Grandes déviations et convergence du spectre de matrices aléatoires

Jonathan Husson

Numéro National de Thèse : 2019LYSEN067

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée par
l'Ecole Normale Supérieure de Lyon

École Doctorale N° 512
École Doctorale en Informatique et Mathématiques de Lyon

Discipline : **Mathématiques**

Soutenue publiquement le 05/12/2019, par :
Jonathan HUSSON

Grandes déviations et convergence du spectre de matrices aléatoires

---

Après l'avis de :

Zeitouni, Ofer	Weizmann Institute
Donati-Martin, Catherine	Université de Versailles

Devant le jury composé de :

DONATI-MARTIN, Catherine	Prof. Univ.	Université de Versailles	Rapporteure
MAÏDA, Mylène	Prof. Univ.	Université de Lille	Examinatrice
BORDENAVE, Charles	C.R.	Université d'Aix-Marseille	Examinateur
CHAFAI, Djalil	Prof. Univ.	Université Paris-Dauphine	Examinateur
AUBRUN, Guillaume	M.C.F.	Université Lyon 1	Examinateur
DABROWSKI, Yoann	M.C.F.	Université Lyon 1	Examinateur
GUIONNET, Alice	D.R.	E.N.S. de Lyon	Directrice

A mes grands-parents.

Remerciements

Je tiens ici à remercier toutes les personnes qui m'ont accompagné, aidé et soutenu durant ma thèse et, plus largement, durant toute ma scolarité à L'ENS de Lyon.

Pour commencer, j'aimerai remercier ma directrice Alice Guionnet pour son encadrement régulier et exigeant et ce malgré de lourdes obligations. Je la remercie également pour sa patience et la richesse de son enseignement mathématique. Sans un tel encadrement d'une telle qualité, l'issue de ces trois années de thèse aurait été beaucoup plus incertaine. Je tiens également à remercier nos collaborateurs, Nicholas Cook et Fanny Augeri pour leur temps et leur motivation. J'ai une gratitude particulière pour Fanny pour avoir apporté jusqu'au dernier moment de précieuses corrections à la partie 1.5.

Je remercie également Catherine Donati-Martin et Ofer Zeitouni pour avoir accepté de rapporter ma thèse ainsi que Mylène Maïda, Charles Bordenave, Djalil Chafaî, Yoann Dabrowski et Guillaume Aubrun pour avoir accepté de faire partie de mon jury. Des remerciements particuliers à Ofer Zeitouni pour avoir suggéré quelques ajouts et corrections bienvenues notamment dans les hypothèses de l'article [45]. Un remerciement spécial à Charles Bordenave pour m'avoir donné l'occasion de mon premier exposé devant un public de chercheurs et chercheuses.

Je remercie également tout le personnel et les chercheurs et chercheuses de l'UMPA où j'ai effectué ma thèse. Tout particulièrement, je remercie Mikael De La Salle, mon tuteur de thèse, qui m'a toujours prêté une oreille dans des moments de doutes tant personnels que mathématiques. De toutes nos rencontres, je suis toujours sorti avec l'esprit plus clair.

Au sein du département de mathématiques j'aimerais également remercier Gregory Miermont pour m'avoir confier le cours de probabilités pour la préparation à l'agrégation. Ces quatre années ont été pour moi une opportunité unique de parfaire ma maîtrise pédagogique des probabilités ainsi que d'élaborer mon propre cours sur le sujet. Enfin je remercie tout mes élèves en séance de TD comme en préparation à l'agrégation qui ont eu à subir mes errements de pédagogue débutant et dont la motivation me pousse à toujours m'améliorer.

Je remercie ensuite les secrétaires du laboratoire, Magalie, Virginia, Farah ainsi que les secrétaires du département, Catherine et Jessica pour leur patience et leur compétence et sans qui nos activités de recherche et d'enseignement ne seraient possibles.

Enfin j'aimerai dire ma gratitude à mes collègues et amis doctorants et AGPR de l'UMPA et de l'ICJ : Arthur camarade nanarophile et progueux et sans qui j'aurais raté mon rendez-vous dans la cour du Roi Cramoisi, Léo exilé dans l'Empire du Milieu dont les blagues nous manquent à tous, Alexandre co-bureau de 2 ans et co-chargé de TD, Thomas le Grand Ancien enfin libre, Benoît pour la team premier degré, mes petits frères de thèse Félix et Ronan, Mickael pour nos nombreuses discussions politico-philosophiques, Simon et Hugo pour l'organisation du séminaire, Samuel infiltré dans la perfide Albion mais aussi Valentine, Auguste, Matthieu, Simon, Marie, Hélène, Florian, Mario, Léo D. ...

Je remercie également ma famille. Il y eu durant ces trois ans et quelques plusieurs moments de grande tristesse familiale et d'épreuves délicates mais à chaque fois j'ai pu compter sur leur soutien et leur réconfort. Un remerciement spécial à mes parents et ma sœur Léa pour être toujours là en cas de besoin (et pour avoir relu ce manuscrit).

Enfin, un grand merci aux nombreux amis très chers que je me suis fait durant toute ma scolarité à l'ENS et merci aussi à Bertrand Carro. Merci à la bande d'Oullins : Pierre Perruche, Toto Gégé et Greg Dossa pour les discussions vidéoludiques, musicales et cinématographiques. Merci à Nicolas et Léonard pour des discussions cinéphiles passionnés. Merci à Adrien et Isa pour leur gentillesse, leur hospitalité, leur soutien pendant mon année d'agreg et les discussions Metalgearsolidesques (malheureusement je ne suis pas encre badass). Merci à Lambin pour m'avoir donné la passion du nanar (quoiqu'une certaine soirée à la cinémathèque française ...). Merci à Martin pour m'avoir donné le goût du rock prog et à toute la coloc : Mickaël, Lambert, Ophélie, Camille pour tant de bonnes soirées passées ensemble. Merci à Agathe pour sa gentillesse et sa cynophilie contagieuse. Merci aussi à tout ceux qui durent me supporter en tant que colocataires : Félix et Nicolas rue du Rhône et Victor (dit Hubert), Nicolas, Antonin et Alla Place des Docteurs Mèrieux.

Enfin last but not least : le Foyer. Le Foyer n'est pas un lieu, le Foyer est un concept. Et ce concept a été un centre de ma vie d'étudiant et de doctorant. Je pourrai tenter de remercier toutes les connaissances et amis que j'y ai rencontré comme il se doit mais cela mériterait probablement une autre thèse. Tout d'abord ceux qui sont ou furent RF : Alexandre A. le Grand Ancien sage et placide, Marine B.-K., Elodie B., Arthur C., Ronan D., Margaux G., Damien R. et tant d'autres ... Les plus vieux : Maxime T., Aurélien S., Antoine B. (dit le président), Solène C., Noé R., Audrey Y., David B., Elie G., Xavier O., Florine D., Remy G. et les plus jeunes : Clément B., Joël F., Alice A.A., Henry D., Jeanelle G. ... Il est certain que j'en oublie et pour cela je m'excuse et comme compensation, si d'aventure vous voulez coincher, je suis prêt à être votre quatrième.

Table des matières

Introduction

La théorie des matrices aléatoires a aujourd'hui de nombreuses ramifications à la fois en mathématiques (probabilités libres, aspects combinatoires, physique statistique, ...) et en dehors (ingénierie des télécommunications, physique quantique, ...). Les matrices aléatoires sont pour la première fois étudiées par Wishart dans les années 1930 dans [78] puis elles ont réémergé dans le années 1950, motivées par la modélisation statistique d'atomes au noyaux lourd [76]. C'est au début des années 1960, sous l'impulsion de Wigner, Dyson, Mehta et leurs co-auteurs que les fondements de la théorie des matrices aléatoires sont jetés [60].

Cette théorie connaît de nombreux développements en particulier dans le cas des modèles gaussiens. En effet, via les formules de Weyl, l'invariance unitaire de ces modèles permet d'expliciter la densité jointe des valeurs propres [60]. Dans les années 1990, l'étude de l'espacement des valeurs propres à l'intérieur du spectre via les déterminants de Fredholm fait apparaître des processus determinantaux et l'écart entre le bord du spectre et la plus grande valeur propre converge à renormalisation près vers la distribution de Tracy-Widom [66, 67]. L'étude de processus matriciels dont les coefficients sont browniens fait apparaître le mouvement brownien de Dyson et permet via le calcul d'Itô de formuler des bornes de grandes déviations [26]. En remplaçant le terme gaussien e^{-x^2} par $e^{-V(x)}$ où V est un potentiel, les densités jointes des valeurs propres du GOE et des GUE peuvent également se généraliser en une famille de modèles appelés β-ensembles sur lesquels il existe une abondante littérature. Citons par exemple des résultats de grandes déviations [17], des résultats centraux limites et des analogues discrets qui apparaissent dans certains modèles de pavage [24]. En théorie des nombres, des liens sont également établis entre le problème de Riemann et les espacements des valeurs propres du GUE [55].

Les années 1990 voient également des liens profonds entre matrices aléatoires et analyse fonctionnelle se tisser grâce aux travaux séminaux de Voiculescu [69]. Celui-ci établit que les moments joints de matrices de Wigner indépendantes convergent vers les moments d'éléments semi-circulaires libres d'un espace de probabilité non-commutatif. Cette approche fait entre autre apparaître de nombreuses analogies entre le comportement limite de ces matrices et les variables gaussiennes dans le cas commutatif. Ces liens permettent d'utiliser les outils de la théorie des probabilités libres pour déterminer le comportement de la mesure empirique d'un polynôme en des matrices indépendantes. La notion d'entropie non-commutative qui généralise la notion d'entropie de Shannon au cadre des probabilités libres a également connu nombre d'utilisations dans des problèmes d'analyse fonctionnelle [75]. Toujours dans ce cadre, les techniques de linéarisations de polynômes permettent de prouver la convergence du rayon spectral d'un polynôme de matrices [50] et la notion de liberté par amalgamation sur une sous-algèbre permet de calculer algorithmiquement la mesure limite d'un polynôme auto-adjoint [14, 15].

Bien d'autres développements et résultats existent en théorie de matrices aléatoires. Citons entre autres le problème de la plus petite valeur singulière pour les matrices à coefficients indépendants [61, 32], les lois circulaires [42, 64], les théorèmes de convergence et les principes d'universalité dans les cas non-hermitiens [65, 46] et les phénomènes de queues lourdes [23, 5]. La théorie des matrices aléatoires s'étend aujourd'hui à un très large spectre mathématique qui va de l'algèbre à la physique statistique.

Cette thèse s'intéresse à deux problèmes de la théorie des matrices aléatoires : le problème des grandes déviations de la plus grande valeur propre ainsi que le problème de la convergence de la mesure empirique pour des polynômes non-hermitiens de matrices aléatoires.

Le premier chapitre est consacré aux grandes déviations. Après un rappel des principaux résultats existants, on expose un résultat de grandes déviations pour la plus grande valeur propre dans le cas de certaines matrices de Wigner, de Wishart ainsi que des matrices à profils de variance, qui vérifient un borne sous-gaussienne sur leur transformée de Laplace. Dans ce chapitre, on expose une méthode nouvelle pour obtenir un principe de grandes déviations dans le cas de matrices aux coefficients non-gaussiens en utilisant une intégrale sphérique. On explore également ce qu'il se passe en dehors des hypothèses sur la transformée de Laplace avec un résultat de grandes déviations locales pour des grandes valeurs de la plus grande valeur propre ainsi que pour des valeurs proches du support de la mesure limite. Les résultats de ce chapitre proviennent de l'article [45] accepté dans Annals of Probability ainsi que sur deux autres articles [7, 54] qui seront bientôt pré-publiés sur ArXiv.

Le second chapitre s'intéresse au problème de la convergence de la mesure empirique pour des polynômes non-hermitiens de matrices de Ginibre. On verra que l'obtention de ce type de résultat revient souvent à produire une borne inférieure sur la plus petite valeur singulière de ces polynômes. On expose une telle minoration dans le cas d'un polynôme de Ginibre de degré 2 et on en déduit le résultat de convergence de mesures empirique qui lui est associé. Ces résultats viennent d'un article en progrès [34].

Enfin, cette thèse comporte une annexe de deux parties rappelant respectivement des bases de la théorie des grandes déviations et les liens entre théorie des probabilités libres et mesures empiriques de matrices aléatoires.

Chapitre 1

Grandes déviations pour la plus grande valeur propre de matrices aléatoires

Résumé du chapitre

Dans ce chapitre, on expose une méthode permettant de démontrer des principes de grandes déviations pour la plus grande valeur propre pour plusieurs modèles de matrices aléatoires : des matrices de Wigner, des matrices de Wishart ainsi que des matrices à profils de variance. Ces résultats, qui s'appliquent dans le cas de coefficients indépendants non-gaussiens, reposent néanmoins sur une hypothèse de majoration de la transformée de Laplace.

On rappelle dans la section 1.1 le contexte général concernant les matrices de Wigner et de Wishart ainsi que les densités jointes des valeurs propres dans les cas gaussiens. On revient sur les principaux principes de grandes déviations connus dans la littérature, tant pour les mesures empiriques que pour la plus grande valeur propre, qui dérivent de l'expression de ces densités.

Dans la section 1.2 on présente les résultats de sections 1.3, 1.4 et 1.5 ainsi qu'un survol de leurs démonstrations.

On expose dans la section 1.3 le principal résultat de ce chapitre tiré de [45]. Ce résultat s'applique au cas de matrices de Wigner et Wishart dont les coefficients ne sont pas gaussiens mais vérifient une borne supérieure sur leurs transformées de Laplace. Cette démonstration repose sur l'étude d'intégrales sphériques du type $\mathbb{E}[\exp(N\theta \langle X_N e, e\rangle)]$ (où e est un vecteur unitaire indépendant des X_N). Dans la section 1.4 on généralise ce résultat à des matrices à profils de variance et on voit qu'il est alors nécessaire de faire des hypothèses supplémentaires sur ces profils. Cette généralisation est tirée de [54].

Enfin dans la section 1.5 on étudie le cas qui se présente lorsque l'hypothèse de domination de la transformée de Laplace n'est plus vérifiée. Dans ce cas, en étudiant le comportement de l'intégrale sphérique, on peut néanmoins arriver à des résultats de grandes déviations locales, notamment pour de grandes valeurs de la plus grande valeur propre. Les résultats de cette partie sont tirés de [7].

1.1 Matrices aléatoires et grandes déviations

1.1.1 Matrices de Wigner et de Wishart

Dans cette section, les modèles de matrices aléatoires dont on étudiera le spectre sont auto-adjoints et donc diagonalisables. Pour capturer le comportement global du spectre de matrices aléatoires, on introduit la notion de mesure empirique :

Définition 1.1.1. *Soit M une matrice auto-adjointe $n \times n$. On note $\lambda_1(M) \leq ... \leq \lambda_n(M)$ ses valeurs propres. On appelle* mesure empirique *de M la mesure de probabilité μ_M définie par*

$$\mu_M := \frac{1}{n}\sum_{i=1}^{n}\delta_{\lambda_i}.$$

La mesure empirique d'une matrice aléatoire est donc une mesure de probabilité elle-même aléatoire. Un modèle de matrices aléatoires est une suite de variables aléatoires $(M_N)_{N\in\mathbb{N}}$ telle que M_N soit à valeurs dans $\mathcal{M}_{u_N,v_N}(\mathbb{K})$ ou $\mathbb{K}=\mathbb{R}$ or $\mathbb{C}$ et où u_N, v_N tendent vers $+\infty$. La théorie des matrices aléatoires s'intéresse au comportement asymptotique des quantités algébriques associées aux M_N (valeurs propres, valeurs singulières, vecteurs propres ...) quand N tend vers $+\infty$.

On va rappeler ici les deux modèles historiques de matrices aléatoires ainsi que les résultats de convergence de mesures empiriques et de plus grandes valeurs propres qui leur sont associés. On définit tout d'abord les matrices de Wigner qui sont un modèle de matrices auto-adjointes avec le maximum d'indépendance possible sur les coefficients :

Définition 1.1.2. *Soit $(a_{i,j})_{i<j}$ une double suite de variables aléatoires i.i.d. centrées à valeurs dans $\mathbb{K}$ et $(d_i)_{i\in\mathbb{N}}$ une autre suite de variables aléatoires i.i.d. centrées à valeurs dans $\mathbb{R}$ indépendante de la première. Si $\mathbb{K}=\mathbb{R}$ on suppose $\mathbb{E}[a_{1,2}^2]=1$ et $\mathbb{E}[d_1^2]=2$. Si $\mathbb{K}=\mathbb{C}$ on suppose $\mathbb{E}[|a_{1,2}|^2]=1$ et $\mathbb{E}[d_1^2]=1$.*

On considère alors pour $N\in\mathbb{N}$ la matrice aléatoire X_N de $\mathcal{M}_{N,N}(\mathbb{K})$ définie par :

$$X_N(i,j) = \begin{cases} \frac{a_{i,j}}{\sqrt{N}} & si\ i<j \\ \frac{d_i}{\sqrt{N}} & si\ i=j \\ \frac{a_{i,j}}{\sqrt{N}} & si\ i>j \end{cases} .$$

On dit que X_N est une matrice de Wigner.

On peut comprendre la renormalisation en $\sqrt{N}^{-1}$ de la manière suivante : on considère le moment d'ordre 2 de la mesure empirique $\mu_N=\mu_{X_N}$ en moyenne, c'est-à-dire $\mathbb{E}[\mu_N(x^2)] = N^{-1}\mathbb{E}[\mathrm{Tr}(X_N^2)]$. En développant la trace, on a que la renormalisation en $\sqrt{N}^{-1}$ est la seule qui donne une limite finie à $\mathbb{E}[\mu_N(x^2)]$. En d'autres termes c'est la seule renormalisation possible si l'on veut faire apparaître un comportement limite du spectre qui ne soit pas trivial.

L'autre principal modèle de matrice aléatoire historiquement étudié est le modèle de Wishart défini comme suit :

Définition 1.1.3. *Soit $(a_{i,j})_{i,j\in\mathbb{N}}$ une double suite de variables aléatoires i.i.d. centrées à valeurs dans $\mathbb{K}$. On suppose $\mathbb{E}[|a_{1,1}|^2]=1$. On considère également $(M(N))_{N\in\mathbb{N}}$ une suite d'entiers telle que :*

$$\lim_{N\to\infty}\frac{M(N)}{N} = \alpha \in [1,+\infty[$$

On considère alors la matrice $N\times M(N)$, X_N définie par :

$$X_N(i,j) = \frac{a_{i,j}}{\sqrt{N}}$$

On dit que $W_N = X_N(X_N)^*$ est une matrice de Wishart.

Le résultat fondamental de la théorie des matrices aléatoires porte sur la convergence de la mesure empirique des matrices de Wigner :

Théorème 1.1.4 (Théorème de Wigner)**.** *Soit $(X_N)_{N\in\mathbb{N}}$ une suite de matrices de Wigner. On note μ_N la mesure empirique de X_N. μ_N converge en probabilité vers la mesure semi-circulaire σ définie sur l'intervalle $[-2,2]$ par :*

$$\sigma := \frac{\sqrt{4-x^2}}{2\pi}dx$$

La démonstration de ce résultat se fait tout d'abord en se ramenant au cas d'une matrice dont les coefficients ont des moments bornés puis par une méthode des moments. En effet la moyenne $\mathbb{E}[\mu_N(x^k)]$ du k-ième moment est égale à $N^{-1}\mathbb{E}[\mathrm{Tr}X_N^k]$. On peut alors développer cette expression en :

$$N^{-1}\mathbb{E}[\mathrm{Tr}X_N^k] := \frac{1}{N^{1+k/2}} \sum_{i_1,\dots,i_{k-1}\in[1,N]} \mathbb{E}[a_{i_1,i_2}...a_{i_k,i_1}]$$

A la limite, les termes prépondérants sont ceux tels que le mot $i_1, ..., i_k$ définisse une partition paire sans croisement de $[1, k]$ (c'est à dire une partition P de $[1, k]$ en paires telles que pour $\{a, b\}, \{c, d\} \in P$, on n'a pas $a < c < b < d$). On en déduit alors que les moments impairs convergent vers 0 et que les moments pairs convergent vers les nombres de Catalan. On conclut alors à la convergence presque sûre des moments en prouvant que $\mathrm{Var}(N^{-1}\mathrm{Tr}X_N^k)$ tend vers 0. Grâce à l'équation sur les nombres de Catalan :

$$C_N = \sum_{k=1}^{N-1} C_k C_{N-k-1}$$

on obtient l'équation fonctionnelle suivante sur la transformée de Stieljes G de la mesure limite :

$$G(z) + \frac{1}{G(z)} = z$$

On a donc $G(z) = (z - \sqrt{z^2 - 4})/2$ et on en déduit que la mesure limite est la mesure semi-circulaire. Avec la même méthode, pour les matrices de Wishart, on a le théorème suivant :

Théorème 1.1.5. *Soit $(W_N)_{N\in\mathbb{N}}$ une suite de matrices de Wishart. On note μ_N la mesure empirique de W_N. μ_N converge en probabilité vers la distribution de Marshenko-Pastur de paramètre α notée MP_α définie par sur l'intervalle $[b_-, b_+]$ par :*

$$MP_\alpha := \frac{\sqrt{(x - b_-)(b_+ - x)}}{2\pi x} dx$$

où $b_- = (1 - \sqrt{\alpha})^2$ et $b_+ = (1 + \sqrt{\alpha})^2$

Ces deux résultats contraignent déjà le comportement des valeurs propres dans l'intérieur du support de la mesure limite. En particulier pour $\alpha \in]0, 1[$, $\lambda_{\lfloor \alpha N \rfloor}$ converge vers le quantile de niveau α de σ. Une question naturelle est alors le comportement des valeurs propres extrêmes λ_1 et λ_N, lesquelles peuvent a priori s'éloigner du support de la mesure limite. Dans le cas où les moments des $a_{i,j}$ vérifient une inégalité de la forme $\mathbb{E}[|a_{i,j}|^k] \leq k^{Ck}$, Füredi et Komlòs ont prouvé que cela ne pouvais pas être le cas, c'est à dire que $\lim_N \lambda_N(X_N) = 2$ presque sûrement si X_N est une matrice de Wigner (voir [41]). Les hypothèses de ce résultat ont ensuite été relaxées par Bai et Yin dans le cas Wigner puis par Bai et Sylverstein de le cas Wishart en la simple existence d'un moment d'ordre 4 :

Théorème 1.1.6. *[8] Avec les notations précédentes, si les $a_{i,j}$ et les d_i ont des lois symétriques et $\mathbb{E}[a_{1,2}^4] < +\infty$, on a presque sûrement que $\lambda_1(X_N)$ et $\lambda_N(X_N)$ convergent respectivement vers -2 et 2.*

Théorème 1.1.7. *[11] Avec les notations précédentes, si les $a_{i,j}$ ont des lois symétriques et $\mathbb{E}[a_{1,1}^4] < +\infty$, on a presque sûrement que $\lambda_1(W_N)$ et $\lambda_N(W_N)$ convergent respectivement vers b_- et b_+.*

Maintenant que l'on connaît le comportement limite des mesures empiriques et de la plus grande valeur propre, on peut s'interroger sur la probabilité de dévier de ces limites. On va exposer dans la sous-section qui suit le cas des modèles gaussiens ainsi que les principes de grandes déviations qui leur sont liés.

1.1.2 Modèles gaussiens, formules de Weyl et grandes déviations

Un cas particulier de matrices de Wigner pour lequel le comportement des valeurs propres à N fixé est entièrement compris (dans le sens ou l'on a une expression explicite de la loi jointe des valeurs propres) est le cas où les $a_{i,j}$ sont des variables gaussiennes. Dans ce cas on dit que les matrices X_N appartiennent au GOE (Gaussian Orthogonal Ensemble) dans le cas réel ou au GUE (Gaussian Unitary Ensemble) dans le cas complexe.

Définition 1.1.8. *On considère X_N une matrice de Wigner*

1. *Dans le cas réel, si la distribution des $a_{i,j}$ est $\mathcal{N}(0,1)$ et la distribution des d_i est $\mathcal{N}(0,2)$, on dit que les X_N appartiennent au* Gaussian Orthogonal Ensemble *(GOE).*
2. *Dans le cas complexe, si la distribution des $a_{i,j}$ est telle que les $\Re(a_{i,j})$ et les $\Im(b_{i,j})$ soient i.i.d. de loi $\mathcal{N}(0,1/2)$ et les d_i soient de loi $\mathcal{N}(0,1)$ on dit que les X_N appartiennent au* Gaussian Unitary Ensemble *(GUE).*

L'appellation vient du fait que dans ce cas, la loi des X_N est invariante par l'action de conjugaison du groupe orthogonal $\mathcal{O}_N$, respectivement du groupe unitaire $\mathcal{U}_N$. Une définition alternative peut être énoncée en donnant la densité de la loi de X_N :

Proposition 1.1.9. *Une matrice aléatoire de $\mathcal{S}_N$ (resp. $\mathcal{H}_N$) appartient au GOE (resp. au GUE) si et seulement si sa loi μ^N a pour densité par rapport à la mesure de Lebesgue sur $\mathcal{S}_N$ (resp. $\mathcal{H}_N$) :*

$$\frac{d\mu^N(H)}{dH} = (Z_N^{(\beta)})^{-1} \exp\left(-\beta N \mathrm{Tr} H^2/4\right)$$

où $\beta = 1$ (resp. $\beta = 2$) et Z_N^β la constante de renormalisation.

On peut également définir des ensembles gaussiens dans le cas de matrices de Wishart. On les appelle alors des ensembles de Laguerre :

Définition 1.1.10. *Soit $W_N = X_N(X_N)^*$ une matrice de Wishart.*

1. *Dans le cas réel, si la distribution des $a_{i,j}$ est $\mathcal{N}(0,1)$ on dit que W_N appartient à l'ensemble de Laguerre réel.*
2. *Dans le cas complexe, si la distribution des $a_{i,j}$ est telle que les $\Re(a_{i,j})$ et les $\Im(b_{i,j})$ sont i.i.d. de loi $\mathcal{N}(0,1/2)$, on dit W_N appartient à l'ensemble de Laguerre complexe.*

Comme dans le cas du GOE/GUE, si $W_N = X_N(X_N)^*$ est un matrice de Laguerre, la densité de la loi μ^N de X_N par rapport à la mesure de Lebesgue peut s'écrire :

$$\frac{d\mu^N(X)}{dX} = \exp\left(\frac{-\beta M(N)}{4}\mathrm{Tr} XX^*\right)$$

On remarque également que la loi de ce modèle est invariante par l'action du groupe de conjugaison unitaire. On a même l'invariance de la distribution de X_N par l'action de $\mathcal{O}_N \times \mathcal{O}_{M(N)}$ dans le cas réel et $\mathcal{U}_N \times \mathcal{U}_{M(N)}$ dans le cas complexe donnée par $(U,V)X = U^*XV$. Enfin on rappelle également l'existence de la mesure de Haar sur le groupe $\mathcal{U}_N$.

Proposition 1.1.11. *Soit G un groupe compact, il existe une unique mesure de probabilité borélienne μ sur G telle que pour tout, $g \in G$ et $A \subset G$ mesurable :*

$$\lambda(gA) = \lambda(A)$$

cette mesure est appelée mesure de Haar sur le groupe G. Dans le cas des groupes $\mathcal{O}_N$,$\mathcal{U}_N$, si U_N est distribué selon la mesure de Haar, on parlera de matrice de Haar.

Pour ces trois cas, on peut exploiter l'invariance unitaire pour déterminer de manière explicite la densité jointe des valeurs propres. Dans la suite, on note pour tout $N \in \mathbb{N}$, $\mathcal{U}_N^1$ le groupe orthogonal $\mathcal{O}_N$ et $\mathcal{U}_N^2$ le groupe unitaire $\mathcal{U}_N$. On rappelle ici les formules d'intégration de Weyl qui permettent de calculer l'intégrale d'une fonction matricielle invariante sous l'action du groupe unitaire :

Théorème 1.1.12 (Formules de Weyl)**.** *Pour $x \in \mathbb{R}^n$, on note $\Delta(x) = \prod_{i<j}(x_j - x_i)$.*

— *En notant $\mathcal{H}_N^\beta$ l'espace $\mathcal{S}_N(\mathbb{R})$ pour $\beta = 1$ et $\mathcal{H}_N(\mathbb{C})$ pour $\beta = 2$, on a pour $\beta = 1, 2$ et $\varphi : \mathcal{H}_N^\beta \to \mathbb{C}$ tel que $\varphi(UAU^*) = \varphi(A)$ pour tout $U \in \mathcal{U}_N^{(\beta)}$:*

$$\int_{\mathcal{H}_N^\beta} \varphi(H)dH = \frac{1}{Z_N} \int_{x \in \mathbb{R}^N} |\Delta(x)|^\beta \varphi(x)dx$$

où l'on note pour $x \in \mathbb{R}^n$, $\varphi(x)$ la valeur de $\varphi(H)$ pour n'importe quelle matrice H dont le spectre est x et où dH est la mesure de Lebesgue sur $\mathcal{H}_N^\beta$.

— *Pour $M \geq N$, en notant $\mathcal{M}_{N,M}^{(\beta)}$ l'espace $\mathcal{M}_{N,M}(\mathbb{R})$ pour $\beta = 1$ et $\mathcal{M}_{N,M}(\mathbb{C})$ pour $\beta = 2$, on a pour $\beta = 1, 2$ et $\varphi : \mathcal{M}_{N,M}^\beta \to \mathbb{C}$ tel que $\varphi(UAV) = \varphi(A)$ pour tout $U \in \mathcal{U}_N^\beta, V \in \mathcal{U}_M^\beta$:*

$$\int_{\mathcal{M}_{M,N}^\beta} \varphi(A)dA = \frac{1}{Z'_N} \int_{x \in (\mathbb{R}^+)^N} \varphi(x)|\Delta(x^2)| \prod_{i=1}^N x_i^{\beta(M-N+1)-1} dx$$

où l'on note pour $x \in (\mathbb{R}^+)^n$, $\varphi(x)$ la valeur de $\varphi(A)$ pour n'importe quelle matrice A dont les valeurs singulières sont x, où dA est la mesure de Lebesgue sur $\mathcal{M}_{M,N}^\beta$ et où $x^2 = (x_1^2, ..., x_n^2)$.

— *Pour $N \in \mathbb{N}$ et $\varphi : \mathcal{U}_N \to \mathbb{C}$ une fonction telle que $\varphi(UVU^*) = \varphi(V)$ pour tout $U \in \mathcal{U}_N$:*

$$\int_{\mathcal{U}_N} \varphi(U)dU = \frac{1}{N!(2\pi)^N} \int_{\theta \in [\prime, \in \pi]^\backslash} |\Delta(\exp(i\theta))|^2 \varphi(\exp(i\theta))d\theta$$

où dU est la mesure de Haar sur $\mathcal{U}_N$ et $\exp(i\theta) = (e^{i\theta_1}, ..., e^{i\theta_n})$.

On a les expressions suivantes pour les constantes de renormalisations :

$$Z_N = \frac{u_N}{N!(u_1)^n}, Z'_N = \frac{u_N u_M 2^{\beta p/2}}{p! u_1^N u_{M-N} 2^{\beta pq/2}}$$

où

$$u_n = \prod_{k=1}^n \frac{2(2\pi)^{\beta k/2}}{2^{\beta/2}\Gamma(\beta k/2)}.$$

Ces formules de Weyl sont en fait une application de la formule de la co-aire qui permet lorsque l'on dispose d'une fonction f régulière entre deux variétés M et N d'intégrer une fonction φ de M en fonction des lignes de niveau de f. Pour plus de précisions au sujet de la formule de la co-aire et des démonstrations des formules de Weyl à un plus haut niveau de généralité, on peut se référer à la section 4.1 de [4] et à l'annexe F de [4].

A partir de ces formules, il est aisé de déduire les distributions jointes des valeurs propres pour le GOE, le GUE, les ensembles de Laguerre et les matrices de Haar :

Théorème 1.1.13. *1. Soit X_N une matrice du GOE (resp. du GUE) et $\lambda_1 \leq ... \leq \lambda_N$ ses valeurs propres. La loi de $(\lambda_1, ..., \lambda_N)$ a la densité suivante par rapport à la mesure de Lebesgue :*

$$\frac{1}{Z_N^G} \mathbb{1}_{x_1 \leq ... \leq x_N} |\Delta(x)|^\beta \exp\left(-\frac{\beta N}{4} \sum_{i=1}^N x_i^2\right)$$

2. Soit $W_N = X_N(X_N)^*$ une matrice de Laguerre et soit $0 \leq \lambda_1 \leq ... \leq \lambda_N$ les valeurs propres de W_N. La loi de $(\lambda_1, ..., \lambda_N)$ a la densité suivante par rapport à la mesure de Lebesgue :

$$\frac{1}{Z_N^L}\mathbb{1}_{0<x_1\leq\ldots\leq x_N}|\Delta(x)|^\beta\prod_{i=1}^N\lambda_i^{\frac{\beta}{2}(N-M(N)+1)-1}\exp\left(-\frac{\beta M(N)}{4}\sum_{i=1}^N x_i\right)$$

3. *Soit U_N une matrice de Haar de $\mathcal{U}_N$ et $\exp(i\theta_1),...,\exp(i\theta_n)$ son spectre avec $0\leq\theta_1\leq\ldots\leq\theta_N<2\pi$. La loi de $(\theta_1,...,\theta_n)$ a la densité suivante :*

$$\frac{1}{(2\pi)^N}\mathbb{1}_{0<\theta_1<\ldots<\theta_n<2\pi}|\Delta(\exp(i\theta))|^\beta d\theta.$$

Ces expressions rendent alors possible la preuve de principes de grandes déviations pour les mesures empiriques ainsi que pour la plus grande valeur propre :

Théorème 1.1.14. *[17] Si $(X_N)_{N\in\mathbb{N}}$ est une suite de matrices d'un ensemble Gaussien, la suite des leurs mesures empiriques μ_N vérifie un principe de grandes déviations avec vitesse N^2 et bonne fonction de taux I_β définie par :*

$$I_\beta(\mu)=\frac{\beta}{2}\int\frac{x^2+y^2}{4}-\log|x-y|d\mu(x)d\mu(y)+\frac{\beta}{4}\log\frac{\beta}{2}-\frac{3\beta}{8}.$$

(I_β est éventuellement infinie).

On pourra se référer à l'annexe A.1 pour des rappels sur les bases de la théorie des grandes déviations, en particulier sur ce que l'on entend par principe de grandes déviations et fonction de taux.

Heuristiquement, la fonction de taux de ce théorème se comprend en remarquant que l'on peut réécrire l'expression de la densité de $(\lambda_1,...,\lambda_N)$ de la façon suivante :

$$\frac{1}{Z_N^G}\mathbb{1}_{x_1\leq\ldots\leq x_N}\exp\left(-N^2\left[\frac{\beta}{4}\int x^2d\mu_N(x)-\frac{\beta}{2}\int_{\mathbb{R}^2-\{(x,x)|x\in\mathbb{R}\}}\log|x-y|d\mu_N(x)d\mu_N(y)\right]\right).$$

Toujours en utilisant l'expression de la densité jointe des valeurs propres, on peut déduire un autre principe de grandes déviations pour la plus grande valeur propre de X_N

Théorème 1.1.15. *Si $(X_N)_{N\in\mathbb{N}}$ est une suite de matrices du GOE ou du GUE et $\lambda_{\max}^N$ la plus grande valeur propre de X_N, les lois de $\lambda_{\max}^N$ suivent un principe de grandes déviations avec vitesse N et bonne fonction de taux J_β définie de manière suivante :*

$$\begin{aligned}J_\beta(x)&=\beta\int_2^x\sqrt{t^2-4}dt\ \text{ si } x\geq 2\\&=+\infty\ \text{ si } x<2.\end{aligned}$$

Ces deux derniers résultats peuvent en fait être généralisés à d'autres modèles où, dans la distribution jointe des λ_i, on remplace le terme $\beta\lambda_i^2/4$ par $V(\lambda_i)$ où la fonction V est continue et telle que $\liminf_{|x|\to\infty}V(x)/(\beta'\log|x|)>1$ où $\beta'\geq\beta$ et $\beta'>1$. On appelle de tels modèles des β-ensembles et ils correspondent au valeurs propres d'une matrice aléatoire distribué selon la loi $Z_N^V\exp(-\mathrm{Tr}(V(H)))dH$. Les formules de Weyl permettent également d'obtenir des principes de grandes déviations pour la mesure empirique (voir [53]) et pour la plus grande valeur propre (voir [68]) de matrices de Laguerre.

Théorème 1.1.16. *Soit $W_N=X_N(X_N)^*$ une matrice de Laguerre réelle ($\beta=1$) ou complexe ($\beta=2$). La mesure empirique de W_N suit un principe de grandes déviations avec pour vitesse N^2 et pour bonne fonction de taux sur $\mathcal{M}(\mathbb{R}^+)$:*

$$I(\mu)=\int\int\frac{1}{2\alpha}\left(\frac{x+y}{2}-\frac{\alpha-1}{\alpha}(\log x+\log y)\right)-\frac{\beta}{\alpha^2}\log|x-y|d\mu(x)d\mu(y)+B$$

(où $B\in\mathbb{R}$ est telle que $\min I=0$)

Théorème 1.1.17. *En reprenant les hypothèses du théorème précédent, la plus grande valeur propre de W_N suit un principe de grandes déviations avec vitesse N et bonne fonction de taux sur $\mathbb{R}^+$:*

$$J(x) = \frac{\beta}{4(1+\alpha)} \int_{b_+}^{x} \frac{\sqrt{(y-b_+)(y-b_-)}}{y} dy$$

Enfin dans le cas de matrices unitaires de Haar, le principe de grandes déviations pour la mesure empirique prend la forme suivante :

Théorème 1.1.18. *[52] Soit U_N une matrice de Haar, la mesure empirique de U_N suit un principe de grandes déviations avec pour vitesse N^2 et pour bonne fonction de taux sur $\mathcal{P}(\mathbb{S}^1)$:*

$$I(\mu) = -\int\int \log|x-y| d\mu(x) d\mu(y) + B$$

(où $B \in \mathbb{R}$ est telle que $\min I = 0$*)*

Ces principes de grandes déviations pour la mesure empirique font apparaître la quantité $\Sigma(\mu) = -\int\int \log|x-y|d\mu(x)d\mu(y)$. Dans le case d'une mesure sur $\mathbb{R}$, $\Sigma(\mu)$ est appelée *entropie non-commutative* de μ et on peut voir les résultats de grandes déviations de la mesure empirique pour le GUE et le GOE comme un équivalent matriciel du théorème de Sanov dans le cas d'une suite de variables gaussiennes. Cette entropie non-commutative a été introduite et discutée par Voiculescu dans la série d'articles [70, 71, 73, 72, 74] et elle généralise l'entropie de Shannon pour des variables aléatoires non-commutatives. Elle joue également un rôle dans plusieurs problèmes d'analyse fonctionnelle.

Pour aller plus loin dans l'étude des grandes déviations pour des modèles gaussiens, une quantité qui joue un rôle important est l'intégrale d'Itzykson-Zuber I_N définie pour toutes matrices A, B symétriques réelles pour $\beta = 1$ ou hermitiennes pour $\beta = 2$ par :

$$I_N^{(\beta)}(A, B) = \int_{U \in \mathcal{U}_N^\beta} \exp\left(\mathrm{Tr}(AUBU^*)\right) dU$$

où l'on intègre par rapport à la mesure de Haar sur le groupe $\mathcal{U}_N^\beta$. Pour $\beta = 2$, la valeur exacte de cette intégrale est donnée par la formule d'Harish-Chandra [60, Appendix 5] :

Théorème 1.1.19. *En supposant que les valeurs propres de A et B sont simples et en notant $\lambda_1 < ... < \lambda_N$ le spectre de A et $\lambda'_1 < ... < \lambda'_N$ celui de B :*

$$I_N^{(2)}(A, B) := \frac{\det((e^{\lambda_i \lambda'_j})_{i,j \in [1,N]})}{\Delta(A)\Delta(B)}$$

où $\Delta(A)$ est le déterminant de Vandermonde $\prod_{i<j}(\lambda_j - \lambda_i)$ (idem pour $\Delta(B)$)

Bien qu'explicite, la complexité du déterminant au numérateur rend cette formule peu exploitable dans le cas général.

Ces intégrales permettent d'étudier des déformations des modèles gaussiens. En effet, toujours en utilisant les formules de Weyl, on obtient les densité jointes suivantes pour plusieurs modèles déformés :

— Soit X_N est une matrice du GOE ou du GUE, A_N une matrice hermitienne et $\lambda_1 \leq ... \leq \lambda_N$ les valeurs propres de $X_N + A_N$. La loi de $(\lambda_1, ..., \lambda_N)$ a la densité suivante :

$$\frac{N!}{Z_N(A_N)} I_N\left(\frac{\beta}{2} N A_N, D(x)\right) \mathbb{1}_{x_1 \leq ... \leq x_N} |\Delta(x)|^\beta \exp\left(-\frac{\beta N}{4}\sum_{i=1}^N x_i^2\right)$$

où $D(x) = \mathrm{diag}(x_1, ..., x_n)$

— Si $W_N = X_N(X_N)^*$ est une matrice de Laguerre et T_N une matrice définie positive et $\lambda_1 \leq ... \leq \lambda_N$ les valeurs propres de $T_N W_N T_N$. La loi de $(\lambda_1, ..., \lambda_N)$ a la densité suivante :

$$\frac{N!}{Z'_N(T_{M(N)})} I_N(-M(N)(T_N)^{-2}, D(x)) \mathbb{1}_{0<x_1\leq\ldots\leq x_N}$$
$$|\Delta(x)|^\beta \prod_{i=1}^N \lambda_i^{\frac{\beta}{2}(N-M(N)+1)-1} \exp\left(-\frac{\beta M(N)}{4}\sum_{i=1}^N x_i\right)$$

On voit que les grandes déviations pour ces modèles sont intimement liées aux comportements des intégrales d'Itzykson-Zuber. Ainsi Guionnet et Zeitouni établissent dans [49] en utilisant un principe de grandes déviations pour le modèle $X_N + A_N$ que dans le cas où μ_{A_N} et μ_{B_N} convergent faiblement vers des mesures μ_A et μ_B, $N^{-2}\ln I_N(NA_N, B_N)$ converge vers une limite $I^{(\beta)}(\mu_A, \mu_B)$. De plus, ils se servent de ce résultat pour prouver un principe de grandes déviations pour la mesure empirique d'une matrice de Wishart gaussienne déformée du type $X_N T_N (X_N)^*$.

Le cas où μ_{A_N} converge faiblement et où le rang $M(N)$ de B_N est un $o(N)$ est également étudié entre autres dans [31, 47] sous certaines conditions sur la matrice B_N. En particulier, le cas de la limite de $N^{-1}\log I_N(NA_N, B_N)$ où B_N est de rang 1 est entièrement résolu par Maïda et Guionnet dans [47]. Ce résultat, qui sera rappelé dans la section 1.3 est central dans la démonstration des principe de grandes déviations pour la plus grande valeur propre pour $X_N + B_N$ lorsque X_N est dans le GOE ou le GUE [58] et pour $A_N + U_N B_N U_N$ lorsque U_N est une matrice unitaire de Haar [48].

Parmi les autres principes de grandes déviations pour des modèles gaussien présents dans la littérature, on peut citer par exemple :

— Le cas d'une somme d'une matrice gaussienne et d'une matrice diagonale $X_N + D_N$. En considérant les coefficients gaussiens comme la réalisation au temps 1 d'un mouvement brownien et en utilisant le calcul d'Itô A. Guionnet et T. Cabanal-Duvillard montrent une borne supérieure et une borne inférieure de grandes déviations sur la mesure empirique [26].

— Le cas d'un processus de matrices gaussiennes $t \mapsto H_N(t)$ dont les coefficients sont des mouvements browniens. C. Donati-Martin et M. Maïda prouvent un principe de grandes déviations pour le processus de la plus grande valeur propre de $H_N(t)$ [37].

— Le cas de matrices de Wishart non centrées $W_N = (X_N + A_N)(X_N + A_N)^*$, A. Hardy et A.B.J. Kuijlaars prouvent un principe de grandes déviations pour la mesure empirique [51].

1.1.3 Grandes déviations dans les cas non-gaussiens

Pour des modèles sans invariance unitaire, les λ_i étant des fonctions complexes des coefficients de X_N, les formules explicites pour leurs densités jointes sont rares et d'autres stratégies sont donc nécessaires pour obtenir des principes de grandes déviations. Dans le cas de matrices de Wigner non-gaussiennes, d'autres principes de grandes déviations connus concernent le cas où les distributions des coefficients ont des queues lourdes dans le sens suivant :

Définition 1.1.20. *Soit $a, \alpha \in \mathbb{R}^{+,*}$. On dit qu'une variable aléatoire complexe Y est dans la classe $\mathcal{S}_\alpha(a)$ si*

$$\lim_{t\to\infty} -t^{-\alpha}\log \mathbb{P}[|Y| \geq t] = a$$

et si il existe une mesure de probabilité sur le cercle unité v et $t_0 \geq 0$ telle que pour $t \geq t_0$:

$$\mathbb{P}[Y/|Y| \in U \text{ et } |Y| \geq t] = v(U)\mathbb{P}[|Y| \geq t]$$

En d'autres termes $\mathcal{S}_\alpha(a)$ est la classe des variables aléatoires dont la queue du module se comporte en $\exp(-at^\alpha)$ et telles que l'argument et le module soient indépendants quand ce dernier est suffisamment grand. On a alors les deux résultats suivants, dus respectivement à Bordenave et Caputo pour le premier et Augeri pour le second :

Théorème 1.1.21. *[23] Soit $\alpha \in]0,2[$ et $a, b > 0$. Si (X_N) est une suite de matrices de Wigner réelles ou complexes telle que $a_{1,2}$ soit dans la classe $\mathcal{S}_\alpha(a)$ et d_1 dans la classe $\mathcal{S}_\alpha(b)$. Alors la mesure empirique μ_N de X_N vérifie un principe de grandes déviations de vitesse $n^{1+\alpha/2}$ avec la bonne fonction de taux I suivante :*

$$\begin{aligned} I(\mu) &= \Phi(\nu) \textit{ si } \mu = \sigma \boxplus \nu \\ &= +\infty \textit{ sinon} \end{aligned}$$

où Φ est une bonne fonction de taux.

Ici $\boxplus$ désigne la convolution libre de deux lois de probabilités.

Théorème 1.1.22. *[5] Soit $\alpha \in]0,2[$ et $a, b > 0$. Si (X_N) est une suite de matrices de Wigner réelles ou complexes telle que $a_{1,2}$ soit dans la classe $\mathcal{S}_\alpha(a)$ et d_1 dans la classe $\mathcal{S}_\alpha(b)$. Alors la plus grande valeur propre de X_N vérifie un principe de grandes déviations avec vitesse $N^{\alpha/2}$ et bonne fonction de taux :*

$$J(x) = \begin{cases} cG_\sigma(x)^{-\alpha} \textit{ si } x > 2 \\ 0 \textit{ si } x = 2 \\ +\infty \textit{ si } x < 2 \end{cases}$$

Où c est une constante dépendant uniquement de a, b, α et G_σ est la transformée de Stieljes de la mesure semi-circulaire σ, c'est à dire :

$$G_\sigma(x) = \int_{-2}^{2} \frac{1}{x-t} d\sigma(t)$$

Ces deux résultats reposent sur le fait qu'à l'échelle exponentielle considérée, les grandes déviations sont gouvernées par le comportement d'un petit nombre de grands coefficients de (X_N) à cause de phénomènes de queues lourdes. Plus précisément, on prouve par troncature des coefficients qu'à l'échelle exponentielle, le comportement de X_N est proche de celui de $H_N + C_N$ où H_N est une matrice aléatoire dont les coefficients sont inférieurs à $(\log N)^d/\sqrt{N}$, C_N est une matrice de rang fini et H_N et C_N sont indépendantes.

Le problème des grandes déviations de la mesure empirique pour des matrices de Wigner générales sous-gaussiennes reste ouvert. Dans le cas de distributions vérifiant des inégalités de Sobolev ou étant à support dans le même compact, on a néanmoins des inégalités de concentration. On va ici considérer le modèle suivant de matrice aléatoire :

On se donne $(\omega^R_{i,j})_{1\le i\le j}$ et $(\omega^I_{i,j})_{1\le i<j\le N}$ deux familles indépendantes de variables aléatoires indépendantes et on note $(P^R_{i,j})_{1\le i\le j}$ et $(P^I_{i,j})_{1\le i<j\le N}$ leurs distributions respectives. On considère $A = (A_{i,j})_{1\le i,j\le N}$ une matrice hermitienne déterministe dont les coefficients sont bornés par un certain $a \in \mathbb{R}$. On définit alors $\omega_{i,j} = \omega^R_{i,j} + i\omega^I_{i,j}$ si $i < j$ et $\omega_{i,i} = \omega^R_{i,i}$ et on note $P_{i,j}$ sa loi. On définit enfin X_N par :

$$X_N(i,j) = \begin{cases} \frac{A_{i,j}\omega_{i,j}}{\sqrt{N}} \text{ si } i \le j \\ \frac{A_{j,i}\omega_{j,i}}{\sqrt{N}} \text{ si } i > j \end{cases}$$

Rappelons la définition suivante :

Définition 1.1.23. *On dit qu'une mesure ν sur $\mathbb{R}$ vérifie une inégalité log-Sobolev avec constante c si pour toute fonction f dérivable on a :*

$$\int_R f^2 \log \frac{f^2}{\int_R f^2 dv} dv \le 2c \int |f'|^2 d\nu$$

Remarque 1.1.24. *Une catégorie d'exemple de loi qui vérifie une inégalité log-Sobolev avec constante c est donnée par les lois de la forme $d\nu = Z^{-1}\exp(-V(x))dx$ où $z = \int_{\mathbb{R}} \exp(-V(x))dx$ et où $V : \mathbb{R} \to \mathbb{R} \cup \infty$ est tel que $V(x) - x^2/2c$ est convexe. En particulier une loi gaussienne standard vérifie l'inégalité log-Sobolev avec constante 1. [22]*

Les résultats suivants qui s'appliquent quand les distributions $P^I_{i,j}$ et $P^R_{i,j}$ vérifient des inégalités log-Sobolev ou sont supportées dans le même compact sont dus à A. Guionnet et O. Zeitouni dans [49] :

Théorème 1.1.25. *[49, Theorem 1.1]*

1. *Si il existe un compact K de $\mathbb{C}$ tel que les $P_{i,j}$ soient a support dans K (i.e. tel que $P_{i,j}(K^c) = 0$ pour tout i, j), alors pour toute fonction f Lipschitz et convexe, on a pour tout $\delta > \delta_0(N) := 8|K|\sqrt{\pi}a|f|_{\mathcal{L}}/N$*

$$\mathbb{P}[N^{-1}\,|\mathrm{Tr}(f(X_N)) - \mathbb{E}[\mathrm{Tr}(f(X_N))]| \geq \delta] \leq 4\exp\left(-\frac{N^2(\delta - \delta_0(N))^2}{16|K|^2a^2{|f|_{\mathcal{L}}}^2}\right)$$

2. *Si il existe $c > 0$ tel que les $P^R_{i,j}, P^I_{i,j}$ vérifient une inégalité log-Sobolev avec constante c, alors pour tout $\delta > 0$:*

$$\mathbb{P}[N^{-1}\,|\mathrm{Tr}(f(X_N)) - \mathbb{E}[\mathrm{Tr}(f(X_N))]| \geq \delta] \leq 2\exp\left(-\frac{N^2\delta^2}{8ca^2{|f|_{\mathcal{L}}}^2}\right)$$

où

$$|f|_{\mathcal{L}} := \sup_{x\neq y}\frac{|f(x) - f(y)|}{|x - y|}$$

Dans des problématiques de grandes déviations connexes dans des cas non-gaussiens, on peut également mentionner les résultats de Chatterjee et ses coauteurs sur les grandes déviations pour le comptage de sous-graphe de graphes aléatoires. Le liens entre matrices et graphes aléatoires peut être exprimé ainsi : Si $H = \{V, E\}$ est un graphe fixé dont V est l'ensemble des sommets et $E \subset V^2$ est l'ensemble des arêtes et si $G = \{W, F\}$ est un graphe aléatoire de matrice d'incidence $(A_{i,j})_{i,j\leq N}$, le nombre d'homomorphisme $n(H, G)$ de graphe de H dans G (i.e. les fonctions injectives $\varphi : V \to W$ telles que pour tout i, j tel que $(i, j) \in E$), alors $(\varphi(i), \varphi(j)) \in F$) peut s'exprimer de la façon suivante :

$$n(H, G) = \sum_{i_1,\ldots,i_k\in[1,N]\text{ distincts}}\ \prod_{(i,j)\in F} A_{i,j}$$

En particulier lorsque G est un graphe d'Erdös-Renyi, la matrice d'incidence est alors une matrice dont les coefficients sur-diagonaux sont des Bernouilli indépendantes. Pour plus de précisions sur ces sujets, on pourra se référer par exemple à [28, 30, 38, 29].

Dans la section suivante, on présente les contributions de cette thèse dans le domaine des grandes déviations.

1.2 Présentation des résultats et perspectives

1.2.1 Présentation des résultats de la section 1.3

La section 1.3 de cette thèse expose les résultats de [45]. Il s'agit d'un résultat de grandes déviations pour des matrices de Wigner (réelles où complexes) dont la distribution des coefficients n'est pas gaussienne mais telle que la transformée de Laplace des coefficients soit bornée par la transformée de Laplace du cas gaussien. Le résultat s'étend à des matrices de Wigner dont les coefficients ne sont pas identiquement distribués mais dans un souci de concision, on énonce le résultat dans le cas i.i.d. classique. On va faire les deux hypothèses suivantes sur la loi des coefficients :

Hypothèse 1.2.1 (Hypothèse technique)**.** *On suppose que les lois de $a_{1,2}$ et de d_1, sont*

- *soit toutes les deux à support compact,*
- *soit vérifient une inégalité* log-*Sobolev et il existe $z \in \mathbb{C}^*$ tel que $\Im(za_{1,2})$ soit indépendantes de $\Re(za_{1,2})$.*

Hypothèse 1.2.2 (Hypothèse sous-gaussienne). *Dans le cas réel, on suppose que $Var(d_1) = 2$ et pour tout $t \in \mathbb{R}$:*

$$\mathbb{E}[\exp(ta_{1,2})] \leq \exp(t^2/2) \text{ et } \mathbb{E}[\exp(td_1)] \leq \exp(t^2)$$

.

Dans le cas complexe, on suppose que $\mathbb{E}[\Re(a_{1,2})^2] = \mathbb{E}[\Im(a_{1,2})^2] = 1/2$, $\mathbb{E}[|d_1|^2] = 1$ et pour tout $t \in \mathbb{R}, z \in \mathbb{C}$

$$\mathbb{E}[\exp(\Re(\bar{z}a_{1,2}))] \leq \exp(|z|^2/4) \text{ et } \mathbb{E}[\exp(td_1)] \leq \exp(t^2/2)$$

.

On a alors le résultat suivant :

Théorème 1.2.3. *Soit $(X_N)_N$ une suite de matrices de Wigner (réelles ou complexes) vérifiant les hypothèses 1.2.1 et 1.2.2 et $\lambda_{\max}(X_N)$ la plus grande valeur propre de X_N. Alors $\lambda_{\max}(X_N)$ vérifie un principe de grandes déviations avec une bonne fonction de taux I^β identique au cas du GOE/GUE.*

Un exemple de distribution des coefficients qui vérifie les hypothèses 1.2.1 et 1.2.2 est la distribution de Rademacher dans le cas réel, c'est-à-dire la loi $(\delta_{-1} + \delta_1)/2$ pour les coefficients non-diagonaux et la loi $(\delta_{-\sqrt{2}} + \delta_{\sqrt{2}})/2$ pour les coefficients diagonaux .

La démonstration du théorème 1.2.3 est en fait similaire à celle du théorème de Cramer dans $\mathbb{R}^n$ (qui est reproduite dans un cas simple dans la section A.0.3). La différence est qu'au lieu d'utiliser des transformées de Laplace $\mathbb{E}[\exp(\langle \theta, S_N \rangle)]$, on utilise les intégrales sphérique $\mathbb{E}_{X,e}[\exp(N\theta\langle e, X_N e\rangle)]$ où l'espérance porte sur X_N et e un vecteur uniforme de la sphère unité indépendant de X_N. Il s'agit en fait d'un cas particulier d'intégrale d'Itzykson-Zuber. Si l'on se donne une suite A_N de matrices $N \times N$ et $J_N(\theta, A_N) = \frac{1}{N} \ln \mathbb{E}_e[\exp(N\theta\langle e, A_N e\rangle)]$, Maïda et Guionnet ont prouvé que si μ_{A_N} converge faiblement vers une certaine mesure de probabilité μ, que $\lambda_{\max}(X_N)$ la plus grande valeur propre de A_N converge vers un certain λ et que la plus petite valeur propre $\lambda_{\min}(X_N)$ reste bornée inférieurement alors $J_N(\theta, X_N)$ converge vers une certaine quantité $J(\mu, \theta, \lambda)$. On a de plus uniformité de la convergence en λ pour $d(\mu_{B_N}, \mu) \leq N^{-\kappa}$ où d est la distance de Dudley. Mais dans notre cas on peut s'y ramener car suivant les inégalités de concentration de [49] et l'hypothèse technique on a $\mathbb{P}[d(\mu_{X_N}, \sigma) \geq N^{-\kappa}] = O(e^{-N^{1+\kappa'}})$. De plus, suivant l'hypothèse sous-gaussienne, pour tout $K > 0$, il existe $M > 0$ tel que $\mathbb{P}[\lambda_{\min}(X_N) \leq -M] = O(e^{-KN})$. Posons $\mathcal{A}_{x,M} := \{X \in \mathcal{H}_N^\beta : |\lambda_{\max}(X) - x| \leq \delta, d(\mu_X, \sigma) \leq N^{-\kappa}, ||X|| \leq M\}$ et supposons que $\lim_{N\to\infty} N^{-1} \log \mathbb{E}_{X,e}[\exp(N\theta\langle e, X_N e\rangle)]$ existe et soit égale à un certain $F(\theta)$. On a alors :

$$\begin{aligned}
\mathbb{P}[X_N \in \mathcal{A}_{x,M}] &\leq \mathbb{E}_X\Big[\frac{\mathbb{E}_e[\exp(N\theta\langle e, X_N e\rangle)]}{\mathbb{E}_e[\exp(N\theta\langle e, X_N e\rangle)]} \mathbb{1}_{X_N \in \mathcal{A}_{x,M}}\Big] \\
&\leq e^{-NJ(\sigma,\theta,x)+o(1)} \mathbb{E}_X[\mathbb{E}_e[\exp(N\theta\langle e, X_N e\rangle)] \mathbb{1}_{X_N \in \mathcal{A}_{x,M}}] \\
&\leq e^{-N[J(\sigma,\theta,x)-F(\theta)+o(1)]}
\end{aligned}$$

et donc en optimisant sur $\theta > 0$:

$$\lim_{\delta\to 0} \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}[|\lambda_{\max}(X_N) - x| \leq \delta] \leq -\sup_{\theta>0}[F(\theta) - J(\sigma, \theta, x)]$$

.

Un point crucial est donc de montrer la convergence de $N^{-1} \log \mathbb{E}_{X,e}[\exp(N\theta\langle e, X_N e\rangle)]$ et c'est ici qu'intervient l'hypothèse sous-gaussienne. Celle-ci permet en effet de montrer que la principale contribution provient des vecteurs délocalisés (c'est-à-dire dont les coefficients sont un $o(N^{-1/4})$).

Pour prouver la borne inférieure, comme pour le théorème de Cramer, on considère les mesures exponentielles $\mathbb{P}^\theta$ définies par :

$$d\mathbb{P}^\theta(X_N) = \frac{\mathbb{E}_e[\exp(N\theta\langle e, X_N e\rangle)]}{\mathbb{E}_{X,e}[\exp(N\theta\langle e, X_N e\rangle)]} d\mathbb{P}(X_N)$$

et on veut donc prouver qu'il existe θ tel que $\lim_{N\to\infty} N^{-1}\log \mathbb{P}^\theta[|\lambda_{\max}(X_N) - x| \leq \delta] = 0$. Si pour un vecteur e de la sphère, on définit la mesure $\mathbb{P}^{e,\theta}$ par

$$d\mathbb{P}^{e,\theta}(X_N) = \frac{\exp(N\theta\langle e, X_N e\rangle)}{\mathbb{E}_X[\exp(N\theta\langle e, X_N e\rangle)]} d\mathbb{P}(X_N)$$

.

On a donc :

$$\mathbb{P}^\theta = \int_{e\in\mathbb{S}^{\beta N-1}} f^\theta(e)\mathbb{P}^{e,\theta} de \text{ où } f^\theta(e) = \frac{\mathbb{E}_X[\exp(N\theta\langle e, X_N e\rangle)]}{\mathbb{E}_{X,e}[\exp(N\theta\langle e, X_N e\rangle)]}$$

.

Les mesures $\mathbb{P}^{e,\theta}$ ont alors la densité suivante :

$$\mathbb{P}^{e,\theta} = \prod_{i<j} \frac{\exp(2\theta\sqrt{N}\Re(e_i\overline{e_j a_{i,j}}))}{\mathbb{E}[\exp(2\theta\sqrt{N}\Re(e_i\overline{e_j a_{i,j}}))} \prod_i \frac{\exp(\theta\sqrt{N}\Re(|e_i|^2 d_i))}{\mathbb{E}[\exp(\theta\sqrt{N}\Re(|e_i|^2 d_i))}$$

. Ainsi, sous cette nouvelle loi de probabilité, les coefficients de X_N sont toujours indépendants et si on note $\Lambda_{i,j}(z) = \log\mathbb{E}[\exp(\Re(z\overline{a_{i,j}})]$ et $\Lambda_{i,i}(t) = \log\mathbb{E}[\exp(td_i)]$, leurs espérances sont données par :

$$\mathbb{E}[a_{i,j}] = \nabla\Lambda_{i,j}(2\sqrt{N}\theta e_i\overline{e_j}) = \sqrt{N}\frac{2\theta e_i\overline{e_j}}{\beta} + o(\sqrt{N}\theta e_i\overline{e_j})$$

$$\mathbb{E}[d_i] = \sqrt{N}\frac{2\theta|e_i|^2}{\beta} + o(\theta\sqrt{N}|e_i|^2)$$

soit, si e est tel que $|e_i| \leq N^{-1/4-\epsilon}$,

$$\mathbb{E}[X_N] = \frac{2\theta ee^*}{\beta} + o(1)$$

. En fait X_N est approximativement une perturbation de rang 1 d'une matrice de Wigner, c'est à dire qu'il existe une matrice de Wigner $\tilde{X}_N$ sous $\mathbb{P}^{e,\theta}$ telle que :

$$X_N = \tilde{X}_N + \frac{2\theta ee^*}{\beta} + o(1)$$

. La convergence de la plus grande valeur propre de X_N fait apparaître un phénomène de BBP-transition et sa limite x vérifie si $\theta > 2/\beta$:

$$\lim_{N\to\infty} \langle e, (x - \tilde{X}_N)^{-1})e\rangle = \frac{\beta}{2\theta}$$

. En utilisant un théorème de loi locale sur $\tilde{X}_N$, on a que $\lim_N \langle e, (z - \tilde{X_N})^{-1}e\rangle = G_\sigma(z)$ et donc pour $x > 2$ fixé, en choisissant $\theta_0 = \frac{2}{\beta G_\sigma(x)}$, on a que $\lambda_{\max}(X_N)$ converge vers x pour $\mathbb{P}^{e,\theta_0}$ pourvu que les coordonnées de e soient inférieures à $N^{-1/4-\epsilon}$. En revenant à $\mathbb{P}^\theta$, on prouve de même que $N^{-1}\ln\mathbb{P}^{\theta_0}[|\lambda_{\max}(X_N) - x| \leq \delta] = 0$ et on a donc également la borne inférieure.

Notons, qu'une fois que l'hypothèse 1.2.2 est vérifiée, la fonction de taux ne dépend pas des lois des coefficients (et est donc identique à celle obtenue dans le cas gaussien). Un résultat similaire peut être démontré pour des matrices de Wishart. On effectue alors le même type de raisonnement sur les matrices linéarisées :

$$\begin{pmatrix} 0 & X_N \\ (X_N)^* & 0 \end{pmatrix}$$

.

La principale différence intervient dans le calcul de $F(\theta)$ et dans l'équation que vérifie la limite de $\lambda_{\max}$ sous la mesure exponentielle. Ces points sont abordés plus en détails dans la partie suivante.

1.2.2 Présentation des résultats de la section 1.4

Les résultats de cette partie proviennent de [54]. Dans cette section, on s'intéresse à la même question pour des matrices ayant des profils de variance. Ces matrices sont de la forme $X_N = \Sigma_N \odot W_N$ où W_N est une matrice de Wigner qui vérifie les hypothèses de la section précédente et $(\Sigma_N)_{N\in\mathbb{N}}$ une suite de matrices telles que $\Sigma_N(i,j) = \sigma(i/N, j/N)$ où $\sigma : [0,1]^2 \to \mathbb{R}^+$ et $\sigma(x,y) = \sigma(y,x)$ et $\odot$ est le produit coefficient par coefficient. On considérera deux cas : un cas où σ est une fonction continue et un cas où σ est constante par morceaux dans le sens où il existe $\alpha_1, ..., \alpha_n \in [0,1]$ de somme 1 et $(\sigma_{i,j})_{i,j\in[1,n]}$ tel que

$$\sigma(x,y) = \sigma_{i,j} \text{ pour } x \in [\gamma_{i-1}, \gamma_i[, y \in [\gamma_{j-1}, \gamma_j[$$

où $\gamma_i = \alpha_1 + ... + \alpha_i$. Notons que contrairement au cas de matrices de Wigner et de Wishart, les résultats de grandes déviations que l'on va énoncer sont nouveaux même pour le cas gaussien.

Le résultat pour les σ continus s'obtient par approximation par des σ constants par morceaux, et on s'intéresse donc à ce cas là en premier. On peut alors utiliser le même argument que dans la section précédente concernant la borne supérieure. Les principales différences se trouvent dans le calcul de la fonction $F(\theta)$ et dans l'équation qui détermine la limite de la plus grande valeur propre pour les mesures exponentielles $\mathbb{P}^{e,\theta}$.

Si l'on suppose X_N gaussienne, on a alors en utilisant la transformée de Laplace pour les variables gaussiennes :

$$\begin{aligned}\mathbb{E}_{X,e}[\exp(N\theta\langle e, X_N e\rangle)] &= \mathbb{E}_e\left[\exp\left(N\frac{\theta^2}{\beta}\sum_{i,j=1}^{N}\Sigma_N(i,j)^2|e_i|^2|e_j|^2\right)\right] \\ &= \mathbb{E}_e\left[\exp\left(N\frac{\theta^2}{\beta}\sum_{k,l=1}^{n}\sigma_{k,l}^2\beta_k\beta_l\right)\right]\end{aligned}$$

où $\beta_k = \sum_{i/N\in[\gamma_{k-1},\gamma_k[} |e_i|^2$. En particulier $(\beta_1, ..., \beta_n)$ est un n-uplet qui suit une loi de Dirichlet de paramètre $(\lfloor\beta\gamma_1 N\rfloor/2, (\lfloor\beta\gamma_2 N\rfloor - \lfloor\beta\gamma_1 N\rfloor)/2, ..., (\beta N - \lfloor\beta\gamma_{n-1}N\rfloor)/2)$. En utilisant le fait que $(\lfloor\beta\gamma_i N\rfloor - \lfloor\beta\gamma_{i-1}N\rfloor)/N \to \beta\alpha_i$ on a que la loi de $(\beta_1, ..., \beta_n)$ suit un principe de grandes déviations de fonction de taux $(t_1, ..., t_n) \mapsto \frac{\beta}{2}\sum_i^n \alpha_i(\log(t_i) - \log(\alpha_i))$. Par le lemme de Varadhan, on a donc que

$$F(\theta) = \sup_{\psi_i>0, \sum_i^N \psi_i = 1}\left\{\frac{\theta^2}{\beta}\sum_{k,l=1}^{n}\sigma_{k,l}^2\psi_k\psi_l - \frac{\beta}{2}\sum_k^n \alpha_k(\log(\psi_k) - \log(\alpha_k))\right\}$$

.

L'hypothèse 1.2.2 permet de se ramener au cas gaussien, quitte à considérer cette intégrale sphérique uniquement sur des vecteurs délocalisés. Ayant déterminé l'existence de $F(\theta)$, on peut donc conclure la borne supérieure comme précédemment avec la fonction de taux $I(\sigma, .)$ définie par :

$$I(\sigma, x) = \sup_{\theta\geq 0}(J(\mu_\sigma, \theta, x) - F(\theta))$$

où μ_σ est la mesure limite de la suite X_N.

De plus, si pour θ donné, on pose $(\psi_1^\theta, ..., \psi_n^\theta)$ un argument maximal du problème d'optimisation ci-dessus, on a que les vecteurs e qui ont une contribution dominante à l'intégrale $\mathbb{E}_{X,e}[\exp(N\theta\langle e, X_N e\rangle)]$ sont ceux tels que $(\beta_1, ..., \beta_n)$ est proche de $(\psi_1^\theta, ..., \psi_n^\theta)$. En particulier, pour le calcul de la borne inférieure, on cherche à prouver que pour $x > r_\sigma$ (où r_σ est la borne supérieure du support de la mesure limite μ_σ de X_N), il existe $\theta > 0$ tel que $N^{-1}\log\mathbb{P}^\theta[|\lambda_{\max}(X_N) - x| \leq \delta] = 0$. Comme $\mathbb{P}^\theta = \int_e f^\theta(e)\mathbb{P}^{e,\theta}de$, les contributions dominantes seront celles telles que $(\beta_1, ..., \beta_n)$ est proche de $(\psi_1^\theta, ..., \psi_n^\theta)$. Lorsque l'on écrit à présent l'équation de la limite de la plus grande valeur propre sous $\mathbb{P}^{e,\theta}$, pour de tels e, on a :

$$\det(I_n - D(\theta, z)S) = 0$$

où $S = (\sigma_{i,j}^2)$ et $D(\theta, z) = \mathrm{diag}(\psi_1^\theta m_1(z), ..., \psi_n^\theta m_n(z))$ où $m_1, ..., m_n$ apparaissent comme des transformées de Stieljes liées à la mesure limite de X_N. Pour obtenir à z donné une solution à cette équation, on a alors besoin de supposer la continuité de $D(\theta, z)$ en θ et donc supposer l'existence d'une détermination continue $\theta \mapsto (\psi_1^\theta, ..., \psi_n^\theta)$ de l'argument maximum. On est alors amené a formuler l'hypothèse suivante :

Hypothèse 1.2.4. *Supposons qu'il existe $\theta \mapsto \psi^\theta$ une fonction continue de $\mathbb{R}^+$ dans $\{x \in (\mathbb{R}^+)^n; x_1 + ... + x_n = 1\}$ telle que :*

$$\frac{\theta^2}{\beta} \sum_{i,j=1}^n \sigma_{i,j}^2 \psi_i^\theta \psi_j^\theta + \frac{\beta}{2} \sum_{k=1}^n \alpha_k (\log(\psi_k) - \log(\alpha_k)) =$$
$$\sup_{\psi \in (\mathbb{R}^+)^n, \sum \psi_i = 1} \left\{ \frac{\theta^2}{\beta} \sum_{i,j=1}^n \sigma_{i,j}^2 \psi_i^\theta \psi_j^\theta + \frac{\beta}{2} \sum_{k=1}^n \alpha_k (\log(\psi_k) - \log(\alpha_k)) \right\}.$$

Si cette hypothèse est vérifiée, on a :

Théorème 1.2.5. *Soit $X_N = \Sigma_N \odot W_N$ tel que W_N vérifie les hypothèses 1.2.2 et 1.2.1 et $\Sigma_N(i,j) = \sigma(i/N, j/N)$. Si σ est constant par morceaux et vérifie l'hypothèse 1.2.4 alors la plus grande valeur propre $\lambda_{\max}(X_N)$ vérifie un principe de grandes déviations avec la bonne fonction de taux $I(\sigma, .)$.*

On approxime le cas continu uniformément sur $[0,1]^2$ par des σ constants par morceaux. On a alors besoin d'une condition similaire, excepté que l'on ne considérera pas des vecteurs ψ positif de somme 1 mais des mesures de probabilités sur $[0,1]$. On introduit donc l'hypothèse suivante :

Hypothèse 1.2.6. *Supposons qu'il existe $\theta \mapsto \psi^\theta$ une fonction de $\mathbb{R}^+$ dans $\mathcal{P}([0,1])$ continue pour la topologie faible telle que :*

$$\frac{\theta^2}{\beta} \int \int \sigma^2(x,y) d\psi^\theta(x) d\psi^\theta(y) + \frac{\beta}{2} D(Leb||\psi^\theta)$$
$$= \sup_{\psi \in \mathcal{P}([0,1])} \left\{ \frac{\theta^2}{\beta} \int \int \sigma^2(x,y) d\psi(x) d\psi(y) + \frac{\beta}{2} D(Leb||\psi) \right\}$$

où $D(.||.)$ est la déviation de Kullback-Leibler.

Le terme $D(Leb||\psi)$ qui vaut $\int \log \frac{d\psi}{dx}(x) dx$ si la mesure de Lebesgue est continue par rapport à ψ et ∞ sinon est l'analogue continu du terme $\sum_k^n \alpha_k (\log(\psi_k) - \log(\alpha_k))$.

Grâce à cette hypothèse, on peut énoncer notre résultat dans le cas σ continu :

Théorème 1.2.7. *Soit $X_N = \Sigma_N \odot W_N$ tel que W_N vérifie les hypothèses 1.2.2 et 1.2.1 et $\Sigma_N(i,j) = \sigma(i/N, j/N)$. Si σ est continue et vérifie l'hypothèse 1.2.6 alors la plus grande valeur propre $\lambda_{\max}(X_N)$ vérifie un principe de grandes déviations avec une bonne fonction de taux $I(\sigma, .)$.*

Un exemple de profils qui vérifient les hypothèses 1.2.4 et 1.2.6 est la classe des σ tel que $\mu \mapsto \int \int \sigma^2(x,y) d\mu(x) d\mu(y)$ est concave sur $\mathcal{P}([0,1])$. Dans le cas par morceaux cela s'applique par exemple aux matrices de la forme

$$\begin{pmatrix} aX & bY \\ bY^* & cZ \end{pmatrix}$$

où $a, b, c \in \mathbb{R}^+$ et tels que $a^2 + c^2 - 2b^2 \leq 0$. Dans le cas continu, on peut prendre $\sigma(x,y) = |f(x) - f(y)|$ où f est une fonction continue croissant de $[0,1]$ dans $\mathbb{R}$. On verra également que sans hypothèse sur le profil de variance σ, le résultat précédent est faux sur certains exemples car il ne donne pas une fonction de taux correcte.

1.2.3 Présentation des résultats de la section 1.5

Dans cette section, on essaye de généraliser les résultats de la section 1.3 dans les cas où l'hypothèse 1.2.2 n'est plus vérifiée mais où l'on a encore une borne sous-gaussienne du type $\log \mathbb{E}[\exp(\sqrt{N}\theta X_N(i,j))] \leq A\theta^2/2$. Grossièrement, sous certaines hypothèses sur la loi des coefficients, on montre que pour $A < 2$ et des valeurs de x proches de 2, il existe des bornes de grandes déviations au voisinage de x avec la même fonction de taux que le cas gaussien. On montre également que pour de grandes valeurs de x il existe également des bornes de grandes déviations au voisinage de x mais cette fois la fonction de taux dépend de la loi des coefficients. Ces résultats viennent de [7] et sont le fruit d'une collaboration avec A. Guionnet et F. Augeri.

La généralisation de ces résultats s'opère en étudiant la fonction $F(\theta)$ définie précédemment (quand elle existe). On a tout d'abord le résultat de borne supérieure suivant :

Théorème 1.2.8. *Soit X_N une matrice de Wigner réelle telle que la famille $\{\sqrt{2N}X_N(i,j)\}_{i<j} \cup \{\sqrt{N}X_n(i,i)\}_i$ soit i.i.d. et de loi symétrique par rapport à 0. On suppose que X_N vérifie l'hypothèse 1.2.1 et que $\lambda_{\max}(X_N)$ soit exponentiellement tendue. $\lambda_{\max}(X_N)$ vérifie alors une borne supérieure de grandes déviations avec la bonne fonction de taux I égale à $+\infty$ sur $(-\infty, 2)$ et pour $x \geq 2$:*

$$I(x) = \sup_{\theta \geq 0}[J(\sigma,\theta,x) - \overline{F}(\theta)]$$

où $\overline{F}(\theta) = \limsup_N \frac{1}{N} \log \mathbb{E}_{X,e}[\exp(N\theta\langle e, X_N e\rangle)]$.

Ce résultat est essentiellement la borne supérieure de la section 1.3 dans un cas plus général. Si on suppose de plus que la fonction $F(\theta)$ est définie sur le fermé $\mathcal{C}_\mu$, on a la borne inférieure suivante :

Théorème 1.2.9. *On reprend les hypothèses du théorème précédent et on suppose que pour $\theta \in \mathcal{C}_\mu$.*

$$\overline{F}(\theta) = \limsup_N \frac{1}{N} \log \mathbb{E}_{X_N,e}[\exp(N\theta\langle e, X_N e\rangle)] = \liminf_N \frac{1}{N} \log \mathbb{E}_{X_N,e}[\exp(N\theta\langle e, X_N e\rangle)] = \underline{F}(\theta)$$

.

On pose pour $x \geq 2$, $\Theta_x := \{\theta \geq 0, J(\sigma,\theta,x) - \overline{F}(\theta) = I(x)\}$

Soit $x \geq 2$, supposons qu'il existe $\theta \in \Theta_x \cap \mathcal{C}_\mu$ tel que pour tout $y \geq 2$ et $y \neq x$, $\theta \neq \Theta_y$. Alors pour tout $\delta > 0$:

$$\liminf_N \frac{1}{N} \log \mathbb{P}[|\lambda_{\max}(X_N) - x| \leq \delta] \geq -I(x)$$

.

En effet supposons l'hypothèse vraie pour un certain x avec un certain $\theta > 0$. Pour démontrer la borne inférieure, il suffit de prouver que $\lim_{N\to\infty} \mathbb{P}^\theta[|\lambda_{\max}(X_N) - x| \leq \delta] = 1$. Or pour tout $y > 2$ distinct de x, on a $\theta' \in \Theta_y \backslash \Theta_x$ et avec le même $\mathcal{A}^M_{y,\delta}$ que celui défini à la section 1.2 :

$$\begin{aligned}
\mathbb{P}^\theta[X_N \in \mathcal{A}^M_{y,\delta}] &\leq \frac{\mathbb{E}[1_{X_N \in \mathcal{A}^M_{y,\delta}} I_N(X_N,\theta)]}{\mathbb{E} I_N(X_N,\theta)} \\
&= \frac{\mathbb{E}[1_{X_N \in \mathcal{A}^M_{y,\delta}} \frac{I_N(X_N,\theta')}{I_N(X_N,\theta')} I_N(X_N,\theta)]}{\mathbb{E} I_N(X_N,\theta)} \\
&= e^{-NJ(\sigma,\theta',y) - NF(\theta) + NJ(\sigma,\theta,y) + No(\delta)} \mathbb{E}[1_{X_N \in \mathcal{A}^M_{y,\delta}} I_N(X_N,\theta')] \\
&\leq e^{-NJ(\sigma,\theta',y) - NF(\theta) + NJ(\sigma,\theta,y,) + N\overline{F}(\theta') + No(\delta)}, \\
&= O(e^{-cN})
\end{aligned}$$

avec une certaine constante $c > 0$. Pour $y < 2$, on obtient le même résultat par concentration de la mesure. On a $\mathbb{P}^\theta[|\lambda_{\max}(X_N) - y| \leq \delta] = O(e^{-cN})$ pour $x \neq y$ et donc par tension exponentielle on conclue.

Étant que Θ_x est contenu dans l'ensemble des points critique de $\theta \mapsto J(\sigma, \theta, x) - F(\theta)$ si F est différentiable, la question de la régularité de $F(\theta)$ et du comportement de son sous-différentiel est crucial. En particulier, on a :

Théorème 1.2.10. *Dans les hypothèses du théorème précédent, si pour $x \geq 2$, il existe $\theta \in \Theta_x \cap \mathcal{C}_\mu$ tel que F est différentiable en θ, alors la borne inférieure est vérifiée pour x.*

On s'intéresse donc à l'existence ainsi qu'à la différentiabilité de F. On pose alors :

$$L(x) = \log \mathbb{E}[x\sqrt{N} X_N(1,2)], \psi(x) = L(x)/x^2 \text{ et } A/2 = \sup_{x \in \mathbb{R}^+} \psi(x)$$

.

Pour des valeurs de A inférieures à 2, on a que $[0, 1/(2\sqrt{A-1})] \subset \mathcal{C}_\mu$ et pour $\theta \in [0, 1/(2\sqrt{A-1})]$, $F(\theta) = \theta^2$. En utilisant le théorème précédent, on a le résultat suivant :

Théorème 1.2.11. *Si $A < 2$, la borne inférieure de grandes déviations est vérifiée pour $x \in [2, \frac{1}{\sqrt{A-1}} + \sqrt{A-1}]$. De plus sur cet intervalle, la fonction de taux I est égale à la fonction de taux du cas GOE.*

Pour obtenir l'existence et la différentiabilité de F pour de grandes valeurs de θ, on va faire l'hypothèse suivante :

Hypothèse 1.2.12. *On suppose que $\psi(x) = L(x)/x^2$ est une fonction croissante sur $\mathbb{R}^+$ et a pour limite $B/2$ en $+\infty$.*

Un exemple pour la loi des coefficients qui vérifie l'hypothèse précédente est la loi de $b\gamma$ ou b est une Bernoulli de paramètre p et γ une gaussienne centrée de variance $1/\sqrt{p}$ indépendante. On a alors $B = 1/(2p)$

Avec les notations ci-dessus, on a pour un vecteur e de la sphère :

$$\begin{aligned}
\frac{1}{N} \log \mathbb{E}_X[\exp(N\theta\langle e, X_N e\rangle)] &= \sum_{i<j} \frac{L(2\sqrt{N}\theta e_i e_j)}{N} + \sum_{i=1}^{N} \frac{L(\sqrt{2N}\theta e_i^2)}{N} \\
&= 2\theta^2 \sum_{i \neq j} \psi(2\sqrt{N}\theta e_i e_j) e_i^2 e_j^2 + 2\theta^2 \sum_{i=1}^{N} \psi(\sqrt{2N}\theta e_i^2) e_i^4.
\end{aligned}$$

Lorsque l'on examine ensuite les vecteurs e ayant une contribution dominante dans l'intégrale, l'hypothèse 1.2.12 permet de se ramener au cas de e ayant un coefficient d'ordre 1 et les autres coefficients d'ordre $1/\sqrt{N}$. On peut alors obtenir l'expression suivante :

Théorème 1.2.13. *Supposons l'hypothèse 1.2.12. Pour tout $\theta \geq 0$, $\overline{F}(\theta) = \underline{F}(\theta) = F(\theta)$ et :*

$$F(\theta) = \sup_{\substack{\alpha_1+\alpha_3=1 \\ \alpha_1,\alpha_3 \geq 0}} \sup_{\substack{\nu_1 \in \mathcal{P}(\mathbb{R}) \\ \int x^2 d\nu_1(x) = \alpha_1}} \left\{ \theta^2\alpha_1^2 + B\theta^2\alpha_3^2 + \int L(2\theta\sqrt{\alpha_3}x) d\nu_1(x) - H(\nu_1) - \frac{1}{2}\log(2\pi) - \frac{1}{2} \right\},$$

.

Par changement de variable on a alors :

$$F(\theta) = \sup_{\substack{\alpha \in]0,1], \nu \in \mathcal{P}([0,1]) \\ \int x^2 d\nu(x) = \alpha}} H_\theta(\alpha, \nu).$$

Où pour $\theta \geq 0$, $\alpha \in [0,1]$, et $\nu \in \mathcal{P}(\mathbb{R})$,

$$H_\theta(\alpha, \nu) = \theta^2\alpha^2 + \theta^2 B(1-\alpha)^2 + \int L(2\theta\sqrt{1-\alpha}x) d\nu(x) - H(\nu) - \frac{1}{2}\log(2\pi) - \frac{1}{2}.$$

Avec α fixé, on peut alors optimiser sur ν et on obtient que :

$$\sup_{\substack{\nu \in \mathcal{P}([0,1]) \\ \int x^2 d\nu(x) = \alpha}} H_\theta(\alpha, \nu) = \theta^2(\alpha^2 + B(1-\alpha)^2) \\ + 4\theta^2\alpha(1-\alpha)\zeta_{a,\theta} + G(\zeta_{a,\theta}) - \frac{1}{2}\log(1-\alpha) - \log 2\theta - \frac{1}{2}\log 2\pi - \frac{1}{2}$$

où G est la fonction définie sur $\mathbb{R}^+$ par $G(\zeta) = \log \int \exp(L(x) - \zeta x^2)dx$ et $\zeta_{\alpha,\theta}$ est défini comme étant la solution de $G'(\zeta) = -4\theta^2\alpha(1-\alpha)$ si $4\theta^2\alpha(1-\alpha) \leq -\lim_{\zeta \to B/2} G'(\zeta)$ et $B/2$ sinon. Il suffit donc d'étudier la fonction

$$K_\theta(\alpha) = \theta^2(\alpha^2 + B(1-\alpha)^2) + 4\theta^2\alpha(1-\alpha)\zeta_{a,\theta} + G(\zeta_{a,\theta}) - \frac{1}{2}\log(1-\alpha) - \log 2\theta - \frac{1}{2}\log 2\pi - \frac{1}{2}$$

pour $\alpha \in [0,1]$. En examinant les symétrie de K_θ ainsi que ses points critiques, on a que si $4\theta^2(B-1) \leq 1$, le maximum est alors atteint en $\alpha = 1$ et $F(\theta) = \theta^2$. Si $\theta^2(B-1) > 1$, on a en étudiant K_θ que son maximum sur $[0,1]$ est atteint soit en $\alpha = 1$, soit en $\alpha_\theta \in]0, 1/2[$ l'unique point critique de K_θ sur cet intervalle et ainsi $F(\theta) = \max\{\theta^2, K_\theta(\alpha_\theta)\}$. On établit de plus que $\theta \mapsto \alpha_\theta$ est différentiable, ce qui permet de démontrer que F est bien différentiable pour de grandes valeurs de θ.

On peut alors en déduire le résultat suivant :

Théorème 1.2.14. *Supposons l'hypothèse 1.2.12 vraie. Alors il existe $\theta_0 > 0$ tel que F soit différentiable sur $[1/\sqrt{B-1}, +\infty[$ sauf éventuellement en θ_0. On a de plus l'existence d'une constante $\kappa > 0$ telle que $\theta_0 \leq \kappa/\sqrt{B-1}$. De plus pour $\theta \leq 1/2\sqrt{B-1}$, $F(\theta) = \theta^2$.*

Grâce au premier point du théorème précédent on peut démontrer la borne inférieure de grandes déviations pour de grandes valeurs de x :

Théorème 1.2.15. *Supposons les hypothèse du théorème 1.2.8 et l'hypothèse 1.2.12 vérifiées. Alors il existe une constante universelle κ telle que la borne inférieure de grandes déviations soit vérifiée pour $x \geq \kappa B$ et la bonne fonction de taux I.*

On peut enfin considérer le cas où ψ atteint son maximum en un point de $]0, +\infty[$:

Hypothèse 1.2.16. *On suppose que $\psi(x) = L(x)/x^2$ admet pour limite $B/2$ en $+\infty$, que ψ atteint son maximum en un unique point m_* de $\mathbb{R}^+$ et que $\psi''(m_*) < 0$.*

Dans le cas où cette hypothèse est vérifiée (ce qui est notamment le cas pour des lois μ à support compact qui vérifient la condition d'unicité du maximum de ψ car $B = 0$), les contributions dominantes pour de grandes valeurs de θ proviennent de vecteurs dont environ $N^{1/4}$ coordonnées sont de l'ordre de $N^{-1/4}\sqrt{m_*/\theta}$ et le reste de l'ordre de $N^{-1/2}$. On a alors le résultat suivant :

Théorème 1.2.17. *Supposons l'hypothèse 1.2.16. Il existe θ_0 tel que $[\theta_0, +\infty[\subset \mathcal{C}_\mu$ et tel que sur cet intervalle :*

$$F(\theta) = \sup \left[\theta^2(A-1)\alpha^2 + \theta^2 + \frac{1}{2}\log(1-\alpha)\right]$$

De plus, il existe x_μ tel que pour tout $x \geq x_\mu$ la borne inférieure de grandes déviations est vérifiée.

1.2.4 Perspectives

Une des généralisations de la méthode développée dans la section 1.3 serait de regarder les grandes déviations jointes du k-uplet des k plus grandes valeurs propres de X_N. Le seul obstacle théorique de taille à cette approche est le calcul du comportement asymptotique des intégrales d'Itzykson-Zuber $\mathbb{E}[\exp(N\mathrm{Tr}(D_N(\theta)UA_NV))]$ où $D_N(\theta) = \mathrm{diag}(\theta_1, ..., \theta_k, 0, ...0)$ qui est particulièrement complexe pour $\theta_1, ..., \theta_k$ quelconque.

Concernant les matrices à profils de variances, l'obstacle principal dans le cas général semble être le fait que l'intégrale sphérique ne capture pas la structure naturelle de la matrice. Par exemple, dans le cas d'un profil de variance constant par morceaux, au lieu de prendre e uniforme sur la sphère, il semble être plus judicieux d'essayer de déterminer le comportement asymptotique de l'intégrale sphérique quand les blocs de e ont une norme fixée et sont pris uniforme sur la sphère. Il s'agirait alors de comprendre de quelles quantités les limites de ces intégrales dépendent. D'autre part, on peut également essayer de trouver une expression de la fonction de taux I qui soit une fonction explicite du profil de variance σ. On s'attend en particulier qu'à l'image des cas Wishart et Wigner, dans le cas de σ constant par morceaux cette fonction de taux soit l'intégrale d'une fonction algébrique.

Enfin concernant le cas général développé dans la section 1.5, on peut essayer d'étendre l'étude à d'autres lois μ, notamment les cas où ψ atteint son maximum en $+\infty$ sans pour autant croître. On peut également étudier la différentiabilité de F plus en détail. En particulier, quand la fonction F est bien définie, on peut se poser plusieurs question sur la forme qu'a l'ensemble des points de non-différentiabilité de F, ce qui pourrait permettre d'obtenir une description plus précise du domaine sur lequel la borne inférieure de grandes déviations est vérifiée. Enfin on peut tenter d'appliquer cette même méthode à d'autres modèles de matrices aléatoires pour lesquels on peut étudier la fonction F.

1.3 Large deviations for the largest eigenvalue of Rademacher matrices

1.3.1 Introduction

Very few large deviation principles could be proved so far in random matrix theory. Indeed, the natural quantities of interest such as the spectrum and the eigenvectors are complicated functions of the entries. Hence, even if one considers the simplest model of Wigner matrices which are self-adjoint with independent identically distributed entries above the diagonal, the probability that the empirical measure of the eigenvalues or the largest eigenvalue deviates towards an unlikely value is very difficult to estimate. A well known case where probabilities of large deviations can be estimated is the case where the entries are Gaussian, centered and well chosen covariances, the so-called Gaussian ensembles. In this case, the joint law of the eigenvalues has an explicit form, independent of the eigenvectors, displaying a strong Coulomb gas interaction. This formula could be used to prove a large deviations principle for the empirical measure in [17] and for the largest eigenvalue [16] (see also [68] for further discussions of the Wishart case, and [35]). More recently, in a breakthrough paper, C.Bordenave and P. Caputo [23] tackled the case of matrices with heavy tails, that is Wigner matrices with entries with stretched exponential tails, going to zero at infinity more slowly than a Gaussian tail. The driving idea to approach this question is to show that large deviations are in this case created by a few large entries, so that the empirical measure deviates towards the free convolution of the semi-circle law and the limiting spectral measure of the matrix created by these few large entries. This idea could be also used to grasp the large deviations of the largest eigenvalue by F.Augeri [5]. Generalization to subgraphs counts and the eigenvalues of random graphs are given in [6, 33, 19]. In the Wishart case, [40] considered the large deviations for the largest eigenvalue of very thin Wishart matrices $W = GG^*$, in the regime where the matrix G is $L \times M$ with L much smaller than M. In the case of Bernoulli entries with parameter $p \ll 1$ precise large deviations could be derived recently for the largest and second largest eigenvalues [19, 33]. Hence large deviations for bounded entries, or simply entries with sub-Gaussian tails, remained mysterious in the case of Wigner matrices or Wishart matrices with L of order M. In this article we analyze the large deviations of the largest eigenvalue of Wigner matrices with Rademacher or uniformly distributed random variables. More precisely our result holds for any independent identically distributed entries with distribution with Laplace transform bounded above by the Laplace transform of the Gaussian law with the same variance. We then prove a large deviation principle with the same rate function than in the Gaussian case: large deviations are universal in this class of measures. We show that this result generalizes to complex entries Wigner matrices as well as to Wishart matrices. We are considering the case of general sub-Gaussian entries in a companion paper with F. Augeri. We show in particular that the rate function is different from the rate function of the Gaussian case, at least for deviations towards very large values.

1.3.1.1 Statement of the results

We consider a family of independent real random variables $(a_{i,j}^{(1)})_{0\le i\le j\le N}$, such that the variables $a_{i,j}^{(1)}$ are distributed according to the laws $\mu_{i,j}^N$. We moreover assume that the $\mu_{i,j}^N$ are centered :

$$\mu_{i,j}^N(x)=\int x d\mu_{i,j}^N(x)=0$$

and with covariance:

$$\mu_{i,j}^N(x^2)=\int x^2 d\mu_{i,j}^N(x)=1, \forall 1\le i<j\le N, \qquad \mu_{i,i}^N(x^2)=2, \quad \forall 1\le i\le N\,.$$

We say that a probability measure μ has a sharp sub-Gaussian Laplace transform iff

$$\forall t\in\mathbb{R}, T_\mu(t)=\int \exp\{tx\}d\mu(x)\le \exp\{\frac{t^2\mu(x^2)}{2}\}\,. \tag{1.1}$$

The terminology "sharp" comes from the fact that for t small, we must have

$$T_\mu(t)\ge \exp\{\frac{t^2\mu(x^2)}{2}(1+o(t))\}\,.$$

Then we assume that :

Assumption 1.3.1 (A0). *We assume that the $\mu_{i,j}^N$ satisfy a sharp sub-Gaussian Laplace transform.*

Remark 1.3.2. *From the sub-Gaussian bound, we have the following bound on the moments of $\mu_{i,j}^N$ if Assumption 1.3.1 is verified and X is a random variable of distribution $\mu_{i,j}^N$:*

$$\mathbb{E}[X^{2i}]\le (2i)!(T_{\mu_{i,j}^N}(1)+T_{\mu_{i,j}^N}(1))/2\le (2i)!e^{\mu_{i,j}^N(x^2)/2}$$

and

$$\mathbb{E}[|X|^{2i+1}]\le \mathbb{E}[X^{2i+2}]^{2i+1/2i+2}\le ((2i+2)!e^{\mu_{i,j}^N(x^2)/2})^{2i+1/2i+2}.$$

We have a bound of the form :

$$\mathbb{E}[|X|^k]\le Ck!$$

for some universal constant C. From this bound, we have if the $\mu_{i,j}^N(x^2)$ are bounded, for every $\delta>0$, there exists $\epsilon>0$ that does not depend on the laws $\mu_{i,j}^N$ such that for $|t|\le\epsilon$.

$$T_{\mu_{i,j}^N}(t)\ge \exp\{\frac{(1-\delta)t^2\mu_{i,j}^N(x^2)}{2}\}\,.$$

We have also that the $T_{\mu_{i,j}^N}$ are uniformly C^3 in a neighborhood of the origin: for $\epsilon>0$ small enough $\sup_{|t|\le\epsilon}\sup_{i,j,N}|\partial_t^3\ln T_{\mu_{i,j}^N}(t)|$ is finite.

Remark 1.3.3. *We could assume a weaker upper bound on the Laplace transform for the diagonal entries such as the existence of A finite such that*

$$\int e^{tx} d\mu_{i,i}^N(x) \le \exp\{t^2 + A|t|\}, \quad \forall 1 \le i \le N,$$

see the proof of Theorem 1.3.22.

Example 1.3.4.
1. *Clearly a centered Gaussian variable has a sharp sub-Gaussian Laplace transform.*
2. *The Rademacher law $B = \frac{1}{2}(\delta_{-1} + \delta_1)$ satisfies a sharp sub-Gaussian Laplace transform since for all real number t*
$$T_B(t) = \cosh(t) \le e^{t^2/2}.$$
3. *U, the uniform law on the interval $[-\sqrt{3}, \sqrt{3}]$, satisfies a sharp sub-Gaussian Laplace transform since we have*
$$\int x^2 dU(x) = 1,$$
and
$$T_U(t) = \frac{1}{t\sqrt{3}} \sinh(t\sqrt{3}) = \sum_{n \ge 0} \frac{t^{2n} 3^n}{(2n+1)!}.$$
Since for all $n \ge 0$, $\frac{3^n}{(2n+1)!} \le \frac{1}{2^n n!}$, it follows that $T_U(t) \le e^{\frac{t^2}{2}}$.
4. *If X, Y are two independent variables with distribution μ and μ', two probability measures which have a sharp sub-Gaussian Laplace transform, for any $a \in [0,1]$, the distribution of $\sqrt{a}X + \sqrt{1-a}Y$ has a sharp sub-Gaussian Laplace transform.*

Note that many measures do not have a sharp sub-Gaussian Laplace transform, e.g. the sparse Gaussian law obtained by multiplying a Gaussian variable by a Bernoulli variable, or the well chosen sum of Rademacher laws. We will also need that the empirical measure of the eigenvalues concentrates in a stronger scale than N, see Lemma 1.3.16. To this end we will also make the following classical assumptions to use standard concentration of measure tools.

Assumption 1.3.5. *There exists a compact set K such that the support of all $\mu_{i,j}^N$ is included in K for all $i, j \in \{1, \dots, N\}$and all integer number N, or all $\mu_{i,j}^N$ satisfy a log-Sobolev inequality with the same constant c independent of N. More precisely, in the later case and when the entries are complex, we assume that they are of the form $z(x+iy)$ with a complex number z, and independent real variables x, y which satisfy log-Sobolev inequality with the same constant c independent of N.*

Remark 1.3.6. *All the examples of Example 1.3.4 satisfy Assumption 1.3.5, except possibly for sums of Gaussian variables and bounded entries.*

We then construct for all $N \in \mathbb{N}$, a real Wigner matrix $N \times N$ $X_N^{(1)}$ by setting :

$$X_N^{(1)}(i,j) = \begin{cases} \frac{a_{i,j}^{(1)}}{\sqrt{N}} \text{ when } i \le j, \\ \frac{a_{j,i}^{(1)}}{\sqrt{N}} \text{ when } i > j. \end{cases}$$

We denote $\lambda_{min}(X_N^{(1)}) = \lambda_1 \le \lambda_2 \cdots \le \lambda_N = \lambda_{\max}(X_N^{(1)})$ the eigenvalues of $X_N^{(1)}$. It is well known [77] that under our hypotheses the empirical distribution of the eigenvalues $\hat{\mu}^N_{X_N^{(1)}} = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}$ converges weakly towards the semi-circle distribution σ: for all bounded continuous function f

$$\lim_{N\to\infty} \int f(x) d\hat{\mu}^N_{X_N^{(1)}}(x) = \int f(x) d\sigma(x) = \frac{1}{2\pi} \int_{-2}^{2} f(x)\sqrt{4-x^2} dx \qquad a.s.$$

It is also well known that the eigenvalues stick to the bulk since we assumed the entries have sub-Gaussian moments [41, 4] :

$$\lim_{N\to\infty} \lambda_{min}(X_N^{(1)}) = -2 \qquad \lim_{N\to\infty} \lambda_{\max}(X_N^{(1)}) = 2, \qquad a.s.$$

Our main result is a large deviation principle from this convergence.

Theorem 1.3.7. *Suppose Assumptions 1.3.1 and 1.3.5 hold. Then, the law of the largest eigenvalue $\lambda_{\max}(X_N^{(1)})$ of $X_N^{(1)}$ satisfies a large deviation principle with speed N and good rate function $I^{(1)}$ which is infinite on $(-\infty, 2)$ and otherwise given by*

$$I^{(1)}(\rho) = \frac{1}{2}\int_2^\rho \sqrt{x^2-4}dx\,.$$

In other words, for any closed subset F of $\mathbb{R}$,

$$\limsup_{N\to\infty} \frac{1}{N} \ln P\left(\lambda_{\max}(X_N^{(1)}) \in F\right) \leq -\inf_F I^{(1)}\,,$$

whereas for any open subset O of $\mathbb{R}$

$$\liminf_{N\to\infty} \frac{1}{N} \ln P\left(\lambda_{\max}(X_N^{(1)}) \in O\right) \geq -\inf_O I^{(1)}\,.$$

The same result holds for the opposite of the smallest eigenvalue $-\lambda_{min}(X_N^{(1)})$.

Therefore, the large deviations principles are the same as in the case of Gaussian entries as soon as the entries have a sharp sub-Gaussian Laplace transforms and are bounded, for instance for Rademacher variables or uniformly distributed variables. Hereafter we show how this result generalizes to other settings. First, this result extends to the case of Wigner matrices with complex entries as follows. We now consider a family of independent random variables $(a_{i,j}^{(2)})_{1\leq i\leq j\leq N}$, such that the variables $a_{i,j}^{(2)}$ are distributed according to a law $\mu_{i,j}^N$ when $i \leq j$, which are centered probability measures on $\mathbb{C}$ (and on $\mathbb{R}$ if $i=j$). We write $a_{i,j}^{(2)} = x_{i,j} + iy_{i,j}$ where $x_{i,j} = \Re(a_{i,j}^{(2)})$ and $y_{i,j} = \Im(a_{i,j}^{(2)})$. We suppose that for all $i \in [1,N]$, $y_{i,i} = 0$. In this context, for a probability measure on $\mathbb{C}$, we will consider its Laplace transform to be the function

$$T_\mu(z) := \int \exp\{\Re(a\bar{z})\}d\mu(a)\,.$$

We assume that

Assumption 1.3.8 (A0c)**.** *For all $i<j$*

$$\forall t \in \mathbb{C}, T_{\mu_{i,j}^N}(t) \leq \exp(|t|^2/4)\,,$$

and for all i

$$\forall t \in \mathbb{R}, T_{\mu_{i,i}^N}(t) \leq \exp(t^2/2)\,.$$

The Remark 1.3.2 stays true in the complex framework. Observe that the above hypothesis implies that for all $i<j$, $2\mathbb{E}[x_{i,j}^2] = 2\mathbb{E}[y_{i,j}^2] = \mathbb{E}[x_{i,i}^2] = 1$ and $\mathbb{E}[x_{i,j}y_{i,j}] = 0$. Examples of distributions satisfying Assumption 1.3.8 are given by taking $(x_{i,j}, y_{i,j})$ centered independent variables with law satisfying a sharp sub-Gaussian Laplace transform. Hereafter, we extend naturally Assumption 1.3.5 by assuming that the compact K is a compact subset of $\mathbb{C}$ or log-Sobolev inequality holds in the complex setting.

We then construct for all $N \in \mathbb{N}$, $X_N^{(2)}$ a complex Wigner matrix $N \times N$ by letting :

$$X_N^{(2)}(i,j) = \begin{cases} \frac{a_{i,j}^{(2)}}{\sqrt{N}} \text{ when } i \leq j, \\ \\ \frac{\overline{a_{j,i}^{(2)}}}{\sqrt{N}} \text{ when } i > j\,. \end{cases}$$

Again, it is well known that the spectral measure of $X_N^{(2)}$ converges towards the semi-circle distribution σ and that the eigenvalues stick to the bulk [4].

Theorem 1.3.9. *Assume that Assumptions 1.3.8 and 1.3.5 hold. Then, the law of the largest eigenvalue $\lambda_{\max}(X_N^{(2)})$ of $X_N^{(2)}$ satisfies a large deviation principle with speed N and good rate function $I^{(2)}$ which is infinite on $(-\infty, 2)$ and otherwise given by*

$$I^{(2)}(\rho) = 2I^{(1)}(\rho) = \int_2^\rho \sqrt{x^2-4}dx\,.$$

We finally generalize our result to the case of Wishart matrices. We let L, M be two integers with $N = L + M$. Let $G^{(\beta)}_{L,M}$ be an $L \times M$ matrix with independent entries $(a^{(\beta)}_{i,j})_{\substack{1\le i\le L\\ 1\le j\le M}}$ with laws $\mu^{L,M}_{i,j}$ on the real line if $\beta = 1$ and on the complex plane if $\beta = 2$. The $\mu^{L,M}_{i,j}$ satisfy a sharp sub-Gaussian Laplace transform (with real or complex values) for all $i, j \in [1, L] \times [1, M]$, and its complementary uniform lower bound (Assumption 1.3.1, or Assumption 1.3.8), are centered and have covariance one. We set $W^{(\beta)}_{L,M} = \frac{1}{L} G^{(\beta)}_{L,M}(G^{(\beta)}_{L,M})^*$. When M/L converges towards α, the spectral distribution of $W^{(\beta)}_{L,M}$ converges towards the Pastur-Marchenko law [59]: for any bounded continuous function f

$$\lim_{N\to\infty} \int f(x) d\hat{\mu}^L_{W^{(\beta)}_{L,M}}(x) = \int f(x) d\pi_\alpha(x) \qquad a.s.,$$

where if $\alpha \ge 1$ and $a_\alpha = (1-\sqrt{\alpha})^2, b_\alpha = (1+\sqrt{\alpha})^2$,

$$\pi_\alpha(dx) = \frac{\sqrt{(b_\alpha - x)(x-a_\alpha)}}{2\pi x} \mathbb{1}_{[a_\alpha, b_\alpha]} dx\,.$$

When $\alpha < 1$, the limiting spectral measure has aditionnally a Dirac mass at the origin with mass $1-\alpha$. We hereafter concentrates on the case $M \ge L$ up to replace $W^{(\beta)}_{L,M}$ by $(G^{(\beta)}_{L,M})^* G^{(\beta)}_{L,M}/M$. Again, the extreme eigenvalues were shown to stick to the bulk [11]. We prove a large deviation principle from this convergence:

Theorem 1.3.10. *Assume that the $\mu^N_{i,j}$ satisfy Assumption 1.3.5. Assume they satisfy a sharp Gaussian Laplace transform 1.3.1 when $\beta = 1$ or 1.3.8 when $\beta = 2$, and a uniform lower bounded Laplace transform 1.3.1 when $\beta = 1$ or 1.3.8 when $\beta = 2$. Assume that there exists $\alpha \ge 1$ and $\kappa > 0$ so that $\frac{M}{L} - \alpha = o(N^{-\kappa})$. Then, the law of the largest eigenvalue $\lambda_{\max}(W^{(\beta)}_{L,M})$ of $W^{(\beta)}_{L,M}$ satisfies a large deviation principle with speed N and good rate function $J^{(\beta)}$ which is infinite on $(-\infty, b_\alpha)$ and otherwise given by*

$$J^{(\beta)}(x) = \frac{\beta}{4(1+\alpha)} \int_{b_\alpha}^x \frac{\sqrt{(y-b_\alpha)(y-a_\alpha)}}{y} dy\,.$$

where $\beta = 1$ in the case of real entries, and $\beta = 2$ in the case of complex entries.

This problem can be recasted in terms of Hermitian models with independent entries in the sense that if we consider the $N \times N$ matrix

$$X_N^{(w_\beta)} = \begin{pmatrix} 0 & \frac{1}{\sqrt{N}} G^{(\beta)}_{L,M} \\ \frac{1}{\sqrt{N}} (G^{(\beta)}_{L,M})^* & 0 \end{pmatrix}$$

the spectrum of the $N \times N$ matrix $X_N^{(w_\beta)}$ is given by L eigenvalues $\sqrt{\frac{L}{N}\lambda}$, L eigenvalues $-\sqrt{\frac{L}{N}\lambda}$, where λ are the eigenvalues of $W^{(\beta)}_{L,M}$, and $M - L$ vanishing eigenvalues. Hence, the largest eigenvalue of $W^{(\beta)}_{L,M}$ is the square of the largest eigenvalue of $X_N^{(w_\beta)}$ multiplied by N/L. $X_N^{(w_\beta)}$ is, like X_N^β a Hermitian matrix which is linear in the random entries, but it has more structure with its zero entries. It is therefore equivalent to show a large deviation principle for the largest eigenvalue of $X_N^{(w_\beta)}$ with speed N and rate function

$$I^{(w_\beta)}(x) = J^{(\beta)}((1+\alpha)x^2)\,.$$

This amounts to consider a Wigner matrix with some entries set to zero. We denote $a_{i,j}^{(w_\beta)}$ the entries of $\sqrt{N}X_N^{(w_\beta)}$:

$$\begin{aligned} a_{i,j}^{(w_\beta)} &= 0, \qquad \text{if } i,j \le L \text{ or } i,j \ge L+1,\\ a_{i,j}^{(w_\beta)} &= a_{i-L,j}^{(\beta)}, \quad i \ge L+1, j \le L,\\ a_{i,j}^{(w_\beta)} &= \overline{a}_{j-L,i}^{(\beta)}, \quad j \ge L+1, i \le N. \end{aligned}$$

Again, we denote by $\mu_{i,j}^N$ the law of the i,jth entry of this matrix. Hereafter, we denote by σ_w the limiting spectral distribution of $X_N^{(w_\beta)}$ given for any test function f by

$$\int f(x)d\sigma_w(x) = \frac{1}{1+\alpha}\left(\int f(\sqrt{\frac{x}{1+\alpha}})d\pi_\alpha(x) + \int f(-\sqrt{\frac{x}{1+\alpha}})d\pi_\alpha(x)\right) + \frac{\alpha-1}{\alpha+1}f(0)\,. \tag{1.2}$$

Therefore, we shall prove Theorem 1.3.10 by showing that

Theorem 1.3.11. *Assume that the $\mu_{i,j}^N$ satisfy Assumption 1.3.5. Assume they satisfy a sharp Gaussian Laplace transform 1.3.1 when $\beta = 1$ or 1.3.8 when $\beta = 2$, and a uniform lower bounded Laplace transform 1.3.1 when $\beta = 1$ or 1.3.8 when $\beta = 2$. Assume that there exists $\alpha \ge 1$ and $\kappa > 0$ so that $\frac{M}{L} - \alpha = o(N^{-\kappa})$. Then, the law of the largest eigenvalue $\lambda_{\max}(X_N^{(w_\beta)})$ of $X_N^{(w_\beta)}$ satisfies a large deviation principle with speed N and good rate function $I^{(w_\beta)}$ which is infinite on $(-\infty, \tilde{b}_\alpha)$, if $\tilde{b}_\alpha = \sqrt{(1+\alpha)^{-1}b_\alpha}$ and otherwise given by*

$$I^{(w_\beta)}(x) = \frac{\beta}{1+\alpha}\int_{\tilde{b}_\alpha}^x \frac{1}{y}\sqrt{(1+\alpha)^2(y^2-1)^2 - 4\alpha}dy\,.$$

where $\beta = 1$ in the case of real entries, and two in the case of complex entries.

Acknowledgments: Alice Guionnet wishes to thank A. Dembo for long discussions about large deviations for the largest eigenvalue for sub-Gaussian matrices in Abu Dhabi in 2011. The idea to tilt measures by the spherical integral came out magically from a discussion with M. Potters in UCLA in 2017 and we wish to thank him for this beautiful inspiration. We also benifited from many discussions with M. Maïda with whom one of the author is working on a companion paper on unitarily invariant ensembles, as well as with Fanny Augeri with whom we are working on a follow up paper tackling the general sub-Gaussian case. Finally, we are very grateful for stimulating discussions with O. Zeitouni and N. Cook.

This work was supported by the LABEX MILYON (ANR-10-LABX-0070) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX- 0007) operated by the French National Research Agency (ANR).

1.3.1.2 Scheme of the proof

The idea of the proof is reminiscent of Cramér's approach to large deviations: we appropriately tilt measures to make the desired deviations likely. The point is to realize that it is enough to shift the measure in a random direction and use estimates on spherical intergrals obtained by one of the author and M. Maïda [47]. To be more precise, we shall follow the usual scheme to prove first exponential tightness:

Lemma 1.3.12. *For $\beta = 1, 2, w_1, w_2$, assume that the distribution of the entries $a_{i,j}^{(\beta)}$ satisfy Assumption 1.3.1 for $\beta = 1, w_1$ and Assumption 1.3.8 for $\beta = 2, w_2$. Then:*

$$\lim_{K\to+\infty} \limsup_{N\to\infty} \frac{1}{N} \ln \mathbb{P}[\lambda_{\max}(X_N^{(\beta)}) > K] = -\infty .$$

Similar results hold for $\lambda_{\min}(X_N^{(\beta)})$.

This result is proved in Section 1.3.2. Note that in fact the proof of Lemma 1.3.12 requires only sub-Gaussian tails. Therefore it is enough to prove a weak large deviation principle.

In the following we summarize the assumptions on the distribution of the entries as follows :

Assumption 1.3.13. *Either the $\mu_{i,j}^N$ are uniformly compactly supported in the sense that there exists a compact set K such that the support of all $\mu_{i,j}^N$ is included in K, or the $\mu_{i,j}^N$ satisfy a uniform log-Sobolev inequality in the sense that there exists a constant c independent of N such that for all smooth function f*

$$\int f^2 \ln \frac{f^2}{\mu_{i,j}^N(f^2)} d\mu_{i,j}^N \le c\mu_{i,j}^N(\|\nabla f\|_2^2) .$$

More precisely, in the later case and when the entries are complex, we assume that they are of the form $z(x+iy)$ with a complex number z, and independent real variables x, y which satisfy log-Sobolev inequality with the same constant c independent of N. When $\beta = 1, w_1$ $\mu_{i,j}^N$ satisfy Assumption 1.3.1, when $\beta = 2, w_2$, they satisfy Assumption 1.3.8. In the case of Wishart matrices, $\beta = w_1$ or w_2, we assume that there exists $\alpha > 0$ and $\kappa > 0$ so that $|\frac{M}{L} - \alpha| \le N^{-\kappa}$ for N large enough.

We shall first prove that we have a weak large deviation upper bound:

Theorem 1.3.14. *Assume that Assumption 1.3.13 holds. Let $\beta = 1, 2, w_1, w_2$. Then, for any real number x,*

$$\limsup_{\delta\to 0} \limsup_{N\to\infty} \frac{1}{N} \ln \mathbb{P}\left(\left|\lambda_{\max}(X_N^{(\beta)}) - x\right| \le \delta\right) \le -I_\beta(x) .$$

We shall then obtain the large deviation lower bound.

Theorem 1.3.15. *Assume that Assumption 1.3.13 holds. Let $\beta = 1, 2, w_1, w_2$. Then, for any real number x,*

$$\liminf_{\delta\to 0} \liminf_{N\to\infty} \frac{1}{N} \ln \mathbb{P}\left(\left|\lambda_{\max}(X_N^{(\beta)}) - x\right| < \delta\right) \ge -I_\beta(x) .$$

To prove Theorem 1.3.14, we first show that the rate function is infinite below the right edge of the support of the limiting spectral distribution. To this end, we use that the spectral measure $\hat{\mu}_N$ converges towards its limit which much larger probability. We denote this limit σ_β: $\sigma_1 = \sigma_2 = \sigma$ is the semi-circle law and $\sigma_{w_1} = \sigma_{w_2} = \sigma_w$ is the symmetrization of Pastur-Marchenko law (1.2). We let d denote the Dudley distance:

$$d(\mu, \nu) = \sup_{\|f\|_L \le 1} \left| \int f(x) d\mu(x) - \int f(x) d\nu(x) \right| ,$$

where $\|f\|_L = \sup_{x\neq y} \left|\frac{f(x)-f(y)}{x-y}\right| + \sup_x |f(x)|$.

Lemma 1.3.16. *Assume that the $\mu_{i,j}^N$ are uniformly compactly supported or satisfy a uniform log-Sobolev inequality, as well as, in the case w_1, w_2, that there exists $\kappa > 0$ such that $|\frac{M}{N} - \alpha| \le N^{-\kappa}$. Then, for $\beta = 1, 2, w_1, w_2$, there exists $\kappa' \in (0, \frac{1}{10} \wedge \kappa)$ such that*

$$\limsup_{N\to\infty} \frac{1}{N} \ln \mathbb{P}\left(d(\hat{\mu}_{X_N^{(\beta)}}^N, \sigma_\beta) > N^{-\kappa'}\right) = -\infty .$$

The proof of this lemma is given in the appendix. As a consequence, we deduce that the extreme eigenvalues cannot deviate towards a point inside the support of the limiting spectral measure with probability greater than e^{-CN} for any $C>0$ and therefore

Corollary 1.3.17. *Under the assumption of Lemma 1.3.16, For $\beta = 1,2$ let x be a real number in $(-\infty,2)$ or, for $\beta = w_1, w_2$, take $x \in (-\infty, \tilde{b}_\alpha)$. Then, for $\delta > 0$ small enough,*

$$\limsup_{N\to\infty} \frac{1}{N} \ln \mathbb{P}\left(|\lambda_{\max}(X_N^{(\beta)}) - x| \le \delta \right) = -\infty\,.$$

Indeed, as soon $\delta > 0$ is small enough so that $x+\delta$ is smaller than $2-\delta$ for $\beta = 1,2$ (resp. $b_\alpha - \delta$ for $\beta = w_1, w_2$), $d(\hat{\mu}_N, \sigma_\beta)$ is bounded below by some $\kappa(\delta) > 0$ on the event that $|\lambda_{\max}(X_N^{(\beta)}) - x| \le \delta$. Hence, Lemma 1.3.16 implies the Corollary.

In order to prove the weak large deviation bounds for the remaining x's, we shall tilt the measure by using spherical integrals:

$$I_N(X,\theta) = \mathbb{E}_e[e^{\theta N \langle e, Xe\rangle}]\,,$$

where the expectation holds over e which follows the uniform measure on the sphere $\mathbb{S}^{N-1}$ with radius one. The asymptotics of

$$J_N(X,\theta) = \frac{1}{N} \ln I_N(X,\theta)$$

were studied in [47] where it was proved that

Theorem 1.3.18. *If $(E_N)_{N\in\mathbb{N}}$ is a sequence of $N\times N$ real symmetric matrices when $\beta = 1$ and complex Hermitian matrices when $\beta = 2$ such that :*

— *The sequence of empirical measures $\hat{\mu}^N_{E_N}$ weakly converges to a compactly supported measure μ,*

— *There are two reals $\lambda_{\min}(E), \lambda_{\max}(E)$ such that $\lim_{N\to\infty} \lambda_{\min}(E_N) = \lambda_{\min}(E)$ and $\lim_{N\to\infty} \lambda_{\max}(E_N) = \lambda_{\max}(E)$,*

and $\theta \ge 0$, then :

$$\lim_{N\to\infty} J_N(E_N,\theta) = J(\mu,\theta,\lambda_{\max}(E))\,.$$

The limit J is defined as follows. For a compactly supported probability measure we define its Stieltjes transform G_μ by

$$G_\mu(z) := \int_{\mathbb{R}} \frac{1}{z-t} d\mu(t)\,.$$

We assume hereafter that μ is supported on a compact $[a,b]$. Then G_μ is a bijection from $\mathbb{R}\setminus[a,b]$ to $]G_\mu(a), G_\mu(b)[\setminus\{0\}$ where $G_\mu(a), G_\mu(b)$ are taken as the (possibly infinite) limits of $G_\mu(t)$ when $t \to a^-$ and $t\to b^+$. We denote by K_μ its inverse and let $R_\mu(z) := K_\mu(z) - 1/z$ be its R-transform as defined by Voiculescu in [72] (defined on a neighborhood of the origin, but also on $]G_\mu(a), G_\mu(b)[$). In the sequel, for any compactly supported probability measure μ, we denote by $r(\mu)$ the right edge of the support of μ. In order to define the rate function, we now introduce, for any $\theta \ge 0$, and $\lambda \ge r(\mu)$,

$$J(\mu,\theta,\lambda) := \theta v(\theta,\mu,\lambda) - \frac{\beta}{2} \int \log\left(1 + \frac{2}{\beta}\theta v(\theta,\mu,\lambda) - \frac{2}{\beta}\theta y\right) d\mu(y), \tag{1.3}$$

with

$$v(\theta,\mu,\lambda) := \begin{cases} R_\mu(\frac{2}{\beta}\theta), & \text{if } 0 \le \frac{2\theta}{\beta} \le H_{\max}(\mu,\lambda) := \lim_{z\downarrow\lambda} \int \frac{1}{z-y} d\mu(y), \\ \lambda - \frac{\beta}{2\theta}, & \text{if } \frac{2\theta}{\beta} > H_{\max}(\mu,\lambda). \end{cases}$$

We shall later use that spherical integrals are continuous. We recall here Proposition 2.1 from [58] (see [58] for an erratum) and Theorem 6.1 from [47]. We denote by $\|A\|$ the operator norm of the matrix A given by $\|A\| = \sup_{\|u\|_2=1} \|Au\|_2$ where $\|u\|_2 = \sqrt{\sum |u_i|^2}$.

Proposition 1.3.19. *For every $\theta > 0$, every $\kappa \in]0, 1/2[$, every $M > 0$, there exist a function $g_\kappa : \mathbb{R}^+ \to \mathbb{R}^+$ going to 0 at 0 such that for any $\delta > 0$ and N large enough, with B_N and B'_N such that $d(\hat{\mu}^N_{B_N}, \hat{\mu}^N_{B'_N}) < N^{-\kappa}$, $|\lambda_{\max}(B_N) - \lambda_{\max}(B'_N)| < \delta$ and $\sup_N ||B_N|| \leq M$, $\sup_N ||B'_N|| \leq M$:*

$$|J_N(B_N, \theta) - J_N(B'_N, \theta)| < g_\kappa(\delta) .$$

From Theorem 1.3.18 and Proposition 1.3.19, we deduce that :

Corollary 1.3.20. *For every $\theta > 0$, every $\kappa \in]0, 1/2[$, every $M > 0$, for any $\delta > 0$ and μ a probability measure supported in $[-M, M]$, if we denote by $\mathcal{B}_N$ the set of symmetric matrices B_N such that $d(\mu_{B_N}, \mu) < N^{-\kappa}$, $|\lambda_{\max}(B_N) - \rho| < \delta$, and $\sup_N ||B_N|| \leq M$, for N large enough, we have :*

$$\limsup_{N\to\infty} \sup_{B_N \in \mathcal{B}_N} |J_N(B_N, \theta) - J(\mu, \theta, \rho)| \leq 2g_\kappa(\delta)$$

where g_κ is the function in Proposition 1.3.19.

By Lemma 1.3.12 and Lemma 1.3.16, it is enough to study the probability of deviations on the set where J_N is continuous:

Corollary 1.3.21. *Suppose Assumption 1.3.5 holds. For $\delta > 0$, take a real number x and set for M large (larger than $x + \delta$ in particular), $\mathcal{A}^M_{x,\delta}$ to be the set of $N \times N$ self-adjoint matrices given by*

$$\mathcal{A}^M_{x,\delta} = \{X : |\lambda_{\max}(X) - x| < \delta\} \cap \{X : d(\hat{\mu}^N_X, \sigma_\beta) < N^{-\kappa'}\} \cap \{X : \|X\| \leq M\} ,$$

where κ' is chosen as in Lemma 1.3.16 . Let x be a real number, $\delta > 0$ and κ' as in Lemma 1.3.16. Then, for any $L > 0$, for M large enough

$$\mathbb{P}\left(\left|\lambda_{\max}(X_N^{(\beta)}) - x\right| < \delta\right) = \mathbb{P}\left(X_N^{(\beta)} \in \mathcal{A}^M_{x,\delta}\right) + O(e^{-NL}) .$$

We are now in position to get an upper bound for $\mathbb{P}\left(X_N^{(\beta)} \in \mathcal{A}^M_{x,\delta}\right)$. In fact, by the continuity of spherical integrals of Corollary 1.3.20, for any $\theta \geq 0$,

$$\begin{aligned}
\mathbb{P}\left(X_N^{(\beta)} \in \mathcal{A}^M_{x,\delta}\right) &= \mathbb{E}\left[\frac{I_N(X_N^{(\beta)}, \theta)}{I_N(X_N^{(\beta)}, \theta)} 1_{\mathcal{A}^M_{x,\delta}}\right] \\
&\leq \mathbb{E}[I_N(X_N^{(\beta)}, \theta)] \exp\{-N \inf_{X \in \mathcal{A}^M_{x,\delta}} J_N(X, \theta)\} \\
&\leq \mathbb{E}[I_N(X_N^{(\beta)}, \theta)] \exp\{N(2g_\kappa(\delta) - J(\sigma_\beta, \theta, x))\} \qquad (1.4)
\end{aligned}$$

where we used that $x \to J(\sigma_\beta, \theta, x)$ is continuous and took N large enough. It is therefore central to derive the asymptotics of

$$F_N(\theta, \beta) = \frac{1}{N} \ln \mathbb{E}[I_N(X_N^{(\beta)}, \theta)] ,$$

and we shall prove in section 1.3.3 that

Theorem 1.3.22. *Suppose Assumption 1.3.13 holds. For $\beta = 1, 2, w_1, w_2$ and $\theta \geq 0$,*

$$\lim_{N\to\infty} F_N(\theta, \beta) = F(\theta, \beta)$$

with $F(\theta, \beta) = \theta^2/\beta$ if $\beta = 1, 2$ and when $\beta = w_i$, $i = 1, 2$:

$$F(\theta, w_i) = \sup_{x\in[0,1]} \{\frac{2\theta^2}{i} x(1-x) + \frac{i}{2(1+\alpha)} \ln(1-x) + \frac{i\alpha}{2(1+\alpha)} \ln x\} - iC_\alpha ,$$

where $C_\alpha = \frac{1}{2(1+\alpha)} \ln(\frac{1}{1+\alpha}) + \frac{\alpha}{2(1+\alpha)} \ln \frac{\alpha}{1+\alpha}$.

We therefore deduce from (1.4), Corollaries 1.3.21 and 1.3.20 , and Theorem 1.3.22, by first letting N going to infinity, then δ to zero and finally M to infinity, that

$$\limsup_{\delta\to 0}\limsup_{N\to\infty}\frac{1}{N}\ln\mathbb{P}\left(\left|\lambda_{\max}(X_N^{(\beta)})-x\right|<\delta\right)\le F(\theta,\beta)-J(\sigma_\beta,\theta,x)\,.$$

We next optimize over θ to derive the upper bound:

$$\limsup_{\delta\to 0}\limsup_{N\to\infty}\frac{1}{N}\ln\mathbb{P}\left(\left|\lambda_{\max}(X_N^{(\beta)})-x\right|<\delta\right)\le -\sup_{\theta\ge 0}\{J(\sigma_\beta,\theta,x)-F(\theta,\beta)\}\,. \tag{1.5}$$

To complete the proof of Theorem 1.3.14, we show in section 1.3.4 that, with the notations of Theorems 1.3.10,1.3.9, and 1.3.11,

Proposition 1.3.23. *For $\beta=1,2,w_1,w_2$,*

$$I_\beta(x)=\sup_{\theta\ge 0}\{J(\sigma_\beta,\theta,x)-F(\theta,\beta)\}\,.$$

To prove the complementary lower bound, we shall prove that

Lemma 1.3.24. *For $\beta=1,2$, for any $x>2$ and for $\beta=w_1,w_2$ for any $x>\tilde b_\alpha$, there exists $\theta=\theta_x\ge 0$ such that for any $\delta>0$ and M large enough,*

$$\liminf_{N\to\infty}\frac{1}{N}\ln\frac{\mathbb{E}[\mathbb{1}_{X_N^{(\beta)}\in\mathcal{A}_{x,\delta}^M}I_N(X_N^{(\beta)},\theta)]}{\mathbb{E}[I_N(X_N^{(\beta)},\theta)]}\ge 0\,.$$

This lemma is proved by showing that the matrix whose law has been tilted by the spherical integral is approximately a rank one perturbation of a Wigner matrix, from which we can use the techniques developed to study the famous BBP transition [13]. The conclusion of Theorem 1.3.15 follows since then

$$\begin{aligned}\mathbb{P}\left(X_N^{(\beta)}\in\mathcal{A}_{x,\delta}^M\right) &\ge \frac{\mathbb{E}[\mathbb{1}_{X_N^\delta\in\mathcal{A}_{x,\delta}^M}I_N(X_N^{(\beta)},\theta_x)]}{\mathbb{E}[I_N(X_N^{(\beta)},\theta_x)]}\mathbb{E}[I_N(X_N^{(\beta)},\theta_x)]\exp\{-N\sup_{X\in\mathcal{A}_{x,\delta}^M}J_N(X,\theta_x)\}\\ &\ge \exp\{N(-2g_\kappa(\delta)+F(\theta_x,\beta)-J(\sigma_\beta,\theta_x,x)+o(1))\}\\ &\ge \exp\{-NI_\beta(x)-N(o(1)+2g_{\kappa(\delta)})\}\end{aligned}$$

where we finally used Theorem 1.3.22 and Lemma 1.3.24.

1.3.2 Exponential tightness

In this section we prove Lemma 1.3.12. We will use a standard net argument that we recall for completness. We denote by $\mathbb{S}^{N-1}$ the unit sphere in $\mathbb{C}^N$ if $\beta=2$ or $\mathbb{R}^N$ if $\beta=1$. For $N\in\mathbb{N}$, let R_N be a 1/2-net of the sphere (i.e. a subset of the sphere $\mathbb{S}^{N-1}$ such as for all $u\in\mathbb{S}^{N-1}$ there is $v\in R_N$ such that $||u-v||_2\le 1/2$). Here the sphere is inside $\mathbb{R}^N$ for $\beta=1,w_1$ and $\mathbb{C}^N$ for $\beta=2,w_2$. We know that we can take R_N with cardinality smaller than 3^{2N}. We notice that for $M>0$

$$\mathbb{P}[||X_N^{(\beta)}||\ge 4K]\le 9^{2N}\sup_{u,v\in R_N}\mathbb{P}[\langle X_N^{(\beta)}u,v\rangle\ge K]\,. \tag{1.6}$$

Indeed, if we denote, for $v\in\mathbb{S}^{N-1}$, u_v to be an element of R_N such that $\|u_v-v\|_2\le 1/2$,

$$\|X_N^{(\beta)}\|=\sup_{v\in\mathbb{S}^{N-1}}\|X_N^{(\beta)}v\|_2\le\sup_{v\in\mathbb{S}^{N-1}}(\|X_N^{(\beta)}u_v\|_2+\frac{1}{2}\|X_N^{(\beta)}\|)$$

so that

$$\|X_N^{(\beta)}\| \le 2 \sup_{u\in R_N} \|X_N^{(\beta)} u\|_2 \,. \tag{1.7}$$

Similarly, taking $v = \frac{X_N^{(\beta)} u}{\|X_N^{(\beta)} u\|_2}$, we find

$$\|X_N^{(\beta)} u\|_2 = \langle v, X_N^{(\beta)} u\rangle \le \langle u_v, X_N^{(\beta)} u\rangle + \|v-u_v\|_2 \|X_N^{(\beta)} v\|_2$$

from which we deduce that

$$\|X_N^{(\beta)}\| \le 4 \sup_{u,v\in R_N} \langle X_N^{(\beta)} u, v\rangle$$

and (1.6) follows. We next bound the probability of deviations of $\langle X_N^{(\beta)} v, u\rangle$ by using Tchebychev's inequality. For $\theta \ge 0$ we indeed have

$$\begin{aligned}
\mathbb{P}[\langle X_N^{(\beta)} u, v\rangle \ge K] &\le \exp\{-\theta N K\}\mathbb{E}[\exp\{N\theta\langle X_N^{(\beta)} u, v\rangle\}]\\
&\le \exp\{-\theta N K\}\mathbb{E}[\exp\left\{\sqrt{N}\left(2\sum_{i<j}\Re(a_{i,j}^{(\beta)} u_i \bar v_j) + \sum_i a_{i,i} u_i v_i\right)\right\}]\\
&\le \exp\{-\theta N K\}\exp\left(\frac{\theta^2 N}{\beta'}(2\sum_{i<j}|u_i|^2|v_j|^2 + \sum_i |u_i|^2|v_i|^2)\right) \qquad (1.8)
\end{aligned}$$

where we used that the entries have a sharp sub-Gaussian Laplace transform. In the case of Wishart matrices, we bounded above some vanishing contributions by a non-negative term. When $\beta = w_i$, $\beta' = i$, otherwise $\beta' = \beta$. We can now complete the upper bound:

$$\begin{aligned}
\mathbb{P}[\langle X_N^{(\beta)} u, v\rangle \ge K] &\le \exp\left(\frac{\theta^2 N}{\beta'}\frac{||u||_2^2||v||_2^2 + \langle u,v\rangle^2}{2} - \theta N K\right)\\
&\le \exp\left(N\left(\frac{1}{\beta'} - K\right)\right)
\end{aligned}$$

where we took $\theta = 1$. We conclude that :

$$\mathbb{P}[\langle X_N^{(\beta)} u, v\rangle \ge K] \le \exp\left(N(1-K)\right)\,.$$

This complete the proof of the Lemma with (1.6).

1.3.3 Proof of Theorem 1.3.22

We consider in this section a random unit vector e taken uniformly on the sphere $\mathbb{S}^{N-1}$ and independent of $X_N^{(\beta)}$. We define F_N by setting, for $\theta > 0$:

$$F_N(\theta,\beta) = \frac{1}{N}\ln \mathbb{E}_{X_N^{(\beta)}}\mathbb{E}_e[\exp(N\theta\langle e, X_N^{(\beta)} e\rangle)]$$

where we take both the expectation $\mathbb{E}_e$ over e and the expectation $\mathbb{E}_{X_N^{(\beta)}}$ over $X_N^{(\beta)}$. In this section we derive the asymptotics of $F_N(\theta,\beta)$. $F(\theta,\beta)$ is as in Theorem 1.3.22. We prove a refinment of Theorem 1.3.22, which shows that under our assumption of sharp sub-Gaussian tails, the random vector e stays delocalized under the tilted measure.

Proposition 1.3.25. *Suppose Assumption 1.3.1 holds if $\beta = 1, w_1$ and Assumption 1.3.8 holds if $\beta = 2, w_2$. Denote by $V_N^\epsilon = \{e \in \mathbb{S}^{N-1} : \forall i, |e_i| \le N^{-1/4-\epsilon}\}$. Then, for $\epsilon \in (0, \frac{1}{4})$,*

$$F(\theta,\beta) \quad = \quad \lim_{N\to+\infty} F_N(\theta,\beta) = \lim_{N\to\infty} \frac{1}{N} \ln \mathbb{E}_e[\mathbb{1}_{e\in V_N^\epsilon} \mathbb{E}_{X_N^{(\beta)}}[\exp(N\theta\langle e, X_N^{(\beta)} e\rangle)]] \,.$$

We first consider the case of Wigner matrices and then the case of Wishart matrices: in both cases the proof shows that the above delocalization holds (i.e we can restrict ourselves to vectors e in V_N^ϵ) and we shall not mention it in the following statements.

1.3.3.1 Wigner matrices

In this section we prove Theorem 1.3.22 in the case of Wigner matrices, namely:

Lemma 1.3.26. *Suppose Assumption 1.3.1 holds if $\beta = 1$ and Assumption 1.3.8 holds if $\beta = 2$. Then for any $\theta \ge 0$*

$$\lim_{N\to+\infty} F_N(\theta,\beta) = F(\theta,\beta) = \frac{\theta^2}{\beta} \,.$$

Proof. By denoting $L_\mu = \ln T_\mu$, we have :

$$\begin{aligned}
\mathbb{E}_{X_N^{(\beta)}}[\exp(N\theta\langle e, X_N^{(\beta)} e\rangle)] &= \mathbb{E}_{X_N^{(\beta)}}[\exp\{\sqrt{N}\theta(2\sum_{i<j} \Re(a_{i,j}^{(\beta)} e_j \bar{e}_i) + \sum_i a_{i,i}^{(\beta)} |e_i|^2)\}] \\
&= \exp\{\sum_{i<j} L_{\mu_{i,j}^N}(2\theta\bar{e}_i e_j \sqrt{N}) + \sum_i L_{\mu_{i,i}^N}(\theta|e_i|^2\sqrt{N})\}
\end{aligned}$$

where we used the independence of the $(a_{i,j}^{(\beta)})_{i\le j}$. Using that the entries have a sharp sub-Gaussian Laplace transform (using on the diagonal the weaker bound $L_{\mu_{i,i}^N}(t) \le \frac{1}{\beta}t^2 + A|t|$) and $\sum e_i^2 = 1$, we deduce that:

$$\begin{aligned}
\mathbb{E}_{X_N^{(\beta)}}[\exp(N\theta\langle e, X_N^{(\beta)} e\rangle)] &\le \mathbb{E}_e[\exp\{\frac{2N\theta^2}{\beta}\sum_{i<j}|e_i|^2|e_j|^2 + \frac{N\theta^2}{\beta}\sum_i |e_i|^4 + A\sqrt{N}\theta \sum_i e_i^2\}] \\
&\le \exp(N\frac{\theta^2}{\beta} + A\sqrt{N}\theta) \,.
\end{aligned}$$

Hence, we have proved the following upper bound

$$\limsup_{N\to\infty} F_N(\theta,\beta) \le \limsup_{N\to\infty} \sup_{e\in\mathbb{S}^{N-1}} \frac{1}{N} \ln \mathbb{E}_{X_N^{(\beta)}}[\exp(N\theta\langle e, X_N^{(\beta)} e\rangle)] \le \frac{\theta^2}{\beta} \,. \tag{1.9}$$

We next prove the corresponding lower bound. The idea is that the expectation over the vector e concentrates on delocalized eigenvectors with entries so that $\sqrt{N}e_i\bar{e}_j$ is going to zero for all i,j. As a consequence we will be able to use the uniform lower bound on the Laplace transform (see Remark 1.3.2) to lower bound $F_N(\theta,\beta)$. Let V_N^ϵ be as in Proposition 1.3.25. We have that :

$$\mathbb{E}[\exp(N\theta\langle e, X_N^{(\beta)} e\rangle)] \ge \mathbb{E}_e[\mathbb{1}_{e\in V_N^\epsilon} \prod_{i<j} \exp\{L_{\mu_{i,j}^N}(2\sqrt{N}\theta\bar{e}_i e_j/\beta)\} \prod_i \exp\{L_{\mu_{i,i}^N}(\sqrt{N}\theta|e_i|^2/\beta)\}]$$

For $e \in V_N^\epsilon$, $2\sqrt{N}\theta|e_i e_j| \le 2\theta N^{-\epsilon}$ so that :

$$\lim_{N\to+\infty} \sup_{e\in V_N^\epsilon} |2\sqrt{N}\theta e_i e_j| = 0 \,.$$

By the uniform lower bound on the Laplace transform of Assumptions 1.3.1 or 1.3.8, we deduce that for any $\delta > 0$ and N large enough

$$\mathbb{E}[\exp(N\theta\langle e, X_N^{(\beta)} e\rangle)] \geq \mathbb{P}_e[V_N^\epsilon] e^{N\frac{\theta^2}{\beta}(1-\delta)} . \tag{1.10}$$

We shall use that

Lemma 1.3.27. *For any $\epsilon \in (0, 1/4)$ we have*

$$\lim_{N\to\infty} \mathbb{P}_e[e \in V_N^\epsilon] = 1 .$$

As a consequence, we deduce from (1.10) that for any $\delta > 0$ and N large enough

$$\liminf_{N\to\infty} F_N(\theta,\beta) \geq (1-\delta)\frac{\theta^2}{\beta} .$$

Hence, together with (1.9), we have proved the announced limit

$$\lim_{N\to\infty} F_N(\theta,\beta) = \frac{\theta^2}{\beta}$$

which completes the proof of Lemma 1.3.26. Finally we prove Lemma 1.3.27. To this end we use the well known representation of the vector e as a renormalized (real or complex) Gaussian vector:

$$e = \frac{g}{\|g\|_2}$$

where $g = (g_1, ..., g_N)$ is a Gaussian vector of covariance matrix I_N. By the law of large numbers, we have the following almost sure limit :

$$\lim_{N\to\infty} \frac{||g||_2}{\sqrt{N}} = 1$$

We also have by the union bound

$$\mathbb{P}[\exists i \in [1,N], |g_i| > N^{1/4-\epsilon}/2] \leq N\mathbb{P}[|g_1| > N^{1/4-\epsilon}/2] \leq N\exp\{-\frac{1}{4}N^{1/2-2\epsilon}\}$$

from which the result follows.

□

1.3.3.2 Wishart matrices

In this subsection we prove Theorem 1.3.22 in the case of Wishart matrices, namely:

Lemma 1.3.28. *Let $\beta = w_1$ or w_2. Suppose Assumption 1.3.13 holds. Then for any $\theta \geq 0$, for $i = 1, 2$*

$$\lim_{N\to\infty} F_N(\theta, w_i) = F(\theta, w_i) = \sup_{x\in[0,1]}\{\frac{2\theta^2}{i}x(1-x) + \frac{i}{2(1+\alpha)}\ln(x) + \frac{i\alpha}{2(1+\alpha)}\ln(1-x)\} - iC_\alpha ,$$

where $C_\alpha = \frac{1}{2(1+\alpha)}\ln(\frac{1}{1+\alpha}) + \frac{\alpha}{2(1+\alpha)}\ln\frac{\alpha}{1+\alpha}$. Moreover, the supremum is achieved at a unique $x_{\theta,\alpha}$ in $[0,1]$ (as it maximizes a strictly concave function). $x_{\theta,\alpha}$ is the almost sure limit of $\|e^{(1)}\|_2^2$, the norm of the first L entries of e, under the tilted law

$$d\mathbb{P}^\theta(e) = \frac{\mathbb{E}_X[\exp\{\theta N\langle e, X_N^{(\beta)} e\rangle\}]}{\mathbb{E}_e[\mathbb{E}_X[\exp\{\theta N\langle e, X_N^{(\beta)} e\rangle\}]]} d\mathbb{P}(e) .$$

Proof. We have, with the same notations than in the previous case :

$$\mathbb{E}_{X_N^{w_i}}[\exp(N\theta\langle X_N^{(w_i)}e, e\rangle)] \quad = \quad \exp\left\{ \sum_{1\le i\le M, 1\le j\le L} L_{\mu_{i,j}^N}(\sqrt{N}2\theta e_i^{(1)}\bar{e}_j^{(2)}) \right\}$$

where $e = (e^{(1)}, e^{(2)})$, that is $e^{(1)}$ is the vector made of the L first entries of e and $e^{(2)}$ the vector made of the M last entries of e. Using that the $\mu_{i,j}^N$ have a sharp sub-Gaussian Laplace transform and a uniform lower bounded Laplace transform, we deduce that with V_N^ϵ as in Proposition 1.3.25 we find that for any $\delta > 0$ and N large enough

$$\mathbb{E}_e[\mathbb{1}_{V_N^\epsilon} e^{(1-\delta)\frac{2\theta^2}{i}N\|e^{(1)}\|_2^2\|e^{(2)}\|_2^2}] \le \mathbb{E}_{X_N^{w_i}}[I_N(\theta, w_i)] \le \mathbb{E}_e[e^{\frac{2\theta^2}{i}N\|e^{(1)}\|_2^2\|e^{(2)}\|_2^2}] \tag{1.11}$$

where $\|e^{(1)}\|_2^2 = 1 - \|e^{(2)}\|_2^2$ follows a Beta law with parameters $(iL/2, iM/2)$. Hence, its distribution is given by

$$\mathrm{Beta}_{iM/2,iL/2}(dx) = C_{M,L}x^{iL/2-1}(1-x)^{iM/2-1}\mathbb{1}_{x\in[0,1]}dx\,,$$

with $C_{M,L} = \Gamma(iN/2)/\Gamma(iM/2)\Gamma(iL/2)$. Therefore, Laplace method implies that

$$\lim_{N\to\infty}\frac{1}{N}\ln\mathbb{E}_e[\exp\{\frac{2\theta^2}{i}N\|e^{(1)}\|_2^2\|e^{(2)}\|_2\}] \tag{1.12}$$

$$= \sup_{x\in[0,1]}\{\frac{2\theta^2}{i}x(1-x) + \frac{i\alpha}{2(1+\alpha)}\ln(1-x) + \frac{i}{2(1+\alpha)}\ln(x)\} - iC_\alpha\,.$$

(1.12) thus yields the expected upper bound. To get the lower bound in (1.11), observe that conditioning by $\|e^{(1)}\|_2$, $e^{(1)}$ and $e^{(2)}$ follow uniform laws on spheres of appropriate radius so that Lemma 1.3.27 applies. Hence, V_N^ϵ has probability going to one under this conditionnal measure and we can remove its indicator function in the lower bound of (1.11). We then apply Laplace method under the Beta law to conclude. Finally, we see from the above that for any set A, any $\delta > 0$

$$\mathbb{P}^\theta(\|e^{(1)}\|_2^2 \in A) \le \exp\{-NF(\theta, w_i) + N\delta\}\int_A x^{iL/2-1}(1-x)^{iM/2-1}\exp\{\frac{2\theta^2}{i}Nx(1-x)\}dx$$

from which it follows by Laplace method that the law of $\|e^{(1)}\|_2^2$ satisfies a large deviation upper bound with speed N and good rate function which is infinite outside $[0,1]$ and otherwise given by

$$-\frac{2\theta^2}{i}x(1-x) - \frac{i\alpha}{2(1+\alpha)}\ln(1-x) - \frac{i}{2(1+\alpha)}\ln x + F(\theta, w_i)\,.$$

In particular $\|e^{(1)}\|_2^2$ converges almost surely towards the unique minimizer $x_{\theta,\alpha}$ of this strictly convex function (which vanishes there).

□

1.3.4 Identification of the rate function

To complete the proof of the large deviation upper bound of Theorem 1.3.14, we need to identify the rate function, that is prove Proposition 1.3.23. This could a priori be done by saying that the rate function corresponds to the one that is well known for the Gaussian case. But for the sake of completness, we verify directly that we have the same result.

1.3.4.1 Wigner matrices

We first consider the case of Wigner matrices. Recall that we want to prove that for $\beta = 1, 2$

$$I_\beta(x) = \frac{\beta}{2}\int_2^x \sqrt{y^2-4}dy = \max_{\theta>0}\left(J(\sigma,\theta,x) - \frac{\theta^2}{\beta}\right) \tag{1.13}$$

where $J(\mu,\theta,\lambda)$ was defined in (1.3). Note that when $\mu = \sigma$, $R_\sigma(x) = x$ and $H_{max}(\sigma,\lambda) = G_\sigma(\lambda) = \frac{1}{2}(\lambda - \sqrt{\lambda^2-4})$. To prove (1.13), observe first that the function

$$\varphi(\theta,x) = J(\sigma,\theta,x) - \frac{\theta^2}{\beta}$$

vanishes when $\frac{2\theta}{\beta} \le H_{max}(\sigma,x) = G_\sigma(x)$, since then $J(\sigma,\theta,x) = \frac{\beta}{2}\int_0^{\frac{2}{\beta}\theta} R_\sigma(u)du = \frac{\theta^2}{\beta}$ (see [47, Page 4]). For $\frac{2\theta}{\beta} > G_\sigma(x)$, the critical points of $\varphi(.,x)$ satisfy

$$\frac{2\theta}{\beta} = (\partial_\theta J)(\sigma,\theta,x)\,.$$

Computing the derivative of J shows that they are solution of

$$\frac{2\theta_x}{\beta} = x - \frac{\beta}{2\theta_x}$$

which has a unique solution $\theta_x > \beta G_\sigma(x)/2$ which is given by:

$$\frac{2\theta_x}{\beta} = \frac{1}{2}(x+\sqrt{x^2-4}) = \frac{1}{G_\sigma(x)}\,.$$

Therefore, $I_\beta(x) = \varphi(\theta_x, x)$. We can compute the derivative of I_β and since θ_x is a critical point of φ, we find that

$$\partial_x I_\beta(x) = (\partial_x\varphi)(\theta_x,x) = (\partial_x J)(\sigma,\theta_x,x) = \theta_x - \frac{\beta}{2}G_\sigma(x) = \frac{\beta}{2}\sqrt{x^2-4}\,.$$

This proves the claim since $I_\beta(2) = 0$.

1.3.4.2 Wishart matrices

Let us now consider Wishart matrices and compute

$$I_{w_\beta}(x) = \max_{\theta>0}\left(J(\sigma_w,\theta,x) - F(\theta,w_\beta)\right)\,.$$

As in the previous proof, it is enough to compute the point θ_x where $\varphi(\theta,x) = J(\sigma_w,\theta,x) - F(\theta,w_\beta)$ achieves its maximal value since then we can compute

$$\partial_x I_{w_\beta}(x) = \partial_x J(\sigma_w,\theta_x,x) = \theta_x - \frac{\beta}{2}G_{\sigma_w}(x)\,.$$

Note that θ_x exists as φ is continuous in θ, going to $-\infty$ at infinity. To identify θ_x we remark that when it is larger than $\frac{\beta}{2}G_{\sigma_w}(x)$, it must satisfy the equation of the critical points of φ :

$$x = \partial_\theta F(\theta,w_\beta) + \frac{\beta}{2\theta} =: K(\theta)\,.$$

Our goal is therefore to identify K and in fact its inverse $\theta_x = K^{-1}(x)$.

We first show that K is analytic away from the origin and equals the inverse $K_{\sigma_w}(\frac{2\theta}{\beta})$ of the Stieljes transform for small θ. Indeed, we claim that $\theta \to F(\theta, w_\beta)$ is analytic in a neighborhood of $]0, +\infty[$. We recall that it is given in Lemma 1.3.28 by

$$F(\theta, w_i) = \sup_{x\in[0,1]} \Psi(\theta, x), \quad \Psi(\theta, x) = \frac{2\theta^2}{i} x(1-x) + \frac{i}{2(1+\alpha)} \ln(x) + \frac{i\alpha}{2(1+\alpha)} \ln(1-x)\,.$$

The maximiser $x_{\theta,\alpha}$ of $\Psi(\theta, x)$ is solution of

$$(\partial_x \psi)(x, \theta) = \frac{1}{i^2}\theta^2(1-2x) + \frac{1}{(1+\alpha)x} - \frac{\alpha}{(1+\alpha)(1-x)} = 0\,.$$

Clearly $x \to (\partial_x \psi)(x, \theta)$ takes its zeroes away from $0, 1$ and is analytic in a complex neighborhood of $[\epsilon, 1-\epsilon]$ for any $\epsilon > 0$. Moreover, $\partial\psi$ can only vanish in a small neighborhood of $x = 1/2$ when θ is large. But for $\Re(\theta) > \delta$, the real part of $-\partial_x^2\psi(\theta, x)$ is bounded below uniformly by some $c(\epsilon) > 0$ uniformly a complex neighborhood U_ϵ of $[\epsilon, 1-\epsilon]$ provided the imaginary part of θ is smaller than some $\kappa_{\epsilon,\delta} > 0$. Hence, the implicit function theorem implies that $\theta \to x_{\theta,\alpha}$, and so $F(., w_\beta)$, is analytic in a neighborhood of $\Re(\theta) \geq \delta$ for any $\delta > 0$. We next show that for θ small enough,

$$F(\theta, w_\beta) = \frac{\beta}{2} \int_0^{\frac{2}{\beta}\theta} R_{\sigma_w}(u) du\,. \tag{1.14}$$

F is clearly lower bounded by this value as for any M

$$F(\theta, w_\beta) \geq \liminf_{N\to\infty} \frac{1}{N} \ln \mathbb{E}_{X_N^{(w_\beta)}}[1_{|\lambda_{\max}(X_N^{(w_\beta)})|\leq M} I_N(X_N^{(w_\beta)}, \theta)]$$

so that for $\frac{2\theta}{\beta} \leq G_{\sigma_w}(M)$,[47, Theorem 1.6] gives the lower bound. The upper bound is obtained similarly by using the exponential tightness which permits to restrict oneself to $\{|\lambda_{\max}| \leq M\}$. Therefore, we conclude that K is analytic in $\Re(\theta) > \delta$ and equals

$$K_{\sigma_w}(\frac{2\theta}{\beta}) = R_{\sigma_w}(\frac{2\theta}{\beta}) + \frac{\beta}{2\theta}$$

for small θ.

We want to find the inverse of K. We thus look for an analytic extension of K_{σ_w}. But in fact K_{σ_w} satisfies an algebraic equation. Indeed, observe that

$$G_{\sigma_w}(x) = 2xG_{\pi_\alpha}((1+\alpha)x^2) + \frac{\alpha - 1}{(1+\alpha)x}$$

where it is well known that G_{π_α}, the Stieltjes transform of the Wishart matries, is solution of

$$(2z)^2 G_{\pi_\alpha}(z)^2 - 4z(z+1-\alpha)G_{\pi_\alpha}(z) + 4z - 8\alpha = 0\,.$$

We deduce that at least for small x, K_{σ_w} is solution of

$$((1+\alpha)K_{\sigma_w}(x)x + 1 - \alpha)^2 - 2(K_{\sigma_w}(x) + 1 - \alpha)((1+\alpha)xK_{\sigma_w}(x) + 1 - \alpha) + 4(1+\alpha)K_{\sigma_w}(x)^2 - 8\alpha = 0\,.$$

As a consequence, K is also solution of this equation for all x, by analyticity. Now, we are looking for the inverse of K and so we deduce that $\theta_x = K^{-1}(x)$ is solution of the equation

$$(\frac{2}{\beta}(1+\alpha)x\theta_x + 1 - \alpha)^2 - 2(x + 1 - \alpha)(\frac{2}{\beta}(1+\alpha)x\theta_x + 1 - \alpha) + 4(1+\alpha)x^2 - 8\alpha = 0\,. \tag{1.15}$$

For $\frac{2\theta_x}{\beta} \leq G_{\sigma_w}(x)$, the solution is

$$\frac{2}{\beta}\theta_x = \frac{2\alpha}{1+\alpha} \frac{x^2 + 1 - \alpha - \sqrt{(x^2 - 1 - \alpha)^2 - 4\alpha}}{2x^2} + \frac{1-\alpha}{1+\alpha}\frac{1}{x} = G_{\sigma_w}(x)\,.$$

but when $\frac{2\theta_x}{\beta} > G_{\sigma_w}(x)$ we have to take the other solution of the quadratic equation

$$\frac{2}{\beta}\theta_x = \frac{2\alpha}{1+\alpha}\frac{x^2+1-\alpha+\sqrt{(x^2-1-\alpha)^2-4\alpha}}{2x^2} + \frac{1-\alpha}{1+\alpha}\frac{1}{x}.$$

As a result, we then have

$$\partial_x I_{w_\beta}(x) = \theta_x - \frac{\beta}{2}G_{\sigma_w}(x) = \frac{\beta\alpha}{1+\alpha}\frac{\sqrt{(x^2-1-\alpha)^2-4\alpha}}{x^2},$$

which completes the proof.

1.3.5 Large deviation lower bounds

Recall that we need to prove Lemma 1.3.24, that is find for any $x > 2$ (or $\tilde{b}_\alpha$ for Wishart matrices) a $\theta = \theta_x \geq 0$ such that for any $\delta > 0$ and M large enough,

$$\liminf_{N\to\infty} \frac{1}{N}\ln\frac{\mathbb{E}[\mathbb{1}_{X_N^{(\beta)}\in\mathcal{A}_{x,\delta}^M} I_N(X_N^{(\beta)},\theta)]}{\mathbb{E}[I_N(X_N^{(\beta)},\theta)]} \geq 0,$$

where we recall that

$$\mathcal{A}_{x,\delta}^M = \{X : |\lambda_{\max}(X) - x| < \delta\} \cap \{d(\hat{\mu}_X^N, \sigma_\beta) < N^{-\kappa'}\} \cap \{\|X\| \leq M\}.$$

For a vector e of the sphere $\mathbb{S}^{N-1}$ and X a random symmetric matrix, we denote by $\mathbb{P}_N^{(e,\theta)}$ the probability measure defined by :

$$d\mathbb{P}_N^{(e,\theta)}(X) = \frac{\exp(N\theta\langle Xe, e\rangle)}{\mathbb{E}_X[\exp(N\theta\langle Xe,e\rangle)]}d\mathbb{P}_N(X)$$

where $\mathbb{P}_N$ is the law of $X_N^{(\beta)}$. We have

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_{X_N^{(\beta)}\in\mathcal{A}_{x,\delta}^M} I_N(X_N^{(\beta)},\theta)] &= \mathbb{E}_e[\mathbb{P}_N^{(e,\theta)}(\mathcal{A}_{x,\delta}^M)\mathbb{E}_X[\exp(N\theta\langle Xe,e\rangle)]] \\
&\geq \mathbb{E}_e[1_{e\in V_N^\epsilon}\mathbb{P}_N^{(e,\theta)}(\mathcal{A}_{x,\delta}^M)\mathbb{E}_X[\exp(N\theta\langle Xe,e\rangle)]] \qquad (1.16)
\end{aligned}$$

where we recall that $V_N^\epsilon = \{e \in \mathbb{S}^{N-1} : |e_i| \leq N^{-1/4-\epsilon}\}$. The main point to prove the lower bound will be to show that $\mathbb{P}_N^{(e,\theta)}(\mathcal{A}_{x,\delta}^M)$ is close to one for delocalized vectors $e \in V_N^\epsilon$ and then proceed as before to show that V_N^ϵ has probability close to one under the tilted measure. More precisely, we will show that for $\epsilon \in (\frac{1}{8}, \frac{1}{4})$, we can find θ so that for any $x > 2$ (resp. $x > \tilde{b}_\alpha$) and $\delta > 0$ we can find $\theta_x \geq 0$ so that for M large enough,

$$\lim_{N\to\infty}\inf_{e\in V_N^\epsilon}\mathbb{P}_N^{(e,\theta_x)}(\mathcal{A}_{x,\delta}^M) = 1. \qquad (1.17)$$

This gives the desired estimate since we then deduce from (1.16) that for N large enough so that the above is greater than $1/2$

$$\mathbb{E}[\mathbb{1}_{X_N^{(\beta)}\in\mathcal{A}_{x,\delta}^M} I_N(X_N^{(\beta)},\theta)] \geq \frac{1}{2}\mathbb{E}_e[1_{e\in V_N^\epsilon}\mathbb{E}_{X_N^{(\beta)}}[\exp(N\theta\langle X_N^{(\beta)}e,e\rangle)]]$$

so that the desired estimate follows from Proposition 1.3.25. To prove (1.17), the first point is to show that

Lemma 1.3.29. *Take $\epsilon \in (0, \frac{1}{4})$. There exists $\kappa' > 0$ and K large enough so that for any θ,*

—
$$\lim_{N\to\infty}\sup_{e\in V_N^\epsilon}\mathbb{P}_N^{(e,\theta)}\left(\lambda_{\max}(X_N^{(\beta)})\geq K\right)=0\,.$$

—
$$\limsup_{N\to\infty}\sup_{e\in V_N^\epsilon}\mathbb{P}_N^{(e,\theta)}\left(d(\hat{\mu}^N_{X_N^{(\beta)}},\sigma_\beta)>N^{-\kappa'}\right)=0\,.$$

Proof. We hereafter fix a vector e on the sphere. The proof of the exponential tightness is exactly the same as for Lemma 1.3.12. Indeed, by Jensen's inequality, we have

$$\mathbb{E}_X[\exp(N\theta\langle X_N^{(\beta)}e,e\rangle)]\geq\exp\{N\theta\mathbb{E}_X[\langle X_N^{(\beta)}e,e\rangle]\}=1$$

Moreover, by Tchebychev's inequality, for any $u,v,e\in\mathbb{S}^{N-1}$, we have

$$\begin{aligned}\int 1\!\!1_{\langle X_N^{(\beta)}u,v\rangle\geq K}\exp(N\theta\langle X_N^{(\beta)}e,e\rangle)d\mathbb{P}_N &\leq \exp\{-NK\}\mathbb{E}_X[\exp(N\theta\langle X_N^{(\beta)}e,e\rangle+N\langle X_N^{(\beta)}u,v\rangle)]\\ &\leq \exp\{-NK\}\exp\{N(\theta^2+1)\sum_{i,j}|e_i\bar{e}_j+u_i\bar{v}_j|^2\}\\ &\leq \exp\{-NK+4(\theta^2+1)N\}\end{aligned}$$

from which we deduce after taking u,v on a δ-net as in Lemma 1.3.12 that

$$\mathbb{P}_N^{(e,\theta)}\left(\lambda_{\max}(X_N^{(\beta)})\geq K\right)\leq 9^{2N}\exp\{-\frac{1}{4}NK+4(\theta^2+1)N\}$$

which proves the first point. The second is a direct consequence of Lemma 1.3.16 and the fact that the log density of $\mathbb{P}_N^{(e,\theta)}$ with respect to $\mathbb{P}_N$ is bounded by $\theta N(|\lambda_{\max}(X)|+|\lambda_{min}(E)|)$ which is bounded by θKN with overwhelming probability by the previous point (recall that $\lambda_{\min}(X)$ satisfies the same bounds than $\lambda_{\max}(X)$).

$\square$

Hence, the main point of the proof is to show the following lemma.

Lemma 1.3.30. *Pick $\epsilon\in]\frac{1}{8},\frac{1}{4}[$. For any $x>2$ if $\beta=1,2$ and $x>\tilde{b}_\alpha$ if $\beta=w_1,w_2$, there exists θ_x such that for every $\eta>0$,*

$$\lim_{N\to\infty}\sup_{e\in V_N^\epsilon}\mathbb{P}_N^{(e,\theta_x)}[|\lambda_{\max}-x|\geq\eta]=0$$

Again, we first consider the simpler Wigner matrix case and then the case of Wishart matrices.

1.3.5.1 Proof of Lemma 1.3.30 for Wigner matrices

For $e\in V_N^\epsilon$ fixed, let $X^{(e),N}$ be a matrix with law $\mathbb{P}_N^{(e,\theta)}$. We have :

$$X^{(e),N}=\mathbb{E}[X^{(e),N}]+(X^{(e),N}-\mathbb{E}[X^{(e),N}])$$

where $\mathbb{E}[X]$ denotes the matrix with entries given by the expectation of the entries of the matrix X. We first show that $\mathbb{E}[X^{(e),N}]$ is approximately a rank one matrix.

Lemma 1.3.31. *For $\epsilon\in]\frac{1}{8},\frac{1}{4}[$, there exists $\kappa(\epsilon)>0$ so that for $e\in V_N^\epsilon$:*

$$\mathbb{E}[X^{(e),N}]=2\theta e\,e^*+\Delta^{(e),N}$$

where the spectral radius of $\Delta^{(e),N}$ is bounded by $N^{-\kappa(\epsilon)}$ uniformly on $e\in V_N^\epsilon$.

Proof of the lemma. We can express the density of $\mathbb{P}_N^{(e,\theta)}$ as the following product :

$$\frac{d\mathbb{P}_N^{(e,\theta)}}{d\mathbb{P}_{X_N}}(X) = \prod_{i\leq j} \exp(2^{1_{i\neq j}}\theta\sqrt{N}\Re(e_i\bar{e}_j a_{i,j}) - L_{\mu_{i,j}^N}(2^{1_{i\neq j}}\theta\sqrt{N}e_i\bar{e}_j))$$

where the $a_{i,j}$ are defined as in the introduction, basically a rescaling of the entries by multiplication by $\sqrt{N}$.

So since we took our $a_{i,j}$ independent (for $i \leq j$), the entries $X_{i,j}^{(e),N}$ remain independent and their mean is given in function of the Taylor expansion of L as follows :

$$(\mathbb{E}[X^{(e),N})])_{i,j} = \frac{L'_{\mu_{i,j}^N}(2\sqrt{N}\theta e_i\bar{e}_j)}{\sqrt{N}} = \frac{2\theta}{\beta}e_i\bar{e}_j + \frac{\delta_{i,j}(2\sqrt{N}\theta e_i\bar{e}_j)N\theta^2|e_i|^2|e_j|^2}{\sqrt{N}}$$

if $i \neq j$, and if $i = j$

$$(\mathbb{E}[X^{(e),N}])_{i,i} = \frac{L'_{\mu_{i,i}^N}(\sqrt{N}\theta|e_i|^2)}{\sqrt{N}} = \frac{2\theta}{\beta}e_i\bar{e}_i + \frac{\delta_{i,i}(2\sqrt{N}\theta|e_i|^2)N\theta^2|e_i|^4}{\sqrt{N}}$$

where we used that by centering and variance one, $L'_{\mu_{i,j}^N}(0) = 0$, $HessL_{\mu_{i,j}^N}(0) = \frac{1}{\beta}Id$ for all $i \neq j, N$, $L''_{\mu_{i,i}^N}(0) = \frac{2}{\beta}$ for all i, N, and where

$$|\delta_{i,j}(t)| \leq 4 \sup_{|u|<t} \max_{i,j,N}\{|L^{(3)}_{\mu_{i,j}^N}(u)|\}\,. \tag{1.18}$$

Hence, we have

$$\Delta_{i,j}^{(e),N} = \delta_{i,j}(2\sqrt{N}\theta e_i\bar{e}_j)\sqrt{N}\theta^2|e_i|^2|e_j|^2, 1 \leq i,j \leq N\,.$$

In order to bound the spectral radius of this remainder term, we use the following lemma

Lemma 1.3.32. *Let A be an Hermitian matrix and B a real symmetric matrix such that :*

$$\forall i,j, |A_{i,j}| \leq B_{i,j}\,.$$

Then the spectral radius of A is smaller than the spectral radius of B.

Proof. Indeed, if we take u on the sphere such that $||Au||_2 = ||A||$, then, by denoting A' the matrix $(|A_{i,j}|)$ and u' the vector $(|u_i|)$, we have by the triangular inequality

$$||A|| = ||Au||_2 \leq ||A'u'||_2 \leq ||Bu'||_2 \leq ||B||\,.$$

□

Therefore, if we choose C so that $C \geq \sup_{N,i,j}\delta_{i,j}(2\sqrt{N}\theta e_i\bar{e}_j)\theta^2$ and set $|e|^2$ to be the vector with entries $(|e_i|^2)_{1\leq i\leq N}$, we have

$$||\Delta^{(e),N}|| \leq C\sqrt{N}|||e|^2(|e|^2)^*||$$

Since $|||e|^2(|e|^2)^*|| = |||e|^2||_2^2 = \sum_i e_i^4 \leq N^{-4\epsilon}$, we deduce that if we take $\epsilon' \in]1/8, 1/4[$ we have with $\kappa(\epsilon) = 1/2 - 4\epsilon$:

$$||\Delta^{(e),N}|| = N^{-\kappa(\epsilon)}\,.$$

□

Remark 1.3.33. *F. Augeri noticed that a maybe more elegant proof of this point would be to use Latala's theorem [56]:*

$$\mathbb{E}[\|Y\|] \le C \sup_j \left(\mathbb{E}\sum_i |Y_{i,j}|^2\right)^{\frac{1}{2}}.$$

Now we denote :

$$\overline{X}^{(e),N} := X^{(e),N} - \mathbb{E}[X^{(e),N}].$$

The entries of $\overline{X^{(e),N}}$ are independent, centered of variance $\partial_z\partial_{\bar z}L_{\mu_{i,j}^N}(\theta e_i\bar e_j\sqrt{N})/N$. Recall that $\partial_z\partial_{\bar z}L_{\mu_{i,j}^N}(0) = 1$ and that the third derivative of the Laplace transform of the entries are uniformly bounded so that

$$\partial_z\partial_{\bar z}L_{\mu_{i,j}^N}(\theta e_i\bar e_j\sqrt{N}) = 1 + \delta_{i,j}(\sqrt{N}|e_ie_j|) = 1 + O(N^{-2\epsilon})$$

uniformly on V_N^ϵ. We can then consider $\widetilde{X}^{(e),N}$ defined by : :

$$\widetilde{X}_{i,j}^{(e),N} = \frac{\overline{X}_{i,j}^{(e),N}}{\sqrt{\partial_z\partial_{\bar z}L_{\mu_{i,j}^N}(\theta e_i\bar e_j\sqrt{N})}}.$$

Set $Y^{(e),N} = \overline{X}^{(e),N} - \widetilde{X}^{(e),N}$. So, we have

$$(Y^{(e),N})_{i,j} = \overline{X}_{i,j}^{(e),N}\left(1 - \frac{1}{\sqrt{\partial_z\partial_{\bar z}L_{\mu_{i,j}^N}(\theta e_i\bar e_j\sqrt{N})}}\right).$$

We next show that for all $\delta > 0$:

$$\lim_{N\to+\infty}\sup_{e\in V_N^\epsilon}\mathbb{P}[||Y^{(e),N}|| > \delta] = 0. \tag{1.19}$$

Indeed, we have the following lemma which is a variant of [4, Theorem 2.1.22] :

Lemma 1.3.34. *Consider for all $N \in \mathbb{N}$ a random Hermitian matrix A^N with independent subdiagonal entries which are centered and for all $k \in \mathbb{N}$:*

$$r_k^N = \max_{i,j} N^{-k/2}\mathbb{E}[|A_{i,j}^N|^k].$$

Suppose that there exists $N_0 \in \mathbb{N}, C > 0$ such that for $N \ge N_0$:

$$r_2^N \le 1, \qquad r_k^N \le k^{Ck}.$$

Then for all $\delta > 0$, $\mathbb{P}[\lambda_{\max}(A^N) > 2 + \delta]$ goes to zero as N goes to infinity.

The proof of this lemma is strictly identical to Theorem 2.1.22 in [4] as we only need to estimate large moments of the matrix, which only requires upper bounds on moments of the entries (and not equality as assumed in [4]) as soon as the entries are centered.

We apply this lemma to the matrices $Y^{(e),N}/\delta$ to derive (1.19): note that the hypothesis on the upper bound on moments is a clear consequence of our bounds on Laplace transform and are uniform for $e \in V_N^\epsilon$. Indeed, first remark that for all i, j

$$\mathbb{E}[\exp(\theta\overline{X}_{i,j}^{(e),N})] = T_{\mu_{i,j}^N}(\theta + \sqrt{N}e_ie_j)\exp(-\sqrt{N}e_ie_j - L'_{\mu_{i,j}^N}(\sqrt{N}e_ie_j\theta))$$

As a consequence, for $e \in V_N^\epsilon$, we have

$$\begin{aligned}\frac{\mathbb{E}[(\sqrt{N}X_{i,j}^{(e),N})^{2k}]}{2k!} &\le \frac{\mathbb{E}[\exp(\sqrt{N}X_{i,j}^{(e),N})] + \mathbb{E}[\exp(-\sqrt{N}X_{i,j}^{(e),N})]}{2} \\ &\le \sup_{t\in[-(1+N^{-2\epsilon}),1+N^{-2\epsilon}]} T_{\mu_{i,j}^N}(t)\exp(N^{-2\epsilon})\exp(\sup_{t\in[-N^{-2\epsilon},N^{-2\epsilon}]}|L'_{\mu_{i,j}^N}(t)|)\end{aligned}$$

which is uniformly bounded by Assumption 1.3.8. Therefore we have a found a finite constant $C > 0$ such that

$$\sup_{e\in V_N^\epsilon}\sup_{i,j}\mathbb{E}[(\sqrt{N}X_{i,j}^{(e),N})^2k] \le C(2k)! \tag{1.20}$$

Furthermore, by the same arguments as in (1.18), we get

$$\lim_{N\to\infty}\sup_{e\in V_N^\epsilon}\left(1 - \frac{1}{\sqrt{\partial_z\partial_{\bar z}L_{\mu_{i,j}^N}(\theta e_i\bar e_j\sqrt{N})}}\right) = 0$$

which readily implies that for any $\delta > 0$, for N large enough, all $k \ge 1$,

$$\sup_{e\in V_N^\epsilon}\sup_{i,j}\mathbb{E}[(\frac{\sqrt{N}Y_{i,j}^{(e),N}}{\delta})^2k] \le C(2k)! \tag{1.21}$$

Hence, we can apply the previous lemma to deduce that

$$\lim_{N\to\infty}\sup_{e\in V_N^\epsilon}\mathbb{P}[||Y^{(e),N}/3\delta|| \ge 3] = 0$$

Hence, since

$$X^{(e),N} = \widetilde{X}^{(e),N} + \frac{2\theta}{\beta}ee^* + \Delta^{(e),N} + Y^{(e),N}\,,$$

we conclude by combining (1.19) and Lemma 1.3.31 that for $\epsilon \in]1/4, 1/8[$ and all $\delta > 0$

$$\lim_{N\to\infty}\sup_{e\in V_N^\epsilon}\mathbb{P}_N^{(e,\theta)}[||X^{(e),N} - (\widetilde{X}^{(e),N} + \frac{2\theta}{\beta}ee^*)|| > \delta] = 0 \tag{1.22}$$

since all estimates were clearly uniform on $e \in V_N^\epsilon$.

And so, to conclude we need only to identify the limit of $\lambda_{\max}(\widetilde{X}^{(e),N} + \frac{2\theta}{\beta}ee^*)$. It is given by the well known BBP transition. We collect below the main elements of the argument for completness. To identify this limit, we easily see as in [18] that the eigenvalues of $\widetilde{X}^{(e),N} + \frac{2\theta}{\beta}ee^*$ satisfy

$$0 = \det(z - \widetilde{X}^{(e),N} - \frac{2\theta}{\beta}ee^*) = \det(z - \widetilde{X}^{(e),N})\det(1 - \frac{2\theta}{\beta}(z - \widetilde{X}^{(e),N})^{-1}ee^*)$$

and therefore z is an eigenvalue away from the spectrum of $\widetilde{X}^{(e),N}$ iff

$$\langle e, (z - \widetilde{X}^{(e),N})^{-1}e\rangle = \frac{\beta}{2\theta}\,.$$

To take the large N limit we will use here the very powerful (and hard) results of [21] but will follow a more pedestrian moment approach in the Wishart case for completness. It was indeed shown in Theorem 2.15 of [21] that for all z with $\Im z > 0$, all $v \in \mathbb{S}^{N-1}$, $\langle v, (z - \widetilde{X}^{(e),N})^{-1}v\rangle$ converges almost surely towards $G_\sigma(z)$ This convergence extends to $z > 2$ first by noticing that G_σ is continuous when z goes to the real axis. And moreover, for any $\epsilon > 0$, N large enough the operator norm of $\widetilde{X}^{(e),N}$ is bounded by $2 + \epsilon$

by Kòmlos-Füredi's theorem [4] so that $\langle v, (z-\widetilde{X}^{(e),N})^{-1}v\rangle$ is continuous in $B(0,2+\epsilon)^c$.Therefore we conclude that the largest eigenvalue $\lambda_{\max}(\widetilde{X}^{(e),N}+\frac{2\theta}{\beta}ee^*)$, must converge towards the solution ρ_θ to

$$G_\sigma(\rho_\theta)=\frac{\beta}{2\theta}$$

as soon as it is strictly greater than 2. We find a unique solution to this equation: it is given by

$$\rho_\theta=\frac{2\theta}{\beta}+\frac{\beta}{2\theta}\,.$$

Reciprocally, for any $x>2$, we can find $\theta_x=\frac{\beta}{2}(x+\sqrt{x^2-4})$ so that $x=\rho_{\theta_x}$. Hence, we have proved that for any sequence of vectors $e\in V_N^\epsilon$ we have the desired estimate for any $\eta>0$

$$\lim_{N\to\infty}\sup_{e\in V_N^\epsilon}\mathbb{P}_N^{(e,\theta_x)}[|\lambda_{\max}-x|\geq\eta]=0$$

which also entails the convergence of the supremum over V_N^ϵ and thus the Lemma.

1.3.5.2 Proof of Lemma 1.3.30 for Wishart matrices

We next prove Lemma 1.3.30 for Wishart matrices and fix $e=(e^{(1)},e^{(2)})\in V_N^\epsilon$. We decompose as in the previous proof

$$X^{(e),N}=\widetilde{X}^{(e),N}+\mathbb{E}[X^{(e),N}]+Y^{(e),N}\,,$$

where the entries of $\widetilde{X}^{(e),N}$ are centered and with covariance $1/N$ and $Y^{(e),N}$ goes to zero in norm. We then find by the same argument that

$$\mathbb{E}[X^{(e),N}]=\frac{2\theta}{i}\begin{pmatrix}0 & e^{(1)}(e^{(2)})^*\\ e^{(2)}(e^{(1)})^* & 0\end{pmatrix}+\Delta^{(e),N}$$

where $\|\Delta^{(e),N}\|\leq N^{-\kappa(\epsilon)}$ and $e^{(1)}$ (resp. $e^{(2)}$) is the vector made of the first L (resp. M last) coordinates of e. Letting

$$S^{(e)}=\begin{pmatrix}e^{(1)} & 0\\ 0 & e^{(2)}\end{pmatrix}\quad\text{and } T^{(e)}=\begin{pmatrix}0 & (e^{(2)})^*\\ (e^{(1)})^* & 0\end{pmatrix}$$

we notice that

$$\begin{pmatrix}0 & e^{(1)}(e^{(2)})^*\\ e^{(2)}(e^{(1)})^* & 0\end{pmatrix}=S^{(e)}T^{(e)}\,.$$

Therefore, we need to find $z>\tilde{b}_\alpha$ such that

$$0=\det(z-\widetilde{X}^{N,(e)}-\frac{2\theta}{i}S^{(e)}T^{(e)})=\det(z-\widetilde{X}^{N,(e)})\det(1-\frac{2\theta}{i}T^{(e)}(z-\widetilde{X}^{N,(e)})^{-1}S^{(e)})\,. \tag{1.23}$$

By writing $R_{\widetilde{X}^{N,(e)}}(z)=(z-\widetilde{X}^{N,(e)})^{-1}$ by blocks with $\widetilde{X}^{N,(e)}$ with upper right $L\times M$ block $\widetilde{G}^{N,(e)}$, we get :

$$R_{\widetilde{X}^{N,(e)}}(z)=\begin{pmatrix}R_{1,1}(z) & R_{1,2}(z)\\ R_{2,1}(z) & R_{2,2}(z)\end{pmatrix}=\begin{pmatrix}zR_{\widetilde{G}^{N,(e)}(\widetilde{G}^{N,(e)})^*}(z^2) & \widetilde{G}^{N,(e)}R_{(\widetilde{G}^{N,(e)})^*\widetilde{G}^{N,(e)}}(z^2)\\ R_{(\widetilde{G}^{N,(e)})^*\widetilde{G}^{N,(e)}}(z^2)(\widetilde{G}^{N,(e)})^* & zR_{(\widetilde{G}^{N,(e)})^*\widetilde{G}^{N,(e)}}(z^2)\end{pmatrix}$$

where $R_{1,1}$ is $L\times L$, $R_{1,2}$ $L\times M$, $R_{2,2}$ $M\times M$, we get the simpler equation

$$\det\left(I-\frac{2\theta}{i}\begin{pmatrix}\langle e^{(2)},R_{2,1}(z)e^{(1)}\rangle & \langle e^{(2)},R_{2,2}(z)e^{(2)}\rangle\\ \langle e^{(1)},R_{1,1}(z)e^{(1)}\rangle & \langle e^{(1)},R_{1,2}(z)e^{(2)}\rangle\end{pmatrix}\right)=0$$

Therefore, we need to find z such that

$$|1-\frac{2\theta}{i}\langle e^{(2)},R_{2,1}(z)e^{(1)}\rangle|^2-\frac{4\theta^2}{i^2}\langle e^{(2)},R_{2,2}(z)e^{(2)}\rangle\langle e^{(1)},R_{1,1}(z)e^{(1)}\rangle=0 \tag{1.24}$$

We are going to prove that

Lemma 1.3.35. *For any $\delta,\varepsilon>0$*

$$\begin{aligned}
\limsup_{N\to\infty}\sup_{e\in V_N^\epsilon}\mathbb{P}_N^{(e,\theta)}\left(\sup_{z\geq\tilde{b}_\alpha+\varepsilon}|\langle e^{(1)},R_{1,1}(z)e^{(1)}\rangle-z(1+\alpha)||e^{(1)}||_2^2G_{MP(\alpha)}((1+\alpha)z^2)|>\delta\right) &= 0\,,\\
\limsup_{N\to\infty}\sup_{e\in V_N^\epsilon}\mathbb{P}_N^{(e,\theta)}\left(\sup_{\Im z\geq\tilde{b}_\alpha+\varepsilon}|\langle e^{(2)},R_{2,2}(z)e^{(2)}\rangle-z(1+\alpha)||e^{(2)}||_2^2G_{MP(1/\alpha)}((1+\alpha)z^2)|>\delta\right) &= 0\,,\\
\limsup_{N\to\infty}\sup_{e\in V_N^\epsilon}\mathbb{P}_N^{(e,\theta)}\left(\sup_{\Im z\geq\tilde{b}_\alpha+\varepsilon}|\langle e^{(2)},R_{2,1}(z)e^{(1)}\rangle|>\delta\right) &= 0\,,
\end{aligned}$$

where $G_{MP(\alpha)}$ is the Stieltjes transform of a Pastur Marchenko law with parameter α.

We first derive Lemma 1.3.30 assuming that Lemma 1.3.35 holds. We have seen in Lemma 1.3.28 that $\|e^{(1)}\|_2$ converges towards $x_{\theta,\alpha}$ almost surely. Therefore, we arrive to the limiting equation

$$(1+\alpha)^2z^2G_{MP(\alpha)}((1+\alpha)z^2)G_{MP(1/\alpha)}((1+\alpha)z^2)=\frac{i^2}{4\theta^2x_{\theta,\alpha}(1-x_{\theta,\alpha})}\,.$$

Now, we claim that $\varphi(\theta)=\theta^2x_{\theta,\alpha}(1-x_{\theta,\alpha})$ is continuous, increasing, going from 0 to $+\infty$. As $x_{\theta,\alpha}$ is a complicated solution of θ (solution of a degree three polynomial equation), we use the following asymptotic characterization which easily follows from the previous large deviation considerations, see Lemma 1.3.28:

$$\frac{4\theta}{i}x_{\theta,\alpha}(1-x_{\theta,\alpha})=\partial_\theta F(\theta,w_i)\,,$$

where we use that the derivatives of $x_{\theta,\alpha}$ vanishes as it is a critical point of the maximum. We moreover notice that $G(\theta)=F(\sqrt{\theta},w_i)$ is convex in θ (as a supremum of convex functions). Hence,

$$\varphi(\theta)=\frac{i}{4}\theta\partial_\theta F(\theta,w_i)=\frac{i}{2}\theta^2G'(\theta^2)\,.$$

It follows that φ is smooth as F is and moreover

$$\varphi'(\theta)=i(\theta G'(\theta^2)+\theta^3G''(\theta))\,.$$

But since φ is non negative, G' is non negative and so φ' is non negative for all $\theta\geq 0$. The fact that φ goes to infinity at infinity is clear as $x_{\theta,\alpha}$ then goes to 1/2. Moreover, for $z>\tilde{b}_\alpha$, $z\mapsto zG_{MP(\alpha)}((1+\alpha)z^2)$ and $z\mapsto zG_{MP(1/\alpha)}((1+\alpha)z^2)$ are positive and decreasing, and therefore so are their product. Hence, there exist a $\theta_\alpha>0$ so that for every $\theta\geq\theta_\alpha$, the equation above has a unique solution on $[\tilde{b}_\alpha,+\infty[$. Moreover, if we denote ρ_θ this solution, $\theta\mapsto\rho_\theta$ is a bijection from $[\theta_\alpha,+\infty[$ onto $[\tilde{b}_\alpha,+\infty[$.

Proof of Lemma 1.3.35. The first two results could be derived from [21] but we here provide a more pedestrian moment approach to achieve this result, a strategy that we could have also followed in the Wigner case. We recall that $\tilde{G}=G_{L,M}$ is a $L\times M$ matrix with centered entries with covariance one and sub-Gaussian tails, $e=(e^{(1)},e^{(2)})$ a unit vector and

$$R_{1,1}(z)=(z^2-\tilde{G}\tilde{G}^*)^{-1},R_{22}(z)=(z^2-\tilde{G}^*\tilde{G})^{-1},R_{1,2}(z)=\tilde{G}(z^2-\tilde{G}^*\tilde{G})^{-1}.$$

In fact $\tilde{G}$ depends a priori on e but we will derive uniform bounds in $e \in V_N^\epsilon$ in the following and we will work conditionnally to e. Moreover to simplify the notations we denote $\tilde{G}$ by G.

The first two points of the Lemma are direct consequences of [21, Theorem 2.5]. It remains to see that $\langle e^{(2)}, R_{2,1}(z)e^{(1)}\rangle$ goes to 0 as N goes to infinity. Because $R_{2,1}(z) = G(z - G^*G)^{-1}$ is not the resolvent of the Wishart matrix, but its multiplication by G, we cannot apply directly [21, Theorem 2.5]. We will give an elementary proof of this result,based on classical moment computations. Indeed, for $\varepsilon > 0$, we have already seen thanks to (1.21) and Lemma 1.3.34 that

$$\lim_{N\to\infty} \sup_{e\in V_N^\epsilon} \mathbb{P}\left(\|G^*G\| \geq b_\alpha + \varepsilon\right) = 0\,.$$

Moreover, on the set where $\{\|G^*G\| \leq b_\alpha + \varepsilon\}$, for $z > b_\alpha + 2\varepsilon$ we can expand

$$\begin{aligned}
\langle e^{(2)}, R_{2,1}(z)e^{(1)}\rangle &= -\sum_{k=0}^{\infty} \frac{\langle e^{(1)}, G(G^*G)^k e^{(2)}\rangle}{z^{2k}} \\
&= -\sum_{k=0}^{K} \frac{\langle e^{(1)}, G(G^*G)^k e^{(2)}\rangle}{z^{2k+2}} + O\left(\frac{1}{\varepsilon}\left(\frac{b_\alpha+\varepsilon}{b_\alpha+2\varepsilon}\right)^{K+1}\right)
\end{aligned}$$

and hence it is enough to get the convergence in probability of K moments with $K \geq 2\varepsilon^{-1}\ln\varepsilon^{-1}$, uniformly in e :

$$\lim_{N\to\infty} \sup_{e\in V_N^\epsilon} \mathbb{P}\left(\sup_{0\leq k\leq K} |\langle e^{(1)}, G(G^*G)^k e^{(2)}\rangle| \geq \delta\right) = 0\,.$$

To this end, it is enough to prove by Tchebycjhev's inequality that :

$$\lim_{N\to\infty} \sup_{e\in V_N^\epsilon} |\mathbb{E}[\langle e^{(1)}, G(G^*G)^k e^{(2)}\rangle]| = 0\,, \tag{1.25}$$

and then

$$\lim_{N\to\infty} \sup_{e\in V_N^\epsilon} Var(\langle e^{(1)}, G(G^*G)^k e^{(2)}\rangle) = 0\,. \tag{1.26}$$

We first prove (1.25). It is clearly true for $k = 0$ by centering of the entries and so we consider $k \geq 1$. Let's call $\mathcal{W}_{2k+1}$ the set of words $(v_1, ..., v_{2k+2})$ of length $2k+1$ so that $v_{2j} \in \{1, ..., L\}$ and $v_{2j+1} \in \{1, ..., M\}$. We use the following notation :

$$E_v = \mathbb{E}[a_{v_1,v_2}a_{v_2,v_3}...a_{v_{2k+1},v_{2k+2}}]\,.$$

We have

$$\mathbb{E}[\langle e^{(1)}, G(G^*G)^k e^{(2)}\rangle] = \frac{1}{N^{k+1/2}} \sum_{v\in\mathcal{W}_{2k+1}} e^{(1)}_{v_1} E_v e^{(2)}_{v_{2k+2}}\,.$$

Given a word v, we can construct a bipartite graph G_v whose vertices are the $\{v_1, v_3, ...\} \cup \{L + v_2, L + v_4, ...\}$ of whose edges (occasionally multiple) are the $(L + v_{2i}, v_{2i-1})$ and $(L + v_{2i}, v_{2i+1})$. We denote $V^{(1)}(v)$ the number of vertices in G_v lying in $\{1, ..., L\}$, $V^2(v)$ the number of vertices in G_v lying in $\{L + 1, ..., L + M\}$ and $V(v) = V^{(1)}(v) + V^2(v)$ and $A(v)$ the number of edges of G_v. If e is an edge of G_v, we denote $n_v(e)$ the multiplicity of this edge.

Let's recall that here the entries $a_{i,j}$ of G are independent but not identically distributed, with distribution eventually depending on e. Nevertheless their variance are 1 and their moments are bounded uniformly i.e. for every k there exists $C_k < +\infty$, independent of $e \in V_N^\epsilon$ (see (1.21)) such that :

$$\sup_{N,i,j} \mathbb{E}[|a_{i,j}|^k] \leq C_k\,.$$

For every word v of length k, we can define $C_v = \prod_{j\leq k} C_j^{l(v,j)}$ where $l(v,j)$ is the number of edge of multiplicity j in G_v. we then have

$$|E_v| \leq C_v$$

We say that two words v, w are equivalent if there exists a bijection $\varphi : \{1, ..., L\} \to \{1, ..., M\}$ and a bijection $\psi : \{1, ..., M\} \to \{1, ..., M\}$ such that $v_{2j} = \varphi(w_{2j})$ and $v_{2j+1} = \psi(w_{2j+1})$. If two words v and w are equivalent then $C_v = C_w$.

Let $\mathcal{T}_{2k+1}$ be a the quotient set of words of length $2k+1$ for this equivalency relationship. We have

$$\mathbb{E}[\langle e^{(1)}, G(G^*G)^k e^{(2)}\rangle] \quad = \quad \frac{1}{N^{k+1/2}} \sum_{j=2}^{2k+2} \sum_{t\in\mathcal{T}_{2k+1}, V(v)=j} \sum_{v|v\sim t} e^{(1)}_{v_1} E_v e^{(2)}_{v_{2k+2}} .$$

Notice that if G_v has an edge of multiplicity 1, then $E_v = 0$ (since the $a_{i,j}$ are independant and centered). So for E_v to be non zero we need that $A(v) \leq (2k+1)/2$ so $A(v) \leq k$. Since G_v is connected $V(v) \leq A(v)+1 \leq k+1$. If $v \in \mathcal{W}_{2k+1}$, there exists $N_v := (L-1)...(L-V^{(1)}(v)+1)(M-1)...(M-V^2(v)+1) \leq N^{V(v)-2}$ equivalent words w_1 provided we fix v_1 and v_{2k+2} so we have the following bound :

$$\mathbb{E}[\langle e^{(1)}, G(G^*G)^k e^{(2)}\rangle] \leq \qquad \frac{1}{N^{k+1/2}} \sum_{j=2}^{k+1} \sum_{t\in\mathcal{T}_{2k+1}, V(t)=j} C_t N_t \sum_{1\leq v_1\leq L, 1\leq v_{2k+2}\leq M} |e^{(1)}_{v_1} e^{(2)}_{v_{2k+2}}| .$$

By using the Cauchy Schwartz inequality, we have that :

$$\sum_{1\leq i\leq L, 1\leq j\leq M} |e^{(1)}_i e^{(2)}_j| \leq N \|e^{(1)}||_2 \times ||e^{(2)}||_2 \leq N ,$$

which yields

$$\mathbb{E}[\langle e^{(1)}, G(G^*G)^k e^{(2)}\rangle] \quad \leq \quad \frac{1}{N^{k-1/2}} \sum_{j=2}^{k+1} \sum_{t\in\mathcal{T}_{2k+1}, V(t)=j} C_t N^{j-2} .$$

The leading order term here is in $N^{-1/2}$ for $k \geq 1$ and so

$$\lim_{N\to\infty} \sup_{\|e\|_2=1} |\mathbb{E}[\langle e^{(1)}, G(G^*G)^k e^{(2)}\rangle]| = 0 .$$

We proceed similarly for the covariance (1.26):

$$\mathrm{Var}(\langle e^{(1)}, G(G^*G)^k e^{(2)}\rangle) = \frac{1}{N^{2k+1}} \sum_{v\in\mathcal{W}_{2k+1}, w\in\mathcal{W}_{2k+1}} e^{(1)}_{v_1} e^{(1)}_{w_1} T_{v,w} e^{(2)}_{v_{2k+2}} e^{(2)}_{w_{2k+2}} ,$$

where $T_{v,w} = E_{v,w} - E_v E_w$ and $E_{v,w} = \mathbb{E}[a_{v_1,v_2} a_{v_2,v_3} ... a_{v_k,v_{k+1}} a_{w_1,w_2} a_{w_2,w_3} ... a_{w_k,w_{k+1}}]$ We extend naturally the previous definitions to couples of words. Let us now do the same analysis than before with couples of words. Let's take $\tilde{\mathcal{T}}_{2k+1}$ the quotient set for the equivalency relationship for couples of words. Let $(v, w) \in \tilde{\mathcal{T}}_{2k+1}$

First, if $G_{v,w}$ is not connected, since it is the union of two connected graphs G_v and G_w, we have that G_v and G_w don't have any edges in common and so, by independence of the entries $T_{v,w} = 0$. So we can assume that $G_{v,w}$ is connected.

Then several cases arise :

First, if $v_1 \neq w_1$ and $v_{2k+2} \neq w_{2k+2}$, then if one edge of $G_{v,w}$ is of multiplicity 1, then $T_{v,w} = 0$. So we can assume that all edges are of multiplicity at least 2. We deduce that $A(v, w) \leq 2k+1$ and $V(v, w) \leq$

$2k+2$. Let $N_{v,w}$ be the number of couple of words equivalent to (v,w) provided $(v_1, w_1, v_{2k+2}, w_{2k+2})$ are fixed, we have $N_{v,w} \leq N^{2k-2}$. Hence

$$\sum_{(u,t)\sim(v,w)} e^{(1)}_{u_1} e^{(1)}_{t_1} T_{v,w} e^{(2)}_{u_{2k+2}} e^{(2)}_{t_{2k+2}} \leq N^{2k}(C_{v,w} - C_v C_w)\,.$$

Then, if $v_1 = w_1$ and $v_{2k+2} \neq w_{2k+2}$ or if $v_1 \neq w_1$ and $v_{2k+2} = w_{2k+2}$, the same reasoning concerning the edges holds. So, we have $V(v,w) \leq 2k+2$ and if $N_{v,w}$ is the number of couple of words equivalent to (v,w) provided $(v_1, w_1, v_{2k+2}, w_{2k+2})$ are fixed, we have $N_{v,w} \leq N^{2k-1}$. If we are in the case $v_1 = w_1$:

$$\sum_{(u,t)\sim(v,w)} e^{(1)}_{u_1} e^{(1)}_{t_1} T_{v,w} e^{(2)}_{u_{2k+2}} e^{(2)}_{t_{2k+2}} \leq N^{2k}||e^{(1)}||^2(C_{v,w} - C_v C_w)\,.$$

And lastly, if $v_1 = w_1$ and $v_{2k+2} = w_{2k+2}$, we have again $N_{v,w} \leq N^{2k}$ and

$$\sum_{(u,t)\sim(v,w)} e^{(1)}_{u_1} e^{(1)}_{t_1} T_{v,w} e^{(2)}_{u_{2k+2}} e^{(2)}_{t_{2k+2}} \leq N^{2k}||e^{(1)}||^2||e^{(2)}||^2(C_{v,w} - C_v C_w)\,.$$

Hence, we conclude that

$$\mathrm{Var}(\langle e^{(1)}, G(G^*G)^k e^{(2)}\rangle) = O\left(\frac{1}{N}\right)\,.$$

□

1.3.6 Appendix: Proof of Lemma 1.3.16

In this section, we want to prove that if $\mu_{i,j}$ are supported inside a common compact K or satisfy a log-Sobolev inequality with a uniformly bounded constant c, the empirical measure of the eigenvalues of the matrices $X_N^{(1)}, X_N^{(2)}, X_N^{(w_1)}, X_N^{(w_2)}$ concentrates as stated in Lemma 1.3.16. To this end, we will use two concentration results respectively from [49] and [9].

Theorem 1.3.36. *By [49, Theorem 1.4)] (for the compact case) and [49, Corollary 1.4 b)] (for the logarithmic Sobolev case), we have for $\beta = 1, 2, w_1, w_2$, and for N large enough*

$$\limsup_{N\to\infty} \frac{1}{N^{7/6}} \ln \mathbb{P}[d(\mu_{X_N^{(\beta)}}, \mathbb{E}[\mu_{X_N^{(\beta)}}]) > N^{-1/6}] < 0$$

where d is the Dudley distance.

We therefore only need to show that

Theorem 1.3.37. *([9, Theorem 4.1]) If we let for every N :*

$$F_{X_N^{(1)}}(x) = \mu_{X_N^{(1)}}(]-\infty, x])\,,$$
$$F_{\sigma_1}(x) = \sigma_1(]-\infty, x])\,,$$

we have that

$$\sup_{x\in\mathbb{R}} |F_{\sigma_1}(x) - \mathbb{E}[F_{X_N^{(1)}}(x)]| = O(N^{-1/4})\,.$$

In order to conclude, we need only to use Lemma 1.3.12 to see that $F_{X_N^{(1)}}(-M)$ and $1 - F_{X_N^{(1)}}(M)$ decay exponentially fast in N for some fixed M so that

$$d(\mathbb{E}[\mu_{X_N^{(1)}}], \sigma_1) \leq 4e^{-N}\|f\|_\infty + 2M\|f\|_L \sup_{x\in\mathbb{R}} |F(x) - \mathbb{E}[F_{X_N^{(1)}}(x)]| = o(N^{-1/6})\,.$$

The same results hold in the complex case, see e.g. [12, (8.1.3)]. For Wishart matrices, we rely on [10, Theorem w.1 and w.2]. Recall that $W_N = G_{L,M}(G_{L,M})^*$.

Theorem 1.3.38. *([10]) Assume that $M/N \in (1, \epsilon^{-1})$ for some fixed ϵ and M/N converges towards α. Then*

$$\sup_{x\in\mathbb{R}} |F_{\pi_\alpha}(x) - \mathbb{E}[F_{W_N}(x)]| = O(N^{-1/10}).$$

We can as well use Lemma 1.3.12 to conclude that $1 - \mathbb{E}[F_{W_N}(M)]$ goes to zero like e^{-N} for M large enough. Finally, we conclude by noticing that since

$$\int f(x) d\mathbb{E}[\hat{\mu}_{X_N^w}](x) = \frac{N}{N+M} \int (f(\sqrt{\lambda}) + f(-\sqrt{\lambda})) d\hat{\mu}_{W_N}(\lambda) + \frac{M-N}{N} f(0),$$

we have

$$\begin{aligned} \left| \int f(x) d(\mathbb{E}[\hat{\mu}_{X_N^w}] - \sigma_w)(x) \right| &\leq \|f\|_\infty (|\frac{M}{N} - \alpha| + e^{-N}) \\ &\quad + \int_0^M |\partial_\lambda f(\sqrt{\lambda})| |F_{\pi_\alpha}(\lambda) - \mathbb{E}[F_{W_N}(\lambda)]| d\lambda \\ &\leq \|f\|_L (N^{-\kappa} + e^{-N} + 2MN^{-\frac{1}{10}}). \end{aligned}$$

1.4 Large deviations for the largest eigenvalue of matrices with variance profiles

1.4.1 Introduction

In random matrix theory, large deviation principles for quantities related to the spectrum are usually hard to prove. In the case of random Wigner matrices with Gaussian entries, that is matrices from the GUE, GOE and Gaussian Wishart matrices, the explicit formulas for the joint distributions of the eigenvalues can be used to establish large deviation principles for the empirical measure and for the largest eigenvalue (see [16, 17] for the Wigner case and [68, 35] for the Wishart case). But in the general case, since eigenvalues are complex functions of the entries, such large deviation principles are rather scarce. There has been nevertheless large deviation-type lower bounds in compactly supported and log-Sobolev settings by A. Guionnet and O. Zeitouni [49], several recent breakthroughs related for instance to matrices with entries with heavy-tailed distributions both for the empirical measure and the largest eigenvalue respectively by C. Bordenave and P. Caputo and by F. Augeri ([5, 23]), and finally a more general result for the largest eigenvalue of matrices with Rademacher-distributed entries by A. Guionnet and the author in [45]. In this article, we will use the techniques developed in [45] and apply them to Wigner matrices with variance profiles (also called sometimes "Wigner-type matrices"). The entries of these matrices verify the same hypothesis as Wigner matrices except their variances may not be equal to $N^{-1/2}$ or $2N^{-1/2}$ and the matrices of the entries variance converges macroscopically to a function on $[0,1]^2$. Such matrices and the limit behaviour of their empirical measures have been thoroughly examined for instance in [43]. In particular the Stieljes transform of this limit measure is controlled by the so-called canonical equation. This canonical equation has also been extensively studied for instance in [43, 3]. In the course of this paper, we will adapt the methods of [45] for the setting of Wigner-type matrices and prove a large deviation principle for the largest eigenvalue under some assumptions for the variance profile. First we will recall in section 1.4.2 the convergence results of the empirical measure we will need during the proof of the large deviation principle. Then in section 1.4.3 we will introduce the rate function and the assumption on the variance profile we will need in order for our result to work. In sections 1.4.4 to 1.4.7 we will treat the case of matrices with piecewise constant variance profile which bears the most similarities with the models treated in [45]. In section 1.4.8 we will approximate the case of a continuous variance profile with piecewise constant ones. In section 1.4.9 we will illustrate the cases where our result applies in the simple context of a piecewise constant variance profile with four blocks. In the same section we will illustrate the limits of our approach and the necessity to make some assumptions concerning the variance profiles, with an example of a matrix whose variance profile does not verifies our assumptions and such that the rate function for the large deviations of the largest eigenvalue does not match our rate function.

1.4.1.1 Variance profiles

In the rest of the article, a real x is said to be positive if $x \geq 0$ and $\mathbb{R}^+$ is the set $\{x \in \mathbb{R} : x \geq 0\}$. Our random matrix matrix model will be of the form $W_N \odot \Sigma_N$, where W_N is either a real or a complex Wigner matrix, Σ_N is a real symmetric matrix and $\odot$ is the entrywise product. First of all, we describe the matrices Σ_N we will be using. These matrices will converge as piecewise constant functions of $[0,1]^2$ to some function σ on $[0,1]^2$ we will call a variance profile. We will consider here two cases of variance profiles : the case where it is piecewise constant and the case where it is continuous.

Piecewise constant variance profile : We consider a variance profile piecewise constant on rectangular blocks. Let $n \in \mathbb{N}^*$, $\Sigma = (\sigma_{i,j})_{i,j \in [1,n]}$ a real symmetric $n \times n$ matrix with positive coefficients and $\vec{\alpha} = (\alpha_1, ..., \alpha_n) \in \mathbb{R}^n$ such that $\alpha_i > 0$ and $\alpha_1 + ... + \alpha_n = 1$. In this context we'll consider Σ_N defined by block by :

$$\Sigma_N(i,j) = \sigma_{k,l} \text{ if } Ni \in I_k \text{ and } Nj \in I_l$$

where :

$$\gamma_j := \sum_{i=1}^{j} \alpha_i \text{ and } I_i = [\gamma_{i-1}, \gamma_i[$$

.

We shall also denote $\sigma : [0,1]^2 \to \mathbb{R}^+$ the function piecewise constant defined by

$$\sigma(x,y) = \sigma_{k,l} \text{ if } (x,y) \in I_k \times I_l$$

.

This case will be referred as the case of a piecewise constant variance profile associated to the parameters Σ and $\vec{\alpha}$.

Continuous variance profile : In this case, we will consider a real non-negative symmetric continuous function $\sigma : [0,1]^2 \to \mathbb{R}^+$ and

$$\Sigma_N(i,j) := \sigma\left(\frac{i}{N}, \frac{j}{N}\right)$$

.

In both cases, we will call σ the variance profile of the matrix model.

1.4.1.2 The generalized Wigner matrix model

The real symmetric case : We consider a family of independent real random variables $(a_{i,j}^{(1)})_{0 \leq i \leq j \leq N}$, such that the variables $a_{i,j}^{(1)}$ are distributed according to the laws $\mu_{i,j}^N$. We moreover assume that the $\mu_{i,j}^N$ are centered :

$$\mu_{i,j}^N(x) = \int x d\mu_{i,j}^N(x) = 0$$

and with covariance :

$$\mu_{i,j}^N(x^2) = \int x^2 d\mu_{i,j}^N(x) = 1, \forall 1 \leq i < j \leq N, \qquad \mu_{i,i}^N(x^2) = 2, \quad \forall 1 \leq i \leq N\,.$$

We say that a probability measure μ has a sharp sub-Gaussian Laplace transform iff

$$\forall t \in \mathbb{R}, T_\mu(t) = \int \exp\{tx\} d\mu(x) \le \exp\{\frac{t^2\mu(x^2)}{2}\}. \quad (1.27)$$

The terminology "sharp" comes from the fact that for t small, we must have

$$T_\mu(t) \ge \exp\{\frac{t^2\mu(x^2)}{2}(1+o(t))\}.$$

Hypothèse 1.4.1 (A0). *We assume that the $\mu^N_{i,j}$ have a sharp sub-Gaussian Laplace transform.*

Again thanks to the remark 1.3.2, we have that if the $\mu^N_{i,j}(x^2)$ are bounded, for every $\delta > 0$, there exists $\epsilon > 0$ that does not depend on the laws $\mu^N_{i,j}$ such that for $|t| \le \epsilon$.

$$T_{\mu^N_{i,j}}(t) \ge \exp\{\frac{(1-\delta)t^2\mu^N_{i,j}(x^2)}{2}\}.$$

and also that the $T_{\mu^N_{i,j}}$ are uniformly C^3 in a neighbourhood of the origin : for $\epsilon > 0$ small enough $\sup_{|t|\le\epsilon}\sup_{i,j,N}|\partial_t^3 \ln T_{\mu^N_{i,j}}(t)|$ is finite.

We will also need that the empirical measure of the eigenvalues concentrates in a stronger scale than N. To this end we will also make the following classical assumptions to use standard concentration of measure tools.

Hypothèse 1.4.2. *There exists a compact set K such that the support of all $\mu^N_{i,j}$ is included in K for all $i,j \in \{1,\dots,N\}$ and all integer number N, or all $\mu^N_{i,j}$ satisfy a log-Sobolev inequality with the same constant c independent of N. In the complex case, we will suppose also that for all (i,j), if Y is a random variable of law $\mu_{i,j}$, there is a complex $a \neq 0$ such that $\Re(aY)$ and $\Im(aY)$ are independent.*

Given the family $(a^{(1)}_{i,j})$, we define the following Wigner matrices :

$$W^{(1)}_N(i,j) = \begin{cases} \frac{a^{(1)}_{i,j}}{\sqrt{N}} \text{ when } i \le j, \\ \frac{a^{(1)}_{j,i}}{\sqrt{N}} \text{ when } i > j. \end{cases}$$

From this definition we define $X^{(1)}_N$ a real matrix with variance profile V_N as :

$$X^{(1)}_N := W^{(1)}_N \odot \Sigma_N$$

where for two matrices $A = (a_{i,j})_{i,j\in[1,n]}, B = (b_{i,j})_{i,j\in[1,n]}$, $A \odot B$ is the matrix $(a_{i,j}b_{i,j})_{i,j\in[1,n]}$.

We denote $\lambda_{min}(X^{(1)}_N) = \lambda_1 \le \lambda_2 \dots \le \lambda_N = \lambda_{\max}(X^{(1)}_N)$ the eigenvalues of $X^{(1)}_N$.

The complex Hermitian case : We now consider a family of independent random variables $(a^{(2)}_{i,j})_{1\le i\le j\le N}$, such that the variables $a^{(2)}_{i,j}$ are distributed according to a law $\mu^N_{i,j}$ when $i \le j$, which are centered probability measures on $\mathbb{C}$ (and on $\mathbb{R}$ if $i = j$). We write $a^{(2)}_{i,j} = x_{i,j} + iy_{i,j}$ where $x_{i,j} = \Re(a^{(2)}_{i,j})$ and $y_{i,j} = \Im(a^{(2)}_{i,j})$. We suppose that for all $i \in [1,N]$, $y_{i,i} = 0$. In this context, for a probability measure on $\mathbb{C}$, we will consider its Laplace transform to be the function :

$$T_\mu(z) := \int \exp\{\Re(a\bar{z})\} d\mu(a).$$

We assume that :

Hypothèse 1.4.3 (A0c). *For all $i < j$*

$$\forall t \in \mathbb{C}, T_{\mu^N_{i,j}}(t) \le \exp(|t|^2/4)$$

and for all i

$$\forall t \in \mathbb{R}, T_{\mu^N_{i,i}}(t) \le \exp(t^2/2).$$

The same uniform lower bounds and $\mathcal{C}^3$ character as in the real case are also implied by this bound. Observe that the above assumption implies that for all $i < j$, $2\mathbb{E}[x_{i,j}^2] = 2\mathbb{E}[y_{i,j}^2] = \mathbb{E}[x_{i,i}^2] = 1$ and $\mathbb{E}[x_{i,j}y_{i,j}] = 0$. Examples of distributions satisfying Assumption 1.4.3 are given by taking $(x_{i,j}, y_{i,j})$ centered independent variables with law satisfying a sharp sub-Gaussian Laplace transform. We then construct for all $N \in \mathbb{N}$, $W_N^{(2)}$ a complex Wigner matrix $N \times N$ by letting :

$$W_N^{(2)}(i,j) = \begin{cases} \frac{a_{i,j}^{(2)}}{\sqrt{N}} \text{ when } i \leq j \\ \frac{\overline{a_{j,i}^{(2)}}}{\sqrt{N}} \text{ when } i > j \end{cases}$$

.

As before, we define the complex matrix with variance profile V_N by :

$$X_N^{(2)} := W_N^{(2)} \odot \Sigma_N$$

.

1.4.1.3 Statement of the results

We denote r_σ the rightmost point of the support of the limit of the empirical measure of $X_N^{(\beta)}$ (the existence of this measure is discussed in section 1.4.2). First of all, we have the following result for the convergence of the largest eigenvalue of $X_N^{(\beta)}$.

Théorème 1.4.4. *Suppose that assumption 1.4.1 holds. Both in the piecewise constant case and the continuous case, we have that $\lambda_{\max}(X_N^{(\beta)})$ converges almost surely toward r_σ.*

The proof of this theorem is in fact contained in corollary 2.10 of [2] for a positive piecewise constant profile. We will remind this result in Theorem 1.4.15. The general result will be proved as a consequence of Lemma 1.4.53.

For the following theorem which states a large deviation principle for $\lambda_{\max}(X_N^{(\beta)})$, we will need Assumptions 1.4.21 and 1.4.25 which are more thoroughly discussed in section 1.4.3. Assumption 1.4.25 states that the following optimization problem for $\psi \in \mathcal{P}([0,1])$:

$$\sup_{\psi \in \mathcal{P}([0,1])} \left\{ \frac{\theta^2}{\beta} \int_{[0,1]^2} \sigma^2(x,y) d\psi(x) d\psi(y) + \frac{\beta}{2} D(Leb||\psi) \right\}$$

has a determination of its maximum argument that is continuous in θ.

Similarly, Assumption 1.4.21 states that the following optimization problem for $\psi \in (\mathbb{R}^+)^n$ such that $\sum \psi_i = 1$:

$$\sup_{\psi \in (\mathbb{R}^+)^n, \sum \psi_i = 1} \left\{ \frac{\theta^2}{\beta} \sum_{i,j=1}^n \sigma_{i,j}^2 \psi_i \psi_j + \frac{\beta}{2} \sum_i^n \alpha_i \left(\log \psi_i - \log \alpha_i \right) \right\}$$

has a determination of its maximum argument that is continuous in θ. Both these assumptions are necessary to obtain the lower large deviation bound.

Théorème 1.4.5. *Suppose Assumptions 1.4.1, 1.4.2 hold. Furthermore suppose that Assumption 1.4.21 holds in the piecewise constant case or that Assumption 1.4.25 holds in the continuous case. Then, the law of the largest eigenvalue $\lambda_{\max}(X_N^{(1)})$ of $X_N^{(1)}$ satisfies a large deviation principle with speed N and good rate function $I^{(1)}$ which is infinite on $(-\infty, r_\sigma)$.*

In other words, for any closed subset F of $\mathbb{R}$,

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(\lambda_{\max}(X_N^{(1)}) \in F\right) \leq -\inf_F I^{(1)} ,$$

whereas for any open subset O of $\mathbb{R}$,

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(\lambda_{\max}(X_N^{(1)}) \in O\right) \geq -\inf_O I^{(1)} .$$

The same result holds for the opposite of the smallest eigenvalue $-\lambda_{min}(X_N^{(1)})$.

The following theorem is the analog for the complex case :

Théorème 1.4.6. *Assume that Assumptions 1.4.3 and 1.4.2 hold. Furthermore suppose that Assumption 1.4.21 holds in the piecewise constant case or that Assumption 1.4.25 holds in the continuous case. Then, the law of the largest eigenvalue $\lambda_{\max}(X_N^{(2)})$ of $X_N^{(2)}$ satisfies a large deviation principle with speed N and good rate function $I^{(2)}$ which is infinite on $(-\infty, r_\sigma)$. Furthermore $I^2 = 2I^1$.*

Both these rate functions are defined in section 1.4.3. Examples of variance profiles that verifies our Assumptions 1.4.21 and 1.4.25 are given in section 1.4.3.

1.4.2 The limit of the empirical measure

In this section, we describe the limit of the empirical measure μ_σ of the matrices X_N. We will also discuss the stability of this measure in function of σ. Under assumptions of positivity for the variance profile, we will prove that the largest eigenvalue converges toward the rightmost point of the support of μ_σ. To describe the limit of the empirical measure we need the following definition for the so-called canonical equation (also called quadratic vector equation). The following definition takes into account both the piecewise constant and the continuous case :

Définition 1.4.7. *Let $\sigma : [0,1]^2 \to \mathbb{R}^+$ be a bounded symmetric measurable function. We call canonical equation K_σ the following equation where m is a function from $\mathbb{H}$ into $\mathcal{H}$, where $\mathbb{H}$ is the complex upper plane $\{z \in \mathbb{C} | \Im z > 0\}$, $\mathcal{H}$ is the set of measurable m functions from $[0,1]$ to $\mathbb{H}$ and where we suppose for every $z \in \mathbb{H}$ that $\sup_{x\in[0,1]} |m(z)(x)| < \infty$,*

$$-\frac{1}{m(z)} = z + Sm(z) \tag{K_σ}$$

.

Where S is the following kernel operator on $\mathcal{H}$:

$$Sf(z)(x) := \int_0^1 \sigma^2(x,y) f(z)(y) dy$$

.

If w is a function from $[0,1]$ to itself, we denote $||w|| = \sup_{x\in[0,1]} |w_x|$. If S is an operator on the space of functions from $[0,1]$ to itself, we denote $||S||$ the corresponding operator norm. If we denote m_x the function $z \mapsto m(z)(x)$, we have the following result concerning the solution of this equation :

Théorème 1.4.8. *The equation K_σ has a unique solution m which is analytic in z. Moreover for every $x \in [0,1]$, the function $m_x = m(.)(x)$ is the Stieljes transform of a probability measure v_x on $\mathbb{R}$.*

This theorem is in fact a direct application of Theorem 2.1 from [3] with $\mathfrak{X} = [0,1]$ and S the kernel operator (which is trivially non-negative).

Remarque 1.4.9. *If σ is a piecewise constant variance profile with parameters $\alpha_1, ..., \alpha_n$ and $(\sigma_{i,j})_{i,j\leq n}$, then the solution of (K_σ) is piecewise constant on the intervals I_j. This can be viewed directly from K_σ by noticing that Sf is always piecewise constant on the intervals I_j.*

In the rest of the paper, we will denote $\mu_\sigma := \int_0^1 v_x dx$ where v_x is given by the preceding theorem. And so we have that the Stieljes transform of μ_σ is $\int_0^1 m_x dx$. Let us denote $\hat{\mu}_N$ the empirical measure of $X_N^{(\beta)}$. The following theorem links the behaviour of the spectrum of $X_N^{(\beta)}$ and K_σ :

Théorème 1.4.10. *Let us denote $\sigma_N : [0,1]^2 \to \mathbb{R}^+$ the function $(x,y) \mapsto \Sigma_N(\lceil Nx \rceil, \lceil Nx \rceil)$. When N tends to infinity, for almost every x we have that with probability 1 :*

$$\lim_{N\to\infty} \left| \hat{\mu}_N(]-\infty, x]) - \int_{-\infty}^{x} d\mu_{\sigma_N}(x) \right| = 0$$

.

Démonstration. This is in fact a reformulation of [43, Theorem 1.1] (with $a_{i,j}$ =0). It is easy to check that the variance profile we consider do satisfy the hypothesis of the theorem. Furthermore, since σ_N is piecewise constant, the solution m of K_{σ_N} is piecewise constant on the intervals $[i/N, (i+1)/N]$. Making the change of variables $c_i(z) = m_{(2i-1)/2}(z)$ we have that the equation K_{σ_N} is equivalent to the matricial equation given in [43, Theorem 1.1]. □

And so we are left with determining the convergence of the measure μ_{σ_N} when N tends to $+\infty$. To that end, we will need the following rough stability results.

Théorème 1.4.11. *Let $\sigma : [0,1]^2 \to \mathbb{R}^+$ be a bounded measurable function. For every open neighbourhood $\mathcal{V}$ of μ_σ there is $\eta > 0$ such that for every $\tilde{\sigma}$ bounded measurable function such that $\sup_{x,y\in[0,1]^2} |\sigma^2(x,y) - \tilde{\sigma}^2(x,y)| \leq \eta$, $\mu_{\tilde{\sigma}} \in \mathcal{V}$.*

Démonstration. This proof is inspired from the proof of [1, Proposition 2.1]. We let $\tilde{S}$ be the kernel operator with kernel $\tilde{\sigma}^2$. Let $\mathbb{H}_\eta = \{z \in \mathbb{C}, \Im z \geq \eta, |z| \leq \eta\}$ and D the function defined on $\mathbb{H}_\eta^2$ by

$$D(\zeta, \omega) = \frac{|\zeta - \omega|^2}{\Im\zeta \Im\omega}$$

then $d := \operatorname{arcosh}(1 + D)$ is the hyperbolic distance on $\mathbb{H}$. For u a function from $\mathbb{H}_\eta \times [0,1] \to \mathbb{C}$ we define Φ and $\tilde{\Phi}$ as follows :

$$\Phi(u)(z) := -\frac{1}{z + Su(z)}$$
$$\tilde{\Phi}(u)(z) := -\frac{1}{z + \tilde{S}u(z)}$$

.

If $\mathcal{B}_\eta := \{u : \mathbb{H}_\eta \to \mathcal{B}([0,1], \mathbb{H}) | \inf_{z\in\mathbb{H}_\eta} \Im u(z) \geq \frac{\eta^3}{(2+\max\{||S||,||\tilde{S}||\})^2}, \sup_{z\in\mathbb{H}_\eta} ||u(z)|| \leq \eta^{-1}\}$, then following the proof of [1, Proposition 2.1], if u maps $\mathbb{H}_\eta \times [0,1]$ into $\mathbb{H}_\eta$, then $\Phi(u) \in \mathcal{B}_\eta$. Then, if $\eta < 2 + \max\{||S||, ||\tilde{S}||\}$, $\frac{\eta^3}{(2+\max\{||S||,||\tilde{S}||\})^2} \leq \eta$, Φ maps $\mathcal{B}_\eta$ onto itself and so on for $\tilde{\Phi}$.

For $x \in [0,1]$, $z \in \mathbb{H}_\eta$ and $\delta \geq \sup_{x,y} |\sigma^2(x,y) - \tilde{\sigma}^2(x,y)|$

$$\begin{aligned}
D(\Phi(u)_x(z), \tilde{\Phi}(u)_y(z)) &= D(z + (Su)_x(z), z + (\tilde{S}u)_x(z)) \\
&\leq \frac{|(S - \tilde{S})u_x(z)|^2}{\Im z^2} \\
&\leq \frac{\delta^2 \sup_{x\in[0,1]} |u_x(z)|^2}{\eta^2} \\
&\leq \frac{\delta^2}{\eta^4}
\end{aligned}$$

.

Let m be the solution of K_σ, that is the fixed point of Φ. For every $n \in \mathbb{N}$ let $v^{(n)} = \tilde{\Phi}^{(n)}(m)$. We have for $z \in \mathbb{H}_\eta$:

$$\sup_{x\in[0,1]} D(m_x(z), v_x^{(1)}(z)) \leq \frac{\delta^2}{\eta^4}$$

and following again [1],

$$\sup_{x\in[0,1]} D(v_x^{(n+1)}(z), v_x^{(n)}(z)) = \sup_{x\in[0,1]} D(\tilde{\Phi}(v^{(n)})_x(z), v_x^{(n)}(z))$$
$$\leq \left(1+\frac{\eta^2}{||S||}\right)^{-2} \sup_{x\in[0,1]} D(v_x^{(n)}(z), v_x^{(n-1)}(z))$$

and so we have :

$$\sup_{x\in[0,1]} D(v_x^{(n+1)}(z), v_x^{(n)}(z)) \leq \frac{\delta^2}{\eta^4}\left(1+\frac{\eta^2}{||\tilde{S}||}\right)^{-2n}$$

.

And so for δ small enough, $(v^{(n)})_{n\in\mathbb{N}}$ is a Cauchy sequence for the distance $d_{\mathbb{H}_\eta}(u,v) = \sup_{x\in[0,1],z\in\mathbb{H}_\eta} \operatorname{arcosh}(1+D(u_x(z), v_x(z)))$ which is converging toward $\tilde{m}$ the fixed point of $\tilde{\Phi}$ and

$$d_{\mathbb{H}_\eta}(m,\tilde{m}) \leq \sum_{i=0}^{+\infty} \operatorname{arcosh}\left(1+\frac{\delta^2}{\eta^4}\left(1+\frac{\eta^2}{||\tilde{S}||}\right)^{-2n}\right)$$

.

Therefore for every $\epsilon > 0$ and $\eta > 0$, there is $\delta > 0$ small enough such thatsup$_{x,y}$ $|\sigma^2(x,y) - \tilde{\sigma}^2(x,y)| \leq \delta$ implies $d_{\mathbb{H}_\eta}(m,\tilde{m}) \leq \epsilon$.

Since a base of neighbourhood of μ_σ for the vague topology is given by :

$$\mathcal{V}_\eta := \{\lambda \in \mathcal{P}(\mathbb{R}) | \forall z \in \mathbb{H}_\eta, |G_{\mu_\sigma}(z) - G_\lambda(z)| \leq \eta\}$$

and since the vague topology and the weak topology are equal on $\mathcal{P}(\mathbb{R})$ we have our result since $G_{\mu_\sigma} = \int_0^1 m_x dx$ and $G_{\mu_{\tilde{\sigma}}} = \int_0^1 \tilde{m}_x dx$

□

Corollaire 1.4.12. *In the case of a continuous variance profile σ, μ_N converges in probability towards μ_σ.*

Démonstration. This is a consequence from Theorem 1.4.10 and the preceding proposition by noticing that $\lim_{N\to\infty} \sup_{x,y\in[0,1]} |(\sigma_N)^2(x,y) - \sigma^2(x,y)| = 0$. □

We will also need a similar result for the piecewise constant case :

Théorème 1.4.13. *Let $s = (s_{i,j})_{i,j=1}^n \in \mathcal{S}_n(\mathbb{R}^+)$ and $\vec{\alpha}, \vec{\beta} \in \mathbb{R}^n$ two vectors of positive coordinates summing to one and let $\gamma_i = \sum_{j=0}^i \alpha_j$ and $\tilde{\gamma}_i = \sum_{j=0}^i \beta_j$. Let σ and $\tilde{\sigma}$ and the two piecewise constant variance profiles associated respectively with the couple $(s,\vec{\alpha})$ and $(s,\vec{\beta})$ and m and $\tilde{m}$ the solutions respectively of K_σ and $K_{\tilde{\sigma}}$. for $i \in \{1,...,n\}$, let also m_i and $\tilde{m}_i$ be the holomorphic functions such that $m_x = \sum_{i=1}^n \mathbb{1}_{\gamma_{i-1}\leq x<\gamma_i} m_i$ and $\tilde{m}_x = \sum_{i=1}^n \mathbb{1}_{\tilde{\gamma}_{i-1}\leq x<\tilde{\gamma}_i} \tilde{m}_i$. Then for every $\eta > 0$ there is $\epsilon > 0$ such that if $\sup_i |\alpha_i - \beta_i| \leq \epsilon$, then for all $z \in \mathbb{H}_\eta$, we have $\sup_i |\tilde{m}_i(z) - \tilde{m}_i(z)| \leq \eta$.*

Démonstration. We use the same notations as in the previous proof. The m_i verify the following system :

$$m_i = \frac{-1}{\sum_{j=1}^n \alpha_j s_{i,j}^2 m_j + z}$$

and

$$\tilde{m}_i = \frac{-1}{\sum_{j=1}^n \beta_j s_{i,j}^2 \tilde{m}_j + z}$$

. We let Φ and $\tilde{\Phi}$ the following operators on the set of holomorphic functions from $\mathbb{H}_\eta$ to $\mathbb{H}_\eta^n$ defined by :

$$\Phi(u)(z) = \frac{-1}{Su + z}$$

and

$$\tilde{\Phi}(u)(z) = \frac{-1}{\tilde{S}u + z}$$

where S and $\tilde{S}$ are the linear applications defined by

$$\forall i = 1, ..., n, (Su)_i = \sum_{i=1}^{n} \alpha_i s_{i,j}^2 u_j \text{ and } (\tilde{S}u)_i = \sum_{i=1}^{n} \beta_i s_{i,j}^2 u_j$$

.

As in the preceding proof, if $\mathcal{B}_\eta := \{u : \mathbb{H}_\eta \to \mathbb{H}^n | \inf_{z\in\mathbb{H}_\eta} \Im u(z) \geq \frac{\eta^3}{(2+\max\{||S||,||\tilde{S}||\})^2}, \sup_{z\in\mathbb{H}_\eta} ||u(z)|| \leq \eta^{-1}\}$, Φ and $\tilde{\Phi}$ maps $\mathcal{B}_\eta$ onto itself for η small enough. For $u \in \mathcal{B}_\eta$, we have as before if $\delta \geq (\sup_{i,j} s_{i,j}^2)(\sup_k |\alpha_k - \beta_k|)$, for all i :

$$\begin{aligned} D(\Phi(u)_i(z), \tilde{\Phi}(u)_i(z)) &\leq \frac{|(S - \tilde{S})(u)_i(z)|}{\Im z^2} \\ &\leq \frac{\delta^2}{\eta^4} \end{aligned}$$

Then, using the same reasoning as in the previous case, we have that for every $\eta > 0$, there is $\delta' > 0$ such that if $\sup_i |\alpha_i - \beta_i| \leq \delta'$ then $\sup_{z\in\mathbb{H}_\eta} \sup_i |\tilde{m}_i(z) - \tilde{m}_i(z)| \leq \eta$.

□

In order to apply the full results of [39, 2, 3], we will need the following assumption for the piecewise constant variance profile :

Hypothèse 1.4.14. *In the piecewise constant case,* $\forall i, j \in [1, n], \sigma_{i,j} > 0$.

Under assumption 1.4.14 we have the following convergence result :

Théorème 1.4.15. *If Assumption 1.4.14 is verified, we have that for any* $D > 0$ *and* $\tau \in \mathbb{R}$

$$\mathbb{P}[|\hat{\mu}_N([-\infty; \tau]) - \mu_\sigma([-\infty; \tau])| \geq N^{-3/4}] \leq N^{-D}$$

for $N \geq N(D)$ *(where* $N(D)$ *does not depend on* τ*). Furthermore, if we* l_N *and* r_N *be respectively the left and right edge of the support of* $\hat{\mu}_N$ *and* l_σ *and* r_σ *the left and right edges of the support of* μ_σ*, we have for every* $\delta > 0$*,* $D > 0$*,*

$$\mathbb{P}[r_N \geq r_\sigma + \delta \text{ or } l_N \leq l_\sigma - \delta] \leq N^{-D}$$

for N *large enough.*

Démonstration. This is in fact an application of corollary 2.10 of [2]. Indeed, up to multiplication by a scalar, our matrix model verifies the condition (A) and 1.4.14 gives us (B) and the sharp-sub Gaussian hypothesis gives (D). For (C) we look to theorem 6.1 of [3], particularly remark 6.2. □

1.4.3 The rate function

In this section we will explicit the rate functions I in 1.4.5 and 1.4.6. This function is in fact defined the same way as in [45] as the supremum $\sup_\theta(J(\mu_\sigma, \theta, x) - F(\theta))$. In this expression, $J(\mu_\sigma, \theta, x)$ is the limit of $N^{-1} \log \mathbb{E}[\exp(N\langle e, A_N e\rangle)]$ where e is a unitary vector taken uniformly on the sphere and A_N is a sequence of matrices such that the empirical measures converge weakly to μ_σ and such that the sequence of the largest eigenvalues of A_N converges to x. $F(\theta)$ is the limit of $N^{-1} \log \mathbb{E}[\exp(N\langle e, X_N e\rangle)]$ where the expectation is taken both in X_N and e. We will first describe the quantity $F(\theta)$.

1.4.3.1 The asymptotics of the annealed spherical integral

For $\sigma : [0,1]^2 \to \mathbb{R}^+$ a bounded measurable function and ψ a probability measure on $[0,1]$, let us denote :

$$P(\sigma, \psi) := \int_0^1 \int_0^1 \sigma^2(x,y) d\psi(x) d\psi(y)$$

and for $\theta > 0$:

$$\Psi(\theta, \sigma, \psi) := \frac{\theta^2}{\beta} P(\sigma, \psi) + \frac{\beta}{2} D(Leb||\psi)$$

where $D(.||.)$ is the Kullback-Leibler deviation, that is :

$$D(\lambda||\mu) = \begin{cases} \int_0^1 \log\left(\frac{d\lambda}{d\mu}(x)\right) d\lambda(x) \text{ if } \lambda \text{ is continuous with respect to } \mu \\ -\infty \text{ if this is not the case} \end{cases}$$

and Leb is the Lebesgue measure on $[0,1]$.

Lemme 1.4.16. *The function :*

$$\mu \mapsto D(Leb||\mu)$$

is upper semi continuous.

Démonstration. This is due to the fact that the function $(\lambda, \mu) \mapsto D(\lambda||\mu)$ is convex in (λ, μ). □

We consider here the following optimization problem on the set $\mathcal{M}_1$ of probability measures on $[0,1]$ with parameter $\theta > 0$:

$$F(\sigma, \theta) := \sup_{\mu \in \mathcal{M}_1} \left\{ \frac{\theta^2}{\beta} P(\sigma, \mu) + \frac{\beta}{2} D(\text{Leb}||\mu) \right\}. \tag{1.28}$$

First, let us study this problem with the following lemma :

Lemme 1.4.17. *In the problem* (1.28)*, if σ is continuous, the* sup *is a* max*. Furthermore, in both the continuous and the piecewise cases, the function F is continuous in θ.*

Démonstration. Let us take μ_n a sequence of measures such that $\frac{\theta^2}{\beta} P(\sigma, \mu_n) + \frac{\beta}{2} D(\text{Leb}||\mu_n)$ converges toward $F(\sigma, \theta)$. By compacity we can assume that this sequence converges weakly to some μ. By weak continuity of P, $\lim_n P(\sigma, \mu_n) = P(\sigma, \mu)$ and by weak upper semi-continuity, $\liminf_n D(\text{Leb}||\mu_n) \geq D(\text{Leb}||\mu)$ so that

$$\liminf_n \left\{ \frac{\theta^2}{\beta} P(\sigma, \mu_n) + \frac{\beta}{2} D(\text{Leb}||\mu_n) \right\} \geq F(\sigma, \theta)$$

. Furthermore, we have for every $\mu \in \mathcal{M}_1$, $|\Psi(\theta, \sigma, \mu) - \Psi(\theta', \sigma, \mu)| \leq ||\sigma^2||_\infty |\theta^2 - \theta'^2|/\beta$ and so $|F(\theta) - F(\theta')| \leq ||\sigma^2||_\infty |\theta^2 - \theta'^2|/\beta$. □

In section 1.4.6 we will prove that the following limit :

$$\lim_{N \to \infty} N^{-1} \log \mathbb{E}_{e, X_N}[\exp(N\theta\langle e, X_N e\rangle)] = F(\sigma, \theta)$$

holds in the piecewise constant case. In the following subsection, we will discuss the simplifications that occurs in the expression of $F(\sigma, \theta)$ in this case.

1.4.3.2 Simplifications for the piecewise constant case

We consider the piecewise constant case, that is when σ is defined with a matrix $(\sigma_{i,j})_{i,j}$ and parameters $\vec{\alpha}$. In this case, the optimization problem that defines F is in fact simpler.

Proposition 1.4.18. *We denote by $\vec{P}(\sigma,.)$ the following quadratic function on $\mathbb{R}^n$:*

$$\vec{P}(\sigma,\psi_1,...,\psi_n) = \sum_{i,j=1}^{n} \sigma_{i,j}^2 \psi_i \psi_j$$

and

$$\vec{\Psi}(\theta,\sigma,\vec{\psi}) := \frac{\theta^2}{\beta}\vec{P}(\sigma,\psi_1,...,\psi_n) + \frac{\beta}{2}\left(\sum_{i=1}^{n} \alpha_i \log\psi_i - \sum_{i=1}^{n}\alpha_i\log\alpha_i\right).$$

We have that

$$F(\sigma,\theta) = \max_{\psi_i\geq 0, \sum_1^n \psi_i = 1} \vec{\Psi}(\theta,\sigma,\vec{\psi}). \tag{1.29}$$

Démonstration. Let ψ be a probability measure on $[0,1]$ and let's define for every $i \in [1,n]$:

$$\psi_i := \psi(I_i)$$

then we have :

$$P(\sigma,\psi) = \vec{P}(\sigma,\psi_1,...,\psi_n).$$

For every $i \in [1,n]$, by concavity of log :

$$\log\left(\alpha_i^{-1}\int_{I_i}\frac{d\psi}{dx}(x)dx\right) \geq \alpha_i^{-1}\int_{I_i}\log\left(\frac{d\psi}{dx}(x)\right)dx.$$

Multiplying by α_i and summing over i, we get :

$$D(Leb||\psi) \leq \sum_{i=1}^{n}\alpha_i\log\psi_i - \sum_{i=1}^{n}\alpha_i\log\alpha_i$$

so we have the inequality $\leq$.

Then if $\psi_1,...\psi_n \in \mathbb{R}^+$ are such that $\psi_1+...+\psi_n = 1$ we define ψ the probability measure defined by its density :

$$\frac{d\psi}{dx}(x) = \alpha_i^{-1}\psi_i \text{ for } x \in I_i.$$

Then we have :

$$P(\sigma,\psi) = \vec{P}(\sigma,\psi_1,...,\psi_n)$$

and

$$D(Leb||\psi) = \sum_{i=1}^{n}\alpha_i\log\psi_i - \sum_{i=1}^{n}\alpha_i\log\alpha_i.$$

and so we have the inequality $\geq$ □

The function $\vec{\Psi}(\sigma,\theta,.)$ we seek to maximize tends to $-\infty$ at the boundary of our domain so this problem has a solution.

1.4.3.3 Definition of the rate functions

Now, in order to introduce our rate functions we need first to introduce the function J. This function is linked to the asymptotics of the following spherical integrals :

$$I_N(X,\theta) = \mathbb{E}_e[e^{\theta N\langle e, Xe\rangle}]$$

where the expectation holds over e which follows the uniform measure on the sphere $\mathbb{S}^{\beta N-1}$ with radius one. Denoting J_N the following quantities :

$$J_N(X,\theta) = \frac{1}{N}\log I_N(X,\theta)$$

the following theorem was proved in [47] :

Théorème 1.4.19. *[47, Theorem 6]*
If $(E_N)_{N\in\mathbb{N}}$ is a sequence of $N\times N$ real symmetric matrices when $\beta=1$ and complex Gaussian matrices when $\beta=2$ such that :
- *The sequence of empirical measures $\hat{\mu}^N_{E_N}$ weakly converges to a compactly supported measure μ,*
- *There are two reals $\lambda_{\min}(E), \lambda_{\max}(E)$ such that $\lim_{N\to\infty}\lambda_{\min}(E_N) = \lambda_{\min}(E)$ and $\lim_{N\to\infty}\lambda_{\max}(E_N) = \lambda_{\max}(E)$,*

and $\theta \geq 0$, then :

$$\lim_{N\to\infty} J_N(E_N,\theta) = J(\mu,\theta,\lambda_{\max}(E))$$

The limit J is defined as follows. For a compactly supported probability measure we define its Stieltjes transform G_μ by

$$G_\mu(z) := \int_{\mathbb{R}} \frac{1}{z-t}d\mu(t)$$

We refer to the Theorem 1.3.18 of the preceding section for the definition of J. In both cases, we introduce our rate function $I^{(\beta)}$ as

$$I^{(\beta)}(\sigma,x) = -\infty \text{ for } x \in]-\infty, r_\sigma[$$

and

$$I^{(\beta)}(\sigma,x) = \max_{\theta\geq 0}\left(J(\mu_\sigma,\theta,x) - F(\sigma,\theta)\right)$$

Lemme 1.4.20. *For $\beta = 1,2$, $I^{(\beta)}(\sigma,.)$ is a good rate function. Furthermore $I^{(2)}(\sigma,.) = 2I^{(1)}(\sigma,.)$*

Démonstration. As a supremum of continuous functions, $I^{(\beta)}(\sigma,.)$ is lower semi-continuous. We want to prove that the level sets of $I^{(\beta)}(\sigma,.)$, that is the $\{x|I^{(\beta)}(\sigma,x)\leq M\}$ are compacts. It is sufficient to show that $\lim_{x\to+\infty} I^{(\beta)}(\sigma,x) = +\infty$. For any fixed $\theta > 0$, we have $\lim_{x\to\infty} J(\mu_\sigma,\theta,x) = +\infty$. And so since we have $I^{(\beta)}(\sigma,x) \geq J(\mu_\sigma,\theta,x) - F(\sigma,\theta)$, $I^{(\beta)}$ is a good rate function. With the change of variables $\theta' = \theta/2$ in the case $\beta = 2$, we have that that $I^{(2)}(\sigma,.) = 2I^{(1)}(\sigma,.)$. □

1.4.3.4 Assumptions on the variance profile σ

In order to prove the lower large deviation bound in the piecewise constant case, we will need the following assumptions on σ :

Hypothèse 1.4.21. *There exists some continuous $\theta \mapsto (\psi_i^\theta)_{i\in[1,n]}$ such that ψ^θ is a maximal argument of the equation 1.29, that is :*

$$\frac{\theta^2}{\beta}\vec{P}(\sigma,\psi_1^\theta,...,\psi_n^\theta) + \frac{\beta}{2}\left(\sum_{i=1}^n \alpha_i \log\psi_i^\theta - \sum_{i=1}^n \alpha_i\log\alpha_i\right) = F(\sigma,\theta)$$

.

As a more practical example, the following assumption implies 1.4.21

Hypothèse 1.4.22. *The function $\vec{P}(\sigma,.)$ restricted to the hyperplane $\{\psi_1+\psi_2+...+\psi_n=0\}$ is concave.*

Remarque 1.4.23. *Some variance profiles that satisfy this assumption are those associated to the parameters $(\alpha_1,...,\alpha_n)$ and $(\delta_{i\neq j})_{i,j\in[1,n]}$. In the case $n=2$ this a linearisation of a Wishart matrix as in [45].*

Indeed, if such is the case then Assumption 1.4.21 holds :

Lemme 1.4.24. *Assumption 1.4.22 implies Assumption 1.4.21.*

Démonstration. The function $\vec{\psi}\mapsto\frac{\theta^2}{\beta}\vec{P}(\sigma,\vec{\psi})+\frac{\beta}{2}\sum_{i=1}^n\alpha_i\log\psi_i$ is strictly concave and since it tends to $-\infty$ on the boundary of the domain, it admits a unique maximal argument ψ^θ which is also the unique solution to the following critical point equation :

$$\frac{2\theta^2}{\beta}\left(\sum_{j=1}^n\sigma_{i,j}^2\psi_j\right)_{i=1,...,n}+\frac{\beta}{2}\left(\frac{\alpha_i}{\psi_i}\right)_{i=1,...,n}\in\mathrm{Vect}(1,...,1).$$

We now want to apply the implicit function theorem to prove that $\theta\mapsto\psi^\theta$ is analytic. Letting $f(\psi)$ be the left hand side of the above equation. We have that

$$\partial f_\psi(u)=\frac{2\theta^2}{\beta}Su-\frac{\beta}{2}\sum_{j=1}^n\frac{u_j}{\psi_j^2}.$$

It suffices to show that $\partial f_\psi(u)\notin\mathrm{Vect}(1,...,1)$ for ψ in the domain and $u\in\{u\in\mathbb{R}^n|u_1+...+u_n=0\}$. For such a u, we have

$$\langle u,\partial f_\psi(u)\rangle=\frac{2\theta^2}{\beta}\langle u,Su\rangle-\frac{\beta}{2}\sum_{j=1}^n\frac{u_j^2}{\psi_j^2}$$

Since $u\in\{u\in\mathbb{R}^n|u_1+...+u_n=0\}$ we have by Assumption 1.4.22 $\langle u,Su\rangle\le 0$ and therefore $\langle u,\partial f_\psi(u)\rangle<0$. So $\partial f_\psi(u)\notin\mathrm{Vect}(1,...,1)$. So we can apply the implicit function theorem. □

Example of variance profiles that satisfies Assumption 1.4.21 but not Assumption 1.4.22 are provided in section 1.4.9. In the same section, we will also show that without any assumptions on σ, the method employed may fail as we can have a large deviation principle but with a rate function different from I.

In the continuous case, we will need the following assumptions :

Hypothèse 1.4.25. *There exists some continuous $\theta\mapsto\psi^\theta$ (for the weak topology) such that ψ^θ is a maximal argument of 1.28 that is :*

$$F(\sigma,\theta)=\Psi(\theta,\sigma,\psi^\theta)$$

As for the piecewise constant case, the following assumption implies 1.4.25

Hypothèse 1.4.26. *The function $P(\sigma,.)$ is concave on the set of probability measure on $[0,1]$.*

Lemme 1.4.27. *Assumption 1.4.26 implies 1.4.25.*

Démonstration. First, since $\mu\mapsto D(Leb||\mu)$ is strictly concave so is $\Psi(\theta,\sigma,.)$ and so the maximum argument ψ_θ is unique. Let us prove now that it is continuous for the weak topology. Let $\theta>0$, $\mathcal{N}_\epsilon:=\{\psi\in\mathcal{P}([0,1])|\Psi(\theta,\sigma,\mu)\ge F(\sigma,\theta)-\epsilon\}$. Since $\mu\to D(Leb||\mu)$ is lower semi continuous, $\Psi(\theta,\sigma,.)$ is upper semi-continuous and so $\mathcal{N}_\epsilon$ is a compact for the weak topology. Let $\mathcal{V}$ a neighborhood of ψ^θ. Then since $\bigcap_{\epsilon>0}\mathcal{N}_\epsilon=\{\psi^\theta\}$ there is an $\epsilon'>0$ such that $\mathcal{N}_{\epsilon'}\subset\mathcal{V}$. We have $|\Psi(\theta',\sigma,\mu)-\Psi(\theta,\sigma,\mu)|\le\frac{|\theta'^2-\theta^2|}{\beta}\max_{x,y}\sigma^2(x,y)$. So for $|\theta'^2-\theta^2|\le\beta\epsilon'/3(\max\sigma^2(x,y))$, $\Psi(\theta',\sigma,\psi^\theta)\ge F(\theta)-\epsilon'/3$ and on $\mathcal{N}_{\epsilon'}^c$, $\Psi(\theta',\sigma,\mu)\le F(\theta)-2\epsilon'/3$. Necessarily, $\psi^{\theta'}\in\mathcal{N}_{\epsilon'}\subset\mathcal{V}$ and so we have the continuity. □

Remarque 1.4.28. *A family of such σ is given by $\sigma^2(x,y) = |f(x) - f(y)| + C$ where f is an increasing continuous function and $C \in \mathbb{R}^+$. Indeed, if f is an increasing and continuous function on $[0,1]$, there is a positive measure ν on $[0,1]$ such that $f(x) - f(0) = \int_0^x d\nu(t)$ and we have $\sigma^2(x,y) = C + \int_x^y d\nu(t) = \int_0^1 \tau_t(x,y)d\nu(t) + C$ where $\tau_t(x,y) = \mathbb{1}_{x \le t < y} + \mathbb{1}_{y \le t < x}$ and so*

$$P(\sigma, \psi) = \int_0^1 P(\tau_t, \psi) d\nu(t) + C$$

. Since $P(\tau_t, \psi) = 2\psi([0,t[)(1 - \psi([0,t[)$, $P(\tau_t, .)$ is concave and so is $P(\sigma, .)$.

1.4.4 Scheme of the proof

The proof of theorems 1.4.5 and 1.4.6 will follow a path similar to [45] for the piecewise constant case and then for σ continuous, we will approximate it by a sequence of piecewise constant profiles. First of all, the following result of exponential tightness will be proved in Section 1.4.5 :

Lemme 1.4.29. *For $\beta = 1,2$, assume that the distribution of the entries $a_{i,j}^{(\beta)}$ satisfy Assumption 1.4.2 for $\beta = 1,2$ and Assumption 1.4.3 for $\beta = 2$. Then :*

$$\lim_{K \to +\infty} \limsup_{N \to \infty} \frac{1}{N} \log \mathbb{P}[\lambda_{\max}(X_N^{(\beta)}) > K] = -\infty.$$

Similar results hold for $\lambda_{\min}(X_N^{(\beta)})$.

Therefore it is enough to prove a weak large deviation principle. In the following we summarize the assumptions on the distribution of the entries as follows :

Hypothèse 1.4.30. *Either the $\mu_{i,j}^N$ are uniformly compactly supported in the sense that there exists a compact set K such that the support of all $\mu_{i,j}^N$ is included in K, or the $\mu_{i,j}^N$ satisfy a uniform log-Sobolev inequality in the sense that there exists a constant c independent of N such that for all smooth function f :*

$$\int f^2 \log \frac{f^2}{\mu_{i,j}^N(f^2)} d\mu_{i,j}^N \le c\mu_{i,j}^N(\|\nabla f\|_2^2)\,.$$

When $\beta = 1$ $\mu_{i,j}^N$ satisfy Assumption 1.4.1, when $\beta = 2$, they satisfy Assumption 1.4.3.

We shall first prove that we have a weak large deviation upper bound :

Théorème 1.4.31. *Assume that we have a piecewise constant variance profile σ and that Assumption 1.4.30 holds. Let $\beta = 1,2$. Then, for any real number x,*

$$\limsup_{\delta \to 0} \limsup_{N \to \infty} \frac{1}{N} \log \mathbb{P}\left(\left|\lambda_{\max}(X_N^{(\beta)}) - x\right| \le \delta\right) \le -I^{(\beta)}(x).$$

We then prove the following large deviation lower bound :

Théorème 1.4.32. *Assume that we are in the case of a piecewise constant variance profile σ and that assumptions 1.4.30 and 1.4.14 holds. Let $E : \mathbb{R}^+ \to \mathbb{R}^+$ be a positive function. Suppose that we have a continuous function $\theta \mapsto (\psi_i^{E,\theta})_{i \in [1,n]}$ such that :*

$$\vec{\Psi}(\theta, \sigma, \psi^{E,\theta}) \ge F(\sigma, \theta) - E(\theta)$$

then, if we let $\tilde{I}(\sigma, x) := \sup_{\theta \ge 0} J(\mu_\sigma, \theta, x) - F(\sigma, \theta) + E(\theta)$, we have for every $x \ge r_\sigma$:

$$\liminf_{N \to \infty} \frac{1}{N} \log \mathbb{P}[|\lambda_{\max}(X) - x| \le \delta] \ge -\tilde{I}(\sigma, x).$$

Then, if Assumption 1.4.21 is verified, we can take $E = 0$ and the result follows. However, when we deal with the continuous case, since Assumption 1.4.25 for σ will not necessarily imply Assumption 1.4.21 for the piecewise constant approximations, the error E will not be zero. However, it will be small enough to be neglected ultimately.

To prove Theorem 1.4.31, we first show that the rate function is infinite below the right edge of the support of the limiting spectral distribution. To this end, we use that the spectral measure $\hat{\mu}_N$ converges towards its limit whith much larger probability. We let d denote the Dudley distance :

$$d(\mu,\nu) = \sup_{\|f\|_L \le 1} \left| \int f(x) d\mu(x) - \int f(x) d\nu(x) \right| ,$$

where $\|f\|_L = \sup_{x\neq y} \left| \frac{f(x)-f(y)}{x-y} \right| + \sup_x |f(x)|$.

Lemme 1.4.33. *Assume that the $\mu_{i,j}^N$ are uniformly compactly supported or satisfy a uniform log-Sobolev inequality. Then, for $\beta = 1,2$, there exists $\kappa' \in (0, \frac{1}{10} \wedge \kappa)$ such that*

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(d(\hat{\mu}^N_{X_N^{(\beta)}}, \mu_\sigma) > N^{-\kappa'} \right) = -\infty .$$

The proof of this lemma is given in the appendix. As a consequence, we deduce that the extreme eigenvalues can not deviate towards a point inside the support of the limiting spectral measure with probability greater than $o(e^{-CN})$ for arbitrarily large C. And therefore :

Corollaire 1.4.34. *Under the assumption of Lemma 1.4.33, For $\beta = 1,2$ let x be a real number in $(-\infty, r_\sigma)$. Then, for $\delta > 0$ small enough,*

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(|\lambda_{\max}(X_N^{(\beta)}) - x| \le \delta \right) = -\infty .$$

Indeed, as soon $\delta > 0$ is small enough so that $x + \delta$ is smaller than $2 - \delta$ for $\beta = 1,2$, $d(\hat{\mu}_N, \mu_\sigma)$ is bounded below by some $\kappa(\delta) > 0$ on $|\lambda_{\max}(X_N^{(\beta)}) - x| \le \delta$. Hence, Lemma 1.4.33 implies the Corollary.

In order to prove the weak large deviation bounds for the remaining x's, we shall tilt the measure by spherical integrals :

$$I_N(X,\theta) = \mathbb{E}_e[e^{\theta N \langle e, Xe\rangle}]$$

where the expectation holds over e which follows the uniform measure on the sphere $\mathbb{S}^{\beta N-1}$ with radius one. The asymptotics of

$$J_N(X,\theta) = \frac{1}{N} \log I_N(X,\theta)$$

were studied in [47] where Theorem 1.4.19 was proved.

We shall later use that spherical integrals are continuous. We recall here Proposition 2.1 from [58] and Theorem 6.1 from [47] where we denote by $\|A\|$ the operator norm of the matrix A given by $\|A\| = \sup_{\|u\|_2=1} \|Au\|_2$ where $\|u\|_2 = \sqrt{\sum |u_i|^2}$.

Proposition 1.4.35. *For every $\theta > 0$, every $\kappa \in]0,1/2[$, every $M > 0$, there exist a function $g_\kappa : \mathbb{R}^+ \to \mathbb{R}^+$ going to 0 at 0 such that for any $\delta > 0$ and N large enough, with B_N and B'_N such that $d(\hat{\mu}^N_{B_N}, \hat{\mu}^N_{B'_N}) < N^{-\kappa}$, $|\lambda_{\max}(B_N) - \lambda_{\max}(B_N)| < \delta$ and $\sup_N ||B_N|| \le M$, $\sup_N ||B'_N|| \le M$:*

$$|J_N(B_N,\theta) - J_N(B'_N,\theta)| < g_\kappa(\delta) .$$

From Theorem 1.4.19 and Proposition 1.4.35, we deduce that :

Corollaire 1.4.36. *For every $\theta > 0$, every $\kappa \in]0,1/2[$, every $M > 0$, for any $\delta > 0$ and μ a probability measure supported in $[-M,M]$, if we denote by $\mathcal{B}_N$ the set of symmetric matrices B_N such that $d(\mu_{B_N}, \mu) < N^{-\kappa}$, $|\lambda_{\max}(B_N) - \rho| < \delta$, and $\sup_N ||B_N|| \le M$, for N large enough, we have :*

$$\limsup_{N\to\infty}\sup_{B_N\in\mathcal{B}_N}|J_N(B_N,\theta)-J(\mu,\theta,\rho)|\le 2g_\kappa(\delta)$$

where g_κ is the function in Proposition 1.4.35.

By Lemma 1.4.29 and Lemma 1.4.33, it is enough to study the probability of deviations on the set where J_N is continuous :

Corollaire 1.4.37. *Suppose Assumption 1.4.2 holds. For $\delta>0$, take a real number x and set for M large (larger than $x+\delta$ in particular), $\mathcal{A}^M_{x,\delta}$ to be the set of $N\times N$ self-adjoint matrices given by*

$$\mathcal{A}^M_{x,\delta}=\{X:|\lambda_{\max}(X)-x|<\delta\}\cap\{X:d(\hat\mu^N_X,\mu_\sigma)<N^{-\kappa'}\}\cap\{X:\|X\|\le M\}\,,$$

where κ' is chosen as in Lemma 1.4.33. Let x be a real number, $\delta>0$ and κ' as in Lemma 1.4.33. Then, for any $L>0$, for M large enough

$$\mathbb{P}\left(\left|\lambda_{\max}(X_N^{(\beta)})-x\right|<\delta\right)=\mathbb{P}\left(X_N^{(\beta)}\in\mathcal{A}^M_{x,\delta}\right)+O(e^{-NL})\,.$$

We are now in position to get an upper bound for $\mathbb{P}\left(X_N^{(\beta)}\in\mathcal{A}^M_{x,\delta}\right)$. In fact, by the continuity of spherical integrals of Corollary 1.4.36, for any $\theta\ge 0$,

$$\begin{aligned}\mathbb{P}\left(X_N^{(\beta)}\in\mathcal{A}^M_{x,\delta}\right) &= \mathbb{E}\left[\frac{I_N(X_N^{(\beta)},\theta)}{I_N(X_N^{(\beta)},\theta)}\mathbb{1}_{\mathcal{A}^M_{x,\delta}}\right]\\ &\le \mathbb{E}[I_N(X_N^{(\beta)},\theta)]\exp\{-N\inf_{X\in\mathcal{A}^M_{x,\delta}}J_N(X,\theta)\}\\ &\le \mathbb{E}[I_N(X_N^{(\beta)},\theta)]\exp\{N(2g_\kappa(\delta)-J(\mu_\sigma,\theta,x))\} \qquad (1.30)\end{aligned}$$

where we used that $x\to J(\mu_\sigma,\theta,x)$ is continuous and took N large enough. It is therefore central to derive the asymptotics of

$$F_N(\theta,\beta)=\frac{1}{N}\log\mathbb{E}[I_N(X_N^{(\beta)},\theta)]$$

and we shall prove in section 1.4.6 that

Théorème 1.4.38. *Suppose Assumption 1.4.30 holds and that σ is a piecewise constant variance profile. For $\beta=1,2$ and $\theta\ge 0$,*

$$\lim_{N\to\infty}F_N(\theta,\beta)=F(\sigma,\theta).$$

We therefore deduce from (1.30), Corollaries 1.4.37 and 1.4.36, and Theorem 1.4.38, by first letting N going to infinity, then δ to zero and finally M to infinity, that

$$\limsup_{\delta\to 0}\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{P}\left(\left|\lambda_{\max}(X_N^{(\beta)})-x\right|<\delta\right)\le F(\sigma,\theta)-J(\mu_\sigma,\theta,x)\,.$$

We next optimize over θ to derive the upper bound :

$$\limsup_{\delta\to 0}\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{P}\left(\left|\lambda_{\max}(X_N^{(\beta)})-x\right|<\delta\right)\le -\sup_{\theta\ge 0}\{J(\mu_\sigma,\theta,x)-F(\sigma,\theta)\}\,. \qquad (1.31)$$

To prove the complementary lower bound, we shall prove that

Lemme 1.4.39. *For $\beta=1,2$, with the assumptions and notations of Theorem 1.4.32, for any $x>r_\sigma$, there exists $\theta=\theta_x\ge 0$ such that for any $\delta>0$ and N large enough,*

$$\liminf_{N\to\infty}\frac{1}{N}\log\frac{\mathbb{E}[\mathbb{1}_{X_N^{(\beta)}\in\mathcal{A}^M_{x,\delta}}I_N(X_N^{(\beta)},\theta)]}{\mathbb{E}[I_N(X_N^{(\beta)},\theta)]}\ge -E(\theta_x)\,.$$

This lemma is proved by showing that the matrix whose law has been tilted by the spherical integral is approximately a finite rank perturbation of a Wigner matrix, from which we can use the techniques developped to study the famous BBP transition [13]. The conclusion follows since then

$$\begin{aligned}
\mathbb{P}\left(X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M\right) &\geq \frac{\mathbb{E}[\mathbb{1}_{X_N^\delta \in \mathcal{A}_{x,\delta}^M} I_N(X_N^{(\beta)},\theta_x)]}{\mathbb{E}[I_N(X_N^{(\beta)},\theta_x)]}\mathbb{E}[I_N(X_N^{(\beta)},\theta_x)]\exp\{-N \sup_{X\in\mathcal{A}_{x,\delta}^M} J_N(X,\theta_x)\} \\
&\geq \exp\{N(g_\kappa(\delta)+F(\theta_x,\beta)-E(\theta_x)-J(\mu_\sigma,\theta_x,x)+o_\delta(1))\} \\
&\geq \exp\{-N\tilde{I}_\beta(x)-No_\delta(1)\}
\end{aligned}$$

where we finally used Theorem 1.4.38 and Lemma 1.4.39.

The theorem follows in the case of piecewise constant variance profile verifying Assumption 1.4.14 by noticing that if Assumption 1.4.21 is verified then we can choose $E=0$. We will relax the Assumption 1.4.14 by approximation as we will treat the continuous case.

1.4.5 Exponential tightness

In this section we prove Lemma 1.4.29. We will in fact prove a stronger and slightly more quantitative result that will also be useful when we will approximate continuous variance profiles using piecewise constant ones (we recall that $\odot$ is the entrywise product of matrices) :

Lemme 1.4.40. *Let $\beta = 1,2$ and $\mathcal{A}_N$ be the following subset of matrices :*

$$\mathcal{A}_N := \{A \in \mathcal{S}_N(\mathbb{R}^+) | \forall i,j, A(i,j) \leq 1\}.$$

For every $M>0$ there exists $B>0$ such that :

$$\limsup_N \frac{1}{N} \sup_{A\in\mathcal{A}_N} \log \mathbb{P}[||A\odot W_N^{(\beta)}|| \geq B] \leq -M.$$

We will use a standard net argument that we recall for the sake of completeness. Let us denote :

$$Y_N^{(\beta)} := A \odot W_N^{(\beta)}.$$

Where $A \in \mathcal{A}_N$.

For $N\in\mathbb{N}$, let R_N be a 1/2-net of the sphere (i.e. a subset of the sphere $\mathbb{S}^{\beta N-1}$ such as for all $u\in\mathbb{S}^{\beta N-1}$ there is $v\in R_N$ such that $||u-v||_2 \leq 1/2$. Here the sphere is inside $\mathbb{R}^N$ for $\beta=1$ and $\mathbb{C}^N$ for $\beta=2$). We know that we can take R_N with cardinality smaller than $3^{\beta N}$. As in the proof of the exponential tightness in [45], we notice that for $M>0$

$$\mathbb{P}[||Y_N^{(\beta)}|| \geq 4K] \leq 9^{\beta N} \sup_{u,v\in R_N} \mathbb{P}[\langle Y_N^{(\beta)}u,v\rangle \geq K]. \tag{1.32}$$

Indeed, if we denote, for $v\in\mathbb{S}^{\beta N-1}$, u_v to be an element of R_N such that $\|u_v - v\|_2 \leq 1/2$,

$$\|Y_N^{(\beta)}\| = \sup_{v\in\mathbb{S}^{\beta N-1}} \|Y_N^{(\beta)}v\|_2 \leq \sup_{v\in\mathbb{S}^{\beta N-1}} (\|Y_N^{(\beta)}u_v\|_2 + \frac{1}{2}\|Y_N^{(\beta)}\|)$$

so that

$$\|Y_N^{(\beta)}\| \leq 2\sup_{u\in R_N} \|Y_N^{(\beta)}u\|_2. \tag{1.33}$$

Similarly, taking $v = \frac{Y_N^{(\beta)}u}{\|Y_N^{(\beta)}u\|_2}$, we find

$$\|Y_N^{(\beta)}u\|_2 = \langle v, Y_N^{(\beta)}u\rangle \leq \langle u_v, Y_N^{(\beta)}u\rangle + \|v-u_v\|_2\|Y_N^{(\beta)}v\|_2$$

from which we deduce that

$$\|Y_\beta^N\| \le 4 \sup_{u,v\in R_N} \langle Y_N^{(\beta)} u, v\rangle$$

and (1.32) follows. We next bound the probability of deviations of $\langle X_N^{(\beta)} v, u\rangle$ by using Tchebychev's inequality. For $\theta \ge 0$ we indeed have

$$\begin{aligned}
\mathbb{P}[\langle Y_N^{(\beta)} u, v\rangle \ge K] &\le \exp\{-\theta N K\}\mathbb{E}[\exp\{N\theta\langle Y_N^{(\beta)} u, v\rangle\}] \\
&\le \exp\{-\theta N K\}\mathbb{E}[\exp\left\{\sqrt{N}\left(2\sum_{i<j}\Re(A(i,j)a_{i,j}^{(\beta)} u_i \bar{v}_j) + \sum_i A(i,i)a_{i,i}u_i v_i\right)\right\}] \\
&\le \exp\{-\theta N K\}\exp\left(\frac{\theta^2 N}{\beta}(2\sum_{i<j}|u_i|^2|v_j|^2 + \sum_i |u_i|^2|v_i|^2)\right) \qquad (1.34)
\end{aligned}$$

where we used that the entries have a sharp sub-Gaussian Laplace transform and that $|A(i,i)| \le 1$. We can now complete the upper bound :

$$\mathbb{P}[\langle Y_N^{(\beta)} u, v\rangle \ge K] \quad \le \quad \exp\left(N\left(\frac{1}{\beta} - K\right)\right)$$

where we took $\theta = 1$. This complete the proof of the Lemma with (1.32).

1.4.6 Proof of Theorem 1.4.38 in the piecewise constant case

We consider in this section a random unitary vector e taken uniformly on the sphere $\mathbb{S}^{\beta N-1}$ and independent of $X_N^{(\beta)}$. We define F_N by setting, for $\theta > 0$:

$$F_N(\theta,\beta) = \frac{1}{N}\log \mathbb{E}_{X_N^{(\beta)}}\mathbb{E}_e[\exp(N\theta\langle e, X_N^{(\beta)} e\rangle)].$$

where we take both the expectation $\mathbb{E}_e$ over e and the expectation $\mathbb{E}_{X_N^{(\beta)}}$ over $X_N^{(\beta)}$. In this section we derive the asymptotics of $F_N(\theta,\beta)$, $F(\sigma,\theta)$ is as in Theorem 1.4.38.

We prove a refinement of Theorem 1.4.38, which shows that under our assumption of sharp sub-Gaussian tails, the random vector e stays delocalized under the tilted measure.

Proposition 1.4.41. *Suppose Assumption 1.4.1 holds if $\beta = 1$ and Assumption 1.4.3 holds if $\beta = 2$.Let $\psi_j := \sum_{i\in I_j}|e_i|^2$ and $V_N^\epsilon = \{e \in \mathbb{S}^{\beta N-1} \,|\, \forall j \in [1,n], \forall i \in I_j, |e_i| \le \sqrt{\psi_j}N^{-1/4-\epsilon}\}$. Then, for $\epsilon \in (0,\frac{1}{4})$,*

$$\begin{aligned}
F(\sigma,\theta) &= \lim_{N\to+\infty} F_N(\theta,\beta) \\
&= \lim_{N\to\infty}\frac{1}{N}\log \mathbb{E}_e[\mathbb{1}_{e\in V_N^\epsilon}\mathbb{E}_{X_N^{(\beta)}}[\exp(N\theta\langle e, X_N^{(\beta)} e\rangle)]].
\end{aligned}$$

Démonstration. By denoting $L_\mu = \log T_\mu$, we have :

$$\begin{aligned}
\mathbb{E}_{X_N^{(\beta)}}[\exp(N\theta\langle e, X_N^{(\beta)} e\rangle)] &= \mathbb{E}_{X_N^{(\beta)}}[\exp\{\sqrt{N}\theta(2\sum_{i<j}\Sigma_N(i,j)\Re(a_{i,j}^{(\beta)} e_j\bar{e}_i) + \sum_i \Sigma_N(i,i)a_{i,i}^{(\beta)}|e_i|^2)\}] \\
&= \exp\{\sum_{i<j} L_{\mu_{i,j}^N}(2\Sigma_N(i,j)\theta\bar{e}_i e_j\sqrt{N}) + \sum_i L_{\mu_{i,i}^N}(\Sigma_N(i,i)\theta|e_i|^2\sqrt{N})\}
\end{aligned}$$

where we used the independence of the $(a_{i,j}^{(\beta)})_{i\leq j}$. Using that the entries have a sharp sub-Gaussian Laplace transform (using on the diagonal the weaker bound $L_{\mu_{i,i}^N}(t) \leq \frac{1}{\beta}t^2 + A|t|$) and , we deduce that :

$$\begin{aligned}\mathbb{E}_{X_N^{(\beta)}}[\exp(N\theta\langle e, X_N^{(\beta)}e\rangle)] \quad &\leq \quad \mathbb{E}_e[\exp\{\frac{2N\theta^2}{\beta}\sum_{i<j}\Sigma_N(i,j)^2|e_i|^2|e_j|^2 \\ &+\frac{N\theta^2}{\beta}\sum_i \Sigma_N(i,i)^2|e_i|^4 + A\sqrt{N}\theta\sum_i \Sigma_N(i,i)e_i^2\}].\end{aligned}$$

Let's denote $\psi_j = \sum_{i\in I_j}|e_i|^2$ We have

$$\mathbb{E}_{X_N^{(\beta)}}[\exp(N\theta\langle e, X_N^{(\beta)}e\rangle)] \quad \leq \quad \exp\left(N\frac{\theta^2}{\beta}\vec{P}(\sigma,\psi_1,...,\psi_n)\right)\exp(A'\theta\sqrt{N}).$$

And so :

$$F_N(\theta,\beta) \leq \frac{1}{N}\log\mathbb{E}_e\left[\exp\left(\frac{\theta^2}{\beta}N\vec{P}(\sigma,\psi_1,...,\psi_n)\right)\right] + \frac{A'\theta}{\sqrt{N}}.$$

But since e is taken uniformly on the sphere, the vector $\psi = (\psi_1,...,\psi_n)$ follows a Dirichlet law of parameters $\left(\frac{\beta\alpha_1 N}{2},...,\frac{\beta\alpha_n N}{2}\right)$. We have the following large deviation principle for the Dirichlet Law :

Lemme 1.4.42. *Let $n \in \mathbb{N}^*$, and $\beta_N = (\beta_{1,N},...,\beta_{n,N}) \in (\mathbb{R}^+)^n$ be a sequence of vector such that $\lim_{N\to\infty}\frac{\beta_N}{N} = (\alpha_1,...,\alpha_n)$ and $\alpha_i > 0$. The sequence of Dirichlet laws $\delta_N = Dir(\beta_{1,N},...,\beta_{n,N})$ verifies a large deviation principle with good rate function $I(x_1,...,x_n) = \sum_{i=1}^n \alpha_i(\log x_i - \log\alpha_i)$.*

Démonstration. We denote f_N and f the functions defined on $D = \{x \in (\mathbb{R}^{+*})^n | x_1 + ... + x_n = 1\}$ by

$$f_N(x) = \sum_i \frac{\beta_{i,N}}{N}\log x_i$$

$$f(x) = \sum_i \alpha_i \log x_i.$$

For $x \in D$, let's denote $\tilde{x} = (x_1,...,x_{n-1})$. We have $d\delta_N(x) = Z_N^{-1}\exp(Nf_N(x_1,...,x_{n-1},(1-x_1-...-x_{n-1})))d\tilde{x}$ where

$$Z_N = \int_{\tilde{D}}\exp(Nf_N(x_1,...,x_{n-1},(1-x_1-...-x_{n-1})))dx_1...dx_{n-1}$$

We have that on every compact of $\tilde{D}$, $f_N(x, 1-\sum x_i)$ converges uniformly toward $f(x, 1-\sum x_i)$ (which is continuous) and furthermore, for every $M > 0$ there is a compact K of $\tilde{D}$ such that for N large enough $f_N(x, 1-\sum x_i) \leq -M$ for $x \notin K$. With both those remarks we deduce via a classical Laplace method that

$$\lim_{N\to\infty}\frac{1}{N}\log Z_N = \max_{x\in D} f(x, x - \sum_i^{n-1} x_i) = \sum_{i=1}^n \alpha_i\log\alpha_i.$$

Using again classical Laplace methods and the fact that $x \mapsto \tilde{x}$ is a homeomorphism between D and $\tilde{D}$, we have that the uniform convergence of f_N and the continuity of the limit f gives a weak LDP with rate function $f(x) - \sum_{i=1}^n \alpha_i\log\alpha_i$ and the bound outside compacts gives the exponential tightness. The LDP is proved. □

Using lemma 1.4.42 and Varadhan's lemma, we have that :

$$\lim_{N\to\infty}\frac{1}{N}\log\mathbb{E}_e[\exp\left(N\frac{\theta^2}{\beta}\vec{P}(\sigma,\psi_1,...,\psi_n)\right)] =$$
$$\sup_{\psi_1,\dots,\psi_n\in[0,1],\psi_1+\dots\psi_n=1}\{\frac{2\theta^2}{\beta}\vec{P}(\sigma,\psi_1,...,\psi_n)-\sum_{i=1}^n\frac{\beta\alpha_i}{2}\log(\psi_i)-\sum_{i=1}^n\frac{\beta\alpha_i}{2}\log(\alpha_i)\}=F(\sigma,\theta).$$

So that we have proved the upper bound that

$$\limsup_{N\to\infty}F_N(\theta,\beta)\le\limsup_{N\to\infty}\sup_{e\in\mathbb{S}^{N-1}}\frac{1}{N}\log\mathbb{E}_{X_N^{(\beta)}}[\exp(N\theta\langle e,X_N^{(\beta)}e\rangle)]\le F(\sigma,\theta). \tag{1.35}$$

We next prove the corresponding lower bound. The idea is that the expectation over the vector e concentrates on delocalized eigenvectors with entries so that $\sqrt{N}e_i\bar{e}_j$ is going to zero for all i,j. As a consequence we will be able to use the uniform lower bound on the Laplace transform to lower bound $F_N(\theta,\beta)$. We have that :

$$\mathbb{E}[\exp(N\theta\langle e,X_\beta^N e\rangle)]\ge\mathbb{E}_e[\mathbb{1}_{e\in V_N^\epsilon}\prod_{i<j}\exp\{L_{\mu_{i,j}^N}(2\sqrt{N}\theta\Sigma_N(i,j)\bar{e}_ie_j)\}\prod_i\exp\{L_{\mu_{i,i}^N}(\Sigma_N(i,i)\sqrt{N}\theta|e_i|^2)\}].$$

For $e\in V_N^\epsilon$, $2\Sigma_N(i,j)\sqrt{N}\theta|e_ie_j|\le2\Sigma_N(i,j)\theta N^{-\epsilon}$ so that :

$$\lim_{N\to+\infty}\sup_{i,j\in[1,N]}\sup_{e\in V_N^\epsilon}|2\sqrt{N}\Sigma_N(i,j)\theta e_ie_j|=0.$$

By the uniform lower bound on the Laplace transform of Assumptions 1.4.1 or 1.4.3, we deduce that for any $\delta>0$ and N large enough :

$$\mathbb{E}[\exp(N\theta\langle e,X_N^\beta e\rangle)]\ge\mathbb{E}_e[\mathbb{1}_{e\in V_N^\epsilon}\prod_{i<j}\exp\{2(1-\delta)N\theta^2\Sigma_N(i,j)^2|e_i|^2|e_j|^2/\beta\}\prod_i\exp\{N(1-\delta)\Sigma_N(i,i)^2\theta^2|e_i|^2/\beta\}] \tag{1.36}$$
$$\ge\mathbb{E}_e[\mathbb{1}_{e\in V_N^\epsilon}e^{N\frac{\theta^2}{\beta}\vec{P}(\sigma,\psi_1,\dots,\psi_n)(1-\delta)}]\,. \tag{1.37}$$

We shall use that

Lemme 1.4.43. *For any $\epsilon\in(0,1/4)$ we have*

$$\lim_{N\to\infty}\mathbb{P}_e[e\in V_N^\epsilon]=1$$

.

And that the event $\{e\in V_N^\epsilon\}$ is independent of the vector ψ. As a consequence, we deduce from (1.37) that for any $\delta>0$ and N large enough

$$\liminf_{N\to\infty}F_N(\theta,\beta)\ge\frac{1}{N}\log\mathbb{E}_e\left[\exp\left((1-\delta)N\frac{\theta^2}{\beta}\vec{P}(\sigma,\psi_1,...,\psi_n)\right)\right].$$

So that together with (1.35) and Lemma 1.4.42 we have proved the announced limit

$$\lim_{N\to\infty}F_N(\theta,\beta)=F(\sigma,\theta)$$

which completes the proof of Lemma 1.4.38.

Finally we prove Lemma 1.4.43. We have the well known fact that if we denote $e^{(j)} = (e_i)_{i\in I_j}$, $f^{(j)} := e^{(j)}/||e^{(j)}||$ is a uniform unitary vector on the sphere of dimension $\beta|I_j|-1$. Furthermore all these f_j are independent.

$$\mathbb{P}[V_N^\epsilon] = \prod_{j=1}^n \mathbb{P}[\forall i \in I_j, |f_i^{(j)}| \leq N^{-1/4-\epsilon}].$$

The result follows since each of these terms converges to 1.

□

1.4.7 Large deviation lower bounds

We will now prove Theorem 1.4.32. For a vector e of the sphere $\mathbb{S}^{\beta N-1}$ and X a random symmetric or hermitian matrix, we denote by $\mathbb{P}_N^{(e,\theta)}$ the tilted probability measure defined by :

$$d\mathbb{P}_N^{(e,\theta)}(X) = \frac{\exp(N\theta\langle Xe, e\rangle)}{\mathbb{E}_X[\exp(N\theta\langle X_N^{(\beta)}e, e\rangle)]} d\mathbb{P}_N(X)$$

where $\mathbb{P}_N$ is the law of $X_N^{(\beta)}$. Let us show that we only need to prove the following lemma :

Lemme 1.4.44. *Suppose that the hypotheses of Theorem 1.4.32 hold. Let $(\delta_N)_{N\in\mathbb{N}}$ be some arbitrary sequence of positive reals converging to 0, W_N the subset of the sphere $\mathbb{S}^{\beta N-1}$ defined by :*

$$W_N := \{e \in \mathbb{S}^{\beta N-1} : \forall i \left| ||e^{(i)}||^2 - \psi_i^{E,\theta} \right| \leq \delta_N\}$$

where E and $\psi^{E,\theta}$ are as in the hypotheses of Theorem 1.4.32 and $e^{(i)}$ is the i-th block of entries of e. For any $x \geq r_\sigma$, there is θ_x such that :

$$\lim_{N\to\infty} \inf_{e\in V_N^\epsilon \cap W_N} \mathbb{P}^{(e,\theta_x)}[X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M] = 1$$

Proof that Lemma 1.4.44 implies Theorem 1.4.32. In fact, we only need to prove that if E, ψ_i^{E,θ_x} are as in the hypotheses of Theorem 1.4.32, we have :

$$\liminf_{N\to\infty} \frac{1}{N} \log \frac{\mathbb{E}[\mathbb{1}_{X_N^{(\beta)}\in\mathcal{A}_{x,\delta}^M} I_N(X_N^{(\beta)}, \theta_x)]}{\mathbb{E}[I_N(X_N^{(\beta)}, \theta_x)]} \geq -E(\theta_x)\,,$$

where we recall that

$$\mathcal{A}_{x,\delta}^M = \{X : |\lambda_{\max}(X) - x| < \delta\} \cap \{d(\hat{\mu}_X^N, \mu_\sigma) < N^{-\kappa'}\} \cap \{\|X\| \leq M\}\,.$$

We have

$$\begin{aligned} \mathbb{E}[\mathbb{1}_{X_N^{(\beta)}\in\mathcal{A}_{x,\delta}^M} I_N(X_N^{(\beta)}, \theta_x)] &= \mathbb{E}_e[\mathbb{P}_N^{(e,\theta_x)}(\mathcal{A}_{x,\delta}^M)\mathbb{E}_X[\exp(N\theta_x\langle X_N^{(\beta)}e, e\rangle)]] \\ &\geq \mathbb{E}_e[1_{e\in V_N^\epsilon}\mathbb{P}_N^{(e,\theta_x)}(\mathcal{A}_{x,\delta}^M)\mathbb{E}_X[\exp(N\theta_x\langle X_N^{(\beta)}e, e\rangle)]] \end{aligned}$$

where we recall that $V_N^\epsilon = \{e \in \mathbb{S}^{N-1} : |e_i| \leq N^{-1/4-\epsilon}\psi_i\}$. Let

$$W_N^\delta := \{e \in \mathbb{S}^{\beta N-1} : \forall i, \left| ||e^{(i)}||^2 - \psi_i^{\theta_x,E} \right| \leq \delta\}$$

We have, using using Lemma 1.4.42 that :

$$\lim_{\delta\to 0} \liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}[e \in W_N^\delta] = \frac{2}{\beta} \sum_{i=1}^n \alpha_i (\log \psi_i^{\theta,E} - \log \alpha_i)$$

let $(\delta_N)_{N\to\infty}$ be a sequence converging to 0 such that :

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}[e \in W_N^{\delta_N}] \geq \frac{2}{\beta} \sum_{i=1}^{n} \alpha_i (\log \psi_i^{\theta_x, E} - \log \alpha_i)$$

and let :

$$W_N := \{e \in \mathbb{S}^{N-1} : \forall i, \left| ||e^{(i)}||^2 - \psi_i^{\theta_x, E} \right| \leq \delta_N \}$$

We have that for $e \in W_N$:

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}_X[\exp(N\theta_x \langle X_N^{(\beta)} e, e\rangle)]] = \vec{P}(\sigma, \psi^{E,\theta_x})$$

This is in fact identical to (1.37).

$$\begin{aligned} \mathbb{E}_e[\mathbb{P}_N^{(e,\theta_x)}(\mathcal{A}_{x,\delta}^M)\mathbb{E}_X[\exp(N\theta_x \langle X_N^{(\beta)} e, e\rangle)]] &\geq \mathbb{E}_e[\mathbb{1}_{e\in V_N^\epsilon \cap W_N} \mathbb{P}_N^{(e,\theta_x)}(\mathcal{A}_{x,\delta}^M)\mathbb{E}_X[\exp(N\theta_x \langle X_N^{(\beta)} e, e\rangle)]] \\ &\geq \mathbb{E}_e[\mathbb{1}_{e\in V_N^\epsilon \cap W_N} \mathbb{P}_N^{(e,\theta_x)}(\mathcal{A}_{x,\delta}^M) e^{-N\theta_x^2 \vec{P}(\sigma,\psi^{E,\theta_x}) + o(N)}] \end{aligned}$$

so we have that :

$$\begin{aligned} \frac{\mathbb{E}[\mathbb{1}_{X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M} I_N(X_N^{(\beta)}, \theta_x)]}{\mathbb{E}[I_N(X_N^{(\beta)}, \theta_x)]} &\geq \mathbb{E}_e[\mathbb{1}_{e\in V_N^\epsilon \cap W_N} \mathbb{P}^{(e,\theta_x)}[\mathcal{A}_{x,\delta}^M]] e^{-N(\theta_x^2 \vec{P}(\sigma,\psi^{E,\theta_x}) - F(\theta_x))} \\ &\geq \mathbb{P}[\mathbb{1}_{e\in V_N^\epsilon \cap W_N}] \inf_{e\in V_N^\epsilon \cap W_N} \mathbb{P}^{(e,\theta_x)}[X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M] e^{-N(\theta_x^2 \vec{P}(\sigma,\psi^{E,\theta_x}) - F(\theta_x) + o(1))} \\ &\geq \mathbb{P}[\mathbb{1}_{e\in V_N^\epsilon}]\mathbb{P}[\mathbb{1}_{e\in W_N}] \inf_{e\in V_N^\epsilon \cap W_N} \mathbb{P}^{(e,\theta_x)}[X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M] e^{-N(\theta_x^2 \vec{P}(\sigma,\psi^{E,\theta}) - F(\theta) + o(1))} \\ &\geq e^{-N(E(\theta_x) + o(1))} \end{aligned}$$

where we used that $\{e \in V_N^\epsilon\}$ and $\{e \in W_N^\epsilon\}$ are independent and that $\frac{1}{N} \log \inf_{e\in V_N^\epsilon \cap W_N} \mathbb{P}^{(e,\theta_x)}[X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M]$ converges to 0. So we have our lower bound. □

And so it remains to prove the lemma 1.4.44. More precisely, we will show that for $\epsilon \in (\frac{1}{8}, \frac{1}{4})$, for any $x > r_\sigma$ and $\delta > 0$ we can find $\theta_x \geq 0$ so that for M large enough,

$$\lim_{N\to\infty} \inf_{e\in V_N^\epsilon \cap W_N} \mathbb{P}_N^{(e,\theta_x)}(\mathcal{A}_{x,\delta}^M) = 1 \,. \tag{1.38}$$

To prove (1.38), the first point is to show that

Lemme 1.4.45. *Take $\epsilon \in (0, \frac{1}{4})$. There exists $\kappa > 0$, such that for any $\theta \geq 0$,*

— *for K large enough :*

$$\lim_{N\to\infty} \sup_{e\in V_N^\epsilon \cap W_N} \mathbb{P}_N^{(e,\theta)} \left(\lambda_{\max}(X_N^{(\beta)}) \geq K\right) = 0$$

—

$$\limsup_{N\to\infty} \sup_{e\in V_N^\epsilon \cap W_N} \mathbb{P}_N^{(e,\theta)} \left(d(\hat{\mu}_{X_N^{(\beta)}}^N, \mu_\sigma) > N^{-\kappa}\right) = 0 \,.$$

Démonstration. We hereafter fix a vector e on the sphere. The proof of the exponential tightness is exactly the same as for Lemma 1.4.29. Indeed, by Jensen's inequality, we have

$$\mathbb{E}_X[\exp(N\theta \langle X_N^{(\beta)} e, e\rangle)] \geq \exp\{N\theta \mathbb{E}_X[\langle X_N^{(\beta)} e, e\rangle]\} = 1$$

Moreover, by Tchebychev's inequality, for any $u, v, e \in \mathbb{S}^{\beta N-1}$, and if $M = \sup_{i,j} \sigma_{i,j}^2$ we have

$$\begin{aligned}\int \mathbb{1}_{\langle X_N^{(\beta)} u,v\rangle \geq K} \exp(N\theta\langle X_N^{(\beta)} e, e\rangle) d\mathbb{P}_N &\leq \exp\{-NK\}\mathbb{E}_X[\exp(N\theta\langle X_N^{(\beta)} e, e\rangle + N\langle X_N^{(\beta)} u, v\rangle)] \\ &\leq \exp\{-NK\}\exp\{NM(\theta^2+1)\sum_{i,j} |e_i\bar{e}_j + u_i\bar{v}_j|^2\} \\ &\leq \exp\{-NK + 4(\theta^2+1)MN\}\end{aligned}$$

from which we deduce after taking u, v on a δ-net as in Lemma 1.4.29 that

$$\mathbb{P}_N^{(e,\theta)}\left(\lambda_{\max}(X_N^{(\beta)}) \geq K\right) \leq 9^{\beta N}\exp\{-\frac{1}{4}NK + 4(\theta^2+1)MN\}$$

which proves the first point. The second is a direct consequence of Lemma 1.4.33 and the fact that the log density of $\mathbb{P}_N^{(e,\theta)}$ with respect to $\mathbb{P}_N$ is bounded by $\theta N(|\lambda_{\max}(X)| + |\lambda_{min}(E)|)$ which is bounded by θKN with overwhelming probability by the previous point (recall that $\lambda_{\min}(X)$ satisfies the same bounds than $\lambda_{\max}(X)$).

□

Hence, the main point of the proof is to show that

Lemme 1.4.46. *Pick $\epsilon \in]\frac{1}{8}, \frac{1}{4}[$. For any $x > r_\sigma$, there exists θ_x such that for every $\eta > 0$,*

$$\lim_{N\to\infty} \sup_{e \in V_N^\epsilon \cap W_N} \mathbb{P}_N^{(e,\theta_x)}[|\lambda_{\max} - x| \geq \eta] = 0.$$

1.4.7.1 Proof of Lemma 1.4.46

For $e \in V_N^\epsilon$ fixed, let $X^{(e),N}$ be a matrix with law $\mathbb{P}_N^{(e,\theta)}$. We have :

$$X^{(e),N} = \mathbb{E}[X^{(e),N}] + (X^{(e),N} - \mathbb{E}[X^{(e),N}])$$

where $\mathbb{E}[X]$ denotes the matrix with entries given by the expectation of the entries of the matrix X. We first show that $\mathbb{E}[X^{(e),N}]$ is approximately a finite rank matrix and $X^{(e),N} - \mathbb{E}[X^{(e),N}]$ is approximately a Wigner matrix with variance profile σ.

Lemme 1.4.47. *For $\epsilon \in]\frac{1}{8}, \frac{1}{4}[$, there exists $\kappa(\epsilon) > 0$ so that for $e \in V_N^\epsilon$:*

$$\mathbb{E}[X^{(e),N}] = \frac{2\theta}{\beta} ESE^* + \Delta^{(e),N}$$

where,

$$E = \begin{pmatrix} e^{(1)} & 0 & \cdots & 0 \\ 0 & e^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{(n)} \end{pmatrix}$$

and :

$$S := (\sigma_{i,j}^2)_{i,j\in[1,n]}$$

The spectral radius of $\Delta^{(e),N}$ is bounded by $N^{-\kappa(\epsilon)}$ uniformly on $e \in V_N^\epsilon$.

Proof of the lemma. We can express the density of $\mathbb{P}_N^{(e,\theta)}$ as the following product :

$$\frac{d\mathbb{P}_N^{(e,\theta)}}{d\mathbb{P}_{X_N}}(X) = \prod_{i\leq j} \exp(2^{1_{i\neq j}}\theta\Sigma_N(i,j)\sqrt{N}\Re(e_i\bar{e}_j a_{i,j}^{(\beta)}) - L_{\mu_{i,j}^N}(2^{1_{i\neq j}}\theta\Sigma_N(i,j)\sqrt{N}e_i\bar{e}_j))$$

where the $a_{i,j}^{(\beta)}$ are defined as in the introduction.

So since we took our $a_{i,j}$ independent (for $i \leq j$), the entries $X_{i,j}^{(e),N}$ remain independent and their mean is given in function of the Taylor expansion of L as follows :

$$(\mathbb{E}[X^{(e),N})])_{i,j} = \frac{\Sigma_N(i,j) L'_{\mu_{i,j}^N}(2\sqrt{N}\Sigma_N(i,j)\theta e_i \bar{e}_j)}{\sqrt{N}} = \frac{2\theta}{\beta}\Sigma_N^2(i,j)e_i\bar{e}_j + \frac{\Sigma_N(i,j)\delta_{i,j}(2\Sigma_N(i,j)\sqrt{N}\theta e_i\bar{e}_j)N\theta^2|e_i|^2|e_j|^2}{\sqrt{N}}$$

if $i \neq j$, and if $i = j$

$$(\mathbb{E}[X^{(e),N}])_{i,i} = \frac{\Sigma_N(i,i) L'_{\mu_{i,i}^N}(\sqrt{N}\Sigma_N(i,i)\theta |e_i|^2)}{\sqrt{N}} = \frac{2\theta}{\beta}\Sigma_N^2(i,i)e_i\bar{e}_i + \frac{\Sigma_N(i,i)\delta_{i,i}(2\sqrt{N}\Sigma_N(i,i)\theta|e_i|^2)N\theta^2|e_i|^4}{\sqrt{N}}$$

where we used that by centering and variance one, $L'_{\mu_{i,j}^N}(0) = 0$, $Hess L_{\mu_{i,j}^N}(0) = \frac{1}{\beta} Id$ for all $i \neq j, N$, $L''_{\mu_{i,i}^N}(0) = \frac{2}{\beta}$ for all i, N, and where

$$|\delta_{i,j}(t)| \leq 4 \sup_{|u|<t} \max_{i,j,N} \{|L^{(3)}_{\mu_{i,j}^N}(u)|\} .$$

Hence, we have

$$\Delta_{i,j}^{(e),N} = \Sigma_N(i,j)\delta_{i,j}(2\sqrt{N}\sigma_{i,j}\theta e_i\bar{e}_j)\sqrt{N}\theta^2|e_i|^2|e_j|^2, 1 \leq i,j \leq N .$$

In order to bound the spectral radius of this remainder term, we use the following result which is Lemma 5.4 from [45].

Lemme 1.4.48. *Let A be an Hermitian matrix and B a real symmetric matrix such that :*

$$\forall i,j, |A_{i,j}| \leq B_{i,j} .$$

Then the spectral radius of A is smaller than the spectral radius of B.

Therefore, if we choose C so that $C \geq \sup_{N,i,j} \delta_{i,j}(2\sqrt{N}\theta e_i \bar{e}_j)\theta^2$ and set $|e|^2$ to be the vector with entries $(|e_i|^2)_{1\leq i \leq N}$, we have

$$||\Delta^{(e),N}|| \leq C\sqrt{N}|||e|^2(|e|^2)^*||.$$

Since $|||e|^2(|e|^2)^*|| = |||e|^2||_2^2 = \sum_i e_i^4 \leq N^{-4\epsilon}$, we deduce that if we take $\epsilon' \in]1/8, 1/4[$ we have with $\kappa(\epsilon) = 1/2 - 4\epsilon$:

$$||\Delta^{(e),N}|| = O(N^{-\kappa(\epsilon)}) .$$

□

Now we denote :

$$\overline{X^{(e),N}} := X^{(e),N} - \mathbb{E}[X^{(e),N}].$$

The entries of $\overline{X^{(e),N}}$ are independent, centered of variance $\Sigma_N(i,j)^2\partial_z\partial_{\bar z}L_{\mu^N_{i,j}}(\theta\Sigma_N(i,j)e_i\bar e_j\sqrt{N})/N$. Recall that $\partial_z\partial_{\bar z}L_{\mu^N_{i,j}}(0)=1$ and that the third derivative of the Laplace transform of the entries are uniformly bounded so that :

$$\partial_z\partial_{\bar z}L_{\mu^N_{i,j}}(\theta\Sigma_N(i,j)e_i\bar e_j\sqrt{N}) = 1+\delta_{i,j}(\sqrt{N}\Sigma_N(i,j)|e_ie_j|) = 1+O(N^{-2\epsilon})$$

uniformly on V_N^ϵ. We can then consider $\widetilde{X}^{(e),N}$ defined by :

$$\widetilde{X}^{(e),N}_{i,j} = \frac{\overline{X}^{(e),N}_{i,j}}{\sqrt{\partial_z\partial_{\bar z}L_{\mu^N_{i,j}}(\theta\Sigma_N(i,j)e_i\bar e_j\sqrt{N})}}$$

Set $Y^{(e),N}=\overline{X}^{(e),N}-\widetilde{X}^{(e),N}$. So, we have

$$(Y^{(e),N})_{i,j} = \overline{X}^{(e),N}_{i,j}\left(1-\frac{1}{\sqrt{\partial_z\partial_{\bar z}L_{\mu^N_{i,j}}(\theta\Sigma_N(i,j)e_i\bar e_j\sqrt{N})}}\right).$$

Next, we have that for all $\delta>0$:

$$\lim_{N\to+\infty}\sup_{e\in V_N^\epsilon}\mathbb{P}[||Y^{(e),N}||>\delta]=0. \tag{1.39}$$

This follows the demonstration as in [45]. Hence, since :

$$X^{(e),N} = \widetilde{X}^{(e),N}+\frac{2\theta}{\beta}ESE^*+\Delta^{(e),N}+Y^{(e),N},$$

we conclude by combining (1.39) and Lemma 1.4.47 that for $\epsilon\in]1/4,1/8[$ and all $\delta>0$

$$\lim_{N\to\infty}\sup_{e\in V_N^\epsilon}\mathbb{P}_N^{(e,\theta)}[||X^{(e),N}-(\widetilde{X}^{(e),N}+\frac{2\theta}{\beta}ESE^*)||>\delta]=0 \tag{1.40}$$

since all estimates were clearly uniform on $e\in V_N^\epsilon$.

And so, to conclude we need only to identify the limit of $\lambda_{\max}(\widetilde{X}^{(e),N}+\frac{2\theta}{\beta}ESE^*)$. The largest eigenvalue of $\widetilde{X}^{(e),N}+\frac{2\theta}{\beta}ESE^*$ satisfy

$$0=\det(z-\widetilde{X}^{(e),N}-\frac{2\theta}{\beta}ESE^*)=\det(z-\widetilde{X}^{(e),N})\det(1-\frac{2\theta}{\beta}(z-\widetilde{X}^{(e),N})^{-1}ESE^*)$$

and therefore z is an eigenvalue away from the spectrum of $\widetilde{X}^{(e),N}$ iff

$$\det(1-\frac{2\theta}{\beta}(z-\widetilde{X}^{(e),N})^{-1}ESE^*)=0.$$

Using the fact that with $A\in\mathcal{M}_{n,p}(\mathbb{K})$ and $A\in\mathcal{M}_{p,n}(\mathbb{K})$ we have $\det(I_n+AB)=\det(I_p+BA)$, we have that the preceding equality is equivalent to

$$\det(I_n-\theta E^*(z-\widetilde{X}^{(e),N})^{-1}ES)=0.$$

Lemme 1.4.49. *For $i,j\in[1,n]$ $\eta>0$, $a>r_\sigma$, we have :*

$$\sup_{e\in V_N^\epsilon\cap W_N}\mathbb{P}\left[\sup_{z\geq a}\left|(E^*(z-\widetilde{X}^{(e),N})^{-1}ES)_{i,j}-\sigma_{i,j}^2\psi_j^\theta m_j(z)\right|\geq\eta\right]\to 0$$

Where m is the solution to K_σ and m_i are defined as in Theorem 1.4.13.

Démonstration. Let $G_N(z) := (z - \widetilde{X}^{(e),N})^{-1}$, $M_N(z) = diag(m_{1,N}(z), ..., m_{N,N}(z))$, where $m_x^N := \sum_{i=0}^{n-1} \mathbb{1}_{Nx \in]i-1,i]} m_{i,N}$ and m^N is the solution of K_{σ^N} (as defined in section 1.4.2). If we denote $\tilde{e}^k = (\mathbb{1}_{j \in NI_k} e_j)_{j=1,...,N}$ the vector e where we zeroed out all entries except for the k-th block.

$$(E^*(z - \widetilde{X}^{(e),N})^{-1} ES)_{i,j} = \sum_{k=1}^{n} (\tilde{e}^i)^* G_N(z) \sigma_{k,j}^2 \tilde{e}^k$$

So since $\lim_{N\to\infty} \sup_{e \in V_N^\epsilon \cap W_N} \left| ||\tilde{e}^k||^2 - \psi_k^\theta \right| = 0$ we only need to prove that for $k, l \in [1, n]$:

$$\lim_{N\to\infty} \sup_{e \in \mathbb{S}^{\beta N-1}} \sup_{z \geq a} \mathbb{P}[|(\tilde{e}^k)^* G_N(z) \tilde{e}^l - \delta_{k,l} m_k(z) \psi_k^\theta| \geq \eta] = 0.$$

To that end, we want to apply the anisotropic local law from [2] but in order to do so, we need to check its assumptions. (A) is verified since the variance profiles are uniformly bounded. (B) is verified with the assumption 1.4.14. (D) is verified with the sub-Gaussian bound. To verify (C), we apply [3, Theorem 6.1]. Thanks to [2, Theorem 1.13], if we fix some $\gamma > 0$, $D > 0$, $\epsilon > 0$, for N large enough :

$$\sup_{e,f \in \mathbb{S}^{\beta N-1}} \sup_{z \in \mathbb{C}, \Im z \geq N^{\gamma - 1}} \mathbb{P}[|e^* G_N(z) f - e^* M_N(z) f| \geq N^{-1/10}] \leq N^{-D}.$$

Furthermore following Corollary 1.7 of [2], we have that for $a' \in]r_\sigma, a[$, $D > 0$, N large enough

$$\mathbb{P}[\lambda_{\max}(\widetilde{X}^{(e),N}) \geq a'] \leq N^{-D}.$$

Let $e, f \in \mathbb{S}^{\beta N-1}$ and $h : z \mapsto e^* G_N(z) f$ and $k : z \mapsto e^* M_N(z) f$. On the event $\{\lambda_{\max}(\widetilde{X}^{(e),N}) < a'\}$, we have that $|h(z)|, |k(z)| \leq \frac{1}{(\Re z - a')}$ and $|h'(z)|, |k'(z)| \leq \frac{1}{(\Re z - a')^2}$ for $\{z | \Re z > a'\}$ and therefore, for $\gamma < 1/10$, we can in fact assume that our bound holds for any z such that $\Re(z) > a$ and in particular for z real. Let

$$A_N := \{a + k/N | k \in [0, N^2]\}.$$

By union bound, we have that for N large enough :

$$\mathbb{P}[\sup_{z \in A_N} |(e^* G_N(z) f - e^* M_N(z) f)| \geq N^{-1/10}] \leq N^{-D+2}$$

Combining this with the Lipschitz bounds and the bound in modulus that is derived from from the bound on $\lambda_{\max}(\widetilde{X}^{(e),N})$, we get :

$$\mathbb{P}[\sup_{z > a} |e^* G_N(z) f - e^* M_N(z) f| \geq N^{-1/10}] \leq N^{-D+2}$$

for N large enough. furthermore, this bound is uniform in e and f. We use Theorem 1.4.13 and the Lipschitz bounds to conclude that for N large enough and $e \in V_N^\epsilon \cap W_N$ we have that :

$$\mathbb{P}[\sup_{z > a} |(\tilde{e}^k)^* (M_N(z)) \tilde{e}^l - \delta_{k,l} \psi_k^\theta m_k(z)| \geq \eta |] \leq N^{-D+2}$$

where m is the solution of K_σ and m_i is the value taken by m on the interval I_i. And so we have :

$$\mathbb{P}[\sup_{z > a} |(E^* G_N(z) SE)_{i,j} - m_j(z) \sigma_{i,j} \psi_j^\theta| > \eta] \leq N^{-D+2} \text{ for every } i, j \in [1, N].$$

□

Let's denote $D(\theta, z)$ the diagonal $n \times n$ matrix $\text{diag}(m_1(z)\psi_1^\theta, ..., m_n(z)\psi_n^\theta)$, we have that the above limit can be rewritten $SD(\theta, z)$. From the preceding lemma we have that for $\eta > 0$ uniformly in $e \in \mathbb{S}^{\beta N-1}$ that

$$\mathbb{P}[\sup_{z > a} |\det(I_n - \theta E^*(z - \widetilde{X}^{(e),N})^{-1} ES) - \det(I_n - SD(\theta, z))| \geq \eta] \leq N^{-D}$$

for N large enough.

So since $\lim_{z\to\infty}\det(I_n - SD(\theta,z)) = 1$, all that remain is to solve the determinantal equation :

$$\det(I_n - SD(\theta,z)) = 0$$

and the largest solution $z > r_\sigma$ if it exists will be the the limit of $\lambda_{\max}$. We can rewrite this equation :

$$\det(I_n - \theta\sqrt{D(\theta,z)}S\sqrt{D(\theta,z)}) = 0. \tag{1.41}$$

Let $\rho(\theta,z)$ be the largest eigenvalue of $\sqrt{D(\theta,z)}S\sqrt{D(\theta,z)}$. Then, the largest z solution of equation 1.41 is the solution of :

$$\theta\rho(\theta,z) = 1. \tag{1.42}$$

Indeed, with θ fixed, if $\theta\rho(\theta,z) = 1$ then z is a solution of equation 1.41. Since the $z \mapsto m_i(z)$ are strictly decreasing, so is $\rho(\theta,.)$. So for $z' > z$, $\theta\rho(\theta,z') < 1$ and so z' cannot be solution of 1.41 for the same θ. Similarly, if z is a solution of 1.41 then $\theta\rho(\theta,z) \geq 1$. If $\theta\rho(\theta,z) > 1$ then since $z \mapsto \theta\rho(\theta,z)$ is continuous and decreasing toward 0, there exists $z' > z$ such that $\theta\rho(\theta,z') = 1$ and z' is therefore a solution of 1.41 strictly larger than z.

Therefore, it suffices to prove that for any $x > r_\sigma$ there is at least one θ_x such that

$$\theta_x\rho(\theta_x,x) = 1$$

First, since $\theta \mapsto D(\theta,x)$ is bounded and continuous on $\mathbb{R}^+$, then $\theta \mapsto \theta\rho(\theta,x)$ is continuous. For $\theta = 0$ the lefthand side is 0 and for $\theta \to \infty$, since $\max_i \psi_i^\theta \geq n^{-1}$ we have that

$$\max_{i,j}(\sqrt{D(\theta,z)}S\sqrt{D(\theta,z)})_{i,j} \geq n^{-1}(\min_i m_i(z))(\min_{i,j}\sigma_{i,j}^2)$$

and so with $M := n^{-1}(\min_i m_i(x))(\min_{i,j}\sigma_{i,j}^2)$, $\rho(\theta,x) \geq M$ and so $\theta\rho(\theta,x) \underset{\theta\to\infty}{\longrightarrow} +\infty$. By continuity, there is at least one θ_x such that $\theta_x\rho(\theta_x,x) = 1$ and so Theorem 1.4.32 is proved.

1.4.8 Case of a continuous variance profile and relaxing Assumption 1.4.14 for the piecewise constant case

We now choose $\sigma : [0,1]^2 \mapsto \mathbb{R}^+$ continuous and symmetric and consider the random matrix model $X_N^{(\beta)} := \Sigma_N \odot W_N^{(\beta)}$ where

$$\Sigma_N(i,j) := \sigma\left(\frac{i}{N}, \frac{j}{N}\right)$$

In order to prove a large deviation principle for X_N, we will approximate the variance profile by a piecewise constant σ. Namely, for $n \in \mathbb{N}$ we let σ^n be the following $n \times n$ matrix :

$$\sigma_{i,j}^n = n^2 \int_{\frac{i-1}{n}}^{\frac{i}{n}} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \sigma(x,y)dxdy + \frac{1}{n+1}.$$

Let's denote $X_N^{(\beta),n}$ the random matrix constructed with the same family of random variables $a_{i,j}^{(\beta)}$ but with the piecewise constant variance profile associated we the the matrix σ^n and the vector of parameters $(\frac{1}{n}, ..., \frac{1}{n})$. Let $F^n = F(\sigma^n,.)$, $\mu^n := \mu_{\sigma^n}$. We will also denote $F = F(\sigma,.)$ and $I = I(\sigma,.)$. Even if we suppose that Assumption 1.4.25 holds in the case of the continuous variance profile σ, we don't necessarily have Assumption 1.4.21 for the variance profiles σ^n and so we don't necessarily have a sharp lower bound. To this end we need to introduce an error term E^n that will be negligible as n tends to ∞ :

Lemme 1.4.50.

$$\lim_{n\to\infty}\sup_{\theta>0}\frac{|F^n(\theta) - F(\theta)|}{\theta^2} = 0$$

Démonstration.

$$|\int_0^1\int_0^1 (\sigma^n)^2(x,y)d\mu(x)d\mu(y) - \int_0^1\int_0^1 \sigma^2(x,y)d\mu(x)d\mu(y)| \leq \int_0^1\int_0^1 |(\sigma^n)^2(x,y) - \sigma^2(x,y)|d\mu(x)d\mu(y).$$

Since $\lim_{n\to\infty}\sup_{x,y}|(\sigma^n)^2(x,y)-\sigma^2(x,y)| = 0$, we have $\lim_{n\to\infty}\sup_\psi|\vec{P}(\sigma,\psi)-\vec{P}(\sigma^n,\psi)| = 0$. The result follows easily.

□

Lemme 1.4.51. *If the Assumption 1.4.25 is true, then for every $\epsilon > 0$, there is a sequence of functions E^n and continuous $\theta \mapsto (\psi_i^{\theta,E^n})_{i\in[1,n]}$ such that :*

$$\vec{\Psi}(\sigma^n,\theta,\psi^{\theta,E^n}) = F^n(\theta) - E^n(\theta)$$

and there is a n_0 such that for $n \geq n_0$:

$$\forall\theta > 0, E^n(\theta) \leq \epsilon\theta^2.$$

Démonstration. Since assumption 1.4.25 is verified, there is some measure valued continuous $\theta \mapsto \psi^\theta$ such that $F(\theta) = \theta^2 P(\sigma,\psi^\theta)/\beta - \beta D(Leb||\psi^\theta)/2$. Let $\psi^{\theta,\epsilon} := K(\psi^\theta * \tau_\epsilon)$ where $*$ is the convolution, τ_ϵ the probability measure whose density is a triangular function of support $[-\epsilon,\epsilon]$ and K the function defined by $K(x) = x$ if $x \in [0,1]$, $K(x) = -x$ if $x \in [-1,0]$ and $K(x) = 2-x$ if $x \in [1,2]$.

$$\psi_i^{\theta,\epsilon,n} := \psi^{\theta,\epsilon}\left(\left[\frac{i-1}{n},\frac{i}{n}\right]\right)$$

we have that for $i = 1,...,n$:

$$\psi_i^{\theta,\epsilon,n} := \int_{\mathbb{R}}(\mathbb{1}_{[(i-1)/n,i/n]} + \mathbb{1}_{[-i/n,(1-i)/n]} + \mathbb{1}_{[2-i/n,2-(1-i)/n]})(x)d(\psi^\theta * \tau_\epsilon)(x).$$

So for $i = 1,...,n$:

$$\psi_i^{\theta,\epsilon,n} := \int_{\mathbb{R}}(\mathbb{1}_{[(i-1)/n,i/n]} + \mathbb{1}_{[-i/n,(1-i)/n]} + \mathbb{1}_{[2-i/n,2-(1-i)/n]}) * \tau_\epsilon(x)d\psi^\theta(x)$$

Since $x \mapsto (\mathbb{1}_{[(i-1)/n,i/n]} + \mathbb{1}_{[-i/n,(1-i)/n]} + \mathbb{1}_{[2-i/n,2-(1-i)/n]}) * \tau_\epsilon(x)$ is continuous and $\theta \mapsto \psi^\theta$ is continuous for the weak topology then $\theta \mapsto \psi_i^{\theta,\epsilon,n}$ is continuous for $i = 1,...,n$.

Let us prove the following lemma :

Lemme 1.4.52. *For every $\eta > 0$, there is $\epsilon > 0, n_0 > 0$ such that for every $\theta > 0, n \geq n_0$*

$$|\sum_{i,j=1}^n(\sigma_{i,j}^n)^2\psi_i^{\theta,\epsilon,n}\psi_j^{\theta,\epsilon,n} - \int_0^1\int_0^1\sigma^2(x,y)d\psi^\theta(x)d\psi^\theta(x)| \leq \eta$$

and,

$$\frac{1}{n}\sum_i^n\left(\log\psi_i^{\theta,\epsilon,n} - \log n\right) \geq D(Leb||\psi^\theta).$$

Proof of the lemma. Let $\eta > 0$ and let us find $\epsilon > 0$ such that :

$$|\int_0^1 \int_0^1 \sigma^2(x,y) d\psi^{\theta,\epsilon}(x) d\psi^{\theta,\epsilon}(x) - \int_0^1 \int_0^1 \sigma^2(x,y) d\psi^{\theta}(x) d\psi^{\theta}(x)| \leq \eta.$$

Let us take X, Y, U_ϵ, V_ϵ independent random variables of law respectively, ψ^θ, ψ^θ and $\tau_\epsilon, \tau_\epsilon$. Then we have

$$|\int_0^1 \int_0^1 \sigma^2(x,y) d\psi^{\theta,\epsilon}(x) d\psi^{\theta,\epsilon}(x) - \int_0^1 \int_0^1 \sigma^2(x,y) d\psi^{\theta}(x) d\psi^{\theta}(x)| = \\ |\mathbb{E}[\sigma^2(K(X+U_\epsilon), K(Y+V_\epsilon)) - \sigma^2(X,Y)]|.$$

Using the uniform continuity of σ^2, and that $|K(X+U_\epsilon) - X|, |K(Y+V_\epsilon) - Y| \leq \epsilon$ almost surely, we have that there exists an $\epsilon > 0$ such that the difference is lower that η. This bound does not depend on θ.

Now, let us find n_0 such that for $n \geq n_0$,

$$|\sum_{i,j=1}^{n} (\sigma_{i,j}^n)^2 \psi_i^{\theta,\epsilon,n} \psi_j^{\theta,\epsilon,n} - \int_0^1 \int_0^1 \sigma^2(x,y) d\psi^{\theta,\epsilon}(x) d\psi^{\theta,\epsilon}(x)| \leq \eta.$$

We have

$$|\sum_{i,j=1}^{n} (\sigma_{i,j}^n)^2 \psi_i^{\theta,\epsilon,n} \psi_j^{\theta,\epsilon,n} - \int_0^1 \int_0^1 \sigma^2(x,y) d\psi^{\theta,\epsilon}(x) d\psi^{\theta,\epsilon}(x)| \\ \leq \int_0^1 \int_0^1 |(\sigma^n(x,y))^2 - \sigma^2(x,y)| d\psi^{\theta,\epsilon}(x) d\psi^{\theta,\epsilon}(x)|$$

where we recall that $(x,y) \mapsto \sigma^n(x,y)$ is the discretized version of σ. There again, using the uniform continuity of σ, we have for every $\epsilon > 0$ the existence of n_0 such that for $n \geq n_0$, for all $x, y \in [0,1] |(\sigma^n(x,y))^2 - \sigma^2(x,y)| \leq \eta$. Combining these two inequalities we get the first point. Then let us show that :

$$D(Leb||\psi^{\theta,\epsilon}) \geq D(Leb||\psi^\theta).$$

Let $f_\epsilon(x) = \max\{0, \epsilon^{-1} - \epsilon^{-2}|x|\}$ and

$$g_\epsilon(x,y) = f_\epsilon(x-y) + f_\epsilon(y+x) + f_\epsilon(2-x+y).$$

We have that :

$$\frac{d\psi^{\theta,\epsilon}}{dx}(x) = \int_{[0,1]} g_\epsilon(x,y) d\psi^\theta(y).$$

Let us notice that $\int_{[0,1]} g_\epsilon(x,y)dy = \int_{[0,1]} g_\epsilon(y,x)dy = 1$. We have

$$
\begin{aligned}
D(Leb||\psi^{\theta,\epsilon}) &= \int_0^1 \log\left(\frac{d\psi^{\theta,\epsilon}}{dx}\right) dx \\
&= \int_0^1 \log\left(\int_0^1 g_\epsilon(x,y) d\psi^\theta(y)\right) dx \\
&\geq \int_0^1 \log\left(\int_0^1 g_\epsilon(x,y) \frac{d\psi^\theta(y)}{dx} dy\right) dx \\
&\geq \int_0^1 \int_0^1 g_\epsilon(x,y) \log\left(\frac{d\psi^\theta(y)}{dx}\right) dy dx \\
&\geq D(Leb||\psi^\theta)
\end{aligned}
$$

where we used the concavity of log. Finally, using again the concavity, we have for every $i \in [1, n]$

$$
\begin{aligned}
n\int_{(i-1)/n}^{i/n} \log\left(\frac{d\psi^{\theta,\epsilon}(x)}{dx}\right) dx &\leq \log\left(n\int_{(i-1)/n}^{i/n} \frac{d\psi^{\theta,\epsilon}(x)}{dx}\right) \\
&\leq \log\left(n\psi_i^{\epsilon,\theta}\right).
\end{aligned}
$$

Summing over i gives us the result.

□

Thererefore, using this lemma for $\epsilon > 0$, there is $\epsilon' > 0$ such that

$$\sup_{\theta \geq 0} \frac{\Psi(\theta, \sigma^n, \psi^{\theta,\epsilon',n}) - F(\theta)}{\theta^2} > -\epsilon$$

for n large enough and so :

$$\sup_{\theta \geq 0} \frac{\Psi(\theta, \sigma^n, \psi^{\theta,\epsilon',n}) - F^n(\theta)}{\theta^2} > -\epsilon.$$

Therefore, taking

$$E^n(\theta) := F^n(\theta) - \vec{\Psi}(\theta, \sigma^n, \psi^{\theta,\epsilon',n})$$

our result is proven. □

We can now introduce I_β^n and $\tilde{I}_\beta^n$ defined on $[r_\sigma, +\infty[$ the rate functions for the upper and lower bound of the piecewise constant approximations

$$
\begin{aligned}
I_\beta^n(x) &:= \sup_\theta \left[J(\mu_{\sigma^n}, \theta, x) - F^n(\theta)\right] \\
\tilde{I}_\beta^n(x) &:= \sup_\theta \left[J(\mu_{\sigma^n}, \theta, x) - F^n(\theta) + E^n(\theta)\right].
\end{aligned}
$$

We will use for each piecewise constant approximation the approximate lower bound of theorem 1.4.32. We will need the following result :

Lemme 1.4.53.

$$\lim_{n\to\infty} \mu^n = \mu$$

and if we denote $r^{(n)}$ the upper bound of the support of μ^n and $l^{(n)}$ its lower bound, we have :

$$\lim_{n\to\infty} r^{(n)} = r_\sigma$$

$$\lim_{n\to\infty} l^{(n)} = l_\sigma.$$

Démonstration. The first point is a consequence of theorem 1.4.11.

Let $\Delta_N^n := X_N - X_N^n$. We have $\Delta_N^n := (\Sigma_N - \Sigma_N^{(n)}) \odot W_N$. Using theorem 1.4.40 and the fact that

$$\max_N \max_{i,j} |(\Sigma_N - \Sigma_N^{(n)})_{i,j}| \underset{n\to\infty}{\longrightarrow} 0$$

we have that for every $\epsilon > 0$ there is n_0 such that for any $n \geq n_0$:

$$\mathbb{P}[||\Delta_N^n|| > \epsilon] \underset{N\to\infty}{\longrightarrow} 0$$

In particular if we denote $\lambda_{1,N} < ... < \lambda_{N,N}$ the eigenvalues of X_N and $\lambda_{1,N}^{(n)} < ... < \lambda_{N,N}^{(n)}$ these of $X_N^{(n)}$, on the event $\{||\Delta_N^n|| \leq \epsilon\}$ we have $\max_{i=1}^N |\lambda_{i,N} - \lambda_{i,N}^{(n)}| \leq \epsilon$.

So, we have $\liminf_n r^{(n)} \geq r_\sigma$. Let us show that $\limsup_n r^{(n)} \leq r_\sigma$.

Suppose that $r := \limsup_n r^{(n)} > r_\sigma$, then up to extraction in n there exists $\epsilon > 0$, n_0 such that $r^{(n)} > r_\sigma + \epsilon$. We can now choose n_1 such that :

$$\mathbb{P}[||\Delta_N^{n_1}|| \leq \epsilon/3] \underset{N\to\infty}{\longrightarrow} 1$$

On the event $\{||\Delta_N^{n_1}|| \leq \epsilon/3\}$, we have that $\max_{i=1}^N |\lambda_{i,N} - \lambda_i^{(n)}| \leq \epsilon/3$. On this event, we can say that

$$\#\{i|\lambda_{i,N} > r^{(n_1)} - 2\epsilon/3\} \geq \#\{i|\lambda_{i,N}^{(n_1)} > r^{(n_1)} - \epsilon/3\}$$

Let's now denote A_N, B_N, C_N the following events :

$$A_N := \{||\Delta_N^{n_1}|| \leq \epsilon/3\}$$

$$B_N := \{\frac{\#\{i|\lambda_{i,N}^{(n_1)} > r^{(n_1)} - \epsilon/3\}}{N} \geq \mu^{(n_1)}([r^{(n_1)} - \epsilon/4, +\infty])\}$$

$$C_N := \{\frac{\#\{i|\lambda_{i,N} > r^{(n_1)} - 2\epsilon/3\}}{N} \geq \mu^{(n_1)}([r^{(n_1)} - \epsilon/4, +\infty])\}.$$

We have $A_N \cap B_N \subset C_N$. Since $\mu_n^{(n_1)}$ converge in probability to $\mu^{(n_1)}$, $\mathbb{P}[B_N]$ converges to 1. Furhermore since μ_n converge in probability to μ_σ, $\mu^{(n_1)}([r^{(n_1)} - \epsilon/4, +\infty]) > 0$ and by hypothesis $r^{(n_1)} - 2\epsilon/3 > r_\sigma$, we have that $\mathbb{P}[C_N]$ converges to 0 which is absurd since $\mathbb{P}[C_N] \geq \mathbb{P}[A_N \cap B_N]$ and $\mathbb{P}[A_N]$ and $\mathbb{P}[B_N]$ converges to 1. We prove the third point identically. □

This result enables us to finally prove the complete version of 1.4.4. Indeed using the Theorem 1.4.33 and its corollary, we have that for every $\epsilon > 0$, $\mathbb{P}[\lambda_{\max}(X_N^{(\beta)}) \leq r_\sigma - \epsilon] = o(e^{-N})$. It suffices to show that for all $\epsilon > 0$, $\mathbb{P}[\lambda_{\max}(X_N^{(\beta)}) \geq r_\sigma + \epsilon] = o(N^{-2})$. In both the continuous and the piecewise constant case that does not satisfies Assumption 1.4.14, we can approximate σ by σ^n strictly positive. And so the results of Theorem 1.4.53 holds, that is for n large enough, we have $r_n \leq r_\sigma + \epsilon/2$. For n large enough, we have $\mathbb{P}[||\Delta_N^n|| \geq \epsilon/2] = o(\exp(-N))$. So we have :

$$\begin{aligned}\mathbb{P}[\lambda_{\max}(X_N^{(\beta)}) \geq r_\sigma + \epsilon] &\leq \mathbb{P}[||\Delta_N^n|| \geq \epsilon/2] + \mathbb{P}[\lambda_{\max}(X_N^{(n)}) \geq r_\sigma + \epsilon/2] \\ &= o(N^{-2})\end{aligned}$$

where we used 1.4.15 for X_N^n. And so 1.4.4 is proved.

Lemme 1.4.54. — *For every $x > r_\sigma$, the function $\theta \mapsto J(\mu^n, \theta, x)$ converges uniformly on all compact set of $\mathbb{R}^+$ towards $\theta \mapsto J(\mu, \theta, x)$.*

— *For every $x > r_\sigma$*

$$I^n(x) \to I(x)$$

.

$$\tilde{I}^n(x) \to I(x).$$

Démonstration. For the first point of the lemma, let's first prove that for every $x \geq r_\sigma$, $\theta \mapsto J(\mu^n, \theta, x)$ converges uniformly on every compact towards the function $\theta \mapsto J(\mu, \theta, x)$. Let $l < r$ be two reals. For μ a probability measure on $\mathbb{R}$ whose support is a subset of $]l, r[$, let Q_μ be the function defined on $\mathcal{D}_{r,\epsilon} = \{(\theta, u) \in \mathbb{R}^+ \times]r, +\infty[| \frac{2\theta}{\beta}(r-u) \leq 1-\epsilon\}$ by

$$Q_\mu(\theta, u) = \int_{y\in\mathbb{R}} \log\left(1 + \frac{2\theta}{\beta}(u-y)\right) d\mu(y)$$

Q_μ is continuous in (θ, u) and for $K \subset \mathcal{D}_{r,\epsilon}$ a compact we have that the function $\mu \to Q_{\mu|K}$ mapping μ to the restriction of Q_μ on K is continuous in μ for the weak topology and μ such that their support is a subset of $]l, r[$ when the arrival space is the set of functions on K endowed with the uniform norm (this is a consequence of Ascoli's theorem). Let $x > r_\sigma$ and r, l such that $l < l_\sigma < r_\sigma < r < x$. For n large enough the support of μ^n is in $]l, r[$. We have that the sequence of functions $\theta \mapsto v(\theta, \mu^n, x)$ converges to $\theta \mapsto v(\theta, \mu, x)$. Indeed if $\frac{2\theta}{\beta} > G_\mu(x)$, then since $\lim_{n\to\infty} G_{\mu^n}(x) = G_\mu(x)$, $\frac{2\theta}{\beta} > G_{\mu^n}(x)$ for n large enough and the result is immediate.

If $\frac{2\theta}{\beta} < G_\mu(x)$ then $\frac{2\theta}{\beta} < G_{\mu^n}(x)$ for n large enough. G_{μ^n} converge towards G_μ on $[r, +\infty[$, for $\epsilon > 0$, K_{μ^n} is defined on $]0, G_\mu(r) - \epsilon]$ for n large enough and K_{μ^n} converges toward K_μ and therefore $R_{\mu^n}(\frac{2\theta}{\beta})$ converges towards $R_\mu(\frac{2\theta}{\beta})$. For $\theta = G_\mu(x)$ we use that $v(\theta, \mu, x) = R_{\mu^n}(\frac{2\theta}{\beta})$ if $\frac{2\theta}{\beta} \leq G_{\mu^n}(x)$ and $v(\theta, \mu, x) = (x - \frac{\beta}{2\theta})$ if $\frac{2\theta}{\beta} \geq G_{\mu^n}(x)$ and that the limits in both cases are $v(\theta, \lambda, x)$.

Then we have that for $2\theta/\beta \leq G_\mu(x)$

$$\begin{aligned}
\frac{2\theta}{\beta}(r - v(\theta, \mu, x)) &= \frac{2\theta}{\beta}(r - R_\mu(2\theta/\beta)) \\
&= \frac{2\theta}{\beta}(r - K_\mu(\beta/2\theta) + \beta/2\theta) \\
&\leq 1 - \frac{2\theta}{\beta}(r - K_\mu(\beta/2\theta)).
\end{aligned}$$

Writing $\frac{2\theta}{\beta} = G_\mu(y)$ with $y > x$ we have

$$\frac{2\theta}{\beta}(r - v(\theta, \mu, x)) \leq 1 - G_\mu(y)(y-r) \leq 1 - G_\mu(x)(x-r)$$

where we used that $y \mapsto G_\mu(y)(y-r)$ is increasing.

For $2\theta/\beta \geq G_\mu(x)$

$$\begin{aligned}
\frac{2\theta}{\beta}(r - v(\theta, \mu, x)) &= \frac{2\theta}{\beta}(r - x) + 1 \\
&\leq 1 - G_\mu(x)(x-r).
\end{aligned}$$

Taking $\epsilon > G_\mu(x)(x-r)$ and using the continuity in μ of $\mu \mapsto G_\mu(x)(x-r)$, we have for n large enough that

$$\sup_{\theta\in K'} \frac{2\theta}{\beta}(r - v(\theta, \mu^n, x)) \leq 1 - \epsilon.$$

Therefore, using the convergence of $v(\theta, \mu^n, x)$ and the uniform convergence of Q_{μ^n}, we have that $J(\mu^n, \theta, x)$ converges towards $J(\mu_\sigma, \theta, x)$. Furthermore, since $\theta \mapsto J(\mu_\sigma, \theta, x)$ are continuous increasing functions, by Dini's theorem the convergence is uniform on all compact.

We now prove the convergence of I^n towards I. Let us prove that there is $A > 0$ and $n_0 \in \mathbb{N}$ such that for $n \geq n_0$ and $\theta > 0$

$$F^n(\theta) - E^n(\theta) \geq A\theta^2.$$

We have

$$\begin{aligned} F^n(\theta) &\geq \frac{\theta^2}{\beta} P(\sigma^n, Leb) \\ &\geq \frac{\theta^2}{\beta} \int_0^1 \int_0^1 \sigma^2(x,y) dx dy \end{aligned}$$

and $-E^n(\theta) \geq \epsilon\theta^2$ for n large enough. Choosing $\epsilon < \int_0^1 \int_0^1 \sigma^2(x,y)dxdy$ we have our result.

Then given that $J(\mu^n, \theta, x) \leq \theta \max(r_n, x)$ we have that for any $r > r_\sigma, x > r_\sigma, \theta > 0$ for n large enough :

$$J(\mu^n, \theta, x) - F^n(\theta) + E^n(\theta) \leq r\theta x - A\theta^2$$

Since $\lim_{\theta\to\infty} r\theta x - A\theta^2 = -\infty$. and that $\theta \mapsto J(\mu^n, \theta, x) - F^n(\theta) + E^n(\theta)$ converges toward $\theta \mapsto J(\mu_\sigma, \theta, x) - F(\theta)$ on every compact of $\mathbb{R}^+$, we deduce that for every $x > r_\sigma$:

$$\lim_{n\to\infty} \tilde{I}^n_\beta(x) = I_\beta(x)$$

and in the same way with $E^n = 0$:

$$\lim_{n\to\infty} I^n_\beta(x) = I_\beta(x).$$

□

We will now prove that the difference between X_N^n and is negligible at the exponential scale.

Lemme 1.4.55. *For every $\epsilon > 0$ and every $A > 0$, there exists some $n_0 \in \mathbb{N}$ such that :*

$$\limsup_N \frac{1}{N} \log \mathbb{P}[||X_N^{n_0} - X_N|| \geq \epsilon] \leq -A$$

Démonstration. We can write that

$$X_N^n - X_N = \Delta_N^n \odot W_N^{(\beta)}$$

where

$$\Delta_N^n = \Sigma_N^n - \Sigma_N$$

Let

$$M_n := \sup_{i,j} |(\Delta_N^n)_{i,j}|$$

We have that :

$$\lim_{n\to\infty} M_n = 0$$

Following lemma 1.4.40, we can write that for every $n \in \mathbb{N}$, $A > 0$ there is $B > 0$ such that

$$\limsup_N \frac{1}{N} \log \mathbb{P}[(M_n)^{-1}||X_N^{n_0} - X_N|| > B] \leq -A.$$

For $n_0 \in \mathbb{N}$ such that $M_n B \leq \epsilon$ for all $n \geq n_0$, our upper bound is verified.

□

Let us now prove that we have a local large deviation principle for $\lambda_{\max}(X_N)$ with rate function I for $x > r_\sigma$ that is for every $\epsilon > 0$, there is $\delta > 0$.

$$-I(x) - \epsilon \leq \liminf_N \frac{1}{N} \log \mathbb{P}[\lambda_{\max}(X_N) \in [x-\delta, x+\delta]] \\ \leq \liminf_N \frac{1}{N} \log \mathbb{P}[\lambda_{\max}(X_N) \in [x-\delta, x+\delta]] \leq -I(x) + \epsilon.$$

Let us begin by choosing n_0 such that $|I^n(x) - I(x)| \leq \epsilon$ and $|\tilde{I}^n(x) - I(x)| \leq \epsilon$ for all $n \geq n_0$. By the large deviation upper and lower bounds in the piecewise constant case (that is theorem 1.4.32), we have the existence of $\delta > 0$ such that :

$$-\tilde{I}^n(x)-\epsilon \leq \liminf_N \frac{1}{N} \log \mathbb{P}[\lambda_{\max}(X_N^n) \in [x-\delta, x+\delta]] \leq \limsup_N \frac{1}{N} \log \mathbb{P}[\lambda_{\max}(X_N^n) \in [x-\delta, x+\delta]] \leq -I^n(x)+\epsilon.$$

Since :

$$\{\lambda_{\max}(X_N) \in [x-2\delta, x+2\delta]\} \supset \{\lambda_{\max}(X_N^n) \in [x-\delta, x+\delta]\} \cap \{||X_N^n - X_N|| < \delta\}$$

and

$$\{\lambda_{\max}(X_N) \in [x-\delta, x+\delta]\} \subset \{\lambda_{\max}(X_N^n) \in [x-2\delta, x+2\delta]\} \cup \{||X_N^n - X_N|| > \delta\}$$

We have

$$-I(x)-2\epsilon \leq \liminf_N \frac{1}{N} \log \mathbb{P}[\lambda_{\max}(X_N^n) \in [x-\delta, x+\delta]] \leq \liminf_N \frac{1}{N} \log \mathbb{P}[\lambda_{\max}(X_N) \in [x-2\delta, x+2\delta]]$$

and for n large enough such that $\limsup_N \frac{1}{N} \log \mathbb{P}[||X_N^n - X_N|| > \delta] < -I(x) - 2\epsilon$:

$$\limsup_N \frac{1}{N} \log \mathbb{P}[\lambda_{\max}(X_N) \in [x-\delta/2, x+\delta/2]] \leq \\ \limsup_N \frac{1}{N} \log \mathbb{P}[\lambda_{\max}(X_N^n) \in [x-\delta, x+\delta]] \leq -I(x) + 2\epsilon \quad (1.43)$$

Taking ϵ to 0, we have :

$$\lim_{\delta \to 0} \liminf_N \frac{1}{N} \log \mathbb{P}[\lambda_{\max}(X_N) \in [x-\delta, x+\delta]] = \lim_{\delta \to 0} \limsup_N \frac{1}{N} \log \mathbb{P}[\lambda_{\max}(X_N) \in [x-\delta, x+\delta]] = I(x).$$

Furthermore since exponential tightness is proved the same way as in the piecewise constant case, our result proves the large deviation principle for continuous variance profiles.

It only remains to relax Assumption 1.4.14 for the piecewise constant case. Let σ be a piecewise variance profile. We can approximate σ by $\sigma^n := \sqrt{\sigma^2 + \frac{1}{n+1}}$. We notice then that

$$\vec{\Psi}(\theta, \sigma^n, \vec{\psi}) = \frac{\theta^2}{(n+1)\beta} + \vec{\Psi}(\theta, \sigma, \vec{\psi})$$

so that if 1.4.21 is verified for σ, it is verified for σ^n. And so, as we have just done for the continuous case, we can prove the same way that the rate functions $I(\sigma^n, .)$ converges to $I(\sigma, .)$ and that the large deviation principle holds with $I(\sigma, .)$.

1.4.9 The case of matrices with 2×2 block variance profiles

In this section, we will discuss the case of piecewise constant variance profiles with 4 blocks (which are not necessarily of equal sizes) and determine what are the cases where the Assumption 1.4.21 holds. In particular, we will provide examples where the maximum argument of Assumption 1.4.21 can be taken continuous without the need for the concavity assumption.

Let's take a piecewise constant variance profile defined by $\vec{\alpha} = (\alpha, 1-\alpha)$ and $\sigma_{1,1} = a, \sigma_{2,2} = b, \sigma_{1,2} = \sigma_{2,1} = c$. In order to apply Theorem 1.4.5 we need to study the maximum arguments of :

$$\psi(x,\theta) = \vec{\Psi}(\sigma, \theta, (x, (1-x))) = \frac{\theta^2}{\beta}[a^2x^2 + b^2(1-x)^2 + 2c^2x(1-x)] + \frac{\beta}{2}[\alpha \log x + (1-\alpha)\log(1-x)].$$

Since we can change α in $1-\alpha$ by switching a and b, we can suppose without loss of generality that $\alpha \leq 1/2$.

We have

$$\partial_x \psi(x,\theta) := \frac{2\theta^2}{\beta}[a^2x - b^2(1-x) + c^2(1-2x)] + \frac{\beta}{2}\left(\frac{\alpha}{x} - \frac{1-\alpha}{1-x}\right).$$

Let

$$x_{\min} := \frac{c^2 - b^2}{2c^2 - a^2 - b^2}$$

$$x_\theta = \operatorname{argmax}_{x \in [0,1]} \psi(x,\theta)$$

and

$$\Phi(x,\theta) := x(1-x)\partial_x\psi(x,\theta) = (a^2 + b^2 - 2c^2)\frac{2\theta^2}{\beta}x(x - x_{min})(1-x) + \frac{\beta}{2}(\alpha - x).$$

1.4.9.1 Case with $(a^2 + b^2 - 2c^2) \leq 0$

In the case $(a^2 + b^2 - 2c^2) \leq 0$ we have the $\psi(.,\theta)$ is strictly concave and therefore $\theta \mapsto x_\theta$ is analytical and assumption 1.4.21 is verified.

From now on, we assume $(a^2 + b^2 - 2c^2) \geq 0$.

1.4.9.2 Case $x_{\min} \geq 1/2$

We look for the zeros of $\Phi(.,\theta)$ in $[0,1]$. To this end, we look for the intersection points of the curve of equation $y = x(1-x)(x - x_{\min})$ and the line $y = A_\theta(x - \alpha)$ where $A_\theta = \frac{\beta^2}{4\theta^2(a^2+b^2-2c^2)}$.

We notice that there is a critical value θ_{crit} such that for $\theta \leq \theta_{crit}$, there is only one critical point x_1^θ which is on $]0,1/2[$. For $\theta > \theta_{crit}$ we have the apparition of two other critical points x_2^θ and x_3^θ that are such that $1/2 < x_2^\theta < x_3^\theta$ with $\psi(x_2^\theta, \theta)$ being a local minimum and $\psi(x_3^\theta, \theta)$ a local maximum. For $x \in]0,1[$, we have :

$$\psi(x,\theta) - \psi(1-x,\theta) = \frac{\beta}{2}(1-2\alpha)(\log(1-x) - \log x) + \frac{\theta^2}{\beta}(a^2 + b^2 - 2c^2)[(x - x_{min})^2 - (1 - x - x_{min})^2].$$

For $x < 1/2$, we have $\psi(x,\theta) > \psi(1-x,\theta)$ and so the absolute maximum of $\psi(.,\theta)$ is attained for $x = x_1^\theta$. Since $\theta \mapsto x_1^\theta$ is analytical, assumption 1.4.21 is verified.

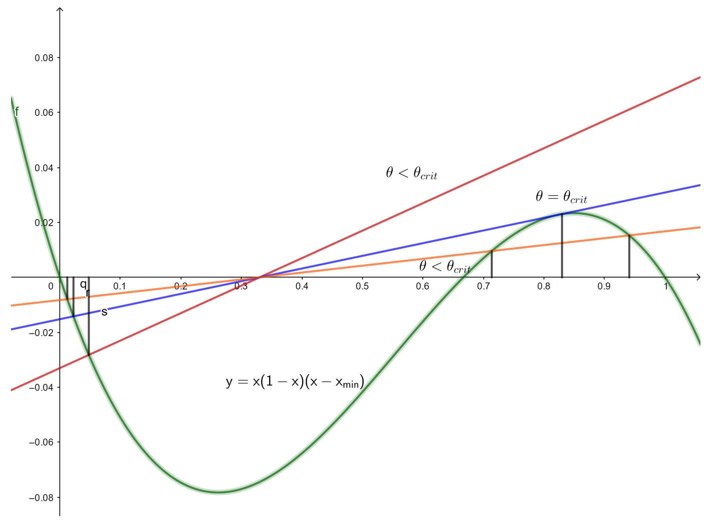

Figure 1.1 – Case $x_{\min} \geq 1/2$

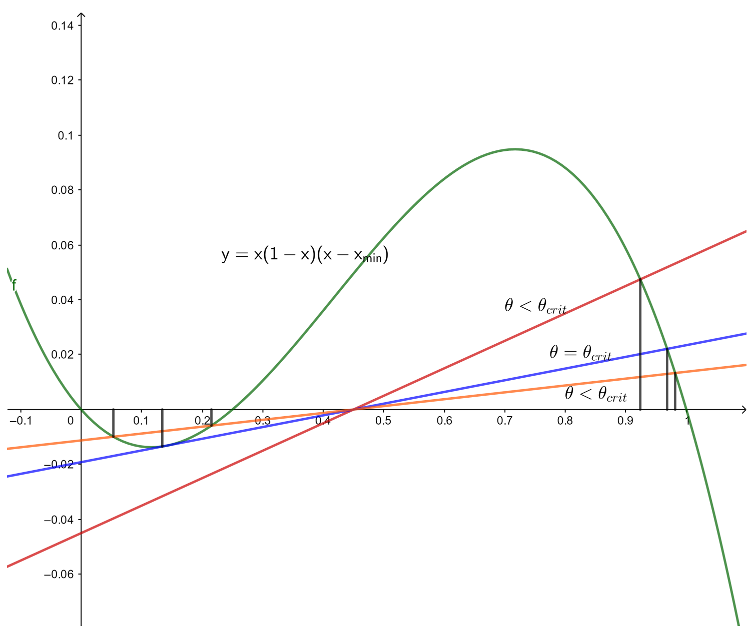

Figure 1.2 – Case $x_{min} \leq \alpha$

1.4.9.3 Case $x_{min} \leq \alpha$

There is again a critical value θ_{crit} such that for $\theta \leq \theta_{crit}$, there is only one critical point x_1^θ which is on $]\alpha, 1[$ and for $\theta > \theta_{crit}$ we have the apparition of two other critical points x_2^θ and x_3^θ such that $x_3^\theta < x_2^\theta \leq \alpha$. We have furthermore :

$$\psi(x,\theta) - \psi(2\alpha - x, \theta) = \frac{\beta}{2}[\alpha \log x - (1-\alpha)\log(1 - 2\alpha + x)] + \frac{\beta}{2}[(1-\alpha)\log(1-x) - \alpha\log(2\alpha - x)] \\ + \frac{\theta^2}{2}(a^2 + b^2 - 2c^2)[(x - x_{min})^2 - (2\alpha - x - x_{min})^2].$$

For $x < \alpha$, $\psi(x,\theta) < \psi(2\alpha - x, \theta)$ and there the absolute maximum of $\psi(.,\theta)$ is attained on $]\alpha, 1[$, so for x_1^θ. Since $\theta \mapsto x_1^\theta$ is analytical, Assumption 1.4.21 is verified.

1.4.9.4 Case $\alpha = 1/2$ and $x_{\min} = 1/2$

Then $x \mapsto \psi(x,\theta)$ is symmetrical in 1/2. Looking at the zeros of $\Phi(.,\theta)$ we have that if we set $\theta_{crit} := \beta\sqrt{2/(a^2+b^2-2c^2)}$ for $\theta \leq \theta_{crit}$ there is only one zero a $x = 1/2$ and for $\theta \geq \theta_{crit}$ there is three zeros in $x = 1/2$ and $x = \frac{1}{2} \pm \delta(\theta)$ where $\delta(\theta) = \frac{\beta}{2\theta}\sqrt{\frac{2}{(a^2+b^2-2c^2)}}$. Furthermore, for $\theta \leq \theta_{crit}$, $\psi(.,\theta)$ has its maximum in $x = 1/2$ and for $\theta \geq \theta_{crit}$, $\psi(.,\theta)$ has its maximum at both points $x = 1/2 \pm \delta(\theta)$ where $\delta(\theta) = \frac{1}{2}\sqrt{1 - \frac{2\beta^2}{\theta^2(a^2+b^2-2c^2)}}$. Therefore the function $\theta \mapsto 1/2$ for $\theta \leq \theta_{crit}$ and $\theta \mapsto 1/2 + \delta(\theta)$ for $\theta \geq \theta_{crit}$ is a continuous determination of the maximal argument of $\psi(.,\theta)$ and so Assumption 1.4.21 is verified and the large deviation principle holds. This gives an example where the maximum argument in Assumption 1.4.21 is neither unique nor $\mathcal{C}^1$ but where we can still derive a large deviation principle.

1.4.9.5 Case $x_{min} \in]\alpha, 1/2[$ and pathological case

The graph we obtain is similar to the graph of the first case. In this case, we also have a θ_{crit} such that for $\theta \leq \theta_{crit}$, there is only one critical point x_1^θ which is in $]0, \alpha[$ and then the apparition of two other critical points x_2^θ and x_3^θ that are such that $1/2 < x_2^\theta < x_3^\theta$, $\psi(x_2^\theta, \theta)$ being a local minimum and $\psi(x_3^\theta, \theta)$ a local maximum. But in this case for high values of θ, we have that the maximum is attained near 1 so x_3^θ is the maximum argument. We have a discontinuity in the maximum argument and so Assumption 1.4.21 is not verified.

Let us now show that if $x_{min} \in]\alpha, 1/2[$ and $c = 0$, the largest eigenvalue satisfies a large deviation principle but with a rate function J different from I.

Our matrix $X_N^{(\beta)}$ then looks like :

$$\begin{pmatrix} \sqrt{a\alpha}T_{\alpha N}^{(\beta)} & 0 \\ 0 & \sqrt{b(1-\alpha)}U_{(1-\alpha)N}^{(\beta)} \end{pmatrix}$$

where $(T_N^{(\beta)})_N$ and $(U_N^{(\beta)})_N$ are both Wigner matrices. We have :

$$\lambda_{\max}(X_N^{(\beta)}) = \max\{\sqrt{a\alpha}\lambda_{\max}(T_N^{(\beta)}), \sqrt{b(1-\alpha)}\lambda_{\max}(U_N^{(\beta)})\}.$$

But both these quantities satisfies large deviation principles, more precisely, if I_β is the rate function for the LDP for a Wigner matrix, $\lambda_{\max}(\sqrt{a\alpha}T_{\alpha N}^{(\beta)})$ follows a LDP with rate $\alpha I_\beta(\frac{x}{\sqrt{a\alpha}})$ and $\lambda_{\max}(\sqrt{b(1-\alpha)}U_{(1-\alpha)N}^{(\beta)})$ follows a LDP with rate $(1-\alpha)I_\beta(\frac{x}{\sqrt{b\alpha}})$. Now $\lambda_{\max}(X_N^{(\beta)})$ follows a LDP with rate J^β which is :

$$J^\beta(x) := \min\{\alpha I_\beta(\frac{x}{\sqrt{a\alpha}}), (1-\alpha)I_\beta(\frac{x}{\sqrt{b\alpha}})\}$$

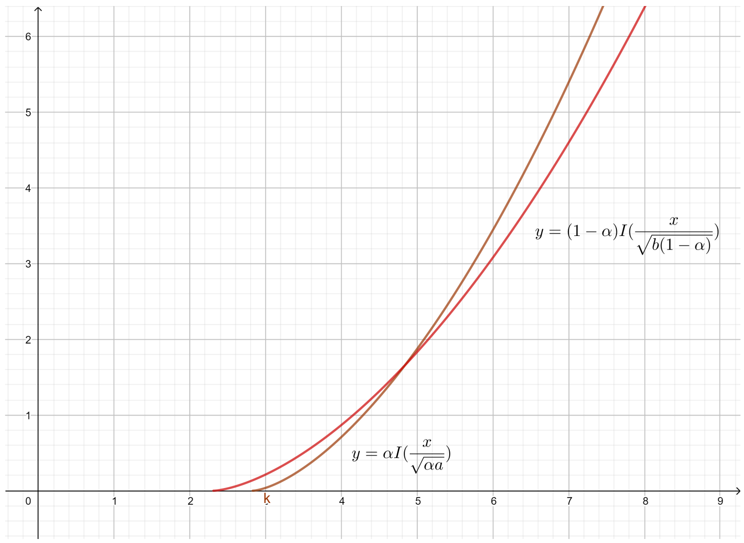

FIGURE 1.3 – Case with $\alpha = 2/3$, $a = 3$, $b = 4$

In particular, if $x_{\min} \in]\alpha, 1/2[$, we have $a\alpha > b(1-\alpha)$ and $b > a$, we notice that $J^\beta(x) = \alpha I_\beta(\frac{x}{\sqrt{a\alpha}})$ for x near $a\alpha$ and $J^\beta(x) = (1-\alpha)I_\beta(\frac{x}{\sqrt{b\alpha}})$ for large x. In this case one can notice that J^β is not a convex function and therefore cannot by obtained as $\sup_\theta F(\theta) - J(x, \mu_\sigma, \theta)$ since it is a sup of convex functions. We have $J \neq I$. For $c > 0$ but small enough we can also conclude that the large deviation principle still does not hold. Indeed, if we denote I_c the rate function we expect and $x_0 \in \mathbb{R}^+$ such that $J(x_0) \leq I_0(x_0) - \eta$ for some $\eta > 0$ (x_0 does exists using the discussion above). Using the same approximation arguments as in section 1.4.8, we have that there exists $\epsilon > 0$ such that for $c < \epsilon$, $I_c(x_0) > I_0(x_0) - \eta/3$ and :

$$\lim_{\delta \to 0} \limsup_N \frac{1}{N} \log \mathbb{P}[\lambda_{\max}(X_N) \in [x_0 - \delta, x_0 + \delta]] \leq -J(x_0) + \eta/3 < -I_c(x_0)$$

Since I_c is continuous in x_0, we have that there cannot be a large deviation principle withe the rate function I_c.

1.4.10 Appendix : Proof of Lemma 1.4.33

This section is devoted to the proof of Lemma 1.4.33. For this, we will use two concentration results respectively from [49] and [9].

Théorème 1.4.56. *By [49, (Theorem 1.4)] (for the compact case) and [49, Corollary 1.4 b)] (for the logarithmic Sobolev case), we have for $\beta = 1, 2$, and for N large enough*

$$\limsup_{N \to \infty} \frac{1}{N^{7/6}} \log \mathbb{P}[d(\mu_{X_N^{(\beta)}}, \mathbb{E}[\mu_{X_N^{(\beta)}}]) > N^{-1/6}] < 0$$

where d is the Dudley distance.

We therefore only need to show that

Théorème 1.4.57. *If we let for every N :*

$$F_{X_N^{(\beta)}}(x) = \mu_{X_N^{(1)}}(]-\infty, x])$$
$$F_{\mu_\sigma}(x) = \mu_\sigma(]-\infty, x])$$

we have that

$$\sup_{x\in\mathbb{R}} |F_{\mu_\sigma}(x) - \mathbb{E}[F_{X_N^{(1)}}(x)]| = O(N^{-1/10})\,.$$

Démonstration. Following the theorem 1.4.15 and the fact that $N(D)$ does not depend on τ we have that for $N \geq N(D)$ and $x \in \mathbb{R}$,

$$\mathbb{P}[|F_{\mu_\sigma}(x) - F_{X_N^{(1)}}(x)| \geq N^{-3/4}] \leq N^{-D}$$

Finally for $x \in \mathbb{R}$ and $N \geq N(3/4)$

$$\begin{aligned}|F_{\mu_\sigma}(x) - \mathbb{E}[F_{X_N^{(1)}}(x)]| \leq \mathbb{E}[|F_{X_N^{(1)}}(x) - F_{\mu_\sigma}(x)|] \qquad &\leq N^{-3/4} + 2\mathbb{P}[|F_{\mu_\sigma}(x) - F_{X_N^{(1)}}(x)| \geq N^{-3/4}] \\ \leq 3N^{-3/4}\end{aligned}$$

and so for $N \geq N(3/4)$

$$|F_{\mu_\sigma}(x) - \mathbb{E}[F_{X_N^{(1)}}(x)]| \leq 3N^{-3/4}$$

□

In order to conclude, we need only to use Lemma 1.4.29 to see that $F_{X_N^{(1)}}(-M)$ and $1 - F_{X_N^{(1)}}(M)$ decay in expectation exponentially fast in N for some fixed M so that

$$d(\mathbb{E}[\mu_{X_N^{(1)}}], \mu_\sigma) \leq 4e^{-N} + 2M \sup_{x\in\mathbb{R}} |F(x) - \mathbb{E}[F_{X_N^{(1)}}(x)]| = o(N^{-1/6})\,.$$

1.5 Large deviations for the largest eigenvalue of sub-Gaussian matrices

1.5.1 Introduction

Understanding the large deviation behavior of the largest eigenvalue of a random matrix is a challenging question, with many applications in statistics and mobile communications systems, e.g [40, 20]. However, it is in general a difficult question and very few results are known. It is known since [77] that the empirical distribution of the eigenvalues of a Wigner matrix converges to the semi-circle law provided the off-diagonal entries have a finite second moment. Following the pioneering work of Kómlos and Fűredi [41], we know by [8] that assuming the Wigner matrix has centered entries, the largest eigenvalue converges to the right edge of the support of the semi-circle law if and only if the fourth moment of the off-diagonal entries is finite. Given these two results, one can wonder what is the probability that the empirical measure or the largest eigenvalue have an unexpected behavior. Large deviation principles were derived for the empirical distribution of the eigenvalues and the largest eigenvalue of classical Gaussian ensembles, as the Gaussian Unitary Ensemble (GUE) and Gaussian Orthogonal Ensemble (GOE) in [17] and [16]. Indeed, in this case, the joint law of the eigenvalues is explicit and large deviations estimates can be derived by Laplace's method, up to taking care of the singularity of the interaction. In a breakthrough paper, Bordenave and Caputo [23] showed that large deviations for the empirical measure of the eigenvalues can be estimated when the tails of the entries are heavier than in the Gaussian case. These large deviations have a smaller speed than in the cases of classical Gaussian Ensembles and are due to a relatively small number of entries of order one. This phenomenon was shown to hold as well for the largest eigenvalue by one of the authors [5].

Yet, the case of sub-Gaussian entries remained still mysterious. Last year, two of the authors showed that if the Laplace transform of the Wigner matrix is pointwise bounded from above by the one of the GUE or GOE, then a large deviations principle holds with the same rate function as in the Gaussian case. This special case of Wigner matrices, which was called *with sharp sub-Gaussian tails*, was shown to include matrices with Rademacher variables and variables uniformly sampled in an interval. Yet, many Wigner matrices with sub-Gaussian entries are not with sharp sub-Gaussian tails, as for Gaussian sparse matrices which are obtained by multiplying entrywise a GOE matrix with an independent Bernoulli random variable. In this article, we investigate this general setting and derive large deviations estimates for the largest eigenvalue of Wigner matrices with sub-Gaussian entries. In particular, we show that the rate function of this large deviations principle is different from the one of the GOE.

We will consider hereafter a $N \times N$ symmetric random matrix X_N with independent entries $(X_{ij})_{i \le j}$ above the diagonal so that $\sqrt{N} X_{ij}$ has law μ for all $i < j$ and $\sqrt{N/2} X_{ii}$ has law μ for all i. In particular, the variance profile is the same as the one of the GOE. We asume that μ is centered and has a variance equal to 1. Let

$$\forall x \in \mathbb{R},\ \psi(x) = \frac{1}{x^2} \log \int e^{xt} d\mu(t),$$

and $\psi(0) = 0$, which defines a continuous function on $\mathbb{R}$. Assume that

$$\frac{A}{2} := \sup_{x \in \mathbb{R}} \psi(x) < +\infty. \tag{1.44}$$

The case where $A = 1$ is the case of *sharp sub-Gaussian tails* which was solved in [45]. We investigate here the case where $A > 1$ and we show the following result.

Theorem 1.5.1. *Denote by λ_{X_N} the largest eigenvalue of X_N. Under some technical assumptions, there exist a good rate function $I_\mu : \mathbb{R} \to [0, +\infty]$ and a set $\mathcal{O}_\mu \subset \mathbb{R}$ such that $(-\infty, 2] \cup [x_\mu, +\infty) \subset \mathcal{O}_\mu$ for some $x_\mu \ge 2$ and such that for any $x \in \mathcal{O}_\mu$,*

$$\lim_{\delta \to 0} \liminf_{N \to +\infty} \frac{1}{N} \log \mathbb{P}\left(|\lambda_{X_N} - x| \le \delta\right) = \lim_{\delta \to 0} \limsup_{N \to +\infty} \frac{1}{N} \log \mathbb{P}\left(|\lambda_{X_N} - x| \le \delta\right) = -I_\mu(x).$$

The rate function I_μ is infinite on $(-\infty, 2)$ and satisfies

$$I_\mu(x) \underset{+\infty}{\sim} \frac{1}{4A} x^2.$$

If $A \in (1,2)$, then $[2, \sqrt{A-1} + 1/\sqrt{A-1}] \subset \mathcal{O}_\mu$ and I_μ coincides on this interval to the rate function of the GOE, that is,

$$I_\mu(x) = \frac{1}{2} \int_2^x \sqrt{y^2 - 4}\, dy = I_{GOE}(x). \tag{1.45}$$

The technical assumptions include the case where ψ is increasing (which holds in the case of sparse Gaussian entries) and the case where the maximum of ψ is achieved on $\mathbb{R}$. In the later case, $I_\mu(x)$ only depends on A for x large enough.

1.5.1.1 Assumptions

We now describe more precisely our assumptions.

Assumption 1.5.2. *Let $\mu \in \mathcal{P}(\mathbb{R})$ be a symmetric probability measure with unit variance. We denote by L its log-Laplace transform,*

$$\forall x \in \mathbb{R},\ L(x) = \log \int e^{xt} d\mu(t),$$

and $\psi(x) = L(x)/x^2$. We assume that μ is sub-Gaussian in the sense that

$$\frac{A}{2} := \sup_{x \in \mathbb{R}} \psi(x) < +\infty,$$

and we define $B \geq 0$ by,

$$\frac{B}{2} := \lim_{|x|\to+\infty} \psi(x).$$

We assume moreover that $L(\sqrt{.})$ is a Lipschitz function and that μ does not have sharp sub-Gaussian tails, meaning that $A > 1$.

We describe below a few examples of probability measures μ which satisfy the above assumptions. In each of these cases, the fact that $L(\sqrt{.})$ is Lipschitz is clear and left to the reader.

Example 1.5.3. — *(Combination of Gaussian and Rademacher laws). Let*

$$\mu(dx) = a\frac{e^{-\frac{1}{2B}x^2}}{\sqrt{2\pi B}}dx + (1-a)\frac{1}{2}(\delta_{-b} + \delta_{+b})$$

where a, b, B are non negative real numbers such that $a \in (0,1)$ and $aB + (1-a)b^2 = 1$. Then, for all $x \in \mathbb{R}$,

$$L_\mu(x) = \log\left(ae^{\frac{B}{2}x^2} + (1-a)\cosh(bx)\right).$$

If $B > 1$ and $b \in (0,1)$ we see that our conditions are fulfilled and $A = B$.

— *(Sparse Gaussian case). Let μ be the law of $\zeta\Gamma$ with ζ a Bernoulli variable of parameter $p \in (0,1)$ and Γ a centered Gaussian variable with variance $1/p$. For any $x \in \mathbb{R}$,*

$$L_\mu(x) = \log\left(pe^{\frac{x^2}{2p}} + 1 - p\right)$$

so that $A = B = \frac{1}{p}$.

— *(Combination of Rademacher laws). Let*

$$\mu = \sum_{i=1}^{p} \frac{\alpha_i}{2}(\delta_{\beta_i} + \delta_{-\beta_i})$$

with $\alpha_i \geq 0$, $\beta_i \in \mathbb{R}$ and $p \in \mathbb{N}$ so that $\sum \alpha_i = 1, \sum \alpha_i\beta_i^2 = 1$. Since μ is compactly supported $B = 0$. The fact that μ does not have sharp sub-Gaussian tails means that there exist some t and $A > 1$ such that

$$\sum_{i=1}^{p} \alpha_i \cosh(\beta_i t) \geq e^{A\frac{t^2}{2}}.$$

The latter is equivalent to

$$\sum_{i=1}^{p} \frac{\alpha_i}{2} e^{\frac{\beta_i^2}{2A}}\left(e^{-\frac{A}{2}(t-\frac{\beta_i}{A})^2} + e^{-\frac{A}{2}(t+\frac{\beta_i}{A})^2}\right) \geq 1.$$

This inequality holds as soon as $\alpha_i e^{\frac{\beta_i^2}{2A}} \geq 2$ for some $i \in \{1,\dots,p\}$ by taking $t = \frac{\beta_i}{A}$. This can be fulfilled if β_i is large enough while $\alpha_i\beta_i^2 < 1$. We also see with this family of examples that A can be taken arbitrarily large (take e.g $p = 2$, $A = \beta_1, t = 1, \alpha_1 = (2\beta_1^2)^{-1}$, $e^{\beta_1/2} \geq 4\beta_1^2, \beta_2^2 = (2-\beta_1^{-2})^{-1}, \alpha_2 = 1-\alpha_1$).

Let $\mathcal{H}_N$ be the set of real symmetric matrices of size N. We denote for any $A \in \mathcal{H}_N$ by λ_A its largest eigenvalue, $||A||$ is spectral radius and by $\hat{\mu}_A$ the empirical distribution of its eigenvalues, that is

$$\hat{\mu}_A = \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i},$$

where $\lambda_1, \dots, \lambda_N$ are the eigenvalues of A. We make the following assumption of exponential tightness of the spectral radius and of concentration of the empirical distribution of the eigenvalues at the scale N.

Assumption 1.5.4. *The spectral radius of X_N, $||X_N||$, is exponentially tight at the scale N:*

$$\lim_{K\to+\infty}\limsup_{N\to+\infty}\frac{1}{N}\log\mathbb{P}\big(||X_N||>K\big)=-\infty. \tag{1.46}$$

Moreover, the empirical distribution of the eigenvalues $\hat{\mu}_{X_N}$ concentrates at the scale N:

$$\limsup_{N\to+\infty}\frac{1}{N}\log\mathbb{P}\left(d(\hat{\mu}_{X_N},\sigma)>N^{-\kappa}\right)=-\infty, \tag{1.47}$$

for some $\kappa>0$, where d is a distance compatible with the weak topology and σ is the semi-circle law, defined by

$$\sigma=\frac{1}{2\pi}\sqrt{4-x^2}\mathbb{1}_{|x|\le 2}dx.$$

Remark 1.5.5.
1. *From [45, Lemmas 1.8, 1.11], we know that Assumption 1.5.4 is fulfilled if μ is either compactly supported, or if μ satisfies a logarithmic Sobolev inequality in the sense that there exists $c>0$ so that for any smooth function $f:\mathbb{R}\to\mathbb{R}$, such that $\int f^2d\mu=1$,*

$$\int f^2\log f^2 d\mu\le c\int\|\nabla f\|_2^2 d\mu\,.$$

2. *If μ is a symmetric sub-Gaussian probability measure on $\mathbb{R}$ with log-concave tails in the sense that $t\mapsto\mu(|x|\ge t)$ is a log-concave function, then the Wigner matrix X_N satisfies Assumption 1.5.4. In particular, if B is a Wigner matrix with Bernoulli entries with parameter p and Γ is a GOE matrix, then the sparse Gaussian matrix $B\circ\Gamma/\sqrt{p}$, where $\circ$ the Hadamard product, satisfies Assumption 1.5.4. We refer the reader to section 1.5.7.1 of the appendix for more details.*

1.5.1.2 Statement of the results and scheme of the proof

As in [45], our approach to derive large deviations estimates is based on a tilting of the law of the Wigner matrix X_N by spherical integrals. Let us recall the definition of spherical integrals. For any $\theta\ge 0$, we define

$$I_N(X_N,\theta)=\mathbb{E}_e[e^{\theta N\langle e,X_Ne\rangle}]$$

where e is uniformly sampled on the sphere $\mathbb{S}^{N-1}$ with radius one. The asymptotics of

$$J_N(X_N,\theta)=\frac{1}{N}\log I_N(X_N,\theta)$$

were studied in [47] where the following result was proved.

Theorem 1.5.6. *[47, Theorem 6] Let $(E_N)_{N\in\mathbb{N}}$ be a sequence of $N\times N$ real symmetric matrices such that:*
- *The sequence of empirical measures $\hat{\mu}_{E_N}$ converges weakly to a compactly supported measure μ.*
- *There is a real number λ_E such that the sequence of largest eigenvalues λ_{E_N} converges to λ_E.*
- $\sup_N||E_N||<+\infty$.

For any $\theta\ge 0$,

$$\lim_{N\to+\infty}J_N(E_N,\theta)=J(\mu,\theta,\lambda_E)$$

The limit J is defined as follows. For a compactly supported probability measure $\mu\in\mathcal{P}(\mathbb{R})$ we define its Stieltjes transform G_μ by

$$\forall z\notin\operatorname{supp}(\mu),\ G_\mu(z):=\int_{\mathbb{R}}\frac{1}{z-t}d\mu(t),$$

where supp(μ) is the support of μ. For any compactly supported probability measure μ, we denote by r_μ the right edge of the support of μ. Then G_μ is a bijection from $(r_\mu, +\infty)$ to $\big(0, G_\mu(r_\mu)\big)$ where

$$G_\mu(r_\mu) = \lim_{t \downarrow r_\mu} G_\mu(t).$$

Let K_μ be the inverse of G_μ on $(0, G_\mu(r_\mu))$ and let

$$\forall z \in (0, G_\mu(r_\mu)),\ R_\mu(z) := K_\mu(z) - 1/z,$$

be the R-transform of μ as defined by Voiculescu in [72]. Then, the limit of spherical integrals is defined for any $\theta \geq 0$ and $x \geq r_\mu$ by,

$$J(\mu, \theta, x) := \theta v(\mu, \theta, x) - \frac{1}{2} \int \log\big(1 + 2\theta v(\mu, \theta, x) - 2\theta y\big) d\mu(y),$$

with

$$v(\mu, \theta, x) := \begin{cases} R_\mu(2\theta) & \text{if } 0 \leq 2\theta \leq G_\mu(x), \\ x - \frac{1}{2\theta} & \text{if } 2\theta > G_\mu(x). \end{cases}$$

In the case of the semi-circle law, we have

$$G_\sigma(x) = \frac{1}{2}(x - \sqrt{x^2 - 4}),\ R_\sigma(x) = x.$$

We denote by $J(x, \theta)$ as a short-hand for $J(\sigma, \theta, \lambda)$. In the next lemma we compute explicitly $J(x, \theta)$, whose proof is left to the reader.

Lemma 1.5.7. *Let $\theta \geq 0$ and $x \geq 2$. For $\theta \leq \frac{1}{2} G_\sigma(x)$,*

$$J(\theta, x) = \theta^2.$$

Whereas for $\theta \geq \frac{1}{2} G_\sigma(x)$,

$$J(x, \theta) = \theta x - \frac{1}{2} - \frac{1}{2} \log 2\theta - \frac{1}{2} \int \log(x - y) d\sigma(y).$$

Moreover, $J(x, .)$ is continuously differentiable.

To derive large deviations estimates using a tilt by spherical integrals, it is central to obtain the asymptotics of the annealed spherical integral $F_N(\theta)$ defined as,

$$F_N(\theta) = \frac{1}{N} \log \mathbb{E}_{X_N} \mathbb{E}_e [\exp(N\theta \langle e, X_N e \rangle)] .$$

In the following lemma, we obtain the limit of F_N as the solution of a certain variational problem. We denote by $\overline{F}(\theta)$ and $\underline{F}(\theta)$ its upper and lower limits:

$$\overline{F}(\theta) = \limsup_{N \to +\infty} F_N(\theta),$$

$$\underline{F}(\theta) = \liminf_{N \to +\infty} F_N(\theta).$$

For any measurable subset $I \subset \mathbb{R}$, we denote by $\mathcal{M}(I)$ and $\mathcal{P}(I)$ respectively the set of measures and the set of probability measures supported on I.

Proposition 1.5.8. *Assume X_N satisfies Assumptions 1.5.2 and 1.5.4.*

$$\overline{F}(\theta) = \limsup_{\substack{\delta\to 0, K\to+\infty \\ \delta K\to 0}} \sup_{\substack{\alpha_1+\alpha_2+\alpha_3=1 \\ \alpha_i\geq 0}} \limsup_{N\to+\infty} \mathcal{F}^N_{\alpha_1,\alpha_2,\alpha_3}(\delta, K),$$

$$\underline{F}(\theta) = \sup_{\substack{\alpha_1+\alpha_2+\alpha_3=1 \\ \alpha_i\geq 0}} \liminf_{\substack{\delta\to 0, K\to+\infty \\ \delta K\to 0}} \limsup_{N\to+\infty} \mathcal{F}^N_{\alpha_1,\alpha_2,\alpha_3}(\delta, K)\,.$$

$\mathcal{F}^N_{\alpha_1,\alpha_2,\alpha_3}(\delta, K)$ is the function given by:

$$\begin{aligned}
\mathcal{F}^N_{\alpha_1,\alpha_2,\alpha_3}(\delta, K) &= \theta^2\big(\alpha_1^2 + 2\alpha_1\alpha_2 + B\alpha_3^2\big) \\
&+ \sup_{\substack{\nu_2\in\mathcal{M}(I_2) \\ \int x^2 d\nu_2=\alpha_2}} \sup_{\substack{s_i\in I_3, i\leq k \\ |\sum_i s_i^2 - N\alpha_3|\leq \delta N}} \Big\{\sum_{i=1}^k \int L\Big(\frac{2\theta s_i x}{\sqrt{N}}\Big)d\nu_2(x) + \frac{1}{2}\int L\Big(\frac{2\theta xy}{\sqrt{N}}\Big)d\nu_2(x)d\nu_2(y) \\
&+ \sup_{\substack{\nu_1\in\mathcal{P}(I_1) \\ \int x^2 d\nu_1(x)=\alpha_1}} \Big\{\sum_{i=1}^k \int L\Big(\frac{2\theta s_i x}{\sqrt{N}}\Big)d\nu_1(x) - H(\nu_1)\Big\} - \frac{1}{2}\log(2\pi) - \frac{1}{2}\Big\},
\end{aligned}$$

where $I_1 = \{x : |x| \leq \delta N^{1/4}\}$, $I_2 = \{x : \delta N^{1/4} \leq |x| \leq KN^{1/4}\}$, $I_3 = \{KN^{1/4} \leq |x| \leq \sqrt{N\alpha_3}\}$, and

$$H(\nu) = \int \log\frac{d\nu}{dx}d\nu(x),$$

if ν is absolutely continuous with respect to the Lebesgue measure, whereas $H(\nu)$ is infinite otherwise.

Remark 1.5.9. *Note that $\underline{F}$ and $\overline{F}$ are convex by Hölder inequality. Since the entries of X_N are sub-Gaussian, $\overline{F}(\theta) \leq A\theta^2$. In particular $\overline{F}, \underline{F}$ are finite convex functions and therefore they are continuous on $\mathbb{R}_+$.*

The above proposition gives quite an intricate definition for the limit of the annealed spherical integrals. Yet, for small enough θ it can be computed explicitly.

Lemma 1.5.10. *For any $\theta \leq \frac{1}{2\sqrt{A-1}}$,*

$$\underline{F}(\theta) = \overline{F}(\theta) = \theta^2.$$

Note that for large θ this formula is not valid anymore when $A > 1$ since $\underline{F}$ grows like $A\theta^2$ at infinity (see the proof of Proposition 1.5.12).

Proof. Using the bound $L(x) \leq Ax^2/2$ for any $x \geq 0$ and the notation of Proposition 1.5.8, we have

$$\mathcal{F}^N_{\alpha_1,\alpha_2,\alpha_3}(\delta, K) \leq \sup_{\alpha_1+\alpha_2+\alpha_3=1} \Big\{\theta^2\big(\alpha_1^2 + 2\alpha_1\alpha_2 + B\alpha_3^2 + 2A\alpha_3\alpha_2 + A\alpha_2^2 + 2A\alpha_1\alpha_3\big) + \frac{1}{2}\log\alpha_1\Big\},$$

where we used the fact that the infimum

$$\inf\{H(\nu_1) : \int x^2 d\nu_1 = \alpha_1, \nu_1 \in \mathcal{P}(\mathbb{R})\},$$

is achieved for $\nu_1 = (2\pi\alpha_1)^{-1/2}e^{-\frac{x^2}{2\alpha_1}}dx$. As $A \geq 1$ and $B \leq A$, we have the upper bound,

$$\begin{aligned}
\overline{F}(\theta) &\leq \sup_{\alpha\in[0,1]} \Big\{\theta^2\big(\alpha^2 + 2A\alpha(1-\alpha) + A(1-\alpha)^2\big) + \frac{1}{2}\log\alpha\Big\} \\
&= \sup_{\alpha\in[0,1]} \Big\{\theta^2\big(A - (A-1)\alpha^2\big) + \frac{1}{2}\log\alpha\Big\}
\end{aligned}$$

We deduce that if $2\theta\sqrt{A-1} \leq 1$ then the function

$$\alpha \mapsto \theta^2\big(A-(A-1)\alpha^2\big)+\frac{1}{2}\log\alpha,$$

is increasing on $[0,1]$. Thus the supremum is achieved at $\alpha = 1$, and $\overline{F}(\theta) \leq \theta^2$. Moreover, taking $\alpha_1 = 1, \alpha_2 = \alpha_3 = 0$, and ν_1 the standard Gaussian law, we find that

$$\underline{F}(\theta) \geq \theta^2. \tag{1.48}$$

Thus, if $2\theta\sqrt{A-1} \leq 1$, we get that $\overline{F}(\theta) = \underline{F}(\theta) = \theta^2$.

□

Although the limit of the annealed spherical integrals may not be explicit for all θ, we can still use it to obtain large deviations upper bounds as we describe now in the following theorem.

Theorem 1.5.11. *Under the Assumptions 1.5.2 and 1.5.4, the law of the largest eigenvalue λ_{X_N} satisfies a large deviation upper bound with good rate function which is infinite on $(-\infty,2)$ and otherwise given by:*

$$\forall y \geq 2,\ \bar{I}(y) = \sup_{\theta\geq 0}\{J(y,\theta) - \overline{F}(\theta)\}\,. \tag{1.49}$$

Proof. From Assumption 1.5.4 we know that the law of the largest eigenvalue is exponentially tight at the scale N. Therefore, it is sufficient to prove a weak large deviations upper bound by [36, Lemma 1.2.18]. Let $\delta > 0$. We have,

$$\mathbb{P}(\lambda_{X_N} < 2-\delta) \leq \mathbb{P}(\hat{\mu}_{X_N}(f) = 0),$$

where f is a smooth compactly supported function with support in $(2-\delta, 2)$. Since $\mathrm{supp}(\sigma) = [-2,2]$, we deduce that,

$$\mathbb{P}(\lambda_{X_N} < 2-\delta) \leq \mathbb{P}(d(\hat{\mu}_{X_N},\sigma) > \epsilon),$$

for some $\epsilon > 0$. As the empirical distribution of the eigenvalues concentrates at the scale N according to (1.47), we get

$$\lim_{N\to+\infty}\frac{1}{N}\log\mathbb{P}(\lambda_{X_N} < 2-\delta) = -\infty.$$

Let now $x \geq 2$ and $\delta > 0$. Recall from (1.48) that $\underline{F}(\theta) \geq \theta^2$ for any $\theta \geq 0$. Therefore,

$$\bar{I}(x) \leq \sup_{\theta\geq 0}\{J(x,\theta) - \theta^2\}.$$

From [45, Section 4.1], we know that

$$\sup_{\theta\geq 0}\{J(x,\theta) - \theta^2\} = I_{GOE}(x),$$

where I_{GOE} is the rate function of the largest eigenvalue of a GOE matrix. In particular $\bar{I}(2) = 0$ since $I_{GOE}(2)$. Therefore we only need to estimate small ball probabilities around $x \neq 2$. As $\hat{\mu}_{X_N}$ concentrates at the scale N, and $||X_N||$ is exponentially tight at the scale N by Assumption 1.5.4 it is enough to show that for any $K > 0$,

$$\limsup_{\delta\to 0}\limsup_{N\to+\infty}\frac{1}{N}\log\mathbb{P}(X_N \in V_{\delta,x}^K) \leq -\bar{I}(x),$$

where $V_{\delta,x}^K = \{Y \in \mathcal{H}_N : |\lambda_Y - x| < \delta, d(\hat{\mu}_Y,\sigma) < N^{-\kappa}, ||Y|| \leq K\}$, for some $\kappa > 0$. Let $\theta \geq 0$. From [58, Proposition 2.1], we know that the spherical integral is continuous, more precisely, for N large enough and any $X_N \in V_{\delta,x}$,

$$|J_N(X_N,\theta) - J(x,\theta)| < g(\delta),$$

for some function $g(\delta)$ going to 0 as $\delta \to 0$. Therefore,

$$\mathbb{P}(X_N \in V_{\delta,x}^K) = \mathbb{E}\Big(\mathbb{1}_{X_N \in V_{\delta,x}^K} \frac{I_N(X_N,\theta)}{I_N(X_N,\theta)}\Big) \leq \mathbb{E}[I_N(X_N,\theta)]e^{-NJ(\theta,x)-Ng(\delta)}.$$

Taking the limsup as $N \to 0$ and $\delta \to 0$ at the logarithmic scale, we deduce

$$\limsup_{\delta\to 0}\limsup_{N\to+\infty}\frac{1}{N}\log\mathbb{P}(X_N \in V_{\delta,x}^K) \leq \overline{F}(\theta) - J(\theta,x).$$

Opimizing over $\theta \geq 0$, we get the claim. □

Proposition 1.5.12. *Under Assumption 1.5.2, the rate function $\bar{I}$ defined in Theorem 1.5.11 is lower semi-continuous, and growing at infinity like $x^2/4A$. In particular, $\bar{I}$ is a good rate function.*

Proof. $\bar{I}$ is lower semi-continuous as a supremum of continuous functions (recall here that $J(\theta,.)$ is continuous by Lemma 1.5.7 and $\overline{F}$ is continuous by Remark 1.5.7). It remains to show that its level sets are compact, for which it is sufficient to prove that $\bar{I}$ goes to infinity at infinity. From Assumption 1.5.2, we know that

$$\forall\theta \geq 0,\ \overline{F}(\theta) \leq \theta^2 A.$$

Let $x > 2$. Let now $C > 0$ a constant to be chosen later such that $Cx \geq 1/2$. We have by taking $\theta = Cx$

$$\begin{aligned}\bar{I}(x) &\geq J(Cx,x) - \overline{F}(Cx)\\ &\geq Cx^2 - \frac{1}{2} - \frac{1}{2}\log\frac{x}{A} - \frac{1}{2}\log x - AC^2x^2. \end{aligned}\tag{1.50}$$

Taking $C = 1/2A$, we obtain that there is constant $\kappa > 0$ such that

$$\bar{I}(x) \geq \frac{x^2}{4A} - o(x^2). \tag{1.51}$$

To get the converse bound, we show that as θ goes to infinity, $\overline{F}$ goes to infinity like $A\theta^2$. We distinguish two cases. First, we consider the case $A = B$. Using Proposition 1.5.8, we get the lower bound for $\theta \geq 1$,

$$\underline{F}(\theta) \geq A\theta^2\Big(1 - \frac{1}{\theta^2}\Big) - \frac{1}{4}\log\theta,$$

by taking $\alpha_2 = 0$, $\alpha_3 = 1-\theta^{-2}$, $\alpha_1 = \theta^{-2}$ and ν_1 the Gaussian law with variance α_1. In the case $A > B$, we define m_* such that $\psi(m_*) = A/2$. Taking $\alpha_3 = 0$, $\alpha_2 = 1-\theta^{-2}$, $\alpha_1 = \theta^{-2}$, $\nu_2 = \frac{2\theta\alpha_2}{m_*}\delta_{\sqrt{\frac{m_*}{2\theta}}}$, and ν_1 the Gaussian law with variance α_1, we obtain,

$$\underline{F}(\theta) \geq A\theta^2\Big(1 - \frac{1}{\theta^2}\Big) - \log\theta. \tag{1.52}$$

It follows that for any $\epsilon > 0$, there exists $M < \infty$ such that for $\theta \geq M$,

$$\underline{F}(\theta) \geq (1-\epsilon)A\theta^2.$$

Therefore

$$\bar{I}(x) \leq \max\left\{\sup_{\theta\geq M}\{J(\theta,x) - (1-\epsilon)A\theta^2\}, \sup_{\theta\leq M}\{J(\theta,x) - \overline{F}(\theta)\}\right\}.$$

But from Lemma 1.5.7 one can see that the second term in the above right-hand side is bounded by $Mx + C$ where C is a numerical constant. Besides, using the same argument as in (1.50), we get

$$\sup_{\theta\geq M}\{J(\theta,x) - (1-\epsilon)A\theta^2\} \geq \frac{x^2}{4(1-\epsilon)A} - o(x^2).$$

Hence, for x large enough,

$$\bar{I}(x) \leq \sup_{\theta \geq M} \{J(\theta, x) - (1-\epsilon)A\theta^2\}.$$

But, for x large enough and $\theta \geq 1/2$, $J(\theta, x) \leq \theta x$. Thus,

$$\sup_{\theta \geq M} \{J(\theta, x) - (1-\epsilon)A\theta^2\} \leq \sup_{\theta \geq 0} \{\theta x - (1-\epsilon)A\theta^2\} = \frac{x^2}{4(1-\epsilon)A},$$

which ends the proof.

□

Proposition 1.5.13. *For any $\theta \geq 0$, $J(\theta, .)$ is a convex function. Therefore, $\bar{I}$ is also convex.*

Proof. Let $x, y \geq 2$ and $t \in (0,1)$. Let E_N be a sequence of diagonal matrices such that $||E_N|| \leq 2$ and such that $\hat{\mu}_{E_N}$ converges weakly to σ. Let E_N^x and E_N^y be such that $(E_N^x)_{i,i} = (E_N^y)_{i,i} = (E_N)_{i,i}$ for any $i \in \{1, \dots, N-1\}$, and

$$(E_N^x)_{N,N} = x \quad (E_N^y)_{N,N} = y.$$

We have $\lambda_{E_N^x} = x$ and $\lambda_{E_N^y} = y$. Then, $H_N = tE_N^x + (1-t)E_N^y$ is such that its empirical distribution of eigenvalues converges to σ, and $\lambda_{H_N} = tx + (1-t)y$. By Hölder's inequality we have,

$$\log I_N(H_N, \theta) \leq t \log I_N(E_N, \theta) + (1-t) \log I_N(D_N, \theta).$$

Taking the limit as $N \to +\infty$, we get,

$$J(tx + (1-t)y, \theta) \leq tJ(x, \theta) + (1-t)J(y, \theta).$$

Therefore, $J(\theta, .)$ is convex and I is convex as a supremum of convex functions. □

To derive the large deviation lower bound, we denote by $\mathcal{C}_\mu$ the set of $\theta \in \mathbb{R}^+$ such that

$$\underline{F}(\theta) = \overline{F}(\theta) =: F(\theta).$$

By Lemma 1.5.10, $\mathcal{C}_\mu$ is not empty. We observe also that by continuity of both $\underline{F}$ and $\overline{F}$ (see Remark 1.5.9), $\mathcal{C}_\mu$ is closed. Let

$$\forall x \geq 2, \; I(x) = \sup_{\theta \in \mathcal{C}_\mu} \{J(x, \theta) - F(\theta)\}$$

Theorem 1.5.14. *For any $x \geq 2$, denote by*

$$\Theta_x = \{\theta \geq 0 : \bar{I}(x) = J(x, \theta) - F(\theta)\},$$

where $\bar{I}$ is defined in (1.49). *Let $x \geq 2$ such that there exists $\theta \in \Theta_x \cap \mathcal{C}_\mu$ and $\theta \notin \Theta_y$ for any $y \neq x$. Then, $I(x) = \bar{I}(x)$ and*

$$\lim_{\delta \to 0} \liminf_{N \to +\infty} \frac{1}{N} \log \mathbb{P}\left(|\lambda_{X_N} - x| \leq \delta\right) \geq -I(x).$$

We apply this general theorem in two cases. We first investigate the case where the function ψ is increasing, case for which we can check that our hypotheses on the sets Θ_x holds for x large enough. This includes the case where μ is the sparse Gaussian law, see Example 1.5.3.

Proposition 1.5.15. *Suppose that Assumptions 1.5.2 and 1.5.4 hold. If ψ is increasing on $\mathbb{R}_+$, then $\mathcal{C}_\mu = \mathbb{R}^+$. Moreover, there exists $x_\mu \geq 2$ such that for any $x \geq x_\mu$, the large deviation lower bound holds with rate function I.*

We then consider the case where μ is such that $B < A$. This includes any compactly supported measure μ since $B = 0$. We prove in this case the following result.

Proposition 1.5.16. *Suppose that Assumptions 1.5.2 and 1.5.4 hold. If μ is such that $B < A$ and such that the maximum of ψ is attained on $\mathbb{R}^+$ for a unique m_* such that $\psi''(m_*) < 0$, then there exists a positive finite real number θ_0 such that for $[\theta_0, +\infty[\subset \mathcal{C}_\mu$. Therefore, there exists a finite constant x_μ such that for $x \geq x_\mu$, the large deviation lower bound holds with rate function I. Furthermore on the interval $[x_\mu, +\infty)$ the rate function I depends only on A.*

In the case where A is sufficiently small, we can show without any additional assumption that the large deviation lower bound holds in a vicinity of 2 and the rate function I is equal to the one of the GOE. This contrasts with Proposition 1.5.12 which shows that the rate function $\bar{I}$ goes to infinity like $x^2/4A$ at infinity and therefore depends on A. In other words the "heavy tails" only kicks in above a certain threshold.

Proposition 1.5.17. *Assume $A < 2$. The large deviation lower bound holds with rate function $\bar{I}$ on $[2, 1/\sqrt{A-1} + \sqrt{A-1}]$. Moreover, $\bar{I}$ coincides with the rate function in the GOE case I_{GOE}, defined in (1.45), on this interval. As a consequence, for all $x \in [2, 1/\sqrt{A-1} + \sqrt{A-1}]$*

$$\lim_{\delta\to 0}\liminf_{N\to+\infty} \frac{1}{N}\log \mathbb{P}\left(|\lambda_{X_N} - x| \leq \delta\right) = \lim_{\delta\to 0}\limsup_{N\to+\infty} \frac{1}{N}\log \mathbb{P}\left(|\lambda_{X_N} - x| \leq \delta\right) = -I_{GOE}(x).$$

In the next section 1.5.2, we detail our approach to prove large deviations lower bounds. Since Proposition 1.5.8 is crucial to all our results, we prove it in the next section 1.5.2. Then, we will apply these results to prove the large deviations lower bounds close to the bulk in section 1.5.4, that is, we give a proof of Proposition 1.5.17. To prove the large deviations lower bounds for large x, we consider first the case of increasing ψ in section 1.5.5 and then the case of $B < A$ in section 1.5.6. Indeed, the variational formulas for the limiting annealed spherical integrals differ in these two cases, as $B = A$ in the first case whereas $B < A$ in the second.

1.5.2 A general large deviation lower bound

We first prove Theorem 1.5.14 and will then give more practical descriptions of the sets Θ_x in order to apply it.

Proof of Theorem 1.5.14. Introducing the spherical integral with parameter $\theta \geq 0$, we have

$$\mathbb{P}\left(|\lambda_{X_N} - x| \leq \delta\right) \geq \mathbb{E}\Big(\mathbb{1}_{X_N\in V_{\delta,x}^K}\frac{I_N(X_N,\theta)}{I_N(X_N,\theta)}\Big),$$

where $V_{\delta,x}^K = \{Y \in \mathcal{H}_N : |\lambda_Y - x| \leq \delta, d(\hat{\mu}_Y, \sigma) < N^{-\kappa}, ||X_N|| \leq K\}$ for some $K > 0$ and $\kappa > 0$. Using the continuity of the spherical integral (see [58, Proposition 2.1]), we get

$$\mathbb{P}\left(|\lambda_{X_N} - x| \leq \delta\right) \geq \frac{\mathbb{E}\big(\mathbb{1}_{X_N\in V_{\delta,x}^K} I_N(X_N,\theta)\big)}{\mathbb{E} I_N(X_N,\theta)} e^{N\underline{F}(\theta) - NJ(x,\theta) - Ng(\delta) - o(N)},$$

where g is a function such that $g(\delta) \to 0$ as $\delta \to 0$. It remains to show that we can find $\theta_x \in \mathcal{C}_\mu$ such that

$$\liminf_{N\to+\infty} \frac{1}{N}\log\frac{\mathbb{E}\big(\mathbb{1}_{X_N\in V_{\delta,x}^K} I_N(X_N,\theta_x)\big)}{\mathbb{E} I_N(X_N,\theta_x)} \geq 0.$$

To this end we will use our large deviation upper bound. Since $\hat{\mu}_{X_N}$ concentrates at the scale N, and $||X_N||$ is exponentially tight at the scale N by Assumption 1.5.4, it suffices to prove we can find $\theta = \theta_x$ such that for all $y \neq x$, for δ small enough, and K large enough,

$$\limsup_{N} \frac{1}{N}\log\frac{\mathbb{E}[\mathbb{1}_{X_N\in V_{\delta,y}^K} I_N(X_N,\theta)]}{\mathbb{E} I_N(X_N,\theta)} < 0.$$

By assumption, there exists $\theta \in \Theta_x \cap \mathcal{C}_\mu$ such that $\theta \notin \Theta_y$ for $y \neq x$. We introduce a new spherical integral with argument θ' and use again the continuity of J_N to show that:

$$\begin{aligned}
\frac{\mathbb{E}[\mathbb{1}_{X_N \in V^K_{\delta,y}} I_N(X_N,\theta)]}{\mathbb{E} I_N(X_N,\theta)} &= \frac{\mathbb{E}[\mathbb{1}_{X_N \in V^K_{\delta,y}} \frac{I_N(X_N,\theta')}{I_N(X_N,\theta')} I_N(X_N,\theta)]}{\mathbb{E} I_N(X_N,\theta)} \\
&= e^{-NJ(y,\theta')-N\underline{F}(\theta)+NJ(y,\theta)+N\epsilon(\delta)} \mathbb{E}[1_{X_N \in V^K_{\delta,y}} I_N(X_N,\theta')] \\
&\leq e^{-NJ(y,\theta')-N\underline{F}(\theta)+NJ(y,\theta)+N\overline{F}(\theta')+N\epsilon(\delta)},
\end{aligned}$$

where $\epsilon(\delta) \to 0$ as $\delta \to 0$. We can conclude that

$$\begin{aligned}
\limsup_{\substack{N\to+\infty \\ \delta \to 0}} \frac{1}{N} \log \frac{\mathbb{E}[\mathbb{1}_{X_N \in V^K_{\delta,y}} I_N(X_N,\theta)]}{\mathbb{E} I_N(X_N,\theta)} &\leq -\sup_{\theta'}\{J(y,\theta') - \overline{F}(\theta')\} + J(y,\theta) - \underline{F}(\theta) \\
&= -\bar{I}(y) + J(y,\theta) - \underline{F}(\theta) \qquad (1.53)
\end{aligned}$$

By assumption, $\theta \notin \Theta_y$, and $\theta \in \mathcal{C}_\mu$ so that $\underline{F}(\theta) = \overline{F}(\theta)$ hence

$$-\bar{I}(y) + J(y,\theta) - \overline{F}(\theta) < 0$$

and the conclusion follows from (1.53). □

In a first step, we identify a subset defined in terms of the subdifferential sets of F at the points of non-differentiability where the large deviation lower bound holds. Let $\mathcal{D}$ be the set of θ such that $\overline{F}$ is differentiable at θ.

Lemma 1.5.18. *The lower bound holds for any $x > 2$ such that $I(x) = \bar{I}(x) > 0$ and*

$$x \notin E := \bigcup_{\theta \in \mathcal{D}^c} \Big(\frac{1}{2\theta} + \partial \overline{F}(\theta) \Big), \qquad (1.54)$$

where $\partial\overline{F}(\theta)$ denotes the subdifferential of $\overline{F}$ at θ

Note that since $\overline{F}$ is a convex function, its subdifferentials are well defined.

Proof. Let $x > 2$ such that $I(x) = \bar{I}(x) > 0$ and $x \notin E$. Since $\overline{F}(\theta) \geq \theta^2$ for any $\theta \geq 0$ by (1.48) and $\overline{F}$ is continuous by Remark 1.5.9, we deduce from Lemma 1.5.7 that $\theta \mapsto J(x,\theta) - \overline{F}(\theta)$ is continuous and goes to $-\infty$ as θ goes to $+\infty$. Since $\mathcal{C}_\mu$ is closed, the supremum,

$$\sup_{\theta \in \mathcal{C}_\mu} \big\{ J(\theta,x) - \overline{F}(\theta) \big\}, \qquad (1.55)$$

is achieved at some $\theta \in \mathcal{C}_\mu$. We will show that $\theta \in \mathcal{D}$.

As $\bar{I}(x) \neq 0$ we must have $\theta > \frac{1}{2}G_\sigma(x)$. Indeed, $\overline{F}(\theta) \geq \theta^2$ and $J(\theta,x) = \theta^2$ for $\theta \leq \frac{1}{2}G_\sigma(x)$ by Lemma 1.5.7 so that

$$\sup_{\theta \leq \frac{1}{2}G_\sigma(x)} \{J(\theta,x) - \overline{F}(\theta)\} = 0.$$

Since $\bar{I}(x) = I(x)$, we deduce by Fermat's rule that θ satisfies the condition:

$$0 \in \frac{\partial J}{\partial \theta}(\theta,x) - \partial\overline{F}(\theta) = x - \frac{1}{2\theta} - \partial\overline{F}(\theta)\,.$$

Since $x \notin E$, we deduce that $\overline{F}$ is differentiable at θ.

According to Theorem 1.5.14, to prove that the lower bound holds at x, it suffices to show that $\theta \notin \Theta_y$ for any $y \neq x$. Let $y \geq 2$ such that $\theta \in \Theta_y$. As $\overline{F}$ is differentiable at θ, it should be a critical point of both $J(y,.) - \overline{F}$ and $J(x,.) - \overline{F}$. We deduce that,

$$\frac{\partial}{\partial \theta} J(y,\theta) = \frac{\partial}{\partial \theta} J(x,\theta).$$

If $G_\sigma(y) < 2\theta$, then we obtain by Lemma 1.5.7 and the fact that $G_\sigma(x) < 2\theta$ that $x = y$. If $G_\sigma(y) \geq 2\theta$, then we have

$$x - \frac{1}{2\theta} = 2\theta.$$

Since $G_\sigma^{-1}(\zeta) = \zeta + 1/\zeta$ for any $\zeta \in (0,1]$, we get $G_\sigma(x) = 2\theta$. As we assumed that $2\theta > G_\sigma(x)$, we conclude that $\theta \notin \Theta_y$ for any $G_\sigma(y) \geq 2\theta$, which completes the proof. $\square$

We are now ready to prove the following result:

Proposition 1.5.19. *Assume that there exists $\theta_0 > 0$ such that $[\theta_0, +\infty[\subset \mathcal{C}_\mu$ and such that $\overline{F}$ is differentiable on $(\theta_0, +\infty)$. There exists $x_\mu \geq 2$ such that for any $x \geq x_\mu$, $I(x) = \bar{I}(x)$ and the large deviation lower bound holds for any $x \geq x_\mu$ with rate function $I(x)$.*

Proof. On one hand,

$$\sup_{\theta \leq \theta_0} \{J(\theta, x) - \overline{F}(\theta)\} \leq \theta_0 x + C,$$

where C is some positive constant. Since $\bar{I}(x) \geq x^2/4A - o(x^2)$ by (1.51), we deduce that there exists $x_\mu \geq 2$ such that for $x \geq x_\mu$, $\bar{I}(x) > 0$ and the supremum of $J(.,x) - \overline{F}$ is achieved on $[\theta_0, +\infty)$, in particular on $\mathcal{C}_\mu$. Thus, for any $x \geq x_\mu$, $\bar{I}(x) = I(x) > 0$. In view of Lemma 1.5.18, it remains to show that E, defined in (1.54), is a bounded set. From the fact that $\overline{F}$ is differentiable on $(\theta_0, +\infty)$ and Lemma 1.5.10, we deduce that

$$\mathcal{D}^c \subset \Big[\frac{1}{2\sqrt{A-1}}, \theta_0\Big].$$

We observe that since $0 \leq \overline{F}(\theta) \leq A\theta^2$, we have for any $\zeta \in \partial\overline{F}(\theta)$,

$$\zeta\theta \leq \overline{F}(2\theta) - \overline{F}(\theta) \leq 4A\theta^2,$$

and thus $\zeta \leq 4A\theta$. Therefore, E is a bounded set, which ends the proof. $\square$

1.5.3 Asymptotics of the annealed spherical integral

In this section we prove Proposition 1.5.8. Taking the expectation first with respect to X_N, we have

$$\begin{aligned} F_N(\theta) &= \frac{1}{N} \log \mathbb{E}_{X_N} \mathbb{E}_e[\exp(N\theta\langle e, X_N e\rangle)] \\ &= \frac{1}{N} \log \mathbb{E}_e \exp\big(f(e)\big), \end{aligned}$$

where

$$f(e) = \sum_{i<j} L\big(2\sqrt{N}\theta e_i e_j\big) + \sum_i L(\sqrt{2N}\theta e_i^2).$$

In a first step, we will prove the following variational representation of the upper and lower limits of $F_N(\theta)$.

Lemma 1.5.20. *Let X_N be a Wigner matrix satisfying Assumptions 1.5.2 and 1.5.4. Then for any $\theta > 0$,*

$$\underline{F}(\theta) \leq \liminf_{N\to+\infty} F_N(\theta) \leq \limsup_{N\to+\infty} F_N(\theta) \leq \overline{F}(\theta)$$

with

$$\overline{F}(\theta) = \limsup_{\substack{\delta\to 0, K\to+\infty \\ \delta K\to 0}} \sup_{\substack{c=c_1+c_2+c_3 \\ c_i\geq 0}} \limsup_{N\to+\infty} F^N_{c_1,c_2,c_3}(\delta, K),$$

$$\underline{F}(\theta) = \sup_{\substack{c=c_1+c_2+c_3 \\ c_i\geq 0}} \liminf_{\substack{\delta\to 0, K\to+\infty \\ \delta K\to 0}} \liminf_{N\to+\infty} F^N_{c_1,c_2,c_3}(\delta, K),$$

where

$$F^N_{c_1,c_2,c_3}(\delta, K) = \sup_{\substack{s_i\geq\sqrt{cK}N^{1/4} \\ |\sum s_i^2 - c_3 N|\leq \delta N}} \sup_{\substack{\sqrt{c\delta}\leq t_i N^{-1/4}\leq\sqrt{cK} \\ |\sum t_i^2 - Nc_2|\leq\delta N}} \Big\{ \frac{\theta^2}{c^2}\big(c_1^2 + 2c_1c_2 + Bc_3^2\big) - \frac{1}{2}\big(c_2+c_3\big)$$
$$+ \frac{1}{N}\sum_{i,j} L\Big(\frac{2\theta s_i t_j}{\sqrt{N}c}\Big) + \frac{1}{2N}\sum_{i,j} L\Big(\frac{2\theta t_i t_j}{\sqrt{N}c}\Big) + \sup_{\substack{\nu\in\mathcal{P}(I_1) \\ \int x^2 d\nu_1(x)=c_1}} \big\{\Phi(\nu,s) - H(\nu|\gamma)\big\}\Big\},$$

and

$$\Phi(\nu, s) = \sum_{i=1}^k \int L\Big(\frac{2\theta x s_i}{\sqrt{N}c}\Big) d\nu(x),$$

with $I_1 = [-\sqrt{c}\delta N^{1/4}, \sqrt{c}\delta N^{1/4}]$.

Proof. We use the representation of the law of the vector e uniformly distributed on the sphere as a renormalized Gaussian vector $g/\|g\|_2$ where g is a standard Gaussian vector in $\mathbb{R}^N$, to write

$$\mathbb{E}_e \exp\big(f(e)\big) = \mathbb{E}\left[\exp(\Sigma(g))\right],$$

where

$$\Sigma(g) = \sum_{i<j} L\Big(2\sqrt{N}\theta\frac{g_i g_j}{\sum_{i=1}^N g_i^2}\Big) + \sum_i L\Big(\sqrt{2N}\theta\frac{g_i^2}{\sum_{i=1}^N g_i^2}\Big).$$

To study the large deviation of $\Sigma(g)$, we split the entries of g into three possible regime: the regime where $g_i \ll N^{1/4}$, an intermediate regime where $g_i \simeq N^{1/4}$ and finally $g_i \gg N^{1/4}$. Fix some $c, K, \delta > 0$ such that $0 < K^{-1} \leq c \leq K$, and $0 < 4\delta < K^{-1}$. Define I_1, I_2, I_3 as

$$\begin{aligned} I_1 &= [0, \sqrt{c\delta}N^{\frac14}] \\ I_2 &= [\sqrt{c\delta}N^{\frac14}, \sqrt{cK}N^{\frac14}] \\ I_3 &= [\sqrt{cK}N^{\frac14}, \sqrt{N(c+3\delta)}]. \end{aligned}$$

Let for $i = 1,2,3$, $J_i = \{j : |g_j| \in I_i\}$ and $\hat{c}_i^N = \sum_{j\in J_i} g_j^2/N$. We will fix the empirical variances $\hat{c}_i^N$ in a first step. To this end let $c_1, c_2, c_3 > 0$ such that $c_1 + c_2 + c_3 = c$. We will compute the asymptotics of

$$F^N_{c_1,c_2,c_3}(\theta,\delta) = \mathbb{E}[\exp(\Sigma(g))\mathbb{1}_{\mathcal{A}^\delta_{c_1,c_2,c_3}}].$$

where

$$\mathcal{A}^\delta_{c_1,c_2,c_3} := \bigcap_{1\leq i\leq 3} \{|\hat{c}_i^N - c_i| \leq \delta\}.$$

Let

$$\Sigma_c(g) = \sum_{i<j} L\Big(2\theta\frac{g_i g_j}{\sqrt{N}c}\Big) + \sum_i L\Big(\sqrt{2}\theta\frac{g_i^2}{\sqrt{N}c}\Big).$$

Using the fact the $L(\sqrt{.})$ is Lipschitz, we prove in the next lemma that on the event $\mathcal{A}^\delta_{c_1,c_2,c_3}$, $\Sigma_c(g)$ is a good approximation of $\Sigma(g)$.

Lemma 1.5.21. *On the event* $\mathcal{A}^{\delta}_{c_1,c_2,c_3}$,

$$\Sigma(g) - \Sigma_c(g) = N o_{\delta K}(1), \quad \text{as } \delta K \to 0. \tag{1.56}$$

Moreover

$$|J_3| \leq \frac{2\sqrt{N}}{K}, \qquad |J_2 \cup J_3| \leq \frac{2\sqrt{N}}{\delta}.$$

Proof. Note that since μ is symmetric, $L(x) = L(|x|)$ and since we assumed $L(\sqrt{.})$ Lipschitz, for any $x, y \in \mathbb{R}$, $|L(x) - L(y)| \leq L|x^2 - y^2|$ for some finite constant L. Therefore,

$$\begin{aligned}\sum_{1 \leq i \neq j \leq N} \left| L\Big(2\sqrt{N}\theta \frac{g_i g_j}{\sum_{i=1}^N g_i^2}\Big) - L\Big(2\theta \frac{g_i g_j}{\sqrt{N} c}\Big)\right| &\leq \frac{4L\theta^2}{N} \sum_{i,j} g_i^2 g_j^2 \left| \frac{1}{(\hat{c}_1^N + \hat{c}_2^N + \hat{c}_3^N)^2} - \frac{1}{c^2} \right| \\ &\leq CNL\theta^2 (c + 3\delta)^2 \frac{\delta}{(c - 3\delta)^3},\end{aligned}$$

where C is a numerical constant. Since $K^{-1} < c < K$, and $4\delta < K^{-1}$ we get

$$\sum_{1 \leq i \neq j \leq N} \left| L\Big(2\sqrt{N}\theta \frac{g_i g_j}{\sum_{i=1}^N g_i^2}\Big) - L\Big(2\theta \frac{g_i g_j}{\sqrt{N} c}\Big)\right| = O(N\delta K),$$

and similarly for the diagonal terms. The estimates on $|J_3|$ and $|J_2|$ are straightforward consequences of Tchebychev's inequality.

□

We next fix the positions of the set of indices J_1, J_2, J_3. Using the invariance under permutation of the coordinates of the Gaussian measure, we can write

$$F^N_{c_1,c_2,c_3}(\theta, \delta) = \sum_{\substack{0 \leq k \leq 2\sqrt{N\delta} \\ 0 \leq l \leq 2\sqrt{NK}}} \binom{N}{k} \binom{N-k}{l} F^{k,l}_{c_1,c_2,c_3},$$

where

$$F^{k,l}_{c_1,c_2,c_3} = \mathbb{E}\Big[\exp(\Sigma_c(g)) \mathbb{1}_{\mathcal{A}^{\delta}_{c_1,c_2,c_3} \cap \mathcal{I}_{k,l}}\Big]$$

and

$$\mathcal{I}_{k,l} = \big\{J_3 = \{1, \ldots, k\}, J_2 = \{k+1, \ldots, k+l\}, J_1 = \{k+l+1, \ldots, N\}\big\}.$$

As the number of all the possible configurations of J_2 and J_3 are sub-exponential in N by Lemma 1.5.21, that is, for any $k \leq 2\sqrt{N}/K$ and $l \leq 2\sqrt{N}/\delta$,

$$\max\Big(\binom{N}{k}, \binom{N-k}{l}\Big) = e^{O(\frac{\sqrt{N}}{\delta} \log N)},$$

we are reduced to compute $F^{k,l}_{c_1,c_2,c_3}$ for fixed k, l. More precisely, we obtain the following result.

Lemma 1.5.22.

$$\log F^N_{c_1,c_2,c_3}(\theta, \delta) = \max_{\substack{k \leq 2\sqrt{N}/K \\ l \leq 2\sqrt{N}/\delta}} \log F^{k,l}_{c_1,c_2,c_3} + O\Big(\frac{\sqrt{N}}{\delta} \log N\Big),$$

To simplify the notations, we denote for any $a, b \in \{1, 2, 3\}$,

$$\forall x, y \in \mathbb{R}^N, \ \Sigma_{a,b}(x, y) = \frac{1}{2N} \sum_{i \in J_a, j \in J_b} L\left(2\theta \frac{x_i y_j}{\sqrt{N} c}\right),$$

if $a \neq b$, and

$$\forall x, y \in \mathbb{R}^N,\ \Sigma_{a,a}(x,y) = \frac{1}{N}\sum_{i\neq j\in J_a} L\left(2\theta\frac{x_i y_j}{\sqrt{N}c}\right) + \frac{1}{N}\sum_{i\in J_a} L\Big(\sqrt{2}\theta\frac{x_i y_i}{\sqrt{N}c}\Big),$$

where now $J_3 = \{1,\dots,k\}, J_2 = \{k+1,\dots,k+l\}, J_1 = \{k+l+1,\dots,N\}$.

Next, we single out the interaction terms which involve either the quadratic behavior of L at 0 or at $+\infty$.

Lemma 1.5.23. *On the event $\mathcal{A}^\delta_{c_1,c_2,c_3}$,*

$$\Sigma_{1,1}(g,g) + 2\Sigma_{1,2}(g,g) + \Sigma_{3,3}(g,g) = \frac{\theta^2}{c^2}\big(c_1^2 + 2c_1c_2 + Bc_3^2\big) + o_{\delta K}(1) + o_K(1),$$

as $\delta K \to 0$ and $K \to +\infty$.

Proof. Observe that for $i \in J_1$, $j \in J_1 \cup J_2$, $|g_i g_j| \leq \sqrt{NK\delta}c$. Since $L(x) \sim_0 x^2/2$, we get,

$$\Sigma_{1,1}(g,g) + 2\Sigma_{1,2}(g,g) = \theta^2\frac{\hat{c}_1^2}{c^2} + 2\theta^2\frac{\hat{c}_1\hat{c}_2}{c^2} + o_{\delta K}(1), \tag{1.57}$$

as $\delta K \to 0$. Since $g \in \mathcal{A}^\delta_{c_1,c_2,c_3}$, we have $|\hat{c}_i^2 - c_i^2| = O(\delta c)$ for any $i \in \{1,2,3\}$. But $c \geq K^{-1}$, therefore

$$\Sigma_{1,1}(g,g) + 2\Sigma_{1,2}(g,g) = \theta^2\frac{c_1^2}{c^2} + 2\theta^2\frac{c_1c_2}{c^2} + o_{\delta K}(1).$$

For $i,j \in J_3$, $|g_ig_j| \geq K\sqrt{cN}$. Since $L(x) \sim_{+\infty} \frac{B}{2}x^2$, we deduce similarly that

$$\Sigma_{3,3}(g,g) = \Big(\frac{B}{2} + o_\delta(1)\Big)\frac{\theta^2}{N^2c^2}\Big(\sum_{i\in J_3} g_i^2\Big)^2 = B\theta^2\frac{c_3^2}{c^2} + o_K(1), \tag{1.58}$$

as $K \to +\infty$, which gives the claim.

□

From the Lemmas 1.5.21 and 1.5.23, we have on the event $\mathcal{A}^\delta_{c_1,c_2,c_3}$,

$$\Sigma(g) = \frac{\theta^2}{c^2}\big(c_1^2 + 2c_1c_2 + Bc_3^2\big) + \Sigma_{1,3}(g,g) + 2\Sigma_{2,3}(g,g) + \Sigma_{2,2}(g,g) + o_{\delta K}(1) + o_K(1). \tag{1.59}$$

Lemma 1.5.24. *Let $k,l \in \mathbb{N}$ such that $k + l \leq N$.*

$$\begin{aligned}\log\mathbb{E}\exp\Big(\Big[\Sigma_{1,3}(g,g)+&2\Sigma_{2,3}(g,g) + \Sigma_{2,2}(g,g)\Big]\mathbb{1}_{\mathcal{A}^\delta_{c_1,c_2c_3}\cap\mathcal{I}_{k,l}}\Big)\\ &= -\frac{N}{2}(c_2+c_3) + o_\delta(1) + O\Big(\frac{\sqrt{N}}{\delta^2}\log(1/\delta)\Big)\\ &+ \max_{\substack{t_i\in I_2\\ |\sum_i t_i^2 - c_2N|\leq\delta N}}\ \max_{\substack{s_i\in I_3\\ |\sum_i s_i^2 - c_3N|\leq\delta N}} \log\mathbb{E}\Big(\exp\Big[\Sigma_{1,3}(g,s) + 2\Sigma_{2,3}(t,s) + \Sigma_{2,2}(t,t)\Big]\mathbb{1}_{\mathcal{A}^\delta_{c_1}}\Big),\end{aligned}$$

where

$$\mathcal{A}^\delta_{c_1} = \big\{|\sum_{i=1}^{N-k-l} g_i^2 - Nc_1| \leq N\delta,\ g_i \in I_1,\ \forall i \in \{1,\dots,N-k-l\}\big\}.$$

Proof. Integrating on $g_i, i \le k+l$, we have

$$\begin{aligned}
\mathbb{E}\exp\Big(\Big[\Sigma_{1,3}(g,g)&+2\Sigma_{2,3}(g,g)+\Sigma_{2,2}(g,g)\Big]\mathbb{1}_{\mathcal{A}^\delta_{c_1,c_2c_3}\cap\mathcal{I}_{k,l}}\Big)\\
&\le \frac{1}{(2\pi)^{\frac{k+l}{2}}}\int \mathbb{1}_{\{|\frac{1}{N}\sum_{i=1}^{k+l} g_i^2-(c_2+c_3)|\le 2\delta\}} e^{-\frac{1}{2}\sum_{i=1}^{k+l} g_i^2}\prod_{i=1}^{k+l} dg_i\\
&\times \max_{\substack{t\in I_2^l\\ |\sum_i t_i^2-c_2N|\le\delta N}}\ \max_{\substack{s\in I_3^k\\ |\sum_i s_i^2-c_3N|\le\delta N}} \mathbb{E}\Big(\exp\Big[\Sigma_{1,3}(g,s)+2\Sigma_{2,3}(t,s)+\Sigma_{2,2}(t,t)\Big]\mathbb{1}_{\mathcal{A}^\delta_{c_1}}\Big),
\end{aligned}$$

But,

$$\begin{aligned}
\frac{1}{(2\pi)^{\frac{k+l}{2}}}&\int \mathbb{1}_{\{|\frac{1}{N}\sum_{i=1}^{k+l} g_i^2-(c_2+c_3)|\le 2\delta\}} e^{-\frac{1}{2}\sum_{i=1}^{k+l} g_i^2}\prod_{i=1}^{k+l} dg_i\\
&\le \frac{1}{(2\pi)^{\frac{k+l}{2}}} e^{-\frac{1}{2}(1-\delta)(c_2+c_3-2\delta)N}\int e^{-\frac{\delta}{2}\sum_{i=1}^{k+l} g_i^2}\prod_{i=1}^{k+l} dg_i\\
&\le e^{-\frac{N}{2}(c_2+c_3)+(O(\delta)+O(\delta K))N}\big(\sqrt{2\pi}\delta^{-1}\big)^{\frac{k+l}{2}},
\end{aligned}$$

where we used the fact that $c \le K$. By Lemma 1.5.21 we have $k+l = O(\sqrt{N}/\delta)$. Therefore,

$$\begin{aligned}
\log\mathbb{E}\exp&\Big(\Big[\Sigma_{1,3}(g,g)+2\Sigma_{2,3}(g,g)+\Sigma_{2,2}(g,g)\Big]\mathbb{1}_{\mathcal{A}^\delta_{c_1,c_2,c_3}\cap\mathcal{I}_{k,l}}\Big)\\
&\le -\frac{N}{2}(c_2+c_3)+O(\delta)N+O(\delta K)N+O\Big(\frac{\sqrt{N}}{\delta}\log(1/\delta)\Big)\\
&+\max_{\substack{t_i\in I_2\\ |\sum_i t_i^2-c_2N|\le\delta N}}\ \max_{\substack{s_i\in I_3\\ |\sum_i s_i^2-c_3N|\le\delta N}} \log\mathbb{E}\Big(\exp\Big[\Sigma_{1,3}(g,s)+2\Sigma_{2,3}(t,s)+\Sigma_{2,2}(t,t)\Big]\mathbb{1}_{\mathcal{A}^\delta_{c_1}}\Big).
\end{aligned}$$

□

Hence, we are left to estimate

$$\Lambda_1^N = \mathbb{E}\big[\exp\big(N\Sigma_{1,3}(g,s)\big)\mathbb{1}_{\mathcal{A}^\delta_{c_1}}\big],$$

where $s \in I_3^k$ such that $|\sum_i s_i^2 \le \delta N - c_3 N|$. Let $\hat{\mu}_N = \frac{1}{|J_1|}\sum_{i\in J_1}\delta_{g_i}$. We can write

$$\Sigma_{1,3}(g,s) = \frac{|J_1|}{N}\sum_{i=1}^{k}\int L\left(\frac{2\theta x s_i}{\sqrt{N}c}\right)d\hat{\mu}_N(x).$$

The first difficulty in estimating Λ_1^N lies in the fact that the function

$$x \mapsto \sum_{i=1}^{k} L\Big(\frac{2\theta x s_i}{\sqrt{N}c}\Big),$$

is not bounded so that Varadhan's lemma (see [36, Theorem 4.3.1]) cannot be applied directly. The second issue is that we need a large deviation estimate which is uniform in the c|hoice of $s \in I_3^k$ such that $|\sum_{i=1}^k s_i^2 - Nc_3| \le \delta N$. In the next lemma, we prove a uniform large deviation estimate of the type of Varadhan's lemma. The proof is postponed in the appendix 1.5.7.2.

Lemma 1.5.25. *Let $f : \mathbb{R} \to \mathbb{R}$ such that $f(0) = 0$ and $f(\sqrt{.})$ is a L-Lipschitz function. Let M_N, m_N be sequences such that $M_N = o(\sqrt{N})$ and $m_N \sim N$. Let $g_1, \dots, g_{m_N}$ be independent Gaussian random* *variables conditioned to belong to* $[-M_N, M_N]$. *Let* $\delta, K, c > 0$ *such that* $K^{-1} < c < K$, *and* $2\delta < K^{-1}$. *Then,*

$$\Big|\frac{1}{N}\log \mathbb{E} e^{\sum_{i=1}^{m_N} f\big(\frac{g_i}{\sqrt{c}}\big)} \mathbb{1}_{\{|\sum_{i=1}^{m_N} g_i^2 - cN| \le \delta N\}} - \sup_{\substack{\nu \in \mathcal{P}([-M_N, M_N]) \\ \int x^2 d\nu = c}} \Big\{ \int f\Big(\frac{x}{\sqrt{c}}\Big) d\nu(x) - H(\nu|\gamma) \Big\} \Big| \le g_{L,K}(N) + h_L(\delta),$$

where $g_{L,K}(N) \to 0$ *as* $N \to +\infty$ *and* $h_L(\delta) \to 0$ *as* $\delta \to 0$.

Let $s \in I_3^l$ such that $|\sum_i s_i^2 - Nc_3| \le \delta N$. We consider the function,

$$f : x \mapsto \sum_{i=1}^{k} L\Big(\frac{2\theta x s_i}{\sqrt{N}c}\Big).$$

One can observe that $f(\sqrt{.})$ is $4\theta^2 L$-Lipschitz. Using the fact that

$$\frac{1}{N}\log \mathbb{P}(g_i \in I_1, \forall i \le N-k-l) = o_N(1),$$

and the previous lemma, we deduce that for any $c_1 \ge K^{-1}$,

$$\Big|\frac{1}{N}\log \Lambda_1^N - \sup_{\substack{\nu \in \mathcal{P}(I_1) \\ \int x^2 d\nu(x) = c_1}} \Big\{ \sum_{i=1}^{k} \int L\Big(\frac{2\theta x s_i}{\sqrt{N}c}\Big) d\nu(x) - H(\nu|\gamma) \Big\}\Big| \le g_K(N) + h_L(\delta K), \tag{1.60}$$

where $g_K(N) \to 0$ as $N \to +\infty$. Putting together (1.59) and (1.60), we obtain

$$\Big|\frac{1}{N}\log \mathbb{E}\Big(\exp(\Sigma(g))\mathbb{1}_{\mathcal{A}_{c_1,c_2,c_3} \cap \mathcal{I}_{k,l}}\Big) - \Psi_{k,l}(c_1,c_2,c_3)\Big| \le g_{L,K}(N) + o_{\delta K}(1) + o_K(1),$$

where

$$\begin{aligned}\Psi_{k,l}(c_1,c_2,c_3) &= Q(c_1,c_2,c_3)\Big) - \frac{1}{2}(c_2+c_3) \\ &+ \max_{\substack{t_i \in I_2 \\ |\sum_i t_i^2 - c_2 N| \le \delta N}} \max_{\substack{s_i \in I_3 \\ |\sum_i s_i^2 - c_3 N| \le \delta N}} \Big\{ \Sigma_{2,3}(t,s) + \Sigma_{2,2}(t,t) + \sup_{\substack{\nu \in \mathcal{P}(I_1) \\ \int x^2 d\nu_1(x) = c_1}} \{\Phi(\nu,s) - H(\nu|\gamma_1)\}\Big\},\end{aligned}$$

with

$$Q(x,y,z) = \theta^2 \frac{x^2 + 2xy + Bz^2}{(x+y+z)^2},$$

and

$$\Phi(\nu,s) = \sum_{i=1}^{k} \int L\Big(\frac{2\theta x s_i}{\sqrt{N}c}\Big) d\nu(x).$$

By Lemma 1.5.22, we obtain as $\delta < K^{-1}$,

$$\Big|\frac{1}{N}\log F^N_{c_1,c_2,c_3}(\theta,\delta) - \max_{\substack{k \le 4\sqrt{N}/K \\ l \le 4\sqrt{N}/\delta}} \Psi_{k,l}(c_1,c_2,c_3)\Big| \le \tilde{g}_K(N) + o_{\delta K}(1) + o_K(1), \tag{1.61}$$

where $\tilde{g}_K(N) \to 0$ as $N \to +\infty$.

Let $c_1, c_2, c_3 \ge 0$ such that $c_1 > 0$. Let $c = c_1 + c_2 + c_3$. There exists $K > 0$ such that $K^{-1} \le c_1 \le c \le K$. We have for any $0 < 4\delta < K^{-1}$,

$$\liminf_{N\to+\infty} F_N(\theta) \ge \liminf_{N\to+\infty} \frac{1}{N}\log F^N_{c_1,c_2,c_3}(\theta,\delta).$$

Therefore,

$$\liminf_{N\to+\infty} F_N(\theta) \geq \liminf_{N\to+\infty} \max_{\substack{k\leq 2\sqrt{N}/K \\ l\leq 2\sqrt{N}/\delta}} \Psi_{k,l}(c_1,c_2,c_3) - o_{\delta K}(1) - o_K(1).$$

To complete the proof of the lower bound of Lemma 1.5.24, one can observe that

$$\Sigma_{2,2}(t,t) = \frac{1}{N}\sum_{i,j} L\Big(2\theta \frac{t_i t_j}{\sqrt{N}c}\Big) + o_N(1),$$

uniformly in $t_i \in I_2$ such that $|\sum_i t_i^2 - c_2 N| \leq \delta N$. Indeed, the diagonal terms are negligible in this case since

$$\frac{1}{N}\sum_{t_i\in J_2} \frac{t_i^4}{Nc^2} \leq c\delta^{-1}N^{-1/2}\frac{1}{N}\sum_{i\in J_2} t_i^2 = O(N^{-1/2}).$$

To conclude the proof of the upper bound, we will use the exponential tightness of $||g||^2$ and of $||g||_{J_1}^{-2}$. More precisely, we claim that

$$\lim_{K\to+\infty}\limsup_{N\to+\infty}\frac{1}{N}\mathbb{P}(||g||^2 \geq KN, ||g||_{J_1}^2 \leq K^{-1}N) = -\infty. \tag{1.62}$$

Indeed, it is clear by Chernoff's inequality that

$$\mathbb{P}(||g||^2 \geq KN) \leq Ce^{-CNK^2},$$

where C is a positive numerical constant. Whereas, using Lemma 1.5.21 and a union bound,

$$\mathbb{P}(||g||_{J_1}^2 \leq K^{-1}N) \leq \sum_{m\leq 2\sqrt{N}/\delta} \binom{N}{m}\mathbb{P}\Big(\sum_{i=1}^{N-m} g_i^2 \leq K^{-1}N\Big).$$

Let $\Lambda(\zeta) = \log \mathbb{E}e^{\theta g_1^2}$ for any $\zeta \in \mathbb{R}$. By Chernoff's inequality, we have for any $m \leq N$,

$$\mathbb{P}\Big(\sum_{i=1}^{N-m} g_i^2 \leq K^{-1}N\Big) \leq e^{-N\Lambda^*(K^{-1})},$$

where Λ^* is the Legendre transform of Λ. Since 0 is not in the support of the law g_1^2, we have $\Lambda^*(0) = +\infty$. Since Λ^* is lower semi-continuous we have $\Lambda^*(K^{-1}) \to +\infty$ as $K \downarrow +\infty$. Using the fact that for any $m \leq 2\sqrt{N}/\delta$,

$$\binom{N}{m} \leq e^{o(N)},$$

we get the claim (1.62).

Using that $L(x) \leq Bx^2/2$, we have

$$\Sigma(g) \leq B\theta^2 N.$$

From the exponential tightness (1.62), we deduce that there exists $\delta > 0$ such that,

$$\mathbb{E}\exp(\Sigma(g))\mathbb{1}_{\{\frac{1}{N}||g||^2\geq K, \frac{1}{N}||g||_{J_1}^2 \leq K^{-1}\}} \leq 1. \tag{1.63}$$

Let $\mathcal{E}_K = \{\frac{1}{N}||g||^2 \geq K, \frac{1}{N}||g||_{J_1}^2 \leq K^{-1}\}$. We have

$$0 \leq \limsup_{N\to+\infty} F_N(\theta) \leq \max\Big(\limsup_{N\to+\infty}\frac{1}{N}\log\mathbb{E}\big(e^{\Sigma(g)}\mathbb{1}_{\mathcal{E}_K}\big), \limsup_{N\to+\infty}\frac{1}{N}\log\mathbb{E}\big(e^{\Sigma(g)}\mathbb{1}_{\mathcal{E}_K^c}\big)\Big)$$

Since we took K so that (1.63) holds, we have

$$\limsup_{N\to+\infty} F_N(\theta) \leq \limsup_{N\to+\infty}\frac{1}{N}\log\mathbb{E}\big(e^{\Sigma(g)}\mathbb{1}_{\mathcal{E}_K}\big). \tag{1.64}$$

Let now $\mathcal{C}_\delta$ be a δ-net for the ℓ^∞-norm of the set

$$\{(c_1, c_2, c_3) \in \mathbb{R}_+^3 : c_1 + c_2 + c_3 \le K, c_1 \ge K^{-1}\}.$$

As $|\mathcal{C}_\delta| = O(K/\delta)$, we have

$$\limsup_{N\to+\infty} F_N(\theta) \le \limsup_{N\to+\infty} \max_{(c_1,c_2,c_3)\in\mathcal{C}_\delta} \frac{1}{N} \log F^N_{c_1,c_2,c_3}(\theta,\delta).$$

$$\limsup_{N\to+\infty} F_N(\theta) \le \max_{(c_1,c_2,c_3)\in\mathcal{C}_\delta} \limsup_{N\to+\infty} \frac{1}{N} \log F^N_{c_1,c_2,c_3}(\theta,\delta).$$

From (1.61), we deduce

$$\limsup_{N\to+\infty} F_N(\theta) \le \max_{\substack{c=c_1+c_2+c_3\\ c_1\ge K^{-1}}} \limsup_{N\to+\infty} \max_{\substack{k\le 2\sqrt{N}/K\\ l\le 2\sqrt{N}/\delta}} \Psi_{k,l}(c_1,c_2,c_3) + o_{\delta K}(1) + o_K(1).$$

Taking now the limit as $\delta \to 0$ and $K \to +\infty$ such that $\delta K \to 0$, we obtain the upper bound of Lemma 1.5.20. □

We are now ready to give a proof of Proposition 1.5.8 . Building on Lemma 1.5.20, we show that we can optimize on the total norm c in order to simplify the variational problem, and replace the supremum over discrete measures $\frac{1}{N}\sum_i \delta_{t_i}$ by a supremum over all measures with a certain constraint of support and second moment.

Proof of Proposition 1.5.8 . We use the notation of Lemma 1.5.20. We make the following changes of variables:

$$\forall i = 1,2,3,\ \alpha_i = \frac{c_i}{c},\ c = c_1 + c_2 + c_3,$$

$$\nu = \nu_1 \circ (x \mapsto x/\sqrt{c})^{-1}.$$

For any $\nu_1 \in \mathcal{P}(I_1)$ such that $\int x^2 d\nu_1(x) = c_1$ we have,

$$H(\nu_1|\gamma) = H(\nu_1) + \frac{1}{2}c_1 + \frac{1}{2}\log(2\pi).$$

Moreover, the density of ν and ν_1 are linked by the relation

$$\forall x \in \mathbb{R},\ \frac{d\nu}{dx}(x) = \sqrt{c}\frac{d\nu_1}{dx}(\sqrt{c}x).$$

Therefore,

$$H(\nu) = \int \log\frac{d\nu_1}{dx} d\nu_1 = H(\nu_1) + \frac{1}{2}\log c. \tag{1.65}$$

We obtain

$$\overline{F}(\theta) = \limsup_{\substack{\delta\to0,K\to+\infty\\ \delta K\to 0}} \sup_{\substack{\alpha_1+\alpha_2+\alpha_3=1\\ \alpha_i\ge0}} \sup_{c\ge0} \limsup_{N\to+\infty} \mathcal{F}^N_{\alpha_1,\alpha_2,\alpha_3,c}(\delta,K),$$

and similarly for $\underline{F}(\theta)$, where

$$\mathcal{F}^N_{\alpha_1,\alpha_2,\alpha_3,c}(\delta,K) = \sup_{\substack{s_i\ge KN^{1/4}\\ |\sum s_i^2 - N\alpha_3|\le N\delta}} \sup_{\substack{\delta\le t_i N^{-1/4}\le K\\ |\sum t_i^2-N\alpha_2|\le N\delta}} \Big\{\theta^2\big(\alpha_1^2 + 2\alpha_1\alpha_2 + B\alpha_3^2\big) - \frac{1}{2}c + \frac{1}{2}\log c$$
$$+ \frac{1}{N}\sum_{i,j} L\Big(\frac{2\theta s_i t_j}{\sqrt{N}}\Big) + \frac{1}{2N}\sum_{i,j} L\Big(\frac{2\theta t_i t_j}{\sqrt{N}}\Big) + \sup_{\substack{\nu\in\mathcal{P}(I_1)\\ \int x^2 d\nu_1(x)=\alpha_1}} \{\Phi(\nu,s) - H(\nu)\} - \frac{1}{2}\log(2\pi)\Big\},$$

and

$$\Phi(\nu, s) = \sum_{i=1}^{k} \int L\Big(\frac{2\theta x s_i}{\sqrt{N}}\Big) d\nu(x),$$

with $I_1 = [-\delta N^{1/4}, \delta N^{1/4}]$. We see that we can optimize in c, and find that the maximum is achieved at $c = 1$ by concavity of the log. We now fix $s_i \in S_3$ such that $\sum_i s_i^2 = N\alpha_3$ and we consider now the problem of optimizing over the parameters $t_i \in [\delta N^{1/4}, KN^{1/4}]$, $i \leq l \leq N$ such that $\sum_i t_i^2 = N\alpha_2$, the function

$$\mathcal{M}(l, (t_i, i \leq l)) = \frac{1}{N} \sum_{i,j} L\Big(\frac{2\theta s_i t_j}{\sqrt{N}}\Big) + \frac{1}{2N} \sum_{i,j} L\Big(\frac{2\theta t_i t_j}{\sqrt{N}}\Big).$$

We show that we can replace the supremum by the supremum over measures ν_2 using the change of variable

$$\nu_2 = \frac{1}{N} \sum_{j=1}^{l} \delta_{t_j}.$$

Clearly we have the upper bound,

$$\begin{aligned}
&\sup_{\substack{\delta \leq t_i N^{-1/4} \leq K \\ |\sum_i t_i^2 - \alpha_2 N| \leq \delta N}} \mathcal{M}(l, (t_i, i \leq l)) \\
&\leq \sup_{\substack{\nu_2 \in \mathcal{M}(S_2) \\ |\int x^2 d\nu_2(x) - \alpha_2| \leq \delta}} \Big\{ \sum_{i=1}^{k} \int L\Big(\frac{2\theta s_i x}{\sqrt{N}}\Big) d\nu_2(x) + \frac{1}{2} \int L\Big(\frac{2\theta xy}{\sqrt{N}}\Big) d\nu_2(x) d\nu_2(y) \Big\}.
\end{aligned}$$

We claim that

$$\begin{aligned}
&\sup_{\substack{\nu_2 \in \mathcal{M}(S_2) \\ |\int x^2 d\nu_2(x) - \alpha_2| \leq \delta}} \Big\{ \sum_{i=1}^{k} \int L\Big(\frac{2\theta s_i x}{\sqrt{N}}\Big) d\nu_2(x) + \frac{1}{2} \int L\Big(\frac{2\theta xy}{\sqrt{N}}\Big) d\nu_2(x) d\nu_2(y) \Big\} \\
&= \sup_{\substack{\nu_2 \in \mathcal{M}(S_2) \\ \int x^2 d\nu_2(x) = \alpha_2}} \Big\{ \sum_{i=1}^{k} \int L\Big(\frac{2\theta s_i x}{\sqrt{N}}\Big) d\nu_2(x) + \frac{1}{2} \int L\Big(\frac{2\theta xy}{\sqrt{N}}\Big) d\nu_2(x) d\nu_2(y) \Big\} + o_\delta(1). \qquad (1.66)
\end{aligned}$$

First, we show that the supremum on the right-hand side of the inequality (1.66) can be restricted to the subset of measures with no atoms. Indeed, as $\nu_2 \mapsto \nu_2 \otimes \nu_2$ is continuous for the weak topology, the function

$$\nu_2 \mapsto \sum_{i=1}^{k} \int L\Big(\frac{2\theta s_i x}{\sqrt{N}}\Big) d\nu_2(x) + \frac{1}{2} \int L\Big(\frac{2\theta xy}{\sqrt{N}}\Big) d\nu_2(x) d\nu_2(y$$

is continuous for the weak topology. Therefore we can restrict the supremum on the left-hand side of (1.66) to the dense subset of measures with no atoms, and thus also in the supremum on the right-hand side of (1.66). Fix now $\nu_2 \in \mathcal{M}(S_2)$ without any atoms such that $|\int x^2 d\nu_2 - \alpha_2|$. There exists $t \in (\delta N^{1/4}, KN^{1/4})$ such that

$$\int_{[t, KN^{1/4}]} x^2 d\nu_2(x) = |\alpha_2 - \int x^2 d\nu_2(x)|.$$

If we let $\tilde{\nu}_2 = \nu_2 + \nu_2 \mathbb{1}_{[t, KN^{1/4}]}$, if $\int x^2 d\nu_2(x) \leq \alpha_2$ and $\tilde{\nu}_2 = \nu_2 - \nu_2 \mathbb{1}_{[t, KN^{1/4}]}$ otherwise, we obtain that $\tilde{\nu}_2(S_2) = \alpha_2$ and for any function $f : S_2 \to \mathbb{R}$ such that $f(\sqrt{\cdot})$ is 1-Lipschitz,

$$\int f d\nu(x) \leq \int f d\tilde{\nu}_2 + O(\delta),$$

which proves (1.66).

We now prove that

$$\sup_{\substack{\delta \le t_i N^{-1/4} \le K \\ |\sum_i t_i^2 - \alpha_2 N| \le \delta N}} \mathcal{M}(l,(t_i, i \le l))$$
$$\ge \sup_{\substack{\nu_2 \in \mathcal{M}(S_2) \\ |\int x^2 d\nu_2(x) - \alpha_2| \le \delta}} \Big\{ \sum_{i=1}^k \int L\Big(\frac{2\theta s_i x}{\sqrt{N}}\Big) d\nu_2(x) + \frac{1}{2}\int L\Big(\frac{2\theta xy}{\sqrt{N}}\Big) d\nu_2(x) d\nu_2(y)\Big\} - o_\delta(1).$$

Let $\nu_2 \in \mathcal{M}(S_2)$ such that $\int x^2 d\nu_2(x) = \alpha_2$. As we saw earlier, we can assume without loss of generality that ν_2 does not have any atoms. Let $\delta N^{1/4} = t_0 \le t_2 \le \ldots \le t_{m+1} = KN^{1/4}$ such that for any $1 \le i \le m+1$,

$$t_i = \max\big\{x \in [\delta N^{1/4}, KN^{1/4}] : \nu_2([t_{i-1}, x]) \le \frac{1}{N}\big\}.$$

As ν_2 does not have atoms, the maximum above is well defined. Define $\nu_2^N = \frac{1}{N}\sum_{i=1}^m \delta_{t_i}$. Since ν_2 is supported on $[\delta N^{1/4}, KN^{1/4}]$ and $\int x^2 d\nu_2(x) \le 1$, we have $\nu_2(\mathbb{R}) \le \sqrt{N}\delta^{-2}$, and therefore $m \le \sqrt{N}\delta^{-2}$. One can check that for any $1 \le i \le m$,

$$\sup\big\{\big|\nu_2([\delta N^{1/4}, x]) - \nu_2^N([\delta N^{1/4}, x])\big| : x \in [\delta N^{1/4}, KN^{1/4}]\big\} = O\Big(\frac{1}{N}\Big) \tag{1.67}$$

We deduce that

$$\sup\Big\{\int f d\nu_2 - \int f d\nu_2^N : ||f||_{BV} \le 1\Big\} = O\Big(\frac{1}{N}\Big), \tag{1.68}$$

where $||f||_{BV}$ is the bounded variation norm defined by

$$||f||_{BV} = \sup \sum_{k=1}^m |f(x_{k+1}) - f(x_k)|,$$

where the supremum runs over all finite sequences $(x_i)_{1\le i \le m}$ such that $\delta = x_1 \le x_2 \ldots \le x_m = K$. For any $s_i \in [KN^{1/4}, \sqrt{N\alpha_3}]$, $i \le k$, if we denote by $f_s : x \in S_2 \mapsto \sum_i L(2\theta s_i x/\sqrt{N})$, we get

$$||f_s||_{BV} = \int_{\delta N^{1/4}}^{KN^{1/4}} |f_s'(x)| dx = \frac{2\theta}{\sqrt{N}} \int_{\delta N^{1/4}}^{KN^{1/4}} \sum_{i=1}^k s_i L'\Big(\frac{2\theta s_i x}{\sqrt{N}}\Big) dx.$$

But $L(x) \le Ax^2/2$ for any $x \ge 0$. Therefore, for any $x \ge 0$, $L'(x) \le Cx$, where C is some positive constant, we deduce that $||f_s||_{BV} = O(K^2)$ uniformly in $s_i \in I_3$. Thus, using (1.68), we get

$$\sum_{i=1}^k \int L\Big(\frac{2\theta s_i x}{\sqrt{N}}\Big) d\nu_2(x) - \sum_{i=1}^k \int L\Big(\frac{2\theta s_i x}{\sqrt{N}}\Big) d\nu_2^N(x) = O\Big(\frac{K^2}{N}\Big). \tag{1.69}$$

Let $x \in S_2$ and define $g_x : y \in I_2 \mapsto L(2\theta xy/\sqrt{N})$. A similar argument as above shows that $||g_x||_{BV} = O(K^4/\sqrt{N})$ uniformly in $x \in S_2$, therefore

$$\int L(2\theta xy) d\nu_2(y) - \int L(2\theta xy) d\nu_2^N(y) = O\Big(\frac{K^4}{N^{3/2}}\Big), \tag{1.70}$$

uniformly in $x \in I_2$. From (1.67) and the fact that $\nu_2(I_2) \le \delta^{-2}\sqrt{N}$, we get that $\nu_2^N(I_2) = O(\delta^{-2}\sqrt{N})$. Thus, integrating the above equation alternatively with respect to ν_2 and ν_2^N, we get

$$\int L(2\theta xy) d\nu_2(y) d\nu_2(x) - \int L(2\theta xy) d\nu_2^N(y) d\nu_2^N(x) = O\Big(\frac{K^4}{\delta^2 N}\Big). \tag{1.71}$$

Since by construction $\int x^2 d\nu_2^N(x) \le \int x^2 d\nu_2(x) = \alpha_2$, we deduce from (1.68) that

$$0 \le \alpha_2 - \int x^2 d\nu_2^N = O\Big(\frac{K^2}{\sqrt{N}}\Big).$$

We can conclude using (1.71) and (1.69) that

$$\sum_{i=1}^k \int L\Big(\frac{2\theta s_i x}{\sqrt{N}}\Big) d\nu_2(x) + \frac{1}{2}\int L\Big(\frac{2\theta xy}{\sqrt{N}}\Big) d\nu_2(x) d\nu_2(y) \le \mathcal{M}(m,(t_i, i \le m)) + o_\delta(1),$$

which ends the proof.

□

1.5.4 The large deviations close to the bulk

We prove in this section Proposition 1.5.17. By Theorem 1.5.14, the large deviation lower bound holds at every $x > 2$ such that $I(x) = \bar{I}(x) \neq 0$ so that there exists $\theta \in \Theta_x$ which does not belong to any Θ_y for $y \neq x$. In the following lemma, we prove that if $\overline{F}(\theta) = \underline{F}(\theta) = \theta^2$ on a interval $(0, b)$ with $b > 1/2$, then the large deviation lower bound holds in a neighborhood of 2 and the rate function I is equal to the one of the GOE.

Lemma 1.5.26. *If for $\theta \in \big(0, \frac{1}{2\epsilon}\big)$, for some $\epsilon \in (0,1)$, $\overline{F}(\theta) = F(\theta) = \theta^2$, then for any $x \in [2, \epsilon + \frac{1}{\epsilon})$,*

$$\bar{I}(x) = I(x) = I_{GOE}(x).$$

As a consequence, $\forall x > 2, \bar{I}(x) > 0$. Moreover, for $x \in [2, \epsilon + \frac{1}{\epsilon})$ the optimizer in I is taken at $\theta_x = 1/2G(x)$ and $\theta_x \notin \Theta_y$ for all $y \neq x$.

Proof. As $\underline{F}(\theta) \ge \theta^2$ for any $\theta \ge 0$, we have that

$$\sup_{\theta \in [0, 1/2\epsilon)} \big\{J(\theta, x) - \theta^2\big\} \le I(x) \le \bar{I}(x) \le \sup_{\theta \ge 0} \big\{J(\theta, x) - \theta^2\big\}.$$

But if $x \in [2, \epsilon + \frac{1}{\epsilon})$,

$$I_{GOE}(x) = \sup_{\theta \ge 0} \big\{J(\theta, x) - \theta^2\big\},$$

is achieved at $\theta = 1/2G(x) \in (0, 1/2\epsilon)$ since $G^{-1}(\epsilon) = \epsilon + 1/\epsilon$. Therefore, if $x \in [2, 2\epsilon + \frac{1}{2\epsilon})$, then we obtain

$$I(x) = \bar{I}(x) = I_{GOE}(x).$$

In particular, $I(x) > 0$ for any $x \in (2, 2\epsilon + 1/2\epsilon)$. As I is convex by Proposition 1.5.13 and $I(2) = 0$, we conclude that $I(x) > 0$ for any $x > 2$. Moreover, the optimizer in $I(x)$ is taken at $\theta_x = 1/2G(x)$ for which $\underline{F}(\theta) = \overline{F}(\theta)$ is differentiable and clearly $\theta_x \notin \Theta_y$ for any $y \neq x$ as G is invertible on $[2, +\infty)$.

□

The result of Proposition 1.5.17 then follows from Lemma 1.5.26 and Lemma 1.5.10. We now study the convergence of the annealed spherical integrals for large values of θ, in which case we need to make additional assumptions on μ.

1.5.5 Case where ψ is an increasing function

In this section we make the additional assumption that ψ is non-decreasing.

Example 1.5.27 (Sparse Gaussian distribution)**.** *Let μ be the law of $\xi\Gamma$ where ξ is a Bernoulli variable of pararmeter $p \in (0,1)$ and Γ is a standard Gaussian random variable. In that case we have for any $x \in \mathbb{R}$,*

$$\psi(x) = \frac{\log[(1-p) + p\exp(x^2/2p)]}{x^2} = \int_0^1 \frac{t\exp((xt)^2/2p)}{(1-p) + p\exp((xt)^2/2p)} dt.$$

One can observe that this last expression is indeed increasing in x.

1.5.5.1 Simplification of the variational problem

We prove in this section that when ψ is increasing on $\mathbb{R}^+$, $\mathcal{C}_\mu = \mathbb{R}^+$ and we can simplify the limit $F(\theta)$ as follows.

Proposition 1.5.28. *For any $\theta \geq 0$, $\overline{F}(\theta) = \underline{F}(\theta) = F(\theta)$ where*

$$F(\theta) = \sup_{\alpha\in[0,1]} \sup_{\substack{\nu\in\mathcal{P}(\mathbb{R}) \\ \int x^2 d\nu(x)=\alpha}} \Big\{\theta^2\alpha^2 + B\theta^2(1-\alpha)^2 + \int L(2\theta\sqrt{1-\alpha}x)d\nu(x) - H(\nu) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\Big\},$$

Proof. Recall that we set for any $\delta, K > 0$, and $\alpha_1 + \alpha_2 + \alpha_3 = 1$,

$$\begin{aligned}\mathcal{F}^N_{\alpha_1,\alpha_2,\alpha_3}(\delta, K) = &\sup_{\substack{\nu_2\in\mathcal{M}(I_2) \\ \int x^2 d\nu_2=\alpha_2}} \sup_{\substack{s\in I_3^k, k\geq 1 \\ |\sum s_i^2 - N\alpha_3|\leq\delta N}} \Big\{\theta^2\big(\alpha_1^2 + 2\alpha_1\alpha_2 + B\alpha_3^2\big) \\ &+ \sum_{i=1}^k \int L\Big(\frac{2\theta s_i x}{\sqrt{N}}\Big)d\nu_2(x) + \frac{1}{2}\int L\Big(\frac{2\theta xy}{\sqrt{N}}\Big)d\nu_2(x)d\nu_2(y) \\ &+ \sup_{\substack{\nu_1\in\mathcal{P}(I_1) \\ \int x^2 d\nu_1(x)=\alpha_1}} \Big\{\sum_{i=1}^k \int L\Big(\frac{2\theta s_i x}{\sqrt{N}}\Big)d\nu_1(x) - H(\nu_1)\Big\} - \frac{1}{2}\log(2\pi) - \frac{1}{2}\Big\},\end{aligned}$$

where $I_1 = \{x : |x| \leq \delta N^{1/4}\}$, $I_2 = \{x : \delta N^{1/4} \leq |x| \leq KN^{1/4}\}$, $I_3 = \{KN^{1/4} \leq |x| \leq \sqrt{N\alpha_3}\}$. Since ψ is non-decreasing, we have on one hand for any $s \in I_3^k$ such that $|\sum_i s_i^2 - \alpha_3 N| \leq N\delta$,

$$\begin{aligned}\sum_{i=1}^k \int L\Big(\frac{2\theta s_i x}{\sqrt{N}}\Big)d\nu_2(x) &= \frac{4\theta^2}{N}\sum_{i=1}^k s_i^2 \int x^2\psi\Big(\frac{2\theta s_i x}{\sqrt{N}}\Big)d\nu_1(x) \\ &\leq \int L\Big(2\theta\sqrt{\alpha_3}x\Big)d\nu_1(x) + o_\delta(1).\end{aligned}$$

On the other hand, since ψ is bounded above by $B/2$,

$$\frac{4\theta^2}{N}\sum_{i=1}^k s_i^2 \int x^2\psi\Big(\frac{2\theta s_i x}{\sqrt{N}}\Big)d\nu_2(x) + \frac{1}{2}\int L\Big(\frac{2\theta xy}{\sqrt{N}}\Big)d\nu_2(x)d\nu_2(y) \leq \theta^2 B(\alpha_2^2 + 2\alpha_3\alpha_2) + o_\delta(1).$$

Therefore, we have the upper bound,

$$\begin{aligned}\overline{F}(\theta) \leq &\sup_{\alpha_1+\alpha_2+\alpha_3=1} \sup_{\substack{\nu_1\in\mathcal{P}(\mathbb{R}) \\ \int x^2 d\nu_1(x)=\alpha_1}} \Big\{\theta^2\big(\alpha_1^2 + 2\alpha_1\alpha_2\big) + \theta^2 B(\alpha_3 + \alpha_2)^2 \\ &+ \int L(2\theta\sqrt{\alpha_3}x)d\nu_1(x) - H(\nu_1) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\Big\}.\end{aligned}$$

We can further simplify this optimization problem by showing that the assumption on the monotonicity of ψ entails that we can take $\alpha_2 = 0$. Indeed, note that $\psi(0) = 1/2$. Therefore, for any $x \geq 0$, $\psi(x) \geq 1/2$. Hence, we deduce that

$$\begin{aligned}2\theta^2\alpha_1\alpha_2 + \int L(2\theta\sqrt{\alpha_3}x)d\nu_1(x) &= 2\theta^2\alpha_1\alpha_2 + 4\theta^2\alpha_3\int x^2\psi(2\theta\sqrt{\alpha_3}x)d\nu_1(x) \\ &\leq 4\theta^2(\alpha_2 + \alpha_3)\int x^2\psi(2\theta\sqrt{\alpha_2+\alpha_3}x)d\nu_1(x).\end{aligned}$$

Thus, with the change of variables $\alpha_3 + \alpha_2 \to \alpha_3$, we deduce

$$\overline{F}(\theta) \leq \sup_{\alpha_1+\alpha_3=1} \sup_{\substack{\nu_1\in\mathcal{P}(\mathbb{R}) \\ \int x^2 d\nu_1(x)=\alpha_1}} \Big\{\theta^2\alpha_1^2 + \theta^2 B\alpha_3^2 + \int L(2\theta\sqrt{\alpha_3}x)d\nu_1(x) - H(\nu_1) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\Big\}.$$

To prove that $\underline{F}(\theta)$ is bounded from below by the same quantity, we fix $\alpha_1, \alpha_2, \alpha_3$ such that $\alpha_1+\alpha_2+\alpha_3 = 1, \alpha_2 = 0$, and $\nu_1 \in \mathcal{P}(\mathbb{R})$ such that $\int x^2 d\nu_1(x) = \alpha_1$. We take in the definition of $\mathcal{F}^N_{\alpha_1,\alpha_2,\alpha_3}(K,\delta)$, $k = 1$ and $s = \sqrt{\alpha_3}N^{1/4}$,

$$\tilde{\nu}_1 = \frac{1}{\nu_1(I_1)}\nu_1(.\cap I_1)\circ h^{-1}_{\sqrt{\lambda}}, \text{ and } \nu_2 = \delta_{\sqrt{K}},$$

where $h_{\sqrt{\lambda}} : x \mapsto \sqrt{\lambda}x$, and

$$\frac{1}{\lambda} = \frac{1}{\nu_1(I_1)\alpha_1}\int_{I_1} x^2 d\nu_1(x).$$

We have, on one hand,

$$2\alpha_3 \int \psi(2\theta sx)d\nu_2(x) + \alpha_2^2 \int \psi(2\theta xy)d\nu_2(x)d\nu_2(y) = 0.$$

On the other hand,

$$4\theta^2\alpha_3 \int x^2\psi\Big(\frac{2\theta sx}{N^{1/4}}\Big)d\tilde{\nu}_1(x) - H(\tilde{\nu}_1) = \int L(2\theta\sqrt{\alpha_3}x)d\tilde{\nu}_1(x) - H(\tilde{\nu}_1).$$

Thus,

$$\begin{aligned}\mathcal{F}^N_{\alpha_1,\alpha_2,\alpha_3}(K,\delta) &\geq \theta^2\big(\alpha_1^2 + B\alpha_3\big)\\ &\quad + \int L(2\theta\sqrt{\alpha_3}x)d\tilde{\nu}_1(x) - H(\tilde{\nu}_1) - \frac{1}{2}\log(2\pi) - \frac{1}{2}.\end{aligned}$$

which proves the claim. □

We next simplify the supremum over ν in the definition of F in the last proposition. To this end, we denote by $G : [B/2, +\infty) \to \mathbb{R}\cup\{+\infty\}$ the function given by

$$\forall \zeta \in [B/2, +\infty),\ G(\zeta) = \log \int \exp(L(x) - \zeta x^2)dx\,. \tag{1.72}$$

Lemma 1.5.29. *Let*

$$l = -\lim_{\zeta\to B/2} G'(\zeta) \in (0, +\infty]. \tag{1.73}$$

For any $C \in (0, l)$, there exists a unique $\zeta \in (B/2, +\infty)$ solution to the equation

$$G'(\zeta) = -C\,.$$

It is denoted by ζ_C. For $C \geq l$, we set $\zeta_C = B/2$. Then,

$$\begin{aligned}\sup_{\substack{\nu\in\mathcal{P}(\mathbb{R})\\ \int x^2 d\nu(x) = C}} \Big[\int L(x)d\nu(x) - H(\nu)\Big] &= \sup_{\substack{\nu\in\mathcal{P}(\mathbb{R})\\ \int x^2 d\nu(x)\leq C}} \Big[\frac{BC}{2} + \int \Big(L(x) - \frac{B}{2}x^2\Big)d\nu(x) - H(\nu)\Big]\\ &= C\zeta_C + G(\zeta_C)\end{aligned}$$

Proof. Define the function

$$\forall \nu \in \mathcal{P}(\mathbb{R}),\ E(\nu) = H(\nu) + \int \Big(\frac{B}{2}x^2 - L(x)\Big)d\nu(x).$$

We first show that

$$\inf_{\substack{\nu\in\mathcal{P}(\mathbb{R})\\ \int x^2 d\nu(x) = C}} E(\nu) = \inf_{\substack{\nu\in\mathcal{P}(\mathbb{R})\\ \int x^2 d\nu(x)\leq C}} E(\nu). \tag{1.74}$$

To prove this equality, we will show that for any $\nu \in \mathcal{P}(\mathbb{R})$ such that $\int x^2 d\nu(x) \leq C$, there exists ν_ϵ such that $\int x^2 d\nu_\epsilon(x) = C$, and

$$\lim_{\epsilon \to 0} \nu_\epsilon = \nu, \quad \lim_{\epsilon \to 0} E(\nu_\epsilon) = E(\nu).$$

We set $\nu_\epsilon = (1-\epsilon)\nu + \epsilon\gamma_\epsilon$ where γ_ϵ is a Gaussian measure of variance 1 and mean m_ϵ, defined by,

$$m_\epsilon = \sqrt{\frac{C-(1-\epsilon)D}{\epsilon} - 1}$$

With this choice of m_ϵ, one can check that $\int x^2 d\nu_\epsilon(x) = C$. Moreover,

$$\nu_\epsilon \underset{\epsilon \to 0^+}{\longrightarrow} \nu.$$

for the weak topology. Therefore, for any continuous function $f : \mathbb{R} \to \mathbb{R}$ such that $f = o_{\pm\infty}(x^2)$,

$$\int f d\nu_\epsilon \underset{\epsilon \to 0^+}{\longrightarrow} \int f d\nu.$$

Since $\frac{B}{2}x^2 - L(x) = o_{\pm\infty}(x^2)$, we have

$$\lim_{\epsilon \to 0} \int \Big(\frac{B}{2}x^2 - L(x)\Big) d\nu_\epsilon(x) = \int \Big(\frac{B}{2}x^2 - L(x)\Big) d\nu(x). \tag{1.75}$$

Besides, since H is convex,

$$H(\nu_\epsilon) \leq (1-\epsilon)H(\nu) + \epsilon H(\gamma_\epsilon).$$

But,

$$H(\gamma_\epsilon) = H(\gamma),$$

where γ is a standard Gaussian distribution. Therefore,

$$\limsup_{\epsilon \to 0^+} H(\nu_\epsilon) \leq H(\nu).$$

As H is lower semi-continuous with respect to the weak topology, we can conclude together with (1.75) that,

$$\lim_{\epsilon \to 0^+} E(\nu_\epsilon) = E(\nu),$$

which ends the proof of the claim (1.96). The point is that now we are optimizing an upper semi-continuous function on a compact set and therefore we know that this supremum is achieved. We next identify this maximizer

For any $\zeta \in [B/2, +\infty)$ such that $G(\zeta) < +\infty$, we define ν_ζ the probability measure given by

$$d\nu_\zeta = \frac{\exp(L(x) - \zeta x^2)}{\int \exp(L(y) - \zeta y^2) dy} dx.$$

We will show that

$$\inf_{\substack{\nu \in \mathcal{P}(\mathbb{R}) \\ \int x^2 d\nu(x) \leq C}} E(\nu) = E(\nu_{\zeta_C}). \tag{1.76}$$

Let μ be a probability measure such that $H(\mu) < +\infty$ and $\int x^2 d\mu(x) = D \leq C$. As $H(\mu) < +\infty$, we can write,

$$\mu = (1+\varphi) d\nu_{\zeta_C},$$

where φ is some measurable function such that $\varphi \geq -1$ ν_{ζ_C}-a.s. and $\int \varphi d\nu_{\zeta_C} = 0$. We have,

$$\begin{aligned} E(\mu) = E(\nu_{\zeta_C}) &+ \int \Big(\frac{B}{2}x^2 - L(x)\Big)\varphi(x) d\nu_{\zeta_C}(x) \\ &+ \int (1+\varphi)\log(1+\varphi) d\nu_{\zeta_C} + \int \log \frac{d\nu_{\zeta_C}}{dx} \varphi d\nu_{\zeta_C}, \end{aligned}$$

where we used the fact that $\int \varphi d\nu_{\zeta_C} = 0$. By convexity of $x \mapsto x \log x$, we have

$$\int (1+\varphi)\log(1+\varphi) d\nu_{\zeta_C} \geq (1 + \int \varphi d\nu_{\zeta_C}) \log(1 + \int \varphi d\nu_{\zeta_C}) = 0.$$

Therefore,

$$\begin{aligned} E(\mu) &\geq E(\nu_{\zeta_C}) + \int \left(\frac{B}{2} x^2 - L(x) + \log \frac{d\nu_{\zeta_C}}{dx} \right) \varphi(x) d\nu_{\zeta_C}(x) \\ &\geq E(\nu_{\zeta_C}) \end{aligned}$$

where we used again that $\int \varphi d\nu_{\zeta_C} = 0$. This shows that ν_{ζ_C} achieves the infimum in (1.76), and ends the proof of Lemma 1.5.29. □

1.5.5.2 Differentiability of the limit of the annealed spherical integral

This section is devoted to the proof of the following proposition.

Proposition 1.5.30. *F is continuously differentiable on $(1/\sqrt{B-1}, +\infty)$ except possibly at the point θ_0 such that*

$$\theta_0 = \inf \left\{ \theta : F(\theta) > \theta^2 \right\}.$$

Moreover, for any $\theta \leq 1/2\sqrt{B-1}$,

$$F(\theta) = \theta^2. \tag{1.77}$$

The second part of the claim of the above proposition (1.77) is due to Proposition 1.5.10 and the fact that $A = B$. From now on, we assume that $\theta^2(B-1) > 1$ and wish to prove the first part of Proposition 1.5.30. We define for any $\alpha \in [0,1]$, and $\nu \in \mathcal{P}(\mathbb{R})$,

$$H_\theta(\alpha, \nu) = \theta^2\alpha^2 + \theta^2 B(1-\alpha)^2 + \int L(2\theta\sqrt{1-\alpha}x) d\nu(x) - H(\nu) - \frac{1}{2}\log(2\pi) - \frac{1}{2}. \tag{1.78}$$

By Proposition 1.5.28, we have

$$F(\theta) = \sup_{(\alpha,\nu)\in S} H_\theta(\alpha, \nu), \tag{1.79}$$

where

$$S = \{(\alpha, \nu) \in [0,1] \times \mathcal{P}(\mathbb{R}) : \int x^2 d\nu(x) = \alpha\}.$$

We first show that we can can restrict the set of parameters (α, ν) to $\alpha \in [0, \frac{1}{2}] \cup \{1\}$, as described in the following lemma.

Lemma 1.5.31. *If $\theta^2(B-1) \geq 1$.*

$$F(\theta) = \max\Big(\sup_{\substack{(\alpha,\nu)\in S \\ \alpha \leq \frac{1}{2}}} H_\theta(\alpha, \nu), \theta^2 \Big).$$

Proof. Making the change of variable which consists in replacing ν by the push-forward of ν by the map $x \mapsto x/\sqrt{\alpha}$, we get

$$\sup_{(\alpha,\nu)\in S} H_\theta(\alpha, \nu) = \sup_{\substack{\alpha\in(0,1] \\ \int x^2 d\nu = 1}} \{\tilde{H}_\theta(\alpha, \nu) - H(\nu)\},$$

where for any $\alpha \in (0,1]$, and $\nu \in \mathcal{P}(\mathbb{R})$,

$$\tilde{H}_\theta(\alpha, \nu) = \theta^2\alpha^2 + \theta^2 B(1-\alpha)^2 + \frac{1}{2}\log\alpha + \int L(2\theta\sqrt{(1-\alpha)\alpha}x) d\nu(x) - \frac{1}{2}\log(2\pi) - \frac{1}{2}.$$

We claim that for any $\nu \in \mathcal{P}(\mathbb{R})$ such that $\int x^2 d\nu(x) = 1$,

$$\tilde{H}_\theta(\alpha, \nu) \leq \max\Big(\tilde{H}_\theta\Big(\frac{1}{2}, \nu\Big), \tilde{H}_\theta(1, \nu)\Big). \tag{1.80}$$

Indeed, first notice that since ψ is increasing, for all $\alpha \in [0,1]$

$$2\int x^2 \psi(2\theta\sqrt{\alpha(1-\alpha)}x) d\nu(x) \leq 2\int x^2 \psi(\theta x) d\nu(x)\,.$$

Denote by $m = 2\int x^2\psi(\theta x)d\nu(x)$. As ψ takes its values in $[1/2, B/2]$, $m \in [1, B]$, and we have for any $\alpha \in (0,1]$,

$$\tilde{H}_\theta(\alpha,\nu) \leq \theta^2\alpha^2 + \theta^2 B(1-\alpha)^2 + \frac{1}{2}\log\alpha + 2\theta^2\alpha(1-\alpha)m - \frac{1}{2}\log(2\pi) - \frac{1}{2} =: f_{\theta,m}(\alpha)\,.$$

We find that

$$f'_{\theta,m}\Big(\frac{1}{2}\Big) = \theta^2(1-B) + 1, \quad f''_{\theta,m}(\alpha) = 2\theta^2(B+1-2m) - \frac{1}{2\alpha^2}.$$

We deduce that $f'_{\theta,m}$ is either increasing, decreasing or decreasing and then increasing. Since $f'_{\theta,m}(1/2) \leq 0$, we conclude that

$$\max_{\alpha\in[\frac{1}{2},1]} f_{\theta,m}(\alpha) = \max\Big(f_{\theta,m}\Big(\frac{1}{2}\Big), f_{\theta,m}(1)\Big)\,,$$

which yields the claim (1.80) since $f_{\theta,m}(\alpha) = \tilde{H}_\theta(\alpha,\nu)$ at the two points $\alpha = 1/2$ and 1. To conclude the proof we observe that since $\tilde{H}_\theta(1,\nu) = \theta^2$ for any $\nu \in \mathcal{P}(\mathbb{R})$, we have

$$\sup_{\int x^2 d\nu(x)=1}\{\tilde{H}_\theta(1,\nu) - H(\nu)\} = \theta^2 + \frac{1}{2} + \frac{1}{2}\log(2\pi) - \inf_{\int x^2 d\nu(x)=1} H(\nu) = \theta^2.$$

$\square$

Due to Lemma 1.5.29, we can further simplify the optimisation problem defining $F(\theta)$ in (1.79) by optimizing on $\nu \in \mathcal{P}(\mathbb{R})$ such that $\int x^2 d\nu(x) = \alpha$, given $\alpha \in (0,1)$.

Corollary 1.5.32. *Let R be the function*

$$R : C \in (0, +\infty) \mapsto C\zeta_C + G(\zeta_C),$$

where ζ_C is defined as in Lemma 1.5.29. Denote by

$$K_\theta(\alpha) = \theta^2(\alpha^2 + B(1-\alpha)^2) + R(4\theta^2\alpha(1-\alpha)) - \frac{1}{2}\log(1-\alpha) - \log(2\theta) - \frac{1}{2}\log(2\pi) - \frac{1}{2}.$$

Then, for any $\theta \geq 1/\sqrt{B-1}$,

$$F(\theta) = \sup_{\alpha\in(0,1]} K_\theta(\alpha)\,.$$

Proof. When $\alpha < 1$, we make the following change of variables which consists in replacing ν by its pushforward by $x \mapsto 2\theta\sqrt{1-\alpha}x$. Using (1.65), we find that

$$H(\nu) = \int \log\frac{d\nu}{dx}d\nu = H(\nu_1) - \frac{1}{2}\log(1-\alpha) - \log(2\theta)$$

and ν has covariance $4\alpha(1-\alpha)\theta^2$. Thus,

$$F(\theta) = \max\Big(\theta^2, \sup_{(\alpha,\nu)\in S'} K_\theta(\alpha,\nu)\Big), \tag{1.81}$$

where

$$S' = \{(\alpha,\nu) \in (0,1)\times\mathcal{P}(\mathbb{R}) : \int x^2 d\nu(x) = 4\alpha(1-\alpha)\theta^2\},$$

and for any $\alpha \in (0,1), \nu \in \mathcal{P}(\mathbb{R})$,

$$K_\theta(\alpha,\nu) = \theta^2(\alpha^2 + B(1-\alpha)^2) + \int L(x)d\nu(x) - H(\nu) - \frac{1}{2}\log(1-\alpha) - \log(2\theta) \\ - \frac{1}{2}\log(2\pi) - \frac{1}{2}.$$

By Lemma 1.5.29, we obtain for any $\alpha \in (0,1)$,

$$\sup_{\int x^2 d\nu(x) = 4\alpha(1-\alpha)\theta^2} K_\theta(\alpha,\nu) = \theta^2(\alpha^2 + B(1-\alpha)^2) + 4\theta^2\alpha(1-\alpha)\zeta_{\alpha,\theta} + G(\zeta_{\alpha,\theta}) \\ - \frac{1}{2}\log(1-\alpha) - \log(2\theta) - \frac{1}{2}\log(2\pi) - \frac{1}{2},$$

where $\zeta_{\alpha,\theta} = \zeta_{4\theta^2\alpha(1-\alpha)}$. Hence, if we set, for $\alpha \in (0,1)$,

$$K_\theta(\alpha) = \theta^2(\alpha^2 + B(1-\alpha)^2) + R(4\theta^2\alpha(1-\alpha)) \\ - \frac{1}{2}\log(1-\alpha) - \log(2\theta) - \frac{1}{2}\log(2\pi) - \frac{1}{2},$$

we deduce from (1.81) that

$$F(\theta) = \max\Big(\theta^2, \sup_{\alpha\in(0,1)} K_\theta(\alpha)\Big). \tag{1.82}$$

To study the maximum of K_θ, we will need the following result on the limit of R at 0, which will allow us to compute the limit of K_θ at 1.

Lemma 1.5.33. *When $C \to 0^+$,*

$$R(C) = \frac{1}{2} + \frac{1}{2}\log(2\pi C) + o(1).$$

Proof. For $C < l$, we have

$$G'(\zeta_C) = -C. \tag{1.83}$$

Since we have,

$$\lim_{C\to+\infty} G'(C) = 0,$$

and G' is invertible as $G''(x) > 0$ on $(-\infty, l)$, we get

$$\lim_{C\to 0} \zeta_C = +\infty.$$

From the inequalities, $0 \le L(x) \le Bx^2/2$, we deduce that by the definition (1.72) of G we have the bounds

$$\frac{1}{2}\log\frac{\pi}{\zeta} \le G(\zeta) \le \frac{1}{2}\log\frac{\pi}{\zeta - \frac{B}{2}},$$

which yields,

$$G(\zeta) \sim_{+\infty} \frac{1}{2}\log\frac{\pi}{\zeta}. \tag{1.84}$$

On the other hand, inserting these bounds in the numerator and the denominator of the derivative, we obtain

$$\sqrt{\frac{\zeta - \frac{B}{2}}{\zeta}}\frac{1}{2\zeta} \le -G'(\zeta) \le \sqrt{\frac{\zeta}{\zeta - \frac{B}{2}}}\frac{1}{2(\zeta - \frac{B}{2})}.$$

We deduce, since $\zeta_C \to +\infty$ as $C \to 0$,

$$G'(\zeta_C) = -\frac{1}{2\zeta_C} + o\Big(\frac{1}{\zeta_C}\Big).$$

Therefore, we get from the definition of ζ_C (1.83) that

$$\zeta_C \sim_0 \frac{1}{2C}.$$

Using (1.84), we can conclude that

$$R(C) = \frac{1}{2} + o(1) + \frac{1}{2}\log\frac{\pi}{\frac{1}{2C} + o\Big(\frac{1}{2C}\Big)} = \frac{1}{2} + \frac{1}{2}\log(2\pi C) + o(1).$$

□

From the previous Lemma 1.5.33, we deduce that

$$\lim_{\alpha \to 1} K_\theta(\alpha) = \theta^2,$$

so that we can continuously extend K_θ to 1. Therefore, we can write,

$$F(\theta) = \sup_{\alpha \in (0,1]} K_\theta(\alpha).$$

□

We now study K_θ and show that it is continuously differentiable on $(0,1)$. This amounts to prove that R is continuously differentiable on $(0,1)$. On $(0,l)$, it is clear that R is continuously differentiable due to the implicit function theorem. Indeed, ζ_C is by definition the unique solution of the equation

$$G'(\zeta) = -C,$$

and G is strictly convex. On $(l,+\infty)$, R is an affine function, therefore it is sufficient to prove that

$$\lim_{C \to l^-} R'(C) = \frac{B}{2}. \tag{1.85}$$

We have for any $C < l$,

$$R'(C) = C\partial\zeta_C + \zeta_C + \partial\zeta_C G'(\zeta_C) = \zeta_C,$$

which gives (1.85). We deduce that K_θ is continuously differentiable on $(0,1)$ and

$$\forall \alpha \in (0,1),\ K'_\theta(\alpha) = 2\theta^2(\alpha + B(\alpha-1)) + 4\theta^2\zeta_{\alpha,\theta}(1-2\alpha) + \frac{1}{2(1-\alpha)}.$$

From Lemma 1.5.33, we get,

$$\lim_{\alpha \to 0} K_\theta(\alpha) = -\infty.$$

Thus, the supremum of K_θ on $(0,1]$ is achieved either at 1 or on $(0,1)$. From Lemma 1.5.31, we have

$$F(\theta) = \max\big(K_\theta(1), \sup_{\alpha \le \frac{1}{2}} K_\theta(\alpha)\big).$$

Let us assume that the maximum of K_θ is achieved on $(0,1)$. We deduce that the maximum of K_θ is achieved on $(0,\frac{1}{2}]$ at a critical point since K_θ is differentiable. The critical points α of K_θ satisfy the equation,

$$2\theta^2(\alpha + B(\alpha-1)) + 4\theta^2\zeta_{\alpha,\theta}(1-2\alpha) + \frac{1}{2(1-\alpha)} = 0, \tag{1.86}$$

As $\theta^2(B-1) > 1$, $1/2$ does not satisfy the above equation so that the critical points of K_θ are the $\alpha \neq 1/2$, such that

$$\zeta_{\alpha,\theta} = \frac{2\theta^2(\alpha + B(\alpha-1)) + \frac{1}{2(1-\alpha)}}{4\theta^2(2\alpha-1)} := \varphi(\alpha). \tag{1.87}$$

We have,

$$\varphi(\alpha) = \frac{q(\alpha)}{8\theta^2(1-\alpha)(2\alpha-1)},$$

with

$$q(\alpha) = -4\theta^2(1+B)\alpha^2 + 4\theta^2(1+2B)\alpha - 4\theta^2 B + 1.$$

We find,

$$\varphi(\alpha) \geq \frac{B}{2} \Longleftrightarrow \begin{cases} P(\alpha) \geq 0 & \text{if } \alpha \geq \frac{1}{2} \\ P(\alpha) \leq 0 & \text{if } \alpha \leq \frac{1}{2}, \end{cases}$$

with $P(\alpha) = 4\theta^2(B-1)\alpha^2 - 4\theta^2(B-1)\alpha + 1$. As $\zeta_{\alpha,\theta} \geq B/2$, we obtain that the maximum of K_θ is achieved at $\alpha \in (0, 1/2)$ such that $P(\alpha) \leq 0$. The roots of P are

$$\alpha_\pm = \frac{1 \pm \sqrt{1 - [\theta^2(B-1)]^{-1}}}{2}. \tag{1.88}$$

Thus, the maximum of K_θ is achieved on $[\alpha_-, 1/2]$. We will show that K_θ is strictly concave on $(0, \frac{1}{2})$. Note that,

$$4\theta^2\alpha(1-\alpha) \geq l \Longleftrightarrow \alpha \in [\beta_-, \beta_+],$$

with

$$\beta_\pm = \frac{1 \pm \sqrt{1 - l\theta^{-2}}}{2}.$$

For any $\alpha \in (\beta_-, \frac{1}{2})$, we must have $\zeta_C = B/2$ and therefore

$$K_\theta(\alpha) = \theta^2(\alpha^2 + B(1-\alpha)^2) + 2B\theta^2\alpha(1-\alpha) + G\Big(\frac{B}{2}\Big) - \frac{1}{2}\log(1-\alpha) + C_\theta,$$

where C_θ is some constant depending on θ. Thus, for $\alpha \in (\beta_-, \frac{1}{2})$,

$$K_\theta''(\alpha) = 2\theta^2(1-B) + \frac{1}{2(1-\alpha)^2} < 2\theta^2(1-B) + 2 < 0.$$

For any $\alpha \in (\alpha_-, \beta_-)$, we have

$$\begin{aligned} K_\theta(\alpha) = {} & \theta^2(\alpha^2 + B(1-\alpha)^2) + 4\theta^2\alpha(1-\alpha)\zeta_{\alpha,\theta} + G(\zeta_{\alpha,\theta}) \\ & - \frac{1}{2}\log(1-\alpha) + C_\theta, \end{aligned}$$

where $\zeta_{\alpha,\theta}$ is such that

$$G'(\zeta_{\alpha,\theta}) = -4\theta^2\alpha(1-\alpha).$$

As G is strictly convex, we deduce by the implicit function theorem that $\alpha \in (0, \beta_-) \mapsto \zeta_{\alpha,\theta}$ is differentiable, and we have

$$\partial_\alpha \zeta_{\alpha,\theta} G''(\zeta_{\alpha,\theta}) = -4\theta^2(1-2\alpha).$$

We get that $\partial_\alpha \zeta_{\alpha,\theta} < 0$, for any $\alpha \in (0, \beta_-)$. Therefore, for $\alpha \in (0, \beta_-)$, we obtain

$$K_\theta''(\alpha) = 2\theta^2(B+1) - 8\theta^2\zeta_{\alpha,\theta} + 4\theta^2\partial_\alpha\zeta_{\alpha,\theta}(1-2\alpha) + \frac{1}{2(1-\alpha)^2}.$$

Using that $\zeta_{\alpha,\theta} > B/2$ and that $\partial_\alpha \zeta_{\alpha,\theta} < 0$ for $\alpha \in (0, \beta_-)$, we deduce,

$$\begin{aligned}\forall \alpha \in (0,\beta_-),\ K''_\theta(\alpha) &\leq 2\theta^2(B+1) - 4\theta^2 B + \frac{1}{2(1-\alpha)^2}\\ &< 2\theta^2(1-B) + 2 \leq 0.\end{aligned}$$

Thus, K'_θ is decreasing on $(0,\beta_-)$ and $(\beta_-, \frac{1}{2})$. Since K'_θ is continuous, we deduce that K'_θ is decreasing on $(0, \frac{1}{2})$ and K_θ is strictly concave on $(0, \frac{1}{2})$. Therefore, the maximum is achieved at a unique point on $(0, \frac{1}{2})$ which we denote by α_θ. We distinguish two cases.

1^{st} **case:** $l \leq \frac{1}{B-1}$**.**

We have,

$$\beta_- \leq \alpha_- \leq \alpha_+ \leq \beta_+.$$

We know that on one hand $P(\alpha_-) = 0$, so that

$$\varphi(\alpha_-) = \frac{B}{2}.$$

On the other hand $\zeta_{\alpha_-,\theta} = B/2$ since $\alpha_- \in [\beta_-, \beta_+]$. We deduce by (1.87) that α_- is a critical point of K_θ which lie in $(0, \frac{1}{2})$. Therefore

$$\alpha_\theta = \alpha_-.$$

2^{nd} **case:** $l > \frac{1}{B-1}$**.**

We have,

$$\alpha_- < \beta_- < \beta_+ < \alpha_+.$$

Note that $0 \leq \alpha_- < \frac{1}{2} < \alpha_+ \leq 1$. Since $\varphi(\alpha) \neq B/2$ for any $\alpha \in [\beta_-, \beta_+]^c$, we deduce that $\alpha_\theta \in [\alpha_-, \beta_-)$, and in particular $K''_\theta(\alpha_\theta) < 0$. We deduce by the implicit function theorem that $\theta \mapsto \alpha_\theta$ is C^1, and therefore $\theta \mapsto K_\theta(\alpha_\theta)$ is continuously differentiable on $(1/\sqrt{B-1}, +\infty)$.

In conclusion, we have shown that for any $\theta^2(B-1) \geq 1$, if $l \leq \frac{1}{B-1}$,

$$F(\theta) = \max\big(\theta^2, K_\theta(\alpha_-)\big),$$

where α_- is defined in (1.88), whereas if $l \geq \frac{1}{B-1}$,

$$F(\theta) = \max\big(\theta^2, K_\theta(\alpha_\theta)\big),$$

where α_θ is the unique solution in $(0, \beta_-)$ such that

$$G'(\zeta_{\alpha,\theta}) = -4\theta^2\alpha(1-\alpha).$$

Moreover, for $\theta \in (0, 1/\sqrt{2(B-1)})$, we have by Lemma 1.5.10

$$F(\theta) = \theta^2.$$

To conclude that F is continuously differentiable on $\mathbb{R}\backslash[1/(2\sqrt{B-1}), 1/\sqrt{B-1}]$ except at most at one point, we show that there exists θ_0 such that

$$\forall \theta \leq \theta_0,\ F(\theta) = \theta^2, \text{ and } \forall \theta > \theta_0,\ F(\theta) > \theta^2.$$

Since $F(\theta) \geq \theta^2$ for any $\theta \geq 0$, it suffices to prove that $\theta \mapsto F(\theta) - \theta^2$ is non-decreasing. Recall that

$$F(\theta) = \lim_{N\to+\infty} F_N(\theta),$$

where

$$F_N(\theta) = \frac{1}{N} \log \mathbb{E}_e \exp\Big(\sum_i L\big(\sqrt{2N}\theta e_i^2\big) + \sum_{i<j} L\big(2\sqrt{N}\theta e_i e_j\big)\Big),$$

and e is uniformly sampled on $\mathbb{S}^{N-1}$. Therefore,

$$F_N(\theta) - \theta^2 = \frac{1}{N} \log \mathbb{E}_e \exp\Big(\sum_i 2N\theta^2\Big(\psi\big(\sqrt{2N}\theta e_i^2\big) - \frac{1}{2}\Big)e_i^4 + \sum_{i<j} 4N\theta^2\Big(\psi\big(2\sqrt{N}\theta e_i e_j\big) - \frac{1}{2}\Big)e_i^2 e_j^2\Big).$$

As ψ is increasing and $\psi(0) = 1/2$, $\theta \mapsto F_N(\theta) - \theta^2$ is non-decreasing, and therefore $\theta \mapsto F(\theta) - \theta^2$ is non-decreasing as well.

Proposition 1.5.34. *Suppose that $\theta_0 = \sup\{\theta \in \mathbb{R}^+ : F(\theta) = \theta^2\} \geq 1/\sqrt{B-1}$. Then F is not differentiable at θ_0.*

Proof. We let $\zeta(\theta) = \zeta_{\alpha_{\theta_0},\theta_0}$ and as in the proof of Lemma 1.5.29 $\nu_\theta = \nu_{\zeta(\theta)}$. We also let :

$$H^1_\theta(\alpha,\nu) = \theta^2(\alpha^2 + B(1-\alpha)^2) + \frac{1}{2}\log\alpha + \int \overline{L}(2\theta\sqrt{1-\alpha}x)d\nu(x) + \\ 2\theta^2\alpha(1-\alpha)B + H(\nu) - \frac{1}{2}\log 2\pi - \frac{1}{2} \quad (1.89)$$

where $\overline{L}(x) = L(x) - Bx^2/2$. Using again the proof of Lemma 1.5.29 we have that :

$$K_\theta(\alpha) = \max_{\substack{\nu\in\mathcal{P}(\mathbb{R}^+)\\ \int x^2 d\nu(x)\leq\alpha}} H^1_\theta(\alpha,\nu).$$

We have for $\theta^2(B-1) \geq 1$:

$$K_\theta(\alpha_\theta) = H^1_\theta(\alpha_\theta,\nu_\theta).$$

And so for $\theta \geq \theta_0$:

$$F(\theta) = K_\theta(\alpha_\theta) = \sup_{\alpha\in]0,1/2[} K_\theta(\alpha) \geq H^1_\theta(\alpha_{\theta_0},\nu_{\theta_0}).$$

For f a convex function, we denote $\partial^+ f$ the right derivative and $\partial^- f$ the left derivative. For $\theta \geq \theta_0$, we have if g is the function $\theta \mapsto H^1_\theta(\alpha_{\theta_0},\nu_{\theta_0})$ that :

$$\begin{aligned}
\partial^+ g(\theta) &= 2\theta(\alpha_{\theta_0}^2 + B(1-\alpha_{\theta_0})^2) + 4\theta\alpha_{\theta_0}(1-\alpha_{\theta_0})\left[2\alpha_{\theta_0}^{-1}\int x^2\overline{\psi}(2\theta\sqrt{1-\alpha_{\theta_0}}x)d\nu_{\theta_0}(x) + B\right] \\
&\quad + 4\theta^2\alpha_{\theta_0}(1-\alpha_{\theta_0})^{3/2}\int x^2\psi'(2\theta\sqrt{1-\alpha_{\theta_0}}x)d\nu_{\theta_0}(x) \\
&\geq 2\theta(\alpha_{\theta_0}^2 + B(1-\alpha_{\theta_0})^2) + 4\theta\alpha_{\theta_0}(1-\alpha_{\theta_0}) \\
&\geq 2\theta + 2(B-1)(1-\alpha_{\theta_0})^2\theta.
\end{aligned}$$

Where we denote $\overline{\psi} = \psi - B/2$ and where we used that $\psi' \geq 0$. We have $\alpha_{\theta_0} < 1/2$, we have $\partial^+ g(\theta_0) > 2\theta$. For $\theta \leq \theta_0$, $F(\theta^2) = \theta^2$ and so $\partial^- F(\theta_0) = 2\theta$. And so, we have :

$$\partial^+ F(\theta_0) \geq \partial^+ g(\theta_0) > \partial^- F(\theta_0)$$

□

1.5.5.3 Proof of Proposition 1.5.15

By Proposition 1.5.30, we know that F is differentiable on $(1/\sqrt{B-1}, +\infty)$ except possibly at θ_0. Using Proposition 1.5.19 we deduce that there exists x_μ such that the lower large deviation lower bound holds with rate function $I(x) = \overline{I}(x)$ for any $x \geq x_\mu$.

1.5.6 The case $B < A$

We consider the case where B exists and is strictly inferior to A and assume that ψ achieves its maximum A at a unique point m^* such that $\psi''(m_*) < 0$. This includes in particular the case where the law of the entries have a compact support (since in this case $B = 0$) and the maximum of ψ is non-degenerated.

Studying the variational problem arising from the limit of the annealed spherical integral $\overline{F}(\theta)$ and $\underline{F}(\theta)$ defined in Proposition 1.5.8, we will show that for θ large enough we can give an explicit formula as stated in the following proposition.

Proposition 1.5.35. *There exists $\theta_0 > 1/\sqrt{A-1}$ such that for any $\theta \geq \theta_0$, $\overline{F}(\theta) = \underline{F}(\theta) = F(\theta)$ where*

$$F(\theta) = \sup_{\alpha \in (0,1]} V(\alpha),$$

with

$$\forall \alpha > 0, \;\; V(\alpha) = \theta^2(A-1)\alpha^2 + \theta^2 + \frac{1}{2}\log(1-\alpha).$$

More explicitly,

$$F(\theta) = \frac{\theta^2}{4}(A-1)\Big(1 + \sqrt{1 - \frac{1}{\theta^2(A-1)}}\Big)^2 + \theta^2 + \frac{1}{2}\log\Big(1 - \sqrt{1 - \frac{1}{\theta^2(A-1)}}\Big) - \frac{1}{2}\log 2.$$

1.5.6.1 Conditions on the optimizers

With the notation of Proposition 1.5.8, we perform the change of variable ν_2 into μ_2 defined as the unique measure such that for any bounded continuous function $f : \mathbb{R} \to \mathbb{R}$,

$$\frac{1}{\alpha_2}\int x^2 f\Big(\frac{\sqrt{2\theta} x}{N^{1/4}}\Big) d\nu_2(x) = \int f d\mu_2.$$

By Proposition 1.5.8, we have the following formulation of the limits $\overline{F}(\theta)$ and $\underline{F}(\theta)$. Recall that $\psi(x) = \frac{L(x)}{x^2}$.

$$\overline{F}(\theta) = \limsup_{\substack{\delta \to 0, K \to +\infty \\ \delta K \to 0}} \sup_{\substack{\alpha_1+\alpha_2+\alpha_3=1 \\ \alpha_i \geq 0}} \limsup_{N \to +\infty} \mathcal{F}^N_{\alpha_1,\alpha_2,\alpha_3}(\delta, K),$$

$$\underline{F}(\theta) = \sup_{\substack{\alpha_1+\alpha_2+\alpha_3=1 \\ \alpha_i \geq 0}} \liminf_{\substack{\delta \to 0, K \to +\infty \\ \delta K \to 0}} \limsup_{N \to +\infty} \mathcal{F}^N_{\alpha_1,\alpha_2,\alpha_3}(\delta, K),$$

where

$$\begin{aligned}\mathcal{F}^N_{\alpha_1,\alpha_2,\alpha_3}(\delta, K) = \sup_{\mu_2 \in \mathcal{P}(S_2)} & \sup_{\substack{s \in S_3^k, k \geq 1 \\ |\sum_i s_i^2 - \alpha_3\sqrt{N}| \leq \delta\sqrt{N}}} \Big\{\theta^2\big(\alpha_1^2 + 2\alpha_1\alpha_2 + B\alpha_3^2\big) \\ &+ \frac{4\theta^2\alpha_2}{\sqrt{N}}\sum_i s_i^2 \int \psi\big(\sqrt{2\theta} s_i x\big) d\mu_2(x) + 2\theta^2\alpha_2^2 \int \psi(xy) d\mu_2(x) d\mu_2(y) \\ &+ \sup_{\substack{\nu_1 \in \mathcal{P}(I_1) \\ \int x^2 d\nu_1(x) = \alpha_1}} \Big\{\frac{4\theta^2}{\sqrt{N}}\sum_i s_i^2 \int x^2 \psi\Big(\frac{2\theta s_i x}{N^{\frac{1}{4}}}\Big) d\nu_1(x) - H(\nu_1)\Big\} - \frac{1}{2}\log(2\pi) - \frac{1}{2}\Big\},\end{aligned}$$

where $I_1 = \{x : |x| \leq \delta N^{1/4}\}$, $S_2 = \{x : \delta/\sqrt{\theta} \leq |x| \leq K/\sqrt{\theta}\}$, $S_3 = \{x : K \leq |x| \leq N^{1/4}\sqrt{\alpha_3}\}$. To study the regime zhere θ is large, we will consider a simplifed variational problem $\hat{F}(\theta)$, defined as

$$\hat{F}(\theta) = \sup_{\substack{\alpha_1+\alpha_2+\alpha_3=1 \\ \alpha_i \geq 0}} \limsup_{\substack{\delta \to 0, K \to +\infty \\ \delta K \to 0}} \limsup_{N \to +\infty} \hat{\mathcal{F}}^N_{\alpha_1,\alpha_2,\alpha_3}(\delta, K),$$

where

$$\begin{aligned}\hat{\mathcal{F}}^N_{\alpha_1,\alpha_2,\alpha_3}(\delta,K) = &\sup_{\mu_2\in\mathcal{P}(S_2)}\sup_{s\in S_3}\Big\{\theta^2\big(\alpha_1^2+2\alpha_1\alpha_2+B\alpha_3^2\big)\\ &+4\theta^2\alpha_3\alpha_2\int\psi(\sqrt{2\theta}sx)d\mu_2(x)+2\theta^2\alpha_2^2\int\psi(xy)d\mu_2(x)d\mu_2(y)\\ &+\sup_{\substack{\nu_1\in\mathcal{P}(I_1)\\ \int x^2d\nu_1(x)=\alpha_1}}\Big\{4\theta^2\alpha_3\int x^2\psi\Big(\frac{2\theta sx}{N^{\frac14}}\Big)d\nu_1(x)-H(\nu_1)\Big\}-\frac12\log(2\pi)-\frac12\Big\}.\end{aligned}$$

One can observe that for any $\theta\geq 0$,

$$\overline{F}(\theta)\leq\hat{F}(\theta). \tag{1.90}$$

Indeed, for any $\mu_2\in\mathcal{P}(S_2)$, $\nu_1\in\mathcal{P}(I_1)$, and $s\in S_3^k$ such that $\sum_i s_i^2=\alpha_3\sqrt{N}$,

$$\frac{1}{\sqrt{N}}\sum_i s_i^2\int\psi(\sqrt{2\theta}s_ix)d\mu_2(x)+\frac{1}{\sqrt{N}}\sum_i s_i^2\int x^2\psi\Big(\frac{2\theta s_ix}{N^{\frac14}}\Big)d\nu_1(x)\leq\alpha_3\int\psi(\sqrt{2\theta}sx)d\mu_2(x),$$

where s is a maximizer of the function

$$s\in S_3\mapsto\alpha_2\int\psi(\sqrt{2\theta}sx)d\mu_2(x)+\int x^2\psi\Big(\frac{2\theta sx}{N^{1/4}}\Big)d\nu_1(x).$$

However, one can notice that even if one takes $k=1$ in the definition of $\underline{F}(\theta)$, it is not possible in general to provide a similar variational formulation for $\underline{F}(\theta)$. Still, we will see that under our assumptions that $B<A$ and that the maximum of ψ is uniquely achieved at m^* such that $\psi''(m^*)<0$, the upper bound $\hat{F}(\theta)$ is sharp when θ is large.

The starting point of our analysis of the variational problem defining $\hat{F}(\theta)$ in the regime where θ is large is the fact that $\underline{F}(\theta)$ and $\hat{F}(\theta)$ behave like $A\theta^2$. More precisely, we know from (1.52) that there exists $\theta_0>0$ depending on A for all $\theta\geq\theta_0$,

$$\hat{F}(\theta)\geq\underline{F}(\theta)\geq A\theta^2-\kappa\log\theta, \tag{1.91}$$

where $\kappa>0$ is a numerical constant.

As a consequence, we can localize the suprema over $(\alpha_1,\alpha_2,\alpha_3)$ and μ_2 in the definitions of $\hat{F}(\theta)$ in some subset of the constraint set, denoted by $\mathcal{S}$, and defined as follow,

$$\mathcal{S}=\Big\{(\underline{\alpha},\mu_2)\in[0,1]^3\times\mathcal{P}(S_2):\alpha_1+\alpha_2+\alpha_3=1\Big\}.$$

Lemma 1.5.36. *There exists a constant $\theta_0>0$ depending on A such that for any $\theta\geq\theta_0$, the suprema defining $\hat{F}(\theta)$ can be restricted to the set $\mathcal{A}_\theta\times\mathcal{B}_\theta\subset\mathcal{S}$ defined by,*

$$\underline{\alpha}\in\mathcal{A}_\theta\Longleftrightarrow\alpha_2\geq 1-\frac{C\sqrt{\log\theta}}{\theta},\ \alpha_1\leq\frac{C\sqrt{\log\theta}}{\theta},\ \alpha_3\leq\frac{C\sqrt{\log\theta}}{\theta},$$

and

$$\mu_2\in\mathcal{B}_\theta\Longleftrightarrow\int\Big(\frac{A}{2}-\psi(xy)\Big)d\mu_2(x)d\mu_2(y)\leq\frac{C\log\theta}{\theta^2}.$$

where C is a some positive constant depending also on A.

Proof. From (1.91) we deduce that we can restrict the suprema in the definitions of $\hat{F}(\theta)$ to the parameters $\underline{\alpha},s,\nu_1,\mu_2$ with $\alpha_1+\alpha_2+\alpha_3=1$, $s\in S_3$, $\int x^2d\mu_1(x)=\alpha_1$ such that,

$$\begin{aligned}&(A-1)(\alpha_1^2+2\alpha_1\alpha_2)+(A-B)\alpha_3^2+4\alpha_2\alpha_3\int\Big(\frac{A}{2}-\psi(2\theta sy)\Big)d\mu_2(y)\\ &+2\alpha_2^2\int\Big(\frac{A}{2}-\psi(2\theta xy)\Big)d\mu_2(y)d\mu_2(x)+4\alpha_3\int y^2\Big(\frac{A}{2}-\psi(2\theta sy)\Big)d\nu_1(y)\\ &+\frac{1}{\theta^2}\big(H(\nu_1)+\log\sqrt{2\pi}\big)\leq\frac{2\kappa\log\theta}{\theta^2}.\end{aligned}$$

But

$$4\alpha_3 \int y^2\Big(\frac{A}{2} - \psi(2\theta s y)\Big) d\nu_1(y) + \frac{1}{\theta^2}\big(H(\nu_1) + \log\sqrt{2\pi}\big) \geq \frac{1}{2}\log\frac{1}{\alpha_1} \geq 0.$$

Therefore,

$$(A-1)(\alpha_1^2 + 2\alpha_1\alpha_2) + (A-B)\alpha_3^2 + 4\alpha_2\alpha_3 \int y^2\Big(\frac{A}{2} - \psi(2\theta s y)\Big) d\mu_2(y)$$
$$+ 2\alpha_2^2 \int \Big(\frac{A}{2} - \psi(xy)\Big) d\mu_2(y) d\mu_2(x) \leq \frac{2\kappa\log\theta}{\theta^2}.$$

Since each term is non-negative, they are all bounded by $2\kappa\log\theta/\theta^2$. Note that this first yields with $C = 2\kappa/\min\{(A-B), A-1\}$,

$$\alpha_1^2 \leq \frac{C\log\theta}{\theta^2},\ \alpha_3^2 \leq \frac{C\log\theta}{\theta^2},\ \alpha_1\alpha_2 \leq \frac{C\log\theta}{\theta^2}. \tag{1.92}$$

The two first estimates imply since $\alpha_2 = 1 - \alpha_1 - \alpha_3$,

$$\alpha_2 \geq 1 - 2\frac{\sqrt{C\log\theta}}{\theta}.$$

□

In the next lemma we prove that the optimizers of the optimization problem over μ_2 are concentrated around $\sqrt{m_*}$ if ψ takes its maximum value at m^* only.

Lemma 1.5.37. *Assume that ψ achieves its maximum value at m^* only and that it is strictly concave in an open neighborhood of this point. There exists $\epsilon_0 > 0$ such that for any $\mu_2 \in \mathcal{B}_\theta$,*

$$\forall 0 < \epsilon < \epsilon_0,\ \mu_2\big(|x - \sqrt{m^*}| \geq \epsilon\big) \leq \frac{C\sqrt{\log\theta}}{\theta\epsilon},$$

where C is a positive constant depending on ψ.

Proof. Let $\mu_2 \in \mathcal{B}_\theta$. By Lemma 1.5.36 we have,

$$\int \Big(\frac{A}{2} - \psi(xy)\Big) d\mu_2(x) d\mu_2(y) \leq \frac{C\log\theta}{\theta^2}.$$

Since ψ is strictly concave in a neighborhood of m_*, and m_* is its unique maximizer, we deduce that there exists $\eta_0 > 0$ such that for all $0 < \eta < \eta_0$,

$$\forall |x - m_*| \in [\sqrt{\eta}, \sqrt{\eta_0}], \quad \frac{A}{2} - \psi(x) \geq \eta/c,$$

for some constant $c > 0$. As ψ is analytic, it admits a finite number of local maximum. Therefore, we can find $\eta_0 > 0$ such that for all $0 < \eta < \eta_0$,

$$\forall |x - m_*| \geq \sqrt{\eta}, \quad \frac{A}{2} - \psi(x) \geq \eta/c,$$

Since $\frac{A}{2} - \psi$ is non-negative, we obtain

$$\forall \eta < \eta_0,\ \mu_2^{\otimes 2}\big(|xy - m^*| \geq \sqrt{\eta}\big) \leq \frac{C'\log\theta}{\eta\theta^2},$$

where $C' \geq 1$ is a constant depending on ψ. For $\epsilon < \sqrt{m_*}/2$, we have

$$\mu_2([0, \sqrt{m_*} - \epsilon])^2 \leq \mu_2^{\otimes 2}\big(xy \leq m_* - \sqrt{m_*}\epsilon\big).$$

Therefore, for ϵ small enough,

$$\mu_2([0, \sqrt{m_*} - \epsilon])^2 \leq \frac{C'\log\theta}{m_*\theta^2\epsilon^2}.$$

On the other hand,

$$\mu_2\big([\sqrt{m_*} + \epsilon, +\infty)\big)^2 \leq \mu_2^{\otimes 2}\big(xy \geq m_* + 2\sqrt{m_*}\epsilon\big).$$

Using a union bound, we obtain the claim. □

1.5.6.2 Resolution of the variational problem for large θ

Using Lemma 1.5.37, we will show that the optimization problem over μ_2 is asymptotically solved by $\delta_{\sqrt{m_*}}$, with an error which vanishes when K, the lower boundary point of S_3, goes to $+\infty$.

Lemma 1.5.38. *There exists θ_0 depending on ψ such that for $\theta \geq \theta_0$, such that for any $\underline{\alpha} \in \mathcal{A}_\theta$ and $s \in S_3$, if the measure μ_2 that realizes the maximum:*

$$\sup_{\mu_2 \in \mathcal{P}(S_2)} \Big\{ 2\alpha_3 \int \psi\big(\sqrt{2\theta} sx\big) d\mu_2(x) + \alpha_2 \int \psi(xy) d\mu_2(x) d\mu_2(y) \Big\}$$

is in $\mathcal{B}_\theta$, then

$$\sup_{\mu_2 \in \mathcal{B}_\theta} \Big\{ 2\alpha_3 \int \psi\big(\sqrt{2\theta} sx\big) d\mu_2(x) + \alpha_2 \int \psi(xy) d\mu_2(x) d\mu_2(y) \Big\} = \frac{A\alpha_2}{2} + \frac{B\alpha_3}{2} + o_K(1).$$

Proof. Letting $\overline{\psi}(x) = \psi(x) - \frac{B}{2}$, it is equivalent to show that :

$$\sup_{\mu_2 \in \mathcal{B}_\theta} \Big\{ 2\alpha_3 \int \overline{\psi}\big(\sqrt{2\theta} sx\big) d\mu_2(x) + \alpha_2 \int \overline{\psi}(xy) d\mu_2(x) d\mu_2(y) \Big\} = \frac{(A-B)\alpha_2}{2} + o_K(1).$$

Let us fix $\theta \geq \theta_0$ where θ_0 is given by Lemma 1.5.36. Observe that since ψ is bounded continuous,

$$Z : \mu \in \mathcal{P}(S_2) \mapsto 2\alpha_3 \int \overline{\psi}\big(\sqrt{2\theta} sx\big) d\mu(x) + \alpha_2 \int \overline{\psi}(xy) d\mu(x) d\mu(y)$$

achieves its maximum value in the closed set $\mathcal{B}_\theta$. Let μ_2 be an optimizer, and therefore a critical point of this function. Writing that $Z(\mu_2) \geq Z(\mu_2 + \epsilon\nu)$ for all measures ν on S_2 such that $\mu_2 + \epsilon\nu$ is a probability measure for small ϵ, we deduce that there exists a constant $C > 0$ such that,

$$\forall x \in S_2, \ \alpha_3 \overline{\psi}(\sqrt{\theta} sx) + \alpha_2 \int \overline{\psi}(xy) d\mu_2(y) \leq C, \tag{1.93}$$

with equality μ_2-almost surely. Using Lemma 1.5.37, we get for any ϵ small enough,

$$\int \overline{\psi}(xy) d\mu_2(y) = \overline{\psi}(\sqrt{m_*} x) + \int_{[\sqrt{m_*}-\epsilon, \sqrt{m_*}+\epsilon]} \big(\overline{\psi}(xy) - \overline{\psi}(\sqrt{m_*} x)\big) d\mu_2(x) + O\Big(\frac{\sqrt{\log\theta}}{\theta\epsilon}\Big).$$

Where we notice that our $O\Big(\frac{\sqrt{\log\theta}}{\theta\epsilon}\Big)$ is a function that does not depend on δ, K or N. As L is the log-Laplace transform of a sub-Gaussian distribution, we have that $|L'|$ is bounded. In particular, $|\psi'|$ is bounded and thus ψ is Lipschitz. Therefore, for any $x \leq M$,

$$\int \overline{\psi}(xy) d\mu_2(y) = \overline{\psi}(\sqrt{m^*} x) + O\Big(\epsilon M + \frac{\sqrt{\log\theta}}{\theta\epsilon}\Big).$$

Again, the $O\Big(\epsilon M + \frac{\sqrt{\log\theta}}{\theta\epsilon}\Big)$ does not depend on δ, K or N. We choose $\epsilon = \theta^{-1/2}$ and $M = \theta^{1/4}$ so that the two error term above goes to zero when θ goes to ∞, so that we have for any $x \geq 0$,

$$\int \overline{\psi}(xy) d\mu_2(y) = \overline{\psi}(\sqrt{m_*} x) + o_\theta(1). \tag{1.94}$$

In particular,

$$\alpha_3 \overline{\psi}(\sqrt{\theta} sx) + \alpha_2 \int \overline{\psi}(xy) d\mu_2(y) = \overline{\psi}(\sqrt{m_*} x) + o_\theta(1).$$

Taking $x = \sqrt{m_*}$ in (1.93), we get

$$C \geq \frac{A-B}{2} + o_\theta(1), \tag{1.95}$$

since $s \geq K$ and $1-\alpha_2 \leq O(\frac{\sqrt{\log\theta}}{\theta})$. The terms $o_\theta(1)$ above do not depend on K, δ or N. We claim that there exists θ_0 such that for any $\theta \geq \theta_0$,

$$\mu_2([0, \sqrt{m_*}/2]) = 0.$$

Indeed, if $x \leq \sqrt{m_*}/2$, we have by (1.94) and the fact that α_2 goes to 1 as θ goes to infinity,

$$\alpha_3 \overline{\psi}(\sqrt{\theta} s x) + \alpha_2 \int \overline{\psi}(xy) d\mu(y) \leq \sup_{t \leq \sqrt{m_*}/2} \overline{\psi}(\sqrt{m_*} t) + o_\theta(1),$$

with $\sup_{t\leq\sqrt{m_*}/2} \overline{\psi}(\sqrt{m_*}t) < (A-B)/2$ since the maximum of ψ is uniquely achieved at m_*. From (1.95) and the fact that equality in (1.93) holds μ_2-a.s, we deduce that for θ large enough (and not depending on δ, K or N) $[0, \sqrt{m_*}/2] \cap \mathrm{supp}(\mu_2) = \emptyset$. Therefore,

$$\begin{aligned}
2\alpha_3 \int \overline{\psi}\big(\sqrt{2\theta} s x\big) d\mu_2(x) + \alpha_2 \int \overline{\psi}(xy) d\mu_2(x) d\mu_2(y) &\leq \frac{(A-B)\alpha_2}{2} + 2 \sup_{y \geq K\frac{\sqrt{m_*\theta}}{2}} \overline{\psi}(y) \\
&= \frac{(A-B)\alpha_2}{2} + o_K(1).
\end{aligned}$$

Thus,

$$\sup_{\mu_2 \in \mathcal{B}_\theta} \Big\{ 2\alpha_3 \int \overline{\psi}\big(\sqrt{2\theta} s x\big) d\mu_2(x) + \alpha_2 \int \overline{\psi}(xy) d\mu_2(x) d\mu_2(y) \Big\} \leq \frac{(A-B)\alpha_2}{2} + o_K(1).$$

The reverse inequality is achieved by taking $\mu_2 = \delta_{\sqrt{m_*}}$, which completes the proof. □

We deduce that for θ large enough, $\overline{F}(\theta) = \underline{F}(\theta) = F(\theta)$, where F is a function described in the following proposition.

Proposition 1.5.39. *There exists θ_0 depending on ψ such that for any $\theta \geq \theta_0$, $\overline{F}(\theta) = \underline{F}(\theta) = F(\theta)$, where*

$$F(\theta) = \sup_{(\underline{\alpha}, s, \nu) \in \mathcal{S}'} \mathcal{F}(\underline{\alpha}, s, \nu),$$

with

$$\begin{aligned}
\mathcal{F}(\underline{\alpha}, s, \nu) = \theta^2\big(\alpha_1^2 + 2\alpha_1\alpha_2\big) + \theta^2 A \alpha_2^2 + \theta^2 B(\alpha_3^2 + 2\alpha_3\alpha_2) \\
+ 4\theta^2 \alpha_3 \int x^2 \psi(2\theta s \sqrt{\alpha_3} x) d\nu(x) - H(\nu) - \frac{1}{2}\log(2\pi) - \frac{1}{2},
\end{aligned}$$

and

$$\mathcal{S}' = \Big\{ (\underline{\alpha}, s, \nu) \in [0,1]^3 \times [0,1] \times \mathcal{P}(\mathbb{R}) : \alpha_1 + \alpha_2 + \alpha_3 = 1, \int x^2 d\nu(x) = \alpha_1 \Big\}.$$

Proof. By Lemmas 1.5.36 and 1.5.38, we know that

$$\hat{F}(\theta) = \sup_{\underline{\alpha} \in \mathcal{A}_\theta} \limsup_{\substack{\delta \to 0, K \to +\infty \\ \delta K \to 0}} \limsup_{N \to +\infty} \hat{\mathcal{F}}^N_{\alpha_1,\alpha_2,\alpha_3}(\delta, K),$$

where

$$\begin{aligned}
\hat{\mathcal{F}}^N_{\alpha_1,\alpha_2,\alpha_3}(\delta, K) = \sup_{s \in S_3} \sup_{\substack{\nu_1 \in \mathcal{P}(I_1) \\ \int x^2 d\nu_1(x) = \alpha_1}} \Big\{ 4\theta^2 \alpha_3 \int x^2 \psi\Big(\frac{2\theta s x}{N^{\frac{1}{4}}}\Big) d\nu_1(x) - H(\nu_1) \Big\} \\
+ \theta^2\big(\alpha_1^2 + 2\alpha_1\alpha_2\big) + A\alpha_2^2 + B(\alpha_3^2 + 2\alpha_2\alpha_3) - \frac{1}{2}\log(2\pi) - \frac{1}{2},
\end{aligned}$$

$S_3 = [K, N^{1/4}\sqrt{\alpha_3}]$. Using the change of variable $s \mapsto sN^{-1/4}$ we have the upper bound,

$$\hat{F}(\theta) \leq \sup_{(\underline{\alpha},s,\nu)\in\mathcal{S}'} \mathcal{F}(\underline{\alpha}, s, \nu).$$

Let now $(\underline{\alpha}, s, \nu) \in \mathcal{S}'$. Taking $\mu_2 = \delta_{\sqrt{m_*}}$, $s' = N^{1/4}s$ and ν_1 to be the unique probability measure on I_1 such that for any continuous bounded function $f : \mathbb{R} \to \mathbb{R}$,

$$\int f d\nu_1 = \frac{1}{\nu(I_1)} \int_{I_1} f(\lambda x) d\nu(x),$$

where λ is such that $\int x^2 d\nu_1(x) = \alpha_1$. With a similar argument as in the proof of Proposition 1.5.28, we can deduce that

$$\underline{F}(\theta) \geq \mathcal{F}(\underline{\alpha}, s, \nu).$$

Since $\underline{F}(\theta) \leq \overline{F}(\theta)$ and $\overline{F}(\theta) \leq \hat{F}(\theta)$ by (1.90), this ends the proof of the claim. $\square$

We next prove that the supremum is taken at $\alpha_3 = 0$.

Proposition 1.5.40. *There exists θ_0 depending on A such that for any $\theta \geq \theta_0$,*

$$\sup_{(\underline{\alpha},s,\nu)\in\mathcal{S}'} \mathcal{F}(\underline{\alpha}, s, \nu) = \sup_{(\alpha_1,\alpha_2,0,s,\nu)\in\mathcal{S}'} \mathcal{F}((\alpha_1, \alpha_2, 0), s, \nu).$$

Proof. We claim that for any $((\alpha_1, \alpha_2, \alpha_3), s, \nu) \in \mathcal{S}'$ such that $\alpha_2 \geq \frac{A-1}{2A-B-1}$, we have

$$\mathcal{F}((\alpha_1, \alpha_2, \alpha_3), s, \nu) \leq \sup_{\nu\in\mathcal{P}(\mathbb{R})} \mathcal{F}((\alpha_1, \alpha_2 + \alpha_3, 0), \nu). \tag{1.96}$$

Note that

$$\sup_{\nu\in\mathcal{P}(\mathbb{R})} \mathcal{F}((\alpha_1, \alpha_2 + \alpha_3, 0), \nu) = \theta^2(\alpha_1 + 2\alpha_1\alpha_2 + 2\alpha_1\alpha_3) + \theta^2 A(\alpha_2 + \alpha_3)^2 + \frac{1}{2}\log\alpha_1.$$

Now, for any $((\alpha_1, \alpha_2, \alpha_3), s, \nu) \in \mathcal{S}'$, using the fact that $\psi(x) \leq A/2$ for any $x \in \mathbb{R}$, we have

$$\mathcal{F}((\alpha_1, \alpha_2, \alpha_3), s, \nu) \leq \theta^2(\alpha_1 + 2\alpha_1\alpha_2) + \theta^2 A\alpha_2^2 + 2\theta^2 A\alpha_1\alpha_3 + \theta^2 B(\alpha_3^2 + 2\alpha_2\alpha_3) + \frac{1}{2}\log\alpha_1.$$

Therefore, it suffices to prove that for α_2 sufficiently near 1:

$$(A - B)(2\alpha_2\alpha_3 + \alpha_3^2) \geq 2(A - 1)\alpha_1\alpha_3$$

that is,

$$2(A - 1)\alpha_1 \leq (A - B)(\alpha_3 + 2\alpha_2).$$

A sufficient condition for this inequality to be true is that $(A - 1)(1 - \alpha_2) \leq (A - B)\alpha_2$, which ends the proof of the claim (1.96). By Lemma 1.5.36, we know that for $\theta \geq \theta_0$,

$$\sup_{(\underline{\alpha},s,\nu)\in S} \mathcal{F}(\underline{\alpha}, s, \nu) = \sup_{\substack{(\underline{\alpha},s,\nu)\in S \\ \alpha_1,\alpha_3\leq C\sqrt{\log\theta}/\theta}} \mathcal{F}(\underline{\alpha}, s, \nu).$$

For θ such that

$$1 - 2C\frac{\sqrt{\log\theta}}{\theta} \geq \frac{A-1}{2A-B-1},$$

we obtain from (1.96) that

$$\sup_{\substack{(\underline{\alpha},s,\nu)\in S \\ \alpha_1,\alpha_3\leq C\sqrt{\log\theta}/\theta}} \mathcal{F}(\underline{\alpha}, s, \nu) \leq \sup_{((\alpha_1,\alpha_2,0),s,\nu)\in S} \mathcal{F}((\alpha_1, \alpha_2, 0), s, \nu).$$

We deduce that for θ large enough,

$$\sup_{(\underline{\alpha},s,\nu)\in S} \mathcal{F}(\underline{\alpha},s,\nu) \leq \sup_{((\alpha_1,\alpha_2,0),s,\nu)\in S} \mathcal{F}(\alpha_1,\alpha_2,0,s,\nu),$$

which ends the proof. $\square$

We can now conclude the proof of Proposition 1.5.35.

Proof of Proposition 1.5.35. By Propositions 1.5.39 and 1.5.40, and optimizing over ν (at the centered Gaussian law with covariance α_1), we deduce that there exists $\theta_0 > 0$ depending on ψ such that for $\theta \geq \theta_0$,

$$F(\theta) = \sup_{\alpha\in[0,1)} V(\alpha),$$

where

$$\forall \alpha > 0,\ V(\alpha) = \theta^2(A-1)\alpha^2 + \theta^2 + \frac{1}{2}\log(1-\alpha).$$

We can therefore give an explicit expression for $F(\theta)$ for θ large enough such that $\theta \geq 1/\sqrt{A-1}$, using the fact that the solution of the optimization problem $\alpha = \alpha_\theta$ should satisfy $\alpha_\theta = 1 - o_\theta(1)$ according to Lemma 1.5.36. $\square$

1.5.7 Appendix

1.5.7.1 Concentration for Wigner matrices with sub-Gaussian log-concave entries

Proposition 1.5.41. *Let μ be a symmetric probability measure on $\mathbb{R}$ which has log-concave tails in the sense that $t \mapsto \mu(x : |x| \geq t)$ is concave, and which is sub-Gaussian in the sense that* (1.44) *holds. Let X_N be a symmetric random matrix of size N such that $(X_{i,j})_{i\leq j}$ are independent random variables. Assume $\sqrt{N}X_{i,j}$ and $\sqrt{N/2}X_{i,i}$ have law μ for any $i \neq j$. There exists a numerical constant $\kappa > 0$ such that for any convex* 1*-Lipschitz function $f : \mathbb{R} \to \mathbb{R}$, and $t \geq 0$,*

$$\mathbb{P}\Big(\big|\frac{1}{n}\mathrm{Tr} f(X_N) - \frac{1}{n}\mathbb{E}\mathrm{Tr} f(X_N)\big| > t\Big) \leq 2e^{-\frac{\kappa}{A}N^2t^2}. \tag{1.97}$$

Moreover, for any $t > 0$,

$$\mathbb{P}\big(|\lambda_{X_N} - \mathbb{E}\lambda_{X_N}| > t\big) \leq 2e^{-\frac{\kappa}{A}Nt^2}. \tag{1.98}$$

One can take $\kappa = 1/8\beta^2$ with $\beta = 1680e$.

From these concentration inequalities, one can deduce that a Wigner matrix with entries having sub-Gaussian and log-concave laws satisfy Assumptions 1.5.4.

Corollary 1.5.42. *Assume μ satisfies the assumptions of Proposition 1.5.41 and has variance* 1*. Then the matrix X_N satisfies the Assumptions 1.5.4.*

Proof. Using the concentration inequality (1.98) and the convergence in expectation of the spectral radius of X_N (see [4, Theorem 2.1.22, Exercise 2.1.7]), we obtain that the spectral radius of X_N is exponentially tight at the scale N in the sense of (1.46).

Let $\mathcal{K} \subset \mathbb{R}$ be a compact subset of $\mathbb{R}$ and denote by $|\mathcal{K}|$ its diameter. Let $\mathcal{F}_{\mathrm{Lip},\mathcal{K}}$ be the set of 1-Lipschitz functions with support in $\mathcal{K}$. From the concentration inequality (1.97), we can deduce by arguing as in the proof of [49, Theorem 1.3] that for any $\delta > 0$,

$$\mathbb{P}\Big(\sup_{f\in\mathcal{F}_{\mathrm{Lip},\mathcal{K}}} \big|\frac{1}{N}\mathrm{Tr} f(X_N) - \mathbb{E}\frac{1}{N}\mathrm{Tr} f(X_N)\big| > \delta\Big) \leq C\frac{|\mathcal{K}|}{\delta}\exp\Big(-C\frac{\delta^2N^2}{|\mathcal{K}|^2}\Big). \tag{1.99}$$

Let $d_{\mathcal{K}}$ be defined by

$$\forall \nu,\mu \in \mathcal{P}(\mathbb{R}),\ d_{\mathcal{K}}(\mu,\nu) = \sup_{f\in\mathcal{F}_{\mathrm{Lip},\mathcal{K}}} \Big|\int f d\nu - \int f d\mu\Big|,$$

and let d denote the bounded-Lipschitz distance defined by

$$\forall \nu, \mu \in \mathcal{P}(\mathbb{R}),\ d(\mu,\nu) = \sup_{f \in \mathcal{F}_{LU}} \Big| \int f d\nu - \int f d\mu \Big|,$$

where $\mathcal{F}_{LU}$ is the set of 1-Lipschitz functions uniformly bounded by 1. Let fix $\mathcal{K} = [-2,2]$. Since $\sigma(\mathcal{K}) = 1$, we have

$$d(\mu_{X_N}, \sigma) \leq d_{\mathcal{K}}(\mu_{X_N}, \sigma) + \mu_{X_N}(\mathcal{K}^c).$$

But $\mu_{X_N}(\mathcal{K}^c) \leq d_{\mathcal{K}}(\mu_{X_N}, \sigma)$ by choosing $f = \mathbb{1}_{\mathcal{K}}$. Therefore,

$$d(\mu_{X_N}, \sigma) \leq 2 d_{\mathcal{K}}(\mu_{X_N}, \sigma).$$

From (1.99), we have that for any $\delta > 0$,

$$\mathbb{P}\big(d(\mu_{X_N}, \sigma) > \delta + 2 d_{\mathcal{K}}(\mathbb{E}\mu_{X_N}, \sigma)\big) \leq \frac{4C}{\delta} e^{-C\frac{\delta^2 N^2}{16}}. \tag{1.100}$$

But, we know by [10, Theorem 4.1] that,

$$d_{KS}(\mathbb{E}\mu_{X_N}, \sigma) = O(N^{-1/4}),$$

where d_{KS} denotes the Kolmogorov-Smirnov distance, defined by

$$\forall \nu, \mu \in \mathcal{P}(\mathbb{R}),\ d_{KS}(\mu,\nu) = \sup\big\{ \big|\mu\big((-\infty, x]\big) - \nu\big((-\infty, x]\big)\big|, x \in \mathbb{R} \big\}.$$

It is a standard fact that the Kolmogorov-Smirnov distance controls the integrals with respect to functions with bounded variations, that is, for any $\nu, \mu \in \mathcal{P}(\mathbb{R})$,

$$d_{KS}(\mu,\nu) = \sup\Big\{ \int f d\mu - \int f d\nu : ||f||_{BV} \leq 1 \Big\},$$

with $||f||_{BV} = \sup \sum_{i=1}^m |f(x_{i+1}) - f(x_i)|$, where the supremum holds over all families $(x_i)_{1 \leq i \leq m+1}$ with $m \in \mathbb{N}$ such that $x_1 < x_2 < \ldots < x_{m+1}$. Since for any 1-Lipschitz function f supported on $\mathcal{K}$, $||f||_{BV} = O(1)$, we deduce that

$$d_{\mathcal{K}}(\mathbb{E}\mu_{X_N}, \sigma) = O(N^{-1/4}).$$

From (1.100), we deduce that

$$\lim_{N \to +\infty} \frac{1}{N} \log \mathbb{P}\big(d(\mu_{X_N}, \sigma) > N^{-\kappa}\big) = -\infty,$$

for any $\kappa < 1/4$. □

We now give a proof of Proposition 1.5.41. It will be a direct consequence of Klein's lemma (see [4, Lemma 4.4.12]) and the following concentration of convex Lipschitz functions under μ^n.

Proposition 1.5.43. *Let μ be a symmetric probability measure on $\mathbb{R}$ which has log-concave tails in the sense that $t \mapsto \mu(x : |x| \geq t)$ is concave, and which is sub-Gaussian in the sense that (1.44) holds. For any lower-bounded convex 1-Lipschitz function $f : \mathbb{R} \to \mathbb{R}$ such that $\int f d\mu^n = 0$ and any $t > 0$,*

$$\mu^n\big(x : |f(x)| > t\big) \leq 2e^{-\frac{t^2}{4\beta^2 A}},$$

where β is numerical constant. One can take $\beta = 1680e$.

Proof. By [63, Corollary 2.2], we know that there exists a numerical constant β such that μ^n satisfies a convex infimum convolution inequality with cost function $\Lambda^*(./\beta)$, where Λ^* is the Legendre transfom of Λ defined by,

$$\forall \theta \in \mathbb{R}^n, \ \Lambda(\theta) = \log \int e^{\langle \theta, x\rangle} d\mu^n(x).$$

Moreover, β can be taken to be $1680e$. More precisely, for any convex lower-bounded function $f : \mathbb{R} \to \mathbb{R}$,

$$\Big(\int e^{f \Box \Lambda^*(./\beta)} d\mu^n \Big) \Big(\int e^{-f} d\mu^n \Big) \leq 1, \tag{1.101}$$

where $\Box$ denotes the infimum convolution operator, defined by

$$f \Box \Lambda^*(./\beta)(x) = \inf_{y \in \mathbb{R}^n} \big\{ f(y) + \Lambda^* \Big(\frac{y-x}{\beta} \Big) \big\}.$$

Since μ is sub-Gaussian in the sense of (1.44), for any $x \in \mathbb{R}^n$,

$$\Lambda^*(x) \geq \frac{1}{2A} ||x||^2,$$

where $||\ ||$ denotes the Euclidean norm in $\mathbb{R}^n$. Therefore,

$$f \Box \Lambda^*(./\beta)(x) \geq \inf_{y \in \mathbb{R}^n} \big\{ f(y) + \frac{1}{2\beta^2 A} ||y-x||^2 \big\}.$$

Assume f is L-Lipschitz for some $L > 0$. Reproducing the arguments of [57, section 1.9, p19] we have for any $x \in \mathbb{R}^n$,

$$\begin{aligned} f \Box \Lambda^*(./\beta)(x) &\geq f(x) + \inf_{y \in \mathbb{R}^n} \big\{ -L||y-x|| + \frac{1}{2\beta^2 A} ||y-x||^2 \big\} \\ &\geq f(x) - \frac{1}{2}\beta^2 A L^2. \end{aligned}$$

Thus, by (1.101) we deduce that

$$\Big(\int e^{f} d\mu^n \Big) \Big(\int e^{-f} d\mu^n \Big) \leq e^{\frac{1}{2}\beta^2 A L^2}. \tag{1.102}$$

Assume now that f is 1-Lipschitz and $\int f d\mu^n = 0$. Using Jensen's inequality, we get for any $\lambda > 0$,

$$\int f d\mu^n \leq e^{\frac{1}{2}\beta^2 A \lambda^2}.$$

Using Chernoff inequality we obtain that for any $t > 0$,

$$\mu^n \big(x : f(x) \geq t \big) \leq e^{-\frac{t^2}{4\beta^2 A}}.$$

Using the symmetry in (1.102) between f and $-f$, we get similarly that for any $t > 0$,

$$\mu^n \big(x : f(x) \leq -t \big) \leq e^{-\frac{t^2}{4\beta^2 A}},$$

which gives the claim. $\square$

1.5.7.2 A Uniform Varadhan's lemma

Lemma 1.5.44. *Let $f : \mathbb{R} \to \mathbb{R}$ such that $f(0) = 0$ and $f(\sqrt{\cdot})$ is L-Lipschitz for some $L > 0$. Let M_N, m_N be sequences such that $M_N = o(\sqrt{N})$ and $m_N \sim N$. Let $g_1, \dots, g_{m_N}$ be independent Gaussian random variables conditioned to belong to $[-M_N, M_N]$. Let $\delta \in (0,1)$ and $c > 0$ such that $K^{-1} < c < K$ and $2\delta < K^{-1}$. Then,*

$$\Big|\frac{1}{N}\log \mathbb{E} e^{\sum_{i=1}^{m_N} f\left(\frac{g_i}{\sqrt{c}}\right)} \mathbb{1}_{|\sum_{i=1}^{m_N} g_i^2 - cN| \le \delta N} - \sup_{\substack{\nu \in \mathcal{P}([-M_N, M_N]) \\ \int x^2 d\nu = c}} \Big\{ \int f\Big(\frac{x}{\sqrt{c}}\Big) d\nu(x) - H(\nu|\gamma)\Big\}\Big|$$
$$\le g_{L,K}(N) + h_L(\delta K),$$

where $g_{L,K}(N) \to +\infty$ as $N \to +\infty$ and $h_L(x) \to 0$ as $x \to 0$.

Proof. Let $\epsilon = 1/N$ and l_0 be the smallest integer such that $(1+\epsilon)^{-l_0} \le \epsilon$. Define $I_{l_0} = [-(1+\epsilon)^{-l_0}, (1+\epsilon)^{-l_0}]$ and $B_{l_0} = \{i : g_i \in I_{l_0}\}$. For any $k > -l_0$, we set

$$I_k = \{x \in \mathbb{R} : (1+\epsilon)^{k-1} \le |x| \le (1+\epsilon)^k\} \text{ and } B_k = \big\{i : g_i \in I_k\big\}. \tag{1.103}$$

Let $\mu_k = |B_k|/m_N$. Let k_0 be the smallest integer such that $(1+\epsilon)^{k_0} \ge M_N$. If $g_i \in [-M_N, M_N]$ for all i, then for any $k > k_0$, $B_k = \emptyset$.

Lemma 1.5.45. *On the event $\big\{|\sum_{i=1}^{m_N} g_i^2 - cN| \le \delta N\big\}$,*

$$\Big|\frac{1}{m_N}\sum_{i=1}^{m_N} f\Big(\frac{g_i}{\sqrt{c}}\Big) - \sum_{k=-l_0}^{k_0} \mu_k f\Big(\frac{(1+\epsilon)^k}{\sqrt{c}}\Big)\Big| \le g_{L,K}(N),$$

where $g_{L,K}$ is function such that $g_{L,K}(x) \to 0$ as $x \to +\infty$.

Proof. As $f(\sqrt{\cdot})$ is L-Lipschitz, we have

$$\begin{aligned}\Big|\frac{1}{m_N}\sum_{i=1}^{m_N} f\Big(\frac{g_i}{\sqrt{c}}\Big) - \sum_{k=-l_0}^{k_0} \mu_k f\Big(\frac{(1+\epsilon)^k}{\sqrt{c}}\Big)\Big| &\le \frac{1}{m_N}\sum_{k=-l_0}^{k_0}\sum_{i\in B_k}\Big|f\Big(\frac{g_i}{\sqrt{c}}\Big) - f\Big(\frac{(1+\epsilon)^k}{\sqrt{c}}\Big)\Big| \\ &\le \frac{L}{c}\sum_{k=-l_0+1}^{k_0} \mu_k (1+\epsilon)^{2k}\big(1-(1+\epsilon)^{-2}\big) \\ &+ \frac{L}{c}\mu_{-l_0}(1+\epsilon)^{-2l_0}.\end{aligned}$$

Using the fact that $(1+\epsilon)^{-l_0} \le \epsilon$, we get,

$$\Big|\frac{1}{m_N}\sum_{i=1}^{m_N} f\Big(\frac{g_i}{\sqrt{c}}\Big) - \sum_{k=-l_0}^{k_0} \mu_k f\Big(\frac{(1+\epsilon)^k}{\sqrt{c}}\Big)\Big| \le \frac{3\epsilon L}{c}\Big(\sum_{k=-l_0+1}^{k_0} \mu_k (1+\epsilon)^{2k} + 1\Big)$$

But, on the other hand

$$\frac{1}{m_N}\sum_{i=1}^{m_N} g_i^2 \ge \sum_{k=-l_0+1}^{k_0} \mu_k (1+\epsilon)^{2(k-1)}.$$

Thus,

$$\Big|\frac{1}{m_N}\sum_{i=1}^{m_N} f\Big(\frac{g_i}{\sqrt{c}}\Big) - \sum_{k=-l_0}^{k_0} \mu_k f\Big(\frac{(1+\epsilon)^k}{\sqrt{c}}\Big)\Big| \le \frac{3\epsilon L}{c}\Big((1+\epsilon)^2\Big(\frac{1}{m_N}\sum_{i=1}^{m_N} g_i^2\Big) + 1\Big),$$

which, as $m_N \sim N$ and $c > K^{-1}$ gives the claim. $\square$

Let $I = \{-l_0, \dots, k_0\}$ and $\mathcal{L}_N$ be the set,

$$\mathcal{L}_N = \big\{y \in \mathbb{R}_+^I : \sum_{k\in I} y_k = 1,\ \forall k \in I, m_N y_k \in \mathbb{N}\big\}.$$

We know from [36, Lemma 2.1.6], that for any $y \in \mathcal{L}_N$,

$$(m_N+1)^{-n} e^{-m_N H(y|\gamma_{m_N})} \leq \mathbb{P}(\mu_k = y_k,\ \forall k \in I) \leq e^{-m_N H(y|\gamma_{m_N})}, \tag{1.104}$$

where $n = |I| = l_0 + k_0 + 1$,

$$H(y|\gamma_{m_N}) = \sum_{k\in I} y_k \log \frac{y_k}{\gamma_{m_N}(k)}, \quad \text{with } \gamma_{m_N}(k) = \frac{\gamma(I_k)}{\gamma([-M_N, M_N])},$$

and $I_k = [(1+\epsilon)^{k-1}, (1+\epsilon)^k]$.

Let $\mu = (\mu_k)_{k\in I}$. There exists a function h_c such that

$$\begin{aligned}\big\{c - \delta - h_K(\epsilon) \leq \sum_{k=-l_0}^{k_0} (1+\epsilon)^{2k} \mu_k \leq c + \delta\big\} \\ \subset \big\{|\sum_{i=1}^{m_N} g_i^2 - cN| \leq \delta N\big\} \subset \{c - \delta \leq \sum_{k=-l_0}^{k_0} (1+\epsilon)^{2k} \mu_k \leq c + \delta + h_K(\epsilon)\},\end{aligned} \tag{1.105}$$

and $h_c(\epsilon) \to 0$ as $\epsilon \to 0$. We introduce the sets

$$\mathcal{A}_+ = \big\{y \in \mathcal{L}_N : c - \delta \leq \sum_{k=-l_0}^{k_0} (1+\epsilon)^{2k} y_k \leq c + \delta + h_K(\epsilon)\big\},$$

and

$$\mathcal{A}_- = \big\{y \in \mathcal{L}_N : c - \delta - h_c(\epsilon) \leq \sum_{k=-l_0}^{k_0} (1+\epsilon)^{2k} y_k \leq c + \delta\big\}.$$

Using (1.104), we have for the upper bound,

$$\mathbb{E} e^{m_N \sum_{k=-l_0}^{k_0} \mu_k f\left(\frac{(1+\epsilon)^k}{\sqrt{c}}\right)} \mathbb{1}_{\mu \in \mathcal{A}_+} \leq \sum_{y \in \mathcal{A}_+} e^{m_N \sum_{k=-l_0}^{k_0} y_k f\left(\frac{(1+\epsilon)^k}{\sqrt{c}}\right)} e^{-m_N H(y|\gamma_{m_N})}, \tag{1.106}$$

whereas for the lower bound,

$$\mathbb{E} e^{m_N \sum_{k=-l_0}^{k_0} \mu_k f\left(\frac{(1+\epsilon)^k}{\sqrt{c}}\right)} \mathbb{1}_{\mu \in \mathcal{A}_-} \geq (m_N+1)^{-n} \sum_{y \in \mathcal{A}_-} e^{m_N \sum_{k=-l_0}^{k_0} y_k f\left(\frac{(1+\epsilon)^k}{\sqrt{c}}\right)} e^{-m_N H(y|\gamma_{m_N})}. \tag{1.107}$$

Let $y \in \mathcal{A}_+$ and define $\nu \in \mathcal{P}(\mathbb{R})$ by $\nu = \varphi d\gamma$, where

$$\varphi(x) = \sum_{k=-l_0}^{k_0} \mathbb{1}_{x\in I_k} \frac{y_k}{\gamma(I_k)}.$$

With this notation, we have

$$H(y|\gamma_{m_N}) = H(\nu|\gamma) - \log \gamma([-M_N, M_N]).$$

With the same argument as in Lemma 1.5.45, we also have as $y \in \mathcal{A}_+$,

$$\Big|\sum_{k=-l_0}^{k_0} y_k f\Big(\frac{(1+\epsilon)^k}{\sqrt{c}}\Big) - \int f\Big(\frac{x}{\sqrt{c}}\Big) d\nu(x)\Big| \leq g_{L,K}(N), \tag{1.108}$$

and

$$c-\delta-h_K(N)\leq\int x^2d\nu(x)\leq c+\delta+h_K(N),$$

where $g_{L,K}(x)$ and $h_K(x)\to 0$ as $x\to+\infty$. From (1.106), we deduce that

$$\frac{1}{N}\log\mathbb{E}e^{m_N\sum_{k=-l_0}^{k_0}\mu_k f\left(\frac{(1+\epsilon)^k}{\sqrt{c}}\right)}\mathbb{1}_{\mu\in\mathcal{A}_+}\leq\frac{m_N}{N}\sup_{\substack{\nu\in\mathcal{P}([-M_N,M_N]\\ |\int x^2d\nu(x)-c|\leq\delta+h_K(N)}}\Big\{\int f\Big(\frac{x}{\sqrt{c}}\Big)d\nu(x)-H(\nu|\gamma)\Big\}$$
$$+\frac{1}{N}\log|\mathcal{L}_N|+g_{L,K}(N).$$

But, $|\mathcal{L}_N|\leq m_N^n$, and $n=|I|\leq C\log N$, where $C>0$ is a numerical constant, so that

$$\frac{1}{N}\log|\mathcal{L}_N|\leq C\frac{(\log N)^2}{N}.$$

To complete the proof of the upper bound, one can check that as $c>K^{-1}$,

$$\sup_{|\int x^2d\nu(x)-c|\leq\delta+h_K(N)}\Big\{\int f\Big(\frac{x}{\sqrt{c}}\Big)d\nu(x)-H(\nu|\gamma)\Big\}$$
$$\leq\sup_{\substack{\nu\in\mathcal{P}([-M_N,M_N]\\ \int x^2d\nu(x)=c}}\Big\{\int f\Big(\frac{x}{\sqrt{c}}\Big)d\nu(x)-H(\nu|\gamma)\Big\}+\tilde{g}_{L,K}(N)+h_L(\delta K),$$

where $h_L(x)\to 0$ as $x\to 0$ and $\tilde{g}_{L,K}(N)\to 0$ as $N\to+\infty$.

For the lower bound, fix ν a probability measure on $\mathbb{R}$ such that $\nu\ll\gamma$. We set $\epsilon=\epsilon_N$ such that $M_N^2/m_N\epsilon_N\to 0$, and we define I_k and B_k as in (1.103). Define, for $k\in\{-l_0+1,\dots,k_0\}$,

$$y_k=\frac{1}{m_N}\lfloor m_N\nu(I_k)\rfloor,$$

and $y_{-l_0}=1-\sum_{k=-l_0+1}^{k_0}y_k$. We claim that for N large enough,

$$\int x^2d\nu(x)=c\Longrightarrow y\in\mathcal{A}_-.$$

Indeed, one can check that on one hand

$$\sum_{k=-l_0}^{k_0}y_k(1+\epsilon)^{2k}\leq(1+\epsilon)^2\int x^2d\nu(x)+\epsilon^2,$$

and on the other hand

$$\sum_{k=-l_0}^{k_0}y_k(1+\epsilon)^{2k}\geq\int x^2d\nu(x)-\frac{(1+\epsilon)^2M_N^2}{m_N\epsilon}.$$

Let $\int x^2d\nu(x)\in[c-\delta+\tilde{h}_K(N),c+\delta-\tilde{h}_K(N)]$. We obtain from (1.107),

$$\log\mathbb{E}e^{m_N\sum_{k=-l_0}^{k_0}\mu_k f\left(\frac{(1+\epsilon)^k}{\sqrt{c}}\right)}\mathbb{1}_{\mu\in\mathcal{A}_-}\geq(m_N+1)^{-n}e^{m_N\sum_{k=-l_0}^{k_0}y_k f\left(\frac{(1+\epsilon)^k}{\sqrt{c}}\right)}e^{-m_NH(y|\gamma_{m_N})}.\tag{1.109}$$

In the next lemma we compare $H(y|\gamma_{m_N})$ and $H(\nu|\gamma)$.

Lemma 1.5.46.

$$H(y|\gamma_{m_N})\leq H(\nu|\gamma)+o_N(1).$$

Proof. By definition we have,

$$H(y|\gamma_{m_N}) = \sum_{k=-l_0}^{k_0} y_k \log \frac{y_k}{\gamma(I_k)} + \log \gamma([-M_N, M_N]). \tag{1.110}$$

Let $f(x) = x \log x$ for $x > 0$ and $f(0) = 0$. We claim that

$$\forall 0 \le x < y,\ f(x) \le f(y) + (y - x). \tag{1.111}$$

Indeed, either $x > e^{-1}$ and $f(x) \le f(y)$ since f is increasing on $[e^{-1}, +\infty)$. Either $x < e^{-1}$ and by convexity,

$$f(x) \le f(y) + f'(x)(x - y).$$

Since $|f'(x)| \le 1$ we get the claim. Note that we have for any $k > -l_0$,

$$\nu(I_k) - \frac{1}{m_N} < y_k \le \nu(I_k),$$

and

$$\nu(I_{-l_0}) \le y_{-l_0} < \nu(I_{-l_0}) + \frac{k_0 + l_0}{m_N}.$$

Thus we deduce from (1.111) that

$$\begin{aligned} H(y|\gamma_{m_N}) &\le \sum_{k=-l_0}^{k_0} \nu(I_k) \log \frac{\nu(I_k)}{\gamma(I_k)} + \sum_{k=-l_0+1}^{k_0} \gamma(I_k) \frac{1}{\gamma(I_k) m_N} + \gamma(I_{l_0}) \frac{k_0 + l_0}{\gamma(I_{-l_0}) m_N} + o_N(1) \\ &\le \sum_{k=-l_0}^{k_0} \nu(I_k) \log \frac{\nu(I_k)}{\gamma(I_k)} + \frac{2(k_0 + l_0)}{m_N} + o_N(1). \end{aligned}$$

We have $k_0 = O(\log(M_N)/\epsilon_N)$ and $l_0 = O(\log(1/\epsilon_N)/\epsilon_N)$. Since $M_N^2/m_N\epsilon_N \to 0$, we get

$$H(y|\gamma_{m_N}) \le \sum_{k=-l_0}^{k_0} \nu(I_k) \log \frac{\nu(I_k)}{\gamma(I_k)} + \frac{2(k_0 + l_0)}{m_N} + o_N(1).$$

Since $f : x \mapsto x \log x$ is a convex function, we have by Jensen's inequality

$$\sum_{k=-l_0}^{k_0} \nu(I_k) \log \frac{\nu(I_k)}{\gamma(I_k)} = \sum_{k=-l_0}^{k_0} \gamma(I_k) f\Big(\frac{1}{\gamma(I_k)} \int_{I_k} \frac{d\nu}{d\gamma} d\gamma\Big) \le \sum_{k=-l_0}^{k_0} \int_{I_k} \frac{d\nu}{d\gamma} \log \frac{d\nu}{d\gamma} d\gamma,$$

which ends the proof. □

Next, we claim that we can compare $\int f(x/\sqrt{c}) d\nu(x)$ and $\sum_{k=-l_0} y_k f((1+\epsilon)^k/\sqrt{c})$.

Lemma 1.5.47.

$$\Big| \int f\Big(\frac{x}{\sqrt{c}}\Big) d\nu(x) - \sum_{k=-l_0}^{k_0} y_k f\Big(\frac{(1+\epsilon)^k}{\sqrt{c}}\Big) \Big| \le g_{L,K}(N),$$

where $g_{L,K}(N) \to 0$ *as* $N \to +\infty$.

Proof. As $f(\sqrt{\cdot})$ is L-Lipschitz, we have on one hand using the same argument as in the proof of Lemma 1.5.45,

$$\Big| \int f\Big(\frac{x}{\sqrt{c}}\Big) d\nu(x) - \sum_{k=-l_0}^{k_0} \nu(I_k) f\Big(\frac{(1+\epsilon)^k}{\sqrt{c}}\Big) \Big| \le \frac{3L\epsilon}{c} \Big(\int x^2 d\nu(x) + 1 \Big).$$

Therefore,

$$\Big|\int f\Big(\frac{x}{\sqrt{c}}\Big)d\nu(x) - \sum_{k=-l_0}^{k_0} \nu(I_k) f\Big(\frac{(1+\epsilon)^k}{\sqrt{c}}\Big)\Big| \leq h_{L,K}(N), \tag{1.112}$$

where $h_{L,K}(N) \to 0$ as $N \to +\infty$. On the other hand, as $|\nu(I_k) - y_k| \leq 1/m_N$ for any $k > -l_0$ and $|\nu(I_{-l_0}) - y_{-l_0}| \leq (k_0 + l_0)/m_N$, we get

$$\sum_{k=-l_0}^{k_0} \Big|(y_k - \nu(I_k)) f\Big(\frac{(1+\epsilon)^k}{\sqrt{c}}\Big)\Big| \leq \frac{L}{cm_N} \sum_{k=-l_0+1}^{k_0} (1+\epsilon)^{2k} + \frac{L(k_0+l_0)}{cm_N}(1+\epsilon)^{-2l_0},$$

where we used the fact that $f(0) = 0$ and $f(\sqrt{.})$ is L-Lipschitz. There exists a numerical constant $\kappa > 0$ such that

$$\sum_{k=-l_0}^{k_0} |y_k - \nu(I_k)| f\Big(\frac{(1+\epsilon)^k}{\sqrt{c}}\Big) \leq \frac{\kappa L}{cm_N}\Big(\frac{M_N^2}{\epsilon_N} + (k_0+l_0)\epsilon_N^2\Big),$$

As $M_N^2/m_N \to 0$ and $k_0 = O(\log(M_N)/\epsilon_N)$ and $l_0 = O(\log(1/\epsilon_N)/\epsilon_N)$, we deduce

$$\frac{(k_0+l_0)}{m_N}\epsilon_N = o_N(1).$$

Since we choose ϵ_N such that $M_N^2/m_N\epsilon_N \to 0$, we can conclude that

$$\sum_{k=-l_0}^{k_0} \Big|(y_k - \nu(I_k)) f\Big(\frac{(1+\epsilon)^k}{\sqrt{c}}\Big)\Big| \leq g_{L,K}(N),$$

where $g_{L,K}(N) \to 0$ as $N \to +\infty$. Combining the above estimate with (1.108), we get the claim. $\square$

Coming back to (1.109), using the results of Lemmas 1.5.46 and 1.5.47, we deduce

$$\mathbb{E}e^{m_N \sum_{k=-l_0}^{k_0} \mu_k f\left(\frac{(1+\epsilon)^k}{\sqrt{c}}\right)} \mathbb{1}_{\mu\in\mathcal{A}_-} \geq (m_N+1)^{-n} e^{m_N \int f(\frac{x}{\sqrt{c}})d\nu(x) - m_N g_{L,\delta}(N)} e^{-m_N(H(\nu|\gamma)+o_N(1))},$$

which gives at the logarithmic scale,

$$\frac{1}{N}\log \mathbb{E}e^{m_N \sum_{k=-l_0}^{k_0} \mu_k f\left(\frac{(1+\epsilon)^k}{\sqrt{c}}\right)} \mathbb{1}_{\mu\in\mathcal{A}_-} \geq \frac{m_N}{N}\Big(\int f\Big(\frac{x}{\sqrt{c}}\Big)d\nu(x) - H(\nu|\gamma)\Big) - \tilde{g}_{L,K}(N),$$

where $\tilde{g}_{L,K}(N) \to 0$ as $N \to +\infty$. As this inequality is true for any $\nu \in \mathcal{P}([-M_N, M_N])$, such that $\int x^2 d\nu = c$,

$$\frac{1}{N}\log \mathbb{E}e^{m_N \sum_{k=-l_0}^{k_0} \mu_k f\left(\frac{(1+\epsilon)^k}{\sqrt{c}}\right)} \mathbb{1}_{\mu\in\mathcal{A}_-} \geq \frac{m_N}{N} \sup_{\substack{\nu\in\mathcal{P}([-M_N,M_N]) \\ \int x^2 d\nu = c}} \Big(\int f\Big(\frac{x}{\sqrt{c}}\Big)d\nu(x) - H(\nu|\gamma)\Big)$$
$$- \tilde{g}_{L,K}(N).$$

$\square$

Chapitre 2

Convergence de la mesure empirique de polynômes de matrices de Ginibre

Résumé du chapitre

Lorsqu'on s'intéresse non plus à une mais à plusieurs matrices aléatoires indépendantes $X_N^1, X_N^2, \cdots$, on peut alors regarder le comportement de la mesure spectrale du polynôme $P_N = P(X_N^1, X_N^2, ...)$. Notons que dans ce cas le polynôme P est non-commutatif étant donné que l'on n'a pas nécessairement $X_N^i X_N^j = X_N^j X_N^i$. En particulier, à l'instar du cas où l'on prend un polynôme en des variables aléatoires réelles ou complexes indépendantes, on peut se poser la question suivante : La mesure limite de P_N dépend-t-elle uniquement de celles de $X_1^N, X_2^N, ...$ et si oui comment la décrire ? Dans [69], Voiculescu répond à cette question si P est un polynôme auto-adjoint et les X_N^i des matrices de Wigner. Dans ce cas la mesure empirique de P_N converge vers la mesure spectrale de $P(s_1, ..., s_n)$ où $s_1, ..., s_n$ sont des éléments semi-circulaires libres d'un espace de probabilité non-commutatif. On pourra se référer à l'Annexe A.2 de cette thèse ou à la section 5 de [4] pour se rappeler les liens entre matrices aléatoires et probabilités libres. La question qui subsiste alors est : que se passe-t-il dans le cas où P est non-hermitien ? Dans ce chapitre, on apporte une réponse à cette question dans le cas d'un polynôme P d'une certaine forme en des matrices gaussiennes :

Théorème 2.0.1 (Corollaire 2.3.3). *Soit $(X_N^1, ..., X_N^k)$ un k-uplet de matrices de Ginibre complexes indépendantes et P un polynôme de degré 2. Alors la mesure empirique de $P(X_N^1, ..., X_N^k)$ converge en probabilité vers la mesure de Brown de $P(c_1, ..., c_k)$ où $c_1, ..., c_k$ sont des éléments circulaires $*$-libres d'un W^*-espace de probabilité.*

Dans la section 2.1, on définit la mesure de Brown qui généralise la notion de mesure spectrale à des éléments non-normaux d'un W^*-espace de probabilité et on cherche à la calculer. Pour cela on examine tout d'abord le problème du calcul de la mesure spectrale d'un polynôme auto-adjoint en des variables non-commutatives libres et on expose une méthode algorithmique reposant sur la linéarisation qui y répond. On généralise ensuite cette méthode pour pouvoir calculer cette mesure de Brown. Enfin on expose les obstacles théoriques qui empêchent d'obtenir la convergence des mesures empiriques de matrices aléatoires non-hermitiennes et les stratégies employées pour contourner ces obstacles.

Dans la section 2.2 on présente les résultats de la section 2.3, c'est-à-dire un résultat de minoration pour les plus petites valeurs singulières d'un polynômes de degré 2 en des matrices de Ginibre qui permet de démontrer la convergence des mesures empiriques de ces polynômes. Ce résultat s'appuie à la fois sur la philosophie de linéarisation employée par Haagerup et Thorbjørnsen et sur les preuve de résultat de ce type sur des matrices à coefficients indépendants dans la veine de [61, 32]. Ces résultats sont tirés de l'article en progrès [34].

2.1 Mesure de Brown et matrices non-hermitiennes

2.1.1 Linéarisation et mesure spectrale d'un polynôme en des variables libres

Le problème de la détermination de la mesure spectrale de $P(x_1, ..., x_n)$ où $x_1, ..., x_n$ sont des variables non-commutatives libres et auto-adjointes est lié à la théorie des matrices aléatoires. En effet si $(X_N^1, ..., X_N^n)_{N\in\mathbb{N}}$ une suite de n-uplets de matrices de Wigner indépendantes alors si P est auto-adjoint, la mesure empirique de $P_N = P(X_N^1, ..., X_N^n)$ converge vers la mesure spectrale de $P(s_1, ..., s_n)$ où $s_1, ..., s_n$ sont des variables semi-circulaires libres d'un C^*-espace de probabilité. On pourra se référer à l'annexe B ou à la section 5 de [4] pour un retour sur les liens entre matrices aléatoires et probabilités libres. Pour étudier les propriétés du polynôme P, on le linéarise, c'est-à-dire que l'on construit une matrice dont les coefficients sont des polynôme de degré 1 en les X_i liée à P et plus facile à étudier. Une linéarisation de $P \in \mathcal{P}$ est une matrice à coefficients dans $\mathcal{P}$ qui vérifie la définition suivante :

Définition 2.1.1. *Soit $P \in \mathcal{P}$, on dit qu'une matrice $M \in \mathcal{M}_{l+1,l+1}(\mathcal{P})$ est une linéarisation de P si les coefficients de M sont de degrés inférieurs ou égaux à 1 et :*

$$M = \begin{pmatrix} a & u \\ v & Q \end{pmatrix}$$

où $a \in \mathcal{P}, u \in \mathcal{M}_{1,l}(\mathcal{P}), v \in \mathcal{M}_{l,1}(\mathcal{P}), Q \in \mathcal{M}_{l,l}(\mathcal{P})$, Q est inversible et $P = a - uQ^{-1}v$

Une linéarisation d'un polynôme P permet de relier le spectre de P_N à celui d'une matrice linéaire en les X_N^i dont les propriétés sont plus faciles à étudier. Cette technique est utilisée par Haagerup et Thorbjørnsen afin de prouver le résultat suivant :

Théorème 2.1.2. *Soit $(X_N^1, ..., X_N^n)_{N\in\mathbb{N}}$ une suite de n-uplets de matrices indépendantes du GOE ou du GUE, $P \in \mathbb{C}\langle X_1, ..., X_n\rangle$ un polynôme non-commutatif auto-adjoint et $s_1, ..., s_n$ des variables aléatoires semi-circulaires libres d'un C^*-espace de probabilité. Alors presque sûrement :*

$$\lim_{N\to\infty} ||P(X_N^1, ..., X_N^n)|| = ||P(s_1, ..., s_n)||_\infty$$

où $||a||_\infty = \lim_{k\to\infty} \varphi(a^{2k})^{1/2k}$

L'idée de la preuve repose sur la linéarisation suivante de P définie par bloc :

$$L_P := \begin{pmatrix} 0 & V_1 & & \\ & -I_{i_2} & \ddots & \\ & & & V_{d-1} \\ V_d & & & -I_{i_d} \end{pmatrix}$$

où $V_1 \in \mathcal{M}_{1,i_1}(\mathcal{P})$, $V_1 \in \mathcal{M}_{i_1,i_2}(\mathcal{P})$,...,$V_d \in \mathcal{M}_{i_d,1}(\mathcal{P})$ avec $\mathcal{P} = C\langle X_1, ..., X_n\rangle$, $P = V_1...V_d$ et où les coefficients des V_i sont des fonctions affines en les X_i.

En posant pour $\lambda \in \mathbb{C}$

$$\Lambda(z) := \begin{pmatrix} \lambda & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}$$

on a la propriété suivante :

$$(L_P - \Lambda(z))^{-1} := \begin{pmatrix} (P-z)^{-1} & \cdots \\ \vdots & \ddots \end{pmatrix}$$

La linéarisation permet de déduire du comportement des résolvantes $(L_P(X_N^1,...,X_N^n)-\Lambda(\lambda))^{-1}$ le comportement de $\lambda \mapsto (P(X_N^1,...,X_N^n)-\lambda)^{-1}$.

On va exposer dans cette sous-section une méthode de linéarisation qui permet de calculer la mesure spectrale d'un polynôme P en des variables aléatoires non-commutatives libres. La linéarisation permet pour étudier le comportement d'un polynôme de se ramener à l'étude d'une matrice dont les blocs sont des variables aléatoires non-commutatives. Tout d'abord, la propriété de la linéarisation utilisée par Haagerup et Thorbjørnsen est en fait valable quelque soit la linéarisation de P qu'on prend. En effet, en posant alors pour tout $z \in \mathbb{C}$ $\Lambda(z)$ la matrice :

$$\Lambda(z) = \begin{pmatrix} z & 0 \\ 0 & I_l \end{pmatrix}$$

on a alors la propriété suivante sur les "résolvantes" de M_p et p :

Proposition 2.1.3. *Soit $p \in \mathcal{P}$ non constant, $z \in \mathbb{C}$ et M_p une linéarisation de p. Alors $(\Lambda(z)-M_p)$ est inversible dans $\mathcal{M}_{l+1,l+1}(Frac(\mathcal{P}))$ et $((\Lambda(z)-M_p)^{-1})_{1,1} = (z-p)^{-1}$ où $Frac(\mathcal{P})$ est le corps de fraction de $\mathcal{P}$.*

Si $\mathcal{A}$ est une $\mathbb{C}$-algèbre (unitaire) et $a_1,...,a_n \in \mathcal{A}$, en notant $\tilde{p} = p(a_1,...,a_n)$ et $\tilde{M}_p = M_p(a_1,...,a_n)$, on a que $(\Lambda(z)-\tilde{M}_p)$ est inversible si et seulement si $z-\tilde{p}$ l'est et $((\Lambda(z)-\tilde{M}_p)^{-1})_{1,1} = (z-p)^{-1}$

Cette proposition est en fait une conséquence directe de la formule du complément de Schur. On va à présent montrer que tout polynôme admet une linéarisation. Cette preuve tient aux deux observations suivantes :

1. X_i a la linéarisation $\begin{pmatrix} 0 & X_i \\ -1 & 1 \end{pmatrix}$ et le monôme $X_{i_1},\cdots,X_{i_l}$ a la linéarisation suivante :

$$\begin{pmatrix} & & & X_i \\ & & X_{i_2} & -1 \\ & \iddots & \iddots & \\ X_{i_n} & -1 & & \end{pmatrix}$$

2. Si les polynômes R,S ont les linéarisations respectives $\begin{pmatrix} a_1 & u_1 \\ v_1 & Q_1 \end{pmatrix}$, et $\begin{pmatrix} a_2 & u_2 \\ v_2 & Q_2 \end{pmatrix}$, alors $R+S$ a la linéarisation suivante :

$$\begin{pmatrix} a_1+a_2 & u_1 & u_2 \\ v_1 & Q_1 & 0 \\ v_2 & 0 & Q_2 \end{pmatrix}$$

On peut donc linéariser n'importe quel polynôme non-commutatif en empilant les linéarisations des monômes.

Remarque 2.1.4. *Cette méthode de linéarisation tirée de [14] est légèrement différente de celle exposée au début de la section employée dans [50]. Pour un polynôme donné, il existe plusieurs linéarisations possibles. Si cela n'a guerre d'importance dans l'algorithme que l'on va exposer, trouver une linéarisation ayant de bonnes propriétés est crucial pour le résultat démontré dans la section 2.3.*

Toutefois même si P est hermitien, la linéarisation L_P ne l'est pas nécessairement. Or, pour pouvoir utiliser certains outils de probabilités libres comme une généralisation matricielle de la R-transformée, il est nécessaire d'utiliser des linéarisations qui soient elles-même auto-adjointes :

Proposition 2.1.5. *On rappelle que $\mathcal{P} = \mathbb{C}\langle X_1,\cdots,X_n\rangle$. Soit $q \in \mathcal{P}$. Si q a la linéarisation $\begin{pmatrix} a & u \\ v & Q \end{pmatrix}$ alors*

$$M = \begin{pmatrix} a+a^* & u & v^* \\ u^* & 0 & Q^* \\ v^* & Q & 0 \end{pmatrix}$$

est une linéarisation de $q + q^*$

Si M_q est une linéarisation de q, on peut donc l'écrire $M_q = b_0 \otimes 1_{\mathcal{A}} + b_1 \otimes X_1 + ... + b_n \otimes X_n$ où $b_0, ..., b_n \in \mathcal{M}_{l+1,l+1}(\mathbb{C})$ où pour $m \in \mathcal{M}_{l+1,l+1}(\mathbb{C}), a \in \mathcal{A}$, $m \otimes a$ est la matrice $(m_{i,j}a)_{i,j}$ de $\mathcal{M}_{l+1,l+1}(\mathcal{A})$.

Ainsi, une fois que l'on dispose d'une linéarisation d'un polynôme P, pour obtenir la transformée de Stieljes de la loi de $P(s_1, ..., s_n)$ où $s_1, ..., s_n \in \mathcal{A}$ sont des variables semi-circulaires libres, il suffit de connaître la "transformée de Stieljes matricielle" de $\tilde{M} = M_P(a_1, ..., a_n)$, c'est à-dire la fonction $b \mapsto (\varphi(((b \otimes 1_{\mathcal{A}}) - \tilde{M})^{-1}_{i,j}))_{i,j \in [1,l+1]}$ de $\mathcal{M}_{l+1,l+1}(\mathbb{C})$ dans $\mathcal{M}_{l+1,l+1}(\mathbb{C})$ (où φ est l'espérance).

Pour étudier ce type objet, on va étendre la théorie des espaces de probabilités non-commutatifs a des espérances φ qui sont à valeurs dans une sous-algèbre $\mathcal{B} \subset \mathcal{A}$

Définition 2.1.6. *Un C^*-espace de probabilité à valeurs dans $\mathcal{B}$ est un triplet $(\mathcal{A}, \mathcal{B}, \varphi)$ où $\mathcal{A}$ est une C^*-algèbre unitaire, $\mathcal{B} \subset \mathcal{A}$ une sous-algèbre unitaire de Banach et φ une espérance conditionnelle, c'est à dire une application linéaire $\mathcal{A} \to \mathcal{B}$ complètement positive, préservant l'unité et telle que pour tout $a \in \mathcal{A}, b_1, b_2 \in \mathcal{B}$, $\varphi(b_1 a b_2) = b_1 \varphi(a) b_2$.*

Dans ce cas pour $x \in \mathcal{A}$, on définit la distribution de x comme étant la collection des fonctions multilinéaires $m_x^{(n)}$ définies de $\mathcal{B}^n \to \mathcal{B}$ par $m_x^{(n)}(b_1, ..., b_n) = \varphi(b_1 x b_2 ... x b_n)$.

Dans notre cas, si $(\mathcal{A}, \varphi)$ est l'algèbre qui contient $s_1, ...s_n$ des variables semi-circulaires libres, on prendra l'espace $\mathcal{M}_{l+1,l+1}(\mathcal{A})$ et $\mathcal{B} = \mathcal{M}_{l+1,l+1}(\mathbb{C})$. Pour espérance conditionnelle, on prendra la fonction $Id \otimes \varphi$ c'est à dire la fonction qui a une matrice d'éléments de $\mathcal{A}$ associe la matrice des images des coefficients par φ. Dans ce cas, on peut étendre la définition de la transformée de Stieljes de la façon suivante comme fonction d'un domaine de $\mathcal{B}$ dans $\mathcal{B}$.

Définition 2.1.7. *Dans le contexte de la définition précédente, on note pour $b \in \mathcal{B}$, $\Re b = (b + b^*)/2, \Im b = (b - b^*)/(2i)$ et $\mathbb{H}^+(\mathcal{B}) = \{b \in \mathcal{B} | \Im b > 0\}$ (où l'on note $b > 0$ si $\mathcal{B}$ si b est positif et inversible). Si $x \in \mathcal{A}$, on note G_x la transformée de Stieljes de x à valeur dans $\mathbb{H}^+(\mathcal{B})$:*

$$G_x(b) = \varphi((b - x)^{-1})$$

$F_x(b) = G_x(b)^{-1}$ l'inverse de la transformée de Stieljes et $h_x(b) = F_x(b) - b$

Enfin, pour étendre les résultats sur l'additivité de la R-transformée pour des variables libres, on a besoin d'étendre la notion de liberté aux espaces de probabilité à valeurs dans B.

Définition 2.1.8. *Soit $(\mathcal{A}, \mathcal{B}, \varphi)$ un C^*-espace de probabilité à valeurs dans $\mathcal{B}$ et $\mathcal{A}_1, ...\mathcal{A}_n$ des sous-algèbres contenant $\mathcal{B}$. $\mathcal{A}_1, ...\mathcal{A}_n$ sont dites libres par amalgamation sur $\mathcal{B}$ si pour tout $l \in \mathbb{N}$, $i_1, ..., i_l \in [1, n]$ tels que $i_j \neq i_{j+1}$ pour $j \in [1, l-1]$ et tout $x_k \in \mathcal{A}_{i_k}$ tels que $\varphi(x_k) = 0$ on a*

$$\varphi(x_1 ... x_l) = 0$$

$x_1, ..., x_n \in \mathcal{A}$ sont dites libres par amalgamation sur $\mathcal{B}$ si les sous-algèbres $\mathcal{B}(x_1), ..., \mathcal{B}(x_n)$ engendrées par les x_i et $\mathcal{B}$ sont libres par amalgamation.

En particulier dans notre cas, il est aisé de voir que si $x_1, ..., x_n \in \mathcal{A}$ sont libres et $b_1, ..., b_n \in \mathcal{M}_{l+1,l+1}(\mathbb{C})$ alors $b_1 \otimes x_1, ..., b_n \otimes x_n$ sont libres par amalgamation sur $\mathcal{B} = \mathcal{M}_{l+1,l+1}(\mathbb{C})$. On a alors le résultat suivant sur les transformées de Stieljes :

Proposition 2.1.9. *[14] Soit $(\mathcal{A}, \mathcal{B}, \varphi)$ un C^*-espace de probabilité à valeurs dans $\mathcal{B}$ et $x, y \in \mathcal{A}$ auto-adjointes et libres par amalgamation sur $\mathcal{B}$, il existe ω_1, ω_2 deux fonctions de $\mathbb{H}^+(\mathcal{B})$ dans $\mathbb{H}^+(\mathcal{B})$ Fréchet-différentiables telles que :*

1. *$\Im \omega_j(b) \geq \Im b$ pour $j = 1, 2$ et $b \in \mathbb{H}^+(\mathcal{B})$*
2. *Pour tout $b \in \mathbb{H}^+(\mathcal{B})$, $F_x(\omega_1(b)) + b = F_y(\omega_2(b)) + b = \omega_1(b) + \omega_2(b)$*
3. *Pour tout $b \in \mathbb{H}^+(\mathcal{B})$, $G_x(\omega_1(b)) = G_y(\omega_1(b)) = G_{x+y}(b)$*

De plus pour $b \in \mathbb{H}^+(\mathcal{B})$, $\omega_1(b)$ est l'unique point fixe de f_b définie par

$$\forall w \in \mathbb{H}^+(\mathcal{B}), f_b(w) = h_y(h_x(w) + b) + b$$

Enfin on a $\omega_1(b) = \lim_{n\to\infty} f^{\circ n}(w)$ pour tout $w \in \mathbb{H}^+(\mathcal{B})$ on $f^{\circ n}$ désigne les itérées de f_b.

On peut voir ce résultat comme une généralisation de l'additivité de la R-transformée dans le cas classique.

On peut alors répondre au problème du calcul de la distribution de $P(a_1, ..., a_n)$ où P est un polynôme auto-adjoint et $a_1, ..., a_n \in \mathcal{A}$ des variables non-commutatives auto-adjointes libres en appliquant l'algorithme suivant (tiré de [14]) :

Etape 1 : On se donne une linéarisation auto-adjointe M_P de P et l'on écrit $\tilde{M}_P = M_P(a_1, ..., a_n) = b_0 \otimes 1_{\mathcal{A}} + ... + b_n \otimes a_n$.

Etape 2 : Pour tout $j \in [1, n]$, on peut calculer la transformée de Stieljes matricielle de $b_j \otimes a_j$ grâce à la distribution μ_j de a_j du fait que

$$G_{b_j\otimes a_j}(b) = \int (b - tb_j)^{-1} d\mu_j(t) = \lim_{\epsilon` \to 0} \frac{-1}{\pi} \int (b - tb_j)^{-1} \Im(G_{\mu_j}(t + i\epsilon))dt$$

Etape 3 : Comme $b_0 \otimes 1, ..., b_n \otimes a_n$ sont libres par amalgamation, on calcule la transformée de Stieljes matricielle de $\tilde{M}_P$ en se servant de la proposition précédente.

Etape 4 : Comme M_P est une linéarisation de P, on peut finalement écrire avec $p = P(a_1, ..., a_n)$ et $z \in \mathbb{C}$ tel que $\Im z > 0$

$$G_p(z) = (G_{\tilde{M}_P}(\Lambda(z)))_{1,1} = \lim_{\epsilon\to 0} (G_{\tilde{M}_P}(\Lambda_\epsilon(z)))_{1,1}$$

Où :

$$\Lambda_\epsilon(z) = \begin{pmatrix} z & \\ & i\epsilon I_l \end{pmatrix} \in \mathbb{H}^+(\mathcal{M}_{l+1,l+1}(\mathbb{C}))$$

On a donc ainsi répondu à la question de la détermination de la distribution de $P(a_1, ..., a_n)$ et donc à la question du calcul de la limite de la distribution empirique de $P(X_N^1, ..., X_N^n)$ pour $X_N^1, ..., X_N^n$ des matrices de Wigner dans le cas hermitien. Dans les deux sections suivantes on va s'intéresser au même problème dans le cas où P est non-hermitien.

2.1.2 Mesure de Brown

Par analogie avec le cas hermitien, dans le cas d'un polynôme non-hermitien en des matrices de Wigner $P(X_N^1, ..., X_N^n)$, on aimerait pouvoir dire que la suite des mesures empiriques converge vers la mesure spectrale de $P(s_1, ..., s_n)$ où $s_1, ..., s_n$ sont des éléments semi-circulaires libres. Toutefois comme $P(s_1, ..., s_n)$ n'est pas à priori normal, sa mesure spectrale n'est pas bien définie. On va donc d'abord définir dans le cas d'un élément non-normal un analogue de la mesure spectrale, la mesure de Brown. Pour cela on se placera dans le cadre d'un W^*-espace de probabilité, c'est-à-dire d'une algèbre de Von-Neumann munie d'une trace que l'on supposera fidèle et normale. On va également généraliser la méthode de linéarisation précédente pour la calculer.

Définition 2.1.10. *Soit $(\mathcal{V}, \varphi)$ un W^*-espace de probabilité. pour $x \in \mathcal{V}$ on définit le déterminant de Flugede-Kadison de x par :*

$$\log \Delta(x) = \varphi(\log(xx^*))$$

Si x n'est pas inversible, la notation précédente doit être interprétée comme la limite de $\log \Delta_\epsilon(x) = \varphi(\log(xx^* + \epsilon^2))$ quand $\epsilon \to 0$. On appelle mesure de Brown *de x et on note μ_x le laplacien suivant :*

$$\mu_x = \frac{1}{2\pi} \left(\frac{\partial^2}{\partial \Re z} + \frac{\partial^2}{\partial \Im z} \right) \log \Delta(x - z)$$

où ce laplacien est pris au sens des distribution.

Étant donné que $z \mapsto \log \Delta(x-z)$ est sous-harmonique, μ_x est bien une mesure et on peut de plus montrer qu'il s'agit bien d'une mesure de probabilité (voir [25]). Pour x un élément normal, on a par calcul fonctionnel que $\log \Delta(x-z) = \int \log |x-z| d\nu_x(z)$ où ν_x est la mesure spectrale de x et donc μ_x coïncide bien avec la mesure spectrale de x. On peut de plus approximer μ_x de la façon suivante. Pour $\epsilon > 0$, on définit $G_{\epsilon,x}$ l'approximation de la transformée de Stieljes de la façon suivante :

$$G_{\epsilon,x}(z) = \varphi((z-x)^*((z-x)(z-x)^* + \epsilon^2)^{-1})$$

et le déterminant approché Δ_ϵ de la façon suivante :

$$\log \Delta_\epsilon = \varphi(\log(xx^* + \epsilon^2))$$

On a alors la formule suivante pour approximer la mesure de Brown de x

$$\mu_{\epsilon,x} = \frac{1}{2\pi}\left(\frac{\partial^2}{\partial \Re z^2} + \frac{\partial^2}{\partial \Im z^2}\right) \log \Delta_\epsilon(x-z) = \frac{1}{\pi}\frac{\partial}{\partial z} G_{\epsilon,x}(z)$$

On peut alors écrire :

$$\left[\begin{pmatrix} i\epsilon & z \\ \bar{z} & -i\epsilon \end{pmatrix} - \begin{pmatrix} 0 & x \\ x^* & 0 \end{pmatrix}\right]^{-1} = \begin{pmatrix} -i\epsilon\left((z-x)(z-x)^* + \epsilon^2\right)^{-1} & (z-x)\left((z-x)^*(z-x) + \epsilon^2\right)^{-1} \\ (z-x)^*\left((z-x)(z-x)^* + \epsilon^2\right)^{-1} & -i\epsilon\left((z-x)^*(z-x) + \epsilon^2\right)^{-1} \end{pmatrix}$$

De cette manière, si l'on est capable de calculer la transformée de Stieljes matricielle suivante allant de $\mathbb{H}^+(\mathcal{M}_{2,2}(\mathbb{C}))$ dans $\mathbb{H}^+(\mathcal{M}_{2,2}(\mathbb{C}))$:

$$b \mapsto Id \otimes \varphi\left(b - \begin{pmatrix} 0\epsilon & x \\ x^* & 0\epsilon \end{pmatrix}\right)^{-1}$$

on peut alors déterminer $G_{\epsilon,x}$ en regardant le coefficient $(2,1)$. Ainsi, si $x = P(a_1, ..., a_n)$ un polynôme non auto-adjoint, il suffit de linéariser la matrice :

$$\begin{pmatrix} 0 & P(a_1, ..., a_n) \\ P^*(a_1, ...a_n) & 0 \end{pmatrix}$$

On peut a partir de ce point reprendre la même méthode que dans le cas auto-adjoint et il suffit alors de linéariser les "monômes" matriciels de la forme :

$$\begin{pmatrix} 0 & a_{i_1}...a_{i_l} \\ a_{i_l}...a_{i_1} & 0 \end{pmatrix}$$

puis d'empiler ces linéarisations.

Or, on peut écrire en posant

$$X_j = \begin{pmatrix} 0 & a_j \\ a_j & 0 \end{pmatrix}, Y_j = \begin{pmatrix} a_j & 0 \\ 0 & 1 \end{pmatrix},$$

$$\begin{pmatrix} 0 & a_{i_1}...a_{i_l} \\ a_{i_l}...a_{i_1} & 0 \end{pmatrix} = Y_1...Y_{k_1}X_kY_{k_1}^*...Y_1$$

On obtient alors une linéarisation auto-adjointe et on peut donc réutiliser les outils de la section précédente pour calculer la transformée de Stieljes matricielle de

$$\begin{pmatrix} 0 & P(a_1, ..., a_n) \\ P^*(a_1, ...a_n) & 0 \end{pmatrix}$$

en fonction des transformées de Stieljes des a_i.

2.1.3 Mesures empiriques de matrices non-hermitiennes

Tout d'abord, rappelons que les moments d'une variable aléatoire non commutative x qui n'est pas normale ne permettent pas de définir directement une loi sur $\mathbb{C}$ étant donné que x et x^* peuvent ne pas commuter. Ainsi, si $P_n = P(X_N^1, ..., X_N^n)$ est un polynôme non-hermitien en des matrices de Wigner, on a la convergence de P_n en $*$-moments, c'est-à-dire la convergence de la distribution non-commutative de (P_n, P_n^*) mais pas a priori la convergence de la mesure empirique de P_n. Pire, il existe des suites de matrices M_n convergeant en $*$-moments vers une variable aléatoire non-commutative normale M mais telles que les mesures empiriques μ_{M_n} convergent vers une limite μ distincte de la mesure spectrale de μ. Prenons M_n la suite suivante :

$$M_n = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}$$

On a alors que M_n converge en $*$-moments vers un élément unitaire dont la mesure spectrale est la mesure uniforme sur le cercle. Toutefois $\mu_{M_n} = 0$ pour tout $n \in \mathbb{N}$.

Pour démontrer des résultats de convergence de mesures empiriques d'une suite de matrices aléatoires M_n vers une certaine mesure μ, on utilise alors la suite des potentiels empiriques $h_N : z \mapsto \int_{w\in\mathbb{C}} \log|z - w| d\mu_{M_N}(w)$. Ces preuves procèdent alors en plusieurs étapes :

1. On prouve que pour presque tout $z \in \mathbb{C}$, h_N converge en probabilité vers une certaine fonction h.
2. On prouve qu'en probabilité, Δh_N converge vers Δh au sens des distributions, c'est à dire que pour tout fonction ψ $\mathcal{C}^\infty$ à support compact, on a en probabilité :
$$\lim_{N\to\infty} \int_{z\in\mathbb{C}} \Delta\psi(z) h_N(z) dz = \int_{z\in\mathbb{C}} \Delta\psi(z) h(z) dz$$
3. On identifie h comme étant le potentiel associé à la mesure μ, i.e. $h(z) = \int_{w\in\mathbb{C}} \log|z - w| d_\mu(w)$

Le point (1) est le plus technique et sa démonstration requiert des bornes sur le comportement des petites valeurs singulières de $z - M_n$. En effet, on a

$$h_N(z) = \frac{1}{N} \log|\det(M_N - z)| = \frac{1}{2N} \log|\det((M_N - z)^*(M_N - z))| = \frac{1}{2} \int_{z\in\mathbb{C}} \log|x| d\nu_N^z(x)$$

où ν_N^z est la mesure empirique de $(M_N - z)^*(M_N - z)$. Étant donné que les $(M_N - z)^*(M_N - z)$ sont auto-adjoints, on peut alors utiliser une méthode des moments pour prouver que les ν_N^z converge faiblement en probabilité vers une famille de mesures $(\nu^z)_{z\in\mathbb{C}}$. On veut alors prouver que :

$$\lim_{N\to\infty} \frac{1}{2} \int_0^\infty \log|x| d\nu_N^z(x) = \frac{1}{2} \int_0^\infty \log|x| d\nu^z(x)$$

en probabilité. Comme log n'est pas une fonction bornée, la convergence faible ne permet pas de conclure immédiatement, on a en particulier besoin de contrôler le comportement de ν_N^z près de la singularité du log en $+\infty$, ce qui est en général aisé car cela revient à borner en probabilité la norme de M_N, et en 0, ce qui revient à regarder le comportement des petites valeurs singulières de $M_N - z$. C'est cette dernière estimée qui est la plus ardue, et qui est en général la clé pour démontrer la convergence de la mesure spectrale de matrices aléatoires. C'est grâce à ce type d'argument que Tao et Vu prouvent le résultat suivant pour des perturbations de matrices à coefficients i.i.d. [65][Corollary 1.15] :

Théorème 2.1.11. *Soit X_N une suite de matrices aléatoires dont les coefficients sont i.i.d., d'espérance 0 et de variance 1 et A_N une suite de matrices déterministes de rang $o(N)$ et telles que $\sup_N N^{-2}||A_N||_2^2 < \infty$. On pose $M_N = X_N + A_N$. Alors la mesure empirique de $M_N/\sqrt{N}$ converge presque sûrement vers la mesure uniforme sur le disque unité.*

Le point crucial est la preuve que pour tout $B > 0$ il existe de $A, C > 0$ telles que :

$$\mathbb{P}[\sigma_{\min}(M_N - z) \leq N^{-A}] \leq CN^{-B}$$

où $\sigma_{\min}(M_N - z)$ est la plus petite valeur singulière de $(M_N - z)$. On a en effet $\{\sigma_{\min}(M_N - z) \leq N^{-A}\}$:

$$\int_0^\infty \log|x| d\nu_N^z(x) = \int_{N^{-A}}^\infty \log|x| d\nu_N^z(x)$$

et on peut alors écrire :

$$|\int_{N^{-A}}^\infty \log|x| d\nu_N^z(x) - \int_{N^{-A}}^\infty \log|x| d\nu^z(x)| \leq A \log N \sup_{x \in \mathbb{R}^+} |F_N(x) - F(x)|$$

où $F(x) = \nu^z([0, x])$ et $F_N(x) = \nu_N^z([0, x])$. Et il reste à montrer que $\sup_{x\in\mathbb{R}^+} |F_N(x) - F(x)| = o(N^{-c})$ en probabilité pour n'importe quel c ce qui est possible grâce au résultat de [12].

On peut déduire du comportement de la fonction potentiel h des propriétés sur la mesure limite μ. C'est par exemple ce que font Guionnet, Krishnapur et Zeitouni dans [46] où ils prouvent la convergence des mesures empiriques pour des matrices du type $U_N T_N V_N$ où U_N et V_N sont des matrices de Haar indépendantes, et où les mesures empiriques des T_N convergent. Ils font de plus apparaître à la limite un phénomène de "single ring", c'est-à-dire que la limite μ est à support dans une couronne du plan complexe $\{z \in \mathbb{C} : 0 < a \leq |z| \leq b < 0\}$:

Théorème 2.1.12. *[46, Theorem 1] Soit $(U_N), (V_N)$ des matrices aléatoires unitaires de Haar indépendantes et (T_N) une suite de matrices aléatoires diagonales réelles indépendantes de U_N, V_N telles que :*

1. *La suite des mesures empiriques μ_{T_N} converge faiblement vers une mesure de probabilité Θ supportée sur $\mathbb{R}$*
2. *Il existe une constante $M > 0$ telle que :*

$$\lim_{N\to\infty} \mathbb{P}[||T_N|| \geq M] = 0$$

3. *Il existe une suite d'événements $(\mathcal{G}_N)_{N\in\mathbb{N}}$ telle que $\lim_{N\to\infty} \mathbb{P}[\mathcal{G}_N^c] = 0$ et des constantes $\delta, \delta' > 0$ telles que*

$$\mathbb{E}[\mathbb{1}_{G_N} (\log \sigma_n^z)^2 \mathbb{1}_{\sigma_n^z < n^{-\delta}}] \leq \delta'$$

où σ_n^z est la plus petite valeur singulière de $z - U_N T_N V_N$.

4. *Il existe des constantes $\kappa_1, \kappa_2 > 0$ telles que :*

$$|\Im \mathbb{E}[G_{\mu_{T_N}}(z)]| \leq \kappa_1 \text{ sur } \{z \in \mathbb{C} : \Im(z) \geq N^{-\kappa_2}\}$$

alors, en posant $A_N = U_N T_N V_N$, on a qu'en probabilité, μ_{A_N} converge faiblement vers une mesure de probabilité μ_A qui vérifie les propriétés suivantes :

1. *μ_A a une densité radialement symétrique ρ_A par rapport à la mesure de Lebesgue sur $\mathbb{C}$ où $\rho_A = \Delta_z(\int \log|x| d\nu^z(x))$ où $\nu^z = \tilde{\Theta} \boxplus \lambda_{|z|}$ où $\tilde{\Theta}$ et $\lambda_{|z|}$ sont les symétrisées de Θ et du Dirac $\delta_{|z|}$ par rapport à 0.*
2. *le support de μ_A est une couronne de la forme $\{z \in \mathbb{C} : a \leq |z| \leq b < \infty\}$ et $a = 0$ si et seulement si $\int x^{-2} d\Theta(x) = \infty$.*

La conjecture suivante est formulée par Belinschi, Sniady et Speicher dans [15].

Conjecture : Si $(X_N^1, ..., X_N^n)$ est une suite de n-uplet de matrices de Wigner indépendantes et $P \in \mathbb{C}\langle X_1, ..., X_n\rangle$. En posant $P_N = P(X_N^1, ..., X_N^n)$, on a que μ_{P_N} converge en probabilité vers μ_p où $p = P(s_1, ..., s_2)$ et $s_1, ..., s_n$ sont des éléments semi-circulaires libres.

Les résultats de la section 2.3 présentent dans un cas particulier une réponse affirmative à cette conjecture.

2.2 Présentation des résultats et perspectives

2.2.1 Présentation des résultats de la section 2.3

Dans la section suivante, on présente un résultat de convergence des mesures empiriques dans le cas d'un polynôme $P_N = P(G_N^1, ..., G_N^n)$ de degré 2 en des matrices de Ginibre (c'est-à-dire des matrices dont les coefficients sont des variables aléatoires gaussiennes i.i.d.) vers la mesure de Brown de $P(c_1, ..., c_n)$ où $c_1, ..., c_n$ sont des éléments circulaires libres en $*$-moments (i.e $c_j = s_{2j-1} + is_{2j}$ où $s_1, ..., s_{2n}$ sont des semi-circulaires libres). Pour ce faire on va comme dans [64] prouver que pour tout $z \in \mathbb{C}$, il existe $\beta, \beta' > 0$ tels que :

$$\mathbb{P}[\sigma_{\min}(z - P_N) \leq N^{-\beta}] \leq N^{-\beta'}$$

Pour démontrer ce résultat, on va suivre l'esprit de la preuve de [61] qui prouve une borne similaire pour des matrices X_N ayant des coefficients indépendants, de variance au moins 1 et de quatrième moment borné. La démonstration de notre résultat se décompose comme suit :

1. On réduit notre polynôme sous la forme $P = \sum_{j=1}^k R_j S_j + T$ où les R_j les S_j et T sont des des formes linéaires de rang k. On aura également également, en identifiant naturellement $Vect(X_1, ..., X_n)$ à l'espace euclidien $\mathbb{R}^n$, que les $\{R_j\}$ soit un famille de vecteurs unitaires orthogonaux. On peut alors faire un changement de variables et remplacer les $R_1, ..., R_k$ par $X_1, ..., X_k$ en utilisant les propriétés des vecteurs gaussiens.

2. On se sert de cette réduction pour linéariser P en une matrice L qui est $(k+1)N \times (k+1)N$. On peut alors considérer cette linéarisation comme une matrice de $N \times N$ blocs $(k+1) \times (k+1)$ indépendants.

3. En utilisant un argument d'incompressibilité similaire à [61], on se ramène à l'obtention d'une borne inférieure sur le déterminant $\det(Q_{(1)}^* L_{1\bullet})$, où $L_{1\bullet}$ est le premier bloc de colonnes de L et $Q_{(1)}$ un système orthonormal de k vecteurs orthogonaux à l'espace des colonnes de la matrice L à laquelle on a retiré les colonnes de $L_{1\bullet}$.

4. On remarque alors que comme $Q_{(1)}$ et $L_{1\bullet}$ sont indépendants, conditionnellement à $Q_{(1)}$, la matrice $k \times k$ $Q_{(1)}^* L_{1\bullet}$ est un vecteur gaussien que l'on peut représenter sous la forme $H\Xi + C$ où Ξ est un vecteur gaussien standard et C un vecteur constant. $\det(Q_{(1)}^* L_{1\bullet})$ est donc un polynôme en des variables gaussiennes. L'inégalité de Carbery-Wright stipule que pour tout polynôme réel $f(g_1, ..., g_n)$ de degré d en des variables gaussiennes standard i.i.d. telles que $\mathrm{Var}(f(g_1, ..., g_n)) = 1$, on a

   $$\sup_{t \in \mathbb{R}} \mathbb{P}[|f(g_1, ..., g_n) - t| \leq \delta] \leq C(d)\delta^{1/d}$$

   Enfin, en écrivant f dans la base de Hermite, on peut écrire, si un des coefficients dominants de f est égal à a que

   $$\mathrm{Var} f(g_1, ..., g_n) \geq c(d)|a|^2$$

   Ainsi en écrivant $\det(Q_{(1)}^* L_{1\bullet})$ sous la forme $f(g_1, ..., g_n)$ il suffit de montrer qu'avec grande probabilité au moins un des coefficients dominants de f est de l'ordre au moins de $N^{-\alpha}$ pour ensuite prendre $\delta = N^{-\alpha}$ avec α assez grand.

5. On calcule la matrice H et le polynôme f explicitement. On a alors que les coefficients dominants de f s'expriment comme des déterminants $\Delta_j(Q)$ en les colonnes des blocs de $Q_{(1)}$ (où j décrit un certain ensemble d'indices J). On veut alors prouver qu'il existe un $\alpha > 0$ tel que l'événement $\{\forall j \in J, \Delta_j(Q) \leq N^{-\alpha}\}$ soit de faible probabilité.

6. De manière générale, pour prouver que la probabilité que Q soit dans un certain ensemble $\mathcal{E}$ est faible, on procède comme suit. On se donne pour $\delta > 0$ un δ-réseau $\mathcal{N}(\delta)$ de $\mathcal{E}$ pour la norme euclidienne. En notant $L_{j\bullet}$ les blocs des colonnes de la linéarisation L, on a par définition de $Q_{(1)}$ que $Q_{(1)}^* L_{j\bullet} = 0$ pour tout $j \geq 2$. On a donc

$$\begin{aligned}\mathbb{P}[Q_{(1)} \in \mathcal{E}] &= \mathbb{P}[Q_{(1)} \in \mathcal{E}, \forall j \geq 2 ||Q_{(1)}^* L_{j\bullet}||_2 = 0] \\ &\leq \sum_{Q' \in \mathcal{N}(\delta)} \mathbb{P}[\forall j \geq 2 ||Q'^* L_{j\bullet}||_2 \leq \delta] + O(e^{-cN}) \\ &\leq |\mathcal{M}_\beta(\delta)| \prod_{j=2}^{N} \mathbb{P}[||Q'^* L_{j\bullet}|| \leq \delta] + O(e^{-cN})\end{aligned}$$

Où l'on s'est servi du fait qu'avec probabilité $1 - O(e^{-cN})$, la norme d'opérateur de la matrice L est bornée, puis de fait que les $L_{j\bullet}$ sont indépendants. Pour borner $\mathbb{P}[||Q'^* L_{j\bullet}||_2 \leq \delta]$, on utilise la borne d'anti-concentration suivante valable pour tout vecteur gaussien standard Ξ dans $\mathbb{C}^d$ et toute matrice $M \in \mathcal{M}_{d,d}(\mathbb{C})$:

$$\sup_{C \in \mathbb{C}^d} \mathbb{P}[||M\Xi + C||_2 \leq \epsilon] \leq \frac{(K\epsilon)^d}{\det(MM^*)}$$

On peut alors écrire, si Π est un projecteur orthogonal sur un espace de dimension d

$$\mathbb{P}[Q_{(1)} \in \mathcal{E}] \leq |\mathcal{M}_\beta(\delta)| \sup_{Q_{(1)} \in \mathcal{E}} \{ \frac{(C\delta)^d}{\det(\Pi H H^* \Pi)} \}$$

On a donc une compétition entre l'entropie de l'ensemble $\mathcal{E}$ à travers le cardinal $|\mathcal{M}_\beta(\delta)|$ et la borne d'anti-concentration en $\frac{(C\delta)^d}{\det(\Pi H H^* \Pi)}$.

7. Dans un premier temps, on cherche à minimiser la borne d'anti-concentration, c'est a dire qu'on veut se ramener au cas où $\det(HH^*) \geq N^{-C_1}$ pour un $C_1 > 0$ arbitraire. Pour cela on fait une disjonction de cas selon le rang effectif de H, c'est à dire l'entier maximum d tel qu'il existe un projecteur Π de rang d tel que $\det(\Pi H H^* \Pi) \geq N^{-C_2}$. Pour chaque rang $d < (k+1)^2$ fixé, on constate alors que la borne d'anti-concentration bat l'entropie. C'est en particulier à cet endroit de la preuve que sert crucialement la forme sous laquelle on a réduit P. Celle-ci donne en effet une matrice H dont le rang est déjà borné inférieurement a priori.
8. S'étant donc ramené au cas $d = (k+1)^2$ maximal, on se sert des contraintes $\Delta_j(Q) \leq N^{-\alpha}$ avec α assez grand pour réduire l'entropie du δ-réseau de façon à ce qu'elle soit de nouveau battue par la borne d'anti-concentration. On prouve donc qu'avec probabilité $1 - O(e^{-cN})$ un des déterminants $\Delta_j(Q)$ est supérieur à $N^{-\alpha}$. On conclut alors en se servant l'inégalité de Carbery-Wright.

Une fois cette borne démontrée, on conclut à la convergence des mesures empiriques en utilisant la stratégie exposée dans la section précédente.

2.2.2 Perspectives

Le fait de prendre un polynôme de degré 2 permet de prendre une réduction qui, une fois linéarisée donne H qui est a priori bien conditionnée, ce qui n'est pas automatique lorsque l'on prend une linéarisation quelconque pour un polynôme quelconque. Pour pouvoir espérer généraliser la méthode précédente à d'autres polynômes de degré quelconque, trouver une linéarisation L qui puisse donner une matrice H bien conditionnée est donc crucial.

Si l'on veut d'autre part pouvoir traiter le cas de polynômes en des matrices du GUE, l'autre obstacle à franchir est alors le fait que la linéarisation d'un tel polynôme ne conserverait pas l'indépendance des colonnes qui est centrale pour pouvoir affirmer que $Q_{(1)}$ est indépendant de $L_{(1\bullet)}$.

2.3 Lower bound on the smallest singular value for polynomials of degree 2 in Ginibre matrices

2.3.1 Introduction

A central question in random matrix theory is the convergence of the empirical measure. The seminal works of Wigner, Marschenko and Pastur answer this question in the case of Wigner matrices and Wishart matrices via a moment method. For random matrix models involving several independent Wigner matrices, Voiculescu showed that the theory of free probability gives powerful tools to understand the limit of the empirical measure of a non-commutative polynomial in those matrices [69]. Indeed, if $X_1^N, ..., X_n^N$ are independent Wigner matrices and P is a non-commutative self-adjoint polynomial, then the empirical measure of $P(X_1^N, ..., X_n^N)$ converges towards the spectral distribution of $P(s_1, ..., s_n)$ where $s_1, ..., s_n$ are n freely independent semi-circular elements of a Von Neumann Algebra. The problem is more involved in the case of a non-hermitian polynomial since the convergence in $*$-moments does not yield the convergence of the empirical measure. In this case, the analogue of the spectral distribution for a non-normal element of a Von Neumann Algebra is the Brown measure and the result one can expect for usual matrix models is the convergence of the empirical measure toward the Brown measure of $P(s_1, ..., s_n)$. There has been recent progress by Speicher, Mai, Belinschi and Sniady who found an algorithm that gives such Brown measures using linearization techniques and subordination results [14, 15]. In the case of a matrix with i.i.d coefficients this result is known as Girko's circular law [42] which was proven in [44, 64]. In order to prove this kind of result, one needs to control the behavior of the smallest singular value of $(P - z)(P - z)^*$ for almot every z. In this paper we prove such a result for a polynomial of degree 2 in independent Ginibre matrices. Our proof uses the so-called linearization trick incepted by Haagerup and Thorbjørnsen in [50] and then we use the same ideas as in [61] and [32]. As a consequence, we will also prove that the empirical measure of degree 2 polynomial in Ginibre converges toward the Brown measure of the associated polynomial in free circular elements.

2.3.2 Staments of the results and overview of the proof

We recall the following definition of the complex Ginibre ensemble :

Définition 2.3.1. *A random matrix X^N is a complex $N \times N$ Ginibre matrix if the families $(\sqrt{N}\Re(X^N)_{i,j})_{i,j\in[1,N]}$ and $(\sqrt{N}\Im(X^N)_{i,j})_{i,j\in[1,N]}$ are independent i.i.d. families of random variables of law $\mathcal{N}(0, 1/2)$*

We will prove the following lower bound for the smallest singular value of $P(X_1^N, X_2^N ..., X_n^N)$.

Theorem 2.3.2. *Let $X_1^N, X_2^N ..., X_n^N$ be n independent complexs Ginibre matrices and $P \in \mathbb{C}\langle X_1, X_2, ..., X_N\rangle$ a (non-commutative) polynomial of degree 2. For almost every $z \in \mathbb{C}$, There are $\beta > 0$ and $\beta' > 0$ such that :*

$$\mathbb{P}[\sigma(P(X_1^N, ..., X_n^N) - z) \leq N^{-\beta}] = o(N^{-\beta'})$$

where $\sigma(M)$ is the smallest singular value of M.

From this bound, we will prove the following convergence result for the empirical measure

Corollary 2.3.3. *With the notation of the preceding theorem, we have that the empirical measure of $P(X_1^N, ..., X_n^N)$ converges almost surely toward the Brown measure of $P(c_1, ..., c_n)$ where $c_1, ..., c_n$ are n $*$-free independent circular variables of a W^*-probability space.*

The demonstration will proceed in the vein of [61] as follows. First in the section 2.3.4 we reduce P in a way that will be useful later by maximizing independence. In section 2.3.5 we will use a linearization trick similar to [50] to linearize $z - P$ into a larger block matrix $L(z)$ whose smallest singular value is smaller than the smallest singular value of $z-P$. In section 2.3.6 we use the techniques of [61] to show that it suffices to have a lower bound for the probability $\mathbb{P}\left(\{|\det(Q_{(1)}^*(L_{1\bullet}(z)))| \leq \varepsilon\} \cap \{\sigma(Q_1) \geq N^{-C_1}\}\right)$, where $(L_{1\bullet}(z))$ is a block of columns from $L(z)$ and $Q_{(1)}$ a system of columns perpendicular to those of $L(z)$ where we removed $(L_{1\bullet}(z))$.In section 2.3.7 we will prove via anti-concentration inequalities and δ-nets that we can restrict ourselves to the case where $Q_{(1)}^* L_{1\bullet}$ conditioned on $Q_{(1)}$ is a Gaussian vector essentially of maximal range (that is that its covariance matrix is well-conditioned). Finally in section 2.3.8, seeing $\det(Q_{(1)}^*(L_{1\bullet}(z)))$ as a polynomial in Gaussian variable we will use Carbery-Wright inequality to conclude.

The proof of Corollary 2.3.3 will be delayed to the last section and we will use Theorem 2.3.2 to prove the convergence of the logarithmic potentials $z \mapsto \int \log|z-w| d\mu_{P_n}(w)$ for almost every z in probability toward the logarithmic potential of the Brown measure and therefore deduce the convergence of the measures in probability.

2.3.3 Notations

Note :

— We will say that a matrix depending on N is well conditionned if its smallest singular value is greater than $N^{-\alpha}$ for some α independent of N.
For $A \in \mathcal{M}_{n,m}(\mathbb{C}), b \in \mathcal{M}_{p,q}(\mathbb{C})$, we will denote $A \otimes B$ the following matrix of $\mathcal{M}_{np,mq}(\mathbb{C})$:

$$A \otimes B := \begin{pmatrix} a_{1,1}B & a_{1,2}B & \dots & a_{1,m}B \\ a_{2,1}B & a_{2,2}B & \dots & a_{2,m}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1}B & a_{n,2}B & \dots & a_{n,m}B \end{pmatrix}.$$

— $\mathbb{C}\langle X_1, \dots X_n\rangle$ is the set of polynomials in n non-commutative indeterminates.

— $\sigma(M)$ denotes the smallest singular value of a matrix M.

2.3.4 Reduction

Lemma 2.3.4. *Let n be an integer, $X_1, \cdots, X_n$ be independent Ginibre. Let P be a polynomial in $\mathbb{C}\langle X_1, \dots, X_n\rangle$ of degree 2. Then there exists two families of independent linear forms $(R_i)_{i\le k}$ and $(S_i)_{i\le k}$ and a polynmial T of degree 1 or less such that*

$$P = \sum_{i=1}^{k} R_i S_i + T.$$

Moreover, up to a unitary transformation for the variables $X_1, X_2, ..., X_n$, we may assume that $Vect(R_i, i \le k) = \mathbb{C}\langle X_1, ..., X_k\rangle$, (R_i) is orthonormal and the family (S_i) is orthogonal.

In the sequel, given the orthonormality and the properties of Gaussian variables, we may and will replace the R_i by the X_i.

Proof. Indeed, it suffices to show that if P is homogeneous of degree 2, it can be written as a sum $\sum_{i=1}^k R_i S_i$. But we can write

$$P(X_1, \dots, X_n) = \sum_{i,j=1}^{n} M_{ij} X_i X_j$$

where $M = UDV$ by the polar decomposition: U, V are unitaries and D is diagonal with entries given by the singular values s of M in decreasing order. Hence

$$P(X_1, \dots, X_n) = \sum_{l=1}^{n} s_l (\sum_{i=1}^{n} u_{il} X_i)(\sum_{j=1}^{n} v_{lj} X_j)$$

Putting $R_l = \sum_{i=1}^n u_{il} X_i, S_l = s_l \sum_{j=1}^n v_{lj} X_j$ yields the claim where k is the rank of the matrix M. $\square$

2.3.5 Linearization

Let us fix a degre two polynomial P

$$P = \sum_{i=1}^{k} R_i S_i + T$$

Let us introduce the following $(k+1)N \times (k+1)N$ linearization matrix

$$L(z) = \begin{pmatrix} -z+T & R_1 & \cdots & R_k \\ S_1 & -Id & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ S_k & \cdots & \cdots & -Id \end{pmatrix}.$$

Lemma 2.3.5. *For any complex z*

$$\|L(z)^{-1}\|_{op} \geq \|(z-P)^{-1}\|_{op}$$

Proof. We use the Schur complement formula

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I & 0 \\ D^{-1}C & I \end{pmatrix}$$

with $M = L(z)$ and observe that when we take the inverse, the block in position $(1,1)$ is $(z-P)^{-1}$. □

2.3.6 Reduction to invertibility of bounded-dimensional projections

We set $D = k+1$. Then we can view $L(z) = M(z)$ as an element of $\mathcal{M}_D(\mathcal{M}_N(\mathbb{C}))$ or $\mathcal{M}_N(\mathcal{M}_D(\mathbb{C}))$. We denote W_j the vector space spanned by the jth column vector of the D blocks of $M(z)$: these are the column vector

$$L_{j\bullet}(z) = (L_{j,i}(z))_{1\leq i\leq DN}, (L_{N+j,i}(z))_{1\leq i\leq DN}, \cdots (L_{N(D-1)+j,i}(z))_{1\leq i\leq DN}.$$

It is a subspace of $\mathbb{C}^{DN}$. Let $Q_{(j)} \in \mathcal{M}_{DN,D}$ be a matrix whose columns are an orthonormal basis of the orthocomplement of W_j, the span of the $(N-1)D$ columns of $L(z)$ outside $L_{j\bullet}$, and independent of $L_{j\bullet}(z)$. Then, if $\sigma(M)$ denotes the smallest singular value of a matrix M,

Lemma 2.3.6. *There exists a constant $C < \infty$ such that for any $\varepsilon > 0$,*

$$\mathbb{P}(\sigma(L(z)) \leq \varepsilon) \leq \exp\{-N\} + N\mathbb{P}\left(|\det(Q^*_{(1)}(L_{1\bullet}(z)))| \leq \varepsilon(|z|+C)N^{1/2}\right).$$

Proof. Assume that $\sigma(L(z)) \leq \varepsilon$, then there exists $q \in \mathbb{C}^{D\times N}$ with norm one such that

$$\|L(z)q\|_2 \leq \varepsilon.$$

Writing $q = (q_1, \ldots, q_N)$ and noticing that there exists a $j \in \{1, \ldots, N\}$ such that $\|q_j\|_2 \geq 1/\sqrt{N}$ and therefore

$$\|L(z)q\|_2 = \|\sum_{j=1}^{N} L_{j\bullet}(z)q_j\|_2 \geq \|Q^*_{(j)}L_{j\bullet}(z)q_j\|_2 \geq \sigma(Q^*_{(j)}(L_{j\bullet}(z))))/\sqrt{N}$$

whereas

$$|\det(Q^*_{(j)}L_{j\bullet}(z))| \leq \sigma(Q^*_{(j)}(L_{j\bullet}(z)))\|Q^*_{(j)}(L_{j\bullet}(z))\|_{op}^{D-1}.$$

Moreover,

$$\|Q^*_{(j)}(L_{j\bullet}(z))\|_{op} \leq \|L(z)\|_{op} \leq (|z|+1) + k^2\|M\|_{op}\max\|X_i\|_{op}^2$$

is bounded by some constant C with probability greater than $1-e^{-N}$ for C big enough by the large deviations bounds for $\|X_i\|_{op}$. Hence, for $C>0$ big enough

$$|\det(Q_{(j)}^* L_{j\bullet}(z))| \le (C+|z|)\sigma(Q_{(j)}^*(L_{j\bullet}(z))) \le C\sqrt{N}\varepsilon$$

with probability greater than $1-e^{-N}$. Hence we conclude that

$$\mathbb{P}(\sigma(L(z)) \le \varepsilon) \le \exp\{-N\} + \sum_{j=1}^{N} \mathbb{P}\left(|\det(Q_{(j)}^*(L_{j\bullet}(z)))| \le \varepsilon(|z|+C)N^{1/2}\right)$$

and the result follows by symmetry. □

Hereafter we bound from above the probability of small determinant:

$$p = \mathbb{P}\left(|\det(Q_{(1)}^*(L_{1\bullet}(z)))| \le \varepsilon\right).$$

Let's denote $Q_1, ..., Q_D$ the blocks of $Q_{(1)}$ of dimensions $N \times D$.

$$Q_{(1)} = \begin{pmatrix} Q_1 \\ \vdots \\ Q_D \end{pmatrix}$$

To this end we first restrict the set of Q_1 and show that Q_1 is well conditioned in the sense that

Lemma 2.3.7. *For all M finite, there exists $C_1 < \infty$ such that uniformly on $|z| \le M$,*

$$\mathbb{P}\left(\sigma(Q_1) \le N^{-C_1}\right) \le e^{-N}$$

Proof. It is sufficient to show that if $q = (q_1, \dots, q_D)$

$$\mathbb{P}\left(\exists q \in \mathbb{S}_{W_1^\perp} : \|q_1\|_2 \le N^{-C_1}\right) \le e^{-N} \tag{2.1}$$

where we recall that W_j is the span of the $(N-1)D$ columns of $L(z)$ outside $L_{j\bullet}$. To this end recall that q is so that q^* is a left null vector of the $DN \times D(N-1)$ matrix

$$\begin{pmatrix} t & r_1 & \cdots & \cdots & r_k \\ T^{(1,1)} - zI_{N-1} & R_1^{(1,1)} & \cdots & \cdots & R_k^{(1,1)} \\ s_1 & 0_{1\times N-1} & 0 & \cdots & 0 \\ S_1^{(1,1)} & -I_{N-1} & 0 & \cdots & 0 \\ \vdots & \ddots & \cdots & \cdots & 0 \\ s_k & 0 & \cdots & \cdots & 0 \\ S_k^{(1,1)} & 0 & \cdots & \cdots & -I_{N-1} \end{pmatrix}$$

where $M^{(1,1)}$ is the matrix M where we have taken out the first row and column and m denotes the first row of M without its first coefficient. This implies that for $\ell \ge 2$, if $q = (q_1, \dots, q_D)$,

$$r_\ell q_1^*(1) + (q_1^*)_{[2,N]} R_\ell^{(1,1)} = (q_\ell)_{[2,N]}.$$

We deduce that, on the event that R_ℓ is bounded by C in operator norm, $\|r_\ell\|_2$ is bounded by C' and $\{\|q_1\|_2 \le N^{-C_1}\}$

$$\|(q_\ell)_{[2,N]}\|_2 \le N^{-C_1}(C+C').$$

Moreover,

$$tq_1^*(1) + \sum s_i(q_i^*)(1) = -(q_1^*)_{[2,N]}(T^{(1,1)} - zI_{N-1}) - \sum (q_i^*)_{[2,N]} S_i^{(1,1)}.$$

When $T^{(1,1)}, S_i^{(1,1)}$ are also bounded in operator norm by C, and t by C', we deduce that

$$|\sum(s_i(q_i^*)(1))| \leq (C+|z|)N^{-C_1}(C+C'+1)D.$$

Observe that the $k\times(N-1)$ matrix with rows $(s_1,\dots,s_D)$ has a smallest singular values bounded below by $c>0$ with probability greater than $1-e^{-N}$ since its entries are independent Gaussian variables with variance bounded below by s/N for some $s>0$. Note that C and C' can be chosen uniformly bounded with probability greater than $1-e^{-N}$. Hence, we deduce

$$\max_i |q_i(1)| \leq (C+|z|)N^{-C_1}(C+C'+1)D$$

and so with probability greater than $1-e^{-N/2}$

$$\|q_\ell\|_2 \leq 2(C+|z|)N^{-C_1}(C+C'+1)D$$

for all ℓ, which is impossible for $C_1>1/2$.

□

The rest of the proof will be devoted to the bound of

$$p = \mathbb{P}\left(\{|\det(Q_{(1)}^*(L_{1\bullet}(z)))| \leq \varepsilon\} \cap \{\sigma(Q_1) \geq N^{-C_1}\}\right). \tag{2.2}$$

2.3.7 Bound on the probability of small determinant p

Recall

$$p = \mathbb{P}\left(\{|\det(Q_{(1)}^*(L_{1\bullet}(z)))| \leq \varepsilon\} \cap \{\sigma(Q_1) \geq N^{-C_1}\}\right).$$

We denote H the matrix

$$H = \begin{pmatrix} Q_1 & 0 & \dots & 0 & M_2 \\ 0 & Q_1 & \dots & 0 & M_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & Q_1 & M_{k+1} \\ 0 & 0 & \cdots & 0 & M_{k+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & M_{n+1} \end{pmatrix}$$

where $M_{i+1} = \sum_{j=1}^k s_{i,j}Q_{j+1} + t_iQ_1$ where $S_j = \sum_{i=1}^n s_{i,j}X_i$ and the t_i are such that $T = \sum_i t_iX_i$. (We recall here we assumed without loss of generality that $R_i = X_i$).

We can see that :

$$(\tilde{S}\otimes I_N)\begin{pmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_{k+1}\end{pmatrix} = \begin{pmatrix} Q_1 \\ M_2 \\ \vdots \\ M_{n+1}\end{pmatrix}$$

where

$$\tilde{S} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ t_1 & s_{1,1} & \cdots & s_{1,k} \\ \vdots & \vdots & \ddots & \vdots \\ t_n & s_{n,1} & \dots & s_{n,k} \end{pmatrix}.$$

In particular, since we have that the matrix $(s_{i,j})_{i,j}$ is of maximal rank, so is $\tilde{S}$ and so there exists a left inverse $\tilde{W}$ such that $\tilde{W}\tilde{S} = I_{k+1}$. and so we have

$$\begin{pmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_{k+1} \end{pmatrix} = (\tilde{W} \otimes I_N) \begin{pmatrix} Q_1 \\ M_2 \\ \vdots \\ M_{n+1} \end{pmatrix}. \tag{2.3}$$

This expression will be useful later as it shows that fixing the columns and rows of the Q_i is essentially the same thing as fixing the columns and rows of the M_i. Let us identify $\mathcal{M}_{k+1,k+1}(\mathbb{C})$ to $\mathbb{R}^{(k+1)^2}$ with the following application :

$$\begin{pmatrix} c_1 & c_2 & \cdots & c_k \end{pmatrix} \mapsto \begin{pmatrix} c_2 \\ \vdots \\ c_{k+1} \\ c_1 \end{pmatrix}.$$

With this definition, if Ξ is a Nn vector of iid standard Gaussian variables, $H^*\Xi$ is a vector whose entries have the same law as the entries of $\sqrt{N}Q^*_{(1)}(L_{1\bullet}(z))$ after recentering by their mean. Let's take $f : \mathbb{N}^* \to \mathbb{R}^+$ an arbitrary strictly increasing function such that for all $r > 1$, $(f(1)+...f(r-1))/f(r) < 1/2$ and $f(1) = 1$. Let's denote col_i the $k(k+1)+i$-th column of H, V the span of the $k(k+1)$ first columns of H and for $i_1, ...i_l \leq k+1$, $V(i_1, ..., i_l)$ the span of V and the columns $col_{i_1}, ..., col_{i_l}$. for $r = 0, ..., k+1$, and $\epsilon > 0$ we consider the following subsets of possible matrices Q,

$$\mathcal{A}_0(\epsilon) := \{Q | \text{ All the } col_i \text{ are at distance at most } \epsilon \text{ of } V\}$$

$$\mathcal{A}_1(\epsilon) := \{Q | \text{ One column } col_{i_1} \text{ is such that } dist(V, col_{i_1}) \geq \epsilon \text{ and the rest of the columns}$$
$$\text{are at distance at most } \epsilon^{f(2)} \text{ of } V(i_1)\}$$

...

$$\mathcal{A}_r(\epsilon) := \{Q | \text{There are } r \text{ distinct indices } i_1, ..., i_r \text{ such that } dist(V, col_{i_j}) \geq \epsilon^{f(j)} \text{ for every } j \leq r$$
$$\text{and } \forall i \in [1, k+1] \backslash \{i_1, ..., i_r\}, dist(col_i, V(i_1, ..., i_r)) \leq \epsilon^{f(r+1)}\}$$

...

$$\mathcal{A}_{k+1}(\epsilon) := \{Q | \text{There are } k+1 \text{ distinct indices } i_1, ..., i_{k+1}$$
$$\text{such that } dist(V, col_{i_j}) \geq \epsilon^{f(j)} \text{ for every } j \leq k+1\}.$$

We will now bound for $r = 0, 1, ..., k$ the probability of $\{Q \in B_r(\epsilon)\}$ where we define.

$$B_r(\epsilon) = \{Q \in \mathcal{A}_r(\epsilon), \sigma(Q_1) \geq N^{-C_1}\}.$$

We have

$$p \leq \sum_{r=0}^{k+1} \mathbb{P}\left(\{|\det(Q^*_{(1)}(L_{1\bullet}(z)))| \leq \varepsilon\} \cap \{Q \in B_r(\epsilon)\}\right).$$

Lemma 2.3.8. *There is some $\gamma > 0$ and $c > 0$ such that if we take $\varepsilon \leq N^{-\gamma}$, for all $r \leq k$,*

$$\mathbb{P}(\{Q \in B_r(\varepsilon)\}) \leq e^{-cN}.$$

Proof. We let $\mathcal{N}_r(\delta)$ be a δ-net of the set $B_r(\epsilon)$. Recall that $Q^*_{(1)}(L_{i\bullet}(z)) = 0$ for all $i \geq 2$. Hence

$$\mathbb{P}\left(\{Q \in B_r(\varepsilon)\}\right) = \mathbb{P}\left(\{Q \in B_r(\varepsilon)\} \cap_{2\leq i \leq N} \{\|Q^*_{(1)}(L_{i\bullet}(z))\|_2 = 0\}\right).$$

We next introduce the δ-net to fix Q and get uniformly on $|z| \leq M$:

$$\begin{aligned}
\mathbb{P}\left(\{Q \in B_r(\epsilon)\}\right) &\leq \sum_{Q' \in \mathcal{N}_r(\delta)} \mathbb{P}\left(\{\|Q - Q'\|_2 \leq \delta\} \cap_{2\leq i\leq N} \{\|Q^*_{(1)}(L_{i\bullet}(z))\|_2 \leq \delta C\}\right) + e^{-N} \\
&\leq \sum_{Q' \in \mathcal{N}_r(\delta)} \mathbb{P}\left(\cap_{2\leq i\leq N}\{\|{Q'_{(1)}}^*(L_{i\bullet}(z))\|_2 \leq \delta C\}\right) + e^{-N} \\
&= \sum_{Q' \in \mathcal{N}_r(\delta)} \prod_{i=2}^{N} \mathbb{P}\left(\{\|{Q'_{(1)}}^*(L_{i\bullet})(z)\|_2 \leq \delta C\}\right) + e^{-N}.
\end{aligned}$$

Where we used that the operator norms of the $(L_{i\bullet})(z)$ are bounded by C with probability $O(e^{-N})$.

We next use the anti-concentration estimate for any fixed matrix M and Z a vector of iid standard Gaussian:

$$\sup_{w\in\mathbb{C}^d} \mathbb{P}(\|MZ - w\|_2 \leq \varepsilon) \leq \frac{(C\epsilon)^{2d}}{\det(MM^*)} \tag{2.4}$$

Recall that $\|{Q'_{(1)}}^*(L_{2\bullet})(z)\|_2 = \|H^*\Xi + C\|_2$ where C is a constant vector. Let's consider $Q' \in \mathcal{N}_r(\delta)$, there exists $k(k+1)+r$ columns of H that are well conditionned. Let's denote Π the orthogonal projection on $span(e_j | j \in \{1, ..., k(k+1)\} \cup \{i_1, ..., i_r\})$ where e_j is the j-th vector of the canonical base. We have that $\|{Q'_{(1)}}^*(L_{2\bullet})(z)\|_2 = \|H^*\Xi + C\|_2 \leq \|\Pi H^*\Xi + \Pi C\|_2$

$$\begin{aligned}
\mathbb{P}\left(\{\|{Q'_{(1)}}^*(L_{2\bullet})(z)\|_2 \leq \delta C\}\right) &\leq \mathbb{P}\left(\{\|\Pi H^*\Xi + \Pi C\|_2 \leq \sqrt{N}\delta C\}\right) \\
&\leq \frac{(C\sqrt{N}\delta)^{2(k(k+1)+r)}}{(N^{-k(k+1)C_1}\varepsilon^{f(1)+...+f(r-1)})^2}
\end{aligned}$$

If we choose δ to be equal to $\epsilon^{f(r)}$, we have

$$\begin{aligned}
\mathbb{P}\left(\{\|{Q'_{(1)}}^*(L_{2\bullet})(z)\|_2 \leq \delta C\}\right) &\leq \frac{(C\sqrt{N}\delta)^{2(k(k+1)+r)}}{(N^{-2k(k+1)C_1}\delta^{(f(1)+...+f(r-1))/f(r)})^2} \\
&\leq \frac{(C\sqrt{N}\delta)^{2(k(k+1)+r)}}{(N^{-k(k+1)C_1})\delta}
\end{aligned}$$

Now we will evaluate the cardinality of the δ-net $N_r(\delta)$. For this, we will fix successively fix each columns of H and count at each step how many choices we must make to δ-cover the set of possible columns and then by multiplying those choices we will deduce the cardinality of $N_r(\delta)$. We first need to fix Q_1, as it will fix the first $k(k+1)$ columns of H. It takes at most $(1/\delta^2)$ by entry so a total of $(1/\delta^2)^{(k+1)N}$. We need then to choose among the remaining $k+1$ columns r that will be well conditioned and then fix the entries of those columns. Due to (2.3), fixing these entries amounts to fixing the entries for the corresponding columns of $Q_2, ..., Q_{k+1}$, for a cost at most $(K/\delta^2)^{rkN}$. The remaining colums are by construction at distance at most $\epsilon^{f(r)} = \delta$ of the span of the rest. So for each columns, we need only a cost $(1/\delta^2)^{(k(k+1)+r)}$ to fix the remaining columns. So we have the following bound on the cardinality of $N_r(\delta)$:

$$\|N_r(\delta)\| \leq N^{O(N)} \left(\frac{1}{\delta^2}\right)^{(k+1+rk)N+(k(k+1)+r)}$$

And so, we have :

$$
\begin{aligned}
\sum_{Q'\in\mathcal{N}_r(\delta)} \prod_{i=2}^{N-1} \mathbb{P}\left(\{\|{Q'_{(1)}}^*(L_{i\bullet})(z)\|_2 \leq \delta C\}\right) & \\
\leq N^{O(N)} \left(\frac{1}{\delta^2}\right)^{(k+1+rk)N+(k(k+1)+r)} & \frac{(C\sqrt{N}\delta)^{2(k(k+1)+r)(N-1)}}{(N^{-k(k+1)C_1})\delta} \\
& \leq KN^{O(N)}\delta^{O(1)+2((k-1)(k+1-r)-1/2)N}
\end{aligned}
$$

For $r < k+1$ we have that there exists some γ_r such that if $\delta \leq N^{-\gamma_r}$, the above bound is asymptotically lower than e^{-N}. So, if we choose $\epsilon := N^{-\gamma}$ with $\gamma \geq \gamma_r/f(r)$ we have the desired bounds.

□

And so it only remains to bound the probability :

$$\mathbb{P}\left(\{|\det(Q_{(1)})^*(L_{1\bullet}(z))| \leq \varepsilon\} \cap \{\sigma(Q_1) \geq N^{-C_1}\} \cap \{Q \in \mathcal{A}_{k+1}(N^{-\gamma})\}\right).$$

2.3.8 Gaussian polynomials calculus

To bound the former probability that determinants are small, we shall use that these determinants are polynomial in Gaussian variables, for which we know generic results. We first have :

Lemma 2.3.9 (Carbery-Wright inequality). *[27] Let $(g_1, \cdots, g_n)$ be iid standard gaussian variables and let f be a real valued degree d polynomial with variance one. Then there exists a finite constant $C(d)$ which only depends on d such that*

$$\sup_{t\in\mathbb{R}} \mathbb{P}\left(|f(g_1, \ldots, g_n) - t| \leq \delta\right) \leq C(d)\delta^{1/d}.$$

We therefore only need to bound from below the variance of the polynomials to conclude. To this end we remark, by writing a polynomial in the Hermite basis that :

Lemma 2.3.10. *Let $(g_1, \cdots, g_n)$ be iid standard gaussian variables and let f be a real valued degree d polynomial. Let a be the coefficient of one of the degree d monomials of f. Then there exists a positive constant $c(d)$ depending only on d such that*

$$\mathrm{Var}(f(g_1, \ldots, g_n)) \geq c(d)|a|^2.$$

For $i = 1, ..., N$ and $j = 1, ...k+1$ let us denote v_i^j the i-th row of Q_i and for $j = 2, ..., n+1$, w_i^j the i-th row of M_j.

Lemma 2.3.11. *Let $\beta > 0$. Let $\mathcal{D}(\beta)$ be the set of matrices Q such that there exists $i_1, ..., i_k \in [1, N]$ distinct and $j\in [1, k]$,such that :*

$$\Delta^j_{i_1,i_2,\ldots,i_k} := \det(v^1_{i_1}, ..., v^1_{i_j}, ..., v^1_{i_k}, w^{j+1}_{i_j}) \geq N^{-\beta}$$

or such that there exist $i_1, ..., i_k, i_{k+1} \in [1, N]$ distinct and $j \in [k+1, n]$ such that :

$$\Delta^j_{i_1,i_2,\ldots,i_k,i_{k+1}} := \det(v^1_{i_1}, ..., ..., ..., v^1_{i_k}, w^{j+1}_{i_{k+1}}) \geq N^{-\beta}$$

then for β large enough and any β'', there exists some $\beta' > 0$,

$$\sup_{Q\in\mathcal{D}(\beta)} \mathbb{P}[|\det(Q^*_{(1)}(L_{1\bullet}(z)))| \leq N^{-\beta'}] = o(N^{-\beta''}).$$

Proof. Recall that we can represent $Q^*_{(1)}(L_{1\bullet}(z))$ as a vector $H^*\Xi + C$.

We may write $\Xi^* = (g^1_1, ...g^1_N, g^2_1, ..., g^{n+1}_N)$. When we compute $\det(Q^*_{(1)}(L_{1\bullet}(z)))$, we have

$$\det(Q^*_{(1)}(L_{1\bullet}(z))) = N^{(k+1)/2}\det(\sum_l g^1_l v^1_l + c_1, ..., \sum_l g^k_l v^1_l + c_k, \sum_{j=1}^{n}\sum_l g^j_l w^{j+1}_l + c_k).$$

Where $c_1, ..., c_k$ are some constants depending on Q.

By expanding and looking at the term in $g^1_{i_1}...(g^j_{i_j})^2...g^k_{i_k}$, for distincts $i_1, ..., i_k \in [1, N]$ and $j \leq k$ we realize that it is equal to the determinant $\Delta^j_{i_1,i_2,...,i_k}$. Idem if we look at the term $g^1_{i_1}......g^k_{i_k} g^j_{i_{k+1}}$ for distincts $i_1, ..., i_{k+1} \in [1, N]$ and $j > k$ we realize that it is equal to the determinant $\Delta^j_{i_1,i_2,...,i_{k+1}}$. And so, if any of those are $\geq N^{-\beta}$, we have that $\mathrm{Var}(\det(Q^*_{(1)}(L_{1\bullet}(z)))) \geq KN^{-2\beta}$,

By using Carbery-Wright, we have that :

$$\mathbb{P}[|\det(Q^*_{(1)}(L_{1\bullet}(z)))| \leq \epsilon] \leq \frac{C(k+1)}{c(k+1)} N^{\beta+1/2}\epsilon^{1/(k+1)}.$$

By taking $\epsilon \leq N^{-(k+1)(\beta''+\beta+1/2)}$ we have our result.

□

Let us look at :

$$\mathbb{P}\left(\{|\det(Q^*_{(1)}(L_{1\bullet}(z)))| \leq \varepsilon\} \cap \{\sigma(Q_1) \geq N^{-C_1}\} \cap \{Q \in \mathcal{A}_{k+1}(N^{-\gamma})\} \cap \{Q \in \mathcal{D}(\gamma)\}\right).$$

Since Q and $L_{1\bullet}(z)$ are independant, this probability is bounded by

$$\sup_{Q\in\mathcal{D}(\gamma)} \mathbb{P}\left(\{|\det(Q^*_{(1)}(L_{1\bullet}(z)))| \leq \varepsilon\} \cap \{\sigma(Q_1) \geq N^{-C_1}\} \cap \{Q \in \mathcal{A}_{k+1}(N^{-\gamma})\}\right)$$

which is a $o(N^{-\beta'})$ if $\epsilon = N^{-\beta}$ for β large enough.

Now we will it remains to bound :

$$\mathbb{P}\left(\{|\det(Q^*_{(1)}(L_{1\bullet}(z)))| \leq \varepsilon\} \cap \{\sigma(Q_1) \geq N^{-C_1}\} \cap \{Q \in \mathcal{A}_{k+1}(N^{-\gamma})\} \cap \{Q \in \mathcal{D}(\beta)^c\}\right)$$

Lemma 2.3.12. *For β large enough :*

$$\mathbb{P}\left(\{|\det(Q^*_{(1)}(L_{1\bullet}(z)))| \leq \varepsilon\} \cap \{\sigma(Q_1) \geq N^{-C_1}\} \cap \{Q \in \mathcal{A}_{k+1}(N^{-\gamma})\} \cap \{Q \in \mathcal{D}(\beta)^c\}\right) \leq e^{-N}$$

Proof. By using the same method as before, we can use a δ-net on the set $\{Q : \sigma(Q_1) \geq N^{-C_1}\} \cap \mathcal{A}_{k+1}(N^{-\gamma}) \cap \mathcal{D}(\beta)^c$. Here the anticoncentration gives us a bound in :

$$N^{O(N)}\delta^{2(k+1)^2 N}$$

Let $\mathcal{M}_\beta(\delta)$ be a δ-net of $\{\sigma(Q_1) \geq N^{-C_1}\} \cap \mathcal{A}_{k+1}(N^{-\gamma}) \cap \mathcal{D}(\beta)^c$ and let's bound the cardinality of $\mathcal{M}_\beta(\delta)$. Since Q_1 is well conditioned, we can suppose that $v^1_1, .., v^1_k$ are such that $\sigma(Q_{1,[1,k]}) > N^{-2C_2}$ for some $C_2 > 0$ where $Q_{1,[1,k]}$ is the matrix Q_1 where we extract the k first rows.

Let i' be such that the $i'+1$-th row of $\tilde{S}$ is non-null, we then have $w^{i'+1}_j = \sum_{l\in[1,k]} s_{i,l} v^{l+1}_j + t_{i'+1} v^1_j$.

Then if $i' \leq k$, for every γ, if $|\Delta^{i'}_{1,2,..i'-1,j,i'+1...,k}| \leq N^{-\beta}$ for N large enough, then

1. either $dist(v^1_j, span(v^1_1, ..., \widehat{v^1_{i'}}, ..., v^1_k)) \leq N^{-(\beta-(k-1)C_2)/2}$
2. or $dist(w^{i'+1}_j, span(v^1_1, ..., v^1_k)) \leq N^{-(\beta-(k-1)C_2)/2}$.

And so, if we take $\delta \geq N^{-(\beta-(k-1)C_2)/2}$, once we have fixed v_j^i for $i \in [1, k+1], j \in [1, k]$, for each $j > k$, either (1) and then we have a cost of $O(1/\delta^2)^{(k+1)^2-2}$ to pay in order to fix all the v_j^i since v_j^1 lies in a space of dimension $k-1$ or (2) and then we have a cost of $O(1/\delta^2)^{((k+1)^2-1)}$ since the non-trivial linear combination $w_j^{i'+1}$ of the v_j^i lies in a space of dimension k.

If $k+1 \leq i' \leq n$, since the columns $v_1^1, ..., v_k^1$ are well conditionned. We have that $dist(w_j^{i'+1}, span(v_1^1, ..., v_1^k)) \leq N^{-(\beta-kC_2)}$, and N large enough. We are therefore back to (ii).

In total, the cardinality of the net is at most $N^{O(N)}O((1/\delta^2)^{N((k+1)^2-1)})$ and is beaten by the anti-concentration bound.

□

Putting both previous bounds together, we have proved the following proposition :

Proposition 2.3.13. *For any $\beta'' > 0$, we can find some $\beta' > 0$, such that :*

$$\mathbb{P}\left(\{|\det(Q_{(1)})^*(L_{1\bullet}(z))| \leq N^{-\beta'}\} \cap \{\sigma(Q_1) \geq N^{-C_1}\} \cap \{Q \in \mathcal{A}_{k+1}(N^{-\gamma})\}\right) = o(N^{-\beta''}).$$

Using this proposition in conjunction with Lemma 2.3.7 and Lemma 2.3.8, we have :

Proposition 2.3.14. *For any $\beta'' > 0$, we can find some $\beta' > 0$, such that :*

$$\mathbb{P}\left(\{|\det(Q_{(1)})^*(L_{1\bullet}(z))| \leq N^{-\beta'}\}\right) = o(N^{-\beta''}).$$

Finally, using 2.3.6 we prove our result.

2.3.9 Proof of the Corollary 2.3.3

We let for $z \in \mathbb{C}$ $\nu_N^z := \mu_{(zI_N-P_N)(zI_N-P_N)^*}$ where μ_A is the empirical measure of A. Here are our the steps of the proof:

1. For almost every $z \in \mathbb{C}$, we will prove that ν_N^z converges weakly in probability towards some identifiable probability measure ν^z.
2. Using our lower bound in Theorem 2.3.2, we will show that for alomost every $z \in \mathbb{C}$, in probability $\int \ln|x| d\nu_N^z(x)$ converges toward $\int \ln|x| d\nu^z(x)$.
3. We will prove that in probability, $z \mapsto \int \ln|x| d\nu_N^z(x)$ converges in L^1 toward $z \mapsto \int \ln|x| d\nu^z(x)$, and therefore in the sense of distribution μ_{P_N} converges toward μ_p.

Proof of (1). We can write $X_j = Y_1 + iZ_j$ where Y_j, Z_j are independent GUE matrices. So $(zI_n - P_n)(zI_n - P_n)^*$ is a self adjoint polynomial in GUE matrices. Following the convergence in law of independent GUE matrices toward free independent semi-circular elements, we have that ν_N^z converges in probability toward the law ν_z of $(z-P)(z-P)^*$ where $P = P(c_1, ..., c_n)$ where $c_1, ..., c_n$ are free circular elements. □

Proof of (2). The subtlety resides in the fact that log is not a bounded function. There is $M > 0$ such that $\lim_{N\to\infty} \mathbb{P}[||P_N|| \geq M] = 0$. Given that ν_N^z converges weakly in probability towards ν^z, it suffices to prove that for every $\delta, \delta' > 0$ there is $1 > \epsilon > 0$ such that :

$$\limsup_{N\to\infty} \mathbb{P}[|\int_0^\epsilon \log x d\nu_N^z(x)| \geq \delta] \leq \delta'$$

Denoting $\mathcal{G}_N = \{||P_N|| \leq M, \sigma(z - P_N) \geq N^{-\beta}\}$, it suffices to show that

$$\lim_{\epsilon\to 0} \limsup_{N\to\infty} \mathbb{E}[\mathbb{1}_{\mathcal{G}_N} \int_0^\epsilon |\log x| d\nu_N^z(x)] = 0.$$

Indeed by Markov's inequality :

$$\mathbb{P}[\mathbb{1}_{\mathcal{G}_n}\int_0^\epsilon |\log x| d\nu_n^z(x) > \delta] \leq \frac{E[\mathbb{1}_{\mathcal{G}_n}\int_0^\epsilon |\log x| d\nu_n^z(x)]}{\delta}.$$

If we choose $0 < \alpha < \beta$ we can write :

$$\int_0^\epsilon \log x d\nu_n^z(x) = \int_0^{N^{-\beta}} \log x d\nu_n^z(x) + \int_{N^{-\beta}}^{N^{-\alpha}} \log x d\nu_n^z(x) + \int_{N^{-\alpha}}^{\epsilon} \log x d\nu_n^z(x).$$

On the event $\mathcal{G}_N$ the first term is obviously 0 and the second term is bounded by $\beta N^{-\alpha}\log N$, so we only need to find $\alpha > 0$ such that :

$$\limsup_{n\to\infty} \mathbb{E}[|\int_{N^{-\alpha}}^{\epsilon} \log x d\nu_n^z(x)|] \leq \delta'.$$

Denoting $\bar{\nu}_n^z = \mathbb{E}[\nu_n^z]$, g the Stieljes transform of ν^z and g_n the Stieljes transform of $\bar{\nu}_n^z$, we will prove the following lemma :

Lemma 2.3.15. *There exist $c_1, C, N_0 > 0$ and $1 > c_2 > 0$ such that for $N^{-c_1} \leq \eta \leq 1$ and $N \geq N_0$:*

$$\Im(g_N(i\eta)) \leq C\eta^{-c_2}$$

We postpone the proof of this lemma to the end of this section. This lemma implies that $\bar{\nu}_n^z([-x,x]) \leq 2x\Im(g_n(ix)) \leq 2Cx^{1-c_2}$ for $x \geq N^{-c_1}$.

Denoting $F_n(x) = \bar{\nu}_n^z([0,x])$ we have for $\alpha < c_1$ et $\epsilon < 1$:

$$\begin{aligned}
\mathbb{E}[|\int_{N^{-\alpha}}^{\epsilon} \log x d\nu_N^z(x)|] &= -\int_{N^{-\alpha}}^{\epsilon} \log x d\bar{\nu}_N^z(x) \\
&= \int_{N^{-\alpha}}^{\epsilon} \frac{1}{x} F_n(x)dx - \log(\epsilon)F_N(\epsilon) + \log(N^{-\alpha})F_N(N^{-\alpha}) \\
&\leq 2\int_{N^{-\alpha}}^{\epsilon} \frac{1}{x} F_N(x)dx - 2\log(\epsilon)\epsilon^{1-c_2} + 2CN^{-\alpha(1-c_2)}\log(N^{-\alpha}) \\
&\leq 2C\int_{N^{-\alpha}}^{\epsilon} x^{-c_2}dx - 2\log(\epsilon)\epsilon^{1-c_2} + 2CN^{-\alpha(1-c_2)}\log(N^{-\alpha}) \\
&\leq 2C\frac{\epsilon^{1-c_2}}{1-c_2} - 2\log(\epsilon)\epsilon^{1-c_2} + 2CN^{-\alpha(1-c_2)}\log(N^{-\alpha})
\end{aligned}$$

And so $\lim_{\epsilon\to 0}\lim_{N\to\infty} \mathbb{E}[|\int_{N^{-\alpha}}^{\epsilon} \log x d\nu_N^z(x)|] = 0$.

So, denoting $h_N(z) := \int \log|x| d\nu_N^z(x)$ and $h(z) = \int \log|x| d\nu^z(x)$. We have that for almost every $z \in \mathbb{C}$, $h_N(z)$ converges in probability towards z. □

Proof of (3). We can write :

$$h_N(z) = 2\pi \int_{w\in\mathbb{C}} \log|z-w| d\mu_{P_N}(w)$$

and since ν^z is the spectral measure of $(z-p)(z-p)^*$, by definition of the Brown measure :

$$h(z) = 2\pi \int_{w\in\mathbb{C}} \log|z-w| d\mu_p(w)$$

where $p = P(c_1, ..., c_2)$ and μ_p is the Brown measure of μ_p. Let $K \subset \mathbb{C}$ a compact. Since the support μ_p is included in $\{z : |z| \leq ||p||\}$ and on the events $\mathcal{A}_N = \{||P_N|| < M\}$, μ_{P_N}'s support is in $\{z : |z| \leq M\}$ and noticing that the functions $z \mapsto \log|z-w|$ are uniformly bounded on $L^2(K)$ for w such that $|w| \leq \max\{M, ||p||\}$, we have that the family $(h_N \mathbb{1}_{\mathcal{A}_N})_{N\in\mathbb{N}}$ and h, viewed as functions of $\Omega \times K$ endowed with the measure $\mathbb{P}\otimes Leb$ (where $(\Omega, \mathcal{T}, \mathbb{P})$ is the ambient probability space) are uniformly bounded in $L^2(\Omega \times K, \mathbb{P} \otimes Leb)$. Therefore if we prove that $h_N \mathbb{1}_{\mathcal{A}_N}$ converges in measure toward h, we will have convergence in $L^1(\Omega \times K, \mathbb{P} \otimes Leb)$. But for every $\delta > 0$, we have by dominated convergence :

$$\lim_{N \to \infty} \int_{z \in K} \mathbb{P}[|\mathbb{1}_{\mathcal{A}_N} h_N(z) - h(z)| \geq \delta] = 0$$

where we used the convergence in probability of h_N toward h for almost every $z \in \mathbb{C}$. Therefore we have

$$\lim_{N \to \infty} \int_{z \in K} \mathbb{E}[\mathbb{1}_{\mathcal{A}_N} |h_N(z) - h(z)|] = 0$$

that is

$$\lim_{N \to \infty} \mathbb{E}[\mathbb{1}_{\mathcal{A}_N} \int_{z \in K} |h_N(z) - h(z)|] = 0$$

Therefore, in probability h_N converge toward h in $L^1(K, Leb)$ and so $\mu_{P_N} = \Delta h_N$ converge toward μ_p in the sense of distributions on K. Again, taking K that contains the support of μ_p and μ_{P_N}, we have the convergence in the sense of distribution implies the weak convergence and the result is proved. □

It remains to show the lemma 2.3.15

Proof of the lemma 2.3.15. We will use two results. The first one is [62, Theorem 1.1], mainly point (6) which has for consequence that the Stieljes transform G_{ν^z} of ν^z behave on $\{z | \Im z > 0\}$ like $C(r_l(z))^k / z^p$ for some $C \in \mathbb{C}$, $k, p \in \mathbb{N}$ near 0 where r_l is the standard determination of the l-th root and so $\Im g(i\epsilon) \leq K \epsilon^q$ for some $q \in \mathbb{Q}$ and $K > 0$. Then following [62][Corollary 1.2], since ν^z has no atoms, we have that $q > -1$. We will now use [4][Lemma 5.5.4]. We have that for $N^{-c_1} \leq \Im z \leq 1$ and N large enough :

$$|g(z) - g_N(z)| \leq \frac{c_2}{N^2 (\Im z)^{c_3}}$$

And so, for $1 \geq \Im z \geq N^{-\max\{2/c_3, c_1\}}$, and N large enough $|g(z) - g_N(z)| \leq 1$. Combining those results, we get lemma 2.3.15 □

Annexe A

Grandes déviations

Dans cette annexe, on fait quelques rappels sur la théorie générale des grandes déviations. Pour des preuves plus détaillées des résultats de cette section, on pourra se référer au livre [36].

A.0.1 Définitions et premiers résultats

Une fois que l'on a prouvé une loi des grands nombres pour une suite de variables aléatoires $(X_N)_{N\in\mathbb{N}}$, c'est-à-dire un résultat de convergence du type $\lim_N \mathbb{P}[X_N \in \mathcal{V}_x] = 1$ pour tout voisinage $\mathcal{V}_x$ d'un certain point x donné, il est naturel de vouloir aller plus loin. Une direction possible est un résultat central limite où l'on prouve la convergence en loi de $f(N)(X_N - x)$ où $\lim_N f(N) = +\infty$. Par exemple si les $X_N = S_N/N$ où S_N est la somme de N variables réelles i.i.d de variance finie σ^2, le résultat bien connu donne la convergence en loi de $\sqrt{N}(X_N - \mathbb{E}[X])$ vers une gaussienne $\mathcal{N}(0,\sigma^2)$. Mais plutôt que de "zoomer" près de la limite à la bonne échelle pour observer un comportement non trivial, une autre direction possible pour raffiner la loi des grands nombres est de quantifier la probabilité que X_N soit éloigné de sa limite. Un exemple simple est par exemple le cas où $X_N = S_N/N$ où S_N est une somme de gaussiennes centrées indépendantes. X_N est alors une gaussienne de loi $\mathcal{N}(0,\sigma^2/N)$ dont la densité est $x \mapsto N\exp(-Nx^2/(2\sigma^2))/\sqrt{2\pi}\sigma$ et pour x dans $\mathbb{R}$ on a heuristiquement :

$$\mathbb{P}[X_N \sim x] = \exp\left(-N\left(\frac{x^2}{2\sigma^2} + o(1)\right)\right)$$

On dira alors que la variable X_N vérifie un principe de grandes déviations de vitesse N et on appelle $x \mapsto \frac{x^2}{2\sigma^2}$ la fonction de taux.

A partir de cette heuristique, on donne une définition de ce qu'est un principe de grandes déviations en toute généralité :

Définition A.0.1. *On se place dans $(\mathcal{X},\mathcal{T})$ un espace topologique. Soit $I : \mathcal{X} \to [0,\infty]$. On dit que I est un fonction de taux si I est semi-continue inférieurement (i.e pour tout $\alpha \in [0,\infty]$, $I^{-1}([0,\alpha])$ est fermé). On dit que I est une bonne fonction de taux si pour tout $\alpha \in [0,\infty[$, $I^{-1}([0,\alpha])$ est compact.*

Soit $(\mu_N)_{N\in\mathbb{N}}$ une suite de mesures de probabilité sur $(\mathcal{X},\mathcal{B}(\mathcal{T}))$ où $\mathcal{B}(\mathcal{T})$ est la tribu des borélien. On dit que $(\mu_N)_{N\in\mathbb{N}}$ vérifie un principe de grandes déviations de vitesse $v(N)$ et de bonne fonction de taux I si :

$$\forall O \text{ ouvert de } \mathcal{X}, \text{ on a } -\inf_{x\in O} I(x) \leq \liminf_{N\to\infty} \frac{\log \mu_N(O)}{v(N)}$$

$$\forall F \text{ fermé de } \mathcal{X}, \text{ on a } -\inf_{x\in F} I(x) \geq \limsup_{N\to\infty} \frac{\log \mu_N(F)}{v(N)}$$

Dans la définition précédente, on appelle la première inégalité la *borne inférieure* de grandes déviations et la seconde inégalité la *borne supérieure* de grandes déviations. Commençons par énoncer un résultat d'unicité sur la fonction de taux. On rappelle qu'un espace topologique est dit Haussdorf si pour tout x, y distincts, il existe des voisinage de x et y disjoints et qu'un espace topologique est dit régulier si pour tout fermé F et tout point $x \notin F$ il existe deux ouverts O_1 et O_2 disjoints tels que $x \in O_1$ et $F \subset O_2$.

Proposition A.0.2. *Si μ_N vérifie un principe de grandes déviations avec vitesse $v(N)$ pour deux bonnes fonctions de taux I et J et que $(\mathcal{X}, \mathcal{T})$ est Haussdorf et régulier, alors $I = J$.*

Souvent, le degré de généralité de cette définition la rend peu pratique en particulier concernant la borne supérieure qui peut porter sur des fermés très "gros". On introduit donc une définition intermédiaire plus faible où l'on restreint la borne supérieure sur les compacts :

Définition A.0.3. *Soit $(\mu_N)_{N\in\mathbb{N}}$ une suite de mesures de probabilité sur $(\mathcal{X}, \mathcal{B}(\mathcal{T}))$. Supposons que tout les compacts de $(\mathcal{X}, \mathcal{T})$ soient dans $\mathcal{B}(\mathcal{T})$. On dit que $(\mu_N)_{N\in\mathbb{N}}$ vérifie un principe de grandes déviations faible de vitesse $v(N)$ et de fonction de taux I si la borne supérieure est vérifiée pour tout compact K tel que $\sup_{x\in K} I(x) < \infty$ et si la borne inférieure est vérifiée pour tout ouvert.*

Pour pouvoir traduire un principe de grandes déviations faible en un "vrai" principe de grandes déviations, on a donc besoin de contrôler à l'échelle exponentielle la mesure $\mu_N(K^c)$ pour de grands compacts K.

Définition A.0.4. *Soit $(\mu_N)_{N\in\mathbb{N}}$ une suite de mesures de probabilité sur $(\mathcal{X}, \mathcal{B}(\mathcal{T}))$. On dit que $(\mu_N)_{N\in\mathbb{N}}$ est exponentiellement tendue (pour la vitesse $v(N)$) si pour tout $M > 0$, il existe un compact K_M tel que :*

$$\limsup_N \frac{1}{v(N)} \log(\mu_N(K_M^c)) < -M$$

Pour les suites de mesures exponentiellement tendues qui vérifient un principe de grandes déviations faible on a alors le résultat suivant :

Proposition A.0.5. *Si $(\mu_N)_{N\in\mathbb{N}}$ est exponentiellement tendue pour la vitesse $v(N)$ et vérifie un principe de grandes déviations faible pour la fonction de taux I, alors I est une bonne fonction de taux et $(\mu_N)_{N\in\mathbb{N}}$ vérifie un principe de grandes déviations avec (bonne) fonction de taux I.*

En pratique on utilise souvent la caractérisation suivante, qui ne requiert de regarder le comportement de μ_N que sur des voisinages des points de $\mathcal{X}$:

Proposition A.0.6. *Soit $\mathcal{A}$ une base de la topologie $\mathcal{T}$, si on on a pour tout $x \in \mathcal{X}$ l'égalité :*

$$\sup_{A\in\mathcal{A}|x\in A} \left[-\limsup_N \frac{1}{v(N)} \log \mu_N(A) \right] = \sup_{A\in\mathcal{A}|x\in A} \left[-\liminf_N \frac{1}{v(N)} \log \mu_N(A) \right]$$

Alors (μ_N) vérifie un principe de grandes déviations faible avec pour fonction de taux I définie :

$$I(x) = \sup_{A\in\mathcal{A}|x\in A} \left[-\limsup_N \frac{1}{v(N)} \log \mu_N(A) \right]$$

Par exemple, pour vérifier ce critère dans un espace métrique, il suffit de vérifier que pour tout $x \in \mathcal{X}$:

$$\lim_{\delta\to 0} \limsup_N \frac{1}{v(N)} \log \mu_N(B(x,\delta)) = \lim_{\delta\to 0} \liminf_N \frac{1}{v(N)} \log \mu_N(B(x,\delta)) = -I(x)$$

Si l'on dispose d'un principe de grandes déviations pour une certaine suite de mesures de probabilité $\mathcal{X}$ et d'une fonction continue de $\mathcal{X}$ dans $\mathcal{Y}$, alors on peut énoncer un principe de grandes déviations pour les mesures images.

Théorème A.0.7. *Soit* $(\mathcal{X},\mathcal{T})$ *et* $(\mathcal{Y},\mathcal{T}')$ *deux espaces topologiques Hausdorff et* $f : \mathcal{X} \to \mathcal{Y}$ *une fonction continue. Soit* $I : \mathcal{X} \to [0,\infty]$ *une bonne fonction de taux. Alors la fonction* $I' : \mathcal{Y} \to [0,\infty]$ *définie par :*

$$I'(y) = \inf_{x\in\mathcal{X}|f(x)=y} I(x)$$

est une bonne fonction de taux. De plus si I *est la bonne fonction de taux associée à un principe de grandes déviations d'une suite de mesures de probabilité* $(\mu_N)_N$*, alors la suite des mesures images* $(f^*\mu_N)_N$ *vérifie un principe de grandes déviations de bonne fonction de taux* I'*.*

A.0.2 Lemme de Varadhan et lemme inverse de Bryc

Une des principales applications du principes de grandes déviations est l'étude du comportement asymptotique des intégrales de Laplace du type $\int e^{v(N)\varphi(x)}d\mu_N(x)$ où φ est continue.

Proposition A.0.8 (Lemme de Varadhan). *Soit* $(\mu_N)_N$ *satisfaisant un principe de grandes déviations avec bonne fonctions de taux* I *et* φ *une fonction continue de* $\mathcal{X}$ *satisfaisant la condition suivante :*

$$\lim_{M\to\infty}\limsup_N \frac{1}{v(N)}\log\mathbb{E}[e^{v(N)\varphi(X_N)}\mathbb{1}_{\varphi(X_N)\geq M}] = -\infty$$

et la condition de moments suivante pour un $\gamma > 1$ *:*

$$\limsup_N \frac{1}{v(N)}\log\mathbb{E}[e^{\gamma v(N)\varphi(X_N)}] < \infty$$

Alors

$$\lim_{N\to\infty}\frac{1}{v(N)}\log\mathbb{E}[e^{v(N)\varphi(X_N)}] = \sup_{x\in\mathcal{X}}\{\varphi(x) - I(x)\}$$

(où la loi de X_N *est* μ_N*).*

On peut voir ce théorème comme une extension de la méthode de Laplace pour les intégrales réelles. Une version inverse du lemme de Varadhan permet de démontrer que, si l'on connaît le comportement des intégrales de Laplace pour suffisamment de fonctions φ, on peut obtenir un principe de grandes déviations pour μ_N :

Proposition A.0.9 (Lemme inverse de Bryc). *Soit* $(\mu_N)_{N\in\mathbb{N}}$ *une suite de mesures de probabilité sur* $(\mathcal{X},\mathcal{B}(\mathcal{T}))$ *exponentiellement tendue et telle que pour tout* $f \in \mathcal{C}_b(\mathcal{X})$ *la limite suivante existe :*

$$\Lambda_f = \lim_{N\to\infty}\frac{1}{v(N)}\log\int e^{v(N)f(x)}d\mu_N(x)$$

alors $(\mu_N)_{N\in\mathbb{N}}$ *satisfait un principe de grandes déviations avec pour bonne fonction de taux :*

$$I(x) = \sup_{f\in\mathcal{C}_b(\mathcal{X})}\{f(x) - \Lambda_f\}$$

A.0.3 Théorème de Cramer et théorème de Sanov

On va à présent citer deux résultats classiques de la théorie des grandes déviations : le théorème de Cramer qui porte sur la somme renormalisée de variables aléatoires i.i.d. dans un espace vectoriel topologique et le théorème de Sanov qui porte sur la mesure empirique d'une suite i.i.d. de variables aléatoires.

Tout d'abord on se fixe μ une mesure de probabilité borélienne sur $(\mathcal{X},\mathcal{T})$ un espace vectoriel topologique réel Hausdorff et localement convexe et on définit pour $\lambda \in \mathcal{X}^*$ une forme linéaire continue sur $\mathcal{X}$:

$$\Lambda(\lambda) = \log\int_{\mathcal{X}} e^{\langle\lambda,x\rangle}d\mu$$

et Λ^*, la transformée de Fenchel-legendre de Λ c'est-à-dire la fonction définie sur $\mathcal{X}$ par :

$$\Lambda^*(x) = \sup_{\lambda \in \mathcal{X}^*} (\langle \lambda, x \rangle - \Lambda(\lambda)) \in [0, \infty]$$

On peut alors énoncer le théorème de Cramer :

Théorème A.0.10 (Théorème de Cramer)**.** *Soit $(X_i)_{i\in\mathbb{N}}$ une suite de variables aléatoires i.i.d. de loi μ à valeurs dans un espace vectoriel topologique réel $(\mathcal{X}, \mathcal{T})$ Hausdorff localement convexe. On pose μ_N la loi de $(X_1 + ... + X_N)/N$. On suppose :*

1. *Il existe $\mathcal{E} \subset \mathcal{X}$ fermé et convexe tel que $\mu(\mathcal{E}) = 1$ et tel que $\mathcal{E}$ muni de la topologie induite peut être muni d'une distance qui en fait un espace polonais.*
2. *L'enveloppe convexe fermé de tout compact $K \subset \mathcal{E}$ est compact.*

Alors $(\mu_N)_{N\in\mathbb{N}}$ vérifie un principe de grandes déviations faible de vitesse N, avec fonction de taux Λ^* et de plus pour tout ouvert convexe $O \subset \mathcal{X}$,

$$\lim \frac{1}{N} \log \mu_N(O) = - \inf_{x\in O} \Lambda^*(x)$$

Un corollaire du théorème précédent est le cas de variables aléatoires à valeurs dans $\mathbb{R}^n$

Corollaire A.0.11. *Si $\mathcal{X} = \mathbb{R}^d$, $(\mu_N)_{N\in\mathbb{N}}$ vérifie un principe de grandes déviations faible avec fonction de taux Λ^*. Si il existe un voisinage $\mathcal{V}$ de 0 tel que* $\sup_{\lambda\in\mathbb{R}^n} \Lambda(\lambda) < \infty$, *$(\mu_N)_{N\in\mathbb{N}}$ vérifie un principe de grandes déviations avec bonne fonction de taux Λ^*.*

De plus dans ce cas les formules donnant Λ et Λ^* se trouve simplifiées par le fait qu'on peut identifier $\mathcal{X}^*$ à $\mathbb{R}^d$ et $\langle ., . \rangle$ au produit scalaire usuel.

On peut voir ce résultat comme une simple spécialisation du théorème de Cramer général. Sous des hypothèses plus fortes, on peut néanmoins le démontrer de manière beaucoup plus élémentaire. On donne ici une démonstration succincte dans le cas où $\mathcal{D}_\Lambda = \{\theta \in \mathbb{R}^n | \Lambda(\theta) < \infty\} = \mathbb{R}^n$ et $\mathrm{supp}(\mu) = \mathbb{R}^n$ (le support de μ est $\mathbb{R}^n$ tout entier). La démonstration fait entre autre apparaître de manière plus claire l'origine de l'expression de Λ^*.

Démonstration. On se donne $(X_i)_{i\in\mathbb{N}}$ une suite de variables aléatoires i.i.d. de loi μ vérifiant les hypothèses précitées. On pose $S_n = X_1 + ... + X_n$

Tension exponentielle : Notons pour $x \in \mathbb{R}^d$, $|d|_1 := \sum_i |x_i|$. On a

$$\exp(|x|_1) \leq \sum_i (\exp(\langle x, e_i \rangle) + \exp(-\langle x, e_i \rangle))$$

où les e_i sont les vecteurs de la base canonique. On a donc :

$$\mathbb{E}[\exp(|S_n|)] \leq \sum_i (\mathbb{E}[\exp(\langle S_n, e_i \rangle)] + \mathbb{E}[\exp(-\langle S_n, e_i \rangle)])$$

En utilisant le caractère i.i.d. des X_i, on a $\mathbb{E}[\exp(\langle S_n, \theta \rangle)] = \mathbb{E}[\exp(\langle X_1, \theta \rangle)]^n$ d'où :

$$\mathbb{E}[\exp(|S_n|)] \leq \sum_i (\mathbb{E}[exp(\langle X_1, e_i \rangle)]^n + \mathbb{E}[\exp(-\langle X_1, e_i \rangle)]^n)$$

En passant au log et a la limite on a :

$$\limsup_n \frac{1}{n} \log \mathbb{E}[\exp(|S_n|)] \leq \max_i \{\Lambda(e_i), \Lambda(-e_i)\} < +\infty$$

On pose $C := \max_i \{\Lambda(e_i), \Lambda(-e_i)\}$

$$\mathbb{P}[|S_n/n| \geq M] \exp(nM) \leq \mathbb{E}[\exp(|S_n|)]$$

Et en passant au log,

$$\limsup_n \frac{1}{n} \log \mathbb{P}[|S_n/n| \geq M] \leq C - M$$

On vérifie alors la tension exponentielle en prenant M assez grand

Borne supérieure :

Soit K un compact de $\mathbb{R}^n$. Pour tout $\theta \in \mathbb{R}^n$, on a par l'inégalité de Markov

$$\mathbb{P}[|S_n/n - x| \leq \delta] \leq e^{-n \min_{|y-x|\leq\delta} \langle \theta, y\rangle} \mathbb{E}[\exp(\langle \theta, S_n\rangle)]$$

Prenons θ tel que $\langle \theta, x\rangle - \Lambda(\theta) \geq \Lambda^*(x) - \epsilon$. On peut alors choisir $\delta > 0$ tel que $\min_{|y-x|\leq\delta}\langle \theta, y\rangle \geq \langle \theta, x\rangle - \epsilon$, on a alors du fait que $\mathbb{E}[\exp(\langle \theta, S_n\rangle)] = \mathbb{E}[\exp(\langle \theta, X_1\rangle)]^n$:

$$\frac{1}{n} \log \mathbb{P}[|S_n/n - x| \leq \delta] \leq -\langle \theta, x\rangle + \Lambda(\theta) + \epsilon \leq -I(x) + 2\epsilon$$

On a bien

$$\lim_{\delta\to 0} \limsup_n \frac{1}{n} \log \mathbb{P}[|S_n/n - x| \leq \delta] \leq -I(x)$$

Borne inférieure :

Commençons par remarquer que l'hypothèse que $\mathrm{supp}(\mu) = \mathbb{R}^n$ implique que pour tout $\theta \in \mathbb{R}^n$, il existe $A \in \mathbb{R}$ tel que $\Lambda(x) \geq \langle \theta, x\rangle + A$ pour tout $x \in \mathbb{R}^n$. On en déduit que pour tout $x \in \mathbb{R}^n$, $\theta \mapsto \Lambda(\theta) - \langle \theta, x\rangle$ tend vers $+\infty$ quand $|\theta| \to \infty$. De cette façon comme $\theta \mapsto \Lambda(\theta) - \langle \theta, x\rangle$ est différentiable, il existe un point critique θ_0 de cette fonction qui vérifie donc $\vec{\nabla}\Lambda(\theta_0) = x$. Or, on a en calculant le gradient sous l'intégrale que

$$\vec{\nabla}\Lambda(\theta_0) = \frac{\mathbb{E}[X_1 \exp(\langle \theta_0, X\rangle)]}{\mathbb{E}[\exp(\langle \theta_0, X\rangle)]}$$

On pose alors $\mathbb{P}^{\theta_0}$ la mesure définie par rapport à $\mathbb{P}$ de la manière suivante :

$$\frac{d\mathbb{P}^{\theta_0}}{d\mathbb{P}} = \frac{\exp(\langle \theta_0, S_n\rangle)}{\mathbb{E}[\exp(\langle \theta_0, S_n\rangle)]} = \prod_{i=1}^{n} \frac{\exp(\langle \theta_0, X_i\rangle)}{\mathbb{E}[\exp(\langle \theta_0, X_i\rangle)]}$$

Comme la densité de cette nouvelle est un produit de la même fonction en les X_i, ceux-ci restent i.i.d. sous la nouvelle mesure et leur moyenne est donc $\nabla\Lambda(\theta_0) = x$. On a alors par la loi des grands nombres pour tout $\delta > 0$ $\lim_n \mathbb{P}^{\theta_0}[|S_n/n - x| \leq \delta] = 1$. On a de plus

$$\begin{aligned}\mathbb{P}^{\theta_0}[|S_n/n - x| \leq \delta] &= \frac{\mathbb{E}[\mathbb{1}_{|S_n/n - x|\leq\delta} \exp(\langle \theta_0, S_n\rangle)]}{\mathbb{E}[\exp(\langle \theta_0, S_n\rangle)]} \\ &\leq \frac{\mathbb{P}[|S_n/n - x| \leq \delta] \exp(n(\langle \theta_0, x\rangle + o_\delta(1)))]}{\mathbb{E}[\exp(\langle \theta_0, X_1\rangle)]^n}\end{aligned}$$

(où $o_\delta(1)$ désigne une quantité qui tend vers 0 quand δ tend vers 0).

En passant au logarithme :

$$\frac{1}{n} \log \mathbb{P}[|S_n/n - x| \leq \delta] \geq \frac{1}{n} \log \mathbb{P}^{\theta_0}[|S_n/n - x| \leq \delta] + \Lambda(\theta_0) - (\langle \theta_0, x\rangle + o_\delta(1))$$

et donc

$$\lim_{\delta\to 0} \liminf_n \frac{1}{n} \log \mathbb{P}[|S_n/n - x| \leq \delta] \geq \Lambda(\theta_0) - \langle \theta_0, x\rangle \geq -I(x)$$

et la borne inférieure est vérifiée $\square$

L'esprit de cette preuve est très similaire a la démonstration des résultats des sections 1.3 et 1.4, excepté qu'au lieu des transformées de Laplace $\mathbb{E}[\exp(\langle\theta, S_n\rangle)]$ on utilise les intégrales d'Itzykson-Zuber $\mathbb{E}_e[\exp(N\theta\langle e, X_N e\rangle)]$.

Enfin, un autre résultat classique de grandes déviations est le théorème de Sanov concernant la mesure empirique donnée par une suite de variables i.i.d.. On se fixe ici Σ un espace Polonais et μ une mesure de probabilité sur Σ.

Théorème A.0.12 (Theorème de Sanov)**.** *Soit $(X_i)_{i\in\mathbb{N}}$ une suite i.i.d. de loi μ. Pour $N \in \mathbb{N}$, on définit $Y_N = (\delta_{X_1} + ... + \delta_{X_N})/N$. Soit μ_N la loi de Y_N. Alors μ_N vérifie un principe de grandes déviations sur $\mathcal{P}(\Sigma)$ l'ensemble des mesures de probabilité boréliennes muni de la topologie faible de vitesse N et de bonne fonction de taux $H(.|\mu)$ donnée par :*

$$H(\lambda|\mu) := \begin{cases} \int f \log f d\mu \text{ si } \lambda \text{ est absolument continue par rapport à } \mu \text{ et } f = \frac{d\lambda}{d\mu} \\ +\infty \text{ sinon} \end{cases}$$

La quantité $H(\lambda|\mu)$ est aussi appelée entropie relative de λ par rapport à ν ou déviation de Kullback-Leibler. On peut voir le théorème de Sanov comme une simple application du théorème de Cramer sur l'espace des mesures (signés) de masse finie muni de la topologie faible et $\mathcal{E}$ est l'ensemble des mesures de probabilité.

Annexe B

Probabilités libres et matrices aléatoires

B.0.1 Espaces de probabilités non-commutatifs

On peut ramener l'étude de la convergence de la mesure empirique d'une matrice de Wigner à l'étude de la convergence des moments de cette mesure. Ce problème est donc équivalent à la question : quelle est la limite des moments $N^{-1}\mathrm{Tr}(X_N^k)$? Dans le cas des matrices de Wigner, ils convergent vers les nombres de Catalan qui sont les moments de la mesure semi-circulaire. Mais cette question peut être généralisée au cas d'un n-uplet de matrices de Wigner indépendantes : si $(X_N^1, ..., X_N^n)$ est un tel n-uplet, vers quelle quantité converge $N^{-1}\mathrm{Tr}(X_N^{i_1}X_N^{i_2}...X_N^{i_k})$ où $i_1, ..., i_k \in [1, n]$? Pour capturer le comportement asymptotique de ce types de quantités et en particulier le caractère non commutatif des matrices en question, on va faire appel aux outils de la théorie des probabilités libres. Les liens entre matrices aléatoires et probabilités non-commutatives ont été mis au jour par Voiculescu dans [69]. Les résultats de cette sous-section peuvent être trouvés notamment dans [4][section 5].

Théorème B.0.1. *Soit $(X_N^1, ..., X_N^n)$ un n-uplet de matrices de Wigner indépendantes et $i_1, ..., i_k \in [1, n]$. $N^{-1}\mathrm{Tr}(X_N^{i_1}...X_N^{i_k})$ converge vers $\sigma^{(n)}(X_{i_1}...X_{i_k})$ où $\sigma^{(n)}$ est une forme linéaire sur l'espace des polynômes non-commutatifs.*

Tout comme les moments de la loi semi-circulaire, les $\sigma^{(n)}(X_{i_1}...X_{i_k})$ peuvent se comprendre de façon combinatoire de la manière suivante : si l'on donne à chaque entier j de $[1, k]$ la couleur i_j, on peut regarder $NC_2(i_1, ..., i_k)$ l'ensemble des partitions paires sans croisement P qui respectent la coloration de $[1, k]$, c'est à dire telles que si $\{a, b\} \in P$ alors $i_a = i_b$. On a alors $\sigma^{(n)}(X_{i_1}...X_{i_k}) = \#NC_2(i_1, ..., i_k)$. Cela est à comparer au cas des moments joints d'une famille de gaussiennes indépendantes où l'on a la même formule excepté que les partitions paires peuvent avoir des croisements.

Dans la suite on note $\mathbb{C}\langle X_1, ..., X_n\rangle$ l'algèbre des polynômes non-commutatifs dotée de l'involution $*$ qui est définie sur les monômes par $(zX_{i_1}...X_{i_l})^* = \bar{z}X_{i_l}...X_{i_1}$. La forme linéaire $\sigma^{(n)}$ vérifie de plus que $\sigma^{(n)}(1) = 1$ et $\sigma^{(n)}(PP^*) \geq 0$. Elle a donc des propriétés similaires à l'espérance vue comme forme linéaire sur un certain espace des variables aléatoires. Dans cet esprit on va introduire la notion d'espace de probabilité non-commutatif :

Définition B.0.2. *Un espace de probabilité non commutatif est un couple $(\mathcal{A}, \varphi)$ où $\mathcal{A}$ est une $\mathbb{C}$-algèbre unitaire et φ une forme linéaire sur $\mathcal{A}$ telle que $\varphi(1) = 1$. Ses éléments sont appelés des variables aléatoires non-commutatives.*

On va à présent définir la loi d'un k-uplet de variables aléatoires non-commutatives :

Définition B.0.3. *Soit $a_1, ..., a_k \in \mathcal{A}$ des variables aléatoires auto-adjointes non-commutatives d'un espace de probabilité non commutatif. La loi de $(a_1, ..., a_k)$ est la forme linéaire $\mu_{a_1,...,a_k}$ sur $\mathbb{C}\langle X_1, ..., X_k\rangle$ définie par :*

$$\mu_{a_1,...,a_k}(P) = \varphi(P(a_1, ..., a_k))$$

On dit que la suite $a^n = (a_1^n, ..., a_k^n)$ de k-uplet de variables aléatoires auto-adjointes non-commutatives converge en loi vers un k-uplet $a = (a_1, ..., a_k)$ si $\mu_{a^n}(P)$ cnverge vers $\mu_a(P)$ pour tout P.

Là encore dans le cas des probabilités classiques, cela revient à caractériser la loi d'une variable bornée par ses moments. On introduit ensuite la notion de C^*-espace de probabilité non-commutatif (en rappelant d'abord la notion de C^*-algèbre) :

Définition B.0.4. *Une C^*-algèbre $(\mathcal{A}, ||.||, *)$ est une triplet où $(\mathcal{A}, ||.||)$ est une algèbre de Banach et $*$ est une application semi-linéaire qui vérifie*

1. $(ab)^* = b^*a^*$
2. $(a^*)^* = a$
3. $||a^*|| = ||a||$
4. $||aa^*|| = ||a||^2$

Une forme linéaire φ sur $(\mathcal{A}, ||.||, *)$ est un état si $\varphi(e) = 1$ (où e est l'élément neutre de l'algèbre) et $\varphi(aa^*) \geq 0$. Un C^*-espace de probabilité $(\mathcal{A}, ||.||, *, \varphi)$ est une C^*-algèbre munie d'un état φ. Si φ vérifie de plus $\varphi(ab) = \varphi(ba)$ pour tout $a, b \in \mathcal{A}$, on dit que φ est une trace.

On dit de plus que a est auto-adjointe si $a = a^*$ et a est normale si $aa^* = a^*a$. Par comparaison avec le cas des probabilités classiques cette approche revient à aborder la théorie des probabilités en étudiant l'algèbre des variables aléatoires complexes bornées sur un certain espace de probabilité et à munir cette algèbre de l'espérance $\mathbb{E}$. La structure supplémentaire apportée par les C^*-algèbres permet d'utiliser entre autres les outils issus du calcul fonctionnel La remarque suivante permet de faire le lien entre loi non-commutative et la loi d'une variable aléatoire dans le cas classique.

Remarque B.0.5. *Si a est une variable aléatoire non-commutative auto-adjointe dans un C^*-espace de probabilité, le calcul fonctionnel donne un isomorphisme de $\mathcal{C}(sp(a))$ dans $\mathcal{A}$ qui à une fonction f associe $f(a) \in \mathcal{A}$. On a en particulier une forme linéaire λ_a de $\mathcal{C}(\mathbb{R})$ dans $\mathbb{R}$ donné par $\lambda_a(f) = \varphi(f(a))$. λ_a est positive et $\lambda_a(1) = 1$. Ainsi par le théorème de représentation de Riesz, il existe une unique mesure de probabilité μ sur $\mathbb{R}$ telle que $\lambda_a(f) = \int f d\mu$. Ainsi, pour tout polynôme P, $\varphi(P(a)) = \mu_a(P) = \int P(x)d\mu(x)$.*

Ainsi la loi d'une variable aléatoire non-commutative auto-adjointe est bien une mesure sur $\mathbb{R}$. Voici quelques exemples de C^*-algèbres :

1. **Probabilité classique** : Si $(\Omega, \mathcal{T}, \mu)$ est un espace de probabilité et $\mathcal{A} = L^\infty(\Omega, \mathcal{T}, \mu)$ l'espace des variables aléatoires complexes bornées alors $(\mathcal{A}, ||.||_\infty, z \mapsto \bar{z}, \mathbb{E}[.])$ est un C^*-espace de probabilité . En particulier dans ce cas, les éléments de $\mathcal{A}$ commutent et pour une variable aléatoire réelle X, la loi de X définie dans la cas classique est bien la même que celle définie dans le cas non-commutatif.
2. **Matrices** : Pour tout $n \in \mathbb{N}$ $(\mathcal{M}_{n,n}(\mathbb{C}), ||.||, *, n^{-1}\mathrm{Tr}(.))$ (où $||.||$ est la norme d'opérateur sur l'espace euclidien classique) est un C^*-espace de probabilité. La loi de M hermitienne est la somme renormalisée des Dirac en ses valeurs propres.
3. **Matrices aléatoires** : Si $(\Omega, \mathcal{T}, \mu)$ est un espace de probabilité, $n \in \mathbb{N}$ et $\mathcal{A} = \mathcal{M}_{n,n}(L^\infty(\Omega, \mathcal{T}, \mu))$, alors $(\mathcal{A}, ||.||, *, n^{-1}\mathbb{E}[\mathrm{Tr}(.)])$ est (où $||.||$ est la norme d'opérateur subordonnée à la norme hermitienne $||.||_2$ sur $(L^2(\Omega, \mathcal{T}, \mu))^n$ donnée par $||X||_2^2 = \mathbb{E}[||X||_{eucl}^2]$) est C^*-espace de probabilité. La loi d'une matrice aléatoire hermitienne M est la moyenne des mesures empiriques de M i.e $\mathbb{E}[\mu_M]$.
4. **Algèbre de Von Neumann** : Soit H est un espace de Hilbert et $B(H)$ l'algèbre de Banach des opérateurs bornés de H pour la norme subordonnée à la norme hilbertienne. On dit qu'une $*$-algèbre $\mathcal{A} \subset B(H)$ qui contient l'identité de H est une algèbre de Von Neumann si elle est fermée pour la topologie faible (c'est-à-dire la plus petite topologie qui rend continue la famille d'applications $\xi \mapsto \langle x, \xi y \rangle$ pour $x, y \in H$). Si φ est un état sur $\mathcal{A}$ qui s'écrit $\varphi(\xi) = \langle u, \xi u \rangle$ pour un certain vecteur $u \in H$ alors on dit que $(\mathcal{A}, ||.||, \varphi)$ est une W^*-espace de probabilité. En particulier, il s'agit d'une C^*-espace de probabilité. On reviendra plus tard plus en détails sur cet exemple.

On a de plus la condition suffisante pour qu'une certaine forme linéaire sur $\mathbb{C}\langle X_1,...,X_n\rangle$ soit la loi d'un n-uplet de variables aléatoires non-commutatives auto-adjointes dans un C^*-espace de probabilité.

Théorème B.0.6. *Soit α une forme linéaire sur $\mathbb{C}\langle X_1,...,X_n\rangle$ telle que $\alpha(1)=1$, $\alpha(PP^*)\geq 0$ pour tout P et telle qu'il existe $R\in\mathbb{R}^*$ tel que pour tout $l\in\mathbb{N}$ et $i_1,...,i_l\in[1,n]$*

$$|\alpha(X_{i_1}...X_{i_l})|\leq R^l$$

Alors il existe un C^*-espace de probabilité tel que α est la loi d'un n-uplet d'éléments $a_1,...,a_n$ auto-adjoints

Remarque B.0.7. *Si la notion de C^*-algèbre donne accès aux outils du calcul fonctionnel et permet de définir $f(a)$ pour f une fonction réelle continue bornée et a un élément auto-adjoint, la notion de W^*-espace de probabilité nous permet d'aller plus loin. Si T est un opérateur auto-adjoint de $B(H)$, le théorème spectral nous donne une résolution de l'identité $\chi:\mathcal{B}(\mathbb{R})\to B(H)$ telle que :*

$$T=\int_{\mathbb{R}}\lambda d\chi(\lambda)$$

On peut alors définir $f(T)$ comme étant $\int_{\mathbb{R}}f(\lambda)d\chi(\lambda)$ pour toute f mesurable bornée et la mesure spectrale de T est $\varphi\circ\chi=\langle u,\chi(.)u\rangle$. La notion de W^*-espace de probabilité permet donc de faire du calcul borélien sur les variables aléatoires non-commutatives. En particulier, il est possible d'utiliser la notion d'opérateur affilié et la théorie des opérateurs non-bornés dans ce contexte pour étudier des variables aléatoires non-commutatives non-bornées ainsi que leurs lois. On peut définir des analogues des espaces L^p dans le cas non-commutatif.

On va à présent énoncer un résultat de représentation de C^*-espaces de probabilité. Cette construction montre que l'étude d'un C^*-espace de probabilité peut en fait se ramener à celle d'un W^*-espace de probabilité :

Théorème B.0.8 (Construction de Gelfand-Naimark-Segal)**.** *Soit α un état sur une C^*-algèbre unitaire $(\mathcal{A},||.||,*)$ engendrée par une famille dénombrable $\{a_i\}_{i\in I}$ d'éléments auto-adjoints de $\mathcal{A}$. Alors, il existe un Hilbert séparable $(H,\langle .,.\rangle)$, $\pi:\mathcal{A}\to B(H)$ un $*$-morphisme faisant décroître la norme et $\xi_1\in H$ tels que :*

1. *$\{\pi(a)\xi_1 : a\in\mathcal{A}\}$ est dense dans H.*
2. *En posant $\varphi_\alpha(x)=\langle\xi_1,x\xi_1\rangle$ pour $x\in B(H)$. Alors pour tout a dans $\mathcal{A}$:*

$$\alpha(a)=\varphi_\alpha(\pi(a))$$

3. *La loi non-commutative des $\{a_i\}_{i\in I}$ dans le C^*-espace de probabilité $(\mathcal{A},||.||,*,\alpha)$ est égale à la loi des $\{\pi(a_i)\}_{i\in I}$ dans le W^*-espace de probabilité $(B(H),\varphi_\alpha)$.*
4. *Soit $W^*(\{a_i\}_{i\in I})$ l'algèbre de Von-Neumann engendrée par les $\{\pi(a_i)\}_{i\in I}$. Si α est une trace, alors la restriction de φ_α à $W^*(\{a_i\}_{i\in I})$ est aussi une trace. Celle-ci vérifie de plus les deux propriétés suivantes :*
   - *φ_α est fidèle, c'est-à-dire que pour tout $x\in W^*(\{a_i\}_{i\in I})$, $\varphi_\alpha(xx^*)=0$ implique que $x=0$.*
   - *φ_α est normale, c'est dire que que pour tout treillis (a_β) monotone décroissant vers 0, on a $\lim_\beta\varphi_\alpha(a_\beta)=0$*

B.0.2 Liberté et R-transformée

Dans ce cadre des probabilités non-commutatives, la notion d'indépendance se généralise en la notion de liberté :

Définition B.0.9. *Soit $(\mathcal{A}, \varphi)$ un espace de probabilité non-commutatif et $\mathcal{A}_1, ..., \mathcal{A}_k$ des sous-algèbres de $\mathcal{A}$. On dit que $\mathcal{A}_1, ..., \mathcal{A}_k$ sont libres si pour tout $l \in \mathbb{N}$, $i_1, ..., i_l \in [1, k]$ tels que $i_1 \neq i_2$,...,$i_{k-1} \neq i_k$ et tout $a_1, ..., a_l$ tels que $a_j \in \mathcal{A}_{i_j}$ et $\varphi(a_j) = 0$:*

$$\varphi(a_1...a_l) = 0$$

Les variables aléatoires non-commutatives $(X_1, ..., X_k)$ sont dites libres si les sous-algèbres qu'elles engendrent sont libres.

Là encore, cette définition est à comparer avec l'indépendance dans le cas des probabilités classiques. Dans le cas classique, on a que $X_1, ...X_k$ sont indépendantes si et seulement si pour toutes fonctions continues bornées $f_1, ..., f_k$ telles que $\mathbb{E}[f_j(X_j)] = 0$ on a $\mathbb{E}[f_1(X_1)...f_k(X_k)] = 0$. A l'instar du cas classique dans lequel la loi d'un k-uplet de variables aléatoires indépendantes ne dépend que des marginales, on a la propriété suivante :

Proposition B.0.10. *Si $a_1, ..., a_n \in \mathcal{A}$ sont des variables aléatoires non-commutatives auto-adjointes libres, alors la loi du n-uplet ne dépend que des lois de $a_1, ..., a_n \in \mathcal{A}$.*

Cette proposition peut se démontrer par récurrence sur le degré des moments joints en remarquant que pour tout $l \in \mathbb{N}$, $i_1, ..., i_l \in [1, k]$ tels que $i_1 \neq i_2$,...,$i_{k-1} \neq i_k$ et $P_1, ..., P_l \in \mathbb{C}[X]$, on a :

$$\Phi((P(a_{i_1}) - \varphi(P(a_{i_1})))...(P(a_{i_l}) - \varphi(P(a_{i_l})))) = 0$$

On peut alors obtenir les moments joints de $a_1, ..., a_n$ en fonction de moments de degré plus petit.

Cette notion de liberté permet de décrire plus précisément la loi $\sigma^{(n)}$ qui apparaît dans le théorème B.0.1 :

Proposition B.0.11. *La loi $\sigma^{(n)}$ dans le théorème B.0.1 est la loi de n variables libres chacune de loi semi-circulaire.*

Au matrices de Wigner du théorème B.0.1, on peut également ajouter des matrices constantes qui convergent en loi :

Théorème B.0.12. *Soit $(X_N^1, ..., X_N^n)$ une suite de matrices de Wigner indépendantes et $(D_N^1, ..., D_N^k)$ une suite de matrices constantes telles que la loi du k-uplet $(D_N^1, ..., D_N^k)$ dans $(\mathcal{M}_{N,N}(\mathbb{C}), ||.||, *, N^{-1}\mathrm{Tr})$ converge une certaine loi non-commutative μ_D. De plus on suppose :*

$$D := \sup_{i \in [1,k], l \in \mathbb{N}, N \in \mathbb{N}} \frac{1}{N} \mathrm{Tr}(|D_i^N|^l)^{1/l} < \infty$$

Alors on a :

1. *La loi $\bar{\mu}_N$ de $(X_N^1, ..., X_N^n, D_N^1, ..., D_N^k)$ dans l'espace $(\mathcal{M}_{N\times N}(\mathcal{L}^2(\Omega)), N^{-1}\mathbb{E}[\mathrm{Tr}.])$ converge vers la loi μ de $(s_1, ..., s_n, (d_1, ..., d_k))$ où $s_1, ..., s_n$ et $(d_1, ..., d_k)$ sont libres, la loi de chacun des s_j est une semi-circulaire et la loi du k-uplet $(d_1, ..., d_k)$ est μ_D.*
2. *Presque sûrement, la loi $\hat{\mu}_N$ de $(X_N^1, ..., X_N^n, D_N^1, ..., D_N^k)$ dans l'espace $(\mathcal{M}_{N\times N}(\mathbb{C}), N^{-1}\mathrm{Tr})$ converge vers μ.*

On peut également comprendre la loi limite μ en utilisant la notion de dérivée non-commutative sur l'anneau de polynôme $\mathcal{A} = \mathbb{C}\langle X_1, ..., X_n, D_1, ..., D_k\rangle$. Pour tout $i \in [1, n]$, on définit $\partial_{X_i} : \mathcal{A} \to \mathcal{A} \otimes \mathcal{A}$ comme l'application linéaire définie sur les monômes q par :

$$\partial_{X_i} q = \sum_{\substack{q_1, q_2 \\ q = q_1 X_i q_2}} q_1 \otimes q_2$$

On a alors la caractérisation suivante de la mesure μ :

Proposition B.0.13. *En reprenant les hypothèses du théorème précédent, la loi μ est caractérisée par :*

1. $\forall P, Q \in \mathcal{A}, \mu(PQ) = \mu(QP)$
2. $\forall i \in [1,n], \forall P \in \mathcal{A}, \mu(X_i P) = \mu \otimes \mu(\partial_{X_i} P)$
3. $\forall R \in \mathbb{C}\langle D_1, ..., D_k\rangle, \mu(R) = \mu_D(R)$

Les équations précédentes sont souvent appelées *équations de Schwinger-Dyson*. Là encore, on peut les comparer au cas gaussien classique. En effet si X est une variable gaussienne standard et si f est une fonction $\mathcal{C}^1$ telles que $|f|$ et $|f'|$ soit bornée par un polynôme, on a :

$$\mathbb{E}[Xf(X)] = \mathbb{E}[f'(X)]$$

L'utilisation de ces équations de Schwinger-Dyson ne se limite pas au cas de la distribution limite de matrices de Wigner. Elles décrivent également la distribution limite pour des matrices de Haar indépendantes (voir [4, Theorem 5.4.10]) et pour des modèles avec potentiel. Ainsi la notion de liberté permet de décrire complètement la distribution limite d'un polynôme en des matrices de Wigner. On va à présent exposer les outils classiques utilisés pour déterminer la distribution de la somme de deux variables libres.

En d'autres termes si a et b sont deux variables aléatoires non-commutatives auto-adjointes de lois respectives μ et ν, on cherche a déterminer la loi de la somme $a + b$. On notera cette loi $\mu \boxplus \nu$ et on l'appelle la somme libre des lois μ et ν.

A l'instar de la théorie des cumulants en probabilité classique, il existe une théorie des cumulants non-commutatifs :

Définition B.0.14. *Soit $(\mathcal{A}, \varphi)$ un espace de probabilité non-commutatif. Il existe une collection k_n de fonctions multilinéaires de $\mathcal{A}^n$ dans $\mathbb{C}$ telle que pour tout $a_1, ..., a_n \in \mathcal{A}$:*

$$\varphi(a_1...a_n) = \sum_{\pi \in NC(n)} \prod_{\substack{\{i_1,\dots i_l\}\in\pi \\ i_1<\dots<i_l}} k_l(a_{i_1}, ..., a_{i_l})$$

où $NC(n)$ est l'ensemble des partitions sans croisements de $[1,n]$.

La propriété suivante est l'analogue de la propriété des cumulants classiques dont l'annulation caractérise l'indépendance :

Proposition B.0.15. *Soit $(\mathcal{A}, \varphi)$ un espace de probabilité non-commutatif et $\mathcal{A}_1, ..., \mathcal{A}_n$ des sous-algèbres de $\mathcal{A}$. Alors $\mathcal{A}_1, ..., \mathcal{A}_n$ sont libres si et seulement si pour tout $l \in \mathbb{N}$, $i_1, ..., i_l \in [1,n]$ et $i_1 \neq i_2 ... \neq i_l$ et $a_1, ..., a_l$ tels que $a_j \in \mathcal{A}_{i_j}$:*

$$k_l(a_1, ..., a_l) = 0$$

On peut alors démontrer la relation suivante pour les cumulants d'une somme libre :

Proposition B.0.16. *Soit a et b deux variables aléatoires non-commutatives libres. En posant $k_n(a) = k_n(a, ..., a)$ on a :*

$$k_n(a+b) = k_n(a) + k_n(b)$$

Comme le logarithme de la transformée de Fourier-Laplace qui caractérise la loi dans le cas classique et dont les coefficients dans le développement en série entière sont les cumulants, on construit la R-transformée d'une loi μ de la façon suivante grâce aux cumulants non-commutatifs :

Définition B.0.17. *Soit a une variable aléatoire non commutative de loi μ, on note R_μ la série entière suivante :*

$$R_\mu(z) = \sum_{n=1}^{+\infty} k_n(a) z^n$$

Cette série est la R-transformée de la mesure μ

On déduit alors naturellement de la proposition précédente que pour toute mesure de probabilité μ, ν, on a (au moins au sens des série formelles) :

$$R_{\mu \boxplus \nu} = R_\mu + R_\nu$$

Il reste maintenant à donner une expression de R_μ directement en fonction de μ. Rappelons que si μ est une mesure de probabilité, sa transformée de Stieljes G_μ est définie pour $z \in \mathbb{C}$ en dehors du support de μ par :

$$G_\mu(z) := \int \frac{1}{z-t} d\mu(t)$$

On note K_μ son inverse formel en $+\infty$, c'est à dire la fonction telle que $G_\mu(K_\mu(z)) = z$, on a alors que K_μ est de la forme :

$$K_\mu(z) = \frac{1}{z} + \sum_{n=1}^{\infty} C_n z^{n-1}$$

On a alors la résultat suivant :

Proposition B.0.18. *Si μ a un support compact, les C_n sont les cumulants non-commutatifs de μ et :*

$$R_\mu(z) = K_\mu(z) - \frac{1}{z}$$

On a donc un moyen de calculer la somme libre $\mu \boxplus \nu$. En particulier on a que la R-transformée d'une mesure semi-circulaire de variance σ^2 est :

$$R_\sigma(z) = \sigma^2 z$$

La somme libre de deux semi-circulaires est une semi-circulaire dont la variance est la somme des deux variances. La semi-circulaire est donc un analogue non-commutatif de la loi gaussienne.

Bibliographie

[1] Oskari H. Ajanki, Laszlo Erdos, and Torben Krüger. Singularities of solutions to quadratic vector equations on complex upper half-plane. *arXiv e-prints*, page arXiv :1512.03703, Dec 2015.

[2] Oskari H. Ajanki, László Erdős, and Torben Krüger. Universality for general Wigner-type matrices. *Probability Theory and Related Fields*, 169(3) :667–727, Dec 2017.

[3] Oskari H. Ajanki and Torben Kruger. *Quadratic Vector Equations on Complex Upper Half-plane.* Memoirs of the American Mathematical Society. American Mathematical Society, 2019.

[4] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[5] Fanny Augeri. Large deviations principle for the largest eigenvalue of Wigner matrices without Gaussian tails. *Electron. J. Probab.*, 21 :32–49, 2016.

[6] Fanny Augeri. Nonlinear large deviation bounds with applications to traces of Wigner matrices and cycles counts in Erdos-Renyi graphs. *ArXiv :1810.01558*, 2018.

[7] Fanny Augeri, Alice Guionnet, and Jonathan Husson. Large deviations for the largest eigenvalue of sub-Gaussian matrices. *In preparation*, 2019.

[8] Zhi-Dong Bai and Yong-Qua Yin. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *The Annals of Probability*, pages 1729–1741, 1988.

[9] Zhidong Bai. Convergence rate of expected spectral distributions of large random matrices. I. Wigner matrices. *Ann. Probab.*, 21(2) :625–648, 1993.

[10] Zhidong Bai. Convergence rate of expected spectral distributions of large random matrices. II. Sample covariance matrices. *Ann. Probab.*, 21(2) :649–672, 1993.

[11] Zhidong. Bai and Jack W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.*, 26(1) :316–345, Jan 1998.

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

[13] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5) :1643–1697, 2005.

[14] Serban Belinschi, Tobias Mai, and Roland Speicher. Analytic subordination theory of operator-valued free additive convolution and the solution of a general random matrix problem. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 0, Mar 2013.

[15] Serban T. Belinschi, Piotr Śniady, and Roland Speicher. Eigenvalues of non-hermitian random matrices and brown measure of non-normal operators : Hermitian reduction and linearization method. *Linear Algebra and its Applications*, 537 :48– 83, 2018.

[16] Gérard Ben Arous, Amir Dembo, and Alice Guionnet. Aging of spherical spin glasses. *Probab. Theory Related Fields*, 120(1) :1–67, 2001.

[17] Gérard Ben Arous and Alice Guionnet. Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probability Theory and Related Fields*, 108(4) :517–542, Aug 1997.

[18] Florent Benaych-Georges, Alice Guionnet, and Mylène Maïda. Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Probab.*, 16 :1621–1662, 2011.

[19] Bhaswar Bhattacharya and Shirshendu Ganguly. Upper tails for edge eigenvalues of random graphs. *arXiv : 1811.07554*, Nov 2018.

[20] Pascal Bianchi, Mérouane Debbah, Mylène Maïda, and Jamal Najim. Performance of statistical tests for single-source detection using random matrix theory. *IEEE Trans. Inform. Theory*, 57(4) :2400–2419, 2011.

[21] Alex Bloemendal, László Erdös, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.*, 19 :33–53, 2014.

[22] Sergey G. Bobkov and Michel Ledoux. From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. *Geometric & Functional Analysis GAFA*, 10(5) :1028–1052, Dec 2000.

[23] Charles Bordenave and Pietro Caputo. A large deviation principle for Wigner matrices without Gaussian tails. *Ann. Probab.*, 42(6) :2454–2496, 2014.

[24] Alexei Borodin, Vadim Gorin, and Alice Guionnet. Gaussian asymptotics of discrete beta-ensembles. *Publications mathématiques de l'IHÉS*, 125(1) :1–78, Jun 2017.

[25] L. G. Brown. Lidskiĭ's theorem in the type II case. In *Geometric methods in operator algebras (Kyoto, 1983)*, volume 123 of *Pitman Res. Notes Math. Ser.*, pages 1–35. Longman Sci. Tech., Harlow, 1986.

[26] Thierry Cabanal Duvillard and Alice Guionnet. Large deviations upper bounds for the laws of matrix-valued processes and non-communicative entropies. *Ann. Probab.*, 29(3) :1205–1261, Jul 2001.

[27] Anthony Carbery and James Wright. Distributional and L_q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. *Mathematical research letters*, 8(3) :233–248, May 2001.

[28] Sourav Chatterjee and Amir Dembo. Nonlinear large deviations. *Advances in Mathematics*, 299 :396–450, 2016.

[29] Sourav Chatterjee and S.R.S. Varadhan. The large deviation principle for the Erdös-Rényi random graph. *European Journal of Combinatorics*, 32(7) :1000–1017, 2011. Homomorphisms and Limits.

[30] Sourav Chatterjee and S.R.S. Varadhan. Large deviation for random matrices. *Communications on Stochastic Analysis*, 6(1), 2012.

[31] Benoît Collins and Piotr Sniady. New scaling of Itzykson-Zuber integrals. *Annales de L'Institut Henri Poincare Section (B) Probability and Statistics*, 43(2) :139–146, Mar 2007.

[32] Nicholas Cook. Lower bounds for the smallest singular value of structured random matrices. *Ann. Probab.*, 46(6) :3442–3500, Nov 2018.

[33] Nicholas. Cook and Amir Dembo. Large deviations of subgraph counts for sparse Erdös-Rényi graphs. *arXiv : 1809.11148*, 2018.

[34] Nicholas Cook, Alice Guionnet, and Jonathan Husson. Smallest singular values of polynomials of degree 2 in i.i.d. matrices. *In progress*, 2020.

[35] David S. Dean and Satya N. Majumdar. Large deviations of extreme eigenvalues of random matrices. *Phys. Rev. Lett.*, 97(16) :160201, 4, 2006.

[36] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*, volume 95. Springer, Jan 2010.

[37] Catherine Donati-Martin and Mylène Maïda. Large deviations for the largest eigenvalue of an Hermitian Brownian motion. *arXiv e-prints*, page arXiv :1102.2970, Feb 2011.

[38] Ronen Eldan. Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. *Geometric and Functional Analysis*, 28(6) :1548–1596, 2018.

[39] László Erdős, Torben Krueger, and Dominik Schröder. Cusp universality for random matrices I : Local law and the complex hermitian case. *arXiv e-prints*, page arXiv :1809.03971v3, Sept 2018.

[40] Anne Fey, Remco van der Hofstad, and Marten J. Klok. Large deviations for eigenvalues of sample covariance matrices, with applications to mobile communication systems. *Adv. in Appl. Probab.*, 40(4) :1048–1071, 2008.

[41] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3) :233–241, 1981.

[42] Vyacheslav L. Girko. Circular law. *Theory Probab. Appl.*, 29 :694–706, 1984.

[43] Vyacheslav L. Girko. *Canonical Equation K1*, pages 1–24. Springer Netherlands, Dordrecht, 2001.

[44] Friedrich Götze and Alexander Tikhomirov. The circular law for random matrices. *The Annals of Probability*, 38(4) :1444–1491, 2010.

[45] Alice Guionnet and Jonathan Husson. Large deviations for the largest eigenvalue of Rademacher matrices. *arXiv e-prints*, page arXiv :1810.01188, Oct 2018.

[46] Alice Guionnet, Manjunath Krishnapur, and Ofer Zeitouni. The single ring theorem. *Ann. Math. (2)*, 174(2) :1189–1217, 2011.

[47] Alice Guionnet and Mylène Maïda. A Fourier view on the R-transform and related asymptotics of spherical integrals. *Journal of Functional Analysis*, 222(2) :435–490, 2005.

[48] Alice Guionnet and Mylène Maïda. Large deviations for the largest eigenvalue of the sum of two random matrices. *arXiv e-prints*, page arXiv :1810.02538, Oct 2018.

[49] Alice Guionnet and Ofer Zeitouni. Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.*, 5 :119–136, 2000.

[50] Uffe Haagerup and Steen Thorbjørnsen. A new application of random matrices : $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group. *Ann. of Math. (2)*, 162(2) :711–775, 2005.

[51] Adrien Hardy and Arno B. J. Kuijlars. Large deviations for a non-centered Wishart matrix. *Random Matrices : Theory and Applications*, 02(01), 2013.

[52] Fumio Hia and Dénes Petz. A large deviation theorem for the empirical eigenvalue distribution of random unitary matrices. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 36(1) :71–85, 2000.

[53] Fumio Hiai and Dénes Petz. Eigenvalue density of the Wishart matrix and large deviations. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 01(04) :633–646, 1998.

[54] Jonathan Husson. Large deviations for the largest eigenvalue of matrices with variance profile. *In preparation*, 2019.

[55] Nicholas M. Katz and Peter Sarnak. *Random matrices, Frobenius eigenvalues, and monodromy*, volume 45. American Mathematical Soc., 1999.

[56] Rafal Latala. L^1-norm of combinations of products of independent random variables. *Israel J. Math.*, pages 295–308, 2014.

[57] Michel Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[58] Mylène Maïda. Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles. *Electron. J. Probab.*, 12 :1131–1150, 2007.

[59] Vladimir A. Marčenko and Pastur Leonid A. Distribution of eigenvalues in certain sets of random matrices. *Math. USSR Sb.*, 1 :457–483, 1967. English translation of Mat. Sbornik 72 507–536.

[60] Madan Lal Mehta. *Random matrices*, volume 142. Elsevier, 2004.

[61] Mark Rudelson and Roman Vershynin. The Littlewood–Offord problem and invertibility of random matrices. *Advances in Mathematics*, 218(2) :600–633, 2008.

[62] Dimitri Shlyakhtenko and Paul Skoufranis. Freely independent random variables with non-atomic distributions. *Transactions of the American Mathematical Society*, 367 :6267–6291, May 2013.

[63] Marta Strzelecka, Michał Strzelecki, and Tomasz Tkocz. On the convex infimum convolution inequality with optimal cost function. *Alea*, 14 :903–915, Feb 2017.

[64] Terence Tao and Van Vu. Random matrices : The circular law. In *Communications in Contemporary Mathematics*, pages 261–307.

[65] Terence Tao, Van Vu, and Manjunath Krishnapur. Random matrices : Universality of ESDs and the circular law. *Ann. Probab.*, 38(5) :2023–2065, Sep 2010.

[66] Craig A. Tracy and Harold Widom. Level-spacing distributions and the Airy kernel. *Communications in Mathematical Physics*, 159(1) :151–174, Jan 1994.

[67] Craig A. Tracy and Harold Widom. On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.*, 177(3) :727–754, 1996.

[68] Pierpaolo Vivo, Satya N. Majumdar, and Oriol Bohigas. Large deviations of the maximum eigenvalue in Wishart random matrices. *J. Phys. A*, 40(16) :4317–4337, 2007.

[69] Dan Voiculescu. Limit laws for random matrices and free products. *Inventiones mathematicae*, 104(1) :201–220, Dec 1991.

[70] Dan Voiculescu. The analogues of entropy and of Fisher's information measure in free probability theory. I. *Comm. Math. Phys.*, 155(1) :71–92, 1993.

[71] Dan Voiculescu. The analogues of entropy and of Fisher's information measure in free probability theory, II. *Inventiones mathematicae*, 118(1) :411–440, Dec 1994.

[72] Dan Voiculescu. The analogues of entropy and of fisher's information measure in free probability theory, V. Noncommutative Hilbert transforms. *Inventiones mathematicae*, 132 :189–227, 1994.

[73] Dan Voiculescu. The analogues of entropy and of Fisher's information measure in free probability theory, III. The absence of Cartan subalgebras. *Geometric and Functional Analysis GAFA*, 6 :172–199, 1996.

[74] Dan Voiculescu. The analogues of entropy and of Fisher's information measure in free probability theory, VI. Liberation and mutual free information. *Advances in Mathematics*, 146(2) :101–166, 1999.

[75] Dan Voiculescu. Free entropy. *arXiv Mathematics e-prints*, page math/0103168, Mar 2001.

[76] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, 62(3) :548–564, 1955.

[77] Eugene P. Wigner. On the distribution of the roots of certain symmetric matrices. *Annals Math.*, 67 :325–327, 1958.

[78] John Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A(1) :32–52, 1928.

Résumé :
L'un des principaux objets d'étude de la théorie des matrices aléatoires est le spectre de matrices dont les coefficients sont des variables aléatoires et dont la dimension est grande. Dans cette thèse, on s'intéresse à deux questions concernant le spectre de matrices aléatoires : les grandes déviations de la plus grande valeur propre et la convergence de la mesure empirique.

En dehors des cas où les coefficients sont à distributions gaussiennes ou à queues lourdes, peu de principes de grandes déviations sont connus en théorie des matrices aléatoires, tant pour la mesure empirique que pour la plus grande valeur propre. Dans la première partie de cette thèse on présente une série de principes de grandes déviations pour la plus grande valeur propre des matrices de Wigner, de Wishart et des matrices à profils de variance ayant des coefficients dont la distribution vérifie une borne sous-gaussienne.

Dans la seconde partie de cette thèse, on s'intéresse à la convergence de la mesure empirique pour des polynômes de matrices aléatoires. Dans le cas de polynômes auto-adjoints en des matrices de Wigner indépendantes, les travaux de Voiculescu permettent de voir la limite de la mesure empirique comme la mesure spectrale d'un polynôme en des éléments semi-circulaires libres. Toutefois dans le cas général, il est nécessaire d'avoir un contrôle sur les plus petites valeurs singulières pour conclure. On présente un tel résultat pour le cas particulier d'un polynôme de degré 2 en des matrices de Ginibre ainsi que la preuve de la convergence de la mesure empirique.

Mots-clé : Matrices aléatoires, grandes déviations, plus grande valeur propre, mesure de Brown, polynôme de matrices aléatoires.

Abstract :
One of the main objects of random matrix theory is the spectrum of matrices of large dimension and whose entries are random variables. In this thesis, we concern ourselves with two questions : the large deviations of the largest eigenvalue and the convergence of the empirical measure.

Apart from the case of coefficients with Gaussian or heavy-tailed distributions., few large deviation principles are proved in random matrix theory, for the empirical measure or the largest eigenvalue. In the first part of this thesis, we prove a series of large deviation principles for the largest eigenvalue of Wigner matrices, Wishart matrices and matrices with variance profiles with coefficients whose distributions satisfy a sub-Gaussian bound.

In the second part of this thesis, we examine the question of the convergence of the empirical measure of polynomials of random matrices. In the case of self-adjoint polynomials in independent Wigner matrices, thanks to the results of Voiculescu, we can see the limit measure as the spectral measure of a polynomial in independently free semi-circular elements. But in the general case, it is necessary to control the smallest singular values in order to conclude. We present such a control for polynomials of degree 2 in Ginibre matrices and we prove the convergence of the empirical measure.

Keywords : Random matrices, large deviations, largest eigenvalue, Brown measure, polynomial in random matrices