist an orthogonal $p \times p$ matrix U and an orthogonal $n \times n$ matrix V such that

$$A = URV,$$

where R is a $p \times n$ matrix whose entries $r_{i,j}$ have the following properties:

- The entries of R are independent random variables.
- For all $i \in [p]$ we have that $r_{i,i} \sim x(n - i + 1)$.
- For all $i \in \{2, 3, \ldots, p\}$ we have that $r_{i,i-1} \sim x(p + 1 - i)$.
- All the other entries are almost surely zero.

Here, for any $m \in \mathbb{N}$ the notation $x(m)$ is used for a positive random variable such that $\chi^2(m)$ is a chi-squared random variable with m degrees of freedom.

Likewise one can show that for any $p \leq n$ and any $p \times n$ random matrix B such that all the entries of B are independent random variables which all follow $N_{\mathbb{C}}(0, 1)$ there exist an orthogonal $p \times p$ matrix U and an $n \times n$ orthogonal matrix V such that

$$B = \sqrt{2}USV,$$

where the entries $s_{i,j}$ of S are independent random variables such that:

- For all $i \in [p]$ we have that $s_{i,i} \sim x(2n + 2 - 2i)$.
- For all $i \in \{2, 3, \ldots, p\}$ we have that $s_{i,i-1} \sim x(2p + 2 - 2i)$.
- All the other entries are almost surely zero.

So, we can consider both A (a $2p \times 2n$ random matrix with i.i.d. standard normal entries) and B (a $p \times n$ random matrix with i.i.d. standard complex normal entries) defined on a common probability space in such a way that $s_{i,j} \leq r_{i,j}$ almost surely (where $s_{i,j}$ and $r_{i,j}$ are the entries of the matrices S and R respectively, as defined before). This can be done since $2n + 2 - 2i \leq 2n+1-i$ and $2p+2-2i \leq 2p+1-i$ so we can couple the matrices in a common

probability space in such a way that $r_{i,j}$ is the square root of the sum of the same standard normal random variables as $s_{i,j}$ plus $i-1$ additional standard normal random variables. Hence, since the eigenvalues of A and B are also the eigenvalues of R and $\sqrt{2}S$ respectively, and since both R and S have non-negative entries (so they are both positive), we get that

$$\|B\|_{\text{op}} \leq \frac{1}{\sqrt{2}}\|A\|_{\text{op}} \quad \Longrightarrow \quad \mathbb{E}\|B\|_{\text{op}} \leq \frac{1}{\sqrt{2}}\mathbb{E}\|A\|_{\text{op}}.$$

So it is sufficient to prove that for any $p \times n$ random matrix A with independent $N(0,1)$ entries it is true that

$$\mathbb{E}\|A\|_{\text{op}} \leq \sqrt{n} + \sqrt{p}.$$

In order to do this, we need some definitions and lemmas.

Definition 2.3.20. A Gaussian process $X = (X_t)_{t\in T}$ is simply a family of jointly Gaussian random variables, usually with mean zero, defined on some probability space Ω, which may or may not be specified.

For more details see [18].

Lemma 2.3.21 (Slepian's inequality)**.** *Let $X = (X_t)_{t\in T}$ and $Y = (Y_t)_{t\in T}$ be Gaussian processes such that for any $t, s \in T$*

$$\mathbb{E}(Y_t - Y_s)^2 \leq \mathbb{E}(X_t - X_s)^2$$

and

$$\mathbb{E}Y_t^2 = \mathbb{E}X_t^2.$$

Then for any $x \in \mathbb{R}$

$$\mathbb{P}\left(\sup_{t\in T} X_t \geq x\right) \leq \mathbb{P}\left(\sup_{t\in T} Y_t \geq x\right).$$

Consequently, by stochastic dominance we have

$$\mathbb{E}\left(\sup_t X_t\right) \leq \mathbb{E}\left(\sup_t Y_t\right).$$

Note: To avoid measurability issues, we study random processes through their finite sub-processes meaning that we interpret $\mathbb{E}\sup_t X_t$ as

$$\sup_{\{T_0 \subseteq T : |T_0| < \infty\}} \mathbb{E} \max_{t \in T_0} X_t.$$

Proof. We shall assume that $|T| < \infty$ and then we can generalise the result (since the supremum of a quantity depending on an infinite set is the supremum of the same quantity over all finite subsets).

Suppose that $|T| = n$. Then X and Y are both Gaussian random vectors in $\mathbb{R}^n$. We may also assume that X and Y are independent (by constructing the analogous product space). So we define the Gaussian random vector $Z(u)$ in $\mathbb{R}^n$ that continuously interpolates between $Z(0) = Y$ and $Z(1) = X$:

$$Z(u) := \sqrt{u}X + \sqrt{1-u}Y, \quad u \in [0,1].$$

Fix $d \in \mathbb{R}$. We need to show that the function $\mathbb{E}f(Z(u))$ (where $f : \mathbb{R}^n \to \mathbb{R}$ is defined by $f(x) = \mathbf{1}_{\max_{i\in[n]} x_i \leq d}(x)$) is increasing in u, which will give us

$$\mathbb{E}f(Z(0)) \leq \mathbb{E}f(Z(1)),$$

which implies the desired inequality.

Firstly we will show that:

Lemma 2.3.22. *Let $X \sim N(0,1)$ and $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function. Then,*

$$\mathbb{E}f'(X) = \mathbb{E}\,(Xf(X)).$$

Proof. Let $p(x)$ be the density function of X. Then, by integration by parts, we have

$$\mathbb{E}f'(X) = \int_{\mathbb{R}} f'(x)p(x)dx = -\int_{\mathbb{R}} f(x)p'(x)dx.$$

Combining this equality with the fact $p'(x) = -p(x)x$ we get the lemma. □

Note: For a random variable $X \sim N(0,\sigma)$ it is true that

$$\mathbb{E}(f'(X)) = \sigma^2 \mathbb{E}\,(f(X)X),$$

since $X = \sigma Z$ where $Z \sim N(0,1)$. By what was done before we have the following generalisation.

Lemma 2.3.23. *Let $X \sim N(0, \Sigma)$ be a Gaussian random vector in $\mathbb{R}^n$. Then*

$$\mathbb{E}(Xf(X)) = \Sigma \mathbb{E}(\nabla f(X)).$$

Proof. Let $X = \Sigma^{1/2} Z$, where $Z \sim N(0, I_n)$. So,

$$X_i = \sum_{k=1}^{n} (\Sigma^{1/2})_{i,k} Z_k$$

and

$$\mathbb{E}(X_i f(X)) = \sum_{k=1}^{n} (\Sigma^{1/2})_{i,k} \mathbb{E}(Z_k f(\Sigma^{1/2} Z)).$$

So, using the previous lemma for $\mathbb{E}(Z_k f(\Sigma^{1/2} Z)$ conditionally on all random variables except $Z_k \sim N(0, 1)$ and simplifying we get the desired equality. □

Lemma 2.3.23 is equivalent to the following: For any $i \in [n]$ it is true that

$$\mathbb{E}(X_i f(X)) = \sum_{j=1}^{n} \Sigma_{i,j} \mathbb{E}\left(\frac{df}{dx_j}(X)\right).$$

Lemma 2.3.24 (Gaussian interpolation)**.** *Consider two independent Gaussian random vectors $X \sim N(0, \Sigma^X)$ and $Y \sim N(0, \Sigma^Y)$. Define the interpolating Gaussian random vector*

$$Z(u) = \sqrt{u} X + \sqrt{1-u} Y, \quad u \in [0, 1].$$

Then for any twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ we have that

$$\frac{d}{du} \mathbb{E}(f(Z(u))) = \frac{1}{2} \sum_{i,j=1}^{n} (\Sigma^X_{i,j} - \Sigma^Y_{i,j}) \mathbb{E}\left(\frac{d^2}{dx_i dx_j}(Z(u))\right).$$

Proof. Using the chain rule we have

$$\begin{aligned} \frac{d}{du} \mathbb{E}(f(Z(u))) &= \sum_{i=1}^{n} \mathbb{E}\left(\frac{df}{dx_i}(Z(u)) \frac{dZ_i}{du}\right) \\ &= \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\left(\frac{df}{dx_i}(Z(u)) \frac{X_i}{\sqrt{u}} - \frac{Y_i}{\sqrt{1-u}}\right). \end{aligned}$$

Let us break this sum into two, and first compute the contribution of the terms containing X_i. To this end, we condition on Y and express

$$\frac{1}{2}\sum_{i=1}^{n}\frac{1}{\sqrt{u}}\mathbb{E}\left(\frac{df}{dx_i}(Z(u))X_i\right)=\frac{1}{2}\sum_{i=1}^{n}\frac{1}{\sqrt{u}}\mathbb{E}\left(X_i g(X_i)\right),$$

where $g_i(X)=\frac{df}{dx_i}(\sqrt{u}X+\sqrt{1-u}Y)$. So by the previous lemma we have

$$\mathbb{E}\left(Xg_i(X)\right)=\sum_{j=1}^{n}\Sigma^X_{i,j}\mathbb{E}\left(\frac{dg_i}{dx_j}(X)\right)=\sum_{j=1}^{n}\Sigma^X_{i,j}\mathbb{E}\left(\frac{d^2f}{d_{x_ix_j}}(\sqrt{u}X+\sqrt{1-u}Y)\sqrt{u}\right).$$

So,

$$\frac{1}{2\sqrt{u}}\sum_{i=1}^{n}\mathbb{E}\left(X_i\frac{df}{dx_i}(Z(u))\right)=\frac{1}{2}\sum_{i,j=1}^{n}\Sigma^X_{i,j}\mathbb{E}\left(\frac{d^2f}{d_{x_ix_j}}(\sqrt{u}X+\sqrt{1-u}Y)\right).$$

Taking expectation of both sides with respect to Y, we lift the conditioning on Y.

Similarly we can evaluate the second part of the sum and by that prove the lemma. □

Now we can return to the proof of the Slepian's inequality. Firstly note that for any $f:\mathbb{R}^n\to\mathbb{R}$ such that

$$\frac{df}{dx_i dx_j}\geq 0$$

for all $i\neq j$ we get that $\mathbb{E}f(X)\geq\mathbb{E}f(Y)$ which follows from Lemma 2.3.23 and the fact that the assumptions in Slepian's inequality imply that for any $i\in[n]$ it is true that $\Sigma^X_{i,i}=\Sigma^Y_{i,i}$ and that for any $i,j\in[n]$ it is true that $\Sigma^X_{i,j}\geq\Sigma^Y_{i,j}$.

Now we are going to approximate an indicator function $\mathbf{1}_{x\leq d}$ by a sequence of twice differentiable functions. Let $n\in\mathbb{N}$. Consider the 5-th degree polynomial g_n with the following properties:

- $g_n\left(d-\frac{1}{n}\right)=1$.
- $g_n\left(d+\frac{1}{n}\right)=0$.
- $g_n'\left(d-\frac{1}{n}\right)=g_n''\left(d-\frac{1}{n}\right)=0$.

· $g_n'\left(d+\frac{1}{n}\right) = g_n''\left(d+\frac{1}{n}\right) = 0.$

· $g_n(x) \geq 0$, $x \in \left[d-\frac{1}{n}, d+\frac{1}{n}\right]$.

Now consider the following function

$$h_n(x) = \begin{cases} 1_{x\leq d} & x \in \mathbb{R} \setminus \left(d-\frac{1}{n}, d+\frac{1}{n}\right) \\ g_n(x) & x \in \left[d-\frac{1}{n}, d+\frac{1}{n}\right] \end{cases}.$$

By the definition of g_n we have that h_n is twice differentiable and the sequence of functions $h_n(x)$ approximates the indicator function $\mathbf{1}_{x\leq d}$. Note that for all $n \in \mathbb{N}$ the function h_n is non-increasing.

Fix $m, n \in \mathbb{N}$. By what was done before, we have that the function $f_m : \mathbb{R}^n \to \mathbb{R}$

$$f_m(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} h_m(x_i)$$

is twice differentiable. The sequence of functions $f_m(x)$ is an approximation to the indicator function $\mathbf{1}_{\max_i x_i \leq d}$.

But

$$\frac{df_m}{d_{x_i} d_{x_j}} = h_m'(x_i) h_m'(x_j) \prod_{k\in[n]\setminus\{i,j\}} h_m(x_k).$$

But, by construction, the third part of the product is non-negative and the first two are both non-positive (since h_m is non-increasing). So, the product is non-negative.

As a result we have

$$\mathbb{E}\left(f_m(X)\right) \geq \mathbb{E}\left(f_m(Y)\right).$$

□

Lemma 2.3.25 (Chevet-Gordon inequalities)**.** *Let $B \in M_{p,n}$ be a random matrix with independent $N(0,1)$ entries. Let $K \subseteq \mathbb{R}^n$ and $L \subseteq S^{p-1}$ be compact sets and $r_k > 0$ such that $K \subseteq r_k B_2^n$ (K is a subset of the Euclidean ball of $\mathbb{R}^n$ with center at zero and radius r_k). Then,*

$$\mathbb{E} \max_{u\in L} \max_{t\in K} \langle Bt, u\rangle \leq w_G(K) + r_K w_G(L).$$

Proof. Let G be a Gaussian vector in $\mathbb{R}^n \oplus \mathbb{R}^p$. We are going to compare the following Gaussian processes induced by $(t, u) \in K \times L$:

$$X_{t,u} = \langle Bt, u \rangle,$$
$$Y_{t,u} = \langle G, t \oplus r_k u \rangle.$$

One can check that, for any $(t, u), (t', u') \in K \times L$ it is true that

$$\mathbb{E}(X_{t,u} - X_{t',u'})^2 \leq \mathbb{E}(Y_{t,u} - Y_{t',u'})^2.$$

So, by Slepian's inequality we get

$$\mathbb{E} max_{u \in L} \max_{t \in K} \langle Bt, u \rangle \leq \mathbb{E} \max_{u \in L} \max_{t \in K} Y_{t,u} = w_G(K) + r_k w_G(L).$$

□

Since $\sup_{t \in S^{n-1}} \sup_{u \in S^{p-1}} \langle Bt, u \rangle = \|B\|_{\text{op}}$ for any matrix $B \in M_{n,p}$, for a matrix $A \in M_{n,p}$ with independent $N(0, 1)$ random entries one has

$$\mathbb{E}\|A\|_{\text{op}} \leq w_G(S^{n-1}) + w_G(S^{p-1}).$$

But, by definition, one has that for any $m \in \mathbb{N}$ it is true that $w_G(S^{m-1}) = \mathbb{E}|G|$ where G is the Gaussian vector of $\mathbb{R}^m$. So, by Jensen's inequality,

$$\mathbb{E}\|G\|_2 \leq (\mathbb{E}\|G\|_2^2)^{1/2} = \sqrt{m},$$

which proves the desired inequality. □

So, since the operator norm is convex and 1-Lipschitz with respect to the Hilbert Schmidt norm, if M is the median of $\|B\|_{\text{op}}$ we have that $M \leq \sqrt{n} + \sqrt{p}$ by Lemma 2.1.37. Therefore, as in the proof of the semicircular law we get that

$$\mathbb{P}(\|B\|_{\text{op}} \geq \sqrt{n} + \sqrt{p} + t) \leq 2e^{-t^2/2}.$$

Now we can start the proof of the second part of the proposition. Since s_n is 1-Lipschitz with respect to the Hilbert-Schmidt norm, if M is the median of $s_n(B)$ then for any $t > 0$ we have that

$$\frac{1}{2} \exp(-tM^2) \leq \mathbb{E} \exp(-tBB^*) \leq n \exp(-(\sqrt{n} - \sqrt{p})^2 t^2 + (s + n)^2 t).$$

The first part of the previous inequality is true since $M \leq \mathbb{E}(s_n(B))$ (see Lemma 2.1.37) and for the second part see [19, Lemma 7.2]. So, for $t = \sqrt{(n+p)\ln(2p)}$ and by the inequality $\sqrt{a-b} \geq \sqrt{a} - \frac{b}{\sqrt{a}}$ which is valid for any $a \geq b \geq 0$, we get

$$M \geq \sqrt{n} - \sqrt{p} - 2\frac{\sqrt{p+n}\sqrt{ln(2p)}}{\sqrt{n} - \sqrt{p}}$$

So, for $t \geq \frac{4\sqrt{\ln(2n)}}{\sqrt{\frac{n}{p}}-1}$

$$\mathbb{P}(s_n(B) \leq \sqrt{n} - \sqrt{p} - t) \leq \mathbb{P}(s_n(B) \leq M - \frac{t}{2}) \leq 2e^{-t^2/2}.$$

The last inequality can be proved with the same method as the one that we used for the semicircular law. □

Part III

Quantum information theory

Random matrices in quantum information theory

In this part we study some results about random matrices that are strongly related to quantum information theory. Before that, we "translate" the results in quantum information theory terms.

In this chapter we are going to use the following notation: $M_n^{sa}(\mathbb{R})$ (or $\mathbb{C}$) will denote the subspace of symmetric (or Hermitian) $n \times n$ matrices in M_n.

We denote by $M_n^{sa,0}(\mathbb{F})$ the subspace of matrices in $M_n^{sa}(\mathbb{F})$ with trace equal to zero. Here $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.

For any $A \in M_n$ we will also write $\{\lambda_i(A)\}_{i=1}^n$ for the eigenvalues of A in non increasing order and μ_A for the E.S.D. of A.

3.1 The $\infty-$ Wasserstein distance

Definition 3.1.1. Let μ_1, μ_2 be two probability measures on $\mathbb{R}$. Their $\infty-$Wasserstein distance is defined as

$$d_\infty(\mu_1, \mu_2) := \inf \|\mu_1 - \mu_2\|_{L_\infty},$$

where the infimum is over all couples (X_1, X_2) of random variables with (marginal) laws μ_1,μ_2 defined on a common probability space. Similarly, if Y_1, Y_2 are real random variables, their $\infty-$Wasserstein distance will be meant to be the $\infty-$Wasserstein distance of their laws.

Note: The definition of the $\infty-$Wasserstein distance can be generalised on a metric space (E, d) by replacing $\inf \|X_1 - X_2\|_{L_\infty}$ by the smallest Δ such that $\mathbb{P}(d(X_1, X_2) \leq \Delta) = 1$.

We will now describe an alternative way to compute the ∞–Wasserstein distance.

Lemma 3.1.2. *For any real random variables X, Y we have*

$$d_\infty(X,Y) = \inf\{\epsilon > 0 : F_X(t-\epsilon) \leq F_Y(t) \leq F_X(t+\epsilon),\ \forall t \in \mathbb{R}\}.$$

Proof. Let $(\Omega, \mathcal{A}, \mathbb{P})$ and X', Y' be real random variables such that $X' \sim X$ and $Y' \sim Y$ defined on this probability space. Let $d := \|X - Y\|_{L_\infty}$ denote the L_∞ distance in this probability space. Then, for any $t > 0$

$$\mathbb{P}(X \leq t - d) \leq \mathbb{P}(X + |Y - X| \leq t) \leq \mathbb{P}(Y \leq t),$$

and likewise

$$\mathbb{P}(Y \leq t) \leq \mathbb{P}(X \leq t + d).$$

So

$$\begin{aligned}\Big\{d > 0 : \exists(\Omega, \mathcal{A}, \mathbb{P}) : X', Y' : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R},& \\ \|X - Y\|_{L_\infty((\Omega,\mathcal{A},\mathbb{P}))} = d, X' \sim X, Y' \sim Y\Big\}& \\ \subseteq \{\epsilon > 0 : F_X(t-\epsilon) \leq F_Y(t) \leq F_X(t+\epsilon),\ \forall t \in \mathbb{R}\}.&\end{aligned}$$

It follows that

$$d_\infty(X,Y) \leq \inf\{\epsilon > 0 : \forall t \in \mathbb{R}, F_X(t-\epsilon) \leq F_Y(t) \leq F_X(t+\epsilon)\}.$$

Conversely, let $\epsilon_0 \in \{\epsilon > 0 : \forall t \in \mathbb{R}, F_X(t-\epsilon) \leq F_Y(t) \leq F_X(t+\epsilon)\}$. Consider the probability space $((0,1), B(0,1), \lambda)$ (here λ stands for Lebesque measure) and the random variables $X'(\omega) = \{\inf t : F_X(t) \geq \omega\}$ and $Y'(\omega) = \{\inf t : F_Y(t) \geq \omega\}$.

Fix $\omega \in (0,1)$ and set $I_\omega = \{t : F_X(t) \geq \omega\}$. Note that I_ω is non-empty since $\lim_{n\to\infty} F(n) = 1$. Also, I_ω is an interval because F_X is increasing (if $t \in I_\omega$ then $[t, +\infty) \subseteq I_\omega$). But since $I_\omega \neq \mathbb{R}$ (because $\lim_{n\to\infty} F_X(-n) = 0$) and since any distribution function is right-continuous, which implies that I_ω is closed, we get that I_ω has the form

$$I_\omega = [b(\omega), \infty),$$

and since $\inf I_\omega = X'(\omega)$ we get that $b(\omega) = X'(\omega)$. Note that

$$t \in I_\omega \iff X'(\omega) \leq t.$$

So, for any $t \in \mathbb{R}$,

$$\begin{aligned} F_{X'}(t) &= \mathbb{P}(X' \leq t) = \lambda(\{\omega \in (0,1) : X'(\omega) \leq t\}) = \lambda(\{\omega \in (0,1) : t \in I_\omega\}) \\ &= \lambda(\{\omega \in (0,1) : F_x(t) \geq \omega\}) = F_x(t). \end{aligned}$$

So $X \sim X'$, and similarly $Y \sim Y'$. The functions X' and Y' are called generalised inverse functions of F_X and F_Y respectively. As a result we have, for any $\omega \in (0,1)$,

$$\{t : F_X(t - \epsilon_0) \geq \omega\} \subseteq \{t : F_Y(t) \geq \omega\}\} \subseteq \{\ t : F_X(t + \epsilon_0) \geq \omega\}.$$

But

$$\{t : F_Y(t - \epsilon_0) \geq \omega\} = \{s + \epsilon_0 : F_Y(s) \geq \omega\},$$

and similarly

$$\{t : F_Y(t + \epsilon_0) \geq \omega\} = \{s - \epsilon_0 : F_Y(s) \geq \omega\},$$

which implies that

$$\inf\{t : F_Y(t - \epsilon_0) \geq \omega\} = Y'(\omega) + \epsilon_0$$

$$\inf\{t : F_Y(t + \epsilon_0) \geq \omega\} = Y'(\omega) - \epsilon_0.$$

Therefore, we get that

$$Y'(\omega) - \epsilon_0 \leq X'(\omega) \leq Y'(\omega) + \epsilon_0 \quad \Longrightarrow \quad \|X' - Y'\|_{L_\infty((0,1),B(0,1),\lambda)} \leq \epsilon_0,$$

which shows that

$$d_\infty(X, Y) \geq \inf\{\epsilon > 0 : \forall t \in \mathbb{R}, F_X(t - \epsilon) \leq F_Y(t) \leq F_X(t + \epsilon)\}.$$

□

Lemma 3.1.3. *The d_∞ distance is greater than the Lévy distance d_L which metricizes weak convergence.*

Proof. Let X, Y be two real random variables. For any $\epsilon \in \{\epsilon : F_X(t-\epsilon) \leq F_Y(t) \leq F_X(t+\epsilon)\}$ we get that $\epsilon \in \{\epsilon : F_X(t-\epsilon) - \epsilon \leq F_Y(t) \leq F_X(t+\epsilon) + \epsilon\}$. Hence

$$\{\epsilon : F_X(t-\epsilon) \leq F_Y(t) \leq F_X(t+\epsilon)\} \subseteq \{\epsilon : F_X(t-\epsilon) - \epsilon \leq F_Y(t) \leq F_X(t+\epsilon) + \epsilon\},$$

which implies

$$\begin{aligned} d_\infty(X,Y) &= \inf\{\epsilon : F_X(t-\epsilon) \leq F_Y(t) \leq F_X(t+\epsilon)\} \\ &\geq \inf\{\epsilon : F_X(t-\epsilon) - \epsilon \leq F_Y(t) \leq F_X(t+\epsilon) + \epsilon\} \\ &= d_L(X,Y). \end{aligned}$$

□

Note: Convergence with respect to d_∞ implies weak convergence.

Lemma 3.1.4. *Let Z be a real random variable distributed according to a probability measure ν_Z whose support is a bounded interval $[a,b]$. If $\{Y_n\}_{n\in\mathbb{N}}$ is a sequence of random variables then the following are equivalent:*

1. $d_\infty(Y_n, Z) \to 0$.
2. *$Y_n \to Z$ weakly and* $\sup Y_n \to b$, $\inf Y_n \to a$.

Note: By sup and inf we mean the essential supremum and infimum respectively.

Proof. (i) $\implies$ (ii): We have already proven that convergence with respect to the ∞−Wasserstein distance implies weak convergence. Also, we have

$$\max\{|\sup Y_n - \sup Z|, |\inf Y_n - \inf Z|\} \leq \|Y_n - Z\|_{L_\infty}.$$

This proves this direction.

(ii) $\implies$ (i): Given $\epsilon > 0$ choose $a = x_1 < x_2 < x_3 \cdots x_r = b$ such that

$$x_{j+1} - x_j < \epsilon$$

for all $j \in [r-1]$. Suppose also that $\{x_i\}_{i=1}^{r-1}$ are points of continuity of F_Z (we may assume this, because the points of discontinuity of any distribution

function form a countable set, and hence the set of points of continuity is dense). Since the support of ν_Z is the interval $[a, b]$, we have that F_Z is strictly increasing on $[a, b]$, and hence there exists $c > 0$ such that $F_Z(x_{j+1}) \geq F_Z(X_j) + c$ for all $0 \leqslant j < r$. Also, for large enough n we get that

$$\inf Y_n > a - \epsilon$$

and

$$\sup Y_n < b + \epsilon,$$

and since $\{x_j\}_{j=1}^r$ is a set of points of continuity of F_Z we have

$$|F_Z(x_j) - F_{Y_n}(x_j)| < c.$$

Let $t \geq b + \epsilon$. Then,

$$\begin{aligned} F_{Y_n}(t) = \mathbb{P}(Y_n \leq t) \geq \mathbb{P}(Y_n < b + \epsilon) \geq \mathbb{P}(Y_n \leq \sup Y_n) = 1 \\ \geq \mathbb{P}(Z \leq t - 2\epsilon) = F_Z(t - 2\epsilon). \end{aligned}$$

Let $t \leq x_2$. Since $x_2 - x_1 < 2\epsilon$ we get that $t - 2\epsilon \leq x_2 - 2\epsilon < x_1$. So

$$F_Z(t - 2\epsilon) \leq F_Z(x_1) = F_Z(a) = 0 \leq F_{Y_n}(t).$$

Note that this is true since a is a continuity point of F_Z. Finally, let $t \in (x_2, b + \epsilon)$. Pick j such that $j \in [r]$ satisfies

$$x_{j-1} \leq t \leq x_j.$$

Then, $t \leq x_j \leq x_{j-2} + 2\epsilon$. Hence

$$\begin{aligned} F_Z(t - 2\epsilon) \leq F_Z(x_{j-2}) \leq F_Z(x_{j-1}) - c \leq F_Z(x_{j-1}) - |F_{Y_n}(x_{j-1}) - F_Z(x_{j-1})| \\ \leq F_{Y_n}(x_{j-1}) \leq F_{Y_n}(t). \end{aligned}$$

So, for every $t \in \mathbb{R}$ we have

$$F_Z(t - 2\epsilon) \leq F_{Y_n}(t).$$

Likewise, we can prove the inequality

$$F_{Y_n}(t) \leq F_Z(t + 2\epsilon).$$

As a result, for sufficiently large n we get

$$d_\infty(Z, Y_n) \leq 2\epsilon.$$

□

3.2 Wishart matrices and random induced states

Definition 3.2.1. Consider the space $M_n^{sa}(\mathbb{C})$ of complex Hermitian $n \times n$ matrices equipped with the Hilbert-Schmidt norm. We say that a matrix $A \in M_n^{sa}(\mathbb{C})$ is a GUE(n)-matrix (Gaussian Unitary Ensemble) if

$$a_{i,j} \sim N_{\mathbb{C}}(0,1) \quad i \neq j$$
$$a_{i,j} \sim N_{\mathbb{R}}(0,1) \quad i = j$$

and the entries of A are pairwise independent.

Note that if U is a unitary matrix then $UAU^* \sim A$.

Definition 3.2.2. Consider the space $M_n^{sa,0}(\mathbb{C})$ equipped with the Hilbert-Schmidt norm. Then if A is in $GUE(n)$ we say that the matrix $B = A - \frac{\mathrm{tr}A}{n} I$ is in $GUE_0(n)$.

Note that the coefficient $\frac{\mathrm{tr}A}{n}$ has distribution $N(0, 1/n)$ and is independent from B.

Definition 3.2.3. Let $n, s \in \mathbb{N}$. Consider the space $M_{n\times s}$ and let $B \in M_{n\times s}$ be a random matrix whose entries are independent random variables all following $N_{\mathbb{C}}(0,1)$. Then the matrix $W = BB^*$, which is in M_n^{sa}, is called Wishart matrix and its distribution is denoted by Wishart(n, s).

There are several models that can be used to study random states. Next we are presenting two of them.

Definition 3.2.4. (i) A random $n \times s$ state is a matrix generated by the following procedure. Consider independent unit vectors $\{\psi_i\}_{i\in[s]}$ distributed uniformly on the sphere of $\mathbb{C}^n$ and consider the average of the corresponding pure states, i.e.

$$\rho = \frac{1}{s} \sum_{i=1}^{s} |\psi_i\rangle\langle\psi_i|.$$

We are now going to present some results about tensor products that can lead to a closely related and often better model of random states. A fundamental concept in quantum information theory is the partial trace which we define bellow:

Definition 3.2.5. Let $H_1 \otimes H_2$ be a bipartite Hilbert space (meaning that it is the tensor product of two finite dimensional Hilbert spaces). A function $\mathrm{tr}_{H_2} : B(H_1) \otimes B(H_2) \to B(H_1)$ is called partial trace over H_2 if and only if

$$\mathrm{tr}_{H_2}(A \otimes B) = \mathrm{tr}(B)A$$

for all $A \in B(H_1)$ and $B \in B(H_2)$.

Lemma 3.2.6. *If $y \in \mathbb{C}^n \otimes \mathbb{C}^m$ then by Corollary 1.5.6 there exists $M \in M_{n,m}(\mathbb{C})$ such that $M = y$. So*

$$\mathrm{tr}_{C^m} |y\rangle\langle y| = MM^*.$$

Proof. In order to prove the lemma we need the following useful tool.

Lemma 3.2.7 (Schmidt decomposition)**.** *Let H_1, H_2 be two Hilbert spaces. Then, every pure state $|\psi\rangle \in H_1 \otimes H_2$ can be written as a linear combination*

$$|\psi\rangle = \sum_{k=1}^{d} \lambda_k |\phi_k^1\rangle |\psi_k^2\rangle,$$

where $d = \min\{\dim(H_1), \dim(H_2)\}$, $\{|\phi_k^1\rangle\} \subseteq H_1$ and $\{|\psi_k^2\rangle\} \subseteq H_2$ are orthonormal sets, and $\{\lambda_k\}_{k=1}^d$ are non-negative real coefficients with $\sum_{k=1}^d \lambda_k^2 = 1$.

Proof. We denote $d_1 = \dim(H_1)$ and $d_2 = \dim(H_2)$ and assume that $d_1 \geq d_2$. We can write a vector $|\psi\rangle \in H_1 \otimes H_2$ in terms of orthonormal bases $\{i_k^1\}_{k=1}^{d_1}$ and $\{j_l^2\}_{l=1}^{d_2}$

$$|\psi\rangle = \sum_{k,l=1}^{d_1,d_2} a_{i,j} |i_k^1\rangle |j_l^2\rangle.$$

Let $E = [a_{i,j}] \in M_{d_1,d_2}$ be the corresponding matrix. Now we can apply the singular value decomposition to the matrix E, which implies that there exist unitary matrices $U \in M_{d_1}$, $V \in M_{d_2}$ and a positive diagonal matrix $\Sigma \in M_{d_2}$ whose entries $\{\lambda_k\}_{k=1}^{d_2}$ are the singular values of E (since the non-zero singular values of E are at most d_2), such that

$$E = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^*.$$

It follows that

$$|\psi\rangle = \sum_{i,j,k} u_{i,k} \lambda_k v_{k,j} |i^1\rangle |j^2\rangle = \sum_{k=1}^{d_2} \lambda_k |\phi_k^1\rangle |\psi_k^2\rangle,$$

where the vectors $\{|\phi_k^1\rangle\}$ constitute an orthonormal set in H_1, and the same for $\{|\psi_k^2\rangle\}$ in H_2, due to the fact that U, V are unitary. Finally, since $|\psi\rangle$ is a unit vector, the corresponding matrix has Hilbert-Schmidt norm equal to one: in other words, $\|E\|_2^2 = \sum_{i,j} |a_{i,j}|^2 = 1$ which in turn implies that $\sum_k \lambda_k^2 = 1$. □

Now, let $y \in \mathbb{C}^n \otimes \mathbb{C}^m$. Write $\{x_i\}_{i\in[n]}$ for the n-dimensional standard basis and $\{\phi_j\}_{j\in[m]}$ for the m-dimensional standard basis. From the Schmidt decomposition we have that there exist $\lambda_{i,j}$ such that

$$y = \sum_i \lambda_i x_i \otimes \phi_i.$$

So,

$$|y\rangle\langle y| = \sum_{i,j} \lambda_i \lambda_j |x\rangle_i \langle x|_j \otimes y_i \otimes y_j.$$

Let $\delta_i(j) = 1_{j=i}$. By linearity of the partial trace, and since the set $\{\phi_i\}_{i\in[m]}$ is orthogonal, we have

$$\begin{aligned} \mathrm{tr}_{\mathbb{C}^m} |y\rangle\langle y| &= \sum_{i,j} \lambda_i \lambda_j |x\rangle_i \langle x|_j \langle \phi|_i |\phi\rangle_j = sum_{i,j} \lambda_i \lambda_j |x\rangle_i \langle x|_j \delta_i(j) \\ &= \sum_i \lambda_i^2 |x\rangle_i \langle x|_i, \end{aligned}$$

which is exactly the matrix BB^*, where B is the matrix which is equivalent to the state y. □

Now we are ready to give the definition of another random state model, which is slightly different from the previous one, and sometimes better. Recall that the first one was given in Definition 3.2.4.

Definition 3.2.8. Let $m, n \in \mathbb{N}$ and let y be uniformly distributed on the sphere of $\mathbb{C}^n \otimes \mathbb{C}^m$. Then, the partial trace $\mathrm{tr}_{\mathbb{C}^n} |y\rangle\langle y|$ of y over $\mathbb{C}^n$ (likewise on $\mathbb{C}^m$) is called random $n \times m$ induced state.

We also use the notation $\mu_{n,s}$ for the distribution of the random $n \times s$ induced state.

In the rest of the thesis we will work on with random induced states. The next lemma shows that a random induced state is a normalization of Wishart matrices. More precisely:

Lemma 3.2.9. *The distribution of a random $n \times s$ induced state is $\frac{W}{\mathrm{tr}W}$ where $W \sim \mathrm{Wishart}(n, s)$ and the random induced state is independent from $\mathrm{tr}W_{n,s}$.*

Proof. The first part of the lemma is merely a combination of the fact that a random vector uniformly distributed on the sphere of $\mathbb{C}^n \otimes \mathbb{C}^s$ is an $n \times s$-matrix whose entries are independent random variables all following $N_{\mathbb{C}}(0, 1)$ and Lemma 3.2.6 which implies that the partial trace of y over $\mathbb{C}^n$ is exactly what the lemma says.

For the second part, the proof is a simple consequence of Remark 1.4.20.

□

The results that we have presented for the Empirical Spectral Distribution have applications to random induced states. In order to state and prove them, we need the following concentration lemmas.

Lemma 3.2.10. *Let $X \sim \chi^2(n)$ .Then*

$$\mathbb{P}(|X - n| \geq \epsilon n) \leq 2 \exp\left(-\frac{n\epsilon^2}{4 + 8\epsilon/3}\right).$$

Proof. By the definition of the $\chi^2(n)$-distribution there exist i.i.d. random variables $\{Z_i\}_{i\in[n]}$ such that $Z_i \sim N(0, 1)$ and $X = \sum_{i=1}^n Z_i^2$. So, for any $s \in (0, \frac{1}{2})$, by independence and isonomy we have that

$$\mathbb{E}(\exp(sX)) = \mathbb{E}\left(\exp \sum_{i=1}^n sZ_i^2\right) = \mathbb{E}\left(\prod_{i=1}^n \exp(sZ_i^2)\right) = \prod_{i=1}^n \mathbb{E}\exp(sZ_i^2).$$

But

$$\begin{aligned}
\mathbb{E}\exp(sZ^2) &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp(sx^2) \exp\left(-\frac{1}{2}x^2\right) dx \\
&= \frac{1}{(1-2s)^{1/2}} \int \frac{(1-2s)^{1/2}}{(2\pi)^{1/2}} \exp\left(-\frac{1}{2}x^2(1-2s)\right) dx \\
&= (1-2s)^{-1/2},
\end{aligned}$$

since the probability density function of a random variable $Y \sim N(0,(1-2s))$ is $f_Y(x) = \frac{(1-2s)^{1/2}}{(2\pi)^{1/2}} \exp\left(-\frac{1}{2}x^2(1-2s)\right)$. It follows that

$$\mathbb{E}(\exp(sX)) = (1-2s)^{-n/2}.$$

So, given $\epsilon > 0$ and for $s = \frac{\epsilon}{2(1+\epsilon)}$, by Markov's inequality we get

$$\begin{aligned}\mathbb{P}(X \geq (1+\epsilon)n) &= \mathbb{P}(sX \geq s(1+\epsilon)n) = \mathbb{P}(\exp(sX) \geq \exp(1+\epsilon)sn) \\ &\leq [(1+\epsilon)\exp(-\epsilon)]^{n/2}.\end{aligned}$$

From the inequality $1+\epsilon \leq \exp(\epsilon - (\epsilon^2 - \epsilon^3)/2)$ we see that

$$\mathbb{P}(X \geq (1+\epsilon)n) \leq \exp(-(\epsilon^3 - \epsilon^2)n/2).$$

Similarly, for $s = \frac{\epsilon}{2(1-\epsilon)}$ we get

$$\mathbb{P}(X \leq (1-\epsilon)) = \mathbb{P}(\exp(-sX) \geq \exp(-s(1-\epsilon)n)) \leq \exp(-(\epsilon^3 - \epsilon^2)n/4).$$

Therefore, for any $\epsilon \in (0,1)$ we get the desired inequality. □

Lemma 3.2.11. *Let W be a* Wishart(n,s) *matrix. Then, for any $t > 0$ we have that*

$$\mathbb{P}(|\mathrm{tr}W - ns| \geq tns) \leq 2\exp\left(-\frac{nst}{2+4t/3}\right).$$

Proof. If $W = BB^*$ then

$$2\mathrm{tr}W = \sum_{i,j=1}^{n} 2|\mathrm{Re}(B_{i,j})|^2 + 2|\mathrm{Im}(B_{i,j})|^2.$$

This sum is exactly the sum of ns squared independent $N(0,1)$ random variables. So, combining this observation with the previous lemma we conclude the proof. □

Corollary 3.2.12. *For any $\epsilon > 0$*

$$\mathbb{P}\left(\left|\frac{\mathrm{tr}W_{n,s}}{ns} - 1\right| \geq \epsilon\right) \to 0.$$

Moreover, the convergence is stronger (meaning almost surely) by the Borel-Cantelli lemma.

We are now ready to translate all the random matrix theory results of Part II to the language of random induced states.

Theorem 3.2.13. *Given $n, s \in \mathbb{N}$, let $\rho_{n,s}$ be a random induced state with probability distribution $\mu_{n,s}$. Then*

(i) *If n is fixed and s tends to infinity then $C\sqrt{s}(\rho_{n,s} - \frac{I}{n})$ converges in distribution towards a $\mathrm{GUE}_0(n)$ matrix, where C is an absolute constant.*

(ii) *If $\lim s/n = \lambda \in (0, \infty)$ then $\mu_{ns}(s\rho_{n,s})$ (the E.S.D.) converges weakly in distribution towards $\mu_{MP(\lambda)}$ (the Marchenko-Pastur distribution). If $\lambda \geq 1$ then the convergence is also true for the $\infty-$Wasserstein distance.*

(iii) *If $s/n, s \to \infty$ then $\mu_{sn}(\sqrt{ns}(\rho_{n,s} - \frac{I}{n}))$ converges in probability with respect to the ∞-Wasserstein distance towards the semicircular law μ_{SC}.*

Proof. (i) By the multivariate central limit theorem for the vector space $M_n^{sa,0}$, if $\{G_i\}_{i\in\mathbb{N}}$ is a sequence of standard normal vectors in $\mathbb{C}^n$ and $A_i = |G_i\rangle\langle G_i|$, and since $\mathrm{tr}W_{n,s}$ can be virtually treated as a constant, we have

$$\frac{\sum_{i=1}^{s} A_i - sI}{\sqrt{s}} \to \mathrm{GUE}_0(n) \quad \Longrightarrow \quad n\sqrt{s}\left(\rho_{n,s} - \frac{I}{n}\right) \to \mathrm{GUE}_0(n).$$

(ii) In the general case, where $\lambda \in (0, \infty)$, the proof follows from the next facts:

Fact 1: If X_n and Y_n are sequences of random variables defined on a common probability space that converge in probability towards the random variables X and Y respectively, then X_nY_n converges in probability towards XY.

Fact 2: From the results of the section on the Marchenko-Pastur theorem, if $W_{n,s}$ is a Wishart matrix then $\mu(W_{n,s}/n)$ converges in probability towards the Marchenko-Pastur distribution (see 2.2.1).

Fact 3: By 3.2.12 we have that $\mathrm{tr}W_{n,s}/ns$ converges weakly towards 1.

In the case where $\lambda \geq 1$ we have that the extreme eigenvalues of $W_{n,s}$ also converge in probability towards the infimum and the maximum of

the Marchenko-Pastur density function. So, by the characterization of the $\infty-$Wasserstein distance and Facts 1 and 3, the proof is complete. (see 2.2.9).

(iii) The proof is similar with the one of (ii) by Proposition 2.3.18 and Theorem 2.3.1. □

Random quantum states

4.1 Miscellaneous tools

In the first section of this chapter we give some necessary definitions and prove some important tools.

4.1.1 Majorization inequalities

Definition 4.1.1. Let $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$. Let also $\sigma \in S_n$ be a permutation of n elements such that the coordinates of x via σ are being arranged in decreasing order, and let d be a permutation of n elements such that the coordinates of y via d are also being arranged in decreasing order. Let $x' = (x'_1, \ldots, x'_n)$, $y' = (y'_1, \ldots, y'_n)$ be the n-dimensional vectors that we obtain when we apply the permutations σ and d to x and y respectively. If $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$ we will say that x is majorized by y and write $x \prec y$ if

$$\text{for all } k \in [n] \text{ we have that } \sum_{i=1}^{k} x'_i \leq \sum_{i=1}^{k} y'_i.$$

The next lemma provides some simple properties of majorization.

Lemma 4.1.2 (properties of majorization)**.** *Let $n \in \mathbb{N}$ and $x, y, z \in \mathbb{R}^n$. Then:*

(i) *If $x \prec y$ and $y \prec z$ then $x \prec z$.*

(ii) *If $\lambda \in (0, \infty)$ and $x \prec y$ then $\lambda x \prec \lambda y$.*

(iii) *If $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ are such that $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i = 0$ and $x \prec y$ then for any $\lambda \in (0, 1)$ we have that $\lambda x \prec y$.*

(iv) *If $x \prec z$ and $y \prec z$ then for any $\lambda \in (0,1)$ we have that*

$$\lambda x + (1-\lambda)y \prec z.$$

Proof. Let $x', y', z' \in \mathbb{R}^n$ be the vectors with the same elements as x, y, z respectively, arranged in decreasing order.

(i) By the assumption we have that for any $k \in [n]$

$$\sum_{i=1}^{k} x_i' \leq \sum_{i=1}^{k} y_i' \leq \sum_{i=1}^{k} z_i'$$

which proves that $x \prec z$ since it is also true that $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} z_i$.

(ii) Since $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$ we get that $\sum_{i=1}^{n}(\lambda x_i) = \sum_{i=1}^{n}(\lambda y_i)$. Also, since

$$\sum_{i=1}^{k} x_i' \leq \sum_{i=1}^{k} y_i'$$

for every $k \in [n]$ and $\lambda > 0$, we see that

$$\lambda \sum_{i=1}^{k} x_i' \leq \lambda \sum_{i=1}^{k} y_i'$$

for every $k \in [n]$.

(iii) Firstly note that $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_1 = \sum_{i=1}^{n}(\lambda x_i) = 0$.

Secondly since $\lambda \in (0,1)$ the rearrangement of the elements of the vector x in decreasing order will arrange the elements of λx in decreasing order as well. Also, for any $k \in [n]$, we have $\sum_{i=1}^{k} x_i' \geq 0$, and hence

$$\sum_{i=1}^{k}(\lambda x_i') \leq \sum_{i=1}^{k} x_i'.$$

So, $\lambda x \prec x$. Then, using (i) for the vectors $\lambda x, x, y$ we get $\lambda x \prec y$.

(iv) Let $\lambda \in (0,1)$. Note that

$$\sum_{i=1}^{n} \lambda x_i + (1-\lambda)y_i = \lambda \sum_{i=1}^{n} z_i + (1-\lambda)\sum_{i=1}^{n} z_i = \sum_{i=1}^{n} z_i.$$

Then, by (ii), for any $k \in [n]$ we have that

$$\sum_{i=1}^{k} \lambda x_i' + (1-\lambda)y_i' \leq \sum_{i=1}^{k} \lambda z_i + \sum_{i=1}^{k}(1-\lambda)z_i = \sum_{i=1}^{k} z_i.$$

□

Next, we present several characterizations of majorization.

Lemma 4.1.3. *Let $x, y \in \mathbb{R}^n$. Suppose also that $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$. Then, the following are equivalent:*

(i) $x \prec y$.

(ii) *x can be written as a convex combination of coordinate-wise permutations of y.*

(iii) *There exists a doubly stochastic $n \times n$ matrix B such that $Bx = y$.*

(iv) *If ϕ is a permutation invariant convex function on $\mathbb{R}^n$ then $\phi(x) \leq \phi(y)$.*

(v) *For every $t \in \mathbb{R}$ we have that $\sum_{i=1}^n |x_i - t| \leq \sum_{i=1}^m |y_i - t|$.*

(vi) *For every $t \in \mathbb{R}$ we have that $\sum_{i=1}^n (x_i - t)^+ \leq \sum_{i=1}^m (y_i - t)^+$ where, for any $z \in \mathbb{R}$ we use the notation $z^+ = \max\{z, 0\}$.*

Proof. Given a vector $a \in \mathbb{R}^n$ we will use the notation $a^\downarrow$ for the vector in $\mathbb{R}^n$ which the same elements as a but in decreasing order.

For the equivalence of (i) and (ii) consider the set

$$A_y = \{z \in \mathbb{R}^n : z \prec y\}.$$

Note that A_y is convex and its extreme points are permutations of y, meaning those $z \in \mathbb{R}^n$ that satisfy $z^\downarrow = y^\downarrow$. Now, the equivalence follows by the Krein-Milman theorem.

For the equivalence of (ii) and (iii), similarly, we use the classical Birkhoff theorem, which asserts that the extreme points of the set of doubly stochastic matrices are exactly the permutation matrices. So the equivalence is implied by using the same permutation whose convex combination is B to receive x from extreme elements of A_y and vice-versa.

The implications (ii) $\implies$ (iv) $\implies$ (v) are obvious.

For the equivalence of (v) and (vi) we just combine the facts that $|x| = 2x^+ - x$ and $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$

Finally for the implication (vi) $\Longrightarrow$ (i) set $t = y_k^{\downarrow}$ for some $k \in [n]$. Then

$$\sum_{i=1}^{n}(y_i - t)^+ = \sum_{i=1}^{k} y_i^{\downarrow} - kt,$$

but

$$\sum_{i=1}^{n}(x_i - t)^+ = \sum_{i=1}^{n}(x_i^{\downarrow} - t)^+ \geq \sum_{i=1}^{k} x_i^{\downarrow} - kt,$$

which ends the proof. □

Definition 4.1.4. We will use the notation $\mathbb{R}^{n,0}$ for the hyperplane

$$\mathbb{R}^{n,0} = \left\{ x \in \mathbb{R}^n : \sum_{i=1}^{n} x_i = 0 \right\}.$$

Lemma 4.1.5. *Let $x, y \in \mathbb{R}^{n,0}$. Assume that $\|y\|_\infty \leq 1$ and $\|y\|_1 \geq an$ for some $a \in (0, 1]$. Then*

$$x \prec \left(\frac{2}{a} - 1\right)\|x\|_\infty y.$$

Proof. By homogeneity and property (ii) of majorization we may assume that $\|x\|_\infty \leq 1$. So we need to show that for any $x \in \mathbb{R}^{n,0}$ with $\|x\|_\infty \leq 1$ it is true that $x \prec (\frac{2}{a} - 1)y$. Consider the set

$$A_{(\frac{2}{a}-1)y} = \left\{ z \in \mathbb{R}^{n,0} : z \prec \left(\frac{2}{a} - 1\right) y \right\}.$$

We will also use the notation $B_\infty^{n,0}$ for the n-dimensional unit ball with respect to the infinity norm restricted on the hyperplane $\mathbb{R}^{n,0}$.

By the properties of majorization we get that the set $A_{(\frac{2}{a}-1)y} \cap B_\infty^{n,0}$ is convex. So, if we show that

$$\mathrm{ext}(B_\infty^{n,0}) \subseteq A_{(\frac{2}{a}-1)y} \cap B_\infty^{n,0},$$

then by the Krein-Milman theorem we will have

$$B_\infty^{n,0} = A_{(\frac{2}{a}-1)y} \cap B_\infty^{n,0},$$

which will prove the lemma. Here for a convex set $D \subseteq \mathbb{R}^n$ we use the notation $\mathrm{ext}(D)$ for the set of extreme points of D.

In order to do this, we need first to specify which are the extreme points of $B_\infty^{n,0}$.

Lemma 4.1.6. *If $n \in 2\mathbb{N}$ then*

$$\text{ext}(B_\infty^{n,0}) = \left\{ x \in B_\infty^{n,0} : |\{i \in [n] : x_i = 1\}| = |\{i \in [n] : x_i = -1\}| = \frac{n}{2} \right\},$$

and if $n \in 2\mathbb{N} + 1$ then

$$\text{ext}(B_\infty^{n,0}) = \left\{ x \in B_\infty^{n,0} : |\{i \in [n] : x_i = 1\}| = |\{i \in [n] : x_i = -1\}| = \frac{n-1}{2} \right\}.$$

Note: In the case where n is odd, if $x \in \text{ext}(B_\infty^{n,0})$ then the coordinates of x which are not equal to 1 or -1 must be equal to zero since $\sum_{i\in[n]} x_i = 0$.

Proof. We assume that $n \in 2\mathbb{N}$. The case where n is odd is very similar and is omitted.

($\supseteq$) Let $d \in \left\{ x \in B_\infty^{n,0} : |\{i \in [n] : x_i = 1\}| = |\{i \in [n] : x_i = -1\}| = \frac{n}{2} \right\}$. Suppose that there exist $y, z \in B_\infty^{n,0}$ and $\lambda \in (0,1)$ such that

$$d = \lambda y + (1-\lambda) z.$$

We will prove that $y = z$. If there exists $i \in [n]$ such either $|y_i| < 1$ or $|z_i| < 1$ then

$$|\lambda y_i + (1-\lambda) z_i| \leq \lambda |y_i| + (1-\lambda)|z_i| < 1,$$

which is a contradiction.

So $|y_i| = |z_i| = 1$ for all $i \in [n]$. Suppose now that there exists $i \in [n]$ such that $z_i = -y_i$. Then $2\lambda y_i + z_i \in \{-1, 1\}$. But then, $\lambda \in \{-1, 0, 1\}$, which is a contradiction. Therefore, $z = y$ and as a result d is an extreme point of $B_\infty^{n,0}$.

($\subseteq$) Let $d \in \text{ext}(B_\infty^{n,0})$. Suppose that there exists $j \in [n]$ such that $0 < |d_i| < 1$. Since $\sum_{i=1}^n d_i = 0$ there must exist another coordinate $k \in [n]$, $k \neq j$ such that $0 < |d_k| < 1$ and $d_k d_j < 0$. Without loss of generality we may assume that $d_j > 0 > d_k$. We may find an $\epsilon > 0$ such that $d_j + \epsilon < 1$, $d_j - \epsilon > 0$, $d_k - \epsilon > -1$, $d_k + \epsilon < 0$. Then, consider the vectors z and y with

$$\forall i \in [n] \setminus \{j, k\} \quad z_i = d_i,\ z_k = d_k - \epsilon,\ z_j = d_j + \epsilon$$

and

$$\forall i \in [n] \setminus \{j, k\} \quad y_i = d_i,\ y_k = d_k + \epsilon,\ y_j = d_j - \epsilon.$$

Then,

$$d = \frac{1}{2}y + \frac{1}{2}z,$$

and by the way ϵ was chosen $z, y \in B_\infty^{n,0}$ which shows that d can not have a non-zero coordinate different from 1 and -1.

Suppose that there exists j such that $d_j = 0$. By what was done before and since $\sum_{i=1}^{n} d_i = 0$ there must exist $k \in [n]$, $k \neq j$ such that $d_k = 0$. But then, for the vectors z and y with

$$\forall i \in [n] \setminus \{j, k\} \quad z_i = d_i,\ z_j = -z_k = 1$$

and

$$\forall i \in [n] \setminus \{j, k\} \quad z_i = d_i,\ z_j = -z_k = -1,$$

we have that $z, y \in B_\infty^{n,0}$ and, as before,

$$d = \frac{1}{2}z + \frac{1}{2}y.$$

So d cannot be an extreme point.

The conclusion is that an extreme point of $B_\infty^{n,0}$ must have all its coordinates non-zero and with absolute value 1. But since the sum of all its coordinates must be zero we see that the cardinality of the set of coordinates equal to 1 must be equal to the cardinality of the set of coordinates equal to -1. So, the proof is complete. □

Returning now to the proof of the main lemma we need to show that for any $x \in \text{ext}(B_\infty^{n,0})$ we have

$$x \prec \left(\frac{2}{a} - 1\right) y.$$

We may also assume that $\|y\|_1 = an$, by the properties of majorization, and that $n \in 2\mathbb{N}$ (the proof in the case where n is odd is very similar and is omitted).

Let $x \in \text{ext}(B_\infty^{n,0})$. Consider the sets

$$C = \{i \in [n] : y_i \geq 0\}$$

and

$$D = \{i \in [n] : y_i < 0\}.$$

Both sets are non-empty, since $\sum_{i\in[n]} |y_i| > 0$ so there exists at least one non-zero coordinate and $y \in \mathbb{R}^{n,0}$ and as a result there exists at least another coordinate with different value than the previous one. Also it is obvious that

$$C \cup D = [n].$$

Since $y \in \mathbb{R}^{n,0}$, it is true that

$$\sum_{i\in C} y_i = -\sum_{i\in D} y_i,$$

and as a result, since $\|y\|_1 = an$, we get

$$\sum_{i\in C} y_i + \sum_{i\in D} -y_i = an \quad \Longrightarrow \quad \sum_{i\in C} y_i = \frac{an}{2} = -\sum_{i\in D} y_i.$$

Finally, since $\|y\|_\infty \leq 1$,

$$\frac{an}{2} \leq |C| \quad \text{and} \quad |D| \leq 1 - \frac{an}{2}.$$

Suppose without loss of generality that $|D| \leq |C|$. Now consider the vector y' with

$$y'_i = \frac{\sum_{i\in C} y_i}{|C|} \text{ for all } i \in |C| \quad \text{and} \quad y'_i = \frac{\sum_{i\in D} y_i}{|D|} \text{ for all } i > |C|.$$

Obviously, $y' \in \mathbb{R}^{n,0}$. By the previous inequalities we have that

$$\frac{a}{2-a} \leq |y'_i|.$$

Now consider the vector d with

$$d_i = y'_i \text{ for all } i \in [n/2] \quad \text{and} \quad d_i = \frac{\sum_{i>\frac{n}{2}} y'_i}{n/2} \text{ for all } i > n/2.$$

Finally we have constructed a vector with the first $n/2$ coordinates equal and positive and the last $n/2$ coordinates equal and negative. So, we have that $\frac{a}{2-a}x \prec |d_1|x = d$. Note that, by construction, the elements of both d and y' are in decreasing order. Now, let $k \in [n]$.

If $k \leq \frac{n}{2}$ then

$$\sum_{i=1}^{k} d_i = \sum_{i=1}^{k} y'_i.$$

If $k \geq \frac{n}{2}$ then obviously

$$n - k \leq |C| \quad \Longrightarrow \quad \sum_{i=k+1} y'_i \leq \frac{(n-k)}{n/2} \sum_{i>n/2} y'_i,$$

which implies that

$$\sum_{i=1}^{k} d_i \leq \sum_{i=1}^{k} y'_i.$$

So, we have proven that $d \prec y'$.

Let z be a vector of $\mathbb{R}^n$ with the same coordinates as y but arranged in decreasing order.

Note: Let $\{a_m\}_{m\in\mathbb{N}}$ be a sequence in decreasing order. Then, for any $N \in \mathbb{N}$,

$$a_{N+1} \leq \frac{\sum_{i=1}^{N} a_i}{N}.$$

Using the same method inductively for the coordinates of the vector y we have that for any $k \in [|C| - 1]$

$$\sum_{i=k+1}^{|C|} \frac{y_i}{|C| - k} \leq \sum_{i=1}^{k} \frac{y_i}{k} \quad \Longrightarrow \quad \sum_{i=1}^{k} y'_i \leq \sum_{i=1}^{k} z_i.$$

Obviously, we have that $\sum_{i=1}^{|C|} y'_i = \sum_{i=1}^{|C|} z_i$.

Likewise, for any $|C| < k < n$ using again the note inductively we get

$$\sum_{i=|C|+1}^{k} y'_i \leq \sum_{i=1}^{k} z_i \quad \Longrightarrow \quad \sum_{i=|C|+1}^{k} y'_i \leq \sum_{i=1}^{k} z_i.$$

Then we must have that $y' \prec z$, which is equivalent to $y' \prec y$. So we have proven that

$$\frac{a}{2-a} x \prec d \prec y' \prec y.$$

Using the properties of majorization we get $\frac{a}{2-a} x \prec y$, and this completes the proof. □

Lemma 4.1.7. *Let $x, y \in \mathbb{R}^{n,0}$ and $y \neq 0$. Then*

$$x \prec \frac{2n\|x\|_\infty}{\|y\|_1} y.$$

Proof. By homogeneity, we may assume that $\|y\|_\infty = 1$ and the result follows from the previous lemma. □

Proposition 4.1.8. *Let $x, y \in \mathbb{R}^{n,0}$. Assume that $\|x - y\|_\infty < \epsilon$ and $\|y\|_1 \geq an$ for some $a > 0$. Then*

$$x \prec \left(1 + \frac{2\epsilon}{a}\right) y.$$

Proof. We use the following elementary property of majorization: If $x_1 \prec \lambda_1 y$ and $x_2 \prec \lambda_2 y$ for some positive λ_1, λ_2 then

$$x_1 + x_2 \prec (\lambda_1 + \lambda_2) y.$$

Setting $x_1 = y$,$x_2 = x - y$, $\lambda_1 = 1$ and using the previous lemma we see that we can also set $\lambda_2 = \frac{2\epsilon}{a}$ and the proposition follows. □

4.1.2 Spectra and norms of unitarily invariant random matrices

At this point we are going to work on the spectra of norms of unitarily invariant matrices in order to approximate a specific gauge or norm.

It is convenient to work in the hyperplane $M_n^{sa,0}$ of self-adjoint complex $n \times n$ matrices with trace zero. We say that a $M_n^{sa,0}$-valued random variable A is unitarily invariant if, for any $U \in U(n)$, the random matrices A and UAU^* have the same distribution. We will also use the notation μ_{SC} for the semicircular distribution, $\mu_{sp}(A)$ for the empirical spectral distribution of a self-adjoint matrix A, and d_∞ for the ∞-Wasserstein distance.

Proposition 4.1.9. *Let A and B be two $M_n^{sa,0}$-valued random variables which are unitarily invariant and satisfy the conditions*

$$\mathbb{P}(d_\infty(\mu_{sp}(A), \mu_{SC}) \leq \epsilon) \geq 1 - p$$

and

$$\mathbb{E}(d_\infty(\mu_{sp}(A), \mu_{SC}(A)) \leq \epsilon$$

for some $\epsilon, p \in (0, 1)$ and similarity for B. Then, for any convex and compact set $K \subseteq M_n^{sa,0}$ containing the origin in its interior, we have that

$$\frac{1-p}{1+C\epsilon} \mathbb{E}\,\|A\|_K \leq \mathbb{E}\,\|B\|_K \leq \frac{1+C\epsilon}{1-p} \mathbb{E}\,\|A\|_K,$$

where $C > 0$ is an absolute constant.

Proof. Consider a (product) probability space $(\Omega, A, \mathbb{P})$ such that both A and B are defined there and are independent.

Let $\phi : \mathbb{R}^{n,0} \to \mathbb{R}$ be the function defined by

$$\phi(x) = \mathbb{E}_U \|U\mathrm{Diag}(x)U^*\|_K,$$

where $U \in U(n)$ denotes a Haar-distributed random unitary matrix (independent of everything else) and $\mathrm{Diag}(x)$ is the diagonal matrix whose ii-th entry is x_i. By unitary invariance we have that

$$\mathbb{E}\,\|A\|_K = \mathbb{E}\,\phi(\mathrm{spec}(A))$$

and similarly for B. Consider the event

$$E = \{d_\infty(\mu_{SP}(B), \mu_{SC}) \leq \epsilon\}.$$

In order to continue we need the following lemma.

Lemma 4.1.10. *Let $f : \mathbb{R} \to \mathbb{R}^+$ be an L-Lipschitz function and set $g = (f - L\epsilon)^+$. If $d_{\infty(X,Y)} \leq \epsilon$ then*

$$\mathbb{E}f(Y) \geq \mathbb{E}g(X).$$

Proof. If $\|X - Y\|_{L_\infty} \leq$ then, obviously, $f(X) \geq g(Y)$. So by the definition of the $\infty-$Wasserstein distance the lemma follows. □

Assume for the moment that the event E holds. Then by Lemma 4.1.10 we have

$$\begin{aligned}\|B\|_1 = n\int |x|d\mu_{SP}(B) &\geq n\int_{-2}^{2}(|x| - \epsilon)^+ d\mu_{sp}(x)\\ &\geq n\int_{-2}^{2}(|x| - 1)^+ d\mu_{SC}(x) = na.\end{aligned}$$

Applying Lemma 4.1.5 for $C = \frac{2}{a}$ we get

$$\mathrm{spec}(A) \prec (1 + Cd_\infty(\mu_{SP}(A), \mu_{SP}(B)))\mathrm{spec}(B).$$

But since ϕ is convex and permutation invariant, by the characterization of majorization we get

$$\phi(\mathrm{spec}(A)) \leq (1 + Cd_\infty(\mu_{SP}(A), \mu_{SP}(B)))\phi(\mathrm{spec}(B)).$$

So by taking expectation over A and then over B (recall that we have assumed that they are defined on a product probability space), and taking into consideration the fact that $d_\infty(\mu_{SP}(A), \mu_{SP}(B)) \leq \epsilon + d_\infty(\mu_{SP}(B), \mu_{SC})$, we get

$$\mathbb{E}\phi(\mathrm{spec}(A)) \leq (1 + 2C\epsilon)\mathbb{E}\phi(\mathrm{spec}(B)).$$

So, in the general case we get by independence

$$\mathbb{E}\phi(\mathrm{spec}(B)) \geq \mathbb{E}\phi(\mathrm{spec}(B))\mathbf{1}_E \geq (1 + 2C\epsilon)^{-1}\mathbb{P}(E)\mathbb{E}\phi(\mathrm{spec}(A)).$$

Since $\mathbb{P}(E) \geq 1 - p$, the proof of the inequality is complete. The other inequality follows by symmetry. □

A very similar result is the following.

Proposition 4.1.11. *Let A, B be two $M_n^{sa,0}$-valued random matrices which are unitarily invariant. Assume that*

$$\mathbb{P}(\|A\|_1 \geq c_1 n) \geq 1 - p$$

and

$$\mathbb{E}\|A\|_\infty \leq C_2.$$

Let $K \subseteq M_n^{sa,0}$ be a convex body containing the origin in its interior. Then

$$C^{-1}\mathbb{E}\|A\|_K \leq \mathbb{E}\|B\|_K \leq C\mathbb{E}\|A\|_K,$$

where $C = (1 - p)^{-1}2C_2/c_1$.

Proof. The proof is similar with the proof of the previous lemma. First define A and B in the same probability space. Then, suppose that the event $E = \{\|B\|_1 \geq c_1 n\}$ holds. Finally, use Proposition 4.1.8 and continue with the same method as in the previous proposition. □

4.1.3 Gaussian approximation to induced states

We are going to investigate typical properties of random induced states, in the large dimension regime. Their spectral properties were discussed in previous sections, and are described either by the Marcenko-Pastur distribution (when s is proportional to n) or by the semicircular distribution (when $s \gg n$).

However, we are also interested in properties that cannot be inferred from the spectrum (the main example being separability vs. entanglement on a bipartite system). In this context, it is useful to compare induced states with their Gaussian approximation. Indeed, the Gaussian model allows us to connect with tools from convex geometry, such as the mean width.

It is convenient to work in the hyperplane $M_n^{sa,0}$ and to consider the shifted operators $\rho - I/n$, which we compare with a $\mathrm{GUE}_0(n)$ random matrix. The following proposition compares the expected value of any norm (or gauge) computed for both models. First we give a definition.

Definition 4.1.12. Let H be a Hilbert space and let $A \subseteq H$ be a convex body. We say that $r > 0$ is the inradius of A if it is the largest radius of a Euclidean ball contained in A.

Likewise, we say that r is the outer radius of A if it is the smallest radius of a Euclidean ball that contains A.

Proposition 4.1.13. *Let $n, s \in \mathbb{N}$ and write $\rho_{n,s}$ for a random induced state on $\mathbb{C}^n$ with distribution $\mu_{n,s}$ and G_n for an $n \times n$ GUE_0 random matrix. Let $C_{n,s}$ be the smallest constant such that the following holds: For any convex body $K \in M_n^{sa,0}$ containing 0 in its interior,*

$$C_{n,s}^{-1}\mathbb{E}\left\|\frac{G_n}{n\sqrt{s}}\right\|_K \leq \mathbb{E}\left\|\rho - \frac{I}{n}\right\|_K \leq C_{n,s}\mathbb{E}\left\|\frac{G_n}{n\sqrt{s}}\right\|_K .$$

Then, if (n_k) and (s_k) are two sequences such that $\lim_{k\to\infty} n_k = \lim_{k\to\infty} s_k/n_k = \infty$, we have that $\lim C_{n_k,s_k} = 1$.

Proof. Firstly we will prove the following lemma.

Lemma 4.1.14. *Let X and Y be two $\mathbb{R}^n$-valued random vectors with the property that for any $t \in S^{n-1}$*

$$0 < \mathbb{E}\,|\langle X, t\rangle|, \mathbb{E}\,|\langle Y, t\rangle| < \infty.$$

There exists a constant C depending on (n, X, Y) such that, for any convex body K containing 0 in its interior,

$$\mathbb{E}\|X\|_K \leq C\mathbb{E}\|Y\|_K.$$

Proof. Let Z be the random variable defined by

$$\mathbb{P}(Z = e_i) = \mathbb{P}(Z = -e_i) = \frac{1}{2n},$$

where $\{e_i\}_{i\in[n]}$ is the standard basis in $\mathbb{R}^n$. Assume also that Z is independent from X and Y.

Then, by the assumptions we made for X and by independence we get that

$$\begin{aligned}\mathbb{E}\|X\|_K &\leq \sum_{i=1}^n \mathbb{E}|X_i|\,\|e_i \mathrm{sign}\,(X_i)\|_K = 2n\sum_{i=1}^n \mathbb{E}|X_i|\,\|e_i \mathrm{sign} X_i\|_K \mathbf{1}_{Z=e_i \mathrm{sign} X_i}\\ &= 2n\sum_{i=1}^n \mathbb{E}|X_i|\,\|Z\|_K \mathbf{1}_{Z=e_i \mathrm{sign} X_i} \leq 2n\sum_{i=1}^n \mathbb{E}|X_i|\,\|Z\|_K = 2n\mathbb{E}|X|_1\mathbb{E}\|Z\|_K.\end{aligned}$$

So, we may set $C_1 = 2n\mathbb{E}|X|_1$.

For the second part let $\mathcal{A} = \{\mathbb{E}(Y\mathbf{1}_A)$, where A is measurable$\}$. Note that for any $y \in \mathrm{conv}(\mathcal{A})$ we have $\|y\|_K \leq \mathbb{E}\|Y\|_K$. Note also that the interior of $\mathrm{conv}(\mathcal{A})$ contains 0 in its interior, otherwise there would exist $t_0 \in S^{n-1}$ such that

$$\mathbb{E}|\langle t_0, Y\rangle| = 0,$$

which is a contradiction.

So, there exists $\epsilon > 0$ such that

$$\pm\epsilon e_i \in \mathrm{conv}(\mathcal{A}) \text{ for all } i \in [n].$$

Then $\epsilon\mathbb{E}\|Z\|_K \in \mathrm{conv}(\mathcal{A})$. So, for $C = \frac{1}{C_1\epsilon}$ we get the desired inequality. □

We can now continue with the proof of the proposition. Assume that the sequences $s := s_k$ and $n := n_k$ have the property that both n_k and s_k/n_k tend to infinity. Then, let $A_k = \sqrt{ns}\,(\rho_{n,s} - I/n)$ and $B_k = G_n/\sqrt{n}$.

We will also use the notation

$$X_k = d_\infty(\mu_{sp}(A_k), \mu_{SC})$$

and

$$Y_k = d_\infty(\mu_{sp}(B_k), \mu_{SC}).$$

Firstly note that

$$X_k \leq \|A_k\|_\infty + 2$$

and

$$Y_k \leq \|B_k\|_\infty + 2.$$

We have proven in Theorem 2.1.38, Proposition 2.1.33, Lemma 2.3.19 and Proposition 2.3.18 that the means of $\|A_k\|_\infty$ and $\|B_k\|_\infty$ are bounded by absolute constants and that both A_k and B_k tend almost surely to 2. We will prove that

$$\lim \mathbb{E}X_k = \lim \mathbb{E}Y_k = 0.$$

We write

$$\mathbb{E}\|B_k\|_\infty \leq 2 \quad \Longrightarrow \quad \liminf \mathbb{E}\|B_k\|_\infty \leq 2,$$

and since $\|B_k\|_\infty \to 2$ in probability we have $\liminf \|B_K\|_\infty = 2$. By Fatou's lemma,

$$\liminf \mathbb{E}\|B_k\|_\infty = 2.$$

Now let $f_k = 2 + \|B_k\|_\infty - Y_k$. Applying again Fatou's lemma we get

$$\mathbb{E}\liminf f_k \leq \liminf \mathbb{E}f_k \quad \Longrightarrow \quad \mathbb{E}\limsup Y_k \geq \limsup \mathbb{E}Y_k.$$

Finally, since Y_k converges in probability to zero we get that $\limsup Y_k = 0$. We argue in the same way for X_k.

Note that equivalently if we had only assumed convergence in probability then we could have used Skorohod's theorem.

So we can apply Proposition 4.1.9 for two sequences $\{\epsilon_k\}$ and $\{p_k\}$ with $\epsilon_k \to 0$ and $p_k \to 0$ to conclude the proof. □

Remark 4.1.15. The previous result can be generalised by removing the assumption of $n, s/n \to \infty$ and then one may show that for any $a > 0$ we have $\sup\{C_{n,s} : s \geq an\} < \infty$. For a proof see [7, Chapter 10].

Remark 4.1.16. We emphasize that the quantity $\mathbb{E}\|G_n\|_K$ appearing in Proposition 4.1.13 is exactly the Gaussian mean width of the polar set K°. Indeed, consider G_n in the space $M_n^{sa,0}$ (equipped with the Hilbert-Schmidt scalar product, as always) which is exactly a $\mathrm{GUE}_0(n)$ random matrix. We could have equivalently formulated Proposition 4.1.13 using the usual mean width: if $C'_{n,s}$ denotes the smallest constant such that the inequalities

$$w(K^\circ)\frac{C'^{-1}_{n,s}}{\sqrt{s}} \leq \mathbb{E}\left\|\rho_{n,s} - \frac{I}{n}\right\|_K \leq \frac{C'_{n,s}}{\sqrt{s}} w(K^\circ),$$

then the conclusion of Proposition 4.1.13 holds for $C'_{n,s}$ as well.

4.1.4 Concentration for gauges of induced states

We present a concentration result which is valid for any gauge evaluated on random induced states. We start with some concentration inequalities.

Lemma 4.1.17 (Lévy's lemma)**.** *Let $n > 2$. If $f : S^{n-1} \to \mathbb{R}$ is an $L-$Lipschitz function and M_f is the median of f then, for every $t > 0$,*

$$s^{n-1}(|f - M_f| > t) \leq \exp(-nt^2/2L^2).$$

Proof. Let $A = \{f < M_f\}$ and set $\epsilon = \frac{t}{L}$. Since f is L-Lipschitz it is easy to prove that

$$A_\epsilon \subseteq \{f \leq M_f + t\}.$$

Likewise, one can show that if $B = \{f > M_f\}$ then

$$B_\epsilon \subseteq \{f \geq M_f + t\}.$$

So by Corollary 1.4.6 we get the desired inequality. □

Definition 4.1.18. Let $f : S^{n-1} \to \mathbb{R}$. A value M will be called central value of f if either it is the mean value of f or

$$s^{n-1}(f \geq M) \text{ and } s^{n-1}(f \leq M) \geq \frac{1}{4}.$$

An equivalent way to define the central values is via the first and the third quartile of a random variable Y defined on an probability space $(\Omega, A, \mathbb{P})$.

The first quartile of Y is defined by

$$\mu = \inf\left\{t \in \mathbb{R} : \mathbb{P}(Y \geq t) \geq \frac{1}{4}, \mathbb{P}(Y \leq t) \geq \frac{1}{4}\right\},$$

and the third quartile of Y is defined by

$$M = \sup\left\{t \in \mathbb{R} : \mathbb{P}(Y \geq t) \geq \frac{1}{4}, \mathbb{P}(Y \leq t) \geq \frac{1}{4}\right\}.$$

So a central value of a random variable Y on a probability space $(\Omega, A, \mathbb{P})$ (in particular on the sphere) is defined to be either the mean value of Y or some $t \in \mathbb{R}$ such that $\mu \leq t \leq M$.

The goal is to generalise Lévy's lemma for any central value of a function f defined on the probability metric space $(S^{n-1}, s^{n-1}, g, B(S^{n-1}))$, where g is the geodesic metric.

Proposition 4.1.19. *If f is an L-Lipschitz function with median M_f and M is any central value of f then*

$$|M - M_f| \leq \sqrt{2\log(2)}n^{-1/2}$$

and

$$\mathbb{P}(f \geq M + t) \leq \exp(-nt^2/4L^2).$$

In order to prove this generalisation of Lévy's lemma we need the following lemma.

Lemma 4.1.20. *Let Y be a real random variable and let M be any central value of Y. Let $a \in \mathbb{R}$ and let $A \geq 1/2$ and $\lambda > 0$ be constants such that for any $t > 0$ it is true that*

$$\max\{\mathbb{P}(Y > a + t), \mathbb{P}(Y < a - t)\} \leq A\exp(-\lambda t^2).$$

Then, $|M - a| \leq \sqrt{\lambda^{-1}\log(4A)}$ and consequently. for any $t \geq \sqrt{\lambda\log(4A)}$,

$$\max\{\mathbb{P}(Y > M + t), \mathbb{P}(Y < M - t)\} \leq 4A^2\exp(-\lambda^{-1}t^2/2).$$

Proof. First note that if $|M-a| \leq \sqrt{\lambda^{-1}\log(4A)}$ then for $t \geq \sqrt{\log(4A)\lambda^{-1}}$ by hypothesis we have that

$$\mathbb{P}(Y \geq M+t) \leq \mathbb{P}(Y \geq a - \sqrt{\lambda^{-1}\log(4A)} + t) \leq A\exp(-\lambda(t - \sqrt{\lambda^{-1}\log(4A)})^2).$$

Since for any $c, d \in \mathbb{R}$ it is true that $4cd \leq c^2 + 4d^2$ we then get that

$$\mathbb{P}(Y \geq M+t) \leq 4A^2\exp(-\lambda t^2/2).$$

Likewise, one can show that the quantity $\mathbb{P}(Y \leq M-t)$ is bounded by $4A^2\exp(-\lambda t^2/2)$. So we get the desired inequality.

Thus, in order to prove the lemma we just need to prove that

$$|M-a| \leq \sqrt{\lambda^{-1}\log(4A)}. \tag{4.1.1}$$

Firstly we will prove the inequality (4.1.1) for the mean of Y. For simplicity we will assume that $\lambda = 1$ and by linearity the result will be true in general.

Let Y_0 be a random variable such that $\mathbb{P}(Y_0 \geq t) \leq A\exp(-t^2)$. Then by the properties of the mean we get

$$\begin{aligned}\mathbb{E}Y_0 \leq \mathbb{E}Y_0^+ &= \int_0^\infty \mathbb{P}(Y_0^+ \geq t)dt \leq A\int_0^\infty \exp(-t^2)dt \\ &= A\sqrt{\pi}/2\int_0^\infty f_z(t)dt,\end{aligned}$$

where f_Z is the density function of a random variable $Z \sim N(0, 1/2)$, taking also into account the inequality $\int_u^\infty \exp(-t^2)dt \leq (\sqrt{u^2+1} - u)\exp(-u^2)$ which holds true for any $u \geq 0$. So, for $u = \sqrt{\log^+(A)}$ we get

$$\mathbb{E}Y_0 \leq \int_0^{\sqrt{\log^+ A}} \exp(-t^2)dt + \int_{\sqrt{\log^+(A)}} \leq \sqrt{1+\log^+(A)}.$$

This shows that

$$\mathbb{E}Y_0 \leq \min\{\sqrt{1+\log^+(A)}, A\sqrt{\pi}/2\} \leq \log(4A).$$

Then, setting $Y_0 = Y - a$ and $Y_0 = a - Y$ we get the inequality

$$|\mathbb{E}Y - a| \leq \sqrt{\log(4A)}.$$

Likewise it easy to prove that for a random variable Y_0 such that $\mathbb{P}(Y_0 \geq t) \leq A\exp(-t^2)$, if M_3 is its third quartile then $M_3 \leq \sqrt{\log(A)}$.

So setting $Y_0 = Y - a$, $Y_0 = a - Y$ (and likewise for the first quartile) we get that for any value M between the first and the third quartile of Y it is true that

$$|M - a| \leq \sqrt{\log(A)},$$

which ends the proof. □

Remark 4.1.21. The proof of the previous lemma shows that if a is the median of the random variable then no restrictions on t is needed.

Proof of Proposition 4.1.19. Combining Lemma 4.1.20 with Lévy's lemma we get the desired inequality. □

Corollary 4.1.22 (Lévy's lemma – local version)**.** *Let $f : S^{n-1} \to \mathbb{R}$ and $\Omega \subseteq S^{n-1}$ such that $\mathbb{P}(\Omega) \geq \frac{3}{4}$ and the restriction of f on Ω is L-Lipschitz. Also, let M_f be the median of f. Then, for every $\epsilon > 0$,*

$$\mathbb{P}(|f - M_f| \geq \epsilon) \leq \mathbb{P}(S^{n-1} \setminus \Omega) + 2\exp(-n\epsilon^2/4L^2).$$

The proof of this fact is very similar to the one of Lévy's lemma and is based on it.

Proof. Let $f' = \inf_{y\in\Omega} f(y) + Ld(x, y)$. Then, M is a central value of f'. We split the set into its intersection with the sets $\{f \neq f'\}$ and $\{f = f'\}$ and apply Lévy's lemma in the version of Proposition 4.1.19 to get the desired inequality. □

Lemma 4.1.23. *Let $M_{n,s}(\mathbb{C})$ denote the set of $n \times s$ matrices. Consider the sphere $S_{\text{HS}} \subseteq M_{n,s}$ equipped with the Hilbert-Schmidt norm. Consider also the function*

$$g : M \in S_{\text{HS}} \mapsto M^*M.$$

Let $\Omega_t = \{M \in S_{\text{HS}} : \|M\|_{\text{op}} \leq t\}$. Then the restriction of g onto Ω_t is $2t$-Lipschitz.

Proof. Let $M, N \in \Omega_t$. Then

$$\begin{aligned}\|MM^* - NN^*\|_{\mathrm{HS}} &\leq \|M(M^* - N^*) - (M-N)N)\|_{\mathrm{HS}} \leq (\|M\|_{\mathrm{op}} + \|N\|_{\mathrm{op}})\|M-N\|_{\mathrm{HS}} \\ &\leq 2t\|M-N\|_{\mathrm{op}}.\end{aligned}$$

□

Now we can present the main result of this paragraph.

Proposition 4.1.24. *Let $s \geq n$ and let $K \subseteq D(\mathbb{C}^n)$ be a convex body with inradius r. Let ρ be a random induced state with distribution $\mu_{n,s}$. Let M be the median of $\left\|\rho - \frac{I}{n}\right\|_{K_0}$, where $K_0 = K - \frac{I}{n}$. Then, for every $\epsilon > 0$,*

$$\mathbb{P}\left(\left|\left\|\rho - \frac{I}{n}\right\|_{K_0} - M\right| \geq \epsilon\right) \leq 2\exp(-s) + 2\exp(-n^2 s r^2 \epsilon^2/72).$$

Proof. We have already proved in Lemma 3.2.9 that a random induced state has the same distribution as $\frac{\mathrm{Wishart}(n,s)}{\mathrm{tr}(\mathrm{Wishart}(n,s))}$ or equivalently as a matrix DD^* where D is uniforlmy distributed on the Hilbert-Schmidt sphere of $M_{n,s}$.

So, consider the function $f : S_{HS} \to \mathbb{R}$ defined by

$$f(A) = \left\|AA^* - \frac{I}{n}\right\|_{K_0}.$$

Also, for every $t > 0$, let $\Omega_t = \{A \in S_{\mathrm{HS}} : \|A\|_{\mathrm{op}} \leq t\}$.

The function f is a composition with several properties:

- The map $A \mapsto \|A\|_{K_0}$ is by definition $1/r$-Lipschitz with respect to the Hilbert-Schmidt norm. This fact is true since $\|A\|_{\mathrm{HS}}^{-1} r(A - I/n) \in K_0$ for all $A \in S_{\mathrm{HS}}$.
- The function $A \mapsto A - I/n$ is an isometry for the Hilbert-Schmidt.
- The map $A \mapsto AA^*$ is $2t$-Lipschitz with respect to the Hilbert-Schmidt norm in Ω_t (see Lemma 4.1.23).

From the facts above we obtain that the function f is $2t/r$ Lipschitz with respect to the Hilbert-Schmidt norm on Ω_t.

So we can apply the local version of Lévy's (Lemma 4.1.22). For every $\epsilon > 0$ we have that

$$\mathbb{P}(|f - M_f| \geq \epsilon) \leq \mathbb{P}(S_{HS} \setminus \Omega_t) + 2\exp(-nsr^2\epsilon^2/8t^2).$$

From the fact that D is uniformly distributed on the sphere, the above probabilities can be equivalently expressed as

$$\mathbb{P}(|f - M_f| \geq \epsilon) = \mathbb{P}\left(\left|\left\|AA^* - \frac{I}{n}\right\|_{K_0} - M\right| \geq \epsilon\right)$$

and

$$\mathbb{P}(S_{HS} \setminus \Omega_t) = \mathbb{P}(\|D\|_{op} \geq t).$$

We can now complete the proof of the proposition, using Proposition 2.3.18 and Corollary 3.2.12 (the norm of the matrix can be treated as constant) which imply that

$$\mathbb{P}\left(\|D\|_{op} \geq \frac{1}{\sqrt{n}} + \frac{1+\epsilon}{\sqrt{s}}\right) \leq \exp(-n\epsilon^2).$$

Choosing $\epsilon = \sqrt{t/n}$ we conclude the proof. □

Remark 4.1.25. The previous argument for $t = 1$ shows that the global Lipschitz constant is bounded by $1/r$. Moreover by (4.1.1) we get that any two central values differ by at most $C/r\sqrt{ns}$.

4.2 Separability of random states

Assume now that we work in a bipartite Hilbert space, and for simplicity consider the case of $C^d \otimes C^d$ where both parties play a symmetric role. Throughout this section we write Sep for $\mathrm{Sep}(C^d \otimes C^d)$ and consider random induced states on $C^d \otimes C^d$ with distribution $\mu_{d^2,s}$.

4.2.1 Almost sure Entanglement for low-dimensional environment

In order to prove the main proposition of this subsection we need the following very important theorem.

Theorem 4.2.1. *Let* $H = \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2} \otimes \cdots \mathbb{C}^{d_k}$. *Also let* $n_0 = \prod_{i=1}^{k} d_i - \sum_{i=1}^{k} d_i + k - 1$.

(i) *If* $m > n_0$ *then any* m*-dimensional subspace of* H *contains a (non-zero) product vector.*

(ii) *If* $m \leq n_0$ *then any* m*-dimensional subspace of* H *contains no (non-zero) product vector.*

Proof. We will prove the second part of the theorem (the first assertion can be proved via the projective dimension theorem from algebraic geometry; for proof see [20]).

For simplicity we will prove the theorem when $H = \mathbb{C}^d \times \mathbb{C}^d$ (so $m \leq n_0 = (d-1)^2$)). The general case is similar. The theorem will be proven by probabilistic dimensional counting. First we give some definitions and prove some necessary lemmas.

Definition 4.2.2. We denote by $P(\mathbb{C}^d)$ the *complex projective space* of $\mathbb{C}^d$, i.e., the quotient of $S_{\mathbb{C}^d}$ under the identification of the elements $\phi, \psi \in S_{\mathbb{C}^d}$ if and only if

$$\phi = \exp(i\theta)y \ , \theta \in \mathbb{R}.$$

We also equip $P(\mathbb{C}^d)$ with the following metric (called *Fubini-Study* metric, or *Bures* metric):

$$d([y],[x]) = \arccos|\langle y, x\rangle|.$$

Moreover, if $H = \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$ we consider the *Sagre* variety of H

$$\text{Seg} = \{\phi \otimes \psi \ , \phi \in S_{\mathbb{C}^{d_1}} \ , \psi \in S_{\mathbb{C}^{d_2}}\}.$$

One can show that $\text{Seg} \subseteq P(\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$.

Definition 4.2.3. The space $Gr(k, V)$ is the family of all k-dimensional subspaces of an n-dimensional vector space V. It is called the *k-Grassmann manifold* of V. Since the properties depend only on the dimension of V we will work on the spaces $Gr(k, \mathbb{R}^n)$ and $Gr(k, \mathbb{C}^n)$.

Remark 4.2.4. Let $O(n)$ denote the space of $n \times n$ orthogonal matrices. Fix $0 < k < n$ and consider the canonical action of $O(n)$ to $Gr(k, \mathbb{R}^n)$. We note that the stabilizer sub-group of $O(n)$ that fixes $\mathbb{R}^k$ consists of block matrices of the form

$$\begin{bmatrix} O_1 & 0 \\ 0 & O_2 \end{bmatrix},$$

where $O_1 \in O(k)$ and $O_2 \in O(n-k)$, and hence it can be identified with $O(k) \times O(n-k)$. Since the action of $O(n)$ on $Gr(k, \mathbb{R}^n)$ is transitive, it follows that $Gr(k, \mathbb{R}^n)$ is a homogeneous space for $O(n)$ and can be identified with the quotient space $O(n)/O(k) \times O(n-k)$.

Moreover similar results can be proven for $Gr(k, \mathbb{C}^n)$ and the $n \times n$ unitary matrices, denoted by $U(n)$.

So each Grassmann manifold carries a natural probability measure which can be constructed as the push-forward of the Haar measure on $O(n)$ via the quotient map $O(n) \to O(n)/O(k) \times O(n-k)$, and likewise $U(n) \mapsto U(n)/U(k) \times U(n-k)$.

Definition 4.2.5. Let K be a compact subset of a metric space (M, d). We will say that a finite subset $N \subseteq K$ is an $\epsilon-$net of K if and only if for all $x \in K$ we have that $d(x, N) < \epsilon$. We will write $N(\epsilon, K)$ for the minimal cardinality of an $\epsilon-$net of K.

Proposition 4.2.6. *Let $M \in Gr(k, \mathbb{R}^n)$ or $M \in Gr(k, \mathbb{C}^n)$ equipped with a metric generated by the Shatten p-norm for some $p \in [1, \infty]$. Then, for any $\epsilon \in (0, \operatorname{diam}(M)]$,*

$$\left(\frac{c \operatorname{diam}(M)}{\epsilon}\right)^{\dim M} \leq N(M, \epsilon) \leq \left(\frac{C \operatorname{diam}(M)}{\epsilon}\right)^{\dim M}$$

for some constants $c, C > 0$ independent of n, k, p and ϵ.

Proof. For a proof see [7, Theorem 5.11]. □

Now we can proceed with the proof of the theorem. We are going to work on $P(H)$ with the Bures metric. The ball with center ψ and radius ϵ will be denoted by $B(\psi, \epsilon)$.

Let F be a random m-dimensional sub-vector of H with respect to the Haar measure on $Gr(m, H)$. More concretely, from what was mentioned before, we may assume that $M = U(F_0)$ where U is Haar-distributed on $U(d^2)$ and F_0 is some fixed $m-$dimensional subspace of H.

We are going to prove that the event $D = \{\text{Seg} \cap F = \emptyset\}$ has probability 1. Given $\epsilon > 0$, by Proposition 4.2.6 let M_ϵ be an $\epsilon-$net inside the projective space $P(F_0)$ with $\text{card}(M_\epsilon) \leq (C'/\epsilon)^{2m-2}$. Next let N_ϵ be an ϵ-net inside $P(H)$ such that, again by Proposition 4.2.6, $\text{card}(N_\epsilon) \leq (C'/2\epsilon)^{2d-2}$. We can check that $N_\epsilon^2 := N_\epsilon \otimes N_\epsilon$ is an $2\epsilon-$net inside Seg. Therefore,

$$\begin{aligned}
\mathbb{P}(D^c) &\leq \mathbb{P}(\cup_{\phi \in N_\epsilon^2} B(\phi, 2\epsilon) \cap U(\cup_{\psi \in M_\epsilon} B(\psi, \epsilon)) \neq \emptyset) \\
&\leq \sum_{\phi \in N_\epsilon^2, \psi \in M_\epsilon} \mathbb{P}(B(\phi, 2\epsilon) \cap U(B(\psi, \epsilon) \neq \emptyset) \\
&\leq \sum_{\phi \in N_\epsilon^2, \psi \in M_\epsilon} \mathbb{P}(d(\phi, U\psi) < 3\epsilon).
\end{aligned}$$

But the quantity $\mathbb{P}(d(\phi, U\psi) < 3\epsilon)$ does not depend on ϕ or ψ and is bounded by $(C''\epsilon)^{d^2-2}$ (by the definition of $P(H)$; for a detailed proof see [7, Ex. 5.11]). It follows that

$$\mathbb{P}(D^c) \leq (C''\epsilon)^{2d^2-2} \text{card}(N_\epsilon^2) \text{card}(M_\epsilon) \leq C\epsilon^{2d^2-2-(2m-2)-2(2d-2)}.$$

So, provided that $m \leq (d-1)^2$, the last quantity tends to zero as $\epsilon \to 0$. This shows that the event D^c has probability 0 and as a result D has probability 1. □

The next proposition is just a consequence of the previous theorem.

Proposition 4.2.7. *Let d, s be integers such that $s \leq (d-1)^2$. Then*

$$\mu_{d^2,s}(\text{Sep}) = 0.$$

Proof. Let $S \subseteq \mathbb{C}^d \otimes \mathbb{C}^d$ be the range of ρ (meaning the image of the corresponding matrix transformation). Obviously, S is almost surely s-dimensional. In order for ρ to be separable, S must contain product vectors. But, by the previous theorem, this cannot be true. So,

$$\mathbb{P}(\rho \text{ is separable}) = \mu_{d^2,s}(\text{Sep}) = 0.$$

□

4.2.2 The threshold theorem

In this subsection we can achieve the main goal of this chapter:

> Consider a system of N identical particles (e.g., N qubits) in a random pure state. For some $k \leq N/2$, let A and B be two subsystems, each consisting of k particles. There exists a threshold function $k_0(N)$ which satisfies $k_0(N) \sim N$ as $N \to \infty$ and such that the following holds. If $k \leq k_0(N)$, then with high probability the two subsystems A and B share entanglement. Conversely, if $k > k_0(N)$, then with high probability the two subsystems A and B do not share entanglement.

In order to continue we need the following very important theorem.

Theorem 4.2.8. *Let* $s_0(d) = w(\mathrm{Sep}(\mathbb{C}^d \otimes \mathbb{C}^d)^\circ)^2$ *for some* $d \in \mathbb{N}$. *Then*

$$cd^3 \leq s_0(d) \leq Cd^3 \log^2 d$$

where C,c are absolute constants.

Proof. We will give a sketch of the proof. By Lemma 1.4.23 an equivalent way to express the desired inequality is the following:

$$cd^{7/2} \leq \mathbb{E}\|G\|_{S_0} \leq Cd^{7/2} \log d.$$

Here we consider the equivalence of Corollary 1.5.6, so we work on $\mathrm{Sep}(\mathbb{C}^n)$, where $n = d^2$, $S_0 = \mathrm{Sep}(\mathbb{C}^n) - \frac{\mathbb{I}}{n}$ and G is a $\mathrm{GUE}_0(n)$ matrix.

Now set

$$S_{\mathrm{sym}} = S_0 \cap (-S_0).$$

Obviously, S_{sym} is a symmetric convex body in $M_n^{sa,0}$ containing 0 in its interior. Firstly note that

$$\|G\|_{S_0} \leq \|G\|_{S_{\mathrm{sym}}} = \max\{\|G\|_{S_0}, \|-G\|_{S_0}\} \leq \|G\|_{S_0} + \|-G\|_{S_0}.$$

Since $-G \sim G$, we have

$$\mathbb{E}\|G\|_{S_0} \leq \mathbb{E}\|G\|_{S_{\mathrm{sym}}} \leq 2\mathbb{E}\|G\|_{S_0},$$

which implies that we can work with $\|G\|_{S_{sym}}$ instead.

One may show that $w(S_0) \cong \text{vrad}(S_0) \cong n^{-3/4}$, meaning that after multiplication with appropriate absolute constants these quantities are all comparable [21].

Now notice that the following chain of inequalities is true

$$w(S_{sym}) \leq w(S_0) \cong \text{vrad}(S_0) \leq 2^{-n}\text{vrad}(S_{sym}) \leq w(S_{sym}).$$

Here the first inequality comes from the fact that $S_{sym} \subseteq S_0$, the second one comes from Lemma 1.4.10 and the last one by Uryshon's inequality (Lemma 1.4.22).

So we conclude that

$$w(S_{sym}) \cong n^{-3/4}.$$

Using again Lemma 1.4.23 we get that

$$\mathbb{E}\|G\|_{S^\circ_{sym}} \cong n^{1/4}.$$

Finally, we will show that, for some absolute constants c, C,

$$cn^2 \leq \mathbb{E}\|G\|_{S_{sym}}\mathbb{E}\|G\|_{S^\circ_{sym}} \leq Cn^2 \log n, \tag{4.2.1}$$

which will end the proof.

Let $E \subseteq M_n^{sa,0}$ be the subspace spanned by the operators $\sigma_1 \otimes \sigma_2$, where σ_i $i = 1, 2$ are self-adjoint operators on $\mathbb{C}^d$. Let F be the orthogonal complement of E, i.e.

$$F = \{\sigma \otimes \mathbb{I},\ tr(\sigma) = 0\} \oplus \{\mathbb{I} \otimes \sigma\ ,\ tr(\sigma) = 0\}.$$

Note that $\dim(F) = 2n - 2$.

By Lemma 1.4.28 we have that there exists a linear map $u : M_n^{sa,0} \to M_n^{sa,0}$ such that $u(S_{sym})$ is in the ℓ–position and has the form

$$u = P_E + \mathbf{0} \oplus v,$$

where $v : F \to F$ is a positive semi-definite operator.

By Lemma 1.4.28 and by the ideal property of the ℓ-norm we get that

$$\ell_{S_{sym}}(P_E) = \ell_{S_{sym}}(uP_E) \leq \ell_{S_{sym}}(u)$$

and likewise for $\ell_{S_{\text{sym}}^\circ}$.

Similarly, since $u^{-1} = P_E + \mathbf{0} \oplus v^{-1}$ we get that

$$\ell_{S_{\text{sym}}}(P_E)\ell_{S_{\text{sym}}^\circ}(P_E) \leq \ell_{S_{\text{sym}}}(u)\ell_{S_{\text{sym}}^\circ}(u^{-1}).$$

Moreover, by Theorem 1.4.27 and by the ideal property of the ℓ–norm we get

$$\ell_{S_{\text{sym}}}(P_E)\ell_{S_{\text{sym}}^\circ}(P_E) \leq Cn^2 \log n,$$

where C is an absolute constant.

Now note that for every $A \in S_0$ we have that

$$A \geq -\frac{1}{n} I,$$

which implies that, for every $A \in S_{\text{sym}}$,

$$\|A\|_\infty \leq 1/n,$$

and hence the outer radius of S_{sym} is bounded by $1/\sqrt{n}$.

Moreover, the inradius of S_{sym} is bounded by the inradius of S_0 which (by [21]) is known to be equal to $(n(n-1))^{-1/2}$. So, by the properties of the ℓ–norm and by Lemma 1.4.23 we get

$$\ell_{S_{\text{sym}}}(P_F) = w_G((S_{\text{sym}} \cap F)^\circ) \leq n\sqrt{2n-2} \leq C'n,$$

and likewise

$$\ell_{S_{\text{sym}}^\circ}(P_F) \leq C'',$$

where C' and C'' are both absolute constants.

Finally, by the triangle inequality,

$$w_G(S_{\text{sym}}) = \ell_{S_{\text{sym}}}(I) \leq \ell_{S_{\text{sym}}(P_E)} + \ell_{S_{\text{sym}}}(P_F),$$

and similarly for $w_G(S_{\text{sym}}^\circ)$.

So, we get (4.2.1) which ends the proof. □

We are now ready to present a threshold theorem.

Theorem 4.2.9. *Consider the function* $s_0(d) = w(\mathrm{Sep}(\mathbb{C}^d \otimes \mathbb{C}^d)^\circ)^2$, *which satisfies*

$$cd^3 \leq s_0(d) \leq Cd^3 \log d^3$$

for some absolute constants C, c *as proved in the previous theorem. Then,* $s_0(d)$ *is the threshold between separability and entanglement in the following sense.*

If ρ *is a random state on* $\mathbb{C}^d \otimes \mathbb{C}^d$ *induced by the environment* $\mathbb{C}^s$ *then, for any* $\epsilon > 0$,

(i) *If* $s \leq (1-\epsilon)s_0(d)$ *then we have*

$$\mathbb{P}(\rho \text{ is entangled }) \geq 1 - 2\exp(-c(\epsilon)d^3). \tag{4.2.2}$$

(ii) *If* $s \geq (1+\epsilon)$ *then we have*

$$\mathbb{P}(\rho \text{ is separable}) \geq 1 - 2\exp(-c(\epsilon)s), \tag{4.2.3}$$

where $c(\epsilon)$ *is a constant depending only on* ϵ.

Proof. (ii) Let $\rho_{d^2,s}$ be a random induced state with distribution $\mu_{d^2,s}$. Denote

- $\mathrm{Sep}_0 = \mathrm{Sep} - \frac{I}{d^2}$,
- $f(\rho) = \|\rho - \frac{I}{d^2}\|_{\mathrm{Sep}_0}$,
- $\mathbb{E}_{s,d} = \mathbb{E}f(\rho_{d^2,s})$.

Now fix $\epsilon > 0$ and let s, d be such that $s \geq (1+\epsilon)s_0(d)$. Note that by the assumption on s we have (if we consider a sequence of s, d) $d^2, s/d^2 \to \infty$.

So, if d is appropriately large enough, we can apply Proposition 4.1.13 (i) (in the version given in Remark 4.1.16) to get

$$\mathbb{E}_{d,s} \leq C'_{d^2,s}\frac{w(\mathrm{Sep}_0^\circ)}{\sqrt{s}} \leq \frac{C'_{d^2,s}}{\sqrt{1+\epsilon}},$$

where $C'_{n,s}$ is very close to 1.

Now let $M_{d,s}$ be the median of $f(\rho_{d,s})$. We have already mentioned that the inradius of Sep is $O(d^2)$. So, from Proposition 4.1.24 we get

$$\mathbb{P}(f(\rho_{d,s}) \geq \epsilon + M_{d,s}) \leq 2\exp(-s) + 2\exp(-cs\epsilon^2).$$

Moreover, since $E_{d,s}$ is a central value (see Remark 4.1.25) we get that there exists $h > 0$ depending only on ϵ such that $M_{d,s} + h \leq 1$. Now the proof of (4.2.3) ends if one notices that, for any state ρ,

$$\rho \text{ is separable} \iff \rho \in \text{Sep} \iff \rho - \frac{I}{d^2} \in \text{Sep}_0 \iff f(\rho) \leq 1.$$

Note that for small values of d we can adjust using an appropriate constant.

(i) The proof of (4.2.2) is similar with the proof of (ii) since Proposition 4.1.24 gives a similar bound for $\mathbb{P}(f(\rho_{d,s}) < M_{d,s} - h)$. □

Bibliography

[1] E. P. Wigner, "On the distribution of the roots of certain symmetric matrices," *Annals of Mathematics*, pp. 325–327, 1958.

[2] O. van Gaans, "Probability measures on metric spaces," 2002-2003.

[3] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Mathematics of the USSR-Sbornik*, vol. 1, no. 4, p. 457, 1967.

[4] Z. D. Bai and Y. Q. Yin, "Convergence to the semicircle law," *The Annals of Probability*, vol. 16, no. 2, p. 863{875, 1988.

[5] T. Tao, *Topics in random matrix theory*. American Mathematical Soc., 2012, vol. 132.

[6] J. W. Silverstein *et al.*, "The smallest eigenvalue of a large dimensional wishart matrix," *The Annals of Probability*, vol. 13, no. 4, pp. 1364–1368, 1985.

[7] G. Aubrun and S. J. Szarek, *Alice and Bob meet Banach: the interface of asymptotic geometric analysis and quantum information theory*. American Mathematical Society, 2017.

[8] M. D. De Chiffre, "The haar measure," Ph.D. dissertation, Ph. D. thesis, Department of Mathematical Sciences, University of Copenhagen, 2011.

[9] R. Durrett, *Probability: theory and examples*. Cambridge Univ. Press, 2010.

[10] K. Fan, "On the krein-milman theorem," *Convexity*, vol. 7, pp. 211–220, 1963.

[11] Αριστείδης Κατάβολος, Ταυστικά γινόμενα γραμμικών χώρων και χώρων Hilbert,' *Σημειώσεις Θεωρίας Τελεστών*, 2019-2020. [Online]. Available: https://eclass.uoa.gr/modules/document/file.php/MATH175/1%29%20%CE%A3%CE%B7%CE%BC%CE%B5%CE%B9%CF%8E%CF%83%CE%B5%CE%B9%CF%82%202019-20/tensors.pdf

[12] R. Diestel, "Graph theory. 2005," *Grad. Texts in Math*, vol. 101, 2005.

[13] T. Kemp, "Introduction to random matrix theory," *UCSD Lecture Notes for Math A*, vol. 247, 2013.

[14] Μαρία Μαστροθεοδώρου , Ισοπεριμετρικές ανισότητες για το μέτρο του Gauss," Master's thesis, Μαθηματικό, ΕΚΠΑ, 2018. [Online]. Available: \protect\protect\edefOT1{OT1}\let\enc@update\relax\protect\xdef\OT1/mak/m/n/12{\OT1/mak/m/n/12}\OT1/mak/m/n/12\size@update\enc@update\defOT1{OT1}{https://pergamos.lib.uoa.gr/uoa/dl/frontend/el/browse/2658788#contents}

[15] M. L. Mehta, *Random matrices*. World Publishing Corporation, 2006.

[16] H. Bateman and A. Erdelyi, *Higher transcendental function: based, in part, on notes left by Harry Bateman, and compiled by the staff of the Bateman Manuscript Project*. McGraw-Hill, 1953.

[17] Z. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*. Springer, 2010, vol. 20.

[18] R. Vershynin, *High-dimensional probability: an introduction with applications in data science*. Cambridge University Press, 2018.

[19] U. Haagerup and S. Thorbjørnsen, "Random matrices with complex gaussian entries," *Expositiones Mathematicae*, vol. 21, no. 4, p. 293{337, 2003.

[20] N. R. Wallach, "An unentangled gleason's theorem," *Contemporary Mathematics*, vol. 305, pp. 291–298, 2002.

[21] G. Aubrun, S. J. Szarek, and D. Ye, "Entanglement thresholds for random induced states," *Communications on Pure and Applied Mathematics*, vol. 67, no. 1, p. 129{171, May 2013. [Online]. Available: http://dx.doi.org/10.1002/cpa.21460

[22] D. Χελιώτης, "Ένα δεύτερο μάθημα στις πιθανότητες," 2015.

[23] V. Milman and A. Pajor, "Entropy and asymptotic geometry of non-symmetric convex bodies," *Advances in Mathematics*, vol. 152, pp. 314–335, 06 2000.

[24] A. R. Feier, "Methods of proof in random matrix theory," Ph.D. dissertation, Harvard University, 2012.

[25] B. Collins and I. Nechita, "Random matrix techniques in quantum information theory," *Journal of Mathematical Physics*, vol. 57, no. 1, p. 015215, Jan 2016. [Online]. Available: http://dx.doi.org/10.1063/1.4936880

[26] P. Billingsley, *Probability and measure: third edition*. Shi jie tu shu chu ban gong si Beijing gong si, 2007.

[27] A. Ehrhard, "Symétrisation dans l'espace de gauss." *MATHEMATICA SCANDINAVICA*, vol. 53, pp. 281–301, Dec. 1983. [Online]. Available: https://www.mscand.dk/article/view/12035

[28] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

[29] G. G. Roussas, *Contiguity of probability measures: some applications in statistics*. Cambridge university press, 1972, vol. 63.

[30] O. Kallenberg, *Random measures, theory and applications*. Springer, 2017.

[31] ——, *Foundations of modern probability*. Springer Science & Business Media, 2006.

[32] T. Koshy, *Catalan numbers with applications*. Oxford University Press, 2008.

[33] Y. M. Bishop, S. E. Fienberg, and P. W. Holland, *Discrete multivariate analysis: theory and practice*. Springer Science & Business Media, 2007.

[34] S. Jukna, *Extremal combinatorics: with applications in computer science*. Springer Science & Business Media, 2011.

[35] S. T. Rachev, *Probability metrics and the stability of stochastic models*. John Wiley & Son Ltd, 1991, vol. 269.

[36] G. Pisier, *The volume of convex bodies and Banach space geometry*. Cambridge University Press, 1999, vol. 94.

[37] S. Geman, "A limit theorem for the norm of random matrices," *Ann. Probab.*, vol. 8, no. 2, pp. 252–261, 04 1980. [Online]. Available: https://doi.org/10.1214/aop/1176994775

[38] V. D. Milman and A. Pajor, "Entropy and asymptotic geometry of non-symmetric convex bodies," *Advances in Mathematics*, vol. 152, no. 2, pp. 314–335, 2000.