université PARIS-SACLAY

Spherical integrals and their applications to random matrix theory

Les intégrales sphériques et leurs applications à la théorie des matrices aléatoires

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°564, Physique en Île-de-France (PIF)
Spécialité de doctorat: Physique
Graduate school: Physique , Référent : Faculté des sciences d'Orsay

Thèse préparée au Laboratoire de Physique Théorique et Modèles Statistiques (LPTMS Université Paris-Saclay, CNRS), sous la direction de Satya MAJUMDAR, directeur de recherche, et le co-encadrement de Marc POTTERS, chercheur à Capital Fund Management.

Thèse soutenue à Paris, le 12 octobre 2022, par

Pierre MERGNY

Composition du jury

Giulio BIROLI Professeur, École Normale Supérieure de Paris	Président
Adam W. MARCUS Professeur assistant, École Polytechnique Fédérale de Lausanne	Rapporteur & Examinateur
Pierpaolo VIVO Maître de conférence, King's College London	Rapporteur & Examinateur
Florent BENAYCH-GEORGES Chercheur, Capital Fund Management	Examinateur
Alice GUIONNET Directrice de Recherche, École Normale Supérieure de Lyon	Examinatrice
Jorge KURCHAN Directeur de Recherche, École Normale Supérieure de Paris	Examinateur
Satya MAJUMDAR Directeur de Recherche, Université Paris-Saclay Laboratoire de Physique Théorique et Modèles Statistiques	Directeur de thèse

THESE DE DOCTORAT
NNT : 2022UPASP103

université PARIS-SACLAY

ÉCOLE DOCTORALE
Physique en Île-de-France (PIF)

Titre: Les intégrales sphériques et leurs applications à la théorie des matrices aléatoires .....
Mots clés: matrices aléatoires, intégrales sphériques, probabilités libres, grandes déviations

Résumé: La théorie des matrices aléatoires a des applications dans de nombreux domaines de la physique (systèmes désordonnés, stabilité de systèmes dynamiques, modèles d'interfaces, transport électronique,...) et des mathématiques (algèbre d'opérateurs, combinatoire énumérative, théorie des nombres,...). Un cas de figure récurrent dans de nombreux domaines consiste à comprendre comment les spectres de deux matrices aléatoires se recombinent quand on effectue leur somme ou leur produit. Dans cette thèse, on étudie ce problème par le prisme des intégrales sphériques et à l'aide d'outils issus de la physique statistique. Ces intégrales sphériques jouent le rôle de la transformée de Fourier dans la théorie des matrices aléatoires et leur étude permet de mieux comprendre les propriétés de la densité limite des valeurs propres de ces modèles de matrices ainsi que le comportement de leur plus grande valeur propre.

Title: Spherical integrals and their applications to random matrix theory .....
Keywords: random matrices, spherical integrals, free probability, large deviations

Abstract: Random matrix theory has found applications in many fields of physics (disordered systems, stability of dynamical systems, interface models, electronic transport,...) and mathematics (operator algebra, enumerative combinatorics, number theory,...). A recurrent problem in many domains is understanding how the spectra of two random matrices recombine when we perform their sum or product. In this thesis, we study this problem through the prism of spherical integrals and with the help of statistical physics tools. These spherical integrals play the role of the Fourier transform in random matrix theory and their study allows us to better understand the properties of both the limiting spectral density and the largest eigenvalue of these matrix models.

Remerciements

Comme le veut la tradition, les premières lignes de ce manuscrit sont des remerciements envers les personnes qui m'ont accompagné pendant ces trois belles années de doctorat.

En premier lieu, je tiens à remercier Marc Potters et Satya Majumdar d'avoir accepté d'encadrer ma thèse durant ces trois années. Ce fut un privilège et un honneur d'avoir pu écouter, apprendre, discuter et collaborer avec eux. J'exprime toute ma gratitude envers Satya qui a tout d'abord rendu ce projet possible. Satya a été un directeur bienveillant et généreux tout au long de ma thèse avec qui il a été un plaisir d'échanger. Je tiens ensuite à remercier Marc pour sa disponibilité, sa pédagogie, son encadrement. Marc m'a grandement aidé à développer mon intuition et ma curiosité sur des problèmes de recherche. Je le remercie également pour la liberté qu'il m'a laissée, ses conseils judicieux et pertinents pour orienter la direction de ma recherche. Je mesure pleinement les progrès conséquents que j'ai fait à leurs côtés et je leur en suis très reconnaissants.

Je souhaiterais par ailleurs remercier toutes les personnes qui ont gravité autour de mon projet de thèse. En particulier, je remercie Jean-Philippe Bouchaud pour sa grande disponibilité et son enthousiasme sur mon projet de recherche. Je remercie aussi Michael Benzaquen, du côté de la chaire EconophysiX et Alberto Rosso, côté LPTMS, qui ont participé au bon déroulement de ma thèse. Je remercie enfin Claudine Le Vaou, Véronique Terras, Nathalie Bilmanoff et Vincent Smajda qui ont permis le bon déroulement administratif et logistique de ma thèse, de mon inscription jusqu'à ma soutenance.

Je remercie Giulio Biroli, Florent Benaych-Georges, Alice Guionnet, Jorge Kurchan, Adam Marcus et Pierpaolo Vivo de m'avoir fait l'honneur de participer au jury de ma thèse. Je suis également très honoré que Jean-Bernard Zuber et Jean-Philippe Bouchaud ait pu assister à ma soutenance en tant qu' invités. Je souhaiterais en particulier remercier Adam Marcus et Pierpaolo Vivo d'avoir accepté d'être les rapporteurs de ma thèse.

J'ai effectué la majeure partie de ma thèse dans les locaux de CFM et je souhaiterais remercier tous les membres de la chaire EconophysiX et tout particulièrement Antoine, José, Mehdi, Théo, Michele, Jérôme, Samy, Johannes, Cécilia, Rudy, Karl, Salma, Max et Gianluca avec qui j'ai partagé mon bureau, échangé sur des problèmes scientifiques et aussi (et surtout !) partagé de bons moments en dehors du travail. Je remercie pareillement les membres du LPTMS que j'ai rencontré lors de mes visites à Orsay et plus personnellement Francesco, Benjamin et Ana. Je pense également à toutes les personnes que j'ai rencontrées en dehors de la thèse et qui ont été là pendant ses trois années, mes amis d'enfance, d'école, de master et d'ailleurs.

Je garde une place toute particulière dans ces remerciements pour Sofiane, qui partage mon

quotidien et m'a toujours soutenu durant cette thèse. Elle m'a encouragé à être la meilleure version de moi-même. Enfin, je tiens à remercier mes parents et mon frère qui ont éveillé et aiguisé ma curiosité dès le plus jeune âge. C'est grâce à eux que je me suis dirigé vers des études scientifiques dont cette thèse marque l'aboutissement. Merci pour tout.

Résumé en français

Cette thèse est consacrée à l'étude de certains problèmes liés à la somme et au produit de matrices aléatoires et à leurs liens avec les intégrales sphériques. Les matrices aléatoires ont été découvertes en 1928 par WISHART sur son travail des matrices de covariance empirique. Bien que le travail de Wishart soit maintenant reconnu comme un pilier fondamental de la théorie des matrices aléatoires (abrégé par RMT pour *random matrix theory* en anglais) et des statistiques, il a fallu plus de vingt ans pour que les matrices aléatoires apparaissent dans un contexte très différent, à savoir la modélisation du spectre d'énergie des atomes de noyaux lourds avec le travail séminal de Wigner [188]. Wigner a proposé de modéliser l'Hamiltonien d'un noyau composé d'un grand nombre de nucléons par une matrice symétrique dont les éléments sont des variables aléatoires indépendantes (à la contrainte de symétrie près) et identiquement distribuées. Cela a conduit WIGNER, DYSON [56] et MEHTA [139] et leurs co-auteurs à étudier la théorie des matrices aléatoires à part entière. Depuis leurs travaux fondateurs dans les années 1960, la théorie des matrices aléatoires a connu un vaste développement et des ramifications dans de nombreux domaines de la physique (théorie des verres de spin et paysages rugueux [104, 131, 9, 8, 19, 156], étude d'interfaces en croissance [106, 162, 159], physique des hautes énergies [96, 179], information et transport quantiques [18, 61, 148]...), des mathématiques (théorie des nombres et statistiques des zéros de la fonction zêta de Riemann [157, 99, 41], combinatoire et géométrie énumératives [12, 60, 82] ...) et des sciences appliquées (communication sans fil [178, 44, 98],estimation de matrice de covariance et apprentissage statistique [37, 151, 114], ...), pour en citer quelques-uns.

En bref, la théorie des matrices aléatoires vise à comprendre les propriétés statistiques des matrices à éléments aléatoires et un intérêt particulier a été consacré à l'étude des valeurs propres des grandes matrices aléatoires symétriques ou Hermitiennes. Lorsqu'on traite des valeurs propres, deux échelles naturelles apparaissent, la première est l'échelle macroscopique ("Quelle est la forme de la distribution des valeurs propres ?") et la seconde est l'échelle microscopique ("Comment les valeurs propres se comportent-elles comme un processus ponctuel ?"). Dans cette thèse, nous nous intéresserons presque exclusivement au régime macroscopique où la dimension N de la matrice considérée est grande. Étant donné le spectre de deux matrices, une question naturelle est de comprendre comment ces distributions spectrales sont recombinées lorsqu'on effectue la somme ou le produit de ces deux matrices, car de tels modèles apparaissent dans une variété de contextes différents. Pour des matrices de invariantes en loi par conjugation avec une matrice orthogonale et dont la dimension tend vers l'infini, la description de cette densité spectrale limite est donnée par la *théorie des probabilités libres* développée par VOICULESCU [183, 182]. Quand la dimension des matrices en question est finie, leur spectre est profondément lié à la théorie des *intégrales sphériques*, introduite pour la première fois par HARISH-CHANDRA [88] pour construire un analogue non-commutatif de l'analyse harmonique.

Par conséquent, comprendre le comportement de ces intégrales sphériques est d'une grande utilité pour répondre à certaines questions concernant le spectre de la somme ou du produit de matrices aléatoires et comme expliqué dans le premier paragraphe, c'est précisément l'objectif de cette thèse. En particulier, nous utiliserons ces asymptotiques pour aborder le problème de dit de *grandes déviations* de la valeur propre la plus grande d'une telle matrice, c'est-à-dire l'estimation de la probabilité d'avoir cette valeur propre la plus grande loin de sa valeur typique. Un autre aspect développé dans cette thèse est la construction d'une famille de convolution (la convolution à haute température ou *high-temperature convolution* en anglais) qui peut être vue comme une opération entre les densités de β-ensembles pris dans un régime très spécifique.

Vue d'ensemble du manuscript

Cette thèse est divisée en six chapitres. Les deux premiers chapitres de cette thèse sont des revues de résultats connus de la théorie des matrices aléatoires et de leur liens avec les β-ensembles, tandis que le contenu des quatre derniers chapitres est basés sur des articles ([141, 144, 143]) et des pré-publications ([142, 140]) que j'ai écrits au cours de ma thèse. À la fin de chaque chapitre, un résumé des principaux résultats développés dans le chapitre est donné. En particulier, Le lecteur déjà familiarisé avec la théorie des matrices aléatoires et la théorie des probabilités libres devrait pouvoir lire relativement rapidement le contenu de ces deux premiers chapitres voir même directement lire leur résumé. Néanmoins, je pense que l'inclusion de ces deux chapitres est hautement bénéfique pour le lecteur (familier ou non avec la théorie des matrices aléatoires) afin de mieux appréhender les résultats développés dans les quatre chapitres suivants.

Il me semble important de souligner que les résultats de cette thèse sont obtenus à l'aide d'outils et de méthodes issus de la physique statistique et ne s'inscrivent pas dans un cadre mathématique strictement rigoureux (en particulier, je continuerai à les désigner par "résultats" plutôt que par "théorèmes" pour cette raison). Cependant, nombre de ces résultats sont *presque* rigoureux, dans le sens où il ne faut pas *beaucoup* de travail pour modifier les preuves correspondantes afin de les rendre complètement rigoureuses. Pour cette raison, je pense (ou du moins j'espère) qu'un lecteur ayant une formation mathématique ne devrait pas avoir trop de difficultés à suivre cette thèse.

La suite du manuscrit est organisé comme suit .

- Le chapitre 1 est une introduction générale aux propriétés des matrices aléatoires *individuelles* et en particulier des matrices invariantes par conjugation. Pour ces matrices, la loi des valeurs propres peut être obtenue explicitement et contient un terme de répulsion logarithmique entre celles-ci. Le paramètre $\beta = 1, 2$ ou 4 qui encode la symétrie de ces matrices peut être naturellement étendu à tout $\beta > 0$ et joue le rôle de l'inverse de la température dans le langage de la physique statistique et dans ce cas là on parle de β-ensembles. Dans ce chapitre, un intérêt particulier est consacré au principe de grandes déviations de la plus grande valeur propre de tels ensembles et à l'étude de ces β-ensembles dans (1) le régime standard où le nombre N de valeurs propres est envoyé à l'infini à $\beta > 0$ fixé (2) le régime dit de basse température (*low-temperature regime* en anglais) où $\beta \to \infty$ avec N fixé et (3) le régime dit de haute température où $N \to \infty$ et β tend vers zéro avec N tel que le produit $N\beta/2 \to c > 0$. On montre en particulier que

ces derniers régimes satisfont une certaine dualité, c'est-à-dire que les propriétés de l'un sont obtenus à partir de celles de l'autre par prolongement analytique de leur paramètre respectif.

- Le chapitre 2 est une revue de certaines propriétés de (1) la somme de matrices aléatoires auto-adjointes, (2) le produit de matrices aléatoires auto-adjointes positives et (3) la somme de matrices rectangulaires. Pour chacune de ces trois opérations, nous décrivons d'abord le comportement de la plus grande valeur propre dans le cas auto-adjoint ou de la plus grande valeur singulière dans le cas rectangulaire lors des perturbations de rang un (ce que l'on appelle les *transitions de phase BBP*) et le processus des valeurs propres obtenues après une marche aléatoire sur l'espace matriciel correspondant (le *mouvement Brownien de Dyson* et ses variantes). Nous nous intéressons ensuite au comportement de la distribution spectrale limite dans le cas générique décrit par la *convolution libre* et sa contrepartie finie, la *convolution libre finie*.

- Le chapitre 3 est une description des *intégrales sphériques* associées aux trois opérations précédentes. Ces intégrales sphériques jouent le rôle de la transformée de Fourier dans le cadre des matrices aléatoires. Nous montrons tout d'abord que l'on peut naturellement étendre la définition de ces intégrales pour une valeur $\beta > 0$ quelconque et ainsi *extrapoler* l'opération de somme/produit pour deux β-ensembles au-delà de la restriction $\beta \in \{1, 2, 4\}$. En particulier, le cas $\beta \to \infty$ correspond à la convolution libre finie décrite dans le chapitre précédent. Nous décrivons ensuite le comportement asymptotique de ces intégrales sphériques pour de grands arguments matriciels dans les deux régimes distincts où (1) la variable conjuguée est de 'rang plein' et (2) cette variable conjuguée est de 'rang un'. Ce deuxième régime est lié à la convolution libre du chapitre précédent, et en particulier le cas de l'intégrale sphérique multiplicative est basé sur les résultats de la Réf.[142].

- Dans le chapitre 4, basé sur les résultats décrits dans la Ref.[141], nous utilisons les propriétés dérivées dans les chapitres précédents et en particulier les propriétés du mouvement Brownien de Dyson pour étudier les propriétés de stabilité d'un système linéaire avec des interactions symétriques aléatoires entre les espèces et où le taux de relaxation intrinsèque est hétérogène entre les différentes espèces. Ce modèle est une extension naturelle du modèle-jouet développé par Robert May [137] dans les années 70 et on montre qu'il existe également une transition de stabilité selon la force desinteractions entre les espèces. Cette transition critique peut être interprété comme le temps de contact d'un mouvement Brownian de Dyson. On s'intéresse ensuite à un choix précis de la distribution taux de relaxation pour lequel on a une description explicite de la loi jointe des valeurs propres de la matrice de stabilité.

- Le chapitre 5, basé sur les résultats obtenus dans la Réf. [144] est une description du principe de grandes déviations (à droite) associé à la plus grande valeur propre (ou la plus grande valeur singulière dans le cas de matrices rectangulaires) de la somme ou du produit de deux matrices aléatoires. Ces principes de grandes déviations sont obtenus en pondérant les lois jointes par les intégrales sphériques du Chapitre 3. On peut alors obtenir l'expression de la fonction de taux (*rate function* en anglais) caractérisant les grandes déviations à partir du comportement asymptotique des intégrales sphériques.

- Dans le chapitre 6, basé sur les résultats obtenus dans les références [143, 140], nous construisons la *convolution à haute température*, une famille de convolutions à un paramètre interpolant entre les convolutions classiques et libres. Cette convolution à haute température peut être vu comme la somme naturelle de β-ensembles dans le régime à haute température $N\beta/2 \to c$ et nous montrons qu'elle satisfait une dualité avec la convolution libre finie du Chapitre 2.

Comme énoncé précédemment, on peut distinguer deux axes développés dans cette thèse : le premier concerne les problèmes liés aux principes de grandes déviations de la plus grande valeur propre/valeur singulière d'une matrice aléatoire (ou d'un β-ensemble) et le second concerne l'étude de la convolution haute température. Les deux résultats sont profondément liés aux comportements asymptotiques des intégrales sphériques. Bien que j'encourage la lecture de ce manuscrit de manière 'linéaire', le lecteur intéressé uniquement par le problème des grandes déviations peut sauter en première lecture les Secs. 1.6, 1.7 du chapitre 1 traitant des β-ensembles dans les régimes de basse et haute températures, la Sec.2.7 introduisant la convolution libre finie et le chapitre 6. De même, le lecteur intéressé uniquement par la convolution à haute température peut sauter la lecture de la Sec.1.5 du chapitre 1 et des chapitres 4 et 5.

General introduction

This thesis is devoted to the study of some problems related to the sum and product of random matrices and their link with spherical integrals. Random matrices were first discovered in 1928 by Wishart [189] to describe empirical covariance matrices. Although the work of Wishart is now recognized as a fundamental pillar of random matrix theory (RMT) and statistics, it took more than twenty years for random matrices to appear in a very different context, namely the modeling of the energy spectrum of heavy nucleus atoms with the seminal work of Wigner [188]. Wigner proposed to model the Hamiltonian of atoms composed of numerous nucleons by a real symmetric matrix with elements given as independent (up to the symmetry constraint) and identically distributed random variables. This led Wigner, Dyson [56] and Mehta [139] and their co-authors to start the study of random matrices in their own right. Since their founding work in the 1960s, RMT has undergone numerous developments ramifications to many fields of physics (spin glass and landscape complexity [104, 131, 9, 8, 19, 156], growing interfaces [106, 162, 159], high-energy physics [96, 179], quantum information and transport [18, 61, 148]...), mathematics (number theory and zeros of the Riemann's zeta function [157, 99, 41], combinatorics and enumerative geometry [12, 60, 82] ...) applied sciences (wireless communications [178, 44, 98], covariance matrix estimation and machine learning [37, 151, 114], ...) to cite a few.

In a nutshell, RMT aims at understanding the statistical properties of matrices with random elements and particular interest has been devoted to the study of eigenvalues of large symmetric or Hermitian random matrices. When dealing with eigenvalues, two natural scalings appear, the first one is the macroscopic scaling ("What is the shape of the distribution of eigenvalues ?") and the second one is the microscopic regime ("How do the eigenvalues behave as a point process ?"). In this thesis, we will be almost exclusively interested in the macroscopic regime where the dimension N of the matrix under consideration is large. Given the spectrum of two matrices, a natural question is to understand how these spectral distributions recombine when one does the sum or the product of these two matrices, as such models of matrices appear in a variety of different contexts. For large matrices invariant by orthogonal conjugation, the asymptotic description of the spectrum is given by the theory of *free probability* developed by Voiculescu [183, 182]. At finite size, the spectrum of this type of matrices is deeply connected to the theory of *spherical integrals*, first introduced by Harish-Chandra [88] to construct a non-commutative analog of the harmonic analysis. As a consequence, understanding the behavior of these spherical integrals is of great use to answer questions related to the spectrum of sum or product of random matrices and as explained in the first paragraph, this is exactly the purpose of this thesis. In particular, we will use these asymptoticsto tackle the problem of the large deviation of the top eigenvalue of a matrix and to construct a family of convolution (the high-temperature convolution) which can be seen as the operation between the spectrum

of the so-called β-ensembles taken in a very specific regime.

Overview of this thesis

This thesis is divided into six chapters. The first two chapters of this thesis are reviews of the results of RMT, while the last four chapters contain materials based on articles ([141, 144, 143]) and pre-publications ([142, 140]) I have conducted during my thesis. Each chapter starts with a description of its structure and ends with a summary of the main results developed in the chapter. The reader already familiar with RMT and free convolution will probably already know most of the results presented in the first two chapters. Nevertheless, I believe that the inclusion of these two chapters is highly beneficial to the reader (familiar or not with RMT) in order to better grasp the results developed in the following four chapters.

Importantly, the results of this thesis are obtained using tools from statistical physics and are not set under a strictly rigorous mathematical framework (in particular I will continue to denote them by 'results' rather than by 'theorems' for this reason). However, many of these results are *almost* rigorous, in the sense that it does not require *much* work to modify the proofs to make them completely rigorous (the devil is of course in the words 'almost' and 'much'). For this reason, I believe (or at least hope) that a reader from a mathematical background should not have too much of difficulty following this thesis.

The rest of this thesis is organized as follows.

- Chapter 1 is a general introduction to the properties of *individual* random matrices and in particular of β-ensembles. A particular interest is devoted to the large deviation principle of the top eigenvalue of such ensembles and the study of β-ensembles in (1) the standard regime where the number N of eigenvalues is sent to infinity at fixed $\beta > 0$ (2) the low-temperature regime where $\beta \to \infty$ with N fixed and (3) the high-temperature regime where $\beta \to 0$, $N \to \infty$ with $N\beta/2 \to c > 0$.

- Chapter 2 is a review of some properties of (1) the sum of self-adjoint random matrices, (2) the product of positive self-adjoint random matrices, and (3) the sum of rectangular matrices. For each of these three operations, we first describe the behavior of the top eigenvalue/singular value for rank-one perturbations (the so-called *BBP phase transitions*) and the process of the eigenvalues obtained after a random walk on the corresponding matrix space (the *Dyson Brownian Motion* and its variants). We then turn to the behavior of the limiting spectral distribution in the generic case described by *the free convolution* and its finite counterpart, the *finite free convolution*.

- Chapter 3 is a description of the *spherical integrals* associated to the three previous operations. These spherical integrals play the role of the Fourier transform in the random matrix setting, and we describe their asymptotic behavior for large matrix arguments in the two distinct regimes where (1) the conjugate variable is 'full-rank' and (2) this conjugate variable is of 'rank-one'. This second regime is related to the free convolution of the previous chapter, and in particular, the case for the multiplicative spherical integral is based on the results of Ref. [142].

- In Chapter 4, based on Ref. [141], we make use of the properties derived in the previous chapters to study the stability property of a linear system with random symmetric

interactions and heterogeneous intrinsic rate of decay.

- Chapter 5, based on Ref. [144], is a description of the (right) large deviation principle associated with the top eigenvalue/singular value of the sum or product of random matrices.

- In Chapter 6, based on the materials of Refs. [143, 140], we construct the *high-temperature convolution*, a one-parameter family of convolutions interpolating between the classical and free convolutions. This high-temperature convolution can be understood as the natural sum of β-ensembles in the high-temperature regime.

As I have already explained, one can distinguish two axes developed in this thesis: the first one concerns problems related to the large deviation principle of the top eigenvalue/singular value of a random matrix (or β-ensemble) and the second one concerns the high-temperature convolution. Both results are deeply linked with the asymptotic behavior of spherical integrals. Even though I encourage reading this manuscript in a 'linear' way, the reader only interested in the large deviation problem may skip at first reading Secs. 1.6, 1.7 of Chapter 1 dealing with β-ensembles in the low-temperature and high-temperature regimes, Sec. 2.7 introducing the finite free convolution and Chapter 6. Similarly, the reader only interested in the high-temperature convolution may skip at first reading Sec. 1.5 of Chapter 1 and Chapters 4 and 5.

Contents

Chapter 1

Random matrix theory and β-ensembles in different regimes

1.1 Introduction

In this chapter, we provide a comprehensive overview of classical results of random matrix theory (RMT) and in particular of β-ensembles. First in Sec. 1.2, we introduce the notations and objects, we will use in the rest of the thesis. In Sec. 1.3, we describe the concepts of *self-averaging* and *universality* through the famous classes of RMT. In particular, in this section, we introduce the Gaussian, Laguerre, and Jacobi ensembles, which we will encounter many times in the rest of the thesis. In Sec. 1.4, we introduce the β-ensembles, where now the parameter β of RMT takes an arbitrary positive value. In the standard regime where $N \to \infty$ with β fixed, we describe the behavior of the limiting density, and in the next section 1.5, the behavior of its top eigenvalue. In Sec. 1.6 and in Sec. 1.7, we describe the behaviors of β-ensembles in respectively the low-temperature regime where $\beta \to \infty$ and N is fixed and in the high-temperature regime where $\beta \to 0$, $N \to \infty$ but $N\beta/2 \to c$.

1.2 Preliminary notations on Matrix spaces

1.2.1 Dyson's index and other notations

In this thesis, we will consider matrices with real, complex, or quaternionic entries. To fix things, let's first introduce **Dyson's threefold way index** $\beta = 1, 2, 4$, which encodes both the dimensions of the field $\mathbb{K}_\beta = \mathbb{R}, \mathbb{C}, \mathbb{H}$ of the entries of the matrix and their symmetry, as we will see in the next paragraph. As we will see during the course of this thesis, one of our main goals will be to make sense of *appropriate* objects without the constraint $\beta \in \{1, 2, 4\}$ to allow an analytical extension for $\beta > 0$. For the time being, we have $\beta = 1, 2$ or 4. A complex number $A = A^{(1)} + \mathrm{i}A^{(2)}$ is a complex random variable if both its real part $A^{(1)}$ and imaginary part $A^{(2)}$ are random variables, and the definition for quaternionic random variables is similar.

Real, complex and quaternionic numbers are denoted by a lowercase ($x \in \mathbb{K}_\beta$), vectors are denoted by a bold lowercase letter $\boldsymbol{x} = (x_1, \dots, x_N)$ and matrices $\mathbf{A} = (A_{ij})_{1 \le i,j \le N}$ by a bold uppercase letter. When varying the dimensions of a matrix $\mathbf{A}$, we implicitly mean taking

a sequence $(\mathbf{A}_N)_{1\leq N}$ of such matrices. To lighten notations, we have dropped the dependency in N for $\mathbf{A} \equiv \mathbf{A}_N$ and large matrices or 'large N limit' refers to the asymptotic limit $N \to \infty$ of this sequence.

1.2.2 Self-adjoint, positive self-adjoint, and rectangular matrices

In the rest of this thesis, we will study the spaces of (1) self-adjoint matrices (2) positive self-adjoint matrices, and (3) rectangular matrices and in this section, we describe these three sets.

Self-adjoint matrices -

Our main attention is devoted to **self-adjoint matrices**, that is square matrices with a symmetry constraint depending on the parameter β. For $\beta \in \{1, 2, 4\}$, the set of self-adjoint matrices is denoted by $\mathrm{Herm}_\beta(N)$.

- For $\beta = 1$, the set of self-adjoint matrices corresponds to the set of symmetric matrices

$$\mathrm{Herm}_{\beta=1}(N) := \{\mathbf{A} \in \mathsf{M}_N(\mathbb{R}) \text{ s.t } \mathbf{A} = \mathbf{A}^{\mathsf{T}}\}, \tag{1.1}$$

  where $\cdot^{\mathsf{T}}$ is the transpose operator, that is $\left[\mathbf{A}^{\mathsf{T}}\right]_{ij} = A_{ji}$.

- For $\beta = 2$, the set of self-adjoint matrices corresponds to the set of Hermitian matrices:

$$\mathrm{Herm}_{\beta=2}(N) := \{\mathbf{A} \in \mathsf{M}_N(\mathbb{C}) \text{ s.t } \mathbf{A} = \mathbf{A}^{\dagger}\}, \tag{1.2}$$

  where $\cdot^{\dagger}$ is the conjugate transpose operator, that is $\left[\mathbf{A}^{\dagger}\right]_{ij} = (A^{(1)}_{ji} - \mathrm{i}\, A^{(2)}_{ji})$.

- For $\beta = 4$, the set of self-adjoint matrices corresponds to the set of quaternionic self-dual matrices:

$$\mathrm{Herm}_{\beta=4}(N) := \{\mathbf{A} \in \mathsf{M}_N(\mathbb{H}) \text{ s.t } \mathbf{A} = \mathbf{A}^{\mathsf{R}}\}, \tag{1.3}$$

  where $\cdot^{\mathsf{R}}$ is the quaternionic conjugate transpose operator, that is $\left[\mathbf{A}^{\mathsf{R}}\right]_{ij} = (A^{(1)}_{ji} - \mathrm{i}\, A^{(2)}_{ji} - \mathrm{j}\, A^{(3)}_{ji} - \mathrm{k}\, A^{(4)}_{ji})$.

In the following, we will use the unified notation $\cdot^* = \cdot^{\mathsf{T}}, \cdot^{\dagger}, \cdot^{\mathsf{R}}$ to denote respectively the transpose operator, the conjugate transpose operator, and the quaternionic conjugate transpose operator for $\beta = 1, 2, 4$ respectively and when considering a generic matrix $\mathbf{A}$, we implicitly mean a self-adjoint matrix, unless otherwise stated.

A random self-adjoint matrix is simply a matrix with random entries constrained to satisfy the symmetry $\mathbf{A} = \mathbf{A}^*$. For $\beta \in \{1, 2, 4\}$, the Lebesgue measure $\mathrm{d}\mathbf{A}$ associated with this space is simply given by the product of the independent components of the standard one-dimensional Lebesgue measure, that is

$$\mathrm{d}\mathbf{A} := \prod_{i=1}^{N} \mathrm{d}A_{ii} \prod_{i<j} \prod_{b=1}^{\beta} \mathrm{d}A^{(b)}_{ij}\,. \tag{1.4}$$

Positive and positive semi-definite self-adjoint matrices -

The second ensemble of matrices encountered in this thesis consists of a subset of self-adjoint matrices determined by an inequality constraint. We distinguish the case where the inequality is strict from the one where it is not. Specifically,

- a self-adjoint matrix is **positive semi-definite** and denoted by $\mathbf{A} \in \mathrm{Herm}_\beta^+(N)$ if $\boldsymbol{x}^*\mathbf{A}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in (\mathbb{K}_\beta)^N$.
- Similarly, a self-adjoint matrix is **positive** and denoted by $\mathbf{A} \in \mathrm{Herm}_\beta^{++}(N)$, if the equality is strict that if $\boldsymbol{x}^*\mathbf{A}\boldsymbol{x} > 0$ for all $\boldsymbol{x} \in (\mathbb{K}_\beta)^N$ different from the null vector.

Let's briefly give some properties of these two convex cones. Every $\mathbf{A} \in \mathrm{Herm}_\beta^+(N)$ admits a Cholesky decomposition $\mathbf{A} = \mathbf{R}_A\mathbf{R}_A{}^*$ where $\mathbf{R}_A$ is a lower triangular matrix and similarly there exists a positive semi-definite self-adjoint matrix denoted by $\sqrt{\mathbf{A}}$ such that $\mathbf{A} = \sqrt{\mathbf{A}}\sqrt{\mathbf{A}} = \sqrt{\mathbf{A}}(\sqrt{\mathbf{A}})^*$. If $\mathbf{A}$ is further (strictly) positive, its inverse $\mathbf{A}^{-1}$ is well-defined.

The set of positive (self-adjoint) matrices will play an important role in Chapter 2 when considering the product of matrices: for two positive semi-definite matrices $\mathbf{A}$ and $\mathbf{B}$, one can construct the symmetric products $\sqrt{\mathbf{A}}\mathbf{B}\sqrt{\mathbf{A}}$ and $\sqrt{\mathbf{B}}\mathbf{A}\sqrt{\mathbf{B}}$ which are also positive semi-definite matrices.

Rectangular and chiral matrices -

This paragraph deals with rectangular matrices and in the following all other sections and paragraphs dealing with rectangular matrices are put with a small font as our primary interest lies in self-adjoint matrices.

The third 'matrix space' we will consider in this thesis, corresponds to the set of $(N \times M)$ rectangular matrices $\mathsf{M}_{N,M}(\mathbb{K}_\beta)$ with entries in $\mathbb{K}_\beta = \mathbb{R}, \mathbb{C}, \mathbb{H}$, for $\beta = 1, 2, 4$ respectively. In the following, we will consider $M \geq N$ without loss of generality. For such rectangular matrices, the 'large N limit regime' corresponds to the double scaling limit $N \to \infty$ and $M \to \infty$ but their ratio stays finite:

$$\frac{N}{M} \to q \in (0, 1)\,, \tag{1.5}$$

where q is the *shape ratio*.

To distinguish between quantities associated with rectangular matrices to the ones associated with self-adjoint matrices, we will usually add an index $\cdot_q$ in the former.

For $\beta \in \{1, 2, 4\}$ since there is no symmetry constraint for rectangular matrices, the associated Lebesgue measure $\mathrm{d}\mathbf{A}$ is simply given as the product over each component of each entry of the standard Lebesgue measure, that is:

$$\mathrm{d}\mathbf{A} := \prod_{i=1}^{N}\prod_{j=1}^{M}\prod_{b=1}^{\beta} \mathrm{d}A_{ij}^{(b)}\,. \tag{1.6}$$

Let's point out that one can construct a self-adjoint matrix from a rectangular matrix $\mathbf{A} \in \mathsf{M}_{N,M}(\mathbb{K}_\beta)$ in two natural ways.

- First, one can simply multiply $\mathbf{A}$ by its conjugate $\mathbf{A}^*$ to define a new matrix $\mathbf{A}\mathbf{A}^* \in \mathrm{Herm}_\beta^+(N)$.
- Second, one can define $\mathbf{A}^{\mathrm{ch}} \in \mathrm{Herm}_\beta(N+M)$ by

$$\mathbf{A}^{\mathrm{ch}} := \begin{bmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^* & \mathbf{0} \end{bmatrix}, \tag{1.7}$$

  which is sometimes known as a *chiral matrix* due to its relation to high-energy physics, see for example Ref. [179] and references therein.

The two matrices $\mathbf{A}\mathbf{A}^*$ and $\mathbf{A}^{\mathrm{ch}}$ will appear frequently in the study of rectangular matrices in this thesis.

1.2.3 Compact matrix groups and associated Haar measure

For each $\beta = 1, 2, 4$, there exists a natural compact group of matrices defined by

$$\mathsf{O}_\beta(N) := \{\mathbf{V} \in \mathsf{M}_N(\mathbb{K}_\beta) \text{ s.t } \mathbf{V}\mathbf{V}^* = \mathbf{I}\}, \tag{1.8}$$

where $\mathsf{O}_\beta(N)$ is the unified notation for

- $\mathsf{O}(N) := \{\mathbf{O} \in \mathsf{M}_N(\mathbb{R}) \text{ s.t } \mathbf{O}\mathbf{O}^\mathsf{T} = \mathbf{I}\}$ **the group of orthogonal matrices**, for $\beta = 1$,
- $\mathsf{U}(N) := \{\mathbf{U} \in \mathsf{M}_N(\mathbb{C}) \text{ s.t } \mathbf{U}\mathbf{U}^\dagger = \mathbf{I}\}$ **the group of unitary matrices**, for $\beta = 2$,
- $\mathsf{Sp}(N) := \{\mathbf{S} \in \mathsf{M}_N(\mathbb{H}) \text{ s.t } \mathbf{S}\mathbf{S}^\mathsf{R} = \mathbf{I}\}$ **the group of symplectic matrices** for $\beta = 4$.

For these three compact spaces, we recall that there exists an important measure playing the role of uniform distribution.

> For $\beta = 1, 2, 4$, we can endow the compact groups $\mathsf{O}_\beta(N) \equiv \mathsf{O}(N)$, $\mathsf{U}(N)$, $\mathsf{Sp}(N)$ with a unique measure known as the **Haar measure**, denoted by $\mathbb{P}_{\mathsf{Haar}}$, which sums to one $\mathbb{P}_{\mathsf{Haar}}[\mathsf{O}_\beta(N)] = 1$ and is both left and right invariant, that is for any rotation matrix $\mathbf{V} \in \mathsf{O}_\beta(N)$ and any region $S \subset \mathsf{O}_\beta(N)$ we have $\mathbb{P}_{\mathsf{Haar}}(\mathbf{V}S) = \mathbb{P}_{\mathsf{Haar}}(S\mathbf{V}) = \mathbb{P}_{\mathsf{Haar}}(S)$. We denote its infinitesimal density by $\mu_{\mathsf{Haar}}(\mathrm{d}\mathbf{V})$ and by $\mathbf{V} \sim \mathsf{Unif}[\mathsf{O}_\beta(N)]$ a matrix distributed according to this Haar measure.

Note that if $\mathbf{V} \sim \mathsf{Unif}[\mathsf{O}_\beta(N)]$ and $\mathbf{V}' \in \mathsf{O}_\beta(N)$ is a *fixed* matrix, then the product is also taken uniformly at random, $\mathbf{V}\mathbf{V}' \sim \mathsf{Unif}[\mathsf{O}_\beta(N)]$, by definition of the Haar measure, and we will use this simple property many times in this thesis.

1.2.4 Eigenvalues, singular values, associated distributions and change of variables

Eigenvalue decomposition -

For self-adjoint matrices, the most important matrix factorization is undoubtedly given by the eigenvalue decomposition:

$$\text{if } \mathbf{A} \in \mathrm{Herm}_\beta(N) \quad \text{then} \quad \mathbf{A} = \mathbf{V}\,\mathbf{Diag}(\boldsymbol{\lambda})\,\mathbf{V}^* \quad \text{with } \boldsymbol{\lambda} \in \mathbb{R}^N \text{ and } \mathbf{V} \in \mathsf{O}_\beta(N), \tag{1.9}$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)$ is the set of *real* eigenvalues and $\mathbf{V}$ is the matrix of eigenvectors. Throughout this thesis, the eigenvalues are implicitly ranked in descending order $\lambda_1 \geq \cdots \geq \lambda_N$ and λ_1 is referred to as the top eigenvalue.

If furthermore $\mathbf{A} \in \mathrm{Herm}_\beta^+(N)$ (resp. $\mathbf{A} \in \mathrm{Herm}_\beta^{++}(N)$) then all the eigenvalues are non-negative (resp. positive).

One of the goals of RMT is to understand the statistics of the eigenvalues of a random matrix. To do so, one needs to understand how the law of a matrix is modified when one does the change of variable from the elements $A_{ij}^{(b)}$ to the eigenvalues $\boldsymbol{\lambda}$ and eigenvectors. In other words, one needs to compute the *Jacobian* associated with this change of variable. This Jacobian for this change of variable is well-known and is given by the following result (see for example Ref. [113, 153]).

Result 1.1 (*Weyl's formula for eigenvalues*)

The Jacobian of the change of variable from a self-adjoint matrix $\mathbf{A}$ *to its eigenvalue decomposition* $(\boldsymbol{\lambda}, \mathbf{V})$ *is given by*

$$\left| \left[\frac{\partial \mathbf{A}}{\partial \mathbf{V}}, \frac{\partial \mathbf{A}}{\partial \boldsymbol{\lambda}} \right] \right| = C \prod_{i<j} |\lambda_i - \lambda_j|^\beta = C \, |\Delta(\boldsymbol{\lambda})|^\beta \tag{1.10}$$

where C *is a constant and* $\Delta(\boldsymbol{\lambda}) := \prod_{i<j}(\lambda_j - \lambda_i) = \det\left(\lambda_i^{j-1}\right)_{1 \leq i,j \leq N}$ *is* the Vandermonde determinant.

For β-ensembles of Sec. 1.4, this formula will be fundamental to have the explicit joint law of the eigenvalues. An important question is to study the shape of the distribution of eigenvalues.

To this end, let's introduce the **empirical spectral distribution** (ESD):

$$\mu_{\mathbf{A}}(\lambda) := \frac{1}{N} \sum_{i=1}^{N} \delta(\lambda - \lambda_i(\mathbf{A})) \,, \tag{1.11}$$

where $\delta(.)$ is the Dirac delta function.

If $\mathbf{A}$ is random, its ESD is also a random probability measure. However, for a large class of random matrices as $N \to \infty$, one of the key features of RMT is that their ESD *self-averages* (or *concentrates* in the mathematical language) to a *non-random* limit. This means that for large matrices, one can replace the ESD μ_A by their average $\mathbb{E}\,\mu_{\mathbf{A}}$.

This deterministic limit is known as the **limiting spectral density** (LSD) and is denoted by:

$$\mu_A(\lambda) := \lim_{N \to \infty} \mu_{\mathbf{A}}(\lambda) = \lim_{N \to \infty} \mathbb{E}\,\mu_{\mathbf{A}}(\lambda) \,. \tag{1.12}$$

Let us mention that the LSD only captures the asymptotic behavior of the *bulk* of the spectrum. In other words, one cannot say anything about possible *outliers* or *spikes* outside the bulk, from the knowledge of the LSD alone. One of the main goals of RMT is to give an analytical expression - or at least a complete characterization - of the LSD μ_A given the model for the randomness of the matrix $\mathbf{A}$.

Singular value decomposition -

This paragraph deals with rectangular matrices.

Similarly to the eigenvalue decomposition for self-adjoint matrices, all rectangular matrices $\mathbf{A} \in \mathsf{M}_{N,M}(\mathbb{K}_\beta)$ admit a factorization known as the singular value decomposition (SVD) given by:

$$\mathbf{A} = \mathbf{V}_1 \mathbf{Diag}_q(\boldsymbol{s}) \mathbf{V}_2 \quad \text{with } \boldsymbol{s} \in \mathbb{R}_+^N \ , \ \mathbf{V}_1 \in \mathsf{O}_\beta(N) \text{ and } \mathbf{V}_2 \in \mathsf{O}_\beta(M)\,. \tag{1.13}$$

where $\boldsymbol{s} := (s_1 \equiv s_1(\mathbf{A}), \dots, s_N = s_N(\mathbf{A}))$ is the set of non-negative singular values of the matrix $\mathbf{A}$ and $\mathbf{Diag}_q(\boldsymbol{s})$ denotes the $(N \times M)$ matrix whose only non-zero elements are the diagonal ones determined by the vector $\boldsymbol{s}$:

$$\mathbf{Diag}_q(\boldsymbol{s}) := \begin{bmatrix} s_1 & & & \\ & \ddots & & \mathbf{0}_{M-N} \\ & & s_N & \end{bmatrix}. \tag{1.14}$$

Without loss of generality, we assume $\boldsymbol{s}$ to be in decreasing order $s_1 \geq \cdots \geq s_N \geq 0$. By abuse of notation, we will sometimes refer to $\mathbf{Diag}_q(\boldsymbol{s})$ as a diagonal matrix even if $M > N$.

The SVD of a rectangular matrix is a self-adjoint decomposition in disguise since

- the singular values $s_i(\mathbf{A})$ are related to eigenvalues of $\mathbf{A}\mathbf{A}^*$ by

$$s_i(\mathbf{A}) = \sqrt{\lambda_i(\mathbf{A}\mathbf{A}^*)}\,; \tag{1.15}$$

- and they are also equal to the N largest eigenvalues of the self-adjoint chiral matrix $\mathbf{A}^{\mathrm{ch}}$ defined by Eq. (1.7). The N lowest eigenvalues of $\mathbf{A}^{\mathrm{ch}}$ are the opposite of the N largest eigenvalues and the remaining $M - N$ eigenvalues are equal to zero.

Yet the study of singular values turns out to be an interesting problem on its own, especially when dealing with the *sum* of rectangular matrices, as discussed in the following Chapter.

Similarly to self-adjoint matrices, doing the change of variable from a rectangular matrix to its singular value decomposition introduced a Jacobian which is given by the following Weyl's integral formula.

Result 1.2 (*Weyl's formula for singular values*)

The Jacobian of the change of variable from a rectangular matrix $\mathbf{A}$ to its SVD $(\boldsymbol{\lambda}, \mathbf{V}_1, \mathbf{V}_2)$ is given by

$$\left| \left[\frac{\partial \mathbf{A}}{\partial \mathbf{V}_1}, \frac{\partial \mathbf{A}}{\partial \boldsymbol{s}}, \frac{\partial \mathbf{A}}{\partial \mathbf{V}_2} \right] \right| = C \left| \Delta(\boldsymbol{s}^2) \right|^\beta \prod_{i=1}^N s_i^{\beta(M-N+1)-1} = C \prod_{i<j} |s_i^2 - s_j^2|^\beta \prod_{i=1}^N s_i^{\beta(M-N+1)-1}, \tag{1.16}$$

where C is a normalization constant.

As in the self-adjoint case, we aim at understanding the statistics of the singular values thanks to **the empirical singular value distribution** (ESVD):

$$\mu_{\mathbf{A}}(s) := \frac{1}{N}\sum_{i=1}^{N} \delta(s - s_i(\mathbf{A}))\,, \tag{1.17}$$

and in particular, we will be interested in the behavior of the ESVD in the double scaling limit where $N \to \infty$ and $M \to \infty$ with $N/M \to q$.

This limiting deterministic limit is known as the **limiting singular value density** (LSVD) and is denoted by:

$$\mu_A(\lambda) := \lim_{\substack{N,M\to\infty \\ N/M\to q}} \mu_{\mathbf{A}}(\lambda) = \lim_{\substack{N,M\to\infty \\ N/M\to q}} \mathbb{E}\,\mu_{\mathbf{A}}(\lambda)\,. \tag{1.18}$$

The singular values are all non-negative and hence the support of the LSVD μ_A is on the positive real line. As for the self-adjoint case, we denote by $\mathrm{a}_\pm$ the edges of the distribution. Let's note that we use the same notation for the LSVD of a rectangular matrix and the LSD of a self-adjoint matrix.

One can express the LSVD of a rectangular matrix as the LSD of the matrix $\mathbf{A}\mathbf{A}^*$ and $\mathbf{A}^{\mathrm{ch}}$ of the previous section.

- Thanks to the change of variable given by Eq.(1.15), the LSVD μ_A is expressed in terms of the LSD μ_{AA^*} of the matrix $\mathbf{A}\mathbf{A}^*$ by:

$$\mu_A(s) = 2s\,\mu_{AA^*}(s^2) \quad \text{or equivalently} \quad \mu_{AA^*}(\lambda) = \frac{\mu_A(\sqrt{\lambda})}{2\sqrt{\lambda}}\,. \tag{1.19}$$

- Similarly, the LSD $\mu_{A^{\mathrm{ch}}}$ of the chiral matrix $\mathbf{A}^{\mathrm{ch}}$ of Eq. (1.7) is related to the LSVD μ_A by:

$$\mu_{A^{\mathrm{ch}}}(\lambda) = \frac{2}{q+1}\cdot\frac{1}{2}\left(\mu_A(\lambda) + \mu_A(-\lambda)\right) + \left(1 - \frac{2}{q+1}\right)\delta(x-0)\,. \tag{1.20}$$

For later use, it will be also convenient to introduce the **symmetrized distribution**:

$$\widehat{\mu}_A(\lambda) := \frac{1}{2}\left(\mu_A(\lambda) + \mu_A(-\lambda)\right)\,. \tag{1.21}$$

This corresponds to removing the Dirac mass at zero in the LSD of the chiral matrix $\mu_{A^{\mathrm{ch}}}$ and re-scales it such that it sums to one. For $q = 1$, the coefficient in front of the Dirac mass is null in Eq. (1.20) and we have equality between the two quantities:

$$\widehat{\mu}_A(\lambda) \underset{q=1}{=} \mu_{A^{\mathrm{ch}}}(\lambda)\,. \tag{1.22}$$

1.2.5 RMT transforms

The study of spectral distributions in RMT is generally done through various transforms, which we describe in this section.

Stieltjes transform -

For $\mathbf{A} \in \mathrm{Herm}_\beta(N)$ with spectrum $\boldsymbol{\lambda}$, its **Stieltjes transform** $g_{\mathbf{A}}(z)$ is defined for any $z \in \mathbb{C} \setminus \{\boldsymbol{\lambda}\}$ by:

$$g_{\mathbf{A}}(z) := \frac{1}{N}\mathrm{Tr}\,(z\mathbf{I} - \mathbf{A})^{-1} = \int \frac{\mu_{\mathbf{A}}(\lambda)}{z-\lambda}\,\mathrm{d}\lambda\,, \tag{1.23}$$

where $\mu_{\mathbf{A}}$ is the ESD of $\mathbf{A}$.

The Stieltjes transform is analytical in its domain of definition and admits the following first-order behavior at infinity:

$$g_{\mathbf{A}}(z) \underset{|z|\to\infty}{\sim} \frac{1}{z}\,. \tag{1.24}$$

Furthermore, if we restrict the argument z to be on the real line and higher than the top eigenvalue of $\mathbf{A}$, $z > \lambda_1$, the Stieltjes transform is a continuously decreasing function of its argument. From its definition, one can get for $\mathfrak{Re}\, z > \lambda_1$ the expansion at all order of the Stieltjes transform:

$$g_{\mathbf{A}}(z) = \frac{1}{z} + \sum_{k=1}^{\infty} \frac{(\mathrm{Tr}\mathbf{A}^k)/N}{z^{k+1}} = \frac{1}{z} + \sum_{k=1}^{\infty} \frac{m_k[\mu_{\mathbf{A}}]}{z^{k+1}}\,, \tag{1.25}$$

where $m_k[\mu_{\mathbf{A}}] := \int \lambda^k \mu_{\mathbf{A}}(\lambda)\mathrm{d}\lambda = (\sum_{i=1}^N \lambda_i^k)/N$ is the k^{th} moment of the distribution $\mu_{\mathbf{A}}$. Thus, one can think of the Stieltjes transform as a moment generating function for the spectral distribution.

If the ESD $\mu_{\mathbf{A}}$ converges to the LSD μ_A then $g_{\mathbf{A}}$ converges point-wise to the function g_A defined for any $z \in \mathbb{C} \setminus \mathrm{Supp}[\mu_A]$ as the large N limit of Eq. (1.23):

$$g_A(z) \equiv g_{\mu_A}(z) := \int_{\mathrm{Supp}[\mu_A]} \frac{\mu_A(\lambda)}{z-\lambda}\,\mathrm{d}\lambda\,. \tag{1.26}$$

The converse is also true and one can invert the Stieltjes to get the LSD by looking at its imaginary part near the branch cut since we have for $\lambda \in \mathrm{Supp}[\mu_A]$ and $\eta > 0$:

$$\mathfrak{Im}\, g_A(\lambda - \mathrm{i}\eta) = \int_{\mathrm{Supp}[\mu_A]} \frac{\mu_A(\lambda')\,\eta}{(\lambda-\lambda')^2+\eta^2}\,\mathrm{d}\lambda' = \pi\,(\mu_A * K_\eta)\,(\lambda)\,. \tag{1.27}$$

where $*$ denotes the classical convolution operator $(f*g)(x) := \int f(x-y)g(y)\mathrm{d}y$ and $K_\eta(\lambda) := \frac{1}{\pi}\frac{\eta}{\eta^2+\lambda^2}$ is the *Cauchy Kernel* of width η. As $\eta \to 0^+$ we have $K_\eta(\lambda) \to \delta(\lambda)$, from which we deduce the **Sokochi-Plemelj inversion formula**:

$$\boxed{\mu_A(\lambda) = \frac{1}{\pi}\mathfrak{Im}\, g_A(\lambda - \mathrm{i}0^+)\,.} \tag{1.28}$$

T-transform -

A similar transform appears naturally in the study of the product of positive semi-definite matrices

For $\mathbf{A} \in \mathrm{Herm}_\beta^+(N)$ with spectrum $\boldsymbol{\lambda}$, its **T-transform** $t_{\mathbf{A}}(z)$ is defined for any $z \in \mathbb{C}\backslash\{\boldsymbol{\lambda}\}$ by:

$$t_{\mathbf{A}}(z) := \frac{1}{N}\mathrm{Tr}\left[\mathbf{A}\left(z\mathbf{I}-\mathbf{A}\right)^{-1}\right] = \int \frac{\mu_{\mathbf{A}}(\lambda)\lambda}{z-\lambda}\,\mathrm{d}\lambda\,, \tag{1.29}$$

and similarly for the LSD:

$$t_A(z) \equiv t_{\mu_A}(z) := \int_{\mathrm{Supp}[\mu_A]} \frac{\mu_A(\lambda)\lambda}{z-\lambda}\,\mathrm{d}\lambda\,. \tag{1.30}$$

The T-transform is also a continuously decreasing function for $z > \mathrm{a}_+$ where a_+ is the upper limit of the support of the distribution. The T-transform is related to the Stieltjes transform by:

$$t_A(z) = z g_A(z) - 1\,. \tag{1.31}$$

D-transform -

This paragraph deals with rectangular matrices

In the study of the singular values of the sum of two rectangular matrices, the following transform will be useful:

For $\mathbf{A} \in \mathrm{M}_{N,M}(\mathbb{K}_\beta)$ with LSVD μ_A, the **D-transform** is defined for any $z \in \mathbb{C} \setminus \mathrm{Supp}[\mu_A]$ by:

$$d_A(z) \equiv d_{\mu_A}(z) := \sqrt{\left(\int \frac{z}{z^2-s^2}\mu_A(s)\mathrm{d}s\right)\left(q\int \frac{z}{z^2-s^2}\mu_A(s)\mathrm{d}s + \frac{1-q}{z}\right)}\,. \tag{1.32}$$

- Using the relation (1.19), This transform can also be expressed in terms of the large N Stieltjes transform g_{AA^*} of the matrix $\mathbf{A}\mathbf{A}^*$:

$$d_A(z) = \sqrt{qz^2\left(g_{AA^*}(z^2)\right)^2 + (1-q)g_{AA^*}(z^2)}\,. \tag{1.33}$$

- Similarly, using the identity

$$\frac{1}{z^2-s^2} = \frac{1}{2z}\left(\frac{1}{z-s}+\frac{1}{z+s}\right)\,. \tag{1.34}$$

one can express the D-transform in terms of the Stieltjes transform $\widehat{g}_A(z) \equiv g_{\widehat{\mu}_A}(z) := \int (z-\lambda)^{-1}\widehat{\mu}_A(\mathrm{d}\lambda)$ of the symmetrized density of Eq. (1.21):

$$d_A(z) = \sqrt{\widehat{g}_A(z)\left(q\cdot\widehat{g}_A(z) + \frac{1-q}{z}\right)}\,. \tag{1.35}$$

We conclude this paragraph on the D-transform with two remarks concerning the simplifications in the limiting cases $q \to 0$ and $q \to 1$ which will be useful later on:

Remark *(D-transform for long matrices ($q \to 0$))*. For $q \to 0$, corresponding to a *long matrix*, the D-transform of Eq. (1.32) is simply given as:

$$d_A(z) \xrightarrow[q\to 0]{} \sqrt{g_{AA^*}(z^2)}\,, \tag{1.36}$$

as one can see from Eq. (1.33). ⌟

Remark *(D-transform for square matrices (q=1) and symmetrized density)*. For q=1, corresponding to an (asymptotic) square matrix, the D-transform of Eq. (1.32) considerably simplifies (for $z > \mathrm{a}_+$) into:

$$d_A(z) \xrightarrow[q\to 1]{} \widehat{g}_A(z) \qquad\qquad (\text{for } z > \mathrm{a}_+)\,, \tag{1.37}$$

as one can see from Eq. (1.35). ⌟

1.3 Famous classes of random matrices

For a random matrix $\mathbf{A}$, a natural question is to know what is the *structure of dependency* between its entries (in addition to the trivial one from the symmetry constraint). This *structure of dependency* is what we call the *class of random matrix* and in this section, we review the most famous and important ones of RMT.

1.3.1 Wigner matrices, Gaussian ensembles, and the semi-circle distribution

The most simple structure of dependency between the entries one can think of is to choose them independently up to the symmetric constraint. This leads to the most studied class of random matrices:

For $\beta = 1, 2, 4$, a real (resp. complex, quaternionic) matrix is in the Wigner class or in short, is a generalized **Wigner matrix**, if it is a symmetric (resp. Hermitian, quaternionic self-dual) matrix $\mathbf{A} = (\tilde{A}_{ij}/\sqrt{N})$, such that diagonal and off-diagonal elements are independent, and

1. the diagonal elements $\tilde{A}_{ii}/\sqrt{N}$ are real iid random variables, where $\tilde{A}_{ii}$ have mean zero $\mathbb{E}\,\tilde{A}_{ii} = 0$ and variance $\mathbb{E}\,\tilde{A}_{ii}^2 = 2\sigma^2/\beta$,
2. the off-diagonal elements $(i < j)$ $\tilde{A}_{ij}/\sqrt{N}$ are real (resp. complex, quaternionic) iid random variables, such that $\tilde{A}_{ii}$ have mean zero $\mathbb{E}\,\tilde{A}_{ij} = 0$ and variance $\mathbb{E}\,|\tilde{A}_{ij}|^2 = \sigma^2$.

Among all Wigner matrices, a special role is given when the laws of the entries follow the Gaussian distribution $\mathcal{N}(0, \sigma)$ with density:

$$\mu_{\mathcal{N}(m,\sigma)}(x) := \frac{\mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}\,. \tag{1.38}$$

We say that a self-adjoint matrix $\mathbf{A} = (A_{ij})$ is taken from **the Gaussian orthogonal ensemble** (GOE) for $\beta = 1$, resp. **the Gaussian unitary ensemble** (GUE) for $\beta = 2$ and resp. **the Gaussian symplectic ensemble** (GSE) for $\beta = 4$, if all the elements A_{ij} are independent for $i \geq j$ and

1. its diagonal elements are distributed as $A_{ii} \sim \mathcal{N}(0, 2\sigma/\sqrt{\beta N})$
2. its off-diagonal elements $(i \neq j)$ are distributed as $A_{ji} = A_{ij} \sim \mathcal{N}_\beta(0, \sigma/\sqrt{N})$

The word orthogonal (resp. unitary and symplectic) refers to a certain invariance in the law of the matrix due to the Gaussian nature of its entries, which will be discussed in more detail in Sec. 1.4.

Since we have $\mathrm{Tr}\mathbf{A}^2 = \sum_i A_{ii}^2 + 2\sum_{i<j} |A_{ij}|^2$ we can write the probability to observe a GOE (resp. GUE, GSE) matrix $\mathbf{A}$ in a region R of $\mathrm{Herm}_\beta(N)$ in the following compact form:

$$\mathbb{P}_G^{(\beta)}[\mathbf{A} \in R] = \frac{1}{Z_{N,\beta}} \int_R \mathrm{e}^{-\frac{N\beta}{4\sigma^2}\mathrm{Tr}\,\mathbf{A}^2} \mathrm{d}\mathbf{A}\,, \tag{1.39}$$

where $Z_{N,\beta} = \int_{\mathrm{Herm}_\beta(N)} \mathrm{e}^{-\frac{N\beta}{4\sigma^2}\mathrm{Tr}\,\mathbf{A}^2} \mathrm{d}\mathbf{A}$ is the normalization constant which ensures that the probability measure sums to one. The distribution of the eigenvalues of a matrix taken from these Gaussian ensembles will be discussed in more detail in Sec. 1.4.

One of the key features of RMT is a concept of is *universality*, borrowed from statistical physics, which loosely speaking means that in the 'thermodynamic limit' $(N \to \infty)$, the macroscopic observables of a system of N particles only depend on few microscopic quantities. In the context of RMT, there exist several *flavors* of this concept, but universality will be here referred to the fact that matrices belonging to the same class have the same LSD, independent of the specific law we put on each entry

Result 1.3 (*Universality for LSD of Wigner matrices [170, 169]*)

If $\mathbf{A}$ is an arbitrary Wigner matrix then its empirical spectrum converges to ***the semi-circle distribution****, that is:*

$$\mu_{\mathbf{A}}(\lambda) \underset{N\to\infty}{\to} \mu_{\mathrm{sc}(\sigma)}(\lambda) := \frac{\sqrt{4\sigma^2 - \lambda^2}}{2\pi\sigma^2} \mathbb{I}_{[-2\sigma,2\sigma]}\,, \tag{1.40}$$

where $\mathbb{I}_{[a,b]}$ is the indicator function, which is equal to 1 if $\lambda \in [a,b]$ and is null otherwise.

Furthermore, if the off-diagonal elements admit a fourth moment $\mathbb{E}\,|A_{ij}|^4 < \infty$, then there are no outliers outside the bulk, that is as $N \to \infty$, the top and bottom eigenvalues converge to the edges: $\lambda_1(\mathbf{A}) \to 2\sigma$ and $\lambda_N(\mathbf{A}) \to -2\sigma$.

The universality for Wigner matrices due to Tao and Vu, see Ref. [170, 169] is actually a much stronger result since it concerns not only the macroscopic LSD but also the local statistics not discussed in this thesis. The first step to get a universal result of this type is generally to start by proving the result for the *most simple ensemble* of the given class and then try to extend it to the whole class. For generalized Wigner matrices, we argue that the most simple ensemble is the Gaussian one and there exist a vast variety of techniques, more or less rigorous, to prove the result in this case. In particular one relies on the special properties of invariance in law of

the Gaussian distribution to derive a *variational principle* for the LSD and this will be described in detail in Sec. 1.4.

Another way to get the LSD, is to use the famous *moment method* which consists in first proving the convergence as $N \to \infty$ of the average moments $\mathbb{E}\, m_k[\mu_{\mathbf{A}}]$ to the moments of the semi-circle distribution (the so-called *Catalan numbers*) and then show that there is self-averaging (or concentration) of the moments, that is loosely speaking $m_k[\mu_{\mathbf{A}}] \approx \mathbb{E}\, m_k[\mu_{\mathbf{A}}]$ for large N.

Another famous and important method is the *Stieltjes method*, where one shows that the Stieltjes transform is self-averaging $g_{\mathbf{A}}(z) \approx \mathbb{E}\, g_{\mathbf{A}}(z)$ and $\mathbb{E}\, g_{\mathbf{A}}(z)$ converges to the Stieltjes transform of the semi-circle distribution given by:

$$g_{\mathrm{sc}(\sigma)}(z) := \frac{z - z\sqrt{1 - 4\sigma^2/z^2}}{2\sigma^2}\,. \tag{1.41}$$

Now going *beyond* the Gaussian ensembles requires (much) more work but these two methods (and in particular the Stieltjes method) can be tuned to tackle the generic case, and we refer to Ref. [58] and references therein for more details.

1.3.2 Sample covariance matrices, Laguerre ensembles and the Marčenko-Pastur distribution

The second most important class of random self-adjoint matrices corresponds to the ones first studied by Wishart:

A real (resp. complex, quaternionic) **sample covariance matrix with iid entries** is a matrix of the form $\mathbf{A} = \mathbf{X}\mathbf{X}^*/M$ where $\mathbf{X} = (X_{ij})$ is a real (resp. complex, quaternionic) matrix of size $(N \times M)$ where the entries X_{ij} are iid random variable with zero mean $\mathbb{E}\, X_{ij} = 0$ and variance one $\mathbb{E}\,|X_{ij}|^2 = 1$.

As in the Wigner case, a special role is given when the entries are Gaussian.

> We say that a sample covariance matrix is a Wishart matrix or is taken from **the Laguerre orthogonal ensemble** (LOE) for $\beta = 1$, from **the Laguerre unitary ensemble** (LUE) for $\beta = 2$ and from **the Laguerre symplectic ensemble** (LSE) for $\beta = 4$, if $X_{ij} \sim \mathcal{N}_\beta(0, 1)$.

Similarly to Gaussian ensembles, for $\beta = 1, 2, 4$ one can express the probability to find a matrix taken from one of the three Laguerre ensembles in a region $R \subset \mathrm{Herm}_\beta^+(N)$ as:

$$\mathbb{P}_L^{(\beta)}\left[\mathbf{A} \in R\right] = \frac{1}{Z_{N,\beta}} \int_R \mathrm{e}^{-\frac{M\beta}{2}\mathrm{Tr}\,\mathbf{A}} \left(\det \mathbf{A}\right)^{-1+\frac{\beta}{2}(M-N+1)} \mathrm{d}\mathbf{A}\,. \tag{1.42}$$

The LSD of a sample covariance matrix with iid entries is also universal in the sense:

Result 1.4 (*Universality for LSD of Wishart matrices [123, 152]*)

> *If $\mathbf{A}$ is an arbitrary sample covariance matrix with iid entries taken in a double scaling limit where $N \to \infty$ and $M \equiv M_N \to \infty$ but the ratio stays bounded, $N/M \to q \in (0, \infty)$, then it is another well-known result of RMT that its empirical spectrum converges to the*

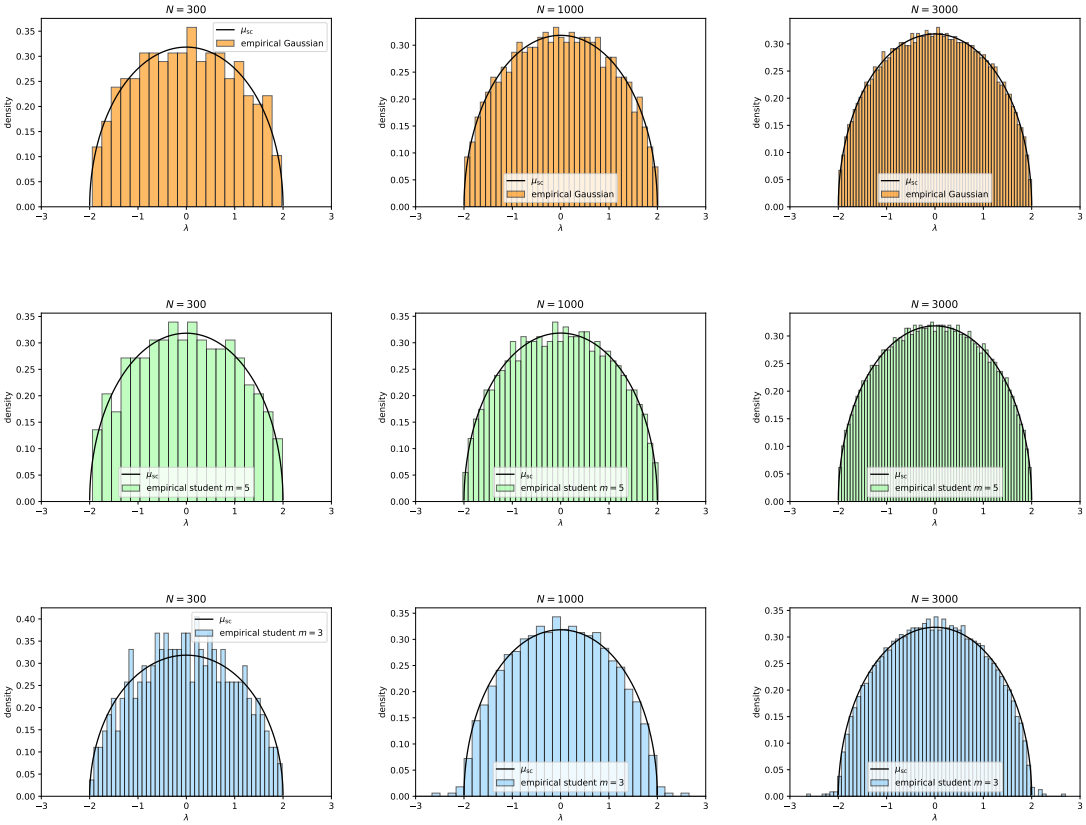


Figure 1.1: Illustration of the self-averaging and universality properties of general Wigner matrices. The top line represents the histogram of eigenvalues of GOE matrices with $\sigma = 1$ and increasing values of N from left to right. The second line represents the eigenvalues of Wigner matrices with Student-t distributed entries with parameter $m = 5$ (that is they admit moments up to order five), normalized to have the same variance as in the Gaussian case. The bottom line represents the Student case with $m = 3$, in this case, the LSD is also a semi-circle distribution but the top and bottom eigenvalues do not convergence to the edges of the distribution.

so-called ***Marčenko-Pastur distribution****:*

$$\mu_{\mathbf{A}}(\lambda) \underset{N\to\infty}{\to} \mu_{\mathrm{MP}(q)}(\lambda) = \frac{\sqrt{(\mathrm{a}_+ - \lambda)(\lambda - \mathrm{a}_-)}}{2\pi q \lambda} \mathbb{I}_{[\mathrm{a}_-,\mathrm{a}_+]} + \left(1 - q^{-1}\right) \delta(x)\, \mathbb{I}_{[q>1]}\,, \quad (1.43)$$

where the edges are given by $\mathrm{a}_\pm = (1 \pm \sqrt{q})^2$.

Furthermore, if the iid entries X_{ij} *admit a fourth moment* $\mathbb{E}\,|X_{ij}|^4 < \infty$*, then there are no outliers outside the bulk, that is the top and bottom eigenvalues converge to the edges:* $\lambda_1(\mathbf{A}) \to (1 + \sqrt{q})^2$ *and* $\lambda_N(\mathbf{A}) \to (1 - \sqrt{q})^2$.

A plot of the Marčenko-Pastur distribution is given in Fig. 1.2. For Wishart matrices, we will be mainly interested in the case $q \in (0, 1)$ where there is no Dirac mass at zero. To conclude this

section on Wishart matrices, we give the Stieltjes transform of the Marčenko-Pastur distribution:

$$g_{\mathrm{MP}(q)}(z) := \frac{z + q - 1 - \sqrt{z - \mathrm{a}_+}\sqrt{z - \mathrm{a}_-}}{2qz} \,. \tag{1.44}$$

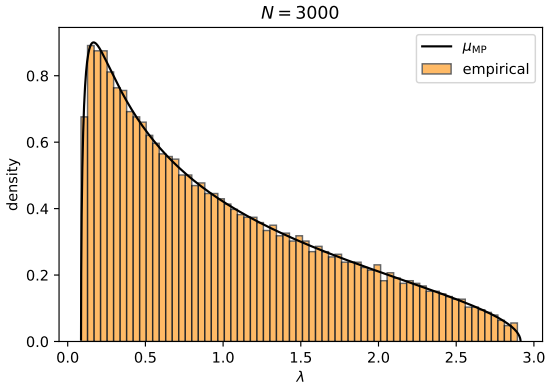


Figure 1.2: The Marčenko-Pastur distribution for $q = 1/2$, compared with the histogram of the eigenvalues of a Wishart matrix with $N = 3000$ and $T = 6000$.

1.3.3 General Manova matrices, Jacobi ensembles, and the Watcher distribution

Wigner and Wishart's matrices play a fundamental role in RMT. The third and last class of self-adjoint matrices we will consider is lesser known and appears in the multivariate analysis of variance (MANOVA) in statistics:

A matrix $\mathbf{A}$ is said to be a **General Manova matrix** if it is of the form:

$$\mathbf{A} = \left(\mathbf{I} + \frac{M_1}{M_2}\sqrt{\mathbf{A}_2}\mathbf{A}_1^{-1}\sqrt{\mathbf{A}_2}\right)^{-1} \tag{1.45}$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are two independent sample covariance matrices with iid entries with parameters (N, M_1) and (N, M_2) respectively and with $M_1 \leq N$ to ensure the matrix $\mathbf{A}_1$ to be invertible.

The simpler element of the class of MANOVA is given by the following definition.

> We say that a MANOVA matrix is taken from **the Jacobi orthogonal ensemble** (JOE) for $\beta = 1$, from **the Jacobi unitary ensemble** (JUE) for $\beta = 2$ and from **the Jacobi symplectic ensemble** (JSE) for $\beta = 4$, if $\mathbf{A}_1$ and $\mathbf{A}_2$ are taken from the corresponding Laguerre ensemble.

In this case, by computing the Jacobian of the change of transformation $(\mathbf{A}_1, \mathbf{A}_2) \to \mathbf{A}$, one can get an explicit expression for the probability of the matrix $\mathbf{A}$. For any region R such that $\mathbf{A}$ and $\mathbf{I} - \mathbf{A}$ are semi-definite positive, we have:

$$\mathbb{P}_{\mathrm{J},\beta}\left[\mathbf{A} \in R\right] = \frac{1}{Z_{\mathrm{J},\beta}} \int_R \left(\det \mathbf{A}\right)^{-1+\frac{\beta}{2}(M_1-N+1)} \left(\det\left[\mathbf{I} - \mathbf{A}\right]\right)^{-1+\frac{\beta}{2}(M_2-N+1)} \mathrm{d}\mathbf{A} \,, \tag{1.46}$$

Similarly to Wigner and Wishart matrices, one should expect to have a universal limit for the spectrum of General Manova matrices and the following result goes in this direction.

Result 1.5 (*Universality for MANOVA matrices [59]*)

Under mild assumptions on the moments of the entries of the Wishart matrices $\mathbf{A}_1$ and $\mathbf{A}_2$, the ESD of a general MANOVA matrix converges as $N, M_1, M_2 \to \infty$ with $N/M_1 \to q_1$ and $N/M_2 \to q_2$, to ***the Watcher distribution****:*

$$\mu_{\mathbf{A}}(\lambda) \underset{N\to\infty}{\to} \mu_{\mathrm{Wat}(q_1,q_2)}(\lambda) = \frac{1}{2\pi(q_1^{-1}+q_2^{-1})} \frac{\sqrt{\mathrm{a}_+ - \lambda}\sqrt{\lambda - \mathrm{a}_-}}{\lambda(1-\lambda)} \mathbb{I}_{[\mathrm{a}_-,\mathrm{a}_+]} + \left(1 - q_2^{-1}\right) \delta(x-1)\, \mathbb{I}_{[q_2>1]} \tag{1.47}$$

where the edges are given by $\mathrm{a}_\pm = \frac{q_2(q_2+q_1(1+q_1-q_2\pm 2\sqrt{q_1+q_2-q_1q_2}))}{(q_1+q_2)^2}$.

One may note that for $q_2 \in (0,1)$ there is no Dirac mass in the Watcher distribution.

Remark. The specific assumptions on the entries are given in Ref. [59] and are sub-optimal. As discussed in Ref. [59] one should expect to have the same type of universal result as in the Wigner and Wishart cases. ⌟

Remark *(Watcher distribution and arcsine law)***.** In the limiting case where $q_1, q_2 \to 1$, the Watcher distribution becomes the *arcsine law*:

$$\mu_{\mathrm{Wat}(1,1)}(\lambda) \equiv \mu_{\mathrm{as}}(\lambda) := \frac{1}{\pi} \frac{1}{\sqrt{\lambda(1-\lambda)}} \mathbb{I}_{[0,1]}\,. \tag{1.48}$$

which we will encounter in several instances in this thesis. ⌟

To conclude, we give the Stieltjes transform of the Watcher distribution:

$$g_{\mathrm{Wat}(q_1,q_2)}(z) := \frac{q_1^{-1} - 1 - (q_1^{-1}+q_2^{-1})z + (q_1^{-1}+q_2^{-1})\sqrt{z-\mathrm{a}_+}\sqrt{z-\mathrm{a}_-}}{2z(1-z)}\,. \tag{1.49}$$

1.3.4 Gaussian rectangular random matrices, Ginibre matrices and associated limiting singular value distribution

This section deals with rectangular matrices.

For rectangular matrices, the most important class, and the only one mentioned in this thesis consists of the matrix with iid entries.

In particular, if $\mathbf{A} = (A_{ij}) \in \mathsf{M}_{N,M}(\mathbb{K}_\beta)$ with $A_{ij} \overset{\text{iid}}{\sim} \mathcal{N}_\beta\left(0, \sigma/\sqrt{M}\right)$, we say that $\mathbf{A}$ is taken from the **Gaussian rectangular ensemble**. For $M = N$, $\mathbf{A}$ is said to be a **Ginibre matrix**.

The law of such matrices is given with respect to the Lebesgue measure $\mathrm{d}\mathbf{A} = \prod_{i,j,b} \mathrm{d}A_{ij}^{(b)}$ by:

$$\mathbb{P}_{q,G}^{(\beta)}[\mathbf{A} \in R] = \frac{1}{Z_{N,\beta}} \int_R \mathrm{e}^{-\frac{M\beta}{2\sigma^2}\mathrm{Tr}\,\mathbf{A}\mathbf{A}^*} \mathrm{d}\mathbf{A}\,. \tag{1.50}$$

Since the LSVD of the matrix $\mathbf{A}$ is related to the spectrum of Marčenko-Pastur distribution by Eq. (1.19) one gets:

$$\mu_A(s) = \frac{\sqrt{4q\sigma^4 - (s^2 - \sigma^2(1+q))^2}}{\pi\sigma^2 qs}\mathbb{I}_{[\sigma(1-\sqrt{q}),\sigma(1+\sqrt{q})]}\,. \tag{1.51}$$

In particular, for $q = 1$ corresponding to Ginibre random matrices, the LSVD becomes the so-called *quarter-circle distribution:*

$$\mu_A(s) = \frac{\sqrt{4\sigma^2 - s^2}}{\pi\sigma^2}\mathbb{I}_{[0,2\sigma]}\,. \tag{1.52}$$

1.4 β-invariant random matrices, β-ensembles and tridiagonal matrices

1.4.1 β-invariant random matrices and Haar measure

We describe the invariance in law mentioned earlier satisfied by the Gaussian, Laguerre, and Jacobi ensembles.

A matrix is orthogonally (resp. unitary, symplectic)-invariant or in short ***β*-invariant** if its law is left invariant by conjugation over $\mathsf{O}_\beta(N)$:

$$\mathbf{A} \overset{\text{in law}}{=} \mathbf{V}'\mathbf{A}\mathbf{V}'^* \quad \text{for any} \quad \mathbf{V}' \in \mathsf{O}_\beta(N)\,. \tag{1.53}$$

One can think of this definition as the matrix counterpart of the rotationally invariance in law satisfied by a Gaussian vector: if $\boldsymbol{x} \overset{\text{iid}}{\sim} \mathcal{N}_\beta(0,1)$ then for any $\mathbf{V} \in \mathsf{O}_\beta(N)$ we have $\mathbf{V}\boldsymbol{x} \overset{\text{in law}}{=} \boldsymbol{x}$, and for Gaussian, Laguerre and Jacobi ensembles Eq. (1.53) is actually a consequence of this property for Gaussian vectors.

Thus for β-invariant random matrix, we can always rotate it by a matrix $\mathbf{V}' \in \mathsf{O}_\beta(N)$ without changing its law. This implies that if $\mathbf{V}$ is the matrix of the eigenvector of a β-invariant random matrix, then $\mathbf{V}$ is distributed uniformly at random over $\mathsf{O}_\beta(N)$.

1.4.2 β-ensembles and Coulomb gas at inverse temperature β

For each of these three ensembles, one has a rather simple expression for the joint density of the elements. In fact, using the identity $\det\cdot = \exp[\operatorname{Tr}\log\cdot]$, we can always write the density in the form of:

$$\mathcal{P}(\mathbf{A}) \propto \exp\left[-\frac{N\beta}{2}\operatorname{Tr} V(\mathbf{A})\right]\,, \tag{1.54}$$

for a function V proportional to the square function for the Gaussian ensemble, to a linear combination of linear and logarithmic functions for the Laguerre ensemble, and to a linear combination of logarithmic functions for the Jacobi ensemble. We recall that for an analytical smooth function V, $V(\mathbf{A})$ is a matrix in the same basis $\mathbf{V}$ of $\mathbf{A}$ but the eigenvalues $\boldsymbol{\lambda}$ are modified according to the potential: $V(\mathbf{A}) := \mathbf{V}\mathbf{Diag}\left(V(\boldsymbol{\lambda})\right)\mathbf{V}^*$, with $V(\boldsymbol{\lambda}) = (V(\lambda_1), \ldots, V(\lambda_N))$.

Now density in the form of Eq. (1.54) makes perfect sense for *any* smooth function V, provided Eq. (1.54) is integrable. This leads to the following definition.

If $V(.)$ is some given smooth function with sufficient growth condition at infinity, then we say that a random matrix is taken from a **β-ensemble with potential** V - or is taken from a β-ensemble in short - if its law is given by

$$\mathbb{P}_V^{(\beta)}\left[\mathbf{A} \in R\right] = \frac{1}{Z_{V,\beta}} \int_R \mathrm{e}^{-\frac{N\beta}{2}\operatorname{Tr} V(\mathbf{A})} \mathrm{d}\mathbf{A}\,, \tag{1.55}$$

For positive semi-definite matrices, we implicitly assume the potential to be defined on $\mathbb{R}_+$ and is infinite on the negative real line.

The Gaussian, Laguerre and Jacobi ensembles are known as the *classical ensembles* since these potentials correspond to ones of the three classical laws of probability, namely the Gaussian law, the gamma law, and the beta law. We give explicitly their potential in the following examples.

Examples *(Potential for the Classical ensembles)***.** Explicitly, we have

- for GOE/GUE/GSE matrices the corresponding potential is given by:

$$V_G(\lambda) = \frac{\lambda^2}{2\sigma^2}\,. \tag{1.56}$$

- for LOE/LUE/LSE matrices the corresponding potential is given for $\lambda \in (0, \infty)$ by:

$$V_L(\lambda) = \frac{M}{N}\lambda + \left(1 - \frac{M}{N} - \frac{1}{N} + \frac{2}{\beta N}\right)\log\lambda\,, \tag{1.57}$$

  and is infinite otherwise. Its limiting behavior (in the limit $N/M \to q$) is given by:

$$\lim_{N,M\to\infty} V_L(\lambda) = \frac{\lambda}{q} + \left(1 - \frac{1}{q}\right)\log\lambda\,. \tag{1.58}$$

- for JOE/JUE/JSE matrices, the corresponding potential is given for $\lambda \in (0, 1)$ by:

$$V_J(\lambda) = \left(1 - \frac{M_1}{N} - \frac{1}{N} + \frac{2}{\beta N}\right)\log\lambda + \left(1 - \frac{M_2}{N} - \frac{1}{N} + \frac{2}{\beta N}\right)\log(1-\lambda)\,, \tag{1.59}$$

  and is infinite otherwise. Its limiting behavior (in the limit $N/M_1 \to q_1$ and $N/M_2 \to q_2$) is given by:

$$\lim_{N,M_1,M_2\to\infty} V_J(\lambda) = (1 - q_1^{-1})\log\lambda + (1 - q_2^{-1})\log(1-\lambda)\,. \tag{1.60}$$

■

By the cyclical property of the trace ($\operatorname{Tr}\mathbf{ABC} = \operatorname{Tr}\mathbf{CAB}$) it is clear that if $\mathbf{A}$ is taken from a β-ensemble, then it is also β-invariant. Note that the converse is not true in general.

In this case, one can hope to get the joint law of its eigenvalues $\boldsymbol{\lambda}$ by integrating out the dependency in the eigenvectors $\mathbf{V}$ after the change of variable $\mathbf{A} \to (\mathbf{V}, \boldsymbol{\lambda})$. Since the eigenvalue decomposition is volume preserving with respect to the Lebesgue measure $\mathrm{d}\mathbf{A}$, we have:

$$\mathrm{e}^{-\frac{N\beta}{2}\operatorname{Tr} V(\mathbf{A})}\mathrm{d}\mathbf{A} \propto \mathrm{e}^{-\frac{N\beta}{2}\sum_{i=1}^{N} V(\lambda_i)} \left|\left[\frac{\partial \mathbf{A}}{\partial \mathbf{V}}, \frac{\partial \mathbf{A}}{\partial \boldsymbol{\lambda}}\right]\right| \mu_{\mathrm{Haar}}(\mathrm{d}\mathbf{V})\mathrm{d}\boldsymbol{\lambda}\,. \tag{1.61}$$

where the Jacobian $\left|\left[\frac{\partial \mathbf{A}}{\partial \mathbf{V}}, \frac{\partial \mathbf{A}}{\partial \boldsymbol{\lambda}}\right]\right|$ is given by Weyl's integral formula of Eq. (1.10). Integrating out the law of the eigenvectors in Eq. (1.61) with the expression of Eq. (1.10) for the Jacobian gives the following expression for the joint density of eigenvalues:

$$\mathcal{P}_V^{(\beta)}(\boldsymbol{\lambda})\mathrm{d}\boldsymbol{\lambda} \propto \mathrm{e}^{-\frac{N\beta}{2}\|V(\boldsymbol{\lambda})\|_1} |\Delta(\boldsymbol{\lambda})|^{\beta}\, \mathrm{d}\boldsymbol{\lambda}\,, \tag{1.62}$$

with $\|V(\boldsymbol{\lambda})\|_1 = \sum_{i=1}^{N} V(\lambda_i)$.

The simple yet fundamental remark is to notice that while Eq. (1.62) is the joint law of eigenvalues of a matrix taken from a β-ensemble, with $\beta \in \{1, 2, 4\}$, it can be naturally defined for *any* $\beta > 0$.

For any $\beta > 0$, we say that a vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$ **is taken from a β-ensemble with potential** V and denote $\boldsymbol{\lambda} \sim \mathcal{P}_V^{(\beta)}$ if its law is given by:

$$\boxed{\mathcal{P}_V^{(\beta)}(\boldsymbol{\lambda}) = \frac{1}{Z_{V,\beta}} \exp\left[-\frac{N\beta}{2}\left(\sum_{i=1}^{N} V(\lambda_i) - \frac{1}{N}\sum_{i,j|i\neq j} \log|\lambda_i - \lambda_j|\right)\right]}\,, \tag{1.63}$$

since the term $\sum_{i,j|i\neq j} \log|\lambda_i - \lambda_j| = 2\sum_{i,j|i<j} \log|\lambda_i - \lambda_j|$ is up to a factor 2 nothing else than the logarithm of the Vandermonde determinant.

By abuse of notation, we will sometimes refer to the vector $\boldsymbol{\lambda}$ as the vector of eigenvalues even if $\beta \notin \{1, 2, 4\}$. Similarly, we will denote by $\mu_{\mathbf{A}}(\lambda) = \sum_{i=1}^{N} \delta(\lambda - \lambda_i)/N$ its empirical 'spectral' distribution.

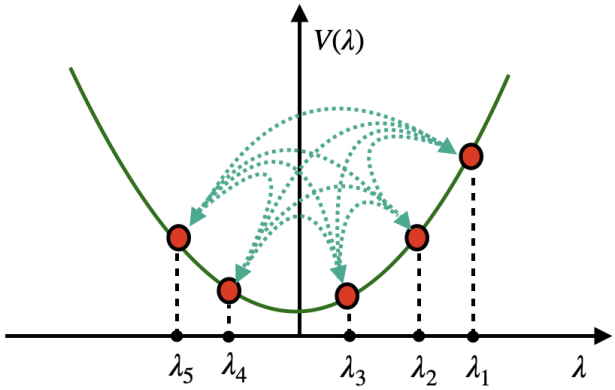


Figure 1.3: A sketch of the Coulomb gas interpretation. The eigenvalues λ_i's can be seen as the position of N particles ($N = 5$ in this figure) in a potential V interacting with a pairwise logarithmic repulsive interaction.

It is then convenient to think of the parameter β as an **inverse temperature** of a Gibbs-Boltzmann distribution, that is re-write the joint law as:

$$\mathcal{P}_V^{(\beta)}(\boldsymbol{\lambda}) = \frac{\mathrm{e}^{-\beta\,\tilde{\mathcal{E}}_N(\boldsymbol{\lambda})}}{Z_{V,\beta}} \text{ with } \tilde{\mathcal{E}}_N(\boldsymbol{\lambda}) := \sum_{i=1}^{N} \frac{N}{2} V(\lambda_i) - \frac{1}{2} \sum_{i,j|i\neq j} \log|\lambda_i - \lambda_j|\,, \tag{1.64}$$

where the **electrostatic energy** $\tilde{\mathcal{E}}_N(.)$ consists of a **confining potential** V and **a pairwise logarithmic repulsive interaction** between the λ_i, preventing them to all collapse at the minima of the potential V in the large N limit, see Fig. 1.3 for an illustration. As it is standard in physics, the energy $\tilde{\mathcal{E}}_N(.)$ can be defined up to an additive constant, since this constant can be absorbed in the normalization $Z_{V,\beta}$, without changing the law. This logarithmic pairwise interaction is exactly the one of a two-dimensional Coulomb gas of N charged particles. The components of the vector $\boldsymbol{\lambda} \sim \mathcal{P}_V^{(\beta)}$ taken from a β-ensemble can be thought of as the positions of N charged particles of a $2D$-Coulomb gas, restricted to be on the real line. One may not that for λ_i of order one if the potential V does not scale with N, the term $N(\sum V(\lambda_i))$ is order $\mathcal{O}(N^2)$ and so does the logarithmic repulsion and the energy $\tilde{\mathcal{E}}_N(.)$. For this reason, it is convenient to introduce a re-scaling in the following way:

$$\mathcal{P}_V^{(\beta)}(\boldsymbol{\lambda}) = \frac{\mathrm{e}^{-\frac{N^2\beta}{2}\,\mathcal{E}_N(\boldsymbol{\lambda})}}{Z_{V,\beta}} \quad \text{with} \quad \mathcal{E}_N(\lambda) := \frac{2}{N^2}\tilde{\mathcal{E}}_N(\boldsymbol{\lambda})\,. \tag{1.65}$$

The (normalized) **energy** $\mathcal{E}_N$ is now of order one for any N and so it is particularly well-suited in the large N limit. It can be explicitly written in terms of a *functional* of the discrete measure $\mu_{\mathbf{A}}(\lambda) = \sum_{i=1}^{N} \delta(\lambda - \lambda_i)/N$ as:

$$\mathcal{E}_N(\boldsymbol{\lambda}) = \int V(\lambda)\mu_{\mathbf{A}}(\mathrm{d}\lambda) + \int\int_{x\neq y} \log\frac{1}{|\lambda-\lambda'|}\mu_{\mathbf{A}}(\mathrm{d}\lambda)\mu_{\mathbf{A}}(\mathrm{d}\lambda') =: \tilde{\mathcal{E}}_N\left[\mu_{\mathbf{A}}\right]. \tag{1.66}$$

As we increase $N \to \infty$ if the argument $\mu_{\mathbf{A}} \to \mu$ of the energy is an absolutely continuous distribution μ, the logarithmic singularity coming from the Coulomb gas interaction is integrable, and we can remove the constraint over the diagonal to define the asymptotic energy functional as:

$$\mathcal{E}[\mu] := \lim_{N\to\infty} \mathcal{E}_N[\mu_{\mathbf{A}}] = \int V(\lambda)\mu(\lambda)\mathrm{d}\lambda + \int\int \log\frac{1}{|\lambda-\lambda'|}\mu(\lambda)\mu(\lambda')\mathrm{d}\lambda\mathrm{d}\lambda'\,. \tag{1.67}$$

If $V(\lambda) \equiv V_N(\lambda)$ depends explicitly on the dimension N (as for the Laguerre or Jacobi ensembles for example), we implicitly mean $V(\lambda) = \lim_{N\to\infty} V_N(\lambda)$ in Eq. (1.67). The repulsive term

$$\Sigma[\mu_A] := \int\int \log|\lambda-\lambda'|\mu(\lambda)\mu(\lambda')\mathrm{d}\lambda\mathrm{d}\lambda'\,, \tag{1.68}$$

is sometimes referred to as the *free entropy*, as it has a somehow analog interpretation as the classical entropy for RMT/free probability.

Now to understand the behavior of the limiting density of a β-ensemble, we will use the powerfull (but non-rigorous) path integral representation by doing the change of variable from the vector $\boldsymbol{\lambda}$ to its empirical distribution $\mu_{\mathbf{A}}$. To do so, we need to compute the Jacobian of this transformation, or equivalently, using the language of statistical physics one needs to

count how many micro-states $\boldsymbol{\lambda}$ are compatible with the density $\mu_{\mathbf{A}}$. The reader familiar with statistical physics will immediately recognize the definition of the entropy of the micro-state and the following result (which has to be understood in the large N limit) should come with no surprise:

$$\mathrm{Jac}[\boldsymbol{\lambda} \to \mu] \approx \mathrm{e}^{-N\mathcal{S}[\mu]} \,, \tag{1.69}$$

where $\mathcal{S}$ is **the entropy** of the distribution, which is defined by

$$\mathcal{S}[\mu] := \int \mu(\lambda) \log\left(\mu(\lambda)\right) \mathrm{d}\lambda \,. \tag{1.70}$$

We are now ready to write down the density to find a given distribution μ as

$$\mathcal{P}_V^{(\beta)}[\mu]\, \mathcal{D}\mu \approx \frac{1}{Z_{V,\beta}} \exp\left[-\frac{N^2\beta}{2}\mathcal{E}[\mu] - \frac{N}{c}\mathcal{S}[\mu]\right] \mathcal{D}\mu \,, \tag{1.71}$$

where $\mathcal{D}\mu$ denotes the ill-defined infinitesimal uniform increment over the space $\mathcal{M}_1$ of probability measure. As $N \to \infty$, the entropy term becomes subdominant, and the probability to find a distribution is dictated by *a large deviation principle*.

Informally, a random variable X_N is dictated by **a large deviation principle** with **speed of convergence** $v(N)$ and **rate function** $\Psi(.) \geq 0$ if for large N the probability to find X_N in a region R, decays exponentially as:

$$\mathrm{Prob}\left[X_N \in R\right] \asymp \exp\left[-v(N) \inf_{x \in R} \Psi(x)\right] \,, \tag{1.72}$$

where $\asymp$ denotes equality at leading exponential order (in N). Importantly, the rate function also encodes the *typical events* since if this rate function is null, the corresponding probability is of order one.

In the context of empirical distributions of β-ensembles, this means that we have the following result which can be proved (and stated) in a rigorous manner, see Ref. [86]:

Result 1.6 (*LDP for the distribution of a β-ensemble*)

In the large N limit, the empirical distribution of β-ensemble satisfies a large deviation principle with speed $\frac{N^2\beta}{2}$ and rate function given by:

$$\mathcal{E}[\mu] - \mathcal{E}\left[\mu_A\right] \qquad \textit{where} \quad \mu_A := \operatorname*{argmin}_{\mu' \in \mathcal{M}_1} \mathcal{E}[\mu'] \,. \tag{1.73}$$

In particular, this implies that if $\boldsymbol{\lambda} \sim \mathcal{P}_V^{(\beta)}$ then its LSD μ_A is independent of β and given by the minimizer of the energy functional.

Let us now characterize the limiting distribution μ_A more precisely. We can write down the variational principle for μ_A as

$$\left(\frac{\delta}{\delta\mu}\mathcal{E}[\mu]\right)\Big|_{\mu_A} = 0 \,, \tag{1.74}$$

that is using Eq. (1.67) for the expression of the electrostatic energy, μ_A is the solution of the *Tricomi's problem*:

$$⨍ \frac{\mu_A(\lambda')}{\lambda - \lambda'} \mathrm{d}\lambda' = \frac{V'(\lambda)}{2} \quad \text{for } \lambda \in \mathrm{Supp}\,[\mu_A]\ , \tag{1.75}$$

where $⨍$ is the Cauchy principal value of the integral.

The case of convex potential and Tricomi's formula -

Whenever V is convex and 'confining enough'[1], the distribution μ_A is supported on an interval of the form $[\mathrm{a}_-, \mathrm{a}_+]$ and the Tricomi problem can be directly inverted thanks to the **Tricomi formula** [174]:

$$\boxed{\mu_A(\lambda) = \frac{1}{\pi\sqrt{\lambda - \mathrm{a}_-}\sqrt{\mathrm{a}_+ - \lambda}} \left(1 - \frac{1}{\pi} ⨍_{\mathrm{a}_-}^{\mathrm{a}_+} \frac{\sqrt{\lambda' - \mathrm{a}_-}\sqrt{\mathrm{a}_+ - \lambda'}}{\lambda - \lambda'} \frac{V'(\lambda)}{2\pi} \mathrm{d}\lambda'\right)}\ . \tag{1.76}$$

The edges $\mathrm{a}_\pm$ can be either determined self-consistently thanks to the constraints $\mu_A(\mathrm{a}_-) = \mu_A(\mathrm{a}_+) = 0$ or equivalently by solving the system of equations:

$$\int_{\mathrm{a}_-}^{\mathrm{a}_+} \frac{V'(\lambda)}{\sqrt{\lambda - \mathrm{a}_-}\sqrt{\mathrm{a}_+ - \lambda}} \mathrm{d}\lambda = 0 \quad \text{and} \quad \int_{\mathrm{a}_-}^{\mathrm{a}_+} \frac{V'(\lambda)\lambda}{\sqrt{\lambda - \mathrm{a}_-}\sqrt{\mathrm{a}_+ - \lambda}} \mathrm{d}\lambda = 2\pi\,. \tag{1.77}$$

After a few algebraic operations, one can transform Tricomi's formula into *Pastur's form*, see for example Ref. [62]:

$$\mu_A(\lambda) = \frac{h(\lambda)}{\pi} \sqrt{(\lambda - \mathrm{a}_-)(\mathrm{a}_+ - \lambda)}\,. \tag{1.78}$$

where the function h is given by

$$h(\lambda) := \frac{1}{2\pi} \int_{\mathrm{a}_-}^{\mathrm{a}_+} \frac{V'(\lambda) - V'(\lambda')}{\lambda - \lambda'} \frac{\mathrm{d}\lambda'}{\sqrt{(\mathrm{a}_+ - \lambda')(\lambda' - \mathrm{a}_-)}}\,. \tag{1.79}$$

> The function h is regular near the edges and as a consequence for a general convex potential V, the density $\mu_A(.)$ is **non-critical**, by which we mean that it behaves as a square root near the edge a_+,
>
> $$\mu_A(x) \underset{x \nearrow \mathrm{a}_+}{\sim} \frac{\gamma_0^{3/2}}{\pi} \sqrt{\mathrm{a}_+ - x}\,, \tag{1.80}$$
>
> where γ_0 is a constant and a similar result holds for the behavior near the bottom edge.

One can compute the function h and the edges $\mathrm{a}_\pm$ in the cases where V is given by Eq. (1.56), Eq. (1.58) with $q \in (0,1)$ and Eq. (1.60) with $q_1, q_2 \in (0,1)$, and the results for the corresponding density will be given respectively by the semi-circle distribution of Eq. (1.40), the continuous part of the Marčenko-Pastur distribution of Eq. (1.43), and the continuous part of the Watcher distribution of Eq. (1.47), as expected.

[1] that is such that $V(x)/\log(1+x^2) \underset{|x|\to\infty}{\to} \infty$.

Non-convex potential and BIPZ formula -

Now if the potential V is analytical and confining but not convex, one cannot directly invert the Tricomi problem. However, if we add and subtract $V'(z)$ in the RHS of Eq. (1.75) and then multiply the whole equation by $\mu_A(\lambda)(z-\lambda)^{-1}$ and then finally integrate over λ, we get the following equation:

$$2 \fint \int \frac{\mu_A(\lambda')\mu_A(\lambda)}{(z-\lambda)(\lambda-\lambda')} \mathrm{d}\lambda' \mathrm{d}\lambda = -\int \frac{V'(z)-V'(\lambda)}{(z-\lambda)} \mu_A(\lambda) \mathrm{d}\lambda + V'(z) \left(\int \frac{\mu_A(\lambda)}{z-\lambda} \mathrm{d}\lambda \right) . \tag{1.81}$$

The LHS of Eq. (1.81) is the square of the Stieltjes transform in disguise, and we get the following simple algebraic equation for the Stieltjes transform g_A of the LSD μ_A:

$$\boxed{g_A^2 - V'(z) g_A(z) + \Pi_V(z) = 0}, \tag{1.82}$$

where the function Π_V is given by:

$$\Pi_V(z) := \int_{\mathrm{Supp}[\mu_A]} \frac{V'(z)-V'(\lambda)}{z-\lambda} \mu_A(\lambda) \mathrm{d}\lambda \, . \tag{1.83}$$

Now the key point is to notice that if V' is a polynomial with the highest degree equal to k, the a priori unknown function Π_V is also a polynomial but with a lower degree $k-1$ and its leading coefficient is the same as the one of V'. This means that one can determine self-consistently the coefficients of Π_V a posteriori. One can actually generalize the situation to the case where V' is a Laurent polynomial or a rational function. In the end, the Stieltjes transform g_A is given by the **Brezin-Itzykson-Parisi-Zuber** (BIPZ) **formula**:

$$g_A(z) = \frac{V'(z)}{2} \pm \frac{\sqrt{V'^2(z) - 4\Pi_V(z)}}{2} \, , \tag{1.84}$$

where one must choose the sign of the Stieltjes transform in accordance with the limiting behavior of the Stieltjes transform near infinity given by Eq.(1.24).

> Interestingly, one can naturally look at the other non-physical solution of Eq. (1.82) given by:
>
> $$\bar{g}_A(z) = \frac{V'(z)}{2} \mp \frac{\sqrt{V'^2(z) - 4\Pi_V(z)}}{2} \, , \tag{1.85}$$
>
> which is known as the **second branch of the Stieltjes transform**.

Summing Eq. (1.84) and Eq. (1.85) we have the following simple relation between the two branches of the Stieltjes transform and the derivative of the potentials:

$$\bar{g}_A(z) + g_A(z) = V'(z) \, . \tag{1.86}$$

The second branch of the Stieltjes transform will appear quite naturally in the study of the large deviation of the top eigenvalue.

1.4.3 β-bi-invariant rectangular random matrices and associated β-ensembles

This section deals with rectangular matrices.

We briefly summarize the analogous formulation in the context of rectangular matrices.

Similarly to the self-adjoint cases, we say that a rectangular matrix $\mathbf{A} \in \mathsf{M}_{N,M}(\mathbb{K}_\beta)$ is **β-bi-invariant** if its law is left unchanged by conjugation over $\mathsf{O}_\beta(N) \times \mathsf{O}_\beta(M)$, that is:

$$\mathbf{A} \overset{\text{in law}}{=} \mathbf{V}_1' \mathbf{A} \mathbf{V}_2' \quad \text{for any } \mathbf{V}_1' \in \mathsf{O}_\beta(N) \text{ and } \mathbf{V}_2' \in \mathsf{O}_\beta(M)\,. \tag{1.87}$$

A β-bi-invariant random matrix admits a singular value decomposition $\mathbf{A} = \mathbf{V}_1 \mathbf{Diag}_q(\boldsymbol{s}) \mathbf{V}_2$ where the left and right eigenmatrix are uniformly Haar distributed, $\mathbf{V}_1 \sim \mathsf{Unif}[\mathsf{O}_\beta(N)]$ and $\mathbf{V}_2 \sim \mathsf{Unif}[\mathsf{O}_\beta(M)]$.

We say that a rectangular matrix is taken from **β-bi-invariant ensemble**, if its law is given as:

$$\mathbb{P}^{(\beta)}_{q,V}[\mathbf{A} \in R] = \frac{1}{Z_{V,\beta}} \int_R \mathrm{e}^{-\frac{N\beta}{2} \operatorname{Tr} V(\mathbf{A}\mathbf{A}^*)} \mathrm{d}\mathbf{A}\,, \tag{1.88}$$

for an analytic potential $V(.)$ such that the density is integrable.

If $\mathbf{A} \sim \mathbb{P}^{(\beta)}_{q,V}$ then it is also β-bi-invariant, and one can get the joint law of its singular values thanks to the Jacobian of Eq. (1.16). If we introduce the modified potential

$$\tilde{V}_q(x) := V(x) + \left(1 - \frac{1}{q} - 1 + \frac{1}{\beta}\right) \log x\,, \tag{1.89}$$

with $q = N/M$, the joint density for the singular values can be compactly written as:

$$\mathcal{P}^{(\beta)}_{q,\tilde{V}_q}(\boldsymbol{s}) = \frac{1}{Z_N} \exp\left[-\frac{N\beta}{2}\left(\sum_{i=1}^N \tilde{V}_q(s_i^2) - \frac{1}{N} \sum_{i,j|i\neq j} \log|s_i^2 - s_j^2|\right)\right]\,. \tag{1.90}$$

and this density makes sense for any $\beta > 0$.

Note that if we do the change of variable $(\boldsymbol{s}(\mathbf{A}) \to \boldsymbol{\lambda}(\mathbf{A}\mathbf{A}^*))$ given by Eq. (1.15) in the joint law of Eq. (1.90), we have that the matrix $\mathbf{A}\mathbf{A}^*$ is taken from an invariant ensemble with the modified potential $\tilde{V}_q(.)$ (plus a vanishing term coming from the change of variable). In the large double scaling limit of Eq. (1.5), the empirical singular value distribution converges to a smooth limit:

$$\mu_{\mathbf{A}}(s) := \frac{1}{N} \sum_{i=1}^N \delta(s - s_i(\mathbf{A})) \to \mu_A(s) = 2s\, \mu_{AA^*}(s^2)\,, \tag{1.91}$$

where μ_{AA^*} is solution of the Tricomi problem of Eq. (1.75) with the potential $V(.)$ replaced by $\tilde{V}_q(.)$ of Eq. (1.89):

$$⨍ \frac{\mu_{AA^*}(\lambda')}{\lambda - \lambda'} \mathrm{d}\lambda' = \frac{\tilde{V}_q'(\lambda)}{2}\,. \tag{1.92}$$

The edges $\mathrm{a}_\pm$ of the LSVD μ_A are the square root of the edges of the distribution μ_{AA^*}. Equivalently, the LSVD μ_A is the solution of the following equation:

$$⨍ \frac{\mu_A(s')}{s^2 - s'^2} \mathrm{d}s' = \frac{\tilde{V}_q'(s^2)}{2}\,, \tag{1.93}$$

which can be directly seen from the joint law of Eq (1.90).

1.4.4 Tridiagonal random matrix model for β-ensembles

For $\beta = 1, 2, 4$, a vector $\boldsymbol{\lambda} \sim \mathcal{P}_V^{(\beta)}$ taken from a β-ensemble can be seen as the set of eigenvalues of a matrix $\mathbf{A}$ taken from the corresponding ensemble. A natural question is to give a 'random matrix' interpretation for other values of β. The idea developed by Dumitriu and Edelman in Ref. [54] is the following: for $\beta = 1, 2, 4$, any self-adjoint matrix $\mathbf{A} \in \mathrm{Herm}_\beta(N)$ can be put in tridiagonal form: $\mathbf{A} = \mathbf{H}\,\mathbf{T}\,\mathbf{H}^*$ with $\mathbf{H} \in \mathsf{O}_\beta(N)$ and

$$\mathbf{T} \equiv \mathbf{T}(\boldsymbol{a}, \boldsymbol{b}) := \begin{bmatrix} a_1 & b_1 & & \\ b_1 & \ddots & \ddots & \\ & \ddots & \ddots & b_{N-1} \\ & & b_{N-1} & a_N \end{bmatrix}, \tag{1.94}$$

with $\boldsymbol{a} := (a_1, \dots, a_N) \in \mathbb{R}^N$ and $\boldsymbol{b} := (b_1, \dots, b_{N-1}) \in \left(\mathbb{R}_+^*\right)^{N-1}$. The tridiagonal matrix $\mathbf{T}$ has the same set of eigenvalues as the matrix $\mathbf{A}$, but unlike $\mathbf{A}$ which has entries in $\mathbb{K}_\beta$, its entries are real for any $\beta \in \{1, 2, 4\}$. Thus, if we take $\mathbf{A} \sim \mathbb{P}_V^{(\beta)}$ for $\beta = 1, 2, 4$, one should expect to leverage the dependency in β as being simply a parameter of the joint distribution of the *real* random vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. Then, the hope is that once this analytical extension to any $\beta > 0$ is done, the eigenvalues of $\mathbf{T}$ are distributed according to $\mathcal{P}_V^{(\beta)}$. As a consequence, one needs first to compute the joint distribution of $(\boldsymbol{a}, \boldsymbol{b})$ associated with the transformation $\mathbf{A} \to (\mathbf{H}, \mathbf{T}(\boldsymbol{a}, \boldsymbol{b}))$, extend the law to any $\beta > 0$ and then compute the joint distribution of the vector $\boldsymbol{\lambda}$ associated to the change of variable $\mathbf{T}(\boldsymbol{a}, \boldsymbol{b}) \to (\mathbf{Q}, \boldsymbol{\lambda})$ where $\mathbf{Q} \in \mathsf{O}(N)$ and $\boldsymbol{\lambda} \in \mathbb{R}^N$ are the eigenmatrix and eigenvalues of $\mathbf{T} = \mathbf{Q}\,\mathbf{Diag}(\boldsymbol{\lambda})\,\mathbf{Q}^\mathsf{T}$.

The change of variable from a self-adjoint matrix $\mathbf{A}$ to its tridiagonal form $\mathbf{T}$ is done by *Householder reflections*, and one can obtain the joint law of the elements $(\boldsymbol{a}, \boldsymbol{b})$ by this procedure. The final result reads:

$$\mathcal{P}_{\mathsf{tri},V}^{(\beta)}(\boldsymbol{a}, \boldsymbol{b}) := \frac{\mathrm{e}^{-\frac{N\beta}{2} \operatorname{Tr} V(\mathbf{T}(\boldsymbol{a},\boldsymbol{b}))}}{Z_{V,\beta}^{\mathsf{tri}}} \prod_{i=1}^{N} b_i^{\beta(N-i)-1}\,. \tag{1.95}$$

> Since this density makes sense for any $\beta > 0$, we say that $\mathbf{T} \sim \mathcal{P}_{\mathsf{tri},V}^{(\beta)}$ is taken from a **tridiagonal β-ensemble** if the law of its diagonal and upper-diagonal element is given by the joint density $\mathcal{P}_{\mathsf{tri},V}^{(\beta)}$.

Next, performing the change of variable from $\mathbf{T}$ to its eigenvalue decomposition and using identities for orthogonal polynomials associated with tridiagonal matrices, one can indeed prove that the joint density of eigenvalues is given by one of the corresponding β-ensemble, that is:

Result 1.7 (*Eigenvalues of a tridiagonal β-ensemble matrix*)

*for **any** $\beta > 0$,*

$$\text{if} \quad \mathbf{T} \sim \mathcal{P}^{(\beta)}_{\text{tri},V} \quad \text{then} \quad \boldsymbol{\lambda}(\mathbf{T}) \sim \mathcal{P}^{(\beta)}_{V} \,, \tag{1.96}$$

where $\mathcal{P}^{(\beta)}_{V}$ is the joint law of Eq. (1.63).

Thus for any $\beta > 0$, one can really think of $\boldsymbol{\lambda} \sim \mathcal{P}^{(\beta)}_{V}$ as the eigenvalues of a random matrix. Note that the eigenvalues of a tridiagonal matrix of the form of Eq. (1.94) are also the roots of the N^{th} monic polynomial defined by the recurrence relation

$$xP_n(x) = b_n^2 P_{n-1}(x) + a_{n+1}P_n(x) + P_{n+1}(x)\,, \tag{1.97}$$

with the first two terms given by $P_{-1} = 0$ and $P_0 = 1$. Hence, for $\boldsymbol{\lambda} \sim \mathcal{P}^{(\beta)}_{V}$, one can see the eigenvalues as the zeroes of the N^{th} monic polynomial obtained by this *random* recurrence.

However, for an arbitrary confining potential V, the description of the law of the corresponding tridiagonal β-ensemble given by Eq. (1.95) is not very enlightening and one may wonder if, for specific choices of V, its description greatly simplifies. This is the case for the three classical ensembles and in particular, the tridiagonal models associated with the Gaussian weight and the Laguerre weight were the first introduced tridiagonal models studied in Ref. [54]. The case of the Jacobi ensemble was studied in Ref. [101].

Tridiagonal matrix model for the three classical ensembles -

A matrix $\mathbf{T}$ is taken from the **Hermite β-ensemble** (HβE), if it is tridiagonal of the form of Eq.(1.94) and all the entries $\boldsymbol{a}, \boldsymbol{b}$ are distributed independently according to:

- $a_i \sim \mathcal{N}\left(0, \frac{\sigma}{\sqrt{\frac{N\beta}{2}}}\right)$ for $i = 1, \dots, N$
- $b_i^2 \sim \text{Gamma}\left(\frac{\beta}{2}(N-i), \frac{\sigma^2}{\frac{N\beta}{2}}\right)$ for $i = 1, \dots, N-1$

For $\mathbf{T} \sim$ HβE the eigenvalues are distributed according to $\boldsymbol{\lambda}(\mathbf{T}) \sim \mathcal{P}^{(\beta)}_{V_G}$ where V_G is the harmonic potential of the Gaussian ensemble given by Eq. (1.56).

A matrix $\mathbf{T}$ is taken from the **Laguerre β-ensemble** (LβE), if it is tridiagonal of the form of Eq.(1.94) and the entries $\boldsymbol{a}, \boldsymbol{b}$ are distributed according to

- $a_1 \overset{\text{in law}}{=} \gamma_1$ and $a_i \overset{\text{in law}}{=} \gamma_{2i-2} + \gamma_{2i-1}$ for $i = 2, \dots, N$,
- $b_i^2 \overset{\text{in law}}{=} \gamma_{2i-1}\gamma_{2i}$ for $i = 1, \dots, N-1$,

where the γ_i's are independent Gamma random variables distributed according to:

- $\gamma_{2i} \sim \text{Gamma}(\frac{\beta}{2}(N-i), \frac{2}{M\beta})$ for $i = 1, \dots, N-1$,
- $\gamma_{2i-1} \sim \text{Gamma}(\frac{\beta}{2}(M+1-i), \frac{2}{M\beta})$ for $i = 1, \dots, N$.

For $\mathbf{T} \sim \mathrm{L}\beta\mathrm{E}$ the eigenvalues are distributed according to $\boldsymbol{\lambda}(\mathbf{T}) \sim \mathcal{P}_{V_L}^{(\beta)}$ where V_L is given by Eq. (1.57) and is the potential associated to the Laguerre ensemble.

A matrix $\mathbf{T}$ is taken from the **Jacobi β-ensemble** (JβE), if it is tridiagonal of the form of Eq.(1.94) and the entries $\boldsymbol{a}, \boldsymbol{b}$ are distributed according to

- $a_1 \overset{\text{in law}}{=} B_1$ and $a_i \overset{\text{in law}}{=} (1 - B_{2i-3})B_{2i-2} + (1 - B_{2i-2})B_{2n-i}$ for $i = 2, \dots, N$,
- $b_1^2 \overset{\text{in law}}{=} B_1(1-B_1)B_2$ and $b_i^2 \overset{\text{in law}}{=} (1-B_{2i-2})B_{2i-1}(1-B_{2i-1})B_{2i}$ for $i = 2, \dots, N-1$,

where the B_i are independent Beta random variables with parameters:

- $B_{2i} \sim \mathrm{Beta}\left(\frac{\beta}{2}(N-i), \frac{\beta}{2}(M_1 + M_2 + 1 - N)\right)$ for $i = 1, \dots, N-1$,
- $B_{2i-1} \sim \mathrm{Beta}\left(\frac{\beta}{2}(M_1 + 1 - i), \frac{\beta}{2}(M_2 + 1 - i)\right)$ for $i = 1, \dots, N$.

For $\mathbf{T} \sim \mathrm{J}\beta\mathrm{E}$ the eigenvalues are distributed according to $\boldsymbol{\lambda}(\mathbf{T}) \sim \mathcal{P}_{V_J}^{(\beta)}$ where V_J is given by Eq. (1.59) and is the potential associated to the Jacobi ensemble.

1.5 Behavior of the largest eigenvalue of a β-ensemble: from Tracy-Widom to large deviations

In this section, we describe the typical and large fluctuations of the top eigenvalue. For a β-ensemble with a confining potential V for $\beta > 0$, if we order the eigenvalues in decreasing order $\lambda_1 \geq \cdots \geq \lambda_N$ then the top eigenvalue converges to:

$$\lambda_1 \equiv \lambda_{1_{(N)}} \xrightarrow[N\to\infty]{\text{a. s}} \mathrm{a}_+ \,, \tag{1.98}$$

where a_+ is the right edge of the limiting distribution μ_A, and our goal in this section is to describe the fluctuations around this limiting value.

1.5.1 Typical fluctuations and the Tracy-Widom distribution

To fix things, we consider the case where μ_A is non-critical, which we recall means that the density has a square root behavior near the edge. If the λ_i's were sampled independently from μ_A, then the fluctuations of its maximum would be described by the *Weibull distribution*, one of the three fixed point distributions of the *Fisher–Tippett–Gnedenko theorem* of extreme value theory. Yet, due to the long-range pairwise logarithmic interaction, the eigenvalues $\boldsymbol{\lambda}$ are *far* from being independent and the fluctuation of the top eigenvalue falls into *another class*:

Result 1.8 (*Tracy-Widom distribution and typical fluctuation [172]*)

If λ_1 is the top eigenvalue of β-ensemble whose LSD μ_A has a square-root behavior near

its top edge a_+*, then*

$$\frac{N^{2/3}}{\gamma_0}\,(\lambda_1 - \mathrm{a}_+) \xrightarrow[N\to\infty]{\text{in law}} T \sim \mathrm{TW}(\beta)\,, \tag{1.99}$$

where γ_0 *is the same constant as in Eq.* (1.80) *and* $\mathrm{TW}(\beta)$ *is the (generalized)* ***β-Tracy-Widom distribution****.*

The Tracy-Widom distribution was first studied for $\beta = 2$ in Ref. [172] and then for $\beta = 1, 4$ in Ref. [173]. In these three cases, the corresponding cumulative distribution admits a rather simple expression in terms of the *Hastings–McLeod solution* of the **Painlevé II equation**:

$$q''(s) - sq(s) - 2q(s)^3 = 0 \quad \text{with boundary condition } q(s) \underset{s\to\infty}{\sim} \mathrm{Ai}(s)\,, \tag{1.100}$$

where $\mathrm{Ai}(s) := \pi^{-1} \cdot \int_0^\infty \cos(t^3/3 + st)\mathrm{d}t$ is the *Airy function of the first kind*. If we denote by $\mathcal{F}_\beta(x) := \mathbb{P}_{T\sim \mathrm{TW}_\beta}[T \le x]$ the cumulative distribution, we have:

$$\mathcal{F}_\beta(x) = \begin{cases} \exp\left[-\frac{1}{2}\left(\int_x^\infty (s-x)q^2(s) + q(s)\mathrm{d}s\right)\right] & \text{for } \beta = 1\,, \\ \exp\left[-\int_x^\infty (s-x)q^2(s)\mathrm{d}s\right] & \text{for } \beta = 2\,, \\ \cosh\left[\frac{1}{2}\int_x^\infty (s-x)q(s)\mathrm{d}s\right]\,\exp\left[-\frac{1}{2}\int_x^\infty (s-x)q^2(s)\mathrm{d}s\right] & \text{for } \beta = 4\,. \end{cases} \tag{1.101}$$

For other values of $\beta > 0$, one can interpret the Tracy-Widom as the distribution of the lowest eigenvalue of a random operator known as the *Stochastic Airy Operator*, see Ref. [155], but there is not anymore a simple expression for the cumulative distribution $\mathcal{F}_\beta$. Yet, it is possible to obtain the behavior of the tail of the distribution for any $\beta > 0$:

$$\mathcal{F}_\beta(x) \asymp \mathrm{e}^{-\frac{\beta}{24}|x|^3} \quad \text{for } x \to -\infty\,, \tag{1.102}$$

and

$$1 - \mathcal{F}_\beta(x) \asymp \mathrm{e}^{-\frac{2\beta}{3}x^{3/2}} \quad \text{for } x \to \infty\,, \tag{1.103}$$

and $\asymp$ denotes equality at leading exponential order.

The Tracy-Widom distribution is ubiquitous and describes the fluctuations of the maximum 'long-range' correlated elements far beyond the realm of RMT. It appears in the well-known Kardar-Parisi-Zhang equation for specific initial conditions, see for example Ref. [106] and references therein, and it also describes the fluctuations of the length of the longest increasing sub-sequence [11] to cite a few, and we refer the reader to Refs. [25, 26, 119] for more on extreme value statistics and related topics.

1.5.2 Right large deviation and the pulled Coulomb gas

The small deviations of λ_1 around the limiting value a_+ are given by the Tracy-Widom law for fluctuations of order $|\lambda_1 - \mathrm{a}_+| \sim \mathcal{O}(N^{-\frac{2}{3}})$. To get the behavior of the large fluctuations,

that is at a value x far from the edge a_+, one is outside the scope of the Tracy-Widom regime describing typical fluctuations and one needs to estimate a *large deviation principle*. The scaling (or speed of convergence) of the large deviation principle is different if x is either above or below the edge a_+ and we first describe the case $x > \mathrm{a}_+$, that is the *right large deviation*.

If one integrates the joint density in Eq. (1.63) with a Dirac delta function $\delta(x-\lambda_1)$ one has that the probability can be expressed in terms of the difference of energy between the configuration of a $2d$-Coulomb gas where the top particle is *pulled* at the position x and the configuration of the unperturbed $2d$-Coulomb gas. For the perturbed gas, we are just moving one particle away from the bulk, and thus we expect that this perturbation does not change the equilibrium density μ_A of the $N-1$ other particles inside the bulk. This leads to the following result, see also Refs. [159, 121].

Result 1.9 (*right LDP for the top eigenvalue of a β-ensemble*)

As $N \to \infty$, the probability to find the top eigenvalue λ_1 as close as we want to a position x above the upper edge a_+, is given by a large deviation principle with speed $N\beta$:

$$\mathbb{P}[\lambda_1(\mathbf{A}) \simeq x] \asymp \exp\left[-N\beta\,\Psi(x)\right] \qquad (\textit{for } x \geq \mathrm{a}_+)\,, \tag{1.104}$$

and ***the right rate function*** Ψ *is given as (half) the difference of energy:*

$$\Psi(x) = \frac{1}{2}\left[V(x) - V(\mathrm{a}_+) - 2\int \log(x-\lambda)\mu_A(\lambda)\mathrm{d}\lambda + 2\int \log(\mathrm{a}_+ - \lambda)\mu_A(y)\mathrm{d}\lambda\right]. \tag{1.105}$$

This rate right function can also be written in integral form as:

$$\Psi(x) = \int_{\mathrm{a}_+}^{x}\left(\frac{V'(t)}{2} - g_A(t)\right)\mathrm{d}t\,, \tag{1.106}$$

where $g_A(.)$ is the *Stieltjes transform* of μ_A. If we introduce the second branch of the Stieltjes transform of Eq. (1.85), we can interpret the rate function as (half) the area between the two branches of the Stieltjes transform up to the position x:

$$\Psi(x) = \frac{1}{2}\int_{\mathrm{a}_+}^{x}\left(\bar{g}_A(t) - g_A(t)\right)\mathrm{d}t\,. \tag{1.107}$$

Note that for $g_A(\mathrm{a}_+) = \infty$, the rate function is finite.

Example *(Rate function for Gaussian ensembles)*. For the Gaussian ensemble, we recall that the potential is given by Eq. (1.56) and the Stieltjes transform is given by Eq. (1.41) and if we use the integral representation of Eq. (1.106) for the rate function, we get after simplification:

$$\Psi_{\mathsf{G}}(x) = \frac{x\sqrt{x^2-4\sigma^2}}{4\sigma^2} + \log\left(\frac{2\sigma}{\sqrt{x^2-4\sigma^2}+x}\right). \tag{1.108}$$

The two branches of the Stieltjes transform, and the rate function are given in Fig. 1.5 (Left). ∎

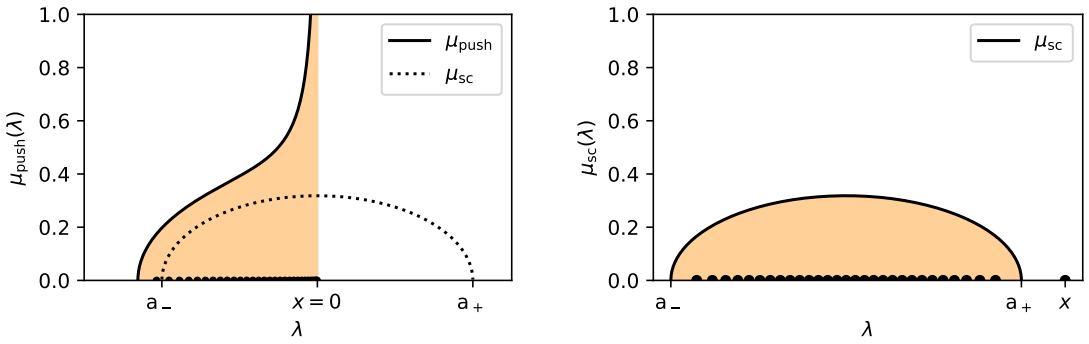


Figure 1.4: On the left, a representation of a 'typical' configuration of a pushed-to-the-origin Coulomb gas in a harmonic potential. To have the top eigenvalue at the position $x = 0$, one needs to also push all the other eigenvalues to the right and in the large N limit, this creates a different limiting density (with a black solid line) compared to the unperturbed semi-circle distribution (with dotted line). On the right, a 'typical' configuration corresponding to a Coulomb gas in a harmonic potential pulled at the position $x = 2.5 > \mathrm{a}_+ = 2$. Only the top eigenvalue pops out of the limiting distribution.

Example *(Rate function for Laguerre ensemble)***.** For the Laguerre ensemble, whose limiting spectrum is the Marčenko-Pastur distribution of Eq. (1.43), the potential is given by Eq. (1.57) and the Stieltjes transform is given by Eq. (1.44). The two branches of the Stieltjes transform are illustrated in Fig. 1.5 (Right). The right rate function is given by:

$$\Psi_{\mathsf{L}}(x) = \int_{\mathrm{a}_+}^{x} \frac{\sqrt{(t-\mathrm{a}_+)(t-\mathrm{a}_-)}}{2qt} \mathrm{d}t \,, \tag{1.109}$$

and is also represented in Fig. 1.5 (Right). Note that the integral in Eq. (1.109) can be computed analytically, but the result is not very enlightening. For $q = 1$ the rate function simplifies considerably, and we have:

$$\Psi_{\mathsf{L}}(x) = \frac{\sqrt{x(x-4)}}{2} + \log\left(\frac{x-2-\sqrt{x(x-4)}}{2}\right) \qquad (\text{for } q=1)\,. \tag{1.110}$$

In this case, one may notice the following identity:

$$\Psi_{\mathsf{L}}(x^2) = 2\Psi_{\mathsf{G}}(x) \qquad (\text{for } q=1)\,, \tag{1.111}$$

which will be discussed in more detail later on. ∎

Remark *(Behavior near the edge and the Tracy-Widom '3/2' scaling)***.** If one is looking at a non-critical density satisfying the condition of Eq. (1.80) near the edge, then both the Stieltjes and its second branch behave near the top edge as:

$$\begin{aligned} g_A(\mathrm{a}_+ + \epsilon) &= g_A(\mathrm{a}_+) - \gamma_0^{\frac{3}{2}}\sqrt{\mathrm{a}_+ + \epsilon} + o(\epsilon^{1/2})\,, \\ \bar{g}_A(\mathrm{a}_+ + \epsilon) &= g_A(\mathrm{a}_+) + \gamma_0^{\frac{3}{2}}\sqrt{\mathrm{a}_+ + \epsilon} + o(\epsilon^{1/2})\,, \end{aligned} \tag{1.112}$$

so approximating the integral of Eq. (1.107) by the Euler method, at first order one has for the rate function:

$$\Psi(\mathrm{a}_+ + \epsilon) = \frac{2\gamma^{3/2}}{3}\epsilon^{3/2} + o\left(\epsilon^{3/2}\right)\,, \tag{1.113}$$

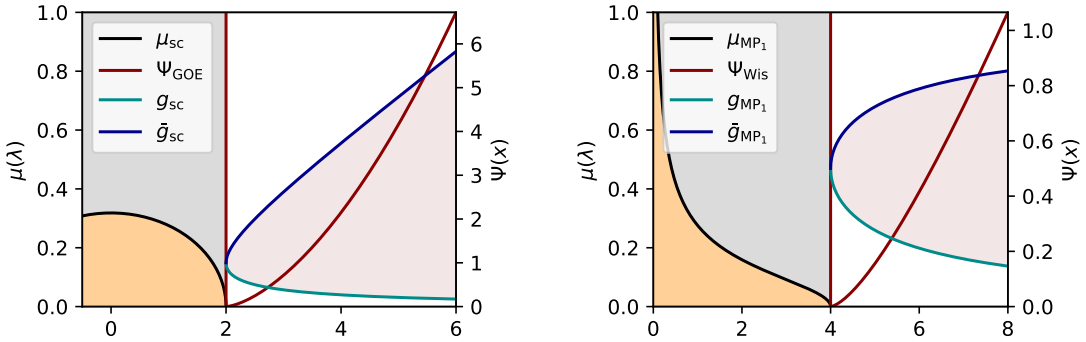


Figure 1.5: On the left, the rate function (in red) for the largest eigenvalue of a GOE matrix with $\sigma = 1$, whose expression is given by Eq. (1.108). On the right, the rate function (in red) for the largest eigenvalue of a Wishart matrix with $q = 1$, given by Eq. (1.110). In each case, the rate function is infinite for values below the edge of the limiting distribution and is otherwise given as half the area between the curve of the second branch of the Stieltjes transform (in blue) and the curve of the Stieltjes transform (in cyan), see Eq. (1.107).

and hence the probability behaves as

$$\mathbb{P}\left[\lambda_1 \simeq x\right] \approx \exp\left[-\frac{2\beta}{3}u^{3/2} + o(N)\right] \qquad \text{where } u = \gamma N^{2/3}(x - \mathrm{c}_+)\,. \tag{1.114}$$

The scaling of the reduced variable u and the asymptotic behavior matches the large argument behavior of the Tracy-Widom regime given by Eq. (1.103) which describes the probability of finding an eigenvalue near the edge. ⌟

1.5.3 Right large deviation of the top singular value

This section deals with rectangular matrices.

As in the self-adjoint case, the goal is to estimate for large N, the probability of finding the top singular value at a position x above its typical value given by the upper edge a_+ of the LSVD μ_A. By an almost identical argument of the previous section, this probability satisfies a large deviation principle with speed $N\beta$

$$\mathbb{P}\left[s_1 \simeq x\right] = \exp\left[-N\beta\,\Psi(x)\right]\,. \tag{1.115}$$

and the rate function Ψ is given as:

$$\Psi(x) = \int_{\mathrm{a}_+}^{x} t\left(\tilde{V}_q'(t^2) - \frac{1}{2}\int_{\mathrm{a}_-}^{\mathrm{a}_+}\frac{\rho_A(s)}{t^2 - s^2}\mathrm{d}s\right)\mathrm{d}t\,. \tag{1.116}$$

Remark *(square matrix and symmetrized density).* In the case where $\boldsymbol{s} \sim \mathcal{P}^{(\beta)}_{q,\tilde{V}_q}$ with $q = 1$, which corresponds for $\beta \in \{1, 2, 4\}$ to the case where $\mathbf{A} \sim \mathbb{P}^{(\beta)}_{q,V}$ is an (asymptotic) square matrix but not self-adjoint, there exist a nice relation with the LDP for the top eigenvalue of β-ensemble of the previous section. For $q = 1$ since $\tilde{V}_{q=1}(s^2) = V(s^2)$ (see Eq. (1.89)), the Tricomi problem of Eq. (1.93) for the LSVD μ_A reads:

$$⨍\frac{\mu_A(s')}{s^2 - s'^2}\mathrm{d}s' = \frac{V'(s^2)}{2}\,. \tag{1.117}$$

If we use again the identity of Eq. (1.34) and introduced a *new potential* $\hat{V}$ according to:

$$\hat{V}(\lambda) := \frac{V(\lambda^2)}{2}\,, \tag{1.118}$$

then we can re-write Eq. (1.117) as a Tricomi for the *symmetrized density*:

$$⨍ \frac{\widehat{\mu}_A(\lambda')}{\lambda - \lambda'} \mathrm{d}\lambda' = \frac{\hat{V}'(\lambda)}{2}\,. \tag{1.119}$$

where we recall that the symmetrized density $\widehat{\mu}_A(\lambda)$ is given by $\widehat{mu}_A(\lambda) = (\mu_A(\lambda) + \mu_A(-\lambda))/2$. Using again the identity Eq. (1.34) in the expression of Eq. (1.116) for the rate function Ψ with $q = 1$, together with the definition of $\hat{V}(.)$ given by Eq. (1.118), we can write the rate function as:

$$\Psi(x) = \int_{\mathrm{a}_+}^{x} \left(\hat{V}'(t) - 2\widehat{g}_A(t)\right) \mathrm{d}t = 2\Psi_{\widehat{A}}(x)\,, \tag{1.120}$$

where $\widehat{g}_A$ is the Stieltjes transform of $\widehat{\mu}_A$, and $\Psi_{\widehat{A}}$ is the rate function associated to the corresponding β-ensemble with the new potential $\hat{V}$. In other words, for a β-bi-invariant ensemble with potential $\tilde{V}_{q=1}$, the rate function associated with the largest singular value s_1 is *twice* the one associated with the largest eigenvalue λ_1 of a β-ensemble in a potential $\hat{V}$. Furthermore, this new potential is the one of the symmetrized distribution of μ_A. This can be heuristically guessed by remarking that for $q = 1$, we can express Eq. (1.90) as:

$$\mathcal{P}^{(\beta)}_{q,V}(\boldsymbol{s}) = \frac{1}{Z_N} \exp\left[-\frac{N\beta}{2}\left(\sum_{i=1}^{N}\left[\hat{V}(s_i) + \hat{V}(-s_i)\right] + \sum_{i,j|i\neq j} \frac{1}{2N}\log|s_i - s_j| + \frac{1}{2N}\log|s_i + s_j|\right)\right]\,, \tag{1.121}$$

where the potential $\hat{V}(x)$ is symmetric by construction. The joint law of the N (positive) singular values can be interpreted as the law of $2N$ eigenvalues following the usual Eq. (1.63) where the first N variables are constrained to be positive and each of the last N is constrained to equal minus its positive counterpart. In the large N limit, these constraints are irrelevant: the two problems have the same density (which does not depend on N) and differ by a factor of two for the rate function (which has an explicit N factor).

⌟

Example *(Rate function for Ginibre matrices).* If $\mathbf{A}$ is a Ginibre matrix, the LSVD is the quarter-circle law of Eq. (1.52) and its symmetrized density is the semi-circle law of Eq. (1.40). As a consequence, the rate function Ψ_{Gin} of the largest singular value of a Ginibre matrix is given by:

$$\Psi_{\mathsf{Gin}}(x) = 2\Psi_{\mathsf{G}}(x) = \frac{x\sqrt{x^2 - 4\sigma^2}}{2\sigma^2} + 2\log\left(\frac{2\sigma}{\sqrt{x^2 - 4\sigma^2} + x}\right)\,. \tag{1.122}$$

For $\sigma = 1$, $\mathbf{A}\mathbf{A}^*$ is a Wishart with shape parameter $q = 1$, hence $\Psi_{\mathsf{Gin}}(x) = \Phi_{\mathsf{L}}(x^2)$ in this case, and we retrieve the relation (1.111). ∎

1.5.4 A word on the left large deviation and the pushed Coulomb gas

For $x < \mathrm{a}_+$, one can still use the Coulomb gas analogy, but now the perturbed $2d$-Coulomb gas is compressed such that its top particle is at the position x. Unlike the case $x > \mathrm{a}_+$, the

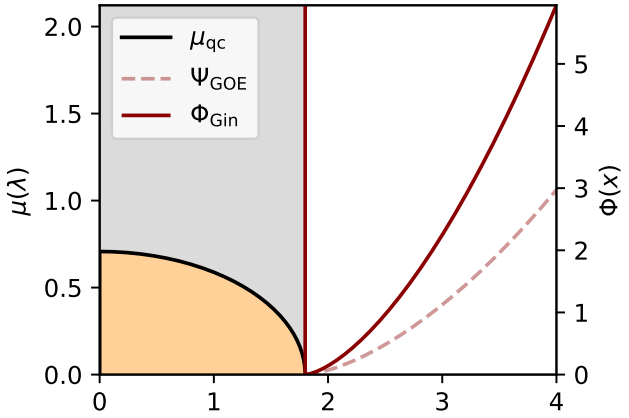


Figure 1.6: The Rate function (in red solid line) for the largest singular value of a Ginibre random matrix with $\sigma = 1$. This rate function is twice the rate function of the largest eigenvalue of a GOE matrix with also $\sigma = 1$, see Eq. (1.122).

equilibrium measure in the bulk is modified since one needs to *push* a large fraction of the particles to satisfy this constraint. Loosely speaking, this makes the probability to observe the event $\{\lambda_1 \simeq x\}$ equals to the probability to observe the event $\{\mu_{\mathbf{A}} \simeq \mu_{wall}(.,x)\}$ where the distribution $\mu_{wall}(.,x)$ is given as the solution of the *constrained Tricomi problem*:

$$-\!\!\!\!\!\!\int \frac{\mu_{wall}(\lambda,x)}{\lambda-\lambda'}\mathrm{d}\lambda' = \frac{V'(\lambda)}{2} \qquad \text{with} \quad \mu_{wall}(\lambda,x) = 0 \quad \text{for } \lambda > x\,. \tag{1.123}$$

Since the distribution $\mu_{\mathbf{A}}$ satisfies the LDP given by Res. 1.6, we deduce the following result for the top eigenvalue:

Result 1.10 (*Left LDP for the top eigenvalue of a β-ensemble*)

As $N \to \infty$, the probability to find the top eigenvalue λ_1 as close as we want to a position x below the upper edge a_+, is given by a large deviation principle with speed $\frac{N\beta}{2}$:

$$\mathbb{P}[\lambda_1 \simeq x] \asymp \exp\left[-\frac{N^2\beta}{2}\,\Psi_{\textit{left}}(x)\right] \qquad (\textit{for } x \geq \mathrm{a}_+)\,, \tag{1.124}$$

and ***the left rate function*** *$\Psi_{\textit{left}}$ is given as the difference of energy:*

$$\Psi_{\textit{left}}(x) = \mathcal{E}\left[\mu_{wall}(.,x)\right] - \mathcal{E}\left[\mu_A\right]\,, \tag{1.125}$$

with the energy $\mathcal{E}$ given by Eq. (1.67) and $\mu_{wall}(.,x)$ is given as the solution of the constrained Tricomi problem of Eq. (1.123) and μ_A as the solution of the unconstrained one of Eq. (1.75).

Let's insist on the fact that the speed of convergence is different for the right and for the left large deviation principle. For a given potential V, the left rate function is *much* harder to compute than the right one since one needs to first solve the constrained Tricomi problem of Eq. (1.123), and in general, there is no simple analytical solution for this problem.

Example *(Left rate function for Gaussian ensemble).* For the Gaussian ensemble with $\sigma =$

1, $V_{\mathsf{G}}(\lambda) = \lambda^2/2$, this left rate function has been computed in Ref. [47] and is given by:

$$\Psi_{left,\mathsf{G}}(w) = \frac{1}{108}\left(72w^2 - 4w^4 - (15\sqrt{2}w + 2\sqrt{2}w^3)\sqrt{2w^2+6} + 27\left(\log 18 - 2\log\left(\sqrt{2}w + \sqrt{2w^2+6}\right)\right)\right). \quad (1.126)$$

Near the upper edge $\mathrm{a}_+ = 2$, this function behaves as:

$$\Psi_{left,\mathsf{G}}(w) \propto (2-w)^3 \text{ for } w \to 2 \text{ and } w < 2\,, \quad (1.127)$$

and this asymptotic behavior matches smoothly with the Tracy-Widom tail of Eq. (1.102). ∎

The left large deviation for the Laguerre ensemble has been derived in Ref. [181]. Our primary interest lies in the right large deviation, and we will (almost) not discuss this left large deviation principle in the rest of this thesis and we refer to Refs. [118, 117, 122, 46, 180, 134] for more on large deviations for β-ensembles.

Summary of the behavior -

These three different regimes can be summarized by the following large N behavior:

$$\mathbb{P}\left[\lambda_1 \simeq x\right] \approx \begin{cases} \exp\left[-\frac{N^2\beta}{2}\Psi_{left}(x) + o(N^2)\right] & \text{for } x < \mathrm{a}_+ \text{ and } |x-\mathrm{a}_+| \sim \mathcal{O}_N(1)\,, \\ \gamma_0 N^{2/3}(\mathcal{F}_\beta)'\left(\gamma_0 N^{2/3}(x-\mathrm{a}_+)\right) & \text{for } |x-\mathrm{a}_+| \sim \mathcal{O}(N^{-\frac{2}{3}})\,, \\ \exp\left[-N\beta\Psi(x) + o(N)\right] & \text{for } x > \mathrm{a}_+ \text{ and } |x-\mathrm{a}_+| \sim \mathcal{O}_N(1)\,, \end{cases} \quad (1.128)$$

where $\mathcal{F}'_\beta$ is the density of the Tracy-Widom distribution, Ψ is given by Eq. (1.105) and Ψ_{left} is given by Eq. (1.125), see Fig. 1.7 for an illustration.

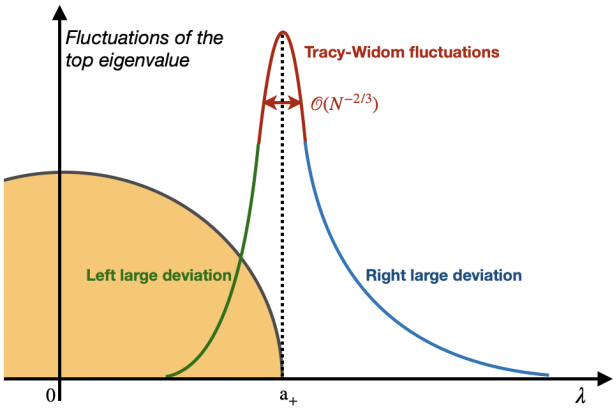


Figure 1.7: A sketch of the different regimes of the fluctuation of the top eigenvalue for a β-ensemble with square root behavior near the edge, as described by Eq. (1.128).

1.6 β-ensembles in the low-temperature regime, average characteristic polynomials, and zeroes of orthogonal polynomials

1.6.1 Characterization of the ground state in the low-temperature limit

In the large N limit, the limiting distribution of a β-ensemble with $\beta > 0$ fixed, is described by the Tricomi problem of Eq. (1.75). If we now **send $\beta \to \infty$ with N being fixed**, that is we look at **β-ensemble in the low-temperature regime**, it is immediate from the joint law of Eq. (1.63) that the eigenvalues will *crystallize* on the deterministic positions $\boldsymbol{\lambda}^\star$ given as the ground state of the energy $\mathcal{E}_N$ of Eq. (1.66):

$$\boldsymbol{\lambda}^\star := \underset{\boldsymbol{\lambda}\in\mathbb{R}^N}{\operatorname{argmin}}\, \mathcal{E}_N(\boldsymbol{\lambda}) \qquad , \tag{1.129}$$

where the potential V in the definition of $\mathcal{E}_N$ has to be understood as the low temperature limit $V \equiv \lim_{\beta\to\infty} V_{N,\beta}$, whenever it depends explicitly on the inverse temperature β.

Up to this possible replacement for the potential, this low-temperature limit corresponds to *the most likely configuration* of a β-ensemble with the number N of eigenvalues being fixed, and it is a natural question is to give an interpretation of the set $\boldsymbol{\lambda}^\star$ in terms of well-known mathematical objects. To this end, let's look at the solution of:

$$\nabla\, \mathcal{E}_N(\boldsymbol{\lambda})\,|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^\star} = \mathbf{0}\,, \tag{1.130}$$

using the expression of Eq. (1.66) for $\mathcal{E}_N$, this gives:

$$\frac{1}{N}\sum_{j|j\neq i}\frac{1}{\lambda_i^\star - \lambda_j^\star} = \frac{V'(\lambda_i^\star)}{2} \qquad \text{for } i = 1,\dots,N\,, \tag{1.131}$$

which is nothing else than the finite version of the Tricomi problem of Eq. (1.75). If we adapt our derivation of the BIPZ equation to this finite setting by adding and subtracting $V'(z)$ to the RHS of Eq. (1.131) and then multiply Eq. (1.131) by $(z-\lambda_i^\star)^{-1}N^{-1}$ and sum over the index i, we get:

$$\frac{2}{N^2}\sum_{i=1}\sum_{j|j\neq i}\frac{1}{(z-\lambda_i^\star)(\lambda_i^\star-\lambda_j^\star)} = V'(z)\left(\frac{1}{N}\sum_{i=1}^N\frac{1}{z-\lambda_i^\star}\right) - \Pi_{N,V}(z) \tag{1.132}$$

where $\Pi_{N,V}(z) := \frac{1}{N}\sum_{i=1}^N \frac{V'(z)-V(\lambda_i^\star)}{z-\lambda_i^\star}$ is the finite counterpart of $\Pi_V(z)$ given by Eq. (1.83). Now if we introduce the finite Stieltjes transform:

$$g_N^\star(z) := \frac{1}{N}\sum_{i=1}^N \frac{1}{z-\lambda_i^\star}\,, \tag{1.133}$$

the LHS of Eq. (1.132) is given by:

$$\frac{2}{N^2}\sum_{i=1}\sum_{j|j\neq i}\frac{1}{(z-\lambda_i^\star)(\lambda_i^\star-\lambda_j^\star)} = g_N^\star(z)^2 + \frac{1}{N}g_N^{\star\prime}(z)\,. \tag{1.134}$$

Note that there is now an additional derivative term (which goes to zero as $N \to \infty$) compared to the BIPZ case. The final equation for the Stieltjes is:

$$V'(z)g_N^\star(z) - \Pi_{N,V}(z) - g_N^\star(z)^2 - \frac{1}{N}g_N^{\star\prime}(z) = 0\,. \tag{1.135}$$

This is a non-linear differential equation for the Stieltjes transform and one would like to find the change of variable which transform this non-linear equation into an easier-to-solve linear equation.

To do, let's introduce the **monic polynomial with roots $\boldsymbol{\lambda}^\star$**:

$$P_N^\star(z) := \prod_{i=1}^{N}(z - \lambda_i^\star)\ . \tag{1.136}$$

This polynomial is related to the Stieltjes transform by a *Cole-Hopf* (or Riccati) transform:

$$g_N^\star(z) = \frac{1}{N}\left(\log P_N^\star(z)\right)' = \frac{1}{N}\,\frac{P_N^{\star\prime}(z)}{P_N^\star(z)}\,, \tag{1.137}$$

and if we insert this expression in the differential equation (1.135) for the Stieltjes, we get the following result:

Result 1.11 (*Differential equation for the polynomial with roots $\boldsymbol{\lambda}^\star$*)

The monic polynomial with roots $\boldsymbol{\lambda}^\star$ is the solution of the second order differential equation

$$P_N^{\star\prime\prime}(z) - NV'(z)\,P_N^{\star\prime}(z) + N^2\Pi_{N,V}(z)\,P_N^\star(z) = 0\,, \tag{1.138}$$

with $\Pi_{N,V}(z) := \frac{1}{N}\sum_{i=1}^{N}\frac{V'(z)-V(\lambda_i^\star)}{z-\lambda_i^\star}$.

Because Eq. (1.138) is *linear*, one can think of $P_N^\star$ as the 'natural quantity' to describe the finite N setting.

1.6.2 Relation to zeroes of orthogonal polynomials

This is equation is reminiscent of the differential equation satisfied by *orthogonal polynomials*. Let's introduce a continuous measure on (a, b) where the bounds a, b may or not be finite, with a density of the form:

$$w(x) := \mathrm{e}^{-NW(x)}\mathbb{I}_{(a,b)}\,, \tag{1.139}$$

for some function W.

The associated **monic orthogonal polynomials** $(\mathsf{P}_{W,i})_{i\in\mathbb{N}}$ are defined by the orthogonal relation:

$$\int \mathsf{P}_{W,i}(x)\mathsf{P}_{W,j}(x)\,\mathrm{e}^{-NW(x)}\mathbb{I}_{(a,b)}\mathrm{d}x = 0 \quad \text{for } i \neq j\,, \tag{1.140}$$

and normalized such that the coefficient in front of the leading term is one:

$$\mathsf{P}_{W,i}(x) = x^i + \sum_{j=0}^{i-1} \alpha_j x^j \ . \tag{1.141}$$

Then if we introduce the function

$$A_N(x) := \frac{\mathrm{e}^{-NW(y)}\mathsf{P}_{W,i}(y)}{y-x}\bigg|_{y=a}^{y=b} + N\int_a^b \frac{W'(x)-W'(y)}{x-y}\mathsf{P}^2_{W,N}(y)\mathrm{e}^{-NW(y)}\mathbb{I}_{(a,b)}\mathrm{d}y\,, \tag{1.142}$$

we have the following result, see for example Ref. [95] and references therein.

Result 1.12 (*electrostatic model for the zeros of orthogonal polynomials, [95]*)

Under some convexity and boundary conditions on W and $\log(A_N)$, the N^{th} monic polynomial $\mathsf{P}_{W,N}$ with respect to the measure given by Eq. (1.139), satisfies the differential equation:

$$\mathsf{P}''_{W,N}(z) - N\frac{\mathrm{d}}{\mathrm{d}z}\left(W(z) + \frac{1}{N}\log A_N(z)\right)\mathsf{P}'_{W,N}(z) + N^2\,\Pi_{N,W+\frac{1}{N}\log A_N}(z)\,\mathsf{P}_{W,N}(z) = 0\,. \tag{1.143}$$

In other words, the zeros of the N^{th} monic orthogonal polynomial with respect to the measure $\mathrm{e}^{-NW(x)}\mathbb{I}_{(a,b)}$ are the minimizers of the Coulomb energy $\mathcal{E}_N$ given by Eq. (1.66) with a potential $V = W(x) + \frac{1}{N}\log A_N(x)$. The proof of this statement together with the precise assumptions can be found in Ref. [95].

Conversely, this suggests that $P_N^\star$ is the N^{th} monic orthogonal polynomial of some measure of the form $\mathrm{e}^{-NV(x)+\log A_N(x)}$, provided the assumptions for Eq. (1.143) to hold are satisfied. The potential $\frac{1}{N}\log A_N(x)$ is a *local potential* whose contribution to the overall energy becomes negligible at large N.

$P_N^\star$ for classical ensembles and classical orthogonal polynomials -

If we now restrict to the cases where V is the potential of one of the classical ensembles, then by either identifying the differential equation (1.138) satisfied by $P_N^\star$ or using results dating back to Stieltjes characterizing the *classical orthogonal polynomials* in terms of an electrostatic model of the form of Eq. (1.143), one has:

- If V is the harmonic potential associated with the Gaussian ensemble, given by Eq. (1.56), then we have:

$$\boxed{P_N^\star(z) = C_{H,N}\,\mathsf{He}_N\left(\frac{\sqrt{N}}{\sigma}z\right)} \tag{1.144}$$

  where He_N is the N^{th} (probabilist's) **Hermite polynomial**

$$\mathsf{He}_N(x) := N!\sum_{k=0}^{\lfloor N/2\rfloor}\frac{(-1)^k}{k!(N-2k)!}\cdot\frac{x^{N-2k}}{2^k}\,, \tag{1.145}$$

  and $C_{H,N} := (N/\sigma^2)^{-N/2}$ is the constant which makes this re-scaled Hermite polynomial monic.

- If V is the limit $\beta \to \infty$ of the potential associated with the Laguerre ensemble, given by Eq. (1.57), then we have:

$$\boxed{P_N^\star(z) = C_{L,N}\, \mathsf{La}_N^{(M-N)}\,(Mz)} \tag{1.146}$$

where $\mathsf{La}_N^{(\alpha)}$ is the N^{th} **generalized Laguerre polynomial**:

$$\mathsf{La}_N^{(\alpha)}(x) := \sum_{k=0}^{N} (-1)^k \binom{N+\alpha}{N-k} \frac{x^k}{k!}\,, \tag{1.147}$$

and $C_{L,N} := (-1)^N M^{-N} N!$ is the constant which makes this re-scaled Laguerre polynomial monic.

- If V is the limit $\beta \to \infty$ of the potential associated with the Jacobi ensemble given by Eq. (1.59), then we have:

$$\boxed{P_N^\star(z) = C_{J,N}\, \mathsf{Ja}_N^{(M_1-N, M_2-N)}\,(z)} \tag{1.148}$$

where $\mathsf{Ja}_N^{(a,b)}$ is the N^{th} **unit Jacobi polynomial**, which is related to the standard Jacobi polynomial by $\mathsf{Ja}_N^{(a,b)}(x) = \mathsf{P}_N^{(b,a)}(2x-1)$[2]:

$$\mathsf{Ja}_N^{(a,b)}(x) := \frac{\Gamma(b+N+1)}{N!\,\Gamma(a+b+N+1)} \sum_{k=0}^{N} \binom{N}{k} \frac{\Gamma(a+b+n+k+1)}{\Gamma(b+k+1)} \cdot (x-1)^k\,, \tag{1.149}$$

and the constant $C_{J,N} := N! \frac{\Gamma(M_1+M_2+1-N)}{\Gamma(M_1+M_2+1)}$ makes this unit Jacobi polynomial monic.

1.6.3 Relation to average characteristic polynomials (ACP)

Relation to the ACP of classical ensembles at any $\beta > 0$ -

For $\boldsymbol{\lambda} \sim \mathcal{P}_V^{(\beta)}$, the **average characteristic polynomial** (ACP) is defined as the following quantity

$$\mathbb{E}\left[\prod_{i=1}^{N}(z-\lambda_i)\right] = \frac{1}{Z_{V,\beta}} \int \mathrm{e}^{-\frac{N\beta}{2}\sum_{i=1}^{N} V(\lambda_i)} \prod_{i=1}^{N}(z-\lambda_i)\,|\Delta(\boldsymbol{\lambda})|^\beta \mathrm{d}\boldsymbol{\lambda}\,. \tag{1.150}$$

In the case where V is the potential of one of the three classical ensembles one can compute the corresponding ACP for *any* $\beta > 0$, thanks to the recurrence relation of Eq. (1.97) for tridiagonal matrices. For Gaussian and Laguerre ensembles, one can directly average this relation with respect to the law of $(\boldsymbol{a}, \boldsymbol{b})$ to get a recurrence relation for the corresponding ACP. The case of the Jacobi ensemble requires slightly more work and has been tackled in Ref.

[2] One may note that we have switched the order of the parameters of the two Jacobi polynomials. This is not a typo, the reason is that unit Jacobi polynomials $(\mathsf{Ja}_i^{(a,b)})_{i\in\mathbb{N}}$ are orthogonal polynomials for the measure associated to a $\mathrm{Beta}(a,b)$ random variable (and not a $\mathrm{Beta}(b,a)$ random variable), hence this order seems more appropriate.

[101]. Importantly, one recovers the classical orthogonal polynomials and with our convention, the result is independent of the parameter β, and thus equal to the one of the low-temperature limit $\beta \to \infty$! Explicitly this means:

- If $\mathbf{A} \sim \mathsf{H}\beta\mathsf{E}$ with eigenvalues $\boldsymbol{\lambda}$ or equivalently if $\boldsymbol{\lambda} \sim \mathcal{P}_{V_G}^{(\beta)}$ with V_G given by Eq. (1.56), then we have:

$$\boxed{\text{for } \textbf{any } \beta > 0\,, \qquad \mathbb{E}\left[\prod_{i=1}^{N}(z-\lambda_i)\right] = P_N^\star(z) = C_{H,N}\,\mathsf{He}_N\left(\frac{\sqrt{N}}{\sigma}z\right)}\,. \tag{1.151}$$

- If $\mathbf{A} \sim \mathsf{L}\beta\mathsf{E}$ with eigenvalues $\boldsymbol{\lambda}$ or equivalently if $\boldsymbol{\lambda} \sim \mathcal{P}_{V_L}^{(\beta)}$ with V_L given by Eq. (1.57), then we have:

$$\boxed{\text{for } \textbf{any } \beta > 0\,, \qquad \mathbb{E}\left[\prod_{i=1}^{N}(z-\lambda_i)\right] = P_N^\star(z) = C_{L,N}\,\mathsf{La}_N^{(M-N)}\,(Mz)}\,. \tag{1.152}$$

- If $\mathbf{A} \sim \mathsf{J}\beta\mathsf{E}$ with eigenvalues $\boldsymbol{\lambda}$ or equivalently if $\boldsymbol{\lambda} \sim \mathcal{P}_{V_J}^{(\beta)}$ with V_J given by Eq. (1.59), then we have:

$$\boxed{\text{for } \textbf{any } \beta > 0\,, \qquad \mathbb{E}\left[\prod_{i=1}^{N}(z-\lambda_i)\right] = P_N^\star(z) = C_{J,N}\,\mathsf{Ja}_N^{(M_1-N,M_2-N)}\,(z)}\,. \tag{1.153}$$

Remark. For the Laguerre and the Jacobi ensembles, the fact that the ACP is independent of the parameter β is specific to our convention for the associated potentials V_L, V_J which depend explicitly on β. ⌟

Average of characteristic polynomial of unitary invariant ensembles $(\beta = 2)$ -

For arbitrary confining potential, as far as I know, there is no simple relation at any $\beta > 0$ between the ACP and the low-temperature polynomial $P_N^\star$. Yet, for $\beta = 2$, there exist another well-known relation between orthogonal polynomial and ACP.

Let's look at the ACP of a β=2-ensemble:

$$\mathbb{E}\left[\prod_{i=1}^{N}(z-\lambda_i)\right] = \frac{1}{Z_{V,\beta=2}}\int \mathrm{e}^{-N\sum_{i=1}^{N}V(\lambda_i)}\prod_{i=1}^{N}(z-\lambda_i)\,\Delta(\boldsymbol{\lambda})^2\mathrm{d}\boldsymbol{\lambda}\,, \tag{1.154}$$

and fix $z > \lambda_1 = \max\boldsymbol{\lambda}$ without loss of generality. We can absorb the product coming from the characteristic polynomial with one Vandermonde determinant to produce a Vandermonde determinant of $N+1$ variables, that is the ACP writes:

$$\mathbb{E}\left[\prod_{i=1}^{N}(z-\lambda_i)\right] = \frac{1}{Z_{V,\beta=2}}\int \mathrm{e}^{-N\sum_{i=1}^{N}V(\lambda_i)}\Delta(\boldsymbol{\lambda},z)\,\Delta(\boldsymbol{\lambda})\mathrm{d}\boldsymbol{\lambda}\,. \tag{1.155}$$

Next, determinants are left unchanged by adding to a column a multiple of another column, we have in particular the identity:

$$\Delta(\boldsymbol{\lambda}) = \det\left(\lambda_i^{j-1}\right)_{1\le i,j\le N} = \det\left[\mathsf{P}_{V,j-1}(\lambda_i)\right]_{1\le i,j\le N}\,, \tag{1.156}$$

and similarly for $\Delta(\boldsymbol{\lambda}, z)$, where we recall that $\mathrm{P}_{V,i}$ is the i^{th} monic orthogonal polynomial with respect to the measure e^{-NV}. If we now use the Leibniz identity to express each determinant as an alternating sum of products, we have:

$$\Delta(\boldsymbol{\lambda})\Delta(\boldsymbol{\lambda}, z) = \sum_{\substack{\sigma \in \mathfrak{S}_N \\ \sigma' \in \mathfrak{S}_{N+1}}} \mathrm{s}(\sigma)\mathrm{s}(\sigma')\, \mathrm{P}_{V,\sigma(0)}(\lambda_1) \dots \mathrm{P}_{V,\sigma(N-1)}(\lambda_N) \times \mathrm{P}_{V,\sigma'(0)}(\lambda_1) \dots \mathrm{P}_{V,\sigma'(N-1)}(\lambda_N)\, \mathrm{P}_{V,\sigma'(N)}(z)\,, \quad (1.157)$$

where $s(\sigma)$ is the sign of the permutation σ. Importantly, each variable λ_i appears in exactly two polynomials of the product of the RHS of Eq. (1.157). As a consequence, if we now multiply this equation by $\mathrm{e}^{-N\sum_{i=1}^{N} V(\lambda_i)}\mathrm{d}\boldsymbol{\lambda}$ and integrate, by the orthogonality relation of Eq. (1.140), the only pairs (σ, σ') of permutations contributing to the sum are the ones satisfying $\sigma(i) = \sigma'(i)$ for *every* $i = 0, \dots, N-1$. This clearly imposes $\sigma'(N) = N$. After integration all such terms are equals and since both the ACP and $\mathrm{P}_{V,N}$ are monic, we have the following result due to Szegő [167]:

Result 1.13 (*Average characteristic polynomial at $\beta = 2$*)

$$\textit{if} \quad \boldsymbol{\lambda} \sim \mathcal{P}_V^{(\beta=2)} \quad \textit{then} \quad \mathbb{E}\left[\prod_{i=1}^{N}(z - \lambda_i)\right] = \mathrm{P}_{V,N}(z)\,, \quad (1.158)$$

where $\mathrm{P}_{V,N}(z)$ is the N^{th} monic orthogonal with respect to the measure e^{-NV}.

Remark *(Vandermonde and orthogonal polynomials).* Eq. (1.156) expressing the Vandermonde as the determinant of the orthogonal polynomials is a well-known identity of RMT and has been used in a variety of different contexts. In particular, this is the starting point to get the well-known *determinantal formula* for the correlation functions in terms of the *Christoffel-Darboux kernel* at $\beta = 2$, see for example Ref. [108]. ⌟

As a consequence we have a relation between three monic polynomials of degree N: the monic polynomial $P_N^\star$ whose roots $\boldsymbol{\lambda}^\star$ corresponds to the limit $\beta \to \infty$ of a β-ensemble with N fixed, the monic orthogonal polynomial $\mathrm{P}_{V,N}$, and the average characteristic polynomials $\mathbb{E}\left[\prod_{i=1}^{N}(z - \lambda_i)\right]$ of a β-ensemble. In particular, for the three classical ensembles, these three objects are equal.

For β-ensemble in the low temperature, as we insert a new particle, we slightly perturb the Coulomb gas and the associated energy increases. In the large N limit, we recover the LSD, and we have $\mathcal{E}_N(\boldsymbol{\lambda}^\star) \underset{N\to\infty}{\nearrow} \mathcal{E}[\mu_A]$. This makes a (well-known) beautiful bridge between the three 'standard' laws of probability (the Gaussian, the Gamma, and the Beta distributions) and the LSD of the three classical ensembles (the semi-circle, Marčenko-Pastur, and Watcher distributions). Up to a re-scaling, the distribution of the zeroes of the N^{th} monic orthogonal polynomials of a standard law converges as $N \to \infty$ to the LSD of the corresponding β-ensemble.

1.7 β-ensembles in the high-temperature regime and the high-low temperature duality

For β-ensemble with a fixed inverse temperature $\beta > 0$, the limiting density μ_A is the same for all β. In essence, this is because the entropic term is of order $\mathcal{O}(N)$ and is, therefore, subdominant compared to the potential and the pairwise logarithmic interaction terms, which are of order $\mathcal{O}(N^2)$ in Eq. (1.71). If we choose to *tune* β with N such that all terms are now of the same order, one should now account for entropic contributions, and the limiting distribution will depend on the weight of this term, and hence on the temperature. For all terms to be of the same order, by a rule of thumbs one sees that the inverse temperature should go slowly to zero as $\beta_N \sim \mathcal{O}(1/N)$.

As a consequence, we are interested in the double scaling limit known as the **high-temperature regime** where:

$$\frac{N\beta_N}{2} \xrightarrow[N\to\infty]{} c \in \mathbb{R}_+^* \,. \tag{1.159}$$

In practice, this high-temperature regime is obtained by first setting $\beta_N = 2c/N$ and then taking the limit $N \to \infty$. One can then look at β-ensembles under such scaling:

A vector $\boldsymbol{\lambda}$ is said to be taken from **β-ensemble at high temperature** (with potential V and parameter c), $\boldsymbol{\lambda} \sim \mathcal{P}_V^{[c]}$, if its joint density is given by Eq. (1.64) with $\beta = 2c/N$,that is:

$$\mathcal{P}_V^{[c]}(\boldsymbol{\lambda}) := \frac{\mathrm{e}^{-c\sum_{i=1}^N V(\lambda_i)}\,|\Delta(\boldsymbol{\lambda})|^{\frac{2}{cN}}}{Z_{V,c}} = \frac{\mathrm{e}^{-Nc\,\mathcal{E}_N(\boldsymbol{\lambda})}}{Z_{V,c}} \tag{1.160}$$

and the energy $\mathcal{E}_N(.)$ is defined as before by Eq. (1.66).

Note that we use the bracket notations $[c]$ to differentiate quantities from the high-temperature regime from their counterpart in the usual regime. The following result is then an immediate consequence of this re-scaling:

Result 1.14 (*LDP for the distribution of high-temperature ensemble*)

In the large N limit, the empirical distribution of β-ensemble at high temperature satisfies a large deviation principle with speed Nc and rate function given by:

$$\mathcal{E}[\mu] + \frac{1}{c}\mathcal{S}[\mu] - \left(\mathcal{E}\left[\mu_A^{[c]}\right] + \frac{1}{c}\mathcal{S}\left[\mu_A^{[c]}\right]\right) \qquad \textit{where} \quad \mu_A^{[c]} := \operatorname*{argmin}_{\mu'\in\mathcal{M}_1}\left(\mathcal{E}[\mu'] + \frac{1}{c}\mathcal{S}[\mu']\right). \tag{1.161}$$

In particular, this implies that if $\boldsymbol{\lambda} \sim \mathcal{P}_V^{[c]}$ then $\mu_{\mathbf{A}} := \sum_{i=1}^N \delta(x-\lambda_i)/N \to \mu_A^{[c]}$ as $N \to \infty$.

Remark *(Re-scaling of the high-temperature ensembles).* For $\boldsymbol{\lambda} \sim \mathcal{P}_V^{[c]}$, as $c \to \infty$ the entropy term in the variational formulation of Eq. (1.161) becomes negligible, and we retrieve $\mu_A^{[c\to\infty]} \to \mu_A$, with μ_A the limiting density of a classical β-ensemble with potential V. However, as $c \to 0^+$ we have the trivial limit $\mu_A^{[c\to 0^+]} \to \delta(.)$, while one would naturally like to recover the classical weight e^{-V} in this limit. This leads to modify the potential

according to $V \to \frac{c+1}{c} V$ to get the following new re-scaled joint density:

$$\tilde{\mathcal{P}}_V^{[c]}(\boldsymbol{\lambda}) \propto \mathrm{e}^{-(c+1)\sum_{i=1}^N V(\lambda_i)} |\Delta(\boldsymbol{\lambda})|^{\frac{2}{cN}} , \tag{1.162}$$

for which the corresponding limiting density naturally interpolates between the classical distribution e^{-V} and μ_A as the parameter c varies from zero to infinity. ⌟

1.7.1 Description of the equilibrium measure in the high-temperature regime and U-function

The limiting (or equilibrium) distribution $\mu_A^{[c]}$ is given as a solution of

$$\frac{\delta}{\delta \mu} \left(\mathcal{E}[\mu] + \frac{1}{c} \mathcal{S}[\mu] \right) \Big|_{\mu_A^{[c]}} = 0 , \tag{1.163}$$

that is using Eq. (1.66) for the expression of the electrostatic energy and Eq. (1.70) for the entropy, $\mu_A^{[c]}$ is solution of:

$$\frac{V'(\lambda)}{2} - \fint \frac{\mu_A^{[c]}(\lambda')}{\lambda - \lambda'} \mathrm{d}\lambda' + \frac{1}{c} \frac{\mathrm{d}}{\mathrm{d}\lambda} \log \left(\mu_A^{[c]}(\lambda) \right) = 0 . \tag{1.164}$$

Since the derivative of the logarithm of a function is simply the derivative of the function over the function itself, one can multiply Eq. (1.164) by $\mu_A^{[c]}(\lambda)/(z-\lambda)$ and integrate over the real line to get an equation for the Stieltjes transform $g_A(z) = \int \mu_A^{[c]}(\lambda)(z-\lambda)^{-1} \mathrm{d}\lambda$. The first two terms of Eq. (1.164) are the usual terms of a β-ensemble and so after this procedure they will give the LHS of Eq. (1.82) and we have:

$$V'(z) g_A(z) - \Pi_{[c],V}(z) - g_A(z)^2 + \frac{1}{c} \int \frac{(\mu_A^{[c]}(\lambda))'}{z - \lambda} \mathrm{d}\lambda = 0 , \tag{1.165}$$

where $\Pi_{[c],V}$ is given by Eq. (1.83) with μ_A replaced by $\mu_A^{[c]}$. By integration by part, the last term of Eq. (1.165) is nothing else than the derivative of the Stieltjes transform. As a consequence, the Stieltjes transform satisfies the following non-linear differential equation:

$$V'(z) g_A(z) - \Pi_{[c],V}(z) - g_A(z)^2 + \frac{1}{c} g_A'(z) = 0 . \tag{1.166}$$

> Interestingly, Eq. (1.166) is exactly Eq. (1.135) with $N \in \mathbb{N}$ replaced by $-c$, with $c \in \mathbb{R}_+^*$. This is the first appearance of the **high-low temperature duality** $c \leftrightarrow N$, a concept that will be re-discovered in Chapter 6. Roughly speaking, this duality is a correspondence between the low-temperature world ($\beta \to \infty$) and the high-temperature world ($N\beta/2 \to c$) where one can go from one world to another by analytically continuing (in N, resp. in c) the appropriate observable of interest.

Note that while this differential equation is not new, (see for example Ref. [1] after simplification of the integral), its correspondence with the low-temperature regime does not seem to be well-known for arbitrary potential V. Let's mention that this duality is different from the one developed in Ref. [49].

Eq. (1.166) is a non-linear differential equation for the Stieltjes transform and as in the low-temperature case described in Sec. 1.6, we would like to find the *natural* change of variable for which Eq. (1.166) transforms into a linear equation.

To this end, for a measure μ_A, with a_+ the upper limit of its support, we define for any $z \in \mathbb{C} \setminus (-\infty, \mathrm{a}_+)$ its **U-function** (of parameter c) as:

$$U_A^{[c]}(z) \equiv U_{\mu_A}^{[c]}(z) := \exp\left[-c \int_{\mathrm{Supp}[\mu_A]} \mathrm{d}\mu_A(\lambda) \log(z-\lambda)\right]. \tag{1.167}$$

This function admits the following large z behavior:

$$U_A^{[c]}(z) \underset{|z|\to\infty}{\sim} z^{-c}, \tag{1.168}$$

The U-function is related to the Stieltjes transform by a Cole-Hopf transform:

$$g_A(z) = -\frac{1}{c}\left(\log U_A^{[c]}(z)\right)' = -\frac{1}{c}\frac{\left(U_A^{[c]}\right)'(z)}{U_A^{[c]}(z)}. \tag{1.169}$$

Note that we can define the U-function up to a multiplicative constant without changing the Stieltjes transform. Thus, if we denote by $U_A^{[c]}$ the U-function of the equilibrium measure $\mu_A^{[c]}$, the change of variable given by Eq. (1.169) kills the non-linearity in Eq. (1.166) and we have:

Result 1.15 (*Differential equation for the U-function*)

$$U_A^{[c]\prime\prime}(z) + c\,V'(z)\,U_A^{[c]\prime}(z) + c^2\,\Pi_V^{[c]}(z)\,U_A^{[c]}(z) = 0\,. \tag{1.170}$$

If the potential V has a simple expression, one can hope to get a rather 'simple' analytical expression for U and then deduce a (more complicated) expression for the density $\mu_A^{[c]}$ by computing the Stieltjes transform thanks to Eq. (1.169) and then using the Sokochi-Plemelj inversion formula of Eq. (1.28).

The function $U_A^{[c]}$ is for the high temperature regime what the polynomial $P_N^\star$ is for the low temperature regime and Eq. (1.170) is Eq. (1.138) for $P_N^\star$, up to the duality $c \leftrightarrow N$. It is convenient to think of the U-function as the most simple quantity describing the high-temperature regime. The high-low temperature duality can be summarized in the following table.

	Low-Temperature ensemble	High-temperature ensemble
limit	$\beta \to \infty$ with N fixed	$N \to \infty, \beta \to 0$ with $N\beta/2 \to c$
parameter	$N \in \mathbb{N}^*$	$c \in \mathbb{R}_+$
support of dist.	N atoms	unconstrained
'natural' object	$P_N^\star$, def. by (1.136), sol. of Eq. (1.138)	$U_A^{[c]}$, def. by (1.167), sol of Eq. (1.170)
Stieltjes	$g_N^* = (\log P_N^\star)'/N$, sol. of Eq. (1.135)	$g_A = -(\log U_A^{[c]})'/c$, sol. of Eq. (1.166)

Remark. For the re-scaled joint law of Eq. (1.162), the differential equation for the U-function of the corresponding limiting measure can be obtained by replacing $V \to \frac{c+1}{c} V$ in Eq. (1.170). ⌟

Remark *(Poisson statistics for high-temperature ensembles).* For high-temperature ensembles, one can show that the local behavior is given by Poisson statistics, see Ref. [24]. ⌟

1.7.2 Classical β-ensembles at high temperature

The previous section deals with high-temperature ensembles for a general potential V yet they were introduced in the case where V is the potential of a classical ensemble, and in particular, the first appearance of a high-temperature ensemble corresponds to the case where V is a harmonic potential, see Ref. [3].

For classical ensembles, one can show that the corresponding joint density is for any $\beta > 0$, the stationary measure of a N-dimensional stochastic process, which also depends explicitly on β. This is the Dyson Brownian motion (DBM) (which will be discussed in the following chapter) with an Ornstein-Uhlenbeck drift for the harmonic potential, the Wishart-Bru process (see Ref. [35]) for the Laguerre ensemble, and the Jacobi-Demni process (see Ref. [48]) for the Jacobi ensemble. One can then take $\beta_N = 2c/N$ and use Ito-calculus to derive a differential equation for the Stieltjes transform of the associated process. Taking a large time and large N limit of this equation will reproduce (up to rescaling depending on the convention) the non-linear differential equation (1.166), see Refs. [3, 4, 175].

Another way to tackle classical-ensembles in the high-temperature regime is to study the corresponding tridiagonal model of Sec. 1.4.4 under the scaling $\beta_N = 2c/N$ and $N \to \infty$, since the laws of these tridiagonal models are relatively simple. Quite remarkably, this is done (see Refs. [55, 176, 175]) by showing that the moments of those ensembles satisfy the duality $c \leftrightarrow N$, and using identities for the easier-to-compute 'low-temperature'-moments (that is the ones of the same ensembles but with $\beta \to \infty$ and N fixed), see also Ref. [69].

In the following, we give the limiting distribution for each classical ensemble. Importantly, our results are given for the re-scaled joint law of Eq. (1.162) (and hence are also re-scaled from the ones of the literature) such as we vary the parameter c, we change the shape of the distribution $\mu_A^{[c]}$ but not its scale.

Gaussian ensemble at high-temperature -

For eigenvalues taken from the Gaussian ensemble in the high-temperature regime with the following convention:

$$\boldsymbol{\lambda} \sim \tilde{\mathcal{P}}_G^{[c]}(\boldsymbol{\lambda}) \propto \mathrm{e}^{-(1+c)\sum_{i=1}^N \frac{\lambda_i^2}{2}} |\Delta(\boldsymbol{\lambda})|^{\frac{2}{cN}}\,, \tag{1.171}$$

the ESD of the λ_i converges as $N \to \infty$ to the ***c*-Gaussian distribution**, also known as *the Gauss-Wigner crossover* or *the Askey-Wimp-Kerov distribution*, with density given by:

$$\boxed{\mu_G^{[c]}(\lambda) := \frac{\sqrt{c+1}}{\sqrt{2\pi}\Gamma(c+1)} \frac{1}{\left|D_{-c}\left(\mathrm{i}\sqrt{c+1}\lambda\right)\right|^2}\,.} \tag{1.172}$$

The function D_{-c} in Eq. (1.172) is the *parabolic cylinder function*, which can be defined by its integral representation:

$$D_{-c}(z) := \frac{\mathrm{e}^{-z^2/4}}{\Gamma(c)} \int t^{c-1}\mathrm{e}^{-zt-t^2/2}\mathrm{d}t\,. \tag{1.173}$$

The c-Gaussian distribution is a centered symmetric distribution continuously interpolating between the density of the unit normal distribution at $c \to 0^+$ and the semi-circle distribution with $\sigma = 1$ at $c \to \infty$. Importantly, with our convention, the variance of the c-Gaussian is independent of c and is equal to one. A plot of the density of the c-Gaussian is given in Fig. 1.8.

The analytical expression for the density can be deduced by noticing in this case the derivative of the potential is a linear function, and so the differential equation satisfied by the U-function of the c-Gaussian $U_G^{[c]}(z) := \exp\left[-c\int \mu_G^{[c]}(\mathrm{d}\lambda)\log(z-\lambda)\right]$ is of *Kummer's type*. The solution is given by:

$$U_G^{[c]}(z) = A_2\,\Psi\left(\frac{c}{2},\frac{1}{2},-\frac{c+1}{2}\,z^2\right) = 2^{c/2}A_2\,\mathrm{e}^{-\frac{(c+1)z^2}{4}}D_{-c}\left(\mathrm{i}\sqrt{c+1}z\right)\,, \tag{1.174}$$

for some constant A_2 and Ψ is the *Tricomi (confluent hypergeometric) function*, which can be compactly defined by its integral representation:

$$\Psi(a,b,z) := \frac{1}{\Gamma(a)}\int_0^\infty \mathrm{e}^{zt}t^{a-1}(1+t)^{b-a-1}\mathrm{d}t \qquad (\text{for} \quad \mathfrak{Re}\,a>0)\,. \tag{1.175}$$

From there, one can easily get the Stieltjes transform $g_G^{[c]}(z) := \int \mu_G^{[c]}(\mathrm{d}\lambda)(z-\lambda)^{-1}$ with Eq. (1.169) and identity for the derivative of the Tricomi function. This gives:

$$g_G^{[c]}(z) = \frac{(c+1)z}{2}\frac{\Psi\left(1+\frac{c}{2},\frac{3}{2},-\frac{c+1}{2}\,z^2\right)}{\Psi\left(\frac{c}{2},\frac{1}{2},-\frac{c+1}{2}\,z^2\right)}\,. \tag{1.176}$$

The density of the c-Gaussian is then obtained by the Plemelj inversion formula Eq. (1.28) and identities for hypergeometric functions near their branch cut.

Laguerre ensemble at high-temperature -

For eigenvalues taken from the Laguerre ensemble at high-temperature with the following convention:

$$\boldsymbol{\lambda} \sim \tilde{\mathcal{P}}_L^{[c]}(\boldsymbol{\lambda}) \propto \mathrm{e}^{-(1+c)\sum_{i=1}^N\left(\frac{\lambda_i^2}{q}+(1-q^{-1})\log(\lambda_i)\right)}|\Delta(\boldsymbol{\lambda})|^{\frac{2}{cN}}\,, \tag{1.177}$$

the ESD of the λ_i converges as $N\to\infty$ to the ***c*-Laguerre distribution**, also known as *the Gamma-Wishart crossover* with density given by:

$$\boxed{\mu_{L(q)}^{[c]}(\lambda) := K_L\frac{\lambda^{(c+1)(q^{-1}-1)}\,e^{-\frac{c+1}{q}\lambda}}{\left|\Psi\left(c,(c+1)(1-q^{-1}),-\frac{c+1}{q}\,\lambda\right)\right|^2}\mathbb{I}_{(0,\infty)}\,.} \tag{1.178}$$

The constant K_L in Eq. (1.178) is given by:

$$K_L := \frac{\left(\frac{c+1}{q}\right)^{(c+1)(q^{-1}-1)}}{\Gamma(c+1)\Gamma\left(\frac{c+1}{q}\right)}\,. \tag{1.179}$$

The c-Laguerre distribution continuously interpolates between the $\mathrm{Gamma}(q^{-1}, q)$ distribution at $c \to 0^+$ and the Marčenko-Pastur distribution with shape ratio q at $c \to \infty$. Its mean and variance are independent of c and given respectively by 1 and q. A plot of the density of the c-Laguerre is given in Fig. 1.8.

Similarly to the Gaussian case, the differential equation satisfied by the U-function $U^{[c]}_{L(q)}(z) := \exp\left[-c \int \mu^{[c]}_{L(q)}(\mathrm{d}\lambda) \log(z-\lambda)\right]$ is of Kummer's type and the U-function is given by:

$$U^{[c]}_{L(q)}(z) = A_2 \, \Psi\left(c, (c+1)(1-q^{-1}), -\frac{c+1}{q} z\right) . \tag{1.180}$$

for some constant A_2.

Correspondingly, its Stieltjes transform $g^{[c]}_{L(q)}(z) := \int \mu^{[c]}_{L(q)}(\mathrm{d}\lambda)(z-\lambda)^{-1}$ is given by:

$$g^{[c]}_{L(q)}(z) = -\frac{c+1}{q} \frac{\Psi\left(c+1, 1+(c+1)(1-q^{-1}), -\frac{c+1}{q} z\right)}{\Psi\left(c, (c+1)(1-q^{-1}), -\frac{c+1}{q} z\right)} , \tag{1.181}$$

and inverting this Stieltjes transform gives the analytical expression of Eq. (1.178) for the density.

Jacobi ensemble at high-temperature -

For eigenvalues taken from the Jacobi ensemble at high temperatures with the following convention:

$$\boldsymbol{\lambda} \sim \tilde{\mathcal{P}}^{[c]}_J(\boldsymbol{\lambda}) \propto \mathrm{e}^{-(1+c)\sum_{i=1}^N \left((1-q_1^{-1})\log(\lambda_i) + (1-q_2^{-1})\log(1-\lambda_i)\right)} |\Delta(\boldsymbol{\lambda})|^{\frac{2}{cN}} , \tag{1.182}$$

the ESD of the λ_i converges as $N \to \infty$ to the ***c*-Jacobi distribution**, with density given by:

$$\boxed{\mu^{[c]}_{J(q_1,q_2)}(\lambda) = K_J \, \frac{\lambda^{(c+1)(q_1^{-1}-1)}(1-\lambda)^{(c+1)(q_1^{-1}-1)}}{|U(\lambda) + \mathrm{e}^{\mathrm{i}\pi(c+1)(q_2^{-1}-1)} V(\lambda)|^2} \mathbb{I}_{(0,1)} .} \tag{1.183}$$

The constant K_J is given by:

$$K_J = \frac{\Gamma(c+1)\Gamma\left((c+1)(q_1^{-1}+q_2^{-1}-1)+1)\right)}{\Gamma\left(\frac{c+1}{q_1}\right)\Gamma\left(\frac{c+1}{q_2}\right)} , \tag{1.184}$$

and for $a = (c+1)(q_1^{-1}-1)$ and $b = (c+1)(q_2^{-1}-1)$, the two functions U and V of Eq. (1.183) are defined by

$$U(x) := \frac{\Gamma(c+1)\Gamma(a+1)}{\Gamma(1+c+a)} {}_2F_1(c, -c-a-b-1, -a, x) , \tag{1.185}$$

$$V(x) := -\frac{\pi c \Gamma(c+a+b+2)\, x^{1+a}(1-x)^{1+b}}{\sin(\pi a)\Gamma(2+a)\Gamma(1+c+b)} {}_2F_1(1-c, 2+c+a+b, 2+a, x) . \tag{1.186}$$

The c-Jacobi distribution continuously interpolates between the $\mathrm{Beta}(q_1^{-1}, q_2^{-1})$ distribution at $c = 0$ and the Watcher distribution $\mathrm{Wat}(q_1, q_2)$ of Eq. (1.47) at $c \to \infty$, with a constant mean given by $\frac{q_1^{-1}}{q_1^{-1}+q_2^{-1}}$.

The differential equation for the U-function $U^{[c]}_{J(q_1,q_2)}(z) := \exp\left[-c \int \mu^{[c]}_{J(q_1,q_2)}(\mathrm{d}\lambda) \log(z-\lambda)\right]$ is up to a change of variable of *Euler's type* and we have:

$$U^{[c]}_{J(q_1,q_2)}(z) := A_2 z^{-c} {}_2F_1\left(c, \frac{c+1}{q_1}, (c+1)(q_1^{-1}+q_2^{-1}), \frac{1}{z}\right), \tag{1.187}$$

where

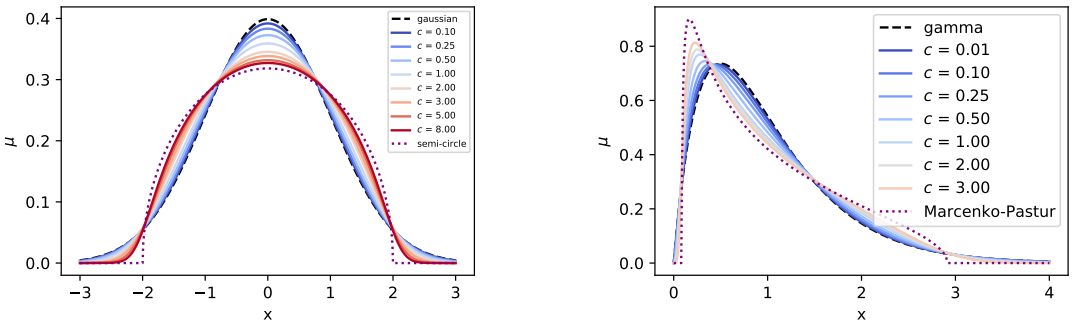
$${}_2F_1(a,b;c;z) := \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k\, k!} z^k, \tag{1.188}$$

is the *(Gauss) hypergeometric function* and $(a)_k := \Gamma(a+k)/\Gamma(a)$ is the *Pochhammer symbol*.

The Stieltjes transform $g^{[c]}_{J(q_1,q_2)}(z) := \int \mu^{[c]}_{J(q_1,q_2)}(\mathrm{d}\lambda)(z-\lambda)^{-1}$ is given by:

$$g^{[c]}_{J(q_1,q_2)}(z) = \frac{1}{z} \frac{{}_2F_1\left(c+1, \frac{c+1}{q_1}, (c+1)(q_1^{-1}+q_2^{-1}), \frac{1}{z}\right)}{{}_2F_1\left(c, \frac{c+1}{q_1}, (c+1)(q_1^{-1}+q_2^{-1}), \frac{1}{z}\right)}. \tag{1.189}$$

To conclude, let's mention that β-ensembles on the real line at high temperature have been recently shown to be connected to the hydrodynamical limit of the classical Toda chain with periodic conditions, see Refs. [163, 138].



(a) c-Gaussian **(b)** c-Laguerre

Figure 1.8: Plots of the c-Gaussian distribution of Eq. (1.172) $(\mathbf{a})$ and the c-Laguerre distribution of Eq. (1.178) $(\mathbf{b})$ for different values of the parameter c.

1.8 Summary and conclusion of Chapter 1

In this chapter, we have reviewed different results of random matrix theory and β-ensembles taken in different regimes. For β-ensembles in the classical regime where $\beta > 0$ is fixed and

$N \to \infty$, the limiting distribution is described by a competition between a confining potential V and a repulsive $2D$-Coulomb gas pairwise interaction. This Coulomb gas picture can also be used to derive the large deviation of the top eigenvalue of a β-ensemble. The other two regimes of β-ensembles: the low-temperature regime where $\beta \to \infty$ and N is fixed, and the high-temperature regime where $N \to \infty$ with $N\beta/2 \to c \geq 0$, share an intriguing high-low temperature duality $c \leftrightarrow N$ between the parameters of the two regimes.

Chapter 2

Sum and product of random matrices and spherical integrals

2.1 Introduction and notations

In the previous chapter, we reviewed some important results concerning the properties of *individual* random matrices. The goal of this chapter is to study the case where a matrix $\mathbf{C}$ is given as the sum or product of two (possibly but not necessarily random) matrices $\mathbf{A}$ and $\mathbf{B}$. Precisely, for $\beta \in \{1, 2, 4\}$ we will consider the three following cases:

- **The additive self-adjoint case** that is $\mathbf{C} = \mathbf{A} + \mathbf{B}$ with $\mathbf{A}, \mathbf{B} \in \mathrm{Herm}_\beta(N)$ two $(N \times N)$ self-adjoint matrices.
- **The multiplicative self-adjoint case**, that is $\mathbf{C} = \sqrt{\mathbf{A}}\mathbf{B}\sqrt{\mathbf{A}}$ where $\mathbf{A}, \mathbf{B} \in \mathrm{Herm}_\beta^{++}(N)$, that is $\mathbf{A}$ and $\mathbf{B}$ are two positive definite matrices (in particular their eigenvalues are all positives). The case of semi-definite matrices can be obtained as a limiting case but will not be discussed here for simplicity.
- **The additive rectangular case**, that is $\mathbf{C} = \mathbf{A} + \mathbf{B}$ with $\mathbf{A}, \mathbf{B} \in \mathrm{M}_{N,M}(\mathbb{K}_\beta)$, two $(N \times M)$ rectangular matrices with entries in $\mathbb{K}_\beta = \mathbb{R}, \mathbb{C}, \mathbb{H}$ for $\beta = 1, 2, 4$.

For each of these three cases, our goal is to describe the spectrum (resp. the singular values in the rectangular case) of the matrix $\mathbf{C}$ based on the knowledge of the spectrum (resp. singular values) of the matrices $\mathbf{A}$ and $\mathbf{B}$ and possible additional 'information' - or *model* on the structure of $\mathbf{A}$ and $\mathbf{B}$. For example, what can be said if we assume one of the matrices (say $\mathbf{A}$ for example) to be invariant in law? or to be additionally of low rank? what happens in the large N limit?

In order to lighten notations, we will often denote by $\boldsymbol{c} = (c_1, \ldots, c_N) \equiv \boldsymbol{\lambda}(\mathbf{C})$ (and resp. $\boldsymbol{c} \equiv \boldsymbol{s}(\mathbf{C})$ in the rectangular case) the vector of eigenvalues (resp. singular values) of the matrix $\mathbf{C}$ and similarly $\boldsymbol{a}, \boldsymbol{b}$ denote the set of eigenvalues/singular values of the matrices $\mathbf{A}, \mathbf{B}$.

Before jumping to the problem, let's mention several important remarks.

Remark *(Symmetrized product and usual product of matrices)***.** We have denoted the multiplicative case as the *symmetrized product* $\mathbf{C} = \sqrt{\mathbf{A}}\mathbf{B}\sqrt{\mathbf{A}}$ instead of the usual matrix product $\tilde{\mathbf{C}} = \mathbf{A}\mathbf{B}$. This is because if $\mathbf{A}$ and $\mathbf{B}$ are self-adjoint (positive definite) matrices,

then the matrix resulting from their symmetrized product is also a **self-adjoint** matrix while the result from the usual matrix product is not. However, because the two matrices $\mathbf{C}$ and $\tilde{\mathbf{C}}$ are *similar*, they have the same set of eigenvalues. As our concerns in this thesis lie only in the behavior of the eigenvalues (and not the eigenvectors), for the product of **two** matrices, we can use either convention. ⌟

Remark *(the additive case as a limit of the multiplicative case)*. Every positive definite matrix $\mathbf{A} \in \mathrm{Herm}_\beta^{++}(N)$ can be written as the exponential of a self-adjoint matrix $\mathbf{A} = \exp[\mathbf{A}_0]$, where the matrix $\mathbf{A}_0$ is simply a matrix in the same basis of $\mathbf{A}$ but with eigenvalues $\log(\boldsymbol{a}) \equiv (\log(a_1), \dots, \log(a_N))$. Note that for every $i = 1, \dots, N$ we have $a_i > 0$ for a matrix $\mathbf{A} \in \mathrm{Herm}_\beta^{++}(N)$ and there is no issue with the logarithm function. We can similarly define the matrix $\mathbf{B}_0$ and $\tilde{\mathbf{C}}_0$ and the usual matrix product $\tilde{\mathbf{C}} = \mathbf{AB}$ can be written in terms of the self-adjoint matrices $\mathbf{A}_0, \mathbf{B}_0, \tilde{\mathbf{C}}_0$ as:

$$\exp[\tilde{\mathbf{C}}_0] = \exp[\mathbf{A}_0] \exp[\mathbf{B}_0] \,. \tag{2.1}$$

Now in general $\mathbf{A}_0$ and $\mathbf{B}_0$ do not commute and $\exp[\mathbf{A}_0] \exp[\mathbf{B}_0] \neq \exp[\mathbf{A}_0 + \mathbf{B}_0]$. The matrix $\tilde{\mathbf{C}}_0$ is given in terms of $\mathbf{A}_0$ and $\mathbf{B}_0$ by the non-trivial Baker-Campbell-Hausdorff formula, see Refs. [15, 38]. However, if we now add a *tunable parameter* ϵ and perform the re-scaling $(\mathbf{A}_0, \mathbf{B}_0, \tilde{\mathbf{C}}_0) \to (\epsilon\mathbf{A}_0, \epsilon\mathbf{B}_0, \epsilon\tilde{\mathbf{C}}_0)$ then at first order in ϵ we have the simple relation:

$$\exp[\epsilon\tilde{\mathbf{C}}_0] = \exp\left[\epsilon(\mathbf{A}_0 + \mathbf{B}_0) + o(\epsilon)\right] \,. \tag{2.2}$$

In other words, one can think of the sum of matrices as a limiting case of the product of matrices. We will encounter this property in several instances of this thesis, where quantities related to the spectrum of the sum of two matrices can be seen as the limiting case of their multiplicative counterpart. ⌟

Remark *(Unit/quantum multiplicative case)*. The multiplicative self-adjoint problem is closely related to another problem that will *not* be discussed in this thesis, which concerns the product of matrices of the form $\mathbf{W} = \mathbf{UV}$ where $\mathbf{U}, \mathbf{V}$ (and hence $\mathbf{W}$) are in $\mathsf{O}_\beta(N)$. The eigenvalues of matrices in $\mathsf{O}_\beta(N)$ are on the unit (complex) circle $\mathbb{T} := \{z \in \mathbb{C} \text{ s.t. } |z| = 1\}$, and one may ask how the eigenvalues of $\mathbf{W}$ depends on the ones of $\mathbf{U}, \mathbf{V}$. ⌟

The rest of this chapter is organized as follows: In Sec. 2.2, we introduce a toy model concerning the norm of the sum of two randomly rotated vectors. This simple example will turn out to be useful to understand the matrix analog developed in the rest of the chapter. In Sec. 2.3, we briefly describe *Horn's problem*, that is, the description of the set of all possible values of the sum of two self-adjoint matrices with given spectra. In Sec. 2.4, we look at the simple model for the sum/product of random matrices, where one performs a rank-one perturbation of a given matrix. In Sec. 2.5, we describe another simple model of the sum/product of matrices where one constructs a matrix process made of infinitesimal perturbation. In Sec. 2.6, we tackle the generic case and its large N limit, describes by the famous *free convolution*. In Sec. 2.7, we describe the finite free convolution, which can be seen as the low-temperature counterpart ($\beta \to \infty$ with N fixed) of the previous free convolution.

2.2 A heuristic toy model: the norm of the sum of two vectors

Let's look at a simple instructive vector problem, where we have two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ belonging to $\mathbb{R}^N$ but one only knows their norms given respectively by $x_0 > 0$ and $y_0 > 0$. We then construct a third vector $\boldsymbol{z}$ made of the sum of these two vectors:

$$\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}\,, \tag{2.3}$$

and ask the following simple question:

What can we say about the norm z_0 of the vector $\boldsymbol{z}$?

In this *worst case scenario*, by which we mean that one does not have any other prior information on the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, thanks to the triangle inequality, one only knows that z_0 belongs to the segment:

$$\Omega_{\rm v} \equiv \Omega_{\rm v}(x_0, y_0) := \{z \in \mathbb{R}_+ \quad \text{s.t} \quad |x_0 - y_0| \leq z \leq x_0 + y_0\} = \mathbb{I}_{[|x_0-y_0|, x_0+y_0]}\,, \tag{2.4}$$

where the index $._{\rm v}$ indicates quantities referring to this vector toy model.

Let's now assume that one has *additional information* on the law of the relative position vectors. The most natural example one can think of is a *random model* where one of the vectors is in an arbitrary position with respect to the other or more explicitly:

- the direction of one of the vectors (say $\boldsymbol{x}$) is taken uniformly at random that is

$$\boldsymbol{x} = x_0\boldsymbol{\sigma} \quad \text{with} \quad \boldsymbol{\sigma} \sim \mathrm{Unif}\left[\mathbb{S}^{N-1}\right]\,, \tag{2.5}$$

- while the direction of the other vector is fixed

$$\boldsymbol{y} = y_0\boldsymbol{e}_1 \quad \text{with} \quad \boldsymbol{e}_1 = (1, 0, \ldots, 0)\,. \tag{2.6}$$

The goal is then to describe the law $p_{\rm v}(z_0)$ which is supported on $\Omega_{\rm v}$ given by Eq. (2.4) of the norm of $\boldsymbol{z}$. Note that since we are only interested in the norm z_0, one can take $\boldsymbol{y}$ to be equal to y_0 time any unit vector without changing the law $p_{\rm v}(.)$. The description of the law will be done by two different methods, as both are instructive for the matrix model described later on.

The most straightforward way to get the density is simply to compute the change of variable from the $\boldsymbol{\sigma}$ to z_0:

$$z_0 = \|\boldsymbol{x} + \boldsymbol{y}\|\,, \tag{2.7}$$

$$z_0 = \sqrt{x_0^2 + y_0^2 + 2x_0y_0\,\boldsymbol{\sigma}^{\mathsf{T}}\boldsymbol{e}_1}\,. \tag{2.8}$$

Next, the law of the first coordinate $\sigma_1 := \boldsymbol{\sigma}^{\mathsf{T}}\boldsymbol{e}_1 \sim \mu$ is given by:

$$\mu(x) = \frac{(1-x^2)^{N/2-1}}{B\left(\frac{N}{2}, \frac{1}{2}\right)}\mathbb{I}_{[-1,1]}\,, \tag{2.9}$$

where $B(.,.)$ is the Beta function. Note that the square of a random variable distributed according to Eq. (2.9) is a beta random variable $\mathrm{Beta}(a,b)$ with parameters $a = 1/2$ and $b = (N-1)/2$. One fancy way to describe the norm z_0 is then to say that it is equal in law to the solution of the random equation:

$$z_0 - \sqrt{x_0^2 + y_0^2 + 2x_0 y_0\, d} = 0\,, \tag{2.10}$$

with $d \sim \mu$ given by Eq. (2.9). If we now perform the change of variable given by Eq. (2.8) and compute the Jacobian associated with it, we get the following law:

$$p_{\mathrm{v}}(z_0) = \frac{1}{C} \cdot \left(z_0^2 - |x_0 - y_0|^2\right)^{N/2-1/2} \left((x_0+y_0)^2 - z_0^2\right)^{N/2-1/2} \mathbb{I}_{[|x_0-y_0|, x_0+y_0]}\,, \tag{2.11}$$

with the normalization constant $C = 2^{N-2}\, B(N/2, 1/2)\, (x_0 y_0)^{N-1}$. A plot of the density is given in Fig. 2.1. As $N \to \infty$, the distribution μ of Eq. (2.9) converges to the Dirac distribution at zero, this is the well-known high-dimensional phenomenon where a rotationally invariant vector is almost surely orthogonal to any other (independent) vector in the large N limit. If we use this in the random equation (2.10) we immediately get the following large N asymptotic:

$$z_0 \underset{N\to\infty}{\to} \sqrt{x_0^2 + y_0^2}\,, \tag{2.12}$$

which is nothing else than the Pythagorean theorem.

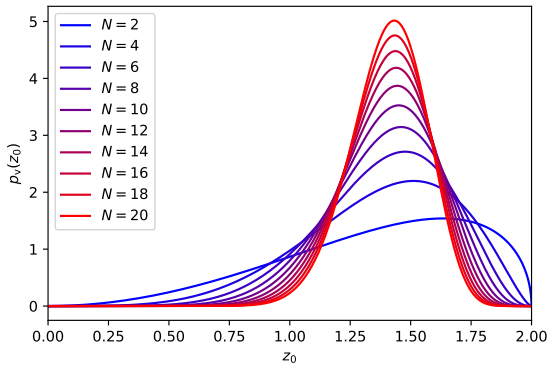


Figure 2.1: A plot of the density of the norm of the sum of two rotationally invariant vectors with norms $x_0 = y_0 = 1$ for different values of the dimension N.

Now let's look at the same problem but from a different angle. Without loss of generality, we can replace $\boldsymbol{y} = y_0 \boldsymbol{e}_1$ by $\boldsymbol{y} = y_0 \boldsymbol{\sigma'}$ where $\boldsymbol{\sigma'} \sim \mathrm{Unif}\left[\mathbb{S}^{N-1}\right]$ without changing the law $p_{\mathrm{v}}(z_0)$. Clearly, by rotationally invariance of $\boldsymbol{x}$ and $\boldsymbol{y}$, $\boldsymbol{z}$ is also of the form $\boldsymbol{z} = z_0 \boldsymbol{\sigma''}$ with $\boldsymbol{\sigma''} \sim \mathrm{Unif}\left[\mathbb{S}^{N-1}\right]$. By classical harmonic analysis, since $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent, we have multiplication of the Fourier transforms (or characteristic function):

$$\mathcal{F}_{\boldsymbol{z}}(\boldsymbol{t}) = \mathcal{F}_{\boldsymbol{x}}(\boldsymbol{t}) \mathcal{F}_{\boldsymbol{y}}(\boldsymbol{t})\,, \tag{2.13}$$

with the Fourier transform given by:

$$\mathcal{F}_{\boldsymbol{q}}(\boldsymbol{t}) := \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\boldsymbol{q}^{\mathsf{T}}\boldsymbol{t}}\right] = \mathrm{E}\left[\mathrm{e}^{\mathrm{i}\sum_{k=0}^{N} q_k t_k}\right]\,. \tag{2.14}$$

If we first look at the RHS of Eq. (2.13), the Fourier transform of $\boldsymbol{x}$ is simply an average over the angular coordinate which can be expressed in terms of Bessel functions:

$$\mathcal{F}_{\boldsymbol{x}}(\boldsymbol{t}) = \int_{\mathbb{S}^{N-1}} \mathrm{e}^{\mathrm{i}\, x_0 \boldsymbol{\sigma}^\intercal \boldsymbol{t}} \mu_{\mathrm{Unif}}(\mathrm{d}\boldsymbol{\sigma}) \tag{2.15}$$

$$\mathcal{F}_{\boldsymbol{x}}(\boldsymbol{t}) = 2^{N/2-1}\Gamma(N/2)\,(x_0\|\boldsymbol{t}\|)^{1-N/2} J_{N/2-1}(x_0\|\boldsymbol{t}\|)\,, \tag{2.16}$$

where $J_\nu(x) := \sum_{m=0}^{\infty} \frac{(-1)^m}{m!\Gamma(m+\nu+1)} \left(\frac{x}{2}\right)^{2m+\nu}$ is *the Bessel function of the first kind* of order ν, and similarly for the vector $\boldsymbol{y}$.

In order to ease notation, let's denote the **angular integral** obtained as the previous Fourier transform:

$$I_{\mathrm{v}}(\lambda, t) := 2^{N/2-1}\Gamma(N/2)\,(\lambda t)^{1-N/2} J_{N/2-1}(\lambda t)\,. \tag{2.17}$$

Now, on the one hand, the distribution f of the *vector* $\boldsymbol{z}$ (not its norm!) is given by an inverse radial Fourier transform which can be also expressed in terms of Bessel functions:

$$f(\boldsymbol{z}) = \frac{1}{(2\pi)^{N/2}} \int \mathrm{e}^{-\mathrm{i}\,\boldsymbol{t}^\intercal \boldsymbol{z}} \left(\mathcal{F}_{\boldsymbol{x}}(\boldsymbol{t})\mathcal{F}_{\boldsymbol{y}}(\boldsymbol{t})\right) \mathrm{d}\boldsymbol{t}\,, \tag{2.18}$$

$$f(\boldsymbol{z}) = \frac{1}{(2\pi)^{N/2}} \int \mathrm{e}^{-\mathrm{i}\,\boldsymbol{t}^\intercal \boldsymbol{z}} I_{\mathrm{v}}\left(x_0, \|\boldsymbol{t}\|\right) I_{\mathrm{v}}\left(y_0, \|\boldsymbol{t}\|\right) \mathrm{d}\boldsymbol{t}\,, \tag{2.19}$$

$$f(\boldsymbol{z}) = \int_0^\infty \frac{t^{N/2}}{\|\boldsymbol{z}\|^{N/2-1}} J_{N/2-1}(\|\boldsymbol{z}\|t) I_{\mathrm{v}}\left(x_0, t\right) I_{\mathrm{v}}\left(y_0, t\right) \mathrm{d}t\,. \tag{2.20}$$

On the other hand, the law $p_{\mathrm{v}}(.)$ of the norm z_0 of the rotationally invariant vector $\boldsymbol{z}$ is given in terms of the vector distribution $f(.)$ by the Jacobian of the spherical change of coordinate and reads:

$$f(\boldsymbol{z}) = \frac{\Gamma(N/2)}{2\pi^{N/2}} \frac{1}{\|\boldsymbol{z}\|^{N-1}} p_{\mathrm{v}}(\|\boldsymbol{z}\|)\,, \tag{2.21}$$

where $\frac{\Gamma(N/2)}{2\pi^{N/2}}$ is the volume of the unit sphere. All in all, this give the following (Bessel) *integral representation* for the law of z_0:

$$p_{\mathrm{v}}(z_0) = \frac{z_0^{N-1}}{C_N} \int I_{\mathrm{v}}(x_0, t) I_{\mathrm{v}}(y_0, t) I_{\mathrm{v}}(z_0, t) t^{N-1} \mathrm{d}t\,, \tag{2.22}$$

with $C_N := 2^{N-2}\Gamma(N/2)^2$. Using *Sonine's formula* (see for example Ref. [187]), this integral representation can be simplified to give the expression of Eq. (2.11). Note that this harmonic approach can be easily generalized to the sum of an arbitrary number of rotationally invariant vectors.

For each of the three operations, there exists a dictionary between the matrix operation and this toy model:

- The norms x_0, y_0, z_0 of the vectors $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ in this toy model are replaced by the set of eigenvalues (or singular values) $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ of the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and similarly the spherical change of coordinate is replaced by Weyl's formula for eigenvalues (or singular values). The notion of a rotationally invariant vector is replaced by the notation of β-invariance in law as described in the previous chapter.

- The segment $\Omega_{\rm v}$ of all possible values of the norm of the sum of two vectors in this toy model is replaced by the *Horn convex hull* briefly discussed in the next section.
- The angular integral $I_{\rm v}$ obtained as an average of the Fourier transform over the unit sphere is replaced by spherical integrals corresponding to an average over the group $\mathsf{O}_\beta(N)$ (and also $\mathsf{O}_\beta(M)$ in the rectangular case) of the Fourier transform associated to the underlying operation.
- The property of almost sure orthogonality of rotationally-invariant vectors in the large N is replaced by the notion of *asymptotic freeness* for β-invariant matrices. Just like one computes the norm of the sum of orthogonal vectors by the simple Pythagorean theorem of Eq. (2.12), one can compute the LSD of the sum and product of free matrices thanks to the free probability transforms described in Sec. 2.6.3.

There is however a crucial difference in the random matrix setting: while one can obtain an analog of the integral representation of Eq. (2.22) for the joint density associated with each operation, there is in general no known analog of Sonine's formula to simplify it.

2.3 A few words on Horn's problem

Let's start with the *worst case scenario*, where one knows the spectrum $\boldsymbol{a}, \boldsymbol{b}$ of two matrices $\mathbf{A}, \mathbf{B} \in \mathrm{Herm}_\beta(N)$ but does *not* have any other prior information on $\mathbf{A}, \mathbf{B}$. What can one says about the unknown spectrum $\boldsymbol{c}$ of $\mathbf{C} = \mathbf{A} + \mathbf{B}$? Note that the multiplicative/rectangular case will be discussed shortly after. In other words, one wants to find a set of *relations* constraining the set of possible values of the sum of two self-adjoint matrices. Since we are only interested in the eigenvalues, this is also equivalent to finding all acceptable values of the eigenvalues of $\mathbf{Diag}(\boldsymbol{a}) + \mathbf{V}\mathbf{Diag}(\boldsymbol{b})\mathbf{V}^*$ for *all* $\mathbf{V} \in \mathsf{O}_\beta(N)$.

The simplest relation one can think of is the linearity of the trace, $\mathrm{Tr}\,(\mathbf{A}+\mathbf{B}) = \mathrm{Tr}\,\mathbf{A} + \mathrm{Tr}\,\mathbf{B}$ which constrained the set of possible vectors $\boldsymbol{c}$ according to:

$$\sum_{i=1}^N c_i = \sum_{i=1}^N (a_i + b_i)\,. \tag{2.23}$$

Next, another idea is to use the sup norm characterization of the top eigenvalue:

$$a_1 = \sup_{\boldsymbol{x}\in\mathbb{K}_\beta^N, \|x\|=1} \boldsymbol{x}^* \mathbf{A} \boldsymbol{x}\,, \tag{2.24}$$

which constrained the top eigenvalue of the sum to be below the sum of the top eigenvalues:

$$c_1 \leq a_1 + b_1\,. \tag{2.25}$$

The famous *Weyl's inequalities* provides another set of inequalities constraining each eigenvalue c_i individually:

$$a_j + b_k \leq c_i \leq a_r + b_s \qquad \text{for} \quad j + k - N \geq i \geq r + s - 1\,. \tag{2.26}$$

Now there are many other complex inequalities (for example the Ky Fan inequalities and the Lidskii-Wietland inequalities to cite a few, see for example Ref. [72]), and a natural question is to know 'when does it stop' and if this type of linear inequalities completely describes the set of possible eigenvalues of the sum of two matrices. Interested in the development of this problem, A. Horn conjectured in 1962 in Ref. [92], that the eigenvalues of the sum of two Hermitian ($\beta = 2$) satisfy the trace constraint and a set of *linear inequalities* of the form

$$\Omega \equiv \Omega(\boldsymbol{a}, \boldsymbol{b}) := \left\{ \boldsymbol{c} \in \mathbb{R}^N_{\geq} \quad \text{s.t} \quad \sum_{i \in I} c_i \leq \sum_{j \in J} a_j + \sum_{k \in K} b_k \quad \text{for} \quad (I, J, K) \in T^{(N)} \right.$$
$$\left. \text{and} \quad \sum_{i=1}^{N} c_i = \sum_{j=1}^{N} a_j + b_j \right\}, \quad (2.27)$$

where the *admissible sets* $T^{(N)}$ can be defined by a sophisticated induction but is not described here for simplicity. Importantly, Horn further conjectured the set of all these linear inequalities for the vector $\boldsymbol{c}$ forms a *convex polyhedron* - now known as **the Horn convex hull** - and this convex polyhedron completely describes the set of all possible eigenvalues of the matrix $\mathbf{C}$.

It took several decades to prove Horn's conjecture and the first real breakthrough was made by A. Klaychko who proved in Ref. [102] that this conjecture can be reduced to another conjecture (the so-called *saturation conjecture*) related to *Littlewood-Richardson coefficients* and the latter conjecture was then proved in Ref. [103] by A. Knutson and T. Tao by introducing a mathematical object known as the *honeycomb model*. The details of the original proof used advanced tools from algebraic geometry and representation theory and are far outside the scope of this thesis. Let's mention however that the latter result can be generalized to $\beta = 1, 4$ and the corresponding Horn convex hull is the same.

Similar to the additive case, one can ask if there is a multiplicative counterpart and a rectangular counterpart of the Horn convex Hull and the answer is positive, and we refer the reader to Ref. [72] for the details. In the rectangular case, the first two simplest relations are again given by Eq. (2.23) and (2.24) while for the multiplicative case, the trace equality is replaced by the determinant equality which gives:

$$\prod_{i=1}^{N} c_i = \prod_{j=1}^{N} (a_j \, b_j) \,, \quad (2.28)$$

and the supnorm inequality gives:

$$c_1 \leq a_1 \, b_1 \,. \quad (2.29)$$

2.4 Rank-one perturbation and BBP-phase transition

2.4.1 Introduction

In this section we consider the case of a rank-one perturbation of a matrix, that is:

- $\mathbf{A} = \gamma \boldsymbol{v}\boldsymbol{v}^*$ with $\boldsymbol{v} \in \mathbb{K}_\beta^N$ and $\|\boldsymbol{v}\|^2 = 1$ for the (self-adjoint) additive case,
- $\mathbf{A} = \mathbf{I} + \gamma \boldsymbol{v}\boldsymbol{v}^*$ for the multiplicative case,
- $\mathbf{A} = \gamma \boldsymbol{v}\boldsymbol{u}^*$ for the (additive) rectangular case.

Without loss of generality, we will assume $\gamma > 0$ (resp. $\gamma > 1$ in the multiplicative case), since one retrieves the case $\gamma < 0$ (resp. $0 < \gamma < 1$) by setting $\mathbf{B} \to -\mathbf{B}$ (resp. $\mathbf{B} \to \mathbf{B}^{-1}$). We start with the description of the additive case in the generic setting where do not assume a model for the description of the vector of perturbation $\boldsymbol{v}$, we will then quickly consider the β-invariant setting.

It should be mentioned that in statistics, the problem is usually looked at the other way: the rank-one matrix corresponds to a *planted signal* corrupted by a noise matrix $\mathbf{B}$ and the goal is to retrieve this signal, see for example Ref. [110] and references therein.

In the following to differentiate each operation, the joint density for the eigenvalues of the random (self-adjoint) additive perturbation will be denoted by $\mathcal{P}^{(\beta)}(\cdot|\gamma, \boldsymbol{b})$, the one for the product will denote by $\mathcal{P}_\times^{(\beta)}(\cdot|\gamma, \boldsymbol{b})$ and the one for the singular values of the rectangular case will be denoted by $\mathcal{P}_q^{(\beta)}(\cdot|\gamma, \boldsymbol{b})$.

2.4.2 Additive rank-one perturbation

In the case of rank-one additive perturbation, Horn inequalities reduce to Weyl's inequalities and the trace constraint, that is the set $\boldsymbol{c}$ of decreasing eigenvalues of $\mathbf{C} = \mathbf{B} + \gamma \boldsymbol{v}\boldsymbol{v}^*$ *interlaces* between the eigenvalues of $\mathbf{B}$:

$$\Omega^{\mathrm{rk1}} \equiv \Omega^{\mathrm{rk1}}(\gamma, \boldsymbol{b}) := \Big\{ \boldsymbol{c} \in \mathbb{R}^N \quad \text{s.t} \quad b_N \leq c_N \leq b_{N-1} \leq \cdots \leq c_1 \leq b_1 + \gamma \quad \text{and} \quad \sum_{i=1}^N c_i = \gamma + \sum_{i=1}^N b_i \Big\}, \tag{2.30}$$

where b_i are the eigenvalues of $\mathbf{B}$ in decreasing order. Now to understand the behavior of the set $\boldsymbol{c}$, we will use the following result known as the Sherman-Morrison formula, to compute the inverse of a rank-one update of an invertible matrix:

Sherman-Morrison formula: *if $\mathbf{M} \in \mathsf{M}_{N,N}(\mathbb{K}_\beta)$ is invertible and $\boldsymbol{u}_0, \boldsymbol{v}_0 \in \mathbb{K}_\beta^N$ then $(\mathbf{M} + \boldsymbol{u}_0\boldsymbol{v}_0^*)$ is invertible if and only if $1 + \boldsymbol{v}_0^*\mathbf{M}^{-1}\boldsymbol{u}_0 \neq 0$ and in this case it is given by:*

$$(\mathbf{M} + \boldsymbol{u}_0\boldsymbol{v}_0^*)^{-1} = \mathbf{M}^{-1} - \frac{\mathbf{M}^{-1}\boldsymbol{u}_0\boldsymbol{v}_0^*\mathbf{M}^{-1}}{1 + \boldsymbol{v}_0^*\mathbf{M}^{-1}\boldsymbol{u}_0}. \tag{2.31}$$

If we take the determinant of this equation and invert it we have:

$$\det(\mathbf{M} + \boldsymbol{u}_0\boldsymbol{v}_0^*) = \left(1 + \boldsymbol{v}_0^*\mathbf{M}^{-1}\boldsymbol{u}_0\right) \cdot \det \mathbf{M}, \tag{2.32}$$

Now setting $\mathbf{M} = z\mathbf{I} - \mathbf{B}$, $\boldsymbol{u}_0 = -\gamma\boldsymbol{v}$, $\boldsymbol{v}_0 = \boldsymbol{v}$ and denote by $\mathbf{V}$ the eigenmatrix of $\mathbf{B}$, we get the following relation between the characteristic polynomial of $\mathbf{C}$ and the one of $\mathbf{B}$:

$$\det(z\mathbf{I} - \mathbf{C}) = \left(1 - \gamma \sum_{i=1}^N \frac{|\boldsymbol{e}_i^*\mathbf{V}\boldsymbol{v}|^2}{z - b_i}\right) \cdot \det(z\mathbf{I} - \mathbf{B}), \tag{2.33}$$

with $\boldsymbol{e}_i$ the vector with 1 on the i^{th} coordinate and zero elsewhere. Since the characteristic polynomial cancels for values of z equal to the eigenvalues of its matrix we have that the eigenvalues c_i of $\mathbf{C}$ which are not eigenvalues of $\mathbf{B}$ are solutions of the following *secular equation*:

$$1 - \gamma \sum_{j=1}^{N} \frac{\left|\boldsymbol{e}_j^* \mathbf{V} \boldsymbol{v}\right|^2}{c_i - b_j} = 0 \qquad (\text{for } c_i \notin \{\boldsymbol{b}\}). \tag{2.34}$$

This describes the general setting where one does not have any prior information on $\mathbf{B}$ or its rank-one perturbation. Now as in the toy model, we move to a random version of this problem where one of the matrices is β-invariant in law, which in the case of rank-one perturbation amounts to studying the following problem:

The **random rank-one additive Horn problem** is defined by

$$\mathbf{C} = \mathbf{Diag}(\boldsymbol{b}) + \gamma \, \boldsymbol{\sigma}\boldsymbol{\sigma}^* \,, \tag{2.35}$$

where $\gamma > 0$ and $\boldsymbol{\sigma} \sim \text{Unif}\left[\mathbb{S}_\beta^{N-1}\right]$. We recall that for $\beta = 1$, $\mathbb{S}_{\beta=1}^{N-1} \equiv \mathbb{S}^{N-1}$ and for $\beta = 2, 4$, $\mathbb{S}_\beta^{N-1}$ is respectively the complex and quaternionic sphere. We will also assume the spectrum of $\mathbf{Diag}(\boldsymbol{b})$ to be simple, that is $b_N < \cdots < b_1$.

Note again that we have fixed the matrix $\mathbf{B} \equiv \mathbf{Diag}(\boldsymbol{b})$ to be diagonal without loss of generality for the eigenvalue problem since if $\mathbf{B}$ is not diagonal, we can always absorb its eigenmatrix in $\boldsymbol{\sigma}$ by rotationally invariance. Next, to describe the law of $\boldsymbol{c}$ we need the law of the square of the components of $\boldsymbol{\sigma}$.

If $\boldsymbol{\sigma} \sim \text{Unif}\left[\mathbb{S}_\beta^{N-1}\right]$ then the vector $(|\sigma_1|^2, \ldots, |\sigma_N|^2) \sim \text{Dir}(\beta/2)$ where $\text{Dir}(\alpha)$ denotes the **Dirichlet distribution** with uniform parameters $(\alpha, \ldots, \alpha)$ with density given over the simplex $\Delta^N := \{\boldsymbol{x} \in \mathbb{R}^N \text{ s.t } 0 \leq x_i \leq 1 \text{ for } i = 1, \ldots, N \text{ and } \sum x_i = 1\}$ by:

$$\mu_{\text{Dir}(\alpha)}(\boldsymbol{x}) := \frac{\Gamma(\alpha)^N}{\Gamma(N\alpha)} \prod_{i=1}^{N} x_i^{\alpha-1} \mathbb{I}_{\Delta^N} \,. \tag{2.36}$$

From this result and Eq.(2.37) we have a complete characterization for this problem.

Result 2.1 (*random secular equation for the rank-one additive problem*)

The eigenvalues $\boldsymbol{c}$ of the matrix $\mathbf{C}$ given by the random rank-one additive Horn problem of Eq. (2.35) are equal in law to the solutions with unknown c_i of the following ***random secular equation***

$$1 - \gamma \sum_{j=1}^{N} \frac{d_j}{c_i - b_j} = 0 \,, \tag{2.37}$$

where $\boldsymbol{d} \sim \text{Dir}(\beta/2)$.

Note that while Eq. (2.37) has been derived for $\beta \in \{1, 2, 4\}$, solutions of Eq. (2.37) can be naturally extended for any $\beta > 0$ even if we do not have a model for the (additive) rank-one perturbation of a matrix for $\beta \notin \{1, 2, 4\}$.

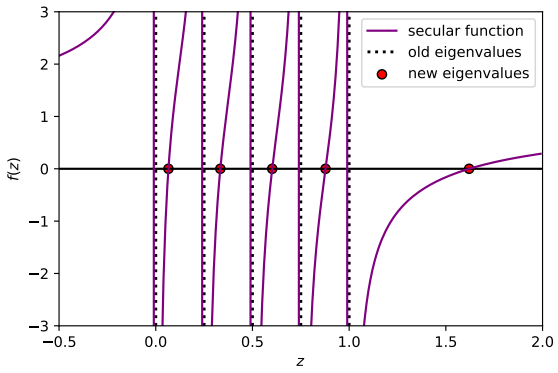


Figure 2.2: Illustration of the solution of a secular equation (2.37) for $N = 5$ and the positions of the b_i's are equidistributed between 0 and 1 (represented by dotted vertical lines). The new eigenvalues $c_1, \ldots, c_N$ (in red) are the zeros of the function $f(z) := 1 - \frac{\gamma}{N} \sum_{j=1}^{N} \frac{d_j}{z - b_j}$, where $\boldsymbol{d} \sim \mathrm{Dir}(1/2)$.

Joint density -

Next, we briefly describe how one can get the joint density $\mathcal{P}^{(\beta)}(.|\gamma, \boldsymbol{b})$ of the eigenvalues $\boldsymbol{c}$ of the rank-one perturbation from Eq. (2.37) and refer to Ref. [67] for details. The LHS of Eq. (2.37) for c_i replaced by $\lambda \in \mathbb{C}$ can be expressed as a ratio of polynomials:

$$1 - \gamma \sum_{j=1}^{N} \frac{d_j}{\lambda - b_j} = \frac{\prod_{i=1}^{N} (\lambda - c_i)}{\prod_{j=1}^{N} (\lambda - b_j)}, \tag{2.38}$$

and if one evaluates the residue at $\lambda = b_j$ of this equation, one gets the following formula for the change of variable from $\boldsymbol{d}$ to $\boldsymbol{c}$:

$$-\gamma d_j = \frac{\prod_{i=1}^{N} (b_j - c_i)}{\prod_{i=1 | i \neq j}^{N} (b_j - b_i)}. \tag{2.39}$$

Next, the Jacobian of this change of variable can be expressed as a modified *Cauchy double alternant determinant*:

$$\left| \left[\frac{\partial \boldsymbol{d}}{\partial \boldsymbol{c}} \right] \right| = \left| \det \left[\frac{1}{b_j - c_i} - \frac{1}{b_j - c_N} \right]_{i,j=1}^{N-1} \right|. \tag{2.40}$$

Now if ones perform the change of variable given by Eq. (2.39) in Eq. (2.36) for $\alpha = \beta/2$ and then uses known identities for the Cauchy double alternant evaluation given by Eq. (2.40), one gets after simplification the following result for the joint law

Result 2.2 (*joint density for the additive rank-one perturbation [190, 67]*)

The joint density of the eigenvalues $\boldsymbol{c}$ of the matrix $\mathbf{C}$ given by the random rank-one additive Horn problem of Eq. (2.35) is given by:

$$\mathcal{P}^{(\beta)}(\boldsymbol{c}|\gamma, \boldsymbol{b}) = \frac{\Gamma\left(\frac{N\beta}{2}\right)}{\Gamma(\beta/2)^N \gamma^{\frac{N\beta}{2}-1} |\Delta(\boldsymbol{b})|^{\beta-1}} |\Delta(\boldsymbol{c})| \prod_{j=1}^{N-1} \prod_{p=1}^{N} |c_j - b_p|^{\frac{\beta}{2}-1} \mathbb{I}_{\Omega^{\mathrm{rk1}}} \tag{2.41}$$

where we recall that Ω^{rk1} is defined by Eq. (2.30).

Let's mention again that this joint law can be naturally defined for any $\beta > 0$ even though our derivation started with a rank-one perturbation model, which only makes sense for $\beta \in \{1, 2, 4\}$. The constant γ in Eq. (2.41) only appears in the convex hull Ω^{rk1} (and hence also in the normalization of the density).

Remark *(Simplification for $\beta = 2$).* For $\beta = 2$, the mixed products involving c_j and b_p in Eq. (2.41) are equal to one, and we get the following simple expression:

$$\mathcal{P}^{(\beta=2)}(\boldsymbol{c}|\gamma, \boldsymbol{b}) = \frac{\Gamma(N)}{\gamma^{N-1}} \frac{|\Delta(\boldsymbol{c})|}{|\Delta(\boldsymbol{b})|} \mathbb{I}_{\Omega^{\mathrm{rk1}}}, \tag{2.42}$$

and one can think of the $\beta = 2$ rank-one modification as the law of N particles, each of them constrained to be in a box $c_i \in [b_i, b_{i-1}]$ (with the convention $b_0 := b_1 + \gamma$), and the particles are interacting with a $2d$ logarithm repulsion **and** the trace constraint. ⌟

BBP phase transition -

Let's now turn to the large N limit of this additive rank-one perturbation. We consider a setting where the ESD of the matrix $\mathbf{B} = \mathbf{Diag}(\boldsymbol{b})$ converges to a smooth limit:

$$\mu_{\mathbf{B}}(\lambda) := \frac{1}{N} \sum_{i=1}^{N} \delta(\lambda - b_i) \xrightarrow[N\to\infty]{} \mu_B(\lambda), \tag{2.43}$$

where the top and bottom edge of the support of μ_B are finite and, given respectively by b_+ and b_-. We further assume that the top/bottom eigenvalue converges to the top/bottom edge: $b_1 \to \mathrm{b}_+$, $b_N \to \mathrm{b}_-$. Clearly, by the interlacing conditions of Eq. (2.30), the ESD of the matrix $\mathbf{C}$, given as the random rank-one perturbation of $\mathbf{B}$ by Eq. (2.35), also converges to the same LSD μ_B and the top eigenvalue c_1 of $\mathbf{C}$ is the only eigenvalue which may pop out of the bulk since it is asymptotically restricted to belong to the segment $[\mathrm{b}_+, \mathrm{b}_+ + \gamma]$. Our goal is therefore to understand the - hopefully deterministic - limit of this top eigenvalue. While this asymptotic behavior can be technically done thanks to the joint law of Eq. (2.41) it is actually *much* easier to consider the large N behavior of the secular equation (2.37).

For $\boldsymbol{d} \sim \mathrm{Dir}(\beta/2)$, the mean of each d_i is given by $1/N$ and its variance by:

$$\mathbb{E}\left[\left(d_i - \frac{1}{N}\right)^2\right] = \frac{(N-1)}{N^2\left(\frac{N\beta}{2} + 1\right)} = \mathcal{O}(N^{-2}), \tag{2.44}$$

this means that in the large N limit the top eigenvalue c_1 is not in the spectrum of $\mathbf{B}$ if it is (approximately) the solution of:

$$\frac{1}{N} \sum_{j=1}^{N} \frac{1}{c_1 - b_j} \approx \frac{1}{\gamma}, \tag{2.45}$$

where from a purely rigorous point of view, one needs to understand Eq. (2.45) as an event with high probability. The LHS of Eq. (2.45) is nothing else than the Stieltjes transform $g_{\mathbf{B}}$ of

the distribution $\mu_{\mathbf{B}}$ and so asymptotically one needs to understand the behavior of the solution (if there is any) of the equation:

$$g_B(c_1) = \frac{1}{\gamma}\,, \tag{2.46}$$

for $c_1 \in (\mathrm{b}_+, \mathrm{b}_+ + \gamma]$, where we have removed the approximate sign and $g_B(z) := \int (z - \lambda)^{-1} \mu_B(\mathrm{d}\lambda)$ is the Stieltjes transform of μ_B. Now, we know that $g_B(.)$ is a strictly decreasing function for an argument higher than the top edge b_+ and it goes to zero at infinity. If we denote by:

$$g_B(\mathrm{b}_+) := \lim_{\epsilon \to 0^+} g_B(\mathrm{b}_+ + \epsilon)\,, \tag{2.47}$$

the right limit of the Stieltjes near the edge, then this means that there is always a unique solution of Eq. (2.46) *provided that the inequality* $1/\gamma \leq g_B(\mathrm{b}_+)$ *is true*. If this inequality holds, then c_1 converges to the *inverse of the Stieltjes transform* evaluated at $1/\gamma$. On the other hand, if $1/\gamma > g_B(\mathrm{b}_+)$, there is no solution to Eq. (2.46) on $(\mathrm{b}_+, \mathrm{b}_+ + \gamma]$ and this means that the top eigenvalue c_1 necessarily converges to the top edge b_+. Thus, we have a phase transition depending on the value of γ, which can be summarized in the following result.

Result 2.3 (*BBP phase transition for the sum [57, 10, 22]*)

In the large N limit, if the ESD of a matrix $\mathbf{B}$ converges to a LSD μ_B in a such way that the top eigenvalue converges to the upper edge of the support of the LSD $b_1 \to \mathrm{b}_+$, then the position of the top eigenvalue of the rank-one perturbation $\mathbf{C}$ given by Eq. (2.35) admits a continuous phase transition depending on the value of γ:

- *For $\gamma \leq \frac{1}{g_B(\mathrm{b}_+)}$: the top eigenvalue sticks to the edge b_+,*

$$c_1 \to \mathrm{b}_+\,. \tag{2.48}$$

- *For $\gamma > \frac{1}{g_B(\mathrm{b}_+)}$: the top eigenvalue pops out of the bulk and is equal to*

$$c_1 \to \lambda^* := g_B^{\langle -1 \rangle}\left(\frac{1}{\gamma}\right) > \mathrm{b}_+\,, \tag{2.49}$$

where g_B is the Stieltjes transform of the LSD μ_B defined by Eq. (1.26), $g_B(\mathrm{b}_+) := \lim_{\epsilon \to 0^+} g_B(\mathrm{b}_+ + \epsilon)$ and $g_B^{\langle -1 \rangle}$ is the functional inverse of the Stieltjes transform.

The BBP transition first appeared in the seminal work of Edwards and Jones in Ref. [57] (and for this reason, some authors argue that it should be named the EJ-BBP phase transition !). The acronym BBP stands for the first letters of the names of the three mathematicians (J. Baik, G. Ben Arous and S. Péché) who study the local behavior of this phase transition. The BBP transition still applies if the deterministic matrix $\mathbf{B}$ is replaced by a random matrix, provided the assumptions of Res. 2.3 holds, and in particular in this case where $\mathbf{B}$ is taken from β-ensemble described in Sec. 1.4. There exists a similar phase transition for the overlap between the top eigenvalue of the matrix $\mathbf{C}$ and the direction of the rank-one perturbation, and one can also generalize the setting to small-rank perturbation, see Ref. [22].

Remark *(Notation for the inverse of the Stieltjes transform).* We have denoted by $g_B^{\langle -1 \rangle}$ the inverse of the Stieltjes transform g_B, which satisfies $g_B^{\langle -1 \rangle}(g_B(z)) = z$ for $z > \mathrm{b}_+$. It

is also customary (especially in the physics literature) to write this transform as $\mathfrak{z}_B(.) \equiv g_B^{\langle -1 \rangle}(.)$ because if the Stieltjes transform is a solution of a fixed point equation (typically an algebraic equation) $f(z, g_B(z)) = 0$ then by definition the inverse transform satisfies the same equation for a value of $g_B(z) = \mathrm{g}$ being fixed: $f(\mathfrak{z}_B(\mathrm{g}), \mathrm{g}) = 0$ and the correct solution of this new fixed point equation is given by the behavior near zero $\mathfrak{z}_B(\mathrm{g}) \underset{\mathrm{g}\to 0}{\sim} 1/g$. For example, the Stieltjes transform of the semi-circle distribution given by Eq. (1.41) is the solution of:

$$\sigma^2 g_{\mathrm{sc}(\sigma)}^2(z) - z g_{\mathrm{sc}(\sigma)}(z) + 1 = 0\,, \tag{2.50}$$

from which one can immediately deduce the expression for the inverse Stieltjes transform:

$$g_{\mathrm{sc}(\sigma)}^{\langle -1 \rangle}(\theta) = \frac{1}{\theta} + \sigma^2 \theta\,, \tag{2.51}$$

valid for $\theta \in (0, 1/\sigma)$ since $\lim_{\epsilon \to 0^+} g_{\mathrm{sc}(\sigma)}(2\sigma + \epsilon) = 1/\sigma$. The limiting value $1/\sigma$ can also be deduced by computing the derivative of the inverse of the Stieltjes transform and noticing that it changes sign after this point hence the inverse transform is no more invertible. ⌟

Example *(rank-one perturbation of* GOE *matrices).* If we take $\mathbf{C} = \mathbf{A} + \gamma \boldsymbol{e}_1 \boldsymbol{e}_1^{\mathsf{T}}$ with $\mathbf{A}$ a GOE matrix with law given by Eq. (1.39) and $\boldsymbol{e}_1 = (1, 0, \ldots, 0)$, then since the LSD of GOE matrices is the semi-circle distribution with inverse Stieltjes transform given by Eq. (2.51), asymptotically the top eigenvalue of $\mathbf{C}$ is given by:

$$c_1 \to \begin{cases} 2\sigma & \text{if } \gamma \leq \sigma\,, \\ \\ \gamma + \frac{\sigma^2}{\gamma} & \text{if } \gamma \geq \sigma\,. \end{cases} \tag{2.52}$$

This is illustrated in Fig. 2.3. ■

We conclude this section on the rank-one additive perturbation with two remarks and refer to [22] for a complete survey of this problem.

Remark *(typical fluctuations and large deviations).* Similarly to individual matrix taken from a β-ensemble, one can ask what are the typical fluctuations of c_1 around the deterministic limit given by Eq. (2.48) and Eq. (2.49). If $\mathbf{B}$ is taken from a β-invariant and is non-critical (that is, it has a square-root behavior near the top edge), then one can show that for $\gamma < 1/g_B(\mathrm{b}_+)$, the fluctuations are of Tracy-Widom type (and hence of order $N^{-2/3}$), while for $\gamma > 1/g_B(\mathrm{b}_+)$ the fluctuations are Gaussian and of order $N^{-1/2}$. Exactly at the transition $\gamma = 1/g_B(\mathrm{b}_+)$, the fluctuations are given by *another* class of distribution. In particular, for $\beta = 2$, this is given by a modified Airy-Kernel determinantal process, as shown in Ref. [10]. Similarly, one can ask what is the large deviation principle associated with this top eigenvalue and this is discussed in detail in Chapter 2, Sec. 5.3. ⌟

Remark *(absence of phase transition in a specific case).* If the LSD μ_B is critical and behaves near the top edge as:

$$\mu_B(\lambda) \underset{\lambda \nearrow \mathrm{b}_+}{\sim} \frac{1}{K}(\mathrm{b}_+ - \lambda)^\theta\,, \tag{2.53}$$

with an exponent $\theta \leq 0$ (for example this is the case for the uniform distribution), then one can show that $g_B(\mathrm{b}_+) = +\infty$ and hence for $\gamma > 0$ there is no phase transition: there is always an outlier outside the bulk. ⌟

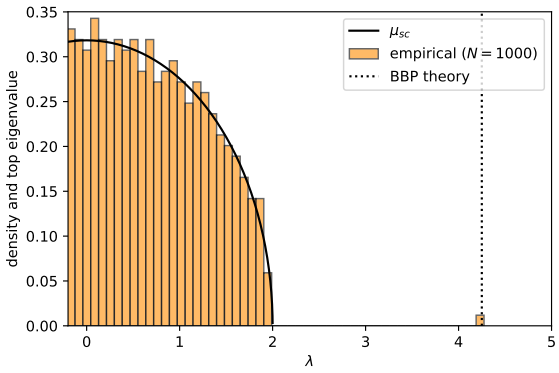


Figure 2.3: Illustration of an outlier for a rank-one deformation of a GOE matrix with $N = 1000$, $\sigma = 1$ and $\gamma = 4$ compared with the asymptotic behavior given by the bottom line of Eq. (2.52).

2.4.3 Multiplicative rank-one perturbation

We next turn to the multiplicative counterpart of the previous rank-one perturbation:

The **random rank-one multiplicative Horn problem** is defined by

$$\mathbf{C} = \sqrt{\mathbf{I} + \gamma\, \boldsymbol{\sigma}\boldsymbol{\sigma}^*}\mathbf{Diag}(\boldsymbol{b})\sqrt{\mathbf{I} + \gamma\, \boldsymbol{\sigma}\boldsymbol{\sigma}^*}\,, \tag{2.54}$$

where $\gamma > 1$ and $\boldsymbol{\sigma} \sim \text{Unif}\left[\mathbb{S}_\beta^{N-1}\right]$. We will also assume the spectrum of $\mathbf{Diag}(\boldsymbol{b})$ to be simple and positive, that is $b_1 > \cdots > b_N > 0$.

The eigenvalues $\boldsymbol{c}$ of $\mathbf{C}$ belong to the set

$$\Omega_\times^{\text{rk1}} := \Big\{\boldsymbol{c} \in \mathbb{R}^N \quad \text{s.t} \quad b_N \le c_N \le b_{N-1} \le \cdots \le c_1 \le b_1(1+\gamma) \\ \text{and} \quad \prod_{i=1}^N c_i = (1+\gamma)\prod_{i=1}^N b_i \Big\}\,. \tag{2.55}$$

As the study of this multiplicative rank-one perturbation is very similar to the one of the additive case, we only sketch the proof of the main results. If we denote by $\mathbf{B} = \mathbf{Diag}(\boldsymbol{b})$, the eigenvalues of the matrix $\mathbf{C}$ given by Eq. (2.54) are the same as the ones of the matrix $\sqrt{\mathbf{B}}(\mathbf{I}+\gamma\,\boldsymbol{\sigma}\boldsymbol{\sigma}^*)\sqrt{\mathbf{B}} = \mathbf{B}+\gamma(\sqrt{\mathbf{B}}\boldsymbol{\sigma})(\sqrt{\mathbf{B}}\boldsymbol{\sigma})^*$. One can use the Sherman-Morrison determinant lemma of Eq. (2.32) with $\boldsymbol{u}_0 = -\gamma(\sqrt{\mathbf{B}}\boldsymbol{\sigma})$, $\boldsymbol{v}_0 = (\sqrt{\mathbf{B}}\boldsymbol{\sigma})$ and the law of Eq. (2.36) of the square of the absolute value of the coordinate of a vector taken uniformly over the sphere to get the following result.

Result 2.4 (*random secular equation for the rank-one multiplicative problem*)

*The eigenvalues $\boldsymbol{c}$ of the matrix $\mathbf{C}$ given by the random rank-one multiplicative Horn problem of Eq. (2.54) are equal in law to the solutions with unknown c_i of the following **random secular equation**:*

$$1 - \gamma\sum_{i=1}^N \frac{d_i b_i}{c_i - b_i} = 0\,, \tag{2.56}$$

where $\boldsymbol{d} \sim \mathrm{Dir}(\beta/2)$ is the Dirichlet distribution with density given by Eq. (2.36).

Note that the secular equation of the multiplicative case only differs from the additive case by an extra factor b_i in the fraction entering the sum. The latter case can be seen as a limiting case of the multiplicative case thanks to the following remark.

Remark *(additive rank-one case as a limit of the multiplicative rank-one case).* As explained in the introductory section of the chapter, one should recover the additive case as a limit of the multiplicative case under the rescaling given by Eq. (2.2) and the limit $\epsilon \to 0^+$. Indeed, if we perform the change of variable $(\gamma = \mathrm{e}^{\epsilon\gamma_0} - 1, \boldsymbol{b} = \mathrm{e}^{\epsilon \boldsymbol{b}_0}, \boldsymbol{c} = \mathrm{e}^{\epsilon \boldsymbol{c}_0})$ than Eq. (2.56) writes:

$$1 - (\mathrm{e}^{\epsilon\gamma_0} - 1) \sum_{i=1}^{N} \frac{d_i \mathrm{e}^{\epsilon b_{0,i}}}{\mathrm{e}^{\epsilon c_{0,i}} - \mathrm{e}^{\epsilon b_{0,i}}} = 0\,, \tag{2.57}$$

which gives at zero order in ϵ:

$$1 - \gamma_0 \sum_{i=1}^{N} \frac{d_i}{c_{0,i} - b_{0,i}} + \mathcal{O}(\epsilon) = 0\,. \tag{2.58}$$

which nothing else than the secular equation Eq. (2.37) for the additive case. ⌟

Joint density -

The derivation of the expression for the joint density of the multiplicative rank-one perturbation is also very similar to the ones of the additive case, and we only give the final result.

Result 2.5 (*joint density for the multiplicative rank-one perturbation*)

The joint density of the eigenvalues $\boldsymbol{c}$ of the matrix $\mathbf{C}$ given by the random rank-one multiplicative Horn problem of Eq. (2.54) is given by:

$$\mathcal{P}_{\times}^{(\beta)}(\boldsymbol{c}|\gamma, \boldsymbol{b}) = \frac{\Gamma\left(\frac{N\beta}{2}\right)}{\Gamma(\beta/2)^N \gamma^{\frac{N\beta}{2}-1} |\Delta(\boldsymbol{b})|^{\beta-1}} |\Delta(\boldsymbol{c})| \prod_{j=1}^{N-1} \prod_{p=1}^{N} |c_j - b_p|^{\frac{\beta}{2}-1} \, \mathbb{I}_{\Omega_{\times}^{\mathrm{rk1}}} \tag{2.59}$$

where we recall that $\Omega_{\times}^{\mathrm{rk1}}$ is defined by Eq. (2.55).

Note that the expression for the joint density of the multiplicative case only differs from the ones of the additive case by the constraint Ω^{rk1} replaced by $\Omega_{\times}^{\mathrm{rk1}}$. It should be mentioned that the expression appears in Ref. [190] for $\beta = 2$ but does not seem to have appeared before for $\beta = 1$ or $\beta = 4$ but the result is actually an immediate consequence of prop 4.2.1 in [67].

BBP phase transition -

We now consider the large N limit of this multiplicative rank-one perturbation, where the ESD of the matrix $\mathbf{B} = \mathbf{Diag}(\boldsymbol{b})$ converges to a smooth LSD μ_B with support included in $\mathbb{R}_+^*$ and such that the top eigenvalue converges to the top edge: $b_1 \to \mathrm{b}_+$. The LSD of the multiplicative

rank-one perturbation matrix $\mathbf{C}$ is also given by μ_B. For N large, the top eigenvalue c_1 is not in the spectrum of $\mathbf{B}$ if it is approximately solution of:

$$\frac{1}{N}\sum_{j=1}^{N}\frac{b_j}{c_1-b_j}\approx\frac{1}{\gamma}\,. \tag{2.60}$$

The LHS of Eq. (2.45) is nothing else than the T-transform transform $t_{\mathbf{B}}(z):=\int\lambda\,(z-\lambda)^{-1}\mu_{\mathbf{B}}(\mathrm{d}\lambda)$ of the distribution $\mu_{\mathbf{B}}$ and since the T-transform is a decreasing function for $z>\mathrm{b}_+$, we deduce by a similar argument as in the additive case, the following result.

Result 2.6 (*BBP phase transition for the product [22]*)

In the large N limit, if the ESD of the b_i converges to a LSD μ_B in a such way that the top eigenvalue converges to the upper edge of the support of the LSD $b_1\to\mathrm{b}_+$, then the position of the top eigenvalue of the rank-one perturbation $\mathbf{C}$ given by Eq. (2.35) admits a phase transition depending on the value of γ:

- *for $\gamma<\frac{1}{t_B(\mathrm{b}_+)}$, the top eigenvalue sticks to the edge b_+:*

$$c_1\to\mathrm{b}_+\,, \tag{2.61}$$

- *for $\gamma>\frac{1}{t_B(\mathrm{b}_+)}$, the top eigenvalue pops out of the bulk and is equal to:*

$$c_1\to\lambda^*:=t_B^{\langle-1\rangle}\left(\frac{1}{\gamma}\right)\,. \tag{2.62}$$

We give below an explicit example that has seen a lot of application in high-dimensional PCA.

Example *(BBP transition for spiked Wishart).* For example, let's consider the case where one samples M realizations of a N-dimensional Gaussian vector $\boldsymbol{x}_m\sim\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})$ for $m=1,\dots,M$, with covariance given by $\boldsymbol{\Sigma}=\mathbf{I}+\gamma\boldsymbol{e}_1\boldsymbol{e}_1^{\mathsf{T}}$. By property of Gaussian vectors, $\boldsymbol{x}_m\overset{\text{in law}}{=}\sqrt{\mathbf{I}+\gamma\boldsymbol{e}_1\boldsymbol{e}_1^{\mathsf{T}}}\,\hat{\boldsymbol{x}}_m$ with $\hat{\boldsymbol{x}}_m\sim\mathcal{N}(\mathbf{0},\mathbf{I})$ and hence the *empirical covariance estimator* is given by:

$$\mathbf{C}:=\frac{1}{M}\sum_{m=1}^{M}\boldsymbol{x}_m\boldsymbol{x}_m^{\mathsf{T}}\overset{\text{in law}}{=}\sqrt{\mathbf{I}+\gamma\boldsymbol{e}_1\boldsymbol{e}_1^{\mathsf{T}}}\,\mathbf{B}\,\sqrt{\mathbf{I}+\gamma\boldsymbol{e}_1\boldsymbol{e}_1^{\mathsf{T}}}\,, \tag{2.63}$$

where $\mathbf{B}$ is LOE (or Wishart) matrix with parameters N and M. In the double scaling where $N,M\to\infty$ with $N/M\to q\in(0,1)$, the ESD of $\mathbf{B}$ converges to the Marčenko-Pastur distribution of Eq. (1.43) and since $\mathbf{C}$ is a rank-one perturbation of a Wishart matrix - known as a *spiked Wishart* - by computing the inverse of the T-transform of the Marčenko-Pastur distribution we have that the top eigenvalue of $\mathbf{C}$ is asymptotically given by:

$$c_1\to\begin{cases}(1+\sqrt{q})^2 & \text{if } \gamma\le\sqrt{q}\,,\\[2ex] (1+\gamma)\left(1+\frac{q}{\gamma}\right) & \text{if } \gamma\ge\sqrt{q}\,.\end{cases} \tag{2.64}$$

Thus, if one thinks of the rank-one vector as a *signal*, then this signal is asymptotically detectable in the spectrum of the matrix $\mathbf{C}$ only if $\gamma>\sqrt{q}$. ∎

2.4.4 Rectangular rank-one perturbation

This section deals with rectangular matrices.

The **random rank-one rectangular Horn problem** is defined by

$$\mathbf{C} = \mathbf{Diag}_q(\boldsymbol{b}) + \gamma\, \boldsymbol{\sigma}\boldsymbol{\sigma'}^* \,, \tag{2.65}$$

where $\boldsymbol{\sigma} \sim \text{Unif}\left[\mathbb{S}_\beta^{N-1}\right]$,$\boldsymbol{\sigma'} \sim \text{Unif}\left[\mathbb{S}_\beta^{M-1}\right]$ and we recall that $\mathbf{Diag}_q(.)$ denotes the $(N \times M)$ diagonal matrix as given by Eq. (1.14). We will also assume $\gamma > 0$ without loss of generality.

For the sum or the product of rank-one perturbation, the starting point is the determinant lemma of Eq. (2.32) obtained from the Sherman-Morrison formula of Eq. (2.31). For the rectangular case, one needs to use a generalization of the Sherman-Morrison formula known as Woodbury's formula from which one can show that the singular values of $\mathbf{C}$ (different from the ones of $\mathbf{B} = \mathbf{Diag}_q(\boldsymbol{b})$) are the zeros of the determinant of a (2×2) matrix. Expanding the determinant gives the following final result, and we refer to Ref. [23] for details.

Result 2.7 (*random secular equation for the rank-one rectangular problem*)

*The singular values $\mathbf{c}$ of the matrix $\mathbf{C}$ given by the random rank-one (additive) rectangular Horn problem of Eq. (2.65) are equal in law to the solutions of the following **random secular equation***

$$\left(\sum_{i=1}^{N} \frac{z|\sigma_i|^2}{z^2 - s_i^2}\right)\left(\sum_{i=1}^{N} \frac{z|\sigma'_i|^2}{z^2 - s_i^2} + \frac{1}{z}\sum_{i=N+1}^{M} |\sigma'_j|^2\right) - \left(\sum_{i=1}^{N} \frac{s_i\sigma_i\sigma_i'^*}{z^2 - s_i^2} - \frac{1}{\gamma}\right)\left(\sum_{i=1}^{N} \frac{s_i\sigma_i^*\sigma'_i}{z^2 - s_i^2} - \frac{1}{\gamma}\right) = 0\,, \tag{2.66}$$

where $\boldsymbol{\sigma} \sim \text{Unif}\left[\mathbb{S}_\beta^{N-1}\right]$ and $\boldsymbol{\sigma}' \sim \text{Unif}\left[\mathbb{S}_\beta^{M-1}\right]$.

As far as I am aware, unlike the additive and multiplicative self-adjoint case, there is no known simple expression for the joint density of the singular values of a rank-one rectangular perturbation in the literature.

Next for large N, M, the sums containing the cross-terms $\sigma_i\sigma_i'^*$ and $\sigma_i^*\sigma'_i$ do not contribute, and we have that c_1 is approximately the solution of

$$\left(\frac{1}{N}\sum_{i=1}^{N} \frac{z}{z^2 - s_i^2}\right)\left(\frac{1}{M}\sum_{i=1}^{N} \frac{z}{z^2 - s_i^2} + \frac{1}{z}\frac{M-N}{M}\right) \approx \frac{1}{\gamma^2}\,, \tag{2.67}$$

and the LHS is nothing else than the square of the D-transform of the ESVD $\mu_{\mathbf{B}}$ defined by Eq. (1.32) and this leads to the following result.

Result 2.8 (*BBP for the rectangular rank-one perturbation [23]*)

In the large N limit, if the ESVD of the b_i converges to a LSVD μ_B in a such way that the top singular value converges to the upper edge of the support of the LSVD $b_1 \to \mathrm{b}_+$, then the position of the top eigenvalue of the rank-one perturbation $\mathbf{C}$ given by Eq. (2.65) admits a phase transition depending on the value of γ:

- *for* $\gamma < \frac{1}{d_B(\mathrm{s}_+)}$, *the top singular value sticks to the edge* b_+*:*

$$c_1(\mathbf{C}) \to \mathrm{b}_+ \, ; \tag{2.68}$$

- *for* $\gamma > \frac{1}{d_B(\mathrm{s}_+)}$, *the top singular value pops out of the bulk and is equal to:*

$$c_1 \to s^* := d_B^{\langle -1 \rangle}\left(\frac{1}{\gamma}\right) \, ; \tag{2.69}$$

where $d_B^{\langle -1 \rangle}$ *is the functional inverse of the D-transform of* μ_B *defined by Eq.* (1.32)*.*

2.5 Random walks on matrix spaces, Dyson Brownian motions, and their hydrodynamical limits

In this section, we consider a setting where for each of the three cases (additive, multiplicative, rectangular) $\mathbf{C}$ is a full-rank perturbation of a matrix $\mathbf{B}$ obtained by an iterative procedure of *small* perturbations. In the limit of infinitesimal perturbations, one obtains a matrix process and the goal is to understand how the spectrum of $\mathbf{B}$ is deformed with the time t. As usual, we start with the additive (self-adjoint) case, for which we give a detailed account of the derivation of the main results, and then turn to the multiplicative and rectangular cases.

The joint law of eigenvalues for the additive (self-adjoint) case is denoted by $\mathcal{P}_t^{(\beta)}(\boldsymbol{c}|\boldsymbol{b})$, for the multiplicative case by $\mathcal{P}_{\times,t}^{(\beta)}(\boldsymbol{c}|\boldsymbol{b})$ and for the singular values of the rectangular case by $\mathcal{P}_{q,t}^{(\beta)}(\boldsymbol{c}|\boldsymbol{b})$.

2.5.1 Iterative infinitesimal additive perturbation and Dyson Brownian Motion

For the additive case, the random walk on the space $\mathrm{Herm}_\beta(N)$ of self-adjoint matrices is obtained by adding element-wise (independently and up to the symmetry) a random variable with variance scaling as $\mathcal{O}(\delta t)$. Since we will be interested in the scaling where δt becomes infinitesimal, thanks to the Central Limit Theorem we can consider the element-wise perturbations to be Gaussian. Thus, we first start with $\mathbf{C}(t_0 = 0) = \mathbf{B}$ with known eigenvalues $\boldsymbol{b}$ and then at each new time step $t_{k+1} = t_k + \delta t$, the matrix $\mathbf{C}$ is modified according to

$$\mathbf{C}(t_{k+1}) = \mathbf{C}(t_k) + \sqrt{\delta t}\, \mathbf{X}_k \, , \tag{2.70}$$

where $\mathbf{X}_k \sim \mathbb{P}_G^{(\beta)}$ is a matrix taken from the Gaussian ensemble with $\sigma = 1$. Note that the variance of each element of the matrix $\mathbf{X}_k$ is order $\mathcal{O}(1/N)$ in order to have a non-trivial limit for the spectrum in the large N limit.

Joint law and the additive spherical integral -

Thus after K steps we have:

$$\mathbf{C}(t_K) = \mathbf{B} + \sum_{k=1}^{K} \sqrt{\delta t}\mathbf{X}_k \,, \tag{2.71}$$

and since the sum of Gaussian random variables is again a Gaussian random variable with variance given as the sum of the variances, in the continuous limit we have:

$$\mathbf{C}(t) = \mathbf{B} + \sqrt{t}\mathbf{X} \,, \tag{2.72}$$

with $\mathbf{X}$ a GOE (resp. GUE, GSE) with $\sigma = 1$ for $\beta = 1$ (resp. $\beta = 2, 4$). Since the joint of $\mathbf{X} = (\mathbf{C} - \mathbf{B})/\sqrt{t}$ is given by Eq. (1.39), we have that the joint density of the matrix $\mathbf{C}$ knowing the matrix $\mathbf{B}$ is given by:

$$P^{(\beta)}(\mathbf{C}|\mathbf{B}) = \frac{1}{Z} \mathrm{e}^{-\frac{N\beta}{2}\mathrm{Tr}\frac{(\mathbf{C}-\mathbf{B})^2}{2t}} \,, \tag{2.73}$$

Expanding the product, this density can be written in the form of a *matrix model with external source*

$$P^{(\beta)}(\mathbf{C}|\mathbf{B}) = \frac{1}{Z'} \mathrm{e}^{-\frac{N\beta}{2}\left(\mathrm{Tr}V(\mathbf{C}) - \alpha\,\mathrm{Tr}(\mathbf{CB}) + \alpha \mathrm{Tr}\,\mathbf{B}^2\right)} \,, \tag{2.74}$$

with $V(\lambda) = \lambda^2/(2t)$ and $\alpha = 1/t$. Matrix models with an external source have been extensively studied in the literature (in particular for $\beta = 2$) and we refer the reader to the monograph [34] for example. Since we are interested in the eigenvalues of $\mathbf{C}$, we can use Weyl's formula (1.10) and integrate the dependency in the eigenvectors to obtain the following form for the joint law of the eigenvalues:

$$\mathcal{P}_t^{(\beta)}(\boldsymbol{c}|\boldsymbol{b}) \propto |\Delta(\boldsymbol{c})|^\beta \mathrm{e}^{-\frac{N\beta}{2}\sum_{i=1}^N V(c_i)} \int_{\mathsf{O}_\beta(N)} \mathrm{e}^{+\frac{N\beta\alpha}{2}\mathrm{Tr}(\mathbf{Diag}(\boldsymbol{c})\mathbf{V}\mathbf{Diag}(\boldsymbol{b})\mathbf{V}^*)} \mu_{\mathrm{Haar}}(\mathrm{d}\mathbf{V}) \,, \tag{2.75}$$

with $\alpha = 1/t$.

The integral over the group $\mathsf{O}_\beta(N)$ is an important object in this thesis. For two matrices $\mathbf{A}, \mathbf{T} \in \mathrm{Herm}_\beta(N)$, with eigenvalues $\boldsymbol{a}, \boldsymbol{t}$, we define the **additive spherical integral** or **Harish-Chandra-Itzykson-Zuber** (HCIZ) **integral** with parameter $\beta \in \{1, 2, 4\}$ as

$$I^{(\beta)}(\mathbf{A}, \mathbf{T}) \equiv \mathcal{I}^{(\beta)}(\boldsymbol{a}, \boldsymbol{t}) := \int_{\mathsf{O}_\beta(N)} \mathrm{e}^{\mathrm{Tr}\,\mathbf{AVTV}^*} \mu_{\mathrm{Haar}}(\mathrm{d}\mathbf{V}) \,. \tag{2.76}$$

Note that the HCIZ integral only depends on the eigenvalues of its matrix arguments since one can always absorb the eigenmatrix in the Haar measure. This spherical integral will be studied in more detail in Chapter 3. As a consequence we can write the joint law of the eigenvalues of the matrix $\mathbf{C}$ at time t as:

$$\boxed{\mathcal{P}_t^{(\beta)}(\boldsymbol{c}|\boldsymbol{b}) \propto |\Delta(\boldsymbol{c})|^\beta \mathrm{e}^{-\frac{N\beta}{4t}\left(\sum_{i=1}^N c_i^2 + b_i^2\right)} \mathcal{I}^{(\beta)}\left(\boldsymbol{c}, \frac{N\beta}{2t}\boldsymbol{b}\right) \,.} \tag{2.77}$$

It turns out that for $\beta = 2$ (as we will see in Chapter 3) there exists a *determinantal formula* for the HCIZ integral and in this case, the joint law can be simplified. However, even for $\beta = 2$, it is *difficult* to characterize the limiting spectral distribution from this joint law. On the other hand, this relation between the joint law and this spherical integral will be particularly convenient to study the large N limit of this spherical integral, as we will see in the following chapter.

Stochastic differential equation for the eigenvalues -

Another way to characterize the set of eigenvalues $\boldsymbol{c}$ is to find its associated (random) equation of evolution, from which it is easier to get the behavior at large N. To do this, let's start with the finite setting of Eq. (2.70), using second-order perturbation theory (also known as Hadamard's variational principle), at first order in δt, the eigenvalues evolve after a time step δt according to:

$$c_i(t+\delta t) = c_i(t) + \sqrt{\delta t}\, \mathbf{v}_i(t)^* \mathbf{X}_k \mathbf{v}_i(t) + \delta t \sum_{j|j\neq i} \frac{|\mathbf{v}_j(t)^* \mathbf{X}_k \mathbf{v}_i(t)|^2}{c_i(t) - c_j(t)} + \mathcal{O}(\delta t^{3/2})\,, \tag{2.78}$$

where $\boldsymbol{v}_i(t)$ is the i^{th} eigenvector of the matrix $\mathbf{C}(t)$ at time t. Now since $\mathbf{X}_k$ is β-invariant, we can always set this vector to be $\boldsymbol{v}_i(t) = \boldsymbol{e}_i$ without changing the law of the eigenvalues $\boldsymbol{c}(t+\delta t)$ and since the entries of $\mathbf{X}_k$ are Gaussian, we have:

$$c_i(t+\delta t) - c_i(t) = \frac{1}{N} \sum_{j|j\neq i} \frac{|\xi_{ij}|^2}{c_i(t) - c_j(t)} \delta t + \sqrt{\frac{2}{N\beta}}\, \sqrt{\delta t}\, \xi_{ii} + \mathcal{O}(\delta t^{3/2})\,, \tag{2.79}$$

with ξ_{ij} a Gaussian random variable with variance one. At first order in δt, the first term of the RHS of Eq. (2.79) is dominated by its mean and since the term $\sqrt{\delta t}\xi_{ii}$ is a Gaussian increment with variance $\sqrt{\delta t}$, one obtains in the informal limit $\delta t \to \mathrm{d}t$ the following stochastic differential equation known as the **Dyson Brownian Motion** (DBM), which was first proposed by Dyson in Ref. [56]:

$$\boxed{\mathrm{d}c_i(t) = \frac{1}{N} \sum_{j|j\neq i} \frac{1}{c_i(t) - c_j(t)} \mathrm{d}t + \sqrt{\frac{2}{N\beta}} \mathrm{d}B_t^{(i)}\,,} \tag{2.80}$$

where $(\mathrm{d}B_t^{(1)}, \dots, \mathrm{d}B_t^{(N)})$ are N independent Brownian motions. The *force* of the interactions between the c_i in the DBM of (2.80) is the derivative of the interaction term of the $2d$ Coulomb gas, and hence the eigenvalues repel each other. For $\beta = 2$, one can show that the DBM is equal in law as the process of N Brownian particles (or *vicious walker*) conditioned to never intersect. A plot of the trajectories of a DBM is given in Fig. 2.4.

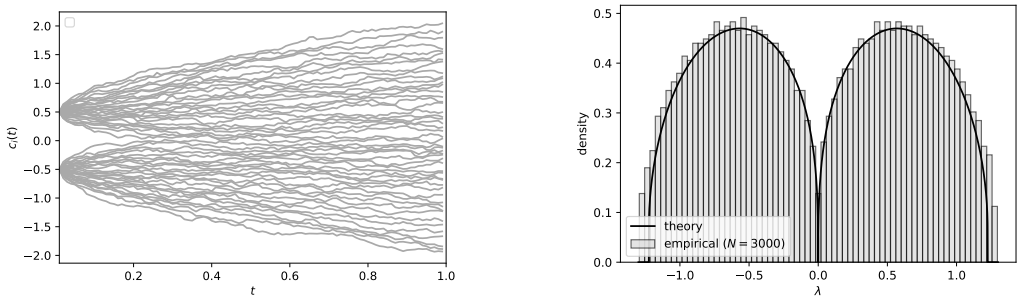
Large N limit and the Burger's equation -

Now, this DBM is particularly well-suited for the study of the large N limit. To do so, let's first perform a change of variable from the c_i to the associated Stieltjes transform $g_{\mathbf{C}} \equiv g(c_1, \dots, c_N, z) = 1/N \sum_{i=1}^N (z - c_i)^{-1}$, which is given by Ito's lemma as:

$$\mathrm{d}g_{\mathbf{C}} = \sum_{i=1}^N \partial_i g_{\mathbf{C}} \cdot \mathrm{d}c_i + \frac{1}{2} \sum_{i,j} \partial_i \partial_j g_{\mathbf{C}} \cdot \left(\frac{2}{N\beta}\right) \mathrm{d}t\,, \tag{2.81}$$

where $\partial_i \equiv \partial_{c_i}$ and these derivatives are given by $\partial_i g_{\mathbf{C}} = (z - c_i)^{-2}/N$ and $\partial_i \partial_j g_{\mathbf{C}} = \delta_{ij} \cdot 2(z - c_i)^{-3}/N$ where $\delta_{ij} = 1$ if and only if $i = j$. This leads to the following equation for the Stieltjes transform:

$$\mathrm{d}g_{\mathbf{C}} = \left(\frac{2}{N^2\beta} \sum_{i=1}^N \frac{1}{(z - c_i)^3} + \frac{1}{N} \sum_{i,j|j\neq i} \frac{1}{(z - c_i)^2} \frac{1}{c_i - c_j} \right) \mathrm{d}t + \sqrt{\frac{2}{N^3\beta}} \sum_{i=1}^N \frac{\mathrm{d}B_t^{(i)}}{(z - c_i)^2}\,. \tag{2.82}$$



(a) **(b)**

Figure 2.4: Illustration of the trajectories of particles following a DBM starting at time $t = 0$ at the positions $1/2$ and $-1/2$ with $N = 50$ and $\beta = 1$ in $(\mathbf{a})$. In $(\mathbf{b})$, a plot of the limiting spectral density for $t = 1/4$ and the same initial condition, as given by Eq. (2.94).

Now the two sums inside the brackets of Eq. (2.82) can be after a few algebraic operations identified as partial derivatives of the Stieltjes transform:

- for the first sum we have:

$$\frac{2}{N^2\beta}\sum_{i=1}^{N}\frac{1}{(z-c_i)^3} = \frac{1}{N\beta}\partial_z^2 g_{\mathbf{C}}\ , \tag{2.83}$$

- while for the second we have:

$$\frac{1}{N}\sum_{i,j|j\neq i}\frac{1}{(z-c_i)^2}\frac{1}{c_i-c_j} = \frac{1}{N}\sum_{i,j|j\neq i}\frac{1}{(z-c_i)^2}\frac{1}{z-c_j} = g_{\mathbf{C}}\partial_z g_{\mathbf{C}} - \frac{1}{2N}\partial_z^2 g_{\mathbf{C}}\ . \tag{2.84}$$

Combining these two results in Eq. (2.82), we get the following (stochastic) partial differential equation:

$$\partial_t g_{\mathbf{C}} + g_{\mathbf{C}}\partial_z g_{\mathbf{C}} = \frac{1}{N}\left(\frac{2-\beta}{2\beta}\right)\partial_z^2 g_{\mathbf{C}} + \sqrt{\frac{2}{N^3\beta}}\sum_{i=1}^{N}\frac{\eta_t^{(i)}}{(z-c_i)^2}\ , \tag{2.85}$$

where $(\eta_t^{(1)},\dots,\eta_t^{(N)})$ are N independent white noise process in time (that is the formal derivative of the Brownian motion). Now in the large N limit, the RHS of Eq. (2.85) vanishes, and we have the following results.

Result 2.9 (*Inviscid Burgers equation for the DBM*)

The Stieltjes transform $g_C(z,t)$ of the limiting distribution $\mu_C(z,t)$ of the particles evolving according to the DBM of Eq. (2.80), is solution of the complex inviscid Burger's equation*:*

$$\partial_t g_C(z,t) + g_C(z,t)\partial_z g_C(z,t) = 0 \tag{2.86}$$

with initial condition $g_C(z,0) = g_B(z) = \int \mu_B(\mathrm{d}\lambda)(z-\lambda)^{-1}$ and with μ_B the limiting spectral distribution of the matrix $\mathbf{B}$.

Using the method of characteristics, one can transform this equation into an implicit equation for $g_C(z,t)$:

$$g_C(z,t) = g_B\left(z - t\, g_C(z,t)\right)\,, \tag{2.87}$$

In particular, if the Stieltjes transform g_B admits a simple inverse, one can look at Eq. (2.87) for large z and apply the inverse:

$$g_B^{\langle -1\rangle}\left(g_C(z,t)\right) + t\, g_C(z,t) - z = 0\,, \tag{2.88}$$

then solve Eq. (2.88) with the proper behavior at infinity $g_C(z,t) \sim 1/z$ and then extend the result to $\mathbb{C}\setminus \mathrm{Supp}[\mu_C(x,t)]$ by analytical property of the Stieltjes transform. As a sanity check, one can see that if $\mathbf{B}$ is the null matrix, then $g_B(z) = g_B^{\langle -1\rangle}(z) = 1/z$ and Eq. (2.88) gives back the fixed point Eq. (2.50) of the Stieltjes transform of the semi-circle distribution with $\sigma = \sqrt{t}$, as one should expect.

One can also look at the behavior of Burgers' equation near the branch cut. For $x \in \mathrm{Supp}[\mu_C(x,t)]$, we have:

$$g_C(x - \mathrm{i}0^+, t) = v(x,t) + \mathrm{i}\,\pi\mu_C(x,t)\,, \tag{2.89}$$

where the imaginary part is nothing else than the Sokochi-Plemelj inversion formula of Eq. (1.28) and the real part v, is given by the *Hilbert transform* of $\mu_C(x,t)$:

$$v(x,t) := \pi\mathcal{H}_{\mu_C}(x,t) := \fint \frac{\mu_C(\lambda,t)}{x-\lambda}\mathrm{d}\lambda\,. \tag{2.90}$$

If we now inject Eq. (2.89) in the Burger's Eq. (2.86) then the associated equations for the real part and imaginary parts give that the couple (μ_C, v) is the solution of *Euler's equations*

$$\partial_t \mu_C(x,t) + \partial_x\left(\mu_C(x,t)v(x,t)\right) = 0\,, \tag{2.91}$$

$$\partial_t v(x,t) + v(x,t)\partial_x v(x,t) - \pi^2\mu_C(x,t)\partial_x\,\mu_C(x,t) = 0\,, \tag{2.92}$$

with initial conditions $\mu_C(x,0) = \mu_B(x)$ and $v(x,0) = \mathcal{H}_{\mu_B}(x)$. In the language of hydrodynamics, Eq. (2.91) is the equation of conservation of mass for a one-dimensional fluid with density μ_C and velocity v and Eq. (2.92) is the equation of motion. In full generality, one needs to understand the couple (μ_C, v) as a weak solution of this set of equations due to its potential singular behavior.

In Chapter 4, we will encounter the DBM with an initial condition being given by the uniform (or 'flat') distribution, while in Chapter 3 we will briefly discuss the case of a DBM constrained at both ends. We conclude this section on the DBM with an explicit example given for another initial condition.

Example *(Bernoulli initial condition).* Let's consider the case where the matrix $\mathbf{B}$ has half its eigenvalues being equal to $+a$ and the other half is equal to $-a$. In the large N, the LSD of $\mathbf{B}$ is given by the discrete measure $\mu_B(\lambda) = \frac{1}{2}\delta(\lambda - a) + \frac{1}{2}\delta(\lambda + a)$ with Stieltjes transform $g_B(z) = \int (z-\lambda)^{-1}\mu_B(\mathrm{d}\lambda) = z(z^2 - a^2)^{-1}$. If we use this expression in the implicit equation (2.87), we have that the Stieltjes transform $g_C \equiv g_C(z,t)$ is the solution of the following cubic equation:

$$t\, g_C^3 - 2zt\, g_C^2 + \left(z^2 - a^2 + t\right) g_C - z = 0\,. \tag{2.93}$$

The correct solution is chosen such that one has the correct asymptotic $g_C(z,t) \sim \frac{1}{z}$ for $|z| \to \infty$ and is given by Cardano's formula. The expression for the density is then given by the inversion formula of Eq. (1.28) and it turns out that it has a relatively simple expression (the result has been obtained thanks to `Mathematica`):

$$\mu_C(x,t) = \frac{-6a^2 + 6t - 2x^2 + 2^{\frac{1}{3}} F_a\left(|x|,t\right)^{\frac{2}{3}}}{2\,\pi\,\sqrt{3}\,2^{\frac{2}{3}}\,t\,F_a\left(|x|,t\right)^{\frac{1}{3}}} \mathbb{I}_J\,, \tag{2.94}$$

with the function

$$F_a\left(x,t\right) := 9(2a^2+t)x - 2x^3 + 3\sqrt{3}\sqrt{-4(a^2-t)^3 + (8a^4+20a^2t-t)x^2 - 4a^2x^4}\,. \tag{2.95}$$

For $t < \sqrt{a}$, the support J of the distribution is made of two disjoint intervals $J = (-\mathrm{c}_+, -\mathrm{c}_{in}) \cup (\mathrm{c}_{in}, \mathrm{c}_+)$ which merges at $t = a^2$. ∎

2.5.2 Multiplicative perturbation and the Dyson Geometric Brownian motion

In this section, we consider the multiplicative analogous of the additive infinitesimal perturbation of the previous section. Because every positive matrix $\mathbf{A} \in \mathrm{Herm}_\beta^{++}(N)$ can be written as $\mathbf{A} = \exp\left[\mathbf{A}_0\right]$, with $\mathbf{A}_0 \in \mathrm{Herm}_\beta(N)$, a natural candidate is to construct an iterative multiplicative procedure with $\mathbf{A}_0 = \sqrt{\delta t}\mathbf{X}_k$ where $\mathbf{X}_k$ is as in the previous section a matrix taken from the Gaussian ensemble. In other words, we look at the matrix process starting at $\mathbf{C}(t_0 = 0) = \mathbf{B} \in \mathrm{Herm}_\beta^{++}(N)$ and at a time step $t_{k+1} = t_k + \delta t$, we update the matrix $\mathbf{C}$ according to:

$$\mathbf{C}(t_{k+1}) = \left(\exp\left[\sqrt{\delta t}\,\mathbf{X}_k\right]\right)^{1/2} \mathbf{C}(t_k) \left(\exp\left[\sqrt{\delta t}\,\mathbf{X}_k\right]\right)^{1/2}. \tag{2.96}$$

Now because we are interested in the small 'step-size' limit where δt is infinitesimal, at first order in δt this can also be written as:

$$\mathbf{C}(t_{k+1}) = (\mathbf{I} + \sqrt{\delta t}\,\mathbf{X}_k)^{1/2}\ \mathbf{C}(t_k)\ (\mathbf{I} + \sqrt{\delta t}\,\mathbf{X}_k)^{1/2} + o(\delta t)\,. \tag{2.97}$$

If now one uses second-order perturbation theory as in the previous section and takes the informal limit "$\delta t \to \mathrm{d}t$", then one can show that the eigenvalues of this continuous process evolve according to the following **Dyson Geometric Brownian Motion** (DGBM):

$$\boxed{\mathrm{d}c_i(t) = \frac{1}{N}\sum_{j|j\neq i} \frac{c_i(t)c_j(t)}{c_i(t)-c_j(t)}\mathrm{d}t + \sqrt{\frac{2}{N\beta}}c_i(t)\,\mathrm{d}B_t^{(i)}\,,} \tag{2.98}$$

with initial conditions, $c_i(0) = b_i$ and where $(\mathrm{d}B_t^{(1)}, \ldots, \mathrm{d}B_t^{(N)})$ are N independent Brownian motions and the equation has to be understood with the Ito convention.

This stochastic differential equation can be put in *hyperbolic form*. If we first do the change of variable $c_{0,i}(t) = \log c_i(t)$, then by Ito calculus, we have:

$$\mathrm{d}c_{0,i}(t) = \frac{1}{N}\sum_{j|j\neq i} \frac{\mathrm{e}^{c_{0,j}(t)}}{\mathrm{e}^{c_{0,i}(t)} - \mathrm{e}^{c_{0,j}(t)}}\mathrm{d}t + \sqrt{\frac{2}{N\beta}}\,\mathrm{d}B_t^{(i)} - \frac{1}{N\beta}\mathrm{d}t\,. \tag{2.99}$$

Next, using the hyperbolic identity $\frac{1}{2}\coth\left(\frac{a-b}{2}\right) = \frac{\mathrm{e}^b}{\mathrm{e}^a-\mathrm{e}^b} + \frac{1}{2}$ and since there are $(N-1)$ terms in the sum $\sum_{j|j\neq i}\ldots$, we get:

$$\mathrm{d}c_{0,i}(t) = \left(-\frac{1}{2} + \frac{1}{N}\left(\frac{1}{2}-\frac{1}{\beta}\right) + \frac{1}{2N}\sum_{j|j\neq i}\coth\left(\frac{c_{0,i}(t)-c_{0,j}(t)}{2}\right)\right)\mathrm{dt} + \sqrt{\frac{2}{N\beta}}\,\mathrm{d}B_t^{(i)}\,. \tag{2.100}$$

The interaction with the hyperbolic cotangent function can be seen as a force derived from a repulsive potential of the form $\sum_{j|j\neq i}\log\sinh(x_i-x_j)$ which we will also encounter in Chapter 4. Let's also mention that one can retrieve the usual DBM from its geometric version which is again a consequence of Eq. (2.2), see the following remark.

Remark *(DGBM at short time and DBM).* If one first re-scales time in the following way $t=\epsilon^2\tau$, we have:

$$\mathrm{d}c_i(\epsilon^2\tau) = \frac{1}{N}\sum_{j|j\neq i}\frac{c_i(\epsilon^2\tau)c_j(\epsilon^2\tau)}{c_i(\epsilon^2\tau)-c_j(\epsilon^2\tau)}\epsilon^2\mathrm{d}\tau + \sqrt{\frac{2}{N\beta}}c_i(\epsilon^2\tau)\epsilon\,\mathrm{d}B_\tau^{(i)}\,. \tag{2.101}$$

Next if one does the change of variable defined by $l_i(\tau) = \epsilon^{-1}\log c_i(\epsilon^2\tau)$ then the l_i's follow the stochastic equation:

$$\mathrm{d}l_i(\tau) = \frac{1}{N}\sum_{j|j\neq i}\frac{\mathrm{e}^{\epsilon l_j(\tau)}}{\mathrm{e}^{\epsilon l_i(\tau)}-\mathrm{e}^{\epsilon l_j(\tau)}}\epsilon\mathrm{d}\tau + \sqrt{\frac{2}{N\beta}}\,\mathrm{d}B_\tau^{(i)} - \frac{\epsilon}{N\beta}\mathrm{d}\tau\,, \tag{2.102}$$

which gives at zero order:

$$\mathrm{d}l_i(\tau) = \frac{1}{N}\sum_{j|j\neq i}\frac{1}{l_i(\tau)-l_j(\tau)}\mathrm{d}\tau + \sqrt{\frac{2}{N\beta}}\,\mathrm{d}B_\tau^{(i)} + \mathcal{O}(\epsilon)\,, \tag{2.103}$$

and this is nothing else than the usual DBM of Eq. (2.80). ⌟

The large N asymptotic can be obtained similarly as in the additive case, the Stieltjes transform is replaced by (a modification of) the T-transform, see the following result.

Result 2.10 (*Burger's equation for the T-transform of the DGBM*)

in the limit $N\to\infty$, the T-transform defined by Eq. (1.29) and evaluated at $y=\mathrm{e}^z$, $\tilde{t}_C(y,t) := t_C(e^z,t)$ of the limiting distribution $\mu_C(z,t)$ of the particles evolving according to the DGBM of Eq. (2.98), is solution of the complex inviscid Burger's equation*:*

$$\partial_t\tilde{t}_C(y,t) + \tilde{t}_C(y,t)\partial_y\tilde{t}_C(y,t) = 0 \tag{2.104}$$

with initial condition $\tilde{t}_C(y,0) = \tilde{t}_B(y) = t_B(\log y)$, where t_B is the T-transform of the limiting spectral distribution μ_B of the matrix $\mathbf{B}$.

In particular, since this is the same Burgers equation as in the previous section, we also have the following implicit representation for its solution:

$$\tilde{t}_C(y,t) = \tilde{t}_B\left(y - t\,\tilde{t}_C(y,t)\right)\,. \tag{2.105}$$

Example *(Multiplicative counterpart of the semi-circle distribution).* For the usual DBM, if the initial matrix is the null matrix $\mathbf{B} = \mathbf{0}$, then the limiting density of the DBM is given by a semi-circle distribution with variance $\sigma^2 = t$. Thus, the natural multiplicative counterpart is to look at the DGBM starting at $\mathbf{B} = \exp[\mathbf{0}] = \mathbf{I}$, that is the identity matrix. Since the T-transform of the identity matrix is $t_1(z) = \frac{1}{z-1}$, one gets that the Stieltjes transform g_C of the DGBM starting at the identity is asymptotically given as a solution of the fixed point equation:

$$g_C(z,t) = \frac{1}{z - \mathrm{e}^{t(g_C(z,t)-1)}}\,, \tag{2.106}$$

which unfortunately does not have a closed expression in terms of classical functions. ∎

2.5.3 Rectangular perturbation and the Dyson Bessel Process

This section deals with rectangular matrices.

For the rectangular case, we consider the matrix process starting at $\mathbf{C}(t_0 = 0) = \mathbf{B} \in \mathsf{M}_{N,M}(\mathbb{K}_\beta)$ and with an update given at each time step by:

$$\mathbf{C}(t_{k+1}) = \mathbf{C}(t_k) + \sqrt{\delta t}\,\mathbf{G}_k\,, \tag{2.107}$$

where $\mathbf{G} = 1/q \cdot \mathbf{X}$ where $q = N/M$ and $\mathbf{X}$ is a Gaussian rectangular matrix with law given by Eq. (1.50). Thus, for any time t, the joint law of the matrix $\mathbf{C}$ is given by:

$$P_q^{(\beta)}(\mathbf{C}|\mathbf{B}) = \frac{1}{Z}\mathrm{e}^{-\frac{N\beta}{2}\mathrm{Tr}\,\frac{(\mathbf{C}-\mathbf{B})(\mathbf{C}-\mathbf{B})^*}{t}}\,, \tag{2.108}$$

and so if we expand the product and use Weyl's formula for singular values given by Eq. (1.16), we get for the joint law of $\boldsymbol{c}$:

$$\boxed{\mathcal{P}_{q,t}^{(\beta)}(\boldsymbol{c}|\boldsymbol{b}) \propto \mathrm{e}^{-\frac{N\beta}{2t}\cdot\sum_{i=1}^N c_i^2+b_i^2}\prod_{i=1}^N c_i^{\beta(M-N+1)-1}\Delta(\boldsymbol{c}^2)^\beta\;\mathcal{I}_q^{(\beta)}(\boldsymbol{c}, N\beta/t\,\boldsymbol{b})\,,} \tag{2.109}$$

where for two rectangular matrices $\mathbf{A}, \mathbf{T} \in \mathsf{M}_{N,M}(\mathbb{K}_\beta)$, with singular values $\boldsymbol{a}$ and $\boldsymbol{t}$, the **rectangular spherical integral** $\mathcal{I}_q^{(\beta)}(\boldsymbol{a},\boldsymbol{t})$ is defined as:

$$I_q^{(\beta)}(\mathbf{A},\mathbf{T}) = \mathcal{I}_q^{(\beta)}(\boldsymbol{a},\boldsymbol{t}) := \int_{\mathsf{O}_\beta(N)}\int_{\mathsf{O}_\beta(M)} \mathrm{e}^{\mathfrak{Re}\,\mathrm{Tr}\left(\mathbf{Diag}_q(\boldsymbol{a})\mathbf{V}_1\mathbf{Diag}_q(\boldsymbol{t})^\mathsf{T}\mathbf{V}_2\right)}\mu_{\mathrm{Haar}}(\mathrm{d}\mathbf{V}_1)\mu_{\mathrm{Haar}}(\mathrm{d}\mathbf{V}_2)\,, \tag{2.110}$$

with $\mathbf{V}_1 \in \mathsf{O}_\beta(N)$ and $\mathbf{V}_2 \in \mathsf{O}_\beta(M)$.

As in the additive (self-adjoint) case, it is difficult to exploit this joint distribution to get the large N limit behavior of the limiting singular value distribution of the matrix $\mathbf{C}$ but on the other hand, it will be useful in order to compute the large N asymptotic of the rectangular spherical integral, which will be done in the following chapter.

The standard way to characterize the distribution is to find the corresponding stochastic differential equations satisfied by the singular values of the matrix $\mathbf{C}$. It can be shown (see for

example [83] and references therein) that they evolved according to the **Dyson Bessel process** (DBP) given by:

$$\boxed{\mathrm{d}c_i(t) = \left(\frac{\alpha_{N,M,\beta}}{c_i(t)} + \frac{1}{2N} \sum_{j|j\neq i} \frac{1}{c_i(t) - c_j(t)} + \frac{1}{2N} \sum_{j|j\neq i} \frac{1}{c_i(t) + c_j(t)} \right) \mathrm{d}t + \sqrt{\frac{1}{N\beta}}\, \mathrm{d}B_t^{(i)} \,,} \tag{2.111}$$

with initial condition $c_i(0) = b_i$ and

$$\alpha_{N,M,\beta} := \frac{M}{N} - 1 + \left(1 - \frac{1}{\beta}\right) \frac{1}{N} \xrightarrow[N/M\to q]{} \frac{1-q}{q} \,. \tag{2.112}$$

Remark *(Simplification for $q = 1$).* For square non-Hermitian matrices, corresponding to $q = 1$, the parameter $\alpha_{N,M,\beta} = \mathcal{O}(N^{-1})$. If we introduce N fictitious particles $c_{-i} := -c_i$ then one has the identity $(c_i + c_j)^{-1} = (c_i - c_{-j})^{-1}$. If we now add and subtract the term corresponding to $j = -i$, we can write the two sums in Eq. (2.111) as a single sum, and this leads to the following dynamics for the $2N$ particles:

$$\mathrm{d}c_i(t) = \frac{1}{2N} \sum_{j=-N|j\neq i,0}^{N} \frac{1}{c_i(t) - c_j(t)} \mathrm{d}t + \sqrt{\frac{1}{N\beta}}\, \mathrm{d}B_t^{(i)} + \mathcal{O}\left(\frac{1}{N}\right) \,, \tag{2.113}$$

where the term of order $\mathcal{O}(N^{-1})$ is explicitly given by $c_i^{-1} \cdot \left((1/2 - \beta^{-1})\right)/N$. Up to a correction of order $\mathcal{O}(N^{-1})$ (which is exactly zero for $\beta = 2$), one retrieves the dynamics of the usual DBM of $2N$ particles. ⌟

The appropriate transform in the large N limit is given by the Stieltjes transform of the symmetrized distribution, as shown by the following result given in Ref. [70, 83].

Result 2.11 (*Differential equation for the symmetrized Stieltjes transform of the DBP*)

in the limit $N \to \infty$, the Stieltjes transform $\widehat{g}_C(z) = \int (z-x)^{-1} \widehat{\mu}_C(x,t)\mathrm{d}x$ of the symmetrized LSVD $\widehat{\mu}_C(x,t) = \mu_C(x,t)/2 + \mu_C(-x,t)/2$ of the particles evolving according to the DBP of Eq. (2.111), is solution of:

$$\partial_t\, \widehat{g}_C(z,t) - \left(\frac{q^{-1}-1}{2z^2}\right) \widehat{g}_C(z,t) + \left(\widehat{g}_C(z,t) + \frac{q^{-1}-1}{2z}\right) \partial_z\, \widehat{g}_C(z,t) = 0 \tag{2.114}$$

with initial condition $\hat{g}_C(z,0) = \hat{g}_B(z)$.

2.6 Sum of Large β-invariant matrices and free probability in a nutshell

2.6.1 Randomized Horn problems, joint laws, and spherical integrals

We now consider the general randomized Horn problem associated with each operation. For each operation, one can express the joint density of the eigenvalues/singular values as an integral of the product of the associated spherical integrals. The properties of these densities have been

recently studied in Refs. [191, 43, 42, 107, 100] and references therein, and we briefly describe these joint densities in this section.

The additive case -

We consider the following **additive randomized Horn's problem**

$$\mathbf{C} = \mathbf{V}\mathbf{Diag}(\boldsymbol{a})\mathbf{V}^* + \mathbf{V}'\mathbf{Diag}(\boldsymbol{b})\mathbf{V}'^* \qquad \text{with } \mathbf{V}, \mathbf{V}' \sim \mathrm{Unif}[\mathsf{O}_\beta(N)] \tag{2.115}$$

and the goal is to compute the law $\mathcal{P}^{(\beta)}(\boldsymbol{c}|\boldsymbol{a}, \boldsymbol{b})$ of the eigenvalues of $\mathbf{C}$ from the knowledge of $\boldsymbol{a}$ and $\boldsymbol{b}$.

Remark *(simplification of the problem).* Since we are interested only in the eigenvalues, one can remove one of the two random eigenmatrices $\mathbf{V}$ or $\mathbf{V}'$ without changing the law $\mathcal{P}^{(\beta)}(\boldsymbol{c}|\boldsymbol{a}, \boldsymbol{b})$. ⌟

To get the joint density, we will repeat the derivation of the vector problem of Sec. 2.2. For the sum $\mathbf{C} = \mathbf{A} + \mathbf{B}$ of two independent self-adjoint matrices, we are adding an operation element-wise and therefore the corresponding Fourier transform of the matrix $\mathbf{A}$ is given by introducing a conjugate variable $T_{ij}^{(\beta)}$ for each component $A_{ij}^{(\beta)}$ of each *independent* element A_{ij}. Thus, by the self-adjoint symmetry, this corresponds to do the average over the entries A_{ij} for $i \leq j$ of $\exp\left[\mathrm{i}\sum_{i\leq j}\sum_{b=1}^{\beta} A_{ij}^{(\beta)} T_{ij}^{(\beta)}\right]$. If we introduce a self-adjoint $\mathbf{T}$, this matrix Fourier transform can be simply expressed as:

$$\mathcal{F}_{\mathbf{A}}(\mathbf{T}) := \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\,\mathrm{Tr}\,\mathbf{AT}}\right], \tag{2.116}$$

and we have for the sum of independent matrices $\mathbf{C} = \mathbf{A} + \mathbf{B}$:

$$\mathcal{F}_{\mathbf{C}}(\mathbf{T}) = \mathcal{F}_{\mathbf{A}}(\mathbf{T})\,\mathcal{F}_{\mathbf{B}}(\mathbf{T})\,. \tag{2.117}$$

Now for the two matrices $\mathbf{A} = \mathbf{V}\mathbf{Diag}(\boldsymbol{a})\mathbf{V}^*$ and $\mathbf{B} = \mathbf{V}'\mathbf{Diag}(\boldsymbol{b})\mathbf{V}'^*$ of the additive randomized Horn problem, this simply corresponds to do an average over the group $\mathsf{O}_\beta(N)$ and up to a *Wick rotation* this by definition the additive spherical integral:

$$\mathcal{F}_{\mathbf{A}}(\mathbf{T}) = I^{(\beta)}(\mathbf{A}, \mathrm{i}\mathbf{T}) = \mathcal{I}^{(\beta)}(\boldsymbol{a}, \mathrm{i}\boldsymbol{t})\,, \tag{2.118}$$

with $I^{(\beta)}(\mathbf{A}, \mathbf{T})$ is defined by Eq. (2.76).

If we now do the inverse Fourier transform, we can write the distribution $f(\mathbf{C}|\boldsymbol{a}, \boldsymbol{b})$ of the *matrix elements* as:

$$f(\mathbf{C}|\boldsymbol{a}, \boldsymbol{b}) = \frac{1}{C_N}\int \mathrm{e}^{-\mathrm{i}\,\mathrm{Tr}\,\mathbf{CT}} I^{(\beta)}(\mathbf{A}, \mathrm{i}\mathbf{T}) I^{(\beta)}(\mathbf{A}, \mathrm{i}\mathbf{T})\mathrm{d}\mathbf{T}\,, \tag{2.119}$$

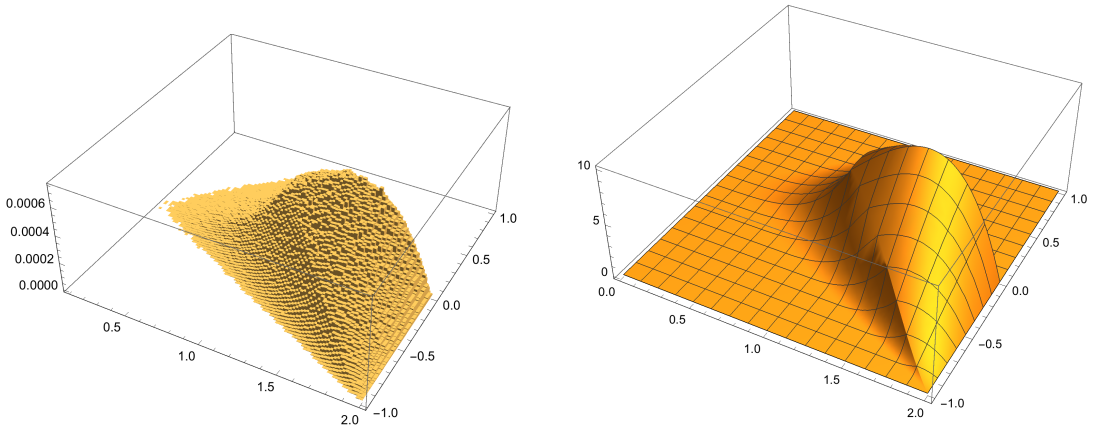
and if we now perform the change of variable from $\mathbf{T}$ to its eigenvalue decomposition, by Weyl's formula of Eq. (1.10) we get:

$$f(\mathbf{C}|\boldsymbol{a}, \boldsymbol{b}) \propto \int \mathcal{I}^{(\beta)}(\boldsymbol{a}, \mathrm{i}\boldsymbol{t})\mathcal{I}^{(\beta)}(\boldsymbol{b}, \mathrm{i}\boldsymbol{t}) I^{(\beta)}(\mathbf{C}, -\mathrm{i}\mathbf{Diag}(\boldsymbol{t}))\,|\Delta(\boldsymbol{t})|^\beta \mathrm{d}\boldsymbol{t}\,, \tag{2.120}$$

and by Weyl's formula again we get for the density of the eigenvalues the following expression

$$\boxed{\mathcal{P}^{(\beta)}(\boldsymbol{c}|\boldsymbol{a},\boldsymbol{b}) \propto |\Delta(\boldsymbol{c})|^{\beta} \int \mathcal{I}^{(\beta)}(\boldsymbol{a},\mathrm{i}\boldsymbol{t})\mathcal{I}^{(\beta)}(\boldsymbol{b},\mathrm{i}\boldsymbol{t})\mathcal{I}^{(\beta)}(\boldsymbol{c},-\mathrm{i}\boldsymbol{t})|\Delta(\boldsymbol{t})|^{\beta}\mathrm{d}\boldsymbol{t}\,.} \tag{2.121}$$

This integral representation is the matrix counterpart of Eq. (2.22) for the toy-model of the norm of the sum of two vectors. However, one does not have Sonine's formula to simplify this density in this case. In particular, because the integral is N-dimensional and the spherical integral is highly oscillatory, it is unclear how one can get the large N behavior of the distribution of the sum (the so-called free convolution of the following section) from this joint density. An illustration of this joint density is given in Fig. 2.5.



(a) empirical **(b)** theoretical

Figure 2.5: Plot of the empirical (**a**) and theoretical (**b**) joint densities of the eigenvalues of the sum of Eq. (2.115) for $N = 3$ and $\beta = 2$ with $\boldsymbol{a} = \boldsymbol{b} = (1, 0, -1)$, represented in the (c_1, c_2) plane. Figures taken from Ref. [191].

The multiplicative case -

The **multiplicative randomized Horn's problem** is the problem of finding the joint law $\mathcal{P}_{\times}^{(\beta)}(\boldsymbol{c}|\boldsymbol{a},\boldsymbol{b})$ of the eigenvalues of the matrix

$$\mathbf{C} = \sqrt{\mathbf{V}\mathbf{Diag}(\boldsymbol{a})\mathbf{V}^*}\ \mathbf{V}'\mathbf{Diag}(\boldsymbol{b})\mathbf{V}'^*\ \sqrt{\mathbf{V}\mathbf{Diag}(\boldsymbol{a})\mathbf{V}^*} \quad \text{with } \mathbf{V},\mathbf{V}' \sim \mathrm{Unif}[\mathsf{O}_{\beta}(N)] \tag{2.122}$$

from the knowledge of $\boldsymbol{a}$ and $\boldsymbol{b}$.

To tackle the multiplicative counterpart, one needs to define the corresponding multiplicative spherical. This object appeared in Refs. [90, 100, 190] and references therein and will be studied in more detail in Chapter 3. It is defined by:

$$\mathcal{I}_{\times}^{(\beta)}(\boldsymbol{a},\boldsymbol{t}) = \int_{\mathsf{O}_{\beta}(N)} \Delta_{\boldsymbol{t}}\left(\mathbf{V}\mathbf{A}\mathbf{V}^*\right)\mu_{\mathrm{Haar}}(\mathrm{d}\mathbf{V})\,, \tag{2.123}$$

where the generalized power function $\Delta_{\boldsymbol{t}}$ is given by

$$\Delta_{\boldsymbol{t}}(\mathbf{A}) := \left(\det \mathbf{A}_{(1)}\right)^{t_1-t_2} \dots \left(\det \mathbf{A}_{(N-1)}\right)^{t_{N-1}-t_N} (\det \mathbf{A})^{t_N} \ , \tag{2.124}$$

where $\mathbf{A}_{(i)}$ is the top left $(i \times i)$ corner of the matrix $\mathbf{A}$. One can express the joint density from this spherical, see Ref. [190].

The rectangular case -

This section deals with rectangular matrices.

The **rectangular randomized Horn's problem** is the problem of finding the joint law $\mathcal{P}_q^{(\beta)}(\boldsymbol{c}|\boldsymbol{a},\boldsymbol{b})$ of the singular values of the matrix

$$\mathbf{C} = \mathbf{V_1}\mathbf{Diag}_q(\boldsymbol{a})\mathbf{V}_2 + \mathbf{V}_1'\mathbf{Diag}_q(\boldsymbol{b})\mathbf{V}_2' \quad \text{with } \mathbf{V}_1, \mathbf{V}_1' \sim \mathrm{Unif}[\mathsf{O}_\beta(N)] \text{ and } \mathbf{V}_2, \mathbf{V}_2' \sim \mathrm{Unif}[\mathsf{O}_\beta(M)]\,, \tag{2.125}$$

from the knowledge of the singular values $\boldsymbol{a}$ and $\boldsymbol{b}$.

The corresponding Fourier transform is given by:

$$\mathcal{F}_{\mathbf{A}}(\mathbf{T}) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\mathrm{Tr}\,(\mathbf{A}\mathbf{T}^* + \mathbf{A}^*\mathbf{T})}\right] = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\mathfrak{Re}\mathrm{Tr}\,(\mathbf{A}\mathbf{T}^*)}\right] \ , \tag{2.126}$$

where the conjugate matrix $\mathbf{T} \in \mathsf{M}_{N,M}(\mathbb{K}_\beta)$. For the rectangular randomized Horn's of Eq. (2.125), the Fourier transform of $\mathbf{A} = \mathbf{V}_1\mathbf{Diag}_q(\boldsymbol{a})\mathbf{V}_2$ (and similarly for $\mathbf{B} = \mathbf{V}_1'\mathbf{Diag}_q(\boldsymbol{b})\mathbf{V}_2'$) is up to a factor 'i' simply given by the rectangular spherical integral of Eq. (2.110). If we repeat the same computation as in the additive case with now the Weyl's formula of Eq. (1.16), one gets for the joint laws of the singular values $\boldsymbol{c}$ of $\mathbf{C}$:

$$\mathcal{P}_q^{(\beta)}(\boldsymbol{c}|\boldsymbol{a},\boldsymbol{b}) \propto \quad |\Delta(\boldsymbol{c}^2)|^\beta \prod_{i=1}^N c_i^{\beta(M-N+1)-1} \times \int \mathcal{I}_q^{(\beta)}(\boldsymbol{a}, \mathrm{i}\boldsymbol{t})\mathcal{I}_q^{(\beta)}(\boldsymbol{b}, \mathrm{i}\boldsymbol{t})\mathcal{I}_q^{(\beta)}(\boldsymbol{c}, -\mathrm{i}\boldsymbol{t})|\Delta(\boldsymbol{t}^2)|^\beta \prod_{i=1}^N t_i^{\beta(M-N+1)-1} \mathrm{d}\boldsymbol{t}\,.$$

where $\mathcal{I}_q^{(\beta)}$ is given by Eq. (2.110).

2.6.2 Large β-invariant matrices and asymptotic freeness

In order to understand the large N behavior of the spectrum of the sum of two β-invariant matrices, let's consider the following illustrative example where

- $\mathbf{A} = \mathbf{Diag}(\boldsymbol{a})$ where the $a_i \overset{\text{i.i.d}}{\sim} \mu_A$ with μ_A a compact distribution with zero mean ;
- $\mathbf{B} = \mathbf{V}\mathbf{Diag}(\boldsymbol{b})\mathbf{V}^*$ where $\mathbf{V} \sim \mathrm{Unif}\,[\mathsf{O}_\beta(N)]$ and the $b_i \overset{\text{i.i.d}}{\sim} \mu_B$, where μ_B is also a compact distribution with zero mean.

As before, we construct a third matrix made of their sum:

$$\mathbf{C} = \mathbf{A} + \mathbf{B}\,, \tag{2.127}$$

and ask what is the behavior of the distribution μ_C of the eigenvalues of $\mathbf{C}$ for large N. A naive (but instructive way) to characterize this distribution is to compute *all* the average spectral moments of $\mathbf{C}$, that is the quantities

$$\tau\left(\mathbf{C}^k\right) = \tau\left(\mathbf{A}+\mathbf{B}\right)^k := \frac{1}{N}\mathbb{E}\,\mathrm{Tr}\left(\mathbf{A}+\mathbf{B}\right)^k\,, \tag{2.128}$$

for large N. Let's look at what happens for $k=4$, as it turns out to be the first non-trivial case. By linearity and cyclicity of the trace, we have:

$$\tau\left(\mathbf{C}^4\right) = \tau\left(\mathbf{A}^4\right) + \tau\left(\mathbf{B}^4\right) + 4\tau\left(\mathbf{A}^2\mathbf{B}^2\right) + 2\tau\left(\mathbf{ABAB}\right) + 4\left[\tau\left(\mathbf{A}^3\mathbf{B}\right) + \tau\left(\mathbf{AB}^3\right)\right]\,, \tag{2.129}$$

The first two terms of the RHS of Eq. (2.129) are simply the fourth moments of the distribution μ_A and μ_B. Let's look at the two following terms:

- By an elementary computation we have for the first one:

$$\tau(\mathbf{A}^2\mathbf{B}^2) = \frac{1}{N}\mathbb{E}\sum_{i,k} a_i^2 b_k^2 |V_{ik}|^2\,, \tag{2.130}$$

$$\tau(\mathbf{A}^2\mathbf{B}^2) = \tau(\mathbf{A}^2)\tau(\mathbf{B}^2)\,\mathbb{E}\left(\frac{1}{N}\sum_{i,k}|V_{ik}|^2\right)\,, \quad \text{(by linearity of } \tau \text{ and independence)} \tag{2.131}$$

$$\tau(\mathbf{A}^2\mathbf{B}^2) = \tau(\mathbf{A}^2)\tau(\mathbf{B}^2)\,, \qquad \text{(since } \mathbf{V}\mathbf{V}^* = \mathbf{I}\text{)}\,. \tag{2.132}$$

- Similarly, for the second term we have:

$$\tau\left(\mathbf{ABAB}\right) = \frac{1}{N}\sum_{i,j,k,l}\mathbb{E}\left[a_i a_j\right]\mathbb{E}\left[b_k b_l\right]\mathbb{E}V_{ik}V_{jk}V_{jl}V_{il}\,, \tag{2.133}$$

$$\tau\left(\mathbf{ABAB}\right) = \tau(\mathbf{A}^2)\tau(\mathbf{B}^2)\,\mathbb{E}\left[\frac{1}{N}\sum_{i,k}|V_{ik}|^4\right] \quad \text{(by independence and zero mean)}\,, \tag{2.134}$$

$$\tau\left(\mathbf{ABAB}\right) \to 0\,, \tag{2.135}$$

  where to get the last asymptotic result, we have used the fact that the column vectors $\boldsymbol{v}^{(i)}$ of an orthogonal/unitary matrix are asymptotically *delocalized* that is their *Inverse Participation Ratio* (IPR) of order $k>1$ goes to 0 as $N\to\infty$:

$$\mathrm{IPR}_k\left[\boldsymbol{v}^{(i)}\right] := \sum_{j=1}^{N}|v_j^{(i)}|^{2k} \to 0\,. \tag{2.136}$$

Similarly, one can easily show that the terms in the bracket of Eq. (2.129) are always zero since the distributions μ_A and μ_B have mean zero and as a result, one can compute the fourth moment of $\mathbf{C}$. Importantly, this simple derivation has shed light on the (highly) non-commutative nature of the matrices $\mathbf{A}$ and $\mathbf{B}$ for large N, since we have:

$$\tau(\mathbf{A}^2\mathbf{B}^2) \neq \tau\left(\mathbf{ABAB}\right) \approx 0\,, \tag{2.137}$$

regardless of the specific distributions μ_A and μ_B (provided they are centered). As a result, one can think of two matrices $\mathbf{A}$ and $\mathbf{B}$, one of them being β-invariant, as two *maximally non-commutative objects*.

If now one wants to push this computation to higher orders $k > 4$, one first needs to expand the k^{th} average moments in terms of non-commutative products of $\mathbf{A}$ and $\mathbf{B}$ - and for large k, this is a daunting task - and then understand which terms contribute to the sum and which terms vanish. It appears that this problem can be greatly simplified thanks to the power-full theory of *free probability*, whose link with random matrices is given by the following theorem due to VOICULESCU

Result 2.12 (*Asymptotic freeness of β-invariant matrices [183, 182, 184]*)

*If $\mathbf{A}, \mathbf{B}$ are β-invariant in law (that is we recall $\mathbf{A} \overset{\text{in law}}{=} \mathbf{V}\mathbf{A}\mathbf{V}^*$ for any $\mathbf{V} \in \mathsf{O}_\beta(N)$), then $\mathbf{A}$ and $\mathbf{B}$ are **asymptotically free**: for any positive integer n and any polynomials $P_1, \dots, P_{2n}$ if we denote by $\tau(.) := \mathbb{E}\mathrm{Tr}\,(.)/N$, we have*

$$\tau\left(P_1(\mathbf{A})\, P_2(\mathbf{B}) \dots P_{2n-1}(\mathbf{A}) P_{2n}(\mathbf{B})\right) \to 0 \tag{2.138}$$

whenever $\tau(P_{2j-1}(\mathbf{A})) \to 0$ and $\tau(P_{2j}(\mathbf{B})) \to 0$ for all $j \in \{1, \dots, N\}$.

Loosely speaking, freeness can be seen as the matrix counterpart of the orthogonality relation for vectors, with the notion of 'rotationally invariant vectors' being replaced by 'β-invariant matrices'. In the following of this section we will review the practical consequences of the freeness of two matrices and refers to Refs. [183, 145, 149] for the theoretical foundations of free probability. In particular, we will always think of free random variables as large β-invariant random matrices, rather than using the abstract language of non-commutative algebra. Freeness can be seen as the non-commutative analog of the usual notion of *independence* between classical random variables on the real line. For independent random variables, the 'rule' to compute the distribution of their sum (resp. their product if they are positive) is given by classical convolution (resp. Mellin convolution), which can be seen as a consequence of the multiplicative property of their moment generating function/Mellin transform. For free matrices there exist analogous *free probability transforms* which give a 'rule' to compute the distribution of the sum or the product of free matrices, and the corresponding operation is known as the *free convolution* and is described in the rest of this section.

2.6.3 Free probability transforms, free cumulants, and non-crossing partitions

Sum of free matrices and R-transform -

For a compactly supported measure μ_A, its **R-transform** denoted by $\mathcal{R}_A \equiv \mathcal{R}_{\mu_A}$ is the function defined on a neighborhood of the origin by:

$$\mathcal{R}_A(y) := g_A^{\langle -1 \rangle}(y) - \frac{1}{y}\,, \tag{2.139}$$

where $g_A^{\langle -1 \rangle}$ is the inverse of the Stieltjes transform g_A of the distribution μ_A.

Remark *(Elementary transformation of distribution and corresponding R-transform)*. Using properties of the Stieltjes transform, one can easily deduce the following properties for the R-transform.

- **Shift**: If one does a shift of the distribution $\mu_A \to \mu_A(. - \delta)$ which corresponds to the asymptotic operation of adding δ times the identity matrix: $\mathbf{A} \to \mathbf{A} + \delta\mathbf{I}$, the corresponding R-transform is given by:

$$\mathcal{R}_{A+\delta 1}(y) = \mathcal{R}_A(y) + \delta\,. \tag{2.140}$$

- **Scaling**: If one re-scales a matrix $\mathbf{A} \to \delta\mathbf{A}$, then the corresponding R-transform is given by:

$$\mathcal{R}_{\delta A}(y) = \delta\mathcal{R}_A(\delta y)\,. \tag{2.141}$$

- **monotonicity**: the R-transform is a continuously increasing function in its domain of definition.

⌟

We give here a few explicit examples of the R-transform of some distributions, which we will encounter several times in this thesis.

Examples. The following result can be easily obtained by computing the Stieltjes transforms (and the inverse) of each distribution:

- the R-transform of the semi-circle distribution $\mu_{\mathrm{sc}(\sigma)}$ given by Eq. (1.40) with variance σ^2 is given by:

$$\mathcal{R}_{\mathrm{sc}(\sigma)}(y) = \sigma^2 y\,. \tag{2.142}$$

- The R-transform of the Marčenko-Pastur distribution of $\mu_{\mathrm{MP}(q)}$ given by Eq. (1.43) is given by

$$\mathcal{R}_{\mathrm{MP}(q)}(y) = (1 + qy)^{-1}\,. \tag{2.143}$$

- the R-transform of the distribution $\mu_{\mathrm{Ber}}(\lambda) = \delta(\lambda + 1)/2 + \delta(\lambda - 1)/2$ is given by:

$$\mathcal{R}_{\mathrm{Ber}}(y) = \frac{\sqrt{1 + 4y^2} - 1}{2y}\,. \tag{2.144}$$

■

The R-transform plays the role of the logarithm of the moment generating function of classical probability theory since we have the following result.

Result 2.13 (*Sum of free matrices and free convolution*)

If $(\mathbf{A},\mathbf{B})$ *is a couple of two self-adjoint matrices being asymptotically free, with LSD given respectively by* μ_A *and* μ_B*, then the LSD* μ_C *of their free sum* $\mathbf{C} = \mathbf{A} + \mathbf{B}$ *is given as the*

unique probability measure solution of:

$$\mathcal{R}_C(y) = \mathcal{R}_A(y) + \mathcal{R}_B(y) \tag{2.145}$$

for all y in the complex plane close enough to the origin. The distribution μ_C is known as the ***free convolution*** *of μ_A and μ_B and is denoted by $\mu_A \boxplus \mu_B$.*

To illustrate this result, let's look at two practical examples:

Example *(Sum with a matrix from a Gaussian ensemble and Burger's equation).* If we come back to the 'DBM' case where $\mathbf{C} = \mathbf{B} + t\mathbf{X}$ with $\mathbf{X}$ taken from a Gaussian ensemble with unit variance, then since $\mathbf{X}$ is β-invariant, the LSD μ_C is given by the free convolution of μ_B with the semi-circle distribution. Since the R-transform of the semi-circle distribution is given by Eq. (2.142) we have:

$$\mathcal{R}_C(y) = ty + \mathcal{R}_B(y)\,, \tag{2.146}$$

and if we use the definition Eq. (2.139) of the R-transform, this can be equivalently expressed as:

$$g_C^{\langle -1\rangle}(y) - ty = g_B^{\langle -1\rangle}(y)\,, \tag{2.147}$$

which if we apply $g_B(.)$ gives back Eq. (2.88), as expected. ∎

Example *(Free sum of Bernoulli distribution and the arcsine law).* Next, consider the case where $\mathbf{A}$ is a deterministic diagonal matrix with half of its eigenvalues being given by $1/2$ and the other half are equal to $-1/2$ and similarly let's define by $\mathbf{B} = \mathbf{VAV}^*$ and construct their free sum $\mathbf{C} = \mathbf{A} + \mathbf{B}$, then since $\mathbf{A}$ corresponds to scaling by $\delta = 1/2$ of the discrete measure on 1 and -1 with R-transform given by Eq. (2.144), we have:

$$\mathcal{R}_C(y) = \frac{\sqrt{1+y^2}-1}{y}\,, \tag{2.148}$$

which gives after inversion, that the free convolution is given by the arcsine law:

$$\left(\frac{1}{2}\delta(x-1/2) + \frac{1}{2}\delta(x+1/2)\right) \boxplus \left(\frac{1}{2}\delta(x-1/2) + \frac{1}{2}\delta(x+1/2)\right) = \frac{1}{\pi}(1-\lambda^2)^{-1/2}\mathbb{I}_{[-1,1]}\,. \tag{2.149}$$

An illustration of this free convolution is given in Fig. 2.6. ∎

Free cumulants -

Now that we have defined the free convolution and its linearizing transform (the R-transform) it appears also natural to define the **free cumulants**, that is the polynomials κ_k in the moments $m_1, \ldots, m_k$ of a distribution μ_A such that they satisfy the three following properties:

- **additivity**: $\kappa_k\left[\mu_A \boxplus \mu_B\right] = \kappa_k\left[\mu_A\right] + \kappa_k\left[\mu_B\right]$,
- **homogeneity**: $\kappa_k\left[\frac{1}{\delta}\mu_A\left(\frac{\cdot}{\delta}\right)\right] = \delta^k \kappa_k\left[\mu_A\right]$,
- **leading term**: κ_k is a polynomial in the first k moments with leading term m_k.

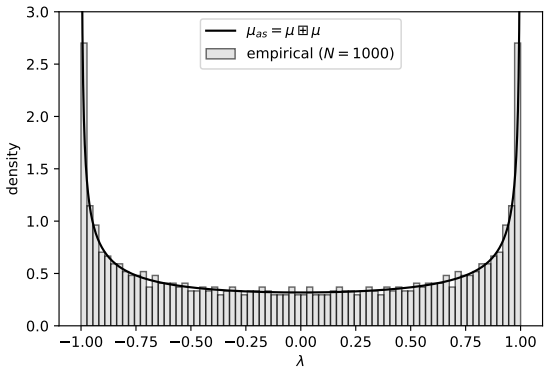


Figure 2.6: Plot of the free convolution of $\mu = \frac{1}{2}\delta(\cdot - 1/2) + \frac{1}{2}\delta(\cdot + 1/2)$ with itself compared with the histogram of the eigenvalues of $\mathbf{A} + \mathbf{OAO}^{\mathsf{T}}$, where $\mathbf{O} \sim \mathrm{Unif}\,[O(N)]$, with $N = 1000$ and $\mathbf{A}$ is a diagonal matrix with half its entries equal to $1/2$ and the other half is equal to $-1/2$.

It turns out that these free cumulants are given as the coefficients of the power series of the R-transform:

$$\mathcal{R}_A(y) = \sum_{k=1}^{\infty} \kappa_k y^{k-1}\,, \tag{2.150}$$

where $\kappa_k \equiv \kappa_k\,[\mu_A]$ is the free cumulant of order k.

Conversely the k^{th} moment, m_k of a distribution is also a polynomial in the first k free cumulants. The m_k, seen as a multivariate polynomial in the κ_i, can be written as a sum over index $j_1, \ldots, j_k$ involving product of the form $\kappa_1^{j_1} \ldots \kappa_k^{j_k}$ where the j_i are constrained by $j_1 + \cdots + j_k = k$ by homogeneity of the moments. As one recognizes a natural property of a *partition*, one may hope that the relation between moment and free cumulant is given by a sum over a certain type of *partitions* and this is indeed the case, see for example Ref. [149]:

Result 2.14 (*Moment-free-cumulant relation*)

The moment m_k of a distribution is given in terms of the first k cumulants by:

$$m_k = \sum_{\boldsymbol{\pi} \in \mathcal{NC}[k]} \kappa_{\boldsymbol{\pi}} \tag{2.151}$$

where $\mathcal{NC}[k]$ denotes the set of all non-crossing partitions of k.

Let's recall that $\pi = \{B_1, \ldots B_r\}$ is a partition of k - denoted by $\pi \in \mathcal{P}[k]$ - of length r, if its blocks $B_1, \ldots, B_r$ are pairwise distinct and non-empty sets of $\{1, \ldots, k\}$ such that $B_1 \cup \cdots \cup B_r = \{1, \ldots, k\}$ and $f_\pi := \prod_{i=1}^{r} f_{|B_i|}$. A partition is said to be crossing whenever there exist at least two blocks B_i, B_j with $a, b \in B_i$ and $c, d \in B_j$ such that $a < c < b < d$. Note that this involved definition admits a natural graphical interpretation, see Fig. 2.7.

This is a different combinatorial formula than the one of classical probability, where the sum is run over *all* partitions of k.

The first four moments are given in terms of the first four free cumulants by:

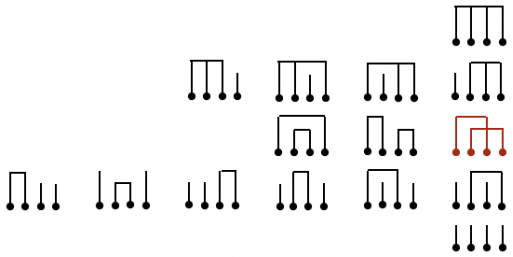


Figure 2.7: Representation of the 15 partition of 4 where elements of the same block are linked. For example, the top partition of this figure has only one block and is given by $\{1,2,3,4\}$ while the bottom partition has four individual blocks and is given by $\{\{1\},\{2\},\{3\},\{4\}\}$. The partition in red is $\{\{1,3\},\{2,4\}\}$ and is the only crossing partition of 4 and so it does not count in the moment-free cumulant relation.

- $m_1 = \kappa_1$,
- $m_2 = \kappa_2 + \kappa_1^2$,
- $m_3 = \kappa_3 + 3\kappa_2\kappa_1 + \kappa_2^3$,
- $m_4 = \kappa_4 + 4\kappa_3\kappa_1 + \mathbf{2}\cdot\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4$.

Conversely, the first four free cumulants are given in terms of the first four moments m_k by:

- $\kappa_1 = m_1$,
- $\kappa_2 = m_2 - m_1^2$,
- $\kappa_3 = m_3 - 3m_1m_2 + 2m_1^3$,
- $\kappa_4 = m_4 - \mathbf{2}\cdot m_2^2 + 10m_2m_1^2 - 5m_1^4$.

In the classical setting, all the coefficients of the first four cumulants are the same, except for the ones in bold (2) which are replaced by 3. We refer the reader to the books of SPEICHER [149] for more on this combinatorial interpretation of the free convolution.

Product of free matrices and S-transform -

We now turn to the free product of matrices and define the analog of the R-transform for this operation.

For a measure μ_A with support included on the positive real line, its **(modified) S-transform** denoted by $\tilde{\mathcal{S}}_A \equiv \tilde{\mathcal{S}}_{\mu_A}$ is the function defined on a neighborhood of the origin by:

$$\tilde{\mathcal{S}}_A(y) := \frac{y}{y+1} t_A^{\langle -1\rangle}(y)\,, \tag{2.152}$$

where $t_A^{\langle -1\rangle}$ is the inverse of the T-transform t_A of the distribution μ_A, defined by Eq. (1.29).

Remark *(Convention for the S-transform).* The usual S-transform is usually defined in the literature as the inverse of the (modified) S-transform defined in this thesis:

$$\mathcal{S}_A(y) := \frac{1}{\tilde{\mathcal{S}}_A(y)} = \frac{y+1}{y\, t_A^{\langle -1\rangle}(y)}\,. \tag{2.153}$$

Our choice of the convention for defining the S-transform with Eq. (2.152) rather than Eq. (2.153) comes from [125] and turns out to be more appropriate in our study. ⌟

The S-transform is multiplicative for the (asymptotic) product of free matrices:

Result 2.15 (*Product of free matrices and the multiplicative free convolution*)

If $(\mathbf{A},\mathbf{B})$ is a couple of two positive self-adjoint matrices being asymptotically free, with LSD given respectively by μ_A and μ_B , then the LSD μ_C of their free product $\mathbf{C} = \sqrt{\mathbf{A}}\mathbf{B}\sqrt{\mathbf{A}}$ is given as the unique probability measure solution of:

$$\tilde{\mathcal{S}}_C(y) = \tilde{\mathcal{S}}_A(y)\tilde{\mathcal{S}}_B(y)\,, \tag{2.154}$$

*for all y in the complex plane close enough to the origin. The distribution μ_C is known as the **multiplicative free convolution** of μ_A and μ_B and is denoted by $\mu_A \boxtimes \mu_B$.*

Let's mention that this can be equivalently as:

$$\log \tilde{\mathcal{S}}_C(y) = \log \tilde{\mathcal{S}}_A(y) + \log \tilde{\mathcal{S}}_B(y)\,. \tag{2.155}$$

In other words, the logarithm of the S-transform is the linearizing transform of the free multiplicative convolution and this transformation will appear somehow more naturally in this thesis. We give below a few examples of the free multiplicative convolution.

Example *(S-transform of the DGBM).* For the limiting density μ_{M_t} of the DGBM of Eq. (2.98), one can show that its S-transform is given by

$$\tilde{\mathcal{S}}_{M_t}(y) = \mathrm{e}^{ty} \tag{2.156}$$

and so its logarithm is simply given by ty. Thus, one can think of this distribution as the multiplicative counterpart of the semi-circle distribution with variance t, since for the latter its R-transform is given ty. ∎

Example *(Watcher and Marčenko-Pastur distributions).* A matrix taken from the Jacobi ensemble is obtained as a shift and the inverse of the product of a Wishart matrix with the inverse of another independent Wishart matrix, see Eq. (1.45). Since Wishart matrices are β-invariant, this product is given asymptotically by the free multiplicative convolution, and we can write the Watcher distribution of Eq. (1.47) (the LSD of a Jacobi ensemble) as:

$$\mu_{\mathrm{Wat}(q_1,q_2)}(\lambda) = \Big(q_1\mu_{\mathrm{MP}(q_1)}(q_1.) \boxtimes \big(q_2\mu_{\mathrm{MP}(q_2)}(q_2.)\big)^{[-1]} \boxplus \delta(.-1)\Big)^{[-1]} \tag{2.157}$$

where if μ is the (limiting) distribution of $\mathbf{A}$, $\mu^{[-1]}$ is the (limiting) distribution of $\mathbf{A}^{-1}$.

∎

We conclude this paragraph on the S-transform with an identity relating it to the R-transform, which can be seen as a consequence of the fact that the sum is a limiting case of the product for self-adjoint matrices.

Result 2.16 (*Relation between R-transform and S-transform*)

If $\mathcal{R}_A$ is the R-transform of the distribution μ_A and $\tilde{\mathcal{S}}_{\exp(\epsilon A)}$ the S-transform of $\mu_{\exp(\epsilon A)}$ (that is the push-forward of μ_A by $x \mapsto \exp(\epsilon\, x)$), then we have:

$$\frac{1}{\epsilon} \log \tilde{\mathcal{S}}_{\exp(\epsilon A)} \left(\frac{\theta}{\epsilon} \right) \xrightarrow[\epsilon \to 0^+]{} \mathcal{R}_A(\theta)\,. \tag{2.158}$$

As far as I know, this relation does not seem to be well known, so I give below a short proof.

PROOF: We can write the definition (2.152) of the S-transform as the implicit equation:

$$\theta = t_A \left(\tilde{\mathcal{S}}_A(\theta) \left(1 + \frac{1}{\theta} \right) \right)\,, \tag{2.159}$$

and if we now do the changes $\theta \to \theta/\epsilon$ and $A \to \exp(\epsilon A)$, this gives:

$$\frac{\theta}{\epsilon} = t_{\exp(\epsilon A)} \left(\tilde{\mathcal{S}}_{\exp(\epsilon A)} \left(\frac{\theta}{\epsilon} \right) \left(1 + \frac{\epsilon}{\theta} \right) \right)\,. \tag{2.160}$$

Now since our goal is to show that the function $P_{\epsilon,A}(.) := \epsilon^{-1} \log \tilde{\mathcal{S}}_{\exp(\epsilon A)}(./\epsilon)$ converges point-wise to the R-transform, we can write Eq. (2.160) in terms of $P_{\epsilon,A}$:

$$\theta = \epsilon\, t_{\exp(\epsilon A)} \left(\mathrm{e}^{\epsilon P_{\epsilon,A}(\theta)} \left(1 + \frac{\epsilon}{\theta} \right) \right)\,, \tag{2.161}$$

$$\theta = \epsilon \int \frac{\mathrm{e}^{\epsilon\lambda}}{\mathrm{e}^{\epsilon P_{\epsilon,A}(\theta)} \left(1 + \frac{\epsilon}{\theta} \right) - \mathrm{e}^{\epsilon\lambda}} \mu_A(\mathrm{d}\lambda)\,, \qquad \text{(by definition (1.29) of the T-transform)} \tag{2.162}$$

$$\theta = \int \frac{1}{P_{\epsilon,A}(\theta) + 1/\theta - \lambda} \mu_A(\mathrm{d}\lambda) + \mathcal{O}(\epsilon)\,, \tag{2.163}$$

$$\theta = g_A \left(P_{\epsilon,A}(\theta) + 1/\theta \right) + \mathcal{O}(\epsilon)\,. \qquad \text{(by definition (1.26) of the Stieltjes transform)} \tag{2.164}$$

Now if we invert this last equation (that is, apply $g_A^{\langle -1 \rangle}(.)$ to this equation), we get:

$$\lim_{\epsilon \to 0^+} P_{\epsilon,A}(\theta) = g_A^{\langle -1 \rangle}(\theta) - \frac{1}{\theta}\,, \tag{2.165}$$

which is nothing else than the definition (2.139) of the R-transform and this concludes the proof. □

There exist other relations known as *subordination relations* relating the two transforms, and I refer to [153] for their precise expression.

We summarize the dictionary between the 'classical/commutative world' and the 'free world' in the following table

Classical Probability	Free probability
X, Y real random variables	A, B infinite random matrices
distribution μ_X, μ_Y	LSD μ_A, μ_B
independent	free
$X + Y \Leftrightarrow \mu_X * \mu_Y$	$A + B \Leftrightarrow \mu_A \boxplus \mu_B$
$C_X(t) := \log \mathbb{E}[\mathrm{e}^{tX}]$	$\mathcal{R}_A(t)$ def. by Eq. (2.139)
$C_X(t) = \sum_{k=1}^{\infty} c_k/k!\, t^k$	$\mathcal{R}_A(t) = \sum_{k=1}^{\infty} \kappa_k\, t^{k-1}$
$m_k = \sum_{\boldsymbol{\pi} \in \mathcal{P}[k]} c_{\boldsymbol{\pi}}$	$m_k = \sum_{\boldsymbol{\pi} \in \mathcal{NC}[k]} \kappa_{\boldsymbol{\pi}}$
$X \cdot Y \Leftrightarrow \mu_{\log X} * \mu_{\log Y}$	$\sqrt{B}A\sqrt{B} \Leftrightarrow \mu_A \boxtimes \mu_B$
$\mathcal{M}e_X(t) := \mathbb{E}\left[X^t\right]$	$\tilde{\mathcal{S}}_A(t)$ def. by Eq. (2.152)

where the c_k's are the cumulants for the classical convolution $(f * g)(x) = \int f(x-y)g(y)\mathrm{d}y$ and when considering the product of random variables, we have implicitly imposed $X, Y > 0$ (and similarly the support of μ_A and μ_B is on the positive real line for the free case).

Bi-free sum of rectangular matrices and the rectangular C-transform -

This section deals with rectangular matrices.

In this paragraph, we briefly explain the rectangular case. To do so, we first need to introduce the function $U(.)$ defined by:

$$U(y) := \frac{-1 - q + \sqrt{(1-q)^2 + 4qy^2}}{2q}\,. \tag{2.166}$$

Note that $U(.)$ is an increasing function of y whose inverse transform is given by the simple formula:

$$U^{\langle -1\rangle}(z) = \sqrt{(1+z)(1+qz)}\,. \tag{2.167}$$

The rectangular C-transform (with shape ratio q) of the LSVD μ_A of a rectangular matrix $\mathbf{A}$ is defined as:

$$\tilde{\mathcal{C}}_A^{(q)}(y) := \frac{U\left(y\, d_A^{\langle -1\rangle}(y)\right)}{y}\,, \tag{2.168}$$

with U given by Eq. (2.166), and $d_A^{\langle -1\rangle}$ is the inverse of the D-transform defined by Eq. (1.32).

Remark *(Convention for the C-transform).* The standard convention for the *rectangular R-transform*, $\mathcal{R}_A^{(q)}$, is related to our rectangular C-transform $\tilde{\mathcal{C}}_A^{(q)}$, via,

$$\tilde{\mathcal{C}}_A^{(q)}(t) =: \mathcal{R}_A^{(q)}(t^2)/t\,, \tag{2.169}$$

As we will see, the convention $\tilde{\mathcal{C}}_A^{(q)}$ will appear more naturally in this thesis. ⌟

Although rectangular matrices are not free, one can also determine the LSVD of their 'bi-free' sum, thanks to the rectangular C-transform, see Ref. [20]:

Result 2.17 (*Bi-free sum of rectangular matrices and the rectangular free convolution*)

If $\mathbf{A}$ and/or $\mathbf{B}$ are two $(N\times M)$ matrices bi-β-invariant in law, with LSVD given respectively by μ_A and μ_B, then the LSVD μ_C of their bi-free sum $\mathbf{C}=\mathbf{A}+\mathbf{B}$ is given as the unique probability measure solution of:

$$\tilde{\mathcal{C}}_C^{(q)}(y)=\tilde{\mathcal{C}}_A^{(q)}(y)+\tilde{\mathcal{C}}_B^{(q)}(y)\,. \tag{2.170}$$

*The distribution μ_C is known as the **rectangular free convolution** of μ_A and μ_B and is denoted by $\mu_A \boxplus\!\!\boxplus_q \mu_B$.*

In the following two remarks, we briefly describe how this rectangular convolution degenerates in the limiting cases $q\to 0$ and $q\to 1$.

Remark *(long matrices ($q\to 0$) and additive free convolution).* In the limit $q\to 0$, corresponding to the case of $(N\times M)$ rectangular *long* matrices with $1\ll N\ll M$, we have for the function U and the D-transform,

$$U(y)\underset{q\to 0}{\to} y^2-1\,, \tag{2.171}$$

and in this case, since we have Eq. (1.36), the inverse of the D-transform is given by:

$$d_A^{\langle -1\rangle}(y)\underset{q\to 0}{\to}\sqrt{g_{AA^*}^{\langle -1\rangle}(y^2)}\,. \tag{2.172}$$

where g_{AA^*} is the Stieltjes transform of $\mu_{AA^*}(.)=\frac{\mu_A(\sqrt{.})}{2\sqrt{.}}$. As a consequence for long matrices, the rectangular C-transform is related to the R-transform by

$$\tilde{\mathcal{C}}_A^{(0)}(y)=y\,\mathcal{R}_{AA^*}\left(y^2\right)\,, \tag{2.173}$$

and so by the linearizing property of the C-transform and the R-transform we have:

$$\tilde{\mathcal{C}}_C^{(0)}(y)=y\,\mathcal{R}_{AA^*+BB^*}\left(y^2\right)\,. \tag{2.174}$$

In other words for long matrices we have $s_i(\mathbf{A}+\mathbf{B})\approx\sqrt{\lambda_i\left(\mathbf{A}\mathbf{A}^*+\mathbf{B}\mathbf{B}^*\right)}$, which is of course not true for $q\neq 0$. ⌟

Remark *(square matrices ($q=1$) and symmetrized density).* For $q=1$, corresponding to (asymptotic) square matrices, the function U is simply given by:

$$U(y)\underset{q\to 1}{\to} y-1\,, \tag{2.175}$$

and the D-transform considerably simplifies into the Stieltjes $\mu_{\hat{A}}$ is the symmetrized density of ρ_A see Eq. (1.37). The C-transform of Eq. (2.168) reads in this case:

$$\tilde{\mathcal{C}}_A^{(1)}(y)=\mathcal{R}_{\widehat{A}}(y)\,, \tag{2.176}$$

and by the linearizing property, we have that the LSVD of the matrix $\mathbf{C}$ is given as the unique probability measure on $\mathbb{R}_+$ such that:

$$\tilde{\mathcal{C}}_C^{(1)}(y)=\mathcal{R}_{\widehat{A}+\widehat{B}}(y)\,. \tag{2.177}$$

In other words, the singular values of the sum of two (bi-free) square matrices are given asymptotically by the additive free convolution of their respective symmetrized singular value densities. ⌟

We now give the C-transform for Gaussian rectangular matrices.

Example *(Gaussian rectangular random matrices).* Let's consider the case of Gaussian rectangular matrices with LSVD given Eq. (1.51). Using the expression of Eq. (1.44) for the Stieltjes transform of the Marčenko-Pastur distribution, one gets the following expression for the D-transform (for $z > 0$):

$$d_A(z) = \frac{1}{\sigma^2}\sqrt{\frac{z^2 - (1+q)\sigma^2 - \sqrt{z^4 - 2(1+q)\sigma^2 z^2 + (1-q)^2\sigma^4}}{2q}}\,, \tag{2.178}$$

whose inverse is given by

$$d_A^{\langle -1\rangle}(y) = \frac{\sqrt{(1+\sigma^2 y^2)(1+q\sigma^2 y^2)}}{y}\,. \tag{2.179}$$

The argument inside the square-root function is nothing else than the inverse $U^{\langle -1\rangle}$ evaluated at $\sigma^2 y^2$, see Eq. (2.167). Using Eq. (2.168), the rectangular C-transform of the Gaussian rectangular matrix is given simply by:

$$\tilde{\mathcal{C}}_A^{(q)}(y) = \sigma^2 y\,. \tag{2.180}$$

■

2.6.4 Free central limit theorems

In this section, we would like to establish the limit theorems associated with free convolution. Let's recall that for *classical probability*, if the X_i's are i.i.d random variables taken from a measure μ with mean zero and variance one, then for large n the sum $S_n := (X_1 + \dots X_n)/\sqrt{n}$ converges to a standard Gaussian distribution. This is due to the fact that the classical cumulants of S_n are given as the sum of the ones of the X_i's and the scaling by n makes only the second (classical) cumulant $c_2 = 1$ of S_n non-vanishing for large n, and this corresponds to the cumulants of a standard Gaussian distribution. Equivalently, this can be written in terms of classical convolutions as

$$\underbrace{\sqrt{n}\mu\left(\sqrt{n}x\right) * \cdots * \sqrt{n}\mu\left(\sqrt{n}x\right)}_{n \text{ times}} \xrightarrow[n\to\infty]{} \mu_{\mathcal{N}(0,1)}(x) := \frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,. \tag{2.181}$$

Now for the free-convolution, the distribution which has only $\kappa_2 \neq 0$ as a non-zero *free* cumulant is the semi-circle distribution, from which we immediately deduce the following result.

Result 2.18 (*Free central limit theorem*)

For μ a distribution with zero mean and variance one, if we re-scale it such that its variance is $1/n$ and performs the free convolution with itself n times, we have asymptotically in n:

$$\underbrace{\sqrt{n}\mu\left(\sqrt{n}x\right) \boxplus \cdots \boxplus \sqrt{n}\mu\left(\sqrt{n}x\right)}_{n \text{ times}} \xrightarrow[n\to\infty]{} \mu_{\mathrm{sc}(1)}(x) := \frac{\sqrt{4-x^2}}{2\pi}\mathbb{I}_{[-2,2]}\,. \tag{2.182}$$

Another well-known limit theorem of classical probability is the law of rare events or the Poisson limit theorem, which states that the sum of independent Bernoulli random variables with a vanishing rate of success, converges to a Poisson random variable, or equivalently if we denote by $\mu_{\mathrm{Ber}(\lambda/n)} := \left(1-\frac{\lambda}{n}\right)\delta(x-0)+\frac{\lambda}{n}\delta(x-1)$, we have:

$$\underbrace{\mu_{\mathrm{Ber}(\lambda/n)} * \cdots * \mu_{\mathrm{Ber}(\lambda/n)}}_{n \text{ times}} \xrightarrow[n\to\infty]{} \mu_{\mathrm{Poi}}(x) := \sum_{k=0}^{\infty} \frac{\lambda^k \mathrm{e}^{-\lambda}}{k!}\delta(x-k)\,. \tag{2.183}$$

The free counterpart of the Poisson distribution is given by the Marčenko-Pastur distribution, see the following result.

Result 2.19 (*Free Poisson limit theorem*)

For $\mu := \left(1-\frac{1}{qn}\right)\delta(x-0)+\frac{1}{qn}\delta(x-q)$ a re-scaled Bernoulli distribution with a small rate of success, we have:

$$\underbrace{\mu \boxplus \cdots \boxplus \mu}_{n \text{ times}} \xrightarrow[n\to\infty]{} \mu_{\mathrm{MP}(q)}(x) \tag{2.184}$$

where $\mu_{\mathrm{MP}(q)}$ is given by Eq. (1.43).

2.7 Finite free convolution

The free convolution of the previous section is an asymptotic ($N \to \infty$) operation on the spectrum of two random matrices, and we now describe its 'finite' counterpart, known as the finite free convolution (FFC in short). The FFC has been introduced by Marcus, Spielman and Srivastava to tackle combinatorial problems in linear algebra, see Refs. [124, 126, 129, 127] and has then been noticed to share many similarities with the free convolution, see Refs. [128, 125]. As we will briefly see, the FFC can be understood as the sum of β-ensembles in the low-temperature regime ($\beta \to \infty$, with N fixed) and just like the description of low-temperature is naturally encoded in a monic polynomial, see Sec. 1.6, the FFC can be described as a simple operation on such monic polynomials.

2.7.1 Introduction and preliminary definitions

To fix things, let's denote by

$$P_{\boldsymbol{a}}(x) := \prod_{i=1}^{N}(x-a_i)\,, \tag{2.185}$$

the monic polynomial of degree N with roots $\boldsymbol{a} \in \mathbb{R}^N$. The discrete measure associated with its roots is denoted by:

$$\mu_{\boldsymbol{a}}(\lambda) := \frac{1}{N}\sum_{i=1}^{N}\delta(\lambda-a_i)\,. \tag{2.186}$$

We can represent the monic polynomial $P_{\boldsymbol{a}}$ as:

$$P_{\boldsymbol{a}}(x) = \sum_{k=0}^{N} \alpha_k (-1)^k x^{N-k} \,, \tag{2.187}$$

Because $P_{\boldsymbol{a}}$ is monic, we have $\alpha_0 = 1$. For other values of k, this coefficient is explicitly given in terms of the roots a_i by the k^{th} *elementary symmetric polynomial*:

$$\alpha_k \equiv \alpha_k(\boldsymbol{a}) = \mathsf{e}_k(\boldsymbol{a}) := \sum_{1 \le j_1 < \cdots < j_k \le N} a_{j_1} \dots a_{j_k} \,. \tag{2.188}$$

The coefficients α_k can be expressed in terms of the moments $m_k := \sum_{i=1}^{N} a_i^k / N$ of the measure $\mu_{\boldsymbol{a}}$ thanks to *Newton's identities* by either a recurrence relation:

$$\alpha_k = \frac{N}{k} \sum_{i=1}^{k} (-1)^{i-1} \alpha_{k-i}\, m_i \,, \tag{2.189}$$

or equivalently by the following combinatorial formula:

$$\alpha_k = \sum_{1 j_1 + \cdots + k j_k = k} (-N)^{j_1 + \cdots + j_k} \prod_{i=1}^{k} \frac{{m_k}^{j_i}}{i^{j_i} j_i!} \,. \tag{2.190}$$

Remark *(elementary transformations of roots of monic polynomials:).*

- **Shift**: if one performs a shift of the roots: $\boldsymbol{a} \to \boldsymbol{a} + \delta \mathbf{1} := (a_1 + \delta, \dots, a_N + \delta)$, then the new monic polynomial is given in terms of the old one by:

$$P_{\boldsymbol{a}+\delta\mathbf{1}}(x) = P_{\boldsymbol{a}}(x - \delta) \,, \tag{2.191}$$

  and the new coefficients α_k in the representation of Eq. (2.187) are changed by:

$$\alpha_k(\boldsymbol{a} + \delta \mathbf{1}) = \sum_{i=0}^{k} \frac{(N-k+i)!}{i!\,(N-k)!} \delta^i\, \alpha_{k-i}(\boldsymbol{a}) \,, \tag{2.192}$$

- **Scaling**: if one multiplies all the roots: $\boldsymbol{a} \to \delta \boldsymbol{a} = (\delta\alpha_1, \dots, \delta\alpha_N)$, then the new monic polynomial is given by:

$$P_{\delta \boldsymbol{a}}(x) = \delta^N P_{\boldsymbol{a}}(x/\delta) \,, \tag{2.193}$$

  and the new coefficients α_k in the representation of Eq. (2.187) are changed by:

$$\alpha_k(\delta \boldsymbol{a}) = \delta^k \alpha_k(\boldsymbol{a}) \,. \tag{2.194}$$

- **Inversion**: If all the roots $\boldsymbol{a}$ are non-zero and one takes their inverse $\boldsymbol{a} \to 1/\boldsymbol{a} = (1/a_1, \dots, 1/a_N)$, then the new monic polynomial is given by:

$$P_{\boldsymbol{a}}^{[-1]}(x) := P_{1/\boldsymbol{a}}(x) = \frac{x^N}{\alpha_N} P_{\boldsymbol{a}}(1/x) \,, \tag{2.195}$$

  and the new coefficients α_k in the representation of Eq. (2.187) are changed by:

$$\alpha_k(1/\boldsymbol{a}) = \frac{\alpha_{N-k}(\boldsymbol{a})}{\alpha_N(\boldsymbol{a})} \,. \tag{2.196}$$

⌟

Next, let's introduce objects which will play an important role in the description of the finite free convolutions. Every monic polynomial can be seen as a differential operator acting on the monomial x^N:

$$P_{\boldsymbol{a}}(x) = \widehat{P_{\boldsymbol{a}}}(D_x)\, x^N\,, \tag{2.197}$$

Where $\widehat{P_{\boldsymbol{a}}}$ is a power series in the differential operator D_x and $D_x := \frac{\mathrm{d}}{\mathrm{d}x}$ acts on x^N by $D_x^k x^N = N!/(N-k)!\, x^{N-k}$. Now since differentiating more than N times the monomial x^N always gives zero: $D_x^{N+k} x^N = 0$ for $k \geq 1$, only the first $N+1$ coefficient in the power series of $\widehat{P_{\boldsymbol{a}}}(.)$ matters in Eq. (2.197) and we have:

$$\widehat{P_{\boldsymbol{a}}}(x) := \sum_{k=0}^{N} \frac{(N-k)!}{N!} \alpha_k\, (-1)^k x^k \qquad \mod x^{N+1}\,, \tag{2.198}$$

where α_k is the coefficient of $P_{\boldsymbol{a}}$ in Eq. (2.187) and $\mod x^{N+1}$ indicates that $\widehat{P_{\boldsymbol{a}}}$ is defined "modulo x^{N+1}" meaning that higher order terms can be set to zero, and in this case $\widehat{P_{\boldsymbol{a}}}$ is also a polynomial. The operator $\widehat{P_{\boldsymbol{a}}}$ is the **additive form** of the monic polynomial $\widehat{P_{\boldsymbol{a}}}$ and will play the role of the Fourier transform in the definition of the finite free (additive) convolution.

Similarly, one can define the **multiplicative form** $\widehat{P_{\boldsymbol{a}}^{(\times)}}$ of the polynomial $P_{\boldsymbol{a}}$ with positive roots as the operator acting on $(x-1)^N$:

$$P_{\boldsymbol{a}}(x) = \widehat{P_{\boldsymbol{a}}^{(\times)}}(xD_x)\, (x-1)^N\,, \tag{2.199}$$

which will naturally appear in the description of the finite free multiplicative convolution.

The last ingredient to introduce before jumping to the description of the finite free convolutions is the following observation: every monic polynomial with real roots can be equivalently written as the sum of N shifted monomial $(x - t_i^A)^N/N$ where $\boldsymbol{t}^A = (t_1^A, \ldots, t_N^A) \in \mathbb{C}^N$:

$$\mathbb{E}\left\{(x - t^A)^N\right\} := \frac{1}{N}\sum_{i=1}^{N} (x - t_i^A)^N = P_{\boldsymbol{a}}(x)\,, \tag{2.200}$$

where we have used the convenient notation $\mathbb{E}\{.\}$ to denotes the average over the set $\boldsymbol{t}^A$. The vector $\boldsymbol{t}^A$ is uniquely defined up to a permutation of its entries and is referred to as the **negative Markov-Krein transform**[1] of $\boldsymbol{a}$. Eq. (2.200) can be written in terms of the discrete complex measure $\nu_{\boldsymbol{t}^A}(\lambda) := \sum_{i=1}^{N} \delta(\lambda - t_i^A)/N$ as:

$$\int \frac{\nu_{\boldsymbol{t}^A}(\mathrm{d}x)}{(z-x)^{-N}} = \exp\left[-(-N)\int \mu_{\boldsymbol{a}}(\mathrm{d}\lambda) \log(z-\lambda)\right], \tag{2.201}$$

which will have a clear analog in the description of the high-temperature convolution of Chapter 6. If we apply the binomial theorem to the LHS of Eq. (2.200), one can express the 'moments' of $\nu_{\boldsymbol{t}^A}$ in terms of the coefficient α_k of $P_{\boldsymbol{a}}$ as:

$$\tau_k := \frac{1}{N}\sum_{i=1}^{N} \left(t_i^A\right)^k = \frac{(N-k)!k!}{N!}\alpha_k\,. \tag{2.202}$$

[1] In Ref. [125] this set is named as the U-transform of $\boldsymbol{a}$. Our convention for denoting this object as the negative Markov-Krein transform rather than the U-transform will appear clear in Chapter 6 when we will introduce the (positive) Markov-Krein transform.

If now one injects this relation into the expression (2.198) for the additive form, one can express the latter as a *moment generating function* (evaluated at $-x$) of the negative Markov-Krein transform:

$$\widehat{P_{\boldsymbol{a}}}(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \tau_k \, x^k \qquad \mathrm{mod}\, x^{N+1}\,, \tag{2.203}$$

$$\widehat{P_{\boldsymbol{a}}}(x) = \mathbb{E}\left\{ \mathrm{e}^{-x t_A} \right\} \qquad \mathrm{mod}\, x^{N+1}\,, \tag{2.204}$$

where we recall that $\mathrm{mod}\, x^{N+1}$ indicates equality up to the $N+1$ first terms in the power series.

Similarly, one can show that the multiplicative form given by the relation (2.199) can be set to be the unique monic polynomial of degree N fixed by the conditions:

$$\widehat{P_{\boldsymbol{a}}^{(\times)}}(N-k) = \mathbb{E}\left\{ t_A^k \right\} \qquad \text{for} \quad k \in \{1, \dots, N\}\,, \tag{2.205}$$

where the set t_A is the same as in the additive case. The RHS of Eq. (2.205) is nothing else than the *Mellin transform* of the negative Markov-Krein transform, evaluated at $x = k$.

It turns out to be convenient to introduce a shifted version of the additive and multiplicative forms such that they match the moment generating function/Mellin-transform. That is we define the polynomial $\widehat{Q_{\boldsymbol{a}}}$ as:

$$\widehat{Q_{\boldsymbol{a}}}(x) := \widehat{P_{\boldsymbol{a}}}(-x) = \mathbb{E}\left\{ \mathrm{e}^{x t_A} \right\} \qquad \mathrm{mod}\, x^{N+1}\,, \tag{2.206}$$

and $\widehat{Q_{\boldsymbol{a}}^{(\times)}}$ as the unique polynomial of degree N given by:

$$\widehat{Q_{\boldsymbol{a}}^{(\times)}}(x) = \widehat{P_{\boldsymbol{a}}^{(\times)}}(N-x)\,, \tag{2.207}$$

$$\widehat{Q_{\boldsymbol{a}}^{(\times)}}(x) = \mathbb{E}\left\{ t_A^x \right\} \qquad \text{for} \quad x \in \{1, \dots, N\}\,. \tag{2.208}$$

Examples. I give here some practical examples which will be useful later on.

- If we define the *rank-one polynomial* as the monic polynomial with only one non-zero root equals to γ:

  $$P_\gamma^{(\text{rk-1})}(x) := x^{N-1}(x-\gamma) = x^N - \gamma x^{N-1}\,. \tag{2.209}$$

  then one can immediately deduce its corresponding additive form:

  $$P_\gamma^{(\text{rk-1})}(x) = \left(1 - \frac{\gamma}{N} \mathrm{D}_x\right) x^N\,. \tag{2.210}$$

- For the Hermite polynomial, which we recall is closely related to Gaussian ensemble in the $\beta \to \infty$ see Sec. 1.6, one can check that we have:

  $$C_{H,N}\, \mathsf{He}_N\left(\frac{\sqrt{N}}{\sigma} x\right) = \exp\left[-\frac{\sigma^2}{2N} \mathrm{D_x}^2 \right] x^N\,, \tag{2.211}$$

  where the constant $C_{H,N}$ is such that this polynomial is monic.

- Similarly, for the Laguerre polynomial appearing in the low temperature of the Laguerre ensembles, we have

$$C_{L,N}\,\mathsf{La}_N^{(M-N)}\left(Mx\right) = \left(1-\frac{1}{M}\mathrm{D}_x\right)^M x^N\,, \tag{2.212}$$

where $C_{L,N}$ is the constant that ensures this polynomial is monic. Note that if we look at a slightly modified version where ones multiply the roots by M/N:

$$\mathfrak{L}^{(M)}(x) := \tilde{C}_{L,N}\,\mathsf{La}_N^{(M-N)}\left(Nx\right) = \left(1-\frac{1}{N}\mathrm{D}_x\right)^M x^N\,, \tag{2.213}$$

its additive form is given by the additive form of the rank-one polynomial raised to the power M, see Eq. (2.210).

∎

Finite free additive and multiplicative convolutions -

For $P_{\boldsymbol{a}}, P_{\boldsymbol{b}}$ two N degree monic polynomials with roots $\boldsymbol{a},\boldsymbol{b}\in\mathbb{R}^N$, their **finite free convolution**(FFC) $P_{\boldsymbol{a}}\boxplus_N P_{\boldsymbol{b}}$ is the bi-linear commutative operation whose result is a monic polynomial of degree N **with real roots**, defined equivalently as:

$$\left(P_{\boldsymbol{a}}\boxplus_N P_{\boldsymbol{b}}\right)(x) := \mathbb{E}_{\mathbf{R}}\left[\det\left(x\mathbf{I}-\left(\mathbf{Diag}(\boldsymbol{a})+\mathbf{R}\,\mathbf{Diag}(\boldsymbol{b})\,\mathbf{R}^*\right)\right)\right]\,, \tag{2.214}$$

$$\left(P_{\boldsymbol{a}}\boxplus_N P_{\boldsymbol{b}}\right)(x) = \sum_{k=0}^{N}\left[\sum_{i+j=k}\frac{(N-i)!(N-j)!}{N!(N-k)!}\alpha_i(\boldsymbol{a})\alpha_j(\boldsymbol{b})\right](-1)^k x^{N-k}\,, \tag{2.215}$$

$$\left(P_{\boldsymbol{a}}\boxplus_N P_{\boldsymbol{b}}\right)(x) = \mathbb{E}\left\{\left(x-(t^A+t^B)\right)^N\right\}\,, \tag{2.216}$$

$$\left(P_{\boldsymbol{a}}\boxplus_N P_{\boldsymbol{b}}\right)(x) = \widehat{P_{\boldsymbol{a}}}(D_x)\widehat{P_{\boldsymbol{b}}}(D_x)\,x^N\,. \tag{2.217}$$

where in Eq. (2.214) the average over $\mathbf{R}$ can be either taken over $\mathsf{U}(N)$, $\mathsf{O}(N)$, or the (discrete) group of signed permutation matrices. The notations $\alpha_k(\boldsymbol{a})$, t_A and $\widehat{P_{\boldsymbol{a}}}$ denote respectively the coefficients of $P_{\boldsymbol{a}}$ in the representation of Eq. (2.187), the negative Markov-Krein transform given by Eq. (2.200) and the additive form given by Eq. (2.217).

By abuse of notations, we will also denote by

$$\mu_{\boldsymbol{a}}\boxplus_N\mu_{\boldsymbol{b}}\,, \tag{2.218}$$

the discrete measure of the roots of $\left(P_{\boldsymbol{a}}\boxplus_N P_{\boldsymbol{b}}\right)(x)$.

The proofs of these equivalent statements can be found in [130, 125]. If one looks at Eq. (2.214) without the expectation with $\mathbf{R}\sim\mathrm{Unif}\left[\mathsf{O}_\beta(N)\right]$ then this is nothing else than the characteristic polynomial of the additive random Horn problem of Sec. 2.6.1. Thus, FFC can be seen as the expected/typical value of the characteristic polynomial of this random Horn problem. Informally as $N\to\infty$, one should expect that have a *self-averaging property* and be able to remove the expectation value if $\mathbf{R}\sim\mathrm{Unif}\left[\mathsf{O}_\beta(N)\right]$ and since the LSD of

$\mathbf{Diag}(\boldsymbol{a}) + \mathbf{R}\,\mathbf{Diag}(\boldsymbol{b})\,\mathbf{R}^*$ converges to the free convolution, one can therefore really think of the FFC as a finite counterpart of the free convolution (hence its name) and we will see that this is indeed the case. The fact that one can reduce the average over the compact Haar group $\mathrm{O}_\beta(N)$ to an average over the discrete group of sign permutations is a non-trivial phenomenon known as *quadrature* and can be seen as a consequence of important results in representation theory. The second non-trivial fact is that this operation gives a real-rooted polynomial since the sum of real-rooted polynomials is *a priori* not real-rooted. Eq. (2.214) can be easily obtained from Eq. (2.214) with the average taken over the group of signed permutation and encodes how the coefficients of the FFC depend on the coefficients of the two polynomials $P_{\boldsymbol{a}}$ and $P_{\boldsymbol{b}}$. Eq. (2.216) and Eq. (2.217) provide another natural interpretation of the FFC: the FFC corresponds to adding the negative Markov-Krein transform of each polynomial and *in spirit* this corresponds to do the (classical) convolution of their discrete measures and hence multiply their moment generating functions.

The multiplicative counterpart of this FFC is defined similarly:

For $P_{\boldsymbol{a}}, P_{\boldsymbol{b}}$ two N degree monic polynomials with roots $\boldsymbol{a}, \boldsymbol{b} \in {\mathbb{R}_+^*}^N$, their **finite free multiplicative convolution** (FFMC) is defined equivalently as:

$$(P_{\boldsymbol{a}} \boxtimes_N P_{\boldsymbol{b}})(x) := \mathbb{E}_{\mathbf{R}}\left[\det\left(x\mathbf{I} - \mathbf{Diag}(\boldsymbol{a})\mathbf{R}\,\mathbf{Diag}(\boldsymbol{b})\,\mathbf{R}^*\right)\right], \tag{2.219}$$

$$(P_{\boldsymbol{a}} \boxtimes_N P_{\boldsymbol{b}})(x) = \sum_{k=0}^{N}\left[\frac{\alpha_k(\boldsymbol{a})\alpha_k(\boldsymbol{b})}{\binom{N}{k}}\right](-1)^k x^{N-k}, \tag{2.220}$$

$$(P_{\boldsymbol{a}} \boxtimes_N P_{\boldsymbol{b}})(x) = \mathbb{E}\left\{\left(z - (t^A\, t^B)\right)^N\right\}, \tag{2.221}$$

$$(P_{\boldsymbol{a}} \boxtimes_N P_{\boldsymbol{b}})(x) = \widehat{P_{\boldsymbol{a}}^{(\times)}}(xD_x)\,\widehat{P_{\boldsymbol{b}}^{(\times)}}(xD_x)\,(x-1)^N. \tag{2.222}$$

furthermore this operation preserves **real-rootness** and **positivity**.

Note that since $\sqrt{\mathbf{A}}\mathbf{B}\sqrt{\mathbf{A}}$ has the same eigenvalues as $\mathbf{AB}$, one can also interpret this FFMC as the expected value of the characteristic polynomial of the multiplicative random Horn problem of Eq. (2.122). Interestingly, the FFMC is now the product of the *same* negative Markov-Krein transforms entering the definition of the FFC and thus this corresponds to multiplying their Mellin transform.

Remark *(finite free rectangular convolution).* One can also define the rectangular counterpart of this finite free convolution, see for example Ref. [78, 77]. However, this operation will not be discussed in the rest of this thesis. ⌟

We conclude this section with a simple example to illustrate the FFC.

Example *(FFC of rank-one polynomials).* Let's consider two rank-one polynomials as defined by Eq. (2.209) with non-zero roots given by γ_1 and γ_2, then using the definition (2.217) and Eq. (2.210) we have for their FFC:

$$P_{\gamma_1}^{(\text{rk-1})}(x) \boxplus_N P_{\gamma_2}^{(\text{rk-1})}(x) = \left(1 - \frac{\gamma_1}{N}\mathrm{D}_x\right)\left(1 - \frac{\gamma_2}{N}\mathrm{D}_x\right)x^N, \tag{2.223}$$

$$P_{\gamma_1}^{(\text{rk-1})}(x) \boxplus_N P_{\gamma_2}^{(\text{rk-1})}(x) = x^N - (\gamma_1+\gamma_2)x^{N-1} + \frac{N-1}{N}\gamma_1\gamma_2 x^{N-2}, \tag{2.224}$$

$$P_{\gamma_1}^{(\text{rk-1})}(x) \boxplus_N P_{\gamma_2}^{(\text{rk-1})}(x) = x^{N-2}\,(x-\gamma_-)\,(x-\gamma_+), \tag{2.225}$$

with

$$\gamma_\pm = \frac{\gamma_1+\gamma_2}{2} \pm \frac{\sqrt{4\gamma_1\gamma_2+(\gamma_1-\gamma_2)^2 N}}{2\sqrt{N}}\,. \tag{2.226}$$

Thus, their FFC is a 'rank-two' polynomial with non-zero roots given asymptotically (in N) as the non-zero roots γ_1 and γ_2. Note that this asymptotic behavior is what we would obtain if we would sum two rotationally invariant rank-one matrices with non-zero eigenvalues given respectively by γ_1 and γ_2. ∎

2.7.2 Finite free transforms and associated finite free cumulants

Linearizing transforms for the finite free convolutions -

We introduce the finite counterpart of the R-transform and (log)-S-transform of the previous section.

The **finite R-transform** $\mathcal{R}_{\boldsymbol{a}}^{(N)}$ of a polynomial $P_{\boldsymbol{a}}$ is defined as:

$$\mathcal{R}_{\boldsymbol{a}}^{(N)}(y) := -\frac{1}{N} D_y \log \widehat{Q_{\boldsymbol{a}}}(-Ny) \qquad \bmod y^N\,, \tag{2.227}$$

with $\widehat{Q_{\boldsymbol{a}}}$ defined by Eq. (2.206).

The (logarithm of the) **finite S-transform** $\tilde{\mathcal{S}}_{\boldsymbol{a}}^{(N)}$ of a polynomial $P_{\boldsymbol{a}}$ with positive roots is defined as:

$$\log \tilde{\mathcal{S}}_{\boldsymbol{a}}^{(N)}(y) := -\frac{1}{N} D_y \log \widehat{Q_{\boldsymbol{a}}^{(\times)}}(-Ny)\,, \tag{2.228}$$

with $\widehat{Q_{\boldsymbol{a}}^{(\times)}}$ defined by Eq. (2.207).

Note that $\widehat{Q_{\boldsymbol{a}}}$ and $\widehat{Q_{\boldsymbol{a}}^{(\times)}}$ corresponds respectively to the moment generating function and Mellin transform of the negative Markov-Krein transform. Since by Eq. (2.216) and (2.221), one knows that the FFC and FFMC correspond respectively to the sum/product of the negative Markov-Krein transform, it should come as no surprise that the linearizing transforms of the FFC and FFMC are given as the logarithm of $\widehat{Q_{\boldsymbol{a}}}$ and $\widehat{Q_{\boldsymbol{a}}^{(\times)}}$. The following results given in Ref. [125] goes into this direction:

Additivity of the finite R-transform and the finite log-S-transform: *If we denote by $\boldsymbol{c}$ the roots of $P_{\boldsymbol{a}} \boxplus_N P_{\boldsymbol{b}}$, then we have*

$$\mathcal{R}_{\boldsymbol{c}}^{(N)}(y) = \mathcal{R}_{\boldsymbol{a}}^{(N)}(y) + \mathcal{R}_{\boldsymbol{a}}^{(N)}(y) \qquad \bmod y^N\,. \tag{2.229}$$

Similarly if we denote by $\boldsymbol{c}$ the roots of $P_{\boldsymbol{a}} \boxtimes_N P_{\boldsymbol{b}}$, we have:

$$\log \tilde{\mathcal{S}}_{\boldsymbol{c}}^{(N)}(y) = \log \tilde{\mathcal{S}}_{\boldsymbol{a}}^{(N)}(y) + \log \tilde{\mathcal{S}}_{\boldsymbol{b}}^{(N)}(y)\,, \tag{2.230}$$

for $y = -k/N$ with $k \in \{1,\dots,N\}$.

Finite free cumulants -

The measure $\mu_{\boldsymbol{a}}$ is discrete on (at most) N atoms, hence it is completely determined by its first N moments. This means that one only needs the knowledge of the first N **finite free cumulants** to compute the finite free sum. One can show (see Ref. [7]) that up to a normalization those cumulants are given as the first N coefficients in the power series of the finite R-transform:

$$\mathcal{R}_{\boldsymbol{a}}^{(N)}(y) := \sum_{k=0}^{N} \left(\frac{(N-k)!}{N!} N^k \right) \kappa_k^{(N)} y^{k-1} , \tag{2.231}$$

where the coefficients $N!/(N-k)!N^{-k}$ ensure the leading term of $\kappa_k^{(N)}$ to be $m_k := \sum_{i=1}^{N} a_i^k/N$, the k^{th} moment of the discrete distribution. By linearity of the finite R-transform we have for $\boldsymbol{c}$ the roots of $P_{\boldsymbol{a}} \boxplus_N P_{\boldsymbol{b}}$:

$$\kappa_k^{(N)}(\boldsymbol{c}) = \kappa_k^{(N)}(\boldsymbol{a}) + \kappa_k^{(N)}(\boldsymbol{b}) \qquad \text{for } k \in \{1, \dots, N\} . \tag{2.232}$$

As in the free case, one would like to relate these finite free cumulants to the moments m_k of the discrete measure $\mu_{\boldsymbol{a}}$ associated with the roots. Using Eq. (2.227) and Eq. (2.203), one can see that the two quantities are related by the two equations:

$$\begin{cases} \tau_k = \left(\frac{(N-k)!}{N!} k! \right) \sum_{1 j_1 + \cdots + k j_k = k} (-N)^{j_1 + \cdots + j_k} \prod_{i=1}^{k} \frac{{m_k}^{j_i}}{i^{j_i} j_i!} & \text{for } k \leq N , \\ \log\left(1 + \sum_{k=0}^{N} \frac{\tau_k}{k!} y^k \right) = \sum_{k=1}^{N} \left(\frac{(N-k)!}{N!} \frac{(-N)}{k} \right) \kappa_k^{(N)} y^k & \mod y^N . \end{cases} \tag{2.233}$$

where the $\tau_k := \sum_{i=1}^{N} (t_i^A)^k/N$ are the moments of the negative Markov-Krein transform. Thus, in practice, one can first solve for the τ_k using either the top equation of Eq. (2.233) or the recursive relation Eq. (2.189) with Eq. (2.202) and then expand the Taylor series of logarithm in the bottom equation of Eq. (2.233) to express the cumulants in terms of the moments. Doing so, one obtains that the first four finite cumulants are given in terms of the moments of $\mu_{\boldsymbol{a}}$ by:

- $\kappa_1^{(N)} = m_1$,
- $\kappa_2^{(N)} = m_2 - m_1^2 \qquad (\text{if } N \geq 2)$,
- $\kappa_3^{(N)} = m_3 - 3 m_1 m_2 + 2 m_1^3 \qquad (\text{if } N \geq 3)$,
- $\kappa_4^{(N)} = m_4 - \frac{2N-3}{N-1} m_2^2 + \frac{10N-12}{N-1} m_2 m_1^2 - \frac{5N-6}{N-1} m_1^4 \qquad (\text{if } N \geq 4)$.

After consequent work (see Ref. [7]) one can eliminate the dependency in τ_k to obtain a sophisticated combinatorial formula just between the finite free cumulants and the moments. This relation can also be interpreted as a topological expansion (see Ref. [6]) and put under the general form:

$$m_k = \sum_{\boldsymbol{\pi} \in \mathcal{P}[k]} W_{\mathrm{FFC}}^{(N)}(\boldsymbol{\pi})\, \kappa_{\boldsymbol{\pi}}^{(N)} \qquad \text{for } k \in \{1, \dots, N\} , \tag{2.234}$$

where $W_{\mathrm{FFC}}^{(N)}(\cdot)$ is a weight factor that can be shown to penalize crossing partitions in a certain way. Asymptotically in N, this weight factor is given by:

$$W_{\mathrm{FFC}}^{(N)}(\boldsymbol{\pi}) \xrightarrow[N\to\infty]{} \begin{cases} 1 & \text{if } \boldsymbol{\pi} \text{ is non-crossing}\,, \\ \\ 0 & \text{otherwise}\ , \end{cases} \tag{2.235}$$

such that one retrieves the moment-cumulant formula of free convolution, see Eq. (2.151). In other words, the finite R-transform converges to the R-transform and one can prove a similar result for the finite S-transform:

Asymptotic property of finite transforms: *If the discrete $\mu_{\boldsymbol{a}}$ associated to the roots of a polynomial $P_{\boldsymbol{a}}$ converges for large N to a smooth distribution μ_A then we have*

$$\mathcal{R}_{\boldsymbol{a}}^{(N)}(y) \underset{N\to\infty}{\to} \mathcal{R}_A(y) \qquad \textit{and} \qquad \tilde{\mathcal{S}}_{\boldsymbol{a}}^{(N)}(y) \underset{N\to\infty}{\to} \tilde{\mathcal{S}}_A(y)\,, \tag{2.236}$$

if the roots are positive and where $\mathcal{R}_A$ and $\tilde{\mathcal{S}}_A$ are respectively the R-transform of μ_A defined by Eq. (2.139) and the S-transform of μ_A defined by Eq. (2.152).

This leads immediately to the following result:

Result 2.20 (*Finite free convolutions tend to free convolutions*)

For two monic polynomials $P_{\boldsymbol{a}}, P_{\boldsymbol{b}}$ of degree N with roots $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^N$, if we denote by $\mu_{\boldsymbol{a}} \boxplus_N \mu_{\boldsymbol{b}}$ the discrete measure with atoms given by the roots of $P_{\boldsymbol{a}} \boxplus_N P_{\boldsymbol{b}}$, then in the large N limit where $\mu_{\boldsymbol{a}} = 1/N \sum_i \delta(x - a_i) \to \mu_A$ and $\mu_{\boldsymbol{b}} = 1/N \sum_i \delta(x - b_i) \to \mu_B$, we have:

$$\mu_{\boldsymbol{a}} \boxplus_N \mu_{\boldsymbol{b}} \xrightarrow[N\to\infty]{} \mu_A \boxplus \mu_B\,. \tag{2.237}$$

Similarly, for the multiplicative case we have:

$$\mu_{\boldsymbol{a}} \boxtimes_N \mu_{\boldsymbol{b}} \xrightarrow[N\to\infty]{} \mu_A \boxtimes \mu_B\,, \tag{2.238}$$

where $\mu_{\boldsymbol{a}} \boxtimes_N \mu_{\boldsymbol{b}}$ is the discrete measure associated to the positive roots of $P_{\boldsymbol{a}} \boxtimes_N P_{\boldsymbol{b}}$.

2.7.3 Finite free central limit theorems

In order to construct an equivalent of the central limit theorem for the finite free convolution, one needs to determine which polynomial has only the second finite free cumulant $\kappa_2^{(N)}$ as being a non-zero cumulant. From the expression (2.211) of the additive form of the Hermite polynomials together with the expression of the finite R-transform of Eq. (2.227), one can show that Hermite polynomials given by Eq. (2.211) as a second cumulant given by $(1-1/N)\sigma^*$, thus in order to have a second cumulant independent of N, one needs to re-scale this polynomial and this gives the following result:

Result 2.21 (*Finite free central limit theorem*)

If $\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(n)}$ are n vectors of size N, such that their associated measure $\mu_{\boldsymbol{a}^{(i)}}$ all have mean zero: $m_1^{(i)} = \sum_{j=1}^N a_j^{(i)}/N = 0$, and second moment $m_2^{(i)} = \sum_{j=1}^N (a_j^{(i)})^2/N = \sigma^2$,

then in the large n limit we have:

$$P_{\boldsymbol{a}^{(1)}/\sqrt{n}}(x) \boxplus_N \cdots \boxplus_N P_{\boldsymbol{a}^{(n)}/\sqrt{n}}(x) \xrightarrow[n\to\infty]{} \mathfrak{H}^{(\sigma^2)}(x) := \tilde{C}_{H,N}\, \mathsf{He}_N\left(\sqrt{\frac{N-1}{\sigma^2}}x\right) \tag{2.239}$$

where $\tilde{C}_{H,N}$ is the constant which makes this polynomial monic.

The Poisson limit theorem for the finite free convolution can be easily deduced from the example of Eq. (2.213) where it was noticed that its additive form is given as M times the one of a rank-one polynomial with non-zero root given by one, thus we have:

Result 2.22 (*Finite free Poisson theorem*)

Let $P_1^{(\text{rk-1})}(x) := x^{N-1}(x-1)$, then we have:

$$\underbrace{P_1^{(\text{rk-1})}(x) \boxplus_N \cdots \boxplus_N P_1^{(\text{rk-1})}(x)}_{M \text{ times}} = \mathfrak{L}^{(M)}(x) := C_{L,N}\, \mathsf{La}_N^{(M-N)}(Nx) \tag{2.240}$$

where $C_{L,N}$ is the constant which makes this polynomial monic.

2.7.4 Finite free convolution as the sum of low-temperature ensembles

For $\beta > 0$, even though one does not have a matrix model, one can *extrapolate* what is the corresponding 'β-sum' of two sets of 'eigenvalues' $\boldsymbol{a}$ and $\boldsymbol{b}$ and this will be shortly described in the next chapter and is based on the theory of special symmetric polynomials known as *Jack polynomials* which depend explicitly on the parameter β. As $\beta \to \infty$, one can show that these Jack polynomials degenerate in a certain way into elementary symmetric polynomials, which as we have seen are closely related to the finite free convolution. From there, one can interpret (also in a certain way) the sum/product of β-ensembles in the low-temperature limit $\beta \to \infty$ with N fixed as being given by the FFC/FFMC. We refer the reader to Ref. [75] for a precise statement of this result and we give in this section three simple examples illustrating this phenomenon.

Rank-one perturbation of polynomial and secular equation -

Let's first consider the *rank-one perturbation* of a polynomial:

$$P_{\boldsymbol{c}}(x) = \left(P_{\boldsymbol{b}} \boxplus_N P_{\gamma}^{(\text{rk-1})}\right)(x)\,, \tag{2.241}$$

where $P_{\gamma}^{(\text{rk-1})}$ is given by Eq. (2.209) with $\gamma > 0$ and we assume the roots of $P_{\boldsymbol{b}}$ to be simple. Now since the additive form of the rank-one polynomial is given by Eq. (2.210), we have:

$$P_{\boldsymbol{c}}(x) = P_{\boldsymbol{b}}(x) - \frac{\gamma}{N}(P_{\boldsymbol{b}})'(x)\,, \tag{2.242}$$

$$P_{\boldsymbol{c}}(x) = \prod_{i=1}^{N}(x-b_i) - \frac{\gamma}{N}\sum_{j=1}^{N}\prod_{i|i\neq j}(x-b_i)\,. \tag{2.243}$$

Hence if we factorize by $\prod_{i=1}^{N}(x-b_i)$ we have that the roots c_i of $P_{\boldsymbol{c}}(x)$ which are different from b_i are given as the solution of the *deterministic* secular equation:

$$\boxed{1-\frac{\gamma}{N}\sum_{j=1}^{N}\frac{1}{c_i-b_j}=0}\,, \tag{2.244}$$

which is nothing else than the $\beta\to\infty$ version of the additive rank-one deformation of Eq. (2.37) since the mean of the d_i is given by $1/N$ and their variance goes to zero as $\beta\to\infty$, see Eq. (2.44).

Infinitesimal perturbation and the heat flow equation for zeroes of an orthogonal polynomial -

We now turn to the 'DBM' counterpart for the finite free convolution. In the finite free world, Gaussian ensembles are replaced by Hermite polynomials and one has two natural choices for their normalization: either choose the convention defined by Eq. (2.211) or the convention from the finite free CLT of Eq. (2.239). We start with the first convention and then briefly describe what changes for the other one. We consider the following deterministic process:

$$P_{\boldsymbol{c}}(t,x)=\left(P_{\boldsymbol{b}}(x)\boxplus_N C_{H,N}\,\mathsf{He}_N\left(\frac{\sqrt{N}}{t}x\right)\right)\,, \tag{2.245}$$

which can be expressed thanks to Eq. (2.211) as:

$$P_{\boldsymbol{c}}(t,x)=\exp\left(-\frac{t}{2N}{\mathrm{D_x}}^2\right)P_{\boldsymbol{b}}(x)\,. \tag{2.246}$$

If we differentiate this equation since we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\exp\left(-\frac{t}{2N}{\mathrm{D_x}}^2\right)=-\frac{1}{2N}{\mathrm{D_x}}^2\exp\left(-\frac{t}{2N}{\mathrm{D_x}}^2\right)\,, \tag{2.247}$$

we get the following **backward heat equation** (BHE) for $P_{\boldsymbol{c}}(t,x)$:

$$\partial_t P_{\boldsymbol{c}}(t,x)=-\frac{1}{2N}\partial_{xx}P_{\boldsymbol{c}}(t,x)\,. \tag{2.248}$$

Note that since FFC preserves real-rootedness, so does the BHE. The roots $c_i(t)$ of $P_{\boldsymbol{c}}(t,x)$ satisfy by definition:

$$P_{\boldsymbol{c}}(t,c_i(t))=0\,, \tag{2.249}$$

and by differentiating this equation with respect to the variable t, we get after a few simplifications, the following equation for the roots:

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}t}c_i(t)=\frac{1}{N}\sum_{j|j\neq i}\frac{1}{c_i(t)-c_j(t)}}\,, \tag{2.250}$$

which is nothing else than the usual DBM without the noise term, that is the $\beta\to\infty$ limit of the DBM. Correspondingly, if we define the Stieltjes transform:

$$g_{\boldsymbol{c}}(t,z)=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{z-c_i(t)}\,, \tag{2.251}$$

it is the solution of the following complex Burgers equation with *negative* coefficient of viscosity:

$$\partial_t g_{\boldsymbol{c}} + g_{\boldsymbol{c}} \partial_z g_{\boldsymbol{c}} = -\frac{1}{2N} \partial_{zz} g_{\boldsymbol{c}} \,, \tag{2.252}$$

which is again the $\beta \to \infty$ of the stochastic partial differential equation Eq. (2.85).

Now if we use instead the Hermite polynomial $\mathfrak{H}^{(t)}$ of the finite free central limit theorem of Eq. (2.239) and define $P_{\tilde{\boldsymbol{c}}}(t,x) = (P_{\boldsymbol{b}} \boxplus_N \mathfrak{H}^{(t)})(x)$, then one can show that the associated Stieltjes transform is the solution of:

$$\partial_t g_{\tilde{\boldsymbol{c}}} + \frac{N}{N-1} g_{\tilde{\boldsymbol{c}}} \partial_z g_{\tilde{\boldsymbol{c}}} = -\frac{1}{2(N-1)} \partial_{zz} g_{\tilde{\boldsymbol{c}}} \,. \tag{2.253}$$

Relation between Jacobi and Laguerre Polynomials -

We conclude this section on the finite free convolution with the finite counterpart of the free relation between the Watcher distribution and the Marčenko-Pastur distribution of Eq. (1.43). To ease notation, let's denote by

$$\mathfrak{Ja}^{(M_1,M_2)}(x) := C_{J,N}\, \mathrm{Ja}_N^{(M_1-N,M_2-N)}(x) = C_{J,N}\, \mathsf{P}_N^{(M_2-N,M_1-N)}(2x-1) \,, \tag{2.254}$$

the monic Jacobi polynomial on the unit interval, then we have

$$\boxed{\mathfrak{Ja}^{(M_1,M_2)}(x) = \left(\left(\mathfrak{L}^{(M_1)} \boxtimes_N \mathfrak{L}^{(M_2)^{[-1]}} \right) \boxplus_N P_{\mathbf{1}} \right)^{[-1]}(x)} \,, \tag{2.255}$$

where $.^{[-1]}$ indicates the operation of taking the inverse of the roots of a polynomial, $P_{\mathbf{1}}(x) = (x-1)^N$ and $\mathfrak{L}$ is the Laguerre polynomial of Eq. (2.240). This result is part of an ongoing work of the author and can be obtained by a 'brute-force' computation thanks to the properties of the FFC.

2.8 Summary and Conclusion of Chapter 2

In this chapter, we have reviewed results concerning the sum and the product of random matrices for $\beta = 1, 2, 4$. The description of the joint density of eigenvalues/singular values is expressed in terms of spherical integrals which play the role of the Fourier transform and at large N, the limiting spectral distribution is given by the theory of free probability. For rank-one and infinitesimal perturbations, we have seen that one can naturally extend the operation to any $\beta > 0$ and as we will see in the following chapter, one can in fact *extrapolate* the sum/product for any $\beta > 0$. The limit $\beta \to \infty$ (with N fixed) of this extrapolation is the finite free convolution of Sec. 2.7, an operation acting on monic polynomials. A natural question is to extend the large deviation principle of the top eigenvalue of Sec. 1.5 (obtained for individual β-ensembles) to the sum/product of random matrices, and this is done in Chapter 5. Another natural question is to give a meaning for the sum of high-temperature ensembles ($N\beta/2 \to c$) and this operation, the high-temperature convolution, is described in Chapter 6. One open question is to know if the support of the finite free convolution is always included in Horn's convex hull. Surely this is what one should expect since the FFC corresponds in a sense to the typical value of Horn's problem, but it is unclear how one proves this statement based on the definitions of the FFC.

Chapter 3

Spherical integrals and their large N limits

Part of the materials presented in this chapter are based on the preprint [142].

3.1 Introduction

In this chapter, we study in detail the spherical integrals appearing in the different flavors of Horn's problem of the previous chapter, and in particular their large N limits. In Sec. 3.2 we recall the main definitions and properties of the spherical integrals obtained for $\beta \in \{1, 2, 4\}$, give the determinantal formulae in the special case of $\beta = 2$ and then explain how one can naturally extend the definitions of these spherical integrals to $\beta > 0$. This extension allows one to extrapolate the sum of β-ensemble to any $\beta > 0$. In the two following sections, we look at two different limits of these spherical integrals. The first one, developed in Sec. 3.3 concerns the case where the two arguments of the spherical integrals are 'full-rank', that is, *most* of their entries are non-zero) while in Sec. 3.4, we look at the exact opposite case, where the conjugate variable $\boldsymbol{t}$ has all its entries zero except for one. This limit is closely related to the free probability transforms described in the previous chapter. Eventually in Sec. 3.4.7, we consider the *annealed limit* of this 'rank-one' integrals.

3.2 More on spherical integrals

Reminder on the definitions of the spherical integrals -

For $\beta = 1, 2, 4$, if we denote as per usual $\mathsf{O}_\beta(N) = \mathsf{O}(N), \mathsf{U}(\mathsf{N}), \mathsf{Sp}(N)$, the corresponding compact group, then for ease of reading, we recall that for each β, the three spherical integrals introduced in the previous chapter are given by the following definitions.

- For $\boldsymbol{a} \in \mathbb{R}^N$ the additive spherical integral (or HCIZ integral) is defined by

$$\mathcal{I}^{(\beta)}(\boldsymbol{a}, \boldsymbol{t}) := \int_{\mathsf{O}_\beta(N)} \mathrm{e}^{\mathrm{Tr}\, \mathbf{Diag}(\boldsymbol{a}) \mathbf{V} \mathbf{Diag}(\boldsymbol{t}) \mathbf{V}^*} \mu_{\mathrm{Haar}}(\mathrm{d}\mathbf{V}) \,, \tag{3.1}$$

- For $\boldsymbol{a} \in \mathbb{R}_+^N$ the multiplicative spherical integral is defined by

$$\mathcal{I}_\times^{(\beta)}(\boldsymbol{a}, \boldsymbol{t}) := \int_{\mathsf{O}_\beta(N)} \Delta_{\boldsymbol{t}}(\mathbf{V}\mathbf{Diag}(\boldsymbol{a})\mathbf{V}^*)\, \mu_{\mathrm{Haar}}(\mathrm{d}\mathbf{V})\,, \tag{3.2}$$

  where we recall that for $\mathbf{A}$ a self-adjoint matrix, $\Delta_{\boldsymbol{t}}(\mathbf{A}) := \left(\det\left[A_{(1)}\right]\right)^{t_1-t_2} \ldots \left(\det\left[A_{(N-1)}\right]\right)^{t_{N-1}-t_N} (\det\mathbf{A})^{t_N}$ is the generalized power function and $\mathbf{A}_{(i)}$ denotes the top left $(i \times i)$ corner of the matrix $\mathbf{A}$.

- For $\boldsymbol{a} \in \mathbb{R}_+^N$, the rectangular spherical integral is defined by:

$$\mathcal{I}_q^{(\beta)}(\boldsymbol{a}, \boldsymbol{t}) := \int_{\mathsf{O}_\beta(N)} \int_{\mathsf{O}_\beta(M)} \mathrm{e}^{\mathfrak{Re}\,\mathrm{Tr}\left(\mathbf{Diag}_q(\boldsymbol{a})\mathbf{V}_1\mathbf{Diag}_q(\boldsymbol{t})^\mathsf{T}\mathbf{V}_2\right)} \mu_{\mathrm{Haar}}(\mathrm{d}\mathbf{V}_1)\mu_{\mathrm{Haar}}(\mathrm{d}\mathbf{V}_2)\,, \tag{3.3}$$

  where $\mathbf{Diag}_q(\boldsymbol{a})$ denotes the $(N \times M)$ matrix with N (a priori) non-zero entries over the diagonal given by the components of $\boldsymbol{a}$.

In each case, we can recall that one can replace $\mathbf{Diag}(\boldsymbol{a})$ (resp. $\mathbf{Diag}_q(\boldsymbol{a})$) by a self-adjoint (resp. rectangular) matrix $\mathbf{A}$ with eigenvalues (resp. singular values) $\boldsymbol{a}$ without changing the result since one can always absorb the matrix of eigenvectors in the eigenvalue (singular value) decomposition of $\mathbf{A}$ in the Haar measure.

In the previous chapter, we have seen that these spherical integrals play the role of the Fourier transform for their corresponding Horn problem since we have the following results. For the additive case, we have:

Result 3.1 (*Averaging property for the additive spherical integral*)

$$\mathbb{E}_{\mathbf{V}\sim\mathrm{Unif}\left[\mathsf{O}_\beta(N)\right]}\left[I^{(\beta)}(\mathbf{A} + \mathbf{V}\mathbf{B}\mathbf{V}^*, \cdot)\right] = I^{(\beta)}(\mathbf{A}, \cdot)\, I^{(\beta)}(\mathbf{B}, \cdot)\,, \tag{3.4}$$

or equivalently in terms of the joint law $\mathcal{P}^{(\beta)}(\boldsymbol{c}|\boldsymbol{a}, \boldsymbol{b})$:

$$\int \mathcal{I}^{(\beta)}(\boldsymbol{c}, \cdot)\, \mathcal{P}^{(\beta)}(\boldsymbol{c}|\boldsymbol{a}, \boldsymbol{b})\mathrm{d}\boldsymbol{c} = \mathcal{I}^{(\beta)}(\boldsymbol{a}, \cdot)\, \mathcal{I}^{(\beta)}(\boldsymbol{b}, \cdot)\,. \tag{3.5}$$

Similarly for the multiplicative case, we have:

Result 3.2 (*Averaging property for the multiplicative spherical integral*)

$$\mathbb{E}_{\mathbf{V}\sim\mathrm{Unif}\left[\mathsf{O}_\beta(N)\right]}\left[I_\times^{(\beta)}\left(\mathbf{B}^{1/2}\mathbf{V}\mathbf{A}\mathbf{V}^*\mathbf{B}^{1/2}, \cdot\right)\right] = I_\times^{(\beta)}(\mathbf{A}, \cdot)\; I_\times^{(\beta)}(\mathbf{B}, \cdot)\,, \tag{3.6}$$

or equivalently,

$$\int \mathcal{I}_\times^{(\beta)}(\boldsymbol{c}, \cdot)\, \mathcal{P}_\times^{(\beta)}(\boldsymbol{c}|\boldsymbol{a}, \boldsymbol{b})\mathrm{d}\boldsymbol{c} = \mathcal{I}_\times^{(\beta)}(\boldsymbol{a}, \cdot)\, \mathcal{I}_\times^{(\beta)}(\boldsymbol{b}, \cdot)\,. \tag{3.7}$$

Eventually, for the rectangular case, we have:

Result 3.3 (*Averaging property for the rectangular spherical integral*)

$$\mathbb{E}_{\mathbf{V},\mathbf{V}'}\left[I_q^{(\beta)}(\mathbf{A}+\mathbf{V}\mathbf{B}\mathbf{V}',\cdot)\right] = I_q^{(\beta)}(\mathbf{A},\cdot)\, I_q^{(\beta)}(\mathbf{B},\cdot)\,, \tag{3.8}$$

or equivalently:

$$\int \mathcal{I}_q^{(\beta)}(\boldsymbol{c},\cdot)\, \mathcal{P}_q^{(\beta)}(\boldsymbol{c}|\boldsymbol{a},\boldsymbol{b})\mathrm{d}\boldsymbol{c} = \mathcal{I}_q^{(\beta)}(\boldsymbol{a},\cdot)\, \mathcal{I}_q^{(\beta)}(\boldsymbol{b},\cdot)\,. \tag{3.9}$$

We now give short proofs of these properties. For the additive and rectangular cases, this is a direct consequence of the Haar property.

Proof for the additive and rectangular cases (Res. 3.1 and 3.3): by definition of the spherical integral, one has:

$$\mathbb{E}_{\mathbf{V}}\left[I^{(\beta)}(\mathbf{A}+\mathbf{V}\mathbf{B}\mathbf{V}^*,\mathbf{T})\right] = \mathbb{E}_{\mathbf{V}'}\left[\mathrm{e}^{\mathrm{Tr}\,\mathbf{A}\mathbf{V}'\mathbf{T}\mathbf{V}'^*}\mathbb{E}_{\mathbf{V}}\left[\mathrm{e}^{\mathrm{Tr}\,\mathbf{B}(\mathbf{V}^*\mathbf{V}')\mathbf{T}(\mathbf{V}^*\mathbf{V}')^*}\right]\right]\,, \tag{3.10}$$

where to lighten notations we have set $\mathbb{E}_{\mathbf{V}} \equiv \mathbb{E}_{\mathbf{V}\sim\mathrm{Unif}[\mathsf{O}_\beta(N)]}$. If we dot the change of variable $\mathbf{U} = (\mathbf{V}^*\mathbf{V}')$, in the expectation over $\mathbf{V}$, we have $\mathbf{U} \sim \mathrm{Unif}\left[\mathsf{O}_\beta(N)\right]$ and this gives:

$$\mathbb{E}_{\mathbf{V}}\left[I^{(\beta)}(\mathbf{A}+\mathbf{V}\mathbf{B}\mathbf{V}^*,\mathbf{T})\right] = \mathbb{E}_{\mathbf{V}'}\left[\mathrm{e}^{\mathrm{Tr}\,\mathbf{A}\mathbf{V}'\mathbf{T}\mathbf{V}'^*}\right]\cdot\mathbb{E}_{\mathbf{U}}\left[\mathrm{e}^{\mathrm{Tr}\,\mathbf{B}\mathbf{U}\mathbf{T}\mathbf{U}^*}\right]\,, \tag{3.11}$$

which gives the desired property by definition of the spherical integral. The proof for the rectangular case is identical. □

The case of the multiplicative spherical integral is not as straightforward as in the additive/rectangular case and a *readable* proof is hard to find in the literature, so I give here a short proof.

Proof for the multiplicative case (Res. 3.2) : We decompose the proof into two steps.

- *First*, every positive self-adjoint matrix $\tilde{\mathbf{A}} \in \mathrm{Herm}_\beta^{++}(N)$ admits a unique *Cholesky decomposition*:

$$\tilde{\mathbf{A}} = \mathbf{R}\mathbf{R}^*\,, \tag{3.12}$$

where $\mathbf{R}$ is an upper triangular matrix. If we apply this decomposition to $\mathbf{V}\mathbf{A}\mathbf{V}^* = \mathbf{R}_A\mathbf{R}_A^*$ in the definition of the spherical integral, we can replace the average over the compact group by an average over the triangular matrix $\mathbf{R}_A$:

$$I_\times^{(\beta)}\left(\mathbf{A},\boldsymbol{t}\right) = \mathbb{E}_{\mathbf{R}_A}\left[\Delta_{\boldsymbol{t}}\left(\mathbf{R}_A\mathbf{R}_A^*\right)\right]\,. \tag{3.13}$$

Next, since the determinant of a product of matrices is the product of the determinants and since the determinant of a triangular matrix is the product of its diagonal components, we have:

$$\det\left[\left(\mathbf{R}_A\mathbf{R}_A^*\right)_{(i)}\right] = \det\left[\mathbf{R}_{A(i)}\right]^2 = \prod_{k=1}^{i}\left((R_A)_{kk}\right)^2\,, \tag{3.14}$$

such that the multiplicative spherical writes:

$$I_{\times}^{(\beta)}\left(\mathbf{A},\boldsymbol{t}\right)=\mathbb{E}_{\mathbf{R}_A}\left[\prod_{j=1}^{N}\left(\left(R_A\right)_{jj}\right)^{2\,t_N}\cdot\prod_{i=1}^{N-1}\prod_{k=1}^{i}\left(\left(R_A\right)_{kk}\right)^{2(t_i-t_{i+1})}\right]\,, \tag{3.15}$$

which simplifies into:

$$I_{\times}^{(\beta)}\left(\mathbf{A},\boldsymbol{t}\right)=\mathbb{E}_{\mathbf{R}_A}\left[\prod_{k=1}^{N}\left(\left(R_A\right)_{kk}\right)^{2\,t_k}\right]\,. \tag{3.16}$$

- *Second*, if we now look at the average multiplicative spherical integral of the corresponding Horn problem, since $\mathbf{B}^{1/2}=\mathbf{B}^{1/2^*}$, we have:

$$\mathbb{E}_{\mathbf{V}}\left[I_{\times}^{(\beta)}\left(\mathbf{B}^{1/2}\mathbf{V}\mathbf{A}\mathbf{V}^{*}\mathbf{B}^{1/2},\boldsymbol{t}\right)\right]=\mathbb{E}_{\mathbf{V},\mathbf{V}'}\left[\Delta_{\boldsymbol{t}}\left(\left(\mathbf{V}'\mathbf{B}^{1/2}\right)\mathbf{V}\mathbf{A}\mathbf{V}^{*}\left(\mathbf{V}'\mathbf{B}^{1/2}\right)^{*}\right)\right]\,. \tag{3.17}$$

If we use the Cholesky decomposition once again: $\mathbf{V}\mathbf{A}\mathbf{V}^{*}=\mathbf{R}_A\mathbf{R}_A^{*}$ and the *QR decomposition* $\mathbf{V}'\mathbf{B}^{1/2}=\mathbf{Q}\mathbf{R}_B$ with $\mathbf{Q}\in\mathsf{O}_{\beta}(N)$ (such that up to a change of variable we have $\mathbf{V}'\mathbf{B}\mathbf{V}'^{*}=\mathbf{R}_B\mathbf{R}_B^{*}$), we have:

$$\mathbb{E}_{\mathbf{V}}\left[I_{\times}^{(\beta)}\left(\mathbf{B}^{1/2}\mathbf{V}\mathbf{A}\mathbf{V}^{*}\mathbf{B}^{1/2},\boldsymbol{t}\right)\right]=\mathbb{E}_{\mathbf{R}_A,\mathbf{R}_B}\left[\Delta_{\boldsymbol{t}}\left(\left(\mathbf{R}_B\mathbf{R}_A\right)\left(\mathbf{R}_B\mathbf{R}_A\right)^{*}\right)\right]\,. \tag{3.18}$$

The matrix $\mathbf{R}_B\mathbf{R}_A$ is the product of two triangular matrices hence it is also triangular. Its diagonal entries are given by $(R_A)_{kk}\cdot(R_B)_{kk}$ and since $\mathbf{R}_A$ and $\mathbf{R}_B$ are independent, we have:

$$\mathbb{E}_{\mathbf{V}}\left[I_{\times}^{(\beta)}\left(\mathbf{B}^{1/2}\mathbf{V}\mathbf{A}\mathbf{V}^{*}\mathbf{B}^{1/2},\boldsymbol{t}\right)\right]=\mathbb{E}_{\mathbf{R}_A}\left[\prod_{k=1}^{N}(R_A)_{kk}^{2t_k}\right]\cdot\mathbb{E}_{\mathbf{R}_B}\left[\prod_{k=1}^{N}(R_B)_{kk}^{2t_k}\right]\,, \tag{3.19}$$

which concludes the proof of the averaging property by Eq. (3.13). □

Spherical integrals in the abstract setting of Gelfand pair and roots systems -

The three averaging properties of Res. 3.1, 3.2, 3.3 can all be written in the following abstract manner:

$$\int_K \varphi(g_1 k g_2)\mu_{\mathrm{Haar}}(\mathrm{d}k)=\varphi(g_1)\varphi(g_2)\qquad\text{for all } g_1,g_2\in G, \tag{3.20}$$

where $(G//K)$ is a so-called *Gelfand pair* or *symmetric space*: G is a group and K is a compact subgroup of G with certain specific conditions between the twos. The construction of a harmonic analysis theory on this type of generic space has received a lot of attention in the past century with the seminal work of Harish-Chandra [88], and Helgason [90] to cite a few. Depending on the geometry of the space $(G//K)$ (Euclidean, compact or non-compact), the root system associated with the group G and its multiplicity (which depends on β), one can show that these spherical integrals satisfy a certain sophisticated partial differential equation. One can then go one step further by even removing the concept of a Gelfand pair and only consider 'spherical functions' as solutions of this differential equation where now the multiplicity of the associated root system takes an 'arbitrary' value. In our setting, this means that one can naturally extend spherical integrals to any $\beta>0$. In Sec. 3.2.2, we will see how to construct this differential equation without appealing to the theory of root systems. For completeness, I give below the geometry and root system associated with the three examples studied in this thesis and refer to Ref. [90, 64] for further details regarding harmonic analysis on these spaces.

Spherical Integral	Geometry of G.P./symm. space	Root system
$\mathcal{I}^{(\beta)}$	Euclidean (flat)	A_{N-1}
$\mathcal{I}_{\times}^{(\beta)}$	non-compact (negative curvature)	A_{N-1}
$\mathcal{I}_{q}^{(\beta)}$	Euclidean (flat)	B_{N-1}

Let's mention that for the multiplicative case, our choice to fix the multiplicative spherical with the averaging property of Eq. (3.6) corresponds to the normalization:

$$\mathcal{I}_{\times}^{(\beta)}(\boldsymbol{a}, \mathbf{0}) = 1 \qquad \text{for all } \boldsymbol{a} \in (\mathbb{R}_{+}^{*})^{N}\,. \tag{3.21}$$

However unlike the additive and rectangular case, this choice of normalization does not make the spherical integral invariant by permutation of the entries of $\boldsymbol{t}$ and for group theoretic reasons on non-compact spaces, it sometimes appears more natural to consider a shifted version of this multiplicative spherical integral such that one has this property. In particular, the so-called *Heckman-Opdam hypergeometric function* [89] which has seen recent applications in problems arising in RMT, see for example Refs. [31, 75], can be seen as a shifted version of our convention. In the rest of this thesis we will not need the use of this shifted version and stick to our convention and definition of the multiplicative spherical integral.

3.2.1 Determinantal formulae for $\beta = 2$

In the special case $\beta = 2$, one can simplify the integral over the unitary groups $\mathrm{U}(N)$ (and $\mathrm{U}(M)$) to write the spherical integrals as a ratio of the determinant of a specific matrix over two Vandermonde determinants. These *determinantal formulae* can be obtained by different methods. I give here the results without proof.

For the additive spherical integral, this determinantal formula is already (indirectly) present in the original work of Harish-Chandra [88] and has been re-discovered by Itzykson and Zuber in high-energy physics [96]. It is given by the following result.

Result 3.4 (*Itzykson-Zuber determinantal formula for $\beta = 2$ [88, 96]*)

For $\beta = 2$, the additive spherical integral of Eq. (3.1) admits the following determinantal formula:

$$\mathcal{I}^{(\beta=2)}(\boldsymbol{a}, \boldsymbol{t}) = C_N \frac{\det\left[\mathrm{e}^{a_i t_j}\right]}{\Delta(\boldsymbol{a})\Delta(\boldsymbol{t})}\,, \tag{3.22}$$

where $C_N = \prod_{i=1}^{N-1} i!$, and $\Delta(.)$ is the Vandermonde determinant.

Note that when two (or more) components of $\boldsymbol{a}$ or $\boldsymbol{t}$ are equal, one needs to understand Eq. (3.22) with L'Hospital's rule.

The determinantal formula for the multiplicative case has been derived by Gelfand and Naĭmark in Ref. [74] the study of harmonic analysis on symmetric spaces and is given by:

Result 3.5 (*Gelfand-Naimark determinantal formula for $\beta = 2$ [74]*)

For $\beta = 2$, the multiplicative spherical integral of Eq. (3.2) admits the following determi-

nantal formula:

$$\mathcal{I}_{\times}^{(\beta=2)}(\boldsymbol{a},\boldsymbol{t}-\boldsymbol{s}_0)=C_N\frac{\det\left[{a_i}^{t_j}\right]}{\Delta(\boldsymbol{a})\Delta(\boldsymbol{t})}\,, \tag{3.23}$$

where $\boldsymbol{s}_0=(N-1,\dots,1,0)$ *and* $C_N=\prod_{i=1}^{N-1}i!$*, and* $\Delta(.)$ *is the Vandermonde determinant.*

Anticipating a bit, let's mention that this formula is very reminiscent of the determinantal formula for the famous *Schur polynomials*, and as we will see in the next section this comes with no surprise as the two objects are closely related.

The determinantal formula for the rectangular spherical integral has been derived by several authors and is given by the following result.

Result 3.6 (*Schlittgen-Wettig determinantal formula for $\beta=2$ [160]*)

For $\beta=2$*, the rectangular spherical integral of Eq.* (3.3) *admits the following determinantal formula:*

$$\mathcal{I}_{q}^{(\beta=2)}(\boldsymbol{a},\boldsymbol{t})=C_{N,M}\frac{\det\left[I_{M-N}(2a_it_j)\right]}{\Delta(\boldsymbol{a}^2)\Delta(\boldsymbol{t}^2)\prod_{i=1}^{N}(a_it_j)^{M-N}}\,, \tag{3.24}$$

where $\Delta(.)$ *is the Vandermonde determinant,* $C_{N,M}:=\prod_{i=1}^{N}(M-i)!(N-i)!$ *and* I_{M-N} *is the Bessel function* $I_\nu(2x):=x^\nu\sum_{k=0}^{\infty}\frac{y^k}{k!(k+\nu)!}$*.*

One may notice that in each case the argument of the determinant in the numerator is the natural integrand of respectively the moment generating function, the Mellin transform and the Hankel transform and since the corresponding spherical integrals are their 'random matrix counterpart', it would be interesting to know if there is a simple argument for this phenomenon.

For large N, these determinantal expressions are difficult to use in practice since they involved an alternating sum with $N!$ terms. However, in some specific cases (which will be detailed later in this chapter and in the following one), these formulae considerably simplify and are useful to study the large N limit.

3.2.2 Extension of spherical integrals to $\beta>0$ and Jack polynomials

In order to extend the definition of spherical integrals to any values of $\beta>0$, the idea is very similar to the definition of β-ensembles, which have been obtained as an analytical continuation in the parameter $\beta>0$ of the joint law of eigenvalues of matrices taken initially from a β-ensemble with $\beta\in\{1,2,4\}$, as described in Sec. 1.4. For simplicity, we will describe in detail only the case of the additive spherical integral. Thus, one wants to find a natural description of this additive spherical integral, where β appears simply as a parameter. To do so, let's remark that for $\beta=\{1,2,4\}$ and $\mathbf{A},\mathbf{B}\in\mathrm{Herm}_\beta(N)$ we have:

$$\boldsymbol{\Delta}_{\mathbf{A}}\,\mathrm{e}^{\mathrm{Tr}\,\mathbf{AB}}=\left(\mathrm{Tr}\,\mathbf{B}^2\right)\,\mathrm{e}^{\mathrm{Tr}\,\mathbf{AB}}\,, \tag{3.25}$$

where

$$\boldsymbol{\Delta}_{\mathbf{A}} := \mathrm{Tr}\frac{\partial^2}{\partial \mathbf{A}^2} = \sum_{i=1}^{N}\frac{\partial^2}{\partial A_{ii}^2} + \frac{1}{2}\sum_{i<j}\sum_{b=1}^{\beta}\frac{\partial^2}{\partial {A_{ij}^{(b)}}^2}\,, \tag{3.26}$$

is the *Laplace-Beltrami* operator over the (flat) space of self-adjoint matrices. Clearly, since $I^{(\beta)}(\mathbf{A},\mathbf{B})$ can be seen as a uniform average over the matrix of eigenvectors of $\mathbf{B}$, it is a β-invariant solution of the same Helmholtz's equation/eigenvalue problem:

$$\boldsymbol{\Delta}_{\mathbf{A}}\, I^{(\beta)}(\mathbf{A},\mathbf{B}) = \left(\mathrm{Tr}\,\mathbf{B}^2\right)\, I^{(\beta)}(\mathbf{A},\mathbf{B})\,, \tag{3.27}$$

and thus the goal is to understand how this differential operator is modified when we do the change of variable from the matrix arguments ($I^{(\beta)}(\mathbf{A},\mathbf{B})$) to their eigenvalues ($\mathcal{I}^{(\beta)}(\boldsymbol{a},\boldsymbol{b})$), since the latter vectors $\boldsymbol{a},\boldsymbol{b}\in\mathbb{R}^N$ do not depend explicitly on $\beta>0$. The final result of this change of variable is well-known (see Refs. [81, 33]) and I propose here to find it using a variational approach. Let ϕ be a function of the entries of the matrix $\mathbf{A}$ and define the following functional:

$$J\left[\phi\right] := \int\left(\frac{1}{2}\left|\boldsymbol{\nabla}_{\mathbf{A}}\phi\right|^2 + \frac{1}{2}\left(\mathrm{Tr}\,\mathbf{B}^2\right)\phi^2\right)\mathrm{d}\mathbf{A}\,, \tag{3.28}$$

where $\boldsymbol{\nabla}_{\mathbf{A}}$ is the (flat) gradient over the space of self-adjoint matrices. As it is well known, if we look at the extrema of this energy functional we get back the eigenvalue problem since we have:

$$\frac{\delta J\left[\phi\right]}{\delta\phi} = 0 = -\boldsymbol{\Delta}_{\mathbf{A}}\,\phi + \left(\mathrm{Tr}\,\mathbf{B}^2\right)\phi\,. \tag{3.29}$$

Thus if we apply Weyl's formula for the eigenvalue decomposition of $\mathbf{A}=\mathbf{V}\mathbf{Diag}(\boldsymbol{a})\mathbf{V}^*$ we get that $\mathcal{I}^{(\beta)}(\boldsymbol{a},\boldsymbol{b})$ is the solution of

$$\frac{\delta}{\delta\phi}\mathcal{J}[\phi=\mathcal{I}^{(\beta)}(\boldsymbol{a},\boldsymbol{b})] = 0 \quad , \tag{3.30}$$

with

$$\mathcal{J}[\phi] = \mathrm{Cst}\cdot\int\left(\frac{1}{2}\sum_{i=1}^{N}\left(\frac{\partial}{\partial_{a_i}}\phi\right)^2 + \frac{1}{2}\left(\sum_{i=1}^{N}b_i^2\right)\phi^2\right)\,|\Delta(\boldsymbol{a})|^\beta\,\mathrm{d}\boldsymbol{a}\,. \tag{3.31}$$

The extrema of this functional is given by the Euler-Lagrange equation:

$$\frac{\partial L}{\partial\phi} - \nabla_{\boldsymbol{a}}\cdot\frac{\partial L}{\partial(\nabla_{\boldsymbol{a}}\phi)} = 0 \quad \text{with} \quad L[\boldsymbol{a},\phi,\nabla_{\boldsymbol{a}}\phi] := |\Delta(\boldsymbol{a})|^\beta\left(\frac{(\sum_{i=1}^N b_i^2)}{2}\phi^2 + \frac{1}{2}|\nabla_{\boldsymbol{a}}\phi|^2\right)\,, \tag{3.32}$$

that is

$$\left(\sum_{i=1}^{N}b_i^2\right)\,|\Delta(\boldsymbol{a})|^\beta\mathcal{I}^{(\beta)}(\boldsymbol{a},\boldsymbol{b}) - \sum_{i=1}^{N}\partial_{a_i}\left(|\Delta(\boldsymbol{a})|^\beta\partial_{a_i}\mathcal{I}^{(\beta)}(\boldsymbol{a},\boldsymbol{b})\right) = 0\,, \tag{3.33}$$

and since $|\Delta(\boldsymbol{a})|^\beta = \exp\left[\beta/2 \sum_{i,j|i\neq j} \log|a_i - a_j|\right]$, by taking the derivative, we get after simplification the following final eigenvalue problem:

$$\boxed{\left(\sum_{i=1}^N \frac{\partial^2}{\partial_{a_i}^2} + \beta \sum_{i,j|j\neq i} \frac{1}{a_i - a_j}\frac{\partial}{\partial_{a_i}}\right) \mathcal{I}^{(\beta)}(\boldsymbol{a},\boldsymbol{b}) = \left(\sum_{i=1}^N b_i^2\right) \mathcal{I}^{(\beta)}(\boldsymbol{a},\boldsymbol{b})\,.} \tag{3.34}$$

As a consequence, we can define the additive spherical function $\mathcal{I}^{(\beta)}(\boldsymbol{a},\boldsymbol{b})$ for any $\beta > 0$ as the solution of the eigenvalue problem of Eq. (3.34) invariant by permutation of $\boldsymbol{a}$ and $\boldsymbol{b}$, invariant by permutation their entries and normalized such that $\mathcal{I}^{(\beta)}(\boldsymbol{a},\boldsymbol{0}) = 1$. This definition is in accordance with what one would obtain in the abstract setting of root systems. It turns out that one can also show that this spherical integral is the eigenfunction of a countable family of differential operators known as symmetrized Dunkl operator but we will not use this point of view in the rest of this thesis. Note that for $\beta \notin \{1,2,4\}$ we still denote $\mathcal{I}^{(\beta)}(\boldsymbol{a},\boldsymbol{b})$ as a 'spherical integral' even-though one does have an integral representation over a group in this case. For this reason, this function is sometimes also referred to as a multivariate Bessel function in this general setting[1].

One may expect to extend the multiplicative spherical integral $\mathcal{I}_\times^{(\beta)}$ in a similar fashion by looking at how the Laplace-Beltrami operator over the curved space $\mathrm{Herm}_\beta^{++}(N)$ is modified when doing the change of variable from a matrix to its eigenvalues. One can then show that for $\beta \in \{1,2,4\}$, this spherical integral is an eigenfunction of:

$$\left(\sum_{i=1}^N a_i^2 \frac{\partial^2}{\partial_{a_i}^2} + \beta \sum_{i,j|i\neq j} \frac{a_i^2}{a_i - a_j}\frac{\partial}{\partial_{a_i}}\right) \mathcal{I}_\times^{(\beta)}(\boldsymbol{a},\boldsymbol{t}) = e_{\boldsymbol{t}}\, \mathcal{I}_\times^{(\beta)}(\boldsymbol{a},\boldsymbol{t})\,. \tag{3.35}$$

where $e_{\boldsymbol{t}}$ is a (sophisticated) function depending on $\boldsymbol{t}$. If we re-scale $\boldsymbol{t}$ by $\boldsymbol{t}/\epsilon$, for small ϵ, one can show $e_{\boldsymbol{t}/\epsilon} = \sum_{i=1} t_i^2/\epsilon^2 + o(1/\epsilon^2)$. Similarly, if we perform the change of variable $\boldsymbol{a} \to \mathrm{e}^{\epsilon \boldsymbol{a}}$, the differential operator in the LHS of Eq. (3.35) becomes the differential operator of the LHS of Eq. (3.34) divided by ϵ^2. This leads to the following relation between the additive and the spherical integral:

$$\mathcal{I}^{(\beta)}(\boldsymbol{a},\boldsymbol{b}) = \lim_{\epsilon\to 0^+} \mathcal{I}_\times^{(\beta)}(\mathrm{e}^{\epsilon\boldsymbol{a}}, \boldsymbol{b}/\epsilon)\,, \tag{3.36}$$

which again can be seen as the fact that the sum of self-adjoint is a limiting case of the product.

Let us mention that there exists an analogous formulation in the rectangular case, see for example Ref. [67] and references therein.

Relation to the Calogero-Moser-Sutherland (CMS) system -

If we define the function

$$\Psi^{(\beta)}(\boldsymbol{a},\boldsymbol{t}) := \Delta(\boldsymbol{a})^{\beta/2}\, \mathcal{I}^{(\beta)}(\boldsymbol{a},\boldsymbol{t})\,, \tag{3.37}$$

then one can show by injecting the expression of $\Psi^{(\beta)}$ in Eq. (3.34), that it satisfies a differential equation:

$$\mathcal{H}^{\mathrm{CMS}}\, \Psi^{(\beta)} = E_{\boldsymbol{t}}\, \Psi^{(\beta)}\,, \tag{3.38}$$

[1] for $N = 1$ one can show that the additive spherical integral is a Bessel function

where $\mathcal{H}^{\mathrm{CMS}}$ is the quantum rational **Calogero-Moser-Sutherland** Hamiltonian:

$$\mathcal{H}^{\mathrm{CMS}} := \left(\sum_{i=1}^{N} \frac{\partial^2}{\partial_{a_i}^2} - \beta \left(\frac{\beta}{2} - 1 \right) \cdot \sum_{i,j|i\neq j} \frac{1}{(a_i - a_j)^2} \right) , \tag{3.39}$$

and the energy is given by $E_{\boldsymbol{t}} := \sum_{i=1}^{N} t_i^2$. The CMS operator is an important integrable system in physics, which also has a fundamental role in the study of root systems. For the multiplicative case, there exist a variant of this property where the rational CMS is replaced by its hyperbolic version, see for example Ref. [165] and references therein.

Expansion in terms of Jack polynomials -

In order to expand the additive spherical function, let's first consider the case $\beta \in \{1, 2, 4\}$ for which we have the integral representation of Eq. (3.1). As we will shortly see, in each of these three cases we have an expansion in terms of a family of symmetric polynomials which admits a natural extension to all $\beta > 0$. We have:

$$\mathcal{I}^{(\beta)}(\boldsymbol{a}, \boldsymbol{b}) = \mathbb{E}_{\mathbf{V} \sim \mathrm{Unif}[\mathsf{O}_\beta(N)]} \left[\mathrm{e}^{\mathrm{Tr}\, \mathbf{AVBV}^*} \right] \qquad (\text{for } \beta \in \{1, 2, 4\}) , \tag{3.40}$$

which by expanding the exponential gives:

$$\mathcal{I}^{(\beta)}(\boldsymbol{a}, \boldsymbol{b}) = \sum_{k=0}^{\infty} \frac{\mathbb{E}_{\mathbf{V} \sim \mathrm{Unif}[\mathsf{O}_\beta(N)]} \left[\mathrm{Tr}\, (\mathbf{AVBV}^*)^k \right]}{k!} , \tag{3.41}$$

Because we want to disentangle the contribution from $\boldsymbol{a}$ from the one from $\boldsymbol{b}$ in the expression of the additive spherical integral, the idea is to decompose the power-sum polynomial $\mathrm{Tr}\, (\mathbf{AVBV}^*)^k$ into a *basis* of symmetric polynomials which behave nicely with respect to the uniform average over $\mathsf{O}_\beta(N)$. It turns out that for each of the three cases $\beta = 1, 2, 4$, there exist such a family of symmetric polynomials indexed by a partition written as a N-tuple of non-increasing non-negative integers $\boldsymbol{\lambda} = (\lambda_1 \geq \cdots \geq \lambda_N)$, which we generically denote by $J_{\boldsymbol{\lambda}}^{\left(\frac{\beta}{2}\right)}$,

$$J_{\boldsymbol{\lambda}}^{\left(\frac{\beta}{2}\right)}(\boldsymbol{a}) \propto \begin{cases} Z_{\boldsymbol{\lambda}}(\boldsymbol{a}) & \text{for } \beta = 1 , \\ s_{\boldsymbol{\lambda}}(\boldsymbol{a}) & \text{for } \beta = 2 , \\ Z_{\boldsymbol{\lambda}}^{(\mathrm{Q})}(\boldsymbol{a}) & \text{for } \beta = 4 . \end{cases} \tag{3.42}$$

where $Z_{\boldsymbol{\lambda}}$ is the zonal symmetric polynomial, $s_{\boldsymbol{\lambda}}$ is the Schur polynomial and $Z_{\boldsymbol{\lambda}}^{(\mathrm{Q})}$ is the quaternionic zonal polynomial. By abuse of notation, if we write $\mathrm{J}_{\boldsymbol{\lambda}}^{\left(\frac{\beta}{2}\right)}(\mathbf{AVBV}^*)$ for the symmetric polynomials in the eigenvalues of $\mathbf{AVBV}^*$, for each of the three cases $\beta \in \{1, 2, 4\}$, we have:

$$\mathbb{E}_{\mathbf{V} \sim \mathrm{Unif}[\mathsf{O}_\beta(N)]} \left[\mathrm{J}_{\boldsymbol{\lambda}}^{\left(\frac{\beta}{2}\right)}(\mathbf{AVBV}^*) \right] = \frac{\mathrm{J}_{\boldsymbol{\lambda}}^{\left(\frac{\beta}{2}\right)}(\boldsymbol{a})\, \mathrm{J}_{\boldsymbol{\lambda}}^{\left(\frac{\beta}{2}\right)}(\boldsymbol{b})}{\mathrm{J}_{\boldsymbol{\lambda}}^{\left(\frac{\beta}{2}\right)}(\mathbf{1})} , \tag{3.43}$$

and the power sum polynomial is given as a linear combination of such zonal/Schur polynomials:

$$\frac{1}{k!}\operatorname{Tr}\left(\mathbf{AVBV}^*\right)^k = \sum_{|\boldsymbol{\lambda}|=k} \frac{1}{H_{\boldsymbol{\lambda}}\left(\frac{\beta}{2}\right)} \cdot \mathrm{J}_{\boldsymbol{\lambda}}^{\left(\frac{\beta}{2}\right)}\left(\mathbf{AVBV}^*\right) , \tag{3.44}$$

where the coefficient in front of the Jack polynomial is given below.

For $\beta \in \{1,2,4\}$ these polynomials can be equivalently defined thanks to the *generalized Cauchy identity*:

$$\prod_{i=1}^{N}\prod_{j=1}^{N}(1-a_i b_j)^{-\frac{\beta}{2}} =: \sum_{|\lambda|=N} \frac{H'_{\boldsymbol{\lambda}}\left(\frac{\beta}{2}\right)}{H_{\boldsymbol{\lambda}}\left(\frac{\beta}{2}\right)} \cdot \mathrm{J}_{\boldsymbol{\lambda}}^{\left(\frac{\beta}{2}\right)}(\boldsymbol{a})\ \mathrm{J}_{\boldsymbol{\lambda}}^{\left(\frac{\beta}{2}\right)}(\boldsymbol{b}) , \tag{3.45}$$

and they admit the normalization:

$$\mathrm{J}_{\boldsymbol{\lambda}}^{\left(\frac{\beta}{2}\right)}(\mathbf{1}) = \frac{(\alpha)_{\boldsymbol{\lambda};\beta/2}}{H'_{\boldsymbol{\lambda}}\left(\frac{\beta}{2}\right)} , \tag{3.46}$$

where the constants are defined by:

$$H_{\boldsymbol{\lambda}}\left(\frac{\beta}{2}\right) := \prod_{s\in\boldsymbol{\lambda}}\left(a(s)+\frac{\beta}{2}l(s)+1\right) , \tag{3.47}$$

$$H'_{\boldsymbol{\lambda}}\left(\frac{\beta}{2}\right) := \prod_{s\in\boldsymbol{\lambda}}\left(a(s)+\frac{\beta}{2}l(s)+\frac{\beta}{2}\right) , \tag{3.48}$$

$$(\alpha)_{\boldsymbol{\lambda};\beta/2} := \prod_{i,j>0|j<\lambda_i}\left(\alpha+(j-1)-\frac{\beta}{2}(i-1)\right) , \tag{3.49}$$

with $a(s)$ and $l(s)$ the arm-length and leg-length of the box s of the partition $\boldsymbol{\lambda}$ seen as a Young diagram, see Ref. [164, 150] for further details. let's mention that these constants will also not play a major role in the following.

Importantly, because the LHS of Eq. (3.45) makes sense for any $\beta > 0$, one can naturally extend the symmetric polynomials $\mathrm{J}_{\boldsymbol{\lambda}}^{\left(\frac{\beta}{2}\right)}$ to all $\beta > 0$ and under this setting, the $\mathrm{J}_{\boldsymbol{\lambda}}^{\left(\frac{\beta}{2}\right)}$ are known as **Jack polynomials** with index $\beta/2$. The Jack polynomials form a rich one-parameter family of symmetric polynomials, and I refer to [115] and [164] for more properties concerning these polynomials.

All in all, this gives the following expansion for the additive spherical integral valid for any $\beta > 0$:

$$\boxed{\mathcal{I}^{(\beta)}(\boldsymbol{a},\boldsymbol{b}) = \sum_{k=0}^{\infty}\sum_{|\boldsymbol{\lambda}|=k} \frac{H'_{\boldsymbol{\lambda}}\left(\frac{\beta}{2}\right)}{\left(\frac{N\beta}{2}\right)_{\boldsymbol{\lambda};\beta/2} H_{\boldsymbol{\lambda}}\left(\frac{\beta}{2}\right)} \cdot \mathrm{J}_{\boldsymbol{\lambda}}^{\left(\frac{\beta}{2}\right)}(\boldsymbol{a})\ \mathrm{J}_{\boldsymbol{\lambda}}^{\left(\frac{\beta}{2}\right)}(\boldsymbol{b}) ,}\tag{3.50}$$

and one can show that this definition is in accordance with the representation in terms of the differential equation of Eq. (3.34).

For the rectangular case, there exists a similar expansion in terms of Jack polynomials and I refer to Ref. [67] for further details. However, for the multiplicative spherical integral (or equivalently

for the Heckman-Opdam function), there is - as far as I know - no known expansion in terms of symmetric polynomials. Nevertheless, since for $\beta \in \{1, 2, 4\}$, Eq. (3.43) is reminiscent of the property of Eq. (3.6) for the multiplicative spherical integral, one may expect the two quantities to be closely related and indeed if we introduce the normalized Jack polynomials:

$$\hat{\mathrm{J}}^{(\theta)}_{\boldsymbol{\lambda}}(\boldsymbol{a}) := \frac{\mathrm{J}^{(\theta)}_{\boldsymbol{\lambda}}(\boldsymbol{a})}{\mathrm{J}^{(\theta)}_{\boldsymbol{\lambda}}(\mathbf{1})}\,, \tag{3.51}$$

then when the conjugate argument $\boldsymbol{t}$ of the multiplicative spherical integral is a non-increasing sequence of integers $\boldsymbol{t} = \boldsymbol{m}$, we have the following identity:

$$\mathcal{I}^{(\beta)}_{\times}(\boldsymbol{a}, \boldsymbol{m}) = \hat{\mathrm{J}}^{(\beta/2)}_{\boldsymbol{m}}(\boldsymbol{a}) \qquad \text{for } \boldsymbol{m} \in \mathbb{N}^N \text{ with } m_1 \geq \ldots, \geq m_N \geq 0\,, \tag{3.52}$$

such that one can think of the multiplicative spherical integral as the analytical continuation (in the index) of the Jack polynomials.

The situation is somehow very analog to the case of classical probability where the moment generating $\mathbb{E}\left[\mathrm{e}^{tX}\right] = \sum_{k=0}^{\infty} \mathbb{E}[X^k] t^k / k!$ which is multiplicative for the sum of two independent random variables, naturally expands in terms of the 'moments' $\mathbb{E}[X^k]$ and t^k, while the Mellin transform $\mathbb{E}[X^t]$, which is multiplicative for the product of independent random variables, is the analytical continuation of the moments of X for non-integer values. In the context of spherical integrals, what plays the role of the moments are Jack polynomials.

From Eq. (3.36), one can see the additive spherical integral as a limiting case of the multiplicative spherical integral and by Eq. (3.52), one can also see it as a limit of normalized Jack polynomials:

$$\mathcal{I}^{(\beta)}(\boldsymbol{a}, \boldsymbol{t}) = \lim_{\epsilon \to 0^+} \hat{\mathrm{J}}^{\left(\frac{\beta}{2}\right)}_{\lfloor \epsilon^{-1} \boldsymbol{t} \rfloor} \left(\mathrm{e}^{\epsilon \boldsymbol{a}}\right)\,, \tag{3.53}$$

where $\boldsymbol{t}$ is a set of non-decreasing tuples and $\lfloor \boldsymbol{t} \rfloor$ denotes the vector whose components are given as the integer part of the ones of $\boldsymbol{t}$. A similar statement holds for the multiplicative spherical integral, where the Jack polynomial in Eq. (3.53) is replaced by a certain *Macdonald polynomial*, a two-parameter generalization of the former, see Ref. [115] for more details.

Positivity conjecture for the Jack-Littlewood-Richardson coefficients and sum of β-ensembles -

The family of Jack polynomials satisfies a certain closure property for the product since we have:

$$\hat{\mathrm{J}}^{(\beta/2)}_{\boldsymbol{\lambda}}(.) \hat{\mathrm{J}}^{(\beta/2)}_{\boldsymbol{\mu}}(.) = \sum_{\boldsymbol{\nu}} c^{\boldsymbol{\nu}}_{\boldsymbol{\lambda}, \boldsymbol{\mu}}(\beta/2)\ \hat{\mathrm{J}}^{(\beta/2)}_{\boldsymbol{\nu}}(.)\,. \tag{3.54}$$

For $\beta = 2$, Jack polynomials degenerate into Schur polynomials and Eq. (3.54) is nothing else than (a normalized version of) the famous Littlewood-Richardson rule. For $\beta > 0$, The coefficients $c^{\boldsymbol{\nu}}_{\boldsymbol{\lambda}, \boldsymbol{\mu}}(\beta/2)$ are the (normalized) Jack-Littlewood-Richardson (JLR) coefficients and play an important role in the theory of symmetric polynomials and in representation theory, see Ref. [115]. By Eq. (3.53), the sum in Eq. (3.54) becomes a N-dimensional integral for the additive spherical integral, and we have:

$$\mathcal{I}^{(\beta)}(\boldsymbol{a}, .)\, \mathcal{I}^{(\beta)}(\boldsymbol{b}, .) = \int \mathcal{I}^{(\beta)}(\boldsymbol{c}, .)\, \mathcal{P}^{(\beta)}(\boldsymbol{c} | \boldsymbol{a}, \boldsymbol{b})\ \mathrm{d}\boldsymbol{c}\,. \tag{3.55}$$

where we have used the notations $\mathcal{P}^{(\beta)}(\boldsymbol{c}|\boldsymbol{a},\boldsymbol{b})$ since for $\beta \in \{1,2,4\}$, this simply corresponds to the joint distribution of the corresponding Horn problem of Sec. 2.6.1. For other values of β, this quantity is *a priori* not a probability distribution but a signed measure. In fact, there exists a general formula for Macdonald polynomials (which contains the Jack polynomials as special cases) and this induces a similar interpretation for the multiplicative spherical:

$$\mathcal{I}_{\times}^{(\beta)}(\boldsymbol{a},.)\,\mathcal{I}_{\times}^{(\beta)}(\boldsymbol{b},.) = \int \mathcal{I}_{\times}^{(\beta)}(\boldsymbol{c},.)\,\mathcal{P}_{\times}^{(\beta)}(\boldsymbol{c}|\boldsymbol{a},\boldsymbol{b})\,\mathrm{d}\boldsymbol{c}\,, \tag{3.56}$$

where again $\mathcal{P}_{\times}^{(\beta)}$ corresponds to the joint density of the multiplicative Horn of Sec. 2.6.1 for $\beta = 1,2,4$ and is *a priori* a signed measure for other values of β.

For any N, any partitions $\boldsymbol{\lambda},\boldsymbol{\mu}$, one can show that the coefficients of the combinatorial formula of the Macdonald polynomials sum to one and this induces the following result:

$$\sum_{\boldsymbol{\nu}} c_{\boldsymbol{\lambda},\boldsymbol{\mu}}^{\boldsymbol{\nu}}(\beta/2) = \int \mathcal{P}^{(\beta)}(\boldsymbol{c}|\boldsymbol{a},\boldsymbol{b})\,\mathrm{d}\boldsymbol{c} = \int \mathcal{P}_{\times}^{(\beta)}(\boldsymbol{c}|\boldsymbol{a},\boldsymbol{b})\,\mathrm{d}\boldsymbol{c} = 1\,, \tag{3.57}$$

valid for any $\beta > 0$, any partitions $\boldsymbol{\lambda},\boldsymbol{\mu}$ and any vectors $\boldsymbol{a},\boldsymbol{b}$. What is however *much* harder to prove is the following still open - but believed to be true - conjecture first stated (under a different form) by STANLEY in Ref. [164].

Conjecture 3.1 (*Positivity of the JLR coefficients [164, 158, 135]*)

For any N, any $\beta > 0$, any partitions $\boldsymbol{\lambda},\boldsymbol{\mu}$ and any vectors $\boldsymbol{a},\boldsymbol{b}$, all the coefficients $c_{\boldsymbol{\lambda},\boldsymbol{\mu}}^{\boldsymbol{\nu}}(\beta/2)$, $\mathcal{P}^{(\beta)}(\boldsymbol{c}|\boldsymbol{a},\boldsymbol{b})$ and $\mathcal{P}_{\times}^{(\beta)}(\boldsymbol{c}|\boldsymbol{a},\boldsymbol{b})$ are non-negative.

This conjecture will play an important role in the construction of the high-temperature convolution of Chapter 6. If this conjecture holds, this means that even though one does not have a 'matrix-model' for the sum/product of invariant '$\beta > 0$' self-adjoint matrices, one can really think of $\mathcal{P}^{(\beta)}$ and $\mathcal{P}_{\times}^{(\beta)}$ as an *extrapolation* of the associated probability distributions, a property we have already encountered when dealing with rank-one and infinitesimal perturbations in the previous chapter. Note that in this case, one should expect the expression of the joint densities given in terms of the spherical integrals given by Eq. (2.121) to be valid for all $\beta > 0$.

3.3 Extensive rank limit of spherical integrals

In this section, we consider the large N limit of spherical integrals in the scaling where the conjugate variable $\boldsymbol{b} \equiv \boldsymbol{t}$ is of 'full-rank' and its distribution converges to a certain LSD μ_B. Our description of this full-rank limit will be relatively brief since our main focus will be on the other 'rank-one' limit.

3.3.1 Large N limit of HCIZ integral and Matytsin's variational principle

The starting point of the computation is the joint law of the DBM given by Eq. (2.77) from which one can express the additive spherical integral has:

$$\mathcal{I}^{(\beta)}\left(\boldsymbol{a},\frac{N\beta}{2}\boldsymbol{b}\right) \propto |\Delta(\boldsymbol{a})|^{-\beta}\mathrm{e}^{+\frac{N\beta}{4}\left(\sum_{i=1}^{N} a_i^2+b_i^2\right)}\,\mathcal{P}_{t=1}^{(\beta)}(\boldsymbol{c}(1)=\boldsymbol{a}|\boldsymbol{b})\,. \tag{3.58}$$

As a consequence, we need to estimate the probability of the very unlikely event where the DBM finishes at time $t = 1$ at the position $\boldsymbol{c}(1) = \boldsymbol{a}$, rather than its usual value. At large N, this probability is given by a large deviation principle. Formally the probability of the trajectory of a Brownian particle between time $t = 0$ and time $t = 1$ is given by $P(\{B_t\}) \approx \exp \int_0^1 \dot{B}_t^2/2 \, \mathrm{d}t$ and if one does the change of variable from the Brownian motions to the c_i in Eq. (2.80), one has formally for the law of a DBM starting at $\boldsymbol{b}$ and ending at $\boldsymbol{a}$:

$$\mathcal{P}_{t=1}^{(\beta)}(\boldsymbol{c}(1) = \boldsymbol{a}|\boldsymbol{b}) \approx \exp\left[-\frac{N\beta}{2} \cdot N \left(\int_0^1 L\left(\boldsymbol{c}(t), \dot{\boldsymbol{c}}(t)\right) \mathrm{d}t \, + \mathrm{BT} + \mathcal{O}(1/N) \right) \right], \tag{3.59}$$

where the *Lagrangian* L is given by:

$$L\left(\boldsymbol{c}, \boldsymbol{p}\right) := \frac{1}{N} \left(\sum_{i=1}^{N} \frac{p_i^2}{2} + \frac{1}{2N^2} \sum_{i=1}^{N} \frac{1}{(c_i - c_j)^2} \right), \tag{3.60}$$

and the boundary term is given at leading order in N:

$$\mathrm{BT} = \frac{1}{2}\Sigma[\mu_A] - \frac{1}{2}\Sigma[\mu_B], \tag{3.61}$$

where we recall $\Sigma[\mu] := \int\int \log|\lambda - \lambda'| \mu(\mathrm{d}\lambda)\mu(\mathrm{d}\lambda')$. Note that the paths of a Brownian motion are very rough and its derivative with respect to time $\dot{B}_t$ (that is the white noise process) is ill-defined, and hence so does $\dot{\boldsymbol{c}}(t)$. Nevertheless, one can give a rigorous meaning of this probability under the framework of the *Onsager-Maschlup formalism*. Let's also mention that we have also implicitly neglected the Jacobian of the change of variables from the Brownian motion to the c_i's.

The Lagrangian L is nothing else than the Lagrangian of the classical Calogero-Moser-Sutherland system. In the large N limit, one can show that this Lagrangian becomes a functional:

$$L\left(\boldsymbol{c}, \boldsymbol{p}\right) \to \mathcal{L}\left[\rho, v\right] := \int \rho(x, .) \left(\frac{v(x, .)^2}{2} + \frac{\pi^2}{3}\rho(x, .)^2 \right) \mathrm{d}x \,, \tag{3.62}$$

and for large N the action in Eq. (3.62) is dominated by the *instantons* solutions of

$$c_i(t) = -\frac{2}{N^3} \sum_{i=1}^{N} \frac{1}{\left(c_i(t) - c_j(t)\right)^3}, \tag{3.63}$$

under the constraints $c_i(0) = b_i$ and $c_i(1) = a_i$, which translates in the $N \to \infty$ into the coupled equations for ρ and v of Eq. (2.91) under the corresponding boundary value problem (see below for the precise statement and Ref. [36] for more details).

The asymptotic behavior of the full-rank additive spherical integral can be decomposed into two distinct contributions:

- The first one corresponds to a 'decoupled' term $S_{tot}^{\mathrm{BT}} = S^{\mathrm{BT}}[\mu_A] + S^{\mathrm{BT}}[\mu_B]$, which has no interactions between the distributions μ_A and μ_B. It is made of the sum of the asymptotic behavior of BT given by Eq. (3.61) *and* the asymptotic behavior of the factor term in front of the density in the RHS of Eq. (3.58). For $\mu = \mu_A$ or μ_B, we have

$$S^{\mathrm{BT}}[\mu] = \frac{1}{2}m_2[\mu] - \frac{1}{2}\Sigma[\mu], \tag{3.64}$$

or more explicitly

$$S^{\mathrm{BT}}[\mu] = \frac{1}{2}\int \lambda^2 \mu(\lambda)\mathrm{d}\lambda - \frac{1}{2}\int\int \log|\lambda' - \lambda|\mu(\lambda)\mu(\lambda')\,. \tag{3.65}$$

- The second term is the coupling *Euler-Matystin* term, it is given by the hydrodynamical limit of the Lagrangian of Eq. (3.62) under the constrained dynamics of the DBM starting at $\boldsymbol{b}$ and ending at $\boldsymbol{a}$, that is, explicitly:

$$S^{\mathrm{Mat}}[\mu_A, \mu_B] = \frac{1}{2}\int_0^1 \mathrm{d}t \int \mathrm{d}x \frac{\pi^2}{3}\rho^*(x,t)^3 + \rho^*(x,t)v^*(x,t)^2\,, \tag{3.66}$$

where ρ^* and v^* are given as the solutions of:

$$\begin{cases} \partial_t \rho^* + \partial_x\left(\rho^* v^*\right) = 0\,, \\ \partial_t v^* + \frac{1}{2}\partial_x v^{*2} - \frac{\pi^2}{2}\partial_x \rho^{*2} = 0\,, \end{cases} \tag{3.67}$$

under the fixed boundary values:

$$\begin{cases} \rho^*(x,0) = \mu_A(x)\,, \\ \rho^*(x,1) = \mu_B(x)\,. \end{cases} \tag{3.68}$$

All in all this can be summarized in the following result:

Result 3.7 (*Matystin's variational principle for the full-rank HCIZ integral [136, 86]*)

In the limit where $\mu_{\mathbf{A}} := \sum_{i=1}^N \delta(.-a_i)/N \to \mu_A$ *and* $\mu_{\mathbf{B}} := \sum_{i=1}^N \delta(.-b_i)/N \to \mu_B$, *we have:*

$$\frac{2}{N^2\beta}\log \mathcal{I}^{(\beta)}\left(\boldsymbol{a}, \frac{N\beta}{2}\boldsymbol{b}\right) \to \mathcal{F}[\mu_A, \mu_B] := -\frac{3}{4} + S^{\mathrm{BT}}[\mu_A] + S^{\mathrm{BT}}[\mu_B] - S^{\mathrm{Mat}}[\mu_A, \mu_B]\,, \tag{3.69}$$

where S^{BT} *is given by Eq.* (3.65) *and* S^{Mat} *is given by Eq.* (3.66).

The constant $3/4$ corresponds to the asymptotic behavior of the normalization constant term in front of Eq. (3.58) and can be obtained by Stirling formula, see Ref. [36].

This result appeared first in the high-energy physics literature in the work of Matytsin in Ref. [136] and as then be put on a rigorous ground by Guionnet and Zeitouni in Ref. [86].

The difficulty in estimating Eq. (3.69) comes from the Matystin term from which there is no closed formula for arbitrary μ_A and μ_B. However, for two appropriate choices of μ_A (or μ_B by symmetry), one can evaluate the functional $\mathcal{F}[\mu_A, \mu_B]$ exactly and this is briefly described below.

Simplification in specific cases-

.

- First, since the limiting function $\mathcal{F}$ does not depend on the parameter β, we can study the case $\beta = 2$ for which we have the determinantal formula of Eq. (3.22). Second, if we choose the $b_{0,j} = (j-1)/N$ we have:

$$\mathcal{I}\left(\boldsymbol{a}, N\boldsymbol{b}_0\right) = \frac{\det\left[\left(\mathrm{e}^{a_i}\right)^{j-1}\right]}{\Delta(\boldsymbol{a})}\,, \tag{3.70}$$

where we have used the fact that $\Delta(N\boldsymbol{b}_0) = C_N$ exactly cancels the factor in front $\mathcal{I}\left(\boldsymbol{a}, N\boldsymbol{b}\right)$ in the determinantal formula of Eq. (3.22). Next, the term $\det\left[\left(\mathrm{e}^{a_i}\right)^{j-1}\right]$ is nothing else than then Vandermonde determinant of $\mathrm{e}^{\boldsymbol{a}}$ and since the distribution of $\boldsymbol{b}_0$ converges to the uniform distribution μ_{Uni} on $(0,1)$, we get the following final result, valid for any β and any well-behaved LSD μ_A of $\boldsymbol{a}$:

$$\mathcal{F}[\mu_A, \mu_{\mathrm{Uni}}] := \frac{1}{2}\Sigma[\mu_{\exp(A)}] - \frac{1}{2}\Sigma[\mu_A]\,, \tag{3.71}$$

with $\Sigma[\mu] = \int\int \log|\lambda - \lambda'|\mu(\mathrm{d}\lambda)\mu(\mathrm{d}\lambda')$. The limiting distribution of the DBM constrained at both ends by a uniform distribution has been recently investigated in Ref. [76].

- There is another simple case where one can compute explicitly the function $\mathcal{F}$ by noticing that if we *choose* the $\boldsymbol{a}$ such that the LSD μ_A is the one of a DBM evaluated at time $t = 1$, that is:

$$\mu_A = \mu_B \boxplus \mu_{\mathrm{sc}(1)}\,, \tag{3.72}$$

then the dynamics of the constrained DBM are equal to the ones of the unconstrained one, and we have used identities for the free energies [116]:

$$\mathcal{F}[\mu_B \boxplus \mu_{\mathrm{sc}(1)}, \mu_B] := -\frac{3}{4} - \Sigma[\mu_B \boxplus \mu_{\mathrm{sc}(1)}] + \frac{1}{2}m_2[\mu_B] + \frac{1}{2}m_2[\mu_B \boxplus \mu_{\mathrm{sc}(1)}]\,, \tag{3.73}$$

3.3.2 Large N limit of the full rank multiplicative spherical integral for $\beta = 2$

For the full-rank multiplicative spherical function, there is - as far as I know of - no known variational description for *any* $\beta > 0$. Nevertheless, for $\beta = 2$ one can compare the two determinantal formulas of Eq. (3.22) and Eq. (3.23) to express the multiplicative spherical function in terms of the additive one, namely we have:

$$\mathcal{I}_{\times}^{(\beta=2)}\left(\boldsymbol{a}, \boldsymbol{b} - \boldsymbol{s}_0\right) = \frac{\Delta(\log \boldsymbol{a})}{\Delta(\boldsymbol{a})}\mathcal{I}^{(\beta=2)}\left(\log \boldsymbol{a}, \boldsymbol{b}\right)\,, \tag{3.74}$$

from which one can deduce the corresponding variational principle. If we introduce the term:

$$S_{\times}^{\mathrm{BT}}\left[\mu_A\right] = m_2[\mu_A] - \Sigma[\mu_A] + \frac{1}{2}\Sigma[\mu_{\log(A)}]\,, \tag{3.75}$$

we have the following result.

Result 3.8 (*Extensive-rank for the multiplicative spherical integral at $\beta = 2$*)

In the large N limit where $\mu_{\mathbf{A}} \to \mu_A$ and $\mu_{\mathbf{B}} \to \mu_B$, the limiting behavior of the multiplicative spherical integral is given for $\beta = 2$ by:

$$\frac{1}{N^2} \log \mathcal{I}_{\times}^{(\beta=2)} \left(\boldsymbol{a}, N\boldsymbol{b} - \boldsymbol{s}_0\right) \to -\frac{3}{4} + S_{\times}^{\mathrm{BT}} \left[\mu_A\right] + S^{\mathrm{BT}} \left[\mu_B\right] - S^{\mathrm{Mat}} \left[\mu_{\log(A)}, \mu_B\right] \ , \tag{3.76}$$

where $S_{\times}^{\mathrm{BT}}$ is given by Eq. (3.75), S^{BT} is given by Eq. (3.65) and S^{Mat} is given by Eq. (3.66).

Unlike the previous case, this function is not symmetric by permutation of μ_A (or $\mu_{\log(A)}$) and μ_B. For the additive case, the DBM plays a fundamental role in the limit behavior of the spherical integral and it will be interesting to know how if one can build a similar interpretation for the multiplicative case with the DGBM of Eq. (2.98) (or a variant of it) in order to extend the result to any β.

3.3.3 Large N limit of rectangular spherical integral and related variational principle

This section deals with rectangular matrices.

The derivation of the hydrodynamical limit of the rectangular spherical integral in the full rank case where the distribution of the singular values $\mu_{\mathbf{A}} = \sum_{i=1}^{N} \delta(. - a_i)/N$ converges to a LSVD μ_A (and similarly $\mu_{\mathbf{B}} \to \mu_B$) is very similar to the one of the additive case. The final result has been derived in Refs. [70, 83] and can be expressed in terms of the symmetrized LSVD $\widehat{\mu}_A = \left(\mu_A(.) + \mu_A(-.)\right)/2$ and $\widehat{\mu}_B = \left(\mu_B(.) + \mu_B(-.)\right)/2$. It is made of two contributions: a coupling term between $\widehat{\mu}_A$ and $\widehat{\mu}_B$ and a term which depends only on $\widehat{\mu}_A$ and $\widehat{\mu}_B$ individually.

- The decoupled term or 'boundary term' is given for each $\widehat{\mu} = \widehat{\mu}_A, \widehat{\mu_B}$ by

$$S_q^{\mathrm{BT}}[\widehat{\mu}] = \frac{1}{2} m_2[\widehat{\mu}] - \frac{q^{-1} - 1}{2} \mathbb{E}_{X \sim \widehat{\mu}}[\log X] - \frac{1}{2} \Sigma[\widehat{\mu}] \, , \tag{3.77}$$

- The coupling term is given by the following variational principle:

$$S_q^{\mathrm{Mat}} \left[\widehat{\mu}_A, \widehat{\mu}_B\right] = \frac{1}{2} \int_0^1 \int \left(\frac{\pi^2}{3} \widehat{\rho}^*(x,t)^3 + \widehat{\rho}^*(x,t) v^*(x,t)^2 + \frac{\left(1/q - 1\right)^2}{4} \frac{\widehat{\rho}^*(x,t)}{x^2} \right) \mathrm{d}t \mathrm{d}x \tag{3.78}$$

  where $\widehat{\rho}^*$ and v^* are solutions of:

$$\begin{cases} \partial_t \widehat{\rho}^* + \partial_x \left(\widehat{\rho}^* v^*\right) = 0 \\ \partial_t v^* + \frac{1}{2} \partial_x v^{*2} - \frac{\pi^2}{2} \partial_x (\widehat{\rho}^*)^2 - \frac{(q^{-1}-1)^2}{4x^3} = 0 \end{cases} \tag{3.79}$$

  with $\widehat{\rho}^*$ being symmetric and with fixed values at the boundaries:

$$\begin{cases} \widehat{\rho}^*(x,0) = \widehat{\mu}_A(x) \\ \widehat{\rho}^*(x,1) = \widehat{\mu}_B(x) \end{cases} \tag{3.80}$$

In summary we have:

Result 3.9 (*Limit of the full-rank rectangular spherical integral [70, 83]*)

In the limit where $N, M \to \infty$ with $N/M \to q \in (0,1)$ and the symmetrized LSVD are given by $\widehat{\mu}_{\mathbf{A}} \to \widehat{\mu}_A$ and $\widehat{\mu}_{\mathbf{B}} \to \widehat{\mu}_B$, the asymptotic behavior of the rectangular spherical integral is given by:

$$\frac{1}{N^2\beta} \log \mathcal{I}_q^{(\beta)}(\boldsymbol{a}, N\beta \boldsymbol{b}) \to \mathcal{F}_q[\widehat{\mu}_A, \widehat{\mu}_B] := \mathrm{cst} + S_q^{\mathrm{BT}}[\widehat{\mu}_A] + S_q^{\mathrm{BT}}[\widehat{\mu}_B] - S_q^{\mathrm{Mat}}[\widehat{\mu}_A, \widehat{\mu}_B] \tag{3.81}$$

where cst is constant depending on q, S_q^{BT} is given by Eq. (3.77) and S_q^{Mat} is given by Eq. (3.78).

The large N behavior of the rectangular spherical integral has appeared recently in the study of the problem of denoising a matrix corrupted by a Gaussian matrix. In particular in the case where $\mu_B = \mu_A \boxplus_q \mu_{\mathrm{MP}(q)}$, one can also obtain an explicit formula for the spherical integral, see Ref. [177].

Let's mention that when $q = 1$, the expression of both the boundary term of Eq. (3.77) and the coupling term of Eq. (3.78) simplify, and we also have $\mathrm{cst} \to -3/4$. In other words, we have

$$\mathcal{F}_q[\widehat{\mu}_A, \widehat{\mu}_B] \underset{q=1}{=} \mathcal{F}[\widehat{\mu}_A, \widehat{\mu}_B], \tag{3.82}$$

where $\mathcal{F}$ is the asymptotic behavior of the additive spherical integral given by Eq. (3.69), this is once again the manifestation of the close relation between square (but not self-adjoint) matrices and self-adjoint matrices we have encountered in the two previous chapters. Note that the normalizations of the additive and rectangular spherical integrals are slightly different.

3.4 Rank-one limit of spherical integrals and their relation to free probability transforms

3.4.1 Introduction

In this section we consider another limit of spherical integrals where the conjugate vector $\boldsymbol{t}$ is of 'rank-one' which by abuse of notation means that it has only one non-zero component:

$$\boldsymbol{t} = (\theta, 0, \dots, 0). \tag{3.83}$$

In order to ease notation, we denote

- the additive 'rank-one' spherical integral by:

$$\mathcal{I}_{\boldsymbol{a}}^{(\beta)}(\theta) := \mathcal{I}^{(\beta)}(\boldsymbol{a}, (\theta, 0 \dots, 0)), \tag{3.84}$$

- the multiplicative 'rank-one' spherical integral by:

$$\mathcal{I}_{\times,\boldsymbol{a}}^{(\beta)}(\theta) := \mathcal{I}_{\times}^{(\beta)}(\boldsymbol{a}, (\theta, 0 \dots, 0)), \tag{3.85}$$

- and the rectangular 'rank-one' spherical integral by:

$$\mathcal{I}_{q,\boldsymbol{a}}^{(\beta)}(\theta) := \mathcal{I}_q^{(\beta)}(\boldsymbol{a}, (\theta, 0 \dots, 0)). \tag{3.86}$$

Note that while the position of the non-zero entries does not matter in the additive and rectangular case, it does for the multiplicative spherical integral which is not invariant by permutation of the entries of $\boldsymbol{t}$. The non-zero entry is set to be the first component of the vector in this case. In all cases, we will assume $\theta > 0$.

We will be interested in the setting where the empirical distribution of $\boldsymbol{a}$ converges to a LSD/LSVD:

$$\mu_{\mathbf{A}} = \frac{1}{N}\sum_{i=1}^{N} \delta(. - a_i)/N \to \mu_A \,, \tag{3.87}$$

but importantly the position of the top eigenvalue/singular-value a_1 **converges to a position** x **which is equal to or higher than the upper edge** a_+ of the support of μ_A:

$$a_1 \to x \geq \mathrm{a}_+ \,. \tag{3.88}$$

At large N, we will see that each spherical integral is closely related to the linearizing transform of the corresponding operation: for example, the asymptotic behavior of the 'rank-one' additive spherical integral is related to the R-transform linearizing the free additive convolution. This behavior is the analog of the definition of the finite R-transform and S-transform for the FFC of Sec. 2.7.2 and we will comment more on this relation in Chapter 6. In fact, the connection between the rank-one spherical integrals and the corresponding linearizing transform can be guessed heuristically by noting that if we take the logarithm of Eq. (3.4) we have:

$$\log \mathbb{E}_{\mathbf{V}} \left[I^{(\beta)}(\mathbf{A} + \mathbf{V}\mathbf{B}\mathbf{V}^*, \theta) \right] = \log I^{(\beta)}(\mathbf{A}, \theta) \, + \log I^{(\beta)}(\mathbf{B}, \theta) \,, \tag{3.89}$$

whereby abuse of notation we have denoted by $I^{(\beta)}(\mathbf{A}, \theta) \equiv I^{(\beta)}(\mathbf{A}, \mathbf{Diag}(\theta, 0, \dots, 0))$. As $N \to \infty$, the distribution of the eigenvalues of $\mathbf{A} + \mathbf{V}\mathbf{B}\mathbf{V}^*$ converges to the free convolution of μ_A and μ_B and if we expect a *self-averaging property* such that we can remove the expectation in the LHS of Eq. (3.89), we retrieve the linearizing property of the R-transform. However, this result will only be true for a small value of θ: beyond a certain threshold depending on the asymptotic position of the top eigenvalue x and μ_A, the spherical function will saturate and the self-averaging property does not hold. In the language of statistical physics, this change of behavior for the spherical function is reminiscent of a *phase transition*, and in order to build intuition on this transition, I will first make a detour to the case $\beta = 1$ for which the spherical integrals are exactly the *partition functions* of well-known models of disorder systems.

3.4.2 Rank-one spherical integrals and $p = 2$ soft spin models for $\beta = 1$

The additive case -

For $\beta = 1$, the rank-one vector $\boldsymbol{t} = (\theta, 0, \dots, 0)$ can be seen as the eigenvalues of the rank-one matrix $\mathbf{T} = \theta \boldsymbol{e}_1 \boldsymbol{e}_1^{\mathsf{T}}$ where $\boldsymbol{e}_1 = (1, 0, \dots, 0)$ has been set without loss of generality. If we denote by $\mathbf{A} = \mathbf{Diag}(\boldsymbol{a})$, the HCIZ integral writes in this case:

$$\mathcal{I}_{\boldsymbol{a}}^{(\beta=1)}\left(\frac{N}{2}\theta\right) := \int_{\mathsf{O}(N)} \exp\left[\frac{N\theta}{2} \mathrm{Tr}\, \left(\mathbf{O}^{\mathsf{T}}\boldsymbol{e}_1\right)^{\mathsf{T}} \mathbf{A} \left(\mathbf{O}^{\mathsf{T}}\boldsymbol{e}_1\right)\right] \mu_{\mathrm{Haar}}\left(\mathrm{d}\mathbf{O}\right) \,. \tag{3.90}$$

If we perform the change of variable from $\mathbf{O}$ to $\boldsymbol{\sigma} := \mathbf{O}^\mathsf{T} e_1$, we can write this integral as:

$$\mathcal{Z}_{\mathbf{A}}(\theta) := \mathcal{I}_{\boldsymbol{a}}^{(\beta=1)}\left(\frac{N}{2}\theta\right) = \left\langle \mathrm{e}^{\frac{N}{2}\theta\,\mathcal{H}^{\mathrm{SSK}}(\boldsymbol{\sigma})} \right\rangle, \tag{3.91}$$

where to follow the standard notation in statistical physics, we have denoted by $\langle . \rangle \equiv \int_{\mathbb{S}^{N-1}} . \, \mathrm{d}\mu_{\mathrm{Unif}}(\boldsymbol{\sigma})$ the uniform average over the spins living in the sphere of radius one and the *Hamiltonian* $\mathcal{H}^{\mathrm{SSK}}$ is given by:

$$\mathcal{H}^{\mathrm{SSK}}(\boldsymbol{\sigma}) := \boldsymbol{\sigma}^\mathsf{T}\mathbf{A}\boldsymbol{\sigma} = \sum_i^N a_i \sigma_i^2 \, . \tag{3.92}$$

Note that we have chosen $\mathbf{A}$ to be a diagonal matrix here but by rotational invariance, we could have picked any matrix $\mathbf{A}$ with eigenvalues given by $\boldsymbol{a}$. This model is known (see Refs. [104, 79, 45, 13]) as the $(p = 2)$ *Spherical Sherrington-Kirkpatrick* (SSK in short) model[2]. The matrix $\mathbf{A}$ is the *disordered pairwise interaction* and the parameter θ is the *inverse temperature* of the model. The SSK model has been studied in detail in the literature, and in particular, the case where $\mathbf{A}$ is a GOE matrix has received a lot of attention. Qualitatively, for high temperatures, it is known that the system is in a paramagnetic phase, and every eigenvalue of the matrix $\mathbf{A}$ contributes roughly equally to the partition function while for low temperatures, the situation is drastically different and the system is in a ferromagnetic phase where the partition function is dominated by rare configurations which are aligned with the top eigenvector of $\mathbf{A}$.

The multiplicative case -

Similarly, with $\boldsymbol{t} = (\theta, 0, \ldots, 0)$, the multiplicative spherical integral reduces to:

$$\mathcal{I}_{\times,\boldsymbol{a}}^{(\beta=1)}\left(\frac{N}{2}\theta\right) := \int_{\mathsf{O}(N)} \left(\det\left[\left(\mathbf{O}\mathbf{A}\mathbf{O}^\mathsf{T}\right)_{(1)}\right]\right)^{\frac{N\theta}{2}} \mu_{\mathrm{Haar}}\left(\mathrm{d}\mathbf{O}\right) \, . \tag{3.93}$$

where $\mathbf{A}_{(1)}$ is the top left principal corner of the matrix $\mathbf{A}$, that is the projection along e_1: $\mathbf{A}_{(1)} = e_1^\mathsf{T}\mathbf{A}e_1$, thus by the same change of variable as before, we can also write the multiplicative spherical integral as an integral over the unit sphere:

$$\mathcal{I}_{\times,\boldsymbol{a}}^{(\beta=1)}\left(\frac{N}{2}\theta\right) := \int_{\mathbb{S}^{N-1}} \left(\boldsymbol{\sigma}^\mathsf{T}\mathbf{A}\boldsymbol{\sigma}\right)^{\frac{N\theta}{2}} \mu_{\mathrm{Unif}}\left(\mathrm{d}\boldsymbol{\sigma}\right) \, . \tag{3.94}$$

which we rewrite as:

$$\mathcal{Z}_{\mathbf{A}}(\theta) := \mathcal{I}_{\times,\boldsymbol{a}}^{(\beta=1)}\left(\frac{N}{2}\theta\right) = \left\langle \mathrm{e}^{\frac{N}{2}\theta\mathcal{H}^{\mathrm{LSSK}}(\boldsymbol{\sigma})} \right\rangle, \tag{3.95}$$

with the Hamiltonian:

$$\mathcal{H}^{\mathrm{LSSK}}(\boldsymbol{\sigma}) := \log\left(\boldsymbol{\sigma}^\mathsf{T}\mathbf{A}\boldsymbol{\sigma}\right) = \log\left(\sum_{i=1}^N a_i\sigma_i^2\right) \, . \tag{3.96}$$

Since $\mathbf{A}$ is the symmetric product of definite positive matrices, its eigenvalues are positive, so this Hamiltonian is well-defined. Due to the logarithmic term, we denote this model as the *Logarithmic Spherical Sherrington-Kirkpatrick* (LSSK for short) model. As for the original SSK model, one should expect to have similar behavior, with a paramagnetic phase at high temperature and a ferromagnetic phase at low temperature.

[2]The SSK is usually introduced with a different convention by absorbing the N in the spin variable: $\tilde{\boldsymbol{\sigma}} := \sqrt{N}\boldsymbol{\sigma}$ so that the spins lives on a sphere with radius $\sqrt{N}$, which does not change the spherical integral.

The rectangular case -

This rectangular spherical can again be understood as the partition function of a spherical model with inverse temperature θ since by a similar argument we have:

$$\mathcal{I}_{q,\boldsymbol{a}}^{(\beta=1)}(\sqrt{NM}\theta) = \int_{\mathbb{S}^{N-1}} \mu_{\mathrm{Unif}}(\mathrm{d}\boldsymbol{\sigma}_1) \int_{\mathbb{S}^{M-1}} \mu_{\mathrm{Unif}}(\mathrm{d}\boldsymbol{\sigma}_2) \exp\left[\sqrt{NM}\theta \boldsymbol{\sigma}_1^{\mathsf{T}} \mathbf{A} \boldsymbol{\sigma}_2\right] . \tag{3.97}$$

If we introduce the vector

$$\boldsymbol{\sigma} := \frac{1}{\sqrt{2}} [\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2] \in \mathbb{S}^{N+M-1} , \tag{3.98}$$

this spherical integral can be written as:

$$\mathcal{Z}_{\mathbf{A}}(\theta) := \mathcal{I}_{q,\boldsymbol{a}}^{(\beta=1)}(\sqrt{NM}\theta) = \left\langle \mathrm{e}^{\sqrt{NM}\theta \mathcal{H}^{\mathrm{BSSK}}(\boldsymbol{\sigma})} \right\rangle , \tag{3.99}$$

with the Hamiltonian

$$\mathcal{H}^{\mathrm{BSSK}}(\theta) := \boldsymbol{\sigma}^{\mathsf{T}} \begin{pmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^{\mathsf{T}} & \mathbf{0} \end{pmatrix} \boldsymbol{\sigma} , \tag{3.100}$$

and by bi-invariance, this can be also written as:

$$\mathcal{H}^{\mathrm{BSSK}}(\theta) = \sum_{i=1}^{N} a_i \sigma_i \sigma_{N+i}. \tag{3.101}$$

This model is known [9, 16, 14] as the *Bipartite Spherical Sherrington-Kirkpatrick* (BSSK in short) spin model, due to the graph structure of the interaction matrix: each coordinate of one family vector interacts only with members of the other family and similarly to the SSK, this model is known to have a paramagnetic phase at high temperature and a ferromagnetic phase at low temperature.

Quenched and annealed free energies -

In the language of statistical physics, the fundamental quantity to characterize a disordered system is the *quenched free energy* defined by

$$J_A(x, \theta) := \lim_{N\to\infty} \frac{\alpha}{N} \log \mathcal{Z}_{\mathbf{A}}(\theta) , \tag{3.102}$$

where α is a constant of proportionality, usually set to one but will depend here on β. We have put explicitly the dependency in the position of the asymptotic position x of the top eigenvalue. Since $\mathcal{Z}_{\mathbf{A}}(\theta)$ is nothing else than a spherical integral, this is our main quantity of interest in this section. By abuse of notation, for $\beta > 0$ we will also call the (re-scaled) logarithm of the spherical integrals, the quenched free energy.

For disordered systems where the interaction matrix $\mathbf{A}$ is random, there is another related quantity known as the *annealed free energy*, which is defined by:

$$F_A(\theta) := \lim_{N\to\infty} \frac{\alpha}{N} \log \mathbb{E}\mathcal{Z}_{\mathbf{A}}(\theta) , \tag{3.103}$$

in other words, it differs from the quenched free energy by taking an average over the disorder $\mathbf{A}$ *before* taking the logarithm. This quantity will also play an important role in the large deviation of the top eigenvalue of Chapter 5. In disordered systems, one has usually equality between the two quantities $F_A(\theta) = J_A(\theta)$ for high temperatures, but this equality breaks for low temperatures and $J_A(\theta)$ is dominated by 'typical values' of $\mathcal{Z}_{\mathbf{A}}(\theta)$ far from its average.

Eventually, let us mention that we will actually be interested more in the partial derivatives of these two free energies rather than the free energies themselves.

3.4.3 General method to get the result

We briefly describe the steps to compute the asymptotic behavior for $\beta = 1$ of the spherical integrals.

- The first step of the computation is to derive a complex integral representation for the partition function $\mathcal{Z}_{\mathbf{A}}(\theta)$ of each model:

$$\mathcal{Z}_{\mathbf{A}}(\theta) = \frac{1}{2\pi \mathrm{i}} \int_{\mathscr{C}} \mathrm{e}^{\frac{N\beta}{2} G_N(\boldsymbol{a}, z, \theta)} \mathrm{d}z \,, \tag{3.104}$$

  where $\mathscr{C}$ is *Bromwich contour*, that is a vertical line in the complex plane making the integral convergent. Such formulae are obtained by removing the constraint over the sphere by introducing a Lagrange multiplier z and then working on the unconstrained integral to write it as a multivariate Gaussian integral. Let's mention that for the rectangular case, we will have a double complex integral representation instead. The case $\beta \notin \{1, 2, 4\}$ requires slightly more work and can be obtained thanks to identities for symmetric polynomials and corresponds to an analytical continuation in the parameter β of the expression obtained for $\beta = 1, 2, 4$.

- The second step is then to do a saddle-point computation (or *Laplace's method*) over the complex variable z to get the asymptotic behavior.

Because the integration is done over a vertical line in the complex plane the integrals are highly oscillatory and from a purely rigorous point of view, one needs to ensure that the saddle-point approximation is valid. For each of the three spherical integrals, the control of the saddle-point approximation has been verified in the mathematical literature, and I will refer to the corresponding references in each case.

3.4.4 Additive spherical integral and the R-transform

Complex integral representation -

As explained in the previous section, the first step is to remove the constraint over the sphere. For a $f : \mathbb{K}_\beta^N \to \mathbb{R}$, we have:

$$\int_{\mathbb{S}_\beta^{N-1}} f(\boldsymbol{\sigma}) \mu_{\mathrm{Unif}}(\mathrm{d}\boldsymbol{\sigma}) \propto \int_{\mathbb{K}_\beta^N} \delta(\boldsymbol{x}^* \boldsymbol{x} - 1) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \,, \tag{3.105}$$

where the constant of proportionality will be set at the end. If one uses the Inverse Laplace representation of the Dirac 'function':

$$\delta(x) = \frac{1}{2\pi \mathrm{i}} \int_{\mathscr{C}} \mathrm{e}^{zx} \mathrm{d}z = \mathcal{L}^{-1}\left[1\right](x)\,, \tag{3.106}$$

Eq. (3.105) can be written as:

$$\int_{\mathbb{S}_\beta^{N-1}} f(\boldsymbol{\sigma}) \mu_{\mathrm{Unif}}(\mathrm{d}\boldsymbol{\sigma}) = \frac{\Gamma\left(\frac{N\beta}{2}\right)}{\pi^{N\beta/2}} \cdot \frac{1}{2\pi \mathrm{i}} \int_{\mathscr{C}} \mathrm{e}^{z} \left(\int_{\mathbb{K}_\beta^N} f(\boldsymbol{x}) \mathrm{e}^{-z\boldsymbol{x}^*\boldsymbol{x}} \mathrm{d}\boldsymbol{x} \right) \mathrm{d}z\,, \tag{3.107}$$

where the constant of proportionality has been determined by setting $f(.) = 1$ and using $\int_{\mathbb{S}_\beta^{N-1}} \mu_{\mathrm{Unif}}(\mathrm{d}\boldsymbol{\sigma}) = 1$ for the LHS and Gaussian integration and the standard complex integral representation of the inverse of the Gamma function for the RHS.

If we specify this expression to the case corresponding to the additive spherical integral, that is set $f(\boldsymbol{\sigma}) := \exp\left[\frac{N\beta\theta}{2} \boldsymbol{\sigma}^* \mathbf{Diag}(\boldsymbol{a}) \boldsymbol{\sigma}\right]$, we obtain for $\beta \in \{1, 2, 4\}$:

$$\mathcal{I}_{\boldsymbol{a}}^{(\beta)}\left(\frac{N\beta}{2}\theta\right) = \frac{\Gamma\left(\frac{N\beta}{2}\right)}{\pi^{\frac{N\beta}{2}}} \frac{1}{2\pi \mathrm{i}} \int_{\mathscr{C}} \mathrm{e}^{z} \left(\int_{\mathbb{K}_\beta^N} \mathrm{e}^{-\boldsymbol{x}^*(z\mathbf{I} - \frac{N\beta\theta}{2}\mathbf{Diag}(\boldsymbol{a}))\boldsymbol{x}} \mathrm{d}\boldsymbol{x} \right) \mathrm{d}z\,. \tag{3.108}$$

Next, let's do the change of variable $z \to (N\beta\theta/2)\, z$, $\mathrm{d}z \to (N\beta\theta/2)\, \mathrm{d}z$, such that we can write this integral:

$$\mathcal{I}_{\boldsymbol{a}}^{(\beta)}\left(\frac{N\beta}{2}\theta\right) = \frac{\Gamma\left(\frac{N\beta}{2}\right)\left(\frac{N\beta\theta}{2}\right)}{\pi^{\frac{N\beta}{2}}} \cdot \frac{1}{2\pi \mathrm{i}} \int_{\mathscr{C}} \mathrm{e}^{\frac{N\beta\theta}{2}z} \left(\int_{\mathbb{K}_\beta^N} \mathrm{e}^{-\frac{N\beta\theta}{2}\boldsymbol{x}^*(z\mathbf{I} - \mathbf{Diag}(\boldsymbol{a}))\boldsymbol{x}} \mathrm{d}\boldsymbol{x} \right) \mathrm{d}z \tag{3.109}$$

The integral inside the bracket is Gaussian integral which is convergent provided $\mathfrak{Re} z > a_1$, which corresponds to the Bromwich contour $\mathscr{C}$ going to the right of all the eigenvalues from which we deduce:

$$\mathcal{I}_{\boldsymbol{a}}^{(\beta)}\left(\frac{N\beta}{2}\theta\right) = \frac{\Gamma\left(\frac{N\beta}{2}\right)}{\left(\frac{N\beta\theta}{2}\right)^{\frac{N\beta}{2}-1}} \frac{1}{2\pi \mathrm{i}} \int_{\mathscr{C}} \mathrm{d}z\, \mathrm{e}^{\frac{N\beta\theta}{2}z} \prod_{i=1}^{N} (z - a_i)^{-\frac{\beta}{2}}\,. \tag{3.110}$$

Let's mention that while Eq. (3.110) has been derived for $\beta \in \{1, 2, 4\}$, using identities for the Jack polynomials, one can show that this expression is valid for any $\beta > 0$ and in particular we have for all β:

$$\mathcal{I}_{\boldsymbol{a}}^{(\beta)}\left(\theta\right) = \frac{\Gamma\left(\frac{N\beta}{2}\right)}{(\theta)^{\frac{N\beta}{2}-1}} \mathcal{L}_z^{-1}\left[\prod_{i=1}^{N} (z - a_i)^{-\frac{\beta}{2}}\right](\theta)\,. \tag{3.111}$$

where $\mathcal{L}_z^{-1}[\cdot]$ is the inverse Laplace transform.

Next, since we are interested in the large N limit, by the generalized Stirling formula:

$$\log\Gamma(\zeta) \underset{|\zeta|\to\infty}{\sim} \zeta\log\zeta - \zeta - \frac{1}{2}\log\zeta + \frac{1}{2}\log 2\pi + \mathcal{O}\left(\frac{1}{\zeta}\right)\,, \tag{3.112}$$

we can write the additive spherical integral as:

$$\mathcal{I}_{\boldsymbol{a}}^{(\beta)}\left(\frac{N\beta}{2}\theta\right) = \frac{1}{2\pi \mathrm{i}}\int_{\mathscr{C}} \mathrm{e}^{\frac{N\beta}{2}G_N(\boldsymbol{a},z,\theta)}\mathrm{d}z\,, \tag{3.113}$$

where $\mathscr{C}$ is a vertical line in the complex plane that goes to the right of all the eigenvalues and

$$G_N(\boldsymbol{a},z,\theta) := z\theta - 1 - \log\theta - \frac{1}{N}\sum_{k=1}^{N}\log(z-a_i) + \mathcal{O}\left(\frac{1}{N}\right)\,. \tag{3.114}$$

Saddle-point computation -

We are now ready to perform the saddle-point computation. At large N the complex integral is dominated by the point $z^* \equiv z^*(\theta)$ such that:

$$\partial_z G_N(\boldsymbol{a},z,\theta)|_{z=z^*} = 0\,, \tag{3.115}$$

that is z^* is the solution of:

$$g_{\boldsymbol{a}}(z^*) := \frac{1}{N}\sum_{i=1}^{N}\frac{1}{z^*-a_i} = \theta\,. \tag{3.116}$$

Now to invert this equation, one needs to be a bit cautious. First, let's recall that we are interested in the large N regime where $\mu_{\boldsymbol{a}} \to \mu_A$ and $a_1 \to x$.

- For small values of θ, since $g_{\boldsymbol{a}}(z)$ is a positive decreasing function for values of $z > a_1$ going to zero at infinity, one can directly invert this equation and this given asymptotically by the inverse Stieltjes transform of the LSD μ_A:

  $$z^*(\theta) = g_A^{\langle -1\rangle}(\theta)\,. \tag{3.117}$$

  However, this result is only valid for a small value of θ because if $\theta > g_A(x)$, one cannot naively invert the equation.

- For large values of θ ($\theta > g_A(x)$), since the Bromwich contour must pass at the right of all eigenvalues, z is constrained to be higher than a_1 and we have a saturation phenomenon, asymptotically the sum of Eq. (3.116) is dominated by the term $1/(z^*-a_1)$ and since $a_1 \to x$, we have:

  $$z^* = x\,. \tag{3.118}$$

As a consequence, we have

$$\frac{2}{N\beta}\frac{\mathrm{d}}{\mathrm{d}\theta}\log\left[\mathcal{I}_{\boldsymbol{a}}^{(\beta)}\left(\frac{N\beta}{2}\theta\right)\right] \to \frac{\mathrm{d}}{\mathrm{d}\theta}G_A(z^*(\theta),\theta)\,, \tag{3.119}$$

where

$$G_A(z,\theta) := \lim_{N\to\infty} G_N(\boldsymbol{a},z,\theta) = z\theta - 1 - \log\theta - \int \log(z-\lambda)\mu_A(\lambda)\,. \tag{3.120}$$

If we use the chain rule, we get:

$$\frac{2}{N\beta}\frac{\mathrm{d}}{\mathrm{d}\theta}\log\left[\mathcal{I}_{\boldsymbol{a}}^{(\beta)}\left(\frac{N\beta}{2}\theta\right)\right] \to \partial_\theta G_A(z^*(\theta),\theta) + \frac{\mathrm{d}z^*(\theta)}{\mathrm{d}\theta}\cdot\underbrace{\partial_z G_A(z,\theta)|_{z=z^*(\theta)}}_{=0}, \tag{3.121}$$

where the second term is null since z^* is the solution of the saddle-point Eq. (3.115) and as a consequence, we have the following simple formula

$$\frac{2}{N\beta}\frac{\mathrm{d}}{\mathrm{d}\theta}\log\left[\mathcal{I}_{\boldsymbol{a}}^{(\beta)}\left(\frac{N\beta}{2}\theta\right)\right] \to z^*(\theta) - \frac{1}{\theta}\,. \tag{3.122}$$

All in all, using the expression of z^* given by Eq. (3.117) and Eq. (3.118), we can summarize the asymptotic behavior in the following result.

Result 3.10 (*rank-one additive spherical integral and R-transform [131, 84]*)

In the large N limit where $\mu_{\mathbf{A}} \to \mu_A$ and $a_1 \to x$, if we denote by:

$$J_A(x,\theta) := \lim_{N\to\infty}\frac{2}{N\beta}\log\left[\mathcal{I}_{\boldsymbol{a}}^{(\beta)}\left(\frac{N\beta}{2}\theta\right)\right], \tag{3.123}$$

we have that the partial derivatives of this free energy are given by:

$$\partial_\theta J_A(x,\theta) = \begin{cases} \mathcal{R}_A(\theta) & \text{for } \theta \le g_A(x)\,, \\ \\ x - \frac{1}{\theta} & \text{for } \theta \ge g_A(x)\,, \end{cases} \tag{3.124}$$

and

$$\partial_x J_A(x,\theta) = \begin{cases} 0 & \text{for } \theta \le g_A(x)\,, \\ \\ \theta - g_A(x) & \text{for } \theta \ge g_A(x)\,. \end{cases} \tag{3.125}$$

where $\mathcal{R}_A$ is the R-transform of μ_A, see Eq. (2.139) and g_A is the Stieltjes transform of μ_A.

The result first appeared in the literature on spherical spin glass in Ref. [131] and has then been shown to be rigorous in Ref. [84]. The result can in fact be generalized to a 'rank-k' spherical integral, but understanding the crossover with the full-rank regime of the previous section is an open problem. Note that we have treated the asymptotic position x of the top eigenvalue as a variable of the quenched free energy as this point of view will turn out to be very useful in Chapter 5.

3.4.5 Multiplicative spherical integral and the S-transform

The results of this section are based on the preprint [142] -

We now turn to the 'rank-one' multiplicative spherical integral given by Eq. (3.85). We first consider the case $\beta \in \{1,2,4\}$ and then briefly discuss the general setting $\beta > 0$. For $\beta \in \{1,2,4\}$, we can remove the spherical constraint by introducing a Lagrange multiplier and

writing the spherical integral as:

$$\mathcal{I}^{(\beta)}_{\times,\boldsymbol{a}}\left(\frac{N\beta}{2}\theta\right)=\frac{\Gamma\left(\frac{N\beta}{2}\right)}{\pi^{\frac{N\beta}{2}}}\cdot\frac{1}{2\pi\mathrm{i}}\int_{\mathscr{C}_1}\mathrm{e}^{z}\int_{\mathbb{K}_\beta^N}\mathrm{e}^{-z\boldsymbol{x}^*\boldsymbol{x}}\left(\boldsymbol{x}^*\mathbf{Diag}(\boldsymbol{a})\boldsymbol{x}\right)^{\frac{N\beta}{2}\theta}\mathrm{d}\boldsymbol{x}\,\mathrm{d}z\,,\tag{3.126}$$

where $\mathscr{C}_1$ is a vertical line in the complex plane making this integral convergent. Now in order to tackle this integral, we introduce an auxiliary variable s thanks to the inverse Laplace representation of the power function since for $a>0$, we have:

$$a^t=\frac{\Gamma(t+1)}{2\pi\mathrm{i}}\int_{\mathscr{C}_2}s^{-t-1}e^{sa}\,\mathrm{d}s\,,\tag{3.127}$$

where $\mathscr{C}_2$ is another vertical line in the complex plane. If we use this integral representation for $a=\boldsymbol{x}^*\mathbf{Diag}(\boldsymbol{a})\boldsymbol{x}$ and $t=(N\beta/2)\theta$ in Eq. (3.126) we have:

$$\begin{aligned}\mathcal{I}^{(\beta)}_{\times,\boldsymbol{a}}\left(\frac{N\beta}{2}\theta\right)=&\frac{\Gamma\left(\frac{N\beta\theta}{2}+1\right)\Gamma\left(\frac{N\beta}{2}\right)}{\pi^{\frac{N\beta}{2}}}\times\\&\frac{1}{2\pi\mathrm{i}}\int_{\mathscr{C}_1}\mathrm{e}^{z}\int_{\mathbb{K}_\beta^N}e^{-z\boldsymbol{x}^*\boldsymbol{x}}\frac{1}{2\pi\mathrm{i}}\int_{\mathscr{C}_2}s^{-\left(\frac{N\beta\theta}{2}\right)-1}e^{s(\boldsymbol{x}^*\mathbf{Diag}(\boldsymbol{a})\boldsymbol{x})}\,\mathrm{d}z\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}s\,.\end{aligned}\tag{3.128}$$

Next, let's do the change of variable $s=e^{-s}$ and deform the Bromwich contour accordingly such that we have:

$$\begin{aligned}\mathcal{I}^{(\beta)}_{\times,\boldsymbol{a}}\left(\frac{N\beta}{2}\theta\right)=&\frac{\Gamma\left(\frac{N\beta\theta}{2}+1\right)\Gamma\left(\frac{N\beta}{2}\right)}{\pi^{\frac{N\beta}{2}}}\times\\&\left(\frac{1}{2\pi\mathrm{i}}\right)^2\int_{\mathscr{C}_1}\int_{\mathscr{C}_2}\mathrm{e}^{s\left(\frac{N\beta\theta}{2}\right)+z}\left(\int_{\mathbb{K}_\beta^N}\mathrm{d}\boldsymbol{x}\,\mathrm{e}^{-\boldsymbol{x}^*(z-\mathrm{e}^{-s}\mathbf{Diag}(\boldsymbol{a}))\boldsymbol{x}}\right)\mathrm{d}z\,\mathrm{d}s\,,\end{aligned}\tag{3.129}$$

The integral inside the bracket is again a Gaussian integral which gives:

$$\mathcal{I}^{(\beta)}_{\times,\boldsymbol{a}}\left(\frac{N\beta}{2}\theta\right)=\Gamma\left(\frac{N\beta}{2}\right)\Gamma\left(\frac{N\beta\theta}{2}+1\right)\left(\frac{1}{2\pi\mathrm{i}}\right)^2\int_{\mathscr{C}_1}\int_{\mathscr{C}_2}\mathrm{e}^{s\left(\frac{N\beta\theta}{2}\right)+z}\prod_{i=1}^{N}\left(z-\mathrm{e}^{-s}a_i\right)^{-\frac{\beta}{2}}\,\mathrm{d}z\,\mathrm{d}s\,,\tag{3.130}$$

If we do another change of variable $(s\to s-\log z,\mathrm{d}s\to\mathrm{d}s)$ which just shifts the contour of the integral, and then we factorize by $z^{-\frac{N\beta}{2}}$ in the product, we have:

$$\begin{aligned}\mathcal{I}^{(\beta)}_{\times,\boldsymbol{a}}\left(\frac{N\beta}{2}\theta\right)=&\Gamma\left(\frac{N\beta}{2}\right)\Gamma\left(\frac{N\beta\theta}{2}+1\right)\times\\&\left(\frac{1}{2\pi\mathrm{i}}\right)^2\left(\int_{\mathscr{C}_1}\mathrm{d}z\,\mathrm{e}^{z}z^{-\frac{N\beta}{2}-\frac{N\beta\theta}{2}}\right)\int_{\mathscr{C}_2}\mathrm{e}^{s\left(\frac{N\beta\theta}{2}\right)}\prod_{i=1}^{N}\left(1-\mathrm{e}^{-s}a_i\right)^{-\frac{\beta}{2}}\,\mathrm{d}s\,.\end{aligned}\tag{3.131}$$

The integral in the bracket is nothing else than the inverse of the gamma function, see Eq. (3.127), such that we can write the multiplicative spherical as:

$$\mathcal{I}^{(\beta)}_{\times,\boldsymbol{a}}\left(\frac{N\beta}{2}\theta\right)=\frac{\Gamma\left(\frac{N\beta}{2}\right)\Gamma\left(\frac{N\beta\theta}{2}+1\right)}{\Gamma\left(\frac{N\beta}{2}(\theta+1)\right)}\frac{1}{2\pi\mathrm{i}}\int_{\mathscr{C}_2}\mathrm{d}s\,\mathrm{e}^{s\left(\frac{N\beta\theta}{2}\right)}\prod_{i=1}^{N}\left(1-\mathrm{e}^{-s}a_i\right)^{-\frac{\beta}{2}}\,,\tag{3.132}$$

and more generally we have:

$$\mathcal{I}_{\times,\boldsymbol{a}}^{(\beta)}(\theta) = \frac{\Gamma\left(\frac{N\beta}{2}\right)\Gamma(\theta+1)}{\Gamma\left(\frac{N\beta}{2}+\theta\right)}\mathcal{L}_s^{-1}\left[\prod_{i=1}^{N}\left(1-\mathrm{e}^{-s}a_i\right)^{-\frac{\beta}{2}}\right](\theta)\,. \tag{3.133}$$

By Stirling approximation, we have:

$$\frac{2}{N\beta}\log\frac{\Gamma\left(\frac{N\beta}{2}\right)\Gamma\left(\frac{N\beta\theta}{2}+1\right)}{\Gamma\left(\frac{N\beta}{2}(\theta+1)\right)} \underset{N\to\infty}{\sim} \theta\log\theta-(1+\theta)\log(1+\theta)+\mathcal{O}\left(\frac{1}{N}\right), \tag{3.134}$$

from which we deduce the following complex integral representation:

$$\mathcal{I}_{\times,\boldsymbol{a}}^{(\beta)}\left(\frac{N\beta}{2}\theta\right) = \frac{1}{2\pi\mathrm{i}}\int_{\mathscr{C}_2}\mathrm{e}^{\frac{N}{2}G_N^{(\times)}(\boldsymbol{a},z,\theta)}\,\mathrm{d}z\,, \tag{3.135}$$

with

$$G_N^{(\times)}(\boldsymbol{a},z,\theta) = z\theta-\frac{1}{N}\sum_{i=1}^{N}\log(1-a_i\mathrm{e}^{-z})+\theta\log\theta-(1+\theta)\log(1+\theta)+\mathcal{O}\left(\frac{1}{N}\right). \tag{3.136}$$

Saddle-point computation -

The integral of Eq. (3.135) is dominated by the saddle point $z^* \equiv z^*(\theta,x)$ solution of $\partial_z G_N(\boldsymbol{a},z,\theta)|_{z=z^*}=0$ that is:

$$t_{\boldsymbol{a}}\left(\mathrm{e}^{z^*}\right) := \frac{1}{N}\sum_{i=1}^{N}\frac{a_i}{\mathrm{e}^{z^*}-a_i} = \theta\,. \tag{3.137}$$

Again, we have two different behaviors for the solution of this fixed point equation depending on the value of θ.

- For θ small, namely $\theta < t_A(x)$ one can directly invert the T-transform, and the results are given asymptotically by:

$$z^*(\theta) = \log t_A^{(-1)}(\theta)\,. \tag{3.138}$$

- Conversely, for $\theta > t_A(x)$, e^{z^*} saturates at $a_1 \to x$ and we have:

$$z^* = \log x\,. \tag{3.139}$$

As a consequence, in the large N limit, the derivative with respect to θ of the logarithm of the spherical integral is given by:

$$\frac{2}{N\beta}\frac{\mathrm{d}}{\mathrm{d}\theta}\log\left[\mathcal{I}_{\times,\boldsymbol{a}}^{(\beta)}\left(\frac{N\beta}{2}\theta\right)\right] \to \partial_\theta G_A^{(\times)}(z^*(\theta),\theta)\,, \tag{3.140}$$

where the total derivative of the RHS is equal to the partial derivative with respect to θ, since z^* is a saddle-point and $G_A^{(\times)}(z^*(\theta),\theta)$ is given by:

$$G_A^{(\times)}(z,\theta) := \lim_{N\to\infty} G_N^{(\times)}(\boldsymbol{a},z,\theta) = z\theta - \int \log(1-\lambda \mathrm{e}^{-z})\mu_A(\mathrm{d}\lambda) + \theta\log\theta - (1+\theta)\log(1+\theta)\,, \tag{3.141}$$

such that we have:

$$\frac{2}{N\beta}\frac{\mathrm{d}}{\mathrm{d}\theta}\log\left[\mathcal{I}^{(\beta)}_{\times,\boldsymbol{a}}\left(\frac{N\beta}{2}\theta\right)\right] \to z^* + \log\frac{\theta}{\theta+1}\,. \tag{3.142}$$

Using the expression for z^* given by Eq. (3.138) and Eq. (3.139) and the definition of the S-transform we have the following result.

Result 3.11 (*rank-one multiplicative spherical integral and S-transform, [142]*)

In the limit where $\mu_{\mathbf{A}} := \sum_{i=1}^N \delta(.-a_i)/N \to \mu_A$ and the top eigenvalue $a_1 \to x$, if we define the free energy by:

$$J_A(x,\theta) := \lim_{N\to\infty}\frac{2}{N\beta}\log\left[\mathcal{I}^{(\beta)}_{\times,\boldsymbol{a}}\left(\frac{N\beta}{2}\theta\right)\right]\,, \tag{3.143}$$

the partial derivatives of this free energy are given by:

$$\partial_\theta J_A(x,\theta) = \begin{cases} \log\tilde{\mathcal{S}}_A(\theta) & \text{for } \theta \le t_A(x)\,, \\ \\ \log\left(\frac{x\theta}{\theta+1}\right) & \text{for } \theta \ge t_A(x)\,. \end{cases} \tag{3.144}$$

and by:

$$\partial_x J_A(x,\theta) = \begin{cases} 0 & \text{for } \theta \le t_A(x)\,, \\ \\ \frac{\theta+1}{x} - g_A(x) & \text{for } \theta \ge t_A(x)\,. \end{cases} \tag{3.145}$$

where $\tilde{\mathcal{S}}_A$ and t_A are respectively the S-transform and T-transform of the LSD μ_A given respectively by Eq. (2.152) and Eq. (1.29).

Let's mention that the control of the saddle-point approximation has been done in the mathematical literature in Ref. [93].

Asymptotic of symmetric polynomials -

For $\beta = 2$, the asymptotic behavior of the HCIZ integral can be translated as an asymptotic behavior over normalized Schur polynomial: if $\frac{1}{N}\sum_{i=1}^N \delta_{N^{-1}(\lambda_i+N-i)}$ converge toward a deterministic measure μ, then the corresponding normalized Schur polynomial with *index* $\boldsymbol{\lambda}$ and with all its arguments except one fixed, converges (up to an integration term) exponentially towards the integral of the R-transform see Ref. [84]. Since the multiplicative spherical function $\mathcal{I}^{(\beta)}_{\times,\boldsymbol{a}}(\theta)$ of this note is nothing else than the analytical extension of the normalized Jack polynomials, see Eq. (3.52), we have a similar interpretation, except that now it is the vector

in argument of the Jack polynomial which converges towards a deterministic measure while the index is the trivial partition $\boldsymbol{\lambda} = \left(\lfloor \frac{N\beta}{2}\theta \rfloor, 0, \ldots, 0\right) =: \lfloor \frac{N\beta}{2}\theta \rfloor$.

$$\lim_{N\to\infty} \frac{2}{N\beta} \log \frac{\mathrm{J}^{\left(\frac{\beta}{2}\right)}_{\lfloor \frac{N\beta}{2}\theta \rfloor}(\boldsymbol{a})}{\mathrm{J}^{\left(\frac{\beta}{2}\right)}_{\lfloor \frac{N\beta}{2}\theta \rfloor}(\mathbf{1})} = \int_0^\theta \log \tilde{\mathcal{S}}_A(s)\,\mathrm{d}s \qquad \text{for } \theta \leq t_A(x)\,. \tag{3.146}$$

In particular, for $\beta = 2$, the Jack polynomials become Schur polynomials, and Schur polynomials of a trivial partition degenerate into *complete homogeneous polynomials* defined by:

$$\mathrm{h}_k(\boldsymbol{a}) := \sum_{1\leq i_1 \leq \cdots \leq i_k \leq N} a_{i_1} \ldots a_{i_k}\,, \tag{3.147}$$

so that the LHS of Eq. (3.146) has a simple explicit expression in terms of the a_i in this case.

As an illustration of this example, let's take $\boldsymbol{a} = (2/N, \ldots, 2i/N, \ldots, 2)$, such that μ_A is given by the uniform distribution on $[0, 2]$ and $a_1 = \mathrm{a}_+ = 2$. In this case, one can show $t_A(\mathrm{a}_+) = \infty$, such that there is no phase transition and after some calculation one has:

$$\int_0^\theta \log \tilde{\mathcal{S}}_A(s)\,\mathrm{d}s = \theta\left(\log \frac{2\theta}{\left|\theta + 1 + W\left(-(\theta+1)e^{-(\theta+1)}\right)\right|} - 1\right) - \log\left|W\left(-(\theta+1)e^{-(\theta+1)}\right)\right|\,, \tag{3.148}$$

where $W(.)$ is the *Lambert W function*, which we compare with:

$$\frac{1}{N}\log \frac{\mathrm{h}_k(\boldsymbol{a})}{\mathrm{h}_k(1,\ldots,1)} = \frac{1}{N}\log\left[\frac{k!(N-1)!}{(k+N-1)!}\mathrm{h}_k(\boldsymbol{a})\right]\,, \tag{3.149}$$

for different N and $\theta = k/N$, where the a_i are the N equidistributed points between 0 and 2. The results are shown in Fig. 3.1.

3.4.6 Rectangular spherical integral and the rectangular C-transform

This section deals with rectangular matrices.

For the rectangular case, we can remove the constraints on the spheres by introducing two Lagrange multipliers z_1, z_2 and write:

$$\mathcal{I}^{(\beta)}_{q,\boldsymbol{a}}\left(\sqrt{NM}\beta\theta\right) = \frac{\Gamma\left(\frac{N\beta}{2}\right)}{\pi^{\frac{N\beta}{2}}}\frac{\Gamma\left(\frac{M\beta}{2}\right)}{\pi^{\frac{M\beta}{2}}}\left(\frac{1}{4\pi\mathrm{i}}\right)^2 \int_{\mathscr{C}_1,\mathscr{C}_2} \mathrm{d}z_1\mathrm{d}z_2 \int_{\mathbb{K}_\beta^{N+M}} \mathrm{d}\boldsymbol{y}\, \mathrm{e}^{\frac{z_1}{2}+\frac{z_2}{2}-\frac{1}{2}\boldsymbol{y}^\intercal \mathbf{Q}\boldsymbol{y}}\,, \tag{3.150}$$

where the matrix $\mathbf{Q} \equiv \mathbf{Q}(z_1, z_2, \theta, \mathbf{s})$ is given by:

$$\mathbf{Q} = \begin{pmatrix} z_1\mathbf{I}_N & -\sqrt{NM}\beta\,\theta\,\mathbf{Diag}(\boldsymbol{a}) & \mathbf{0}_{M-N,N} \\ -\sqrt{NM}\beta\,\theta\,\mathbf{Diag}(\boldsymbol{a}) & z_2\mathbf{I}_N & \mathbf{0}_{M-N,N} \\ \mathbf{0}_{M-N,N} & \mathbf{0}_{M-N,N} & z_2\mathbf{I}_{M-N} \end{pmatrix},$$

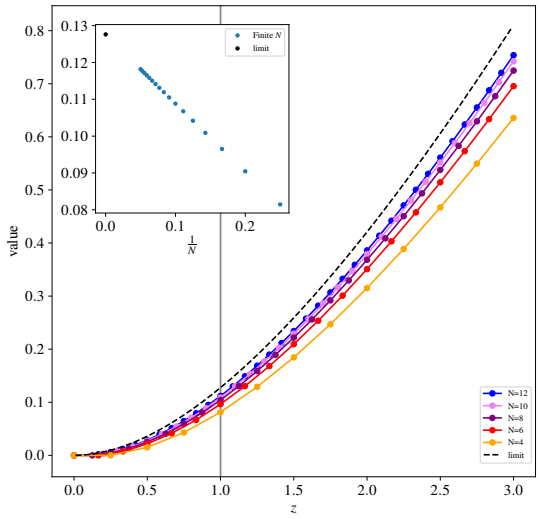


Figure 3.1: Value of the logarithm of the normalized complete homogeneous symmetric polynomials (3.149) for equidistributed entries between 0 and 2 for different N and different $k = Nz$, compared with the limiting behavior (3.148), represented by a dashed line. The inset graph represents the convergence at $k = N$ ($z = 1$) for more values of N, represented as a function of $\frac{1}{N}$.

with $\boldsymbol{a}$ the vector of singular values. By Gaussian integration, we have:

$$\mathcal{I}^{(\beta)}_{q,\boldsymbol{a}}\left(\sqrt{NM}\beta\theta\right) = \Gamma\left(\frac{N\beta}{2}\right)\Gamma\left(\frac{M\beta}{2}\right)\left(\frac{1}{4\pi\mathrm{i}}\right)^2 \int_{\mathscr{C}_1,\mathscr{C}_2} \mathrm{d}z_1\mathrm{d}z_2\, \mathrm{e}^{z_1/2+z_2/2}(\det\mathbf{Q})^{-\frac{\beta}{2}}\,, \tag{3.151}$$

so we need to compute the determinant of this matrix $\mathbf{Q}$. Expanding twice along the right-bottom block, we have:

$$\det\mathbf{Q} = \det(z_2\mathbf{I}_{M-N})\det\begin{pmatrix} z_1\mathbf{I}_N & -\sqrt{NM}\beta\theta\mathbf{Diag}(\boldsymbol{a}) \\ -\sqrt{NM}\beta\theta\mathbf{Diag}(\boldsymbol{a}) & z_2\mathbf{I}_N \end{pmatrix}, \tag{3.152}$$

$$\det\mathbf{Q} = z_2^{M-N}\det(z_2\mathbf{I}_N)\det\left(z_1\mathbf{I}_N - \left(\sqrt{NM}\beta\theta\right)^2\frac{1}{z_2}\mathbf{Diag}(\boldsymbol{a}^2)\right), \tag{3.153}$$

$$\det\mathbf{Q} = z_2^{M-N}\prod_{i=1}^{N}\left(z_1z_2 - NM\beta^2\theta^2a_i^2\right), \tag{3.154}$$

which gives:

$$\mathcal{I}^{(\beta)}_{q,\boldsymbol{a}}\left(\sqrt{NM}\beta\theta\right) = \Gamma\left(\frac{N\beta}{2}\right)\Gamma\left(\frac{M\beta}{2}\right)\times \int_{\mathscr{C}_1,\mathscr{C}_2}\mathrm{d}z_1\mathrm{d}z_2\, z_2^{\frac{N\beta-M\beta}{2}}\,\mathrm{e}^{\frac{z_1}{2}+\frac{z_2}{2}-\frac{\beta}{2}\sum_{i=1}^{N}\log\left(z_1z_2-NM\beta^2\theta^2a_i^2\right)}. \tag{3.155}$$

Let's do the change of variable $(z_1 \to N\beta\theta z_1, \mathrm{d}z_1 \to N\beta\theta\mathrm{d}z_1)$, $(z_2 \to M\beta\theta z_2, \mathrm{d}z_2 \to$

$M\beta\theta \mathrm{d}z_2)$ we have:

$$\mathcal{I}_{q,\boldsymbol{a}}^{(\beta)}\left(\sqrt{NM}\beta\theta\right) = \frac{\Gamma\left(\frac{N\beta}{2}\right)\Gamma\left(\frac{M\beta}{2}\right)}{(N\beta\theta)^{\frac{N\beta}{2}-1}(M\beta\theta)^{\frac{M\beta}{2}-1}} \times \left(\frac{1}{4\pi\mathrm{i}}\right)^2 \int_{\mathscr{C}_1,\mathscr{C}_2} \mathrm{d}z_1\mathrm{d}z_2 \mathrm{e}^{\frac{N\beta\theta}{2}z_1 + \frac{M\beta\theta}{2}z_2 + \frac{N\beta-M\beta}{2}\log(z_2) - \frac{\beta}{2}\sum_{i=1}^{N}\log\left(z_1z_2-a_i^2\right)} . \quad (3.156)$$

If we now consider the large N limit with the double scaling limit $N/M \to q$ and use Stirling approximation to get the behavior of the multiplicative constant, we have the following (double) integral representation:

$$\mathcal{I}_{q,\boldsymbol{a}}^{(\beta)}\left(\sqrt{NM}\beta\theta\right) = \left(\frac{1}{4\pi\mathrm{i}}\right)^2 \int_{\mathscr{C}_1,\mathscr{C}_2} \mathrm{d}z_1\mathrm{d}z_2 \mathrm{e}^{\frac{N\beta}{2}G(z_1,z_2,\theta)} , \quad (3.157)$$

with:

$$G(z_1,z_2,\theta) := \theta z_1 + \frac{\theta}{q}z_2 - \frac{1-q}{q}\log z_2 - \frac{1}{N}\sum_{i=1}^{N}\log\left(z_1z_2 - s_i^2\right) - \frac{1+q}{q}(1+\log\theta) + \mathcal{O}\left(\frac{1}{N}\right) . \quad (3.158)$$

Saddle-point computations -

In the large N limit, the complex integral of Eq. (3.157) is dominated by the saddle points $(z_1^*, z_2^*) \equiv (z_1^*(\theta), z_2^*(\theta))$ solutions of the zero-gradient equations:

$$\begin{cases} \partial_{z_1^*} G(z_1^*, z_2^*, \theta) &= 0 , \\ \\ \partial_{z_2^*} G(z_1^*, z_2^*, \theta) &= 0 . \end{cases} \quad (3.159)$$

That is, the solution of:

$$\theta - \frac{z_2^*}{N}\sum_{i=1}^{N}\left(z_1^*z_2^* - a_i^2\right)^{-1} = 0 , \quad (3.160)$$

and

$$\theta - (1-q)\frac{1}{z_2^*} - \frac{qz_1^*}{N}\sum_{i=1}^{N}\left(z_1^*z_2^* - a_i^2\right)^{-1} = 0 . \quad (3.161)$$

The equations Eq. (3.160) and Eq. (3.161) are coupled but can be easily taken care of by noting that:

1. if ones multiply Eq. (3.160) by qz_1^* and Eq. (3.161) by z_2^* and then subtract the two, one gets

$$\theta(qz_1^* - z_2^*) = (q-1) , \quad (3.162)$$

that is

$$z_2^* = qz_1^* + \frac{1-q}{\theta} , \quad (3.163)$$

2. if we put the variable θ on the RHS in both Eq. (3.160) and Eq. (3.161) and multiply the two equations, we get:

$$\theta^2 = q z_1^* z_2^* \left(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{z_1^* z_2^* - a_i^2} \right)^2 + (1-q) \sum_{i=1}^{N} \frac{1}{z_1^* z_2^* - a_i^2} \,, \tag{3.164}$$

that is following the expression (1.32) of the D-transform:

$$\theta = d_{\mathbf{A}}(\sqrt{z_1^* z_2^*}) \,. \tag{3.165}$$

Eq. (3.163) and Eq. (3.165) allow one to get the behavior of the spherical integral but as in the additive and multiplicative case, we need to be careful before inverting Eq. (3.165) and we have to consider two cases:

- *the case* $\theta \leq d_A(x)$*:* In this case, we can directly invert Eq. (3.165) and we get:

$$z_1^* z_2^* = \left[d_A^{\langle -1 \rangle}(\theta) \right]^2 \text{ for } \theta \leq d_A(x) \,. \tag{3.166}$$

  Injecting Eq. (3.163) in Eq. (3.166), we get the following quadratic equation for z_1^*:

$$q(z_1^*)^2 + \frac{1-q}{\theta} z_1^* - \left[d_A^{\langle -1 \rangle}(\theta) \right]^2 = 0 \,, \tag{3.167}$$

  whose (correct) solution is given by:

$$z_1^*(\theta) = \frac{-(1-q) + \sqrt{(1-q)^2 + 4q\theta^2 \left[d_A^{\langle -1 \rangle}(\theta) \right]^2}}{2q\theta} \,. \tag{3.168}$$

- *the case* $\theta \geq d_A(x)$*:* In this case, we have (again) a saturation. To satisfy Eq. (3.165), one must have:

$$\sqrt{z_1^* z_2^*} = x \,, \tag{3.169}$$

  since Eq. (3.163) is still valid, z_1^* is solution of the same quadratic equation (3.167) except that the term $d_A^{\langle -1 \rangle}(\theta)$ is replaced by x, so that we have:

$$z_1^*(x, \theta) = \frac{-(1-q) + \sqrt{(1-q)^2 + 4q\theta^2 x^2}}{2q\theta} \,. \tag{3.170}$$

We now have all the tools to compute the asymptotic behavior of the rectangular spherical integral, since z_1^*, z_2^* are the saddle points we have:

$$\frac{1}{N\beta} \frac{\mathrm{d}}{\mathrm{d}\theta} \log \left[\mathcal{I}_{q,\boldsymbol{a}}^{(\beta)} \left(\sqrt{NM}\beta\theta \right) \right] \to \partial_\theta G(z_1^*(\theta), z_2^*(\theta), \theta)/2 \,, \tag{3.171}$$

which using Eq. (3.158) gives:

$$\frac{1}{N\beta} \frac{\mathrm{d}}{\mathrm{d}\theta} \log \left[\mathcal{I}_{q,\boldsymbol{a}}^{(\beta)} \left(\sqrt{NM}\beta\theta \right) \right] \to \frac{1}{2} \left[z_1^* + \frac{z_2^*}{q} - \frac{1+q}{q} \frac{1}{\theta} \right] \,, \tag{3.172}$$

and using Eq. (3.163), we can express z_2^* as a function of z_1^* to have:

$$\frac{1}{N\beta}\frac{\mathrm{d}}{\mathrm{d}\theta}\log\left[\mathcal{I}_{q,\boldsymbol{a}}^{(\beta)}\left(\sqrt{NM}\beta\theta\right)\right] \to z_1^* - \frac{1}{\theta}\,. \tag{3.173}$$

If we now inject the expression of z_1^* for $\theta < d_A(x)$ and $\theta > d_A(x)$ given by Eq. (3.168) and Eq. (3.170) and compare it with the expression of the C-transform of Eq. (2.168) and the function U of Eq. (2.166), we get the following result.

Result 3.12 (*rank-one rectangular spherical integral and C-transform, [21]*)

In the large $N \to \infty$ limit where $N/M \to q$ and $\mu_{\mathbf{A}} := \sum_{i=1}^N \delta(. - a_i)/N \to \mu_A$ and the top singular value $a_1 \to x$, if we define the quenched free energy by:

$$J_A(x,\theta) := \lim_{N\to\infty}\frac{1}{N\beta}\log\left[\mathcal{I}_{q,\boldsymbol{a}}^{(\beta)}\left(\sqrt{NM}\beta\theta\right)\right]\,, \tag{3.174}$$

we have that the partial derivatives of this free energy are given by:

$$\partial_\theta J_A(x,\theta) = \begin{cases} \tilde{\mathcal{C}}_A^{(q)}(\theta) & \text{for } \theta \le d_A(x)\,, \\ \\ \frac{U(\theta x)}{\theta} & \text{for } \theta \ge d_A(x)\,. \end{cases} \tag{3.175}$$

and by:

$$\partial_x J_A(x,\theta) = \begin{cases} 0 & \text{for } \theta \le d_A(x)\,, \\ \\ \frac{\sqrt{(1-q)^2+4q\theta^2x^2}-\sqrt{(1-q)^2+4qd_A(x)^2x^2}}{2qx} & \text{for } \theta \ge d_A(x)\,. \end{cases} \tag{3.176}$$

where $\tilde{C}^{(q)}$ and d_A are respectively the C-transform and D-transform of the LSVD μ_A given by Eq. (2.168) and Eq. (1.32), and U is the function defined by Eq. (2.166).

3.4.7 Limit for annealed free energies

The result of this section will only be used in Chapter 5 dealing with the large deviation of the top eigenvalue of the sum/product of random matrices.

The setting -

Instead of looking at the *quenched free energy* of the spherical integral, we look at another limit where we first do an average over the laws of $\mathbf{A}$ (or equivalently over the laws of its eigenvalues $\boldsymbol{a}$), before taking the logarithm and taking the large N limit, that is we want to compute the *annealed free energy*. In order to do so, one must specify what is the law of the eigenvalues. As $N \to \infty$, we denote as usual by μ_A the LSD/LSVD with corresponding upper edge a_+. Importantly, we will consider a rather general setting where:

- the potential V is 'confining enough', that is, it is a convex function for values higher than a_+;

- The potential $V(x)$ is finite for values of x lower than a position of a *wall* $\mathrm{a}_+ \leq w_A \leq +\infty$ and is infinite for values higher than the wall.

The introduction of these invariant ensembles with a wall might seem odd at first, but as we will see later on, it will turn out to be very convenient when considering the problem of the large deviation of the top eigenvalue/singular value of the sum or the product of random matrices in Chapter 5. In order to highlight the presence of such a wall, we denote this law by $\boldsymbol{a} \sim \mathcal{P}^{(\beta)}_{V,w_A}$.

Reminder on the second branch of the Stieltjes transform-

Let's recall for convenience from Sec. 1.4, that the second branch of the Stieltjes transform $\bar{g}_A$ is given as the 'unphysical' solution of the algebraic BIPZ equation (1.84) with unknown g:

$$\mathrm{g}^2 - V'(z)\mathrm{g} + P(z) = 0\,, \tag{3.177}$$

that is for $z > \mathrm{a}_+$ this function is given by:

$$\bar{g}_A(z) = \frac{V'(z)}{2} + \frac{\sqrt{V'^2(z) - 4P(z)}}{2} = V'(z) - g_A(z)\,. \tag{3.178}$$

For $z \geq \mathrm{a}_+$, this function starts at $\bar{g}_A(\mathrm{a}_+) = g_A(\mathrm{a}_+)$ and is then continuously increasing with asymptotic behavior given by:

$$\bar{g}_A(z) \underset{z\to\infty}{\sim} V'(z)\,. \tag{3.179}$$

rather than the decaying behavior of the Stieltjes transform $g_A(z) \underset{z\to\infty}{\sim} 1/z$.

If now we look at the algebraic equation (3.177) the other way by fixing the value of $\mathrm{g} = y$ for some y in $(0, g_A(\mathrm{a}_+))$, the corresponding $z(y) \equiv z$ is by definition the inverse $g_A^{\langle -1\rangle}(y)$. If now the parameter y is higher than $g_A(\mathrm{a}_+)$ (but lower than $\mathrm{r}_A := \lim_{z\to\infty} V'(z)$), Eq. (3.177) is the implicit equation for the analytical continuation of $g^{\langle -1\rangle}$. Since this regime corresponds to the second branch of Stieltjes, we have a natural interpretation for the analytical continuation of $g_A^{\langle -1\rangle}$ beyond the point $g_A(\mathrm{a}_+)$: it is the inverse function of the second branch of the Stieltjes transform.

The additive case -

Based on ideas developed in Ref. [66], let's consider $\boldsymbol{a} \sim \mathcal{P}^{(\beta)}_{V,w_A}$, such that we have:

$$\mathbb{E}\,\mathcal{I}^{(\beta)}_{\boldsymbol{a}}\left(\frac{N\beta}{2}\theta\right) = \frac{1}{2\pi\mathrm{i}}\int_{\mathbb{R}^N\times\mathscr{C}} \mathrm{e}^{\frac{N\beta}{2}H_N(\boldsymbol{a},z,\theta)}\mathbb{I}_{\{a_i\leq w_A\}}\mathrm{d}\boldsymbol{a}\mathrm{d}z\,, \tag{3.180}$$

where $\mathbb{E}$ denotes the average over $\mathcal{P}^{(\beta)}_{V,w_A}$ and

$$H_N(\boldsymbol{a},z,\theta) := -\sum_{i=1}^N V(a_i) + \frac{1}{N}\sum_{i,j|j\neq i}\log|a_i - a_j| - \frac{1}{N}\sum_{i=1}^N \log(z-\lambda_i) + z\theta - 1 - \log\theta + \mathcal{O}\left(\frac{1}{N}\right). \tag{3.181}$$

This can be understood as the Hamiltonian of a system of $N+1$ particles[3]. In the large N limit, we argue that the integral is dominated by the most probable configuration given as the solution of the set of saddle-point equations:

$$\begin{cases} \partial_{a_i^*} H_N(\boldsymbol{a}^*, z^*, \theta) = 0 \quad \text{for } i = 1, \dots, N\,, \\ \\ \partial_{z^*} H_N(\boldsymbol{a}^*, z^*, \theta) = 0\,. \end{cases} \tag{3.182}$$

That is:

$$\begin{cases} V'(a_i^*) = \frac{2}{N} \sum_{j=1|j\neq i}^{N} \frac{1}{a_i^* - a_j^*} + \frac{1}{N} \frac{1}{z^* - a_i^*} \quad \text{for } i = 1, \dots, N\,, \\ \\ \theta = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{z^* - a_i^*}\,. \end{cases} \tag{3.183}$$

These two equations have to be understood with the additional constraints:

$$\begin{cases} a_1^* \leq z^* & , \\ \\ a_N^* \leq \cdots \leq a_1^* \leq w_A & . \end{cases} \tag{3.184}$$

We need to distinguish three different cases:

1. For $\theta \leq g_A(\mathrm{a}_+)$, the bottom line of Eq. (3.183) can be satisfied with the λ_i^*'s in their classical positions. By direct inversion we find:

$$z^*(\theta) = g_A^{\langle -1 \rangle}(\theta)\,, \tag{3.185}$$

   and by the self-averaging property, this gives the same result as in the quenched case.

2. For $g_A(\mathrm{a}_+) \leq \theta \leq \bar{g}_A(w_A)$, for the RHS of the bottom line of Eq. (3.183) to be equal to θ, one has to have the distance between z^* and a_1^* be of order $\mathcal{O}(1/N)$ so Eq. (3.183) (bottom) has to be understood as:

$$\theta \simeq g_A(z^*) + \frac{1}{N} \frac{1}{z^* - a_1^*} + \mathcal{O}\left(\frac{1}{N}\right)\,. \tag{3.186}$$

   Next, since $V(.)$ is analytic, we can approximate the potential and interaction term in Eq. (3.183) for $i = 1$ by:

$$V'(a_1^*) = V'(z^*) + \mathcal{O}\left(\frac{1}{N}\right)\,, \tag{3.187}$$

$$\frac{1}{N} \sum_{j=2}^{N} \frac{1}{a_1^* - a_j^*} \simeq g_A(z^*) + \mathcal{O}\left(\frac{1}{N}\right)\,, \tag{3.188}$$

[3]The variable z superficially looks like another eigenvalue repelled by all the other ones and with its own linear potential. But closer inspection reveals that the force between z and the λ_i is actually attractive. What is even stranger is that the equilibrium position of z is a local minimum of the probability. The reason for this is that in the integral of Eq. (3.180), z is integrated on a vertical line in the complex plane, so the second derivative in that direction should be positive for the integral to converge.

injecting this in the top line of Eq. (3.183) for $i = 1$ one gets:

$$V'(z^*) = 2g_A(z^*) + \frac{1}{N}\frac{1}{z^* - a_1^*} + \mathcal{O}\left(\frac{1}{N}\right). \tag{3.189}$$

Now making the difference of Eq. (3.186) and Eq. (3.189) to eliminate the term $(z^* - a_1^*)^{-1}$ and neglecting term of order $\mathcal{O}(1/N)$ one obtain a simple self-consistent equation for the unknown z^*:

$$V'(z^*) - g_A(z^*) = \theta\,, \tag{3.190}$$

which using Eq. (3.178) reads:

$$\bar{g}_A(z^*) = \theta\,, \tag{3.191}$$

where $\bar{g}_A(.)$ is the second branch of the Stieltjes transform. Following the properties of the second branch of the Stieltjes transform, inverting Eq. (3.191) yields:

$$z^*(\theta) = g_A^{\langle -1\rangle}(\theta)\,, \tag{3.192}$$

where $g_A^{\langle -1\rangle}(.)$ is here the analytical continuation beyond the point $g_A(\mathrm{a}_+)$ of the inverse of the Stieltjes.

3. for $\theta \geq \bar{g}_A(w_A)$, the position of the top eigenvalue becomes fixed at the wall and since the distance between z^* and this top of eigenvalue is infinitely small in the large N limit, we have again a saturation, but now at the position w_A instead of x in the quenched case:

$$z^*(\theta) = w_A\,. \tag{3.193}$$

All in all, we have the following result:

Result 3.13 (*additive annealed free energy*)

For $\boldsymbol{a} \sim \mathcal{P}_V^{(\beta)}$, such that V is strictly increasing beyond the top edge a_+ of the LSD μ_A and infinite after the position w_A, if we denote the annealed free energy by:

$$F_A(w_A, \theta) := \lim_{N\to\infty} \frac{2}{N\beta} \log \mathbb{E}\left[\mathcal{I}_{\boldsymbol{a}}^{(\beta)}\left(\frac{N\beta}{2}\,\theta\right)\right], \tag{3.194}$$

then we have:

$$\partial_\theta F_A(w_A, \theta) = \begin{cases} \mathcal{R}_A(\theta) & \text{for } \theta \leq \bar{g}_A(w_A)\,, \\ \\ w_A - \frac{1}{\theta} & \text{for } \theta \geq \bar{g}_A(w_A)\,, \end{cases} \tag{3.195}$$

where $\mathcal{R}_A$ is the analytical continuation of the R-transform of the LSD μ_A and $\bar{g}_A$ is the second branch of the Stieltjes transform.

Let's point out two fundamental remarks, corresponding to the limiting cases where the wall is either at the edge $w_A = \mathrm{a}_+$ or send to infinity $w_A \to \infty$, which will play an important role later on.

Remark *(wall at the edge)***.** If we choose the position of the wall to be exactly at the edge: $w_A = \mathrm{a}_+$, since we have the relation

$$\bar{g}_A(\mathrm{a}_+) = g_A(\mathrm{a}_+)\,, \tag{3.196}$$

Eq. (3.195) reads in this case:

$$\partial_\theta F_A(w_A = \mathrm{a}_+, \theta) = \begin{cases} \mathcal{R}_A(\theta) & \text{for } \theta \leq g_A(\mathrm{a}_+)\,, \\ \\ \mathrm{a}_+ - \frac{1}{\theta} & \text{for } \theta \geq g_A(\mathrm{a}_+)\,. \end{cases} \tag{3.197}$$

Comparing Eq. (3.197) and Eq. (3.124) (with the index C replaced by the index A and for $x = \mathrm{a}_+$), we see that we have indeed the relation (5.46) such that from the point of view of the annealed free energy, **we can consider a fixed diagonal matrix as an invariant matrix with a wall at the edge of its distribution**. ⌟

Remark *(wall at infinity and classical invariant ensemble)***.** Since classical ensembles are obtained by taking the limit $w_A \to \infty$. If we define by:

$$\mathrm{r}_A := \lim_{x \to +\infty} V'(x)\,, \tag{3.198}$$

we have for the corresponding annealed free energy:

$$\partial_\theta F_A(w_A \to \infty, \theta) = \begin{cases} \mathcal{R}_A(\theta) & \text{for } \theta \leq \mathrm{r}_A\,, \\ \\ \infty & \text{for } \theta \geq \mathrm{r}_A\,, \end{cases} \tag{3.199}$$

One may note that the second line of Eq. (3.199) is removed if $\mathrm{r}_A = \infty$ (which is for example is the case for a GOE matrix) but is present otherwise (which is, for example, is the case for a Wishart/LOE matrix for which $\mathrm{r}_A = 1/q$). ⌟

For classical ensembles, the result can be directly obtained by direct Gaussian integration, see the following two examples.

Example *(*GOE *matrices and Gaussian integration)***.** For $\mathbf{A}$ a GOE matrix with variance σ^2, the potential is convex on the whole real line and the annealed free energy can be directly obtained by Gaussian integration. By rotationally invariance, the average of the partition function is simply the moment generating function of one of the diagonal elements, say A_{11}. Since this element is a Gaussian random variable with variance $2\sigma^2/N$, we have:

$$\mathbb{E}_{\mathbf{A}}\left[\mathcal{I}_{\mathbf{A}}^{(\beta=1)}\left(\frac{N}{2}\,\theta\right)\right] = \int \frac{\mathrm{e}^{\frac{N\theta}{2}A_{11} - \frac{N}{4\sigma^2}A_{11}^2}}{\sqrt{4\pi\sigma^2/N}}\mathrm{d}A_{11} = \mathrm{e}^{\frac{N}{2}\frac{\sigma^2\theta^2}{2}}\,. \tag{3.200}$$

Using the expression of Eq. (2.142) for the R-transform, this gives indeed Eq. (3.195) for the annealed free energy with $w_A = \bar{g}_A(w_A) = \infty$. ∎

Example *(Wishart matrices and Gaussian integration)***.** For $\mathbf{A}$ a real Wishart/LOE matrix of shape parameter $q = N/M$, the annealed free energy can be also directly obtained by Gaussian integration. Indeed, we have $\mathbf{A} \overset{\text{in law}}{=} \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{x}_m \boldsymbol{x}_m^{\mathsf{T}}$, where the $\{\boldsymbol{x}_m\}_{m=1,\ldots,M}$

are M independent N-dimensional standard Gaussian vectors. Since $\mathbf{A}$ is rotationally invariant, we can remove the integral over the sphere and fix $\boldsymbol{\sigma} = \boldsymbol{e}_1 = (1, 0, \dots, 0)$ without loss of generality and this gives:

$$\mathbb{E}_{\mathbf{A}}\left[\mathcal{I}_{\mathbf{A}}^{(\beta=1)}\left(\frac{N}{2}\,\theta\right)\right] = \int \mathrm{e}^{\frac{N\theta}{2M}\boldsymbol{e}_1^{\mathsf{T}}\left(\frac{1}{M}\sum_{m=1}^{M}\boldsymbol{x}_m\boldsymbol{x}_m^{\mathsf{T}}\right)\boldsymbol{e}_1}\prod_{i=1}^{M}\frac{\mathrm{e}^{-\frac{|\boldsymbol{x}_m|^2}{2}}}{(2\pi)^{-\frac{M}{2}}}\mathrm{d}\boldsymbol{x}_m \tag{3.201}$$

$$= \left(\int \frac{\mathrm{e}^{-\frac{1}{2}\boldsymbol{x}^{\mathsf{T}}(\mathbf{I}-q\theta\boldsymbol{e}_1\boldsymbol{e}_1^{\mathsf{T}})\boldsymbol{x}}}{(2\pi)^{-\frac{1}{2}}}\mathrm{d}\boldsymbol{x}\right)^{M}. \tag{3.202}$$

Now for $\theta \geq 1/q$, the integral inside the bracket is diverging and hence we get $F_A'(\theta) = \infty$. Otherwise, we can do the Gaussian integration, and we have:

$$\frac{2}{N}\log\mathbb{E}_{\mathbf{A}}\left[\mathcal{I}_{\mathbf{A}}^{(\beta=1)}\left(\frac{N}{2}\,\theta\right)\right] = -\frac{1}{q}\log(1-q\theta) \qquad \left(\text{for } \theta < \frac{1}{q}\right). \tag{3.203}$$

Using the expression of Eq. (2.143) for the R-transform of a Wishart matrix, this gives indeed Eq. (3.195) for the annealed free energy. A similar computation can be done for generalized Wishart matrices, that is a matrix of the form $(\sqrt{\boldsymbol{\Sigma}}\mathbf{X})(\sqrt{\boldsymbol{\Sigma}}\mathbf{X})^*/T$, where $\mathbf{X}$ is a matrix with iid Gaussian entries and $\boldsymbol{\Sigma}$ is the covariance matrix. ∎

The multiplicative case -

The annealed average in the multiplicative is almost identical to the ones of the additive case, if we introduce the second branch of the T-transform $\bar{t}_A(.)$ satisfying:

$$\bar{t}_A(z) = z\bar{g}_A(z) - 1\,, \tag{3.204}$$

then $\bar{t}_A(.)$ can be seen as the analytical continuation of the inverse of $t_A^{\langle -1\rangle}$ (and conversely) and we have the following result.

Result 3.14 (*multiplicative annealed free energy*)

For $\boldsymbol{a} \sim \mathcal{P}_V^{(\beta)}$, such that V is convex beyond the top edge a_+ of the LSD μ_A and infinite after the position w_A, if we denote the annealed free energy by:

$$F_A(w_A, \theta) := \lim_{N\to\infty}\frac{2}{N\beta}\log\mathbb{E}\left[\mathcal{I}_{\times,\boldsymbol{a}}^{(\beta)}\left(\frac{N\beta}{2}\,\theta\right)\right], \tag{3.205}$$

then we have:

$$\partial_\theta F_A(w_A,\theta) = \begin{cases} \log\tilde{\mathcal{S}}_A(\theta) & \text{for } \theta \leq \bar{t}_A(w_A)\,, \\ \log\left(\frac{w_A\theta}{\theta+1}\right) & \text{for } \theta \geq \bar{t}_A(w_A)\,, \end{cases} \tag{3.206}$$

where $\tilde{\mathcal{S}}_A(\theta)$ is the analytical continuation of the S-transform and $\bar{t}_A$ is the second branch of the T-transform.

The rectangular case -

A similar computation as in the additive case shows that if we introduce the second branch of the D-transform:

$$\bar{d}_A(z) = \sqrt{qz^2 \left(\bar{g}_{AA^\intercal}(z^2)\right)^2 + (1-q)\bar{g}_{AA^\intercal}(z^2)}\,. \tag{3.207}$$

then we have the following result:

Result 3.15 (*rectangular annealed free energy*)

For $\boldsymbol{a} \sim \mathcal{P}_V^{(\beta)}$, such that V is convex beyond the top edge a_+ of the LSVD μ_A and infinite after the position w_A, if we denote the annealed free energy by:

$$F_A(w_A, \theta) := \lim_{N\to\infty} \frac{1}{N\beta} \log \mathbb{E}\left[\mathcal{I}_{q,\boldsymbol{a}}^{(\beta)}\left(\sqrt{NM}\beta\,\theta\right)\right]\,, \tag{3.208}$$

then we have

$$\partial_\theta F_A(w_A,\theta) = \begin{cases} \tilde{\mathcal{C}}_A^{(q)}(\theta) & \text{for } \theta \le \bar{d}_A(w_A)\,, \\ \\ \frac{U(\theta w_A)}{\theta} & \text{for } \theta \ge \bar{d}_A(w_A)\,, \end{cases} \tag{3.209}$$

where $\tilde{\mathcal{C}}_A^{(q)}$ is the Rectangular C-transform of Eq. (2.168).

3.5 Summary and conclusion of Chapter 3

In this chapter, we have reviewed results concerning the additive, multiplicative and rectangular spherical integrals. We first explain how one can naturally extend these spherical integrals to any $\beta > 0$ and this extension allows one to extrapolate the sum/product of matrices at any values of β. Then, we look at two different limits of these spherical integrals: the full-rank limit which is dictated by a variational principle, and the rank-one limit is related to the linearizing transforms of the free convolution of the previous chapter. The result of the additive full-rank limit will be (partially) used in the following chapter concerning the dynamics of large complex systems, while the result for the rank-one limit will be fundamental for the large deviation principle of Chapter 5 and the construction of the high-temperature convolution of Chapter 6. An important question concerning the asymptotic behavior of the spherical integrals is to understand the crossover between the full-rank regime and the small-rank regime. Another important question is the full-rank limit of the multiplicative spherical, which has been obtained here for $\beta = 2$ thanks to the determinantal formulae, but it is unclear how this result generalizes to other values of β. Last, the additive and multiplicative forms of the finite free convolution of Chapter 3, see Sec. 2.7 share many similarities with spherical integrals, and it will be interesting to know if they are also spherical integrals of a specific Gelfand pair.

Chapter 4

Stability of large complex systems with heterogeneous relaxation dynamics

The results of this chapter are based on the paper [141]. Its structure is very similar to the original paper, with minor changes to homogenize the notations with the rest of the thesis and avoid redundancy. The core of the chapter rely on the large deviation principle for the top eigenvalue of a β-ensemble developed in Chapter 1 (Sec. 1.5), the Dyson Brownian Motion of Chapter 2 (Sec. 2.5.1) and on the extensive rank limit (Sec. 3.3.1) and determinant formula of the HCIZ integral of Chapter 3 (Sec. 3.2.1). In particular, this chapter does not require any knowledge of free probability.

4.1 Introduction to May's work on large complex systems

One of the main objectives in the study of large complex systems is to understand their stability properties. A major theoretical contribution to answer this hard question was made by Robert May in 1972 [137]. Using a simple 'toy' model May argued that large complex systems might become unstable as the system complexity (measured by the strength of interactions between different units) increases. The seminal work of May was motivated by ecological questions at his time [2], but even today his results have found resonance among the study of large complex systems arising across disciplines including economical sciences [146], neural networks [161, 186], gene regulations [87] to cite a few. May's approach will be discussed in detail below and consists in approximating the dynamics of the system by a set of linear coupled equations with random coefficients, and we refer to [73, 27, 19] for recent studies going beyond this linear approximation.

In his original toy model, May considered a complex system consisting of N ecological species. To start with, each of the N species is assumed to be in equilibrium with population P_i^* ($i = 1, 2, \cdots, N$). Consider first the case where the species are non-interacting and linearly stable. By linearly stable, one means that if the population size P_i's are slightly perturbed from their equilibrium values, then the deviation $n_i(t) = P_i(t) - P_i^*$ for each i evolves in a

deterministic manner as

$$\frac{\mathrm{d}n_i(t)}{\mathrm{d}t} = -n_i(t) \quad \text{for } i = 1, \ldots, N\,. \tag{4.1}$$

For simplicity, May assumed an identical intrinsic decay rate (set to be unity in Eq. (4.1)) for each species, and this is what we call the *homogeneous* relaxation hypothesis. Imagine now *switching on* a pairwise interaction between the species, such that the dynamics are modified in the following way [137]

$$\frac{\mathrm{d}\,n_i(t)}{\mathrm{d}t} = -n_i(t) - \sqrt{T}\sum_j J_{ij} n_j(t) \quad \text{for } i = 1, \ldots, N\,, \tag{4.2}$$

where J_{ij} represents a pairwise interaction term which denotes the influence of the j^{th} species on the relaxation dynamics of the i^{th} species and $\sqrt{T}$ is a measure of the strength of this interaction. The notation $\sqrt{T}$ for this interaction strength may seem a bit strange at this stage, but we will see later that it corresponds to the 'time' in the associated Dyson Brownian motion picture of Sec. 2.5.1 of Chapter 2. May's further assumption was to model this complex interaction matrix J_{ij} as a random Gaussian matrix with real elements. The dynamics for $\boldsymbol{n}(t) = (n_1(t), \ldots, n_N(t))$ in Eq. (4.2) can be described in a compact matrix form as

$$\frac{\mathrm{d}\boldsymbol{n}(t)}{\mathrm{d}t} = -\left(\mathbf{I} + \sqrt{T}\,\mathbf{J}\right)\,\boldsymbol{n}(t)\,, \tag{4.3}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{J}$ is a real matrix with independent Gaussian entries. To make further progress, May also assumed that the interaction matrix J_{ij} is *symmetric*. In other words, the random matrix $\mathbf{J} = (J_{ij})_{i,j}$ coincides with the GOE matrices of Chapter 1. Note that for a GOE matrix $\mathbf{J}$ has the same distribution as $-\mathbf{J}$, hence we have chosen an overall negative sign in the interaction term in Eq. (4.2) without any loss of generality.

May's equation (4.3) then maps a dynamics question "Is the multi-component system stable?" to a RMT question "Are all the eigenvalues of the random matrix $\mathbf{B} = \mathbf{I} + \sqrt{T}\,\mathbf{J}$ positive?". Using the properties of GOE matrices, May argued that strictly in the large N limit (where all finite size fluctuations disappear), there exists a critical strength T_c where the system undergoes a *stability-instability phase transition* (sometimes known as May-Wigner transition): for $T < T_c$ the system is stable while for $T > T_c$ it is always unstable. Using the well-known Wigner semi-circular law for the average eigenvalue density of GOE eigenvalues as $N \to \infty$, May computed T_c explicitly for this homogeneous model [137]. Thanks to the well-known properties of GOE matrices, one can go beyond May's calculation of T_c and even investigate the regime where N is still large but finite and derive the behaviors of the typical and atypical fluctuations of the stability property of the system [120], thanks to the large deviation principle of the top eigenvalue of a GOE matrix, developed in Chapter 1, and this will be recalled briefly in the next section.

One of the important ingredients in May's model (apart from the fact that $\mathbf{J}$ is a GOE matrix) was to assume a homogeneous decay rate for all species. In this chapter, we address a simple question: assuming that $\mathbf{J}$ is still a GOE matrix, how the May-Wigner transition gets modified if one just makes the intrinsic decay rates for the species *heterogeneous*? This is a natural and simple generalization of May's original toy model. In this heterogeneous version, one just

replaces the identity matrix $\mathbf{I}$ in Eq. (4.3) with an arbitrary diagonal matrix with positive entries $\mathbf{A} = \mathbf{Diag}(a_1, \ldots, a_N)$. Eq. (4.3) now gets modified to

$$\frac{\mathrm{d}\boldsymbol{n}(t)}{\mathrm{d}t} = -(\mathbf{A} + \sqrt{T}\,\mathbf{J})\boldsymbol{n}(t) = -\mathbf{B}\,\boldsymbol{n}(t)\,, \tag{4.4}$$

where the effective relaxation matrix

$$\mathbf{B} = \mathbf{A} + \sqrt{T}\mathbf{J} = \sqrt{T}\left[\mathbf{J} + \frac{1}{\sqrt{T}}\,\mathbf{A}\right]\,, \tag{4.5}$$

can be interpreted as a deformation of the GOE matrix $\mathbf{J}$ by an additive positive diagonal matrix $\mathbf{A}$, with $1/\sqrt{T}$ playing the role of the strength of 'perturbation'. In May's original homogeneous model where $\mathbf{A} = \mathbf{I}$, the matrix $\mathbf{B}$, for any strength parameter $\sqrt{T}$, is just a shifted GOE matrix. However, in the generic case $\mathbf{A} \neq \mathbf{I}$, the spectrum of $\mathbf{B}$ is more complex, and it is described by the Dyson Brownian Motion of Chapter 2 (see Sec. 2.5.1). While deformed GUE (Gaussian unitary ensemble) models have been studied extensively in the recent past with many applications (see e.g. [105] and references therein), here we obtain a natural example of a deformed GOE matrix.

For this heterogeneous May model, we expect again that in the limit $N \to \infty$, where there are no finite size fluctuations, there should a critical value T_c separating the stable ($T < T_c$) and the unstable ($T > T_c$) phases. However, it turns out that the moment the intrinsic diagonal positive rate matrix $\mathbf{A}$ differs from $\mathbf{I}$, computing T_c becomes nontrivial. We will first develop a general method to compute T_c for arbitrary diagonal positive $\mathbf{A}$, and then use it to calculate T_c explicitly for a particularly interesting case where the elements of $\mathbf{A}$ are distributed uniformly over a finite interval (we will refer to this case as the *flat initial condition* since this corresponds to the value of $\mathbf{B}$ at "time" $T = 0$). This is the first main result of this chapter.

Next, for a general positive diagonal matrix $\mathbf{A}$, computing *explicitly* the average density profile or LSD of the eigenvalues of the deformed matrix $\mathbf{B}$ is also hard. However, for the 'flat initial condition' described above, we are able to compute analytically the average density of the eigenvalues of $\mathbf{B}$ in the large N limit (in explicit parametric form), providing the second main result of this chapter.

Finally, for the same choice of $\mathbf{A}$ (the flat initial condition), we make the link with another ensemble, the deformed GUE, for which one can compute the joint density of the eigenvalues, going beyond just the average density. This is possible thanks to the determinantal formula of the additive spherical (HCIZ) integral of Chapter 3,Sec. 3.2.1. To the best of our knowledge, this provides a new RMT ensemble–a Coulomb gas in a harmonic potential, where the repulsive interaction between any pair of eigenvalues is a linear combination of logarithmic (as in the standard GUE) and log-sinh types. The RMT ensemble with only logarithmic (the standard GUE) or only log-sinh interaction [132, 133, 53, 171, 166, 68] have been studied before, but here we obtain naturally a linear combination of them as interaction. Such a mixed Coulomb gas is interesting to study in its own right. Moreover, this Coulomb gas approach also allows us to estimate, how for finite but large N, the probability of stability differs from 1 on the stable side as one decreases T below T_c with $T_c - T \sim \mathcal{O}(1)$ (we recall that strictly in the $N \to \infty$ limit, the probability of stability is exactly 1 for $T < T_c$ by definition).

The rest of this chapter is organized as follows: In Section 4.2, we recall in detail the derivation and properties of May's original model. In particular, we describe in detail the finite size effects

on the May-Wigner transition, in terms of the Tracy-Widom distribution and the large deviation functions describing respectively the typical and atypical fluctuations of the system. We then describe the new main model with heterogeneous relaxation dynamics. In Section 4.3, we recall the main tools to perform the study of the heterogeneous model: the Stieltjes transform, and its link with Dyson Brownian Motion (DBM) and the Burgers' equation. This allows us to get the equation satisfied by the critical strength T_c for a generic matrix $\mathbf{A}$. In Section 4.4, we show how one can get the parametric solution for the density, based again on the Burgers' equation. In Section 4.5, we describe the deformed GUE with flat initial conditions and show how one can get the joint law density thanks to the Harish-Chandra-Itzykson-Zuber (HCIZ) integral, we then make the link with different models and show how one can get the large deviation function in the weakly stable phase for the original deformed GOE model with a flat initial condition.

4.2 May's homogeneous model and its heterogeneous generalization

4.2.1 May's original homogeneous model

The homogeneous May model has already been described in the introduction. In this subsection, we show how T_c for this model is computed in the strict $N \to \infty$ limit and then demonstrate how the probability of stability gets modified when N is large but finite. Also, this recapitulation would be useful for understanding the stability issues in the general setting of the heterogeneous model that we will discuss in the next subsection.

The deviations $n_i(t) = P_i(t) - P_i^*$ evolve via Eq. (4.3) in the original homogeneous model, where $\mathbf{J}$ is a GOE matrix with variance $\sigma^2 = 1$. We recall that the law of the elements of $\mathbf{J}$ is given by Eq. (1.39) and the distribution of the eigenvalues converges as $N \to \infty$ to the semi-circle distribution of Eq. (1.40) whose support is the interval $[-2, 2]$.

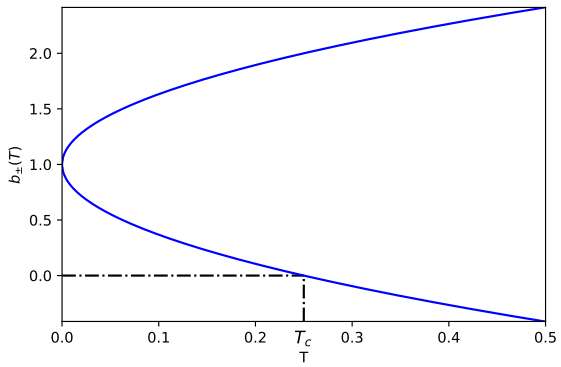


Figure 4.1: Plot of the edges $\mathrm{b}_\pm(T) = 1 \pm 2\sqrt{T}$ of a Wigner semi-circular distribution of variance T as a function of T. The critical strength occurs at $T_c = \frac{1}{4}$, where the lower edge $\mathrm{b}_-(T)$ hits zero.

The matrix form of May's equation (4.3) reads

$$\frac{\mathrm{d}\boldsymbol{n}(t)}{\mathrm{d}t} = -\mathbf{B}\,\boldsymbol{n}(t)\,, \tag{4.6}$$

where the effective relaxation matrix

$$\mathbf{B} = \mathbf{I} + \sqrt{T}\,\mathbf{J}\,, \tag{4.7}$$

is just a shifted GOE. Let $\lambda_1 \geq \ldots \geq \lambda_N$ and similarly $b_1(T) \geq b_2(T) \geq \ldots \geq b_N(T)$ denote the ordered eigenvalues of $\mathbf{J}$ and $\mathbf{B}$ in Eq. (4.7) respectively. Clearly,

$$b_i(T) = 1 + \sqrt{T}\,\lambda_i\,, \;\; \text{for all}\, i = 1, 2, \ldots, N\,. \tag{4.8}$$

One can now write down the condition for stability in terms of the ordered eigenvalues $\{b_i(T)\}$. From Eq. (4.6), it is clear that the system is stable if all eigenvalues of $\mathbf{B}$ are positive. Hence the probability of the stability can be expressed, for fixed T and N, as

$$\mathcal{P}_{\rm stable}(T, N) := \mathrm{Prob}\left[b_1(T) > 0, \ldots, b_N(T) > 0\right]\,, \tag{4.9}$$

or equivalently since we have ordered the eigenvalues

$$\mathcal{P}_{\rm stable}(T, N) = \mathrm{Prob}\left[b_N(T) > 0\right] = \mathrm{Prob}\left[\lambda_N > -\frac{1}{\sqrt{T}}\right]\,, \tag{4.10}$$

where we used $b_N(T) = 1 + \sqrt{T}\,\lambda_N$ from Eq. (4.8). For finite N, the value of λ_N, and hence that of $b_N(T) = 1 + \sqrt{T}\,\lambda_N$ fluctuates from sample to sample. However, strictly in the $N \to \infty$ limit, we have seen before that the eigenvalues of $\mathbf{J}$ converge, almost surely, to Wigner semi-circular law. This means that, as $N \to \infty$, Since λ_N is the lowest eigenvalue, it converges to the lower edge of the semi-circular, i.e., $\lambda_1 \to -2$. Consequently, from Eq. (4.8), the eigenvalues $\boldsymbol{b}(t)$ of $\mathbf{B}$ also converge to a shifted semi-circular law over the finite support $[\mathrm{b}_-(T), \mathrm{b}_+(T)]$, where the edges are given by:

$$\mathrm{b}_-(T) = 1 - 2\sqrt{T} \quad \text{and} \quad \mathrm{b}_+(T) = 1 + 2\sqrt{T}\,. \tag{4.11}$$

In particular, the lowest eigenvalue $b_N(T)$ converges to the lower edge as $N \to \infty$, i.e., $b_N(T) \to \mathrm{b}_-(T) = 1 - 2\sqrt{T}$, see Fig. 4.1. This means that as $N \to \infty$, almost surely, $b_N(T) > 0$ if $T < T_c = 1/4$ and $b_N(T) < 0$ if $T > T_c = 1/4$. Thus, the probability of stability in Eq. (4.10) also converges to an N-independent form as $N \to \infty$

$$\mathcal{P}_{\rm stable}(T, \infty) = \begin{cases} 1 & \text{if } T < T_c = \frac{1}{4}\,, \\ 0 & \text{otherwise.} \end{cases} \tag{4.12}$$

Thus, strictly in the $N \to \infty$ limit, the stability probability, as a function of T, approaches a 'sharp' step function with the step located at $T_c = 1/4$, as shown in Fig. 4.2.

However, for finite but large N, this curve $\mathcal{P}_{\rm stable}(T, N)$ vs. T will deviate from the step function (see Fig. 4.2) and one may ask how does the step function get modified for finite but large N. To extract this information, we see from Eq. (4.10) that we need to know the probability distribution of the lowest (minimum) eigenvalue λ_N of an $(N \times N)$ GOE matrix $\mathbf{J}$. Since for a Gaussian random matrix, the top eigenvalue λ_1 has the same distribution as $-\lambda_N$

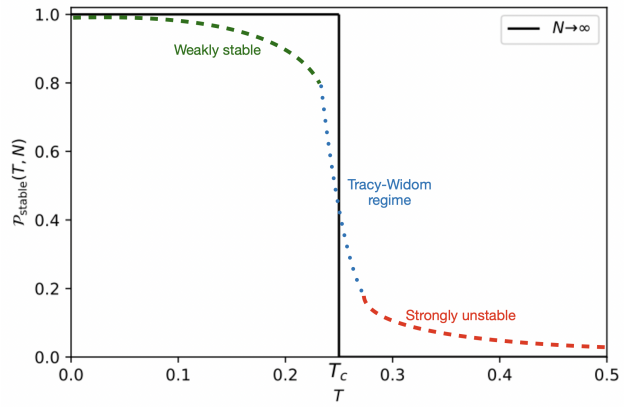


Figure 4.2: Sketch of the stability diagram in May's homogeneous model: the stability probability $\mathcal{P}_{\rm stable}(T,N)$ as a function of T for fixed large N. The solid (black) line corresponds to the strictly $N \to \infty$ limit, where $\mathcal{P}_{\rm stable}(T,N)$ is a step function with the step at $T_c = 1/4$. For finite but large N, this step function gets smoothened, as shown schematically by the dashed and dotted lines. The dotted line around T_c, shows the behavior of $\mathcal{P}_{\rm stable}(T,N)$ on a scale $|T-T_c| \sim \mathcal{O}(N^{-2/3})$, and has the Tracy-Widom form. The dashed lines describe the behavior of $\mathcal{P}_{\rm stable}(T,N)$ when $|T-T_c| \sim \mathcal{O}(1)$ and are described by the two large deviation behaviors on the two sides of T_c.

by symmetry, we can equivalently express the stability probability in Eq. (4.10) in terms of the distribution of the top eigenvalue λ_1 of the GOE matrix, namely

$$\mathcal{P}_{\rm stable}(T,N) = {\rm Prob}\left[b_1(T)>0\right] = {\rm Prob}\left[\lambda_1 > -\frac{1}{\sqrt{T}}\right] = {\rm Prob}\left[\lambda_1 < \frac{1}{\sqrt{T}}\right] . \qquad (4.13)$$

Thus, we need to know how the top eigenvalue λ_1 of an $(N\times N)$ GOE matrix is distributed for finite but large N. At the time of May's original work [137], this information was not available. Currently, however, one knows a great deal about the distribution of the top eigenvalue λ_1 of a $(N\times N)$ GOE matrix for finite but large N, as we have seen in Chapter 1, Sec. 1.5. This information was used to estimate $\mathcal{P}_{\rm stable}(T,N)$ for finite but large N in Ref. [120], which we briefly recall below.

Summary of the large N behavior of the top eigenvalue λ_1 of an $(N\times N)$ GOE matrix -

As mentioned earlier, the largest eigenvalue λ_1 converges to 2 as $N\to\infty$, i.e., coincides with the right edge of the Wigner semi-circular density. However, for finite but large N, the random variable fluctuates around this right edge 2 and the cumulative distribution admits the following

summary behavior, as discussed in Chapter 1:

$$\text{Prob}\left[\lambda_1 < w\right] \approx \begin{cases} \exp\left[-\frac{N^2}{2}\Psi_-(w) + o(N^2)\right] & \text{for } w < 2 \text{ and } |w-2| \sim \mathcal{O}(1)\,, \\ \\ \mathcal{F}^{(1)}\left(N^{2/3}(w-2)\right) & \text{for } |w-2| \sim \mathcal{O}(N^{-\frac{2}{3}})\,, \\ \\ 1 - \exp\left[-\frac{N}{2}\Psi_+(w) + o(N)\right] & \text{for } w > 2 \text{ and } |w-2| \sim \mathcal{O}(1)\,. \end{cases} \tag{4.14}$$

where we recall that $\mathcal{F}^{(1)}$ is the $\beta = 1$ Tracy-Widom distribution and $\Psi_- \equiv \Psi_{left,\mathsf{G}}$, $\Psi_- \equiv \Psi_{\mathsf{G}}$ are the two large deviations functions given by Eq. (1.108) and Eq. (1.126). Let's recall that the two large deviation functions have the following asymptotic behaviors near the edge $w = 2$:

$$\Psi_-(w) \propto (2-w)^3 \text{ for } w \to 2 \text{ and } w < 2\,, \tag{4.15}$$

$$\Psi_+(w) \propto (w-2)^{\frac{3}{2}} \text{ for } w \to 2 \text{ and } w > 2\,. \tag{4.16}$$

which match smoothly with the Tracy-Widom tails of Eq. (1.102) and Eq. (1.103).

Large N behavior of the stability probability $\mathcal{P}_{\text{stable}}(T,N)$ in the homogeneous May model-

Using the relation in Eq. (4.13) one can then translate the large N behavior of the cumulative density function (CDF) of the top eigenvalue $\text{Prob}[\lambda_N < w]$ into the large N behavior of $\mathcal{P}_{\text{stable}}(T,N)$ in May's homogeneous model. Setting $w = 1/\sqrt{T}$ in Eq. (4.14), we see that the Wigner edge $w = 2$ corresponds to $T_c = 1/4$ and the behaviors of the probability of stability $\mathcal{P}_{\text{stable}}(T,N)$ around $T_c = 1/4$ for finite, but large N are described by

$$\mathcal{P}_{\text{stable}}(T,N) \approx \begin{cases} \exp\left[-\frac{N^2}{2}\Phi_+(T) + o(N^2)\right] & \text{for } T > T_c = 1/4 \text{ and } |T-T_c| \sim \mathcal{O}(1)\,, \\ \\ \mathcal{F}^{(1)}\left(N^{2/3}\left(T^{-1/2}-2\right)\right) & \text{for } |T-T_c| \sim \mathcal{O}(N^{-\frac{2}{3}})\,, \\ \\ 1 - \exp\left[-\frac{N}{2}\Phi_-(T) + o(N)\right] & \text{for } T < T_c = 1/4 \text{ and } |T-T_c| \sim \mathcal{O}(1)\,, \end{cases} \tag{4.17}$$

where $\mathcal{F}^{(1)}$ is again the Tracy-Widom (GOE) function and now the rate functions $\Phi_\pm(w)$ are given by:

$$\Phi_\pm(w) = \Psi_\mp\left(w = \frac{1}{\sqrt{T}}\right)\,, \tag{4.18}$$

with $\Psi_\mp$ given by Eq. (1.108) and Eq. (1.126). These behaviors are schematically sketched by the dashed-dotted lines in Fig. 4.2 and describe precisely how the sharp step function (for $N \to \infty$) gets modified for finite but large N. In fact, the critical behavior around $T_c = 1/4$ for finite N in May's homogeneous model is similar to the so-called 'double scaling' limit in various matrix models arising in lattice gauge theory and they all share a 'third order' phase transition around the critical point, as reviewed extensively in Ref. [120].

Let us remark that for finite but large N and $T > T_c$, the stability probability $\mathcal{P}_{\rm stable}(T, N)$ in Eq. (4.17) in the large deviation regime $T - T_c \sim \mathcal{O}(1)$ deviates only very slightly $\sim \exp[-\mathcal{O}(N^2)]$ from its value 0 when $N \to \infty$. Thus, the $N \to \infty$ 'unstable' phase remains 'strongly' unstable when N reduces from ∞. Hence we refer to this phase as 'strongly unstable' in Fig. 4.2. In contrast, for $T < T_c$, the deviation of $\mathcal{P}_{\rm stable}(T, N)$ from its $N \to \infty$ value 1 is of order $\sim \exp[-\mathcal{O}(N)]$ which is much larger than the deviation $\sim \exp[-\mathcal{O}(N^2)]$ on the other side, i.e., for $T > T_c$. Thus, for $T < T_c$, the $N \to \infty$ 'stable' phase, where the system was stable with probability 1 when $N \to \infty$, is likely to change with a relatively higher probability when N is reduced from ∞. Hence, in Fig. 4.2, we refer to the phase $T < T_c$ as the 'weakly stable' phase.

Finally, we remark that these two different N dependence of the large deviation behaviors of $\mathcal{P}_{\rm stable}(T, N)$ on either side of T_c admits a nice physical interpretation in terms of the underlying log-gas picture of the eigenvalues of the relaxation matrix $\mathbf{B} = I + \sqrt{T}\,\mathbf{J}$, see Fig. 4.3. One can view the eigenvalues of the matrix $\mathbf{B}$ as a gas of N particles living on the real line, confined by a harmonic potential and subject to a pairwise logarithmic repulsive interaction. For $T < T_c$, the system is asymptotically stable: this means all the eigenvalues $\{b_i(T)\}$ are above 0 for $T < T_c$ with probability 1 in the $N \to \infty$ limit. To reduce this probability from unity, i.e, to trigger an event that will make the system unstable for $T < T_c$, one needs a rare configuration of charges for which the lowest eigenvalue $b_1(T) < 0$. This can be achieved by *pulling* the lowest eigenvalue $b_N(T)$ from its spectrum (whose lower edge is above 0 for $T < T_c$) to the value 0. This costs energy of order $\mathcal{O}(N)$ since one needs to disturb (pull) only one eigenvalue, without disturbing the rest of the spectrum. Hence, this explains the behavior $1 - \mathcal{P}_{\rm stable}(T, N) \sim \exp[-\mathcal{O}(N)]$ for $T < T_c$. In contrast, for $T > T_c$, the system is asymptotically unstable, i.e., the lower edge of the spectrum of eigenvalues $\{b_i(T)\}$ is already below 0. To increase stability, one needs to create a rare configuration where one *pushes* the whole gas of eigenvalues above 0. Since this involves a re-arrangement of N particles in the Coulomb gas, it will cost energy of $\mathcal{O}(N^2)$ (since each pair will contribute when the whole gas is compressed from its equilibrium configuration). This explains the behavior $\mathcal{P}_{\rm stable}(T, N) \sim \exp[-\mathcal{O}(N^2)]$ for $T > T_c$. This 'pulled' to 'pushed' phase transition occurs also in various lattice gauge models [80, 185, 71] where the 'pulled' phase corresponds to the 'weak coupling' phase in gauge theory, while the 'pushed' phase corresponds to the 'strong coupling' phase in gauge theory (for a review see [120]). Thus, the 'stability-instability' phase transition in May's homogeneous model can also be viewed as a 'pulled-pushed' transition. The 'stable' phase in May's model is the analog of the 'weak coupling' phase of the gauge theory, while the 'unstable' phase is the analog of the 'strong coupling' phase of the gauge theory [120].

4.2.2 Heterogeneous relaxation dynamics

A natural extension of May's work is to drop the assumption that all the damping constants are equal and allow a spread in the distribution of the damping constants a_i's, i.e., modify the evolution equation (4.2) to

$$\frac{\mathrm{d}\, n_i(t)}{\mathrm{d}t} = -a_i\, n_i(t) - \sqrt{T} \sum_j J_{ij} n_j(t) \quad \text{for } i = 1, \ldots, N \ , \tag{4.19}$$

where the $a_i > 0$'s are not necessarily equal. In the matrix form, this can be written as as Eq. (4.4) with $\mathbf{A}$ being a diagonal matrix with positive entries $\{a_1, a_2, \ldots, a_N\}$. To keep the model

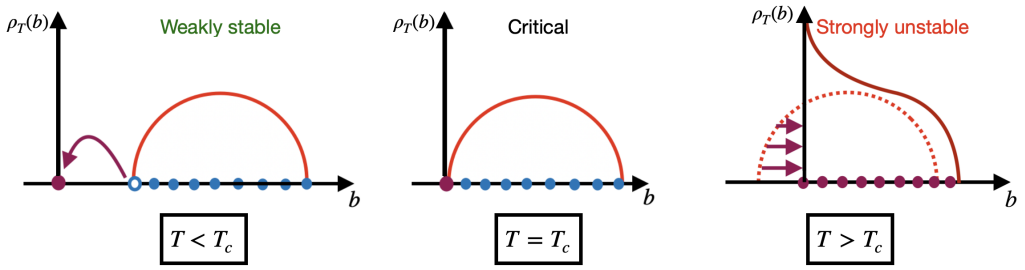


Figure 4.3: Sketch of the two different processes leading to the different scaling in N of the probability (4.17). **(Left)** For $T < T_c$, one needs to pull the lowest eigenvalue to the origin to make the system unstable, which does not change the equilibrium density. **(Center)** For $T = T_c$, the system is at the critical position where its lowest eigenvalue goes to 0. **(Right)** For $T > T_c$, to make the system stable, one needs to push all the eigenvalues below 0 by a wall, in such a case the gas rearranges itself and the equilibrium density is modified.

simple, we will still assume that the matrix $\mathbf{J}$ in Eq. (4.4) is a GOE matrix with the law given by (1.39). Since we will first study this generalized system in the $N \to \infty$ limit, we assume the empirical distribution of the a_i's converges to a continuous distribution $\mu(a)$ whose support is included in the positive real axis (since we have assumed the $a_i > 0$ to ensure stability without interactions). Thus, $\mu(a)$ can be considered as the 'initial' value of the deformed GOE matrix $\mathbf{B}$ at $T = 0$. The homogeneous May model corresponds to the choice of the 'initial' condition

$$\mu(a) = \delta(a - 1)\,. \tag{4.20}$$

Our main goal, in this chapter, is to understand how the May-Wigner transition may get modified when there is a spread or heterogeneity in the 'initial' density $\mu(a)$.

Starting from a given 'initial' density $\mu(a)$ at $T = 0$, the eigenvalues $\{b_i(T)\}$ of $\mathbf{B}$ will evolve in 'time' T. The first natural question is: for a general 'initial' density $\mu(a)$, what is the limiting density $\rho_T(b)$ of the eigenvalues $\{b_i(T)\}$ at time T, in the $N \to \infty$ limit? For the special homogeneous initial condition in Eq. (4.20), we have seen in the previous subsection that $\rho_T(b)$ is a shifted semi-circular law with support over $b \in [1 - 2\sqrt{T}, 1 + 2\sqrt{T}]$ at 'time' T. For a general $\mu(a)$, we will again expect that the limiting density $\rho_T(b)$ will have a finite support $b \in [\mathrm{b}_-(T), \mathrm{b}_+(T)]$ at time T. If one can compute the location $\mathrm{b}_-(T)$ of the lower edge of the support of the limiting density as a function of T, then setting $b_-(T = T_c) = 0$ will give us access to the exact critical strength T_c for an arbitrary 'initial' condition $\mu(a)$.

Computing the limiting density $\rho_T(b)$ at T for arbitrary 'initial' density $\mu(a)$ seems rather hard. However, one can make analytical progress for a specific choice of the 'initial' values $b_i(T = 0) = a_i$,

$$a_{N-i} = 1 + \sigma \frac{i-1}{N} \quad \text{for } i = 1, \dots, N\,, \tag{4.21}$$

which we call the *flat initial condition* since in the limit $N \to \infty$, the distribution of the a_i's given by (4.21) converges towards the flat distribution $\mu(a)$ between 1 and $1 + \sigma$:

$$\mu(a) = \frac{1}{\sigma} \mathbb{I}_{[1,1+\sigma]}(a)\,, \tag{4.22}$$

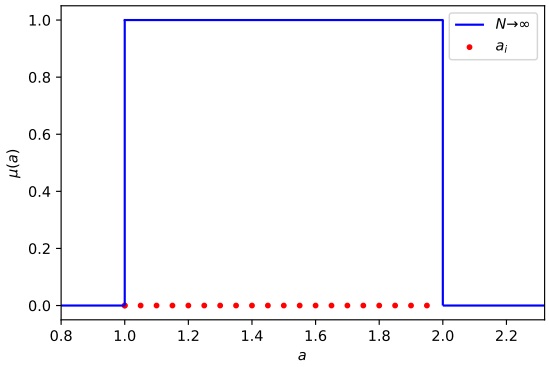


Figure 4.4: Flat discrete initial configuration of the a_i for $N = 20$ (in red) and their limiting flat density as $N \to \infty$ (in blue) for $\sigma = 1$.

where $\mathbb{I}_{[a,b]}(x)$ is the *indicator function*: $\mathbb{I}_{[a,b]}(x) = 1$ if x is in $[a, b]$ and 0 otherwise, see Fig. 4.4. The parameter σ controls the width of this distribution and in particular the limit $\sigma \to 0$ corresponds to the homogeneous limit of May, so we can consider this new model as one parameter extension of May's original model.

For this 'flat initial condition', we are able to compute, in the $N \to \infty$ limit, the limiting density $\rho_T(b)$ for all T. In particular, we will see in the next section that the precise knowledge of the lower edge $\mathrm{b}_-(T)$ of its support will enable us to compute the exact value of T_c in this model. Furthermore, for finite but large N, we expect that for the 'flat initial condition', the stability probability $\mathcal{P}_{\mathrm{stable}}(T, N)$ near its critical point $T = T_c$ will have a qualitatively similar behavior as in its homogeneous counterpart in Eq. (4.17): In particular, from the general universality argument of the top eigenvalue of a GOE, we expect that the typical fluctuation of $\mathcal{P}_{\mathrm{stable}}(T, N)$ will still be described by the ($\beta = 1$) Tracy-Widom scaling function $\mathcal{F}^{(1)}(x)$ in the middle line of Eq. (4.17). However, the large deviation functions in the region $|T - T_c| \sim \mathcal{O}(1)$, respectively in the 'unstable' and the 'stable' side, are expected to be different in this heterogeneous 'flat initial condition' model. We will see in later sections that while we can compute the rate function on the 'weakly stable' side, i.e., for $T < T_c$, computing the rate function in the 'strongly unstable' phase remains a hard challenging problem even for the flat initial condition case.

4.3 Critical strength and the hitting time of a Dyson Brownian Motion

4.3.1 Reminder on DBM and Burger's equation

The idea to characterize the critical strength in the general setting is to think of the parameter T as a (fictitious) 'time' variable, since we know from Chapter 2, Sec. 2.5.1, that the laws of the eigenvalues $\boldsymbol{b}$ of the matrix $\mathbf{B}$ corresponds to joint law of a Dyson Brownian Motion (DBM) with $\beta = 1$ starting at time $t = 0$ at $\boldsymbol{a}$ and evaluated at time T, that is:

$$\frac{\mathrm{d}b_i(T)}{\mathrm{d}T} = \frac{1}{N} \sum_{j:j\neq i} \frac{1}{b_i(T) - b_j(T)} + \sqrt{2D}\, \eta_i(T)\,, \tag{4.23}$$

starting from the initial conditions,

$$b_i(0) = a_i \qquad \text{for } i = 1, \dots, N\,, \tag{4.24}$$

and with $D = 1/N$ and for each i, $\eta_i(T)$ in Eq. (4.23) is an independent Gaussian white noise with zero mean and correlator $\langle \eta_i(t)\eta_j(t')\rangle = \delta_{ij}\,\delta(t-t')$. In the limit $N \to \infty$, we recall that the LSD of the $b_i(T)$

$$\rho_T(b) := \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta(b - b_i)\,, \tag{4.25}$$

is completely determined by its Stieltjes transform

$$g(z,T) := \lim_{N\to\infty} \frac{1}{N} \mathrm{Tr}\,(z\,\mathbf{I} - \mathbf{B})^{-1} = \int_{\mathrm{b}_-(T)}^{\mathrm{b}_+(T)} \frac{\rho_T(b)}{z-b} \mathrm{d}b\,, \tag{4.26}$$

which is the solution of the complex inviscid Burgers' equation:

$$\partial_T g(z,T) + g(z,T)\partial_z g(z,T) = 0\,, \tag{4.27}$$

evolving from the initial condition $g(z,0) = g_0(z) = \int \mathrm{d}a \frac{\mu(a)}{z-a}$. Using the method of characteristics, see for example Eq. (2.87) the solution may be expressed as a fixed point equation, or equivalently in a parametric form as:

$$g(z,T) = g_0(\xi)\,, \tag{4.28}$$

with

$$\xi = z - T g_0(\xi)\,. \tag{4.29}$$

For z and T fixed and given g_0, one first needs to solve (4.29) for ξ and then inject the solution in (4.28). Conversely, from (4.29) with ξ fixed, one can express z as an implicit function of ξ

$$z(\xi) = \xi + T g_0(\xi)\,. \tag{4.30}$$

The idea would be to eliminate ξ from Eqs. (4.28) and (4.30) to obtain $g(z,T)$ as a function of z for fixed T. However, in practice, this is not always easy, as we will see shortly.

4.3.2 Critical strength

To compute the critical strength in the large N limit, one then needs first to compute the lower edge $\mathrm{b}_-(T)$ of the LSD of the DBM and then compute the critical strength (or time) by solving the solution:

$$\mathrm{b}_-(T_c) = 0\,, \tag{4.31}$$

and we first explain the general method to characterize the edges from the knowledge of the Stieltjes transform and then apply this method to Burger's equation.

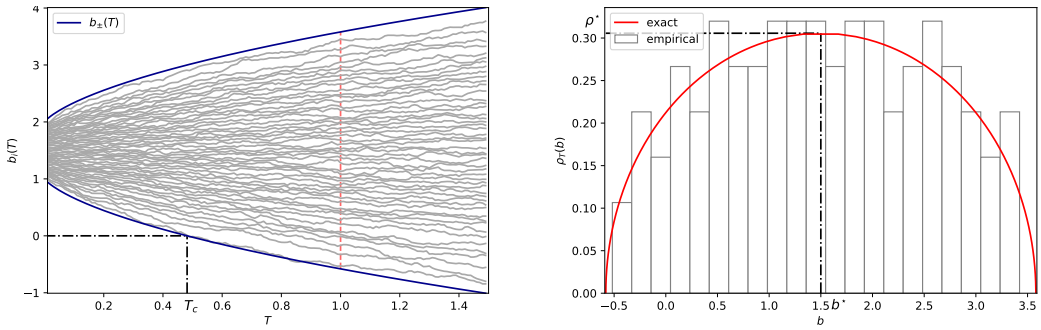


Figure 4.5: (Left) A representation of the DBM described by eq. (4.23) at $\beta = 1$ with the initial flat condition (4.21) for $\sigma = 1$ and $N = 50$. In blue, the limiting curve for the bottom and top edges as $N \to \infty$. The dotted line corresponds to the value of $T = 1$ of the plot of the limiting density on the right. **(Right)** Plot of the limiting density for the flat initial density using the parametric solution (4.68) for $\sigma = T = 1$, compared to a histogram of the positions of the DBM at this time.

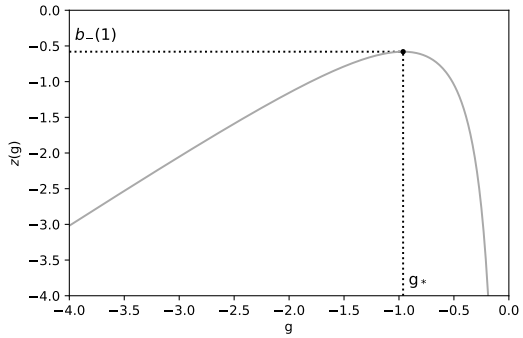


Figure 4.6: Representation of the inverse function $z(\mathrm{g})$ of the Stieltjes transform $g(z,T)$ for $T = \sigma = 1$. If one starts from the origin $g = 0$ and goes to the left, the function $z(\mathrm{g})$ increases until it reaches the point g_* from which one can get the lower edge $\mathrm{b}_-(1) \approx -0.580457$.

Edges from the Stieltjes transform -

The lower and the upper edges $b_\mp(T)$ of the density $\rho_T(b)$ can be extracted from the Stieltjes transform $g(z,T)$ by applying the following general prescriptions.

- First, one needs to compute the inverse Stieltjes transform $g^{\langle -1\rangle}(.)$ of the Stieltjes transform $g(.)$ which we recall is the function satisfying $g^{\langle -1\rangle}(g(z)) = z$ for z large enough. Note that we have suppressed the T dependence of $g(z,T)$ for convenience. Let's recall that for the semi-circle distribution, this inverse has been computed in Eq. (2.51) and is given by $g^{\langle -1\rangle}(\theta) = \theta + 1/\theta$.

- Next, one needs to find the roots g_* solutions of

$$\left(g^{\langle -1\rangle}\right)'(\mathrm{g}_*) = 0\,, \tag{4.32}$$

  where $\left(g^{\langle -1\rangle}\right)'(\theta) := dg^{\langle -1\rangle}(\theta)/d\theta$ is the derivative of the inverse of the Stieltjes. In general, this equation will have multiple roots. For a density confined in a single interval

on the real line, this has typically two roots. For example, for the semi-circular distribution, $\left(g^{\langle -1\rangle}\right)'(\theta) = 1 - 1\theta^2$, which gives two roots $\mathrm{g}_* = -1$ and $\mathrm{g}^* = 1$. The smallest root is $\mathrm{g}_* = -1$ and the largest one is $\mathrm{g}^* = 1$. The solutions g_* of Eq. (4.32) corresponds to the values of the Stieltjes transforms at the edges of the boundaries of the support of the distribution.

- The lower edge $\mathrm{b}_-(T)$ of the support of the density is then given by

$$\mathrm{b}_-(T) = g^{\langle -1\rangle}(\mathrm{g}_*)\,, \tag{4.33}$$

Similarly, the upper edge of the support is given by the other root, i.e., $\mathrm{b}_+(T) = z(\mathrm{g}^*)$. For example, for the semi-circular law, one gets $\mathrm{b}_- = -2$ and $\mathrm{b}_+ = 2$ which indeed are respectively the lower and the upper edge of the support $[-2, 2]$ of the semi-circle distribution.

Lower edge for the DBM and critical strength in the general setting

We now have all the necessary ingredients to compute the critical strength T_c. If for convenience we denote by $z(\theta) \equiv g^{\langle -1\rangle}(\theta)$, The equation $z'(\mathrm{g}) = 0$ is equivalent to

$$\frac{\mathrm{d}z(\xi)}{\mathrm{d}\xi}\frac{\mathrm{d}\xi}{\mathrm{d}g} = 0\,. \tag{4.34}$$

In general the term $\frac{\mathrm{d}\xi}{\mathrm{d}g}$ is non-zero, hence this is equivalent to solve

$$\frac{\mathrm{d}z(\xi)}{\mathrm{d}\xi} = 0\,. \tag{4.35}$$

Using the expression (4.30) for $z(\xi)$, one gets

$$1 + T g_0'\left(\xi_*(T)\right) = 0\,, \tag{4.36}$$

where $\xi_*(T)$ denotes the lowest root of Eq. (4.36). Injecting this $\xi_*(T)$ back into Eq. (4.30) and using (4.33) gives the lower edge

$$\mathrm{b}_-(T) = \xi_*(T) + T\, g_0\left(\xi_*(T)\right)\,, \tag{4.37}$$

where $\xi_*(T)$ is obtained from Eq. (4.36). Finally, setting $\mathrm{b}_-(T_c) = 0$ gives T_c.

This can be summarized in the following result

Result 4.1 (*Stability criterion for the heterogeneous model*)

The probability of stability for May's heterogeneous model with arbitrary initial density $\mu(a)$ is given in the limit $N \to \infty$ by:

$$\mathcal{P}_{\mathrm{stable}}(T, \infty) = \begin{cases} 1 & \text{if } T < T_c\,, \\ 0 & \text{otherwise.} \end{cases} \tag{4.38}$$

where now the critical strength T_c, which implicitly depends on $\mu(a)$, is obtained from the solution of the transcendental equation

$$\xi_*(T_c) + T_c\, g_0\left(\xi_*(T_c)\right) = 0\,, \tag{4.39}$$

with $\xi_*(T_c)$ given by Eq. (4.36).

Our algorithm for determining T_c, for arbitrary initial density $\mu(a)$, thus follows three principal steps:

- Given $\mu(a)$, we first determine the initial Stieltjes transform $g_0(z) = \int \mathrm{d}a \frac{\mu(a)}{z-a}$.
- Once we have $g_0(z)$, we solve Eq. (4.36) and determine $\xi_*(T)$.
- Next we inject this $\xi_*(T)$ in the transcendental equation (4.39) and solve it to determine T_c.

For example, in May's original homogeneous model, we have $\mu(a) = \delta(a-1)$. This gives, $g_0(z) = 1/(z-1)$. Substituting this in Eq. (4.36), we get two roots, and the lowest root gives $\xi_*(T) = 1 - \sqrt{T}$. Substituting this in (4.39) gives $1 - 2\sqrt{T_c} = 0$, and hence $T_c = 1/4$. Our method, outlined above, holds for arbitrary $\mu(a)$ and in the next subsection, we show that for the flat initial condition with $\mu(a)$ given in Eq. (4.22), the general procedure described above can be carried out explicitly, thus providing a nontrivial generalization of May's homogeneous initial condition.

Critical strength for the flat initial condition -

As a nontrivial example, we now consider the flat initial condition with $\mu(a)$ given in Eq. (4.22). In this case, the initial Stieltjes transform is given by:

$$g_0(z) \equiv g_A(z) = \frac{1}{\sigma}\int_1^{1+\sigma} \frac{\mathrm{d}a}{z-a} = \frac{1}{\sigma}\log\left(\frac{z-1}{z-1-\sigma}\right), \tag{4.40}$$

and its derivative is given by

$$g_0'(z) = -\frac{1}{(z-1)(z-1-\sigma)}. \tag{4.41}$$

Using Eq. (4.36), $\xi_*(T)$ satisfies the quadratic equation

$$(\xi_*(T)-1))\,(\xi_*(T)-1-\sigma) = T, \tag{4.42}$$

whose lowest solution is given by

$$\xi_*(T) = 1 + \frac{\sigma}{2} - \frac{\sigma}{2}\sqrt{1+\frac{4T}{\sigma^2}}. \tag{4.43}$$

Using Eq. (4.37), the lower edge at fixed T is given by

$$\mathrm{b}_-(T) = 1 + \frac{\sigma}{2} - \frac{\sigma}{2}\sqrt{1+\frac{4T}{\sigma^2}} + \frac{T}{\sigma}\log\frac{\sqrt{1+\frac{4T}{\sigma^2}}-1}{1+\sqrt{1+\frac{4T}{\sigma^2}}}. \tag{4.44}$$

Setting $\mathrm{b}_-(T_c) = 0$ in Eq. (4.44) gives T_c. However, it is not easy to solve explicitly this transcendental equation. To proceed further, we first write Eq. (4.44) in a more compact form,

$$\mathrm{b}_-(T) = 1 - \sigma h\left(\frac{4T}{\sigma^2}\right), \tag{4.45}$$

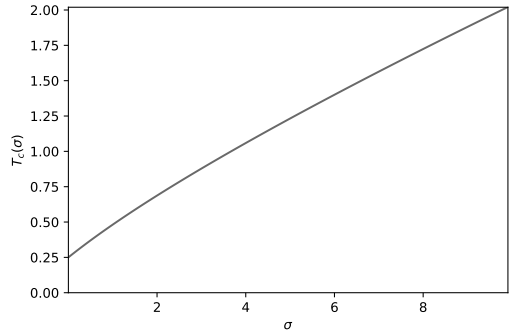


Figure 4.7: Plot of the critical strength T_c given by Eq. (4.49), function of the spread σ.

where the scaling function $h(u)$ is given by

$$h(u) = \frac{\sqrt{1+u}-1}{2} + \frac{u}{2}\log\left(\frac{1+\sqrt{u+1}}{\sqrt{u}}\right). \tag{4.46}$$

This function admits the following asymptotic behaviors near the origin and at infinity:

$$h(u) \sim \begin{cases} \frac{1}{4}\left(1+2\log 2-\log u\right)u & \text{for } u \to 0\,, \\ \\ \sqrt{u}-\frac{1}{2}+\frac{1}{6}u^{-\frac{1}{2}} & \text{for } u \to \infty\,. \end{cases} \tag{4.47}$$

Setting $\mathrm{b}_-(T_c) = 0$ in Eq. (4.45) gives

$$h\left(\frac{4T_c}{\sigma^2}\right) = \frac{1}{\sigma}\,. \tag{4.48}$$

As a result, the critical strength is given in this case by

$$\boxed{T_c(\sigma) = \frac{\sigma^2}{4}u\left(\frac{1}{\sigma}\right)}, \tag{4.49}$$

where $u(h)$ is the inverse function of $h(u)$. Since the function $h(u)$ is explicit in Eq. (4.46), its inverse function $u(h)$ can be easily plotted and hence we can plot T_c in Eq. (4.49) as a function of the spread σ, as shown in Fig. 4.7. The asymptotic behaviors of T_c for small and large σ can also be derived using Eq. (4.47) and are given by

$$T_c(\sigma) \sim \begin{cases} \frac{1+\sigma}{4} & \text{for } \sigma \to 0\,, \\ \\ \frac{\sigma}{\log\left(\frac{\sigma}{4}\right)} & \text{for } \sigma \to \infty\,. \end{cases} \tag{4.50}$$

In particular, we recover as expected the limit $T_c = \frac{1}{4}$ of May's original model for $\sigma \to 0$. In the limit $\sigma \to \infty$, we find $T_c \to \infty$ from Eq. (4.50), which indicates that for large σ, the system is always stable, regardless of the value of the strength parameter T. This is an interesting result that perhaps could not have been guessed a priori.

4.4 Parametric solution for the density with flat initial condition

The goal of this section is to obtain an expression for the limiting density $\rho_T(b)$ of the eigenvalues of the matrix $\mathbf{B}$ for the flat initial condition (4.21), at arbitrary time T. The idea is to rely again on the complex Burgers' equation (4.27) for the Stieltjes transform. As we will see, that the density $\rho_T(b)$ cannot be easily expressed in terms of known analytical functions. However, it can be expressed in an easily plottable parametric form.

We start with the two basic equations satisfied by the Stieltjes transform $g(z,T)$, namely the solution of the complex Burger's equation in Eq. (4.28) and Eq. (4.30). For easy reading, let us re-write these two equations together

$$g(z,T) = g_0(\xi) \tag{4.51}$$

$$z(\xi) = \xi + T\, g_0(\xi)\,. \tag{4.52}$$

The idea is to eliminate the auxiliary variable ξ between these two equations and express g as a function of z, for a fixed T.

To proceed, we start with the initial Stieltjes transform

$$g_0(\xi) = \int \frac{\mu(a)}{\xi - a} \mathrm{d}a\,. \tag{4.53}$$

Suppose we could invert this equation and write ξ as a function of g_0

$$\xi = z_0(g_0(\xi))\,, \tag{4.54}$$

Thus $z_0(.) \equiv g_0^{\langle -1 \rangle}$ is just the inverse function of $g_0(\xi)$ in Eq. (4.53). Substituting (4.51) in Eq. (4.54) gives

$$\xi = z_0(g(z,T))\,. \tag{4.55}$$

Using this relation in Eq. (4.52) and further using $g_0(\xi) = g(z,T)$, Eq. (4.52) reduces to

$$z = T\, g(z,T) + z_0(g(z,T))\,. \tag{4.56}$$

Thus, for fixed T, if we know the initial inverse function $z_0(.)$, we have, in principle, a closed equation for $g(z,T)$. From the expression (4.40) of the initial Stieltjes transform g_0 in the flat initial condition case, its inverse function $z_0(g)$ is given by:

$$z_0(g) = 1 + \sigma + \frac{\sigma}{\mathrm{e}^{\sigma g} - 1}\,, \tag{4.57}$$

Substituting this in Eq. (4.56), we then have a closed equation for the Stieltjes transform $g(z,T)$ at any time T

$$z = T\, g(z,T) + 1 + \sigma + \frac{\sigma}{\mathrm{e}^{\sigma g(z,T)} - 1}\,. \tag{4.58}$$

Solving explicitly $g(z,T)$ from this transcendental equation does not seem feasible, unfortunately. To derive the density $\rho_T(b)$ from this Stieltjes transform $g(z,T)$ using Eq. (1.28),

we set $z = b - \mathrm{i}0^+$, with b between the two edges $\mathrm{b}_\pm(T)$. Then by (1.28) we have $g(b - \mathrm{i}0^+, T) = u + \mathrm{i}\pi\rho$, where u is the real part of the Stieltjes transform. For simplicity, we have used the shorthand notation $u \equiv u(b,T)$ and $\rho \equiv \rho_T(b)$. Identifying the real and the imaginary parts of (4.58), we get a pair of coupled equations

$$\begin{cases} b &= 1 + \sigma + Tu + \sigma\, \mathfrak{Re}\left[\dfrac{1}{\mathrm{e}^{\sigma(u+\mathrm{i}\pi\rho)} - 1}\right], \\ 0 &= T\pi\rho + \sigma\, \mathfrak{Im}\left[\dfrac{1}{\mathrm{e}^{\sigma(u+\mathrm{i}\pi\rho)} - 1}\right]. \end{cases} \tag{4.59}$$

One can multiply the numerator and denominator inside the brackets by $\mathrm{e}^{\sigma(u-\mathrm{i}\pi\rho)} - 1$, to get the real and imaginary parts of the function inside the brackets, and the system can then be written as

$$\begin{cases} b &= 1 + \sigma + Tu + \sigma \dfrac{\cos(\sigma\pi\rho)\mathrm{e}^{\sigma u} - 1}{\mathrm{e}^{2\sigma u} - 2\cos(\pi\sigma\rho)\mathrm{e}^{\sigma u} + 1}, \\ 0 &= T\pi\rho - \sigma \dfrac{\sin(\pi\sigma\rho)\mathrm{e}^{\sigma u}}{\mathrm{e}^{2\sigma u} - 2\cos(\pi\sigma\rho)\mathrm{e}^{\sigma u} + 1}. \end{cases} \tag{4.60}$$

Ideally, the goal would be to eliminate u from these pair of equations and express $\rho \equiv \rho_T(b)$ as a function of b, for fixed T.

Let us first consider the simple case of May's homogeneous model, i.e., the limit $\sigma \to 0$. In this limit, Eq. (4.60) reduces to

$$\begin{cases} b &= 1 + Tu + \frac{u}{u^2 + \pi^2\rho^2} \\ 0 &= T\pi\rho - \frac{\pi\rho}{u^2 + \pi^2\rho^2}. \end{cases} \tag{4.61}$$

Eliminating u from these pair of equations, one immediately gets the shifted Wigner semi-circular density

$$\rho_T(b)|_{\sigma=0} = \frac{1}{2\pi T}\sqrt{4T - (b-1)^2}, \tag{4.62}$$

supported over the interval $b \in \left[1 - 2\sqrt{T}, 1 + 2\sqrt{T}\right]$. Thus, in May's homogeneous model, starting from the initial condition $\mu(a) = \delta(a-1)$, the density of eigenvalues $b_i(T)$'s, at any time $T > 0$, is of the shifted Wigner semi-circular form in Eq. (4.62).

For general $\sigma > 0$, eliminating u from Eq. (4.60) and expressing $\rho_T(b)$ explicitly (as in the $\sigma = 0$ case) seems difficult. Instead, for a general $\sigma > 0$, one can obtain the solution parametrically as follows. We note that the top equation of (4.60) is a parametric expression for $b(u,\rho)$. The idea is to eliminate the dependency on u by working a bit on the bottom equation of (4.60). To do so, let us denote by $w = \mathrm{e}^{\sigma u}$, and then from the bottom equation of (4.60) w satisfies a quadratic equation,

$$\frac{w^2}{2} - w\left(\frac{\sigma^2}{2T}\mathrm{sinc}(\pi\sigma\rho) + \cos(\pi\sigma\rho)\right) + \frac{1}{2} = 0, \tag{4.63}$$

where $\mathrm{sinc}(x) = \frac{\sin(x)}{x}$ is the standard *sinus cardinal* function. Let us introduce further the function

$$f_{\sigma,T}(\rho) := \frac{\sigma^2}{2T}\mathrm{sinc}\left(\pi\sigma\rho\right) + \cos\left(\pi\sigma\rho\right), \tag{4.64}$$

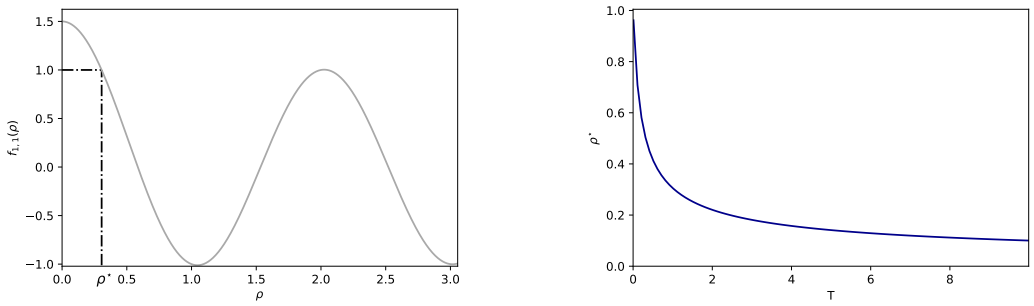


Figure 4.8: (Left) Plot of $f_{\sigma,T}(\rho)$ given by Eq. (4.64), function of ρ for $\sigma = T = 1$. Only the part on the left of the value $\rho^\star \approx 0.305637$ contributes to the parametric solution of the density. **(Right)** Plot of the maximum $\rho^\star$ described by Eq. (4.70) function of T, for $\sigma = 1$.

plotted in Fig. 4.8 (Left).

Let us now imagine that the value of ρ is fixed. Then the two solutions $w_\pm(\rho)$ of the system (4.63) are given in terms of this function $f_{\sigma,T}(\rho)$ by,

$$w_\pm(\rho) = f_{\sigma,T}(\rho) \pm \sqrt{f_{\sigma,T}(\rho)^2 - 1}\,, \tag{4.65}$$

and they satisfy the symmetry relation

$$\frac{1}{w_-(\rho)} = w_+(\rho)\,. \tag{4.66}$$

Injecting this into the top equation of (4.59) we get two solutions $b_\pm(\rho)$:

$$b_\pm(\rho) = 1 + \frac{\sigma}{2} + \frac{T}{\sigma}\log w_\pm(\rho) + \frac{T}{2\sigma\,\mathrm{sinc}(\pi\sigma\rho)}\left(w_\pm - \frac{1}{w_\pm}\right)\,. \tag{4.67}$$

Using the symmetry relation (4.66) and the expression (4.65) for $w_\pm(\rho)$, we get the following result:

Result 4.2 (*parametric solution for the DBM with flat initial condition*)

In the large N limit, the LSD $\rho \equiv \rho_T$ of the DBM of Eq. (4.23) with the flat initial condition of Eq. (4.21) admits a parametric solution in the form

$$b_\pm(\rho) = 1 + \frac{\sigma}{2} \pm \frac{T}{\sigma}\left(\log\left(f_{\sigma,T}(\rho) + \sqrt{f_{\sigma,T}(\rho)^2 - 1}\right) + \frac{\sqrt{f_{\sigma,T}(\rho)^2 - 1}}{\mathrm{sinc}\,(\pi\sigma\rho)}\right)\,. \tag{4.68}$$

where $f_{\sigma,T}$ is given by Eq. (4.64).

In Fig. 4.9, we plot the two branches $b_\pm(\rho)$ as a function of ρ for fixed T. Indeed, if one rotates this plot anticlockwise by $\pi/2$ and then reflects around the vertical axis, one gets the desired density $\rho_T(b)$ as a function of b, as seen in Fig. 4.5 (Right). Apart from being able to plot the density, one can also extract a few additional information from the explicit expression in Eq. (4.68), as discussed below.

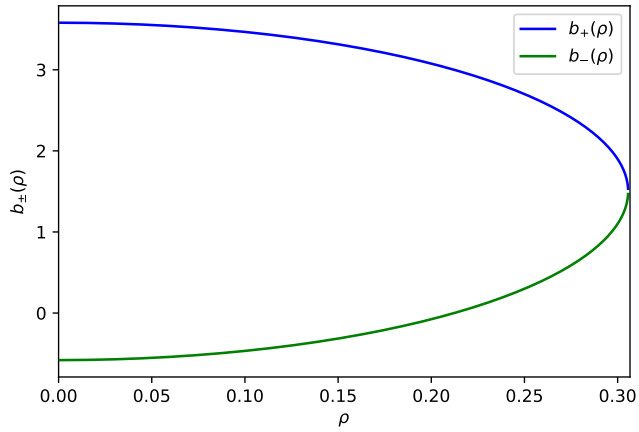


Figure 4.9: Plot of the two inverse functions $b_\pm(\rho)$ given by Eq. (4.67), for $\sigma = T = 1$. Rotating the figure anticlockwise, followed by a reflection around the vertical axis gives $\rho_T(b)$ as a function of b, as shown in Fig. 4.5 (Right).

Maximum value of the density -

From Eq. (4.68), we see that the maximum value of the density $\rho^\star$ is attained at the point $b^\star$ for which $b_+(\rho) = b_-(\rho)$, i.e., $b^\star = 1 + \frac{\sigma}{2}$. The value of the maximum of the density $\rho^\star = \rho_T\,(b^\star)$, is therefore given as the *first positive* solution of

$$f_{\sigma,T}(\rho^\star) = 1\,, \tag{4.69}$$

which using Eq. (4.64) is equivalent to finding the first positive solution of

$$\rho^\star \tan\left(\frac{\sigma\pi\rho^\star}{2}\right) = \frac{\sigma}{2T\pi}\,. \tag{4.70}$$

A plot of the maximum $\rho^\star$ as a function of T for $\sigma = 1$ is given in Fig. 4.8 (Right).

Behavior near the edges -

By Taylor expanding the function $b_\pm(\rho)$ near $\rho = 0$, we have

$$b_\pm(\rho) \sim \mathrm{b}_\pm(T) \mp \frac{T\pi^2}{2}\sqrt{\sigma^2 + 4T}\,\rho^2 \quad \text{as } \rho \to 0\,, \tag{4.71}$$

where the edges $\mathrm{b}_\pm(T)$ are given by

$$\mathrm{b}_\pm(T) = 1 + \frac{\sigma}{2} \pm \left[\frac{\sigma}{2}\sqrt{1 + \frac{4T}{\sigma^2}} + \frac{T}{\sigma}\log\left(1 + \frac{\sigma^2}{2T} + \frac{\sigma^2}{2T}\sqrt{1 + \frac{4T}{\sigma^2}}\right)\right]\,. \tag{4.72}$$

It is easy to verify that the expression for $\mathrm{b}_-(T)$ coincides with Eq. (4.44). Inverting the relation in Eq. (4.71), one finds that the density vanishes as a square root near the edge, with a prefactor that can be computed explicitly

$$\rho_T(b) \sim \frac{1}{\pi}\sqrt{\frac{2}{T}}\frac{1}{(\sigma^2 + 4T)^{\frac{1}{4}}}\sqrt{(\mathrm{b}_\pm(T) - b)_+} \quad \text{as } b \to \mathrm{b}_\pm(T)\,. \tag{4.73}$$

where $(x)_+$ is equal to x for $x > 0$ and 0 otherwise.

4.5 The deformed GUE with flat initial condition and the left large deviation function of the dynamical system

So far, we have computed the average density of eigenvalues $\rho_T(b)$ in the large N limit of the relaxation matrix, $\mathbf{B} = \mathbf{A} + \sqrt{T}\,\mathbf{J}$ for any T, where J is a $N \times N$ GOE matrix and $\mathbf{A}$ is diagonal with positive entries drawn from a flat distribution over $[1, 1+\sigma]$ with width σ. This gives us the exact T_c between the stable to unstable transition. We expect that for finite but large N, the stability probability will have qualitatively similar behavior as in the homogeneous model in Eq. (4.17), see also Fig. 4.2:

$$\mathcal{P}_{\rm stable}(T,N) \approx \begin{cases} \exp\left[-\frac{N^2}{2}\Phi_+(\sigma,T) + o(N^2)\right] & \text{for } T > T_c \text{ and } |T-T_c| \sim \mathcal{O}(1)\,, \\ \mathcal{F}^{(1)}\left(\gamma N^{2/3}\left(T^{-1/2} - T_c^{-1/2}\right)\right) & \text{for } |T-T_c| \sim o(N^{-\frac{2}{3}})\,, \\ 1 - \exp\left[-\frac{N}{2}\Phi_-(\sigma,T) + o(N)\right] & \text{for } T < T_c \text{ and } |T-T_c| \sim \mathcal{O}(1)\,, \end{cases} \tag{4.74}$$

where γ is a constant of order one and the large deviation functions $\Phi_\pm(\sigma,T)$ on either side of T_c would be different. It turns out (see later) that to compute the large deviation functions $\Phi_\pm(\sigma,T)$, we need the information on the full joint distribution of eigenvalues, and not just the one-point function, i.e, the average density.

Hence, our next natural step was to see if we could compute the joint distribution of the eigenvalues of $\mathbf{B}$, where $\mathbf{J}$ is a GOE matrix. For this $\beta = 1$ case, one does not have a closed formula for the joint density. However, it turns out that one can compute the joint distribution of eigenvalues in the Hermitian counterpart of the relaxation matrix, $\tilde{\mathbf{B}} = \mathbf{A} + \sqrt{T}\tilde{\mathbf{J}}$, where $\mathbf{A}$ is still diagonal with a flat distribution, but now $\mathbf{J}$ is Hermitian, i.e., a GUE matrix.

In this section, for this deformed GUE model, we derive an explicit formula for the joint law of eigenvalues for the flat initial condition, thanks to the Itzykson-Zuber determinantal formula. We will see that this leads to a new Coulomb gas, where the eigenvalues can be interpreted as the positions of a gas of particles confined in a harmonic potential and repelling pairwise as in the standard GUE, but with an additional twist that the pairwise interaction here is a *linear combination* of a logarithmic (as in standard GUE) and a log-sinh type interaction. Finally, using this Hermitian modification, we will show how to compute at least the large deviation function $\Phi_-(\sigma,T)$ appearing in Eq. (4.74), in the 'weakly stable' phase ($T < T_c$) in the original deformed GOE model. However, computing the large deviation function $\Phi_+(\sigma,T)$ on the 'strongly unstable' phase ($T > T_c$) still remains out of reach.

4.5.1 The deformed GUE with flat initial condition and its joint law for the eigenvalues

The deformed GUE model is the Hermitian counterpart of the deformed GOE model

$$\tilde{\mathbf{B}} = \mathbf{A} + \sqrt{T}\tilde{\mathbf{J}}\,, \tag{4.75}$$

with the matrix $\mathbf{A} = \mathbf{Diag}(a_1, \ldots, a_N)$ with positive entries as before. The matrix $\tilde{\mathbf{J}}$ is now a GUE matrix whose law is given by Eq. (1.39) with $\beta = 2$.

As we have seen in Chapter 2, the large N limit of the DBM is independent of the Dyson index $\beta > 0$ and is completely characterized by the Burgers' equation (4.27) for the Stieltjes transform for any arbitrary initial density $\mu(a)$ of the a_i's. In particular, for the flat initial condition, this means that the large N limit of the deformed GUE model has the LSD given by the parametric solution of Eq. (4.68).

Furthermore in Chapter 2 (see Eq. (2.77)), we have also seen that the joint law of the eigenvalues $\tilde{\boldsymbol{b}}$ of the deformed GUE can be explicitly expressed in terms of the ($\beta = 2$) additive spherical integral, that is:

$$\mathcal{P}_T^{(\beta=2)}\left(\tilde{\boldsymbol{b}}\middle|\boldsymbol{a}\right) \propto e^{-\frac{N}{2T}\sum_{i=1}^N \tilde{b}_i^2}\,\Delta(\tilde{\boldsymbol{b}})^2\,\mathcal{I}^{(\beta=2)}\left(\boldsymbol{a}, N/T\cdot\tilde{\boldsymbol{b}}\right)\,, \tag{4.76}$$

where we recall that $\Delta(\tilde{\boldsymbol{b}}) = \Delta(\tilde{b}_1, \ldots \tilde{b}_N) = \prod_{i<j}(\tilde{b}_j - \tilde{b}_i)$ is the *Vandermonde product* and $\mathcal{I}^{(\beta=2)}$ is the additive spherical integral of Eq. (3.1) over the unitary group $\mathsf{O}_{\beta=2}(N) \equiv \mathsf{U}(N)$. In Chapter 3, we have also seen that in this $\beta = 2$ case, the spherical integral admits the determinantal formula of Eq. (3.22) which allows us to write the joint density explicitly in terms of several determinants:

$$\mathcal{P}_T^{(\beta=2)}\left(\tilde{\boldsymbol{b}}\middle|\boldsymbol{a}\right) \propto e^{-\frac{N}{2T}\sum_{i=1}^N \tilde{b}_i^2}\Delta(\tilde{\boldsymbol{b}})\,\det\left[e^{\frac{N}{T}a_i\tilde{b}_j}\right]_{1\le i,j\le N}\,. \tag{4.77}$$

At this stage, the equation (4.77) holds for an arbitrary diagonal matrix $\mathbf{A} = \mathbf{Diag}(a_1, \ldots, a_N)$. Let us now take the a_i's to be given by the flat initial condition (4.21). In this case, the determinant appearing in (4.77) considerably simplifies since

$$\det\left[e^{\left(\frac{N}{T}+\frac{\sigma(i-1)}{T}\right)\tilde{b}_j}\right] = \mathrm{e}^{\frac{N}{T}\sum_{i=1}^N \tilde{b}_i}\Delta\left(e^{(\sigma/T)\,\tilde{\boldsymbol{b}}}\right) = \mathrm{e}^{\frac{N}{T}\sum_{i=1}^N \tilde{b}_i}\prod_{i<j}\left(e^{\frac{\sigma}{T}\tilde{b}_j} - e^{\frac{\sigma}{T}\tilde{b}_i}\right)\,. \tag{4.78}$$

Hence, the joint law for the ordered eigenvalues (4.77) simplifies to

$$\mathcal{P}_T^{(\beta=2)}(\tilde{\boldsymbol{b}}) \propto \exp\left\{\sum_{i=1}^N \frac{N}{T}\left(-\frac{\tilde{b}_i^2}{2} + \tilde{b}_i\right)\right\}\,\Delta(\tilde{\boldsymbol{b}})\,\Delta\left(e^{\frac{\sigma}{T}\,\tilde{\boldsymbol{b}}}\right)\,, \tag{4.79}$$

Next, using the identity

$$(\mathrm{e}^x - \mathrm{e}^y)\,\mathrm{e}^{-\frac{(x+y)}{2}} = 2\,\sinh\left(\frac{x-y}{2}\right)\,, \tag{4.80}$$

we can write the second Vandermonde in Eq. (4.79) as

$$\Delta\left(e^{\frac{\sigma}{T}\tilde{\boldsymbol{b}}}\right) \propto \exp\left\{\left(\frac{\sigma}{2T}\sum_{i\neq j}(b_i+b_j)\right)+\frac{1}{2}\sum_{i\neq j}\log\sinh\left(\frac{\sigma}{2T}|b_i-b_j|\right)\right\}, \tag{4.81}$$

$$\Delta\left(e^{\frac{\sigma}{T}\tilde{\boldsymbol{b}}}\right) = \exp\left\{\frac{\sigma(N-1)}{2T}\cdot\sum_{i} b_i+\frac{1}{2}\sum_{i\neq j}\log\sinh\left(\frac{\sigma}{2T}|b_i-b_j|\right)\right\}. \tag{4.82}$$

Using Eq. (4.79) and completing the square, this induced the following final expression for the joint law:

Result 4.3 (*Joint distribution for deformed GUE with flat initial condition*)

The joint distribution of the eigenvalues $\tilde{\boldsymbol{b}}$ of the deformed GUE *model of Eq.* (4.75) *under the initial flat condition of Eq.* (4.21) *for the matrix $\mathbf{A}$ is given by*

$$\mathcal{P}_T^{(\beta=2)}(\tilde{\boldsymbol{b}}) \propto \exp\left\{-N\sum_{i=1}^{N}\frac{\left(\tilde{b}_i-b_N^*\right)^2}{2T}+\frac{1}{2}\sum_{i\neq j}\left(\log|\tilde{b}_i-\tilde{b}_j|+\log\sinh\left(\frac{\sigma}{2T}|\tilde{b}_i-\tilde{b}_j|\right)\right)\right\} \tag{4.83}$$

where $b_N^* := 1+(\sigma/2)(N-1)/N$.

Eq. (4.83) provides a nice Coulomb gas interpretation of the joint law of eigenvalues. The joint distribution in Eq. (4.83) can be written as a Boltzmann distribution $\sim e^{-E(\{\tilde{b}_i\})}$, where the energy function can be read off the argument of the exponential in Eq. (4.83). The eigenvalues $\{\tilde{b}_i\}$'s can be interpreted as the positions of N charges on a line. These charges are subjected to an external harmonic potential centered at $b_N^* = 1+(\sigma/2)(N-1)/N$. In addition, they repel each other pairwise. The pairwise interaction is a linear combination of the logarithmic repulsion (represented by the second term inside the exponential in Eq. (4.83)) and a log-sinh interaction (the third term in Eq. (4.83)). In the limit $\sigma \to 0$ (upon absorbing an overall constant in the normalization), the third term also becomes logarithmic, and hence the system reduces to the standard log-gas of Gaussian random matrices [67]. But for a nonzero $\sigma > 0$, we have a new variety of Coulomb gas with both log and log-sinh interactions that is usually not encountered in RMT models.

Given the joint density of the eigenvalues in the Coulomb gas representation in Eq. (4.83), one can, in principle, obtain the average density in the large N limit by a variational principle, i.e., by employing a saddle point method for large N to evaluate the partition function of the Coulomb gas. This amounts to minimizing the energy function $E(\{\tilde{b}_i\})$. Minimizing this energy in Eq. (4.83) gives the saddle point equation

$$\frac{1}{T}\left(1+\frac{\sigma}{2}\frac{N-1}{N}-\tilde{b}\right)+\frac{1}{N}\sum_{j:j\neq i}\frac{1}{\tilde{b}_i-\tilde{b}_j}+\frac{\sigma}{T}\frac{1}{2N}\sum_{j:j\neq i}\coth\left(\frac{\sigma}{2T}(\tilde{b}_i-\tilde{b}_j)\right)=0. \tag{4.84}$$

For large N, the sums can be replaced by integrals, and one obtains an integral equation satisfied by the density $\rho_T(\tilde{b})$

$$\frac{1}{T}(b^\star-\tilde{b})+\fint\frac{\rho_T(b')}{\tilde{b}-b'}\,\mathrm{d}b'+\frac{\sigma}{2T}\fint\rho_T(b')\,\coth\left(\frac{\sigma}{2T}(\tilde{b}-b')\right)\,\mathrm{d}b'=0, \tag{4.85}$$

where we recall $b^\star = 1+\sigma/2$, and this integral equation holds for all $\tilde{b} \in [\mathrm{b}_-(T), \mathrm{b}_+(T)]$ where $\mathrm{b}_\pm(T)$ denotes the support edges.

In the limit $\sigma \to 0$, the third term coincides with the second term in Eq. (4.85), and one recovers the standard saddle point density of the log-gas [67, 120],

$$\frac{1}{2T}(b^\star - \tilde{b}) + \fint \frac{\rho_T(\tilde{b}')}{\tilde{b}-\tilde{b}'}\,\mathrm{d}\tilde{b}' = 0\,. \tag{4.86}$$

This singular value integral equation can be inverted using Tricomi's formula (see Ref. [120] for details) and one recovers the shifted semi-circular law in Eq. (4.62). For a nonzero σ, we were not able to solve the singular integral equation (4.85). However, remarkably, we actually know the solution $\rho_T(\tilde{b})$, albeit in a parametric form, in Eq. (4.68) via the Stieltjes transform method. Note that the parametric solution in Eq. (4.68) also holds for deformed GUE $\rho_T(\tilde{b})$ which is identical to that of deformed GOE, as shown earlier. It then remains a mathematical challenge to derive this parametric solution (4.68) directly from the singular value integral equation (4.85).

4.5.2 Relations to other models

The matrix $\tilde{\mathbf{B}}$ (and the matrix $\mathbf{B}$ of the original model) as described in the previous section is related to several models of RMT that have appeared before in the literature. The joint density for the matrix $\tilde{\mathbf{B}}$ in Eq. can be written as

$$\mathcal{P}_N(\tilde{\mathbf{B}})\mathrm{d}\tilde{\mathbf{B}} \propto e^{-N\mathrm{Tr}\left[V(\tilde{\mathbf{B}})-\tilde{\mathbf{A}}\tilde{\mathbf{B}}\right]}\mathrm{d}\tilde{\mathbf{B}}\,, \tag{4.87}$$

with $V(x) = \frac{x^2}{2}$ and $\tilde{\mathbf{A}} = \frac{\mathbf{A}}{T}$. The matrix $\tilde{\mathbf{A}}$ in (4.87) plays the role of an external field, and hence models of the type (4.87) are known as random matrices with an external source [34]. A particular interest has been devoted to the case where one half of the eigenvalues of the matrix $\tilde{\mathbf{A}}$ takes the value a and the other half takes the value $-a$, see [32, 28, 5, 29]. The local properties for the case of flat initial condition (4.21) have also been studied in [40] using Riemann-Hilbert techniques.

From Eq. (4.79), one can see that the joint law of eigenvalues exhibits a bi-orthogonal structure of a determinantal point process which resembles somewhat the Muttalib-Borodin ensemble with parameter $\theta > 0$ [147, 30]

$$\mathcal{P}^{(\theta)}_{\mathrm{MB}}(\boldsymbol{\lambda}) \propto \exp\left[-N\sum_{i=1}^{N} V(\lambda_i)\right] \Delta(\boldsymbol{\lambda})\,\Delta\left(\boldsymbol{\lambda}^\theta\right)\,. \tag{4.88}$$

with the difference that in the second Vandermonde, the arguments are exponential in (4.79), while they have a power-law form in (4.88). However, the case with the exponential function in the second Vandermonde, appeared in the randomized multiplicative Horn problem [190], in the DPMK equation for transport in semiconductors [18] and in the multiplicative analog of Dyson Brownian Motion [94].

If one makes the change of variable $x_i = \frac{1}{\sqrt{T}}\left(\tilde{b}_i - (1+\frac{\sigma}{2})\frac{N-1}{N}\right)$ in Eq. (4.83) and writes $r = \frac{\sigma}{2\sqrt{T}}$, the joint distribution of the x_i's is given by:

$$\mathcal{P}(\boldsymbol{x}) \propto \exp\left[-\frac{N}{2}\sum_{i=1}^{N} x_i^2 + \frac{1}{2}\sum_{i\neq j}\log|x_j - x_i| + \frac{1}{2}\sum_{i\neq j}\log\sinh\left(r|x_j - x_i|\right)\right] . \tag{4.89}$$

Thus, we have a Coulomb gas where the pairwise interaction is a linear combination of logarithmic and log-sinh. The case with only log-repulsion (without the log-sinh) corresponds to the standard Gaussian matrices. The case with only log-sinh repulsion (without the log term) appears in the partition function of the Chern-Simons model on S^3 [132, 133], in the theory of Stieltjes-Wigert polynomials [53, 171, 166, 168] and in the recent study of vicious walkers constrained at both ends by a flat initial conditions [76]. The parameter $r = \sigma/\sqrt{4T}$ in (4.89) controls the strength of the second interaction, since for r positive, the function $\log\sinh(r)$ is increasing from 0 to ∞. In the limit $r \to 0$, it reduces to the log-gas as shown before. In the opposite limit $r \to \infty$, Eq. (4.89) to leading order in r reduces to a 1D-one component plasma (OCP) model [111, 154, 17]

$$\mathcal{P}(\boldsymbol{x}) \propto \exp\left[-\frac{N}{2}\sum_{i=1}^{N} x_i^2 + \frac{1}{2}\sum_{i\neq j}|x_j - x_i|\right] , \tag{4.90}$$

for which the equilibrium measure is the flat distribution and the distribution of its largest (lowest) eigenvalue have recently been computed exactly, both for typical fluctuations and also for large deviations [50, 51, 65].

4.5.3 Large deviation below the critical strength T_c for the flat initial condition

We now go back to the original deformed GOE model with flat initial condition (4.21). In the strict $N \to \infty$ limit, the probability of stability $\mathcal{P}_{\rm stable}(N \to \infty, T)$ follows the step function behavior as in Eq. (4.38). We have computed the exact T_c and also the LSD of particles in a parametric form (4.68) for the flat initial condition (4.21). As we have discussed in the introduction, the next step is to derive the behavior of the probability $\mathcal{P}_{\rm stable}(N, T)$ for large but finite N, close to the critical point $T = T_c$. Similar to May's original homogeneous model in Eq. (4.17), one can show [109] that the typical 'small' fluctuations of $\mathcal{O}(N^{-2/3})$ around $T = T_c$, are again described by the Tracy-Widom distribution. This is the middle equation of Eq. (4.74) where the constant γ in Eq. (4.74) is given in [109]. For $\sigma > 0$, The large deviation functions $\Phi_{\pm}(\sigma, T)$ are expected to be different from the homogeneous model $\Phi_{\pm}(\sigma = 0, T) = \Phi_{\pm}(T) = \Psi_{\pm}\left(\frac{1}{\sqrt{T}}\right)$, with $\Psi_{\pm}$ are given by Eqs. (1.108) and (1.126). For values of $T > T_c$ (see Fig. 4.5 (Left)) a finite fraction of the eigenvalues are negative and as explained in the introduction, to access the large deviation regime one needs to push all those eigenvalues leading to a modification of the equilibrium density in the bulk. For the matrix $\mathbf{B}$, the eigenvalues do not behave as a simple 2D Coulomb-gas particles confined on the real line and therefore this equilibrium density in the presence of a pushing wall, needed for the computation of the large deviation function $\Phi_{\pm}(\sigma, T)$ in this regime, is hard to obtain. For this reason, we restrict the discussion only to the weakly stable phase, corresponding to $T < T_c$,

where the bulk density remains unchanged when one pulls a single charge out of the bulk and is still given by ρ_T. To access the large deviation function $\Phi_-(\sigma, T)$ in this regime, we recall using Eq. (4.10) that one has to compute

$$\mathcal{P}_{\rm stable}(T,N) = 1 - {\rm Prob}\left[b_1 < 0\right] , \tag{4.91}$$

where to ease notation, we simply write $b_i \equiv b_i(T)$ in the rest of this section, and the eigenvalues $\{b_i\}$ are in increasing order. To evaluate this probability, we will redo a similar computation as the one in Sec. 4.5.1 to obtain the joint law for the eigenvalues $(b_1, \dots, b_N)$. The main difference with the Hermitian case is that the joint law will involve the $\beta = 1$ HCIZ integral. Instead of the $\beta = 2$ HCIZ integral, there is no simple determinantal formula for the $\beta = 1$ case. It will be possible to overcome this difficulty thanks to the known 'full-rank' asymptotic of the HCIZ integral derived in Chapter 3, Sec. 3.3.1, and we can then compute the probability by integrating the joint law over all eigenvalues and then use of a standard saddle-point approximation.

The laws of the eigenvalues $\boldsymbol{b}$ of the original deformed GOE model of Eq. (4.5) is given by:

$$\mathcal{P}_T^{(\beta=1)}\left(\boldsymbol{b}\middle|\boldsymbol{a}\right) \propto e^{-\frac{N}{4T}\sum_{i=1}^N b_i^2}\, \Delta(\boldsymbol{b})^2\, \mathcal{I}^{(\beta=1)}\left(\boldsymbol{a}, N/(2T)\cdot \boldsymbol{b}\right) , \tag{4.92}$$

where now $\mathcal{I}^{(\beta=1)}$ is the $\beta = 1$ HCIZ integral. There is no simple Itzykson-Zuber formula in this case, but since we are interested in the large N limit, what one only needs is the asymptotic behavior of this integral. For large N and $\beta = 1, 2$, this integral behaves as:

$$\mathcal{I}^{(\beta)}\left(\boldsymbol{a}, \frac{N\beta}{2}\cdot\boldsymbol{b}\right) \approx \exp\left\{\frac{N^2\beta}{2}\mathcal{F}(a_1,\dots,a_N;b_1,\dots,b_N) + o(N^2)\right\} , \tag{4.93}$$

where the function $\mathcal{F}(.)$ has been derived in Chapter 3, see Res. 3.7 . The important point is that this function $\mathcal{F}(.)$ does not depend on β and the β dependence appears just as a prefactor of $\mathcal{F}(.)$ in Eq. (4.93). Thus, for N large, the joint law for the Deformed GOE is asymptotically given by:

$$\mathcal{P}_T^{(\beta=1)}\left(\boldsymbol{b}\middle|\boldsymbol{a}\right) \propto \exp\left[-\frac{N^2}{2}\left(\frac{1}{N}\sum_{i=1}^N \frac{b_i^2}{2T} - \frac{1}{N^2}\sum_{j\neq i}^N \log|b_i - b_j| \right.\right. \\ \left.\left. -\mathcal{F}\left(\frac{a_1}{T},\dots,\frac{a_N}{T};b_1,\dots,b_N\right) + o(1)\right)\right] . \tag{4.94}$$

Similarly, for the deformed GUE, the joint law can be written as:

$$\mathcal{P}_T^{(\beta=2)}\left(\tilde{\boldsymbol{b}}\middle|\boldsymbol{a}\right) \propto \exp\left[-N^2\left(\frac{1}{N}\sum_{i=1}^N \frac{\tilde{b}_i^2}{2T} - \frac{1}{N^2}\sum_{j\neq i}^N \log|\tilde{b}_i - \tilde{b}_j| \right.\right. \\ \left.\left. -\mathcal{F}\left(\frac{a_1}{T},\dots,\frac{a_N}{T};\tilde{b}_1,\dots,\tilde{b}_N\right) + o(1)\right)\right] . \tag{4.95}$$

Comparing Eq. (4.83) and Eq. (4.95), one gets for the flat initial case (4.21) that the function

$\mathcal{F}$ is asymptotically given by:

$$\mathcal{F}\left(\frac{a_1}{T},\dots,\frac{a_N}{T};b_1,\dots,b_N\right) \approx \sum_{i=1}^{N}\frac{b_i b^\star}{2T} + \frac{1}{2N^2}\sum_{j\neq i}\log\sinh\left(\frac{\sigma}{2T}|b_i-b_j|\right) - \frac{1}{2N^2}\sum_{j\neq i}^{N}\log|b_i-b_j| + C + o(1)\,, \quad (4.96)$$

where C is a constant independent of the $\{b_i\}$ and $b^\star = 1+\frac{\sigma}{2}$. Discarding sub-leading term in N in Eq. (4.94) with $\mathcal{F}$ given by Eq. (4.96), one has **asymptotically**:

$$\mathcal{P}_T^{(\beta=1)}(\boldsymbol{b}) \propto \exp\left[-\frac{N^2}{2}\left(\frac{1}{2NT}\sum_{i=1}^{N}(b_i-b^\star)^2 - \frac{1}{2N^2}\sum_{j\neq i}\log|b_i-b_j| - \frac{1}{2N^2}\sum_{j\neq i}\log\sinh\frac{\sigma}{2T}|b_i-b_j| + o(1)\right)\right]. \quad (4.97)$$

From this (asymptotic) behavior, one can deduce the following result for the behavior of the bottom eigenvalue (the result for the top eigenvalue is obtained by symmetry)

Result 4.4 (*LDP for flat initial condition*)

For large N and $w < b_-(T)$ the bottom eigenvalue of the deformed GOE model with flat initial condition satisfy a large deviation principle with speed $N/2$, $\mathbb{P}\left[b_N \simeq w\right] \sim \mathrm{e}^{-N/2\cdot\Psi_{\sigma,T}(w)}$ and the rate function $\Psi_{\sigma,T}(w)$ is given by:

$$\Psi_{\sigma,T}(w) = \frac{(w-b^\star)^2}{2T} - \frac{1}{2}\left(\int_{\mathrm{b}_-(T)}^{\mathrm{b}_+(T)}\left[\log(b-w)+\log\sinh\left(\frac{\sigma}{2T}(b-w)\right)\right]\rho_T(b)\mathrm{d}b\right) - A, \quad (4.98)$$

with the constant A is chosen such that $\Psi_{\sigma,T}(\mathrm{b}_-(T)) = 0$.

The cumulative distribution is given by

$$\mathrm{Prob}\left[b_1<0\right] \approx \int_{-\infty}^{0}\exp\left\{-\frac{N}{2}\Psi_{\sigma,T}(b_1)+o(N)\right\}\mathrm{d}b_1\,, \quad (4.99)$$

and the function $\Psi_{\sigma,T}(w)$ is decreasing on $(-\infty,0)$ and hence takes its minimum at 0, so that the integral (4.99) is dominated at large N by the value at zero, which is nothing else than the left large deviation function we want to compute:

$$\Phi_-(\sigma,T) := \Psi_{\sigma,T}(0)\,. \quad (4.100)$$

Note that unlike the homogeneous case corresponding to $\sigma = 0$, one cannot simplify further this expression. Thus, we have

$$\mathrm{Prob}\left[b_1<0\right] \approx \exp\left\{-\frac{N}{2}\Phi_-(\sigma,T)+o(N)\right\}\,, \quad (4.101)$$

and from Eq. (4.91) the probability of stability writes:

$$\boxed{\mathcal{P}_{\mathrm{stable}}(T,N) \approx 1-\exp\left\{-\frac{N}{2}\Phi_-(\sigma,T)+o(N)\right\}}. \quad (4.102)$$

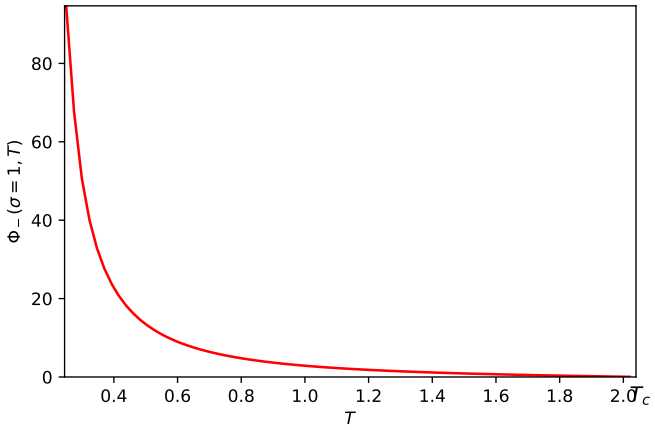


Figure 4.10: Plot of the rate function $\Phi_-(\sigma, T)$ defined by Eq. (4.100) for $\sigma = 1$ and different value of $T < T_c$.

This can be easily computed thanks to Eq. (4.68) for the density $\rho_T(b)$. A plot of the large deviation function for $\sigma = 1$ is given in Fig. 4.10 and a careful analysis of the rate function show that it matches the Tracy-Widom scaling near the edge.

4.6 Summary and conclusion of Chapter 4

In this chapter, we have studied the probability of stability of a large complex system of size N within the framework of a generalized May model, which takes into account a possible heterogeneity $a_i \neq a_j$ in the intrinsic relaxation rates of each species. In this model, the control parameter is T which is the square of the interaction strength of the random pairwise interaction between the different species. For generic distribution $\mu(a)$ of the a_i's, Eq. (4.39) completely characterizes the critical point T_c of the May-Wigner phase transition, where the system undergoes a transition from a 'stable' phase to an 'unstable' phase as T increases. Focusing on the special case where the a_i's follow what we call the flat initial condition (4.21), where σ is the only new parameter of the model controlling the spread of the distribution $\mu(a)$, we are able to (i) characterize how T_c behaves with σ, (ii) to obtain the parametric solution of the eigenvalue density of the stability matrix in the large N limit for any T, and (iii) to obtain the 'left' large deviation function $\Phi_-(\sigma, T)$ that controls the probability of stability for $T < T_c$ on the stable side, for large but finite N. One important challenge is to develop a framework to compute the 'right' large deviation function $\Phi_+(\sigma, T)$ which characterizes the probability of stability in the unstable phase $(T > T_c)$. To compute Φ_+, one needs to find the equilibrium measure of a *pushed-to-the-origin* gas of particles with a mixture of logarithmic and log-sinh pairwise interactions, as given in the joint law of eigenvalues. Finally, another natural question is to investigate the large deviation function for other initial conditions, for which we do not have a simple formula for the joint law of eigenvalues. This is tackled in the next chapter.

Chapter 5

Large deviation for the top eigenvalue of sum and product of invariant random matrices

This chapter is based on the paper [144]. The results are this chapter use the rank-one asymptotics of the spherical integrals of Chapter 3, see Sec. 3.4 and Sec. 3.4.7.

5.1 Introduction

In the previous chapter, we have seen that the large deviation principle for the top eigenvalue of a random matrix can be used to study the stability of a random linear system, and we have used this property for a very specific model: the sum of a diagonal matrix with equidistributed entries with a matrix taken from a Gaussian ensemble. In this chapter, our goal is to get the (right) large deviation principle for the sum and the product of random matrices in a very general setting. As we will see, depending on the model for the sum/product, the rate function may have one or even two second-order discontinuities. In order to have an intuition on this result, let's look at the three simple models

- $\mathrm{GOE}_\sigma + \mathrm{GOE}_{\sigma'}$,
- $\mathrm{GOE}_\sigma + \mathbf{Diag}(\mathrm{sc}_{\sigma'})$,
- $\mathbf{Diag}(\mathrm{sc}_\sigma) + \mathbf{O}\mathbf{Diag}(\mathrm{sc}_{\sigma'})\mathbf{O}^\mathsf{T}$, with $\mathbf{O} \sim \mathrm{Unif}\,[\mathrm{O}(N)]$.

where $\cdot_\sigma$ indicates that the corresponding GOE matrix has variance σ, and $\mathbf{Diag}(\mathrm{sc}_\sigma)$ is a *diagonal matrix* whose entries are taken independently from a semi-circle distribution with variance σ. At large N, the limiting spectral distributions of the three models are the same: it is a semi-circle distribution with variance $\sqrt{\sigma^2 + (\sigma')^2}$. However, the behavior of the top eigenvalue is very different. In the first case, the matrix is again a GOE matrix, and one can immediately get the smooth rate function thanks to the results of Chapter 1, Sec. 1.5. In the third case, the top eigenvalue is, for any N, lower than $2\sigma + 2\sigma'$ by the sup norm inequality. This means that the corresponding rate function must be infinite after this point and hence is very different from the one of the sums of two GOE matrices and its precise description will be given in this chapter.

The rest of this chapter is organized as follows. In Sec. 5.2, we consider the large deviation for the didactic toy model of a rank-one plus rank-one matrix. In Sec. 5.3, we study the large deviation associated with each rank-one perturbation. In Sec. 5.5, Sec. 5.6, Sec. 5.7, we consider respectively the (right) large deviation associated with the sum, product and rectangular sum of 'full-rank' matrices.

5.2 Rank-one plus rank-one toy model

In this section, we consider the 'toy model' of the sum of two rank-one matrices where one is randomly rotated. This example should be read in conjunction with the example of the norm of the sum of two vectors of Sec. 2.2 in Chapter 2. Explicitly, we consider a matrix

$$\mathbf{C} := w_A \, \boldsymbol{e}_1 \boldsymbol{e}_1^\mathsf{T} + w_B \, \boldsymbol{v}\boldsymbol{v}^\mathsf{T} \,, \tag{5.1}$$

with $\boldsymbol{e}_1 = (1, 0, \dots, 0)$ is the unit vector in the first canonical direction and $\boldsymbol{v} \sim \mathrm{Unif}\left[\mathbb{S}^{N-1}\right]$ is a unit vector taken uniformly on the real sphere. The constants w_A and w_B are known, and we take the convention $w_A \geq w_B$ without loss of generality.

An elementary calculus gives that the largest eigenvalue of $\mathbf{C}$ is given by

$$\lambda_1(\mathbf{C}) = \frac{w_A + w_B + \sqrt{(w_A - w_B)^2 + 4 w_A w_B |\boldsymbol{e}_1^\mathsf{T} \boldsymbol{v}|^2}}{2} \,. \tag{5.2}$$

Note that since $0 \leq |\boldsymbol{e}_1^\mathsf{T} \boldsymbol{v}|^2 \leq 1$, we have $w_A \leq \lambda_1(\mathbf{C}) \leq w_A + w_B$ as expected.

In the limit $N \to \infty$, the vector $\boldsymbol{v}$ is almost-surely orthogonal to the vector $\boldsymbol{e}_1$ so that the top eigenvalue of the matrix $\mathbf{C}$ is given by w_A:

$$\lambda_1(\mathbf{C}) \xrightarrow[N \to \infty]{} w_A \,, \tag{5.3}$$

This can be checked by taking the limit $|\boldsymbol{e}_1^\mathsf{T} \boldsymbol{v}| \to 0^+$ in Eq. (5.2).

Now at large but finite N, we can ask *what is the probability of finding $\lambda_1(\mathbf{C})$ at a position x higher than w_A?*

To do so, let's remark, that since $\boldsymbol{v}$ is uniform on the sphere, the square of each of its components - and hence the squared overlap - is known to follow a *Beta distribution* of parameters $(1/2, N/2)$. Its probability density is given by:

$$p\left(|\boldsymbol{e}^\mathsf{T} \boldsymbol{v}|^2 = \phi\right) = \frac{\phi^{-1/2} (1 - \phi)^{N/2 - 1}}{B(1/2, N/2)} \,, \tag{5.4}$$

where $B(1/2, N/2)$ is the Euler Beta function. From this we can do the change of variable from the overlap to the top eigenvalue given by Eq. (5.2) to write the exact probability density $\mathcal{P}_{\mathrm{tm}}$ of the law of $c_1 \equiv \lambda_1(\mathbf{C})$:

$$\mathcal{P}_{\mathrm{tm}}(c_1) = \frac{2\, c_1 - w_A - w_B}{B(1/2, N/2) w_A w_B} \left(1 - \frac{c_1 (w_A + w_B - c_1)}{w_A w_B}\right)^{-1/2} \left(\frac{c_1 (w_A + w_B - c_1)}{w_A w_B}\right)^{N/2 - 1} . \tag{5.5}$$

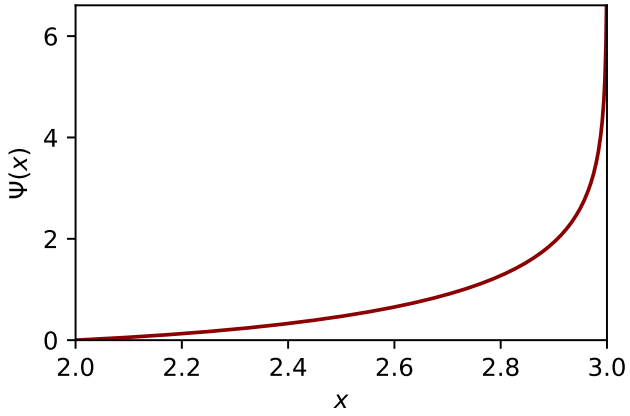


Figure 5.1: The rate function for the eigenvalue of the toy model of the sum of two rank-one matrices with non-zero eigenvalues being respectively given by $w_A = 2$ and $w_B = 1$, as described by Eq. (5.6).

IF we now take the large N limit we obtain the following result for the large deviation of the top eigenvalue.

Result 5.1 (*LDP for a rank-two matrix*)

For large N, the probability to have the top eigenvalue c_1 of the rank-one plus rank-one matrix of Eq. (5.1), at a position $x \geq w_A$, satisfies a large deviation principle with N, $\mathbb{P}[c_1 \simeq x] \approx \mathrm{e}^{-N\Psi_{\mathrm{tm}}(x)}$ and rate function given for $w_A \leq x < w_A + w_B$:

$$\Psi_{\mathrm{tm}}(x) = \frac{1}{2}\left(-\log\left(\frac{w_A + w_B - x}{w_B}\right) - \log\left(\frac{x}{w_A}\right)\right), \tag{5.6}$$

and is infinite for other values of x.

This rate function is represented in Fig. 5.1. Importantly, this rate has a logarithmic divergence near the upper bound $w_A + w_B$ and as we will see, for more sophisticated models of sum/product of matrices, this type of divergence near the upper edge will be universal.

5.3 Large deviation for rank-one deformation

In this section, we consider the case where the square (resp. rectangular) matrix $\mathbf{C}$ is a rank-one deformation of a random matrix $\mathbf{B}$ and study the right large deviation of its largest eigenvalue (resp. singular value). For this type of problem, depending on the norm of the vector of the rank-one deformation, as we have seen in Chapter 2 (see Sec. 2.4) there exists a regime where the top eigenvalue (resp. singular value) sticks to the right edge of the bulk density and another one where it pops out of the bulk density and forms an outlier. The goal of this section is to characterize completely the large deviation of this top eigenvalue/singular value. I will give proof of the result only in the regime where there is an outlier. The proof for the other regime is obtained thanks to the 'tilting method' of Sec. 5.4.

Importantly in order to tackle both the case where the 'full-rank' $\mathbf{B}$ is a fixed diagonal and the case where it is taken from an invariant ensemble, we will consider such matrix $\mathbf{B}$ to be taken

from a β-ensemble with a wall, as described in Sec. 5.4.2. We recall from Sec. 5.4.2 that the case of a fixed diagonal matrix corresponds to having the parameter at the wall $w_B = \mathrm{b}_+$, while for classical β-ensemble with a potential defined on the whole real line, this corresponds to send the wall at infinity $w_B \to \infty$.

5.3.1 Additive rank-one deformation

In this subsection, the matrix $\mathbf{B} \sim \mathbb{P}_{V,w_B}(.)$, with limiting spectral density μ_B and (right) edge b_+ and we consider the rank-one additive deformation:

$$\mathbf{C} := \mathbf{B} + \gamma\, \boldsymbol{v}\boldsymbol{v}^*\,, \tag{5.7}$$

where $\boldsymbol{v}$ is an arbitrary[1] unit vector. In this case, let's recall Result 2.3:

$$\lambda_1(\mathbf{C}) \to \begin{cases} \mathrm{b}_+ & \text{for } \gamma \le 1/g_B(\mathrm{b}_+)\,, \\ \lambda^* := g_B^{\langle -1\rangle}\left(\frac{1}{\gamma}\right) & \text{for } \gamma \ge 1/g_B(\mathrm{b}_+)\,. \end{cases} \tag{5.8}$$

To get the full behavior in the case where the typical value is the BBP outlier at λ^*, the idea is to use the characterization of the top eigenvalue as a solution of the secular equation (2.37). For $\lambda_1(\mathbf{C}) \simeq x$, we recall that this equation takes the form::

$$1 - \gamma \boldsymbol{v}^* \mathbf{G}_{\mathbf{B}}(x)\boldsymbol{v} = 0\,. \tag{5.9}$$

where $\mathbf{G}_{\mathbf{B}}(z) = (z\mathbf{I} - \mathbf{B})^{-1}$ is the resolvent of the matrix $\mathbf{B}$. As a consequence, we can write the probability of finding the top eigenvalue $\lambda_1(\mathbf{C})$ at the position x in terms of an average over a Dirac function:

$$\mathbb{P}\left[\lambda_1(\mathbf{C}) \simeq x\right] = \mathbb{E}\left[\delta(1 - \gamma \boldsymbol{v}^* \mathbf{G}_{\mathbf{B}}(x)\boldsymbol{v} = 0)\right]\,. \tag{5.10}$$

Using the inverse Laplace representation of the Dirac, this can be equivalently written as:

$$\mathbb{P}\left[\lambda_1(\mathbf{C}) \simeq x\right] = \frac{1}{K}\,\mathbb{E}\left[\int_{\mathscr{C}_1} \mathrm{e}^{\frac{N\beta z_1}{2} - \frac{N\beta\gamma z_1}{2}\boldsymbol{v}^*\mathbf{G}_{\mathbf{B}}(x)\boldsymbol{v}}\mathrm{d}z_1\right]\,, \tag{5.11}$$

where K is a (complex) constant whose asymptotic will not contribute to the large deviation in its integral representation. For simplicity, we take the notation K from one line to another, even though this constant might be different. Next, since $\mathbf{B}$ is rotationally invariant, we can either take the vector $\mathbf{v}$ to be either fixed or random. To perform the computation, it will be convenient to take $\mathbf{v}$ uniform over the sphere. Removing the constraint over the sphere by introducing a second Lagrange multiplier z_2, Eq. (5.11) now writes:

$$\mathbb{P}\left[\lambda_1(\mathbf{C}) \simeq x\right] = \frac{1}{K}\,\mathbb{E}\left[\int_{\mathscr{C}_1\times\mathscr{C}_2} \mathrm{e}^{\frac{N\beta z_1}{2} + \frac{N\beta z_1}{2}} \left(\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{N\beta}{2}\boldsymbol{v}^*(z_2\mathbf{Id} + \gamma z_1 \mathbf{G}_{\mathbf{B}}(x))\boldsymbol{v}}\mathrm{d}\boldsymbol{v}\right)\mathrm{d}z_1\mathrm{d}z_2\right]\,. \tag{5.12}$$

By Gaussian integration over the N-dimensional variable $\boldsymbol{v}$, we have the following integral:

$$\mathbb{P}\left[\lambda_1(\mathbf{C}) \simeq x\right] = \mathbb{E}\left[\int_{\mathscr{C}_1\times\mathscr{C}_2} \mathrm{e}^{\frac{N\beta}{2}H(z_1,z_2,x)}\mathrm{d}z_1\mathrm{d}z_2\right]\,, \tag{5.13}$$

[1] If one replaces the matrix $\mathbf{B}$ by a fixed diagonal matrix o then one has to consider the vector $\boldsymbol{v}$ to be taken uniformly over the sphere $\mathbb{S}^{N-1}$ for the problem to be invariant

with:

$$H(z_1, z_2, x) := z_1 + z_2 - \frac{1}{N}\sum_{i=1}^{N} \log\left(\gamma z_1 + z_2(x - b_i)\right) + \frac{1}{N}\sum_{i=1}^{N} \log(x - b_i) + K + \mathcal{O}\left(\frac{1}{N}\right), \tag{5.14}$$

where K is constant independent of z_1, z_2 and x. In the large N limit, this integral over the variables z_1 and z_2 are dominated by the saddle-points z_1^* and z_2^* solutions of:

$$\begin{cases} \frac{\gamma}{N}\sum_{i=1}^{N} \frac{1}{\gamma z_1^* + z_2^*(x-b_i)} & = 1\,, \\ \\ \frac{1}{N}\sum_{i=1}^{N} \frac{(x-b_i)}{\gamma z_1^* + z_2^*(x-b_i)} & = 1\,. \end{cases} \tag{5.15}$$

Combining the two equations, we have that the two saddle points are related to one each another by:

$$z_1^* + z_2^* = 1\,. \tag{5.16}$$

As we will see later on, only the saddle-point z_2^* will contribute to the large deviation. Substituting z_2^* in the top line of Eq. (5.15), we have:

$$\frac{z_2^*}{\gamma} = g_{\mathbf{B}}\left(x - \gamma + \frac{\gamma}{z_2^*}\right). \tag{5.17}$$

where $g_{\mathbf{B}}(z) = 1/N \cdot \sum_{i=1}^{N}(z - b_i)^{-1}$ is the Stieltjes transform of the matrix $\mathbf{B}$. As in the study of the asymptotic behavior of rank-one spherical integral, we need to distinguish two cases: Eq. (5.17) only makes sense if the argument of the RHS does not exceed w_B, that is if we define:

$$x_{c_2} := w_B + \gamma - \frac{1}{\bar{g}_B(w_B)}\,, \tag{5.18}$$

then

- if $x \leq x_{c_2}$ we don't have any saturation. Inverting Eq. (5.17) gives:

$$\frac{z_2^*}{\gamma} = \mathcal{R}_B^{\langle -1\rangle}(x - \gamma)\,, \tag{5.19}$$

- if $x \geq x_{c_2}$, then there is a saturation, that is z_2^* is given by

$$\frac{z_2^*}{\gamma} = \frac{1}{w_B + \gamma - x}\,. \tag{5.20}$$

We have now all the tools to express the rate function in this regime. We have:

$$\Psi_C(x) = -\int_{\lambda^*}^{x} \frac{\mathrm{d}}{\mathrm{d}s} H(z_1^*, z_2^*, s)\mathrm{d}s = -\int_{\lambda^*}^{x} \frac{\partial}{\partial s} H(z_1^*, z_2^*, s)\mathrm{d}s\,, \tag{5.21}$$

since the partial derivatives with respect to z_1^* and z_2^* are exactly zero at the saddle points. The derivative of the function H with respect to s is given thanks to Eq. (5.14) by:

$$\frac{\partial}{\partial s} H(z_1^*, z_2^*, s) = -\frac{1}{N}\sum_{i=1}^{N} \frac{z_2^*}{\gamma z_1^* + z_2^*(x - b_i)} + \frac{1}{N}\sum_{i=1}^{N} \frac{1}{x - b_i}\,. \tag{5.22}$$

The first term in the RHS can be simplified since if we multiply the topline of Eq. (5.15) by $\frac{z_2^*}{\gamma}$ we find that this term is exactly $\frac{z_2^*}{\gamma}$. Taking the large N limit, one has the expression for the rate function in this regime. The proof of the rate function for the regime where there is no outlier can be obtained thanks to the tilting method described in the next Section, and we have the following two results

Result 5.2 (*LDP for the additive rank-one perturbation below the BBP threshold*)

For the additive rank-perturbation of Eq. (5.7), in the regime where there is no outlier, $\gamma \leq \frac{1}{g_B(\mathrm{b}_+)}$, the probability to observe the event $\lambda_1(\mathbf{C}) \simeq x$ satisfies a large deviation principle with speed $N\beta$: $\mathbb{P}\left[\lambda_1(\mathbf{C}) \simeq x\right] \sim \mathrm{e}^{-N\beta\cdot\Psi_C(x)}$ and the rate function is infinite for $x \notin [\mathrm{b}_+, w_B+\gamma)$ and is otherwise given by:

$$\Psi_C(x) = \frac{1}{2}\begin{cases} \int_{\mathrm{b}_+}^{x} \left(\bar{g}_B(t) - g_B(t)\right) \mathrm{d}t & \text{for } \mathrm{b}_+ \leq x \leq x_{c_1}\,, \\ K_1 + \int_{x_{c_1}}^{x} \left(\mathcal{R}_B^{\langle -1\rangle}(t-\gamma) - g_B(t)\right) \mathrm{d}t & \text{for } x_{c_1} \leq x \leq x_{c_2}\,, \\ K_2 + \log\left(\frac{1}{w_B+\gamma-x}\right) - \int_{x_{c_2}}^{x} g_B(t)\mathrm{d}t & \text{for } x_{c_2} \leq x \leq w_B+\gamma\,, \end{cases} \tag{5.23}$$

with $x_{c_1} := g_B^{\langle -1\rangle}\left(\frac{1}{\gamma}\right)$ and $x_{c_2} := w_B + \gamma - \frac{1}{\bar{g}_B(w_B)}$ and the constant K_1 and K_2 are such that Ψ_C is continuous at the points x_{c_1} and x_{c_2}: $K_1 := \int_{\mathrm{b}_+}^{x_{c_1}} \left(\bar{g}_B(t) - g_B(t)\right) \mathrm{d}t$ and $K_2 := K_1 + \log\left(\frac{1}{\bar{g}_B(w_B)}\right) + \int_{x_{c_1}}^{x_{c_2}} \left(\mathcal{R}_B^{\langle -1\rangle}\left(t-\gamma\right) - g_B(t)\right) \mathrm{d}t$.

Result 5.3 (*LDP for the additive rank-one perturbation above the BBP threshold*)

For the additive rank-perturbation of Eq. (5.7), in the regime $\gamma \geq \frac{1}{g_B(\mathrm{b}_+)}$ where there is an outlier at $\lambda^* = g_B^{\langle -1\rangle}(1/\gamma)$, the probability to observe the event $\lambda_1(\mathbf{C}) \simeq x$ satisfies a large deviation principle with speed $N\beta$: $\mathbb{P}\left[\lambda_1(\mathbf{C}) \simeq x\right] \sim \mathrm{e}^{-N\beta\cdot\Psi_C(x)}$ and the rate function is infinite for $x \notin [\mathrm{b}_+, w_B+\gamma)$ and is otherwise given by:

$$\Psi_C(x) = \frac{1}{2}\begin{cases} \int_{\lambda^*}^{x} \left(\mathcal{R}_B^{\langle -1\rangle}(t-\gamma) - g_B(t)\right) \mathrm{d}t & \text{for } \mathrm{b}_+ \leq x \leq x_{c_2}\,, \\ K_2^* + \log\left(\frac{1}{w_B+\gamma-x}\right) - \int_{x_{c_2}}^{x} g_B(t)\mathrm{d}t & \text{for } x_{c_2} \leq x \leq w_B+\gamma\,, \end{cases} \tag{5.24}$$

with $x_{c_2} := w_B + \gamma - \frac{1}{\bar{g}_B(w_B)}$ and $K_2^* = \int_{\lambda^*}^{x_{c_2}} \left(\bar{g}_B(t) - g_B(t)\right) \mathrm{d}t + \log\left(\frac{1}{\bar{g}_B(w_B)}\right)$.

Let's once again recall that the limiting case $w_B = \mathrm{b}_+$ and $w_B \to \infty$ corresponds respectively to the case where $\mathbf{B}$ is a fixed diagonal matrix and the case where $\mathbf{B}$ is taken from a β-ensemble. In the latter case, the critical point $x_{c_2} \to \infty$ such that one can remove the bottom line of both Eqs (5.23) and (5.24).

Remark *(Behavior near the edge).* Below the threshold when there is no outlier, if the density is non-critical, one recovers the Tracy-Widom '3/2' scaling for the rate function since the expression matches the ones of the classical case of Eq. (1.107). However above

the threshold, because both the Stieltjes transform and the R-transform (and hence its shifted inverse) are analytic around λ^*, the rate function as a square behavior close to the outlier. This is expected because the fluctuations of the outliers are known to be Gaussian and of variance $N^{-\frac{1}{2}}$. ⌟

Example *(Large deviation for rank-one perturbation of a* GOE *matrix)***.** Let's consider the case where $\mathbf{A}$ is a GOE matrix. In this case, $\mathrm{b}_+ = 2\sigma$, $g_B(\mathrm{b}_+) = \frac{1}{\sigma}$, and if $\gamma \leq \sigma$ there is no outlier but one critical point $x_{c_1} = \gamma + \frac{\sigma^2}{\gamma}$ and if $\gamma \geq \sigma$ there is one outlier at $\lambda^* = \gamma + \frac{\sigma^2}{\gamma}$. As a consequence, the rate function is given by:

- if $\gamma \leq \sigma$,

$$\Psi_C(x) = \begin{cases} \frac{x\sqrt{x^2-4\sigma^2}}{4\sigma^2} + \log\left(\frac{2\sigma}{\sqrt{x^2-4\sigma^2}+x}\right) & \text{for } 2\sigma \leq x \leq \gamma + \frac{\sigma^2}{\gamma}, \\ \frac{\left(x-(\gamma+\frac{\sigma^2}{\gamma})\right)\left(x-(3\gamma+\frac{\sigma^2}{\gamma})\right)+x\sqrt{x^2-4\sigma^2}}{4\sigma^2} + \log\left(\frac{2\sigma}{x+\sqrt{x^2-4\sigma^2}}\right) & \text{for } x \geq \gamma + \frac{\sigma^2}{\gamma}, \end{cases} \tag{5.25}$$

  One can find a plot of this function for $\gamma = 1/2$ and $\sigma = 1$ in Fig. 5.2 (Left).

- and if $\gamma \geq \sigma$,

$$\Psi_C(x) = \frac{x^2 - 4\gamma x + 2(\gamma^2+\sigma^2) + x\sqrt{x^2-4\sigma^2}}{4\sigma^2} + \log\left(\frac{2\gamma}{x+\sqrt{x^2-4\sigma^2}}\right). \tag{5.26}$$

  One can find a plot of this function for $\gamma = 2$ and $\sigma = 1$ in Fig. 5.2 (Right).

■

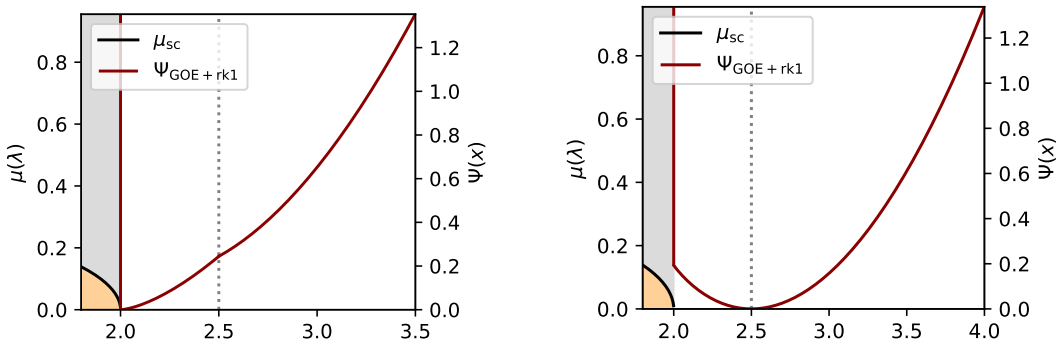


Figure 5.2: On the left, the Rate function (in red) of the largest eigenvalue of the sum of a GOE matrix with $\sigma = 1$ and a rank-one matrix with a non-zero eigenvalue equal to $\gamma = 1/2$, as described by Eq. (5.25). This function admits a phase transition at $x_{c_1} = 2.5$, represented by the vertical dotted line. On the right, the Rate function (in red) of the largest eigenvalue of the sum of a GOE matrix with $\sigma = 1$ and a rank-one matrix with now a non-zero eigenvalue equal to $\gamma = 2$, as described by Eq. (5.26). In this case, one is above the BBP transition and the typical position of the largest eigenvalue of this matrix is represented in a dotted line.

5.3.2 Multiplicative case

Similar to the additive case, we consider in this subsection the 'multiplicative' rank-one deformation of Sec. 2.4.3, that is:

$$\mathbf{C} = \sqrt{\mathbf{I} + \gamma \mathbf{v}\mathbf{v}^{\mathsf{T}}}\mathbf{B}\sqrt{\mathbf{I} + \gamma \mathbf{v}\mathbf{v}^{\mathsf{T}}}\,, \tag{5.27}$$

where $\mathbf{B} \sim \mathbb{P}_{V,w_B}(.)$ and is further assumed to be semi-definite positive. In the large N limit, let's recall form Res. 2.6 the associated BBP transition:

$$\lambda_1(\mathbf{C}) \to \begin{cases} \mathrm{b}_+ & \text{for } \gamma \leq 1/t_B(\mathrm{b}_+)\,, \\ \lambda^* := t_B^{\langle -1\rangle}\left(\frac{1}{\gamma}\right) & \text{for } \gamma \geq 1/t_B(\mathrm{b}_+)\,. \end{cases} \tag{5.28}$$

The proof of the large deviation principle is similar to the additive case. For the case below the BBP transition, this can be done by the tilting method of the next section while when there is an outlier, one may use the characterization with the (multiplicative) secular equation of Eq. (2.56), that is to have λ_1 at x, it must satisfy:

$$1 - \gamma \boldsymbol{v}^* \sqrt{\mathbf{B}}\mathbf{G}_{\mathbf{B}}(x)\sqrt{\mathbf{B}}\boldsymbol{v} = 0\,. \tag{5.29}$$

with $\mathbf{G}_{\mathbf{B}}(x) = (x\mathbf{I} - \mathbf{B})^{-1}$. Using again the delta trick one may compute the associated large deviation function and the results are given by:

Result 5.4 (*LDP for the multiplicative rank-one perturbation below the BBP threshold*)

For the multiplicative rank-perturbation of Eq. (5.27), in the regime where there is no outlier, $\gamma \leq \frac{1}{t_B(\mathrm{b}_+)}$, the probability to observe the event $\lambda_1(\mathbf{C}) \simeq x$ satisfies a large deviation principle with speed $N\beta$: $\mathbb{P}\left[\lambda_1(\mathbf{C}) \simeq x\right] \sim \mathrm{e}^{-N\beta\cdot\Psi_C(x)}$ and the rate function is infinite for $x \notin [\mathrm{b}_+, w_B(1+\gamma))$ and is otherwise given by:

$$\Psi_C(x) = \frac{1}{2}\begin{cases} \int_{\mathrm{b}_+}^{x} \bar{g}_B(t) - g_B(t)\mathrm{d}t & \text{for } \mathrm{b}_+ \leq x \leq x_{c_1}\,, \\ K_1 + \int_{x_{c_1}}^{x}\left(\frac{\tilde{\mathcal{S}}_B^{\langle -1\rangle}\left(\frac{t}{1+\gamma}\right)+1}{t} - g_B(t)\right)\mathrm{d}t & \text{for } x_{c_1} \leq x \leq x_{c_2}\,, \\ K_2 + \log\left(\frac{x}{w_A(1+\gamma)-x}\right) - \int_{x_{c_2}}^{x} g_B(t)\mathrm{d}t & \text{for } x_{c_2} \leq x \leq w_B(1+\gamma)\,. \end{cases} \tag{5.30}$$

with $x_{c_1} := t_B^{\langle -1\rangle}(1/\gamma)$ and $x_{c_2} := \gamma w_B \frac{\bar{t}_B(w_B)}{t_B(w_B)+1}$ and the constant K_1 and K_2 are such that Ψ_C is continuous at the points x_{c_1} and x_{c_2}: $K_1 := \int_{\mathrm{c}_+}^{x_{c_1}} \left(\bar{g}_B(t) - g_B(t)\right)\mathrm{d}t$ and $K_2 := K_1 + \log\left(1/\bar{t}_B(w_B)\right) + \int_{x_{c_1}}^{x_{c_2}}\left((\tilde{\mathcal{S}}_B^{\langle -1\rangle}\left(t/(1+\gamma)\right)+1)/t - g_B(t)\right)\mathrm{d}t$.

Result 5.5 (*LDP for the multiplicative rank-one perturbation above the BBP threshold*)

For the multiplicative rank-perturbation of Eq. (5.27), in the regime $\gamma \geq \frac{1}{t_B(\mathrm{b}_+)}$ where there is an outlier at $\lambda^* = t_B^{\langle -1\rangle}(1/\gamma)$, the probability to observe the event $\lambda_1(\mathbf{C}) \simeq x$ satisfies a large deviation principle with speed $N\beta$: $\mathbb{P}\left[\lambda_1(\mathbf{C}) \simeq x\right] \sim \mathrm{e}^{-N\beta\cdot\Psi_C(x)}$ and the

rate function is infinite for $x \notin [\mathrm{b}_+, w_B(1+\gamma))$ and is otherwise given by:

$$\Psi_C(x) = \frac{1}{2} \begin{cases} \int_{\lambda^*}^{x} \left(\frac{\tilde{\mathcal{S}}_B^{\langle -1 \rangle}\left(\frac{t}{1+\gamma}\right)+1}{t} - g_B(t) \right) \mathrm{d}t & \text{for } \mathrm{b}_+ \leq x \leq x_{c_2}\,, \\ K_2^* + \log\left(\frac{x}{w_B(1+\gamma)-x}\right) - \int_{x_{c_2}}^{x} g_B(t)\mathrm{d}t & \text{for } x_{c_2} \leq x \leq w_B(1+\gamma)\,. \end{cases} \tag{5.31}$$

with $x_{c_2} := (1+\gamma) w_B \frac{\bar{t}_B(w_B)}{t_B(w_B)+1}$ and the constant $K_2^* := \log(1/\bar{t}_B(w_B)) + \int_{\lambda^*}^{x_{c_2}} \left(\left(\tilde{\mathcal{S}}_B^{\langle -1 \rangle}(t/(1+\gamma)) + 1 \right)/t - g_B(t) \right) \mathrm{d}t$ is such that Ψ_C is continuous at x_{c_2}.

Example *(spiked square Wishart)*. Let's consider the case where $\mathbf{B}$ is Wishart, where in order to have a simple analytical formula for the rate function, we consider the shape parameter to be equal to one, $q = 1$. In this case, the density of Eq. (1.43) has a top edge at $\mathrm{b}_+ = 4$. The matrix $\mathbf{C}$ given by Eq. (5.27) is known as a spiked (square) Wishart matrix and the rate function for its largest eigenvalue is given by:

- if $\gamma \leq 1$, then there is no outlier and the rate function is given by:

$$\Psi_C(x) = \begin{cases} \frac{\sqrt{x(x-4)}}{2} + \log\left(\frac{x-2-\sqrt{x(x-4)}}{2}\right) & \text{for } 4 \leq x \leq 2+\gamma+\frac{1}{\gamma}\,, \\ \frac{x-\gamma x+(1+\gamma)\sqrt{x(x-4)}}{4(1+\gamma)} + \frac{1}{2}\log\left(\frac{x\gamma-2\gamma-\gamma\sqrt{x(x-4)}}{2}\right) & \text{for } x \geq 2+\gamma+\frac{1}{\gamma}\,. \end{cases} \tag{5.32}$$

  One can find a plot of this function for $\gamma = 1/2$ and $q = 1$ in Fig. 5.3 (Left).

- if $\gamma > 1$, then there is an outlier at $\lambda^* = 2+\gamma+\frac{1}{\gamma}$ and in this case the rate function is given by:

$$\Psi_C(x) = \frac{1+\frac{1}{\gamma}}{2} + \frac{x-2-\gamma-\frac{1}{\gamma}}{2(1+\gamma)} + \frac{\sqrt{x(x-4)}-x}{4} + \frac{1}{2}\log\left(\gamma \frac{1-\sqrt{\frac{x-4}{x}}}{1+\sqrt{\frac{x-4}{x}}}\right)\,. \tag{5.33}$$

  One can find a plot of this function for $\gamma = 2$ and $q = 1$ in Fig. 5.3 (Right).

■

5.4 The tools to compute large deviations: tilting method and invariant-ensembles with a wall

5.4.1 Tilting method

For an individual matrix taken from a β-ensemble of Sec. 4.5.3 and the toy model of the rank-two matrix of Sec. 5.2 and the rank-one perturbation of the previous section, we were able

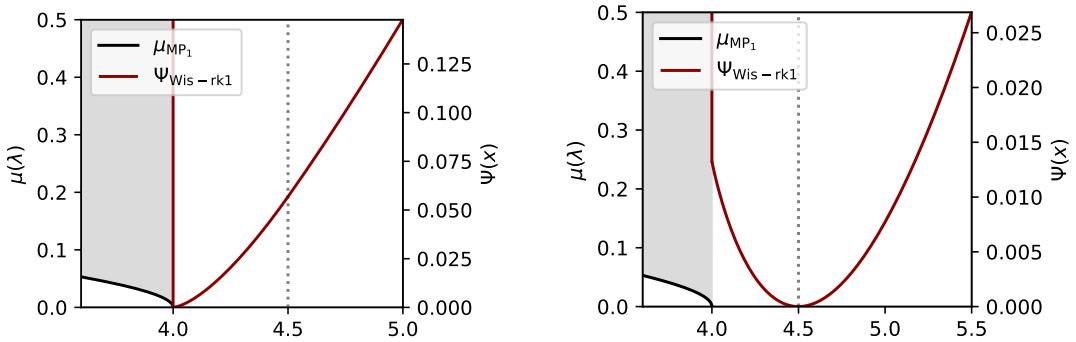


Figure 5.3: On the left, the Rate function (in red) of the largest eigenvalue of a spiked Wishart matrix with $q = 1$ and the value of the spike is $\gamma = 1/2$, as described by Eq. (5.25). This function admits a phase transition at $x_{c_1} = 4.5$ represented by the vertical dotted line. On the right, the Rate function (in red) of the largest eigenvalue of the sum of a spiked Wishart matrix with $q = 1$ and a spike $\gamma = 2$, as described by Eq. (5.26). In this case, one is above the BBP transition and the typical position of the largest eigenvalue of this matrix is represented in a dotted line.

to compute the (right) large deviation of the top eigenvalue thanks to an explicit characterization for the joint laws of the eigenvalues or even the top eigenvalue itself. For the sum or the product of matrices, one does not have a *simple* expression for the joint density of the eigenvalues/singular values, except in some specific cases, see for example the previous chapter. Instead of directly looking at $\mathbf{C}$, the idea introduced in Ref. [85] in the context of RMT, is to look at a *weighted* realization of this matrix. If we denote by $P(.)$ the probability density of the random matrix $\mathbf{C}$ in the space of self-adjoint/rectangular matrix, let's consider *another* random matrix $\mathbf{C}'$ whose probability density is given by:

$$P^{[\theta]}(\mathbf{C}') := \frac{\mathcal{Z}_{\mathbf{C}'}(\theta)}{\mathbb{E}\left[\mathcal{Z}_{\mathbf{C}}(\theta)\right]} P(\mathbf{C}')\,, \tag{5.34}$$

where $\mathcal{Z}_{\mathbf{C}'}(\theta)$ is a *tilting function*. Note that we use the same notation as in Chapter 3, Sec. 3.4.2 for this tilting function since, as we will see, it will correspond to the rank-one spherical integrals. The role of this function is to make the event $\{\zeta_1(\mathbf{C}') \simeq x\}$ of high probability for the new matrix $\mathbf{C}'$ for a given choice of the parameter θ. One can then relate the rate function $\Psi_C(.)$ to its asymptotic behavior, as we will explain below. We first list sufficient conditions for the choice of this tilting function and then explain how based on these properties one can get the large deviations for the top eigenvalue/singular-value.

Sufficient conditions for the tilting function -

(C1) The tilting function $\mathcal{Z}_{\mathbf{C}'}(\theta)$ only depends on the eigenvalues/singular values of $\mathbf{C}'$ (and not its eigenvectors) and for large N this function depends only on the position of the top eigenvalue/singular value c'_1 and becomes self-averaging with respect to the other $c'_2, \ldots, c'_N$ and thus it is independent of them.

(C2) The tilting function $\mathcal{Z}_{\mathbf{C}'}(\theta)$ satisfies the following *decomposition property:*

$$\mathbb{E}\left[\mathcal{Z}_{\mathbf{C}}(\theta)\right] = \mathbb{E}_{\mathbf{A}}\left[\mathcal{Z}_{\mathbf{A}}(\theta)\right] \mathbb{E}_{\mathbf{B}}\left[\mathcal{Z}_{\mathbf{B}}(\theta)\right]\,. \tag{5.35}$$

(C3) For large N, one can compute the *quenched free energy* defined as

$$J_C(x,\theta) :\approx \frac{2}{N\beta}\log \mathcal{Z}_{\mathbf{C}|\{c_1=x\}}(\theta)\,, \tag{5.36}$$

where $\{c_1 = x\}$ indicates that the limit is taken with the constraint that the top eigenvalue (or singular value) is fixed at x. Similarly, for large N, one can compute the two *annealed free energies*:

$$F_A(\theta) :\approx \frac{2}{N\beta}\log \mathbb{E}\mathcal{Z}_{\mathbf{A}}(\theta) \quad \text{and} \quad F_B(\theta) :\approx \frac{2}{N\beta}\log \mathbb{E}\mathcal{Z}_{\mathbf{B}}(\theta)\,. \tag{5.37}$$

(C4) For any $x > \mathrm{c}_+$, there is one optimal $\theta^*(x)$, solution of:

$$\theta^*(x) = \underset{\theta>0}{\operatorname{argsup}}\{I_x(\theta) := J_C(x,\theta) - F_A(\theta) - F_B(\theta)\}\,, \tag{5.38}$$

and this supremum can be computed.

General expression for the rate function with the tilting method -

Under these conditions and for large N, because $\mathcal{Z}_{\mathbf{C}'}$ is a function of the eigenvalues (or singular values), we can integrate out the dependency in the eigenvectors in Eq. (5.34) to obtain a relation between the joint density $\mathcal{P}_N^{(\theta)}(c_1,\ldots,c_N)$ of the eigenvalues/singular values of $\mathbf{C}'$ to the (unknown) joint density $\mathcal{P}_N(c_1,\ldots,c_N)$ of eigenvalues/singular values of $\mathbf{C}$:

$$\mathcal{P}_N^{(\theta)}(c_1',\ldots,c_N') = \mathrm{e}^{\frac{N\beta}{2}\left[J_C(c_1',\theta)-F_A(\theta)-F_B(\theta)\right]+\mathcal{O}(N)}\mathcal{P}_N\left(c_1',\ldots,c_N'\right)\,, \tag{5.39}$$

where we have used the decomposition property (5.35) and the definitions of the annealed and quenched free energies given by Eq. (5.36) and Eq. (5.37). If we now integrate over the variables $c_1,'\ldots c_N'$ with a Dirac delta function $\delta(c_1' - x)$, by self-averaging property (C1) with respect to the variables $c_2',\ldots,c_N'$ for the tilting function, we have:

$$\text{Prob.}^{[\theta]}\left[c_1' \simeq x\right] = \mathrm{e}^{\frac{N\beta}{2}[J_C(x,\theta)-F_A(\theta)-F_B(\theta)]+\mathcal{O}(N)}\,\text{Prob.}\left[\lambda_1(\mathbf{C}) \simeq x\right]\,, \tag{5.40}$$

which is by definition of the rate function

$$\text{Prob.}^{[\theta]}\left[c_1' \simeq x\right] = \mathrm{e}^{\frac{N\beta}{2}[J_C(x,\theta)-F_A(\theta)-F_B(\theta)-\Psi_C(x)]+\mathcal{O}(N)}\,. \tag{5.41}$$

Now to get the rate function $\Psi_C(.)$, let us find the optimal value $\theta^* \equiv \theta^*(x)$ of the parameter θ that makes the event $\{c_1' \simeq x\}$ of probability one. Since the LHS of (5.40) is a probability measure, it is always bounded by one, and so the corresponding optimal θ^* corresponds to maximize the argument of the exponential in the RHS of Eq. (5.40), that is, θ^* is given by Eq. (5.38). By condition (C4), this supremum exists for any $x > \mathrm{c}_+$. As a consequence, we can relate the unknown rate function $\Psi_C(.)$ to the annealed and quenched free energies:

$$\Psi_C(x) = \frac{1}{2}\left(J_C(x,\theta^*(x)) - F_A(\theta^*(x)) - F_B(\theta^*(x))\right)\,. \tag{5.42}$$

Now if the supremum in Eq. (5.38) is a maximum, one has that θ^* is the solution of:

$$I_x'(\theta^*) := \partial_\theta J_C(x,\theta)|_{\theta^*} - F_A'(\theta^*) - F_B'(\theta^*) = 0\,. \tag{5.43}$$

By construction $\Psi_C(\mathrm{c}_+) = 0$. One can therefore take the derivative of Eq. (5.42) with respect to x to get the following integral representation for the rate function:

$$\Psi_C(x) = \frac{1}{2}\int_{\mathrm{c}_+}^{x} \partial_t J_C(t, \theta^*(t)) + \theta^{*\prime}(t)(\partial_\theta J_C(t,\theta)|_{\theta^*} - F'(\theta^*))\mathrm{d}t\,, \tag{5.44}$$

but by Eq. (5.43) the second term is null, so that we have the following final simple formula for the rate function.

$$\boxed{\Psi_C(x) = \frac{1}{2}\int_{\mathrm{c}_+}^{x} \partial_t J_C(t, \theta^*(t))\mathrm{d}t \qquad \text{with } \theta^* \text{ solution of Eq. (5.43)}}\,. \tag{5.45}$$

Spherical Integrals as tilting functions -

Thus, if one has found a tilting function with the four properties (C1)-(C4), one can compute the rate function. We argue that a natural candidate for the tilting function is given by the *spherical integral* of the operation we are considering:

- On the one hand, the quenched free energy associated with each spherical integral has been computed in Chapter 3 (see Sec. 3.4.7) and it satisfies a θ-dependent transition between a phase where it does not depend explicitly on the position of the top eigenvalue/singular value and a phase where it does. In other words, spherical functions satisfy the conditions (C1)-(C2).
- On the other hand, the spherical functions satisfy by the construction of the decomposition property (C3) of Eq. (5.35).

As a consequence, what is left is to prove that spherical functions satisfy the property (C4) and then compute the corresponding optimal parameter $\theta^*(x)$. This is done for each case in Sec. 5.5.2; Sec. 5.6.1 and Sec. 5.7.1. Injecting the expression of the optimal parameter $\theta^*(x)$ in Eq. (5.45), we can then get an expression for the rate function and the results are given in in Sec. 5.5.3, in Sec. 5.6.2 and in Sec. 5.7.2.

5.4.2 Invariant ensembles with a wall

In order to deal with the free sum (or the free product or bi-free sum) of either matrix taken from an invariant ensemble or fixed diagonal matrices or a combination of the two under the same framework, we recall the definition of *invariant ensembles with a wall* of Chapter 3.

We say that a matrix is taken from an invariant ensemble with a wall at w_A, which we denote by $\mathbf{A} \sim \mathbb{P}^{(\beta)}_{V,w_A}$, if $\mathbf{A} = \mathbf{O}\mathrm{Diag}\left(\lambda_1, \dots, \lambda_N\right)\mathbf{O}^\mathsf{T}$ with $\mathbf{O}$ uniform over $\mathrm{O}(N)$ and the $\{\lambda_i\}$ follow the joint law of Eq. (1.63) with $V(.)$ a confining potential such that $V(x > w_A) = \infty$, see Fig. 5.4. The definition naturally extends to rectangular matrices. It is important to notice that the introduction of this wall does not change the limiting equilibrium density $\mu_A(.)$ since $w_A \geq \mathrm{a}_+$ and the solution of the Tricomi problem of Eq. (1.75) only depends on the values of the potential between the two edges $\mathrm{a}_\pm$.

Now if we look at the two limiting cases:

- If $w_A = \infty$, then there is no wall, and we recover the case of an unrestricted invariant ensemble.

- If $w_A = \mathrm{a}_+$, then at finite N the top eigenvalue cannot fluctuate outside the support of the bulk density $\mu_A(.)$. We argue that from the point of the right large deviation it behaves as a fixed diagonal matrix with limiting density μ_A and edge a_+. To justify this statement, we have seen the following property in Chapter 3;

  (C5) the annealed free energy for an invariant ensemble with a wall is related to the quenched free energy by:

  $$\partial_\theta F_A(w_A = \mathrm{a}_+, \theta) = \partial_\theta J_A(\mathrm{a}_+, \theta)\,, \tag{5.46}$$

  where we have made explicit the dependency in the wall w_A for $F_A(w_A, \theta) := \frac{1}{N}\log \mathbb{E}_{\mathbf{A}\sim\mathbb{P}_{V,w_A}} \mathcal{Z}_{\mathbf{A}}(\theta)$. Indeed, if $\mathbf{A}$ is a fixed diagonal matrix with limiting density μ_A and limiting edge a_+ then one can drop the average over $\mathbf{A}$ in the decomposition property (C3) of Eq. (5.35):

  $$\mathbb{E}\left[\mathcal{Z}_{\mathbf{C}}(\theta)\right] = \mathcal{Z}_{\mathbf{A}}(\theta)\,\mathbb{E}_{\mathbf{B}}\left[\mathcal{Z}_{\mathbf{B}}(\theta)\right]\,, \tag{5.47}$$

  which, by definition of the quenched free energy, corresponds to replacing $F'_A(\theta^*)$ by $\partial_\theta J_A(\mathrm{a}_+, \theta)$ in Eq. (5.43). Since no other quantities are changed, if we have (C5) then from the point of the right large deviation of the top eigenvalue, we can safely replace a fixed diagonal matrix with a corresponding matrix taken from an invariant ensemble with a wall at its edge.

As a consequence, invariant ensembles with a wall provide a framework to tackle both cases (invariant and fixed diagonal) at the same time and appear to be interesting objects on their own. By construction, matrices taken from an invariant ensemble with a wall are rotationally invariant, such that we can also drop out the conjugation by Haar matrices.

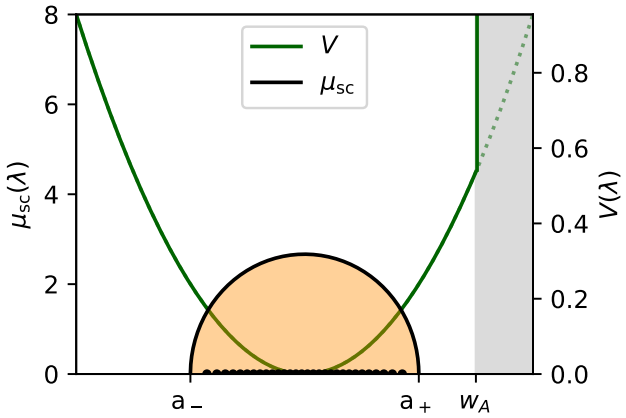


Figure 5.4: Potential and limiting density for a random matrix taken from an invariant ensemble with a wall. Beyond the wall, the potential is infinite. The limiting density is the same as if there were no wall ($w_A \to \infty$) since $w_A \geq \mathrm{a}_+$. The black dots represent 'typical' configurations of the eigenvalues at finite N.

5.5 Large deviation for the sum of self-adjoint matrices

In this section, we consider the case where the matrix $\mathbf{C}$ is given as

$$\mathbf{C} = \mathbf{A} + \mathbf{B}\,, \tag{5.48}$$

where $\mathbf{A} \sim \mathbb{P}^{(\beta)}_{V_A,w_A}$ and $\mathbf{B} \sim \mathbb{P}^{(\beta)}_{V_B,w_B}$ are two *symmetric* matrices, each taken from an invariant ensemble with a wall as defined in Sec. 5.4.2. We recall from the previous section that:

- if we set the wall parameters at the edges, $w_A = \mathrm{a}_+$ and $w_B = \mathrm{b}_+$, then we get the large deviation for the top eigenvalue of the sum of two randomly fixed diagonal matrices:

$$\mathbf{Diag}(\boldsymbol{a}) + \mathbf{V}\mathbf{Diag}(\boldsymbol{b})\mathbf{V}^*\,, \tag{5.49}$$

  with $\mathbf{O}$ a uniform orthogonal matrix and $1/N\sum_i \delta(\lambda - a_i) \to \mu_A(\lambda)$ and $1/N\sum_i \delta(\lambda - b_i) \to \mu_B(\lambda)$.

- If we set $w_A = \mathrm{a}_+$ and $w_B = \infty$, then we get the large deviation for the top eigenvalue of the sum of:

$$\mathbf{Diag}(\boldsymbol{a}) + \mathbf{B}\,, \tag{5.50}$$

  with $1/N\sum_i \delta(\lambda - a_i) \to \mu_A(\lambda)$ and $\mathbf{B} \sim \mathbb{P}^{(\beta)}_{V_B}$ is a matrix taken from an invariant ensemble.

- If we set $w_A = \infty$ and $w_B = \infty$, then we get the large deviation for the top eigenvalue of the sum:

$$\mathbf{A} + \mathbf{B}\,, \tag{5.51}$$

  where $\mathbf{A} \sim \mathbb{P}^{(\beta)}_{V_A}$ and $\mathbf{B} \sim \mathbb{P}^{(\beta)}_{V_B}$ are two matrices taken from (unrestricted) invariant ensembles.

We aim at computing the rate function $\Psi_C(x)$:

$$\mathbb{P}\left[\lambda_1(\mathbf{C}) \simeq x\right] = \exp\left[-N\beta\,\Psi_C(x) + o(N)\right] \qquad \text{for } x > \mathrm{c}_+\,, \tag{5.52}$$

where c_+ is the edge of the limiting spectrum μ_C of $\mathbf{C}$, and $\mu_C = \mu_A \boxplus \mu_B$ is described by the free convolution of Chapter 2 (see Sec. 2.6.3).

5.5.1 Warm-up computation: retrieving the right large deviation for an individual random matrix taken from a β-ensemble

The next step in order to get the rate function is to show that there exists an optimal temperature $\theta^*(x)$ and compute it. Let's consider the case of *one* individual random matrix $\mathbf{A} \sim \mathbb{P}^{(\beta)}_V(.)$ in a classical invariant ensemble, and let's retrieve the expression of Eq. (1.107) with the tilting

method as a warm-up exercise. For any $x > \mathrm{a}_+$, the optimal inverse temperature θ^* is given as the supremum of Eq. (5.38) with $F_B = 0$ and $J_C = J_A$.

One may notice by integrating Eqs. (3.124) and (3.199) with respect to θ, that for θ between 0 and $g_A(x)$, we have:

$$J_A(x,\theta) = F_A(w_A \to \infty, \theta) = \frac{1}{2}\int_0^{\frac{\theta}{2}} \mathcal{R}_A(\theta')\mathrm{d}\theta' \qquad (\text{for } \theta \in (0, g_A(x))\,. \tag{5.53}$$

If one uses the interpretation in terms of the partition function of a soft spin model of Sec. 3.4.2, this corresponds to the high-temperature regime (or *paramagnetic phase*) of the system where both the annealed and quenched free energy are equal. Necessarily, the optimal temperature $\theta^*(x)$, if there is one, cannot be in this region since from Eq. (5.38), we want precisely the difference between the two free energies to be as high as possible. We can therefore restrict the range of possible optimal temperature to be in the *ferromagnetic phase*, $\theta > g_A(x)$:

$$\theta^*(x) = \underset{\theta > g_A(x)}{\mathrm{argsup}} \left\{ I_x(\theta) := J_A(x,\theta) - F_A(w_A \to \infty, \theta) \right\}\,. \tag{5.54}$$

Let's compute the derivative with respect to θ of this function $I_x(\theta)$. According to Eq. (3.124) and Eq. (3.199), and the definition of the R-transform given by Eq. (2.139), it is simply given by:

$$I'_x(\theta) = \left(x - g_A^{\langle -1 \rangle}(\theta) \right) \qquad (\text{for } \theta > g_A(x))\,. \tag{5.55}$$

Here the function $g_A^{\langle -1 \rangle}(\theta) = \mathcal{R}_A(\theta) + 1/\theta$ contains the inverse of both branches of the Stieltjes transform. It is decreasing until it reaches the value $\theta = g_A(\mathrm{a}+)$ and then it is increasing until it reaches the (possibly infinite) value, $\theta = \mathrm{r}_A = \bar{g}_A(w_A \to \infty)$ where it goes to infinity. Conversely, the function $I'_x(\theta)$ of Eq. (5.55), seen as function of θ for x fixed, starts at zero at $\theta = g_A(x)$ and then is increasing until it reaches the point $\theta = g_A(\mathrm{a}+)$ and then decreasing again, and goes to $-\infty$ as $\theta \to \mathrm{r}_A$. Thus, as one varies θ starting at $g_A(x)$, this continuous function is positive and then negative and only crosses the real axis once. As a consequence, the supremum in Eq. (5.54) is a maximum and this maximum is unique. This maximum θ^* is given at the unique point where the function $I'_x(\theta)$ of Eq. (5.55) crosses the real axis in the region $\theta > g_A(\mathrm{a}_+)$. In other words, finding θ^* amounts to solve the equation:

$$x = g_A^{\langle -1 \rangle}(\theta^*(x)) \qquad \text{for } \theta^*(x) > g_A(\mathrm{a}_+)\,, \tag{5.56}$$

which is nothing else than the definition of the second branch of the Stieltjes transform, that is we have:

$$\theta^*(x) = \bar{g}_A(x)\,. \tag{5.57}$$

If we now use the integral representation Eq. (5.45) of the rate function, with the expression of Eq. (3.125) for the partial derivative of the quenched free energy, together with Eq. (5.57) for the expression of θ^*, we recover Eq. (1.107) as expected. A plot of the function I'_x, for $\mathbf{A}$ a GOE matrix, is given in Fig. 5.5.

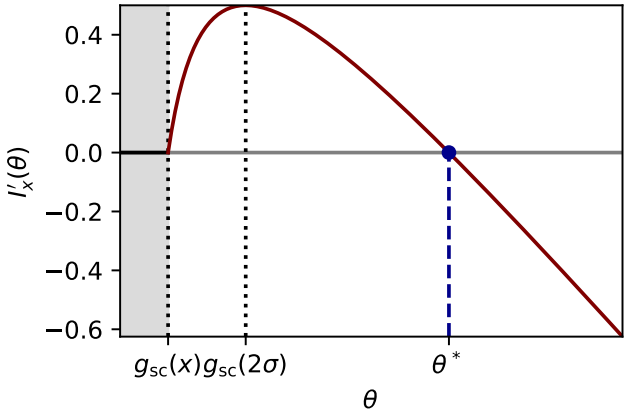


Figure 5.5: Derivative of the difference of the free energy in the case of one GOE random matrix with $\sigma = 1$ and $x = 3$, given as the argument of Eq. (5.54). For $\theta \leq g_{\text{sc}}(x)$, this function (in black) is null since the two free energies are equals, see Eq. (5.53), and this corresponds to the paramagnetic phase. For $\theta \geq g_{\text{sc}}(x)$, this function (in brown) is increasing and then decreasing with a maximum at $g_{\text{sc}}(\text{a}_+ = 2\sigma) = 1$, and this corresponds to the spin glass phase. The optimal inverse temperature (in blue) corresponds to the value where this function crosses the real axis in the spin glass phase.

5.5.2 Optimal inverse temperature for the sum

Let's now consider the general case given by Eq. (5.48). Without loss of generality, we can consider[2]

$$\bar{g}_A(w_A) \leq \bar{g}_B(w_B)\,. \tag{5.58}$$

We further assume the non-trivial condition:

$$g_C(\text{c}_+) < \infty\,, \tag{5.59}$$

as for $g_C(\text{c}_+) = \infty$ (which necessarily implies $g_A(\text{a}_+) = g_B(\text{b}_+) = \infty$ by property of the free convolution), the right large deviation is infinite for any $x > \text{c}_+$. As in the previous section, we first want to show that the supremum in Eq. (5.38) is attained at a unique point, where F_C is given by the sum of Eq. (5.47) and the annealed free energies $F_A(.)$ and $F_B(.)$ are given by Eq. (3.195). Since the function $I_x(\theta)$ is given as the sum of three piece-wise functions, let's first note that we have the following set of inequalities:

$$g_C(x) \leq g_C(\text{c}_+) \leq g_A(\text{a}_+) \leq \bar{g}_A(w_A) \leq \bar{g}_B(w_B)\,. \tag{5.60}$$

The first inequality is due to the fact that the Stieltjes is decreasing for $x > \text{c}_+$. The second inequality is a property of the free convolution. The third is due to the second branch of the Stieltjes transform being monotonically increasing. The fourth is the previously mentioned convention of Eq. (5.58). Using the asymptotics of Eqs. (3.124) (3.195) for the quenched and annealed free energies, together with the linearizing property of the R-transform, one has the

[2]For general $\mathbf{A}$ and $\mathbf{B}$, having $w_A \leq w_B$ does not imply $\bar{g}_A(w_A) \leq \bar{g}_B(w_B)$ nor its converse. One can even come up with examples where $\mathbf{A}$ has a no wall ($w_A \to \infty$) while $\mathbf{B}$ has a finite wall w_B but still $\bar{g}_A(w_A) \leq \bar{g}_B(w_B)$.

following behavior for the difference between the derivative of the annealed and quenched free energy:

$$I'_x(\theta) = \begin{cases} 0 & \text{for } \theta \leq g_C(x)\,, \\ x - g_C^{\langle -1 \rangle}(\theta) & \text{for } g_C(x) \leq \theta \leq \bar{g}_A(w_A)\,, \\ x - w_A - \mathcal{R}_B(\theta) & \text{for } \bar{g}_A(w_A) \leq \theta \leq \bar{g}_B(w_B)\,, \\ x - w_A - w_B + \frac{1}{\theta} & \text{for } \theta \geq \bar{g}_B(w_B)\,, \end{cases} \tag{5.61}$$

where if $\bar{g}_B(w_B) = \infty$ one has to remove the last line and similarly if $\bar{g}_A(w_A) = \infty$, one has to remove the last two lines. This function is represented in Fig. 5.6 for different values of x. Let's look at each interval separately.

1. For $\theta < g_C(x)$, we are in the paramagnetic phase where both the annealed and the quenched free energy are equal. Since for each $x > \mathrm{c}_+$, we want again the difference between the two to be as high as possible, the optimal inverse temperature is not in this region of the phase space.

2. For $g_C(x) \leq \theta \leq \bar{g}_A(w_A)$, as we have seen in the simple case of one invariant matrix, the function $\theta \mapsto x - g_C^{\langle -1 \rangle}(\theta)$ is increasing until it reaches $g_C(\mathrm{c}_+)$ and then decreasing.

3. For $\bar{g}_A(w_A) \leq \theta \leq \bar{g}_B(w_B)$, since the R-transform is increasing, the function $\theta \mapsto x - w_A - \mathcal{R}_B(\theta)$ is decreasing.

4. For $\bar{g}_A(w_A) \leq \theta \leq \bar{g}_B(w_B)$, the function $\theta \mapsto x - w_A - w_B + \frac{1}{\theta}$ is decreasing.

One can easily check that the function of Eq. (5.61) is continuous at each point where its behavior changes. At $\bar{g}_A(w_A)$ it is equal to:

$$I'_x(\bar{g}_A(w_A)) = \frac{1}{2}\left(x - g_C^{\langle -1 \rangle}(\bar{g}_A(w_A))\right)\,, \tag{5.62}$$

and at $\bar{g}_B(w_B)$, it is equal to:

$$I'_x(\bar{g}_B(w_B)) = x - w_A - w_B + \frac{1}{\bar{g}_B(w_B)}\,. \tag{5.63}$$

To summarize, in the spin glass phase $\theta \geq g_C(x)$ the function of Eq. (5.61) is continuously increasing until $\theta = g_C(\mathrm{c}_+)$ and then it is continuously decreasing.

For $x > w_A + w_B$, it is easy to check that this function never crosses the real axis for values of $\theta > g_C(x)$, as a consequence,

$$\theta^*(x) = \infty \qquad \text{for } x > w_A + w_B\,. \tag{5.64}$$

This is expected because for the sum of two matrices we have the classical inequality:

$$\lambda_1(\mathbf{C}) \leq \lambda_1(\mathbf{A}) + \lambda_1(\mathbf{B})\,, \tag{5.65}$$

and since by definition of the walls, $\lambda_1(\mathbf{A}) \leq w_A$ and $\lambda_1(\mathbf{B}) \leq w_B$, the top eigenvalue of $\mathbf{C}$ cannot exceed $w_A + w_B$. So we find that the rate function is infinite for $x > w_A + w_B$.

Otherwise, for values of $x < w_A + w_B$, this function always crosses the real axis once in this region, that is we satisfy the condition (C4) of Sec. 5.4. The correct equation for $\theta^*(x)$ - the point where the function $I'_x(.)$ touches the real axis, see Eq. (5.43) - depends on if the value of this function at $\bar{g}_A(w_A)$ or $\bar{g}_B(w_B)$ is above or below zero, and hence on the value of x. There exist three possible cases, separated by two critical points, x_{c_1} and x_{c_2} defined respectively as the solution of the RHS of Eq. (5.62) and the RHS of Eq. (5.63)) being equal to zero, that is:

$$x_{c_1} := g_C^{\langle -1 \rangle}(\bar{g}_A(w_A)) = w_A + \mathcal{R}_B(\bar{g}_A(w_A))\,, \tag{5.66}$$

and

$$x_{c_2} := w_A + w_B - \frac{1}{\bar{g}_B(w_B)}\,. \tag{5.67}$$

Note that $x_{c_1} \leq x_{c_2}$ as we have postulated Eq. (5.58). We have:

1. for $\mathrm{c}_+ < x < x_{c_1}$, the optimal inverse temperature is attained in the region $g_C(x) \leq \theta \leq \bar{g}_A(w_A)$ and so replacing in Eq. (5.43) the expression of the difference of the free energies by the top line of Eq. (5.61), it is the solution of the same equation (5.56) as the one in the simple one invariant random matrix case (with $g_A^{\langle -1 \rangle}$ replaced by $g_C^{\langle -1 \rangle}$) and thus we have:

$$\theta^*(x) = \bar{g}_C(x)\,, \tag{5.68}$$

   where $\bar{g}_C(.)$ is defined as the inverse of $g_C^{\langle -1 \rangle}(.)$ for values beyond $g_C(\mathrm{c}_+)$.

2. For $x_{c_1} < x < x_{c_2}$, the optimal inverse temperature is attained in the region $\bar{g}_A(w_A) \leq \theta \leq \bar{g}_B(w_B)$ and so from the expression of the second line of the RHS of Eq. (5.61), it is the solution of:

$$\mathcal{R}_B\left(\theta^*(x)\right) = x - w_A\,, \tag{5.69}$$

   since the R-transform is continuously increasing, it has an inverse which we denote by $\mathcal{R}_B^{\langle -1 \rangle}$ so that the optimal temperature is given by:

$$\theta^*(x) = \mathcal{R}_B^{\langle -1 \rangle}\left(x - w_A\right)\,. \tag{5.70}$$

3. For $x_{c_2} < x < w_A + w_B$, $\theta^*(x)$ is attained in the region $\theta \geq \bar{g}_B(w_B)$ and so solving the third line of Eq. (5.61) being equal to zero, we have:

$$\theta^*(x) = \frac{1}{w_A + w_B - x}\,. \tag{5.71}$$

One can check that the piecewise function $\theta^*(x)$ is actually continuously increasing.

5.5.3 Expression for the rate function

Now that we have the expression for the optimal temperature, we can get the expression for the (right) rate function $\Psi_C(.)$ thanks to Eq. (5.45) and this gives the following result:

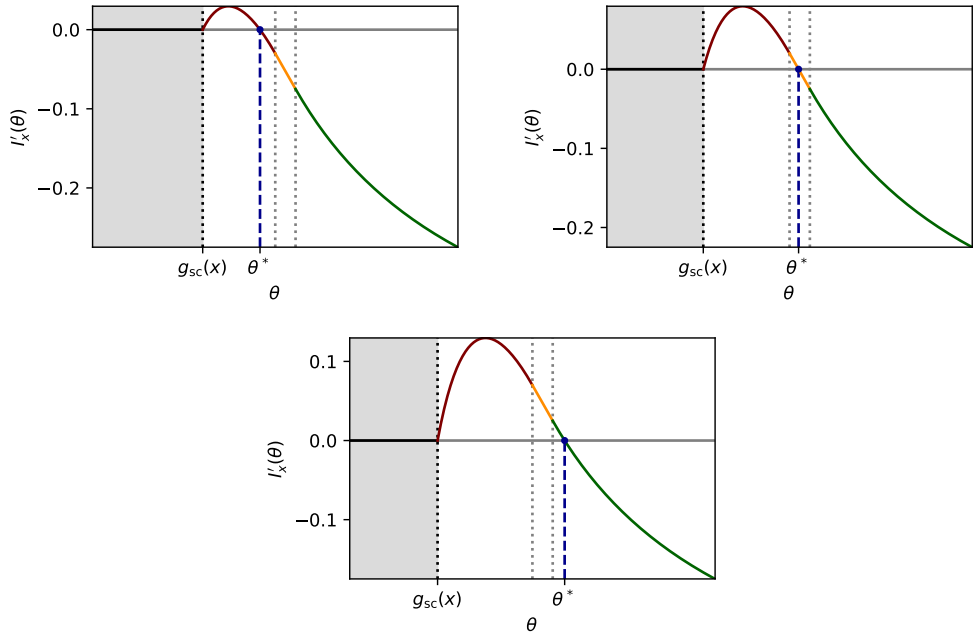


Figure 5.6: Representation of the function of Eq. (5.61) for $\mathbf{A}$ and $\mathbf{B}$ two matrices from a GOE ensemble with a wall at their edge, with $\sigma_A = 1$ and $\sigma_B = 9/10$ and for different values of x. Each color represents the correct expression of this piecewise continuous function in a given interval. In the upper left, $x = 2.75$, and the optimal inverse temperature is attained in the first region where the brown curve crosses the real axis. In the upper right, $x = 2.85$, and the optimal inverse temperature is attained in the second region where the yellow curve crosses the real axis. In the center, $x = 2.95$ and the optimal inverse temperature is attained in the third region where the green curve crosses the real axis.

Result 5.6 (*LDP for the sum*)

In the large N limit, the top eigenvalue of the sum of two 'full-rank' matrices with LSD μ_C satisfies a large deviation principle with speed $N\beta$ and rate function given by

$$\Psi_C(x) = \frac{1}{2}\begin{cases} \displaystyle\int_{\mathrm{c}_+}^{x} (\bar{g}_C(t) - g_C(t))\,\mathrm{d}t & \text{for } \mathrm{c}_+ \le x \le x_{c_1}\,, \\ \displaystyle K_1 + \int_{x_{c_1}}^{x} \left(\mathcal{R}_B^{\langle -1\rangle}(t - w_A) - g_C(t)\right)\mathrm{d}t & \text{for } x_{c_1} \le x \le x_{c_2}\,, \\ \displaystyle K_2 + \log\left(\frac{1}{w_A + w_B - x}\right) - \int_{x_{c_2}}^{x} g_C(t)\mathrm{d}t & \text{for } x_{c_2} \le x \le w_A + w_B\,, \end{cases} \tag{5.72}$$

where x_{c_1} and x_{c_2} are given respectively by Eq. (5.66) and Eq. (5.67) and the two constants $K_1 := \int_{\mathrm{c}_+}^{x_{c_1}} (\bar{g}_C(t) - g_C(t))\,\mathrm{d}t$ and $K_2 := K_1 + \log\left(1/\bar{g}_B(w_B)\right) + \int_{x_{c_1}}^{x_{c_2}} \left(\mathcal{R}_B^{\langle -1\rangle}(t - w_A) - g_C(t)\right)\mathrm{d}t$ are such that Ψ_C is continuous at x_{c_1} and x_{c_2}.

Remark *(Tracy-Widom '3/2'-scaling near the edge)***.** Since the first regime matches the ones of the classical case of one random matrix given by Eq. (1.107), we retrieve in particular the Tracy-Widom '3/2'-scaling of Eq. (1.114) near the edge for the free convolution of non-degenerate densities, as recently investigated in Ref. [97]. ⌟

Remark *(Number of critical points)***.** In general, if one is considering the sum of two random matrices taken from an invariant ensemble with a wall, then the rate function has two possible critical points. However, if $\bar{g}_A(w_A) = \bar{g}_B(w_B)$, then from Eqs. (5.66) (5.67), one can see that two critical points merge, and we have at most one critical point. In particular, this happens when one is considering the free sum $\tilde{\mathbf{A}} + \mathbf{O}\tilde{\mathbf{A}}\mathbf{O}^{\mathsf{T}}$, where $\tilde{\mathbf{A}}$ is a fixed diagonal matrix. Since x_{c_1} diverges if $w_A \to \infty$ while x_{c_2} diverges if either $w_A \to \infty$ or $w_B \to \infty$, the free sum of a two fixed diagonal and a fluctuating matrix (no wall) has at most one critical point while the sum of two fluctuating matrices never has a critical point, in this case, the rate function is the same as that of an invariant ensemble Eq. (1.107). ⌟

Remark *(Interpretation of the three regimes)***.** For $c_+ \leq x \leq x_{c_1}$ the rate function is the same as the one from an invariant ensemble. In this regime, the eigenvalues (including rare large ones) of the matrix $\mathbf{C}$ behave exactly as an invariant ensemble with the potential (and its analytical continuation outside the segment $[c_-, c_+]$) $V_C(.)$ compatible with the limiting density $\mu_C(.)$, that is, $V_C(.)$ is given by Eq. (1.75). In particular, walls (if any) do not modify the rate function in this regime. For $x_{c_1} \leq x \leq x_{c_2}$, the wall w_A starts to matter, and the derivative of the rate function is now the same as if the matrix $\mathbf{A}$ were replaced by a rank-one matrix with eigenvalue w_A but with the still correct $g_C(x)$. Finally, for $x_{c_1} \leq x \leq w_A + w_B$, both walls matter and the derivative of the rate function is now the same as for the sum of two rank-one matrices with eigenvalues w_A and w_B, again up to the correct $g_C(x)$ (see Sec. 5.2). In particular for $\beta = 1$, very close to the maximal value $w_A + w_B$, we have:

$$\mathbb{P}\left[\lambda_1\left(\mathbf{C}\right) \simeq (w_A + w_B)(1-\epsilon)\right] \approx \epsilon^{N/2} \qquad \text{for } \epsilon \ll 1\,, \tag{5.73}$$

which is the asymptotic probability of two random vectors having a squared overlap of order $1-\epsilon$ in dimension N. ⌟

Example *(Free sum of two diagonal semi-circle matrices)***.** In this paragraph, we will compute explicitly the rate function $\Psi_C(x)$ for a matrix $\mathbf{C} = \tilde{\mathbf{A}} + \mathbf{O}\tilde{\mathbf{B}}\mathbf{O}^{\mathsf{T}}$, where $\tilde{\mathbf{A}} = \tilde{\mathbf{B}} = \mathbf{Diag}(\lambda_1, \ldots, \lambda_N)$ are two fixed diagonal matrices with the semi-circle distribution of Eq. (1.40) as their limiting spectrum. Without loss of generality, let's consider that their variances are given by $\sigma_A = 1$ and by $\sigma_B^2 \equiv \sigma^2 \leq 1$ respectively. The computation is equivalent to the sum of two invariant random matrices in quadratic potentials $V_i(x) = x^2/(2\sigma_i^2)$ and a wall at $w_i = 2\sigma_i$ for $i = A, B$ respectively. The free convolution of two semi-circle distributions is again a semi-circle distribution, with variance the sum of the variance. In other words, the limiting law and Stieltjes transform of the matrix $\mathbf{C}$ are given respectively by Eq. (1.40) and Eq. (1.41) with $\sigma_C = \sqrt{1+\sigma^2}$. Since we have $g_A(2) \leq g_B(2\sigma)$ and the R-transform of the semi-circle distribution is given by Eq. (2.142), the optimal inverse

temperature is given by:

$$\theta^*(x) = \begin{cases} \frac{x+\sqrt{x^2-4(1+\sigma^2)}}{2(1+\sigma^2)} & \text{for } 2\sqrt{1+\sigma^2} \le x \le 2+\sigma^2\,, \\ \frac{x-2}{\sigma^2} & \text{for } 2+\sigma^2 \le x \le 2+\sigma\,, \\ \frac{1}{2+2\sigma-x} & \text{for } 2+\sigma \le x \le 2+2\sigma\,, \end{cases} \tag{5.74}$$

and so the rate function is given for $x \in [2\sqrt{1+\sigma^2}, 2+2\sigma]$ by

$$\Psi_{\text{sc+sc}}(x) = \begin{cases} \frac{x\sqrt{x^2-4(1+\sigma^2)}}{4(1+\sigma^2)} + \log\left(\frac{2\sqrt{1+\sigma^2}}{\sqrt{x^2-4(1+\sigma^2)}+x}\right) & \text{for } 2\sqrt{1+\sigma^2} \le x \le 2+\sigma^2\,, \\ \frac{(x-2)^2}{4\sigma^2(1+\sigma)^2} + \frac{4\sigma^2+x^2-8x+12+x\sqrt{x^2-4(1+\sigma^2)}}{8(1+\sigma^2)} \\ \quad + \frac{1}{4}\log\left(\frac{x^2-2(1+\sigma^2)-x\sqrt{x^2-4(1+\sigma^2)}}{2(1+\sigma^2)^2}\right) & \text{for } 2+\sigma^2 \le x \le 2+\sigma\,, \\ \frac{1}{4}\log\left(\frac{(2+\sigma)(2+\sigma-\sqrt{\sigma(4-3\sigma)})-2(1+\sigma^2)}{2}\right) \\ \quad + \frac{6(1+\sigma^2)-x^2+x\sqrt{x^2-4(1+\sigma^2)}}{8(1+\sigma^2)} \\ \quad + \frac{1}{2}\log\left(\frac{\sigma\left(2+\sigma+\sqrt{\sigma(4-3\sigma)}\right)}{(1+\sigma^2)(2(1+\sigma)-x)\left(x+\sqrt{x^2-4(1+\sigma^2)}\right)}\right) & \text{for } 2+\sigma \le x \le 2+2\sigma\,, \end{cases} \tag{5.75}$$

and is infinite otherwise. This function has been plotted in Fig. 5.7 for $\sigma = 9/10$. ∎

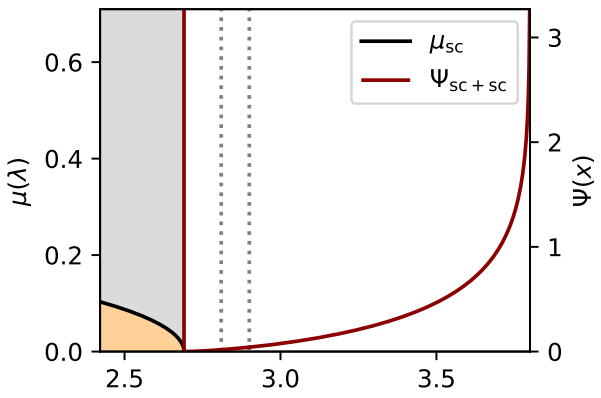


Figure 5.7: The rate function of the largest eigenvalue of the sum of $\mathbf{A}$ and $\mathbf{B}$ from a GOE ensemble with a wall at their edge, with $\sigma_A = 1$ and $\sigma_B = \sigma = 9/10$, see Eq. (5.75).

5.6 Large deviation for the product of self-adjoint matrices

In this section, we consider the case where the matrix $\mathbf{C}$ is given as the symmetric product

$$\mathbf{C} = \sqrt{\mathbf{A}}\mathbf{B}\sqrt{\mathbf{A}}\,, \tag{5.76}$$

where $\mathbf{A} \sim \mathbb{P}_{V_A,w_A}$ and $\mathbf{B} \sim \mathbb{P}_{V_B,w_B}$ are two positive self-adjoint random matrices. In the large N limit, the limiting density $\mu_C = \mu_A \boxplus \mu_B$ of $\mathbf{C}$ is described by the multiplicative free convolution of Chapter 2 (see Sec. 2.6.3) and our goal is to compute the rate function $\Psi_C(x)$:

$$\mathbb{P}\left[\lambda_1(\mathbf{C}) \simeq x\right] \approx \exp\left[-N\beta\,\Psi_C(x) + o(N)\right] \qquad \text{for } x > \mathrm{c}_+\,, \tag{5.77}$$

describing the right large deviation of the top eigenvalue of $\mathbf{C}$ far from its typical value given by the edge c_+ of μ_C. Let us make the following important remark concerning the case of the product of rectangular matrices.

5.6.1 Optimal temperature for the product

Without any loss of generality, let's assume

$$\bar{t}_A(w_A) \leq \bar{t}_B(w_B)\,, \tag{5.78}$$

and

$$\bar{t}_C(\mathrm{c}_+) \leq \infty\,. \tag{5.79}$$

Our goal is to show that the supremum in Eq. (5.38) is attained at a unique point by looking at the derivative $I'_x(.)$ with respect to θ. Paying attention to the bounds in Eqs. (3.144) and (3.206), one has the following behavior:

$$I'_x(\theta) = \begin{cases} 0 & \text{for } \theta \leq t_C(x)\,, \\ \log \dfrac{x}{t_C^{\langle -1\rangle}(\theta)} & \text{for } t_C(x) \leq \theta \leq \bar{t}_A(w_A)\,, \\ \log \dfrac{x}{w_A \tilde{\mathcal{S}}_B(\theta)} & \text{for } \bar{t}_A(w_A) \leq \theta \leq \bar{t}_B(w_B)\,, \\ \log \dfrac{x(\theta+1)}{w_A w_B \theta} & \text{for } \theta \geq \bar{t}_B(w_B)\,. \end{cases} \tag{5.80}$$

Based on a similar monotonous argument as in the additive case of Sec. 5.5.2, one can show that for $\theta \geq t_C(x)$ this function is continuously increasing until it reaches the point $t_C(\mathrm{c}_+)$ and then it is continuously decreasing. For values of $x > w_A\, w_B$, it never crosses the real axis, and we have $\theta^*(x) = \infty$. Otherwise, it crosses the real axis exactly one time, and the equation determining $\theta^*(x)$ depends on the position of x with respect to the two critical points x_{c_1} and x_{c_2} defined by

$$x_{c_1} := t_C^{\langle -1\rangle}(\bar{t}_A(w_A)) = w_A\, \tilde{\mathcal{S}}_B\left(\bar{t}_A(w_A)\right)\,, \tag{5.81}$$

and by:

$$x_{c_2} := w_A w_B \, \frac{\bar{t}_B(w_B)}{\bar{t}_B(w_B)+1} \, . \tag{5.82}$$

1. For $\mathrm{c}_+ < x < x_{c_1}$, θ^* is attained in $(t_C(\mathrm{c}_+), \bar{t}_A(w_A))$ and so setting the second line of the RHS of Eq. (5.80) being equal to zero, gives:

$$\theta^*(x) = \bar{t}_C(x) \, ; \tag{5.83}$$

2. for $x_{c_1} < x < x_{c_2}$, the optimal inverse temperature is attained in the region $\bar{t}_A(w_A) \leq \theta \leq \bar{t}_B(w_B)$ and $\theta^*(x)$ is solution of the third line of the RHS of Eq.(5.80) being equal to zero, that is:

$$\theta^*(x) = \tilde{\mathcal{S}}_B^{\langle -1 \rangle}\left(\frac{x}{w_A}\right) \, ; \tag{5.84}$$

3. for $x_{c_2} < x < w_A w_B$, $\theta^*(x)$ is attained in the region $\theta \geq \bar{t}_B(w_B)$ and so from Eq. (5.80) it is given by:

$$\theta^*(x) = \frac{x}{w_A w_B - x} \, . \tag{5.85}$$

5.6.2 Expression for the rate function

Using the expressions of the previous section for the optimal temperature together with the expression of the partial derivative of the quenched free energy of Eq. (3.145) in Eq. (5.45), we have that the rate function is given by the following result.

Result 5.7 (*LDP for the product*)

In the large N limit, the top eigenvalue of the product of two 'full-rank' matrices with LSD μ_C satisfies a large deviation principle with speed $N\beta$ and rate function given by

$$\Psi_C(x) = \frac{1}{2} \begin{cases} \displaystyle\int_{\mathrm{c}_+}^{x} \left(\bar{g}_C(t) - g_C(t)\mathrm{d}t\right) & \text{for } \mathrm{c}_+ \leq x \leq x_{c_1} \, , \\ \displaystyle K_1 + \int_{x_{c_1}}^{x} \left(\frac{\tilde{\mathcal{S}}_B^{\langle -1 \rangle}\left(\frac{t}{w_A}\right)+1}{t} - g_C(t) \right) \mathrm{d}t & \text{for } x_{c_1} \leq x \leq x_{c_2} \, , \\ \displaystyle K_2 + \log\left(\frac{x}{w_A w_B - x}\right) - \int_{x_{c_2}}^{x} g_C(t)\mathrm{d}t & \text{for } x_{c_2} \leq x \leq w_A w_B \, , \end{cases} \tag{5.86}$$

where x_{c_1} and x_{c_2} are given respectively by Eq. (5.81) and Eq. (5.82) and the two constants $K_1 := \int_{\mathrm{c}_+}^{x_{c_1}} \left(\bar{g}_C(t) - g_C(t)\right) \mathrm{d}t$ and $K_2 := K_1 + \log\left(1/\bar{t}_B(w_B)\right) + \int_{x_{c_1}}^{x_{c_2}} \left(\frac{\tilde{\mathcal{S}}_B^{\langle -1 \rangle}(t/w_A)+1}{t} - g_C(t) \right) \mathrm{d}t$ are such that Ψ_C is continuous at x_{c_1} and x_{c_2}.

Example *(Rate Function for Generalized Wishart).* Let's consider the case in Eq. (5.76) where $\mathbf{B}$ is a (White) Wishart with shape ratio q and $\mathbf{A}$ is a fixed diagonal (positive semi-definite) matrix. In this case, we have $w_A = \mathrm{a}_+$ and $w_B = \bar{t}_B(w_B) = \infty$. As a consequence, there is just one critical point, and using the expression (2.152) for the S-transform of the Wishart matrix, it is given by:

$$x_{c_1} = \mathrm{a}_+(1 + q\, t_A(\mathrm{a}_+))\,, \tag{5.87}$$

and using Eq. (1.44) for the S-transform, we have for the rate function:

$$\Psi_C(x) = \begin{cases} \dfrac{1}{2}\displaystyle\int_{\mathrm{c}_+}^{x} (\bar{g}_C(t) - g_C(t))\,\mathrm{d}t & \text{for } \mathrm{c}_+ \le x \le x_{c_1}\,, \\[2ex] \dfrac{K_1}{2} + \dfrac{1}{2}\left(-\dfrac{1}{q} - t_A(\mathrm{a}_+) + \dfrac{x}{q\mathrm{a}_+} + \left(1 - \dfrac{1}{q}\right)\log\left(\dfrac{x}{x_{c_1}}\right) - \displaystyle\int_{x_{c_1}}^{x} g_C(t)\mathrm{d}t\right) & \text{for } x \ge x_{c_1}\,, \end{cases} \tag{5.88}$$

which is up to a change in the notation, the results obtained in Ref. [116]. ∎

5.7 Large deviation for the top singular value of the sum of rectangular matrices

In this section, we consider the case where the matrix $\mathbf{C}$ is given as

$$\mathbf{C} = \mathbf{A} + \mathbf{B}\,, \tag{5.89}$$

where $\mathbf{A} \sim \mathbb{P}_{V_A, w_A}$ and $\mathbf{B} \sim \mathbb{P}_{V_B, w_B}$ are two *rectangular* matrices, each taken from a bi-invariant ensemble with a wall as defined similarly to the one of Sec. 5.4.2. We aim at computing the rate function $\Psi_C(x)$:

$$\mathbb{P}\left[s_1(\mathbf{C}) = x\right] \approx \exp\left[-N\beta\,\Phi_C(x) + o(N)\right] \qquad \text{for } x > \mathrm{c}_+\,, \tag{5.90}$$

where c_+ is the edge of the limiting density of singular values μ_C of $\mathbf{C}$, described by the *rectangular free convolution* of Chapter 2, see Sec. 2.6.3.

5.7.1 Optimal temperature for the rectangular case

Without loss of generality, we assume:

$$\bar{d}_A(w_A) \le \bar{d}_B(w_B)\,, \tag{5.91}$$

and the non-trivial condition:

$$d_C(\mathrm{c}_+) < \infty\,. \tag{5.92}$$

In this case, the derivative with respect to θ of the function $I_x(.)$ in the supremum of Eq. (5.38) is given by:

$$I'_x(\theta) = \begin{cases} 0 & \text{for } \theta \leq d_C(x)\,, \\ \frac{U(\theta x)}{\theta} - \tilde{\mathcal{C}}_C^{(q)}(\theta) & \text{for } d_C(x) \leq \theta \leq \bar{d}_A(w_A)\,, \\ \frac{U(\theta x)}{\theta} - \frac{U(\theta w_A)}{\theta} - \tilde{\mathcal{C}}_B^{(q)}(\theta) & \text{for } \bar{d}_A(w_A) \leq \theta \leq \bar{d}_B(w_B)\,, \\ \frac{U(\theta x)}{\theta} - \frac{U(\theta w_A)}{\theta} - \frac{U(\theta w_B)}{\theta} & \text{for } \theta \geq \bar{d}_B(w_B)\,. \end{cases} \tag{5.93}$$

By property of the rectangular free convolution, for $\theta > d_C(x)$, this function is first increasing with θ until it reaches the value $\theta = d_C(\mathrm{c}_+)$ and then it is decreasing with θ. One can check that it crosses the real axis if x is in the interval $[\mathrm{c}_+, w_A + w_B]$, and in this case, the position of the optimal inverse temperature θ^* depends on two critical points x_{c_1} and x_{c_2}. The first one is given by

$$x_{c_1} := d_C^{\langle -1\rangle}(\bar{d}_A(w_A))\,. \tag{5.94}$$

Unfortunately, unlike the sum and the product of symmetric matrices, the expression for x_{c_1} in terms of the rectangular C-transform is quite involved. if we introduce the function:

$$f_q(z) := \frac{1}{2}\sqrt{(1-q)^2 + 4qz^2}\,, \tag{5.95}$$

to ease the notation, then we have:

$$x_{c_1} = \sqrt{w_A^2 + \tilde{\mathcal{C}}_B^{(q)}(\bar{d}_A(w_A))\left(q\tilde{\mathcal{C}}_B^{(q)}(\bar{d}_A(w_A)) + \frac{f_q(w_A\bar{d}_A(w_A))}{\bar{d}_A(w_A)}\right)}\,. \tag{5.96}$$

The other critical point x_{c_2} is given by:

$$x_{c_2} := \frac{1}{\bar{d}_B(w_B)} U^{\langle -1\rangle}\left(U(w_A\bar{d}_B(w_B)) + U(w_B\bar{d}_B(w_B))\right)\,, \tag{5.97}$$

where we recall that $U^{\langle -1\rangle}$ is given by Eq. (2.167). Eq. (5.97) can be written in semi-explicit form with the function f_q of Eq. (5.95):

$$x_{c_2} = \frac{\sqrt{\left(f_q\left(w_A\bar{d}_B(w_B)\right) + f_q\left(w_B\bar{d}_B(w_B)\right) - 1\right)\left(f_q\left(w_A\bar{d}_B(w_B)\right) + f_q\left(w_B\bar{d}_B(w_B)\right) - q\right)}}{\sqrt{q}\bar{d}_B(w_B)}\,. \tag{5.98}$$

1. For $\mathrm{c}_+ \leq x \leq x_{c_1}$, θ^* is attained in $(d_C(\mathrm{c}_+), \bar{d}_A(w_A))$ and hence θ^* is the solution of the second line of the RHS of Eq. (5.93) being equal to zero. Using Eq. (2.168) to express the C-transform in terms of the function U, one gets:

$$U(\theta^* x) = U(d_C^{\langle -1\rangle}(\theta^*)\theta^*)\,; \tag{5.99}$$

and so by applying the (monotonous) function $U^{\langle -1\rangle}(.)$ to this equation and dividing by θ^*, one gets to solve the equation:

$$d_C^{\langle -1\rangle}(\theta^*) = x \qquad \left(\text{for } \theta^* \in (d_C(\mathrm{c}_+), \bar{d}_A(w_A))\right) . \tag{5.100}$$

The solution is given by the second branch of the D-transform:

$$\theta^*(x) = \bar{d}_C(x) . \tag{5.101}$$

2. For $x_{c_1} \leq x \leq x_{c_2}$, the optimal inverse temperature is attained in the region $\bar{d}_A(w_A) \leq \theta \leq \bar{d}_B(w_B)$. $\theta^* \equiv \theta^*(x)$ is solution of the third line of the RHS of Eq.(5.93) being equal to zero so that it satisfies:

$$U(x\theta^*) := U(w_A\theta^*) + \tilde{\mathcal{C}}_B^{(q)}(\theta^*)\theta^* . \tag{5.102}$$

If one applies the function $U^{\langle -1\rangle}$ on each side, one gets after simplification an analytical expression for the function $\theta^* \mapsto x(\theta^*)$:

$$x(\theta^*) = \sqrt{w_A^2 + \tilde{\mathcal{C}}_B^{(q)}(\theta^*)\left(q\tilde{\mathcal{C}}_B^{(q)}(\theta^*) + \frac{f_q(w_A\theta^*)}{\theta^*}\right)} , \tag{5.103}$$

which is by definition the inverse of the function $\theta^*(x)$. Now unfortunately for general values of the parameter q, we do not have a simple analytical formula for the optimal inverse temperature function of the position x and so we simply denote by F_1 the solution of Eq. (5.103) with unknown θ^* for x between x_{c_1} and x_{c_2}.

3. For $x_{c_2} \leq x \leq w_A + w_B$, $\theta^*(x)$ is attained in the region $\theta \geq \bar{d}_B(w_B)$ and so from Eq. (5.93) and after simplification, one gets the following analytical expression for the function $\theta^* \mapsto x(\theta^*)$:

$$x(\theta^*) = \frac{\sqrt{(f_q(w_A\theta^*) + f_q(w_B\theta^*) - 1)(f_q(w_A\theta^*) + f_q(w_B\theta^*) - q)}}{\sqrt{q}\theta^*} . \tag{5.104}$$

In full generality, one can isolate one of the radical functions and take the square of the newly obtain equation and repeat the process until it becomes a polynomial equation. In our setting and for general q, w_A, w_B, one would obtain that $(\theta^*)^2$ is one of the zeros of a polynomial of degree 8, and hence there is no hope of finding an analytical expression for θ^*. As a consequence, we simply denote by F_2 the solution (in θ) of Eq. (5.104) for x higher than x_{c_2}. Now for specific values of w_A and w_B, for example $w_A = w_B$, or q, for example $q = 0$ or $q = 1$, Eq. (5.104) becomes, after some work, a quadratic or even linear equation for $(\theta^*)^2$ (or θ), as we will see.

Remark *(Simplification for the case of long ($q = 0$) matrices).* For $q \to 0$, one has:

- For $x_{c_1} \leq x \leq x_{c_2}$, Eq. (5.103) for the optimal inverse temperature simplifies into:

$$x = \sqrt{w_A + \frac{\tilde{\mathcal{C}}_B^{(0)}(\theta^*)}{\theta^*}} , \tag{5.105}$$

and since the rectangular C-transform of a long matrix is related to R-transform by Eq. (2.173), we have:

$$\theta^*(x) = \sqrt{\mathcal{R}_{BB^*}^{\langle -1\rangle}(x^2 - w_A^2)} . \tag{5.106}$$

- For $x_{c_2} \leq x \leq w_A + w_B$, Eq. (5.104) for the optimal inverse temperature simplifies into:

$$x = \frac{\sqrt{\theta^{*2}\left(w_A^2 + w_B^2\right) - 1}}{\theta^*}\,, \tag{5.107}$$

such that the optimal temperature is given by:

$$\theta^*(x) = \frac{1}{\sqrt{w_A^2 + w_B^2 - x^2}}\,. \tag{5.108}$$

⌟

Remark *(Simplification for the case of square ($q = 1$) matrices).* For $q \to 1$, one has:

- For $x_{c_1} \leq x \leq x_{c_2}$, Eq. (5.103) for the optimal inverse temperature simplifies into:

$$x - w_A = \mathcal{C}_B^{(1)}(\theta^*)\,, \tag{5.109}$$

and since for $q = 1$, the rectangular C-transform is the R-transform of the symmetrized density (see Eq. (2.176)), we have:

$$\theta^*(x) = \mathcal{R}_{\widehat{B}}^{\langle -1\rangle}(x - w_A)\,. \tag{5.110}$$

- For $x_{c_2} \leq x \leq w_A + w_B$, Eq. (5.104) for the optimal inverse temperature simplifies into:

$$\left(x\theta^*(x)\right)^2 = \left(\theta^*(x)\left(w_A + w_B\right) - 1\right)^2\,, \tag{5.111}$$

and hence we have:

$$\theta^*(x) = \frac{1}{w_A + w_B - x}\,. \tag{5.112}$$

⌟

5.7.2 Expression for the rate function

Using the general expression of Eq. (5.45) for the rate function with the expression of Eq. (3.176) and the expression of the optimal inverse temperature of the previous section, we have the following result for the rate function.

Result 5.8 (*LDP for the rectangular sum*)

In the large N limit, the top singular value of the sum of two 'full-rank' rectangular matrices with LSVD μ_C satisfies a large deviation principle with speed $N\beta$ and rate function given

by

$$\Psi_C(x) = \begin{cases} \dfrac{1}{2}\displaystyle\int_{\mathrm{c}_+^2}^{x^2} (\bar{g}_{CC^*}(t) - g_{CC^*}(t))\,\mathrm{d}t & \text{for } \mathrm{c}_+ \leq x \leq x_{c_1}\,, \\ K_1 + \displaystyle\int_{x_{c_1}}^{x} \frac{f_q\left(F_1(t)t\right) - f_q\left(d_C(t)t\right)}{qt}\mathrm{d}t & \text{for } x_{c_1} \leq x \leq x_{c_2}\,, \\ K_2 + \displaystyle\int_{x_{c_2}}^{x} \frac{f_q\left(F_2(t)t\right) - f_q\left(d_C(t)t\right)}{qt}\mathrm{d}t & \text{for } x_{c_2} \leq x \leq w_A + w_B\,, \end{cases} \tag{5.113}$$

where x_{c_1}, x_{c_2} are defined by Eq. (5.94) and Eq. (5.97) and F_1 (resp. F_2) is defined as the (correct) solution with unknown θ^* of Eq. (5.103) (resp. Eq. (5.104)) and K_1 and K_2 are the constants such that this rate function is continuous.

Let's conclude with two remarks concerning the limiting cases $q \to 0$ and $q \to 1$.

Remark *(Rate function for the sum of long ($q \to 0$) matrices).* In this case, we have:

$$\Psi_C(x) = \Psi_{AA^*+BB^*}(x^2)\,, \tag{5.114}$$

that is,

$$\Psi_C(x) = \frac{1}{2}\begin{cases} \displaystyle\int_{\mathrm{c}_+^2}^{x^2} (\bar{g}_{AA^*+BB^*}(t) - g_{AA^*+BB^*}(t))\,\mathrm{d}t & \text{for } \mathrm{c}_+ \leq x \leq x_{c_1}\,, \\ K_1 + \displaystyle\int_{x_{c_1}^2}^{x^2} \left(\mathcal{R}^{\langle -1\rangle}_{BB^*}(t - w_A^2) - g_{AA^*+BB^*}(t)\right)\mathrm{d}t & \text{for } x_{c_1} \leq x \leq x_{c_2}\,, \\ K_2 + \log\left(\dfrac{1}{w_A^2 + w_B^2 - x^2}\right) - \displaystyle\int_{x_{c_2}^2}^{x^2} g_{AA^*+BB^*}(t)\mathrm{d}t & \text{for } x_{c_2} \leq x \leq w_A + w_B\,. \end{cases} \tag{5.115}$$

where $g_{AA^*+BB^*}$ is the Stieltjes transform of the free convolution $\mu_{AA^*} \boxplus \mu_{BB^*}$. ⌟

Remark *(Rate function for the sum of square ($q = 1$) matrices).* In this case, we have:

$$\Psi_C(x) = 2\Psi_{\widehat{A}+\widehat{B}}(x) = \begin{cases} \displaystyle\int_{\mathrm{c}_+}^{x} \left(\bar{g}_{\widehat{A}+\widehat{B}}(t) - g_{\widehat{A}+\widehat{B}}(t)\right)\mathrm{d}t & \text{for } \mathrm{c}_+ \leq x \leq x_{c_1}\,, \\ K_1 + \displaystyle\int_{x_{c_1}}^{x} \left(\mathcal{R}^{\langle -1\rangle}_{\widehat{B}}(t - w_A) - g_{\widehat{A}+\widehat{B}}(t)\right)\mathrm{d}t & \text{for } x_{c_1} \leq x \leq x_{c_2}\,, \\ K_2 + \log\left(\dfrac{1}{w_A + w_B - x}\right) - \displaystyle\int_{x_{c_2}}^{x} g_{\widehat{A}+\widehat{B}}(t)\mathrm{d}t & \text{for } x_{c_2} \leq x \leq w_A + w_B\,. \end{cases} \tag{5.116}$$

where $g_{\widehat{A}+\widehat{B}}$ is the Stieltjes transform of the free convolution of the symmetrized distributions $\widehat{\mu}_A \boxplus \widehat{\mu}_B$. ⌟

Example *(Free sum of fixed diagonal quarter-circle distribution)*. Let's consider two diagonal square matrices $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ with LSVD given by the quater-circle distribution of Eq. (1.52) where without loss of generality we take the variances to be respectively equal to $\sigma_A = 1$ and $\sigma_B \leq 1$ then we have that the rate function $\Phi_{\mathrm{qc+qc}}$ associated to the large deviation of the top singular value of the sum $\tilde{\mathbf{A}} + \mathbf{U}\tilde{\mathbf{B}}\mathbf{V}^{\mathsf{T}}$ is given by:

$$\Phi_{\mathrm{qc+qc}}(x) = 2\Psi_{\mathrm{sc+sc}}(x)\,, \tag{5.117}$$

with $\Psi_{\mathrm{sc+sc}}$ given by Eq. (5.75). ∎

5.8 Summary and conclusion of Chapter 5

In this chapter, we have derived the right large deviation function for:

1. the top eigenvalue of the sum of two *generic* self-adjoint matrices;
2. the top eigenvalue of the product of two *generic* self-adjoint matrices;
3. the top singular of the sum of two *generic* rectangular matrices;

whereby 'generic' means that we can take the matrices to be either taken from an invariant ensemble (resp. a bi-invariant for the case of rectangular random matrices) or to be a randomly rotated fixed diagonal matrix (resp. a fixed diagonal rectangular matrix). From the point of view of large deviation, the former corresponds to the study of an invariant ensemble with a wall sent to infinity ($w_A \to \infty$) and the latter to the case where the wall is at the edge ($w_A = \mathrm{a}_+$). The results rely on a direct link with spherical spin models and are summarized in Sec. 5.5.3 for the case of the sum of self-adjoint matrices, in Sec. 5.6.2 for the case of the product of self-adjoint matrices and in Sec. 5.7.2 for the case of the sum of rectangular matrices. In each case, we find that the rate function has up to three different regimes, and we give an interpretation of the behavior in each regime. A natural question is to extend the above construction to tackle the 'left' large deviation, for which the speed of convergence of the large deviation is N^2.

Chapter 6

The High-temperature convolution: interpolating between classical and free convolutions

The results of this chapter are based on the paper [142] and the preprint [140]. They are closely related to the high-temperature ensemble of Chapter 1 (see. Sec. 1.7), the free convolution and finite free convolution of Chapter 2 (see Secs. 2.6.3) and the asymptotic behavior of the rank-one additive spherical integral (see Sec. 3.4).

6.1 Introduction and Motivation

In this chapter, we aim at constructing a family of convolutions, indexed by a continuous parameter c, such that as one varies c, one continuously interpolates between the classical convolution and the free convolution. They are at least two (related) reasons why one would like to construct such a family of convolutions.

1. First, as we have seen in Chapter 2, free convolution (and more generally free probability) deals with objects which are in 'generic position' or 'maximally non-commutative' in the sense that if $\mathbf{A}$ and $\mathbf{B}$ are asymptotically free, with LSD μ_A and μ_B with zero mean, then if we denote by $\tau(.) := \mathrm{Tr}(.)/N$, we have for large dimensions N, $\tau(\mathbf{A}^2\mathbf{B}^2) = \tau(\mathbf{A}^2)\tau(\mathbf{B}^2) = m_2[\mu_A]m_2[\mu_B]$ with $m_2[\mu] := \int x^2 \mathrm{d}\mu$ but $\tau(\mathbf{ABAB}) \simeq 0 \neq \tau(\mathbf{A}^2)\tau(\mathbf{B}^2)$. On the other hand, classical probability naturally deals with 'commuting objects' for example if one looks at the matrix $\mathbf{Diag}(\boldsymbol{a}) + \mathbf{PDiag}(\boldsymbol{b})\mathbf{P}^\mathsf{T}$ where $\mathbf{P} \sim \mathrm{Unif}(\mathsf{S}(N))$ is a (uniform) random permutation matrix, one is summing diagonal matrices and the limiting spectral distribution for the sum is now simply given by the classical convolution $\mu_A * \mu_B$. Thus, classical convolution naturally appears in Random Matrix Theory (RMT) when the eigenbasis of both symmetric matrices are perfectly aligned or said differently when one is looking at the spectrum of large commutative self-adjoint objects. In particular, if now $\mathbf{B} := \mathbf{PDiag}(\mathbf{b})\mathbf{P}^\mathsf{T}$ and μ_A and μ_B have zero mean, we have $\tau(\mathbf{ABAB}) = \tau(\mathbf{A}^2)\tau(\mathbf{B}^2)$. A natural (and difficult) problem is to give meaning to 'intermediate cases', that is, to describe the spectrum of the sum of two large self-adjoint objects for cases where those objects are not commutative nor maximally non-commutative.

2. Second, we have also seen in Chapter 2 that one can give a meaning for the sum of low-temperature ensembles of Chapter 1 (where $\beta \to \infty$ and N is fixed) thanks to the finite free convolution. A natural task is to complete the picture by constructing a similar operation for the high-temperature ensemble of Chapter 1, Sec. 1.7, summarized in the following table.

convolution	operation on corresponding ensemble
free convolution	sum of β-ensembles with $N \to \infty$ and $\beta > 0$ fixed
finite free convolution (FFC)	sum of β-ensembles with N fixed and $\beta \to \infty$
high-temperature convolution	sum of β-ensembles with $N \to \infty$ and $N\beta/2 \to c$

   For *individual* ensemble, we have also seen that the limit $\beta \to \infty$ with N fixed and the double limit $N \to \infty$ with $N\beta/2 \to c$ share a certain high-low temperature duality (see Sec. 1.7) and a natural question is to know if this duality will extend between the high-temperature convolution (HTC) and the FFC.

The precise description of this high-temperature convolution will be done in the rest of the chapter, but for now one can think of it as an operation taking a parameter $c \in (0, \infty)$ and two distributions μ_A and μ_B as inputs and giving a distribution denoted by $\mu_A \oplus_c \mu_B$ as output and such that $\mu_A \oplus_{c\to 0} \mu_B \equiv \mu_A * \mu_B$ and $\mu_A \oplus_{c\to\infty} \mu_B \equiv \mu_A \boxplus \mu_B$. Let us mention that this convolution is done directly at the level of the limiting distributions and finding the corresponding 'linear algebra operation' is a difficult open problem.

The rest of this chapter is organized as follows. In Sec. 6.2, we recall several properties for the rank-one additive spherical integral. In Sec. 6.3, we construct the high-temperature convolution by looking at this spherical integral in the high-temperature regime. In Sec. 6.4 we give a complete analytical example of the high-temperature convolution. In Sec. 6.5 we derive the cumulant-moment relation. In Sec. 6.6 we derive the central limit theorems associated with this high-temperature convolution.

6.2 More on the additive spherical integral for fixed β

6.2.1 Equivalent representations for the rank-one spherical integral at any $\beta > 0$

Before jumping to the high-temperature regime of the additive spherical, let's derive and recall some properties for fixed $\beta > 0$.

Power series representation -

As we have seen in Chapter 3 (see Eq. (3.50)), for any $\beta > 0$ the additive spherical integral admits an expansion in terms of Jack polynomials, which are defined by the generalized Cauchy identity of Eq. (3.45). Let's now look at this expansion in the specific case where one of the

vectors is of rank one $\boldsymbol{t} = (\theta, 0, \dots, 0)$. To do, let's introduce the polynomials $\mathrm{g}_k^{(\beta)}$ defined by the Cauchy identity of Eq. (3.45) when one of the vector is of rank one:

$$\prod_{i=1}^{N}(1 - a_i\theta)^{-\beta/2} := \sum_{k=0}^{\infty} \mathrm{g}_k^{(\beta)}(\boldsymbol{a})\theta^k \,. \tag{6.1}$$

These polynomials can be expressed in terms of the moments $m_k(\boldsymbol{a}) := \sum_{i=1}^{N} a_i^k/N$ of the a_i's thanks to the following combinatorial formula:

$$\mathrm{g}_k^{(\beta)}(\boldsymbol{a}) = \sum_{1j_1+\cdots+kj_k=k} \left(\frac{N\beta}{2}\right)^{j_1+\cdots+j_k} \prod_{i=1}^{k} \frac{m_i(\boldsymbol{a})^{j_i}}{i^{j_i} j_i!} \,, \tag{6.2}$$

and they admit the following recurrence relation:

$$k\,\mathrm{g}_k^{(\beta)}(\boldsymbol{a}) = \frac{N\beta}{2} \sum_{l=1}^{k} \mathrm{g}_{k-l}^{(\beta)}(\boldsymbol{a})\, m_l(\boldsymbol{a}) \,. \tag{6.3}$$

For the rank-one specialization of the additive spherical integral, the sum of Eq. (3.50) over partitions becomes a power series, and we have:

$$\boxed{\mathcal{I}_{\boldsymbol{a}}^{(\beta)}(\theta) = \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{N\beta}{2}\right)}{\Gamma\left(\frac{N\beta}{2}+k\right)} \mathrm{g}_k^{(\beta)}(\boldsymbol{a})\,\theta^k} \,. \tag{6.4}$$

Inverse Laplace representation -

Now if multiply and divide this power-sum representation by $\theta^{N\beta/2-1}$, we have:

$$\mathcal{I}_{\boldsymbol{a}}^{(\beta)}(\theta) = \left(\frac{\Gamma(\frac{N\beta}{2})}{\theta^{\frac{N\beta}{2}-1}}\right) \sum_{k=0}^{\infty} \mathrm{g}_k^{(\beta)}(\boldsymbol{a}) \underbrace{\frac{1}{\Gamma(\frac{N\beta}{2}+k)} \theta^{\frac{N\beta}{2}+k-1}}_{=\mathcal{L}_z^{-1}\left[\frac{1}{z^{\frac{N\beta}{2}+k}}\right](\theta)} \,, \tag{6.5}$$

$$\mathcal{I}_{\boldsymbol{a}}^{(\beta)}(t) = \left(\frac{\Gamma(\frac{N\beta}{2})}{t^{\frac{N\beta}{2}-1}}\right) \sum_{k=0}^{\infty} \mathrm{g}_k^{(\beta)}(\boldsymbol{a})\, \mathcal{L}_z^{-1}\left[\frac{1}{z^{\frac{N\beta}{2}+k}}\right](t) \,, \tag{6.6}$$

$$\mathcal{I}_{\boldsymbol{a}}^{(\beta)}(t) = \left(\frac{\Gamma(\frac{N\beta}{2})}{t^{\frac{N\beta}{2}-1}}\right) \mathcal{L}_z^{-1}\left[\frac{1}{z^{\frac{N\beta}{2}}} \sum_{k=0}^{\infty} \mathrm{g}_k^{(\beta)}(\boldsymbol{a}) \frac{1}{z^k}\right](t) \,. \tag{6.7}$$

Thanks to the Cauchy Eq. (6.1) for the polynomials $g_k(\boldsymbol{a})$, the sum inside the inverse Laplace transform is the characterized polynomial raised to the power $-\beta/2$:

$$\boxed{\mathcal{I}_{\boldsymbol{a}}^{(\beta)}(\theta) = \left(\frac{\Gamma\left(\frac{N\beta}{2}\right)}{\theta^{\frac{N\beta}{2}-1}}\right) \mathcal{L}_z^{-1}\left[U_{\boldsymbol{a}}^{(\beta)}(z)\right](\theta)} \,, \tag{6.8}$$

where the function $U_{\boldsymbol{a}}^{(\beta)}$ is defined by:

$$U_{\boldsymbol{a}}^{(\beta)}(z) := \prod_{i=1}^{N}(z - a_i)^{-\frac{\beta}{2}} \,. \tag{6.9}$$

Differential operator representation -

Now let's compare the expansion of the function $U_{\boldsymbol{a}}^{(\beta)} = \frac{1}{z^{\frac{N\beta}{2}}} \sum_{k=0}^{\infty} \mathrm{g}_k^{(\beta)}(\boldsymbol{a}) \frac{1}{z^k}$ with the one of the spherical integral of Eq. (6.4). They differ by the ratio of Gamma functions and a re-scaling by z. This ratio of Gamma functions exactly corresponds to the coefficient of the rational function obtained after differentiating the function $z^{-N\beta/2}$ since we have:

$$(-1)^k \frac{\mathrm{d}^k}{\mathrm{d}t^k} \left[t^{-\frac{N\beta}{2}} \right] = \frac{\Gamma(\frac{N\beta}{2} + k)}{\Gamma(\frac{N\beta}{2})} t^{-\frac{N\beta}{2} - k} \,. \tag{6.10}$$

As a consequence, we have the following differential operator representation for the additive spherical integral:

$$\boxed{U_{\boldsymbol{a}}^{(\beta)}(z) = \mathcal{I}_{\boldsymbol{a}}^{(\beta)}(-\mathrm{D}_z)\, z^{-\frac{N\beta}{2}}}, \tag{6.11}$$

where $\mathrm{D}_z = \mathrm{d}/\mathrm{d}z$.

Moment generating function representation and finite Markov Krein relation -

Using identities for Dirichlet distribution, one can also express the additive spherical integral as a moment generating function of a convex combination of the a_i' and we have:

$$\boxed{\mathcal{I}_{\boldsymbol{a}}^{(\beta)}(\theta) = \mathbb{E}\left[\mathrm{e}^{\theta X}\right] \qquad \text{where } X \overset{\text{in law}}{=} \boldsymbol{d}^{\mathsf{T}} \boldsymbol{a}} \quad \text{with } \boldsymbol{d} \sim \mathrm{Dir}(\beta/2)\,, \tag{6.12}$$

and we recall that $\mathrm{Dir}(\beta/2)$ denotes the Dirichlet distribution with uniform parameter $\beta/2$, see Eq. (2.36).

Now let's relate the distribution ν_N of X to the a_i. Using the Inverse Laplace transform of Eq. (6.8), we have:

$$U_{\boldsymbol{a}}^{(\beta)}(z) = \frac{1}{\Gamma\left(\frac{N\beta}{2}\right)} \mathcal{L}_\theta \left[\theta^{N\beta/2-1} \mathcal{I}_{\boldsymbol{a}}^{(\beta)}(\theta) \right](z)\,, \tag{6.13}$$

$$U_{\boldsymbol{a}}^{(\beta)}(z) = \frac{1}{\Gamma\left(\frac{N\beta}{2}\right)} \mathcal{L}_\theta \left[\theta^{N\beta/2-1} \mathbb{E}_{X\sim\nu_N}\left[\mathrm{e}^{\theta X}\right] \right](z)\,, \tag{6.14}$$

$$U_{\boldsymbol{a}}^{(\beta)}(z) = \frac{1}{\Gamma\left(\frac{N\beta}{2}\right)} \mathbb{E}_{X\sim\nu_N} \left[\underbrace{\mathcal{L}_\theta \left[\theta^{N\beta/2-1} \mathrm{e}^{\theta X} \right](z)}_{=\Gamma(N\beta/2)\cdot(z-X)^{-N\beta/2}} \right], \tag{6.15}$$

$$U_{\boldsymbol{a}}^{(\beta)}(z) = \mathbb{E}_{X\sim\nu_N}\left[(z-X)^{-N\beta/2} \right], \tag{6.16}$$

or more explicitly:

$$\int \frac{1}{(z-x)^{N\beta/2}} \nu_N(\mathrm{d}x) = \prod_{i=1}^{N} (z-a_i)^{-\beta/2}\,. \tag{6.17}$$

This is the 'finite' counterpart of the Markov-Krein relation we will encounter in the high temperature regime.

Relation to low-temperature ensemble/FFC -

Let's mention that if one analytically extends the Cauchy identity of Eq. (6.1) to $\beta = -2$, then on has $g_k^{(\beta=-2)}(\boldsymbol{a}) = (-1)^k \mathrm{e}_k(\boldsymbol{a})$, where the e_k's are the elementary symmetric polynomials appearing in the definition of the FFC. The same remark holds for the other representations, and in particular, the analytical extension to $\beta = -2$ of the differential operator representation of Eq. (6.11) is nothing else than the definition of the additive form of the FFC. This is the first hint of the high-low temperature duality which will appear when taking the double scaling limit $N \to \infty$ and $N\beta/2 \to c$.

6.2.2 Asymptotic behavior for $\beta > 0$ versus $\beta = 0$

As $N \to \infty$, the asymptotic behavior of the spherical integral depends on if either $\beta > 0$ or $\beta = 0$.

Reminder of the asymptotic behavior for $\beta > 0$ -

As we have seen in Chapter 3, the asymptotic behavior for $\beta > 0$ is given by the R-transform since we have:

$$\frac{2}{N\beta} \log \left[\mathcal{I}_{\boldsymbol{a}}^{(\beta)} \left(\frac{N\beta}{2} \theta \right) \right] \xrightarrow[N\to\infty]{} \int_0^\theta \mathcal{R}_A(t) \mathrm{d}t \qquad \text{for } \theta \text{ small enough} \ , \tag{6.18}$$

and hence the logarithm of the spherical integral is the linearizing transform of the free convolution.

The case $\beta = 0$ and classical convolution -

The situation for the case where we *first* take the limit $\beta \to 0^+$ and then the limit $N \to \infty$ is different. In this limit using either properties of Jack polynomials or of the Dirichlet distribution in the representation of Eq. (6.12), one has:

$$\mathcal{I}_{\boldsymbol{a}}^{(\beta)}(\theta) \xrightarrow[\beta\to 0^+]{} \sum_{k=0} \frac{m_k(\boldsymbol{a})}{k!} \theta^k = \mathrm{E}_{Y\sim\mu_{\boldsymbol{a}}} \left[e^{\theta Y} \right] \ , \tag{6.19}$$

where $\mu_{\boldsymbol{a}}(x) = \sum_{i=1}^N \delta(x - a_i)$ and $m_k(\boldsymbol{a})$ is the k^{th} moment of $\mu_{\boldsymbol{a}}$. Thus, in this setting, the additive spherical integral is nothing else than the moment generating function of the distribution of the a_i's and in the large N limit this is the MGF of μ_A. In other words, for $\beta = 0$, the logarithm of the additive spherical is the linearizing transform of the *classical convolution*.

This sharp change of behavior at $\beta = 0$ suggests scaling β slowly with N in order to smooth this transition to constructing an interpolation between the two convolutions and a natural way to do so is to look at the high-temperature regime $N\beta/2 \to c$, as discussed in the next section.

6.3 Construction of the (additive) high-temperature convolution

The high-temperature convolution is constructed in such a way that the logarithm of the additive spherical taken in the high-temperature regime is its associated linearizing (that is, it plays the role of the R-transform for this convolution). We first describe the high-temperature regime of this spherical integral before jumping to the properties of the associated convolution itself.

6.3.1 Spherical integral in the high-temperature regime

We are interested in the double scaling limit where $N \to \infty$ and $N\beta/2 \to c \geq 0$, and the distribution of the a_i's converges to a smooth distribution with a compact support: $\mu_{\boldsymbol{a}} \to \mu_A$. We next introduce the 'high-temperature' counterpart of the quantities of the previous section. In order to differentiate between quantities defined for $\beta > 0$ and $N\beta/2 \to c \geq 0$ we denote (as usual) the former by $f^{(\beta)}$ and the latter by $f^{[c]}$ such that there should be no confusion between the two. Precisely, we define:

$$\mathrm{g}_k^{[c]}[\mu_A] := \lim_{N\beta/2 \to c} \mathrm{g}_k^{(\beta)}(\boldsymbol{a}) = \sum_{1j_1+\cdots+kj_k=k} c^{j_1+\cdots+j_k} \prod_{i=1}^{k} \frac{m_i[\mu_A]^{j_i}}{i^{j_i} j_i!}\,, \tag{6.20}$$

where $m_i[\mu_A]$ is the i^{th} moment of μ_A and similarly we recall that the U-function introduced in Chapter 1 (see. Eq. (1.167)) is the high-temperature limit of $U_{\boldsymbol{a}}$ of Eq. (6.9):

$$U_{\mu_A}^{[c]}(z) := \lim_{N\beta/2 \to c} U_{\boldsymbol{a}}^{(\beta)}(z) = \exp\left\{ -c \int_{\mathrm{Supp}[\mu_A]} \log(z-x)\mu_A(x)\mathrm{d}x \right\}, \tag{6.21}$$

which is defined for $z \in \mathbb{C} \setminus (-\infty, \mathrm{a}_+)$, where as usual, a_+ is the upper edge of μ_A.

The spherical integral in the high-temperature regime is then defined thanks to the following representations:

$$\mathcal{I}_{\mu_A}^{[c]}(\theta) := \sum_{k=0}^{\infty} \frac{\Gamma(c)}{\Gamma(c+k)} \mathrm{g}_k^{[c]}[\mu_A]\, \theta^k\,, \tag{6.22}$$

$$\mathcal{I}_{\mu_A}^{[c]}(\theta) = \left(\frac{\Gamma(c)}{\theta^{c-1}} \right) \mathcal{L}_z^{-1}\left[U_{\mu_A}^{[c]}(z) \right](\theta)\,, \tag{6.23}$$

$$U_{\mu_A}^{[c]}(z) = \mathcal{I}_{\mu_A}^{[c]}(-\mathrm{D})\, z^{-c}\,, \tag{6.24}$$

$$\mathcal{I}_{\mu_A}^{[c]}(\theta) = \mathbb{E}_{X \sim \nu_A}\left[\mathrm{e}^{\theta X} \right]. \tag{6.25}$$

The distribution ν_A in Eq. (6.25) is the **Markov-Krein Transform of index** c (MKT in short) of the distribution μ_A. It is determined by taking the high-temperature limit of Eq. (6.17), that is ν_A is the unique probability measure satisfying the **Markov-Krein (MK in short) relation** :

$$\boxed{\int_{\mathrm{Supp}\,\nu_A} \frac{\nu_A(\mathrm{d}x)}{(z-x)^c} = \exp\left[-c \int_{\mathrm{Supp}\,\mu_A} \mu_A(\mathrm{d}y) \log(z-y) \right]}. \tag{6.26}$$

The LHS of the MK relation is a generalized Stieltjes and for $c = 1$ this is the standard Stieltjes transform. In Sec. 6.3.3 we will give an expression to compute ν_A from μ_A (that is, to compute the MKT of μ_A). Similarly, in Sec. 6.3.4, we will give an expression to compute μ_A from ν_A, (that is, to compute *inverse Markov-Krein transform* (IMKT) of ν_A).

From the MK relation of Eq. (6.26), one can see that shifting or rescaling a distribution μ introduces the same shift/scaling for the MKT ν. Importantly, one can expand the MK relation for $|z| \to \infty$ to relate the moments of the MKT with the ones of μ. The full combinatorial description of this relation will be given later but let's point out here that the mean of the MKT ν is the same as the one of μ **but** its variance is re-scaled by a factor $(c+1)^{-1}$:

$$m_2[\nu] - m_1[\nu]^2 = \frac{m_2[\mu] - m_1[\mu]^2}{c+1}\,. \tag{6.27}$$

By definition of the high-temperature regime, the parameter c is restricted to be positive. If we extend analytically its definition to negative integers $c = -N$ and change the quantities accordingly, one retrieves the definitions of the functions appearing in the definition of the FFC, see Sec. 2.7. In other words, the function $\mathcal{I}^{[c]}_{\mu_A}(\theta)$ will be for the high-temperature convolution what the additive form $\widehat{Q}_{\boldsymbol{a}}$ is for the FFC.

6.3.2 High-temperature convolution as the classical convolution of Markov-Krein transforms

We now have all the tools to construct the high-temperature convolution of two distributions.

For μ_A and μ_B two distributions with compact support, we *define* their **high-temperature convolution** (HTC in short) denoted by $\mu_A \oplus_c \mu_B$ such that the logarithm of the additive spherical taken in the high-temperature regime is the corresponding linearizing transform:

$$\boxed{\mu_A \oplus_c \mu_B \Leftrightarrow \log \mathcal{I}^{[c]}_{\mu_A \oplus_c \mu_B}(\theta) = \log \mathcal{I}^{[c]}_{\mu_A}(\theta) + \log \mathcal{I}^{[c]}_{\mu_B}(\theta)}\,. \tag{6.28}$$

Positivity conjecture -

A priori, the quantity $\mu_A \oplus_c \mu_B$ obtained by this construction might not be a probability distribution. However, let's recall from Chapter 3, Sec. 3.2.2, that the quantity $\mathcal{P}^{(\beta)}(\boldsymbol{c}|\boldsymbol{a},\boldsymbol{b})$ satisfying:

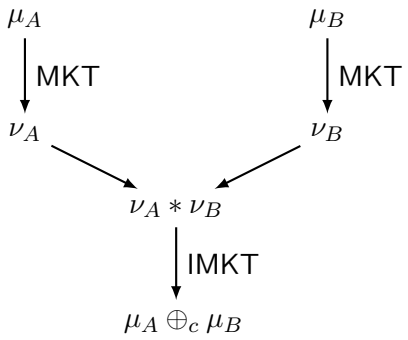
$$\int \mathcal{I}^{(\beta)}_{\boldsymbol{c}}(\theta)\mathcal{P}^{(\beta)}(\boldsymbol{c}|\boldsymbol{a},\boldsymbol{b})\mathrm{d}\boldsymbol{c} = \mathcal{I}^{(\beta)}_{\boldsymbol{a}}(\theta)\mathcal{I}^{(\beta)}_{\boldsymbol{b}}(\theta)\,, \tag{6.29}$$

sums to one and is conjectured to be positive for any $\beta > 0$ such that one can give a meaning of the 'β-sum' for arbitrary β. Since this result is conjectured to be true for **any** β and any N, this is also the case if we *choose* β to depend on N: $\beta_N = 2c/N$. If we now let N go to infinity, the empirical distribution of the c_i's converges to the LSD which is nothing else than the high-temperature convolution $\mu_A \oplus_c \mu_B$ by construction. In other words, the distribution $\mu_A \oplus_c \mu_B$ is conjectured to be a well-defined probability distribution, as a consequence of

the positivity conjecture for $\mathcal{P}^{(\beta)}(\boldsymbol{c}|\boldsymbol{a},\boldsymbol{b})$ (itself obtained as a consequence of the positivity conjecture for Jack-Littlewood-Richardson coefficients, itself obtained as a consequence of a similar conjecture for Macdonald polynomials !). **In the following, we will assume $\mu_A \oplus_c \mu_B$ to be a well-defined probability distribution**. I believe that one should be able to prove this weaker positivity conjecture thanks to the formulae developed in the following of the thesis (without appealing to tools from representation theory) but for now, this remains an open problem.

HTC as the classical convolution of MKT -

By Eq. (6.25) the high-temperature spherical integral of a distribution μ is the moment generating function of its MKT and by the definition of Eq. (6.28), the HTC corresponds to a classical convolution in the 'Markov-Krein space'. This statement can be summarized by the following scheme:



and the details to compute the MKT and IMKT will be given in the next two sections.

Operation on U-functions -

For the finite free convolution, we have seen the convolution is naturally seen as an operation acting on polynomials rather than acting on the measure associated with the real roots of these polynomials. Similarly, for the HTC, one can define it as a compact operation acting on U-functions (rather than on the associated measures). By Eq. (6.23) and the definition of the HTC, we have that the U-function associated with $\mu_A \oplus_c \mu_B$ is given in terms of the ones of μ_A and μ_B by:

$$U^{[c]}_{\mu_A \oplus_c \mu_B}(z) = \Gamma(c)\, \mathcal{L}_\theta \left[\theta^{1-c}\, \mathcal{L}^{-1}_{z_1} \left[U^{[c]}_{\mu_A}(z_1) \right](\theta)\ \mathcal{L}^{-1}_{z_2} \left[U^{[c]}_{\mu_B}(z_2) \right](\theta) \right](z)\,. \tag{6.30}$$

Note that we can also write the U-function of the HTC in terms of the differential operator representation of Eq. (6.24) but this expression is not very useful in practice.

6.3.3 Computing the Markov-Krein Transform (MKT)

In this section, we describe how one can compute the MKT ν of a measure μ thanks to the relation of Eq. (6.26). We decompose this computation into two steps:

1. First we express the imaginary part of the generalized Stieltjes near the branch cut:

$$h^{(c)}(x) := \frac{1}{\pi}\,\mathfrak{Im}\left[\int_{\mathrm{Supp}\,\nu_A}\frac{\nu(\mathrm{d}x')}{(x-x'-\mathrm{i}0^+)^c}\right], \tag{6.31}$$

   in terms of the measure μ

2. Then we establish a generalized Sokochi-Plemelj formula, that is we explain how to compute the measure ν from $h^{(c)}$.

Expression for $h^{(c)}$ -

Using the Markov-Krein relation of Eq. (6.26), we have:

$$h^{(c)}(x) = \frac{1}{\pi}\mathfrak{Im}\left[\exp\left[-c\int_{\mathrm{Supp}\,\mu_A}\log\left(x-y-\mathrm{i}0^+\right)\mu(\mathrm{d}y)\right]\right] \tag{6.32}$$

Next using the behavior of the logarithm function near the branch cut:

$$\log(\xi-\mathrm{i}0^+) = \begin{cases}\log|\xi|-\mathrm{i}\pi & \text{for } \xi<0\,,\\ \log|\xi| & \text{for } \xi>0\,,\end{cases} \tag{6.33}$$

with $\xi = x-y$, gives for $h^{(c)}$:

$$h^{(c)}(x) = \frac{1}{\pi}\,\exp\left[-c\int_{\mathrm{Supp}\,\mu}\log|x-y|\mu(\mathrm{d}y)\right]\cdot\mathfrak{Im}\left[\exp\left[-\mathrm{i}\pi c\int_x^{\mathrm{a}_+}\mu(\mathrm{d}y)\right]\right] \tag{6.34}$$

If for simplicity we denote by

$$\mu\left([x,\mathrm{a}_+]\right) := \int_x^{\mathrm{a}_+}\mu(\mathrm{d}y)\,, \tag{6.35}$$

then since the imaginary part of the exponential is the sinus function, we have the following expression

$$\boxed{h^{(c)}(x) = \frac{1}{\pi}\,\exp\left[-c\int_{\mathrm{Supp}\,\mu}\log|x-y|\mu(\mathrm{d}y)\right]\cdot\sin\left(\pi c\,\mu\left([x,\mathrm{a}_+]\right)\right)}\,. \tag{6.36}$$

Expression for the MKT -

Now to establish the generalized Sokochi-Plemelj formula, let's use once again the behavior of the power function near the branch cut:

$$(\xi-\mathrm{i}0^+)^{-c} = \begin{cases}\cos(-\pi c)|\xi|^{-c}+\mathrm{i}\sin(-\pi c)|\xi|^{-c} & \text{for } \xi<0\,,\\ \xi^{-c} & \text{for } \xi>0\,,\end{cases} \tag{6.37}$$

injecting this behavior with $\xi = x-x'$ in the definition of Eq. (6.31) gives

$$h^{(c)}(x) = -\frac{\sin(\pi c)}{\pi}\int_x^{\infty}\frac{\nu(\mathrm{d}y)}{(y-x)^c}\,. \tag{6.38}$$

Eq. (6.38) is a generalized Abel transform/fractional derivative or integral for which the inverse transforms are known and depend on the parameter c, and this is summarized in the following result

Result 6.1 (*MKT of a measure*)

For μ a distribution, its Markov-Krein transform ν given by Eq. (6.26), can be expressed as an integral operator acting on μ_A:

- *for $0 < c < 1$:*

$$\nu(x) := -\int_x^\infty (y-x)^{c-1} \frac{\mathrm{d}}{\mathrm{d}y} h^{(c)}(y)\, \mathrm{d}y\,, \tag{6.39}$$

- *for $c = 1$,* $\quad \nu(x) = h^{(c=1)}(x)$,
- *for $c > 1$:*

$$\nu(x) = (c-1)\int_x^\infty (y-x)^{c-2} h^{(c)}(y)\, \mathrm{d}y\,. \tag{6.40}$$

where the function $h^{(c)}$ is given in terms of the distribution ν by Eq. (6.38).

We conclude this section with several examples.

Example *(MKT of the Bernoulli distribution).* Let us denote by

$$\mu_{B(p)}(x) := (1-p)\delta(x-0) + p\delta(x-1)\,, \tag{6.41}$$

the *Bernoulli distribution* with probability of success p, then one can show [39] [91] that its MKT follows the law of a *beta distribution* $\mathrm{Beta}\,(cp, c(1-p))$ so that we have:

$$\nu(x) := \mathrm{MKT}[\mu_{B(p)}] = \frac{\Gamma(c)}{\Gamma(cp)\,\Gamma(c(1-p))} x^{cp-1}\,(1-x)^{c(1-p)-1}\,\mathbb{I}_{[0,1]}\,, \tag{6.42}$$

where $\mathbb{I}$ is the indicator function. Let's mention the result can also be derived by first looking at the finite N case for which one has a simple expression for the Jack polynomials and then taking the high-temperature limit, see for example the original paper from which this chapter is based on. The additive spherical writes:

$$\mathcal{I}^{[c]}_{\mu_{B(p)}}(t) = {}_1F_1\,(cp, c, t)\,, \tag{6.43}$$

where ${}_1F_1(a,b,z) := \sum_{k=0}^\infty \frac{(a)_k}{(b)_k k!} z^k$ is Kummer confluent hypergeometric function which is the moment generating function of (6.42). ∎

Example *(MKT of the arcsine distribution).* Another known example in closed form (see for example [91] and reference therein) is given when the original distribution is the *arcsine distribution*:

$$\mu_{\mathrm{As}}(x) := \frac{1}{\pi\sqrt{x(1-x)}}\mathbb{I}_{[0,1]}\,, \tag{6.44}$$

then one may show that its MKT follows the law of a beta distribution $\mathrm{Beta}\left(c+\frac{1}{2}, c+\frac{1}{2}\right)$:

$$\nu(x) = \frac{\Gamma(2c+1)}{\Gamma\left(c+\frac{1}{2}\right)^2} \left(x\,(1-x)\right)^{c-\frac{1}{2}} \mathbb{I}_{[0,1]}\,, \tag{6.45}$$

this can be checked by computing the LHS and RHS of the MK relation of Eq. (6.26) with the corresponding measures. ∎

Example *(MKT of the uniform distribution).* If we now take the original distribution to be the uniform distribution on $[0,1]$:

$$\mu_U := \mathbb{I}_{[0,1]}\,, \tag{6.46}$$

then using the expression for the function $h^{(c)}$, we have:

$$h^{(c)}(x) = \frac{e^c}{\pi} (1-x)^{-c(1-x)}\, x^{-cx} \sin\left(\pi c(1-x)\right) \mathbb{I}_{[0,1]}\,, \tag{6.47}$$

which gives in particular the density for $c=1$ of the corresponding MKT transform. For $c<1$ and $c>1$, one needs to use the formula (6.39) and (6.40) but no analytical expression is known. ∎

Example *(MKT of the Cauchy distribution).* If we extend our setting to non-compact measures, we have for every $c>0$, the Markov-Krein transform of a *Cauchy distribution* with parameters x_0 and b:

$$\mu_{C_{x,b}}(x) := \frac{b}{\pi\left(b^2+(x-x_0)^2\right)}\,, \tag{6.48}$$

is again a Cauchy distribution with the same parameters (which can be seen by computing LHS and RHS of the MK relation of Eq. (6.26), see for example [63] [112]) ∎

6.3.4 Computing the Inverse Markov-Krein Transform (IMKT)

Now conversely, one also needs to compute a distribution μ from its MKT ν. From the Markov-Krein relation, we have that Stieltjes transform g_μ of μ is given in terms of ν by

$$g_\mu(z) = -\frac{1}{c}\frac{\mathrm{d}}{\mathrm{d}z} \log\left[\int \frac{\nu(x)\mathrm{d}x}{(z-x)^c}\right]\,, \tag{6.49}$$

if we now apply the Sokochi-Plemelj inversion $\mu(x) = \mathfrak{Im} g_\mu(x-\mathrm{i}0^+)/\pi$ together with $(\log f)' = f'/f$ we have:

$$\mu(x) = \frac{1}{\pi}\mathfrak{Im}\left[\frac{\int (x-x'-\mathrm{i}0^+)^{-c-1}\,\nu(x')\mathrm{d}x'}{\int (x-x'-\mathrm{i}0^+)^{-c}\,\nu(x')\mathrm{d}x'}\right]\,. \tag{6.50}$$

Now thanks to the property of the power function near the branch cut see Eq. (6.37), we have for $\alpha = c, c+1$,

$$\int (x-x'-\mathrm{i}0^+)^{-\alpha}\,\nu(x')\mathrm{d}x' = I_\alpha^< + \cos(\pi\alpha) I_\alpha^> - \mathrm{i}\sin(\pi\alpha) I_\alpha^>\,, \tag{6.51}$$

where

- the first integral is defined by

$$I_\alpha^< := \int_{-\infty}^{x} (x - x')^{-\alpha} \nu(x') \mathrm{d}x' \,, \tag{6.52}$$

- and the second one by:

$$I_\alpha^> := \int_{x}^{\infty} (x' - x)^{-\alpha} \nu(x') \mathrm{d}x' \,. \tag{6.53}$$

Let's write $\mu = \pi^{-1} \cdot \mathrm{Num}/\mathrm{Den}$, where Den is obtained by conjugation of the denominator in the RHS of Eq. (6.50), that is $\mathrm{Den} := \left| \int (x - x' - \mathrm{i}0^+)^{-c} \nu(x') \mathrm{d}x' \right|^2$ and then look at each term individually

- For the Denominator (Den), we have:

$$\mathrm{Den} = \left(I_c^<\right)^2 + \left(\sin(\pi c) I_c^>\right)^2 \,, \tag{6.54}$$

$$\mathrm{Den} = \left(I_c^<\right)^2 + 2\cos(\pi c) \left(I_c^<\right) \left(I_c^>\right) + \underbrace{\cos(\pi c)^2 + \sin(\pi c)^2}_{=1} \left(I_c^>\right)^2 \,, \tag{6.55}$$

$$\mathrm{Den} = \left(I_c^<\right)^2 + 2\cos(\pi c) \left(I_c^<\right) \left(I_c^>\right) + \left(I_c^>\right)^2 \,. \tag{6.56}$$

- For the numerator (Num), we have:

$$\mathrm{Num} = \sin(\pi c) \left(I_{c+1}^< + \cos(\pi(c+1)) I_{c+1}^>\right) I_c^> - \sin(\pi(c+1)) I_{c+1}^> \left(I_c^< + \cos(\pi c) I_c^>\right) \,, \tag{6.57}$$

$$\mathrm{Num} = \sin(\pi c) \left(I_{c+1}^< I_c^> + I_{c+1}^> I_c^< + \underbrace{\left(\cos(\pi c) - \cos(\pi c)\right)}_{=0} I_{c+1}^> I_c^> \right) \,, \tag{6.58}$$

$$\mathrm{Num} = \sin(\pi c) \left(\left(I_{c+1}^<\right) \left(I_c^>\right) + \left(I_{c+1}^>\right) \left(I_c^<\right) \right) \,. \tag{6.59}$$

In summary, we have the following result.

Result 6.2 (*Expression for the inverse Markov-Krein transform*)

For ν a distribution, its corresponding inverse Markov-Krein transform μ is given by:

$$\mu(x) := \frac{\sin(\pi c)}{\pi} \cdot \frac{\left(I_{c+1}^<\right) \left(I_c^>\right) + \left(I_{c+1}^>\right) \left(I_c^<\right)}{\left(I_c^<\right)^2 + 2\cos(\pi c) \left(I_c^<\right) \left(I_c^>\right) + \left(I_c^>\right)^2} \tag{6.60}$$

where for $\alpha = c, c+1$ the function $I_\alpha^>$ and $I_\alpha^<$ are defined by Eqs. (6.52), (6.53).

Let' conclude this section with important examples of IMKT.

Example *(IMKT of the standard Gaussian distribution)*. For

$$\nu_G(x) := \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \,, \tag{6.61}$$

a *standard Gaussian distribution*, the additive spherical integral in the high-temperature regime of its IMKT $\mu \equiv \mathrm{IMKT}[\nu_G]$ is given by the MGF of a Gaussian distribution that is:

$$\mathcal{I}_{\mu}^{[c]}(\theta) = \mathbb{E}_{X\sim\nu_G}\left[\mathrm{e}^{\theta X}\right] = \mathrm{e}^{\theta^2/2}\,. \tag{6.62}$$

Next, the U-function of the IMKT of the Gaussian distribution is given thanks to Eq. (6.23) by

$$U_{\mu}^{[c]} = \frac{1}{\Gamma(c)}\int_0^{\infty}\theta^{c-1}\mathrm{e}^{\theta^2/2}\mathrm{e}^{z\theta}\mathrm{d}\theta\,. \tag{6.63}$$

One may recognize here the integral representation of the parabolic cylinder function, see Eq. (1.173) and we have:

$$U_{\mu}^{[c]} \propto \mathrm{e}^{-z^2/4}D_{-c}(\mathrm{i}z)\,. \tag{6.64}$$

which is the U-function of the c-Gaussian distribution of Chapter 1, re-scaled to have variance $c+1$, see Eq. (1.174). As a consequence, we have:

$$\mu(x) = \mathrm{IMKT}\left[\nu_G\right] = \frac{1}{\sqrt{2\pi}\Gamma(c+1)}\frac{1}{\left|D_{-c}\left(\mathrm{i}x\right)\right|^2}\,, \tag{6.65}$$

where D_{-c} is the parabolic cylinder function of Eq. (1.173). ∎

Example *(IMKT of the gamma distribution)*. Similarly, one can check that the IMKT of a Gamma distribution is a re-scaled c-Laguerre distribution of Chapter 1 thanks to the integral representation of Eq. (1.175) of the Tricomi function. ∎

6.3.5 Summary on how to compute the High-temperature convolution

We summarize here one way (among many others, thanks to the different formulae derived in the previous section) to compute the high–temperature convolution. This operation can be decomposed into four steps.

1. Compute the MKT of the two distributions μ_A and μ_B and their *moment generating functions* (MGF) $M_{A,B}(s) := \mathbb{E}_{Y\sim\nu_A,\nu_B}\left[\mathrm{e}^{sY}\right]$.

2. Compute the corresponding U-function thanks to:

$$U^{[c]}(z) := \frac{1}{\Gamma(c)}\int_0^{\infty}\mathrm{d}s\,\mathrm{e}^{-zs}s^{c-1}M_A(s)M_B(s)\,, \tag{6.66}$$

   for z high enough, that is higher than the $K = \max(\mathrm{Supp}\,\mu_A) + \max(\mathrm{Supp}\,\mu_B)$ and then extend analytically this function to all $z \in \mathbb{C}\setminus(-\infty, K)$.

3. Compute the Stieltjes transform $g(z) \equiv g_{\mu_A\oplus_c\mu_B}(z) := \int \mathrm{d}(\mu_A \oplus_c \mu_B)(x)\,(z-x)^{-1}$, thanks to the formula:

$$g(z) := -\frac{1}{c}\frac{\mathrm{d}}{\mathrm{d}z}\log U^{[c]}(z) = -\frac{1}{c}\frac{(U^{[c]})'(z)}{U^{[c]}(z)}\,. \tag{6.67}$$

4. Compute the distribution $\mu_A \oplus_c \mu_B$ thanks to the Sokochi-Plemelj formula:

$$(\mu_A \oplus_c \mu_B)(x) = \frac{1}{\pi}\mathfrak{Im}\, g(x - \mathrm{i}0^+)\,, \tag{6.68}$$

   or using the inversion formula of Eq. (6.60).

Each step can be approximated numerically such that one can really think of the entire process as an algorithm to compute the high-temperature convolution of two distributions, see for examples Fig. 6.1 for numerical examples.

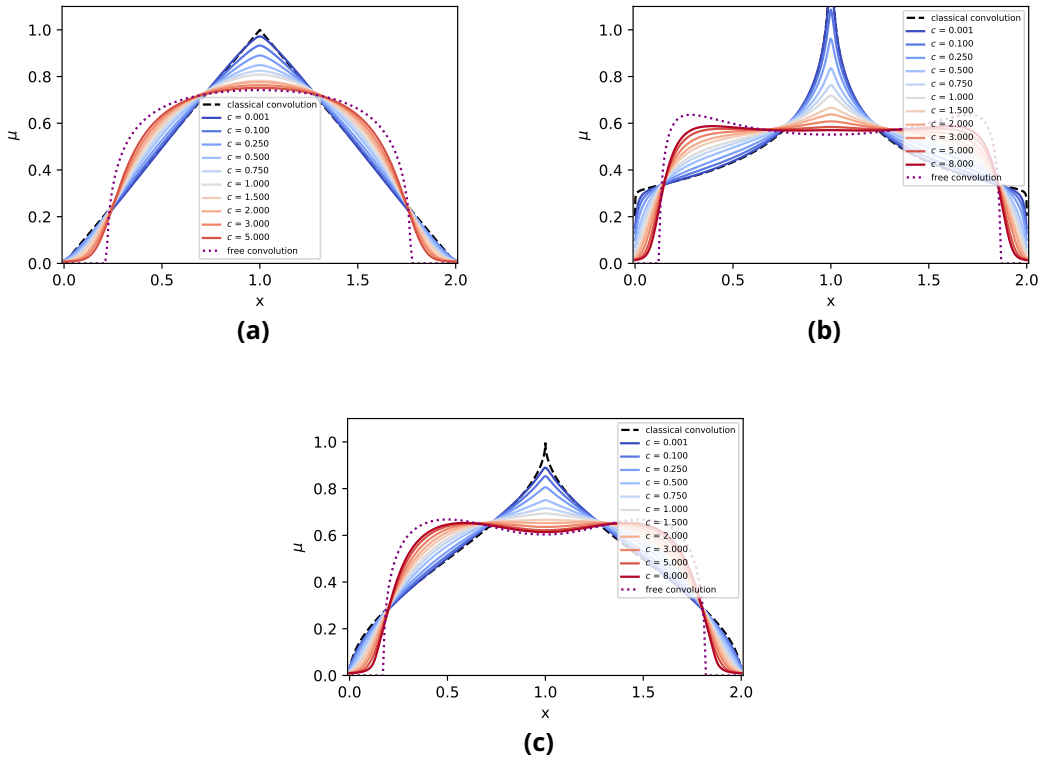


Figure 6.1: Illustration of the numerical approximations of the high-temperature convolution for different values of c of $(\mathbf{a})$ the uniform distribution with itself $(\mathbf{b})$ the arcsine law with itself and $(\mathbf{c})$ the uniform distribution with the arcsine law. The classical convolution is given in a dashed line and the free convolution is given in a dotted line.

6.4 An explicit analytical example: High-temperature convolution of two symmetric Bernoulli distributions

In the previous section, we have explained how one can compute the HTC. However, for a given choice of μ_A and μ_B and the parameter c, finding an *explicit expression* non-trivial example for

the density of their high-temperature convolution is a daunting task. Note that even for the free convolution, one has an analytical expression for the density only for specific choices of the distribution μ_A and μ_B such that one should not expect to have a simple expression for the high-temperature convolution.

This section aims to answer this issue by providing a complete description of $\mu_A \oplus_c \mu_B$ for a specific choice of μ_A and μ_B and *any value of the parameter* c. We consider the case where $\mu_A = \mu_B = \mu$, with,

$$\mu = \frac{1}{2}\delta\left(x + 1/2\right) + \frac{1}{2}\delta\left(x - 1/2\right) , \tag{6.69}$$

since this is a famous case where the density of its free convolution with itself is known analytically and given by the (shifted and re-scaled) arcsine law: for $x \in [-1, 1]$, $(\mu \boxplus \mu)(x) = \frac{1}{\pi\sqrt{1-x^2}}$. For the classical convolution, we have $\mu * \mu = \frac{1}{4}\delta_{-1} + \frac{1}{2}\delta_0 + \frac{1}{4}\delta_1$, which is the re-centered binomial distribution with number of trials $n = 2$ and probability of success $p = 1/2$.

6.4.1 MGF of the MKT

For the symmetric Bernoulli distribution, we have shown in the previous Section, see Eq. (6.42), that the corresponding MKT is the density of the random variable $Y' \sim \mathrm{Beta}(c/2, c/2)$. The distribution μ is the symmetric Bernoulli distribution, shifted by $1/2$. From the Markov-Krein relation of Eq. (6.26), one sees immediately that a shift in the distribution μ introduces the same shift in the MKT. Thus, the MKT of μ is simply the law of $Y = Y' - 1/2$. From well-known properties of the Beta distribution, this means that the MGF of the MKT of μ, $M(.) \equiv M_A(.) = M_B(.)$, is given by:

$$M(s) = \mathrm{e}^{s/2}\, {}_1F_1\left(\frac{c}{2}; c; s\right) , \tag{6.70}$$

where ${}_1F_1(a; b; u) := \sum_{k=0}^{\infty} \frac{(a)_k}{(b)_k} u^k$ is the *confluent hypergeometric function*. Using identities [52] for the confluent hypergeometric, this can also be expressed in terms of the *modified Bessel function of the kind* $I_\alpha(.)$:

$$M(s) = C_1\, s^{(1-c)/2} I_{\frac{c-1}{2}}\left(\frac{s}{2}\right) , \tag{6.71}$$

where $C_1 := 2^{c-1}\Gamma\left(\frac{c+1}{2}\right)$.

6.4.2 Expression for the U-function

Injecting Eq. (6.71) into the definition of $U^{[c]}(z)$ given by Eq. (6.66), one obtains the integral representation:

$$U^{[c]}(z) = C_2 \int_0^\infty \mathrm{d}s\, \mathrm{e}^{-zs} \left(I_{\frac{c-1}{2}}\left(s/2\right)\right)^2 , \tag{6.72}$$

where $C_2 := C_1^2/\Gamma(c)$ is a constant that will not contribute to the expression of the Stieltjes transform (and hence the density). The square of the Bessel function can be expressed as an integrated Bessel function thanks to the formula [52]:

$$\left(I_{\frac{c-1}{2}}\left(s/2\right)\right)^2 = \frac{2}{\pi}\int_0^{\frac{\pi}{2}} \mathrm{d}\theta\, I_{c-1}\left(s\cos\theta\right) . \tag{6.73}$$

If we do the change of variable $s \to s\cos\theta$ in Eq. (6.72) and then $\theta \to \arccos(\cosh\theta)$ we have:

$$U^{[c]}(z) = C_3 \int_0^\infty \mathrm{d}s\, I_{c-1}(s) \left(\int_0^\infty \mathrm{d}\theta\, \mathrm{e}^{-(zs)\cosh\theta} \right), \tag{6.74}$$

with $C_3 = 2\,C_2/\pi$. The integral with respect to the variable θ is the integral representation [52] of the *Bessel function of the second kind* $K_0(.)$, such that we have:

$$U^{[c]}(z) = C_3 \int_0^\infty \mathrm{d}s\, I_{c-1}(s)\, K_0(zs)\,. \tag{6.75}$$

By identities for the integral of the product of two Bessel functions of different kinds, see [52], one can finally express $U^{[c]}$ in terms of a hypergeometric function:

$$\boxed{U^{[c]}(z) = \Gamma(c)\, z^{-c}\, {}_2F_1\left(\frac{c}{2}, \frac{c}{2}; c; \frac{1}{z^2}\right)}. \tag{6.76}$$

6.4.3 Expression for the Stieltjes transform

In order to compute g given by Eq. (6.67), we first need to compute the derivative of $U^{[c]}(z)$. Using the differentiation formula, see Ref. [52] for the hypergeometric function, $(\mathrm{d}/\mathrm{d}u)\ {}_2F_1(a,b;c;u) = (ab/c)\ {}_2F_1(a+1,b+1;c+1;u)$, we get:

$$(U^{[c]})'(z) = \Gamma(c)(-c)z^{-c-1}\left[\frac{c}{2z^2}\,{}_2F_1\left(1+\frac{c}{2}, 1+\frac{c}{2}; 1+c; \frac{1}{z^2}\right) + {}_2F_1\left(\frac{c}{2}, \frac{c}{2}; c; \frac{1}{z^2}\right)\right]. \tag{6.77}$$

Next, using identities see Ref. [52] between *contiguous* hypergeometric functions, the sum inside the brackets simplifies such that the derivative of $U^{[c]}(z)$ writes:

$$(U^{[c]})'(z) = \Gamma(c)(-c)z^{-c}\,\frac{{}_2F_1\left(\frac{c}{2}, 1+\frac{c}{2}; c; \frac{1}{z^2}\right)}{z}\,. \tag{6.78}$$

Injecting the expression of $U^{[c]}(z)$ and its derivative, given respectively by Eq. (6.76) and Eq. (6.78), in Eq. (6.67) gives the following results for the Stieltjes transform. see Eq. (6.79).

Result 6.3 (*Stieltjes transform of HTC of symmetric Bernoulli distributions*)

The Stieltjes transform g of $\mu \oplus_c \mu$ where μ is the symmetric Bernoulli distribution of Eq. (6.69) is given by

$$g(z) = \frac{1}{z}\frac{{}_2F_1\left(\frac{c}{2}, 1+\frac{c}{2}; c; \frac{1}{z^2}\right)}{{}_2F_1\left(\frac{c}{2}, \frac{c}{2}; c; \frac{1}{z^2}\right)}\,, \tag{6.79}$$

where ${}_2F_1(a,b;c;u) := \sum_{k=0}^\infty \frac{(a)_k (b)_k}{(c)_k\, k!} u^k$ is the Gauss hypergeometric function and $(a)_k := \Gamma(a+k)/\Gamma(a)$.

Using the power series of the hypergeometric functions, one gets the large z behavior of the Stieltjes transform:

$$g(z) = \frac{1}{z} + \frac{1}{2z^3} + \frac{4+3c}{8(c+1)z^5} + \frac{8+5c}{16(c+1)z^7} + o(z^{-8})\,, \tag{6.80}$$

from which we deduce that the first even moments of the symmetric distribution $\mu \oplus_c \mu$ are given by $m_2 = 1/2$, $m_4 = (4+3c)/(8c+8)$ and $m_6 = (8+5c)/(16c+16)$.

The Stieltjes transform is expressed as the inverse function times the ratio of two different hypergeometric functions. Such a general form is very reminiscent of the high-temperature Jacobi ensemble of Chapter 1 (see. Eq. (1.189)). Yet, the parameters of the hypergeometric functions here are different such that - to the best knowledge of the author - the family of distributions $\mu \oplus_c \mu$ (and $\tilde{\rho}$) is a new one in RMT.

The case $c \to 0^+$ and classical convolution -

In the limiting case $c \to 0^+$, one should retrieve the classical convolution. This is done by using an expansion for small c in the hypergeometric functions entering the expression of g. For the numerator, we get:

$$ {}_2F_1\left(\frac{c}{2}, 1+\frac{c}{2}; c; \frac{1}{z^2}\right) = 1 + \sum_{k=1}^{\infty} (\frac{1}{2} + \mathcal{O}_c(1)) \frac{1}{z^{2k}}\,, \tag{6.81} $$

that is:

$$ {}_2F_1\left(\frac{c}{2}, 1+\frac{c}{2}; c; \frac{1}{z^2}\right) = 1 + \frac{1}{(z^2-1)} + \mathcal{O}_c(1)\,. \tag{6.82} $$

Similarly, the hypergeometric function in the denominator is equal to $1 + \mathcal{O}_c(1)$. Combining these two asymptotic behaviors, we get for the Stieltjes:

$$ g(z) \xrightarrow[c\to 0^+]{} \frac{1}{z} + \frac{1}{2z(z^2-1)}\,, \tag{6.83} $$

which decomposes into simple elements as:

$$ g(z) \xrightarrow[c\to 0^+]{} \frac{1}{4(z+1)} + \frac{1}{2z} + \frac{1}{4(z-1)}\,, \tag{6.84} $$

as expected for the Stieltjes transform of the centered binomial distribution, $\mu * \mu = \frac{1}{4}\delta_{-1} + \frac{1}{2}\delta_0 + \frac{1}{4}\delta_1$.

The case $c \to \infty$ and free convolution -

The limiting case $c \to \infty$ corresponding to the free convolution requires more work, and we only sketch the main ingredients to recover the Stieltjes transform of the arcsine law. The idea is to use the integral representation [52] of the hypergeometric function:

$$ {}_2F_1\left(a, b; c; u\right) = C_4 \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tu)^{-a}\,\mathrm{d}t\,, \tag{6.85} $$

with $C_4 = \frac{\Gamma(c)}{\Gamma(c-b)\Gamma(b)}$ and $c > b > 0$. If we denote by $F_\eta(c,u) := {}_2F_1\left(c/2, \eta + c/2; c; u\right)$ with $\eta = 1$ for the hypergeometric function in the numerator of Eq. (6.79) and $\eta = 0$ for the denominator, this means that we can write $F_\eta(c,u)$ as:

$$ F_\eta(c,u) \propto \int_0^1 \mathrm{e}^{\frac{c}{2}g(t,u)} h(t)\,\mathrm{d}t\,, \tag{6.86} $$

with $g(t,z) := \log(t(1-t)) - \log(1-tu)$ and $h(t) := t^{\eta-1}(1-t)^{1-\eta}$. As $c \to \infty$, Eq. (6.86) can be approximated by Laplace method and the results write:

$$F_\eta(c,u) \underset{c\to\infty}{\sim} K_c(u)(1-u)^{-\eta/2}\,, \tag{6.87}$$

where $K_c(u)$ is a function independent of the parameter η. Thus, if we inject this asymptotic behavior in Eq. (6.79) we get:

$$g(z) \xrightarrow[c\to\infty]{} \frac{1}{z}\frac{1}{\sqrt{1-\frac{1}{z^2}}}\,, \tag{6.88}$$

which is indeed the Stieltjes transform of $\mu \boxplus \mu$, see Chapter 2, Eq. (1.189).

6.4.4 Expression for the density

The explicit expression for the density is obtained thanks to the Sokochi-Plemelj formula of Eq. (6.68) and the expression of Eq. (6.79) for the Stieltjes transform, by looking carefully at the behavior of the hypergeometric functions near their branch cuts. The situation is very similar for both the numerator and denominator and is also very similar to the derivation of the expression of the IMKT and I only sketch the main steps. The idea is to use both the integral representation of Eq. (6.85) for $a = c/2$ and $b = \eta + c/2$ and the behavior of the power function near its branch cut of Eq. (6.37). As $z \to x - \mathrm{i}0^+$ with $x \in [-1,1]$, we have

$$(1-t/z^2)^{-c/2} = (1-t/x^2 + \mathrm{i}\,\mathrm{sign}(x)0^+)^{-c/2}\,, \tag{6.89}$$

such that we need to differentiate the cases $t < x^2$ and $t > x^2$ in Eq. (6.85). Since the distribution $\mu \oplus_c \mu$ is symmetric, we also fix $x > 0$. Thus, if we introduce the two functions $J_{1,2}(x)$ corresponding respectively to the split of the integral of Eq. (6.85) for $a = c/2$ and $b = c/2$, into the segment $[0, x^2]$ and $[x^2, 1]$:

$$J_1(x) := C_4 \int_0^{x^2} \mathrm{d}t\,(t(1-t))^{-c/2}\left(1-\frac{t}{x^2}\right)^{-c/2}\,, \tag{6.90}$$

and

$$J_2(x) := C_4 \int_{x^2}^1 \mathrm{d}t\,(t(1-t))^{-c/2}\left(\frac{t}{x^2}-1\right)^{-c/2}\,, \tag{6.91}$$

then the real and imaginary parts of the hypergeometric function in the denominator of Eq. (6.79) are given by:

$$\mathfrak{Re}\,{}_2F_1\left(\frac{c}{2},\frac{c}{2};c;\frac{1}{(x-\mathrm{i}0^+)^2}\right) = J_1(x) + \cos\left(\frac{\pi c}{2}\right)J_2(x)\,, \tag{6.92}$$

and

$$\mathfrak{Im}\,{}_2F_1\left(\frac{c}{2},\frac{c}{2};c;\frac{1}{(x-\mathrm{i}0^+)^2}\right) = -\sin\left(\frac{\pi c}{2}\right)J_2(x)\,. \tag{6.93}$$

If we now perform the change of variable $t \to x^2 t$ in Eq. (6.90), we can rewrite $J_1(x)$ as:

$$J_1(x) = C_4\,x^c \int_0^1 \mathrm{d}t\,t^{c/2-1}\,(1-t)^{-c/2}\left(1-x^2t\right)^{-c/2-1}\,. \tag{6.94}$$

By Eq. (6.85) and up to a multiplicative constant we recognize the integral in Eq. (6.94) as the hypergeometric function

$$V_1(x) := {}_2F_1\left(1-\frac{c}{2},\frac{c}{2},1;x\right) . \tag{6.95}$$

The multiplicative constant can be simplified thanks to the complement formula of the Gamma function $\Gamma(1-z)\Gamma(z) = \pi/\sin(\pi z)$ for $z \in \mathbb{C}\setminus\mathbb{Z}$, and we finally obtain:

$$J_1(x) = \frac{\pi\Gamma(c)}{\Gamma\left(\frac{c}{2}\right)^2 \sin\left(\frac{c\pi}{2}\right)} x^c\, V_1(x^2) . \tag{6.96}$$

Note that the integral representation of Eq. (6.94) is actually only valid for $c \in (0,2)$ but by analytic continuation of the hypergeometric function, the result holds for any $c > 0$ such that $2c \notin \mathbb{N}$, due to the presence of the inverse of the sinus function in Eq. (6.96).

Similarly, by the change of variable $t \to x^2 + (1-x^2)t^2$ in Eq. (6.91), $J_2(x)$ can be expressed as:

$$J_2(x) = \frac{\pi\Gamma(c)}{\Gamma\left(\frac{c}{2}\right)^2 \sin\left(\frac{c\pi}{2}\right)} x^c\, V_1(1-x^2) . \tag{6.97}$$

Thanks to Eq. (6.92) and Eq. (6.93), one has the complete behavior near the branch cut for the denominator of Eq. (6.79).

We then sketch the remaining steps to get the analytical expression for density: one can then repeat the exact same previous computation for the numerator of Eq. (6.79), where instead of the function $V_1(.)$, another function $V_2(.)$ will appear (with a different multiplicative constant) when splitting the integral into the segments $[0,x^2]$ and $[x^2,1]$. All in all, one gets the density by taking the imaginary part of the entire expression, divided by π. After simplification with the trigonometric identity $\cos(c\pi/2)^2 + \sin(c\pi/2)^2 = 1$, appearing when one multiplies the denominator of Eq. (6.79) by its conjugate, one gets the desired expression of Eq. (6.98) for the density.

Result 6.4 (*Explicit density for HTC of symmetric Bernoulli distributions*)

For μ the symmetric Bernoulli distribution of Eq. (6.69), for any c such that $2c \notin \mathbb{N}$, the density $\mu \oplus_c \mu$ of its HTC with itself, is given for any $x \in [-1,1]\setminus\{-1,0,1\}$ by:

$$(\mu \oplus_c \mu)(x) = \frac{(2-c)\sin(c\pi/2)}{2\pi} \cdot \frac{|x|(V_1(1-x^2)V_2(x^2) + V_1(x^2)V_2(1-x^2))}{V_1(x^2)^2 + 2\cos\left(\frac{c\pi}{2}\right) V_1(x^2)V_1(1-x^2) + V_1(1-x^2)^2} . \tag{6.98}$$

with $V_1(x) := {}_2F_1\left(1-\frac{c}{2},\frac{c}{2},1;x\right)$ and $V_2(x) := {}_2F_1\left(2-\frac{c}{2},1+\frac{c}{2},2;x\right)$.

Furthermore, one can obtain the cases where c is a positive even integer by carefully taking the limit, such that one can understand Eq. (6.98) as being valid for any $c > 0$, after proper regularization. The distribution $\mu \oplus_c \mu$ diverges at the points $\{-1,0,1\}$ and is otherwise absolutely continuous with no singular parts in $[-1,1]$. A plot of the density of $\mu \oplus_c \mu$ is given in Fig. 6.2.

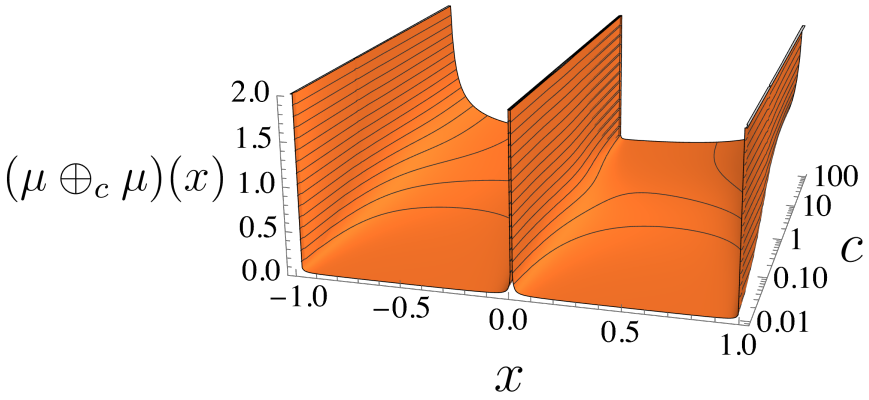


Figure 6.2: Plot of the density $\mu \oplus_c \mu$ for μ the symmetric Bernoulli distribution of Eq. (6.69), for $x \in [-1, 1]$ and different values of the parameter c, in logarithmic scale.

Simplification for special values of c

For special values of the parameter c, this expression of the HTC greatly simplifies. For example for $c = 1$, corresponding in a sense to the mid-point between the classical and the free convolution, we have:

$$(\mu \oplus_{c=1} \mu)(x) = \frac{1}{2|x|(1-x^2)} \frac{1}{K(x^2)^2 + K(1-x^2)^2}\,, \tag{6.99}$$

where $K(.)$ is the complete elliptic integral of the first kind, $K(u) := \int_0^{\pi/2} \mathrm{d}\theta\, (1-u^2 \sin^2\theta)^{-1/2}$.

The expression for $c = 2$ is even simpler since we have:

$$(\mu \oplus_{c=2} \mu)(x) = \frac{1}{|x|(1-x^2)} \frac{1}{\left(\log\left(\frac{1-x^2}{x^2}\right)\right)^2 + \pi^2}\,. \tag{6.100}$$

Note that in practice when c is an even positive integer it is easier to evaluate the Stieltjes transform thanks to Eq. (6.79) and then use the Sokochi-Plemelj formula of Eq. (6.68) rather than taking the limit in the generic expression of Eq. (6.98).

6.5 HTC cumulants and the high-low temperature duality

The goal of this section is to construct the cumulants $\kappa_k^{[c]}$ associated to the HTC and relates them to moments $m_k[\mu]$ of the distribution in question. Just like for the FFC of Sec. 2.7.2, we will relate the cumulants and moments via a third quantity: the moments of the MKT.

To do so, let's first compare the power sum representation of Eq. (6.22) with the moment generating function representation of Eq. (6.25). If we denote by $n_k \equiv m_k[\nu]$ the k^{th} moment

of the MKT, we have:

$$n_k = \frac{\Gamma(c)k!}{\Gamma(c+k)} g_k[\mu]\,, \tag{6.101}$$

from which thanks to Eq. (6.2) one can deduce a first combinatorial relation between the n_k's and the m_k's. The exact combinatorial formula will be given shortly after but one can check that the leading in m_k is given by:

$$n_k = \frac{\Gamma(c+1)(k-1)!}{\Gamma(c+k)} m_k + \text{lin. comb. of product of lower moments}\,. \tag{6.102}$$

Next, the cumulants for the HTC are *up to a factor* the coefficients of the power series of the logarithm of the spherical integral. By definition of the cumulants, the leading term of $\kappa_k^{[c]}$ must be $1 \cdot m_k$ and thanks to the leading term of n_k, one can deduce the factor in the power series of the cumulants in order to have such behavior and the final result reads as follows.

Result 6.5 (*Moment-cumulant relation for the HTC*)

The moments $m_k \equiv m_k[\mu]$ of a distribution μ are related to the cumulants $\kappa_k^{[c]}$ of the HTC by the following coupled equation:

$$\begin{cases} n_k = \left(\frac{\Gamma(c)k!}{\Gamma(c+k)}\right) \sum_{1j_1+\cdots+kj_k=k} (c)^{j_1+\cdots+j_k} \prod_{i=1}^{k} \frac{{m_k}^{j_i}}{i^{j_i} j_i!}\,, \\ \\ \log\left(1 + \sum_{k=1}^{N} \frac{n_k}{k!} y^k\right) = \sum_{k=1}^{N} \left(\frac{\Gamma(c+1)}{\Gamma(c+k)k}\right) \kappa_k^{[c]} y^k\,, \end{cases} \tag{6.103}$$

where n_k is the k^{th} moment of the MKT of μ.

Let's mention that the n_k are 'silent' variables since one can technically inject the top line of Eq. (6.103) in the bottom line of Eq. (6.103) to have a (sophisticated) relation involving only the moments and the HTC cumulants.

The first four HTC cumulants are given by:

$$\kappa_1^{[c]} = m_1\,, \tag{6.104}$$

$$\kappa_2^{[c]} = m_2 - m_1^2\,, \tag{6.105}$$

$$\kappa_3^{[c]} = m_3 - 3m_2m_1 + 2m_1^3\,, \tag{6.106}$$

$$\kappa_4^{[c]} = m_4 - 4m_3m_1 - \left(2 + \frac{1}{c+1}\right) m_2^2 + \left(10 + \frac{2}{c+1}\right) m_2m_1^2 - \left(5 + \frac{1}{c+1}\right) m_1^4\,, \tag{6.107}$$

In particular when the first moment $m_1 = 0$, we have that the 4^{th} cumulant is given by:

$$\kappa_4^{[c]} = m_4 - m_2^2 \left(\frac{2c+3}{c+1}\right)\,, \tag{6.108}$$

from which one can see that the value $c = 1$ corresponds in a sense to the midpoint between the classical case (for which the coefficient in front of m_2^2 is equal to 3) and the free case (for which the coefficient in front of m_2^2 is equal to 2).

Similarly we can obtain the moments in terms of the c-cumulants, we only give here the first four moment-c-cumulant relations:

$$m_1 = \kappa_1^{[c]}\,, \tag{6.109}$$

$$m_2 = \kappa_2^{[c]} + \left(\kappa_1^{[c]}\right)^2\,, \tag{6.110}$$

$$m_3 = \kappa_3^{[c]} + 3\kappa_2^{[c]}\kappa_1^{[c]} + \left(\kappa_2^{[c]}\right)^3\,, \tag{6.111}$$

$$m_4 = \kappa_4^{[c]} + 4\kappa_3^{[c]}\kappa_1^{[c]} + \left(2+\frac{1}{c+1}\right)\left(\kappa_2^{[c]}\right)^2 + 6\kappa_2^{[c]}\left(\kappa_1^{[c]}\right)^2 + \left(\kappa_1^{[c]}\right)^4\,. \tag{6.112}$$

High-low temperature duality for cumulants -

The relation between moments and HTC cumulants is very reminiscent of the one between moments and FFC cumulants, see Eq. (2.233). In fact, one can analytically extend this relation to $c = -N$ thanks to the limit relation for the Gamma function:

$$\lim_{\epsilon\to 0^+} \frac{\Gamma(k-N+\epsilon)}{\Gamma(-N+\epsilon)} = (-1)^k \frac{N!}{(N-k)!}\,. \tag{6.113}$$

Next, if one does the change of variable $y \to -y$, one retrieves the relation between moments and finite free cumulants with $\tau_k \equiv (-1)^k n_k$ and since the τ_k are silent variables we have the following result between FFC and HTC.

Result 6.6 (*High-low temperature duality for cumulants*)

If we denote by $\kappa^{(N)}_{\mathrm{FFC},k}$ the k^{th} cumulant associated to the finite free convolution, and by $\kappa^{[c]}_{\mathrm{HTC},k}$ the k^{th} cumulant associated to the high-temperature convolution, then the two families of cumulants seen as an expansion in terms of moments, satisfy the duality relation

$$\kappa^{(N)}_{\mathrm{FFC},k} = \kappa^{[-N]}_{\mathrm{HTC},k} \tag{6.114}$$

where the RHS has to be understood as an analytical continuation to negative integer values of the parameter c.

In other words, the high-low temperature duality, which was present for *individual* ensembles in Chapter 1, naturally extends to the associated convolution.

Weighted partitions and relation between classical and free convolution -

As for the finite free convolution, after a consequent work, one can eliminate the dependency in the n_k's to express the k^{th} moment as a sum over partitions of HTC cumulants:

$$m_k = \sum_{\boldsymbol{\pi}\in\mathcal{P}[k]} W^{[c]}_{\mathrm{HTC}}(\boldsymbol{\pi})\,\kappa^{[c]}_{\boldsymbol{\pi}}\,, \tag{6.115}$$

where $W_{\mathrm{HTC}}^{[c]}$ is again a weighting factor that can be shown to penalize crossing partitions in a certain way. Asymptotically in c, this weight factor satisfies:

$$W_{\mathrm{HTC}}^{[c]}(\boldsymbol{\pi}) \xrightarrow[c\to 0^+]{} 1 \quad \text{and} \quad W_{\mathrm{HTC}}^{[c]}(\boldsymbol{\pi}) \xrightarrow[c\to\infty]{} \begin{cases} 1 & \text{if } \boldsymbol{\pi} \text{ is non-crossing}\,, \\ \\ 0 & \text{otherwise}\ , \end{cases} \tag{6.116}$$

such that one retrieves the moment-cumulant formula of the classical convolution for $c \to 0^+$ and the moment-cumulant formula of the free convolution in the limiting case $c \to \infty$, see Eq. (2.151).

6.6 High-temperature limit theorems

In this section, we derive the limit theorems for the high-temperature convolution.

6.6.1 High-temperature CLT

Let's consider a measure μ with zero mean and variance one. Its MKT ν is a measure with zero mean and variance $1/(c+1)$. If we now re-scale μ to have variance $1/n$ (that is $\sqrt{n}\mu(\sqrt{n}\cdot)$), the corresponding MKT $\sqrt{n}\nu(\sqrt{n}\cdot)$ has now variance $1/(n\cdot(c+1))$. Now let's perform the HTC of this measure $\sqrt{n}\mu(\sqrt{n}\cdot)$ with itself n times. Since the HTC corresponds to do classical convolution of the MKT, by the (classical) CLT, the MKT of the limiting distribution is a Gaussian distribution with mean zero and variance $1/(c+1)$. Since the IMKT of a *standard* Gaussian is given by Eq. (6.65), we get by scaling the following result.

Result 6.7 (*CLT for the HTC*)

For μ a distribution with zero mean and variance one, if we re-scale it such that its variance is $1/n$ and performs the high-temperature convolution with itself n times, we have asymptotically in n:

$$\underbrace{\sqrt{n}\mu\left(\sqrt{n}x\right) \oplus_c \cdots \oplus_c \sqrt{n}\mu\left(\sqrt{n}x\right)}_{n \text{ times}} \xrightarrow[n\to\infty]{} \mu_G^{[c]}(x) := \frac{\sqrt{c+1}}{\sqrt{2\pi}\Gamma(c+1)} \frac{1}{\left|D_{-c}\left(\mathrm{i}\sqrt{c+1}x\right)\right|^2} \tag{6.117}$$

where D_{-c} is the Dawson function of Eq. (1.173).

The distribution $\mu_G^{[c]}$ is the c-Gaussian distribution of Chapter 1 (see Sec. 1.7 and Fig. 1.8 for an illustration). Its additive spherical $\mathcal{I}_G^{[c]} \equiv \mathcal{I}^{[c]}_{\mu_G^{[c]}}$ is given by the MGF of a Gaussian distribution with variance $(c+1)$:

$$\mathcal{I}_G^{[c]}(u) = \mathrm{e}^{\frac{u^2}{2(c+1)}}\,, \tag{6.118}$$

from which one can easily check that the HTC cumulants are given by:

$$\kappa_k^{[c]}\left[\mu_G^{[c]}\right] = \begin{cases} 1 & \text{if } k=2\,, \\ \\ 0 & \text{otherwise}\ , \end{cases} \tag{6.119}$$

as one should expect for the limiting distribution of a CLT.

DBM at high-temperature -

Let's now consider the high-temperature analog of the DBM of Chapter 2 (see Sec. 2.5.1), that is, we want to describe the distribution:

$$\mu_C(x,t) = \mu_B \oplus_c \mu_G^{[c]}\left(\cdot/\sqrt{t}\right)/\sqrt{t}\,, \tag{6.120}$$

where $\mu_G^{[c]}\left(\cdot/\sqrt{t}\right)/\sqrt{t}$ is the c-Gaussian distribution re-scaled to have variance t. By multiplicative of the spherical integral we have that U-function $U_C^{[c]} \equiv U_{\mu_C}^{[c]}$ is given by:

$$U_C^{[c]} = \mathcal{I}_{G(t)}^{[c]}\left(-\mathrm{D}_z\right)\underbrace{\mathcal{I}_B^{[c]}\left(-\mathrm{D}_z\right)z^{-c}}_{=U_B^{[c]}}, \tag{6.121}$$

$$U_C^{[c]} = \mathrm{e}^{\frac{t}{2(c+1)}\mathrm{D}_z^2}\,U_B^{[c]}\,, \tag{6.122}$$

Next since $\partial_t \exp\left[\frac{t}{2(c+1)}\mathrm{D}_z^2\right] = \mathrm{D}_z^2/(2\cdot(c+1))\cdot\exp\left[\frac{t}{2(c+1)}\mathrm{D}_z^2\right]$, by Differentiating Eq. (6.122) with respect to t, we have the *forward heat equation*:

$$\partial_t U(t,z) = \frac{1}{2(c+1)}\partial_{zz}U(t,z)\,. \tag{6.123}$$

Now perform the change of variable from the U-function to the Stieltjes transform $g(z,t) \equiv g_C(z,t)$, given by Eq. (6.67), we have:

$$\boxed{\partial_t g + \frac{c}{c+1}g\partial_z g = \frac{1}{2}\frac{1}{c+1}\partial_{zz}g}\,, \tag{6.124}$$

under the initial condition $g(0,z) = \int (z-x)^{-1}\mathrm{d}\mu_B$.

As $c \to \infty$ we retrieve the inviscid Burgers equation of Eq. (2.86) while as $c \to 0$ one has the standard heat equation for the Stieltjes. If the parameter c extends analytically to negative values $c = -N$, we get the DBM of $\beta \to \infty$ with N fixed of Eq. (2.252).

6.6.2 High-temperature Poisson limit theorem

We now turn to the Poisson central limit theorem which concerns the limit of 'c-sum' of independent Bernoulli random variables with a probability of success that goes to zero at a speed $1/n$:

$$\mu_{a\cdot\mathrm{Ber}(\lambda/n)} := \left(1-\frac{\lambda}{n}\right)\delta(x-0) + \frac{\lambda}{n}\delta(x-a)\,, \tag{6.125}$$

that is, we want to characterize the limit $n \to \infty$ of the HTC of $\mu_{a\cdot\mathrm{Ber}(\lambda/n)}$ with itself, done n times, which we denote by $\mu_{c-\mathrm{Poi}}$.

We know, from Eq. (6.42), that the Markov-Krein transform of the Bernoulli distribution with a probability of success p is the beta distribution $\mathrm{Beta}(cp, c(1-p))$. Since again c-convolution

corresponds to classical convolution in the MK space, we first need to determine the limiting distribution of:

$$\nu(.) := \mathrm{MKT}[\mu_{c-\mathrm{Poi}}] \quad := \lim_{n\to\infty} \left[\underbrace{\frac{1}{a}\mu_{\mathrm{Beta}\left(\frac{c\lambda}{n},\frac{c(n-\lambda)}{n}\right)}\left(\frac{\cdot}{a}\right) \oplus_c \cdots \oplus_c \mu_{\mathrm{Beta}\left(\frac{c\lambda}{n},\frac{c(n-\lambda)}{n}\right)}\left(\frac{\cdot}{a}\right)}_{n \text{ times}} \right] , \tag{6.126}$$

and then take the IMKT. This kind of distribution does not seem to have appeared before in the literature, and we will characterize it with its moment generating function (as no closed form is known). The moment generating function of the beta distribution is given by Eq. (6.43), so that we have:

$$\mathbb{E}_{X\sim\nu}\left[e^{tX}\right] = \lim_{n\to\infty} {}_1F_1\left(\frac{c\lambda}{n}, c; at\right)^n , \tag{6.127}$$

$$\mathbb{E}_{X\sim\nu}\left[e^{tX}\right] = \lim_{n\to\infty} \left(1 + \frac{\sum_{k=1}^{\infty} \frac{\lambda\Gamma(c+1)}{\Gamma(c+k)\,k}(at)^k}{n} + \mathcal{O}\left(\frac{1}{n^2}\right)\right)^n , \tag{6.128}$$

Next, we use:

$${}_2F_2\left(\{1,1\},\{2,c+1\};t\right) = \sum_{k=0}^{\infty} \frac{\Gamma(c+1)}{\Gamma(c+k+1)\,(k+1)} t^k = \frac{1}{t}\sum_{k=1}^{\infty} \frac{\Gamma(c+1)}{\Gamma(c+k)\,k} t^k , \tag{6.129}$$

where ${}_2F_2$ is the *hypergeometric function*. Together with the classical limit identity for the exponential:

$$e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n , \tag{6.130}$$

we get:

$$\mathbb{E}_{X\sim\nu}\left[e^{tX}\right] = \exp\left\{a\lambda t\, {}_2F_2\left(\{1,1\},\{2,c+1\};at\right)\right\} . \tag{6.131}$$

Since the distribution ν is supported on $\mathbb{R}_+$, we can take the inverse Laplace transform of the moment generating function evaluated at $-t$:

$$\nu(x) = \mathcal{L}_t^{-1}\left[\exp\left\{-a\lambda t\, {}_2F_2\left(\{1,1\},\{2,c+1\};-at\right)\right\}\right](x) . \tag{6.132}$$

One can compute numerically ν using this last formula. This can be summarized in the following result.

Result 6.8 (*High-temperature Poisson distribution*)

For $\mu_{a\cdot\mathrm{Ber}(\lambda/n)} := \left(1-\frac{\lambda}{n}\right)\delta(x-0) + \frac{\lambda}{n}\delta(x-a)$ a re-scaled Bernoulli distribution with small rate of success λ/n, we have:

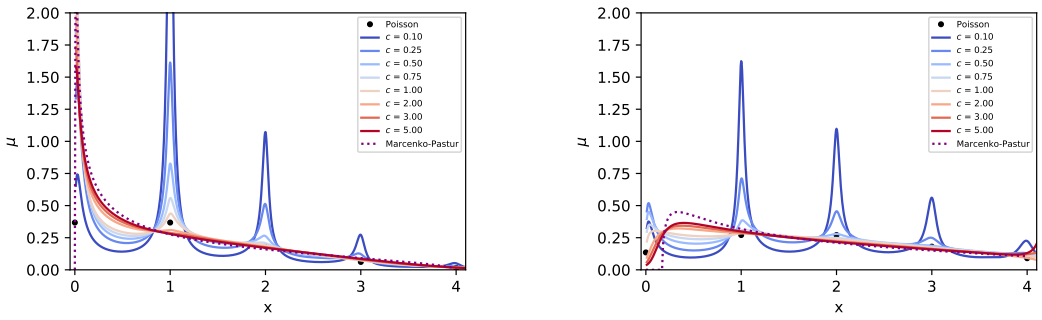
$$\underbrace{\mu_{a\cdot\mathrm{Ber}(\lambda/n)} \oplus_c \cdots \oplus_c \mu_{a\cdot\mathrm{Ber}(\lambda/n)}}_{n \text{ times}} \xrightarrow[n\to\infty]{} \mu_{c-\mathrm{Poi}} \tag{6.133}$$

where $\mu_{c-\mathrm{Poi}}$ is the distribution whose Markov-Krein transform ν is given by Eq. (6.132).

This high-temperature Poisson distribution can be approximated numerically, and we have plotted the different results in Fig 6.3. It interpolates between the (standard) Poisson distribution and the Marčenko-Pastur distribution. Thanks to Eq. (6.131), one can show that its cumulants (for the HTC) are given by:

$$\kappa_k^{(c)} = a^k \lambda \,, \tag{6.134}$$

as one should expect for the limit distribution of a Poisson theorem.



(a) $\lambda = 1$ **(b)** $\lambda = 2$

Figure 6.3: Plots of the numerical approximation of the limiting distribution of the Poisson limit theorem with parameters $a = 1$ and $\lambda = 1$ for the left figure, and $\lambda = 2$ for the right figure, for different values of c, compared to the classical (Poisson) and free (Marčenko-Pastur) limiting distributions.

6.7 Summary and Conclusion of Chapter 6

In this chapter, we have constructed the high-temperature convolution (HTC), a one-parameter interpolation between the classical ($c = 0$) and the free ($c \to \infty$) convolutions. Our construction of this HTC relies heavily on the study of the additive spherical integral in the high-temperature regime $\frac{N\beta}{2} \to c$. It turns out that in this regime the additive spherical integral is the moment generating function of the so-called Markov-Krein transform of the distribution of interest, such that the HTC of two distributions corresponds to a classical convolution of their Markov-Krein transforms. Furthermore, this HTC shares a duality with the FFC introduced in Chapter 2, a property we have already encountered for individual ensembles in Chapter 1.

The HTC is conjectured to be positive-preserving and it will be interesting to know if one can come up with proof of this property using the formula developed in this thesis.

An natural extension of this work is to construct the multiplicative counter of this HTC, thanks to the asymptotic behavior of the multiplicative spherical integral of Chapter 3. This is a problem I am currently working on.

Bibliography

[1] G. Akemann and S.-S. Byun, *The high temperature crossover for general 2d Coulomb gases*, Journal of Statistical Physics, 175 (2019), pp. 1043–1065.

[2] S. Allesina and S. Tang, *The stability–complexity relationship at age 40: a random matrix perspective*, Population Ecology, 57 (2015), pp. 63–75.

[3] R. Allez, J.-P. Bouchaud, and A. Guionnet, *Invariant beta ensembles and the Gauss-Wigner crossover*, Physical Review Letters, 109 (2012).

[4] R. Allez, J.-P. Bouchaud, S. N. Majumdar, and P. Vivo, *Invariant beta-Wishart ensembles, crossover densities and asymptotic corrections to the Marčenko–Pastur law*, Journal of Physics A: Mathematical and Theoretical, 46 (2012), p. 015001.

[5] A. I. Aptekarev, P. M. Bleher, and A. B. Kuijlaars, *Large N limit of Gaussian random matrices with external source, part II*, Communications in mathematical physics, 259 (2005), pp. 367–389.

[6] O. Arizmendi, J. Garza-Vargas, and D. Perales, *Finite free cumulants: Multiplicative convolutions, genus expansion and infinitesimal distributions*, 2021.

[7] O. Arizmendi and D. Perales, *Cumulants for finite free convolution*, Journal of Combinatorial Theory, Series A, 155 (2018), pp. 244–266.

[8] A. Auffinger, G. B. Arous, and J. Černý, *Random matrices and complexity of spin glasses*, Communications on Pure and Applied Mathematics, 66 (2013), pp. 165–201.

[9] A. Auffinger and W.-K. Chen, *Free energy and complexity of spherical bipartite models*, Journal of Statistical Physics, 157 (2014), pp. 40–59.

[10] J. Baik, G. B. Arous, and S. Péché, *Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices*, The Annals of Probability, 33 (2005).

[11] J. Baik, P. Deift, and K. Johansson, *On the distribution of the length of the longest increasing subsequence of random permutations*, Journal of the American Mathematical Society, 12 (1999), pp. 1119–1178.

[12] J. Baik, P. Deift, and T. Suidan, *Combinatorics and random matrix theory*, Graduate studies in mathematics, American Mathematical Society, Providence, RI, June 2016.

[13] J. Baik and J. O. Lee, *Fluctuations of the free energy of the spherical Sherrington–Kirkpatrick model*, Journal of Statistical Physics, 165 (2016), pp. 185–224.

[14] ——, *Free energy of bipartite spherical Sherrington–Kirkpatrick model*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 56 (2020).

[15] H. F. Baker, *Further applications of metrix notation to integration problems*, Proceedings of the London Mathematical Society, s1-34 (1901), pp. 347–360.

[16] A. Barra, P. Contucci, E. Mingione, and D. Tantari, *Multi-species mean field spin glasses. rigorous results*, Annales Henri Poincaré, 16 (2014), pp. 691–708.

[17] R. J. Baxter, *Statistical mechanics of a one-dimensional Coulomb system with a uniform charge background*, Mathematical Proceedings of the Cambridge Philosophical Society, 59 (1963), p. 779–787.

[18] C. W. J. Beenakker, *Random-matrix theory of quantum transport*, Rev. Mod. Phys., 69 (1997), pp. 731–808.

[19] G. Ben Arous, Y. V. Fyodorov, and B. A. Khoruzhenko, *Counting equilibria of large complex systems by instability index*, Proceedings of the National Academy of Sciences, 118 (2021).

[20] F. Benaych-Georges, *Rectangular random matrices, related convolution*, Probability Theory and Related Fields, 144 (2008), pp. 471–515.

[21] F. Benaych-Georges, *Rectangular R-transform as the limit of rectangular spherical integrals*, Journal of Theoretical Probability, 24 (2011), p. 969.

[22] F. Benaych-Georges and R. R. Nadakuditi, *The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices*, Advances in Mathematics, 227 (2011), pp. 494–521.

[23] ——, *The singular values and vectors of low rank perturbations of large rectangular random matrices*, Journal of Multivariate Analysis, 111 (2012), pp. 120–135.

[24] F. Benaych-Georges and S. Péché, *Poisson statistics for matrix ensembles at large temperature*, Journal of Statistical Physics, 161 (2015), pp. 633–656.

[25] G. Biroli, J.-P. Bouchaud, and M. Potters, *Extreme value problems in random matrix theory and other disordered systems*, Journal of Statistical Mechanics: Theory and Experiment, 2007 (2007), pp. P07019–P07019.

[26] G. Biroli, J.-P. Bouchaud, and M. Potters, *On the top eigenvalue of heavy-tailed random matrices*, Europhysics Letters (EPL), 78 (2007), p. 10001.

[27] G. Biroli, G. Bunin, and C. Cammarota, *Marginally stable equilibria in critical ecosystems*, New Journal of Physics, 20 (2018), p. 083051.

[28] P. Bleher and A. B. Kuijlaars, *Large N limit of Gaussian random matrices with external source, part I*, Communications in mathematical physics, 252 (2004), pp. 43–76.

[29] P. M. Bleher and A. B. Kuijlaars, *Large N limit of Gaussian random matrices with external source, part III: double scaling limit*, Communications in mathematical physics, 270 (2007), pp. 481–517.

[30] A. Borodin, *Biorthogonal ensembles*, Nuclear Physics B, 536 (1998), pp. 704–732.

[31] A. Borodin and V. Gorin, *General β-Jacobi corners process and theGaussian free field*, Communications on Pure and Applied Mathematics, 68 (2015), pp. 1774–1844.

[32] E. Brézin and S. Hikami, *Level spacing of random matrices in an external source*, Physical Review E, 58 (1998), p. 7176.

[33] E. Brézin and S. Hikami, *An extension of the HarishChandra-Itzykson-Zuber integral*, Communications in Mathematical Physics, 235 (2003), pp. 125–137.

[34] E. Brézin and S. Hikami, *Random Matrix Theory with an External Source*, Springer Singapore, 2016.

[35] M.-F. Bru, *Wishart processes*, Journal of Theoretical Probability, 4 (1991), pp. 725–751.

[36] J. Bun, J.-P. Bouchaud, S. N. Majumdar, and M. Potters, *Instanton approach to large N Harish-Chandra-Itzykson-Zuber integrals*, Physical review letters, 113 (2014), p. 070201.

[37] J. Bun, J.-P. Bouchaud, and M. Potters, *Cleaning large correlation matrices: Tools from random matrix theory*, Physics Reports, 666 (2017), pp. 1–109.

[38] J. E. Campbell, *On a Law of Combination of Operators bearing on the Theory of Continuous Transformation Groups*, Proceedings of the London Mathematical Society, s1-28 (1896), pp. 381–390.

[39] D. M. Cifarelli and E. Regazzini, *Distribution functions of means of a Dirichlet process*, The Annals of Statistics, (1990), pp. 429–442.

[40] T. Claeys and D. Wang, *Random matrices with equispaced external source*, Communications in Mathematical Physics, 328 (2014), pp. 1023–1077.

[41] B. Conrey, *Riemann's hypothesis*, 2015.

[42] R. Coquereaux, C. McSwiggen, and J.-B. Zuber, *Revisiting Horn's problem*, Journal of Statistical Mechanics: Theory and Experiment, 2019 (2019), p. 094018.

[43] R. Coquereaux and J. B. Zuber, *The Horn Problem for real symmetric and quaternionic self-dual matrices*, Symmetry, Integrability and Geometry: Methods and Applications, (2019).

[44] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*, Cambridge University Press, Cambridge, England, Sept. 2011.

[45] L. F. Cugliandolo and D. S. Dean, *Full dynamical solution for a spherical spin-glass model*, Journal of Physics A: Mathematical and General, 28 (1995), pp. 4213–4234.

[46] F. D. Cunden, P. Facchi, M. Ligabò, and P. Vivo, *Universality of the third-order phase transition in the constrained Coulomb gas*, Journal of Statistical Mechanics: Theory and Experiment, 2017 (2017), p. 053303.

[47] D. S. Dean and S. N. Majumdar, *Large deviations of extreme eigenvalues of random matrices*, Phys. Rev. Lett., 97 (2006), p. 160201.

[48] N. Demni, *Free Jacobi process*, Journal of Theoretical Probability, 21 (2008), pp. 118–143.

[49] P. Desrosiers, *Duality in random matrix ensembles for all β*, Nuclear Physics B, 817 (2009), pp. 224–251.

[50] A. Dhar, A. Kundu, S. N. Majumdar, S. Sabhapandit, and G. Schehr, *Exact extremal statistics in the classical 1d Coulomb gas*, Physical review letters, 119 (2017), p. 060601.

[51] ——, *Extreme statistics and index distribution in the classical 1d Coulomb gas*, Journal of Physics A: Mathematical and Theoretical, 51 (2018), p. 295001.

[52] *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/, Release 1.1.5 of 2022-03-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

[53] Y. Dolivet and M. Tierz, *Chern–Simons matrix models and Stieltjes–Wigert polynomials*, Journal of mathematical physics, 48 (2007), p. 023507.

[54] I. Dumitriu and A. Edelman, *Matrix models for beta ensembles*, Journal of Mathematical Physics, 43 (2002), pp. 5830–5847.

[55] T. H. Dung and T. K. Duy, *Beta Laguerre ensembles in global regime*, Osaka Journal of Mathematics, 58 (2021), pp. 435–450.

[56] F. J. Dyson, *A Brownian-motion model for the eigenvalues of a random matrix*, Journal of Mathematical Physics, 3 (1962), pp. 1191–1198.

[57] S. F. Edwards and R. C. Jones, *The eigenvalue spectrum of a large symmetric random matrix*, Journal of Physics A: Mathematical and General, 9 (1976), pp. 1595–1603.

[58] L. Erdős, *Universality of Wigner random matrices: a survey of recent results*, Russian Mathematical Surveys, 66 (2011), pp. 507–626.

[59] L. Erdős and B. Farrell, *Local eigenvalue density for general manova matrices*, Journal of Statistical Physics, 152 (2013), pp. 1003–1032.

[60] B. Eynard and N. Orantin, *Topological recursion in enumerative geometry and random matrices*, Journal of Physics A: Mathematical and Theoretical, 42 (2009), p. 293001.

[61] P. Facchi, U. Marzolino, G. Parisi, S. Pascazio, and A. Scardicchio, *Phase transitions of bipartite entanglement*, Phys. Rev. Lett., 101 (2008), p. 050502.

[62] J. Faraut, *Logarithmic potential theory, orthogonal polynomials, and random matrices*, Modern methods in multivariate statistics, Lecture Notes of CIMPA-FECYT-UNESCO-ANR. Hermann, (2014).

[63] J. Faraut and F. Fourati, *Markov-Krein transform*, Colloquium Mathematicum, 144 (2016), pp. 137–156.

[64] J. Faraut and A. Korányi, *Analysis on Symmetric Cones (Oxford Mathematical Monographs)*, Clarendon Press, 2 1995.

[65] A. Flack, S. N. Majumdar, and G. Schehr, *Truncated linear statistics in the one dimensional one-component plasma*, 2021.

[66] L. Foini and J. Kurchan, *Annealed averages in spin and matrix models*, preprint arXiv:2104.04363, (2021).

[67] P. J. Forrester, *Log-Gases and Random Matrices (LMS-34) (London Mathematical Society Monographs)*, Princeton University Press, 7 2010.

[68] P. J. Forrester, *Global and local scaling limits for the β= 2 Stieltjes–Wigert random matrix ensemble*, Random Matrices: Theory and Applications, (2021), p. 2250020.

[69] ——, *High–low temperature dualities for the classical β-ensembles*, Random Matrices: Theory and Applications, (2022), p. 2250035.

[70] P. J. Forrester and J. Grela, *Hydrodynamical spectral evolution for random matrices*, Journal of Physics A: Mathematical and Theoretical, 49 (2016), p. 085203.

[71] P. J. Forrester, S. N. Majumdar, and G. Schehr, *Non-intersecting Brownian walkers and Yang–Mills theory on the sphere*, Nuclear Physics B, 844 (2011), pp. 500–526.

[72] W. Fulton, *Eigenvalues, invariant factors, highest weights, and schubert calculus*, Bulletin of the American Mathematical Society, 37 (2000), pp. 209–249.

[73] Y. V. Fyodorov and B. A. Khoruzhenko, *Nonlinear analogue of the May–Wigner instability transition*, Proceedings of the National Academy of Sciences, 113 (2016), pp. 6827–6832.

[74] I. M. Gelfand and M. A. Naĭmark, *Unitary representations of the classical groups*, Trudy Mat. Inst. Steklov., 36 (1950), p. 288.

[75] V. Gorin and A. W. Marcus, *Crystallization of Random Matrix Orbits*, International Mathematics Research Notices, 2020 (2018), pp. 883–913.

[76] J. Grela, S. N. Majumdar, and G. Schehr, *Non-intersecting Brownian Bridges in the Flat-to-Flat Geometry*, Journal of Statistical Physics, 183 (2021), p. 49.

[77] A. Gribinski, *A theory of singular values for finite free probability*, 2022.

[78] A. Gribinski and A. W. Marcus, *A rectangular additive convolution for polynomials*, 2019.

[79] D. J. Gross and M. Mézard, *The simplest spin glass*, Nuclear Physics B, 240 (1984), pp. 431–452.

[80] D. J. Gross and E. Witten, *Possible third-order phase transition in the large–N lattice gauge theory*, Physical Review D, 21 (1980), p. 446.

[81] T. Guhr and H. Kohler, *Recursive construction for a class of radial functions. i. ordinary space*, Journal of Mathematical Physics, 43 (2002), pp. 2707–2740.

[82] A. Guionnet, *Random matrices and enumeration of maps*, Proceedings oh the International Congress of Mathematicians, Vol. 3, 2006-01-01, ISBN 978-3-03719-022-7, pags. 623-636, 3 (2006).

[83] A. Guionnet and J. Huang, *Large deviations asymptotics of rectangular spherical integral*, 2021.

[84] A. Guionnet and M. Maïda, *A Fourier view on the R-transform and related asymptotics of spherical integrals*, Journal of Functional Analysis, 222 (2005), pp. 435–490.

[85] A. Guionnet and M. Maïda, *Large deviations for the largest eigenvalue of the sum of two random matrices*, Electronic Journal of Probability, 25 (2020), pp. 1 – 24.

[86] A. Guionnet and O. Zeitouni, *Large deviations asymptotics for spherical integrals*, Journal of Functional Analysis, 188 (2002), pp. 461–515.

[87] Y. Guo and A. Amir, *Stability of gene regulatory networks*, 2020.

[88] Harish-Chandra, *Differential Operators on a Semisimple Lie Algebra*, American Journal of Mathematics, 79 (1957), p. 87.

[89] G. J. Heckman and E. M. Opdam, *Root systems and hypergeometric functions. i*, Compositio Mathematica, 64 (1987), pp. 329–352.

[90] S. Helgason, *Differential geometry, Lie groups, and symmetric spaces*, Academic press, 1979.

[91] N. L. Hjort and A. Ongaro, *Exact inference for random Dirichlet means*, Statistical Inference for Stochastic Processes, 8 (2005), pp. 227–254.

[92] A. Horn, *Eigenvalues of sums of Hermitian matrices.*, Pacific Journal of Mathematics, 12 (1962), pp. 225 – 241.

[93] J. Husson, *Asymptotic behavior of multiplicative spherical integrals and S-transform*, preprint arXiv:2108.11842, (2021).

[94] J. R. Ipsen and H. Schomerus, *Isotropic Brownian motions over complex fields as a solvable model for May–Wigner stability analysis*, Journal of Physics A: Mathematical and Theoretical, 49 (2016), p. 385201.

[95] M. E. Ismail, *An electrostatics model for zeros of general orthogonal polynomials*, Pacific Journal of Mathematics, 193 (2000), pp. 355–369.

[96] C. Itzykson and J.-B. Zuber, *The planar approximation. II*, Journal of Mathematical Physics, 21 (1980), pp. 411–421.

[97] H. C. Ji and J. Park, *Tracy-Widom limit for free sum of random matrices*, 2021.

[98] A. Karadimitrakis, A. L. Moustakas, and P. Vivo, *Outage capacity for the optical mimo channel*, IEEE Transactions on Information Theory, 60 (2014), pp. 4370–4382.

[99] J. P. Keating and N. C. Snaith, *Random matrix theory and* $\zeta(1/2+it)$, Communications in Mathematical Physics, 214 (2000), pp. 57–89.

[100] M. Kieburg and H. Kösters, *Exact relation between singular value and eigenvalue statistics*, Random Matrices: Theory and Applications, 05 (2016), p. 1650015.

[101] R. Killip and I. Nenciu, *Matrix models for circular ensembles*, International Mathematics Research Notices, 2004 (2004), pp. 2665–2701.

[102] A. A. Klyachko, *Stable bundles, representation theory and hermitian operators*, Selecta Mathematica, 4 (1998), p. 419.

[103] A. Knutson and T. Tao, *The honeycomb model of $GL_n(\mathbb{C})$ tensor products i: Proof of the saturation conjecture*, Journal of the American Mathematical Society, 12 (1999), pp. 1055–1090.

[104] J. M. Kosterlitz, D. J. Thouless, and R. C. Jones, *Spherical model of a spin-glass*, Phys. Rev. Lett., 36 (1976), pp. 1217–1220.

[105] A. Krajenbrink, P. Le Doussal, and N. O'Connell, *Tilted elastic lines with columnar and point disorder, non-Hermitian quantum mechanics, and spiked random matrices: Pinning and localization*, Physical Review E, 103 (2021), p. 042120.

[106] T. Kriecherbauer and J. Krug, *A pedestrian's view on interacting particle systems, KPZ universality and random matrices*, Journal of Physics A: Mathematical and Theoretical, 43 (2010), p. 403001.

[107] A. B. J. Kuijlaars and P. Romàn, *Spherical Functions Approach to Sums of Random Hermitian Matrices*, International Mathematics Research Notices, 2019 (2017), pp. 1005–1029.

[108] M. Lal Mehta, *Random Matrices (Volume 142) (Pure and Applied Mathematics, Volume 142)*, Academic Press, 11 2004.

[109] J. O. Lee and K. Schnelli, *Edge universality for deformed Wigner matrices*, Reviews in Mathematical Physics, 27 (2015), p. 1550018.

[110] M. Lelarge and L. Miolane, *Fundamental limits of symmetric low-rank matrix estimation*, Probability Theory and Related Fields, 173 (2019), pp. 859–929.

[111] A. Lenard, *Exact statistical mechanics of a one-dimensional system with Coulomb forces*, Journal of Mathematical Physics, 2 (1961), pp. 682–693.

[112] G. Letac, M. Piccioni, et al., *Dirichlet curves, convex order and Cauchy distribution*, Bernoulli, 24 (2018), pp. 1–29.

[113] G. Livan, M. Novaes, and P. Vivo, *Introduction to random matrices*, SpringerBriefs in Mathematical Physics, Springer International Publishing, Basel, Switzerland, 1 ed., jan 2018.

[114] C. Louart, Z. Liao, and R. Couillet, *A random matrix approach to neural networks*, The Annals of Applied Probability, 28 (2018), pp. 1190–1248.

[115] I. G. Macdonald, *Symmetric Functions and Hall Polynomials (Oxford Classic Texts in the Physical Sciences: Oxford Mathematical Mongraphs)*, Oxford University Press, December 2015.

[116] A. Maillard, *Large deviations of extreme eigenvalues of generalized sample covariance matrices*, EPL (Europhysics Letters), 133 (2021), p. 20005.

[117] S. N. Majumdar, C. Nadal, A. Scardicchio, and P. Vivo, *Index distribution of Gaussian random matrices*, Phys. Rev. Lett., 103 (2009), p. 220603.

[118] ——, *How many eigenvalues of a Gaussian random matrix are positive?*, Phys. Rev. E, 83 (2011), p. 041105.

[119] S. N. Majumdar, A. Pal, and G. Schehr, *Extreme value statistics of correlated random variables: A pedagogical review*, Physics Reports, 840 (2020), pp. 1–32. Extreme value statistics of correlated random variables: A pedagogical review.

[120] S. N. Majumdar and G. Schehr, *Top eigenvalue of a random matrix: large deviations and third order phase transition*, Journal of Statistical Mechanics: Theory and Experiment, 2014 (2014), p. P01012.

[121] S. N. Majumdar and M. Vergassola, *Large deviations of the maximum eigenvalue for Wishart and Gaussian random matrices*, Phys. Rev. Lett., 102 (2009), p. 060601.

[122] S. N. Majumdar and P. Vivo, *Number of relevant directions in principal component analysis and Wishart random matrices*, Phys. Rev. Lett., 108 (2012), p. 200601.

[123] V. A. Marčenko and L. A. Pastur, *DISTRIBUTION OF EIGENVALUES FOR SOME SETS OF RANDOM MATRICES*, Mathematics of the USSR-Sbornik, 1 (1967), pp. 457–483.

[124] A. Marcus, D. A. Spielman, and N. Srivastava, *Interlacing families I: Bipartite ramanujan graphs of all degrees*, in 2013 IEEE 54th Annual Symposium on Foundations of Computer Science, 2013, pp. 529–537.

[125] A. W. Marcus, *Polynomial convolutions and (finite) free probability*, 2021.

[126] A. W. Marcus, D. A. Spielman, and N. Srivastava, *Interlacing families II: Mixed characteristic polynomials and the kadison—singer problem*, Annals of Mathematics, 182 (2015), pp. 327–350.

[127] A. W. Marcus, D. A. Spielman, and N. Srivastava, *Interlacing families IV: Bipartite Ramanujan graphs of all sizes*, SIAM Journal on Computing, 47 (2018), pp. 2488–2509.

[128] ——, *Finite free convolutions of polynomials*, Probability Theory and Related Fields, 182 (2022), pp. 807–848.

[129] ——, *Interlacing families III: Sharper restricted invertibility estimates*, Israel Journal of Mathematics, 247 (2022), pp. 519–546.

[130] A. W. Marcus and W. Yomjinda, *Analysis of rank 1 perturbations in general β ensembles*, preprint, (2016).

[131] E. Marinari, G. Parisi, and F. Ritort, *Replica field theory for deterministic models. II. a non-random spin glass with glassy behaviour*, Journal of Physics A: Mathematical and General, 27 (1994), pp. 7647–7668.

[132] M. Mariño, *Chern–Simons theory, matrix integrals, and perturbative three-manifold invariants*, Communications in Mathematical Physics, 253 (2005), pp. 25–49.

[133] ——, *Matrix models and topological strings*, in Applications of random matrices in physics, Springer, 2006, pp. 319–378.

[134] R. Marino, S. N. Majumdar, G. Schehr, and P. Vivo, *Index distribution of Cauchy random matrices*, Journal of Physics A: Mathematical and Theoretical, 47 (2014), p. 055001.

[135] K. Matveev, *Macdonald-positive specializations of the algebra of symmetric functions: Proof of the Kerov conjecture*, Annals of Mathematics, 189 (2019), pp. 277 – 316.

[136] A. Matytsin, *On the Large-N limit of the Itzykson-Zuber integral*, Nuclear Physics B, 411 (1994), pp. 805–820.

[137] R. M. May, *Will a large complex system be stable?*, Nature, 238 (1972), pp. 413–414.

[138] G. Mazzuca, *On the mean density of states of some matrices related to the beta ensembles and an application to the Toda lattice*, Journal of Mathematical Physics, 63 (2022), p. 043501.

[139] M. Mehta and M. Gaudin, *On the density of eigenvalues of a random matrix*, Nuclear Physics, 18 (1960), pp. 420–427.

[140] P. Mergny, *An explicit example for the high temperature convolution: crossover between the binomial law $b(2, 1/2)$ and the arcsine law*, 2022.

[141] P. Mergny and S. N. Majumdar, *Stability of large complex systems with heterogeneous relaxation dynamics*, Journal of Statistical Mechanics: Theory and Experiment, 2021 (2021), p. 123301.

[142] P. Mergny and M. Potters, *Asymptotic behavior of the multiplicative counterpart of the Harish-Chandra integral and the S-transform*, preprint arXiv:2007.09421, (2020).

[143] P. Mergny and M. Potters, *Rank one HCIZ at high temperature: interpolating between classical and free convolutions*, SciPost Phys., 12 (2022), p. 22.

[144] ——, *Right large deviation principle for the top eigenvalue of the sum or product of invariant random matrices*, Journal of Statistical Mechanics: Theory and Experiment, 2022 (2022), p. 063301.

[145] J. A. Mingo and R. Speicher, *Free probability and random matrices*, vol. 35, Springer, 2017.

[146] J. Moran and J.-P. Bouchaud, *May's instability in large economies*, Phys. Rev. E, 100 (2019), p. 032307.

[147] K. A. Muttalib, *Random matrix models with additional interactions*, Journal of Physics A: Mathematical and General, 28 (1995), p. L159.

[148] C. Nadal, S. N. Majumdar, and M. Vergassola, *Statistical distribution of quantum entanglement for a random bipartite state*, Journal of Statistical Physics, 142 (2011), pp. 403–438.

[149] A. Nica and R. Speicher, *Lectures on the Combinatorics of Free Probability*, Cambridge University Press, 2006.

[150] A. Okounkov and G. Olshanski, *Shifted Jack polynomials, binomial formula, and applications*, Mathematical Research Letters, 4 (1997), pp. 67–78.

[151] J. Pennington and P. Worah, *Nonlinear random matrix theory for deep learning*, in Advances in Neural Information Processing Systems, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds., vol. 30, Curran Associates, Inc., 2017.

[152] N. S. Pillai and J. Yin, *Universality of covariance matrices*, The Annals of Applied Probability, 24 (2014), pp. 935 – 1001.

[153] M. Potters and J.-P. Bouchaud, *A first course in random matrix theory*, Cambridge University Press, 2020.

[154] S. Prager, *The one-dimensional plasma*, Advances in chemical physics, 4 (1962), pp. 201–224.

[155] J. Ramírez, B. Rider, and B. Virág, *Beta ensembles, stochastic Airy spectrum, and a diffusion*, Journal of the American Mathematical Society, 24 (2011), pp. 919–944.

[156] V. Ros, G. Biroli, and C. Cammarota, *Complexity of energy barriers in mean-field glassy systems*, EPL (Europhysics Letters), 126 (2019), p. 20003.

[157] Z. Rudnick and P. Sarnak, *Zeros of principal l-functions and random matrix theory*, Duke Mathematical Journal, 81 (1996), pp. 269–322.

[158] M. Rösler, *A positive radial product formula for the dunkl kernel*, Transactions of the American Mathematical Society, 355 (2003), pp. 2413–2438.

[159] G. Schehr, *Extremes of n vicious walkers for large n: Application to the directed polymer and kpz interfaces*, Journal of Statistical Physics, 149 (2012), pp. 385–410.

[160] B. Schlittgen and T. Wettig, *Generalizations of some integrals over the unitary group*, Journal of Physics A: Mathematical and General, 36 (2003), pp. 3195–3201.

[161] H. Sompolinsky, A. Crisanti, and H.-J. Sommers, *Chaos in random neural networks*, Phys. Rev. Lett., 61 (1988), p. 259.

[162] H. Spohn, *Exact solutions for kpz-type growth processes, random matrices, and equilibrium shapes of crystals*, Physica A: Statistical Mechanics and its Applications, 369 (2006), pp. 71–99.

[163] ——, *Generalized gibbs ensembles of the classical toda chain*, Journal of Statistical Physics, 180 (2020), pp. 4–22.

[164] R. P. Stanley, *Some combinatorial properties of Jack symmetric functions*, Advances in Mathematics, 77 (1989), pp. 76–115.

[165] Y. Sun, *A new integral formula for Heckman–Opdam hypergeometric functions*, Advances in Mathematics, 289 (2016), pp. 1157–1204.

[166] R. J. Szabo and M. Tierz, *Chern–Simons matrix models, two-dimensional Yang–Mills theory and the Sutherland model*, Journal of Physics A: Mathematical and Theoretical, 43 (2010), p. 265401.

[167] G. Szegő, *Orthogonal Polynomials*, American Mathematical Society, dec 1939.

[168] Y. Takahashi and M. Katori, *Noncolliding Brownian motion with drift and time-dependent Stieltjes–Wigert determinantal point process*, Journal of mathematical physics, 53 (2012), p. 103305.

[169] T. Tao and V. Vu, *Random matrices: Universality of local eigenvalue statistics up to the edge*, Communications in Mathematical Physics, 298 (2010), pp. 549–572.

[170] T. Tao and V. Vu, *Random matrices: Universality of local eigenvalue statistics*, Acta Mathematica, 206 (2011), pp. 127 – 204.

[171] M. Tierz, *Schur polynomials and biorthogonal random matrix ensembles*, Journal of mathematical physics, 51 (2010), p. 063509.

[172] C. A. Tracy and H. Widom, *Level-spacing distributions and the Airy kernel*, Communications in Mathematical Physics, 159 (1994), pp. 151–174.

[173] ——, *On orthogonal and symplectic matrix ensembles*, Communications in Mathematical Physics, 177 (1996), pp. 727–754.

[174] F. G. Tricomi, *Integral Equations (Dover Books on Mathematics)*, Dover Publications, paperback ed., 3 1985.

[175] H. D. Trinh and K. D. Trinh, *Beta Jacobi ensembles and associated jacobi polynomials*, Journal of Statistical Physics, 185 (2021), pp. 1–15.

[176] K. D. Trinh, *Global spectrum fluctuations for Gaussian beta ensembles: A martingale approach*, Journal of Theoretical Probability, 32 (2017), pp. 1420–1437.

[177] E. Troiani, V. Erba, F. Krzakala, A. Maillard, and L. Zdeborová, *Optimal denoising of rotationally invariant rectangular matrices*, 2022.

[178] A. M. Tulino and S. Verdú, *Random matrix theory and wireless communications*, Now Publishers Inc, 2004.

[179] J. Verbaarschot, *Spectrum of the QCD Dirac operator and chiral random matrix theory*, Phys. Rev. Lett., 72 (1994), pp. 2531–2533.

[180] P. Vivo, *Large deviations of the maximum of independent and identically distributed random variables*, European Journal of Physics, 36 (2015), p. 055037.

[181] P. Vivo, S. N. Majumdar, and O. Bohigas, *Large deviations of the maximum eigenvalue in Wishart random matrices*, Journal of Physics A: Mathematical and Theoretical, 40 (2007), pp. 4317–4337.

[182] D. Voiculescu, *Limit laws for random matrices and free products*, Inventiones mathematicae, 104 (1991), pp. 201–220.

[183] D. V. Voiculescu, *Addition of certain non-commuting random variables*, Journal of Functional Analysis, 66 (1986), pp. 323–346.

[184] D. V. Voiculescu, *Free probability theory*, vol. 12, American Mathematical Soc., 1997.

[185] S. R. Wadia, *$N = \infty$ phase transition in a class of exactly soluble model lattice gauge theories*, Physics Letters B, 93 (1980), pp. 403–410.

[186] G. Wainrib and J. Touboul, *Topological and dynamical complexity of random neural networks*, Phys. Rev. Lett., 110 (2013), p. 118101.

[187] G. N. Watson, *A Treatise on the Theory of Bessel Functions (Cambridge Mathematical Library)*, Cambridge University Press, 8 1995.

[188] E. P. Wigner, *On the distribution of the roots of certain symmetric matrices*, Annals of Mathematics, 67 (1958), pp. 325–327.

[189] J. Wishart, *The generalised product moment distribution in samples from a normal multivariate population*, Biometrika, 20A (1928), pp. 32–52.

[190] J. Zhang, M. Kieburg, and P. J. Forrester, *Harmonic analysis for rank-1 randomised Horn problems*, Letters in Mathematical Physics, 111 (2021), p. 98.

[191] J.-B. Zuber, *Horn's problem and Harish-Chandra's integrals. probability density functions*, Annales de l'Institut Henri Poincaré D, 5 (2018), pp. 309–338.