New Applications of Random Matrices in Spin Glass and Machine Learning

by

Hao Wu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Applied and Interdisciplinary Mathematics)
in The University of Michigan
2019

Doctoral Committee:

Professor Jinho Baik, Co-Chair
Associate Professor Raj Rao Nadakuditi, Co-Chair
Professor Alfred Hero
Professor Peter Miller
Professor Mark Rudelson

Hao Wu

lingluan@umich.edu

ORCID iD: 0000-0002-7629-6915

For all the people

ACKNOWLEDGEMENTS

There are many people from whom I get selfless help on my way towards this dissertation. First, I would like to thank my family in China for their love and support, for their understanding and encouragement so that I can fully pursue the best version of myself while finding a balanced life.

I want to show my gratefulness to my advisors Jinho Baik and Raj Rao Nadakuditi for their patience and guidance. I learn a lot from all those moments we work together. They present me with the beauty and elegance of mathematics and science and share with me their enthusiasm and wisdom. They teach me how to do research and how to express in a clean style.

There is a lot of faculty kindly offering me their time and knowledge. Among them, I would like to especially thank Alfred Hero, Peter Miller, Mark Rudelson, Roman Vershynin for serving on my committee and providing valuable comments to my proposal and thesis.

Finally, many thanks to my friends and colleagues, especially Weidong Chen, Siwei Lu, Xinyi Ge, Kaixin Wang, Chencheng Zhou, Yuchen Liao, Arvind Prasadan and Luby Lu. They enlighten my life in Ann Arbor and make my graduate study at the University of Michigan full of fun and joy.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF FIGURES viii

LIST OF TABLES xi

LIST OF APPENDICES xii

ABSTRACT xiii

CHAPTER

I. An overview of this thesis 1

II. Introduction of spin glass 4

2.1 The origin of spin glass 4
2.2 Three spin glass models 5
2.2.1 The Edwards-Anderson model. 5
2.2.2 The Sherrington-Kirkpatrick model. 7
2.2.3 The Spherical Sherrington-Kirkpatrick model. . . . 9
2.3 Outline of Chapter IV and V 11
2.3.1 Paramagnetic-Ferromagnetic phase transition for 2-spin SSK with Curie-Weiss interaction 11
2.3.2 2-spin SSK with external field 13
2.3.3 Method 14

III. Random matrices preliminary 16

3.1 Law of large number 16
3.2 Linear Statistics 18
3.3 Eigenvalue rigidity 19

3.4 Edge Behaviors 20
3.5 Spiked Wigner matrices 21

IV. Ferromagnetic to paramagnetic transition in spherical spin glass with Curie-Weiss interaction 23

4.1 Introduction 23
4.1.1 Model 25
4.1.2 Previous results in each regime 27
4.1.3 Results 29
4.1.4 Organization 33
4.2 Results on Wigner matrices with non-zero mean 33
4.2.1 Partial linear statistics 34
4.2.2 Joint convergence of the largest eigenvalue and linear statistics 36
4.2.3 Proof of Theorem IV.5 37
4.3 Proof of Theorem IV.4 38
4.3.1 Preliminaries 38
4.3.2 Steepest-descent analysis 41
4.3.3 Proof of Theorem IV.4 50
4.4 Partial linear statistics 51
4.4.1 Proof of Theorem IV.6 52
4.4.2 Proof of Proposition IV.17 54
4.4.3 Proof of Lemma IV.19 56
4.5 Joint Distribution of χ_N and $\mathcal{N}_N^{(2)}(\varphi)$ 56
4.5.1 Asymptotic Independence for the GOE case 57
4.5.2 Proof of Theorem IV.8 for general case 64
4.6 Matching 78
4.6.1 $B \to +\infty$ 79
4.6.2 $B \to -\infty$ 80

V. Spherical spin glass model with external field 82

5.1 Introduction 82
5.2 Single integral representations 85
5.3 Results from random matrices 87
5.3.1 Notations 87
5.3.2 Semicircle law 88
5.3.3 Rigidity 88
5.3.4 Edge behavior 89
5.3.5 Central limit theorem of linear statistics 89
5.3.6 Special sums 91
5.3.7 Weighted sums of independent random variables . . 93
5.3.8 Another special sums 94
5.4 Fluctuations of the free energy 95

5.4.1 No external field: $h = 0$ 95
5.4.2 Positive external field: $h > 0$ 98
5.4.3 Comparison between $h > 0$ and $h = 0$ cases 100
5.5 Transition of the free energy in the high temperature regime . 102
5.5.1 Result 102
5.5.2 Comparison with $h = 0$ and $h > 0$ cases 103
5.6 Transition of the free energy in the low temperature regime . 105
5.6.1 Result 105
5.6.2 Comparison with the case when $h = 0$ 108
5.6.3 Comparison with the case when $h > 0$ 108

VI. Introduction of free component analysis 113

6.1 From PCA to ICA via cumulants 114
6.2 Known result: ICA unmixes mixtures of independent random variables 116
6.3 Our contribution: From ICA to FCA via free cumulants . . . 117
6.4 Our main finding: FCA unmixes mixtures of free random variables 119
6.5 Insight: FCA unmixes mixtures of (asymptotically) free random matrices 121
6.6 Insight: FCA can be applied wherever ICA has been applied . 122
6.7 Surprising insight: FCA often better unmixes random variables than ICA 125
6.8 Organization 126

VII. Free component analysis: main result, simulation and proof 129

7.1 Main result: Recovery guarantees for FCA 129
7.1.1 Setup and assumptions under an orthogonal mixing model 129
7.1.2 Free kurtosis based FCA 130
7.1.3 Setup and assumptions under a non-orthogonal mixing model 135
7.1.4 Unmixing mixed free random variables using FCA . 136
7.1.5 Overdetermined and underdetermined FCA 139
7.1.6 Unmixing mixtures of matrices using FCA 140
7.1.7 Numerical algorithms for Free Component Factorization 142
7.2 Numerical Simulations 143
7.2.1 Unmixing of self-adjoint matrices using self-adjoint FCA 145
7.2.2 Unmixing of rectangular matrices with rectangular FCA 147
7.2.3 Unmixing mixed images using rectangular FCA . . 148

7.2.4 Unmixing performance of free kurtosis vs free entropy FCA vs ICA . . . 150
7.2.5 Unmixing mixed waveforms using rectangular FCA 151
7.2.6 Unmixing rectangular matrices using self-adjoint FCA and more . . . 155
7.3 Conclusions and Open Problems . . . 155
7.3.1 Open Problems . . . 156
7.3.2 Open Problem: Using FCA to construct new matrix models for freeness . . . 157
7.3.3 Open Problem: Improving FCA by "more free" submatrix selection . . . 158
7.4 Proof of Proposition VII.13 and VII.14 . . . 160
7.4.1 Proof of Proposition VII.13 . . . 160
7.4.2 Proof of Proposition VII.14 . . . 161
7.5 Proofs of the main results . . . 162
7.5.1 Proof of Theorem VII.1 . . . 162
7.5.2 Proof of Theorem VII.3 . . . 166
7.5.3 Proof of Theorem VII.4 . . . 167
7.5.4 Proof of Theorem VII.8 . . . 170
7.5.5 Proof of Theorem VII.9 . . . 174
7.6 Proof of Theorem VII.19 . . . 180

APPENDICES . . . 184

A. Replica method and replica symmetry breaking . . . 185

A.1 Replica method for p-spin SSK model . . . 185
A.2 Replica symmetry breaking and Parisi formula . . . 189

B. Free probability, matrix embeddings and independent component analysis . . . 193

B.1 What is freeness of random variables? . . . 194
B.1.1 Prologue: What is independence of commuting random variables? . . . 194
B.1.2 Freeness of self-adjoint random variables . . . 197
B.1.3 Freeness of non self-adjoint random variables . . . 201
B.1.4 When are random matrices free? . . . 204
B.2 Matrix Embeddings . . . 207
B.3 Independent Component Factorization . . . 211

BIBLIOGRAPHY . . . 213

LIST OF FIGURES

Figure

2.1 Phase diagram for SSK+CW model. Here, β is the inverse temperature and θ is the coupling constant. 13

4.1 Phase diagram for SSK+CW model. Here, β is the inverse temperature and J is the coupling constant. 24

4.2 (a) Probability density function of $Q(\mathcal{G}_2)$ for $B = -1, 0, 1$, (b) Probability density function of normalized $Q(\mathcal{G}_2)$ resembles a Gaussian density as $B \to +\infty$. 32

6.1 An experiment in image separation using ICA and FCA. Note that subplots (c), (g) (unmixed images via ICA) and (d), (h) (unmixed images via FCA) both recover (a), (e) respectively. Here, $\boldsymbol{A} = [\sqrt{2}, \sqrt{2}; -\sqrt{2}, \sqrt{2}]/2$ in (6.10). The error of ICA is 6.08×10^{-2} while the error of FCA is 2.69×10^{-2}. See (7.32) for the definition of the error. 123

6.2 An experiment in waveform separation using ICA and FCA. Note that subplots (c), (g) (unmixed waves via ICA) and (d), (h) (unmixed waves via FCA) both recover (a), (e). Visually, FCA performs better in this experiment. In this simulation, $\boldsymbol{A} = [\sqrt{2}, \sqrt{2}; -\sqrt{2}, \sqrt{2}]/2$ in (6.10). The errors for ICA and FCA are 9.95×10^{-2} and 6.66×10^{-2} respectively. 124

6.3 An experiment in audios separation via ICA and FCA: Note that subplots (c), (g) (unmixed audio signals via ICA) and (d), (h) (unmixed audio signals via FCA) both recover (a), (e). In this experiment, $\boldsymbol{A} = [\sqrt{2}, \sqrt{2}; -\sqrt{2}, \sqrt{2}]/2$ in (6.10). The errors for ICA and FCA are 1.47×10^{-2} and 1.79×10^{-2} respectively. 125

6.4 An experiment in image denoising via ICA and (kurtosis-based) FCA: Comparing (g) and (f), we observe that FCA does a better job then ICA in this experiment. Here $\boldsymbol{A} = [\sqrt{2}, \sqrt{2}; -\sqrt{2}, \sqrt{2}]/2$ in (6.10). The variance of whitened Gaussian noise is set to equal the empirical variance of original image. 126

6.5 We can regard ICA and FCA with various embedding as "projections" onto corresponding manifolds. Here, the gray surface denotes the manifold of independent pairs. The red and blue surfaces stand for self-adjoint free pairs and rectangular free pairs respectively. In order to achieve the best performance, one shall pick the projection into the closest manifold. For example, if the latent data is $(X_1^{(1)}, X_2^{(1)})$, then rectangular FCA should have the best performance when separating them from the additive mixture. In contrast, for the underlying data $(X_1^{(2)}, X_2^{(2)})$, one should pick ICA. 127

7.1 An experiment in the separation of self-adjoint matrices. The mixing matrix $\boldsymbol{A} = [0.5, 0.5; -0.5, 0.6]$, $N = 800, M = 1600$. The average errors over 200 trials are 8.67×10^{-3} for kurtosis-based FCF and 6.44×10^{-3} for entropy-based FCF. 146

7.2 An experiment in the separation of rectangular random matrices. In this simulation, $\boldsymbol{A} = [0.5, 0.5; -0.5, 0.6]$, $N = 800, M = 1000$ and $f(x) = (x-1)^4$. The average errors over 200 trials are 1.55×10^{-3} for kurtosis-based FCF and 8.81×10^{-4} for entropy-based FCF. . . 148

7.3 (a) Polar graph of $c(\theta)$ (b) Polar graph of $\kappa(\theta)$ (c) Polar graph of $E(\theta)$. The red dashed lines stand for the direction of 45°. Note that the directions of maximum of $\kappa(\theta)$ and $E(\theta)$ agree with the red line well, while the maximum of $c(\theta)$ is off. Because of the randomness of Gaussian noise, for different trials, the polar graph of ICA will vary. However, the polar graphs of FCA are relatively stable. 151

7.4 (a) Averaged (over 200 trials) errors of kurtosis based FCA and ICA for increasing dimension. (b) Averaged error of entropy based FCA and ICA for increasing dimension. (c) CDF of $\text{Error}_{fca}/\text{Error}_{ica}$ for the kurtosis-based method. (d) cdf of $\text{Error}_{fca}/\text{Error}_{ica}$ for the entropy-based method. All methods appear to have a convergence rate of N^{-1}. In this simulation, we set $N/M = 0.8$ and $f(x) = (x-1)^4$. 152

7.5 The spectrogram of signals. For the spectrogram, we adapt the Hamming window of 250 samples, the number of overlapped samples is 125, the number of DFT points is 256. 154

7.6 An experiment in waveform separation using ICA and FCA. Note that subplots (c), (g) (unmixed waves via ICA) and (d), (h) (unmixed waves via FCA) both recover (a), (e). In this simulation, $\boldsymbol{A} = [\sqrt{2}, \sqrt{2}; -\sqrt{2}, \sqrt{2}]/2$ in (6.10). The errors for ICA and FCA are 7.70×10^{-5} and 1.36×10^{-2} respectively. 155

7.7 Patch bases obtained via an ICA (top row) or FCA (bottom row) factorization of 6×6 patches of the panda and the hedgehog images from Figures 6.1a and 6.1e respectively. 158

7.8 An Application of FCA to images not so free. The mixing matrix is $\boldsymbol{A} = [\sqrt{2}, \sqrt{2}; -\sqrt{2}, \sqrt{2}]/2$ 160

7.9 Application of FCA to sub images gives better results. The mixing matrix is $\boldsymbol{A} = [\sqrt{2}, \sqrt{2}; -\sqrt{2}, \sqrt{2}]/2$ 161

7.10 Normalized square of inner products between: (a) left singular vectors of Figure 7.8a and 7.8d, (b) right singular vectors of Figure 7.8d and 7.8a, (c) left singular vectors of Figure 6.1a and 6.1e, (d) right singular vectors of Figure 6.1a and 6.1e. We observe that inner product between the right singular vectors of Figure 7.8d and 7.8a (corresponding to (b)) are clearly not uniform. 162
A.1 Matrices $\widehat{Q}$ corresponding to the "replica symmetry" ansatz and the first two RSB ansatzs. The first matrix represents $\widehat{Q}$ for the "replica symmetry" ansatz. The second and third matrices are $\widehat{Q}$ for the 1RSB and 2RSB scheme respectively. In this diagram, we divide every groups into two subgroups ($m_1 = n/2$ and $m_2 = m_1/2$). In each step, we further breaking matrices on the diagonal into smaller sub matrices while off-diagonal matrices remain unchanged. 191
A.2 Full RSB scheme. Left panel: the overlap between replicas α and β is q_1, which corresponds to the first common level containing both α and β. Right panel: replicas are divided into subsets iteratively. Figure from [93]. 192

LIST OF TABLES

Table

6.1 FCA algorithms and their limits. 120
7.1 Formulas for $\widehat{F}(\cdot)$ in Algorithm 2. Here $\boldsymbol{X}$ is either a self-adjoint or a rectangular matrix. 143
7.2 Euclidean gradients for the setting in Theorem VII.19 144

LIST OF APPENDICES

Appendix

A. Replica method and replica symmetry breaking 185

A.1 Replica method for p-spin SSK model 185
A.2 Replica symmetry breaking and Parisi formula 189

B. Free probability, matrix embeddings and independent component analysis 193

B.1 What is freeness of random variables? 194
B.2 Matrix Embeddings 207
B.3 Independent Component Factorization 211

ABSTRACT

Recent advancement in random matrix theory is beneficial to challenging problems in many disciplines of science and engineering. In another direction, these applications motivate a lot of new questions in random matrix theory. In this thesis, we present two applications of random matrix theory to statistical physics and machine learning.

The first part of this thesis is about the spherical Sherrington-Kirkpatrick (SSK) model in statistical physics. The SSK model is defined by a random probability measure on a high dimensional sphere. The probability measure involves the temperature and a random Hamiltonian. We consider the simplest non-trivial case where the Hamiltonian is a random symmetric quadratic form perturbed by a specific symmetric polynomial of degree one or two. It is interesting to consider the interaction between the quadratic form and the perturbations. In particular, using the obvious connection between random quadratic forms and random matrices, we study the free energies and obtain the limiting law of their fluctuations as the dimension becomes large.

The second part is devoted to an application of the random matrix theory in machine learning. We develope Free component analysis (FCA) for unmixing signals in the matrix form from their linear mixtures with little prior knowledge. The matrix signals are modeled as samples of random matrices, which are further regarded as non-commutative random variables. The counterpart of scalar probability for non-commutative random variables is the free probability. Our principle of separation is

to maximize free independence between the unmixed signals. This is achieved in a manner analogous to the independent component analysis (ICA) based method for unmixing independent random variables from their additive mixtures. We describe the theory, the various algorithms, and compare FCA to ICA. We show that FCA performs comparably to, and often better than, ICA in every application, such as image and speech unmixing, where ICA has been known to succeed.

CHAPTER I

An overview of this thesis

This thesis consists of applications of random matrices in statistical physics and machine learning.

The first part of this thesis, including Chapters II to V, is about spin glass models in statistical physics. The spin glass models were first invented around the early 60s in order to describe the behavior of magnetic alloys. Here, we present the results of the fluctuations of the free energy of two specific spin glass models.

Chapter II is a brief historical review. We focus on the origin of spin glass and discuss the necessary background for a exposure of the spherical Sherrington-Kirkpatrick (SSK) model, which is the model of our interest. Since the spin glass has evolved for 50 years with various branches, this introduction is not intended to be a comprehensive one. A classical approach for free energy is presented in Appendix A.

Chapter III contains the necessary random matrices results. The object of our interest is the behavior of eigenvalues of Wigner random matrices. We will start with the famous Wigner semicircle law and then present more recent results: central limit theorem of eigenvalues, eigenvalue rigidity, edge behavior, fluctuation of leading eigenvalue for spiked Wigner matrice.

With the preparation in Chapter II and Chapter III, we present our contributions in Chapter IV and Chapter V. In these two chapters, we consider the 2-spin SSK

model with Curie-Weiss (CW) interaction and external field respectively. Mathematically, CW interaction can be regarded as a deterministic quadratic perturbation while the external field is linear. Depending on the temperature and the strength of the perturbation,both models exhibit distinct behaviors in different domains (phases). Our contribution is an systematic analysis of fluctuations of the free energy near phase transition regimes using the steepest descent and the random matrix results. The results in Chapter IV are published in a joint paper [15] with Baik and Lee. The results in Chapter V will be a part of a new joint paper [11] with Baik and le Doussal.

The second part, consisting of Chapter VI and Chapter VII, considers the blind source separation (BSS) problem in machine learning. The goal of the BSS problem is to recover source signals from a set of mixed signals, with very little information about the source signals or the mixing process. Here, we propose the a innovate BSS algorithm called Free Component Analysis (FCA). FCA is a method for unmixing mixtures of freely independent random variables. Random matrices play the role of free random variables in this setting. In terms of applications, FCA is designed to separate data whose components are matrices. For example, a typical application of FCA is to separate images from their mixture. The results in Chapter VI will be published in a joint paper with Nadakuditi [123].

In Chapter VI, we give a brief introduction about BSS and some known approaches including Principal Component Analysis (PCA) and Independent component analysis (ICA). Then we introduce FCA as a natural analog of ICA for non-commutative random variables (e.g. random matrices). The probability framework of the non-commutative random variables is called free probability, the intuition of which is given in Chapter VI while a rigorous treatment is provided in Appendix B.1 and the reference therein.

We present the framework of FCA in a thorough, systematic manner in Chapter VII, which contains variations of FCA, recovery guarantee and rigorous proofs. It

worth mentioning that most of the proofs are parallel to its counterpart in ICA, but they are based on results from the free probability and asymptotic freeness of random matrices. Some numerical simulations are also included where we test the FCA algorithms in both theoretical and practical settings.

CHAPTER II

Introduction of spin glass

This part of the thesis focuses on a specific spin glass model. In this chapter, we make a historical presentation of the birth of the spin glass theory and the evolution of the model in which we are interested. For a more general introduction of the spin glass, interested readers are referred to [89, 33].

2.1 The origin of spin glass

The first uncommon experimental phenomenon of the spin glass was detected around the 1960s. In the study of manganese(Mn)-copper(Cu) alloy, people observed a cusp in the susceptibility at a critical temperature [49, 125]. In addition, other physical observables including magnetization and specific heat demonstrated peculiar properties in the low temperature regime. All these pieces of evidence indicated a new magnetic phase, which was called "spin-glass" phase later.

Above compelling phenomenon turned out to be the consequence of the following two facts. First, instead of being allied in a regular pattern, the Mn ions in the alloy are positioned randomly. (This analogy with the positional disorder of the conventional glass gives rise to the term "glass".) Second, the interactions here is of Ruderman-Kittel-Kasuya-Yosida (RKKY) type [124, 85]. Qualitatively, the interactions are sinusoidal functions of distance between two Mn ions with quick oscillation.

In other words, the interactions can be either ferromagnetic (force spins to orient in same direction) or antiferromagnetic (force spins to orient in opposite directions) and they are very sensitive to the (random) distances between ions.

Under the light of the above facts, it is reasonable to model the interactions as random variables, which should take both positive and negative values. In the following, we review three models containing this idea in order of successively simplified forms. The last model, SSK, will be the main objective of this part of the thesis.

2.2 Three spin glass models

2.2.1 The Edwards-Anderson model.

The famous Edwards-Anderson (EA) model was proposed by Edwards and Anderson in 1975 [54, 53]. This model successfully recovered the cusp of susceptibility qualitatively and was regarded as a milestone in the history of spin glass. Consider spins that are arranged on a regular lattice, the Hamiltonian of the EA model is given by

$$\mathcal{H}_{\mathrm{EA}}(\sigma) := \sum_{i \sim j} J_{ij}\sigma_i\sigma_j, \tag{2.1}$$

where N denotes the total number of spins, $\sigma = (\sigma_1, \cdots, \sigma_N) \in \{\pm 1\}^N$ denotes the Ising-type spin variables, and J_{ij} stands for the interaction between spins at site i and site j. Here, $i \sim j$ means that the sum only runs over i and j that are neighbors in the lattice. That is, the EA model only considers nearest neighbor interactions. This is natural due to the fact that the interactions between ions decay rapidly with increasing distance.

The fundamental novelty of the EA model is that the interactions J_{ij}'s are modeled as i.i.d random variables. (J_{ij}'s are considered to be i.i.d standard Gaussian in [54], the specific distribution is not essential for the behavior of the system.) We may call J_{ij} disorders when we want to emphasize their randomness. Since the J_{ij} can

take either positive or negative values, there typically exists a loop on the lattice, where the product of J_{ij} are negative. Thus, even for a spin variable maximizing the Hamiltonian, there are some pairs of spins making negative contributions. This feature of spin glass is called frustration, which was first pointed out by Toulouse [115].

Denote the inverse temperature by

$$\beta = \frac{1}{T}. \tag{2.2}$$

Assuming that the interactions are constant on the time scale of the evolution of spin variable, Edwards and Anderson studied the following quenched free energy:

$$\mathcal{F}_{\mathrm{EA}} = \frac{1}{\beta N} \log \mathcal{Z}_{\mathrm{EA}}, \quad \mathcal{Z}_{\mathrm{EA}} = \sum_{\sigma \in \{\pm 1\}^N} e^{\beta \mathcal{H}_{\mathrm{EA}}(\sigma, J)}, \tag{2.3}$$

where $\mathcal{Z}_{\mathrm{EA}}$ denotes the partition function of the system. The associated Gibbs measure is given by

$$\langle \mathcal{O} \rangle = \frac{1}{Z_{\mathrm{EA}}} \sum_{\sigma \in \{\pm 1\}^N} \mathcal{O}(\sigma) e^{\beta \mathcal{H}(\sigma)} \quad \text{for any observable } \mathcal{O}. \tag{2.4}$$

The spin variable with higher Hamiltonian is assigned with larger probability. Note that the above Gibbs measure here is a random measure, it varies for different realizations of J_{ij}. Consequently, the $\mathcal{F}_{\mathrm{EA}}$ is a random variable depending on all interactions. However, when N is large, it is expected that the system with different realizations of J_{ij} share a similar behavior. Actually, we can divide the lattice into K small lattice with $1 \ll K \ll N$. Then, the free energy of the whole system is the average of the free energy of each sub-system plus the interactions at the interfaces between sub-systems. The later contribution is negligible when $N \to \infty$. Therefore, by the

central limit theorem,

$$\mathbb{E}[\mathcal{F}_{\mathrm{EA}}] - (\mathbb{E}[\mathcal{F}_{\mathrm{EA}}])^2 \to 0, \quad \text{as } N \to \infty. \tag{2.5}$$

Thus we call the free energy of EA model is self-averaging. The limiting free energy $F_{\mathrm{EA}} = \lim_{N\to\infty} \mathcal{F}_{\mathrm{EA}}$ is deterministic.

In order to get the limiting free energy, It is natural to consider mean of the free energy and let $N \to \infty$. Since J_{ij}'s are assumed to be i.i.d Gaussian, it is tempting to average the right-hand side of (2.3) directly. However, note that the formula of $\mathcal{F}_{\mathrm{EA}}$ involves a logarithm, which will be applied to $\mathcal{Z}_{\mathrm{EA}}$ before averaging over the J_{ij}. This fact together with the neighboring interaction makes the evaluation of $\mathbb{E}\,\mathcal{F}_{\mathrm{EA}}$ and precise analysis of the EA model extremely difficult. The existence of the spin glass phase for the EA model is only verified numerically at the moment [20, 4]. Nevertheless, features possessed by the EA model, such as frustration, i.i.d disorders, and the quenched, self-averaging observables, are the distinctive characterizations of spin glass models.

2.2.2 The Sherrington-Kirkpatrick model.

The Sherrington-Kirkpatrick (SK) model was proposed by Sherrington and Kirkpatrick in 1975 [107]. Let $J = (J_{ij})_{i,j=1}^N$ be a real symmetric matrix where $J_{ij}, 1 \leq i < j \leq N$, are independent random variables with mean 0 and variance 1. (The diagonal elements is not essential here, one can assume $J_{ii} = 0$ for $i = 1, \cdots, N$.) The Hamiltonian of 2-spin SK model is given by

$$\mathcal{H}_{\mathrm{SK}}(\sigma) = \frac{1}{2\sqrt{N}} \sum_{i,j=1}^{N} J_{ij}\sigma_j\sigma_j, \tag{2.6}$$

Here, the scaling factor $N^{-1/2}$ is picked such that the free energy per spin is of $O(1)$. Note that instead of a lattice, the underlying graph of the SK model is fully connected.

Thus it is regarded as a mean field version of EA model. The 2-spin model only takes the interactions between pairs of spins into account. One can also consider the p-spin generalization, whose Hamiltonian is given by

$$\mathcal{H}_{\mathrm{SK}} = \frac{1}{p! N^{(p-1)/2}} \sum_{i_1,\cdots,i_p=1}^{N} J_{i_1\cdots i_p}\sigma_{i_1}\cdots\sigma_{i_p}. \tag{2.7}$$

In particular, when $p = 1$, the vector $(J_1, \cdots, J_N)^T$ plays the role of the external field. More generally, people are interested in mixed p-spin SK model whose Hamiltonian is a linear combination of above Hamiltonians for different p.

Given $\mathcal{H}_{\mathrm{SK}}$ defined as above, the free energy of the SK model is defined in the same way as in (2.3). It turned out that the free energy $\mathcal{F}_{\mathrm{SK}}$ is again self-averaging, even through SK is a mean-field model. That is $\mathrm{Var}[\mathcal{F}_{\mathrm{SK}}] \to 0$ as $N \to 0$ and $F_{\mathrm{SK}} = \lim_{N\to\infty} \mathcal{F}_{\mathrm{SK}}$ is non-random. The authors of [107] considered 2-spin SK model where $(J_{ij})_{i<j}$ was assumed to be Gaussian. They calculated the F_{SK} using the replica method (see Appendix A.1). A critical temperature was found, below which the spin glass phase presented. Their result was correct in the high temperature regime, while exhibited a non-physical negative entropy in the low temperature regime. The failure was a consequence of the "replica symmetry" assumption [48, 27]. In 1980, the Parisi proposed the idea of full replica symmetry breaking and conjectured the famous Parisi formula [103] (see Appendix A.2). The Parisi formula described F_{SK} for general mixed p-spin model using a variational problem, and is valid for all temperature. A rigorous proof was given 20 years later by Talagrand [113] for the case where interaction random variables were Gaussian and only even p-spin interactions were considered. The Parisi formula for the case including the odd p-spin (Gaussian) interaction was later proved by Panchenko [100]. The universality of the F_{SK} for non-Gaussian interactions was proved by Carmona and Hu [32].

After getting the limiting free energys, which is the leading term of $\mathcal{F}_{\mathrm{SK}}$ as $N \to \infty$,

it is also interesting to consider the next order fluctuations. Aizenman, Lebowtiz, and Ruelle [1] showed that if the disorder random variables $(J_{ij})_{i<j}$ was assumed to be Gaussian, then

$$N(\mathcal{F}_{\mathrm{SK}} - F_{\mathrm{SK}}) \Rightarrow \frac{1}{\beta}\mathcal{N}(-\frac{1}{2}\alpha, \alpha), \tag{2.8}$$

where

$$F_{\mathrm{SK}} = \frac{\log 2}{\beta} + \frac{\beta}{4} \quad \text{and} \quad \alpha = -\frac{1}{2}\log(1-\beta^2) - \frac{1}{2}\beta^2. \tag{2.9}$$

The similar result held for the non-Gaussian disorder case with a modification of α [1]. The Gaussian fluctuations was also obtained in p-spin SK models for $p \geq 3$ [25]. For the mixed p-spin SK model with the presence of the external field, the limiting Gaussian distribution of the fluctuation was found in [37]. However, a limit theorem for the fluctuations in the low temperature regime still remains an open question.

2.2.3 The Spherical Sherrington-Kirkpatrick model.

The spin variable of the SK model is Ising-type, i.e., the configuration space consists of all vertices of a hypercube and is discrete. This stands as an obstacle in the analysis of the SK model. Aiming to introduce a model easier for analysis, Kosterlitz, Thouless and Jones proposed the spherical Sherrington-Kirkpartick (SSK) model [78]. The Hamiltonian of the SSK model is the same as the SK model (cf. (2.6)):

$$\mathcal{H}_{\mathrm{SSK}}(\sigma) = \frac{1}{2\sqrt{N}} \sum_{i,j=1}^{N} J_{ij}\sigma_i\sigma_j. \tag{2.10}$$

However, the spin variable is now supported on the sphere of radius $\sqrt{N}$:

$$\sigma \in S_{N-1} = \{\sigma \in \mathbb{R}^N | \, \|\sigma\|_2 = \sqrt{N}\}. \tag{2.11}$$

Thus, the free energy of the spherical model is given by

$$\mathcal{F}_{\mathrm{SSK}} := \frac{1}{N\beta} \log \mathcal{Z}_{\mathrm{SSK}}, \quad \mathcal{Z}_{\mathrm{SSK}} := \int_{\sigma \in S_{N-1}} e^{\beta \mathcal{H}_{\mathrm{SSK}}(\sigma)} \mathrm{d}w(\sigma), \tag{2.12}$$

where $\mathrm{d}w(\sigma)$ denotes the normalized uniform measure on S_{N-1}. Generally, people are also interested in the SSK model with the external field and mixed p-spin SSK models.

The pure 2-spin SSK was considered in [78]. Using a heuristic argument, the authors of [78] obtained critical temperature T_c as well as the non-random limiting free energy $F_{\mathrm{SSK}} = \lim_{N\to\infty} \mathcal{F}_{\mathrm{SSK}}$ for all temperature. For the mixed p-spin SSK model, the analogy of the Parisi formula was proposed by Crisanti and Sommers [47], which was again rigorously proved by Talagrand [112] for the Gaussian disorders. The universality of the limiting free energy for 2-spin SSK for general interactions are shown by [63].

We summarize the known results regarding the limiting law of the fluctuations of the free energy for SSK models. In the high temperature regime ($T > T_c$), the fluctuations are of order $\mathcal{O}(N^{-1})$ and converge to a Gaussian distribution [12]. The low temperature regime was considered in the same work. In this case, the authors of [12] showed that the fluctuations are of order $N^{-2/3}$ and established the Tracy-Widom limiting distribution. For pure p-spin model with $p \geq 3$, Subag and Zeitouni studied free energy in the zero temperature, which is $\max_{\sigma \in S_{N-1}} \mathcal{H}_{\mathrm{SSK}}(\sigma)$. They proved that the fluctuations of the free energy are of order N^{-1} and the limiting law was Gumbel [111]. This dichotomy between $p = 2$ and $p \geq 3$ cases can be seen from the number of the critical points of the system. For $p = 2$, the Hamiltonian $\mathcal{H}_{\mathsf{SSK}}(\sigma)$ posseses $2N$ critical points corresponding to the eigenvectors of the disorder $J = (J_{ij})_{i,j=1}^{N}$. And in the low temperature regime $T \leq T_c$, the free energy is governed by the top eigenvalue of J [12]. In contrast, when $p \geq 3$, the number of the critical points is exponential in

N and the free energy converges to the extreme of a Poisson point process [111].

The free energy of the 2-spin SSK model is the main objective of this thesis. From now on, we omit the subscription SSK, the quantities and observables are associated with the SSK model if without extra specification.

2.3 Outline of Chapter IV and V

As mentioned before, the limiting free energies of SK and SSK are given by the Parisi formula and Crisanti-Sommers formula. However, the fluctuations of the free energies, especially in the low temperature regime, are not well studied. Recently, using new progress in the random matrix theory, the authors of [13, 12, 14] obtain the fluctuations of the free energy of the 2-spin SSK model (and some variants) for arbitrary temperature. Their approach relies on the quadratic form of the Hamiltonian and thus only applies to the 2-spin case. In this part of the thesis, we apply this approach to the two other scenarios related to 2-spin SSK models.

2.3.1 Paramagnetic-Ferromagnetic phase transition for 2-spin SSK with Curie-Weiss interaction

The Curie-Weiss (CW) interaction is defined by

$$\mathcal{H}_{\mathrm{CW}}(\sigma) := \frac{\theta}{N}\sum_{i,j=1}^{N}\sigma_i\sigma_j = \frac{\theta}{N}\left(\sum_{i=1}^{N}\sigma_i\right)^2. \tag{2.13}$$

Here, θ is a real number which plays a role of the coupling constant. Adapt the notation $\mathbf{1} = (1, \cdots, 1)^T$, the Hamiltonian $\mathcal{H}_{\mathrm{CW}}(\sigma)$ is maximized when $\sigma = \pm\mathbf{1}$. Thus, the CW interaction forces spins to orient to the same direction and thus is a mean-field ferromagnetic interaction,

Chapter IV considers the Hamiltonian given by

$$\mathcal{H}(\sigma) = \mathcal{H}_{\mathrm{SSK}}(\sigma) + \mathcal{H}_{\mathrm{CW}}(\sigma) \tag{2.14}$$

for $\mathcal{H}_{\mathrm{SSK}}(\sigma)$ defined as in (2.10). The behavior of the spherical spin system with this Hamiltonian depends on the temperature and CW coupling strength. When θ is small, the system resembles the pure 2-spin SSK model: there is critical temperature $T_c = 1$ below which the spin glass phase presents (the high temperature regime is called paramagnetic regime). When θ is large, the spin variables is pulled towards $\pm\mathbf{1}$ and the system falls into the ferromagnetic phase. The limiting free energy of this spin system was obtained by Kosterlitz, Thouless, and Jones [78] through a non-rigorous argument. This limiting free energy, as a function of T and θ, was piecewise analytic. And a phase diagram (See Figure 2.1) was proposed accordingly. These results were rigorously proved in 2017 [13]. In addition, the authors of [13] established the limiting distributions of the fluctuations of the free energies in each regime. The order of the fluctuations are $N^{-2/3}$,N^{-1},$N^{-1/2}$ and the limiting distributions are Tracy-Widom, Gaussian, and Gaussian in the spin glass, paramagnetic regime, ferromagnetic regime, respectively.

Note that the orders and the laws of the fluctuations are different between the three phases, it is interesting to consider the phase transition and the near-critical behavior. The transition between the spin-glass and ferromagnetic regimes has been studied in [13]. Consider N-dependent θ as

$$\theta = 1 + wN^{-2/3}, \quad \text{for } w \in \mathbb{R}. \tag{2.15}$$

Then the fluctuation of the free energy is of order $N^{-2/3}$ and is governed by a family of random variables $\mathrm{TW}_{1,w}$ interpolating Gaussian distribution and Tracy-Widom distribution [22]. In Chapter IV, we focus on the transitional regime between the

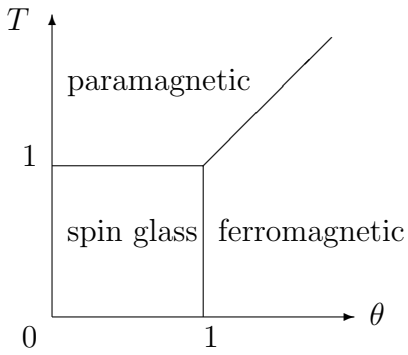


Figure 2.1: Phase diagram for SSK+CW model. Here, β is the inverse temperature and θ is the coupling constant.

ferromagnetic and paramagnetic regime and find the limiting law for the fluctuations. The transition between the paramagnetic and spin glass regimes is still open.

2.3.2 2-spin SSK with external field

In Chapter V, we consider the 2-spin spherical model, whose Hamiltonian is $\mathcal{H}_{\mathrm{SSK}}(\sigma)$ and plus a linear perturbation:

$$\mathcal{H}(\sigma) = \mathcal{H}_{\mathrm{SSK}}(\sigma) + h\sum_{i=1}^{N} J_i\sigma_i = \frac{1}{2\sqrt{N}}\sum_{i,j=1}^{N} J_{ij}\sigma_i\sigma_j + h\sum_{i=1}^{N} J_i\sigma_i. \tag{2.16}$$

In the context of the spin glass, the vector $(J_i)_{i=1}^N$ plays the role of the external field and $h \in \mathbb{R}$ stands for its strength. We focus on the case where $(J_i)_{i=1}^N$ is a standard Gaussian vector, and $(J_{ij})_{i,j=1}^N$ belonging to GOE is independent from $(J_i)_{i=1}^N$.

Like in the case of SSK perturbed by CW, we are interested in the interplay between the SSK Hamiltonian and the external field. Note that $\sum_{i=1}^N J_i\sigma_i$ is maximized for σ orienting to the same direction as $(J_i)_{i=1}^N$. The spin variables will be pulled towards $(J_i)_{i=1}^N$ when the external field presents. Actually, when $h > 0$, there is a uniform formula of the limiting free energy for all temperature [78, 37]. Furthermore, for any $h, T > 0$, we will show that the fluctuations of the free energies are of order

$N^{-1/2}$ and converge to a Gaussian distribution. These pieces of evidence indicate that phase transition only occurs when $h \to 0$.

When $h = 0$, there are paramagnetic regime ($T > 1$) and spin glass regime ($T < 1$). (See Figure 2.1 when $h = 0$ for the phase diagram for pure 2-spin SSK.) The fluctuations of the free energy are of order N^{-1} and $N^{-2/3}$ and the limiting distributions are Gaussian and Tracy-Widom respectively. Therefore, the phase transitions as $h \to 0$ should be considered separately for the cases $T > 1$ and $T < 1$. Our main contributions are the limiting theorems of fluctuations in these transitional regimes.

Due to the time limitation, the results in Chapter V are derived using heuristic calculations. We believe that these results can be proved rigorously. With light modification, our approach also applies to the case where the external field is deterministic. For the case where $J = (J_{ij})_{i,j=1}^N$ is a Wigner matrix, we expect similar results hold given that the eigenvectors of J are delocalized.

2.3.3 Method

Here we outline the applications of the method from [13, 12, 14] to the above two scenarios.

For the case of CW perturbation, we write the Hamiltonian (2.14) as

$$\mathcal{H}(\sigma) = \sigma^T \left(\frac{1}{2\sqrt{N}} J + \frac{\theta}{N} \mathbf{1}\mathbf{1}^T \right) \sigma, \tag{2.17}$$

Note that the above $\mathcal{H}(\sigma)$ is still a quadratic form with the random symmetric matrix $M = \frac{1}{2\sqrt{N}} J + \frac{\theta}{N} \mathbf{1}\mathbf{1}^T$. The matrices of this type are called spiked Wigner matrices, which is well-studied in the random matrix theory (see Section 3.5). On the other hand, Kosterlitz, Thouless, and Jones [78] found an identity formulating the partition function (see (2.12), which is a N-fold integral) as a single integral (see e.g. (4.42)), whose integrand depends on the eigenvalues of M. Using the eigenvalue rigidity

[77, 55], we apply the steepest descent to this single integral. As a result, the free energy can be formulated as a function of eigenvalues. In the transitional regime of our interest, it turns out that the fluctuations of the free energies are governed by a combination of the fluctuations of the leading eigenvalue and the global linear statistics. We show that the latter two fluctuations converges to a bivariate Gaussian distribution jointly.

The story in the case of external field perturbation is quite similar. The main difference is that the single integral formulation of the partition function now involves the eigenvectors of $(J_{ij})_{i,j=1}^N$. Since we assume that $(J_{ij})_{i,j=1}^N$ belongs to GOE, the eigenvector matrix is uniformly distributed on the orthogonal group and is independent of the eigenvalues. This fact together with the rigidity of the eigenvalues makes an application of the steepest descent possible. Consequently, we write the free energies as functions of eigenvalues and eigenvectors of $(J_{ij})_{i,j=1}^N$, which lead us to the limiting laws of the fluctuations. In particular, when temperature is low, we find that the fluctuations in the transitional regime (as $h \to 0$) depend on the top eigenvalues of $(J_{ij})_{i,j=1}^N$. Therefore, the corresponding limit theorem will be formulated using the GOE Airy kernel point process.

CHAPTER III

Random matrices preliminary

A random matrix is a matrix-valued random variable, that is, a matrix whose entries are random variables. The study of the random matrix theory started from the work of Wigner on the spectra of heavy atoms in physics in the 1950's [122]. After decades of evolution, the random matrix is now a broad subject with applications in various areas including number theory, statistical physics, statistics, electrical engineering and finance [28, 52, 2].

In this section, we collect some elementary results from the random matrix theory that is necessary to present the spin glass results in this thesis. The goal is to provide basic knowledge, the results being presented are not the strongest version with minimum conditions. And the more specific and technical discussions are left in the random matrix section in each chapter. Readers are referred to [86, 6, 2] for a comprehensive introduction of the random matrix theory.

3.1 Law of large number

We first introduce the real Wigner matrix, which is the basic model in the random matrix theory and main model considered in this thesis.

Definition III.1. Given two independent family of i.i.d centered, real-valued random

variables $\{J_{ij}\}_{1\le i<j\le N}$ and $\{J_{ii}\}_{i=1}^{N}$, such that $\mathbb{E}[J_{12}^2]=1$ and

$$\max\{\mathbb{E}[|J_{12}|^k],\mathbb{E}[|J_{11}|^k]\}<+\infty \quad \text{for all integer } k\ge 1. \tag{3.1}$$

Consider the (symmetric) $N\times N$ matrix M with entries

$$M_{ij}=M_{ji}=\frac{1}{\sqrt{N}}J_{ij} \quad \text{for all } 1\le i\le j\le N. \tag{3.2}$$

We call such a matrix M a Wigner matrix.

If J_{12} and J_{11} are both Gaussian random variables with variance 1 and 2 respectively, then the M belongs to the Gaussian orthogonal ensemble (GOE), which will be the matrix model considered in Chapter V.

Let $\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_N$ be the eigenvalues of M, we define the empirical distribution of the eigenvalues by

$$\mathrm{d}\sigma_M(x)=\frac{1}{N}\sum_{i=1}^{N}\delta(x-\lambda_i). \tag{3.3}$$

Note that the empirical distribution is a random measure depending on M. Wigner made the following key observation: while entries of M fluctuated wildly from sample to sample, the $\mathrm{d}\sigma_M$ concentrates around a deterministic measure. Actually, define the semicircle law

$$\mathrm{d}\sigma_{scl}(x)=\frac{\sqrt{4-x^2}}{2\pi}\mathbb{1}_{x\in[-2,2]}\mathrm{d}x, \tag{3.4}$$

we have the following theorem.

Theorem III.2 ([122]). *If $f(x)$ is a continuous and bounded function,*

$$\int f(x)\mathrm{d}\sigma_M(x)=\frac{1}{N}\sum_{i=1}^{N}f(\lambda_i)\to\int f(x)\mathrm{d}\sigma_{scl}(x) \quad \textit{as } N\to\infty. \tag{3.5}$$

The convergence above is in probability.

This theorem can be regarded as the start point of the random matrix theory. Note that the asymptotic behaviors of empirical distributions of Wigner matrices are independent of the distribution of the entries. The result of this kind is called universality.

3.2 Linear Statistics

The Theorem III.2 describes the leading term of the linear statistics $\frac{1}{N}\sum_{i=1}^{N} f(\lambda_i)$. Thus it is regarded as the analog of the law of large number for independent random variables in the random matrix theory. This result is the first step in the study of the behavior of the eigenvalue distribution of a random matrix ensemble. It is natural to consider the central limit theorem for fluctuations of linear eigenvalue statistics $\frac{1}{N}\sum_{i=1}^{N} f(\lambda_i)$.

Theorem III.3 (Central limit theorem; [9])**.** *Let f be an analytic function defined in a open neighborhood of $[-2,2]$, then*

$$\sum_{i=1}^{N} f(\lambda_i) - N\int f(x)\mathrm{d}\sigma_{scl}(x) \tag{3.6}$$

converges in distribution to a Gaussian random variable $\mathcal{N}(M(f), V(f))$ as $N \to \infty$. Here $M(f)$ and $V(f)$ are explicit functionals depending on the fourth moments J_{12}.

If λ_i are i.i.d random variables whose probability density function is given semi-circle law, then by classical central limit theorem, $\frac{1}{\sqrt{N}}(\sum_{i=1}^{N} f(\lambda_i) - N\int f(x)\mathrm{d}\sigma_{scl}(x))$ converges to a Gaussian distribution. Note that in Theorem III.3, the normalization factor is 1 instead of $N^{-1/2}$. This indicates that the eigenvalues are highly correlated. As what we will see in the next subsection, the eigenvalues of Wigner matrices are distributed in a way more rigid than the i.i.d random variables.

3.3 Eigenvalue rigidity

The law of large number (Theorem III.2) and the central limit theorem (Theorem III.3) consider the global behavior of the eigenvalues. It is also interesting to consider the local behavior.

By the semicircle law, the expectation of spacings between eigenvalues are of order N^{-1} in the bulk and $N^{-2/3}$ in the edge (due to the square root behavior of semicircle law near ± 2). On the other hand, define the classical locations of the eigenvalues $\widehat{\lambda}_k$ by

$$\int_{\widehat{\lambda}_k}^{2} \mathrm{d}\sigma_{scl}(x) = \frac{k - 1/2}{N}. \tag{3.7}$$

The locations $\widehat{\lambda}_k$ are basically N-quantile of the semicircle law. It turns out that the distances between eigenvalues and the corresponding classical locations are within scales slightly large than the expected spacing between eigenvalues.

Theorem III.4 (Rigidity of eigenvalues; [55])**.** *Let λ_i be eigenvalues of a Wigner matrix and $\widehat{\lambda}_i$ be the classical locations defined as above. Then for any $\epsilon, D > 0$, there exists a positive constant $N_0 \geq 1$ such that*

$$\max_i \mathbb{P}\left(|\lambda_i - \widehat{\lambda}_i| \geq (\min\{i, N - i + 1\})^{-1/3} N^{-2/3+\epsilon}\right) \leq N^{-D} \tag{3.8}$$

for large enough $N \geq N_0$.

In particular, the largest eigenvalue λ_1 is $O(N^{-2/3+\epsilon})$ away from 2 with high probability for any small $\epsilon > 0$. Similarly, the eigenvalues in the bulk are $O(N^{-1+\epsilon})$ away from their classical locations with high probability. Note that if λ_i were i.i.d random variables following the semicircle distribution, by an order statistics result, the λ_i would typically fluctuate around the classical location with scale $N^{-1/2}$. Therefore, the eigenvalue rigidity indicates a strong correlation between eigenvalues. Actually,

there is a strong repulsion between the eigenvalues. The neighboring eigenvalues avoid getting too close to each other, thus the eigenvalue is sticking near the classical location by this effect.

3.4 Edge Behaviors

As mentioned, the λ_i's are associated with the energy levels of heavy atoms system. This motivates the study of local distributions of the eigenvalues. In this section, we state the results regarding the joint distribution of top eigenvalues for GOE.

Consider the rightmost eigenvalue λ_1. By semicircle law and eigenvalue rigidity, we expect that λ_1 fluctuates around 2 in a scale of $N^{-2/3}$. Actually, $N^{2/3}(\lambda_1 - 2)$ converges in distribution to Tracy-Widom GOE distribution TW_{GOE} as $N \to \infty$ [116]. Let $\mathrm{Ai}(x)$ denote the Airy function and $q(s)$ be the solution of the Painléve II differential determined by asymptotics $q(s) \sim \mathrm{Ai}(s)$ as $s \to \infty$, the cumulative probability function of TW_{GOE} is given by

$$\lim_{N\to\infty} \mathbb{P}\left(N^{2/3}(\lambda_1 - 2) \le s\right) = \exp\left(-\frac{1}{2}\int_s^\infty q(x) + (x-s)q^2(x)\mathrm{d}x\right). \tag{3.9}$$

The above result is the $k = 1$ case of the following result considering the joint distribution of top k eigenvalues. Denote the rescaled eigenvalues by $a_i = N^{2/3}(\lambda_i - 2)$ for $i = 1, \cdots, N$. Recall the GOE Airy kernel point process $\{\alpha_i\}_{i=1}^\infty$, which is a Pfaffian point process with a kernel built out of $\mathrm{Ai}(s)$ (see [58] for an explicit formula of the kernel). In particular, α_1 follows Tracy-Widom GOE distribution, i.e., $a_1 \Rightarrow \alpha_1$ in distribution. Actually, this is true for top k eigenvalues for any fixed k.

Theorem III.5. *For any fixed $k \in \mathbb{N}$, we have that*

$$\{a_i\}_{i=1}^k \Rightarrow \{\alpha_i\}_{i=1}^k, \quad \text{as } N \to \infty. \tag{3.10}$$

Here the convergence is in distribution.

These results also apply to the top eigenvalues from Wigner matrices [109]. A similar discussion in the bulk of spectrum can be found in [86, 58].

3.5 Spiked Wigner matrices

In Chapter IV, we will study the spiked Wigner matrices M' given by

$$M' = M + \frac{\theta}{N}\mathbf{1}\mathbf{1}^T, \tag{3.11}$$

where M is a Wigner matrix, $\mathbf{1} = (1, \cdots, 1)^T$ and $\theta \in \mathbb{R}$ denotes the strength of the perturbation. As $-M$ is still a Wigner matrix, we only consider the case where $\theta > 0$. The spiked random matrices (also called as deformed random matrices or Wigner matrices with non-zero mean in the literature) were studied extensively in random matrix theory [10, 57, 106, 13], which is motivated by the study of the empirical covariance matrices with spiked covariance structure.

We first consider the global behavior of the spectrum for spiked Wigner matrices. It is not difficult to show that the eigenvalues of M' and eigenvalues of M are interlacing:

$$\lambda_N(M) \leq \lambda_N(M') \leq \lambda_{N-1}(M) \leq \lambda_{N-1}(M') \leq \cdots \leq \lambda_1(M) \leq \lambda_1(M'). \tag{3.12}$$

Therefore, same as the Wigner matrix, the empirical distribution of M' converges to the semicircle law. Furthermore, by the eigenvalue rigidity of Wigner matrix, $\lambda_i(M')$ is pinned around the classical location $\widehat{\lambda}_i$ for all $i \geq 2$.

On the other hand, it is interesting to consider $\lambda_1 = \lambda_1(M')$. It turned out that

the behavior of λ_1 depends on the strength of perturbation θ:

$$\begin{cases} N^{2/3}\left(\lambda_1 - 2\right) \Rightarrow \mathrm{TW}_{GOE} & \text{if } \theta < 1, \\ N^{1/2}\left(\lambda_1 - \theta - \frac{1}{\theta}\right) \Rightarrow \mathcal{N}(W_3(\theta^{-2} - \theta^{-4}), 2(1-\theta^{-2})) & \text{if } \theta > 1, \end{cases} \tag{3.13}$$

where $W_3 = \mathbb{E}[J_{12}^3]$. That is if θ is not large enough (i.e. $\theta < 1$), then the influence of the perturbation is negligible to the behavior of the top eigenvalue. For $\theta < 1$, the top eigenvalue is close to the 2 with order $O(N^{-2/3+\epsilon})$. In contrast, for $\theta > 1$, the top eigenvalue λ_1 fluctuates around $\theta + 1/\theta$, which is $O(1)$ away from 2. As N becomes large, the fluctuation is of order $N^{-1/2}$ and is governed by a Gaussian distribution with explicit mean and variance. The transitional regime is $\theta = 1 + wN^{-1/3}$. It was shown in [22] that $N^{2/3}(\lambda_1 - 2) \Rightarrow \mathrm{TW}_{1,w}$, where $\mathrm{TW}_{1,w}$ is a family of random variables interpolating TW_{GOE} and Gaussian distribution.

There are also results for spiked Wigner matrices regarding the eigenvalue rigidity and the central limit theorem for the linear statistics. Readers are referred to Chapter IV for further discussion.

CHAPTER IV

Ferromagnetic to paramagnetic transition in spherical spin glass with Curie-Weiss interaction

4.1 Introduction

We consider a disordered system defined by random Gibbs measures whose Hamiltonian is the sum of a spin glass Hamiltonian and a ferromagnetic Hamiltonian. Depending on the strength of the coupling constant and the temperature, the system may exhibit several phases in the large system limit. The chapter is concerned with the fluctuations of the free energy near the boundary between two phases known as ferromagnetic and paramagnetic regimes.

Consider the sum of the pure 2-spin spherical Sherrington-Kirkpatrick (SSK) Hamiltonian and the Curie-Weiss (CW) Hamiltonian. We call this sum the SSK+CW Hamiltonian. We denote the coupling constant by J and the inverse temperature by β. We consider the random Gibbs measure with the SSK+CW Hamiltonian. The focus of this chapter is on the free energy.

The limiting free energy was obtained non-rigorously by Kosterlitz, Thouless, and Jones [78] in 1976. When $J = 0$, this formula is the explicit evaluation of the Crisanti–Sommers formula [47] (which was proved rigorously by Talagrand [112]) in the case of the pure 2-spin SSK. The Crisanti–Sommers formula is the spherical version of the

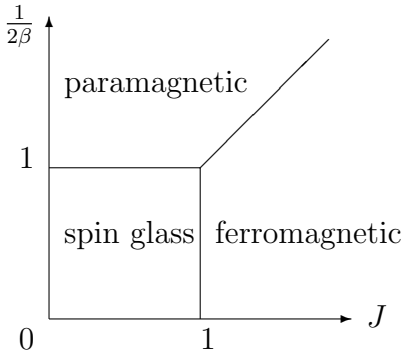


Figure 4.1: Phase diagram for SSK+CW model. Here, β is the inverse temperature and J is the coupling constant.

Parisi formula [103, 113]. The formula of Kosterlitz, Thouless, and Jones shows a two-dimensional phase transition: see Figure 4.1. The three regimes are determined by the condition that $\max\{1, \frac{1}{2\beta}, J\}$ is equal to 1 (spin glass regime), $\frac{1}{2\beta}$ (paramagnetic regime) or J (ferromagnetic regime). The limiting free energy is analytic with respect to both β and J in each regime, but not on the boundary.

Recently, the authors of [13] showed that the result of Kosterlitz, Thouless, and Jones is rigorous. Furthermore, the authors also evaluated the distribution of the fluctuations of the free energy in each regime. (The case when $J = 0$ was obtained earlier in [12].) The order of the fluctuations are $N^{-2/3}, N^{-1}, N^{-1/2}$ and the limiting distributions are Tracy-Widom, Gaussian, and Gaussian in the spin glass, paramagnetic regime, ferromagnetic regime, respectively. In the same paper, the transition between the spin glass regime and the ferromagnetic regime was also studied. However, the other two transitions and the triple point were left open. The goal of this chapter is to describe the transition between the paramagnetic regime and and the ferromagnetic regime.

Another system which combines a spin glass and a ferromagnetic model is the SSK with an external field. The difference between the CW Hamiltonian and an external field is that one is a quadratic function and the other is a linear function

of the spin variables. These two models are related; see [36] for a one-sided inequality. For the spin glass with external field, the fluctuations of the free energy were computed recently in [37, 38] when the coupling constant is positive (for both SSK and SK (Sherrington-Kirkpatrick) cases with general spin interactions). However, the transitions are not obtained except for certain large deviation results [61, 50]. One of the interests of the SSK+CW model is that it is an easier model which can be analyzed in detail in the transitional regimes.

4.1.1 Model

Let

$$S_{N-1} = \{\sigma = (\sigma_1, \cdots, \sigma_N) \in \mathbb{R}^N : \sigma_1^2 + \cdots + \sigma_N^2 = N\} \tag{4.1}$$

be a sphere in $\mathbb{R}^N$ of radius $\sqrt{N}$. Define the SSK+CW Hamiltonian by

$$H_N(\sigma) = H_N^{\mathrm{SSK}}(\sigma) + H_N^{\mathrm{CW}}(\sigma), \qquad \sigma \in S_{N-1} \tag{4.2}$$

where

$$H_N^{\mathrm{SSK}}(\sigma) = \frac{1}{\sqrt{N}} \sum_{i,j=1}^{N} A_{ij}\sigma_i\sigma_j, \qquad H_N^{\mathrm{CW}}(\sigma) = \frac{J}{N} \sum_{i,j=1}^{N} \sigma_i\sigma_j = \frac{J}{N} \left(\sum_{i=1}^{N} \sigma_i \right)^2. \tag{4.3}$$

Here J is the coupling constant. The random coefficients A_{ij} satisfy $A_{ij} = A_{ji}$ and A_{ij}, $i \leq j$, are independent centered random variables. We call A_{ij} disorder variables. The precise conditions are given in Definition IV.1 below. Note that as a function of σ, $H_N^{\mathrm{CW}}(\sigma)$ is large when the coordinates of σ have same sign. On the other hand, the maximizers σ of $H_N^{\mathrm{SSK}}(\sigma)$ depend highly on $\{A_{ij}\}$.

With $\beta > 0$ representing the inverse temperature, the free energy and the partition

function are defined by

$$F_N = \frac{1}{N}\log Z_N, \qquad Z_N = \int_{S_{N-1}} e^{\beta H_N(\sigma)} \mathrm{d}\omega_N(\sigma) \tag{4.4}$$

where ω_N is the normalized uniform measure on S_{N-1}. Note that F_N and Z_N are random variables since they depend on the disorder variables A_{ij}. The free energy and the partition function depend on the parameters β and J,

$$F_N = F_N(\beta, J), \qquad Z_N = Z_N(\beta, J). \tag{4.5}$$

Since the Curie-Weiss Hamiltonian is a quadratic function of the spin variable, we can write the SSK+CW Hamiltonian as $H_N(\sigma) = \sum_{i,j=1}^N M_{ij}\sigma_i\sigma_j$ where $M_{ij} = \frac{1}{\sqrt{N}}A_{ij} + \frac{J}{N}$ are non-centered random variables. In terms of matrix notations,

$$H_N(\sigma) = \sigma^T M \sigma, \qquad M = \frac{1}{\sqrt{N}}A + \frac{J}{N}\mathbf{1}\mathbf{1}^T \tag{4.6}$$

with $A = (A_{ij})_{1\le i,j\le N}$, $\mathbf{1} = (1, \cdots, 1)^T$, $M = (M_{ij})_{1\le i,j\le N}$, and $\sigma = (\sigma_1, \cdots, \sigma_N)^T$. The non-centered random symmetric matrix M is an example of a real Wigner matrix perturbed by a deterministic finite rank matrix. Such matrices are often called spiked random matrices. We will use the eigenvalues of spiked random matrices in our analysis of the free energy.

We assume the following conditions on the disorder variables.

Definition IV.1 (Assumptions on disorder variables)**.** Let A_{ij}, $i \le j$, be independent real random variables satisfying the following conditions:

- All moments of A_{ij} are finite and $\mathbb{E}[A_{ij}] = 0$ for all $i \le j$.
- For all $i < j$, $\mathbb{E}[A_{ij}^2] = 1$, $\mathbb{E}[A_{ij}^3] = W_3$, and $\mathbb{E}[A_{ij}^4] = W_4$ for some constants $W_3 \in \mathbb{R}$ and $W_4 \ge 0$.

- For all i, $\mathbb{E}[A_{ii}^2] = w_2$ for a constant $w_2 \geq 0$.

Set $A_{ij} = A_{ji}$ for $i > j$. Let $A = (A_{ij})_{i,j=1}^N$ and we call it a Wigner matrix (of zero mean).

Definition IV.2 (Eigenvalues of non-zero mean Wigner matrices)**.** Let M be the $N \times N$ symmetric matrix defined in (4.6). We call it a Wigner matrix of non-zero mean [1]. Its eigenvalues are denoted by

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N. \tag{4.7}$$

We introduce the following terminology.

Definition IV.3 (High probability event)**.** We say that an N-dependent event Ω_N holds with high probability if, for any given $D > 0$, there exists $N_0 > 0$ such that

$$\mathbb{P}(\Omega_N^c) \leq N^{-D}$$

for any $N \geq N_0$.

4.1.2 Previous results in each regime

We review the results on the fluctuations in each regime obtained in [13]. We state two types of results: one in terms of the eigenvalues of M and the other in terms of limiting distributions.

Set

$$\tilde{J} := \max\{J, 1\}. \tag{4.8}$$

[1] In [13], the authors consider the case when the diagonal entries of M have mean $\frac{J'}{N}$ and the off-diagonal entries have mean $\frac{J}{N}$ where J and J' are allowed to be different. However, in this case, $M = \frac{1}{\sqrt{N}} + \frac{J}{N}\mathbf{1}\mathbf{1}^T + \frac{J'-J}{N}I$ where I is the identity matrix. This only shifts all eigenvalues by a deterministic small number. As we will see in Remark IV.7, it is not more general than the case with $J' = J$.

It was shown in [13] that the following holds with high probability. In both ferromagnetic and the spin glass regimes (given by $\tilde{J} > \frac{1}{2\beta}$), with any $\epsilon > 0$,

$$F_N = \tilde{F}_N + \left(\beta - \frac{1}{2\tilde{J}}\right)\left(\lambda_1 - \tilde{J} - \frac{1}{\tilde{J}}\right) + O(N^{-1+\epsilon}). \tag{4.9}$$

In the paramagnetic regime (given by $\tilde{J} < \frac{1}{2\beta}$),

$$F_N = \tilde{F}_N - \frac{1}{2N}\sum_{i=1}^{N} \log\left(2\beta + \frac{1}{2\beta} - \lambda_i\right) + O(N^{-2+\epsilon}). \tag{4.10}$$

Here, $\tilde{F}_N$ is a deterministic function of N, β, J. The above results show that the fluctuations of F_N are determined, to the leading order, by the top eigenvalue λ_1 in the ferromagnetic and spin glass regimes, while they are determined by all eigenvalues in the paramagnetic regime.

A limit theorem for F_N follows if we use limit theorems for the eigenvalues of random matrices. The relevant random matrices are Wigner matrices of non-zero mean in (4.6). For such random matrices, the following is known [105, 29] (see [10] for complex matrices):

$$\begin{cases} N^{2/3}\left(\lambda_1 - 2\right) \Rightarrow \mathrm{TW}_1 & \text{if } J < 1, \\ N^{1/2}\left(\lambda_1 - J - \frac{1}{J}\right) \Rightarrow \mathcal{N}(W_3(J^{-2} - J^{-4}), 2(1 - J^{-2})) & \text{if } J > 1, \end{cases} \tag{4.11}$$

where the convergences are in distribution. Here TW_1 denotes the GOE Tracy-Widom distribution and $\mathcal{N}(a, b)$ denotes the Gaussian distribution of mean a and variance b. The dichotomy is due to the effect of the non-zero mean; if J is not large enough (i.e. $J < 1$), then the influence of the non-zero mean is negligible to contribute to the fluctuations of the top eigenvalue. For $J < 1$, the top eigenvalue is close to the second eigenvalue with order $O(N^{-2/3+\epsilon})$. But for $J > 1$, the difference of the top eigenvalue and the second eigenvalue is of order $O(1)$.

On the other hand, the following is also known (see Theorem 1.6 of [13]): if a function φ is smooth in an open interval containing the interval $[-2, \tilde{J} + \tilde{J}^{-1}]$, then

$$\sum_{i=1}^{N} \varphi(\lambda_i) - N \int_{-2}^{2} \varphi(x) \mathrm{d}\sigma_{scl}(x) \Rightarrow \mathcal{N}(f, a), \qquad \mathrm{d}\sigma_{scl}(x) := \frac{\sqrt{4 - x^2}}{2\pi} \mathrm{d}x, \tag{4.12}$$

for some explicit constants f, a. This result is applicable to the paramagnetic regime.

Together, we have the following asymptotic results, which are minor corrections of Theorem 1.4 of [13]:

(i) (Spin glass regime) If $\beta > \frac{1}{2}$ and $J < 1$, then

$$\frac{1}{\beta - \frac{1}{2}} N^{2/3} \left(F_N - F\right) \Rightarrow \mathrm{TW}_1 \,. \tag{4.13}$$

(ii) (Paramagnetic regime) If $\beta < \frac{1}{2}$ and $\beta < \frac{1}{2J}$, then

$$N \left(F_N - F\right) \Rightarrow \mathcal{N} \left(f_1, \alpha_1\right). \tag{4.14}$$

(iii) (Ferromagnetic regime) If $J > 1$ and $\beta > \frac{1}{2J}$, then

$$\sqrt{N} \left(F_N - F\right) \Rightarrow \mathcal{N} \left(f_2', \alpha_2'\right). \tag{4.15}$$

for some deterministic function $F = F(\beta, J)$ and some explicit constants f_1, α_1, f_2' and α_2' depending on β and J.

4.1.3 Results

We state the results on the transition between the paramagnetic regime and the ferromagnetic regime. The boundary between these two regimes is given by the equation $\frac{1}{2\beta} = J$ with $J > 1$. In the transitional regime, the correct scaling turns out

to be the following: let $J > 1$ be fixed and let $\beta = \beta_N$ be given by

$$2\beta = \frac{1}{J} + \frac{B}{\sqrt{N}} \tag{4.16}$$

with fixed $B \in \mathbb{R}$. The following is the first main result of this chapter. This relates the free energy with the eigenvalues of M.

Theorem IV.4. *Let β be given by* (4.16)*. Then, for every $0 < \epsilon < \frac{1}{8}$,*

$$F_N = \tilde{F}_N - \frac{1}{2N}\sum_{i=2}^{N} g(\lambda_i) + \frac{1}{N}Q(\chi_N) + O(N^{-3/2+4\epsilon}), \qquad \chi_N := \sqrt{N}(\lambda_1 - J + J^{-1}), \tag{4.17}$$

with high probability as $N \to \infty$, where

$$\tilde{F}_N = \beta(J+J^{-1}) - \frac{1}{2} - \frac{1}{2}\log(2\beta) + \frac{1}{N}\left(\frac{1}{4}\log N + \log\frac{\beta}{\sqrt{\pi}}\right), \quad g(z) := \log(J+J^{-1}-z). \tag{4.18}$$

Also,

$$Q(x) = \frac{s(x)}{2(s(x)-x)} - \frac{s(x)^2}{4(J^2-1)} + \frac{\log(s(x)-x)}{2} + \log \mathbf{I}\left(\frac{(s(x)-x)^2}{J^2-1}\right) \tag{4.19}$$

with

$$s(x) = \frac{x - B(J^2-1) + \sqrt{(x+B(J^2-1))^2 + 4(J^2-1)}}{2} \tag{4.20}$$

and

$$\mathbf{I}(\alpha) = \int_{-\infty}^{\infty} \frac{e^{-\frac{\alpha}{4}t^2 + \frac{\mathrm{i}t}{2}}}{\sqrt{1+\mathrm{i}t}}\mathrm{d}t, \tag{4.21}$$

where the square root denotes the principal branch.

The formula (4.17) shows a combined contribution from $\lambda_2, \cdots, \lambda_N$ and a distinguished contribution from λ_1. Compare the formula with (4.9) and (4.10).

Now we state a result analogous to (4.14) and (4.15). This follows if we have

limit theorems for $Q(\chi_N)$ and $\sum_{i=2}^N g(\lambda_i)$. From the second part of (4.11), $Q(\chi_N)$ converges to an explicit function of a Gaussian random variable. On the other hand, $\sum_{i=2}^N g(\lambda_i)$ is different from $\sum_{i=1}^N g(\lambda_i)$ by one term. It is not difficult to show that removing one term does not affect the fluctuations much and the fluctuations are still given by a Gaussian random variable similar to (4.12); see Theorem IV.6 in the next section. In random matrix theory, these sums are known as partial linear statistic and linear statistic, respectively. The main technical part of this chapter is to evaluate the joint distribution of $Q(\chi_N)$ and $\sum_{i=2}^N g(\lambda_i)$. We show that jointly they converge in distribution to a bivariate Gaussian variable with an explicit covariance. See the next section for the precise statement. These results are interesting on their own in random matrix theory. Putting together, we obtain the following result.

Theorem IV.5. *We have*

$$N\left(F_N - \frac{1}{4J^2} - \frac{B}{2J\sqrt{N}} - \frac{\log N}{4N} - \frac{B^2J^2}{4N}\right) \Rightarrow \mathcal{G}_1 + Q(\mathcal{G}_2) \tag{4.22}$$

in distribution as $N \to \infty$ where $\mathcal{G}_1$ and $\mathcal{G}_2$ are bivariate Gaussian random variables with

$$\mathbb{E}[\mathcal{G}_1] = \frac{1}{4}\log(J^2-1) + \frac{w_2-2}{4J^2} + \frac{W_4-3}{8J^4} + \log\frac{1}{2\sqrt{\pi}J}, \tag{4.23}$$

$$\mathrm{Var}[\mathcal{G}_1] = -\frac{1}{2}\log(1-J^{-2}) + \frac{w_2-2}{4J^2} + \frac{W_4-3}{8J^4}, \tag{4.24}$$

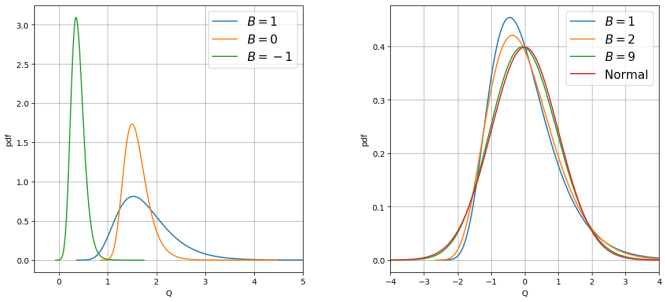
$$\mathbb{E}[\mathcal{G}_2] = W_3(J^{-2}-J^{-4}), \qquad \mathrm{Var}[\mathcal{G}_2] = 2(1-J^{-2}), \tag{4.25}$$

and

$$\mathrm{Cov}(\mathcal{G}_1, \mathcal{G}_2) = \frac{W_3(J^{-2}-J^{-4})}{2}. \tag{4.26}$$

Note that $\mathcal{G}_1$ and $\mathcal{G}_2$ do not depend on B. The function Q is defined in (4.19).

Note that if the third moment W_3 of A_{ij} with $i \neq j$ is zero, then $\mathcal{G}_1$ and $\mathcal{G}_2$ are



(a) pdf of $Q(\mathcal{G}_2)$ (b) pdf of normalized $Q(\mathcal{G}_2)$

Figure 4.2: (a) Probability density function of $Q(\mathcal{G}_2)$ for $B = -1, 0, 1$, (b) Probability density function of normalized $Q(\mathcal{G}_2)$ resembles a Gaussian density as $B \to +\infty$.

independent Gaussians.

The above result is consistent with the results on ferromagnetic and paramgnatic regimes if we let formally $B \to +\infty$ and $B \to -\infty$, respectively. One can show that when $B \to +\infty$, $Q(\mathcal{G}_2)$ dominates $\mathcal{G}_1$. Furthermore, while $Q(\mathcal{G}_2)$ is not Gaussian, upon proper normalization, it converges to a Gaussian as $B \to +\infty$. See Figure 4.2. On the other hand, when $B \to -\infty$, the leading two terms of $Q(\mathcal{G}_2)$ are constants and the random part is smaller than $\mathcal{G}_1$. See Section 4.6 for details.

Let us comment on the other transitions in the phase digram in Figure 4.1. As mentioned before, the transition between the spin glass and ferromagnetic regimes was discussed in [13]. Note that (4.9) is valid in both regimes. It was shown that if we let $\beta > 1/2$ be fixed and consider N-dependent $J = 1 + wN^{-1/3}$, then for each $w \in \mathbb{R}$, (4.9) still holds. Now, for such J, it was shown in [22] that $N^{2/3}(\lambda_1 - 2) \Rightarrow \mathrm{TW}_{1,w}$ where $\mathrm{TW}_{1,w}$ is a one-parameter family of random variables interpolating TW and Gaussian distributions. Hence, we obtain the fluctuations for the transitional regime.

On the other hand, the transition between the spin glass and paramagnetic regimes is an open question. By matching the fluctuation scales in both regimes, we expect that the critical scale is $\beta = \frac{1}{2} + O(\frac{\sqrt{\log N}}{N^{1/3}})$.

4.1.4 Organization

The rest of the chapter is organized as follows. In Section 4.2, we first state new results on random matrices. They are given in Theorem IV.6 (partial linear statistics) and Theorem IV.8 (joint convergence). Using them, we derive Theorem IV.5 from Theorem IV.4. In Section 4.3, we prove Theorem IV.4. In the next two sections, we prove the random matrix results stated in Section 4.2; Theorem IV.6 in Section 4.4 and Theorem IV.8 in Section 4.5. In Section 4.6, we show that Theorem IV.5 is consistent with the previous results on ferromagnetic and paramagnetic regimes.

4.2 Results on Wigner matrices with non-zero mean

In order to prove Theorem IV.5 from Theorem IV.4, we need some new results on random matrices. We need (i) a limit theorem for partial linear statistics $\sum_{i=2}^{N} g(\lambda_i)$ and (ii) a joint convergence of the large eigenvalue and partial linear statistics. These results are interesting on their own in random matrix theory. We state them here and prove them in Section 4.4 and Section 4.5 below. Using these results, we prove Theorem IV.5 in Subsection 4.2.3.

Recall that the $N \times N$ symmetric matrix M is given by $M = \frac{1}{\sqrt{N}} A + \frac{J}{N} \mathbf{1}\mathbf{1}^T$ where $A = (A_{ij})$ is a symmetric matrix with independent entries for $i \leq j$ satisfying the conditions given in Definition IV.1 and $\mathbf{1} = (1, \cdots, 1)^T$. The matrix M is called a Wigner matrix with a non-zero mean $\frac{J}{N}$. Recall that we assume

$$J > 1. \tag{4.27}$$

The eigenvalues of M are denoted by $\lambda_1 \geq \cdots \geq \lambda_N$.

It is known that λ_1 is close to $J + J^{-1}$ with high probability and $\lambda_2, \cdots, \lambda_N$ are in a neighborhood of $[-2, 2]$ with high probability. See Lemma IV.10 below for the precise statement.

4.2.1 Partial linear statistics

A linear statistic is the sum of a function of the eigenvalues. The fluctuations of linear statistics for Wigner matrices and other random matrix ensembles are of central interest in the random matrix theory; see, for example, [76, 9, 83]. For Wigner matrices with non-zero mean, the following result was obtained in Theorem 1.6 and Remark 1.7 of [13]. Set

$$\hat{J} = J + J^{-1}. \tag{4.28}$$

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a function which is analytic in an open neighborhood of $[-2, \hat{J}]$ and has compact support. Then, as $N \to \infty$, the random variable

$$\mathcal{N}_N(\varphi) := \sum_{i=1}^{N} \varphi(\lambda_i) - N \int_{-2}^{2} \varphi(x) \mathrm{d}\sigma_{scl}(x) \Rightarrow \mathcal{N}(M(\varphi), V(\varphi)) \tag{4.29}$$

where

$$\begin{aligned} M(\varphi) =& \frac{1}{4}(\varphi(2) + \varphi(-2)) - \frac{3}{2}\tau_0(\varphi) - J^{-1}\tau_1(\varphi) + (w_2 - 2)\tau_2(\varphi) \\ &+ (W_4 - 3)\tau_4(\varphi) + \varphi(\hat{J}) - \sum_{\ell=2}^{\infty} J^{-\ell}\tau_\ell(\varphi), \\ V(\varphi) =& (w_2 - 2)\tau_1(\varphi)^2 + (W_4 - 3)\tau_2(\varphi)^2 + 2\sum_{\ell=1}^{\infty} \ell\tau_\ell(\varphi)^2. \end{aligned} \tag{4.30}$$

Here, $W_4 = \mathbb{E}[A_{12}^4]$, $w_2 = \mathbb{E}[A_{11}^2]$, and

$$\tau_\ell(\varphi) = \frac{1}{\pi} \int_{-2}^{2} \varphi(x) \frac{T_\ell(x/2)}{\sqrt{4 - x^2}} \mathrm{d}x = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(2\cos(\theta)) \cos(\ell\theta) \mathrm{d}\theta, \tag{4.31}$$

where $T_\ell(t)$ are the Chebyshev polynomials of the first kind.

We are interested in a partial linear statistic, $\sum_{i=2}^{N} \varphi(\lambda_i)$. See [16, 97] for other types of partial linear statistics. The partial linear static $\sum_{i=2}^{N} \varphi(\lambda_i)$ is the linear

statistic minus one term $\varphi(\lambda_1)$. Since $\lambda_1 \to \hat{J}$ in probability (see the second part of (4.11)), by (4.29), Slutsky's theorem [62] implies that

$$\sum_{i=2}^{N} \varphi(\lambda_i) - N \int_{-2}^{2} \varphi(x) \mathrm{d}\sigma_{scl}(x) \Rightarrow \mathcal{N}(M(\varphi) - \varphi(\hat{J}), V(\varphi)).$$

Since this follows from (4.29), this is true assuming that φ is analytic in an open neighborhood of $[-2, \hat{J}]$. However, we are interested in the test function $\varphi(x) = g(x) = \log(\hat{J} - x)$ (see (4.17)). Since this function is not analytic at $x = \hat{J}$, the above simple argument does not apply. Nonetheless, if we adapt the proof of (4.29), one can show that it is enough to assume that the test function is analytic in a neighborhood of the interval $[-2, 2]$, not of $[-2, \hat{J}]$.

Theorem IV.6. *Let $J > 1$. Then for every test function φ which is analytic in a neighborhood of $[-2, 2]$,*

$$\mathcal{N}_N^{(2)}(\varphi) := \sum_{i=2}^{N} \varphi(\lambda_i) - N \int_{-2}^{2} \varphi(x) \mathrm{d}\sigma_{scl}(x) \Rightarrow \mathcal{N}(M^{(2)}(\varphi), V^{(2)}(\varphi)) \tag{4.32}$$

as $N \to \infty$ with

$$\begin{aligned} M^{(2)}(\varphi) =& \frac{1}{4}(\varphi(2) + \varphi(-2)) - \frac{3}{2}\tau_0(\varphi) - J^{-1}\tau_1(\varphi) + (w_2 - 2)\tau_2(\varphi) \\ &+ (W_4 - 3)\tau_4(\varphi) - \sum_{\ell=2}^{\infty} J^{-\ell}\tau_\ell(\varphi), \end{aligned} \tag{4.33}$$

and $V^{(2)}(\varphi) = V(\varphi)$ where $V(\varphi)$ is defined in (4.30).

Note that

$$M^{(2)}(\varphi) = M(\varphi) - \varphi(\hat{J}) \tag{4.34}$$

for φ analytic in a neighborhood of $[-2, \hat{J}]$.

Remark IV.7. We comment on a case when the test function depends on N. Consider

the function φ_N defined by

$$\varphi_N(x) = \varphi(x) + \frac{\phi(x)}{N} + O(N^{-2})$$

uniformly for x in a neighborhood of $[-2, 2]$ for analytic functions φ and ϕ. Define the corresponding linear statistic $\mathcal{N}_N^{(2)}(\varphi_N) = \sum_{i=2}^N \varphi_N(\lambda_i) - N\int_{-2}^2 \varphi_N(x)\mathrm{d}\sigma_{scl}(x)$, then

$$\begin{aligned}\mathcal{N}_N^{(2)}(\varphi_N) =& \sum_{i=2}^N \varphi_N(\lambda_i) - N\int_{-2}^2 \varphi_N(x)\mathrm{d}\sigma_{scl}(x) \\ =& \mathcal{N}_N^{(2)}(\varphi) + \frac{1}{N}\left(\sum_{i=2}^N \phi(\lambda_i) - N\int \phi(x)\mathrm{d}\sigma_{scl}(x)\right) + O(\frac{1}{N}).\end{aligned} \tag{4.35}$$

By Theorem IV.6, the second order term converges to zero in probability. Thus, $\mathcal{N}_N^{(2)}(\varphi_N)$ and $\mathcal{N}_N^{(2)}(\varphi)$ converge to the same Gaussian distribution. The same argument also applies to full linear statistics; this is used in Remark IV.24 below. Now, the claim in footnote[1] (see Page 27) is verified by noting that $\varphi(x + \frac{J'-J}{N}) = \varphi(x) + \frac{\varphi'(x)(J'-J)}{N} + O(N^{-2})$.

4.2.2 Joint convergence of the largest eigenvalue and linear statistics

By Theorem IV.6 and the second part of (4.11), the partial linear statistic and the largest eigenvalue each converge to Gaussian distributions individually. The following theorem shows that they converge jointly to a bivariate Gaussian with an explicit covariance.

Theorem IV.8. *Let $J > 1$. Then for $\varphi(x)$ which is analytic in a neighborhood of $[-2, 2]$, $\mathcal{N}_N^{(2)}(\varphi) := \sum_{i=2}^N \varphi(\lambda_i) - N\int_{-2}^2 \varphi(x)\mathrm{d}\sigma_{scl}(x)$ and $\chi_N := \sqrt{N}(\lambda_1 - \hat{J})$ converges jointly in distribution to a bivariate Gaussian variable with mean*

$$(M^{(2)}(\varphi), W_3(J^{-2} - J^{-4})) \tag{4.36}$$

and covariance

$$\begin{pmatrix} V^{(2)}(\varphi) & 2W_3\tau_2(\varphi)(1-J^{-2}) \\ 2W_3\tau_2(\varphi)(1-J^{-2}) & 2(1-J^{-2}) \end{pmatrix}. \tag{4.37}$$

The proof of this theorem, given in Section 4.5, is the main technical part of this chapter. We prove the theorem first for the Gaussian case, and then use an interpolation argument.

4.2.3 Proof of Theorem IV.5

We now derive Theorem IV.5 from Theorem IV.4 using the results on the eigenvalues stated in the previous two subsections. The term $Q(\chi_N)$ converges to $Q(\mathcal{G}_2)$ in distribution from Theorem IV.8. Consider the rest. It was shown in [12, (A.5)] that for $g(z) = \log(J + J^{-1} - z)$,

$$\int g(z)\mathrm{d}\sigma_{scl}(x) = \frac{1}{2J^2} + \log J. \tag{4.38}$$

Inserting $2\beta = J^{-1} + BN^{-1/2}$ and using the Taylor expansion $\log(1 + \frac{BJ}{\sqrt{N}}) = \frac{BJ}{\sqrt{N}} - \frac{B^2J^2}{2N} + O(N^{-3/2})$,

$$\tilde{F}_N - \frac{1}{2}\int g(z)\mathrm{d}\sigma_{scl}(x) = \frac{1}{4J^2} + \frac{B}{2J\sqrt{N}} + \frac{\log N}{4N} + \frac{1}{N}\left[\frac{B^2J^2}{4} + \log\frac{1}{2\sqrt{\pi}J}\right] + O(N^{-3/2}). \tag{4.39}$$

We can evaluate $M^{(2)}(g)$ defined as in (4.33) using [13, (2.7)] which evaluated the $M(h)$ with $h(x) = \log(2\beta + \frac{1}{2\beta} - x)$: (note that $J' = J$ here)

$$M^{(2)}(g) = \lim_{\beta \to \frac{1}{2J}} \left(M(h) - \log(2\beta + \frac{1}{2\beta} - J - J^{-1}) \right) = -\frac{1}{2}\log(J^2 - 1) - \frac{w_2 - 2}{2J^2} - \frac{W_4 - 3}{4J^4}. \tag{4.40}$$

The variance $V^{(2)}(g) = V(g)$, which is independent of J, is given by 4 times [12, (3.13)] if we replace 2β by J^{-1}:

$$V^{(2)}(g) = -2\log(1 - J^{-2}) + \frac{1}{J^2}(w_2 - 2) + \frac{1}{2J^4}(W_4 - 3). \tag{4.41}$$

For the covariance term, we have $\tau_2(g) = -\frac{1}{2J^2}$ from [12, (A.17)]. Hence, from Theorem IV.6 and IV.8, we obtain the result.

4.3 Proof of Theorem IV.4

The proof follows the steps for the proof of the Theorem 1.5 of [13] for paramagnetic and ferromagnetic regimes with necessary adjustments. The analysis is based on applying a method of steepest-descent to a random integral. The location of the critical point is important. In the transitional regime, the critical point is close to the largest eigenvalue but not as close as the ferromagnetic case. On the other hand, the critical point is away from the largest eigenvalue in the paramagnetic case. See Subsection 4.3.2 below for details.

4.3.1 Preliminaries

The following formula is a simple result in [78].

Lemma IV.9 ([78]; also Lemma 1.3 of [12])**.** *Let M be a real $N \times N$ symmetric matrix with eigenvalue $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. Then for fixed $\beta > 0$,*

$$\int_{S_{N-1}} e^{\beta\sigma^T M\sigma} \mathrm{d}\omega_N(\sigma) = C_N \int_{\gamma-i\infty}^{\gamma+i\infty} e^{\frac{N}{2}G(z)} \mathrm{d}z, \qquad G(z) = 2\beta z - \frac{1}{N}\sum_{i=1}^{N} \log(z - \lambda_i), \tag{4.42}$$

where γ is any constant satisfying $\gamma > \lambda_1$, the integration contour is the vertical line

from $\gamma - \mathrm{i}\infty$ *to* $\gamma + \mathrm{i}\infty$*, the* log *function is defined in the principal branch, and*

$$C_N = \frac{\Gamma(N/2)}{2\pi \mathrm{i}(N\beta)^{N/2-1}}. \tag{4.43}$$

Here $\Gamma(z)$ *denotes the Gamma function.*

Let M be a Wigner matrix with non-zero mean as in (4.6). Then its eigenvalues λ_i are random variables, and hence the above result gives a random integral representation of the partition function. In [13, 12], the above random integral was evaluated using the method of steepest-descent for different choices of random matrices. The key ingredient in controlling the error term is a precise estimate for the eigenvalues which are obtained in the random matrix theory.

Lemma IV.10 (Rigidity of eigenvalues: Theorem 2.13 of [55] and Theorem 6.3 of [77])**.** *For each positive integer* $k \in [1, N]$*, set* $\hat{k} := \min\{k, N+1-k\}$*. Let* γ_k *be the classical location defined by*

$$\int_{\gamma_k}^{\infty} \mathrm{d}\sigma_{scl}(x) = \frac{1}{N}\left(k - \frac{1}{2}\right). \tag{4.44}$$

Then, for every $0 < \epsilon < \frac{1}{2}$*,*

$$|\lambda_k - \gamma_k| \leq \hat{k}^{-1/3} N^{-2/3+\epsilon} \tag{4.45}$$

for all $k = 2, 3, \cdots, N$ *with high probability. Furthermore, for fixed* $J > 1$*, recall* $\hat{J} = J + J^{-1}$*,*

$$|\lambda_1 - \hat{J}| \leq N^{-1/2+\epsilon} \tag{4.46}$$

holds with high probability.

From the rigidity, it is easy to obtain the following law of large numbers for eigenvalues.

Corollary IV.11 (c.f. Lemma 5.1 of [12]). *Fix $\delta > 0$, let $\{f_\alpha\}_{\alpha\in I} \subset C^1[-2-\delta, 2+\delta]$ be a family of monotonic increasing functions satisfying $\sup_{\alpha\in I} \max_x |f_\alpha(x)| \leq C_0$ and $\sup_{\alpha\in I} \max_x |f'_\alpha(x)| \leq C_1$. Then, for every $0 < \epsilon < 1$,*

$$\sup_{\alpha\in I} \left| \frac{1}{N} \sum_{i=2}^{N} f_\alpha(\lambda_i) - \int_{-2}^{2} f_\alpha(x) \mathrm{d}\sigma_{scl}(x) \right| = O(N^{-1+\epsilon}) \tag{4.47}$$

with high probability.

Proof. Let $f = f_\alpha$ for some $\alpha \in I$. The absolute value on the left hand-side is bounded above by

$$\left| \frac{1}{N} \sum_{i=2}^{N} f(\lambda_i) - \frac{1}{N} \sum_{i=2}^{N} f(\gamma_i) \right| + \left| \frac{1}{N} \sum_{i=2}^{N} f(\gamma_i) - \int_{-2}^{2} f(x) \mathrm{d}\sigma_{scl}(x) \right|. \tag{4.48}$$

By Lemma IV.10,

$$\left| \frac{1}{N} \sum_{i=2}^{N} (f(\lambda_i) - f(\gamma_i)) \right| \leq \frac{\max |f'(x)|}{N} \sum_{i=2}^{N} |\lambda_i - \gamma_i| \leq \frac{C_0}{N^{1-\epsilon}} \tag{4.49}$$

with high probability. On the other hand, set $\hat{\gamma}_j$ by

$$\int_{\hat{\gamma}_j}^{2} \mathrm{d}\sigma_{scl}(x) = \frac{j}{N}, \qquad j = 1, 2, \cdots, N, \tag{4.50}$$

and by convention $\hat{\gamma}_0 = 2$. As $f(x)$ is a monotonic increasing function, for $i = 2, 3, \cdots, N-1$,

$$\int_{\hat{\gamma}_{i+1}}^{\hat{\gamma}_i} f(x) \mathrm{d}\sigma_{scl}(x) \leq \frac{1}{N} f(\gamma_i) \leq \int_{\hat{\gamma}_{i-1}}^{\hat{\gamma}_{i-2}} f(x) \mathrm{d}\sigma_{scl}(x). \tag{4.51}$$

Thus,

$$\left| \frac{1}{N} \sum_{i=2}^{N} f(\gamma_i) - \int_{-2}^{2} f(x) \mathrm{d}\sigma_{scl}(x) \right| \leq \frac{3 \max |f(x)|}{N} \leq \frac{3C_1}{N}. \tag{4.52}$$

Since the upper bounds are independent of f, we obtain the result. □

4.3.2 Steepest-descent analysis

We now apply steepest descent analysis to the integral in Lemma 4.42. We deform the contour to pass a critical point and show that the main contribution to the integral comes from a small neighborhood of the critical point. For $G(z)$ given in (4.42), it is easy to check that all solutions of $G'(z) = 0$ are real-valued, and there is a unique critical point γ which lies in the interval (λ_1, ∞) (see Lemma 4.1 of [13]).

Note that since G is random, the critical point is also random. For the paramagnetic regime, it was shown in [13] that $\gamma - \lambda_1 = O(1)$ with high probability. In the same paper, it was also shown that in the ferromagnetic regime, $\gamma - \lambda_1 = O(N^{-1+\epsilon})$ with high probability. The following lemma establishes a corresponding result for the transitional regime; it shows that $\gamma - \lambda_1 = O(N^{-\frac{1}{2}+\epsilon})$ with high probability.

Lemma IV.12 (Critical point)**.** *Recall that (see* (4.16)*)* $J > 1$ *is fixed and* $2\beta = 2\beta_N = \frac{1}{J} + \frac{B}{\sqrt{N}}$ *with fixed* $B \in \mathbb{R}$*. Then, for every* $0 < \epsilon < \frac{1}{4}$*,*

$$\gamma = \lambda_1 + \frac{1}{2\sqrt{N}} \left(-\chi_N - B(J^2-1) + \sqrt{(\chi_N + (J^2-1)B)^2 + 4(J^2-1)} \right) + O(N^{-1+\epsilon}) \tag{4.53}$$

with high probability, where we set $\chi_N := \sqrt{N}(\lambda_1 - \hat{J})$*.*

Note that γ given above is larger than λ_1 with high probability since the term in the big parenthesis is positive.

Proof. Set

$$\theta := \frac{-\chi_N - B(J^2-1) + \sqrt{(\chi_N + (J^2-1)B)^2 + 4(J^2-1)}}{2}. \tag{4.54}$$

Note that $\theta > 0$. By the rigidity of λ_1, we have $|\chi_N| \leq N^{\frac{\epsilon}{4}}$ and hence, $\theta \leq N^{\frac{\epsilon}{3}}$ with high probability. On the other hand, using $-a + \sqrt{a^2+b^2} = \frac{b^2}{\sqrt{a^2+b^2}+a}$,

$$\theta = \frac{2(J^2-1)}{\sqrt{(J^2-1)B+\chi_N)^2+4(J^2-1)}+((J^2-1)B+\chi_N)},$$

and hence $\theta \geq CN^{-\frac{\epsilon}{4}}$ for some constant $C > 0$ with high probability. Hence,

$$N^{-\frac{\epsilon}{3}} \leq \theta \leq N^{\frac{\epsilon}{3}} \tag{4.55}$$

with high probability. Set

$$\gamma_\pm := \lambda_1 + \frac{\theta}{\sqrt{N}} \pm N^{-1+\epsilon}. \tag{4.56}$$

By the above properties of θ, we have $\gamma_\pm > \lambda_1$ with high probability. We will show that $G'(\gamma_-) < 0$ and $G'(\gamma_+) > 0$ with high probability. Since $G'(z)$ is a monotone increasing function for real z in the interval (λ_1, ∞), this shows that $\gamma_- < \gamma < \gamma_+$ with high probability, proving the lemma.

Recall that $\lambda_1 \to \hat{J}$ in probability. Let us write

$$\gamma_\pm = J + \frac{1}{J} + \frac{\phi}{\sqrt{N}} \pm N^{-1+\epsilon}, \qquad \phi := \theta + \chi_N \tag{4.57}$$

where $\chi_N = \sqrt{N}(\lambda_1 - \hat{J})$. Note that $\phi = O(N^{\frac{\epsilon}{3}})$ with high probability. Now, notice that

$$G'(z) = 2\beta - \frac{1}{N}\sum_{i=2}^{N} \frac{1}{z-\lambda_i} - \frac{1}{N(z-\lambda_1)}. \tag{4.58}$$

We apply Corollary IV.11 to the family of functions $\{\frac{1}{z-x}\}_{z>2+c}$ for some constant $c > 0$ and obtain

$$G'(\gamma_\pm) = 2\beta - \frac{\gamma_\pm - \sqrt{\gamma_\pm^2 - 4}}{2} + O(N^{-1+\frac{\epsilon}{3}}) - \frac{1}{N(\gamma_\pm - \lambda_1)}$$

with high probability. By (4.57),

$$\frac{\gamma_\pm - \sqrt{\gamma_\pm^2 - 4}}{2} = \frac{1}{J} - \frac{1}{J^2-1}\left(\frac{\phi}{\sqrt{N}} \pm N^{-1+\epsilon}\right) + O(N^{-1+\frac{2\epsilon}{3}}).$$

By (4.56),

$$\frac{1}{N(\gamma_\pm - \lambda_1)} = \frac{1}{\theta\sqrt{N}}\left(1 \mp \frac{N^{-\frac{1}{2}+\epsilon}}{\theta} + O\left(\frac{N^{-1+2\epsilon}}{\theta^2}\right)\right).$$

Using the formula of 2β and the estimate (4.55) for $\frac{1}{\theta}$, we find that

$$G'(\gamma_\pm) = \frac{1}{\sqrt{N}}\left(B + \frac{\phi}{J^2-1} - \frac{1}{\theta}\right) \pm \left(\frac{1}{J^2-1} + \frac{1}{\theta^2}\right) N^{-1+\epsilon} + O(N^{-1+\frac{2\epsilon}{3}}) \quad (4.59)$$

with high probability since $0 < \epsilon < \frac{1}{4}$. By the definition of θ, the leading term is zero. The coefficient of the second term is positive. Hence we find that $G'(\gamma_-) < 0$ and $G'(\gamma_+) > 0$, and we obtain the lemma. □

Then we have the following lemma.

Lemma IV.13. *Set*

$$s = s_N := \sqrt{N}(\gamma - J - J^{-1}) \text{ and } \Delta = \Delta_N := \sqrt{N}(\gamma - \lambda_1) = s_N - \chi_N. \quad (4.60)$$

Then, for every $\epsilon > 0$,

$$s = \frac{\chi_N - B(J^2-1) + \sqrt{(\chi_N + (J^2-1)B)^2 + 4(J^2-1)}}{2} + O(N^{-\frac{1}{2}+\epsilon}) \quad (4.61)$$

with high probability. We also have

$$|s| \le N^\epsilon \text{ and } N^{-\epsilon} \le \Delta \le N^\epsilon \quad (4.62)$$

with high probability.

Proof. The previous lemma implies (4.61). The first part of (4.62) follows from the

fact that $\chi_N = O(N^\epsilon)$ with high probability. The second part is the estimate (4.55) in the proof of the previous lemma. □

We also need the following lemma.

Lemma IV.14. *For every* $0 < \epsilon < 1$,

$$\frac{1}{N}\sum_{i=2}^{N}\frac{1}{(\gamma-\lambda_i)^2} = \frac{1}{J^2-1} + O(N^{-1+\epsilon}) \tag{4.63}$$

with high probability.

Proof. This follows from Corollary IV.11 applied to $f(x) = \frac{1}{(\gamma-x)^2}$. □

The following auxiliary lemma is used to estimate an error in the steepest descent analysis.

Lemma IV.15. *Define*

$$\mathbf{I}_m(\alpha) := \int_{-\infty}^{\infty} \frac{t^m}{\sqrt{1+\mathrm{i}t}} e^{-\frac{\alpha}{4}t^2+\frac{\mathrm{i}t}{2}}\,\mathrm{d}t \tag{4.64}$$

for non-negative integers m *and* $\alpha > 0$, *where the square root is the defined on the principal branch. We set* $\mathbf{I}(\alpha) := \mathbf{I}_0(\alpha)$; *see* (4.21). *Then,*

$$\mathbf{I}(\alpha) = \sqrt{\frac{4\pi}{\alpha}}(1+O(\alpha^{-1})) \quad \text{as } \alpha \to +\infty, \tag{4.65}$$

$$\mathbf{I}(\alpha) = \sqrt{\frac{8\pi}{e}}(1+O(\alpha)) \quad \text{as } \alpha \to 0_+, \tag{4.66}$$

and for every $m \geq 0$,

$$\mathbf{I}_m(\alpha) \text{ is uniformly bounded for } \alpha \in (0,\infty). \tag{4.67}$$

A particular consequence is that the derivative $\mathbf{I}'(\alpha) = -\frac{1}{4}\mathbf{I}_2(\alpha)$ *is uniformly bounded for* $\alpha > 0$. *Furthermore,* $\mathbf{I}(\alpha) > 0$ *for all* $\alpha > 0$.

Proof. Consider (4.65). Applying the method of steepest-descent to $\mathbf{I}(\alpha) = \int_{-\infty}^{\infty} g(t)e^{\alpha h(t)}\mathrm{d}t$ with $h(z) = -\frac{z^2}{4}$ and $g(z) = \frac{1}{\sqrt{1+\mathrm{i}z}}e^{\frac{\mathrm{i}z}{2}}$, we find that

$$\mathbf{I}(\alpha) = \frac{e^{\alpha h(z_c)}}{\sqrt{\alpha}}\left[\sqrt{\frac{2\pi}{|h''(z_c)|}}g(z_c) + O(\alpha^{-1})\right] = \sqrt{\frac{4\pi}{\alpha}}(1 + O(\alpha^{-1})) \tag{4.68}$$

as $\alpha \to +\infty$. For $\mathbf{I}_m(\alpha)$, using $\int_{-\infty}^{\infty} y^m e^{-\alpha y^2}\mathrm{d}y = O(\alpha^{-(m+1)/2})$, we find that

$$\mathbf{I}_m(\alpha) = O(\alpha^{-\frac{m+1}{2}}) \quad \text{as } \alpha \to +\infty. \tag{4.69}$$

Consider the limit $\alpha \to 0_+$. After the change of the variables $t = z/\alpha$,

$$\mathbf{I}(\alpha) = \frac{e^{-\frac{1}{4\alpha}}}{\sqrt{\alpha}}\int_{-\infty}^{\infty}\frac{e^{-\frac{(z-\mathrm{i})^2}{4\alpha}}}{\sqrt{\alpha+\mathrm{i}z}}\mathrm{d}z. \tag{4.70}$$

The integrand is analytic in the complex plane minus the vertical line from $\mathrm{i}\alpha$ to $\mathrm{i}\infty$. Note that the saddle point is i and it is on the branch cut. We show that the main contribution to the integral comes from the branch point $z = \mathrm{i}\alpha$. We deform the contour so that it consists of the following four line segments: L_1 from $\mathrm{i} - \infty$ to i on the left half-plane, L_2 from i to $\mathrm{i}\alpha$ lying on the left of the branch cut, L_3 from $\mathrm{i}\alpha$ to i lying on the right of the branch cut, and L_4 from i to $\mathrm{i} + \infty$ lying on the right-half plane. On L_4, setting $z = \mathrm{i} + \sqrt{\alpha}x$,

$$\int_{L_4}\frac{e^{-\frac{(z-\mathrm{i})^2}{4\alpha}}}{\sqrt{\alpha+\mathrm{i}z}}\mathrm{d}z = \sqrt{\alpha}\int_0^{\infty}\frac{e^{-\frac{x^2}{4}}}{\sqrt{\alpha-1+\mathrm{i}\sqrt{\alpha}x}}\mathrm{d}x = O(\sqrt{\alpha}) \tag{4.71}$$

as $\alpha \to 0$. Similarly, the integral over L_1 is also of the same order. On the other

hand, setting $z = \mathrm{i}\alpha + \mathrm{i}y$,

$$\int\limits_{L_2 \cup L_3} \frac{e^{-\frac{(z-\mathrm{i})^2}{4\alpha}}}{\sqrt{\alpha + \mathrm{i}z}} \mathrm{d}z = 2 \int\limits_0^{1-\alpha} \frac{e^{\frac{(\alpha+y-1)^2}{4\alpha}}}{\sqrt{y}} \mathrm{d}y = 2e^{\frac{(\alpha-1)^2}{4\alpha}} \int\limits_0^{1-\alpha} \frac{e^{\frac{y}{2}+\frac{y^2-2y}{4\alpha}}}{\sqrt{y}} \mathrm{d}y. \tag{4.72}$$

The function $y^2 - 2y$ decreases as y increases from $y = 0$ to $y = 1$. Hence the main contribution to the integral comes near the point $y = 0$. Using Watson's lemma,

$$\int\limits_0^{1-\alpha} \frac{e^{\frac{y}{2}+\frac{y^2-2y}{4\alpha}}}{\sqrt{y}} \mathrm{d}y = \Gamma(1/2)\sqrt{2\alpha}(1 + O(\alpha)). \tag{4.73}$$

Combining together and using $\Gamma(1/2) = \sqrt{\pi}$, we obtain (4.66). For $\mathbf{I}_m(\alpha)$, the analysis is same except that we use

$$\int\limits_0^{1-\alpha} (\mathrm{i}\alpha + \mathrm{i}y)^m \frac{e^{\frac{y}{2}+\frac{y^2-2y}{4\alpha}}}{\sqrt{y}} \mathrm{d}y = O(\alpha^{m+1/2}). \tag{4.74}$$

Hence, we find that for $m \geq 0$, $\mathbf{I}_m(\alpha) = O(1)$ as $\alpha \to 0_+$. Together with (4.69), this implies the uniform boundness of $\mathbf{I}_m(\alpha)$.

For the positiveness of $\mathbf{I}(\alpha)$, we first write it as

$$\mathbf{I}(\alpha) = \int\limits_{-\infty}^{\infty} \frac{e^{-\frac{\alpha}{4}t^2 + \frac{\mathrm{i}}{2}(t - \arctan t)}}{(1+t^2)^{1/4}} \mathrm{d}t = 2 \int\limits_0^{\infty} \frac{e^{-\frac{\alpha}{4}t^2}}{(1+t^2)^{1/4}} \cos\left(\frac{1}{2}(t - \arctan t)\right) \mathrm{d}t. \tag{4.75}$$

The function $\theta(t) = t - \arctan t$ is monotone increasing. We use the inverse function, $t = t(\theta)$, to change the variables and find that

$$\mathbf{I}(\alpha) = 2 \int\limits_0^{\infty} e^{-\frac{\alpha}{4}t^2} \frac{(1+t^2)^{3/4}}{t^2} \cos\left(\frac{\theta}{2}\right) \mathrm{d}\theta, \qquad t = t(\theta). \tag{4.76}$$

Since $e^{-\frac{\alpha}{4}t(\theta)^2}$ is positive and monotone decreasing in θ, we obtain $I(\alpha) > 0$ for every

$\alpha > 0$ if we show that (i)

$$\int_0^{\pi} \frac{(1+t^2)^{3/4}}{t^2} \cos\left(\frac{\theta}{2}\right) \mathrm{d}\theta \geq -\int_{\pi}^{3\pi} \frac{(1+t^2)^{3/4}}{t^2} \cos\left(\frac{\theta}{2}\right) \mathrm{d}\theta, \tag{4.77}$$

and (ii)

$$(-1)^k \int_{(2k-1)\pi}^{(2k+1)\pi} \frac{(1+t^2)^{3/4}}{t^2} \cos\left(\frac{\theta}{2}\right) \mathrm{d}\theta, \qquad k = 1, 2, 3, \cdots, \tag{4.78}$$

is decreasing in k. (i) can be verified numerically. On the other hand, (ii) follows immediately from the fact $(1+t^2)^{3/4}/t^2$ is a decreasing function of t. This completes the proof. □

We now evaluate the integral in (4.42) using the steepest descent analysis.

Lemma IV.16. *Fix $J > 1$ and let $2\beta = J^{-1} + BN^{-1/2}$. Consider $G(z)$ defined in* (4.42). *Then, for every $0 < \epsilon < \frac{1}{8}$,*

$$\int_{\gamma-\mathrm{i}\infty}^{\gamma+\mathrm{i}\infty} e^{\frac{N}{2}G(z)} \mathrm{d}z = \frac{\mathrm{i}\Delta e^{\frac{N}{2}G(\gamma)}}{\sqrt{N}} \mathbf{I}(F''(\gamma)\Delta^2) \left(1 + O(N^{-\frac{1}{2}+4\epsilon})\right) \tag{4.79}$$

with high probability, where

$$F(z) = 2\beta z - \frac{1}{N}\sum_{i=2}^{N} \log(z - \lambda_i) - \frac{1}{N}\log(\gamma - \lambda_1) - \frac{z-\gamma}{N(\gamma-\lambda_1)} \tag{4.80}$$

and $\mathbf{I}(\alpha)$ is defined in (4.21). *Recall that $\Delta = \sqrt{N}(\gamma - \lambda_1)$ (see Lemma IV.13.)*

Proof. We choose the γ, which defines the contour, as the critical point of $G(z)$. The path of steepest-descent is locally a vertical line near the critical point. It turns out that, instead of using the path of steepest-descent, it is enough to proceed the analysis using the straight line $\gamma + \mathrm{i}\mathbb{R}$ globally. This choice was also made for the analysis in the paramagnetic regime in [13].

We first write, using the function $F(z)$,

$$\int_{\gamma-\mathrm{i}\infty}^{\gamma+\mathrm{i}\infty} e^{\frac{N}{2}G(z)}\mathrm{d}z = e^{\frac{N}{2}G(\gamma)} \int_{\gamma-\mathrm{i}\infty}^{\gamma+\mathrm{i}\infty} e^{\frac{N}{2}(G(z)-F(z))+\frac{N}{2}(F(z)-G(\gamma))}\mathrm{d}z. \tag{4.81}$$

From the definitions of $G(z)$ and $F(z)$,

$$e^{\frac{N}{2}(G(z)-F(z))} = \sqrt{\frac{\gamma-\lambda_1}{z-\lambda_1}} e^{\frac{z-\gamma}{2(\gamma-\lambda_1)}}. \tag{4.82}$$

Changing the variables $z = \gamma + \mathrm{i}tN^{-1/2}$ and using the notation $\Delta = \sqrt{N}(\gamma-\lambda_1)$,

$$\int_{\gamma-\mathrm{i}\infty}^{\gamma+\mathrm{i}\infty} e^{\frac{N}{2}G(z)}\mathrm{d}z = \frac{\mathrm{i}e^{\frac{N}{2}G(\gamma)}}{\sqrt{N}} \int_{-\infty}^{\infty} \frac{e^{\frac{\mathrm{i}t}{2\Delta}}}{\sqrt{1+\frac{\mathrm{i}t}{\Delta}}} e^{\frac{N}{2}(F(\gamma+\mathrm{i}tN^{-1/2})-G(\gamma))}\mathrm{d}t. \tag{4.83}$$

It is easy to check that the part of the integral with $|t| \geq N^{\epsilon}$ is small. To show this, we first note that

$$\Re\left(N\left(F(\gamma+\frac{\mathrm{i}t}{\sqrt{N}})-G(\gamma)\right)\right) = -\Re\sum_{i=2}^{N}\log\left(\frac{\gamma-\lambda_i+\mathrm{i}tN^{-1/2}}{\gamma-\lambda_i}\right) \leq -\frac{N-1}{2}\log\left(1+\frac{c^2t^2}{N}\right)$$

with high probability for some constant $c > 0$, since there is a constant $c > 0$ such that $c \leq \gamma - \lambda_i \leq \frac{1}{c}$ for all $i = 2, \cdots, N$, with high probability. Hence,

$$\begin{aligned}\left|\int_{N^{\epsilon}}^{\infty} \frac{e^{\frac{\mathrm{i}t}{2\Delta}}}{\sqrt{1+\frac{\mathrm{i}t}{\Delta}}} e^{\frac{N}{2}(F(\gamma+itN^{-1/2})-G(\gamma)}\mathrm{d}t\right| &\leq \int_{N^{\epsilon}}^{\infty} e^{-\frac{N-1}{2}\log\left(1+\frac{c^2t^2}{N}\right)}\mathrm{d}t \\ &\leq \int_{N^{\epsilon}}^{N} e^{-\frac{c^2}{8}N^{2\epsilon}}\mathrm{d}t + \int_{N}^{\infty} \frac{1}{(c^2N^{-1}t^2)^{N/4}}\mathrm{d}t = O(e^{-N^{\epsilon}}) + O(N^{-N/8})\end{aligned} \tag{4.84}$$

with high probability.

Consider the part $|t| \leq N^{\epsilon}$. Note that $F(z)$ satisfies $F(\gamma) = G(\gamma)$, $F'(\gamma) = G'(\gamma) = 0$, and for each $m \geq 2$, $F^{(m)}(z) = O(1)$ uniformly for z in a small neighbor-

hood of γ (by Corollary IV.11). For $m = 2$, by Lemma IV.14,

$$c_1 \leq F''(\gamma) \leq c_2 \tag{4.85}$$

for some constants $0 < c_1 < c_2$, uniformly in N. By Taylor expansion, for $|t| \leq N^{\epsilon}$,

$$F(\gamma + \mathrm{i}tN^{-1/2}) - G(\gamma) = -\frac{F''(\gamma)t^2}{2N} - \frac{\mathrm{i}F'''(\gamma)t^3}{6N^{3/2}} + O(N^{-2+4\epsilon}) \tag{4.86}$$

and hence,

$$e^{\frac{N}{2}(F(\gamma+\mathrm{i}tN^{-1/2})-G(\gamma))} = e^{-\frac{F''(\gamma)t^2}{4}}\left(1 - \frac{\mathrm{i}F'''(\gamma)t^3}{12N^{1/2}} + O(N^{-1+6\epsilon})\right). \tag{4.87}$$

Therefore,

$$\begin{aligned}
&\int_{-N^{\epsilon}}^{N^{\epsilon}} \frac{e^{\frac{\mathrm{i}t}{2\Delta}}}{\sqrt{1+\frac{\mathrm{i}t}{\Delta}}} e^{\frac{N}{2}\left(F(\gamma+\mathrm{i}tN^{-1/2})-G(\gamma)\right)}\mathrm{d}t \\
&= \int_{-\infty}^{\infty} \frac{e^{\frac{\mathrm{i}t}{2\Delta}}}{\sqrt{1+\frac{\mathrm{i}t}{\Delta}}} e^{-\frac{F''(\gamma)}{4}t^2}\mathrm{d}t - \frac{\mathrm{i}F'''(\gamma)}{12N^{1/2}}\int_{-\infty}^{\infty} t^3 \frac{e^{\frac{\mathrm{i}t}{2\Delta}}}{\sqrt{1+\frac{\mathrm{i}t}{\Delta}}} e^{-\frac{F''(\gamma)}{4}t^2}\mathrm{d}t + O(N^{-1+6\epsilon}) \\
&= \Delta\, \mathbf{I}(F''(\gamma)\Delta^2) - \frac{\mathrm{i}F'''(\gamma)\Delta^4}{12N^{1/2}}\, \mathbf{I}_3(F''(\gamma)\Delta^2) + O(N^{-1+6\epsilon}).
\end{aligned} \tag{4.88}$$

By (4.85) and Lemma IV.13, $c_1 N^{-\epsilon} \leq F''(\gamma)\Delta^2 \leq c_2 N^{\epsilon}$. Hence, Lemma IV.15 implies that

$$\mathbf{I}(F''(\gamma)\Delta^2) \geq cN^{-\epsilon} \tag{4.89}$$

for some constant $c > 0$. Hence, using Lemma IV.13, Lemma IV.15, and the uniform boundedness of $F'''(\gamma)$, we find that (4.88) is equal to

$$\Delta\, \mathbf{I}(F''(\gamma)\Delta^2)(1 + O(N^{-\frac{1}{2}+4\epsilon})) \tag{4.90}$$

if $0 < \epsilon < \frac{1}{8}$. Thus, using (4.89) and Lemma IV.13 again, we conclude that

$$\int_{\gamma-\mathrm{i}\infty}^{\gamma+\mathrm{i}\infty} e^{\frac{N}{2}G(z)}\mathrm{d}z = \frac{\mathrm{i}\Delta e^{\frac{N}{2}G(\gamma)}}{\sqrt{N}}\,\mathbf{I}(F''(\gamma)\Delta^2)(1+O(N^{-1/2+4\epsilon})). \tag{4.91}$$

□

4.3.3 Proof of Theorem IV.4

Proof of Theorem IV.4. From Lemma IV.9 and Lemma IV.16, for every $0 < \epsilon < \frac{1}{8}$,

$$Z_N = C_N \frac{\mathrm{i}\Delta e^{\frac{N}{2}G(\gamma)}}{\sqrt{N}}\,\mathbf{I}(F''(\gamma)\Delta^2)(1+O(N^{-\frac{1}{2}+4\epsilon})) \tag{4.92}$$

with high probability. Using Stirling's formula,

$$C_N = \frac{\Gamma(N/2)}{2\pi\mathrm{i}(N\beta)^{N/2-1}} = \frac{\sqrt{N}\beta}{\mathrm{i}\sqrt{\pi}(2\beta e)^{N/2}}(1+O(N^{-1})), \tag{4.93}$$

thus we find that $F_N = \frac{1}{N}\log Z_N$ satisfies

$$F_N = \frac{1}{2}(G(\gamma)-1-\log(2\beta))+\frac{1}{N}\left(\log\left(\frac{\beta\Delta}{\sqrt{\pi}}\right)+\log\mathbf{I}(F''(\gamma)\Delta^2)\right)+O(N^{-\frac{3}{2}+4\epsilon}) \tag{4.94}$$

with high probability.

Let us consider $G(\gamma)$. Since γ and $\hat{J} = J + J^{-1}$ are away from $\lambda_2, \cdots, \lambda_N$ with high probability,

$$\begin{aligned}\log(\gamma-\lambda_i) &= \log(\hat{J}-\lambda_i) - \log\left(1-\frac{\gamma-\hat{J}}{\gamma-\lambda_i}\right)\\ &= \log(\hat{J}-\lambda_i) + \frac{\gamma-\hat{J}}{\gamma-\lambda_i} + \frac{(\gamma-\hat{J})^2}{2(\gamma-\lambda_i)^2} + O(|\gamma-\hat{J}|^3)\end{aligned} \tag{4.95}$$

for $i = 2, \cdots, N$, where we also use that $\gamma - \hat{J} = O(N^{-\frac{1}{2}+\epsilon})$ with high probability (see Lemma IV.13). Then, using Lemma IV.14 and the fact that $G'(\gamma) = 2\beta -$

$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\gamma-\lambda_i}=0$,

$$\frac{1}{N}\sum_{i=2}^{N}\log(\gamma-\lambda_i)=\frac{1}{N}\sum_{i=2}^{N}\log(\hat{J}-\lambda_i)+2\beta(\gamma-\hat{J})-\frac{\gamma-\hat{J}}{N(\gamma-\lambda_1)}+\frac{(\gamma-\hat{J})^2}{2(J^2-1)}+O(N^{-\frac{3}{2}+3\epsilon})$$

with high probability. Hence, from the formula of $G(z)$ in (4.42),

$$\begin{aligned}G(\gamma)&=2\beta\hat{J}-\frac{1}{N}\sum_{i=2}^{N}\log(\hat{J}-\lambda_i)-\frac{1}{N}\log(\gamma-\lambda_1)+\frac{\gamma-\hat{J}}{N(\gamma-\lambda_1)}-\frac{(\gamma-\hat{J})^2}{2(J^2-1)}+O(N^{-\frac{3}{2}+3\epsilon})\\&=2\beta\hat{J}-\frac{1}{N}\sum_{i=2}^{N}\log(\hat{J}-\lambda_i)-\frac{1}{N}\log\left(\frac{\Delta}{\sqrt{N}}\right)+\frac{s_N}{N\Delta}-\frac{s_N^2}{2N(J^2-1)}+O(N^{-\frac{3}{2}+3\epsilon})\end{aligned}$$

using the notations $s_N=\sqrt{N}(\gamma-\hat{J})$ and $\Delta=\sqrt{N}(\gamma-\lambda_1)$ in Lemma IV.13. Thus,

$$\begin{aligned}F_N=&\beta\hat{J}-\frac{1}{2}-\frac{1}{2}\log(2\beta)-\frac{1}{2N}\sum_{i=2}^{N}\log(\hat{J}-\lambda_i)+\frac{1}{4N}\log N\\&+\frac{1}{N}\left(\frac{s_N}{2\Delta}-\frac{s_N^2}{4(J^2-1)}+\frac{1}{2}\log\Delta+\log\frac{\beta}{\sqrt{\pi}}+\log\mathbf{I}(F''(\gamma)\Delta^2)\right)+O(N^{-\frac{3}{2}+4\epsilon}).\end{aligned}\tag{4.96}$$

To conclude Theorem IV.4, we use (i) the fact that $\Delta=s_N-\chi_N$, (ii) the asymptotic (4.61) of s_N in terms of χ_N, (iii) the fact that $F''(\gamma)=\frac{1}{J^2-1}+O(N^{-1+\epsilon})$ which follows from Lemma IV.14, and (iv) the fact that $\mathbf{I}'(\alpha)$ is uniformly bounded for $\alpha>0$ (see Lemma IV.15). □

4.4 Partial linear statistics

This section is devoted to a proof of Theorem IV.6 on partial linear statistics. The proof is a simple modification of [13] for the linear statistics of all eigenvalues, which, in turn, follows the proof of [9, 8] for the case when the random matrix has zero mean.

4.4.1 Proof of Theorem IV.6

Recall $\hat{J} := J + J^{-1}$ denotes the classical location of the largest eigenvalue of a Wigner matrix of non-zero mean. Fix (N-independent) constants $a_- < -2$ and $2 < a_+ < \hat{J}$. Let Γ be the rectangular contour whose vertices are $(a_- \pm \mathrm{i}v_0)$ and $(a_+ \pm \mathrm{i}v_0)$ for some $v_0 \in (0, 1]$. The contour is oriented counter-clockwise. For a test function $\varphi(x)$ which is analytic in a neighborhood of $[-2, 2]$, we consider

$$\begin{aligned}\mathcal{N}_N^{(2)}(\varphi) :=& \sum_{i=2}^{N} \varphi(\lambda_i) - N \int_{\mathbb{R}} \varphi(x) \mathrm{d}\sigma_{scl}(x) \\ =& \sum_{i=2}^{N} \frac{1}{2\pi \mathrm{i}} \oint_{\Gamma} \frac{\varphi(z)}{z - \lambda_i} \mathrm{d}z - \frac{N}{2\pi \mathrm{i}} \int_{\mathbb{R}} \oint_{\Gamma} \frac{\varphi(z)}{z - x} \mathrm{d}z \mathrm{d}\sigma_{scl}(x) = -\frac{1}{2\pi i} \oint_{\Gamma} \varphi(z) \xi_N^{(2)} \mathrm{d}z,\end{aligned} \tag{4.97}$$

where

$$\xi_N^{(2)}(z) := \sum_{i=2}^{N} \frac{1}{\lambda_i - z} - N \int_{\mathbb{R}} \frac{1}{x - z} \mathrm{d}\sigma_{scl}(x). \tag{4.98}$$

Decompose Γ into $\Gamma_u \cup \Gamma_d \cup \Gamma_l \cup \Gamma_r \cup \Gamma_0$, where

$$\Gamma_u = \{z = x + \mathrm{i}v_0 : a_- \le x \le a_+\}, \tag{4.99}$$

$$\Gamma_d = \{z = x - \mathrm{i}v_0 : a_- \le x \le a_+\}, \tag{4.100}$$

$$\Gamma_l = \{z = a_- + \mathrm{i}y : N^{-\delta} \le |y| \le v_0\}, \tag{4.101}$$

$$\Gamma_r = \{z = a_+ + \mathrm{i}y : N^{-\delta} \le |y| \le v_0\}, \tag{4.102}$$

$$\Gamma_0 = \{z = a_- + \mathrm{i}y : |y| < N^{-\delta}\} \cup \{z = a_+ + \mathrm{i}y : |y| < N^{-\delta}\} \tag{4.103}$$

for some $\delta > 0$. In the proof of Theorem 1.6 in [13], the authors showed that

$$\xi_N(z) := \sum_{i=1}^{N} \frac{1}{\lambda_i - z} - N \int_{\mathbb{R}} \frac{1}{x - z} \mathrm{d}\sigma_{scl}(x) = \xi_N^{(2)}(z) + \frac{1}{\lambda_1 - z} \tag{4.104}$$

converges weakly to a Gaussian process with mean $b(z) = b^{(2)}(z) + \frac{1}{\hat{J}-z}$ and covariance $\Gamma(z_i, z_j) = \Gamma^{(2)}(z_i, z_j)$ where $b^{(2)}(z)$ and $\Gamma^{(2)}(z_i, z_j)$ are given in the proposition below. Since for each fixed $z \in \mathbb{C}_+$, $\frac{1}{\lambda_1 - z} \to \frac{1}{\hat{J}-z}$ in probability (by Lemma IV.10), it is natural to expect the following result for a partial sum.

Proposition IV.17. *Let*

$$s(z) = \int \frac{1}{x-z} \mathrm{d}\sigma_{scl}(x) = \frac{-z + \sqrt{z^2-4}}{2} \tag{4.105}$$

be the Stieltjes transform of the semicircle measure. Fix a constant $c > 0$ and a path $\mathcal{K} \subset \mathbb{C}_+$ such that $\Im z > c$ for $z \in \mathcal{K}$. Then the process $\{\xi_N^{(2)}(z) : z \in \mathcal{K}\}$ converges weakly to a Gaussian process with the mean

$$b^{(2)}(z) = \frac{s(z)^2}{1-s(z)^2}\left(-\frac{J}{1+Js(z)} + (w_2 - 1)s(z) + s'(z)s(z) + (W_4 - 3)s(z)^3\right) - \frac{1}{\hat{J}-z} \tag{4.106}$$

and the covariance matrix

$$\Gamma^{(2)}(z_i, z_j) = s'(z_i)s'(z_j)\left((w_2 - 2) + 2(W_4 - 3)s(z_i)s(z_j) + \frac{2}{(1-s(z_i)s(z_j))^2}\right). \tag{4.107}$$

Remark IV.18. Note that as $z \to \hat{J}$,

$$\frac{s(z)^2}{1-s(z)^2}\frac{J}{1+Js(z)} = \frac{s'(z)}{\frac{1}{J}+s(z)} = \frac{1}{z-\hat{J}} + \frac{s''(\hat{J})}{s'(\hat{J})} + O(z-\hat{J}). \tag{4.108}$$

Hence, $b^{(2)}(z)$ is analytic near $\hat{J}$ and thus analytic for $z \in \mathbb{C} \setminus [-2, 2]$.

In order to complete the proof of Theorem IV.6, we will prove the following lemma.

Lemma IV.19. *Define the events*

$$\Omega_N := \{\lambda_1 \geq \hat{J} - N^{-1/3}, \lambda_2 \leq 2 + N^{-1/3}\} \tag{4.109}$$

which satisfies $\mathbb{P}(\Omega_N^c) < N^{-D}$ *for any fixed (large)* $D > 0$. *Then for some* $\delta > 0$,

$$\lim_{v_0 \to 0^+} \limsup_{N \to \infty} \int_{\Gamma_\#} \mathbb{E}\, |\xi_N^{(2)}(z) \mathbb{1}_{\Omega_N}|^2 \mathrm{d}z = 0, \tag{4.110}$$

where $\Gamma_\#$ *can be* Γ_r, Γ_l *or* Γ_0.

From the explicit formulas (4.106) and (4.107), it is easy to check that

$$\lim_{v_0 \to 0^+} \int_{\Gamma_\#} \mathbb{E}\, |\xi^{(2)}(z)|^2 \mathrm{d}z = 0. \tag{4.111}$$

Proposition IV.17, Lemma IV.19 and (4.111) imply that $\mathcal{N}_N^{(2)}(\varphi)$ converges in distribution to a Gaussian random variable with the following mean and variance:

$$-\frac{1}{2\pi \mathrm{i}} \oint_\Gamma \varphi(z) b^{(2)}(z) \mathrm{d}z, \qquad \frac{1}{(2\pi \mathrm{i})^2} \oint_\Gamma \oint_\Gamma \varphi(z_1)\varphi(z_2)\Gamma(z_1, z_2) \mathrm{d}z_1 \mathrm{d}z_2. \tag{4.112}$$

It is direct to check that these are equal to $M^{(2)}(\varphi)$ and $V^{(2)}(\varphi)$ (see Section 4.2 in [13]). We thus obtain Theorem IV.6.

4.4.2 Proof of Proposition IV.17

From Theorem 7.1 of [21], we need to show (i) the finite-dimensional convergence of $\xi_N^{(2)}(z)$ to a Gaussian vector with desired mean and variance, and (ii) the tightness of $\xi_N^{(2)}(z)$. We will base our proof on the corresponding properties of $\xi_N(z)$ obtained in [13]. Let us first recall the limit theorem for $\xi_N(z)$.

Lemma IV.20 (Proposition 4.1 in [13])**.** *Let* $s(z)$ *and* $\mathcal{K}$ *defined in the same way as in Proposition IV.17. Then, the process* $\{\xi_N(z) : z \in \mathcal{K}\}$ *converges weakly to a*

Gaussian process $\{\xi(z) : z \in \mathcal{K}\}$ *with the mean*

$$b(z) = \frac{s(z)^2}{1 - s(z)^2} \left(-\frac{J}{1 + Js(z)} + (w_2 - 1)s(z) + s'(z)s(z) + (W_4 - 3)s(z)^3 \right) \tag{4.113}$$

and the covariance matrix

$$\Gamma(z_i, z_j) = s'(z_i)s'(z_j) \left((w_2 - 2) + 2(W_4 - 3)s(z_i)s(z_j) + \frac{2}{(1 - s(z_i)s(z_j))^2} \right). \tag{4.114}$$

Let $z_1, z_2, \cdots, z_p$ are p distinct points in $\mathcal{K}$. The above lemma implies that the random vector $(\xi_N(z_i))_{i=1}^p$ converges weakly to a p-dimensional Gaussian distribution with the mean $(b(z_i))_{i=1}^p$ and the covariance matrix $\Gamma(z_i, z_j)$. Since the distance between $\mathcal{K}$ and λ_1 is bounded below, $\frac{1}{\lambda_1 - z_i} \to \frac{1}{\hat{J} - z_i}$ in probability for $i = 1, \cdots, p$. Hence, by Slutsky's theorem, $(\xi_N^{(2)}(z_i))_{i=1}^p$ converges weakly to a p-dimensional Gaussian distribution vector with the mean $(b^{(2)}(z_i))_{i=1}^p$ and the covariance matrix $\Gamma^{(2)}(z_i, z_j)$, where

$$b^{(2)}(z) = b(z) - \frac{1}{\hat{J} - z}, \tag{4.115}$$

and $\Gamma^{(2)}(z_i, z_j) = \Gamma(z_i, z_j)$.

From Theorem 12.3 of [21], in order to show the tightness of a random process $(\zeta_N(z))_{z \in \mathcal{K}}$, it is sufficient to show that (i) $(\zeta_N(z))_N$ is tight for a fixed z, and (ii) the following Hölder condition holds: for some N-independent constant $K > 0$,

$$\mathbb{E}\,|\zeta_N(z_1) - \zeta_N(z_2)|^2 \le K|z_1 - z_2|^2, \qquad z_1, z_2 \in \mathcal{K}. \tag{4.116}$$

In [13], the authors considered the random process $\zeta_N(z) := \xi_N(z) - \mathbb{E}[\xi_N(z)]$, and proved that it satisfies conditions (i) and (ii). Now, we consider $\xi_N^{(2)}(z) := \zeta_N^{(2)} + \mathbb{E}[\xi_N(z)]$, where $\zeta_N^{(2)}(z) := \zeta_N(z) - \frac{1}{\lambda_1 - z}$. Since $\mathbb{E}[\xi_N(z)]$ converges, it is enough to check that $(\zeta_N^{(2)}(z))_N$ satisfies conditions (i) and (ii). Now for a fixed z, the tightness

of $(\zeta_N(z))_N$ and the boundedness of $\frac{1}{\lambda_1 - z}$ imply that $(\zeta_N^{(2)}(z))_N$ is tight. On the other hand, since $\zeta_N(z)$ satisfies the Hölder condition and $\Im z \geq c$ for $z \in \mathcal{K}$,

$$\begin{aligned} \mathbb{E}\,|\zeta_N^{(2)}(z_1) - \zeta_N^{(2)}(z_2)|^2 &\leq 2\,\mathbb{E}\,|\zeta_N(z_1) - \zeta_N(z_2)|^2 + 2\,\mathbb{E}\left|\frac{1}{\lambda_1 - z_1} - \frac{1}{\lambda_1 - z_2}\right|^2 \\ &\leq 2K|z_1 - z_2|^2 + \frac{2|z_1 - z_2|^2}{c^4} = \left(K + \frac{2}{c^4}\right)|z_1 - z_2|^2. \end{aligned} \tag{4.117}$$

Thus $\{\xi_N^{(2)}(z), z \in \mathcal{K}\}$ is tight. This completes the proof of Proposition IV.17.

4.4.3 Proof of Lemma IV.19

For $z \in \Gamma_0$, we notice that $|\xi_N^{(2)} \mathbb{1}_{\Omega_N}| \leq CN$ and then

$$\int_{\Gamma_0} \mathbb{E}\,|\xi_N^{(2)} \mathbb{1}_{\Omega_N}|^2 \leq CN^{2-\delta}. \tag{4.118}$$

Thus (4.110) holds for Γ_0 with $\delta > 2$. For Γ_r and Γ_l, it is sufficient to show $\mathbb{E}\,|\xi_N^{(2)}|^2 < K$ for some N-independent constant $K > 0$. The authors in [13] showed[2] that $\mathbb{E}\,|\xi_N(z)|^2 < K$. Hence, for $z \in \Gamma_r$,

$$|\xi_N^{(2)}(z)\mathbb{1}_{\Omega_N}|^2 \leq 2|\xi_N(z)\mathbb{1}_{\Omega_N}|^2 + 2\left|\frac{1}{\lambda_1 - z}\mathbb{1}_{\Omega_N}\right|^2. \tag{4.119}$$

The lemma then follows from the fact that $|\frac{1}{\lambda_1 - z}\mathbb{1}_{\Omega_N}|$ is bounded.

4.5 Joint Distribution of χ_N and $\mathcal{N}_N^{(2)}(\varphi)$

As before, let A be a random symmetric matrix of size N whose entries are (up to the symmetry condition) independent centered random variables satisfying Definition IV.1. Let $M = \frac{1}{\sqrt{N}}A + \frac{J}{N}\mathbf{1}\mathbf{1}^T$ where $J > 1$. Let $\lambda_1 \geq \cdots \geq \lambda_N$ be the eigenvalues of M.

[2]Even though it is stated in Lemma 4.2 of [13] that the lemma holds for sufficiently small $\delta > 0$, the proof of it is valid for any $\delta > 0$, and we use $\delta > 2$ for our purpose.

Let $\chi_N = \sqrt{N}(\lambda_1 - \hat{J})$ denoting the rescaled largest eigenvalue. Given an analytic function $\varphi(x)$, recall the partial linear statistics $\mathcal{N}_N^{(2)}(\varphi) = \sum_{i=2}^N \varphi(\lambda_i) - N\int_{-2}^2 \varphi(x)\mathrm{d}\sigma_{scl}(x)$. We saw in the previous sections that χ_N and $\mathcal{N}_N^{(2)}(\varphi)$ converge individually to Gaussian random variables. In this section, we consider the joint distribution and prove Theorem IV.8. In Subsection 4.5.1, we first prove Theorem IV.8 assuming that the disorder variables are Gaussian random variables. In Subsection 4.5.2, the general disorder variables are considered using an interpolation trick.

4.5.1 Asymptotic Independence for the GOE case

Let the off-diagonal entries of A be Gaussian random variables of variance 1 and the diagonal entries be Gaussian random variables of variance 2. In random matrix theory, the random symmetric matrix $H = \frac{1}{\sqrt{N}}A$ is said to belong to the Gaussian orthogonal ensemble (GOE). A special property of GOE, compared with general random symmetric matrices, is that the probability measure of GOE is invariant under orthogonal conjugations.

The following result is basically in [30].

Lemma IV.21. *Let $(\frac{1}{\sqrt{2}}A_{ii}, A_{ij}, y_i)_{1\le i<j\le N}$ be i.i.d. standard Gaussian random variables. Let $H = \frac{1}{\sqrt{N}}A$ with $A = (A_{ij})_{1\le i,j\le N}$ and let $Y = \frac{1}{\sqrt{N}}(y_1, \cdots, y_N)^T$. Define $G(z) = (H - zI)^{-1}$ for $z \in \mathbb{C}\backslash[-2-\delta, 2+\delta]$, which is well defined with high probability for fixed $\delta > 0$. Then, for $z \in \mathbb{R}\backslash[-2-\delta, 2+\delta]$,*

$$n_N(z) := \sqrt{N}(Y^*G(z)Y - \frac{1}{N}\mathrm{Tr}(G(z))) \Rightarrow n(z) \tag{4.120}$$

where $n = n(z) := \mathcal{N}\left(0, 2\int \frac{\mathrm{d}\sigma_{scl}(x)}{(x-z)^2}\right)$ is a Gaussian random variable.

Proof of Lemma IV.21. We follow the idea presented in [30]. By Theorem 5.2 of [30], it is enough to check the following three conditions for G: (i) There exists an N-independent constant a such that $\|G\| \le a$ with high probability, (ii) $\frac{1}{N}\mathrm{Tr}\, G^2$

converges to a constant in probability, and (iii) $\frac{1}{N}\sum_{i=1}^{N} G_{ii}^2$ converges to a constant in probability. They follow from rigidity of eigenvalue (Lemma IV.10), law of large numbers (Corollary IV.11), and local law (Theorem 2.9 of [55]), respectively. □

We are now ready to prove the following property of GOE matrices.

Proposition IV.22. *For H defined in Lemma IV.21, denote its eigenvalues by $\rho_1 \geq \rho_2 \geq \cdots \geq \rho_N$. For fixed k, consider a random vector $(X_N^1, X_N^2, \cdots, X_N^k)$ whose entries are real measurable functions of those eigenvalues, i.e., $X_N^i = X_N^i(\rho_1, \rho_2, \cdots, \rho_N)$ for $i = 1, 2, \cdots, k$. Suppose there is a random vector $(X^i)_{i=1}^k$ such that $(X_N^i)_{i=1}^k \Rightarrow (X^i)_{i=1}^k$ as $N \to \infty$. Then for n_N and n defined as in (4.120), $(X_N^1, X_N^2, \cdots, X_N^k, n_N) \Rightarrow (X^1, X^2 \cdots, X^k, n)$, where n is independent from $(X^1, X^2, \cdots, X^k)$.*

Proof. For the convergence, it is enough to show (i) $(X_N^1, X_N^2 \cdots, X_N^k, n_N)$ is tight, and (ii) convergence of characteristic function. The tightness follows from the tightness of individual random vector (variable), which is a consequence of individual convergence.

For (ii), consider the eigenvalue decomposition $H = OPO^T$, where $P = \mathrm{diag}(\rho_1, \rho_2, \cdots, \rho_N)$ and O is an orthogonal matrix. Since the H is orthogonal invariant, P and O are independent. Set $X = O^T Y$. Then $X = \frac{1}{\sqrt{N}}(x_1, \cdots, x_N)$ where $x_1, \cdots, x_N$ are i.i.d standard Gaussian (X is also independent with P).

Now, $n_N = Y^* G(z) Y - \frac{1}{N}\mathrm{Tr} G(z) = \frac{1}{N}\sum_{i=1}^{N} \frac{x_i^2 - 1}{\rho_i - z}$. Since $\mathbb{E}[e^{tx_1^2}] = \frac{1}{\sqrt{1-2t}}$, we find that for any $t \in \mathrm{i}\mathbb{R}$, the conditional expectation over X given P satisfies

$$\mathbb{E}_X\left[e^{tn_N}\middle|P\right] = \mathbb{E}_X\left[e^{\frac{t}{\sqrt{N}}\sum_{i=1}^{N}\frac{x_i^2-1}{\rho_i - z}}\middle|P\right] = \prod_{i=1}^{N} e^{-\frac{1}{2}\log(1-\frac{2t}{\sqrt{N}(\rho_i - z)}) - \frac{t}{\sqrt{N}(\rho_i - z)}}.$$

Note that $(X_N^1, X_N^2, \cdots, X_N^k)$ only depends on the eigenvalues, and hence it is inde-

pendent of X. Thus, for any $u_1, u_2, \cdots, u_k, t \in \mathrm{i}\mathbb{R}$,

$$\mathbb{E}\left[e^{\sum_{j=1}^k u_j X_N^j + t n_N}\right] = \mathbb{E}\left[e^{\sum_{j=1}^k u_j X_N^j} \prod_{i=1}^N e^{-\frac{1}{2}\log(1-\frac{2t}{\sqrt{N}(\rho_i - z)}) - \frac{t}{\sqrt{N}(\rho_i - z_2)}}\right]. \tag{4.121}$$

Since $-\frac{1}{2}\log(1-2z) - z = z^2 + O(z^3)$ as $z \to 0$, using Corollary IV.11,

$$\begin{aligned}\prod_{i=1}^N e^{-\frac{1}{2}\log(1-\frac{2t}{\sqrt{N}(\rho_i - z)}) - \frac{t}{\sqrt{N}(\rho_i - z)}} &= e^{\frac{1}{N}\sum_{i=1}^N \frac{t^2}{(\rho_i - z)^2} + O(N^{-\frac{1}{2}})} = e^{t^2 \int \frac{1}{(x-z)^2} \mathrm{d}\sigma_{scl}(x) + O(N^{-\frac{1}{2}})} \\ &= \mathbb{E}\left[e^{t n(z)}\right] e^{O(N^{-\frac{1}{2}})}\end{aligned} \tag{4.122}$$

with high probability. Denote this high probability event by Ω_N. Then,

$$\begin{aligned}\lim_{N\to\infty} \mathbb{E}\left[e^{\sum_{j=1}^k u_j X_N^j + t n_N}\right] &= \lim_{N\to\infty} \left(\mathbb{E}\left[e^{\sum_{j=1}^k u_j X_N^j + t n_N} \middle| \Omega_N\right] \mathbb{P}(\Omega_N) + \mathbb{E}\left[e^{\sum_{j=1}^k u_j X_N^j + t n_N} \middle| \Omega_N^c\right] \mathbb{P}(\Omega_N^c)\right) \\ &= \mathbb{E}\left[e^{\sum_{j=1}^k u_j X^j}\right] \mathbb{E}\left[e^{t n(z)}\right],\end{aligned} \tag{4.123}$$

since $t, u_1, u_2, \cdots, u_k \in \mathrm{i}\mathbb{R}$ and hence all exponents are pure imaginary. Note that the characteristic function of $(X^1, \cdots, X^k, n)$ is equal to the product of the characteristic functions of individual random vector (variable). Thus $n(z)$ is independent from $(X^1, \cdots, X^k)$. This completes the proof. □

Corollary IV.23. *Fix $\delta > 0$, consider $z_1 \in \mathbb{C}\backslash\mathbb{R}$ and $z_2 \in \mathbb{R}\backslash[-2-\delta, 2+\delta]$. Recall $s(z)$ defined in* (4.105)*. Then $(\mathrm{Tr}(G(z_1)) - Ns(z_1), n_N(z_2))$ converges in distribution to independent Gaussian random variables.*

Proof. Note that $\mathrm{Tr}(G(z_1)) - Ns(z_1)$ is complex, we consider the random vector $(\Re(\mathrm{Tr}(G(z_1)) - Ns(z_1)), \Im(\mathrm{Tr}(G(z_1)) - Ns(z_1)))$. By Proposition IV.22, it is enough to show that $(\Re(\mathrm{Tr}(G(z_1)) - N(s_1)), \Im(\mathrm{Tr}(G(z_1)) - Ns(z_1))$ converges to a Gaussian random vector. Consider the expression $z_1 = E + i\eta$ for $\epsilon, \eta \in \mathbb{R}$ and $\eta \neq 0$. Recalling

the definition of linear statistics $\mathcal{N}_N(\varphi)$ defined in (4.29), we have

$$\Re(\mathrm{Tr}(G(z_1) - Ns(z_1))) = \mathcal{N}_N(\varphi_r), \qquad \varphi_r(x) = \frac{x-E}{(x-E)^2+\eta^2},$$

and

$$\Im(\mathrm{Tr}(G(z_1) - Ns(z_1)) = \mathcal{N}_N(\varphi_i), \qquad \varphi_i(x) = \frac{\eta}{(x-E)^2+\eta^2}.$$

That is, they are both linear statistics. Then Corollary then follows from Theorem 1.1 of [9]. □

Remark IV.24. When we prove Theorem IV.8 for GOE, we use Proposition IV.22 and Corollary IV.23 with N-dependent z_i. First, for a fixed $z_2 \in \mathbb{R}\backslash[-2-\delta, 2+\delta]$ for some $\delta > 0$, let $\tilde{z}_2 = \tilde{z}_2(N) := \sqrt{\frac{N+1}{N}} z_2$. Using the exactly same argument in the proof of Lemma IV.21, one can show $n_N(\tilde{z}_2) \Rightarrow n(z_2)$. Since the (4.122) still holds for $\tilde{z}_2$ and $n(z_2)$, the asymptotic independence in Proposition IV.22 is still valid, i.e.

$$(X_N^1, X_N^2, \cdots, X_N^k, n_N(\tilde{z}_2)) \Rightarrow (X^1, X^2 \cdots, X^k, n(z_2)),$$

where $n(z_2)$ is independent from $(X^1, X^2, \cdots, X^k)$. Second, for $z_1 \in \mathbb{C}\backslash\mathbb{R}$, consider $\tilde{z}_1 = \tilde{z}_1(N) := \sqrt{\frac{N+1}{N}} z_1$. Notice that

$$\frac{1}{x-\tilde{z}_1} = \frac{1}{x-z_1} + \frac{z_1}{2N(x-z_1)^2} + O(N^{-2}).$$

Then, by the discussion in Remark IV.7, $\mathrm{Tr}(G(\tilde{z}_1)) - Ns(\tilde{z}_1) = \mathcal{N}_N(\frac{1}{x-\tilde{z}_1})$ converges to a Gaussian random variable. Now, putting together, for $\tilde{z}_1$ and $\tilde{z}_2$ defined as above, $(\mathrm{Tr}(G(\tilde{z}_1)) - Ns(\tilde{z}_1), n_N(\tilde{z}_2))$ converge jointly to independent Gaussian random variables.

We now prove Theorem IV.8 for the case where the disorder belongs to GOE.

Proof of Theorem IV.8 when A belongs to GOE. Recall that λ_i are the eigenvalues of

$M = \frac{1}{\sqrt{N}}A + \frac{J}{N}\mathbf{1}\mathbf{1}^T$ with A from the GOE. Since the means and variances follow from [30] and Theorem IV.6, it is enough to prove the asymptotic independence of χ_N and $\mathcal{N}_N^{(2)}(\varphi)$. (Notice that that $W_3 = 0$ for Gaussian A_{ij}.) Now, for any analytic test function φ, the partial linear statistics can be expressed as (see (4.98)) an integral of

$$\xi_N^{(2)}(z) = \sum_{i=2}^{N} \frac{1}{\lambda_i - z} - N \int_{\mathbb{R}} \frac{1}{x - z} \mathrm{d}\sigma_{scl}(x), \quad z \in \mathbb{C}\backslash\mathbb{R}. \tag{4.124}$$

Then according to Lemma IV.19 and what follows, it is enough to prove that χ_N and $\xi_N^{(2)}(z)$ are asymptotically independent for fixed $z \in \mathbb{C} \setminus \mathbb{R}$. Let

$$\xi_N(z) = \xi_N^{(2)}(z) + \frac{1}{\lambda_1 - z} = \mathrm{Tr}(M - zI)^{-1} - Ns(z).$$

Since $\frac{1}{\lambda_1 - z} \to \frac{1}{\hat{J} - z}$ in probability, it is enough to prove that χ_N and $\xi_N(z)$ are asymptotically independent.

Since the GOE is orthogonal invariant, for every deterministic matrix U, the eigenvalues of $A + U$ have the same distribution as $A + OUO^T$ for any orthogonal matrix O. Thus, we may consider the following equivalent model:

$$M = \frac{1}{\sqrt{N}}A + \mathrm{diag}(J, 0, \cdots, 0). \tag{4.125}$$

Following the proof of Theorem 2.2 in [30], we write

$$M = \begin{bmatrix} \frac{A_{11}}{\sqrt{N}} + J & Y^* \\ Y & \hat{M} \end{bmatrix}. \tag{4.126}$$

Since $\det(M - zI) = \det(\hat{M} - zI)\left(\frac{A_{11}}{\sqrt{N}} + J - z - Y^*\hat{G}(z)Y\right)$ with

$$\hat{G}(z) := (\hat{M} - zI_{N-1})^{-1} = (\frac{1}{\sqrt{N}}\hat{A} - zI_{N-1})^{-1}, \tag{4.127}$$

the largest eigenvalue of M satisfies

$$\lambda_1 = J + \frac{A_{11}}{\sqrt{N}} - Y^* \hat{G}(\lambda_1) Y \tag{4.128}$$

if λ_1 is not an eigenvalue of $\hat{M}$, which holds with high probability. Using the resolvent formula twice, we write

$$\begin{aligned}\hat{G}(\lambda_1) &= \hat{G}(\hat{J}) + (\hat{G}(\lambda_1) - \hat{G}(\hat{J})) = \hat{G}(\hat{J}) + (\lambda_1 - \hat{J})\hat{G}(\lambda_1)\hat{G}(\hat{J}) \\ &= \hat{G}(\hat{J}) + (\lambda_1 - \hat{J})\hat{G}(\hat{J})^2 + (\lambda_1 - \hat{J})^2 \hat{G}(\lambda_1)\hat{G}(\hat{J})^2.\end{aligned}$$

Hence,

$$\begin{aligned}\lambda_1 - \hat{J} &= \frac{A_{11}}{\sqrt{N}} - \frac{1}{J} - Y^* \hat{G}(\lambda_1) Y \\ &= \frac{A_{11}}{\sqrt{N}} - \frac{1}{J} - Y^* \hat{G}(\hat{J}) Y + (\lambda_1 - \hat{J}) Y^* \hat{G}(\hat{J})^2 Y + (\lambda_1 - \hat{J})^2 Y^* \hat{G}(\lambda_1)\hat{G}(\hat{J})^2 Y\end{aligned}$$

with high probability. Moving all terms with factor $\lambda_1 - \hat{J}$ to the left and taking it out as a common factor, we arrive at

$$\chi_N = \sqrt{N}(\lambda_1 - \hat{J}) = \frac{A_{11} - \sqrt{N}(\frac{1}{J} + Y^* \hat{G}(\hat{J}) Y)}{1 + Y^* \hat{G}(\hat{J})^2 Y + (\lambda_1 - \hat{J}) Y^* \hat{G}(\lambda_1)\hat{G}(\hat{J})^2 Y} \tag{4.129}$$

with high probability.

Note that $\hat{M}$ and Y satisfy the setting of Corollary IV.23 up to the scaling factor $\sqrt{\frac{N}{N-1}}$. Set

$$\tilde{Y} = \sqrt{\frac{N}{N-1}} Y, \qquad \tilde{G}(z) = \big(\sqrt{\frac{N}{N-1}} \hat{M} - z I_{N-1}\big)^{-1} \tag{4.130}$$

Then, $\tilde{Y}$ and $\tilde{G}$ satisfy the setting of Corollary IV.23, and

$$Y^* \hat{G}(\hat{J}) Y = \sqrt{\frac{N-1}{N}} \tilde{Y}^* \tilde{G}(\tilde{J}) \tilde{Y}, \qquad \tilde{J} := \sqrt{\frac{N}{N-1}} \hat{J}. \tag{4.131}$$

Now, by Corollary IV.11,

$$\frac{1}{N-1}\operatorname{Tr}(\tilde{G}(\tilde{J})) = s(\hat{J}) + O(N^{-1+\epsilon}) = -\frac{1}{J} + O(N^{-1+\epsilon}) \tag{4.132}$$

with high probability. By Lemma IV.21, Corollary IV.11 and Lemma IV.10,

$$Y^*\hat{G}(\hat{J})^2 Y \to \frac{1}{J^2-1}, \qquad (\lambda_1 - \hat{J})Y^*\hat{G}(\lambda_1)\hat{G}(\hat{J})^2 Y \to 0 \tag{4.133}$$

in probability. Using (4.133), (4.132) and denoting the denominator in (4.129) by D_1, we write

$$\chi_N = D_1^{-1}\left(A_{11} - \tilde{n}_{N-1}(\tilde{J}) + O(N^{-\frac{1}{2}+\epsilon})\right), \tag{4.134}$$

where $n_{N-1}(\tilde{J}) = \sqrt{N-1}(\tilde{Y}^*\tilde{G}(\tilde{J})\tilde{Y} - \frac{1}{N-1}\operatorname{Tr}(\tilde{G}(\tilde{J})))$ (see (4.120)) and $D_1 \to \frac{J^2}{J^2-1}$ in probability. Note that A_{11} and $n_{N-1}(\tilde{J})$ are independent, the distribution of χ_N is governed by their convolution.

We now turn to the linear statistic $\xi_N(z)$. Using Schur complement of M with block structure in (4.126), for any $z \in \mathbb{C}\backslash\mathbb{R}$,

$$\operatorname{Tr}(M - zI)^{-1} = (J + \frac{A_{11}}{\sqrt{N}} - z - Y^*\hat{G}(z)Y)^{-1}(1 + Y^*\hat{G}(z)^2 Y) + \operatorname{Tr}(\hat{G}(z)) \tag{4.135}$$

Using Lemma IV.21 and Lemma IV.11,

$$D_2 = D_2(N) := \frac{1 + Y^*\hat{G}(z)^2 Y}{J + \frac{A_{11}}{\sqrt{N}} - z - Y^*\hat{G}(z)Y} \to \frac{1 + s'(z)}{J - z - s(z)}$$

in probability. Then, by setting $\tilde{z} := \tilde{z}(N) = \sqrt{\frac{N}{N-1}}z$, we write

$$\begin{aligned}\xi_N(z) =& \operatorname{Tr}(M - zI)^{-1} - Ns(z) = D_2 + \operatorname{Tr}\hat{G}(z) - Ns(z) + O(N^{-\frac{1}{2}+\epsilon}) \\ =& D_2 - \frac{s(z)}{2} + \frac{zs'(z)}{2} + \sqrt{\frac{N}{N-1}}\left(\operatorname{Tr}\tilde{G}(\tilde{z}) - (N-1)s(\tilde{z})\right) + O(N^{-\frac{1}{2}+\epsilon}).\end{aligned} \tag{4.136}$$

That is, the fluctuation of $\xi_N(z)$ is govern by $\mathrm{Tr}\tilde{G}(\tilde{z}) - (N-1)s(\tilde{z})$. Now using Corollary IV.23 and Remark IV.24, one can conclude that $(\mathrm{Tr}\tilde{G}(\tilde{z}) - (N-1)s(\tilde{z}), n_{N-1}(\tilde{J}))$ converge to independent Gaussian random variables. Furthermore, A_{11} is independent of both Y and $\hat{M}$. Thus by (4.134) and (4.136), $(\xi_N(z), \chi_N)$ converge to independent random variables. Theorem IV.8 then follows. □

4.5.2 Proof of Theorem IV.8 for general case

We prove Theorem IV.8 for general disorders, where the disorder matrix A is a Wigner matrix and satisfies Definition IV.1. Unlike the GOE, Wigner matrices are not orthogonal invariant, hence we cannot apply (4.125) where we replaced the rank-1 perturbation in M by a diagonal matrix. To overcome the difficulty, we use an interpolation method. It has been successfully applied in many works in random matrix theory, where a given matrix and a reference matrix such as GOE are interpolated. We refer to [83] for its application in the analysis of linear eigenvalue statistics.

Let $V = \frac{1}{\sqrt{N}}A$ be a (normalized) Wigner matrix and V^G be a (normalized) GOE matrix independent from V. Define

$$H(t) = V\cos t + V^G \sin t \tag{4.137}$$

so that $H(0) = V$ and $H(\frac{\pi}{2}) = V^G$. Note that $\mathbb{E}[H_{ij}^2] = \frac{1}{N}$ for $i \neq j$. Let

$$\boldsymbol{e} = \frac{1}{\sqrt{N}}\mathbf{1}^T = \frac{1}{\sqrt{N}}(1, 1, \ldots, 1)^T \in \mathbb{R}^N \tag{4.138}$$

and

$$M(t) = H(t) + J\boldsymbol{e}\boldsymbol{e}^T, \tag{4.139}$$

whose eigenvalues are denoted by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. Define the resolvents

$$G(z) = (M - zI)^{-1}, \qquad \hat{G}(z) = (H - zI)^{-1}. \tag{4.140}$$

Here, we omit the dependence on t for the ease of notation. We note that G and $\hat{G}$ are symmetric (not Hermitian). For any (small) fixed $\delta > 0$, $\hat{G}(z)$ is well-defined for $z \in \mathbb{C} \setminus [-2-\delta, 2+\delta]$ with high probability.

For $\chi_N = \sqrt{N}(\lambda_1 - \hat{J})$, we notice that

$$\hat{G}_{\boldsymbol{ee}}(\lambda_1) := \langle \boldsymbol{e}, \hat{G}(\lambda_1)\boldsymbol{e}\rangle = -\frac{1}{J} \tag{4.141}$$

with high probability. The claim holds since

$$\begin{aligned} 0 = \det(M - \lambda_1 I) &= \det(H - \lambda_1 I)\det(I + J\hat{G}(\lambda_1)\boldsymbol{e}\boldsymbol{e}^T) \\ &= \det(H - \lambda_1 I)\det(I + J\boldsymbol{e}^T\hat{G}(\lambda_1)\boldsymbol{e}) = \det(H - \lambda_1 I)\left(1 + J\hat{G}_{\boldsymbol{ee}}(\lambda_1)\right) \end{aligned} \tag{4.142}$$

and λ_1 is not an eigenvalue of H with high probability (See Lemma 6.1 of [77]). Furthermore, by Taylor expansion,

$$-\frac{1}{J} = \hat{G}_{\boldsymbol{ee}}(\lambda_1) = \hat{G}_{\boldsymbol{ee}}(\hat{J}) + \hat{G}'_{\boldsymbol{ee}}(\hat{J})(\lambda_1 - \hat{J}) + O(N^{-1+\epsilon}) \tag{4.143}$$

with high probability, since $|\lambda_1 - \hat{J}| = O(N^{-\frac{1}{2}+\epsilon})$ and $\|\hat{G}''(z)\| = O(1)$ with high probability. From the isotropic local law, Theorem 2.2 of [77], we find that

$$\hat{G}_{\boldsymbol{ee}}(\hat{J}) = s(\hat{J}) + O(N^{-\frac{1}{2}+\epsilon}), \qquad \hat{G}'_{\boldsymbol{ee}}(\hat{J}) = s'(\hat{J}) + O(N^{-\frac{1}{2}+\epsilon}) \tag{4.144}$$

with high probability. Thus, using Lemma IV.21,

$$\chi_N = \sqrt{N}(\lambda_1 - \hat{J}) = -\frac{\sqrt{N}(J^{-1} + \hat{G}_{\boldsymbol{ee}}(\hat{J}))}{s'(\hat{J})} + O(N^{-\frac{1}{2}+2\epsilon}) \tag{4.145}$$

with high probability. That is, the behavior of χ_N is governed by the fluctuation of $\hat{G}_{\boldsymbol{ee}}(\hat{J})$.

To prove the Theorem IV.8, as in the Gaussian disorder case, it is enough to show the convergence of the joint distribution of χ_N and the full linear statistics $\xi_N(z) = \mathrm{Tr}(G(z)) - Ns(z)$ for fixed $z \in \mathbb{C} \setminus \mathbb{R}$. Under the light of (4.145), we set out to calculate the following characteristic function involving $\xi_N(z)$ and $\hat{G}_{\boldsymbol{ee}}(\hat{J})$. Explicitly, for $t_1, t_2, t_3 \in \mathrm{i}\mathbb{R}$ and $z = E + \mathrm{i}\eta$ with $E \in \mathbb{R}$ and $\eta > 0$, we define

$$\mathbb{E}\left[e^{P(t)}\right] := \mathbb{E}\left[e^{t_1 \Re\xi_N + t_2 \Im\xi_N + t_3 n_N}\right], \qquad P(t) := t_1 \Re\xi_N(z) + t_2 \Im\xi_N(z) + t_3 n_N, \tag{4.146}$$

where

$$n_N = \sqrt{N}\left(\hat{G}_{\boldsymbol{ee}}(\hat{J}) + \frac{1}{J}\right). \tag{4.147}$$

Note that n_N is real, the exponent $P(t)$ is pure imaginary and thus $|e^{P(t)}| \leq 1$. For our purpose, it is desired to estimate $\mathbb{E}[e^{P(0)}]$. At $t = \frac{\pi}{2}$, the disorder $H(\frac{\pi}{2})$ reduces to the GOE case. From Subsection 4.5.1, χ_N and ξ_N are asymptotically independent in the GOE case, then

$$\lim_{N\to\infty} \mathbb{E}\left[e^{P(\frac{\pi}{2})}\right] = \mathbb{E}\left[e^{t_1 \Re\xi + t_2 \Im\xi}\right] \cdot \mathbb{E}\left[e^{t_3 n}\right] \tag{4.148}$$

for some Gaussian random variables ξ, n with known mean and variance. Thus, it only remains to estimate the t-derivative of $\mathbb{E}[e^{P(t)}]$. Here, we recall the following identity for the derivative of the resolvent G. For $i, j, a, b = 1, 2, \cdots, N$,

$$\frac{\partial}{\partial M_{ij}} G_{ab} = -\beta_{jk}(G_{aj}G_{kb} + G_{ak}G_{jb}) \tag{4.149}$$

with

$$\beta_{jk} = \begin{cases} 1 & j \neq k, \\ 1/2 & j = k. \end{cases} \tag{4.150}$$

We note that the above identity also holds if one replace G by $\hat{G}$. Thus for any fixed

event Ω,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}\left[e^{P(t)}|\Omega\right] &= \mathbb{E}\left[\sum_{i\le j}\frac{\mathrm{d}M_{ij}}{\mathrm{d}t}\frac{\partial}{\partial M_{ij}}e^{P(t)}\middle|\Omega\right] \\ &= \sum_{i,j}\mathbb{E}\left[\left(V_{ij}\sin t - V_{ij}^{G}\cos t\right)\left(t_1\Re\left(G^2\right)_{ij} + t_2\Im\left(G^2\right)_{ij} + \frac{t_3}{\sqrt{N}}\sum_{p,q}\hat{G}_{pi}\hat{G}_{jq}\right)e^{P(t)}\middle|\Omega\right].\end{aligned} \tag{4.151}$$

The reason for the introduction of Ω will be revealed after the proof of Corollary IV.27. The right hand side of (4.151) motivates us to apply the generalized Stein's lemma. More precisely, we will use Proposition 3.1 of [83] with a small modification as follows:

Proposition IV.25. *Given an event Ω, let X be a random variable such that $\mathbb{E}[|X|^{p+2}|\Omega] < \infty$ for a certain non-negative integer p. Denote the conditional cumulants of X by $\kappa_l := \kappa_l(\Omega)$, $l = 1, \ldots, p+1$. Then for any function $\Phi : \mathbb{R} \to \mathbb{C}$ of the class C^{p+1} with bounded derivatives $\Phi^{(l)}, l = 1, \ldots, p+1$, we have*

$$\mathbb{E}[X\Phi(X)|\Omega] = \sum_{l=0}^{p}\frac{\kappa_{l+1}}{l!}\,\mathbb{E}[\Phi^{(l)}(X)|\Omega] + \epsilon_p, \tag{4.152}$$

where the remainder term ϵ_p admits the bound

$$|\epsilon_p| \le C_p\,\mathbb{E}\left[|X|^{p+2}\left(1 + \max_{1\le j\le p+1}\left(\int_0^1 |\Phi^{(p+1)}(vX)|\mathrm{d}v\right)^{\frac{p+2}{j}}\right)\middle|\Omega\right] \tag{4.153}$$

for some constant C_p that depends only on p.

Proof. We basically follow the proof of Proposition 3.1 of [83]. Let π_p be the degree p Taylor polynomial of Φ and let $r_p = \Phi - \pi_p$. Then, as in the proof of Proposition 3.1 of [83],

$$\mathbb{E}[X\pi_p(X)|\Omega] = \sum_{j=0}^{p}\frac{\kappa_{j+1}}{j!}\,\mathbb{E}[\pi_p^{(j)}(X)|\Omega]. \tag{4.154}$$

Thus

$$\left|\mathbb{E}[X\Phi(X)|\Omega] - \sum_{l=0}^{p} \frac{\kappa_{l+1}}{l!}\,\mathbb{E}[\Phi^{(l)}(X)|\Omega]\right| \le |\mathbb{E}[Xr_p(X)|\Omega]| + \sum_{l=0}^{p} \frac{|\kappa_{l+1}|}{l!}\left|\mathbb{E}\left[r_p^{(l)}(X)|\Omega\right]\right|. \tag{4.155}$$

Since

$$r_p(X) = \frac{X^{p+1}}{p!}\int_0^1 \Phi^{(p+1)}(vX)(1-v)^p \mathrm{d}v, \tag{4.156}$$

by the estimate $|\kappa_j| \le (2j)^j\,\mathbb{E}[|X|^j|\Omega]$ and Hölder's inequality,

$$\begin{aligned}\sum_{l=0}^{p} \frac{|\kappa_{l+1}|}{l!}\left|\mathbb{E}\left[r_p^{(l)}(X)|\Omega\right]\right| &\le \sum_{l=0}^{p} \frac{\kappa_{l+1}}{l!(p-l)!}\,\mathbb{E}\left[|X|^{p+1-l}\int_0^1 |\Phi^{(p+1)}(vX)|\mathrm{d}v\Big|\Omega\right] \\ &\le \sum_{l=0}^{p} \frac{(2l+2)^{l+1}}{l!(p-l)!}\,\mathbb{E}\left[|X|^{p+2}\left(1+\left(\int_0^1 |\Phi^{(p+1)}(vX)|\mathrm{d}v\right)^{\frac{p+2}{p+1-l}}\right)\Big|\Omega\right].\end{aligned} \tag{4.157}$$

As $|\mathbb{E}[Xr_p|\Omega]|$ can also be bounded by the right hand side of (4.157), the proof is complete.

□

In order to apply Proposition IV.25 to (4.151), we need prior bounds of $P(t)$ and its derivatives to bound ϵ_p in (4.152). As we will see later, it is enough to bound G_{ij}, $(G^2)_{ij}$, $\hat{G}_{ij}$ and $\sum_p \hat{G}_{ip}$. In the following, we are going to introduce a high probability event Ω, on which we have the desired bounds.

With the trivial bound $\|G\| \le \frac{1}{\eta}$ (recall that $z = E + i\eta$), we have that $|G_{ij}| \le \frac{1}{\eta}$ and $|(G^2)_{ij}| \le \|G^2\| \le \frac{1}{\eta^2}$. For $\hat{G}_{ij}$, we introduce the high probability event $\Omega_1 = \{\lambda_1 \le (2+\hat{J})/2\}$. It is easy to check that $\|\hat{G}\|\mathbb{1}_{\Omega_1} \le \frac{1}{\hat{J}-2}$ and thus

$$|\hat{G}_{ij}\mathbb{1}_{\Omega_1}| \le \frac{1}{\hat{J}-2}, \tag{4.158}$$

For $\sum_p \hat{G}_{ip}$, we recall the following concentration theorem for the quadratic function

of $\hat{G}$:

Proposition IV.26 (Theorem 2.3 and Remark 2.4 of [77]). *Fix $\Sigma \geq 3$. Set $\varphi = (\log N)^{\log\log N}$. Then there exist constants C_1 and C_2 such that for any*

$$E \in [\Sigma, -2 - \varphi^{C_1} N^{-\frac{2}{3}}] \cup [2 + \varphi^{C_1} N^{-\frac{2}{3}}, \Sigma],$$

and any $\eta \in (0, \Sigma]$, and any deterministic $v, w \in \mathbb{C}^N$,

$$|\langle v, \hat{G}(z) w\rangle - s(z)\langle v, w\rangle| \leq \varphi^{C_2} \sqrt{\frac{\Im s(z)}{N\eta}}\, \|v\|\, \|w\| \tag{4.159}$$

with high probability, uniformly on $z = E + \mathrm{i}\eta$.

Let $\boldsymbol{e}_i := (0, \cdots, 1, \cdots, 0)$. Noting that $\sum_{p=1}^N \hat{G}_{pi} = \sqrt{N}\langle \boldsymbol{e}, \hat{G}\boldsymbol{e}_i\rangle$, we can derive a prior bound for $\sum_{p=1}^N \hat{G}_{pi}$, which is summarized in the following Corollary.

Corollary IV.27. *For any fixed $E \in \mathbb{R}\backslash[-2, 2]$, the tail bound*

$$|\sum_p (\hat{G}(E))_{pi}| \leq N^{\epsilon} \tag{4.160}$$

holds simultaneously for $i = 1, \cdots, N$ with high probability. We also have that

$$|\langle v, \hat{G}(E) w\rangle - s(E)\langle v, w\rangle| \leq \|v\|\, \|w\|\, N^{-\frac{1}{2}+\epsilon} \tag{4.161}$$

with high probability.

Proof. We first prove (4.161). Consider $z = E + \mathrm{i}N^{-1/2}$. Using Proposition IV.26, we

find there exists some $C > 0$ such that

$$\begin{aligned}\langle v, \hat{G}(E)w\rangle - s(E)\langle v, w\rangle| &\leq |\langle v, (\hat{G}(z) - \hat{G}(E)w\rangle| + |\langle v, \hat{G}(z)w\rangle - s(z)\langle v, w\rangle| \\ &\quad + |s(z) - s(E)||\langle v, w\rangle| \\ &\leq CN^{-1/2}\|v\|\|w\| + C\varphi^C N^{-\frac{1}{2}}\|v\|\|w\| + CN^{-1/2}.\end{aligned} \tag{4.162}$$

Here we also use the fact that Ω_1 holds with high probability. Since $\varphi \ll N^\epsilon$, (4.161) then follows. The tail bound (4.160) can be obtained from (4.161) by setting $v = \sqrt{N}\boldsymbol{e}$ and $w = \boldsymbol{e}_i$. □

Based on our discussion above, we are ready to introduce the high probability event as promised. Set $s_1 := s(z)$, $s_1' := s'(z)$ and $s_2 := s(\hat{J}) = -J^{-1}$, the desired high probability event Ω is the intersection of Ω_1 and the following events:

$$\Omega_2 = \{|\sum_p (\hat{G}(\hat{J}))_{pi}| \leq N^\epsilon, \quad \forall i = 1, \cdots, N\} \cap \{|\hat{G}_{\boldsymbol{ee}}(\hat{J}) - s_2| \leq N^{-\frac{1}{2}+\epsilon}\}, \tag{4.163}$$

$$\Omega_3 = \{|\hat{G}_{ij} - \delta_{ij}s_2| \leq N^{-\frac{1}{2}+\epsilon}, \quad \forall i, j = 1, \cdots, N\}, \tag{4.164}$$

$$\Omega_4 = \{|G_{ij} - \delta_{ij}s_1|, |(G^2)_{ij} - \delta_{ij}s_1'| \leq N^{-\frac{1}{2}+\epsilon}, \quad \forall i, j = 1, \cdots, N\}, \tag{4.165}$$

$$\Omega_5 = \{|V_{ij}|, |V_{ij}^G|, |M_{ij}| \leq N^{-\frac{1}{2}+\epsilon}, \quad \forall i, j = 1, \cdots, N\}. \tag{4.166}$$

Here, by Corollary IV.27, Ω_2 is a high probability event. The fact that Ω_3 and Ω_4 are high probability events can be checked from Theorem 2.8 and Theorem 2.9 of [55]. It is easy to check that Ω_5 is a high probability event from the existence of all moments. Furthermore, by the Lipshitz continuity of the resolvents $\hat{G}(z;t)$ w.r.t to t, we also find that Ω holds uniformly on t with high probability.

Applying Proposition IV.25 to Equation (4.151) conditioning on Ω, we claim

$$\begin{aligned}
&\sum_{i,j} \mathbb{E}\left[V_{ij}\left(t_1\Re(G^2)_{ij} + t_2\Im(G^2)_{ij} + \frac{t_3}{\sqrt{N}}\sum_{p,q}\hat{G}_{pi}\hat{G}_{jq}\right)e^P\Big|\Omega\right] \\
&= \sum_{l=1}^{3}\cos^l t\sum_{i,j}\frac{\kappa_{l+1}^{V_{ij}}}{l!}\,\mathbb{E}\left[\left(\frac{\partial}{\partial M_{ij}}\right)^l\left(\left(t_1\Re(G^2)_{ij} + t_2\Im(G^2)_{ij} + \frac{t_3}{\sqrt{N}}\sum_{p,q}\hat{G}_{pi}\hat{G}_{jq}\right)e^P\right)\Big|\Omega\right] \\
&\quad + O(N^{-\frac{1}{2}+\epsilon})
\end{aligned} \tag{4.167}$$

where $\kappa_l^{V_{ij}}$ denotes the l-th cumulant of V_{ij}. Here, it is justified to replace the conditional cumulants by $\kappa_l^{V_{ij}}$, since Ω is a high probability event.

To prove the claim, we begin by controlling the remainder term ϵ_p in (4.152). On Ω, $G_{ij}, \hat{G}_{ij}$ and $(G^2)_{ij}$ are $O(1)$, and

$$N^{-\frac{1}{2}}\sum_{p,q}\hat{G}_{pi}\hat{G}_{jq} = N^{-\frac{1}{2}}\left(\sum_p \hat{G}_{pi}\right)\left(\sum_q \hat{G}_{qj}\right) = O(N^{-\frac{1}{2}+\epsilon}).$$

Thus, $\frac{\partial}{\partial M_{ij}}P = O(1)$ on Ω. From the resolvent identity and the definition of event Ω, we find $\|G(z;vV_{ij}) - G(z;V_{ij})\| = O(N^{-\frac{1}{2}+\epsilon})$ for $0 \le v \le 1$. Thus on Ω, $\frac{\partial}{\partial M_{ij}}P(t;vV_{ij}) = O(1)$ for $0 \le v \le 1$. Furthermore, we notice that

$$\frac{\partial}{\partial M_{ij}}(G^2)_{ij} = \frac{\partial}{\partial M_{ij}}\sum_k G_{ki}G_{jk} = -\beta_{ij}\left(2G_{ij}(G^2)_{ij} + G_{ii}(G^2)_{jj} + G_{jj}(G^2)_{ii}\right), \tag{4.168}$$

and

$$\frac{\partial}{\partial M_{ij}}\sum_p \hat{G}_{pi} = -\beta_{ij}\left(\hat{G}_{ji}\sum_p \hat{G}_{pi} + \hat{G}_{ii}\sum_p \hat{G}_{pj}\right). \tag{4.169}$$

Thus we can obtain similar estimates for higher derivatives of P. Since $V_{ij}^5 = O(N^{-\frac{5}{2}+5\epsilon})$ on Ω_5, we find that

$$|V_{ij}|^5\left(1 + \max_{1\le j\le 5}\left(\int_0^1\left|\left(\frac{\partial}{\partial M_{ij}}\right)^5 P\right|\mathrm{d}v\right)^{\frac{5}{j}}\right) \le CN^{-\frac{5}{2}+C\epsilon} \tag{4.170}$$

on Ω. That is, $\epsilon_3 \leq CN^{-\frac{5}{2}+C\epsilon}$, and after summing over i, j, the claim (4.167) is proved.

We next consider the term in (4.151) containing V^G. Noting that the cumulants of order higher than 2 vanish for Gaussian random variables, it reduces to

$$\begin{aligned}
&\sum_{i,j} \mathbb{E}\left[V_{ij}^G \left(t_1 \Re(G^2)_{ij} + t_2 \Im(G^2)_{ij} + \frac{t_3}{\sqrt{N}} \sum_{p,q} \hat{G}_{pi}\hat{G}_{jq}\right) e^P \Big| \Omega\right] \\
=&(\sin t) \sum_{i,j} \kappa_2^{V_{ij}^G} \mathbb{E}\left[\frac{\partial}{\partial M_{ij}} \left(\left(t_1 \Re(G^2)_{ij} + t_2 \Im(G^2)_{ij} + \frac{t_3}{\sqrt{N}} \sum_{p,q} \hat{G}_{pi}\hat{G}_{jq}\right) e^P\right) \Big| \Omega\right] + O(N^{-\frac{1}{2}+\epsilon}),
\end{aligned} \tag{4.171}$$

where $\kappa_2^{V_{ij}^G}$ denotes the second cumulant of V_{ij}^G. We now put (4.167) and (4.171) into (4.151) conditioning on Ω. This yields

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[e^{P(t)} \big| \Omega\right] = (\sin t) \sum_{l=1}^{3} (\cos^l t) I_l - (\cos t \sin t) I_1^G + O(N^{-\frac{1}{2}+\epsilon}), \tag{4.172}$$

where we define

$$I_l = \sum_{i,j} \frac{\kappa_{l+1}^{V_{ij}}}{l!} \mathbb{E}\left[\left(\frac{\partial}{\partial M_{ij}}\right)^l \left(\left(t_1 \Re(G^2)_{ij} + t_2 \Im(G^2)_{ij} + \frac{t_3}{\sqrt{N}} \sum_{p,q} \hat{G}_{pi}\hat{G}_{jq}\right) e^P\right) \Big| \Omega\right] \tag{4.173}$$

and

$$I_1^G = \sum_{i,j} \kappa_2^{V_{ij}^G} \mathbb{E}\left[\frac{\partial}{\partial M_{ij}} \left(\left(t_1 \Re(G^2)_{ij} + t_2 \Im(G^2)_{ij} + \frac{t_3}{\sqrt{N}} \sum_{p,q} \hat{G}_{pi}\hat{G}_{jq}\right) e^P\right) \Big| \Omega\right]. \tag{4.174}$$

In the following, we will evaluate I_l for $l = 1, 2, 3$ separately. We may omit the conditioning on Ω for the ease of notation.

4.5.2.1 Estimate for $I_1 - I_1^G$

Since $\kappa_2^{V_{ij}^G} = \kappa_2^{V_{ij}} = \frac{1}{N}$ for $i \neq j$, we only need to consider the contribution from the diagonal entries to $I_1 - I_1^G$. By the definition of I_1 and I_1^G,

$$I_1 - I_1^G = \sum_i (\kappa_2^{V_{ii}} - \kappa_2^{V_{ii}^G}) \, \mathbb{E}\left[\frac{\partial}{\partial M_{ii}} \left(\left(t_1 \Re (G^2)_{ii} + t_2 \Im (G^2)_{ii} + \frac{t_3}{\sqrt{N}} \sum_{p,q} \hat{G}_{pi} \hat{G}_{iq} \right) e^P \right) \right]. \tag{4.175}$$

From (4.169), we find that

$$\frac{t_3}{\sqrt{N}} \frac{\partial}{\partial M_{ii}} \sum_{p,q} \hat{G}_{pi} \hat{G}_{iq} = O(N^{-\frac{1}{2}+\epsilon}).$$

Similarly, it can be checked that all terms in the right-hand side of (4.175) involving $\hat{G}$ are $O(N^{-\frac{1}{2}+\epsilon})$. Collecting the terms of order 1 only, we obtain that

$$\begin{aligned} I_1 - I_1^G = \frac{1}{N} \sum_i (w_2 - 2) \, \mathbb{E}\left[\left(2t_1 \Re \left((G^2)_{ii} G_{ii} \right) + 2t_2 \Im \left((G^2)_{ii} G_{ii} \right) + \left(t_1 \Re (G^2)_{ii} + t_2 \Im (G^2)_{ii} \right)^2 \right) e^P \right] \\ + O(N^{-\frac{1}{2}+\epsilon}). \end{aligned} \tag{4.176}$$

Using the estimate $|G_{ij} - \delta_{ij} s_1|, |(G^2)_{ij} - \delta_{ij} s_1'| \leq N^{-\frac{1}{2}+\epsilon}$ on Ω_4, we conclude that

$$I_1 - I_1^G = (w_2 - 2) \left(2t_1 \Re (s_1' s_1) + 2t_2 \Im (s_1' s_1) + \left(t_1 \Re (s_1') + t_2 \Im (s_1') \right)^2 \right) \mathbb{E}\left[e^P \right] + O(N^{-\frac{1}{2}+\epsilon}). \tag{4.177}$$

4.5.2.2 Estimate for I_2

We decompose I_2 into

$$\begin{aligned} I_2 &= \sum_{i,j} \frac{W_3}{2N^{\frac{3}{2}}} \mathbb{E}\left[\left(\frac{\partial}{\partial M_{ij}} \right)^2 \left(\left(t_1 \Re (G^2)_{ij} + t_2 \Im (G^2)_{ij} + \frac{t_3}{\sqrt{N}} \sum_{p,q} \hat{G}_{pi} \hat{G}_{jq} \right) e^P \right) \right] \\ &:= I_{2,0} + 2I_{1,1} + I_{0,2}, \end{aligned} \tag{4.178}$$

where

$$I_{r,2-r} := \sum_{i,j} \frac{W_3}{2N^{\frac{3}{2}}} \mathbb{E}\left[\left(\frac{\partial}{\partial M_{ij}}\right)^r \left(t_1 \Re(G^2)_{ij} + t_2 \Im(G^2)_{ij} + \frac{t_3}{\sqrt{N}} \sum_{p,q} \hat{G}_{pi}\hat{G}_{jq} \right) \cdot \left(\frac{\partial}{\partial M_{ij}}\right)^{2-r} e^P \right]. \tag{4.179}$$

We first consider the case $i \neq j$ in the summand of $I_{r,2-r}$ for $r = 0, 1, 2$. Recall that all terms of $O(N^{-\frac{1}{2}+\epsilon})$ are negligible in the sense that they can be absorbed into the error term in the right-hand side of (4.172).

- For $I_{2,0}$, we note that the terms arising from the derivatives of the G^2 are negligible, which can be checked by following the argument in the proof of Theorem 3.3 in [83], especially the estimate of T_3 in (3.53) of [83]. For example, one of such terms is bounded by

$$\left| N^{-\frac{3}{2}} \sum_{i,j} \frac{W_3}{2} \mathbb{E}\left[t_1 \Re(G_{ii}G_{jj}(G^2)_{ij}) e^P \right] \right| \leq \frac{C}{\eta^4\sqrt{N}}. \tag{4.180}$$

  To prove it, we consider a vector $\boldsymbol{u} = (G_{11}, G_{22}, \ldots, G_{NN})$ and proceed as

$$\left| \sum_{i,j} G_{ii}G_{jj}(G^2)_{ij} \right| = \left| \langle \overline{\boldsymbol{u}}, G^2\boldsymbol{u} \rangle \right| \leq \|G^2\|\|\boldsymbol{u}\|^2 \leq N\|G^2\|\|G\|^2 \leq \frac{N}{\eta^4}.$$

  On the other hand,

$$\begin{aligned} \left(\frac{\partial}{\partial M_{ij}}\right)^2 \hat{G}_{pi}\hat{G}_{jq} = 6(\hat{G}_{pi}\hat{G}_{ji}^2\hat{G}_{jq} + \hat{G}_{pj}\hat{G}_{ii}\hat{G}_{ji}\hat{G}_{jq} + \hat{G}_{pi}\hat{G}_{ji}\hat{G}_{jj}\hat{G}_{iq}) \\ + \hat{G}_{ii}\hat{G}_{jj}(4\hat{G}_{pi}\hat{G}_{jq} + 2\hat{G}_{pj}\hat{G}_{iq}). \end{aligned} \tag{4.181}$$

  From the estimate $|\hat{G}_{ij} - \delta_{ij}s_2| \leq N^{-\frac{1}{2}+\epsilon}$ on Ω_3 the concentration of $\hat{G}_{\boldsymbol{ee}}$ on Ω_2,

we then claim that

$$\begin{aligned} I_{2,0} =& \frac{W_3 t_3}{2N^2} \sum_{i,j} \mathbb{E}\left[6\hat{G}_{ii}\hat{G}_{jj}(\sum_p \hat{G}_{pi})(\sum_q \hat{G}_{qj}) e^P \right] + O(N^{-\frac{1}{2}+\epsilon}) \\ =& 3W_3 t_3 s_2^2 \, \mathbb{E}\left[\hat{G}^2_{\boldsymbol{ee}} e^P \right] + O(N^{-\frac{1}{2}+\epsilon}) = 3W_3 t_3 s_2^4 \, \mathbb{E}[e^P] + O(N^{-\frac{1}{2}+\epsilon}). \end{aligned} \tag{4.182}$$

All the other terms in $I_{2,0}$ arising from $\left(\frac{\partial}{\partial M_{ij}}\right)^2 \sum_{p,q} \hat{G}_{pi}\hat{G}_{jq}$ are negligible. One of such terms is bounded by

$$\begin{aligned} & \left| \frac{W_3 t_3}{2N^2} \sum_{i,j} \mathbb{E}\left[(\sum_p \hat{G}_{pj})\hat{G}_{ii}\hat{G}_{ji}(\sum_q \hat{G}_{jq}) e^P \right] \right| \\ & \leq \frac{2|W_3||t_3|}{(\hat{J}-2)N^{\frac{5}{2}-3\epsilon}} \sum_{i,j} \mathbb{E}\left[|e^P| \right] = O(N^{-\frac{1}{2}+3\epsilon}) \end{aligned} \tag{4.183}$$

where we use the definitions of Ω_1, Ω_2 and Ω_3.

- For $I_{1,1}$, the estimates for the negligible terms can be done by using the argument similar to (4.183) and (4.180). The remaining $O(1)$-terms are

$$\begin{aligned} \frac{W_3 t_3}{N^2} \sum_{i,j} \mathbb{E}\Big[\sum_{p,q} \hat{G}_{pi}\hat{G}_{jq} \big(& t_1 \Re\left(G_{ii}(G^2)_{jj} + G_{jj}(G^2)_{ii} \right) \\ & + t_2 \Im\left(G_{ii}(G^2)_{jj} + G_{jj}(G^2)_{ii} \right) \big) e^P \Big]. \end{aligned}$$

Using the definitions of Ω_2 and Ω_4, we write

$$\begin{aligned} I_{1,1} &= 2W_3 t_3 \left(t_1 \Re(s_1 s_1') + t_2 \Im(s_1 s_1') \right) \mathbb{E}\left[\hat{G}^2_{\boldsymbol{ee}} e^P \right] + O(N^{-\frac{1}{2}+\epsilon}) \\ &= 2W_3 t_3 \left(t_1 \Re(s_1 s_1') + t_2 \Im(s_1 s_1') \right) s_2^2 \, \mathbb{E}\left[e^P \right] + O(N^{-\frac{1}{2}+\epsilon}). \end{aligned} \tag{4.184}$$

- For $I_{0,2}$, from the same analysis as for $I_{1,1}$,

$$I_{0,2} = 2W_3 t_3 \left(t_1 \Re(s_1 s_1') + t_2 \Im(s_1 s_1') \right) s_2^2 \, \mathbb{E}\left[e^P \right] + O(N^{-\frac{1}{2}+\epsilon}). \tag{4.185}$$

Again, the estimate can be done in a similar manner.

For the case $i = j$, since there are only N terms in the summation in I_2, all terms are negligible due to the priori bounds on $\|G\|$ and $\sum_p \hat{G}_{pi}$.

Collecting the terms in (4.182), (4.184), and (4.185), we get

$$\begin{aligned}
&\sum_{i,j} \frac{\kappa_3^{V_{ij}}}{2!} \mathbb{E}\left[\left(\frac{\partial}{\partial M_{ij}}\right)^2 \left(\left(t_1 \Re(G^2)_{ij} + t_2 \Im(G^2)_{ij} + \frac{t_3}{\sqrt{N}} \sum_{p,q} \hat{G}_{pi}\hat{G}_{jq}\right) e^P\right)\right] \\
=&W_3 \left[3t_3 s_2^4 + 6t_1 t_3 \Re(s_1 s_1')(s_2)^2 + 6t_2 t_3 \Im(s_1 s_1')(s_2)^2\right] \mathbb{E}\left[e^P\right] + O(N^{-\frac{1}{2}+\epsilon}).
\end{aligned} \tag{4.186}$$

4.5.2.3 Estimate for I_3

Note that any term in I_3 involving $\hat{G}$ is negligible due to the extra $N^{-\frac{1}{2}}$ factor. Estimating as in the previous subsection, we obtain that

$$\begin{aligned}
I_3 =& \sum \frac{\kappa_4^{V_{ij}}}{3!} \mathbb{E}\left[\left(\frac{\partial}{\partial M_{ij}}\right)^3 \left(\left(t_1 \Re(G^2)_{ij} + t_2 \Im(G^2)_{ij} + \frac{t_3}{\sqrt{N}} \sum_{p,q} \hat{G}_{pi}\hat{G}_{jq}\right) e^P\right)\right] \\
=& -4(W_4 - 3)\left[t_1 \Re(s_1^3 s_1') + t_2 \Im(s_1^3 s_1') + \left(t_1 \Re(s_1 s_1') + t_2 \Im(s_1 s_1')\right)^2\right] \mathbb{E}\left[e^P\right] \\
&+ O(N^{-\frac{1}{2}+\epsilon})
\end{aligned} \tag{4.187}$$

We remark that $O(1)$-terms in I_3 contribute only to the corrections of linear statistics.

4.5.2.4 Proof of Theorem IV.8 for general case

Let

$$\begin{aligned}
\tilde{P}(t) =& P(t) - (W_2 - 2)(\cos t)^2 \left(t_1 \Re(s_1' s_1) + t_2 \Im(s_1' s_1) + \frac{1}{2}\left(t_1 \Re(s_1') + t_2 \Im(s_1')\right)^2\right) \\
&+ W_3 (\cos t)^3 \left(t_3 s_2^4 + 2t_1 t_3 \Re(s_1 s_1') s_2^2 + 2t_2 t_3 \Im(s_1 s_1') s_2^2\right) \\
&- (W_4 - 3)(\cos t)^4 \left(t_1 \Re(s_1^3 s_1') + t_2 \Im(s_1^3 s_1') + \left(t_1 \Re(s_1 s_1') + t_2 \Im(s_1 s_1')\right)^2\right).
\end{aligned} \tag{4.188}$$

Then, substituting (4.177), (4.186), and (4.187) into (4.167), we find that

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}[e^{\tilde{P}}|\Omega] = O(N^{-\frac{1}{2}+\epsilon}), \tag{4.189}$$

which implies that

$$\mathbb{E}[e^{\tilde{P}(0)}|\Omega] = \mathbb{E}[e^{\tilde{P}(\frac{\pi}{2})}|\Omega] + O(N^{-\frac{1}{2}+\epsilon}). \tag{4.190}$$

Thus,

$$\begin{aligned}\lim_{N\to\infty}\mathbb{E}\left[e^{P}(0)\right] =& \lim_{N\to\infty}\left(\mathbb{E}\left[e^{P(0)}|\Omega\right]\mathbb{P}(\Omega) + \mathbb{E}\left[e^{P(0)}|\Omega^c\right]\mathbb{P}(\Omega^c)\right)\\ =& e^{P(0)-\tilde{P}(0)}\lim_{N\to\infty}\mathbb{E}[e^{\tilde{P}(0)}|\Omega] = e^{P(0)-\tilde{P}(0)}\lim_{N\to\infty}\mathbb{E}[e^{P(\frac{\pi}{2})}].\end{aligned} \tag{4.191}$$

Here we use the fact that Ω holds with high probability and $\tilde{P}(\frac{\pi}{2}) = P(\frac{\pi}{2})$. We can now conclude that $(\Re\xi_N(z), \Im\xi_N(z), n_N)$ converges to a multivariate Gaussian vector in distribution as $N \to \infty$. By direct calculation, we also find that

$$\begin{pmatrix}\xi_N(z)\\ n_N\end{pmatrix} \Rightarrow \mathcal{N}\left(\begin{pmatrix}b(z)\\ -W_3 s_2^4\end{pmatrix}, \begin{pmatrix}V(z_1) & -2W_3 s_1 s_1' s_2^2\\ -2W_3 s_1 s_1' s_2^2 & \frac{2}{J^2(J^2-1)}\end{pmatrix}\right) \tag{4.192}$$

with $b(z)$ and $V(z)$ are defined in Lemma IV.20. Now, using (4.145), we arrive at

$$\begin{pmatrix}\xi_N(z)\\ \chi_N\end{pmatrix} \Rightarrow \mathcal{N}\left(\begin{pmatrix}b(z)\\ \frac{W_3}{J^2}(1-\frac{1}{J^2})\end{pmatrix}, \begin{pmatrix}V(z_1) & 2W_3 s_1 s_1'(1-\frac{1}{J^2})\\ 2W_3 s_1 s_1'(1-\frac{1}{J^2}) & 2(1-\frac{1}{J^2})\end{pmatrix}\right). \tag{4.193}$$

Hence, the asymptotic Gaussianity of $(\mathcal{N}_N^{(2)}(\varphi), \chi_N)$ follows. For (4.36) and (4.37), the mean and the variance of $\mathcal{N}^{(2)}[\varphi]$ is given in Theorem IV.6. The limiting covariance is given by

$$-2W_3(1-\frac{1}{J^2})\oint_\Gamma \varphi(z)s(z)s'(z)\frac{\mathrm{d}z}{2\pi\mathrm{i}} = 2W_3(1-\frac{1}{J^2})\tau_1(\varphi). \tag{4.194}$$

where we use the change of variables $z \mapsto s$ mapping $\mathbb{C} \setminus [-2, 2]$ to the disk $|s| < 1$ with $s + \frac{1}{s} = -z$ and (4.16) in [13]. This completes the proof of Theorem IV.8 for general case.

4.6 Matching

In the transitional regime, we took $2\beta = \frac{1}{J} + \frac{B}{\sqrt{N}}$. The ferromagnetic regime and the paramagnetic regime correspond to the limiting cases $2\beta > J$ and $2\beta < J$, respectively. In this section, we will consider formal limits $B \to \pm\infty$ of the formula given in the main result, Theorem IV.5, and check the consistency with the results for ferromagnetic and paramagnetic regimes obtained in [13].

Theorem IV.5 states that the free energy F_N is close to the random variable

$$\mathcal{F}_N^{\text{tran}} := \frac{1}{4J^2} + \frac{B}{2J\sqrt{N}} + \frac{\log N}{4N} + \frac{B^2 J^2}{4N} + \frac{1}{N}\mathcal{G}_1 + \frac{1}{N}Q(\mathcal{G}_2) \tag{4.195}$$

in an appropriate sense. Here, $(\mathcal{G}_1, \mathcal{G}_2)$ is a Gaussian vector independent of B. The function $Q(x)$ is given by (4.19). In ferromagnetic and paramagnetic regimes, [13] shows that the free energy is close to

$$\mathcal{F}_N^{\text{ferro}} := \beta\big(J + \frac{1}{J}\big) - \frac{1}{2}\log(2\beta J) - \frac{1}{4J^2} - \frac{1}{2} + \frac{\beta - \frac{1}{2J}}{\sqrt{N}}\mathcal{N}(f_2', \alpha_2') \tag{4.196}$$

and

$$\mathcal{F}_N^{\text{para}} := \beta^2 + \frac{1}{N}\mathcal{N}(f_1, \alpha_1), \tag{4.197}$$

respectively, where $\mathcal{N}(f, \alpha)$ denotes a Gaussian distribution of mean f and variance α. The parameters for the Gaussians are (see [13, (4)])

$$\begin{aligned} f_2' &= W_3(J^{-2} - J^{-4}), \\ \alpha_2' &= 2(1 - J^{-2}) \end{aligned} \tag{4.198}$$

and (see (1.11) and (1.12) of [13]; we set $J' = J$)

$$\begin{aligned} f_1 &= \frac{1}{4}\log(1-4\beta^2) + \beta^2(w_2-2) + 2\beta^4(W_4-3) - \frac{1}{2}\log(1-2\beta J), \\ \alpha_1 &= -\frac{1}{2}\log(1-4\beta^2) + \beta^2(w_2-2) + 2\beta^4(W_4-3). \end{aligned} \tag{4.199}$$

The function $Q(x)$ in (4.195) is given by

$$Q(x) = \frac{s(x)}{2(s(x)-x)} - \frac{s(x)^2}{4(J^2-1)} + \frac{\log(s(x)-x)}{2} + \log \mathbf{I}\left(\frac{(s(x)-x)^2}{J^2-1}\right) \tag{4.200}$$

where (recall the formula (4.20))

$$s(x) = \frac{x - B(J^2-1) + \sqrt{(x+B(J^2-1))^2 + 4(J^2-1)}}{2}. \tag{4.201}$$

From the formula, for $x = O(1)$,

$$s(x) = \begin{cases} x + \frac{1}{B} + O(B^{-2}) & \text{as } B \to +\infty, \\ -B(J^2-1) - \frac{1}{B} + O(B^{-2}) & \text{as } B \to -\infty. \end{cases} \tag{4.202}$$

Note that since we set $2\beta = \frac{1}{J} + \frac{B}{\sqrt{N}}$ in the transitional regime, we regard $B = O(\sqrt{N})$ when we take $B \to \pm\infty$.

4.6.1 $B \to +\infty$

Using (4.202), we find that for $x = O(1)$,

$$Q(x) = \frac{Bx}{2} + O(\log B). \tag{4.203}$$

Hence, since $\mathcal{G}_1$ does not depend on B, we see that as $B = O(\sqrt{N})$ with $B > 0$,

$$\mathcal{F}_N^{\text{tran}} = \frac{1}{4J^2} + \frac{B}{2J\sqrt{N}} + \frac{B^2J^2}{4N} + \frac{B}{2N}\mathcal{G}_2 + O\left(\frac{\log B}{N}\right) + O\left(\frac{\log N}{N}\right). \tag{4.204}$$

where $O(f(B,N))$ represents a random variable X such that the moments of $\frac{X}{f(B,N)}$ are all bounded by constants independent of B and N.

We compare the above formula with the ferromagnetic case (4.196). If we set $2\beta = \frac{1}{J} + \frac{B}{\sqrt{N}}$, then

$$\mathcal{F}_N^{\text{ferro}} = \frac{1}{4J^2} + \frac{B}{2J\sqrt{N}} + \frac{B^2J^2}{4N} + \frac{B}{2N}\mathcal{N}(f_2', \alpha_2') + O(N^{-3/2}). \tag{4.205}$$

We note that (see (4.198) and (4.25)) the mean and variance are $f_2' = \mathbb{E}[\mathcal{G}_2]$ and $\alpha_2' = \text{Var}[\mathcal{G}_2]$. The above formula of $\mathcal{F}_N^{\text{tran}}$ is thus consistent with $\mathcal{F}_N^{\text{ferro}}$.

4.6.2 $B \to -\infty$

Consider (4.200). Recall that $\mathbf{I}(\alpha) = \sqrt{\frac{4\pi}{\alpha}}(1 + O(\alpha^{-1}))$ as $\alpha \to +\infty$ from (4.65). Hence, if $x = O(1)$ and $s(x) \to \infty$, then

$$Q(x) = -\frac{s(x)^2}{4(J^2-1)} + \log\sqrt{\frac{4\pi(J^2-1)}{s(x)}} + \frac{1}{2} + O\left(\frac{1}{s(x)}\right). \tag{4.206}$$

Using (4.202), we find that for $x = O(1)$,

$$Q(x) = -\frac{B^2(J^2-1)}{4} + \log\sqrt{\frac{4\pi}{|B|}} + O(B^{-1}). \tag{4.207}$$

Hence, the two leading terms of $Q(\mathcal{G}_2)$ do not depend on $\mathcal{G}_2$. Therefore, for $B = O(\sqrt{N})$ with $B < 0$,

$$\mathcal{F}_N^{\text{tran}} = \frac{1}{4}\left(\frac{1}{J} + \frac{B}{\sqrt{N}}\right)^2 + \frac{1}{2N}\log\left(\frac{4\pi\sqrt{N}}{|B|}\right) + \frac{1}{N}\mathcal{G}_1 + O\left(\frac{1}{NB}\right). \tag{4.208}$$

On the other hand, in the paramagnetic regime, if we set $2\beta = \frac{1}{J} + \frac{B}{\sqrt{N}}$ with

$B < 0$, then the parameters in (4.199) satisfy (see (4.23))

$$\begin{aligned} f_1 &= \frac{1}{4}\log(1-J^{-2}) + \frac{1}{4J^2}(w_2-2) + \frac{1}{8J^4}(W_4-3) - \frac{1}{2}\log\left(\frac{|B|J}{\sqrt{N}}\right) + O(N^{-1/2}) \\ &= \mathbb{E}[\mathcal{G}_1] + \frac{1}{2}\log\left(\frac{4\pi\sqrt{N}}{|B|}\right) + O(N^{-1/2}) \end{aligned} \tag{4.209}$$

and

$$\alpha_1 = -\frac{1}{2}\log(1-J^{-2}) + \frac{w_2-2}{4J^2} + \frac{W_4-3}{8J^4} + O(N^{-1/2}) = \mathrm{Var}[\mathcal{G}_1] + O(N^{-1/2}) \tag{4.210}$$

Thus, if we set $2\beta = \frac{1}{J} + \frac{B}{\sqrt{N}}$ with $B < 0$, then

$$\mathcal{F}_N^{\mathrm{para}} = \frac{1}{4}\left(\frac{1}{J} + \frac{B}{\sqrt{N}}\right)^2 + \frac{1}{2N}\log\left(\frac{4\pi\sqrt{N}}{|B|}\right) + \frac{1}{N}\mathcal{N}(\mathbb{E}[\mathcal{G}_1], \mathrm{Var}[\mathcal{G}_1]) + O(N^{-3/2}). \tag{4.211}$$

This is consistent with the formula of $\mathcal{F}_N^{\mathrm{tran}}$.

CHAPTER V

Spherical spin glass model with external field

5.1 Introduction

We consider a spin glass model in which the spin variable σ is in S_{N-1}, the sphere of radius $\sqrt{N}$ in $\mathbb{R}^N$. Let $M = (M_{ij})_{1\leq i,j\leq N}$ is a disorder matrix given by a random symmetric matrix from the Gaussian orthogonal ensemble (GOE). For $i \leq j$, the variables M_{ij} are independent Gaussian random variables with variance $\frac{1}{N}(1+\delta_{ij})$. By the symmetry matrix condition, $M_{ij} = M_{ji}$ for $i > j$. The external field is given by the vector $\mathbf{g} = (g_1, g_2, \cdots, g_N)^T$ which we assume to be a standard Gaussian vector. The analysis of this chapter also applies to the case when $\mathbf{g} = (1, \cdots, 1)^T$ but we do not give any details on this case. The strength of the external field is denoted by a non-negative scalar h.

The 2-spin spherical Sherrington-Kirkpatrick (SSK) model with external field is defined by the Hamiltonian

$$\mathcal{H}(\sigma) = \frac{1}{2}\sum_{i,j=1}^{N} M_{ij}\sigma_i\sigma_j + h\sum_{i=1}^{N} g_i\sigma_i = \frac{1}{2}\sigma^T M\sigma + h\mathbf{g}^T\sigma \tag{5.1}$$

for σ in the sphere S_{N-1}. The associated Gibbs measure is

$$p(\sigma) = \frac{1}{\mathcal{Z}_N}e^{\beta\mathcal{H}(\sigma)} \quad \text{for } \sigma \in S_{N-1} \tag{5.2}$$

where $\beta = 1/T$ denotes the inverse temperature. The partition function and the free energy are

$$\mathcal{Z}_N = \int_{S_{N-1}} e^{\beta \mathcal{H}(\sigma)} \mathrm{d}\omega_N(\sigma) \quad \text{and} \quad \mathcal{F}_N = \frac{1}{N\beta} \log \mathcal{Z}_N, \tag{5.3}$$

where ω_N is the normalized uniform measure on S_{N-1}. Since the disorder variables M_{ij} are random, the Gibbs measure is a random measure and the free energy $\mathcal{F}_N$ is a random variable. We are interested in the fluctuations of the free energy when $h \to 0$ in a critical way.

The free energy for the Hamiltonian above when $h = 0$ converges to a deterministic value and this was computed by Kosterlitz, Thouless and Jones in [78]. The Hamiltonian above is the 2-spin case of more general spherical spin glass models which include interactions between multiple spin coordinates. The limit of the free energy for the general spherical spin glass models which also includes the external field is given by the Crisanti-Sommers formula [47]. This formula is the spherical version of the Parisi formula [103] for the spins in hypercubes. The Parisi formula and Crisanti-formula are proved rigorously by Talagrand in [113, 112]. The result of Kosterlitz, Thouless and Jones shows that when $h = 0$, there are two phases: the spin glass phase when $T < 1$ and the paramagnetic phase when $T > 1$. On the other hand, they argued that when $h > 0$, assuming that the external field is uniform, there is no phase transition.

The next order term of the free energy depends on the disorders and hence it describes the fluctuations of the free energy. For $h = 0$ and $T > 1$, the fluctuation term is of order N^{-1} and has the Gaussian distribution. This is proved for both hypercube case [1, 60, 42] and the spherical case [12]. For $h = 0$ and $T < 1$, for the Hamiltonian above, the fluctuation term is of order $N^{-2/3}$ and has the GOE Tracy-Widom distribution [12]. On the other hand, when $h > 0$, the fluctuation term is of order $N^{-1/2}$ and has the Gaussian distribution for all temperature in [37]. This paper also obtained similar results for general spherical spin glass models and also

the hypercube models.

It is interesting to consider the case when $h \to 0$ scaled with N. When $T = 0$, a large deviation result for the free energy was studied in the scaling $h = O(N^{-1/6})$ in [61]. This computation was proved rigorously in [50]. The paper [61] also considered the number of local minima and maxima of the Hamiltonian when $h = O(N^{-1/6})$ and also $h = O(N^{-1/2})$.

The purpose of this chapter is to study the case $h \to 0$ systematically up to the fluctuation term for the free energy. Our computation is based on a single integral representation of the partition function. The integrand of this integral contains random terms, the disorder variables and the external field. This formula was first observed by Kosterlitz, Thouless and Jones [78] and was used to evaluate the limiting free energy. We adapt their analysis to compute the next order term. This is achieved by utilizing the recent results of random matrix theory. The case when $h = 0$ was obtained in [12] and here we extend it to the cases when $h > 0$ and $h \to 0$.

The computations in this chapter is not carried out in full rigor. Most of our computataions involve an error term bounded with high probability and hence keeping track of them in detail and full rigor is cumbersome. Instead, we focus on discovering the limiting law fluctuations with a detailed and convincing computations. We expect that the computations can be proved rigorously, where extra estimations of eigenvalues of GOE may be required.

The method we use provides a unifying approach for the 2-spin SSK model with external field. Unfortunately this method is restricted only to the 2-spin case since we use a Gaussian integral to obtain a single integral representation for the partition function.

5.2 Single integral representations

The partition function is an N-fold integral over a sphere. Using the Laplace transform and Gaussian integrations, Kosterlitz, Thouless and Jones showed in [78] that this integral can be expressed as a single integral. We state this result and also include its derivation here.

The following computation applies to an arbitrary fixed symmetric matrix M and a vector $\mathbf{g}$. Let $\lambda_1 \geq \cdots \geq \lambda_N$ be the eigenvalues of the matrix M and let $\mathbf{u}_i$ be corresponding unit eigenvectors. The sign ambiguity of $\mathbf{u}_i$ does not affect the computations that follow. Let $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_N)$ be a diagonal matrix and let $O = (\mathbf{u}_1, \cdots, \mathbf{u}_N)$ be an orthogonal matrix so that $M = O\Lambda O^T$. Let S^{N-1} be the sphere of radius 1 in $\mathbb{R}^N$ and let $\mathrm{d}\Omega_{N-1}$ be the surface area element on S^{N-1}. Then, using the changes of variables $\frac{1}{\sqrt{N}} O^T \sigma = x$,

$$\mathcal{Z}_N = \frac{1}{|S^{N-1}|} \int\limits_{S^{N-1}} e^{\frac{\beta N}{2} \sum_{i=1}^N \lambda_i x_i^2 + \beta\sqrt{N} h \sum_{i=1}^N n_i x_i} \mathrm{d}\Omega_{N-1}(x) \tag{5.4}$$

where $n_i = (O^T \mathbf{g})_i = \mathbf{u}_i^T \mathbf{g}$. Note that if $\mathbf{g}$ is a standard Gaussian random vector, then $(n_1, \cdots, n_N)$ is also a standard Gaussian random vector. If, in addition, M is a GOE matrix, then n_i are independent of the eigenvalues. We may write

$$\mathcal{Z}_N = \frac{1}{|S^{N-1}|} I\left(\frac{\beta N}{2}, h\sqrt{2\beta}\right) \quad \text{where} \quad I(t,s) = \int\limits_{S^{N-1}} e^{t \sum_{i=1}^N \lambda_i x_i^2 + s\sqrt{t} \sum_{i=1}^N n_i x_i} \mathrm{d}\Omega_{N-1}(x). \tag{5.5}$$

We take the Laplace transform of $J(t) = t^{N/2-1} I(t,s)$. We find, making a simple change of variables $t = r^2$ and using Gaussian integrals, that the Laplace transform

is equal to

$$L(z) = \int_0^\infty e^{-zt} J(t) \mathrm{d}t = 2 \int_{\mathbb{R}^N} e^{-\sum_{i=1}^N (z-\lambda_i) y_i^2 + s \sum_{i=1}^N n_i y_i} \mathrm{d}^N y = 2 \prod_{i=1}^N e^{\frac{s^2 n_i^2}{4(z-\lambda_i)}} \sqrt{\frac{\pi}{z-\lambda_i}} \tag{5.6}$$

for z satisfying $z > \lambda_1$. We obtain a single integral formula of the partition function by taking the inverse Laplace transform.

Using $|S^{N-1}| = \frac{2\pi^{N/2}}{\Gamma(N/2)}$, we arrive at the formula

$$\mathcal{Z}_N = C_N \int_{\gamma - \mathrm{i}\infty}^{\gamma + \mathrm{i}\infty} e^{\frac{N}{2}\mathcal{G}(z)} \mathrm{d}z \quad \text{where} \quad C_N = \frac{\Gamma(N/2)}{2\pi \mathrm{i} (N\beta/2)^{N/2-1}} \tag{5.7}$$

and

$$\mathcal{G}(z) = \beta z - \frac{1}{N} \sum_{i=1}^N \log(z - \lambda_i) + \frac{h^2 \beta}{N} \sum_{i=1}^N \frac{n_i^2}{z - \lambda_i} \quad \text{with} \quad n_i = (O^T \mathbf{g})_i = \mathbf{u}_i^T \mathbf{g}. \tag{5.8}$$

Here, the integration is over the vertical line $\gamma + \mathrm{i}\mathbb{R}$ where γ is an arbitrary constant satisfying $\gamma > \lambda_1$. Note that there is no ambiguity due to the choice of the sign of the eigenvector since the formula depends only on n_i^2.

In later sections, we evaluate the above integral asymptotically using the method of steepest-descent. Consider

$$\mathcal{G}'(z) = \beta - \frac{1}{N} \sum_{i=1}^N \frac{1}{z - \lambda_i} - \frac{h^2 \beta}{N} \sum_{i=1}^N \frac{n_i^2}{(z - \lambda_i)^2}. \tag{5.9}$$

As a function of real variable z, it is an increasing function taking values from $-\infty$ to β as z moves from λ_1 to ∞. Hence, there is a unique real number $z > \lambda_1$ satisfying $\mathcal{G}'(z) = 0$. We set γ in (5.7) to be this critical point. Using the fact that

$$C_N = \frac{\sqrt{N}\beta}{2\mathrm{i}\sqrt{\pi}(\beta e)^{N/2}} (1 + O(N^{-1})), \tag{5.10}$$

the free energy can be written as

$$\mathcal{F}_N = \frac{1}{2\beta}(\mathcal{G}(\gamma)-1-\log\beta)+\frac{1}{N\beta}\log\left(\int_{\gamma-\mathrm{i}\infty}^{\gamma+\mathrm{i}\infty} e^{\frac{N}{2}(\mathcal{G}(z)-\mathcal{G}(\gamma))}\mathrm{d}z\right)+\frac{1}{N\beta}\log\left(\frac{\sqrt{N}\beta}{2\mathrm{i}\sqrt{\pi}}\right)+O(N^{-2}) \tag{5.11}$$

where $O(N^{-2})$ is a constant which does not depend on M and $\mathbf{g}$. The main task of later sections is to find γ and evaluate $\mathcal{G}(\gamma)$.

5.3 Results from random matrices

In this section, we recall the random matrix results introduced in Chapter III and record calculation for later purpose. Some extra random matrices results required is also introduced here.

5.3.1 Notations

For two N-dependent random variables $A := A_N$ and $B := B_N$, the notation

$$A = \mathcal{O}(B) \tag{5.12}$$

means that $A \leq B$ with high probability, i.e., for every $\epsilon, D > 0$, there exists $N_0 > 0$ such that for all $N \geq N_0$,

$$\mathbb{P}(A \geq BN^{\epsilon}) < N^{-D}. \tag{5.13}$$

The notation $\simeq$ means an asymptotic expansion up to the terms indicated on the right-hand side and the notation $\asymp$ denotes two sides are of the same order.

The convergence in distribution of a sequence of random variables X_N to a random variable X with respect to the disorder variables is denoted by $X_N \Rightarrow X$. We also use the notations $\overset{\mathcal{D}}{=}$ and $\overset{\mathcal{D}}{\simeq}$ to denote an equality and an asymptotic expansion in distribution with respect to the disorder variables and the external field, respectively.

Similarly, the notations $\overset{\mathfrak{D}}{=}$ and $\overset{\mathfrak{D}}{\simeq}$ denote an equality and an asymptotic expansion in distribution with respect to the Gibbs measure, respectively.

5.3.2 Semicircle law

By Theorem III.2, the empirical distribution of eigenvalues of M converges to the semicircle law [86]: for every continuous bounded function $f(x)$,

$$\frac{1}{N}\sum_{i=1}^{N} f(\lambda_i) \to \int f(x)\mathrm{d}\sigma_{scl}(x) \quad \text{where} \quad \mathrm{d}\sigma_{scl}(x) = \frac{\sqrt{4-x^2}}{2\pi}\mathbb{1}_{x\in[-2,2]}\mathrm{d}x \tag{5.14}$$

with probability 1 as $N \to \infty$.

We use the following functions later:

$$s_0(z) = \int \log(z-x)\mathrm{d}\sigma_{scl}(x) \quad \text{and} \quad s_k(z) = \int \frac{\mathrm{d}\sigma_{scl}(x)}{(z-x)^k} \quad \text{for } k = 1, 2, \cdots, \tag{5.15}$$

They are be evaluated explicitly as

$$\begin{aligned} s_0(z) =& \frac{1}{4}z(z-\sqrt{z^2-4}) + \log(z+\sqrt{z^2-4}) - \log 2 - \frac{1}{2}, \\ s_1(z) =& \frac{z-\sqrt{z^2-4}}{2}, \qquad s_2(z) = \frac{z-\sqrt{z^2-4}}{2\sqrt{z^2-4}}, \qquad s_3(z) = \frac{1}{(z^2-4)^{3/2}} \end{aligned} \tag{5.16}$$

for z not in the real interval $[-2, 2]$.

5.3.3 Rigidity

Under the notation introduced in (5.12), the Theorem III.4 about eigenvalue rigidity can be written as

$$|\lambda_k - \widehat{\lambda}_k| \leq (\min\{k, N+1-k\})^{-1/3}\mathcal{O}\left(N^{-2/3}\right) \tag{5.17}$$

uniformly for $k = 1, 2, \cdots, N$, where $\widehat{k} := \min\{k, N+1-k\}$,

5.3.4 Edge behavior

Recall the Theorem III.5, the the rescaled eigenvalues

$$a_i = N^{2/3}(\lambda_i - 2). \tag{5.18}$$

converges to the GOE Airy kernel point process $\{\alpha_i\}_{i=1}^{\infty}$ [116, 109]. The rightmost point α_1 of the GOE Airy kernel process has the GOE Tracy-Widom distribution TW_{GOE},

$$a_1 \Rightarrow \alpha_1 \stackrel{\mathcal{D}}{=} \mathrm{TW}_{GOE}\,. \tag{5.19}$$

The GOE Airy point process satisfies the asymptotic property that

$$\alpha_k \simeq -\left(\frac{3\pi k}{2}\right)^{2/3} \quad \text{as } k \to \infty. \tag{5.20}$$

This asymptotic is due to the fact that the semicircle law is asymptotic to $\frac{\sqrt{2-x}}{\pi}\mathrm{d}x$ as $x \to 2$. The above formula and the rigidity imply that

$$a_k \asymp -k^{2/3} \quad \text{as } k, N \to \infty \text{ satisfying } k \leq N \tag{5.21}$$

with high probability.

5.3.5 Central limit theorem of linear statistics

Theorem III.3 claims that

$$\sum_{i=1}^{N} f(\lambda_i) - N \int f(x)\mathrm{d}\sigma_{scl}(x) \tag{5.22}$$

converges to a Gaussian distribution without any scaling factor for continuous $f(x)$. We use the above result for $f(x) = \log(z - x)$ where real $z > 2$. Set (recall (5.16))

$$\mathcal{L}_N(z) := \sum_{i=1}^{N} \log(z - \lambda_i) - N s_0(z). \tag{5.23}$$

In this case,

$$\mathcal{L}_N(z) \Rightarrow \mathcal{N}(M(z), V(z)) \tag{5.24}$$

where (see Lemma A.1 in [12])

$$M(z) = \frac{1}{2} \log \left(\frac{2\sqrt{z^2 - 4}}{z + \sqrt{z^2 - 4}} \right), \qquad V(z) = 2 \log \left(\frac{z + \sqrt{z^2 - 4}}{2\sqrt{z^2 - 4}} \right). \tag{5.25}$$

For later uses, we record that for $0 < \beta < 1$,

$$M(\beta + \beta^{-1}) = \frac{1}{2} \log(1 - \beta^2), \qquad V(\beta + \beta^{-1}) = -2 \log(1 - \beta^2). \tag{5.26}$$

In Subsubsection 5.4.1.2, we need to evaluate $\sum_{i=1}^{N} \log(z - \lambda_i)$ for $z = 2 + O(N^{-2/3})$. Observe that

$$M(z) = O(\log(z - 2)) \quad \text{and} \quad V(z) = O(\log(z - 2)) \quad \text{as } z \to 2 \tag{5.27}$$

and

$$s_0(z) = \frac{1}{4} z(z - \sqrt{z^2 - 4}) - \frac{1}{2} + \log \left(\frac{z + \sqrt{z^2 - 4}}{2} \right) = \frac{1}{2} + (z - 2) + O((z - 2)^{3/2}). \tag{5.28}$$

Hence, a formal application (5.24) to this case implies that for $z \to 2$ such that $|z - 2| \geq N^{-d}$ for some $d > 0$,

$$\frac{1}{N} \sum_{i=1}^{N} \log(z - \lambda_i) = s_0(z) + \mathcal{O}\left(N^{-1}\right) = \frac{1}{2} + (z - 2) + \mathcal{O}\left(N^{-1}\right) + O((z - 2)^{3/2}). \tag{5.29}$$

5.3.6 Special sums

For $k = 1, 2, \cdots$, consider

$$\frac{1}{N}\sum_{i=2}^{N}\frac{1}{(\lambda_1-\lambda_i)^k}. \tag{5.30}$$

Later we need an estimate for this quantity for $k \geq 2$ and an asymptotic formula for $k = 1$ as $N \to \infty$. The above quantity looks superficially close to the linear statistics (5.22) with $f(x) = \frac{1}{(\lambda_1-x)^k}$ with one term removed but the function $f(x)$ is singular at $x = \lambda_1$. We note that if we replace $f(x)$ by $\frac{1}{(2-x)^k}$ and use the semicircle law, we obtain $s_k(2)$ which diverges for $k \geq 2$. Hence, the result of the previous subsection does not apply. However, using the asymptotic (5.21) of the scaled eigenvalues (5.18), we find that

$$\frac{1}{N}\sum_{i=2}^{N}\frac{1}{(\lambda_1-\lambda_i)^k} = N^{2k/3-1}\sum_{i=2}^{N}\frac{1}{(a_1-a_i)^k} = \mathcal{O}\left(N^{2k/3-1}\right) \qquad \text{for } k \geq 2. \tag{5.31}$$

For $k = 1$, the sum of $\sum_{i=2}^{N}\frac{1}{a_1-a_i}$ does not converge and hence (5.31) is not valid in this case. However, we note that

$$\int_{-2}^{2}\frac{\mathrm{d}\sigma_{scl}(x)}{2-x} = s_1(2) = 1. \tag{5.32}$$

Indeed, it turned out that (5.30) with $k = 1$ converges to this value. For our purpose, we also need the next fluctuation term. This was obtained recently in [80]. Landon and Sosoe proved that

$$\Xi_N := N^{1/3}\left(\frac{1}{N}\sum_{i=2}^{N}\frac{1}{\lambda_1-\lambda_i} - 1\right) \Rightarrow \Xi \tag{5.33}$$

for a random variable Ξ as $N \to \infty$. The limiting random variable Ξ can be expressed

in terms of the GOE Airy kernel point process as

$$\Xi = \lim_{n\to\infty} \left(\sum_{i=2}^{n} \frac{1}{\alpha_1 - \alpha_i} - \frac{1}{\pi} \int_0^{\left(\frac{3\pi n}{2}\right)^{2/3}} \frac{\mathrm{d}x}{\sqrt{x}} \right) \tag{5.34}$$

where the limit exists almost surely. Landon and Sosoe used the result (5.33) to describe the fluctuations of the overlap of two spins when $h = 0$ and $T < 1$. Their analysis also uses an integral representation and a connection to random matrix theory as in this chapter.

The idea behind the proof of (5.33) is the following. For a small $\delta > 0$, let n be the nearest integer of N^{δ}. We write

$$\Xi_N = \left[\sum_{i=2}^{n} \frac{1}{a_1 - a_i} - N^{1/3} \int_{\widehat{\lambda}_n}^{2} \frac{\mathrm{d}\sigma_{scl}(x)}{2-x} \right] + \left[\frac{1}{N^{2/3}} \sum_{i=n+1}^{N} \frac{1}{\lambda_1 - \lambda_i} - N^{1/3} \int_{-2}^{\widehat{\lambda}_n} \frac{\mathrm{d}\sigma_{scl}(x)}{2-x} \right]. \tag{5.35}$$

By the rigidity of the eigenvalues, $\lambda_1 - \lambda_i = 2 - \widehat{\lambda}_i + \mathcal{O}\left(N^{-2/3}\right)$ for all $i \geq 2$. Hence,

$$\frac{1}{N^{2/3}} \sum_{i=n+1}^{N} \frac{1}{\lambda_1 - \lambda_i} = \frac{1}{N^{2/3}} \sum_{i=n+1}^{N} \frac{1}{2 - \widehat{\lambda}_i} + \mathcal{O}\left(\frac{1}{N^{4/3}} \sum_{i=n+1}^{N} \frac{1}{(2-\widehat{\lambda}_i)^2} \right). \tag{5.36}$$

From the definition, it is easy to check that $2 - \widehat{\lambda}_n \asymp \frac{n^{2/3}}{N^{2/3}}$ for all $n = 1, \cdots, N$. Hence, the error term satisfies

$$\frac{1}{N^{4/3}} \sum_{i=n+1}^{N} \frac{1}{(2-\widehat{\lambda}_i)^2} \leq C \sum_{i=n+1}^{\infty} \frac{1}{i^{4/3}} \tag{5.37}$$

for some positive constant C, and hence it converges to zero as $N \to \infty$. Replacing the sum of $\frac{1}{2-\widehat{\lambda}_i}$ in (5.36) by an integral using the semicircle law yields an error of the same order. Therefore, the second bracket term in (5.35) converges to zero with high probability. It remains to show that the integral in the first bracket term can be

replaced by the integral

$$\frac{1}{\pi}\int_0^{\left(\frac{3\pi n}{2}\right)^{2/3}} \frac{\mathrm{d}x}{\sqrt{x}}. \tag{5.38}$$

This follows easily by recalling that $n = [N^\delta]$ and noting that

$$\mathrm{d}\sigma_{scl}(x) = \frac{\sqrt{4-x^2}}{2\pi} \simeq \frac{\sqrt{2-x}}{\pi} \tag{5.39}$$

as $x \to 2$ and

$$N^{2/3}\left(2 - \widehat{\lambda}_n\right) \simeq \left(\frac{3\pi n}{2}\right)^{2/3} \tag{5.40}$$

as $N \to \infty$. This computation implies the result (5.33).

A consequence of (5.33) is that

$$\frac{1}{N}\sum_{i=2}^{N} \frac{1}{\lambda_1 - \lambda_i} = 1 + \mathcal{O}\left(N^{-1/3}\right), \tag{5.41}$$

a fact which we use several times in this chapter.

5.3.7 Weighted sums of independent random variables

Let n_i denotes i.i.d standard normal random variable. In our setting, we have $n_i = (O^T\mathbf{g})_i$. Define

$$\mathcal{S}_N(z;k) := \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \frac{n_i^2 - 1}{(z - \widehat{\lambda}_i)^k} \tag{5.42}$$

for $k \geq 1$. By the central limit theorem and the definition of $\widehat{\lambda}_i$,

$$\mathcal{S}_N(z;k) \Rightarrow \mathcal{N}(0, 2s_{2k}(z)) \tag{5.43}$$

as $N \to \infty$ for $z > 2$.

In later sections, we need the asymptotic behavior of

$$\frac{1}{N}\sum_{i=1}^{N}\frac{n_i^2}{(z-\lambda_i)^k}=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{(z-\lambda_i)^k}+\frac{1}{N}\sum_{i=1}^{N}\frac{n_i^2-1}{(z-\lambda_i)^k}. \tag{5.44}$$

Recall that n_i's and λ_i's are independent of each other. Applying the central limit theorem (5.22) for linear statistics for the first sum and using the rigidity of the eigenvalues for the second sum, we find that

$$\frac{1}{N}\sum_{i=1}^{N}\frac{n_i^2}{(z-\lambda_i)^k}=s_k(z)+\frac{\mathcal{S}_N(z;k)}{\sqrt{N}}+\mathcal{O}\left(N^{-1}\right) \tag{5.45}$$

for $z>2$ and $k\geq 1$.

5.3.8 Another special sums

We also need an estimate that

$$\frac{1}{N}\sum_{i=2}^{N}\frac{n_i^2}{(\lambda_1-\lambda_i)^k}=N^{2k/3-1}\sum_{i=1}^{N}\frac{n_i^2}{(a_1-a_i)^k}=\mathcal{O}\left(N^{2k/3-1}\right). \tag{5.46}$$

for $k\geq 2$. This follows from the fact that the sum converges due to (5.21).

For $k=1$,

$$N^{1/3}\left(\frac{1}{N}\sum_{i=2}^{N}\frac{n_i^2}{\lambda_1-\lambda_i}-1\right)\Rightarrow\lim_{n\to\infty}\left(\sum_{i=2}^{n}\frac{\nu_i^2}{\alpha_1-\alpha_i}-\frac{1}{\pi}\int_0^{\left(\frac{3\pi n}{2}\right)^{2/3}}\frac{\mathrm{d}x}{\sqrt{x}}\right) \tag{5.47}$$

where ν_i are i.i.d Gaussian random variables with mean 0 and variance 1 independent of the GOE Airy kernel point process α_i. This follows from (5.33) and the fact that

$$\sum_{i=2}^{\infty}\frac{\nu_i^2-1}{\alpha_1-\alpha_i} \tag{5.48}$$

converges due to (5.20) and Kolmogorov's three series theorem [51].

5.4 Fluctuations of the free energy

We evaluate the fluctuations of the free energy when the external field strength $h \geq 0$ is fixed. We use the integral formula (5.11) as outlined in Subsection 5.2. For the case when $h = 0$, this computation was done in [78] for the leading deterministic term and in [12] for the fluctuation term using the same integral formula. For $h > 0$, the fluctuations for the SK model were computed in [37] using a different method. In this section, we review the computation of [12] for the case when $h = 0$ and also give a new computation for the case when $h > 0$ using the integral formula. Our result when $h > 0$ seems to be in agreement with the result of [37] assuming that their result is also valid for the SSK model.

The integral formula (5.11) for the free energy involves the function (see (5.8))

$$\mathcal{G}(z) = \beta z - \frac{1}{N}\sum_{i=1}^{N} \log(z - \lambda_i) + \frac{h^2\beta}{N}\sum_{i=1}^{N} \frac{n_i^2}{z - \lambda_i} \quad \text{where} \quad n_i = (O^T \mathbf{g})_i. \tag{5.49}$$

The point γ is the critical point of $\mathcal{G}(z)$ satisfying $\gamma > \lambda_1$. The formula (5.11) also involves an integral of $e^{\frac{N}{2}(\mathcal{G}(z) - \mathcal{G}(\gamma))}$. Since $\mathcal{G}(z) - \mathcal{G}(\gamma) = \mathcal{G}(z) - \mathcal{G}(\gamma) - \mathcal{G}'(\gamma)(z - \gamma)$, the exponent can be written as

$$N(\mathcal{G}(z) - \mathcal{G}(\gamma)) = -\sum_{i=1}^{N} \left[\log(1 + \frac{z - \gamma}{\gamma - \lambda_i}) - \frac{z - \gamma}{\gamma - \lambda_i}\right] + h^2\beta \sum_{i=1}^{N} \frac{n_i^2 (z - \gamma)^2}{(z - \lambda_i)(\gamma - \lambda_i)^2}. \tag{5.50}$$

This expression will also be used in the next two sections.

5.4.1 No external field: $h = 0$

5.4.1.1 High temperature regime: $T > 1$

When $h = 0$ and $T > 1$, we may write

$$\mathcal{G}(z) = \beta z - s_0(z) - \frac{\mathcal{L}_N(z)}{N} \tag{5.51}$$

using the notation (5.23). From (5.24), we see that $\mathcal{L}_N(z) = \mathcal{O}(1)$. Hence, a perturbation argument implies that $\gamma = \gamma_0 + \mathcal{O}(N^{-1})$ where γ_0 is the critical point of $\mathcal{G}_0(z) = \beta z - s_0(z)$. Solving the critical point equation $\mathcal{G}_0'(z) = 0$ explicitly using the formula (5.16) of $s_0(z)$, we obtain

$$\gamma_0 := \beta + \beta^{-1}. \tag{5.52}$$

This solution exists only when $T > 1$ since we need a solution such that $\gamma_0 > 2$ and $s_0'(2) = s_1(2) = 1$. We can check from the perturbation argument that

$$\mathcal{G}(\gamma) = \mathcal{G}(\gamma_0) - \frac{\mathcal{L}_N(\gamma_0)}{N} + \mathcal{O}\left(N^{-2}\right) = \frac{\beta^2}{2} + 1 + \log\beta - \frac{\mathcal{L}_N(\gamma_0)}{N} + \mathcal{O}\left(N^{-2}\right). \tag{5.53}$$

The random variable $\mathcal{L}_N(\gamma_0)$ converges to a Gaussian distribution, whose mean and the variance are given by (5.26).

In order to compute the integral term in (5.11), we use (5.50) with $h = 0$. Setting $z - \gamma = uN^{-1/2}$ with $u = O(1)$ and using a Taylor approximation,

$$N(\mathcal{G}(z) - \mathcal{G}(\gamma)) = \frac{1}{2N}\left[\sum_{i=1}^{N} \frac{1}{(\gamma - \lambda_i)^2}\right] u^2 + \mathcal{O}\left(N^{-1/2}\right) = \frac{s_2(\gamma_0)}{2}u^2 + \mathcal{O}\left(N^{-1/2}\right) \tag{5.54}$$

where we used the semicircle law. Hence,

$$\int_{\gamma-\mathrm{i}\infty}^{\gamma+\mathrm{i}\infty} e^{\frac{N}{2}(\mathcal{G}(z)-\mathcal{G}(\gamma))}\mathrm{d}z = \mathrm{i}\sqrt{\frac{4\pi}{N s_2(\gamma_0)}}\left(1 + \mathcal{O}\left(N^{-1/2}\right)\right). \tag{5.55}$$

It is direct to compute that $s_2(\gamma_0) = \beta^2/(1-\beta^2)$.

In conclusion, from (5.11),

$$\mathcal{F}_N(T, 0) = \frac{\beta}{4} + \frac{1}{2\beta N}\left[\log(1-\beta^2) - \mathcal{L}_N(\gamma_0)\right] + \mathcal{O}\left(N^{-3/2}\right). \tag{5.56}$$

Hence, the fluctuations of the free energy are of order N^{-1} and are governed by $\mathcal{L}_N(\gamma_0)$. Using the fact that $\mathcal{L}_N(\gamma_0)$ converges to a Gaussian distribution (see (5.24) and (5.26)) as $N \to \infty$, we find that

$$\mathcal{F}_N(T,0) \overset{\mathcal{D}}{\simeq} \frac{1}{4T} + \frac{T}{2N}\mathcal{N}(-\alpha, 4\alpha) \qquad \text{for } T > 1 \tag{5.57}$$

where $\alpha = -\frac{1}{2}\log(1 - T^{-2})$.

5.4.1.2 Low temperature regime: $T < 1$

When $h = 0$ and $T < 1$, it turned out that the critical point γ is $\lambda_1 + \mathcal{O}(N^{-1})$. We set $\gamma = \lambda_1 + sN^{-1}$ where s is to be determined. Separating out the term with $i = 1$ and using (5.41), we find that

$$0 = \mathcal{G}'(\gamma) = \beta - \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\gamma - \lambda_i} = \beta - 1 - \frac{1}{s} + \mathcal{O}\left(N^{-1/3}\right). \tag{5.58}$$

Thus $s = \frac{1}{\beta - 1} + \mathcal{O}\left(N^{-1/3}\right)$ which is consistent with our assumption that $s = \mathcal{O}(1)$.

To evaluate $\mathcal{G}(\gamma)$, we use the result (5.29) for the sum of log functions and obtain

$$\mathcal{G}(\gamma) = 2\beta - \frac{1}{2} + (\beta - 1)(\lambda_1 - 2) + \mathcal{O}\left(N^{-1}\right). \tag{5.59}$$

On the other hand, using (5.50) with $z = \gamma + uN^{-1}$ for $u = O(1)$,

$$N(\mathcal{G}(z) - \mathcal{G}(\gamma)) = -\log\left(1 + \frac{u}{s}\right) + \frac{u}{s} + \mathcal{O}\left(N^{-1/3}\right), \tag{5.60}$$

where only the term with $i = 1$ makes the main contribution. Thus, we obtain an estimate

$$\int_{\gamma - \mathrm{i}\infty}^{\gamma + \mathrm{i}\infty} e^{\frac{N}{2}(\mathcal{G}(z) - \mathcal{G}(\gamma))}\mathrm{d}z \asymp N^{-1}. \tag{5.61}$$

Hence, from (5.11), when $T < 1$,

$$\mathcal{F}_N(T,0) = 1 - \frac{3}{4\beta} - \frac{\log\beta}{2\beta} + \frac{1}{2N^{2/3}}\left(1-\frac{1}{\beta}\right)a_1 + \mathcal{O}\left(N^{-1}\right) \tag{5.62}$$

where $a_1 = N^{2/3}(\lambda_1 - 2)$ as in (5.18). The scaled largest eigenvalue a_1 converges to the GOE Tracy-Widom distribution (see (5.19)). Hence, the fluctuations of the free energy are of order $N^{-2/3}$ and are governed by the largest eigenvalue. This is to be contrasted to the high temperature case in which the fluctuations are governed by all eigenvalues. In conclusion, when $h = 0$,

$$\mathcal{F}_N(T,0) \overset{\mathcal{D}}{\simeq} 1 - \frac{3T}{4} + \frac{T\log T}{2} + \frac{1-T}{2N^{2/3}}\,\mathrm{TW}_{GOE} \qquad \text{for } T < 1. \tag{5.63}$$

5.4.2 Positive external field: $h > 0$

When $h > 0$, we use (5.24) and (5.45) to write $\mathcal{G}(z)$ as

$$\mathcal{G}(z) = \beta z - s_0(z) + h^2\beta\left[s_1(z) + \frac{1}{\sqrt{N}}\mathcal{S}_N(z;1)\right] + \mathcal{O}\left(N^{-1}\right) \tag{5.64}$$

where $\mathcal{S}_N(z;k)$ is defined in (5.42) which converges in distribution to $\mathcal{N}(0, 2s_{2k}(z))$. By a perturbation argument, the critical point is given by

$$\gamma = \gamma_0 + \frac{\gamma_1}{\sqrt{N}} + \mathcal{O}\left(N^{-1}\right). \tag{5.65}$$

where γ_0 is the critical point of $\mathcal{G}_0(z) = \beta z - s_0(z) + h^2\beta s_1(z)$. It is easy to check that there is $\gamma_0 > 2$ satisfying the equation

$$\beta - s_1(\gamma_0) - h^2\beta s_2(\gamma_0) = 0. \tag{5.66}$$

We do not need the explicit formula of γ_1 in this section but we record it here since the formula is used in a later section;

$$\gamma_1 = \frac{h^2\beta\mathcal{S}_N(\gamma_0;2)}{s_2(\gamma_0)+2h^2\beta s_3(\gamma_0)} \tag{5.67}$$

where we used the fact that $\frac{\mathrm{d}}{\mathrm{d}z}\mathcal{S}_N(z;1) = -\mathcal{S}_N(z;2)$.

A perturbation argument implies that

$$\mathcal{G}(\gamma) = \beta\gamma_0 - s_0(\gamma_0) + h^2\beta s_1(\gamma_0) + \frac{h^2\beta}{\sqrt{N}}\mathcal{S}_N(\gamma_0;1) + \mathcal{O}\left(N^{-1}\right). \tag{5.68}$$

The integral term in (5.11) can be evaluated in the same manner as in 5.4.1.1 using the fact that $\gamma_0 > 2$. We do not need the exact formula; we only need an estimate and obtain

$$\int_{\gamma-\mathrm{i}\infty}^{\gamma+\mathrm{i}\infty} e^{\frac{N}{2}(\mathcal{G}(z)-\mathcal{G}(\gamma))}\mathrm{d}z \asymp N^{-1/2}. \tag{5.69}$$

In conclusion, for $h > 0$ and $T > 0$,

$$\mathcal{F}_N(T,h) = \left[\frac{\gamma_0}{2} - \frac{s_0(\gamma_0)}{2\beta} + \frac{h^2 s_1(\gamma_0)}{2} - \frac{1+\log\beta}{2\beta}\right] + \frac{h^2\mathcal{S}_N(\gamma_0;1)}{2\sqrt{N}} + \mathcal{O}\left(N^{-1}\log N\right) \tag{5.70}$$

where $\mathcal{S}_N(z;k)$ is defined in (5.43). This implies that the fluctuations of the free energy are of order $N^{-1/2}$ and they are governed by the inner products n_i of the eigenvectors and the external field. By (5.43), $\mathcal{S}_N(\gamma_0;1) \Rightarrow \mathcal{N}(0, 2s_2(\gamma_0))$. Hence,

$$\mathcal{F}_N(T,h) \overset{\mathcal{D}}{\simeq} F(T,h) + \frac{h^2}{2\sqrt{N}}\mathcal{N}\left(0, 2s_2(\gamma_0)\right) \quad \text{for } T > 0 \text{ and } h > 0 \tag{5.71}$$

where

$$F(T,h) = \frac{\gamma_0}{2} - \frac{Ts_0(\gamma_0)}{2} - \frac{T - T\log T}{2} + \frac{h^2 s_1(\gamma_0)}{2} \tag{5.72}$$

and $\gamma_0 > 2$ denotes the unique solution of

$$1 - Ts_1(\gamma_0) - h^2 s_2(\gamma_0) = 0. \tag{5.73}$$

This result shows that for $h > 0$, there is no phase transition between the low and the high temperature regimes, as indicated in [78].

The fluctuations of the free energy of SK model with $h > 0$ was computed in [37] when $\mathbf{g} = \mathbf{1}$ The approach of [37] can be extended to SSK models as well and implies that $\sqrt{N}\left(\mathcal{F}_N(T,h) - \mathbb{E}[F(T,h)]\right)$ converges in distribution to the centered Gaussian distribution as $N \to \infty$ with variance

$$\frac{h^4\beta^4(1-q_0)^4}{2(1-\beta^2(1-q_0))}, \tag{5.74}$$

where q_0 solves the equation

$$\beta^2 q_0 + h^2\beta^2 = \frac{q_0}{(1-q_0)^2}. \tag{5.75}$$

Our result (5.71) above is for the SSK model when $\mathbf{g}$ is a Gaussian vector. One can also compute the case when $\mathbf{g} = \mathbf{1}$ using the same method with some extra work. In this case the variance of the limiting Gaussian distribution changes from $\frac{h^4}{2}s_2(\gamma_0)$ to $\frac{h^4}{2}(s_2(\gamma_0) - (s_1(\gamma_0))^2)$. It is not immediate clear that this formula is same as the (5.74). However, we expect that these two formulas are equal. This was confirmed by numerical evaluations for three different values of (T,h).

5.4.3 Comparison between $h > 0$ and $h = 0$ cases

In the previous two subsections, we considered three different cases: (a) $h = 0$ and $T < 1$, (b) $h = 0$ and $T > 1$, and (c) $h > 0$. The order of the fluctuations are N^{-1}, $N^{-2/3}$, and $N^{-1/2}$, respectively. The fluctuations are governed by all eigenvalues

(via a linear statistic) for (a), the top eigenvalue for (b), and a combination of the external field and the eigenvectors for (c). The limiting distributions are the Gaussian distribution, the GOE Tracy-Widom distribution, and the Gaussian distribution, respectively. Using the results obtained above, we now try to find the transitional scaling of $h \to 0$ by matching the order of the fluctuations when $h > 0$ and when $h = 0$. We need to consider the high temperature case and the low temperature case separately.

Consider the case when $T > 1$. When $h > 0$, the variance of the limiting Gaussian distribution is $\frac{h^4}{2} s_2(\gamma_0)$ where γ_0 solves the equation $1 - T s_1(\gamma_0) - h^2 s_2(\gamma_0) = 0$. If we set $h = 0$, the equation becomes $1 - T s_1(\gamma_0) = 0$. Since $T > 1$, there is a solution given by $T + \frac{1}{T}$. Hence, by a perturbation argument, $\gamma_0 = T + \frac{1}{T} + O(h^2)$ as $h \to 0$. Therefore, the fluctuations of the free energy are of order $\frac{h^2\sqrt{s_2(\gamma_0)}}{\sqrt{N}} = O(\frac{h^2}{\sqrt{N}})$. On the other hand, when $h = 0$, the fluctuations are of order $O(N^{-1})$ (see (5.57)). These two terms, $\frac{h^2}{\sqrt{N}}$ and N^{-1} are of same order if $h = O(N^{-1/4})$.

We now consider the case when $T < 1$. As before, we consider the limit as $h \to 0$ of the algebraic equation $1 - T s_1(\gamma_0) - h^2 s_2(\gamma_0) = 0$ which determines the variance of the fluctuations when $h > 0$. In this case, we expect that $\gamma_0(h) \to 2$. From the formulas (5.16), it is easy to see that, as $z \to 2$,

$$s_2(z) = \frac{1}{2\sqrt{z-2}} + O(1) \quad \text{and} \quad s_1(z) = 1 + O(\sqrt{z-2}). \tag{5.76}$$

Hence, the algebraic equation becomes

$$1 - T - \frac{h^2}{2\sqrt{\gamma_0 - 2}} + O(h^2) + O(\sqrt{\gamma_0 - 2}) = 0 \tag{5.77}$$

and we find that

$$\gamma_0(h) = 2 + \frac{h^4}{4(1-T)^2} + O(h^6). \tag{5.78}$$

This implies that the fluctuations of the free energy are of order $\frac{h^2\sqrt{s_2(\gamma_0)}}{\sqrt{N}} = O(\frac{h}{\sqrt{N}})$. On the other hand, when $h = 0$, the fluctuations are of order $O(N^{-2/3})$ (see (5.63)). The terms $\frac{h}{\sqrt{N}}$ and $N^{-2/3}$ are of same order if $h = O(N^{-1/6})$.

Summarizing, a formal computation suggests that the transitional scalings are

$$\begin{aligned} h &= O(N^{-1/4}) \quad \text{for } T < 1, \\ h &= O(N^{-1/6}) \quad \text{for } T > 1. \end{aligned} \tag{5.79}$$

In next two sections, we compute the fluctuations of the free energy in the above transitional regimes.

5.5 Transition of the free energy in the high temperature regime

5.5.1 Result

Assume that $T > 1$ and we set

$$h = HN^{-1/4} \tag{5.80}$$

for fixed $H > 0$. In this case, using the notations (5.23) and (5.42) and results (5.24) and (5.45),

$$\mathcal{G}(z) = \beta z - s_0(z) - \frac{\mathcal{L}_N(z)}{N} + \frac{H^2\beta}{\sqrt{N}}\left[s_1(z) + \frac{\mathcal{S}_N(z;1)}{\sqrt{N}}\right] + \mathcal{O}\left(N^{-3/2}\right). \tag{5.81}$$

Recall that the terms $\mathcal{L}_N(z)$ and $\mathcal{S}_N(z;1)$ are $\mathcal{O}(1)$. Applying a perturbation argument to the function

$$\mathcal{G}(z) = \beta z - s_0(z) + \frac{H^2\beta}{\sqrt{N}}s_1(z) + \frac{1}{N}\left[H^2\beta\mathcal{S}_N(z;1) - \mathcal{L}_N(z)\right] + \mathcal{O}\left(N^{-3/2}\right), \tag{5.82}$$

we find, using the explicit formulas of $s_0(z)$ and $s_1(z)$, that

$$\gamma = \gamma_0 + \frac{\gamma_1}{\sqrt{N}} + \mathcal{O}\left(N^{-1}\right) \quad \text{with } \gamma_0 = \beta + \beta^{-1} \text{ and } \gamma_1 = H^2\beta, \tag{5.83}$$

and

$$\begin{aligned}\mathcal{G}(\gamma) = &\frac{\beta^2}{2} + 1 + \log\beta + \frac{H^2\beta^2}{\sqrt{N}} \\ &+ \frac{1}{N}\left[-\frac{H^4\beta^4}{2(1-\beta^2)} + H^2\beta\mathcal{S}_N(\gamma_0;1) - \mathcal{L}_N(\gamma_0)\right] + \mathcal{O}\left(N^{-3/2}\right).\end{aligned} \tag{5.84}$$

One can also check that the integral in the formula (5.11) can be evaluated by the same manner as in the case of $h = 0$ and $T > 1$, and the formula (5.55) still holds.

From the above calculations, we find that for $h > 0$ and $T > 1$,

$$\begin{aligned}\mathcal{F}_N(T,h) =&\frac{\beta}{4} + \frac{H^2\beta}{2\sqrt{N}} \\ &+ \frac{1}{2\beta N}\left[\log(1-\beta^2) - \frac{H^4\beta^4}{2(1-\beta^2)} + H^2\beta\mathcal{S}_N(\gamma_0;1) - \mathcal{L}_N(\gamma_0)\right] + \mathcal{O}\left(N^{-3/2}\right).\end{aligned} \tag{5.85}$$

The random variables $\mathcal{S}_N(\gamma_0;1)$ and $\mathcal{L}_N(\gamma_0)$ both converge to Gaussian distributions. Since $\mathcal{S}_N(\gamma_0;1)$ depends only on n_i's and $\mathcal{L}_N(\gamma_0)$ depends only on λ_i, these two random variables are independent. Therefore,

$$\mathcal{F}_N(T,h) \overset{\mathcal{D}}{\simeq} \left[\frac{1}{4T} + \frac{H^2}{2T\sqrt{N}}\right] + \frac{T}{2N}\mathcal{N}(-\alpha, 4\alpha) \quad \text{for } T > 1 \text{ and } h = HN^{-1/4}, \tag{5.86}$$

where

$$\alpha = \frac{H^4}{2T^2(T^2-1)} - \frac{1}{2}\log(1-T^{-2}). \tag{5.87}$$

5.5.2 Comparison with $h = 0$ and $h > 0$ cases

If we set $H = 0$ in (5.86), we recover the result (5.57) for the case when $h = 0$.

We now consier the limit $H \to \infty$. We formally set $H = hN^{1/4}$ in (5.86) with h fixed and N large. The formula (5.87) becomes

$$\alpha = \frac{h^4 N}{2T^2(T^2-1)} + O(1). \tag{5.88}$$

Hence, the right hand side of (5.86) becomes

$$\frac{1}{4T} + \frac{h^2}{2T} - \frac{h^4}{4T(T^2-1)} + \frac{h^2}{\sqrt{2N(T^2-1)}}\mathcal{N}(0,1) \tag{5.89}$$

up to terms of orders N^{-1} or $N^{-1}H^{-2}$. We compare this with the formal limit $h \to 0$ of the result obtained for the case when $h > 0$. The result (5.71) involves the solution γ_0 of the equation $1 - Ts_1(\gamma_0) - h^2 s_2(\gamma_0) = 0$. A perturbation argument implies that

$$\gamma_0 = \left(T + \frac{1}{T}\right) + \frac{h^2}{T} + O(h^4) \tag{5.90}$$

since $1 - Ts_1(\gamma_0) = 0$ has a solution when $T > 1$. Hence, a direct computation shows that (5.72) becomes, as $h \to 0$,

$$F(T,h) = \frac{1}{2T} + \frac{h^2}{2T} - \frac{h^4}{4T(T^2-1)} + O(h^6) \tag{5.91}$$

and (5.71) becomes

$$F(T,h) + \frac{h^2}{\sqrt{2N(T^2-1)}}\mathcal{N}(0,1) \tag{5.92}$$

up to terms of orders h^6 or $h^4 N^{-1/2}$. This formula is same as (5.89).

5.6 Transition of the free energy in the low temperature regime

5.6.1 Result

Assume that $T < 1$ and we set

$$h = HN^{-1/6} \tag{5.93}$$

for fixed $H > 0$. In this case, (5.8) becomes

$$\mathcal{G}(z) = \beta z - \frac{1}{N}\sum_{i=1}^{N} \log(z - \lambda_i) + \frac{H^2\beta}{N^{4/3}} \sum_{i=1}^{N} \frac{n_i^2}{z - \lambda_i}. \tag{5.94}$$

Recall the notation (5.18) for the scaled eigenvalues, $a_i = N^{2/3}(\lambda_i - 2)$. We need to find the critical point $\gamma > \lambda_1$ satisfying $\mathcal{G}'(\gamma) = 0$. When $h = 0$, we had $\gamma = \lambda_1 + \mathcal{O}(N^{-1})$. In the transitional scaling $h = HN^{-1/6}$, we use the ansatz

$$\gamma = \lambda_1 + sN^{-2/3} \tag{5.95}$$

with $s = \mathcal{O}(1)$ to be determined. Using (5.41), the critical point equation becomes

$$0 = \beta - 1 - H^2\beta \sum_{i=1}^{N} \frac{n_i^2}{(s + a_1 - a_i)^2} + \mathcal{O}\left(N^{-1/3}\right). \tag{5.96}$$

Note that since $a_i \asymp i^{2/3}$ as $i \to \infty$ (see (5.21)), the series

$$\sum_{i=1}^{N} \frac{n_i^2}{(s + a_1 - a_i)^k} \tag{5.97}$$

converges with high probability as $N \to \infty$ for $k \geq 2$. Therefore, there is a solution $s > 0$ of the equation (5.96).

We now insert $\gamma = \lambda_1 + sN^{-2/3}$ into (5.94). By (5.29), the sum involving the log function becomes

$$\frac{1}{N}\sum_{i=1}^{N}\log(\gamma-\lambda_i) = \frac{1}{2} + N^{-2/3}(a_1+s) + \mathcal{O}\left(N^{-1}\right). \tag{5.98}$$

The other sum is equal to

$$\frac{H^2\beta}{N^{2/3}}\sum_{i=1}^{N}\frac{n_i^2}{a_1+s-a_i} = \frac{H^2\beta}{N^{2/3}}\left(N^{1/3}+\mathcal{E}_N(s)\right) \tag{5.99}$$

where

$$\mathcal{E}_N(s) := N^{1/3}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{n_i^2}{\lambda_1+sN^{-2/3}-\lambda_i} - 1\right] = \sum_{i=1}^{N}\frac{n_i^2}{s+a_1-a_i} - N^{1/3}. \tag{5.100}$$

This term is $\mathcal{O}(1)$ by (5.47). Using this notation, we obtain

$$\mathcal{G}(\gamma) = \left[2\beta - \frac{1}{2} + \frac{H^2\beta}{N^{1/3}}\right] + \frac{1}{N^{2/3}}\left[(\beta-1)(a_1+s) + H^2\beta\mathcal{E}_N(s)\right] + \mathcal{O}\left(N^{-1}\right) \tag{5.101}$$

where s solves (5.96).

The formula for the free energy in (5.11) also involves an integral. Setting $z-\gamma = uN^{-5/6}$ with $u = O(1)$ in (5.50), we have

$$N(\mathcal{G}(z)-\mathcal{G}(\gamma)) = \frac{H^2\beta u^2}{2}\sum_{i=1}^{N}\frac{n_i^2}{(s+a_1-a_i)^3} + \mathcal{O}\left(N^{-1/3}\right), \tag{5.102}$$

and hence

$$\int_{\gamma-\mathrm{i}\infty}^{\gamma+\mathrm{i}\infty} e^{\frac{N}{2}(\mathcal{G}(z)-\mathcal{G}(\gamma))}\mathrm{d}z \asymp N^{-5/6}. \tag{5.103}$$

Therefore, we find that

$$\mathcal{F}_N = \left[1 - \frac{3}{4\beta} - \frac{\log\beta}{2\beta} + \frac{H^2}{2N^{1/3}}\right] + \frac{(\beta-1)(a_1+s) + H^2\beta\mathcal{E}_N(s)}{2\beta N^{2/3}} + \mathcal{O}\left(N^{-1}\log N\right) \tag{5.104}$$

where s solves the equation (5.96) and $\mathcal{E}_N(s)$ is defined in (5.100). This implies that the order of fluctuations is $N^{-2/3}$, which is same as the $h = 0$ case. The fluctuations are governed by all eigenvalues and n_i's. In conclusion,

$$\mathcal{F}_N(T,h) \simeq F(T,h) + \frac{\widetilde{\mathcal{F}}(T,h)}{N^{2/3}} \qquad \text{for } h = HN^{-1/6} \text{ and } T < 1 \tag{5.105}$$

where the leading term is

$$F(T,h) = 1 - \frac{3T}{4} + \frac{T\log T}{2} + \frac{h^2}{2} \tag{5.106}$$

and the fluctuation term $\widetilde{\mathcal{F}}(T,h)$ is given as follows. The equation (5.96) for s appearing in (5.104) may be replaced by the equation for ς given by

$$1 - T = H^2 \sum_{i=1}^{\infty} \frac{\nu_i^2}{(\varsigma + \alpha_1 - \alpha_i)^2} \qquad \text{for } \varsigma > 0 \tag{5.107}$$

where α_i are the GOE Airy kernel point process and ν_i are independent standard Gaussian random variables. (Recall that $n_i = (O^T\mathbf{g})_i$ are independent standard Gaussian random variables for $i = 1, \cdots, N$.) The term $\mathcal{E}_N(s)$ in (5.104) may be replaced by (cf. (5.47))

$$\mathcal{E}(\varsigma) := \lim_{n\to\infty} \left(\sum_{i=1}^{n} \frac{\nu_i^2}{\varsigma + \alpha_1 - \alpha_i} - \frac{1}{\pi} \int_0^{\left(\frac{3\pi n}{2}\right)^{2/3}} \frac{\mathrm{d}x}{\sqrt{x}} \right). \tag{5.108}$$

Hence, we have

$$\widetilde{\mathcal{F}}(\beta,h) \stackrel{\mathcal{D}}{=} \frac{1}{2}(1-T)(\varsigma+\alpha_1) + \frac{H^2\mathcal{E}(\varsigma)}{2}. \tag{5.109}$$

5.6.2 Comparison with the case when $h = 0$

We show that (5.105) agrees with (5.63) when $H \to 0$. The leading term satisfies

$$F(T,h) = 1 - \frac{3T}{4} + \frac{T\log T}{2} + \frac{H^2}{2N^{1/3}} = F(T,0) + O(\frac{H^2}{N^{1/3}}). \tag{5.110}$$

For the fluctuation term, we first look for the solution ς of (5.107) as $H \to 0$ of the form

$$\varsigma = yH \quad \text{with } y > 0. \tag{5.111}$$

In this case, the equation becomes $1 - T = \frac{\nu_1^2}{y^2} + O(H^2)$ whose solution is given by

$$y = \frac{|\nu_1|}{\sqrt{1-T}} + O(H^2). \tag{5.112}$$

Inserting $\varsigma = yH$ into (5.108), we find that

$$\mathcal{E}(\varsigma) = \frac{\nu_1^2}{yH} + \lim_{n\to\infty}\left(\sum_{i=2}^{n}\frac{\nu_i^2}{yH+\alpha_1-\alpha_i} - \frac{1}{\pi}\int_0^{\left(\frac{3\pi n}{2}\right)^{2/3}}\frac{\mathrm{d}x}{\sqrt{x}}\right) = \mathcal{O}\left(\frac{1}{H}\right). \tag{5.113}$$

Therefore,

$$\widetilde{\mathcal{F}}(T,h) = \frac{1}{2}(1-T)(\varsigma+\alpha_1) + \frac{H^2\mathcal{E}(\varsigma)}{2} = \frac{1-T}{2}\alpha_1 + \mathcal{O}\left(H\right). \tag{5.114}$$

The term $\frac{1-T}{2}\alpha_1$ is exactly the fluctuation term $\widetilde{\mathcal{F}}(\beta,0)$ in (5.63).

5.6.3 Comparison with the case when $h > 0$

We show that the formal limit of (5.105) as $H \to \infty$ is consistent with the formal limit of (5.71) as $h \to 0$.

5.6.3.1 Formal limit of (5.71) as $h \to 0$

The formula involves $\gamma_0 > 2$ which solves the equation

$$1 - Ts_1(\gamma_0) - h^2 s_2(\gamma_0) = 0. \tag{5.115}$$

If we set $h = 0$, there is no solution $\gamma_0 > 2$ since $T < 1$. However as $h \to 0$, γ_0 is given by (see (5.78))

$$\gamma_0 = 2 + \frac{h^4}{4(1-T)^2} + O(h^6). \tag{5.116}$$

Hence, the leading term (5.72) becomes

$$F(T,h) = 1 - \frac{3T}{4} + \frac{T\log T}{2} + \frac{h^2}{2} - \frac{h^4}{8(1-T)} + O(h^6). \tag{5.117}$$

The fluctuation term has the Gaussian distribution with mean zero and variance $\frac{h^4 s_2(\gamma_0)}{2} = \frac{h^2(T-1)}{2} + O(h^4)$. Hence, the formal asymptotic of the case when $h > 0$ as $h \to 0$ is given by

$$\mathcal{F}_N \simeq \left[1 - \frac{3T}{4} + \frac{T\log T}{2} + \frac{h^2}{2} - \frac{h^4}{8(1-T)}\right] + \frac{1}{\sqrt{N}}\mathcal{N}\left(0, \frac{h^2(1-T)}{2}\right) \tag{5.118}$$

up to the terms of orders h^6 and $h^3 N^{-1/2}$.

5.6.3.2 Formal limit of (5.105) as $H \to \infty$

We first determine the solution ς of the equation (5.107),

$$\frac{1-T}{H^2} = \sum_{i=1}^{\infty} \frac{\nu_i^2}{(\varsigma + \alpha_1 - \alpha_i)^2} \tag{5.119}$$

when $H \to \infty$. Since the right-hand side is a decreasing function of ς and it converges to zero as $\varsigma \to \infty$, the solution $\varsigma \to \infty$ as $H \to \infty$. We evaluate the asymptotic of the right-hand side in this limit. From $\alpha_i \simeq -\left(\frac{3\pi i}{2}\right)^{2/3}$ for large i (see (5.20)), we find

using the Riemann sum approximation of an integral that for every $k \geq 2$,

$$\sum_{i=1}^{\infty} \frac{1}{(\varsigma + \alpha_1 - \alpha_i)^k} \simeq \frac{1}{\varsigma^{k-3/2}} \int_0^{\infty} \frac{\mathrm{d}y}{\left(1 + (\frac{3\pi y}{2})^{2/3}\right)^k} = \frac{\Gamma(k-3/2)}{2\sqrt{\pi}\Gamma(k)\varsigma^{k-3/2}} \tag{5.120}$$

as $\varsigma \to \infty$. In particular, when $k = 2$, the right hand side of the above equation is equal to $\frac{1}{2\sqrt{\varsigma}}$ and the equation (5.119) becomes

$$\frac{1-T}{H^2} = \frac{1}{2\sqrt{\varsigma}} + \sum_{i=1}^{\infty} \frac{\nu_i^2 - 1}{(\varsigma + \alpha_1 - \alpha_i)^2} + \mathcal{O}\left(\varsigma^{-3/2}\right) \tag{5.121}$$

where the error term comes from (5.120) with $k = 3$. Since the conditional variance satisfies

$$\mathrm{Var}\left[\sum_{i=1}^{\infty} \frac{\nu_i^2 - 1}{(\varsigma + \alpha_1 - \alpha_i)^2} \middle| \alpha_i\right] = \sum_{i=1}^{\infty} \frac{2}{(\varsigma + \alpha_1 - \alpha_i)^4} = \mathcal{O}\left(\varsigma^{-\frac{5}{2}}\right) \tag{5.122}$$

using (5.120) with $k = 4$, we expect that

$$\sum_{i=1}^{\infty} \frac{\nu_i^2 - 1}{(\varsigma + \alpha_1 - \alpha_i)^2} = \mathcal{O}\left(\varsigma^{-5/4}\right), \tag{5.123}$$

which is larger than the error $\mathcal{O}\left(\varsigma^{-\frac{3}{2}}\right)$ in (5.121). Hence, the solution is given by

$$\varsigma \simeq \varsigma_0 + \mathcal{O}(H) \quad \text{with } \varsigma_0 = \frac{H^4}{4(1-T)^2}. \tag{5.124}$$

Even though it is not needed, we can also compute the next term and find that

$$\varsigma \simeq \varsigma_0 + \frac{H^6}{2(1-T)^3} \sum_{i=1}^{N} \frac{\nu_i^2 - 1}{(\varsigma_0 + \alpha_1 - \alpha_i)^2}. \tag{5.125}$$

Now we consider the asymptotic of $\mathcal{E}(\varsigma)$ defined in (5.108) as $\varsigma \to \infty$. We write

it as

$$\mathcal{E}(\varsigma) = \sum_{i=1}^{\infty} \frac{\nu_i^2 - 1}{\varsigma + \alpha_1 - \alpha_i} + \lim_{n\to\infty} \left[\sum_{i=1}^{n} \frac{1}{\varsigma + \alpha_1 - \alpha_i} - \int_0^{\left(\frac{3\pi n}{2}\right)^{2/3}} \frac{\sqrt{x}\mathrm{d}x}{\pi(\varsigma + x)} \right] + \frac{1}{\pi} \int_0^{\infty} \left(\frac{\sqrt{x}}{\varsigma + x} - \frac{1}{\sqrt{x}} \right) \mathrm{d}x. \tag{5.126}$$

The last term is equal to $-\varsigma^{1/2}$ by evaluating the integral directly. The second term is the difference between a Riemann sum and a Riemann integral. This difference is of the same order as $\sum_{i=1}^{\infty} \frac{1}{(\varsigma+\alpha_1-\alpha_i)^2}$ which was shown to be of order $\varsigma^{-1/2}$ in (5.120). We will see that this term is smaller than the first term which will be shown to be of order $\varsigma^{-1/4}$. Consider the first term. Its convergence is warranted by the Kolmogorov three series theorem. We evaluate its behavior as $\varsigma \to \infty$ by considering the moment generating function: for any fixed ξ, as $\varsigma \to \infty$,

$$\mathbb{E}\left[e^{\xi \sum_{i=1}^{\infty} \frac{\nu_i^2 - 1}{\varsigma + \alpha_1 - \alpha_i}} \middle| \{\alpha_i\} \right] = \prod_{i=1}^{\infty} e^{-\frac{\xi}{\varsigma+\alpha_1-\alpha_i} - \frac{1}{2}\log(1 - \frac{2\xi}{\varsigma - \alpha_1 - \alpha_i})} = e^{\sum_{i=1}^{\infty} \frac{\xi^2}{(\varsigma+\alpha_1-\alpha_i)^2} + O\left(\sum_{i=1}^{\infty} \frac{1}{(\varsigma+\alpha_1-\alpha_i)^3}\right)} \tag{5.127}$$

Using (5.120), we find that

$$\mathbb{E}\left[e^{\xi \sum_{i=1}^{\infty} \frac{\nu_i^2 - 1}{\varsigma + \alpha_1 - \alpha_i}} \middle| \{\alpha_i\} \right] = \exp\left(\frac{\xi^2}{2\varsigma^{1/2}} + \mathcal{O}\left(\varsigma^{-3/2}\right) \right). \tag{5.128}$$

Since the leading term does not depend on α_i, we conclude that

$$\varsigma^{1/4} \sum_{i=1}^{\infty} \frac{\nu_i^2 - 1}{\varsigma + \alpha_1 - \alpha_i} \Rightarrow \mathcal{N}(0, 1) \tag{5.129}$$

as $\varsigma \to \infty$. This shows that the first term of (5.126) is of order $\varsigma^{-1/4}$ and converges to

a standard Gaussian random variable as $\varsigma \to \infty$. Therefore, we find that as $\varsigma \to \infty$,

$$\mathcal{E}(\varsigma) \overset{\mathcal{D}}{=} -\varsigma^{1/2} + \varsigma^{-1/4}\mathcal{N}(0,1) + \mathcal{O}\left(\varsigma^{-1/2}\right). \tag{5.130}$$

By (5.124) and (5.130), the term $\widetilde{\mathcal{F}}(\beta, h)$ with $h = HN^{-1/6}$ in (5.109) becomes, as $H \to \infty$,

$$\widetilde{\mathcal{F}}(T,h) \overset{\mathcal{D}}{=} \frac{1}{2}(1-T)(\varsigma+\alpha_1) + \frac{H^2\mathcal{E}}{2} \overset{\mathcal{D}}{\simeq} -\frac{H^4}{8(1-T)} + H\sqrt{\frac{1-T}{2}}\mathcal{N}(0,1). \tag{5.131}$$

Combining with (5.106), and replacing H by $hN^{1/6}$, we find that the formal asymptotic of (5.105) as $H \to \infty$ is given by

$$\mathcal{F}_N \overset{\mathcal{D}}{\simeq} 1 - \frac{3T}{4} + \frac{T\log T}{2} + \frac{h^2}{2} - \frac{h^4}{8(1-T)} + \frac{h}{N^{1/2}}\sqrt{\frac{1-T}{2}}\mathcal{N}(0,1). \tag{5.132}$$

This agrees with (5.118).

CHAPTER VI

Introduction of free component analysis

Consider the following scenario: there is a group of people talking simultaneously at a cocktail party. Given microphones recording mixed sounds, we want to separate the speeches of each individual. This cocktail problem [39], together with other situations of practical interest, can be formulated as

$$\underbrace{\begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix}}_{=:\boldsymbol{z}} = \underbrace{\begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_s \end{bmatrix}}_{=:\boldsymbol{A}} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}_{=:\boldsymbol{x}}, \tag{6.1}$$

where $\boldsymbol{x} = [x_1, \cdots, x_n]^T$ with x_i representing the source signals (eg. the speech of each person in cocktail party problem). In this context, $\boldsymbol{z} = [z_1, \cdots, z_m]^T$ containing the signals observed (eg. the record in cocktail party problem). Also, the matrix $\boldsymbol{A}$, which is called the mixing matrix, represents the linear transfer matrix between sources and observations.

The blind source separation (BSS) studies how to identify the mixing matrix and recover the underlying signals with limited prior knowledge; most of the information is merely the observations. Due to the lack of information and underdeterminancy of the problem, the idea of BSS is to impose conditions certain principles to the underlying mixing matrix and source signals. The ideal principles should characterize the source

signals and narrow the set of possible solutions.

There are two main approaches of BSS. The idea of the first approaches is to impose structural constraints (eg. nonnegative matrix factorization [98, 81]). The second approach adapts a probabilistic and information-theoretic point of view. For example, the recovered signals by the principal component analysis (PCA) or independent component analysis (ICA) are maximally uncorrelated or independent respectively. For a comprehensive introduction of BSS as well as a historical review, interested readers are referred to [44].

In this chapter, we present a method for unmixing matrix signals in a way analogous to ICA.

6.1 From PCA to ICA via cumulants

Principal component analysis (PCA) [104] is a widely used dimensionality reduction technique in statistical machine learning. The principal components learned by PCA are the directions that maximize the variance, subject to a set of orthogonality constraints. Mathematically speaking, given a (centered) data matrix $\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}_1 & \dots & \boldsymbol{y}_s \end{bmatrix}^T$, the i-th principal component is the solution to the manifold optimization problem

$$\boldsymbol{w}_i^{\mathsf{pca}} = \underset{||\boldsymbol{w}||_2=1}{\arg\max}\, \text{variance}(\boldsymbol{w}^T\boldsymbol{Y}) \text{ subject to } \boldsymbol{w} \perp \boldsymbol{w}_1^{\mathsf{pca}}, \dots, \boldsymbol{w}_{i-1}^{\mathsf{pca}}. \tag{6.2}$$

The variance or the second cumulant [45] of a random variable x is defined as

$$c_2(x) = \text{variance}(x) := \mathbb{E}[x^2] - (\mathbb{E}[x])^2 \tag{6.3}$$

Substituting (6.3) into (6.2) allows us to cast PCA as a maximization of the second

cumulant:

$$\boldsymbol{w}_i^{\mathsf{pca}} = \underset{||\boldsymbol{w}||_2=1}{\arg\max}\, c_2(\boldsymbol{w}^T\boldsymbol{Y}) \text{ subject to } \boldsymbol{w} \perp \boldsymbol{w}_1^{\mathsf{pca}}, \ldots, \boldsymbol{w}_{i-1}^{\mathsf{pca}}. \tag{6.4}$$

Independent component analysis (ICA) [43, 73] is a dimensionality reduction technique that is obtained by replacing (in our notation) $c_2(\cdot)$ on the right hand side of (6.4) by the fourth cumulant $c_4(\cdot)$, thereby yielding the optimization problem

$$\boldsymbol{w}_i^{\mathsf{ica}} = \underset{||\boldsymbol{w}||_2=1}{\arg\max}\, |c_4(\boldsymbol{w}^T\boldsymbol{Y})| \text{ subject to } \boldsymbol{w} \perp \boldsymbol{w}_1^{\mathsf{ica}}, \ldots, \boldsymbol{w}_{i-1}^{\mathsf{ica}}. \tag{6.5}$$

The fourth cumulant $c_4(\cdot)$ of a scalar random variable x is equivalent to its kurtosis [40, 45], and when $\mathbb{E}[x] = 0$ it is given by [108, Eq. (6)]

$$c_4(x) = \text{kurtosis}(x) := \mathbb{E}[x^4] - 3\left(\mathbb{E}[x^2]\right)^2. \tag{6.6}$$

We refer to the formulation in (6.5) as kurtosis, or c_4-ICA in short. Replacing $c_4(\cdot)$ on the right hand side of (6.5) with the $c_j(\cdot)$ for integer $j \geq 3$ yields c_j-ICA. There are other formulations of ICA involving different objective functions, such as for example any non-quadratic, well-behaving even function as in [69, 68]; see [43] for a discussion on other such contrast functions.

6.2 Known result: ICA unmixes mixtures of independent random variables

Suppose we are given a multivariate vector $\boldsymbol{z}$ modeled as

$$\underbrace{\begin{bmatrix} z_1 \\ \vdots \\ z_s \end{bmatrix}}_{=:\boldsymbol{z}} = \underbrace{\begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_s \end{bmatrix}}_{=:\boldsymbol{A}} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_s \end{bmatrix}}_{=:\boldsymbol{x}}, \tag{6.7}$$

where $\boldsymbol{A}$ is a non-singular $s \times s$ mixing matrix and $\boldsymbol{x}$ is a vector of independent scalar-valued random variables. Assume, without loss of generality, that $\mathbb{E}[\boldsymbol{x}] = \boldsymbol{0}$ and $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^T] = \boldsymbol{I}$. Let $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ be the singular value decomposition (SVD) of the mixing matrix. Then, we have that

$$\boldsymbol{C}_{\boldsymbol{z}\boldsymbol{z}} := \mathbb{E}[\boldsymbol{z}\boldsymbol{z}^T] = \boldsymbol{A}\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^T]\boldsymbol{A}^T = \boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^T.$$

The whitened vector $\boldsymbol{y} = \boldsymbol{C}_{\boldsymbol{z}\boldsymbol{z}}^{-1/2}\boldsymbol{z}$ has identity covariance and can be rewritten in terms of the SVD of $\boldsymbol{A}$ as

$$\boldsymbol{y} = \boldsymbol{C}_{\boldsymbol{z}\boldsymbol{z}}^{-1/2}\boldsymbol{z} = \boldsymbol{U}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x} = \underbrace{\left(\boldsymbol{U}\boldsymbol{V}^T\right)}_{=:\boldsymbol{Q}}\boldsymbol{x}. \tag{6.8}$$

Note that $\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{V}^T$ in (6.8) is an orthogonal matrix, because $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices derived from the SVD of $\boldsymbol{A}$. Equation (6.8) thus reveals that the whitened vector $\boldsymbol{y}$ is related to the latent independent random variables that we wish to unmix via an orthogonal transformation. If we can estimate $\boldsymbol{Q}$ from $\boldsymbol{y}$, we can unmix the independent random variables by computing $\widehat{\boldsymbol{W}}^T\boldsymbol{y}$ provided $\widehat{\boldsymbol{W}} = \boldsymbol{Q}\boldsymbol{P}\boldsymbol{S}$ where $\boldsymbol{P}$ is a permutation matrix and $\boldsymbol{S}$ is a diagonal matrix with ± 1 as diagonal elements.

It is a remarkable fact [43, 73] that, generically, for $\boldsymbol{y}$ modeled as in (6.8), c_4-ICA as in (6.5) returns $\boldsymbol{W}_{\mathsf{ica}}$ such that $\boldsymbol{W}_{\mathsf{ica}}^T \boldsymbol{y}$ unmixes the mixed independent random variables. Thus ICA can be viewed as a procedure for unmixing sums of independent random variables from each other.

The caveat of c_4-ICA is that no more than one of the independent random variables is Gaussian, and that the random variables do not all have a kurtosis identically equal to zero. The latter condition rules out the use of c_k-ICA for odd $k > 2$ because the cumulants of a symmetric random variable are identically equal to zero, so that we would not be able to unmix a large class of random variables.

Replacing c_4-ICA with c_k-ICA for even $k > 4$ would still not allow us to unmix more than one Gaussian random variable: this is a fundamental limit of ICA [43, Section 2]. Cardoso [31] discusses aspects related to the use of higher order contrast functions for ICA while Chen and Bickel [35] address the important issue of the statistical efficiency of ICA estimators in the presence of limited samples. In practice, c_4-ICA or kurtosis based ICA is often used for its simplicity.

6.3 Our contribution: From ICA to FCA via free cumulants

Free probability theory is a mathematical theory developed by Voiculescu [118, 119, 120, 121] that is a counterpart of scalar (or classical) probability theory, except that the random variables are non-commutative in a manner that scalar random variables are not. In free probability theory, "freeness" or free independence is the analogue of the classical notion of independence.

We begin by placing ourselves in an abstract setting with a (unital) algebra $\mathcal{X}$ of non-commutative random variables that is equipped with a linear functional $\varphi : \mathcal{X} \to \mathbb{C}$. The important point here is that functional $\varphi(\cdot)$ plays the same role as the expectation operator $\mathbb{E}[\cdot]$ in classical probability theory. The critical difference comes from non-commutative nature of the underlying probability space in free probability

as we illustrate next[1].

Suppose x_1 and x_2 are classically independent random variables. Then the their mixed moment satisfies

$$\mathbb{E}[(x_1x_2)^3] = \mathbb{E}[x_1^3x_2^3] = \mathbb{E}[x_1^3] \cdot \mathbb{E}[x_2^3],$$

since $x_1x_2x_1x_2x_1x_2 = x_1^3x_2^3$ because x_1 and x_2 are scalar random variables and are hence commutative. In contrast, even when x_1 and x_2 are freely independent, the mixed moments

$$\varphi[(x_1x_2)^3] = \varphi[x_1x_2x_1x_2x_1x_2] \neq \varphi(x_1^3) \cdot \varphi(x_2^3),$$

since $x_1x_2x_1x_2x_1x_2 \neq x_1^3x_2^3$ whenever x_1 and x_2 are assumed to be non-commutative.

Free probability, via free independence, provides a recipe for computing such mixed moments of freely independent random variables in a manner that is analogous to but different from classical probability theory. For our purpose here, there is a notion of free cumulants $\kappa(\cdot)$ for integer m which exhibit the same properties as the classical cumulants (see Theorem B.7 and (B.13) in Appendix B.1.2). This allows us to cast FCA analogous to the ICA in (6.5) as a fourth free cumulant maximization problem of the form

$$\boldsymbol{w}_i^{\mathsf{fca}} = \underset{||\boldsymbol{w}||_2=1}{\arg\max} |\kappa_4(\boldsymbol{w}^T\boldsymbol{y})| \text{ subject to } \boldsymbol{w} \perp \boldsymbol{w}_1^{\mathsf{fca}}, \ldots, \boldsymbol{w}_{i-1}^{\mathsf{fca}}, \tag{6.9}$$

where $\kappa_4(\cdot)$ is the fourth free cumulant. We can similarly formulate κ_m-FCA for $m \geq 3$ as we did for ICA.

This is also where we depart from ICA in another crucial sense. We can model the random variables as self-adjoint (or symmetric) or non-self adjoint (or rectangular/non-

[1]See Appendix B.1 for a self-contained introduction to free probability and how it differs from classical probability.

symmetric) and gives us self-adjoint and rectangular variants of FCA, respectively. [118] developed free probability theory for self-adjoint random variables; [19] extended it to rectangular random variables.

In the self-adjoint setting $\kappa_4(\cdot)$ is given by (7.3) while in the non-self adjoint (or rectangular, in a sense we shall shortly see) setting $\kappa_4(\cdot)$ is given by (7.4)

The development and analysis of algorithms for self-adjoint and rectangular FCA is the main contribution of this paper.

6.4 Our main finding: FCA unmixes mixtures of free random variables

If we whiten the vector $\boldsymbol{z}$ as in (6.8) with the covariance matrix defined via the $\varphi(\cdot)$ operator as in Definition VII.11, then we show that κ_4-FCA, just as c_4-ICA, returns $\boldsymbol{W}_{\mathsf{fca}} = \boldsymbol{QPS}$ (see Theorem VII.4), and thus $\boldsymbol{W}_{\mathsf{fca}}^T \boldsymbol{y}$ unmixes the mixed free random variables.

The caveat of κ_4-FCA, analogous to the c_4-ICA algorithm, is that no more than one of the free random variables can be the free probabilistic equivalent of the classical Gaussian random variable, and that the random variables do not all have a free kurtosis equal to zero. In the self-adjoint setting, the free analog of the Gaussian is the free semi-circular element [65] while in the rectangular setting, it is the free Poison element [19].

Just as for ICA, the condition that the free kurtosis of the free random variables cannot all equal to zero rules out the use of κ_m-FCA for odd valued $m \geq 3$ in the self-adjoint setting, because the free cumulants of a symmetric free random variable are identically equal to zero and so we would not be able to unmix a large class of free random variables with symmetric distribution. On the other hand, for rectangular free random variable, cumulants odd orders are zeros by default (see [18, (b), pp. 6]).

We will prove that just as in the ICA setting, replacing κ_4-FCA with κ_m-FCA for even valued $m \geq 4$ would still not allow us to unmix more than one Gaussian analog free random variable: this is a fundamental limit of FCA. Thus FCA fails whenever we have more than one free Gaussian analogs mixed together. This is the fundamental limit of FCA.

The free semi-circular element in the self-adjoint setting, and the Poisson element in the rectangular case, are the only non-commutative random variables with higher order kurtosis equal to zero, analogous to the Gaussian in the scalar setting. Thus, we might say that FCA finds directions that maximize deviation from the semi-circularity (or Poissonity) when the random variables are self-adjoint (or rectangular, respectively).

We also develop an algorithm for FCA based on the maximization of the free entropy for both the self-adjoint [118, 65] and rectangular settings [18], and show that FCA successfully unmixes free random variables in a similar way. Table 6.1 summarizes our results.

Table 6.1: FCA algorithms and their limits.

	self-adjoint FCA (free kurtosis)	self-adjoint FCA (free entropy)	rect. FCA (free rect. kurtosis)	rec FCA (free rect. entropy)
Recovery Guarantee	Theorem VII.4	Theorem VII.8	Theorem VII.4	Theorem VII.8
Identifiability Condition	At most one component with $\kappa_4 = 0$	At most one free semicircular element	At most one component with $\kappa_4 = 0$	At most one free Poisson element

6.5 Insight: FCA unmixes mixtures of (asymptotically) free random matrices

Voiculescu [117, 90] showed that symmetric random matrices are good models for asymptotically free self-adjoint random variables (also see Appendix B.1.4.1). The non-commutativity comes in because matrix multiplication is non-commutative. Benaych-Georges [19] showed that rectangular random matrices are good models for asymptotically free rectangular random variables (also see Appendix B.1.4.2).

In the self-adjoint setting, Voiculescu showed that random matrices $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are asymptotically free whenever $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent of each other if one, or both, of the random matrices have isotropically random (or Haar distributed) eigenvectors. In the non self-adjoint or rectangular setting, Benaych-Georges showed analogous that rectangular random matrix $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are free whenever they are independent of each other and if the singular vectors of one or both of the random matrices are Haar distributed. Since these pioneering works, many authors have relaxed the conditions and broadened the class of random matrices that we now know to be asymptotically free – see, for example the work of [84] and [5]. We can thus consider the matrix mixing model

$$\underbrace{\begin{bmatrix} \boldsymbol{Z}_1 \\ \vdots \\ \boldsymbol{Z}_s \end{bmatrix}}_{=:\boldsymbol{Z}} = \underbrace{\begin{bmatrix} a_{11}\boldsymbol{I} & \dots & a_{1s}\boldsymbol{I} \\ \vdots & \dots & \vdots \\ a_{s1}\boldsymbol{I} & \dots & a_{ss}\boldsymbol{I} \end{bmatrix}}_{=:\boldsymbol{A}\otimes\boldsymbol{I}} \underbrace{\begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_s \end{bmatrix}}_{=:\boldsymbol{X}}, \tag{6.10}$$

When the matrices $\boldsymbol{X}_1, \dots, \boldsymbol{X}_s \in \mathbb{R}^{N\times N}$ are symmetric or Hermitian, then we are in the self-adjoint setting. Voiculescu [117] showed that the appropriate linear function $\varphi(\cdot)$ is exactly the normalized trace function. That is,

$$\varphi(\boldsymbol{X}_i) = \lim_N \frac{1}{N}\mathrm{Tr}(\boldsymbol{X}_i) \tag{6.11}$$

and

$$\varphi(\boldsymbol{X}_i\boldsymbol{X}_j) = \lim_N \frac{1}{N}\mathrm{Tr}(\boldsymbol{X}_i\boldsymbol{X}_j). \tag{6.12}$$

Replacing this with their sample analogs gives us a concrete algorithm for self-adjoint FCA; see Algorithm 1 and Algorithm 2.

When the matrices $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_s \in \mathbb{R}^{N\times M}$ are rectangular, we are in the non self-adjoint setting [19]. Then. the appropriate pair of the linear functionals $\varphi_1(\cdot)$ and $\varphi_2(\cdot)$ are exactly the normalized trace functions in $\mathbb{R}^{N\times N}$ and $\mathbb{R}^{M\times M}$:

$$\varphi_1(\boldsymbol{X}_i\boldsymbol{X}_j^H) = \lim_N \frac{1}{N}\mathrm{Tr}(\boldsymbol{X}_i\boldsymbol{X}_j^H) \tag{6.13}$$

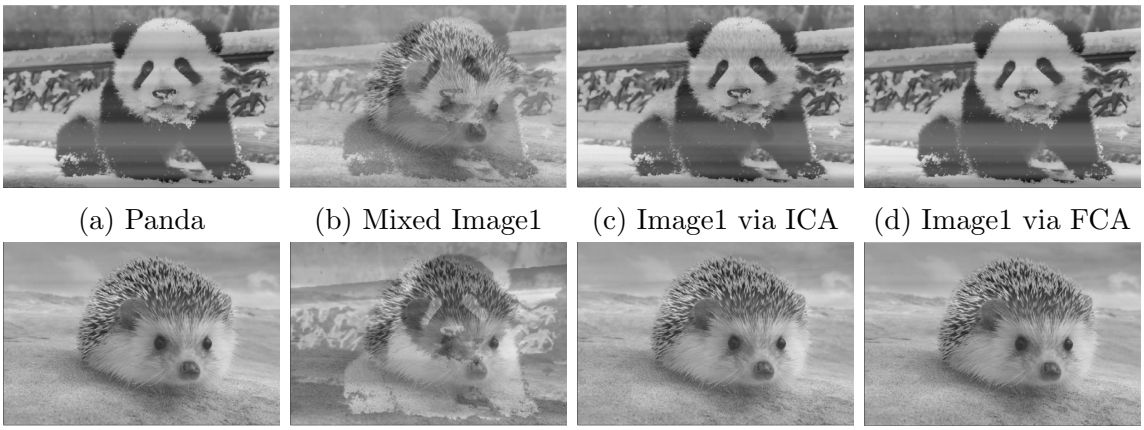
and

$$\varphi_2(\boldsymbol{X}_i^H\boldsymbol{X}_j) = \lim_M \frac{1}{M}\mathrm{Tr}(\boldsymbol{X}_i^H\boldsymbol{X}_j). \tag{6.14}$$

Thus we expect that asymptotically, FCA should unmix asymptotically free random matrices. In the setting where the random matrices are large but finite, we expect FCA to approximately unmix the asymptotically free random matrices, with some non-zero but small unmixing error, analogous to the finite sample unmixing performance of ICA [75, 59, 7]. We will use numerical simulations to demonstrate that FCA can near perfectly unmix mixtures of large, finite sized (asymptotically free) matrices - see Sections 7.2.1 and 7.2.2.

6.6 Insight: FCA can be applied wherever ICA has been applied

ICA has been successfully applied to image unmixing, audio separation and waveform unmixing problem [82, 91]. Here we show that *FCA can be successfully applied wherever ICA has succeeded*, including in settings where there are seemingly no matrices in sight.



(a) Panda (b) Mixed Image1 (c) Image1 via ICA (d) Image1 via FCA

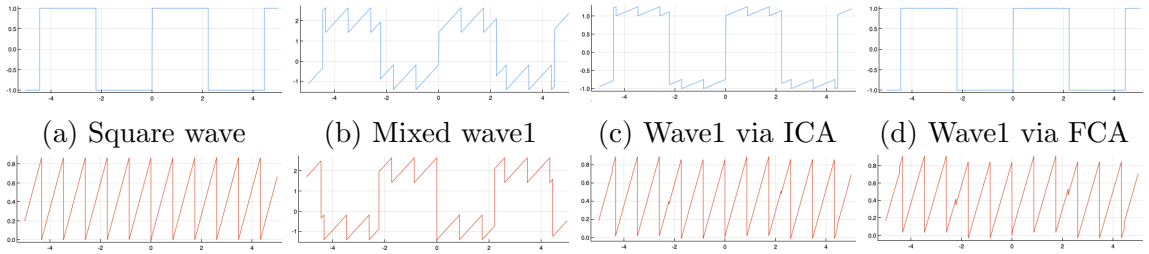
(e) Hedgehog (f) Mixed Image2 (g) Image2 via ICA (h) Image2 via FCA

Figure 6.1: An experiment in image separation using ICA and FCA. Note that subplots (c), (g) (unmixed images via ICA) and (d), (h) (unmixed images via FCA) both recover (a), (e) respectively. Here, $\boldsymbol{A} = [\sqrt{2}, \sqrt{2}; -\sqrt{2}, \sqrt{2}]/2$ in (6.10). The error of ICA is 6.08×10^{-2} while the error of FCA is 2.69×10^{-2}. See (7.32) for the definition of the error.

Figure 6.1 showcases the successful use of FCA for unmixing mixed images. This is a natural place to apply FCA because (grayscale) images are matrices. Applying ICA to unmix the images involves vectorizing the images, and treating them as mixed scalar random variables in a way that ignores the spatial matrix information that FCA uses. Perhaps it is therefore not surprising that FCA outperforms ICA.

What *is* surprising is that the images in Figure 6.1 *are not* textbook examples of asymptotically free random matrices. By this we mean that would not have predicted that the panda and hedgehog matrices are free according to the definition in Appendix B.1. One might even argue that they are not really random matrices. And yet, FCA unmixes them as though they are free. For this and many, many other examples of mixed natural images. It is as though matrices in the wild are free-er than we might initially fear they are not. We hope that experiments with FCA and computational reasoning on its unexpected successes can guide free probabilists looking to expand the class of matrix models for which freeness holds.

Figures 6.2 and 6.3 show examples where we are trying to unmix mixed deter-



(a) Square wave (b) Mixed wave1 (c) Wave1 via ICA (d) Wave1 via FCA

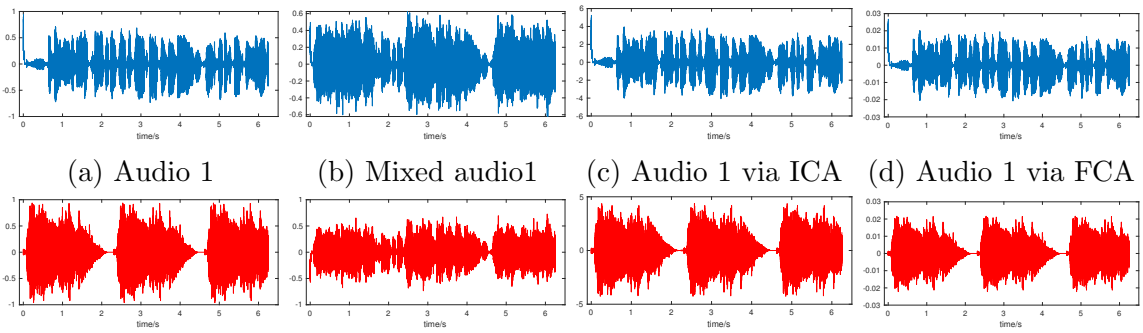
(e) Sawtooth wave (f) Mixed wave2 (g) Wave2 via ICA (h) Wave2 via FCA

Figure 6.2: An experiment in waveform separation using ICA and FCA. Note that subplots (c), (g) (unmixed waves via ICA) and (d), (h) (unmixed waves via FCA) both recover (a), (e). Visually, FCA performs better in this experiment. In this simulation, $\boldsymbol{A} = [\sqrt{2}, \sqrt{2}; -\sqrt{2}, \sqrt{2}]/2$ in (6.10). The errors for ICA and FCA are 9.95×10^{-2} and 6.66×10^{-2} respectively.

ministic waveforms and audio signals respectively. ICA is known to succeed in these examples, and it is natural to apply ICA here since the latent variables are scalar valued. FCA seems unnatural because there are no matrices in sight, let alone mixed matrices!

The surprising insight is that if we compute the spectrogram of the mixed signals, then the matrix mixing model in (6.10) is with respect to the spectrogram matrices of the mixed signals: we can use FCA to unmix the signals! Here, FCA on the spectrogram embdedding outperforms ICA. We might compute other matrix embeddings (say via the short time wavelet transform) and apply FCA there. We do not (yet) have a theory to predict which embedding would lead to better unmixing; nonetheless, the important point is that by embedding scalar valued signals as matrices, we can apply FCA wherever ICA has been applied, and that we can also possibly get better (or worse – see Figure 7.6) unmixing performance by varying the matrix embedding.

Figure 6.5 summarizes our worldview on this and our sense that there is a theory waiting to be fully revealed on the relation between non-asymptotic recovery of mixed variables and a to-be-defined notion of distance to the various notions of freeness and independence that can provide a principled way to reason about whether ICA or FCA will better unmix the mixed variables.



(a) Audio 1 (b) Mixed audio1 (c) Audio 1 via ICA (d) Audio 1 via FCA

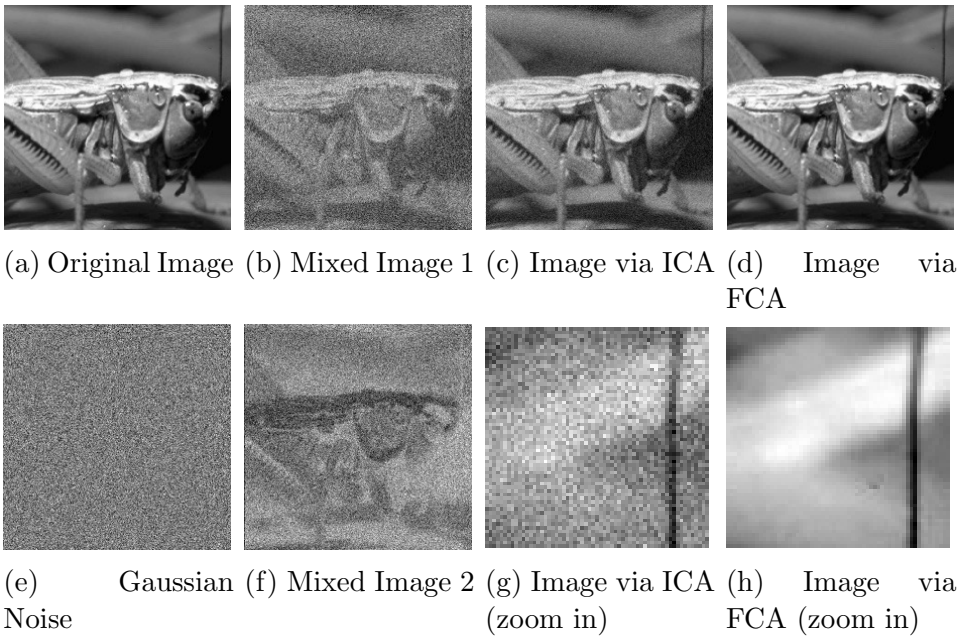
(e) Audio 2 (f) Mixed Audio 2 (g) Audio 2 via ICA (h) Audio 2 via FCA

Figure 6.3: An experiment in audios separation via ICA and FCA: Note that subplots (c), (g) (unmixed audio signals via ICA) and (d), (h) (unmixed audio signals via FCA) both recover (a), (e). In this experiment, $\boldsymbol{A} = [\sqrt{2}, \sqrt{2}; -\sqrt{2}, \sqrt{2}]/2$ in (6.10). The errors for ICA and FCA are 1.47×10^{-2} and 1.79×10^{-2} respectively.

6.7 Surprising insight: FCA often better unmixes random variables than ICA

In the examples in Figures 6.1 and 6.3, FCA did better than ICA in a quantitative sense. Figure 6.4 shows a setting where we are unmixing two mixed images and where one of the images corresponds to a Gaussian random matrix. In this setting, FCA performs better than ICA in a visually perceptible way. We have observed that FCA usually does at least as well as ICA and often better.

In a similar setting, we replace the locust image by a matrix $\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T$ in SVD form, where $\boldsymbol{U}$, $\boldsymbol{V}$ are Discrete Cosine Transformation (DCT) matrices and $\boldsymbol{D}$ is a diagonal matrix (see Section 7.2.2). This matrix model enables us to increase the dimension and compare the asymptotic behavior of ICA and FCA. Our numerical simulations show that κ_4-FCA and c_4-ICA perform similarly. However, we observe that free entropy based FCA significantly outperforms ICA (see Figure 7.4) at the cost of increased computational complexity, since estimating the free entropy involves eigenvalue (or singular value) computation, which are of order $O(N^3)$.



(a) Original Image (b) Mixed Image 1 (c) Image via ICA (d) Image via FCA

(e) Gaussian Noise (f) Mixed Image 2 (g) Image via ICA (zoom in) (h) Image via FCA (zoom in)

Figure 6.4: An experiment in image denoising via ICA and (kurtosis-based) FCA: Comparing (g) and (f), we observe that FCA does a better job then ICA in this experiment. Here $\boldsymbol{A} = [\sqrt{2}, \sqrt{2}; -\sqrt{2}, \sqrt{2}]/2$ in (6.10). The variance of whitened Gaussian noise is set to equal the empirical variance of original image.

6.8 Organization

The remainder of the paper is organized as follows. We will develop FCA for self-adjoint and rectangular non-commutative random variables (corresponding to self-adjoint and rectangular random matrices) in Section 7.1 by describing the objective functions whose maximization, analogous to the ICA setting, leads to successful unmixing of the 'free' components from their additive mixture. Then we describe FCA based algorithms for factorizing data matrices in Section 7.1.6. We illustrate our theorems and ability of FCA to successfully unmix real-world images using numerical simulation in Section 7.2. We present some concluding remarks and highlight some open problems in Section 7.3

A self-contained introduction to the free probability is given in Section B.1.2 and B.1.3 for self-adjoint and rectangular random variables respectively. We build the connection between non-commutative random variables and random matrices in

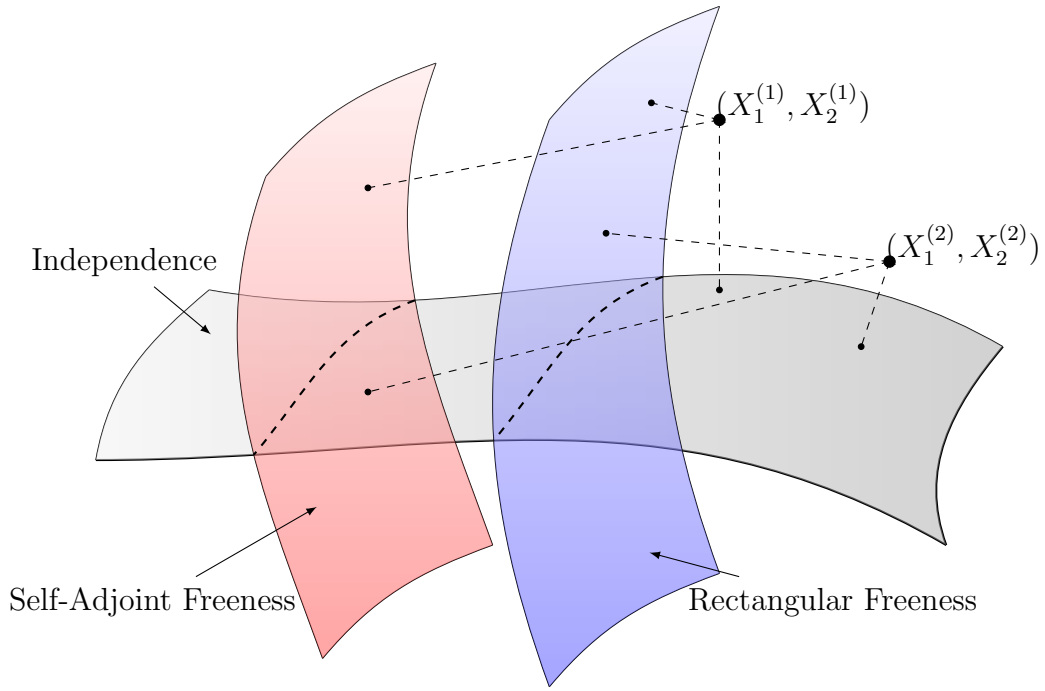


Figure 6.5: We can regard ICA and FCA with various embedding as "projections" onto corresponding manifolds. Here, the gray surface denotes the manifold of independent pairs. The red and blue surfaces stand for self-adjoint free pairs and rectangular free pairs respectively. In order to achieve the best performance, one shall pick the projection into the closest manifold. For example, if the latent data is $(X_1^{(1)}, X_2^{(1)})$, then rectangular FCA should have the best performance when separating them from the additive mixture. In contrast, for the underlying data $(X_1^{(2)}, X_2^{(2)})$, one should pick ICA.

Section B.1.4

CHAPTER VII

Free component analysis: main result, simulation and proof

7.1 Main result: Recovery guarantees for FCA

7.1.1 Setup and assumptions under an orthogonal mixing model

7.1.1.1 The self-adjoint setting

Given a probability space $(\mathcal{X}, \varphi)$, let $x_1, \ldots, x_s$ be s self-adjoint and free random variables (see Appendix B.1.2). Let $\boldsymbol{y}$ denote the vector which contains as its elements the various additive mixtures of $x_1, \ldots, x_s$. We model $\boldsymbol{y}$ as

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_s \end{bmatrix}}_{=:\boldsymbol{y}} = \underbrace{\begin{bmatrix} \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_s \end{bmatrix}}_{=:\boldsymbol{Q}} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_s \end{bmatrix}}_{=:\boldsymbol{x}}, \tag{7.1}$$

where $\boldsymbol{Q}$ is a $s \times s$ orthogonal matrix.

For self-adjoint FCA, we assume that the variables x_i are centered and have unit variance, *i.e.* for $i = 1, \cdots, s$, we have that $\varphi(x_i) = 0$ and that $\varphi(x_i^2) = 1$.

7.1.1.2 The non self-adjoint setting

Given a (ρ_1, ρ_2)-rectangular probability space $(\mathcal{X}, p_1, p_2, \varphi_1, \varphi_2)$ – (see Appendix B.1.3) – we consider a setup similar to that in (7.1) where we model $\boldsymbol{y}$ as

$$\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{x}, \tag{7.2}$$

where $\boldsymbol{Q}$ is an $s \times s$ orthogonal matrix. We assume that for $i = 1, \cdots, s$, x_i, y_i are rectangular random variables (i.e., $x_i, y_i \in \mathcal{X}_{12} := p_1 \mathcal{X} p_2$) and $\varphi_1(x_i x_i^*) = 1$. Note that $\varphi(x_i) = 0$ by default. The fundamental assumption is now that $(x_i)_{i=1}^s$ are free with amalgamation over the linear span of p_1 and p_2. We will simply say that $(x_i)_{i=1}^s$ are free if there is no ambiguity.

7.1.2 Free kurtosis based FCA

The *free kurtosis* of a centered self-adjoint random variable $x \in \mathcal{X}$ is defined as

$$\kappa_4(x) = \varphi(x^4) - 2\,\varphi(x^2)^2. \tag{7.3}$$

The *rectangular free kurtosis* of a rectangular random variable $x \in \mathcal{X}_{12}$ is defined as

$$\kappa_4(x) = \varphi_1((xx^*)^2) - \left(1 + \frac{\varphi(p_1)}{\varphi(p_2)}\right)(\varphi_1(xx^*))^2\,. \tag{7.4}$$

We now state a result on the largest free component

Theorem VII.1 (Largest free component)**.** *Assume $\boldsymbol{x}$ and $\boldsymbol{y}$ are related either via* (7.1) *in the self-adjoint setting or via* (7.2) *in the non self-adjoint setting. Suppose, additionally, without of loss of generality, that*

$$|\kappa_4(x_1)| \geq |\kappa_4(x_2)| \geq \cdots \geq |\kappa_4(x_s)| > 0. \tag{7.5}$$

Let $\boldsymbol{w}^{(1)}$ *denote the solution of the manifold optimization problem*

$$\boldsymbol{w}^{(1)} = \arg\max_{\boldsymbol{w}} \left|\kappa_4(\boldsymbol{w}^T\boldsymbol{y})\right| \ \textit{subject to} \ \|\boldsymbol{w}\| = 1 \tag{7.6}$$

(a) Suppose

$$|\kappa_4(x_1)| > |\kappa_4(x_2)| > \cdots > |\kappa_4(x_s)| > 0. \tag{7.7}$$

Then

$$\boldsymbol{w}^{(1)} = \pm\,\boldsymbol{q}_1 \tag{7.8}$$

(b) Suppose there is an integer $r \in [2..s]$*, such that*

$$|\kappa_4(x_1)| = \cdots = |\kappa_4(x_r)| > |\kappa_4(x_{r+1})| \geq \cdots \geq |\kappa_4(x_s)| > 0. \tag{7.9}$$

Then

$$\boldsymbol{w}^{(1)} \in \{\pm\,\boldsymbol{q}_1, \ldots, \pm\,\boldsymbol{q}_r\}. \tag{7.10}$$

Remark VII.2. (b) of the above theorem considers the case where there are multiple indexes corresponding to the largest absolute kurtosis. In contrast to the principal component analysis, the maximizers of (7.6) (and also of (7.10)) only contains corresponding columns of $\boldsymbol{Q}$, and not their general linear combinations. This is a consequence of that we are using the fourth order statistics of random variables.

Theorem VII.3 (The k-th largest free component)**.** *Assume that* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *are related as in Theorem VII.1. Let* $\boldsymbol{w}^{(k)}$ *denote the solution to the manifold optimization problem*

$$\boldsymbol{w}^{(k)} = \arg\max_{\boldsymbol{w}} \left|\kappa_4(\boldsymbol{w}^T\boldsymbol{y})\right| \ \textit{subject to} \ \|\boldsymbol{w}\| = 1, \boldsymbol{w} \perp \boldsymbol{w}^{(1)}, \cdots, \boldsymbol{w}^{(k-1)} \tag{7.11}$$

Suppose

$$|\kappa_4(x_1)| > |\kappa_4(x_2)| > \cdots > |\kappa_4(x_s)| > 0.$$

Then

$$\boldsymbol{w}^{(k)} = \pm\, \boldsymbol{q}_k. \tag{7.12}$$

Theorem VII.4 (Principal free components)**.** *Assume that* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *are related as in Theorem VII.1. Let* $\boldsymbol{O}(s)$ *denote the set of* $s \times s$ *orthogonal matrices, and consider the manifold optimization problem*

$$\max_{\boldsymbol{W}} \sum_{i=1}^{s} \left|\kappa_4\left([\boldsymbol{W}^T\boldsymbol{y}]_i\right)\right| \;\; \textit{subject to } \boldsymbol{W} \in \boldsymbol{O}(s), \tag{7.13}$$

where $[\boldsymbol{W}^T\boldsymbol{y}]_i$ *denotes the* i*-th element of* $\boldsymbol{W}^T\boldsymbol{y}$*. Suppose that*

$$|\kappa_4(x_1)| > |\kappa_4(x_2)| > \cdots > |\kappa_4(x_s)| > 0.$$

Then $\boldsymbol{W}$ *is an optimum if and only if :*

$$\boldsymbol{W} = \boldsymbol{QPS}, \tag{7.14}$$

for some $\boldsymbol{P}$ *and* $\boldsymbol{S}$ *where where* $\boldsymbol{P}$ *is a permutation matrix and* $\boldsymbol{S}$ *is a diagonal matrix with* ± 1 *as diagonal elements.*

Remark VII.5. Above theorems still hold if there is at one components with zero free kurtosis.

7.1.2.1 Higher-order free cumulant based FCA

Remark VII.6. It can be shown that above theorems still hold with $\kappa_4(\cdot)$ replaced by any $\kappa_{2m}(\cdot)$, for $m \geq 3$.

Remark VII.7. The maximizer of (7.13) is not guaranteed to recover $\boldsymbol{Q}$ when there are multiple components of x with zero free kurtosis. In this case, one may try to use optimization problem based on $\kappa_{2m}(\cdot)$, $m \geq 3$. However, the semicircle elements (for the self-adjoint case; see Appendix B.1.2.4) and the Poisson elements (for the non self-adjoint case; see Appendix B.1.3.4) have all vanishing free cumulants of order higher than 2. In Theorem VII.9, we will prove that $\boldsymbol{Q}$ can be recovered if and only if $\boldsymbol{x}$ contains at most one semicircular element or free Poissonian element for the self-adjoint or non self-adjoint settings, respectively.

7.1.2.2 Free-entropy based FCA

The free entropy $\chi(a_1, \cdots, a_n)$ (see Sections B.1.2.3 and B.1.3.3 for the definitions in the self-adjoint and non self-adjoint settings) of a tuple of free random variables encodes the dependence between the variables a_i. Analogous to the scalar setting, the free entropy is maximized when the random variables are freely independent. Thus we can pose FCA as an entropy maximization problem as stated next.

Theorem VII.8 (FCA based on free entropy)**.** *Assume that $\boldsymbol{x}$ and $\boldsymbol{y}$ are related as in Theorem VII.1 and at most one component of x is semicircular in the self-adjoint setting or a free Poisson in the non-self adjoint setting. Let $\boldsymbol{O}(s)$ denote the set of $s \times s$ orthogonal matrices. Suppose that*

$$\chi(x_i) > -\infty \text{ for } i = 1, \cdots, s.$$

Consider the manifold optimization problem

$$\max_{\boldsymbol{W}} \sum_{i=1}^{s} -\chi\left([\boldsymbol{W}^T\boldsymbol{y}]_i\right) \text{ subject to } \boldsymbol{W} \in \boldsymbol{O}(s), \tag{7.15}$$

where $[\boldsymbol{W}^T\boldsymbol{y}]_i$ denotes the i-th element of $\boldsymbol{W}^T\boldsymbol{y}$. Then $\boldsymbol{W}$ is an optimum if and only

if:

$$\boldsymbol{W} = \boldsymbol{QPS}, \tag{7.16}$$

for some $\boldsymbol{P}$ *and* $\boldsymbol{S}$ *where* $\boldsymbol{P}$ *is a permutation matrix and* $\boldsymbol{S}$ *is a diagonal matrix with* ± 1 *as diagonal elements.*

7.1.2.3 FCA identifiability condition

In the self-adjoint setting c_4-FCA will fail when $\boldsymbol{x}$ contains semicircular elements because free semi-circular elements have a free kurtosis identically equal to zero. Moreover, suppose $\boldsymbol{x} = (x_1, x_2)^T$ where x_i are free semicircular elements with $\varphi(x_i) = 0$ and $\varphi(x_i^2) = 1$. Then, it can be shown that for any $\boldsymbol{Q} \in \boldsymbol{O}(2)$, the components of $\boldsymbol{Qx}$ are still free semicircular elements. Therefore, if there are more two components are semicircular elements, it is impossible to identify $\boldsymbol{Q}$ with the mere knowledge of free independence between the components of $\boldsymbol{x}$. The analog of this holds for the non self-adjoint setting as well.

We now state an FCA identiability condition based on this observation.

Theorem VII.9 (Identifiability Condition)**.** *Consider* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *and* $\boldsymbol{Q} \in \boldsymbol{O}(s)$ *such that* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *are related as in Theorem VII.1. Assume* $\boldsymbol{x}$ *has free elements. Assume that at most one component of* $\boldsymbol{x}$ *is semicircular in the self-adjoint setting or free Poisson in the non self-adjoint setting.*

Now, if there is a $\boldsymbol{W} \in \boldsymbol{O}(s)$ *such that* $\boldsymbol{W}^T\boldsymbol{y}$ *has free components, then*

$$\boldsymbol{W} = \boldsymbol{QPS}. \tag{7.17}$$

for some $\boldsymbol{P}$ *and* $\boldsymbol{S}$ *where* $\boldsymbol{P}$ *is a permutation matrix and* $\boldsymbol{S}$ *is a diagonal matrix with* ± 1 *as diagonal elements. That is,* $\boldsymbol{W}$ *can be obtained by permuting the columns of* $\boldsymbol{Q}$ *with possible sign flips and vice versa.*

Remark VII.10 (Weakness of FCA condition relative to ICA). Note that the FCA

identifiability condition is weaker than the corresponding condition for ICA [43, Theorem 10, pp. 294]. The ICA condition is a consequence of scalar Cramérs Lemma [43, Lemma 9, pp. 294] and a Lemma of Marcinkiewicz-Dugue [43, Lemma 10, pp. 294]. Since the analog of Cramérs Lemma in the free probability does not hold [41], the identifiability condition is correspondingly weaker.

7.1.3 Setup and assumptions under a non-orthogonal mixing model

7.1.3.1 The self-adjoint setting

Given a probability space $(\mathcal{X}, \varphi)$, let $x_1, \ldots, x_s$ be s self-adjoint and free random variables. Let $\boldsymbol{x}$ be a vector of free, but not necessarily centered random variables. Then the variable $\widetilde{x}_i$ defined as

$$\widetilde{x}_i = x_i - \underbrace{\varphi(x_i)\, 1_{\mathcal{X}}}_{=:\overline{x}_i},$$

is centered since $\varphi(\widetilde{x}_i) = 0$. Substituting $x_i = \widetilde{x}_i + \varphi(x_i)$ in (6.7) we obtain the mixed model

$$\underbrace{\begin{bmatrix} z_1 \\ \vdots \\ z_s \end{bmatrix}}_{=:\boldsymbol{z}} = \underbrace{\begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_s \end{bmatrix}}_{=:\boldsymbol{A}} \underbrace{\begin{bmatrix} \widetilde{x}_1 + \overline{x}_1 \\ \vdots \\ \widetilde{x}_s + \overline{x}_s \end{bmatrix}}_{=:\widetilde{\boldsymbol{x}}+\overline{\boldsymbol{x}}} = \boldsymbol{A}\widetilde{\boldsymbol{x}} + \boldsymbol{A}\overline{\boldsymbol{x}}, \tag{7.18}$$

In this general, non-orthogonal mixing setup, we assume, without loss of generality, that $\varphi(\widetilde{x}_i^2) = 1$ and covariance $\boldsymbol{C}_{\boldsymbol{xx}} = \boldsymbol{I}$, where the covariance matrix $\boldsymbol{C}_{\boldsymbol{xx}}$ is defined as following.

Definition VII.11 (Covariance matrix of self-adjoint random variables)**.** Let $\boldsymbol{z} = \begin{bmatrix} z_1 & \ldots & z_s \end{bmatrix}^T$ be a vector of self-adjoint random variables. The covariance $\boldsymbol{C}_{\boldsymbol{zz}}$

matrix of $\boldsymbol{z}$ is the $s \times s$ matrix given by:

$$[\boldsymbol{C_{zz}}]_{ij} = \varphi\left(\widetilde{z}_i \widetilde{z}_j\right) \quad \text{for } i, j = 1, \ldots, s, \tag{7.19}$$

where $\widetilde{z}_i$ is the centered random variable

$$\widetilde{z}_i = z_i - \varphi(z_i) 1_{\mathcal{X}}.$$

7.1.3.2 The non self-adjoint setting

Given a rectangular probability space $(\mathcal{X}, p_1, p_2, \varphi_1, \varphi_2)$, let $x_1, \ldots, x_s$ be s self-adjoint and free random variables. We assume that $\boldsymbol{z}$ is modeled as in (7.18). In the non self-adjoint setting, the variables are centered by construction – we assume additionally that for all $x_i \in \mathcal{X}_{12}$, $\varphi_1(x_i x_i^*) = 1$ and covariance $\boldsymbol{C_{xx}} = \boldsymbol{I}$, where the covariance matrix $\boldsymbol{C_{xx}}$ is defined as following.

Definition VII.12 (Covariance matrix of non self-adjoint random variables)**.** For an arbitrary random vector $\boldsymbol{z}^T = \begin{bmatrix} z_1 & \ldots & z_s \end{bmatrix}$ of rectangular random variables from $\mathcal{X}_{12}$, note that $\varphi(z_i) = 0$ by default, the covariance matrix of z is defined by a $s \times s$ matrix $\boldsymbol{C_{zz}}$ where

$$[\boldsymbol{C_{zz}}]_{ij} = \varphi_1\left(z_i z_j^*\right). \tag{7.20}$$

7.1.4 Unmixing mixed free random variables using FCA

We first establish some properties of the covariance matrices thus computed.

Proposition VII.13. *The covariance matrix as in Definitions VII.11 and VII.12 is positive semi-definite.*

For the covariance of $\boldsymbol{z}$ satisfying (7.18), we have the following stronger result.

Proposition VII.14. *The vector of mixed variables $\boldsymbol{z}$ modeled as in (7.18) has covariance $\boldsymbol{C_{zz}}$ that is real and positive definite.*

This proposition allows us to formulate FCA on the whitened vector and prove a recovery results as stated next.

Theorem VII.15. *Assume $\boldsymbol{x}$ and $\boldsymbol{z}$ are related as in (6.7). Let $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ be the singular value decomposition of $\boldsymbol{A}$. Consider the manifold optimization problem*

$$\max_{\boldsymbol{W}} \sum_{i=1}^{s} \kappa_4 \left([\boldsymbol{W}^T \boldsymbol{y}]_i \right) \text{ subject to } \boldsymbol{W} \in \boldsymbol{O}(s), \tag{7.21}$$

where $\boldsymbol{y}$ is the whitened and centered vector given by:

$$\boldsymbol{y} = \boldsymbol{C}_{\boldsymbol{zz}}^{-1/2} \widetilde{\boldsymbol{z}}, \tag{7.22}$$

where $\boldsymbol{C}_{\boldsymbol{zz}}^{-1/2} = \boldsymbol{U}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^T$ is the inverse of the square root of the covariance matrix $\boldsymbol{C}_{\boldsymbol{zz}}$ and $\boldsymbol{z}$ is the centered vector whose i-th element is given by

$$\widetilde{z}_i = z_i - \varphi(z_i) 1_{\mathcal{X}}.$$

Suppose that

$$|\kappa_4(x_1)| > |\kappa_4(x_2)| > \cdots > |\kappa_4(x_s)| > 0.$$

Then $\boldsymbol{W}$ is an optimum if and only if:

$$\boldsymbol{W} = \left(\boldsymbol{U}\boldsymbol{V}^T \right) \boldsymbol{P}\boldsymbol{S}, \tag{7.23}$$

for some $\boldsymbol{P}$ and $\boldsymbol{S}$ where where $\boldsymbol{P}$ is a permutation matrix and $\boldsymbol{S}$ is a diagonal matrix with ± 1 as diagonal elements.

Proof. It suffices to observe via (6.8) that

$$\boldsymbol{y} = \left(\boldsymbol{U}\boldsymbol{V}^T \right) \widetilde{\boldsymbol{x}}. \tag{7.24}$$

The matrix $\boldsymbol{U}\boldsymbol{V}^T$ is an orthogonal matrix because $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices and so we can recover $\boldsymbol{U}\boldsymbol{V}^T$ from the stated manifold optimization problem via an application of Theorem VII.4. □

Theorem VII.16. *Suppose $\boldsymbol{x}$ and $\boldsymbol{z}$ are related as in Theorem VII.15. Let $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ be the singular value decomposition of $\boldsymbol{A}$. Also suppose at most one elements of $\boldsymbol{x}$ is semicircular in the self-adjoint setting and free Poissionian in the non self-adjoint setting and that*

$$\chi(x_i) > -\infty \qquad \text{for } i = 1, \cdots, s. \tag{7.25}$$

Consider the manifold optimization problem

$$\max_{\boldsymbol{W}} \sum_{i=1}^{s} -\chi\left([\boldsymbol{W}^T\boldsymbol{y}]_i\right) \text{ subject to } \boldsymbol{W} \in \boldsymbol{O}(s), \tag{7.26}$$

where $\boldsymbol{y}$ is the whitened and centered vector given by (7.22)*.*

Then $\boldsymbol{W}$ is an optimum if and only if:

$$\boldsymbol{W} = \left(\boldsymbol{U}\boldsymbol{V}^T\right)\boldsymbol{P}\boldsymbol{S}, \tag{7.27}$$

where $\boldsymbol{P}$ is a permutation matrix and $\boldsymbol{S}$ is a diagonal matrix with ± 1 diagonal elements.

Proof. The proof is exactly same as the the proof of the Theorem VII.15, except for our application of Theorem VII.8 to (7.24) instead of Theorem VII.4. □

Corollary VII.17 (Unmixing via FCA)**.** *Suppose $\boldsymbol{x}$ and $\boldsymbol{z}$ are related as in Theorem VII.15 and that the x_i's satisfy the conditions in Theorem VII.15 and VII.16. Let $\boldsymbol{W}_{\mathsf{opt}}$ denote an optimum of the optimization problem in (7.21) or* (7.26)*. Consider*

the factorization

$$\boldsymbol{z} = \widehat{\boldsymbol{A}}\,\widehat{\boldsymbol{x}}, \tag{7.28}$$

where

$$\widehat{\boldsymbol{A}} = \boldsymbol{C}_{\boldsymbol{zz}}^{1/2}\,\boldsymbol{W}_{\mathsf{opt}}, \tag{7.29a}$$

and

$$\widehat{\boldsymbol{x}} = \widehat{\boldsymbol{A}}^{-1}\boldsymbol{z}. \tag{7.29b}$$

Then $\widehat{\boldsymbol{A}} = \boldsymbol{APS}$ *for some* $\boldsymbol{P}$ *is a permutation matrix and* $\boldsymbol{S}$ *is a diagonal matrix with* ± 1 *diagonal elements. Therefore,* $\widehat{\boldsymbol{x}}$ *recovers* $\boldsymbol{x}$ *up to permutation and sign flips.*

Proof. As $\boldsymbol{C}_{\boldsymbol{zz}}^{1/2} = \boldsymbol{U}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^T$, given an optimum $\boldsymbol{W}$ satisfying $\boldsymbol{W} = \left(\boldsymbol{U}\boldsymbol{V}^T\right)\boldsymbol{PS}$,

$$\boldsymbol{C}_{\boldsymbol{zz}}^{1/2}\boldsymbol{W} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{PS} = \boldsymbol{APS}. \tag{7.30}$$

That is, we recover mixing matrix $\boldsymbol{A}$ up to column permutation and column sign flips. This completes the proof. □

7.1.5 Overdetermined and underdetermined FCA

We now consider same model as in (6.10) for the settings where the mixing matrix $\boldsymbol{A}$ is rectangular. In the over-determined setting where $\boldsymbol{A}$ is a $p \times s$ mixing matrix with $p > s$. Then it can be shown that FCA applied to $\boldsymbol{y} = \boldsymbol{\Sigma}_s^{-1}\boldsymbol{U}_s^T\boldsymbol{z}$ will unmix the free random variables. Here $\boldsymbol{U}_s$ is a $p \times s$ matrix and $\boldsymbol{\Sigma}_s$ is an $s \times s$ diagonal matrix of the singular values of $\boldsymbol{A}$. These matrices can obtained by using eigenvalue decomposition $\boldsymbol{C}_{zz} = \boldsymbol{U}_s\boldsymbol{\Sigma}_s\boldsymbol{\Sigma}_s^T\boldsymbol{U}_s^T$.

In the under-determined setting where $p < s$, then FCA cannot be used to unmix the free random variables just as ICA cannot either.

7.1.6 Unmixing mixtures of matrices using FCA

The multiplication of matrices are non-commutative, therefore we can consider the mixing model in (6.10) where $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_s$ are finite dimensional (asymptotically) free self-adjoint or rectangular matrices (see Definitions B.16 and B.17). The goal is to unmix $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_s$ from their additive mixtures $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_s$.

Corollary VII.17 provides a recipe for unmixing the mixture of matrices by factorizing $\boldsymbol{Z}$ into the matricial analog of (7.28). In the matricial setting, this is equivalent to factorizing $\boldsymbol{Z} = (\widehat{\boldsymbol{A}} \otimes \boldsymbol{I}_N)\widehat{\boldsymbol{X}}$. We shall refer to this factorization of an array of matrices as Free Component Factorization (FCF).

To compute $\widehat{\boldsymbol{A}}$ in FCF as prescribed by Corollary VII.17 we must compute the matricial analog $\boldsymbol{Y}$ of $\boldsymbol{y}$ in (7.22). This involves first computing the matricial $s \times s$ covariance matrix analog as in Algorithm 1 where we have replaced the $\varphi(\cdot)$ and $\varphi_1(\cdot)$ in the self-adjoint and the rectangular settings with their matricial analogs as in (6.11) and (6.13), respectively.

Having computed the whited array of matrices $\boldsymbol{Y}$ we can compute the matrix $\widehat{\boldsymbol{A}}$ via Algorithm 2 where the dot operator is as defined next.

Definition VII.18 (Dot operator)**.** Let $\boldsymbol{Y} = [\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_s]^T$ be an array of matrices where $\boldsymbol{Y}_i \in \mathbb{R}^{N \times M}$. Let a function $F : \mathbb{R}^{N \times M} \mapsto \mathbb{R}$, we have that

$$F.(\boldsymbol{Y}) := \begin{bmatrix} F(\boldsymbol{Y}_1) \\ \vdots \\ F(\boldsymbol{Y}_s) \end{bmatrix}. \tag{7.31}$$

Algorithm 1 Free whitening for random matrices

Input: $\boldsymbol{Z} = [\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_s]^T$ where $\boldsymbol{Z}_i$ are $N \times M$ matrices. $M = N$ if $\boldsymbol{Z}_i$ are self-adjoint.

1. Compute $\overline{\boldsymbol{Z}} = [\overline{\boldsymbol{Z}}_1, \cdots, \overline{\boldsymbol{Z}}_s]^T$, where

$$\overline{\boldsymbol{Z}}_i = \begin{cases} \frac{1}{N}\operatorname{Tr}(\boldsymbol{Z}_i)\boldsymbol{I}_N & \text{if } \overline{\boldsymbol{Z}}_i \text{ are self-adjoint,} \\ \operatorname{mean}(vec(\boldsymbol{Z}_i)) \times \operatorname{ones}(N, M) & \text{if } \overline{\boldsymbol{Z}}_i \text{ are rectangular.} \end{cases}$$

2. Compute $\widetilde{\boldsymbol{Z}} = \boldsymbol{Z} - \overline{\boldsymbol{Z}}$ and $s \times s$ empirical covariance matrix $\boldsymbol{C}$ where for $i, j = 1, \ldots, s$:

$$\boldsymbol{C}_{ij} = \frac{1}{N}\operatorname{Tr}(\widetilde{\boldsymbol{Z}}_i \widetilde{\boldsymbol{Z}}_j^H).$$

3. Compute eigen-decomposition , $\boldsymbol{C} = \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^T$.
4. Compute $\boldsymbol{Y} = ((\boldsymbol{U}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^T) \otimes \boldsymbol{I}_N)\widetilde{\boldsymbol{Z}}$.
5. **return:** $\boldsymbol{Y}, \boldsymbol{\Sigma}$, and $\boldsymbol{U}$.

Algorithm 2 Free Component Factorization (FCF) of an array of matrices

Input: Array of matrices $\boldsymbol{Z} = [\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_s]^T$ where $\boldsymbol{Z}_i$ are $N \times M$ matrices. 1. Compute $\boldsymbol{Y}, \boldsymbol{\Sigma}, \boldsymbol{U}$ by applying Algorithm 1 to $\boldsymbol{Z}$.

2. Compute [1]

$$\widehat{\boldsymbol{W}} = \underset{\boldsymbol{W} \in O(n)}{\arg\min} \sum_{i=1}^{s} |\widehat{F}.\left(\widetilde{\boldsymbol{W}}^T \boldsymbol{Y}\right)|, \quad \text{where } \widetilde{\boldsymbol{W}} = \boldsymbol{W} \otimes \boldsymbol{I}_N.$$

3. Compute $\widehat{\boldsymbol{A}} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^T\widehat{\boldsymbol{W}}$ and $\widehat{\boldsymbol{X}} = (\widehat{\boldsymbol{A}}^{-1} \otimes \boldsymbol{I}_N)\boldsymbol{Z}$.
4. Sort components of $\widehat{\boldsymbol{X}}$ by magnitude of $\widehat{F}(\cdot)$
5. Permute the columns of $\widehat{\boldsymbol{A}}$ such that $\boldsymbol{Z} = (\widehat{\boldsymbol{A}} \otimes \boldsymbol{I}_N)\widehat{\boldsymbol{X}}$.
6. **return:** $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{X}}$

[1] Here $\widehat{F}(\cdot)$ is either the (self-adjoint or rectangular) free kurtosis, the free entropy or a higher (than fourth) order (even valued) free cumulant. See Table 7.1.

7.1.7 Numerical algorithms for Free Component Factorization

The manifold optimization problem in FCF can be solved using a gradient descent with retraction method [24, 92].

Theorem VII.19 (Gradient of the objective function;)**.** *Let* $\boldsymbol{Y} = [\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_s]$ *and* $\boldsymbol{W} = [\boldsymbol{w}_1, \cdots, \boldsymbol{w}_s] \in \mathbb{R}^{s \times s}$*. Suppose*

$$\widetilde{\boldsymbol{W}} = \boldsymbol{W} \otimes \boldsymbol{I}_N,$$

Then the gradient

$$\partial_{\boldsymbol{W}_{k\ell}} \sum_{i=1}^{s} \widehat{F}. \left(\widetilde{\boldsymbol{W}}^T \boldsymbol{Y} \right),$$

depends on whether $\boldsymbol{Y}$ *is an array of self-adjoint or rectangular matrices.*

Suppose $\widehat{F}(\cdot)$ *is chosen to be free kurtosis or free entropy for the self-adjoint or rectangular setting as in Table 7.1, then the gradient is given by the corresponding expression in Table 7.2 where*

$$\boldsymbol{X}_\ell = \widetilde{\boldsymbol{w}}_\ell^T \boldsymbol{Y},$$

and $\widetilde{\boldsymbol{w}}_\ell = \boldsymbol{w}_\ell \otimes \boldsymbol{I}_N$*.*

Armed with these gradients we can compute the free component factorization of an array of matrices using numerical solvers for manifold optimization, such as for example the `manopt` [24] package (for `MATLAB`) or the `Optim.jl` [92] package for `Julia`. Our software implemntation via the `FCA.jl` package [94] does precisely this.

Table 7.1: Formulas for $\widehat{F}(\cdot)$ in Algorithm 2. Here $\boldsymbol{X}$ is either a self-adjoint or a rectangular matrix.

	self-adjoint FCF	rectangular FCF
free kurtosis	$\widehat{F}(\boldsymbol{X}) = -\left\|\widehat{\kappa}_4(\boldsymbol{X})\right\|$, where $\widehat{\kappa}_4(\boldsymbol{X}) = \frac{1}{N}\operatorname{Tr}(\boldsymbol{X}^4) - 2\left[\frac{1}{N}\operatorname{Tr}(\boldsymbol{X}^2)\right]^2$	$\widehat{F}(\boldsymbol{X}) = -\left\|\widehat{\kappa}_4(\boldsymbol{X})\right\|$, where $\widehat{\kappa}_4(\boldsymbol{X}) = \frac{1}{N}\operatorname{Tr}((\boldsymbol{X}\boldsymbol{X}^H)^2) - \left(1+\frac{N}{M}\right)\left[\frac{1}{N}\operatorname{Tr}(\boldsymbol{X}\boldsymbol{X}^H)\right]^2$
free entropy	$\widehat{F}(\boldsymbol{X}) = \widehat{\chi}(\boldsymbol{X})$, where $\widehat{\chi}(\boldsymbol{X}) = \sum_{i\neq j}\frac{\log\lvert\lambda_i-\lambda_j\rvert}{N(N-1)}$, λ_i are eigenvalues of $\boldsymbol{X}$	Set $\alpha = \frac{N}{N+M}$ and $\beta = \frac{M}{N+M}$, $\widehat{F}(\boldsymbol{X}) = \chi(\boldsymbol{X})$, where $\widehat{\chi}(\boldsymbol{X}) = \frac{\alpha^2}{N(N-1)}\sum_{i\neq j}\log\lvert\lambda_i-\lambda_j\rvert + \frac{(\beta-\alpha)\alpha}{N}\sum_{i=1}^{N}\log\lambda_i$, λ_i are the eigenvalues of $\boldsymbol{X}\boldsymbol{X}^H$

7.2 Numerical Simulations

We will now validate the unmixing performance of FCA on additive mixtures on random matrices and compare the unmixing performance with that of ICA. To that end, we first define a permutation invariant unmixing error metric that is also invariant to scaling and sign ambiguities.

Definition VII.20 (Unmixing Error Metric)**.** Let $\boldsymbol{A}$ be the mixing matrix in (6.10) and $\widehat{\boldsymbol{A}}$ be an estimate of the mixing matrix. The scaling and permutation invariant unmixing error is defined as

$$\mathcal{E}(\boldsymbol{A}, \widehat{\boldsymbol{A}}) = \min_{\boldsymbol{D}\in\mathcal{D},\boldsymbol{P}\in\Pi} ||\boldsymbol{P}\boldsymbol{D}\widehat{\boldsymbol{A}}^{-1}\boldsymbol{A} - \boldsymbol{I}||_F, \tag{7.32}$$

where $\mathcal{D}$ denotes the set of non-singular diagonal matrices and Π denotes the set of

Table 7.2: Euclidean gradients for the setting in Theorem VII.19

	self-adjoint FCF	rectangular FCF
free kurtosis	$\partial_{\boldsymbol{W}_{k\ell}} \sum_{i=1}^{s} \widehat{F}.\left(\widetilde{\boldsymbol{W}}^T \boldsymbol{Y}\right) =$ $-\operatorname{sign}(\widehat{\kappa}_4(\boldsymbol{X}_\ell)) \times \left[\frac{4}{N} \operatorname{Tr}(\boldsymbol{Y}_k \boldsymbol{X}_\ell^3) - \frac{8}{N^2} \operatorname{Tr}(\boldsymbol{X}_\ell^2) \operatorname{Tr}(\boldsymbol{Y}_k \boldsymbol{X}_\ell)\right]$	$\partial_{\boldsymbol{W}_{k\ell}} \sum_{i=1}^{s} \widehat{F}.\left(\widetilde{\boldsymbol{W}}^T \boldsymbol{Y}\right) = -\operatorname{sign}(\widehat{\kappa}_4(\boldsymbol{X}_\ell))$ $\times \left[\frac{4}{N} \operatorname{Tr}(\boldsymbol{Y}_k \boldsymbol{X}_\ell^H \boldsymbol{X}_\ell \boldsymbol{X}_\ell^H) - \left(1 + \frac{N}{M}\right) \frac{4}{N^2} \operatorname{Tr}(\boldsymbol{X}_\ell \boldsymbol{X}_\ell^H) \operatorname{Tr}(\boldsymbol{Y}_k \boldsymbol{X}_\ell^H)\right]$
free entropy	Let $\lambda_i^{(\ell)}$ and $v_i^{(\ell)}$ be the $\boldsymbol{X}_\ell$'s eigenvalues and eigenvectors $\partial_{\boldsymbol{W}_{k\ell}} \sum_{i=1}^{s} \widehat{F}.\left(\widetilde{\boldsymbol{W}}^T \boldsymbol{Y}\right) =$ $\sum_{i \neq j} \frac{\partial_{\boldsymbol{W}_{k\ell}} \lambda_i^{(\ell)} - \partial_{\boldsymbol{W}_{k\ell}} \lambda_j^{(\ell)}}{N(N-1)(\lambda_i^{(\ell)} - \lambda_j^{(\ell)})}$ with $\partial_{\boldsymbol{W}_{k\ell}} \lambda_i^{(\ell)} = (v_i^{(\ell)})^H \boldsymbol{Y}_k v_i^{(\ell)}$	Let $\lambda_i^{(\ell)}$ and $v_i^{(\ell)}$ be the eigenvalues and eigenvectors of $\boldsymbol{X}_\ell \boldsymbol{X}_\ell^H$. $\partial_{\boldsymbol{W}_{k\ell}} \sum_{i=1}^{s} \widehat{F}.\left(\widetilde{\boldsymbol{W}}^T \boldsymbol{Y}\right) =$ $\frac{\alpha^2}{N(N-1)} \sum_{i \neq j} \frac{\partial_{\boldsymbol{W}_{k\ell}} \lambda_i^{(\ell)} - \partial_{\boldsymbol{W}_{k\ell}} \lambda_j^{(\ell)}}{\lambda_i^{(\ell)} - \lambda_j^{(\ell)}}$ $+ \frac{(\beta - \alpha)\alpha}{N} \sum_{i=1}^{N} \frac{\partial_{\boldsymbol{W}_{k\ell}} \lambda_i^{\ell}}{\lambda_i^{(\ell)}}$ with $\partial_{\boldsymbol{W}_{k\ell}} \lambda_i^{(\ell)} = (v_i^{(\ell)})^H (\boldsymbol{Y}_k \boldsymbol{X}_\ell^H + \boldsymbol{X}_\ell \boldsymbol{Y}_k^H) v_i^{(\ell)}$.

(square) permutation matrices.

We shall utilize this metric to compare FCA and ICA in what follows.

7.2.1 Unmixing of self-adjoint matrices using self-adjoint FCA

We now verify Theorems VII.15, VII.16 and Corollary VII.17 by showing that self-adjoint FCA can successfully, while not perfectly, unmix mixtures of self-adjoint matrices.

Let $\boldsymbol{G}_1 \in \mathbb{R}^{N\times N}$ and $\boldsymbol{G}_2 \in \mathbb{R}^{N\times M}$ be two independent Gaussian matrices composed of i.i.d. $\mathcal{N}(0,1)$ entries. Define

$$\boldsymbol{X}_1 = \frac{\boldsymbol{G}_1 + \boldsymbol{G}_1^T}{\sqrt{2N}} \quad \text{and} \quad \boldsymbol{X}_2 = \frac{\boldsymbol{G}_2\boldsymbol{G}_2^T}{M}. \tag{7.33}$$

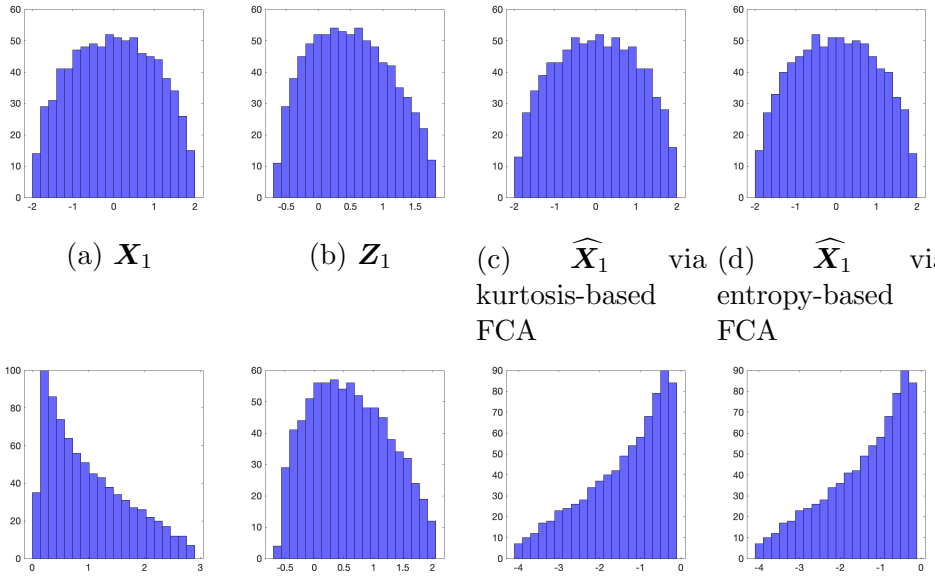
The matrices $\boldsymbol{X}_1, \boldsymbol{X}_2$ are self-adjoint by construction, and their eigen-spectra are displayed in Figures 7.1a and 7.1e respectively. In the parlance of random matrix theory [52], $\boldsymbol{X}_1$ is a matrix drawn from the Gaussian orthogonal ensemble (GOE) and its limiting eigen-distribution obeys the semi-circle distribution, while $\boldsymbol{X}_2$ is a matrix drawn from Laguerre orthogonal ensemble (LOE) and its limiting eigen-distribution obeys the Marčenko-Pastur distribution.

We now mix the matrices as in (6.10) for a non-singular

$$\boldsymbol{A} = \begin{bmatrix} 0.5 & 0.5 \\ -0.5 & 0.6 \end{bmatrix}.$$

The eigen-spectra of the mixed matrices $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are plotted in Figures 7.1b and 7.1f.

The distributions of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are orthogonally invariant, and according to the discussion following Definition B.16, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are asymptotically free. Moreover, only one matrix ($\boldsymbol{X}_1$) has a limiting eigen-distribution that converges to that of an



(a) $\boldsymbol{X}_1$ (b) $\boldsymbol{Z}_1$ (c) $\widehat{\boldsymbol{X}_1}$ via kurtosis-based FCA (d) $\widehat{\boldsymbol{X}_1}$ via entropy-based FCA

(e) $\boldsymbol{X}_2$ (f) $\boldsymbol{Z}_2$ (g) $\widehat{\boldsymbol{X}_2}$ via kurtosis-based FCA (h) $\widehat{\boldsymbol{X}_2}$ via entropy-based FCA

Figure 7.1: An experiment in the separation of self-adjoint matrices. The mixing matrix $\boldsymbol{A} = [0.5, 0.5; -0.5, 0.6]$, $N = 800, M = 1600$. The average errors over 200 trials are 8.67×10^{-3} for kurtosis-based FCF and 6.44×10^{-3} for entropy-based FCF.

abstract free semicircular element. Hence, we can apply self-adjoint FCA to factorize $\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{Z}_2 \end{bmatrix}^T$ using Algorithm 2 and obtain estimates $\widehat{\boldsymbol{A}}$, $\widehat{\boldsymbol{X}_1}$ and $\widehat{\boldsymbol{X}_2}$ which should be good estimates of $\boldsymbol{A}$, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ respectively.

Figures 7.1c and 7.1g display the eigen-spectra of the matrices $\widehat{\boldsymbol{X}_1}$ and $\widehat{\boldsymbol{X}_2}$ returned by self-adjoint free kurtosis-based FCA. Comparing Figures 7.1c, 7.1g and 7.1a, 7.1e reveals that free kurtosis-based self-adjoint FCA successfully unmixes the mixed matrices well. Figures 7.1d and 7.1h show that free entropy based self-adjoint FCA successfully unmixes the mixed matrices. Both free kurtosis and free entropy based unmixing have comparably small but not zero error, which we compute over 200 Monte-Carlo realizations. This is expected since the matrices are asymptotically free and the simulations are with finite dimensional matrices.

7.2.2 Unmixing of rectangular matrices with rectangular FCA

We now show that the rectangular FCA can successfully, while not perfectly, unmix mixtures of rectangular matrices. To that end, we let $\boldsymbol{X}_2$ be an $N \times M$ Gaussian matrix with i.i.d. $\mathcal{N}(0, 1/M)$ entries and set $\boldsymbol{X}_1 = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T$ where

$$\begin{aligned}\boldsymbol{U}(i,j) &= \frac{2}{N}\sin\left[\frac{\pi}{2N}(i+1)(2j+1)\right] \qquad \text{for } 1 \leq i,j \leq N \\ \boldsymbol{V}(i,j) &= \frac{2}{M}\cos\left[\frac{\pi}{2M}i(2j+1)\right] \qquad \text{for } 1 \leq i,j \leq M,\end{aligned}$$

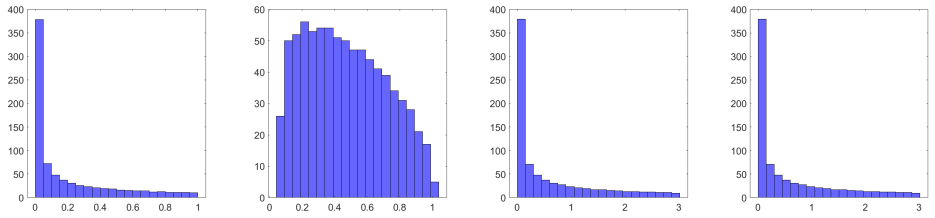
so that $\boldsymbol{U}$ and $\boldsymbol{V}$ thus constructed are orthogonal matrices. We pick a 'nice' function $f(z)$ and set the diagonal matrix $\boldsymbol{D} \in \mathbb{R}^{N\times M}$ such that $\boldsymbol{D}(i,i) = f((i-1/2)/N)), i = 1, \cdots, N$.

The singular value spectra of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are plotted in Figures 7.2a and 7.2e. As before, we mix the matrices as in (6.10). Figures 7.2b and 7.2f display the singular value spectra of $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$.

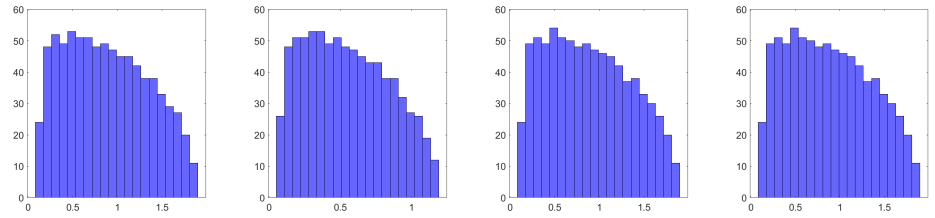
We now note that the singular value distributions of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ converge to a non-random limit and that the distribution of $\boldsymbol{X}_2$ is bi-orthogonally invariant. Thus, following the discussion after Definition (B.17), $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are asymptotically free. Moreover, only $\boldsymbol{X}_2$ has a limiting distribution which converges to that of an abstract free Poisson rectangular element.

Hence, we can apply rectangular FCA to factorize $\boldsymbol{Z} = \begin{bmatrix}\boldsymbol{Z}_1 & \boldsymbol{Z}_2\end{bmatrix}^T$ using Algorithm 2 and obtain estimates $\widehat{\boldsymbol{A}}$, $\widehat{\boldsymbol{X}_1}$ and $\widehat{\boldsymbol{X}_2}$ which should be good estimates of $\boldsymbol{A}$, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ respectively.

Figures 7.2c and 7.2g display the eigen-spectra of the matrices $\widehat{\boldsymbol{X}_1}$ and $\widehat{\boldsymbol{X}_2}$ returned by rectangular free kurtosis-based FCA. Comparing Figures 7.2c, 7.2g and 7.2a, 7.2e reveals that rectangular free kurtosis-based FCA successfully unmixes the mixed matrices well. Figures 7.1d and 7.1h show that rectangular free entropy based FCA successfully unmixes the mixed matrices. Both free kurtosis and free entropy



(a) $\boldsymbol{X}_1$ (b) $\boldsymbol{Z}_1$ (c) $\widehat{\boldsymbol{X}_1}$ via kurtosis-based FCA (d) $\widehat{\boldsymbol{X}_1}$ via entropy-based FCA



(e) $\boldsymbol{X}_2$ (f) $\boldsymbol{Z}_2$ (g) $\widehat{\boldsymbol{X}_2}$ via kurtosis-based FCA (h) $\widehat{\boldsymbol{X}_2}$ via entropy-based FCA

Figure 7.2: An experiment in the separation of rectangular random matrices. In this simulation, $\boldsymbol{A} = [0.5, 0.5; -0.5, 0.6]$, $N = 800, M = 1000$ and $f(x) = (x-1)^4$. The average errors over 200 trials are 1.55×10^{-3} for kurtosis-based FCF and 8.81×10^{-4} for entropy-based FCF.

based unmixing have comparably small but not zero error, which we compute over 200 Monte-Carlo realizations. This is expected since the matrices are asymptotically free but the simulations are with finite dimensional matrices.

7.2.3 Unmixing mixed images using rectangular FCA

We now consider the problem of unmixing mixed images. Grayscale images can be viewed as matrices so rectangular FCA can be used to separate the mixed images. We can also apply ICA to unmix the images via reshaping images to vectors and we shall compare the unmixing performance of FCA with that of ICA. Algorithm 9 in Appendix B.3 describes the independent component factorization (ICF) mirroring the language we used for FCF.

We set $\boldsymbol{X}_1$ to be the grayscale image of the locust in Figure 6.4a. The matrix $\boldsymbol{X}_2$

is a Gaussian random matrix of the same size as $\boldsymbol{X}_1$ with i.i.d. zero mean, uniform variance entries as depicted in Figure 6.4e. We mix the matrices following (6.10) and display the mixed images in Figures 6.4b and 6.4f.

Next, we apply (rectangular) free kurtosis based FCA to the mixed images $\boldsymbol{Z} = [\boldsymbol{Z}_1, \boldsymbol{Z}_2]^T$ and display the unmixed image that is closes to that of the locust in Figure 6.4d. The unmixed image obtained by using (classical) kurtosis based ICA is displayed in Figure 6.4c. Both methods return unmixed images that are close to the original image of the locust. A closer inspection of Figures 6.4d 6.4c reveals that FCA better unmixes the images than ICA as illustrated in Figures 6.4h and 6.4g. Quantitatively speaking, when averaged over 200 Monte-Carlo realizations of the noise, we find that the denoising error for kurtosis-based FCF is 5.35×10^{-3} while the error for kurtosis-based ICF is 2.42×10^{-1}, thereby illustrating the superiority of FCF over ICF for this task.

To gain additional insight on the improved unmixing performance of FCA relative to ICA for this example, we investigate the landscape of their respective objective functions. To that end we first note that the mixing matrix

$$\boldsymbol{A} = \frac{1}{2}\begin{bmatrix} \sqrt{2} & \sqrt{2} \\ \sqrt{2} & -\sqrt{2} \end{bmatrix},$$

is orthogonal and so we can recast the spherical manifold optimization problem underlying FCA and ICA as a 1-D optimization problem in polar coordinates. In other words, we can parameterize the optimization problem in (6.5) and (6.9) in terms of $\boldsymbol{w} := \boldsymbol{w}_\theta = \begin{bmatrix} \cos(\theta) & \sin(\theta) \end{bmatrix}^T$. Similarly, optimization (7.15) can be parameterized with

$\boldsymbol{W} = [\cos(\theta)\sin(\theta); -\sin(\theta), \cos(\theta)]^T$. We compute and display the landscape of the objective functions corresponding to maximization of the classical kurtosis $|c_4(\theta)|$, free kurtosis $|\kappa_4(\theta)|$ and the free entropy $E(\theta)$ for $\theta \in [0, 2\pi]$ in Figures 7.3a, 7.3b and

7.3c, respectively.

The dashed red line in these figures corresponds to the ground truth freely independent component direction associated with $\theta_1 = \pi/4$ associated with the first column of the mixing matrix $\boldsymbol{A}$; the other direction (not displayed) is orthogonal and corresponds to the second column of $\boldsymbol{A}$ and is associated with $\theta_2 = 3\pi/4$.

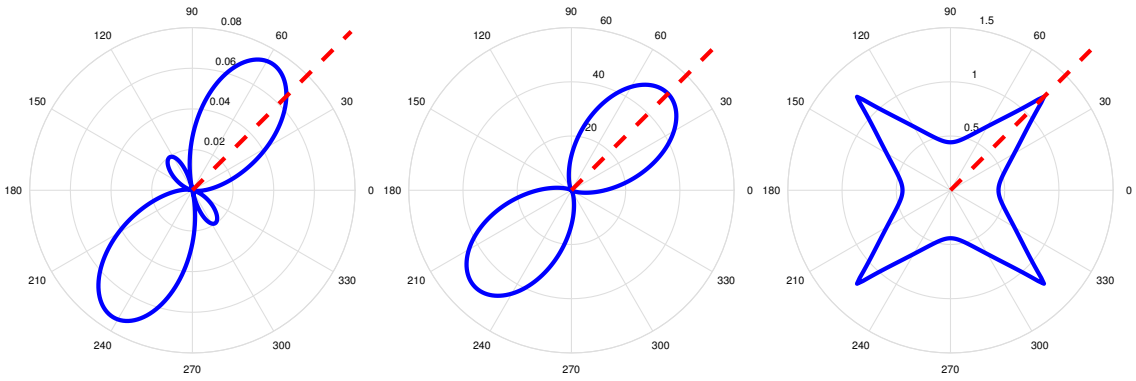
Figures 7.3b and 7.3c, reveal that $|\kappa(\theta)|$ and $E(\theta)$ are maximized at angles θ very close to $\theta_1 = \pi/4$. In contrast, Figure 7.3a reveals that $|c(\theta)|$ is maximized at an angle θ further away from θ_1 than is the case for the FCA algorithms. This is why FCA better unmixes the images than ICA.

There is a more interesting story in these plots. Figure 7.3b shows that the $|\kappa(\theta_2)|$ for $\theta_2 = 3\pi/4$ is very close to zero, as expected because $\boldsymbol{X}_2$ is a Gaussian random matrix and in the large matrix limit the free rectangular kurtosis of its free counterpart is identically zero. The classical kurtosis of a Gaussian random variable is also zero. A closer inspection reveals that the classical kurtosis of the locust image is also close to zero (the scale of the polar plot initially obscures this fact!) while its free kurtosis is significantly greater than zero (or that of $\boldsymbol{X}_2$).

The fact that the locust image $\boldsymbol{X}_1$ and the Gaussian image $\boldsymbol{X}_2$ have a higher "contrast" in their free kurtosis relative to their classic kurtosis is why FCA does better at unmixing them than ICA. Figure 6.5 captures this perspective and suggests a direction for future research in more precisely defining how the "contrasts" between the scalar (or ICA) and matrix (or FCA) embeddings affects the realized unmixing performance.

7.2.4 Unmixing performance of free kurtosis vs free entropy FCA vs ICA

We now compare the performance of FCF and ICF as a function of the dimensionality of the system, since the errors in FCF and ICF are both governed by the deviation from some limiting large sample quantities (or large matrix size). Here,



(a) Polar graph of $|c(\theta)|$ (b) Polar graph of $|\kappa(\theta)|$ (c) Polar graph of $E(\theta)$

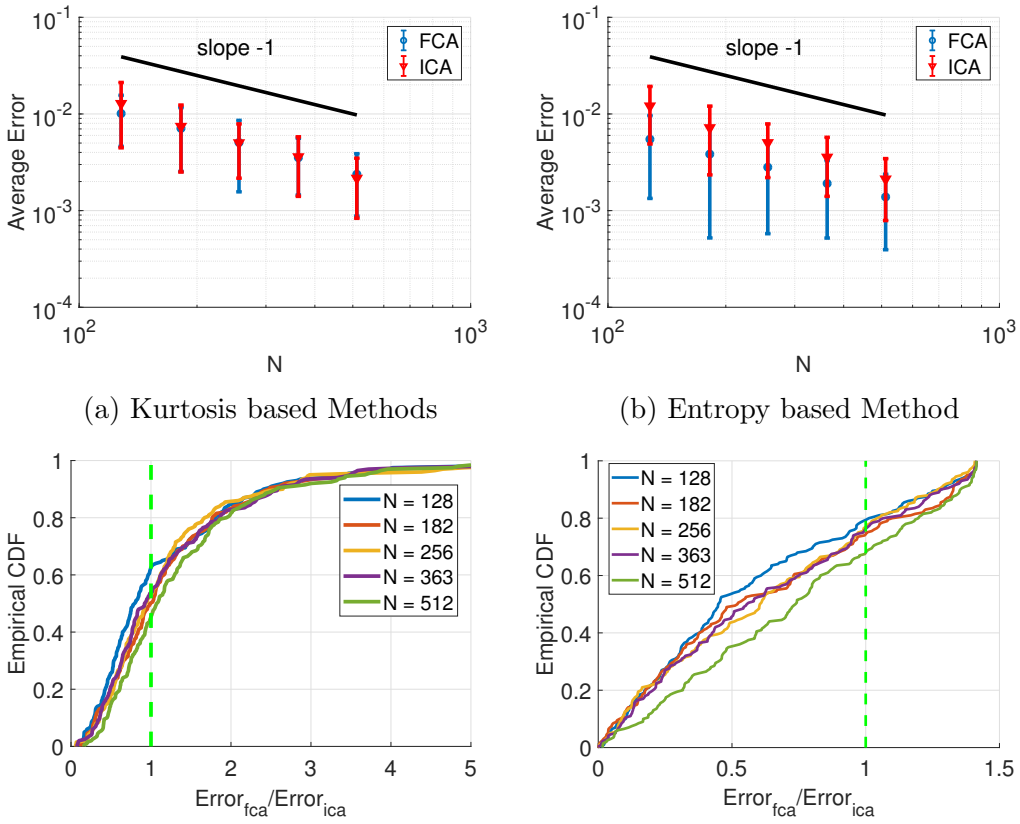
Figure 7.3: (a) Polar graph of $c(\theta)$ (b) Polar graph of $\kappa(\theta)$ (c) Polar graph of $E(\theta)$. The red dashed lines stand for the direction of 45°. Note that the directions of maximum of $\kappa(\theta)$ and $E(\theta)$ agree with the red line well, while the maximum of $c(\theta)$ is off. Because of the randomness of Gaussian noise, for different trials, the polar graph of ICA will vary. However, the polar graphs of FCA are relatively stable.

we adopt the same setup as in Section 7.2.2 with whitened $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ matrics and $\boldsymbol{A} = [\sqrt{2}, \sqrt{2}; -\sqrt{2}, \sqrt{2}]/2$. We increase N, M in a fixed ratio, and obtain an estimate of the unmixing matrix using kurtosis based FCA, entropy based FCA, kurtosis based ICA and entropy based ICA and compute the unmixing error over 200 Monte-Carlo realizations.

Figures 7.4a and 7.4c show that free kurtosis based FCA and kurtosis based ICA realize similar unmixing performance. However, Figures 7.4b and 7.4d show that free entropy based FCA has a lower error than entropy based ICA, even while both have errors that decay at the same rate.

7.2.5 Unmixing mixed waveforms using rectangular FCA

Let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ denote two vectors representing the audio signals whose waveforms are displayed in Figures 6.3a and 6.3e, respectively. Their mixture produces signals whose waveforms are displayed in Figures 6.3b and 6.3f, respectively. This is the famous cocktail party problem [64] and ICA is known to succeed in unmixing the mixed signals. Figures 6.3c and 6.3g confirm that it does.



(a) Kurtosis based Methods

(b) Entropy based Method

(c) $\text{Error}_{fca}/\text{Error}_{ica}$ for Kurtosis methods

(d) $\text{Error}_{ica}/\text{Error}_{fca}$ for Entropy method

Figure 7.4: (a) Averaged (over 200 trials) errors of kurtosis based FCA and ICA for increasing dimension. (b) Averaged error of entropy based FCA and ICA for increasing dimension. (c) CDF of $\text{Error}_{fca}/\text{Error}_{ica}$ for the kurtosis-based method. (d) cdf of $\text{Error}_{fca}/\text{Error}_{ica}$ for the entropy-based method. All methods appear to have a convergence rate of N^{-1}. In this simulation, we set $N/M = 0.8$ and $f(x) = (x-1)^4$.

In this setting the mixed waveforms are modeled as

$$\begin{bmatrix} \boldsymbol{z}_1^T \\ \vdots \\ \boldsymbol{z}_s^T \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_s^T \end{bmatrix}. \tag{7.34}$$

There is no matrix in sight in (7.34), so we can seemingly not apply FCA directly.

The key insight is that we are at liberty to design a linear matrix embedding operator $\mathcal{M} : \boldsymbol{x} \in \mathbb{R}^n \mapsto \boldsymbol{X} \in \mathbb{F}^{M \times N}$ for some integer M and N. A simple example is by reshaping the $n = MN$ vector into an $M \times N$ matrix. Here linearity implies that for any scalars α and β we have that

$$\mathcal{M}(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) = \alpha \mathcal{M}(\boldsymbol{x}) + \beta \mathcal{M}(\boldsymbol{y}).$$

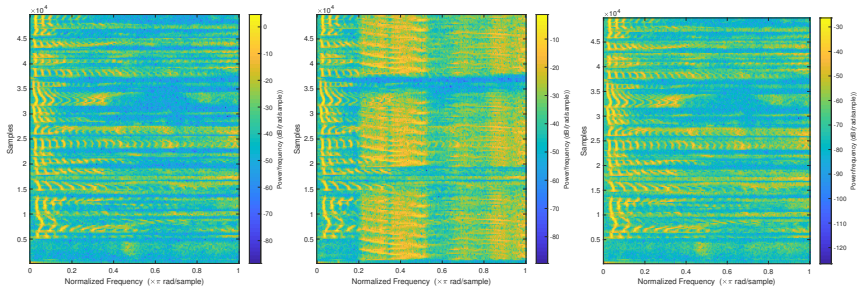
Then as a consequence of the linearity of the embedding operator we have that

$$\begin{bmatrix} \mathcal{M}(\boldsymbol{z}_1)^T \\ \vdots \\ \mathcal{M}(\boldsymbol{z}_s)^T \end{bmatrix} = (\boldsymbol{A} \otimes \boldsymbol{I}_N) \begin{bmatrix} \mathcal{M}(\boldsymbol{x}_1)^T \\ \vdots \\ \mathcal{M}(\boldsymbol{x}_s)^T \end{bmatrix} \tag{7.35}$$

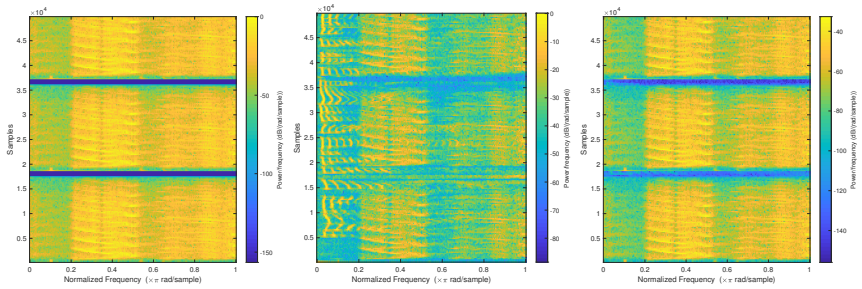
so that it fits (6.10) and we can apply FCA to estimate the mixing matrix and thus unmix the mixed waveforms.

For the cocktail party problem we used a (complex-valued) spectrogram embedding, as described in Algorithm 7, and computed the mixed (complex-valued) spectrogram matrices $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ displayed in Figure 7.5. Since the mixing matrix is real-valued we modified the FCA algorithms slightly by whitening using only real part of the covariance matrix.

Figures 6.3d and 6.3h show that FCA on the complex-valued spectrogram matrices successfully unmixes the complex-valued spectograms of the latent waveforms.



(a) $\boldsymbol{X}_1$ (b) $\boldsymbol{Z}_1$ (c) $\widehat{\boldsymbol{X}}_1$ via FCA with spectrogram embedding

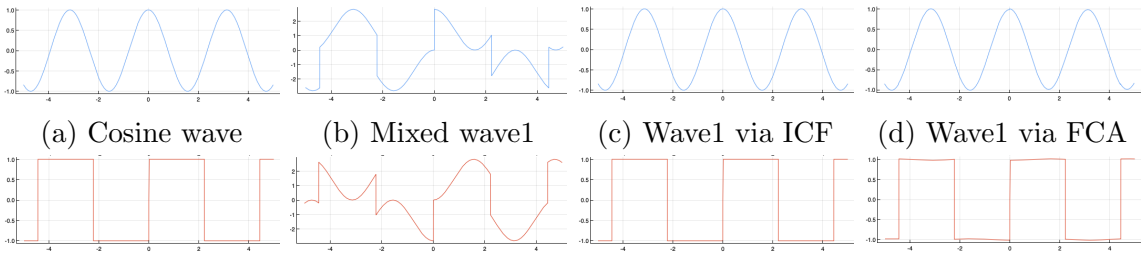


(d) $\boldsymbol{X}_2$ (e) $\boldsymbol{Z}_2$ (f) $\widehat{\boldsymbol{X}}_2$ via FCA with spectrogram embedding

Figure 7.5: The spectrogram of signals. For the spectrogram, we adapt the Hamming window of 250 samples, the number of overlapped samples is 125, the number of DFT points is 256.

Figure 6.2 shows that FCA succeeds in unmixing the waveforms and that FCA better unmixes the waveforms than ICA. Figure 7.6 illustrates a setting where ICA does better.

These experiments illustrate our general point that FCA can be used wherever ICA has been used and that they perform comparably well. The key step is embedding a vector waveform as a matrix in a way that preserves the mixing model. We used the spectrogram embedding here – other linear embeddings could be used as well. Determining the optimal embedding so we can reason about why FCA does better than ICA for the setup in Figure 6.2 but does not for the setup in Figure 7.6 is a natural next question.



(a) Cosine wave (b) Mixed wave1 (c) Wave1 via ICF (d) Wave1 via FCA

(e) Square wave (f) Mixed wave2 (g) Wave2 via ICA (h) Wave2 via FCA

Figure 7.6: An experiment in waveform separation using ICA and FCA. Note that subplots (c), (g) (unmixed waves via ICA) and (d), (h) (unmixed waves via FCA) both recover (a), (e). In this simulation, $\boldsymbol{A} = [\sqrt{2}, \sqrt{2}; -\sqrt{2}, \sqrt{2}]/2$ in (6.10). The errors for ICA and FCA are 7.70×10^{-5} and 1.36×10^{-2} respectively. .

7.2.6 Unmixing rectangular matrices using self-adjoint FCA and more

We can take this embedding idea even further by embedding mixed rectangular matrices modeled as (7.34) and embedding them as self-adjoint matrices as described in Algorithm 6 and then using self-adjoint FCA to unmix them. Or, we may even embed rectangular matrices into another rectangular matrix with a different number of rows and columns as described in Algorithm 6. Determining the right matricial embedding adds another aspect to the question of optimal embedding selection as in Figure 6.5.

7.3 Conclusions and Open Problems

We have developed free component analysis as a non-commutative analog of independent component analysis. We proved that when certain identifiability conditions are met then mixtures of self-adjoint and rectangular variants can be unmixed using self-adjoint and non self-adjoint/rectangular FCA. We developed an algorithm for umixing mixtures of matrices based on FCA and demonstrated how FCA can be used to unmix images (viewed as matrices), speech signals and waveforms (when embedded as spectrogram matrices) and images where it initially fails (via FCA on a free subset of the mixed images).

7.3.1 Open Problems

We now list some directions for future research. These include developing a non-linear extension of FCA analogous to non-linear ICA [56, 26, 67, 72, 3, 74], a fast algorithm for FCA analogous to fast ICA [70, 35, 96] and algorithms for sparse FCA analogous to sparse ICA [44, 23].

A more general line of inquiry is related to the so-called one-unit learning work in ICA. In ICA, it is known that instead of maximizing the kurtosis, we can equivalently maximize a large class of so-called contrast functions $G(\cdot)$ of the form [69, Equation (2)]

$$J_G(\boldsymbol{w}) = \mathbb{E}_x[G(\boldsymbol{w}^T x)] - \mathbb{E}\gamma[G(\gamma)],$$

where $G(\cdot)$ is non-quadratic well-behaving even function and γ is a standardized Gaussian random variable. Developing the analog of this theory for the self-adjoint and rectangular FCA settings will allow for a finer study of the statistical efficiency of the FCA algorithms in the finite matrix setting akin to the work by Arora et al. [7] and facilitate the development of asymptotically consistent and statistically efficient estimators akin to the work by Chen and Bickel [35].

Our simulations showed that free entropy based FCA outperformed free kurtosis based FCA (see Figure 7.4d). Computing the free entropy is computationally more expensive than computing the free kurtosis. In ICA, the mutual information is approximated via a cumulant expansion [43, Section 3.1, pp. 295]. Developing a rapidly converging approximation to free entropy in terms of the free cumulants that converges faster than the approximation in [90, Exercise 5, pp. 190] would lead a faster FCA that we expect to be statistically more efficient than free kurtosis based FCA.

7.3.2 Open Problem: Using FCA to construct new matrix models for freeness

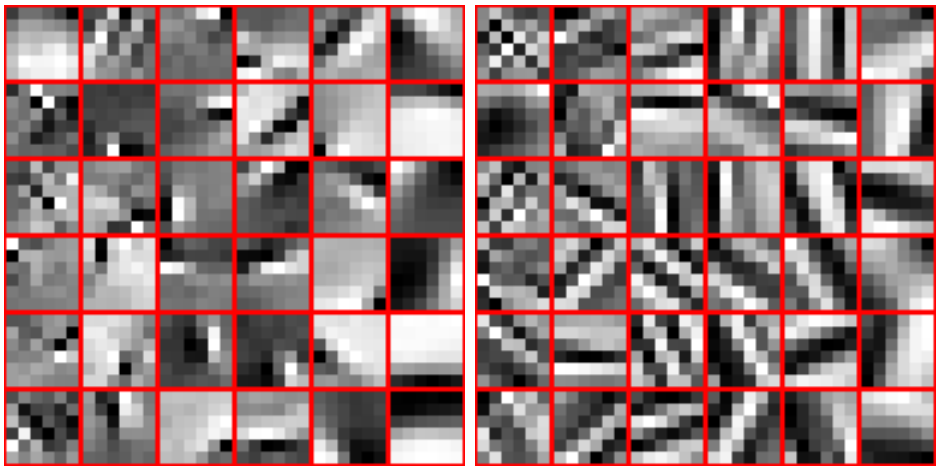
We can use the ICA to decompose small patches of an image into linear independent combinations of ICA basis vectors that can be learned from the data via the ICA factorization [66, 17]. Figure 7.7a displays the 36 ICA bases patches thus obtained by reshaping into 6×6 matrices the 36 ICA bases vectors corresponding to each of the columns of the 36×36 $\boldsymbol{W}_{\mathsf{ica}}$ matrix obtained by an kurtosis based ICA factorization of the 6×6 patches of the panda image in Figure 6.1a.

We can similarly use FCA to decompose the patches of an image into "as free as possible" matrices. Figure 7.7c shows the 36 free patch bases obtained by displaying the matricial elements of the $\boldsymbol{X}_{\mathsf{fca}}$ array of matrices of the panda image. Each sub-image in the panda is a linear combination of these free patches.

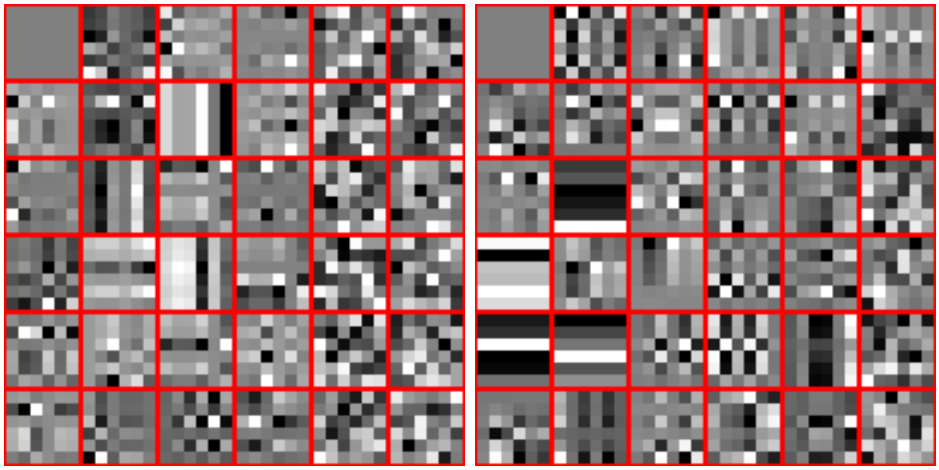
The patch FCA versus patch ICA bases vectors for the hedgehog image in Figure 6.1e are shown in Figures 7.7d and 7.7b. Comparing the ICA bases vectors in Figure 7.7a to the FCA patch bases reveals that the ICA bases contain diagonal elements whereas the FCA bases are more checkerboard like and are even reminiscent of the 2D- DCT matrix. The ICA patch bases seem to depend on the image much more strongly than the FCA patch bases. Both the FCA and ICA patch bases are more structured than we might have expected.

Since FCA worked in unmixing the panda and hedgehog images and since each of these images is composed of a linear combination of FCA extracted free patches, this suggests a way of constructing not-so-random matrix models from random linear combinations of not-so-random (sub) matrices that are asymptotically free. This line of inquiry would complement the recent work [5, 34, 84] in developing not-as-random matrix models that are asymptotically free.

FCA can serve as a valuable computational tool for reasoning and formulating mathematically plausible conjectures about matricial freeness in not-so-random ma-



(a) Patch ICA on panda patches. (b) Patch ICA on hedgehog patches.



(c) Patch FCA on panda patches. (d) Patch FCA on hedgehog patches.

Figure 7.7: Patch bases obtained via an ICA (top row) or FCA (bottom row) factorization of 6×6 patches of the panda and the hedgehog images from Figures 6.1a and 6.1e respectively.

trices.

7.3.3 Open Problem: Improving FCA by "more free" sub-matrix selection

FCA (and ICA) do not always succeed in unmixing images. See, for example Figure 7.8 where applying FCA to the mixed images does not produce a good estimate of the mixing matrix. In Figure 7.9, we show how we can better estimate the mixing matrix by applying FCA to sub-matrices instead. In this example, we can reason

that FCA on the whole matrix fails because the in-alignment faces make the matrices "less free" whereas the sub-matrices are "more free".

We can formalize this idea further by examining how random or not-random the left and right singular vectors of the matrices are. Asymptotically free matrices have left and right singular vectors that are isotropically random relative to each other. Hence, if $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are $N \times N$ right singular vectors of two matrices and if $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are independent and isotropically random, then we expect the entries of $\boldsymbol{V}_1^T * \boldsymbol{V}_2$ to be delocalized and having the values of order $N^{-1/2}$. We can employ a similar argument for the left singular vectors.

We can use this as a heuristic for quantifying how close-to-free two matrices we are trying to unmix are.

For the panda and hedgehog images in Figure 6.1, we can see from Figures 7.10c and 7.10d that the right and left singular vectors respectively are more uniform and so we might FCA to succeed as it indeed did.

In contrast, for the matrices in the Figure 7.8, the right and left singular vectors of the matrices in Figures 7.10a and 7.10b respectively are not that uniform and so we might expect FCA to fail as it did.

The sub-matrices on which we applied FCA in Figure 7.9 are "more free" than the matrices in Figure 7.8 and so FCA worked better in the former case than in the latter. ICA similarly fails as FCA when applied to the whole matrices and similarly succeeds when applied to the sub-matrices.

New algorithmic methods for identifying "more (freely) independent" sub-matrices to improve the unmixing performance of FCA (and ICA) would be invaluable in applications where practitioners have applied FCA (or ICA) and given up because it seemingly did not succeed. Such methods would help make FCA, and ICA, (even) great(er) (again)!

(a) X_1 (b) Z_1 (c) $\widehat{X}_1$ via FCA

(d) X_2 (e) Z_2 (f) $\widehat{X}_2$ via FCA

Figure 7.8: An Application of FCA to images not so free. The mixing matrix is $\boldsymbol{A} = [\sqrt{2}, \sqrt{2}; -\sqrt{2}, \sqrt{2}]/2$

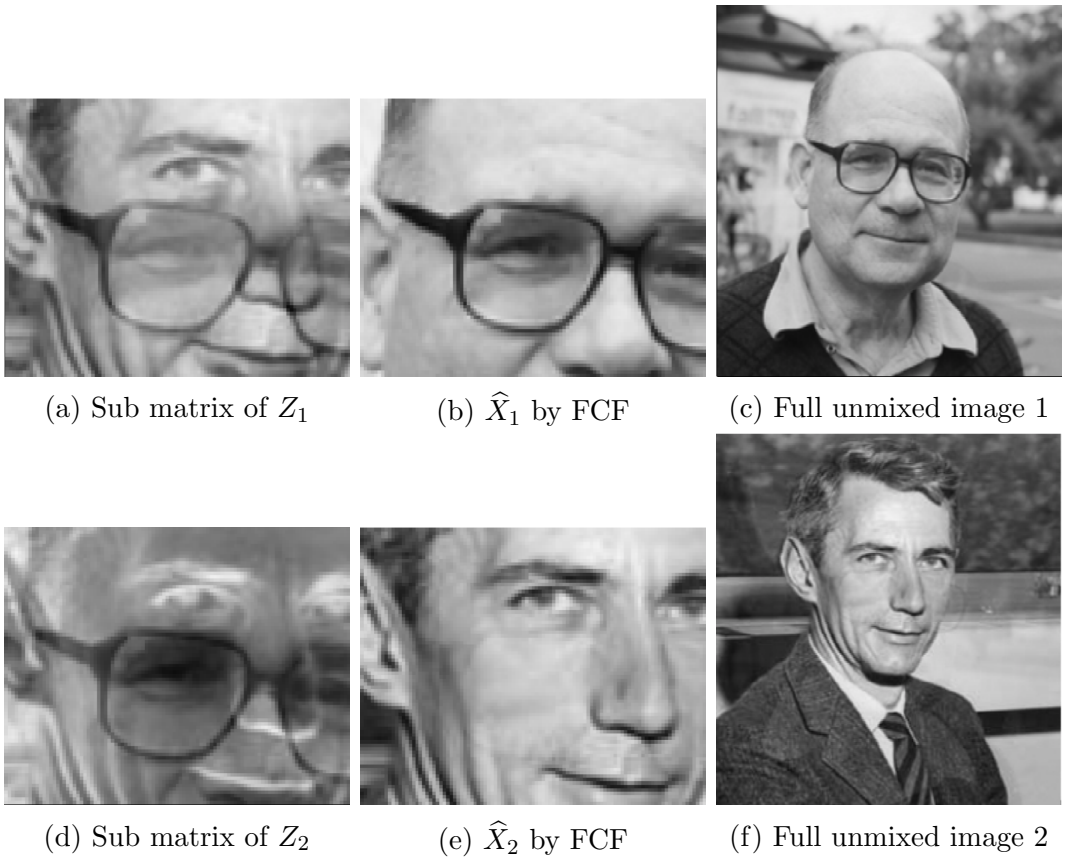
7.4 Proof of Proposition VII.13 and VII.14

We proof Proposition VII.13 and VII.14 for the covariance matrix for rectangular case. The self-adjoint case can be proved with straightforward modification.

7.4.1 Proof of Proposition VII.13

By Remark 1.2 of [95], for any random variable a, $\varphi(a^*) = \overline{\varphi(a)}$. Thus,

$$\begin{aligned} \overline{[\boldsymbol{C}_{\boldsymbol{zz}}]}_{ij} &= \overline{\varphi_1(\tilde{z}_i \tilde{z}_j^*)} \\ &= \varphi_1((\tilde{z}_i \tilde{z}_j^*)^*) \\ &= \varphi_1(\tilde{z}_j \tilde{z}_i^*) = [\boldsymbol{C}_{\boldsymbol{zz}}]_{ji}. \end{aligned} \tag{7.36}$$



(a) Sub matrix of Z_1 (b) $\widehat{X}_1$ by FCF (c) Full unmixed image 1

(d) Sub matrix of Z_2 (e) $\widehat{X}_2$ by FCF (f) Full unmixed image 2

Figure 7.9: Application of FCA to sub images gives better results. The mixing matrix is $\boldsymbol{A} = [\sqrt{2}, \sqrt{2}; -\sqrt{2}, \sqrt{2}]/2$

Therefore, $\boldsymbol{C_{zz}}$ is Hermitian.

We turn to show that $[\boldsymbol{C_{zz}}]$ is positive semi-definite. Actually, as φ is a linear functional, for any column vector $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_s]$,

$$\boldsymbol{\alpha C_{zz} \alpha}^H = \varphi((\sum_{i=1}^{s} \alpha_i \tilde{z}_i)(\sum_{i=1}^{s} \alpha_i \tilde{z}_i)^*) \geq 0 \tag{7.37}$$

where we used that $\varphi(\cdot)$ is positive. This completes the proof.

7.4.2 Proof of Proposition VII.14

Since $\boldsymbol{z} = \boldsymbol{A x}$ and $\boldsymbol{C}_{xx} = \boldsymbol{I}$,

$$\boldsymbol{C_{zz}} = \boldsymbol{A C_{xx} A}^H = \boldsymbol{A A}^H.$$

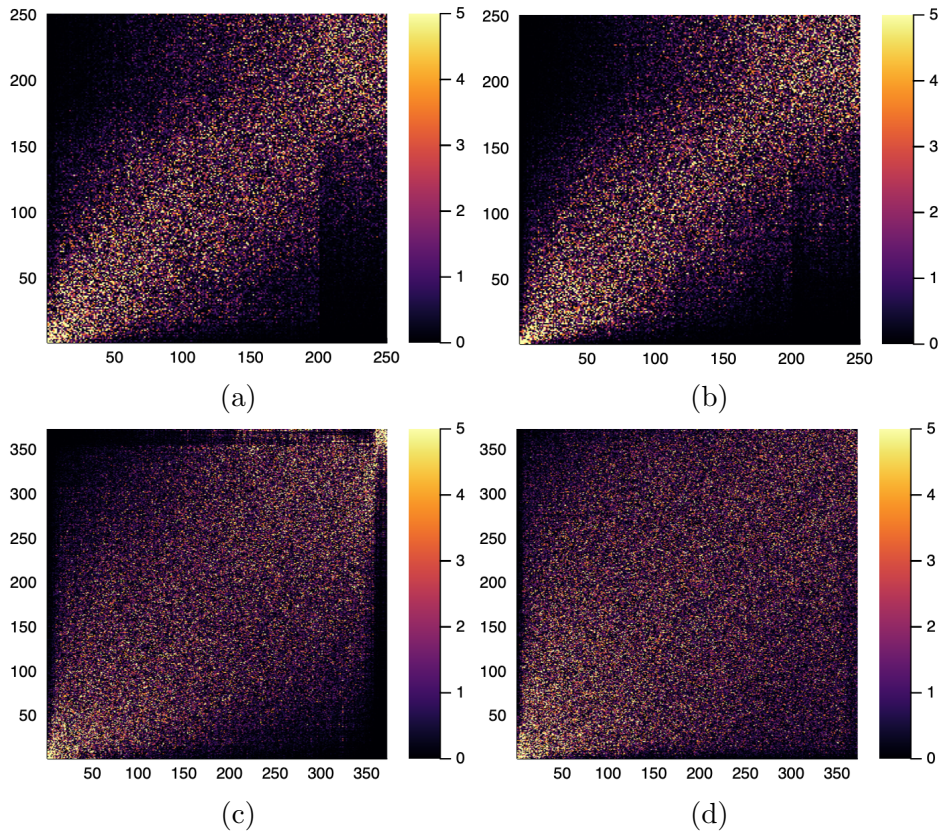


Figure 7.10: Normalized square of inner products between: (a) left singular vectors of Figure 7.8a and 7.8d, (b) right singular vectors of Figure 7.8d and 7.8a, (c) left singular vectors of Figure 6.1a and 6.1e, (d) right singular vectors of Figure 6.1a and 6.1e. We observe that inner product between the right singular vectors of Figure 7.8d and 7.8a (corresponding to (b)) are clearly not uniform.

Note that we assume that $\boldsymbol{A}$ is real and non-singular, $\boldsymbol{C_{zz}}$ is real and positive-definite.

7.5 Proofs of the main results

7.5.1 Proof of Theorem VII.1

In order to prove Theorem VII.1, we first recall the following lemma of free cumulants.

Lemma VII.21. *Given a probability space $(\mathcal{X}, \varphi)$, recall the free cumulants for a single random variable $\kappa_m(a)$ defined in* (B.12)*. We have*

(i) (Free additivity, Proposition 12.3 in [95]) For any $m \geq 1$, if a and b are freely

independent random variables, then

$$\kappa_m(a+b) = \kappa_m(a) + \kappa_m(b). \tag{7.38}$$

(2) For any $m \geq 1$, $\alpha \in \mathbb{C}$ and $a \in \mathcal{X}$, we have that

$$\kappa_m(\alpha a) = \alpha^m \kappa_m(a). \tag{7.39}$$

This immediately follows from the multilinearity of free cumulants (see (B.9)*).*

The above lemma is still valid with $\kappa_m(\cdot)$ denoting rectangular free kurtosis defined in (B.20). The analogy of (7.38) for rectangluar free kurtosis follows from equation (10) in [19]. The analogy of (7.39) is a direct result of (B.19).

We are ready to prove the Theorem VII.1.

7.5.1.1 Proof of Theorem VII.1 (a)

Set $\boldsymbol{g} = \boldsymbol{Q}^T\boldsymbol{w}$, then $\boldsymbol{w} = \boldsymbol{Q}\boldsymbol{g}$. As $\boldsymbol{x}$ and $\boldsymbol{y}$ are related via (7.1), we have that

$$\boldsymbol{w}^T\boldsymbol{y} = \boldsymbol{w}^T\boldsymbol{Q}\boldsymbol{x} = (\boldsymbol{Q}^T\boldsymbol{w})^T\boldsymbol{x} = \boldsymbol{g}^T\boldsymbol{x}. \tag{7.40}$$

Adapt the notation $\boldsymbol{g} = (g_1, \cdots, g_s)^T$. Note that x_i are freely independent, then using (7.38), we have that

$$\kappa_4(\boldsymbol{g}^T\boldsymbol{x}) = \kappa_4\left(\sum_{i=1}^{s} g_i x_i\right) = \sum_{i=1}^{s} \kappa_4(g_i x_i). \tag{7.41}$$

By (7.39), $\kappa_4(g_i x_i) = g_i^4 \kappa_4(x_i)$ for $i = 1, \cdots, s$, thus the above equation becomes

$$\kappa_4(\boldsymbol{g}^T\boldsymbol{x}) = \sum_{i=1}^{s} g_i^4 \kappa_4(x_i). \tag{7.42}$$

Combining (7.40) and (7.42), we get

$$\left|\kappa_4(\boldsymbol{w}^T\boldsymbol{y})\right| = \left|\sum_{i=1}^{s} g_i^4 \kappa_4(x_i)\right|. \tag{7.43}$$

When $\boldsymbol{w}$ runs over all unit vectors, $g = \boldsymbol{Q}^T\boldsymbol{w}$ also runs over all unit vectors. Therefore, if $\boldsymbol{w}^{(1)}$ is a maximizer of (7.6), then $\boldsymbol{w}^{(1)} = \boldsymbol{Q}g^{(1)}$ where $g^{(1)}$ is a maximizer of

$$\max_{\boldsymbol{g}\in\mathbb{R}^s,\ \|u\|=1} \left|\sum_{i=1}^{s} g_i^4 \kappa_4(x_i)\right|. \tag{7.44}$$

Thus in order to prove (a), it is equivalent to show that $\boldsymbol{g}^{(1)}$ is maximizer of (7.44) if and only if $\boldsymbol{g}^{(1)} \in \{(\pm 1, 0, \cdots, 0)^T\}$.

For any unit vector u, since $|g_i| \le 1$, we have that

$$\sum_{i=1}^{s} g_i^4 \le \sum_{i=1}^{s} g_i^2 = 1. \tag{7.45}$$

Note that the equality holds if and only if there is a index i such that $g_i \in \{\pm 1\}$ (thus $g_j = 0$ for all $j \ne i$). Then using (7.5) and (7.45),

$$\begin{aligned} \left|\sum_{i=1}^{s} g_i^4 \kappa_4(x_i)\right| &\le \sum_{i=1}^{s} g_i^4 |\kappa_4(x_i)| \\ &\le \sum_{i=1}^{s} g_i^4 |\kappa_4(x_1)| \\ &\le |\kappa_4(x_1)|. \end{aligned} \tag{7.46}$$

On the other hand, for $\boldsymbol{g} = (\pm 1, 0, \cdots, 0)^T$, it can be checked that all equalities equalities in (7.46) hold. Thus

$$\max_{\boldsymbol{g}\in\mathbb{R}^s,\ \|\boldsymbol{g}\|=1} \left|\sum_{i=1}^{s} g_i^4 \kappa_4(x_i)\right| = |\kappa_4(x_1)| \tag{7.47}$$

and $\boldsymbol{g}^{(1)}$ is a maximizer of (7.44) if $\boldsymbol{g}^{(1)} \in \{(\pm 1, 0, \cdots, 0)^T\}$.

For the other direction, if $\boldsymbol{g}^{(1)}$ is maximizer of (7.44), then the second equality in (7.46) holds for $\boldsymbol{g} = \boldsymbol{g}^{(1)}$. That is,

$$0 = \sum_{i=1}^{s} (g_i^{(1)})^4 \left(|\kappa_4(x_i)| - |\kappa_4(x_1)| \right). \tag{7.48}$$

Due to (7.7), $|\kappa_4(x_i)| - |\kappa_4(x_1)| < 0$ for $i = 2, \cdots, s$. Thus (7.48) implies $g_i^{(1)} = 0$ for $i = 2, \cdots, s$. Since $\boldsymbol{g}^{(1)}$ is a unit vector, $\boldsymbol{g}^{(1)} \in \{(\pm 1, 0, \cdots, 0)^T\}$. This completes the proof.

7.5.1.2 Proof of Theorem VII.1 (b)

In the proof of (a), the arguments upto (7.48) only rely on properties of free kurtosis $\kappa(\cdot)$ and condition (7.5). Thus (7.44), (7.46), (7.47) and (7.48) also apply in the setting of (b). Thus in order to prove (b), it is equivalent to show that $u^{(1)}$ is a maximizer of (7.44) if and only if

(i) $g_i^{(1)} = 0$ for $i = r+1, \cdots, s$,

(ii) there is an index i such that $g_i^{(1)} \in \{\pm 1\}$.

The backward direction can be checking directly using $|\kappa_4(x_1)| = \cdots = |\kappa_4(x_r)|$.

We now prove the forward direction. If $\boldsymbol{g}^{(1)}$ maximizes (7.44), then it satisfies (7.48). By (7.9), $|\kappa_4(x_i)| - |\kappa_4(x_1)| = 0$ for $i = 1, \cdots, r$ and $|\kappa_4(x_i)| - |\kappa_4(x_1)| < 0$ for $i = r+1, \cdots, s$. (i) then follows. On the other hand, as $|\kappa_4(x_1)| = \cdots = |\kappa_4(x_r)|$, enforcing the third equality in (7.46) implies

$$\sum_{i=1}^{r} (g_i^{(1)})^4 = 1. \tag{7.49}$$

By the observation below (7.45), this indicates indicates (ii). This completes the proof.

7.5.2 Proof of Theorem VII.3

Set $\boldsymbol{g} = \boldsymbol{Q}^T\boldsymbol{w}$, we use the notation $\boldsymbol{g} = [g_1, \cdots, g_s]^T$. As $\boldsymbol{w}^{(i)} \in \{\pm\boldsymbol{Q}_i\}$ for $i = 1, \cdots, k-1$,

$$\|\boldsymbol{w}\| = 1, \boldsymbol{w} \perp \boldsymbol{w}^{(1)}, \cdots, \boldsymbol{w}^{(k-1)} \iff \|\boldsymbol{g}\| = 1, g_1 = \cdots = g_{k-1} = 0. \tag{7.50}$$

Using (7.43), if $\boldsymbol{w}^{(k)}$ is a maximizer of (7.11), then $\boldsymbol{w}^{(k)} = \boldsymbol{Q}\boldsymbol{g}^{(k)}$ where $\boldsymbol{g}^{(k)}$ is a maximizer of

$$\max_{\substack{\boldsymbol{g}\in\mathbb{R}^s,\ \|\boldsymbol{g}\|=1 \\ g_1=\cdots=g_{k-1}=0}} \left|\sum_{i=1}^{n} g_i^4 \kappa_4(x_i)\right|. \tag{7.51}$$

Thus in order to prove (a), it is equivalent to show that $\boldsymbol{g}^{(k)} = (g_1^{(k)}, \cdots, g_s^{(k)})^T$ is maximizer of (7.51) if and only if $g_k^{(k)} \in \{\pm 1\}$ (thus $g_j^{(k)} = 0$ for $j \neq k$).

As we are maximizing over unit vector $\boldsymbol{g}$ such that $g_1 = \cdots = g_{k-1} = 0$, again using (7.5) and (7.45)

$$\begin{aligned} \left|\sum_{i=1}^{s} g_i^4 \kappa_4(x_i)\right| &= \left|\sum_{i=k}^{s} g_i^4 \kappa_4(x_i)\right| \\ &\leq \sum_{i=k}^{s} g_i^4 |\kappa_4(x_i)| \\ &\leq \sum_{i=k}^{s} g_i^4 |\kappa_4(x_k)| \\ &\leq |\kappa_4(x_k)|. \end{aligned} \tag{7.52}$$

For $\boldsymbol{g}$ with $g_k \in \{\pm 1\}$, it can be checked that all equalities in (7.52) hold. Thus

$$\max_{\substack{g\in\mathbb{R}^s,\ \|g\|=1 \\ g_1=\cdots=g_{k-1}=0}} \left|\sum_{i=1}^{s} g_i^4 \kappa_4(x_i)\right| = |\kappa_4(x_k)|, \tag{7.53}$$

and $\boldsymbol{g}^{(k)}$ is a maximizer if $g_k^{(k)} \in \{\pm 1\}$.

For the other direction, if $\boldsymbol{g}^{(k)}$ is a maximizer of (7.51), all equalities in (7.52) hold

with $g = g^{(k)}$. In particular, the third equality in (7.52) implies

$$0 = \sum_{i=k}^{s} (g_i^{(k)})^4 \left(|\kappa_4(x_i)| - |\kappa_4(x_k)| \right). \tag{7.54}$$

Due to (7.7), $|\kappa_4(x_i)| - |\kappa_4(x_k)| < 0$ for $i = k+1, \cdots, n$. Thus (7.54) implies that $g_i^{(k)} = 0$ for $i = k+1, \cdots, s$. Since $\boldsymbol{g}^{(k)}$ is a unit vector, $g_k^{(k)} \in \{\pm 1\}$. This completes the proof.

7.5.3 Proof of Theorem VII.4

7.5.3.1 Proof of Theorem VII.4 (a)

We prove (a) by showing that

$$\max_{\boldsymbol{W} \in \boldsymbol{O}(s)} \sum_{i=1}^{s} \left| \kappa_4 \left((\boldsymbol{W}^T y)_i \right) \right| = \sum_{i=1}^{s} |\kappa_4(x_i)| \tag{7.55}$$

and $\boldsymbol{W} = \boldsymbol{Q}$ reaches the maximum. Set $\boldsymbol{G} = \boldsymbol{Q}^T \boldsymbol{W} \in \boldsymbol{O}(s)$. As $\boldsymbol{x}$ and $\boldsymbol{y}$ are related via (7.1),

$$\begin{aligned} \boldsymbol{W}^T \boldsymbol{y} &= \boldsymbol{W}^T \boldsymbol{Q} \boldsymbol{x} \\ &= (\boldsymbol{Q}^T \boldsymbol{W}) \boldsymbol{x} \\ &= \boldsymbol{G}^T \boldsymbol{x}. \end{aligned} \tag{7.56}$$

Adapt the notation $\boldsymbol{G} = (g_{ij})_{i,j=1}^{s}$. Then for all $i = 1, \cdots, n$, $(\boldsymbol{W}^T \boldsymbol{y})_i = (\boldsymbol{G}^T \boldsymbol{x})_i = \sum_{j=1}^{s} g_{ji} x_j$. Together with (7.38) and (7.39), for any $i = 1, \cdots, s$, we have that

$$\begin{aligned} \kappa_4((\boldsymbol{W}^T \boldsymbol{y})_i) =& \kappa_4 \left(\sum_{j=1}^{s} g_{ji} x_j \right) \\ =& \sum_{j=1}^{s} \kappa_4 \left(g_{ji} x_j \right) \\ =& \sum_{j=1}^{s} g_{ji}^4 \kappa_4(x_j). \end{aligned} \tag{7.57}$$

Apply triangular inequality to above equation, we get

$$\left|\kappa_4((\boldsymbol{W}^T\boldsymbol{y})_i)\right| \leq \sum_{j=1}^{s} g_{ji}^4 \left|\kappa_4\left(x_j\right)\right|. \tag{7.58}$$

Note that $(g_{j1}, \cdots, g_{jn})^T$ is a unit vector, by (7.45), $\sum_{j=1}^{s} g_{ji}^4 \leq 1$. Then summing (7.58) over $i = 1, \cdots, n$, we obtain that

$$\begin{aligned}\sum_{i=1}^{s}\left|\kappa_4((\boldsymbol{W}^T y)_i)\right| \leq & \sum_{i=1}^{s}\sum_{j=1}^{s} g_{ij}^4 \left|\kappa_4\left(x_j\right)\right| \\ = & \sum_{j=1}^{s}\left(\sum_{i=1}^{s} g_{ji}^4\right)\left|\kappa_4(x_j)\right| \\ \leq & \sum_{j=1}^{s}\left|\kappa_4(x_j)\right|.\end{aligned} \tag{7.59}$$

Actually, for $\boldsymbol{W} = \boldsymbol{Q}$, $\boldsymbol{Q}^T\boldsymbol{y} = \boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{x}$, thus

$$\sum_{i=1}^{s}\left|\kappa_4((\boldsymbol{Q}^T\boldsymbol{y})_i)\right| = \sum_{i=1}^{s}\left|\kappa_4(x_i)\right|. \tag{7.60}$$

Equations (7.60) and (7.59) together imply (7.55). Then by (7.60), $\boldsymbol{Q}$ is a maximizer of (7.13).

7.5.3.2 Proof of Theorem VII.4 (b)

We first introduce several notations. For a permutation matrix $\boldsymbol{P} = (p_{ji})_{i,j=1}^{s}$, there is a associate permutation σ such that $p_{\sigma(i)i} = 1$ and $p_{ji} = 0$ for all $i = 1, \cdots, s$ and $j \neq \sigma(i)$. For a signature matrix $\boldsymbol{S}$, we denote its i-th diagonal elements by S_i.

Now for any $\boldsymbol{P}$ and $\boldsymbol{S}$, under the light of (7.55), it is desired to show that

$\sum_{i=1}^{s} \left|\kappa_4\left(((\boldsymbol{QPS})^T\boldsymbol{y})_i\right)\right| = \sum_{i=1}^{s} |\kappa_4(x_i)|$. As $\boldsymbol{x}$ and $\boldsymbol{y}$ satisfy (7.1), we have

$$\begin{aligned}(\boldsymbol{QPS})^T\boldsymbol{y} =& \boldsymbol{S}^T\boldsymbol{P}^T\boldsymbol{Q}^T\boldsymbol{y} \\ =& \boldsymbol{S}^T\boldsymbol{P}^T\boldsymbol{x} \\ =& (S_1 x_{\sigma(1)}, \cdots, S_s x_{\sigma(s)})^T.\end{aligned} \tag{7.61}$$

As $S_i \in \{\pm 1\}$, by (7.39)

$$\kappa_4(S_i x_{\sigma(i)}) = S_i^4 \kappa_4(x_{\sigma(i)}) = \kappa_4(x_{\sigma(i)}). \tag{7.62}$$

Combining (7.61) and (7.62) together, we obtain that

$$\begin{aligned}\sum_{i=1}^{s} \left|\kappa_4\left(((\boldsymbol{QPS})^T\boldsymbol{y})_i\right)\right| =& \sum_{i=1}^{s} \left|\kappa_4(S_i x_{\sigma(i)})\right| \\ =& \sum_{i=1}^{s} |\kappa_4(x_{\sigma(i)})| \\ =& \sum_{i=1}^{s} |\kappa_4(x_i)|.\end{aligned} \tag{7.63}$$

This completes the proof of (b).

7.5.3.3 Proof of Theorem VII.4 (c)

By (b), any matrix $\widehat{\boldsymbol{W}}$ of the form $\widehat{\boldsymbol{W}} = \boldsymbol{QPS}$ is a maximizer. For the other direction, we want to show that any maximizer $\widehat{\boldsymbol{W}}$ can be written in the this form.

Actually, if $\widehat{\boldsymbol{W}}$ is a maximizer, we consider $(\widehat{g}_{ij})_{i,j=1}^{s} = \widehat{\boldsymbol{G}} = \boldsymbol{Q}^T\widehat{\boldsymbol{W}}$. The thrid equality of (7.59) holds with $g_{ij} = \widehat{g}_{ij}$. That is,

$$\sum_{j=1}^{s} \left(\sum_{i=1}^{s} \widehat{g}_{ji}^4\right) |\kappa_4(x_j)| = \sum_{j=1}^{s} |\kappa_4(x_j)|. \tag{7.64}$$

Since all the components of x has non-zero free kurtosis and $\sum_{i=1}^{s} \widehat{g}_{ji}^4 \leq 1$ for $j =$

$1, \cdots, s$, (7.64) is equivalent to

$$\sum_{i=1}^{s} \widehat{g}_{ji}^4 = 1, \qquad \text{for } j = 1, \cdots, n. \tag{7.65}$$

By the observation below (7.45), for each j, there is a i such that $\widehat{g}_{ji} \in \{\pm 1\}$ while $\widehat{g}_{jk} = 0$ for $k \neq i$. That is, each column of $\widehat{\boldsymbol{G}}$ has exactly one non-zero entry. By Proposition VII.25, $\widehat{\boldsymbol{G}} \in \boldsymbol{O}_{sp}$ and thus $\widehat{\boldsymbol{G}} = \boldsymbol{PS}$ for some permutation matrix $\boldsymbol{P}$ and signature matrix $\boldsymbol{S}$. Now, recall that $\widehat{\boldsymbol{W}} = \boldsymbol{Q}\widehat{\boldsymbol{G}}$, we have that $\widehat{\boldsymbol{W}} = \boldsymbol{QPS}$. This completes the proof.

7.5.4 Proof of Theorem VII.8

Here, we recall two propositions of free entropy which will be handful in the proof.

Proposition VII.22. *Let $\boldsymbol{x} = (x_1, \cdots, x_s)^T$ where x_i are self-adjoint non-commutative random variables. Then for any $\boldsymbol{Q} = (q_{ij})_{i,j=1}^s \in \boldsymbol{O}(s)$, then*

$$\chi\left((\boldsymbol{Qx})_1, \cdots, (\boldsymbol{Qx})_s\right) = \chi(x_1, \cdots, x_s). \tag{7.66}$$

That is, the free entropy is invariant under the orthogonal transformation.

Proof. This proposition is a special case of a general result. For any matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, we actually have that (see Corollary 6.3.2 in [65]),

$$\chi\left((\boldsymbol{Ax})_1, \cdots, (\boldsymbol{Ax})_s\right) = \chi(x_1, \cdots, x_s) + \log|\det \boldsymbol{A}|. \tag{7.67}$$

Now, for $\boldsymbol{Q} \in \boldsymbol{O}(s)$, $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$, thus

$$(\det \boldsymbol{Q})^2 = \det \boldsymbol{Q}^T \det \boldsymbol{Q} = \det(\boldsymbol{Q}^T\boldsymbol{Q}) = \det \boldsymbol{I} = 1. \tag{7.68}$$

That is, $|\det \boldsymbol{Q}| = 1$ and thus $\log|\det \boldsymbol{Q}| = 0$. Now, set $\boldsymbol{A} = \boldsymbol{Q}$ in (7.67), we obtain

(7.66). □

Proposition VII.23. *Let $x_1, \cdots, x_s$ be self-adjoint non-commutative random variable, then*

$$\chi(x_1, \cdots, x_s) \leq \sum_{i=1}^{s} \chi(x_i). \tag{7.69}$$

Further assume that $\chi(x_i) > -\infty$ for $i = 1, \cdots, n$, then the above equality holds if and only if $x_1, \cdots, x_s$ are freely independent.

Proof. The proof for the inequality can be found in Proposition 6.1.1 in [65]. The equivalence between the equality and freely independence is Theorem 6.4.1 in [65]. □

Proposition VII.22 also holds with $\chi(\cdot)$ denoting the rectangular free entropy. This can be proved using Proposition 5.8 of [18]. On the other hand, Theorem 5.7 and Corollary 5.16 in [18] together prove the analogy Proposition VII.23 for rectangular free entropy $\chi(\cdot)$.

7.5.4.1 Proof of Theorem VII.8 (a)

Set $\boldsymbol{Z} = \boldsymbol{Q}^T\boldsymbol{W}$. As $\boldsymbol{x}$ and $\boldsymbol{y}$ are related via (7.1), $\boldsymbol{W}^T\boldsymbol{y} = (\boldsymbol{Q}\boldsymbol{Z})^T\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{Z}^T\boldsymbol{x}$. Then by (7.69),

$$\sum_{i=1}^{s} \chi\left((\boldsymbol{W}^T\boldsymbol{y})_i\right) = \sum_{i=1}^{s} \chi\left((\boldsymbol{Z}^T\boldsymbol{x})_i\right) \geq \chi\left((\boldsymbol{Z}^T x)_1, \cdots, (\boldsymbol{Z}^T\boldsymbol{x})_s\right). \tag{7.70}$$

On the other hand, note that Z is an orthogonal matrix, then by (7.66),

$$\chi\left((\boldsymbol{Z}^T\boldsymbol{x})_1, \cdots, (\boldsymbol{Z}^T\boldsymbol{x})_s\right) = \chi\left(x_1, \cdots, x_s\right) \tag{7.71}$$

Combining (7.70) and (7.71) together, we obtain that, for any $\boldsymbol{W} \in \boldsymbol{O}(s)$,

$$\sum_{i=1}^{s} \chi\left((\boldsymbol{W}^T\boldsymbol{y})_i\right) \geq \chi\left(x_1, \cdots, x_s\right) \tag{7.72}$$

Now consider $\boldsymbol{W} = \boldsymbol{Q}$. As $\boldsymbol{Q}^T\boldsymbol{y} = \boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{x}$, we have that

$$\sum_{i=1}^{s} \chi\left((\boldsymbol{Q}^T\boldsymbol{y})_i\right) = \sum_{i=1}^{s} \chi\left(x_i\right). \tag{7.73}$$

On the other hand, as x_i are freely independent, then by Proposition VII.23,

$$\sum_{i=1}^{s} \chi(x_i) = \chi(x_1, \cdots, x_s). \tag{7.74}$$

Then (7.73) becomes

$$\sum_{i=1}^{s} \chi\left((\boldsymbol{Q}^T\boldsymbol{y})_i\right) = \chi\left(x_1, \cdots, x_s\right). \tag{7.75}$$

Equations (7.75) and (7.72) together indicate

$$\min_{\boldsymbol{W}\in\boldsymbol{O}(s)} \sum_{i=1}^{s} \chi\left((\boldsymbol{W}^T\boldsymbol{y})_i\right) = \chi\left(x_1, \cdots, x_s\right) \tag{7.76}$$

and $\boldsymbol{Q}$ is a maximizer of (7.15).

7.5.4.2 Proof of Theorem VII.8 (b)

Adapt the notations introduced in the proof of Theorem VII.4 (b). For any permutation matrix $\boldsymbol{P}$ associate with permutation σ and signature matrix $\boldsymbol{S} = \mathrm{diag}(S_1, \cdots, S_s)$, we have that (see (7.61))

$$(\boldsymbol{Q}\boldsymbol{P}\boldsymbol{S})^T y = (S_1 x_{\sigma(1)}, \cdots, S_s x_{\sigma(n)})^T. \tag{7.77}$$

Thus

$$\sum_{i=1}^{s} \chi\left(((\boldsymbol{Q}\boldsymbol{P}\boldsymbol{S})^T y)_i\right) = \sum_{i=1}^{s} \chi(S_i x_{\sigma(i)}). \tag{7.78}$$

As $S_i \in \{\pm 1\}$ can be regard as 1-by-1 orthogonal matrices, then the 1-dimensional verision of (7.67) yields

$$\chi(S_i x_{\sigma(i)}) = \chi(x_{\sigma(i)}), \qquad \text{for } i = 1, \cdots, n. \tag{7.79}$$

Then (7.78) becomes

$$\sum_{i=1}^{s} \chi\left(((\boldsymbol{QPS})^T y)_i\right) = \sum_{i=1}^{s} \chi(x_i). \tag{7.80}$$

Under the light of (7.76), $\boldsymbol{QPS}$ is a maximizer of (7.15).

7.5.4.3 Proof of Theorem VII.8 (c)

By (b), any matrix $\widehat{\boldsymbol{W}}$ of the form $\widehat{\boldsymbol{W}} = \boldsymbol{QPS}$ is a maximizer. For the other direction, it is enough to show that, any maximizer $\widehat{\boldsymbol{W}}$ of (7.15) can be written in the form $\widehat{\boldsymbol{W}} = \boldsymbol{QPS}$ for some permutation matrix $\boldsymbol{P}$ and signature matrix $\boldsymbol{S}$. Actually, if $\widehat{\boldsymbol{W}}$ maximize (7.15), then by (7.76),

$$\sum_{i=1}^{s} \chi\left((\widehat{\boldsymbol{W}}^T y)_i\right) = \chi\left(x_1, \cdots, x_s\right) \tag{7.81}$$

Since $\widehat{\boldsymbol{W}}^T \boldsymbol{Q}$ is a orthogonal matrix, then by (7.66) and (7.1),

$$\begin{aligned} \chi\left(x_1, \cdots, x_s\right) &= \chi\left((\widehat{\boldsymbol{W}}^T \boldsymbol{Q} x)_1, \cdots, (\widehat{\boldsymbol{W}}^T \boldsymbol{Q} x)_s\right) \\ &= \chi\left((\widehat{\boldsymbol{W}}^T y)_1, \cdots, (\widehat{\boldsymbol{W}}^T y)_s\right) \end{aligned} \tag{7.82}$$

Then (7.81) becomes

$$\sum_{i=1}^{s} \chi\left((\widehat{\boldsymbol{W}}^T y)_i\right) = \chi\left((\widehat{\boldsymbol{W}}^T y)_1, \cdots, (\widehat{\boldsymbol{W}}^T y)_s\right) \tag{7.83}$$

By Proposition VII.23, the above equation indicates that $\widehat{\boldsymbol{W}}^T y$ has freely independent components. As we assume that x has at most one semi-circular element, Theorem

VII.9 implies that $\widehat{\boldsymbol{W}} = \boldsymbol{QPS}$ for some permutation matrix $\boldsymbol{P}$ and signature matrix $\boldsymbol{S}$. This completes the proof.

7.5.5 Proof of Theorem VII.9

Definition VII.24. We denote all matrix of size $s \times s$ which are product of a permutation matrix and a signature matrix by

$$\boldsymbol{O}_{ps} = \boldsymbol{O}_{ps}(s) := \{\boldsymbol{PS} \mid \boldsymbol{P} \text{ is a permutation matrix, } \boldsymbol{S} \text{ is a signature matrix}\}. \tag{7.84}$$

Let $\boldsymbol{O} := \boldsymbol{O}(s)$ denotes the sets of orthogonal matrix of size $s \times s$. Note that any permutation matrix $\boldsymbol{P}$ and signature matrix $\boldsymbol{S}$ belong to $\boldsymbol{O}$. Furthermore, it can be checked that $\boldsymbol{O}_{ps}$ is a subgroup of $\boldsymbol{O}$.

We first prove two propositions of $\boldsymbol{O}_{ps}$. An orthogonal matrix must contain at least one nonzero entry in each column (and each row). On the other hand, the matrix belonging to $\boldsymbol{O}_{ps}$ has exactly one nonzero entry in each column (and each row). The following proposition states that this characterizes the matrices contained in $\boldsymbol{O}_{ps}$.

Proposition VII.25. *Fix a positive integer $s \geq 1$, $\boldsymbol{Q} \in \boldsymbol{O}(s)$ has exactly one non-zero entry in each column if and only if $\boldsymbol{Q} \in \boldsymbol{O}_{ps}(s)$.*

Proof. If $\boldsymbol{Q} \in \boldsymbol{O}_{ps}$, then $\boldsymbol{Q} = \boldsymbol{PS}$ for some permutation matrix $\boldsymbol{P}$ and signature matrix $\boldsymbol{S}$. Thus it follows that $\boldsymbol{Q}$ has exactly one non-zero entry in each column.

For the other direction, consider an arbitrary $\boldsymbol{Q} \in \boldsymbol{O}(s)$ with exactly one non-zero entry in each column. Note that $\boldsymbol{Q}$ has totally n non-zero entries. As $\boldsymbol{Q}$ is non-singular, it also has exactly one non-zero entry in each row. As a result, there exists a permutation matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^T\boldsymbol{Q}$ is a diagonal matrix.

On the other hand, note that $(\boldsymbol{P}^T\boldsymbol{Q})^T(\boldsymbol{P}^T\boldsymbol{Q}) = \boldsymbol{Q}^T\boldsymbol{Q} = I$, $\boldsymbol{P}^T\boldsymbol{Q}$ is a diagonal orthogonal matrix. Thus the diagonal entries of $\boldsymbol{P}^T\boldsymbol{Q}$ are either $+1$ or -1. Then

there exists a signature matrix $\boldsymbol{S}$ such that $\boldsymbol{P}^T\boldsymbol{Q} = \boldsymbol{S}$. That is equivalent to $\boldsymbol{Q} = \boldsymbol{P}\boldsymbol{S} \in \boldsymbol{O}_{ps}$. This completes the proof. □

By above proposition, for any $\boldsymbol{Q} \in \boldsymbol{O}\backslash\boldsymbol{O}_{ps}$, there must be a column with more than one non-zero entry. For the later purpose, we prove a stronger result.

Proposition VII.26. *Given any $s \geq 2$, consider matrix $\boldsymbol{Q} = (q_{ij})_{i,j=1}^{s} \in \boldsymbol{O}(s)\backslash\boldsymbol{O}_{ps}(s)$. Then there is a 2×2 submatrix of $\boldsymbol{Q}$ with all 4 entries non-zero. Explicitly, there exist $i, j, k, \ell \in \{1, \cdots, n\}$ $(i \neq j,\ k \neq \ell)$ such that all q_{ik}, $q_{i\ell}$, q_{jk}, and $q_{j\ell}$ are non-zero.*

Proof. We first make the following observation. Two orthogonal vectors either share 0 or more than 2 positions for non-zero entries. Actually, consider any $u = (u_1, \cdots, u_s)^T$ and $v = (v_1, \cdots, v_s)^T$ such that u and v are orthogonal. Assume that there is exactly one index k such that both u_k and v_k are non-zero, then

$$u^T v = \sum_{i=1}^{s} u_i v_i = u_k v_k \neq 0. \tag{7.85}$$

This contradicts the fact that $u^T v = 0$.

Now we are ready to prove the proposition. Denote i-th columns of $\boldsymbol{Q}$ by $\boldsymbol{Q}_i$, for $i = 1, \cdots, s$. Note that the $\{\boldsymbol{Q}_i\}_{i=1}^{s}$ form an orthonormal basis. As $\boldsymbol{Q} \in \boldsymbol{O}(s)\backslash\boldsymbol{O}_{ps}(s)$, there must be a column containing more than two non-zero entries. Without lose of generality, assume it is $\boldsymbol{Q}_1$. If all $\boldsymbol{Q}_2, \cdots, \boldsymbol{Q}_s$ share 0 positions of non-zero entry with $\boldsymbol{Q}_1$, then $\{\boldsymbol{Q}_i\}_{i=2}^{s}$ span a linear space of dimension less than $n-2$. This contradicts with the fact that $\{\boldsymbol{Q}_i\}_{i=2}^{s}$ span a linear space of dimension $s-1$. Thus there must exist a $j \in \{2, \cdots, s\}$ such that $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_j$ share at least one positions for non-zero entry. By the observation we made in the last paragraph, $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_j$ then share at least two positions of non-zero entry. This completes the proof. □

Corollary VII.27. *Fix a positive integer $n \geq 2$ and a $\boldsymbol{Q} \in \boldsymbol{O}(s)\backslash\boldsymbol{O}_{ps}(s)$. There*

exists indexes $i, j, k, \ell \in [1, .., s]$ *(*$i \neq j$ *and* $k \neq \ell$*), such that for any* $m \geq 3$*,*

$$q_{ik}^{m-1} q_{jk} \neq 0, \quad \textit{and} \quad q_{i\ell}^{m-1} q_{j\ell} \neq 0. \tag{7.86}$$

In particular, if $s = 2$*, then for any* $m \geq 3$*,*

$$q_{11}^{m-1} q_{21} \neq 0, \quad \textit{and} \quad q_{12}^{m-1} q_{22} \neq 0. \tag{7.87}$$

Theorem VII.9 can be obtained as a corollary of the following lemma.

Lemma VII.28. *Fix a* $s \geq 2$*, let* $\boldsymbol{x} = (x_1, x_2, \cdots, x_s)^T$ *and* $\boldsymbol{y} = (y_1, y_2, \cdots, y_s)^T$ *be two random vectors such that* $\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{x}$*, where* $\boldsymbol{Q} \in \boldsymbol{O}(s)$*. Assume* $(x_i)_{i=1}^s$ *are freely independent. Now if* $(y_i)_{i=1}^s$ *are freely independent, then at least one of the following happens:*

(i) $\boldsymbol{Q} \in \boldsymbol{O}_{ps}(s)$*.*

(ii) There are at least two components of $\boldsymbol{x}$ *are semicircular (or Poisson in the non self-adjoint setting).*

We first show that Theorem VII.9 follows from Lemma VII.28.

Proof of Theorem VII.9. As $\boldsymbol{x}$ and $\boldsymbol{y}$ satisfy (7.1), $\boldsymbol{x} = (\boldsymbol{Q}^T\boldsymbol{W})\boldsymbol{W}^T\boldsymbol{y}$. Now, by assumption, $\boldsymbol{x}$ and $\boldsymbol{W}^T\boldsymbol{y}$ have free components. Then according to Lemma VII.28, there are two possibilities: (i) $\boldsymbol{Q}^T\boldsymbol{W} \in \boldsymbol{O}_{ps}$ or (ii) $\boldsymbol{x}$ has at least two semicircular components. As (ii) has been excluded, (i) shall happens. That is, there exist a permutation matrix $\boldsymbol{P}$ and a signature matrix $\boldsymbol{S}$ such that $\boldsymbol{Q}^T\boldsymbol{W} = \boldsymbol{P}\boldsymbol{S}$, i.e., $\boldsymbol{W} = \boldsymbol{Q}\boldsymbol{P}\boldsymbol{S}$. □

Proof of Lemma VII.28. If $\boldsymbol{Q} \in \boldsymbol{O}_{ps}(s)$, then the components of $\boldsymbol{y}$ are exactly the components of $\boldsymbol{x}$ with different order and possible sign change. It is not surprising that y_i are freely independent. In the following, we assume that $\boldsymbol{Q} \in \boldsymbol{O}(s) \backslash \boldsymbol{O}_{ps}(s)$,

and $\boldsymbol{x}, \boldsymbol{y}$ has free components, the goal is to show that $\boldsymbol{x}$ has at least two semicircular elements.

We start with the case where $n = 2$. Then it is desired to show x_1 and x_2 are both semicircular elements. Recall the Definition B.8 for the semicircular element, it is enough to show $\kappa_m(x_i) \equiv 0$ for all $m \geq 3$ and $i = 1, 2$.

Fix $m \geq 3$, we consider the mixed cumulants of y_1, y_2 of the specific form $\kappa_m(y_1, \cdots, y_1, y_2, y_p)$ for $p = 1, 2$. As y_1, y_2 are free-independent, these cumulants satisfies the condition of Theorem B.7 by noting that $i(1) = 1 \neq i(m-1) = 2$. Thus these mixed cumulants vanishes, i.e.,

$$\kappa_m(y_1, \cdots, y_1, y_2, y_p) = 0 \quad \text{for } p = 1, 2. \tag{7.88}$$

On the other hand, as $(y_i)_{i=1}^2$ are linear combinations of $(x_i)_{i=1}^2$, using multi-linearity of $\kappa_m(\cdot)$ (see (B.9)), we will express $\kappa_m(y_1, \cdots, y_1, y_2, y_p)$ as linear combinations of $\kappa_m(x_i)$ (recall the notation (B.12)). Adapt the notation $\boldsymbol{Q} = (q_{ij})_{i,j=1}^2$, then $y_i = \sum_{j=1}^2 q_{ij} x_j$. We first derive the expression for $\kappa_m(y_1, \cdots, y_1, y_2, y_1)$ (i.e., $p = 1$),

$$\kappa_m(y_1, \cdots, y_1, y_2, y_1) = \kappa_m\left(\sum_{j=1}^2 q_{1j}x_j, \cdots, \sum_{j=1}^2 q_{1j}x_j, \sum_{j=1}^2 q_{2j}x_j, \sum_{j=1}^2 q_{1j}x_j\right). \tag{7.89}$$

Apply (B.9) to the right hand side of (7.89) to expand the first variable,

$$\kappa_m(y_1, \cdots, y_1, y_2, y_1) = \sum_{j_1=1}^2 q_{1j_1}\kappa_m\left(x_{j_1}, \sum_{j=1}^2 q_{1j}x_j, \cdots, \sum_{j=1}^2 q_{1j}x_j, \sum_{j=1}^2 q_{2j}x_j, \sum_{j=1}^2 q_{1j}x_j\right). \tag{7.90}$$

Again apply (B.9) for the second variable, we obtain that

$$\kappa_m(y_1, \cdots, y_1, y_2, y_1) = \sum_{j_1=1}^2 \sum_{j_2=1}^2 q_{1j_1}q_{1j_2}\kappa_m\left(x_{j_1}, x_{j_2}, \cdots, \sum_{j=1}^2 q_{1j}x_j, \sum_{j=1}^2 q_{2j}x_j, \sum_{j=1}^2 q_{1j}x_j\right). \tag{7.91}$$

Repeat applying (B.9) for the rest $n-2$ variables, we arrive at

$$\kappa_m(y_1,\cdots,y_1,y_2,y_1)=\sum_{j_1=1}^{2}\cdots\sum_{j_n=1}^{2}\left(\prod_{\ell=1}^{s-2}q_{1j_\ell}\right)q_{2j_{s-1}}q_{1j_s}\kappa_m(x_{j_1},\cdots,x_{j_s}). \tag{7.92}$$

There are totally 2^s terms in above summation. Note that x_1 and x_2 are free independent. Then by Theorem B.7, most of these cumulants vanish. For example, $\kappa_s(x_1,x_2,\cdots x_2)=0$ where $j_1=1\neq j_2=2$. Consequently, there are only two terms corresponding to the choices of indexes $j_1=j_2=\cdots=j_s=1$ and $j_1=j_2=\cdots=j_s=2$ survive. Thus using the notation (B.12), (7.92) can be written as

$$\kappa_m(y_1,\cdots,y_1,y_2,y_1)=q_{11}^{m-2}q_{21}q_{11}\kappa_m(x_1)+q_{12}^{m-2}q_{22}q_{12}\kappa_m(x_2). \tag{7.93}$$

Combining (7.93) with (7.88), we obtain that

$$q_{11}^{m-2}q_{21}q_{11}\kappa_m(x_1)+q_{12}^{m-2}q_{22}q_{12}\kappa_m(x_2)=0. \tag{7.94}$$

Repeat (7.88) to (7.94) for $\kappa_m(y_1,\cdots,y_1,y_2,y_2)$ (i.e., $p=2$), we find that

$$q_{11}^{m-2}q_{21}q_{21}\kappa_m(x_1)+q_{12}^{m-2}q_{22}q_{22}\kappa_m(x_2)=0. \tag{7.95}$$

Writing (7.94) and (7.95) in the matrix form, we obtain that

$$\begin{aligned}&\begin{pmatrix}q_{11}^{m-2}q_{21}q_{11} & q_{12}^{m-2}q_{22}q_{12}\\ q_{12}^{m-2}q_{21}q_{21} & q_{12}^{m-2}q_{22}q_{22}\end{pmatrix}\begin{pmatrix}\kappa_m(x_1)\\ \kappa_m(x_2)\end{pmatrix}=\\ &\qquad\begin{pmatrix}q_{11} & q_{12}\\ q_{21} & q_{22}\end{pmatrix}\begin{pmatrix}q_{11}^{m-2}q_{21} & 0\\ 0 & q_{12}^{m-2}q_{22}\end{pmatrix}\begin{pmatrix}\kappa_m(x_1)\\ \kappa_m(x_2)\end{pmatrix}=\vec{0}\end{aligned} \tag{7.96}$$

We actually get a linear equation system for $\kappa_m(x_1)$ and $\kappa_m(x_2)$. Note that $\boldsymbol{Q}=$

$(q_{ij})_{i=1}^2$ is an orthogonal matrix and thus is invertible. Thus (7.96) is equivalent to

$$\begin{pmatrix} q_{11}^{m-2}q_{21} & 0 \\ 0 & q_{12}^{m-2}q_{22} \end{pmatrix} \begin{pmatrix} \kappa_m(x_1) \\ \kappa_m(x_2) \end{pmatrix} = \vec{0}. \tag{7.97}$$

Now, as $\boldsymbol{Q} \in \boldsymbol{O}(2)\backslash\boldsymbol{O}_{ps}(2)$, then by (7.87), above linear equation system has a unique solution, $\kappa_m(x_i) = 0$, $i = 1, 2$. Note that this holds for all $m \geq 3$. Then by Definition B.8, we conclude that x_i for $i = 1, 2$ are semicircular element (or Poisson element in non self-adjoint setting). This conclude the proof for $n = 2$.

For general $n \geq 2$. As $\boldsymbol{Q} \in \boldsymbol{O}\backslash\boldsymbol{O}_{ps}$, by Corollary VII.27, there exist i, j, k, ℓ ($i \neq j$ and $k \neq \ell$) such that (7.86) holds. We will show that x_k, x_ℓ are semicircular elements. For fixed $m \geq 3$, we consider the vanishing mixed cumulants

$$\kappa_m(y_i, \cdots, y_i, y_j, y_p) = 0 \quad \text{for } p = 1, \cdots, s. \tag{7.98}$$

Use relation $y_i = \sum_{j=1}^s q_{ij}x_j$ and multilinearity of κ_m, we can repeat (7.88) to (7.94) for each $\kappa_m(y_i, \cdots, y_i, y_j, y_p)$ and get

$$q_{i1}^{m-1}q_{j1}q_{p1}\kappa_m(x_1) + \cdots + q_{is}^{m-1}q_{js}q_{ps}\kappa_m(x_s) = 0, \quad \text{for } p = 1, \cdots, s. \tag{7.99}$$

Write above equations in the matrix form:

$$\begin{pmatrix} q_{11} & \cdots & q_{1s} \\ \vdots & \ddots & \vdots \\ q_{s1} & \cdots & q_{ss} \end{pmatrix} \begin{pmatrix} q_{i1}^{m-2}q_{j1} & & \\ & \ddots & \\ & & q_{is}^{m-2}q_{js} \end{pmatrix} \begin{pmatrix} \kappa_m(x_1) \\ \vdots \\ \kappa_m(x_s) \end{pmatrix} = \vec{0}. \tag{7.100}$$

Again, $\boldsymbol{Q} = (q_{ij})_{i=1}^s$ is invertible and $q_{ik}^{m-1}q_{jk} \neq 0$ (see (7.86)), thus $\kappa_m(x_k) = 0$. For the same reason, $\kappa_m(x_\ell) = 0$. As these hold for all $m \geq 3$, x_k, x_ℓ are semicircular elements. □

Remark VII.29. We remark that for each $s \geq 2$, there are case there are exactly two semicircular elements. Consider the $\boldsymbol{x}$ with x_1 and x_2 are semicircular elements (or Poisson element in non self-adjoint setting), let the mixing matrix $\boldsymbol{Q}$ be the following:

$$\boldsymbol{Q} = \begin{pmatrix} \cos\theta & -\sin\theta & & & \\ \sin\theta & \cos\theta & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}. \tag{7.101}$$

As mentioned, $(\boldsymbol{Q}\boldsymbol{x})_1$ and $(\boldsymbol{Q}\boldsymbol{x})_2$ are sill free semicircular elements. Also $x_3, \cdots, x_s$ remains unchanged. Thus $\boldsymbol{Q}\boldsymbol{x}$ still has free components.

Not that in the self-adjoint case, identifiability condition (Theorem VII.9) was proved using Corollary VII.27 and Theorem B.7. We can prove Theorem (e) using the same proof with Theorem B.7 replaced by Theorem B.14.

7.6 Proof of Theorem VII.19

Lemma VII.30. *Given $\boldsymbol{Y} = [\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_s]^T \in \mathbb{C}^{Ns \times N}$ with $\boldsymbol{Y}_i \in \mathbb{C}^{N \times N}$ are Hermitian matrices and a vector $\boldsymbol{w} = [w_1, \cdots, w_s] \in \mathbb{R}^s$. For*

$$\boldsymbol{X} = \widetilde{\boldsymbol{w}}^T \boldsymbol{Y}, \quad \text{with } \widetilde{\boldsymbol{w}} = \boldsymbol{w} \otimes \boldsymbol{I}_N,$$

we recall the empirical free kurtosis

$$\widehat{\kappa}_4(\boldsymbol{X}) = \frac{1}{N}\operatorname{Tr}(\boldsymbol{X}^4) - 2\left[\frac{1}{N}\operatorname{Tr}(\boldsymbol{X}^2)\right]^2.$$

Then we have that

$$\frac{\partial \widehat{\kappa}_4(\boldsymbol{X})}{\partial w_k} = \frac{4}{N}\operatorname{Tr}(\boldsymbol{Y}_i\boldsymbol{X}^3) - \frac{8}{N^2}\operatorname{Tr}(\boldsymbol{X}^2)\operatorname{Tr}(\boldsymbol{Y}_i\boldsymbol{X}). \tag{7.102}$$

Proof. As $\operatorname{Tr}(\cdot)$ is a linear function of entries of input matrix,

$$\frac{\partial \widehat{\kappa}_4(\boldsymbol{X})}{\partial w_k} = \frac{1}{N}\operatorname{Tr}\left(\frac{\partial \boldsymbol{X}^4}{w_k}\right) - \frac{4}{N^2}\operatorname{Tr}(\boldsymbol{X}^2)\operatorname{Tr}\left(\frac{\partial \boldsymbol{X}^2}{\partial w_k}\right). \tag{7.103}$$

Note that

$$\boldsymbol{X} = \widetilde{\boldsymbol{w}}^T\boldsymbol{Y} = w_1\boldsymbol{Y}_1 + \cdots + w_s\boldsymbol{Y}_s,$$

thus, for any $k = 1, \cdots, s$,

$$\frac{\partial \boldsymbol{X}}{\partial w_k} = \boldsymbol{Y}_k. \tag{7.104}$$

Therefore,

$$\frac{\partial \boldsymbol{X}^4}{\partial w_k} = \boldsymbol{Y}_k\boldsymbol{X}^3 + \boldsymbol{X}\boldsymbol{Y}_k\boldsymbol{X}^2 + \boldsymbol{X}^2\boldsymbol{Y}_k\boldsymbol{X} + \boldsymbol{X}^3\boldsymbol{Y}_k. \tag{7.105}$$

Using $\operatorname{Tr}(AB) = \operatorname{Tr}(BA)$, we find that

$$\operatorname{Tr}(\boldsymbol{Y}_k\boldsymbol{X}^3) = \operatorname{Tr}(\boldsymbol{X}\boldsymbol{Y}_k\boldsymbol{X}^2) = \operatorname{Tr}(\boldsymbol{X}^2\boldsymbol{Y}_k\boldsymbol{X}) = \operatorname{Tr}(\boldsymbol{X}^3\boldsymbol{Y}_k)$$

and thus

$$\operatorname{Tr}\left(\frac{\partial \boldsymbol{X}^4}{w_k}\right) = 4\operatorname{Tr}(\boldsymbol{Y}_k\boldsymbol{X}^3). \tag{7.106}$$

Repeat (7.105) to (7.106) for $\operatorname{Tr}\left(\frac{\partial \boldsymbol{X}^2}{\partial w_k}\right)$, we get that

$$\operatorname{Tr}\left(\frac{\partial \boldsymbol{X}^2}{\partial w_k}\right) = 2\operatorname{Tr}(\boldsymbol{Y}_k\boldsymbol{X}). \tag{7.107}$$

Plug (7.106) and (7.107) into (7.103), we obtain (7.102).

□

Lemma VII.31. *Given $\boldsymbol{Y} = [\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_s]^T \in \mathbb{C}^{Ns\times N}$ with $\boldsymbol{Y}_i \in \mathbb{C}^{N\times N}$ are Hermitian matrices and a vector $w = [w_1, \cdots, w_s] \in \mathbb{R}^s$. For*

$$\boldsymbol{X} = \widetilde{\boldsymbol{w}}^T \boldsymbol{Y}, \quad \textit{with } \widetilde{\boldsymbol{w}} = w \otimes I_N,$$

with eigenvalues λ_i and corresponding eigenvectors v_i, we recall the empirical free entropy

$$\widehat{\chi}(\boldsymbol{X}) = \frac{1}{N(N-1)} \sum_{i\neq j} \log|\lambda_i - \lambda_j|.$$

Then we have that

$$\frac{\partial \widehat{\chi}(\boldsymbol{X})}{\partial w_k} = \frac{1}{N(N-1)} \sum_{i\neq j} \frac{\partial_{w_k}\lambda_i - \partial_{w_k}\lambda_j}{\lambda_i - \lambda_j} \tag{7.108}$$

with $\partial_{w_k}\lambda_i = v_i^T \boldsymbol{Y}_k v_i$.

Proof. Equation (7.108) is obtained by directly taking derivative. The fact that $\partial_{w_k}\lambda_i = v_i^T \boldsymbol{Y}_k v_i$ follows from (7.104) and perturbation theory of eigenvalues [87]. □

Proof of Theorem VII.19. We first prove part (a). Set $\boldsymbol{X} = [\boldsymbol{X_1}, \cdots, \boldsymbol{X_s}] = \widetilde{\boldsymbol{W}}^T \boldsymbol{Y}$, then

$$\sum_{i=1}^{s} \left| \widehat{\kappa}_4. \left(\widetilde{\boldsymbol{W}}^T \boldsymbol{Y} \right) \right| = \sum_{i=1}^{s} \left| \widehat{\kappa}_4 \left(\boldsymbol{X}_i \right) \right|$$

As only X_ℓ explicitly depends on $\boldsymbol{W}_{k\ell}$,

$$\partial_{\boldsymbol{W}_{k\ell}} \sum_{i=1}^{s} \left| \widehat{\kappa}_4. \left(\widetilde{\boldsymbol{W}}^T \boldsymbol{Y} \right) \right| = \partial_{\boldsymbol{W}_{k\ell}} \left| \widehat{\kappa}_4 \left(\boldsymbol{X}_\ell \right) \right|. \tag{7.109}$$

Further notice that $\boldsymbol{X}_\ell = \widetilde{\boldsymbol{w}}_\ell^T \boldsymbol{Y}$ with $\boldsymbol{w}_\ell = [\boldsymbol{W}_{1\ell}, \cdots, \boldsymbol{W}_{s\ell}]^T$, thus

$$\begin{aligned} \partial_{\boldsymbol{W}_{k\ell}} \left| \widehat{\kappa}_4 \left(\boldsymbol{X}_\ell \right) \right| &= \operatorname{sign}(\widehat{\kappa}_4 \left(\boldsymbol{X}_\ell \right)) \cdot \partial_{\boldsymbol{W}_{k\ell}} \widehat{\kappa}_4 \left(\boldsymbol{X}_\ell \right) \\ &= \operatorname{sign}(\widehat{\kappa}_4 \left(\boldsymbol{X}_\ell \right)) \left(\frac{4}{N} \operatorname{Tr}(\boldsymbol{Y}_k \boldsymbol{X}_\ell^3) - \frac{8}{N^2} \operatorname{Tr}(\boldsymbol{X}_\ell^2) \operatorname{Tr}(\boldsymbol{Y}_k \boldsymbol{X}_\ell) \right), \end{aligned} \tag{7.110}$$

where we used Lemma VII.30 for the last equality. Then (a) is proved by plugging (7.110) into (7.109). The part (b) can be proved in a similar manner by repeating the process from (7.109) to (7.110), where we replace $|\hat{\kappa}_4(\cdot)|$ with $\chi(\cdot)$ and Lemma VII.30 by Lemma VII.31.

We omit the details of the proofs of (c) and (d) since these are straightforward modifications of the proofs of Lemma VII.30, VII.31 and parts (a) and (b). $\square$

APPENDICES

APPENDIX A

Replica method and replica symmetry breaking

The replica method is a popular, however non-rigorous trick in statistical physics. Together with the "replica symmetry breaking", the replica method successfully predicts the correct results in the spin glass theory and motivates the famous Parisi formula. Therefore, we present a brief discussion of the replica method and replica symmetry breaking even though they are not involved in the main content of this thesis.

A.1 Replica method for p-spin SSK model

We mainly follow the calculation in [79]. The applications of replica method to mixed p-spin SK and SSK can be carried out in a similar manner. Recall the Hamiltonian of SSK model $\mathcal{H}(\sigma)$ defined as in (2.10). For simplicity, we assume that the entries of the real symmetric random matrix $J = (J_{ij})_{i,j=1}^N$ are independent Gaussian random variables with variance $(1+\delta_{ij})/N$. Here δ_{ij} denotes the Kronecker delta function. Recall that the free energy and the partition function of SSK model are given by (cf. (2.12))

$$\mathcal{F} := \frac{1}{N\beta}\log\mathcal{Z}, \quad \mathcal{Z} := \int\limits_{\sigma\in S_{N-1}} e^{\beta\mathcal{H}(\sigma)}\mathrm{d}w(\sigma). \tag{A.1}$$

The goal is to calculate the non-random limit $F = \lim_{N\to\infty} \mathcal{F}$. Note that the free energy of SSK is self-averaging thus it is enough to evaluate $\lim_{N\to\infty} \mathbb{E}\,\mathcal{F}$. On the other hand, the formula of $\mathcal{F}$ involves a log function, which is inconvenient for taking the expectation. The idea of replica method is the following simple identity: $\log x = \lim_{n\to 0} \frac{x^n-1}{n}$. That is, the log function can be written as a limit of moments. We write

$$F = \lim_{N\to\infty} \mathbb{E}\,\mathcal{F} = \lim_{N\to\infty} \frac{1}{N\beta} \lim_{n\to 0} \frac{\mathbb{E}\,\mathcal{Z}^n - 1}{n}. \tag{A.2}$$

We now calculate $\mathbb{E}\,\mathcal{Z}^n$ for integer $n \geq 1$. Denoting n replica of σ by $\sigma^{(1)}, \cdots, \sigma^{(n)}$ and using $J_{ij} = J_{ji}$, we find that

$$\begin{aligned} \mathbb{E}\,\mathcal{Z}^n &= \int \prod_{\ell=1}^{n} \mathrm{d}w(\sigma^{(\ell)}) \prod_{i\leq j} \sqrt{\frac{N}{2\pi}} \mathrm{d}J_{ij} \exp\left[-\frac{NJ_{ij}}{2(1+\delta_{ij})} + \beta J_{ij} \sum_{\ell=1}^{n} \sigma_i^{(\ell)} \sigma_j^{(\ell)} \right] \\ &= \int \prod_{\ell=1}^{n} \exp\left[\frac{N\beta^2}{4} \sum_{k,\ell=1}^{n} \left(\frac{1}{N} \sum_{i=1}^{N} \sigma_i^{(k)} \sigma_i^{(\ell)} \right)^2 \right]. \end{aligned} \tag{A.3}$$

The replicas are introduced to eliminate the coupling between $(J_{ij})_{i,j=1}^N$. This works due to the assumption that J_{ij} are Gaussian.

Now we have to deal with the coupling between replicas. Adapt the notation $Q = (Q_{k\ell})_{k,\ell=1}^n$, and write (A.3) as

$$\mathbb{E}\,\mathcal{Z}^n = \int \prod_{\ell=1}^{k} \mathrm{d}w(\sigma^{(\ell)}) \int \prod_{k,\ell}^{n} \mathrm{d}Q_{ij} \delta(Q_{ij} - \frac{1}{N} \sum_{i=1}^{N} \sigma_i^{(k)} \sigma_i^{(\ell)}) \exp\left[\frac{N\beta^2}{4} \sum_{k,\ell}^{n} Q_{k\ell}^2 \right]. \tag{A.4}$$

where $\delta(x)$ denotes the Dirac delta function. That is, the nontrivial contribution comes from

$$Q_{k\ell} = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^{(k)} \sigma_i^{(\ell)}. \tag{A.5}$$

Note that $Q_{k\ell}$ measures the similarity between replicas $\sigma^{(k)}$ and $\sigma^{(\ell)}$. By the spherical constrain, $0 \leq Q_{k\ell} \leq 1$ and $Q_{kk} = 1$ for $k, \ell = 1, \cdots, n$. Changing the order of

integration and noting

$$\int \prod_{i=1}^{k} \mathrm{d}w(\sigma^{(\ell)}) \prod_{k,\ell}^{n} \delta(Q_{k\ell} - \frac{1}{N} \sum_{i=1}^{N} \sigma_i^{(k)} \sigma_i^{(\ell)}) \approx (\det Q)^{-N/2}, \tag{A.6}$$

we find

$$\mathbb{E}\, \mathcal{Z}^n \approx \int \mathrm{d}Q e^{NS(Q)}, \tag{A.7}$$

where

$$S(Q) = \frac{\beta^2}{4} \sum_{k,\ell=1}^{n} Q_{k\ell}^2 + \frac{1}{2} \log \det(Q). \tag{A.8}$$

Here comes the first risky assumption of the replica method. The calculation leading to (A.7) is only valid for integer $n \geq 1$. However, we will assume that (A.7) holds for all $n \to 0$ and thus (A.2) implies

$$F = \lim_{N\to\infty} \lim_{n\to 0} \frac{1}{nN\beta} \left(\int \mathrm{d}Q e^{NS(Q)} - 1 \right). \tag{A.9}$$

It is tempting to apply the steepest descent analysis to the integral. However, the limit $n \to 0$ has to be taken first. Besides, $S(Q)$ depends on n implicitly. In order to avoid these difficulties, we make the second risky assumption that the two limits in (A.9) is exchangeable. Then, an application of steepest descent analysis to (A.9) yields

$$F = \lim_{n\to 0} \frac{1}{n\beta} \max S(Q). \tag{A.10}$$

Here, we note that that the above optimization problem only makes sense for $n \geq 1$ is an integer. As the diagonal elements of Q are fixed to be 1, there are $n(n-1)/2$ variables, we are actually optimizing over a negative number of variables number as $n \to 0$.

Aware of this mathematical weirdness, we optimize $S(Q)$ for $n \geq 1$ and naively

take the $n \to 0$ limit of the maximum. Note that for arbitrary matrix M,

$$\frac{\partial}{\partial M_{k\ell}} \log \det M = (M^{-1})_{k\ell}. \tag{A.11}$$

Therefore, if $\widehat{Q} = \text{argmax}_Q S(Q)$, then $\widehat{Q}_{k\ell}$ satisfies

$$0 = \left.\frac{\partial S}{\partial Q_{k\ell}}\right|_{Q=\widehat{Q}} = \frac{\beta^2 p}{4}\widehat{Q}_{k\ell}^{p-1} + \frac{1}{2}(\widehat{Q}^{-1})_{k\ell} \quad \text{for all } k \neq \ell. \tag{A.12}$$

The above equation involves the inverse of $\widehat{Q}$ and is not easy to solve. Mimicking what Sherrington and Kirkpatrick did in [107], we assume a replica symmetry (RS) form for $\widehat{Q}$. That is, all replicas are assumed to be equivalent, which implies that the overlaps $\widehat{Q}_{k\ell} \equiv q_0$ for $k \neq \ell$, i.e.,

$$\widehat{Q}_{k\ell} = q_0 I + (1 - q_0)\delta_{k\ell}. \tag{A.13}$$

For $\widehat{Q}$ with above structure, one can check that

$$(\widehat{Q}^{-1})_{k\ell} = \frac{1}{1 - q_0}\delta_{k\ell} - \frac{q_0}{(1 - q_0)[1 + (n-1)q_0]}. \tag{A.14}$$

Then the equation (A.12), in the limit of $n \to 0$, becomes

$$\frac{\beta^2 p}{4} q_0^{p-1} - \frac{q_0}{2(1 - q_0)^2} = 0. \tag{A.15}$$

We first discuss the 2-spin case. When $T > 1$ ($\beta < 1$), equation (A.15) has a unique solution $q_0 = 0$. On the other hand, when $T < 1$ ($\beta > 1$), the equation (A.15) possesses two solutions 0 and $1 - \frac{1}{\beta}$. Since, the replicas will be frozen at direction maximizing the Hamiltonian when $T = 0$, we expect that $q_0 \to 1$ as $T \to 0$. Therefore, we take $q_0 = 1 - \frac{1}{\beta}$ for $T < 1$. Observe that there a critical temperature $T_c = 1$.

Plugging the above q_0 back to (A.10), we obtain that

$$F = \begin{cases} \frac{1}{4T} & \text{if } T > 1, \\ 1 - \frac{3T}{4} + \frac{T \log T}{2} & \text{if } T < 1. \end{cases} \tag{A.16}$$

This formula of limiting free energy agrees with the one obtained from the Parisi formula, which indicates that the RS assumption is valid for 2-spin SSK model.

We now consider the cases where $p \geq 3$. There exists a critical temperature T_c such that the equation (A.15) behaves differently between $T > T_c$ and $T < T_c$. In the high temperature ($T > T_c$), the equation (A.15) possesses a unique solution $q_0 = 0$. And plugging $q_0 = 0$ into (A.10) yields the correct limiting free energy. On the other hand, there is a pair of nonzero solutions appear when $T < T_c$. We select the solution increases with decreasing T. However, the limiting free energies associated with this nonzero solution does not agree with the result from the Parisi formula. Furthermore, it exhibits an unphysical negative entropy [107, 89] when $T < T_0$ for some temperature $T_0 < T_c$. From the point of view of steepest descent, "replica symmetric" solutions are unstable when $T < T_c$ [47]; the Hessian of $S(Q)$ at $\widehat{Q}$ has a negative eigenvalue. All these indicate our RS scheme is incorrect in the low temperature regime. As we will see in the next section, for $p \geq 3$, $\widehat{Q}$ in the low temperature should be of a "replica symmetry breaking" (RSB) form.

There are efforts to validate the replica method for SK and SSK models [89, 114]. However a fully rigorous justification is still open. At the moment, we can regard it as a powerful predictive tool.

A.2 Replica symmetry breaking and Parisi formula

In the last section, we found that the RS scheme is insufficient and a RSB solution was necessary when we consider p-spin SSK models for $p \geq 3$ in the low temperature

regime. The simplest possible structure of RSB is the following. The n replicas are divided into groups of size m_1 (thus there are around n/m_1 groups in total). Then it is assumed that the overlaps between replicas from the same group are uniform, whose value is denoted by q_1 satisfying $0 \leq q_1 \leq 1$. On the other hand, the overlaps between replicas from the distinct groups are also uniform and the value is assumed to be q_0 satisfying $0 \leq q_0 \leq q_1$. This ansatz is called 1RSB scheme (see Figure A.1 for corresponding $\widehat{Q}$).

Plugging $\widehat{Q}$ with the 1RSB structure into (A.10) and taking the limit $n \to 0$ carefully (again we are optimizing over a negative number of variables), we find that the corresponding limiting free energy F_{1RSB} satisfies

$$\begin{aligned} F_{\mathsf{1RSB}} =& \frac{\beta[1+(m_1-1)q_1^p - m_1 q_0^p]}{4} + \frac{\log[m_1(q_1-q_0)+(1-q_1)]}{2m_1\beta} \\ &+ \frac{(m_1-1)\log(1-q_1)}{2m_1\beta} + \frac{q_0}{m_1\beta(q_1-q_0)+(1-q_1)}. \end{aligned} \tag{A.17}$$

The explicit values of m_1, q_1 and q_0 is defined by the equation systems $\partial_{m_1} F_{\mathsf{1RSB}} = 0$, $\partial_{q_1} F_{\mathsf{1RSB}} = 0$ and $\partial_{q_0} F_{\mathsf{1RSB}} = 0$ and the constraints $0 \leq m_1, q_1 \leq 1$ and $0 \leq q_0 \leq q_1$. Here, m_1 stands for the probability that a overlap is equal to q_1 when it is uniformly sampled from all pairs of the replicas. It turned out, for pure p-spin SSK models, the F_{1RSB} is stable (the eigenvalues of Hessian are all positive) and agrees with the limiting free energy given by the Parisi formula [47].

However, for p-spin SK and mixed p-spin SSK, the above 1RSB scheme does not lead to the correct limiting free energy; the $S(Q)$ is unstable at the $\widehat{Q}$ with 1RSB structure and F_{1RSB} is inconsistent with the Parisi formula. Nevertheless, 1RSB scheme is better in terms of the entropy. It only exhibits negative entropy when $T < T_1$ for some $T_1 < T_0$. This suggests that further symmetry breaking need to be introduced based on 1RSB scheme, which is exactly what Parisi did in his series of work towards the Parisi formula [101, 102, 103, 89]. Parisi divided each group (of

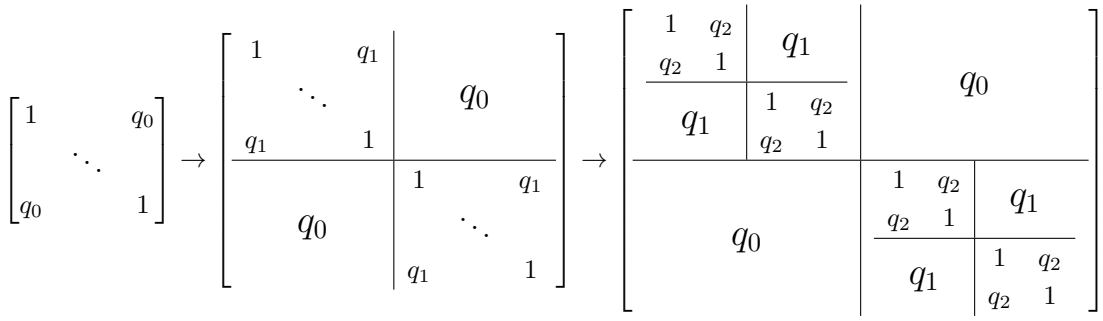


Figure A.1: Matrices $\widehat{Q}$ corresponding to the "replica symmetry" ansatz and the first two RSB ansatzs. The first matrix represents $\widehat{Q}$ for the "replica symmetry" ansatz. The second and third matrices are $\widehat{Q}$ for the 1RSB and 2RSB scheme respectively. In this diagram, we divide every groups into two subgroups ($m_1 = n/2$ and $m_2 = m_1/2$). In each step, we further breaking matrices on the diagonal into smaller sub matrices while off-diagonal matrices remain unchanged.

size m_1) in 1RSB schme into subgroups of size m_2. Similar to assumptions made in symmetry braeking from RS to 1RSB, it is assumed that the overlap between replicas within the same subgroups are some constant $q_2 \geq q_1$, while other overlaps remained unchanged. The resulting ansatz is called 2RSB scheme. The $\widehat{Q}$ of 2RSB scheme is plotted in Figure A.1 and we denote the associated limiting free energy by F_{2RSB}. While F_{2RSB} does not agree with the Parisi formula, it is associated with a better entropy, which becomes negative at even lower temperatures than 1RSB and replica symmetry schemes. This motivated Parisi to continue perform the symmetry breaking based on 2RSB scheme. Actually, Parisi proposed that the process could be iterated infinite times . The limiting ansatz was called full RSB scheme or Parisi ansatz, which is the structure of replicas behind the Parisi formula.

In the full RSB scheme, the replicas are organized in a hierarchical way (see Figure A.2), which is called ultrametric [88, 99]. For mixed p-spin SK and SSK models, full RSB scheme is necessary: any finite step RSB fails to describe the system. Therefore, complicated energy landscapes are expected for these models.

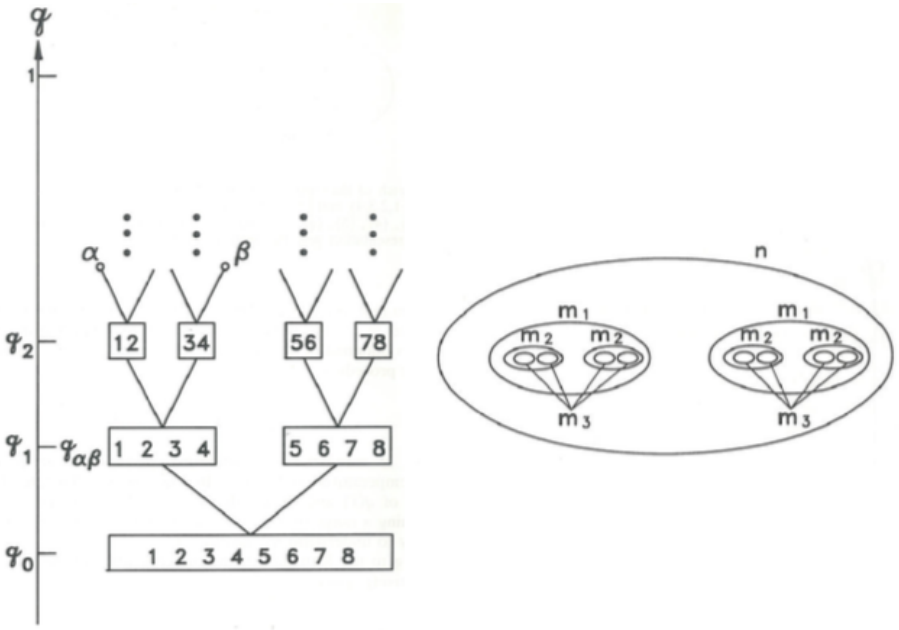


Figure A.2: Full RSB scheme. Left panel: the overlap between replicas α and β is q_1, which corresponds to the first common level containing both α and β. Right panel: replicas are divided into subsets iteratively. Figure from [93].

APPENDIX B

Free probability, matrix embeddings and independent component analysis

B.1 What is freeness of random variables?

The goal of this section is to introduce the freeness of non-commutative random variable. We first discuss the independence (freeness) in the context of the scalar probability, free probability for self-adjoint (non-commutative) random variables and free probability for rectangular (non-commutative) random variables respectively. We focus on the the behavior of (free) cumulants and (free) entropy of independent (free) random variables, which are the basis ICA (FCA). The connection between in independent random matrices and free random variables is given at the end.

For a detailed introduction of free probability, readers are referred to [95, 65, 90]

B.1.1 Prologue: What is independence of commuting random variables?

Here, we briefly review the statistic independence in scalar probability. We state the behavior of cumulants and entropy of independent random variables, which are the basis of ICA. In the end, we discuss the unique role the Gaussian random variables play in ICA.

B.1.1.1 Mixed moments point of view

Let I denotes an index set, and $(x_i)_{i\in I}$ denote random variables. They are independent if for any $n \in \mathbb{N}$ and $m_1, \cdots, m_n \geq 0$,

$$\mathbb{E}[x_{i(1)}^{m_1} \cdots x_{i(n)}^{m_n}] = \mathbb{E}[x_{i(1)}^{m_1}] \cdots \mathbb{E}[x_{i(n)}^{m_n}].$$

if $i(j) \in I$, $j = 1, \cdots n$ are all distinct. An alternative definition is that for any polynomials $P_1, \cdots P_n$ of one variables,

$$\mathbb{E}[P_1(x_{i(1)}) \cdots P_n(x_{i(n)})] = 0 \tag{B.1}$$

if $\mathbb{E}[P_j(x_{i(j)})] = 0$ for all $j = 1, \cdots, n$ and $i(j) \in I$, $j = 1, \cdots n$ are all distinct.

B.1.1.2 Cumulants – kurtosis and higher order – independent additivity

The (joint) cumulants of n random variables $a_1, \cdots, a_k$ is defined by

$$c_n(a_1, \cdots, a_n) = \sum_{\pi} (|\pi| - 1)!(-1)^{|\pi|-1} \prod_{B \in \pi} \mathbb{E}\left[\prod_{i \in B} a_i\right], \tag{B.2}$$

where π runs through all partitions of $\{1, \cdots, m\}$, B runs through all blocks of partition π. Equivalently, $\{c_m\}_{m \geq 1}$ is defined through

$$E(x_1 \cdots x_n) = \sum_{\pi} \prod_{B \in \pi} c_{|B|}(a_i : i \in B) \tag{B.3}$$

The reason that ICA adapts an optimization problem involving cumulants is the following property: if $(x_i)_{i \in I}$ are independent, then for any $n \in \mathbb{N}$

$$c_n(x_{i(1)}, \cdots, x_{i(n)}) = 0 \tag{B.4}$$

whenever there exists $1 \leq \ell, k \leq n$ with $i(\ell) \neq i(k)$. That is, any cumulants involving two (or more) independent random variables is zero. Adapt the notation

$$c_n(x) := c_n(x, \cdots, x).$$

A quick consequence of (B.4) is that for independent x_1 and x_2.

$$c_n(x_1 + x_2) = c_n(x_1) + c_n(x_2). \tag{B.5}$$

B.1.1.3 Entropy – independent additivity

For random variables $a_1, \cdots, a_n$ with joint distribution $f(a_1, \cdots, a_n)$, the (joint) entropy is defined by [46]

$$h(a_1, \cdots, a_n) = -\int f(\alpha_1, \ldots, \alpha_n) \log f(\alpha_1, \cdots, \alpha_n) \mathrm{d}\alpha_1 \cdots \mathrm{d}\alpha_n. \tag{B.6}$$

The joint entropy of a set of variables is less than or equal to the sum of the individual entropies of the variables in the set,

$$h(a_1, \cdots, a_n) \leq h(a_1) + \cdots + h(a_n). \tag{B.7}$$

In particular, the equality in (B.7) holds if and only if $a_1, \cdots, a_n$ are independent. Therefore, the entropy is regard a measure of indepenedence and thus can be used in ICA.

B.1.1.4 Why Gaussians cannot be unmixed? Gaussians have zero higher order cumulants

In ICA, the optimization problem people used finds the independent direction by maximizing the kurtosis (fourth cumulant). However, all cumulants of order larger than 2 for Gaussian random variables vanish. Thus ICA fails to unmix Gaussian random variables. ICA based on the entropy also fails to unmix Gaussian random variables, as nontrivial mixtures of independent Gaussian random variables can still be independent Gaussian. On the other hand, it was shown that this is the only case where ICA does not work [43]. The result of this kind is called identifiability condition.

B.1.2 Freeness of self-adjoint random variables

We first introduce the definition of probability space for non-commutative random variables. The start point is the an unital algebra of non-commutative variables.

Definition B.1. Let $\mathcal{X}$ be a vector space over $\mathbb{C}$ equipped with product $\cdot : \mathcal{X} \times \mathcal{X} \mapsto \mathcal{X}$. Denote the vector space addition by $+$, we call $\mathcal{X}$ an algebra if for all $a, b, c \in \mathcal{X}$ and $\alpha \in \mathbb{C}$,

1. $a(bc) = (ab)c$,
2. $a(b + c) = ab + ac$,
3. $\alpha(ab) = (\alpha a)b = a(\alpha b)$.

We call $\mathcal{X}$ a unital algebra if there is a unital element $1_{\mathcal{X}}$ such that, for all $a \in \mathcal{X}$

$$a = a1_{\mathcal{X}} = 1_{\mathcal{X}}a. \tag{B.8}$$

An algebra $\mathcal{X}$ is called a $*$-algebra if it is also endowed with an antilinear $*$-operation $\mathcal{X} \ni a \mapsto a^* \in \mathcal{X}$, such that $(\alpha a)^* = \bar{\alpha}a^*$, $(a^*)^* = a$ and $(ab)^* = b^*a^*$ for all $\alpha \in \mathbb{C}$, $a, b \in \mathcal{X}$.

Note that $ab = ba$ does not necessarily hold for general $a, b \in \mathcal{X}$, i.e., they are non-commutative.

Definition B.2. A (non-commutative) $*$-probability space $(\mathcal{X}, \varphi)$ consists of a unital $*$-algebra and a linear functional $\varphi : \mathcal{X} \to \mathbb{C}$, which serves as the 'expectation'. with We also require that φ satisfies

(i) (positive). $\varphi(aa^*) \geq 0$ for all $a \in \mathcal{X}$.

(ii) (tracial). $\varphi(ab) = \varphi(ba)$ for all $a, b \in \mathcal{X}$.

(iii) $\varphi(1_{\mathcal{X}}) = 1$.

The elements $a \in \mathcal{X}$ are called non-commutative random variables. (We may omit the word non-commutative if there is no ambiguity.) Given a series of random variables $x_1, \cdots, x_k \in \mathcal{X}$, for any choice of $n \in \mathbb{N}$, $i(1), \cdots, i(n) \in [1..k]$ and $\epsilon_1, \cdots, \epsilon_n \in \{1, *\}$, $\varphi(x_{i(1)}^{\epsilon_1} \cdots x_{i(n)}^{\epsilon_n})$ is a mixed moment of $\{x_i\}_{i=1}^k$. The collection of all moments is called the joint distribution of $x_1, \cdots, x_k$.

The moments of general random variables can be complex-valued; self-adjoint random variables, which are defined in a minute, necessarily have real-valued moments and will be the object of our study.

Definition B.3. Let $(\mathcal{X}, \varphi)$ be a non-commutative probability space, a element $a \in \mathcal{X}$ is self-adjoint if $a = a^*$. In particular, the moments of self-adjoint elements are real (see Remark 1.2 in [95]).

The counterpart of independence in free probability is freely independence or simply free. We now consider the freeness of self-adjoint random variables from various perspectives as in Section B.1.1.

B.1.2.1 Mixed moments point of view

The following official definition of freeness should be compared with (B.1).

Definition B.4. Let $(\mathcal{X}, \varphi)$ be a non-commutative probability space and fix a positive integer $n \geq 1$.

For each $i \in I$, let $\mathcal{X}_i \subset \mathcal{X}$ be a unital subalgebra. The subalgebras $(\mathcal{X}_i)_{i \in I}$ are called freely independent (or simply free), if for all $k \geq 1$

$$\varphi(x_1 \cdots x_k) = 0$$

whenever $\varphi(x_j) = 0$ for all $j = 1, \cdots, k$, and neighboring elements are from diffierent subalgebras, i.e. $x_j \in \mathcal{X}_{i(j)}$, $i(1) \neq i(2), i(2) \neq i(3), \cdots, i(k-1) \neq i(k)$.

In particular, a series of elements $(x_i)_{i\in I}$ are called free if the subalgebras generated by x_i and x_i^* are free.

B.1.2.2 Free cumulants – free additivity

The analog of cumulants for non-commutative random variables is called free cumulants, which was proposed by Roland Speicher [110, 95].

Definition B.5. Given a $*$-probability space $(\mathcal{X}, \varphi)$, the free cumulants refer to a family of multilinear functionals $\{\kappa_m : \mathcal{X}^m \mapsto \mathbb{C}\}_{m\geq 1}$. Here, the multilinearity means that κ_m is linear in one variable when others hold constant, i.e., for any $\alpha, \beta \in \mathbb{C}$ and $a, b \in \mathcal{X}$,

$$\kappa_m(\cdots, \alpha a + \beta b, \cdots) = \alpha\kappa_m(\cdots, a, \cdots) + \beta\kappa_m(\cdots, b, \cdots). \tag{B.9}$$

Explicitly, for $a_1, \cdots, a_n \in \mathcal{X}$, their mixed free cumulant is defined through (cf. (B.3))

$$\varphi(a_1 \cdots a_n) = \sum_{\pi\in NC(n)} \prod_{B\in\pi} \kappa_{|B|}(a_i : i \in B) \tag{B.10}$$

where $NC(n)$ denotes the non-crossing partition of $[1..n]$. Equivalently (cf. (B.2)),

$$\kappa_n(a_1, \cdots, a_n) = \sum_{\pi\in NC(n)} \mu(\pi, \mathbf{1}_n) \sum_{B\in\pi} \varphi\left[\prod_{i\in B} a_i\right], \tag{B.11}$$

where μ is the Möbius function on $NC(n)$.

Example B.6. We have that

$$\kappa_1(a_1) = \varphi(a_1),$$

$$\kappa_2(a_1, a_2) = \varphi(a_1a_2) - \varphi(a_1)\varphi(a_2),$$

$$\begin{aligned}\kappa_3(a_1, a_2, a_3) = {} & \varphi(a_1a_2a_3) - \varphi(a_1)\varphi(a_2a_3) \\ & - \varphi(a_2)\varphi(a_1a_3) - \varphi(a_3)\varphi(a_1a_2) + 2\varphi(a_1)\varphi(a_2)\varphi(a_3).\end{aligned}$$

Recall that in the scalar probability, mixed cumulants of independent random variables vanish (see (B.4)). The same holds for the free cumulants in the free probability.

Theorem B.7 (Theorem 11.16 of [95]). *Let $(\mathcal{X}, \varphi)$ be a non-commutative probability space with associate free cumulants $(\kappa_\ell)_{\ell\in\mathbb{N}}$. Consider random variables $(x_i)_{i\in I}$, assume that they are freely independent. Then for all $n \geq 2$, and $i(1), \cdots, i(n) \in I$, we have $\kappa_n(a_{i(1)}, \cdots, a_{i(n)}) = 0$ whenever there exist $1 \leq l, k \leq n$ with $i(l) \neq i(k)$.*

Consequently, set

$$\kappa_m(a) := \kappa_m(a, a, \cdots, a), \tag{B.12}$$

we have that for $a, b \in \mathcal{X}$ that are free independent,

$$\kappa_m(a + b) = \kappa_m(a) + \kappa_m(b) \quad \forall m \in \mathbb{N}. \tag{B.13}$$

The above equation should be compared with (B.5).

B.1.2.3 Free entropy – free additivity

For non-commutative random variables, the free entropy χ was introduced by Voiculescu [118, 119, 121]. The formulation of free entropy is rather complicated and out of main naretive, readers are referred to Section 6 of [65] for a detailed introduction.

The free entropy shares the similar property with the scalar entropy. We have that (cf. (B.7))

$$\chi(a_1, \cdots, a_n) \leq \chi(a_1) + \cdots + \chi(a_n). \tag{B.14}$$

Again the equality in (B.14) holds if and only if $a_1, \cdots, a_n$ are freely independent (see Proposition VII.23).

B.1.2.4 Analogy of Gaussian random variables in free probability: the free semi-circular element

The analogy of Gaussian random variable in a $*$-probability space is semicircular element. Recall that the Gaussian random variable is characterized by vanishing cumulants of order higher than 2, the semicircular elements can be defined in a similar manner.

Definition B.8. Given a $*$-probability space $(\mathcal{X}, \varphi)$, we call a random variable $a \in \mathcal{X}$ a semicircular element if

$$\kappa_m(a) \equiv 0, \qquad \text{for } m \geq 3, \tag{B.15}$$

and $\kappa_2(a) > 0$ (such that a is not constant).

B.1.3 Freeness of non self-adjoint random variables

We brief introduce the mathematical preliminaries for rectangular probability space. For a detailed introduction, the readers are refered to [19, 18]

Consider a $*$-probablity space $(\mathcal{X}, \varphi)$ with p_1, p_2 of non-zero self-adjoint projectors (i.e. $\forall i, p_i^2 = p_i$) which are pairwise orthogonal (i.e. $\forall i \neq j, p_i p_j = 0$), and such that $p_1 + p_2 = 1_{\mathcal{X}}$. Then any element $a \in \mathcal{X}$ can be represented in the following block form

$$a = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \tag{B.16}$$

where $\forall i, j = 1, 2, a_{ij} = p_i a p_j$ and we define $\mathcal{X}_{ij} := p_i \mathcal{X} p_j$. Note that $\mathcal{X}_{ii}$ is a subalgebra, and we equip it with the functional $\varphi_k = \frac{1}{\rho_k}\varphi|_{\mathcal{X}_{kk}}$, where $\rho_k := \varphi(p_k)$. The functionals φ_i, $i = 1, 2$ are tracial in the sense that $\varphi_k(p_k) = 1$ and for all i, j, $x \in \mathcal{X}_{ij}$, $y \in \mathcal{X}_{ji}$,

$$\rho_i \varphi_i(xy) = \rho_j \varphi_j(yx). \tag{B.17}$$

Definition B.9. Such a family $(\mathcal{X}, p_1, p_2, \varphi_1, \varphi_2)$ is called a (ρ_1, ρ_2)-rectangular prob-

ability space. We call $a \in \mathcal{X}_{12}$ rectangular random variable.

Remark B.10. If a is a rectangular element, then in the matrix decomposition (B.16), only a_{12} is non-zero. Later in Section B.1.4.2, we will model rectangular matrices by embedding them into a_{12} of rectangular random variables.

For such a rectangular probability space, the linear span of p_1, p_2 is denoted by $\mathcal{D}$. Then $\mathcal{D}$ is subalgebra of finite dimension. Define the $\mathbb{E}_{\mathcal{D}}(a) = \sum_{i=1}^{2} \varphi(a_{ii})p_i$. It can be checked that $\mathbb{E}_{\mathcal{D}}(1_{\mathcal{X}}) = 1_{\mathcal{X}}$ and $\forall(d, a, d') \in \mathcal{D} \times \mathcal{X} \times \mathcal{D}$, $\mathbb{E}_{\mathcal{D}}(dad') = d\,\mathbb{E}_{\mathcal{D}}(a)d'$. The map $\mathbb{E}_{\mathcal{D}}(\cdot)$ is regarded as the conditional expectation from $\mathcal{X}$ to $\mathcal{D}$.

We now consider the freeness in rectangular probability space.

B.1.3.1 Mixed moments point of view

The following definition of freeness should be compared with (B.1) and Definition B.4.

Definition B.11. Given a rectangular probability space and subalgebra $\mathcal{D}$ with the corresponding conditional expectation $\mathbb{E}_{\mathcal{D}}$. A family $(\mathcal{X}_i)_{i\in I}$ of subalgebras containing $\mathcal{D}$ is said to be free with amalgamation over $\mathcal{D}$ (we simply use the word free when there is no ambiguity) if for all $k \geq 1$

$$\mathbb{E}_{\mathcal{D}}(x_1 \cdots x_k) = 0 \tag{B.18}$$

whenever $\mathbb{E}_{\mathcal{D}}(x_j) = 0$ for all $j = 1, \cdots, k$, and neighboring elements are from different subalgebras, i.e., $x_j \in \mathcal{X}_{i(j)}$, $i(1) \neq i(2), i(2) \neq i(3), \cdots, i(k-1) \neq i(k)$. In particular, a family of rectangular random variables $\{x_i\}_{i\in I}$ $(i \in I)$ are called free if the subalgebras generated by $\mathcal{D}$, a_i, and a_i^* are free.

B.1.3.2 Rectangular free cumulants – free additivity

The free cumulants are also defined for rectangular probability space [19, 18].

Definition B.12 (Analogy of cumulant in rectangular probability space)**.** Given a (ρ_1, ρ_2)-probability space $(\mathcal{X}, p_1, p_2, \varphi_1, \varphi_2)$, for any $n \geq 1$, we denote n-th tensor product over $\mathcal{D}$ of $\mathcal{X}$ by $\mathcal{X}^{\otimes_{\mathcal{D}^m}}$. We recall a family of linear functions $\{\kappa_m : \mathcal{X}^{\otimes_{\mathcal{D}^m}} \mapsto \mathbb{C}\}_{m\geq 1}$ introduced in [19] (which are denoted as $c^{(1)}$ in [19]). By linearity, we mean that for $m \geq 1$ and any $a, b \in \mathcal{X}$ and $a, b \in \mathbb{C}$,

$$\kappa_m(\cdots \otimes (\alpha a + \beta b) \otimes \cdots) = \alpha\kappa_m(\cdots \otimes a \otimes \cdots) + \beta\kappa_m(\cdots \otimes b \otimes \cdots). \tag{B.19}$$

For convenience, we call $\{\kappa_m\}$ rectangular free kurtosis (or kurtosis when there is no ambiguity). For each $m \geq 1$ and any rectangular random variable a, we put

$$\kappa_{2m}(a) := \kappa_{2m}(a \otimes a^* \otimes \cdots \otimes a \otimes a^*). \tag{B.20}$$

We consider consider the even order as odd order cumulants vanishes for all rectangular elements.

Remark B.13. In [19, 18], the free cumulants refer to a family of linear functions between $\mathcal{X}^{\otimes_{\mathcal{D}^n}}$ and $\mathcal{D}$. The cumulants we introduced here are their coefficient functions of p_1.

The following vanishing lemma holds for the rectangular cumulants defined as in above.

Theorem B.14 (Vanishing of mixed cumulants)**.** *A family $(x_i)_{i\in I}$ of elements in $\mathcal{X}$ is free with amalgamation over $\mathcal{D}$ if and only if for all $n \geq 2$, and $i(1), \cdots, i(n) \in I$, we have $\kappa_n(x_{i(1)} \otimes \cdots \otimes x_{i(n)}) = 0$ whenever there exists $1 \leq l, k \leq n$ with $i(l) \neq i(k)$.*

Therefore, for any $a, b \in \mathcal{X}_{12}$ that are free,

$$\kappa_{2m}(a + b) = \kappa_{2m}(a) + \kappa_{2m}(b), \quad \forall m \in \mathbb{N}.$$

B.1.3.3 Rectangular free entropy – free additivity

The free entropy χ for rectangular free probability space are introduced in [18]. Let $a_1, \cdots, a_n$ denotes rectangular random variables, we have that

$$\chi(a_1, \cdots, a_n) \leq \chi(a_1) + \cdots + \chi(a_n). \tag{B.21}$$

And the equality in (B.21) holds if and only if $a_1, \cdots, a_n$ are free (See Corollary 5.16 in [18]).

B.1.3.4 Analogy of Gaussian random variables in rectangular free probability: the free Poisson element

Definition B.15. Given an rectangular probability space $(\mathcal{X}, \varphi)$. An rectangular random variable $a \in \mathcal{X}_{12}$ is a free Poisson element if

$$\kappa_{2m}(a) \equiv 0, \qquad \text{for } m \geq 2. \tag{B.22}$$

B.1.4 When are random matrices free?

Here, we describe the modelization of independent random matrices as free random variables and the explicit formulas of free kurtosis and entropy as functions of the input matrices.

B.1.4.1 Symmetric random matrix

Given a $N > 0$, we consider the algebra consists of all the real $N \times N$ matrices over scalar random variables $L^2(\Sigma, P)$:

$$\mathcal{X} = \mathbb{M}_N(L^2(\Sigma, P)) \tag{B.23}$$

and the functional φ on it is

$$\varphi(\boldsymbol{X}) = \frac{1}{N}\,\mathbb{E}[\mathrm{Tr}(\boldsymbol{X})]. \tag{B.24}$$

Denote the matrix transpose with complex conjugate by $*$. Then $(\mathcal{X}, \varphi)$ is a $*$-probability space.

We recall the following definition of asymptotic freely independence [95]

Definition B.16 (Asymptotic independence)**.** Let, for each $N \in \mathbb{N}$, $(\mathcal{X}_N, \varphi_N)$ be a non-commutative probability space. Let I be an index set and consider for each $i \in I$ and each $N \in \mathbb{N}$ random variables $a_i(N) \in \mathcal{X}_N$. We say that $\{a_i(N)\}_{i\in I}$ are asymptotically free ($N \to \infty$), if $(a_i(N))_{i\in I}$ converges in distribution towards $(a_i)_{i\in I}$ for some random variables $a_i \in \mathcal{X}$ ($i \in I$) in some non-commutative probability space $(\mathcal{X}, \varphi)$ and if the limits $(a_i)_{i\in I}$ are free in $(\mathcal{X}, \varphi)$.

A pair of symmetric (Hermitian) random matrices with isotropically random eigenvectors that are independent of the eigenvalues (and each other) are asymptotically free [95].

Given the $*$-probability space $(\mathcal{X}, \varphi(\cdot))$ defined as above. Recall the free kurtosis defined in (7.3). Thus for a self-adjoint random matrix $\boldsymbol{X} \in \mathcal{X}$ with $\varphi(\boldsymbol{X}) = 0$, the free kurtosis is explicitly given by

$$\kappa_4(\boldsymbol{X}) = \frac{1}{N}\,\mathbb{E}[\mathrm{Tr}(\boldsymbol{X}^4)] - 2\left(\frac{1}{N}\,\mathbb{E}[\mathrm{Tr}(\boldsymbol{X}^2)]\right)^2. \tag{B.25}$$

Also, denote the eigenvalues density function of X by $\mu(x)$, free entropy is defined by [65]

$$\chi(\boldsymbol{X}) = \int\int \log|x-y|\mathrm{d}\mu(x)\mathrm{d}\mu(y). \tag{B.26}$$

For a large class of random matrices $\boldsymbol{X}$, the free kurtosis and entropy concentrates around a deterministic value when N is large. For example, if $\boldsymbol{X}$ is a Wigner matrix

or Wishart matrix, then $\mathrm{Var}[\kappa_4(\boldsymbol{X})] \to 0$ and $\mathrm{Var}[\chi(\boldsymbol{X})] \to 0$ as $N \to \infty$. Thus single sample gives us an accurate empirical estimate. Given an sample of a random matrix X with $\mathrm{Tr}(\boldsymbol{X}) = 0$, the empirical free kurtosis is

$$\widehat{\kappa}_4(\boldsymbol{X}) = \frac{1}{N}\,\mathrm{Tr}(\boldsymbol{X}^4) - 2\left(\frac{1}{N}\,\mathrm{Tr}(\boldsymbol{X}^2)]\right)^2. \tag{B.27}$$

Also, the empirical free entropy of $\boldsymbol{X}$ is given by

$$\widehat{\chi}(\boldsymbol{X}) = \frac{1}{N(N-1)}\sum_{i\neq j}\log|\lambda_i - \lambda_j|, \tag{B.28}$$

where λ_i denotes the eigenvalue of $\boldsymbol{X}$.

B.1.4.2 Rectangular random matrix

Consider rectangular random matrix of size $N \times M$, and assume that $N \leq M$. In [19], the author embeded $N \times M$ matrix into top right block of a $(N+M)\times(N+M)$ "extension matrix". The algebra of all $(N+M)\times(N+M)$ random matrices together with this block structure structure is defined as an rectangular probability space $(\mathbb{M}_{N+M}(L^2(\Sigma,\mathbb{P})), \mathrm{diag}(I_N, 0_M), \mathrm{diag}(0_N, I_M), \frac{1}{N}\,\mathrm{Tr}, \frac{1}{M}\,\mathrm{Tr})$ [19].

We recall the following definition of asymptotic freely independence in rectangular probability space [18].

Definition B.17 (Asymptotic freely independence)**.** Let, for each $N \in \mathbb{N}$, $(\mathcal{X}_N, p_1(N), p_2(N), \varphi_{1,N}, \varphi_{2,N})$ be a $(\rho_{1,N}, \rho_{2,N})$-rectangular probability space such that

$$(\rho_{1,N}, \rho_{2,N}) \to (\rho_1, \rho_2), \qquad N \to \infty.$$

Let I be an index set and consider for each $i \in I$ and each $N \in \mathbb{N}$ random variables $a_i(N) \in \mathcal{X}_N$. We say that $\{a_i(N)\}_{i\in I}$ are asymptotically free $(N \to \infty)$, if $(a_i(N))_{i\in I}$ converges in $\mathcal{D}$-distribution towards $(a_i)_{i\in I}$ for some random variables $a_i \in \mathcal{X}$ $(i \in I)$

in some (ρ_1, ρ_2)-probability space $(\mathcal{X}, p_1, p_2, \varphi_1, \varphi_2)$ and if the limits $(a_i)_{i\in I}$ are free in $(\mathcal{X}, \varphi)$.

And independent bi-unitary invariant rectangular random matrices with converging singular law are asymptotically freely independent [19, 18].

Following (7.4), the free kurtosis for a single $N \times M$ random matrices $\boldsymbol{X}$ is given by

$$\kappa_4(\boldsymbol{X}) = \frac{1}{N}\,\mathbb{E}[\mathrm{Tr}((\boldsymbol{X}\boldsymbol{X}^H)^2)] - (1+\frac{N}{M})\left(\frac{1}{N}\,\mathbb{E}[\mathrm{Tr}((\boldsymbol{X}\boldsymbol{X}^H))]\right)^2. \tag{B.29}$$

Denote the probability density function of eigenvalues of $\boldsymbol{X}\boldsymbol{X}^H$ by $\mu(x)$, set $\alpha = \frac{N}{N+M}$ and $\beta = \frac{M}{N+M}$, the free entropy is given by [18]

$$\chi(\boldsymbol{X}) = \alpha^2 \int\int \log|x-y|\mathrm{d}\mu(x)\mathrm{d}\mu(y) + (\beta-\alpha)\alpha\int \log x \mathrm{d}\mu(x). \tag{B.30}$$

Again, empirical statistics over a single sample of large dimension give an accurate estimate of limit value. Given a realization of rectangular random matrix $\boldsymbol{X} \in \mathbb{R}^{N\times M}$, the empirical free kurtosis is given by

$$\widehat{\kappa}_4(\boldsymbol{X}) = \frac{1}{N}\,\mathrm{Tr}((\boldsymbol{X}\boldsymbol{X}^H)^2) - (1+\frac{N}{M})\left(\frac{1}{N}\,\mathrm{Tr}(\boldsymbol{X}\boldsymbol{X}^H)\right)^2. \tag{B.31}$$

The empirical free entropy is given by

$$\widehat{\chi}(\boldsymbol{X}) = \frac{\alpha^2}{N(N-1)}\sum_{i\neq j}\log|\lambda_i-\lambda_j| + \frac{(\beta-\alpha)\alpha}{N}\sum_{i=1}^{N}\log\lambda_i, \tag{B.32}$$

where λ_i denote the eigenvalue of $\boldsymbol{X}\boldsymbol{X}^H$.

B.2 Matrix Embeddings

One restriction of ICA is that it only operates on vector-valued components (see Section B.3). In contrast, FCF applies to data whose matrix-valued components

that can be of arbitrary dimensions. Thus, one can embed components into new dimensions potentially obtain a better performance with FCA. In this section, we list several matrix embedding algorithms.

For $\boldsymbol{Z} = [\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_N]^T$ where the $\boldsymbol{Z}_i$ are rectangular matrices, Algorithm 3 embeds Z_i in the upper diagonal parts of a $N' \times N'$ self-adjoint matrices. In practice, the target dimension N' should be picked such that there no loss of information while also avoiding too many artificial zeros. To embed $\boldsymbol{Z}_i$ into rectangular matrices of other dimensions, we introduce Algorithm 5. Putting the above embeddings and appropriate FCF algorithms together, we get Algorithm 4 and Algorithm 6. One easily state the analogs of the above algorithms for data containing self-adjoint matrices; for the sake of brevity, we omit them here.

If the $\boldsymbol{Z}_i$ are vectors, one can use the spectrogram to embed them into matrices. The spectrogram of a vector is the alignment of the discrete Fourier transform of a sliding window. The outcome is a complex rectangular matrix to which we can apply rectangular FCFs. This is summarized in Algorithm 7.

Algorithm 3 Symmetric Embedding

Input: $\boldsymbol{Z} = [\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_N]^T$ where $\boldsymbol{Z}_i \in \mathbb{C}^{N \times M}$.

Input: Target dimension $N' \times N'$.

1. Draw S uniformly from all subsets of $\{1, \cdots, \frac{N'(N'-1)}{2}\}$ with size $\frac{N'(N'-1)}{2} - NM$.
2. **for** $i = 1, \cdots, s$
3. Construct $z' \in \mathbb{R}^{\frac{N'(N'-1)}{2}}$ by setting $z'[S] = 0$ and $z'[S^c] = vec(\boldsymbol{Z}_i)$.
4. Fill the upper diagonal part of zero matrix $\boldsymbol{Z}' \in \mathbb{C}^{N' \times N'}$ by setting $(\boldsymbol{Z}'_{ij})_{j>i} = z'$.
5. Construct self-adjoint matrix $\boldsymbol{Z}'_i = \boldsymbol{Z}' + (\boldsymbol{Z}')^H$.
6. **end for**
7. **return:** $\boldsymbol{Z}' = [\boldsymbol{Z}'_1, \cdots, \boldsymbol{Z}'_s]^T$.

Algorithm 4 Symmetric Embedding FCF

Input: $\boldsymbol{Z} = [\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_N]^T \in \mathbb{C}_{sN \times M}$ where $\boldsymbol{Z}_i \in \mathbb{C}^{N \times M}$.

Input: Target dimension N' such that $\frac{N'(N'-1)}{2} \geq NM$.

1. Apply Algorithm 3 to $\boldsymbol{Z}$ and find $\boldsymbol{Z}'$.

2. Apply Algorithm 2 to $\boldsymbol{Z}'$ and find estimated mixing matrix $\widehat{\boldsymbol{A}}$.

3. Compute $\widehat{\boldsymbol{X}} = (\widehat{\boldsymbol{A}}^{-1} \otimes \boldsymbol{I}_N)\boldsymbol{Z}$ such that $\boldsymbol{Z} = (\widehat{\boldsymbol{A}} \otimes \boldsymbol{I}_N)\widehat{\boldsymbol{X}}$.

4. **return:** $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{X}}$.

Algorithm 5 Rectangular Embedding

Input: $\boldsymbol{Z} = [\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_N]^T \in \mathbb{C}_{sN \times M}$ where $\boldsymbol{Z}_i \in \mathbb{C}^{N \times M}$.

Input: Target dimension N' and M' such that $N'M' \geq NM$.

1. Draw S uniformly from all subsets of $\{1, \cdots, N'M'\}$ with size $N'M' - NM$.

2. **for** $i = 1, \cdots, s$

4. Construct $z' \in \mathbb{R}^{N'M'}$ by setting $z'[S] = 0$ and $z'[S^c] = vec(\boldsymbol{Z}_i)$.

5. Reshape z' to $\boldsymbol{Z}'_i \in \mathbb{C}^{N' \times M'}$.

6. **end for**

7. **return:** Return $\boldsymbol{Z}' = [\boldsymbol{Z}'_1, \cdots, \boldsymbol{Z}'_s]^T$

Algorithm 6 Rectangular Embedding FCF

Input: $\boldsymbol{Z} = [\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_N]^T \in \mathbb{C}_{sN \times M}$ where $\boldsymbol{Z}_i \in \mathbb{C}^{N \times M}$.

Input: Target dimension N' and M' such that $N'M' \geq NM$

1. Apply Algorithm 5 to $\boldsymbol{Z}$ and get $\boldsymbol{Z}'$.

2. Apply Algorithm 2 to $\boldsymbol{Z}' = [\boldsymbol{Z}'_1, \cdots, \boldsymbol{Z}'_N]^T$ and get the estimated mixing matrix $\widehat{\boldsymbol{A}}$.

3. Compute $\widehat{\boldsymbol{X}} = (\widehat{\boldsymbol{A}}^{-1} \otimes \boldsymbol{I}_N)\boldsymbol{Z}$ such that $\boldsymbol{Z} = (\widehat{\boldsymbol{A}} \otimes \boldsymbol{I}_N)\widehat{\boldsymbol{X}}$.

4. **return:** $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{X}}$.

Algorithm 7 Spectrogrm Embedding FCF

Input: $\boldsymbol{Z} = [\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_N]^T \in \mathbb{C}_{s\times N}$ where $\boldsymbol{Z}_i \in \mathbb{C}^{1\times N}$.

Input: Necessary parameters for spectrogram

1. For each $\boldsymbol{Z}_i$, for $i = 1, \cdots, s$, compute the spectrogram $\boldsymbol{Z}_i'$.

2. Apply Algorithm 2 to $\boldsymbol{Z}' = [\boldsymbol{Z}_1', \cdots, \boldsymbol{Z}_N']^T$ and get the estimated mixing matrix $\widehat{\boldsymbol{A}}$.

3. Compute $\widehat{\boldsymbol{X}} = (\widehat{\boldsymbol{A}}^{-1} \otimes \boldsymbol{I}_N)\boldsymbol{Z}$ such that $\boldsymbol{Z} = (\widehat{\boldsymbol{A}} \otimes \boldsymbol{I}_N)\widehat{\boldsymbol{X}}$.

4. **return:** $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{X}}$.

B.3 Independent Component Factorization

We would like to numerically compare FCA with ICA, and begin by providing a summary of the ICA algorithm. Given data whose components are rectangular matrices, we first vectorize them and then apply ICA. We once again perform a whitening process (see Algorithm 8) and solve an optimization problem.

Here, we present Algorithm 9 whose optimization problem is based on the empirical (scalar) kurtosis $\widehat{c}_4(\cdot)$ or the empirical (scalar) negentropy $\widehat{\mathcal{E}}(\cdot)$. We call them kurtosis-based ICF and entropy-based ICF respectively. Given a centered and whitened vector $x \in \mathbb{R}^T$, its empirical kurtosis $\widehat{c}_4(x)$ can be expressed as

$$\widehat{c}_4(x) = \frac{1}{T}\sum_{i=1}^{T} x_i^4 - 3\left(\frac{1}{T}\sum_{i=1}^{T} x_i^2\right)^2. \tag{B.33}$$

The negentropy $\mathcal{E}(x)$ is defined as

$$\mathcal{E}(x) = h(g_x) - h(x), \tag{B.34}$$

where $h(x)$ denotes the entropy of random variable x (see (B.6)) and g_x denote the Gaussian random variable with the same mean and variance as x. It is used as a measure of distance to normality. The empirical negentropy $\widehat{\mathcal{E}}(x)$ involves the empirical distribution of x, which is computationally difficult. Fortunately, it can also be expressed as a infinite sum of cumulants. Thus in practice, $\widehat{\mathcal{E}}(x)$ can be approximated by a finite truncation of that sum [43, Theorem 14 and (3.2) pp. 295].

In the simulation of this paper, we adapt the following approximation (see Section 5 of [71]):

$$\widehat{\mathcal{E}}(x) = \frac{1}{12}\left(\frac{1}{T}\sum_{i=1}^{T} x_i^3\right)^2 + \frac{1}{48}\widehat{c}_4(x) = \text{also cumulants} \tag{B.35}$$

Algorithm 8 Reshape and whitening

Input: $\boldsymbol{Z} = [\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_N]^T \in \mathbb{C}_{sN \times M}$ where $\boldsymbol{Z}_i \in \mathbb{C}_{N \times M}$.

1. For $\boldsymbol{z} = [z_1, \cdots, z_s]^T$, where $z_i = vec(\boldsymbol{Z}_i)$, Compute $\mu_z = \text{mean}(z, 2)$ and $\widetilde{\boldsymbol{z}} = \boldsymbol{z} - \mu_z 1_{NM}^T$.

3. Compute $\boldsymbol{C} = \frac{1}{NM}\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^H$ and the eigenvalue decomposition $\Re\boldsymbol{C} = \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^T$.

4. Compute $\boldsymbol{y} = \boldsymbol{U}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^T\widetilde{\boldsymbol{z}}$.

5. **return:** $\boldsymbol{y}, \boldsymbol{\Sigma}, \boldsymbol{U}$.

Algorithm 9 Prototypical ICF

Input: $\boldsymbol{Z} = [\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_n]^T \in \mathbb{C}_{sN \times M}$ where $\boldsymbol{Z}_i \in \mathbb{C}_{N \times M}$

1. Compute $\boldsymbol{y}, \boldsymbol{\Sigma}, \boldsymbol{U}$ by applying Algorithm 8 to Z.

2. Compute

$$\widehat{\boldsymbol{W}} = \underset{\boldsymbol{W} \in O(n)}{\arg\min} \sum \widehat{F}.\left(\boldsymbol{W}^T\boldsymbol{y}\right),$$

where $\widehat{F}(\cdot)$ is equal to $-|\widehat{c}_4(\cdot)|$ for kurtosis-based ICF or $-\widehat{\mathcal{E}}(\cdot)$ for entropy-based ICF.

3. Compute $\widehat{\boldsymbol{A}} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^T\widehat{\boldsymbol{W}}$ and $\widehat{\boldsymbol{X}} = (\widehat{\boldsymbol{A}}^{-1} \otimes \boldsymbol{I}_N)\boldsymbol{Z}$.

4. Sorting components of $\widehat{\boldsymbol{X}}$ by kurtosis or entropy. Permute the columns of $\widehat{\boldsymbol{A}}$ correspondingly.

5. **return:** $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{X}}$.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] M. Aizenman, J. L. Lebowitz, and D. Ruelle. Some rigorous results on the Sherrington-Kirkpatrick spin glass model. *Comm. Math. Phys.*, 112(1):3–20, 1987.

[2] Gernot Akemann, Jinho Baik, and Philippe Di Francesco. *The Oxford handbook of random matrix theory*. Oxford University Press, 2011.

[3] Luís B Almeida. MISEP–Linear and nonlinear ICA based on mutual information. *Journal of Machine Learning Research*, 4(Dec):1297–1318, 2003.

[4] R. Alvarez Baños, A. Cruz, L. A. Fernandez, J. M. Gil-Narvion, A. Gordillo-Guerrero, M. Guidetti, A. Maiorano, F. Mantovani, E. Marinari, V. Martin-Mayor, J. Monforte-Garcia, A. Muñoz Sudupe, D. Navarro, G. Parisi, S. Perez-Gaviro, J. J. Ruiz-Lorenzo, S. F. Schifano, B. Seoane, A. Tarancon, R. Tripiccione, and D. Yllanes. Static versus dynamic heterogeneities in the $d = 3$ Edwards-Anderson-Ising spin glass. *Phys. Rev. Lett.*, 105:177202, Oct 2010.

[5] Greg W Anderson and Brendan Farrell. Asymptotically liberating sequences of random unitary matrices. *Advances in Mathematics*, 255:381–413, 2014.

[6] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*, volume 118. Cambridge university press, 2010.

[7] Sanjeev Arora, Rong Ge, Ankur Moitra, and Sushant Sachdeva. Provable ica with unknown gaussian noise, with implications for gaussian mixtures and autoencoders. In *Advances in Neural Information Processing Systems*, pages 2375–2383, 2012.

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

[9] Zhidong Bai and Jun Yao. On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli*, 11(6):1059–1092, 2005.

[10] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005.

[11] Jinho Baik, Pierre Le Doussal, and Hao Wu. Spherical spin glass model with external field. in preparation.

[12] Jinho Baik and Ji Oon Lee. Fluctuations of the free energy of the spherical Sherrington-Kirkpatrick model. *J. Stat. Phys.*, 165(2):185–224, 2016.

[13] Jinho Baik and Ji Oon Lee. Correction to: Fluctuations of the free energy of the spherical Sherrington-Kirkpatrick model with ferromagnetic interaction. *Ann. Henri Poincaré*, 18(11):3757–3758, 2017.

[14] Jinho Baik and Ji Oon Lee. Free energy of bipartite spherical Sherrington–Kirkpatrick model. *arXiv preprint arXiv:1711.06364*, 2017.

[15] Jinho Baik, Ji Oon Lee, and Hao Wu. Ferromagnetic to paramagnetic transition in spherical spin glass. *Journal of Statistical Physics*, 173(5):1484–1522, 2018.

[16] Zhigang Bao, Guangming Pan, and Wang Zhou. Central limit theorem for partial linear eigenvalue statistics of Wigner matrices. *J. Stat. Phys.*, 150(1):88–129, 2013.

[17] Anthony J Bell and Terrence J Sejnowski. The independent components of natural scenes are edge filters. *Vision research*, 37(23):3327–3338, 1997.

[18] Florent Benaych-Georges. Rectangular random matrices, entropy, and fisher's information. *Journal of Operator Theory*, pages 371–419, 2009.

[19] Florent Benaych-Georges. Rectangular random matrices, related convolution. *Probability Theory and Related Fields*, 144(3-4):471–515, 2009.

[20] R. N. Bhatt and A. P. Young. Numerical studies of Ising spin glasses in two, three, and four dimensions. *Physical Review B*, 37(10):5606, 1988.

[21] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[22] Alex Bloemendal and Bálint Virág. Limits of spiked random matrices I. *Probab. Theory Related Fields*, 156(3-4):795–825, 2013.

[23] Pau Bofill and Michael Zibulevsky. Underdetermined blind source separation using sparse representations. *Signal processing*, 81(11):2353–2362, 2001.

[24] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre. Manopt, a Matlab toolbox for optimization on manifolds. *Journal of Machine Learning Research*, 15:1455–1459, 2014.

[25] Anton Bovier, Irina Kurkova, Matthias Löwe, et al. Fluctuations of the free energy in the REM and the p-spin sk models. *The Annals of Probability*, 30(2):605–651, 2002.

[26] Philemon Brakel and Yoshua Bengio. Learning independent features with adversarial nets for non-linear ica. *arXiv preprint arXiv:1710.05050*, 2017.

[27] A. J. Bray and M. A. Moore. Replica-symmetry breaking in spin-glass theories. *Physical Review Letters*, 41(15):1068, 1978.

[28] Édouard Brézin, Vladimir Kazakov, Didina Serban, Paul Wiegmann, and Anton Zabrodin. *Applications of random matrices in physics*, volume 221. Springer Science & Business Media, 2006.

[29] M. Capitaine, C. Donati-Martin, and D. Féral. Central limit theorems for eigenvalues of deformations of Wigner matrices. *Ann. Inst. Henri Poincaré Probab. Stat.*, 48(1):107–133, 2012.

[30] Mireille Capitaine, Catherine Donati-Martin, and Delphine Féral. The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations. *Ann. Probab.*, 37(1):1–47, 2009.

[31] Jean-François Cardoso. High-order contrasts for independent component analysis. *Neural computation*, 11(1):157–192, 1999.

[32] Philippe Carmona and Yueyun Hu. Universality in Sherrington–Kirkpatrick's spin glass model. In *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, volume 42, pages 215–222. Elsevier, 2006.

[33] Tommaso Castellani and Andrea Cavagna. Spin-glass theory for pedestrians. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(05):P05012, 2005.

[34] Guillaume Cébron, Antoine Dahlqvist, and Camille Male. Universal constructions for spaces of traffics. *arXiv preprint arXiv:1601.00168*, 2016.

[35] Aiyou Chen and Peter J. Bickel. Efficient independent component analysis. *The Annals of Statistics*, 34(6):2825–2855, 2006.

[36] Wei-Kuo Chen. On the mixed even-spin Sherrington-Kirkpatrick model with ferromagnetic interaction. *Ann. Inst. Henri Poincaré Probab. Stat.*, 50(1):63–83, 2014.

[37] Wei-Kuo Chen, Partha Dey, and Dmitry Panchenko. Fluctuations of the free energy in the mixed p-spin models with external field. *Probab. Theory Related Fields*, 168(1-2):41–53, 2017.

[38] Wei-Kuo Chen and Arnab Sen. Parisi formula, disorder chaos and fluctuation for the ground state energy in the spherical mixed p-spin models. *Comm. Math. Phys.*, 350(1):129–173, 2017.

[39] E Colin Cherry. Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, 25(5):975–979, 1953.

[40] Brad S Chissom. Interpretation of the kurtosis statistic. *The American Statistician*, 24(4):19–22, 1970.

[41] GP Chistyakov and Friedrich Götze. Characterization problems for linear forms with free summands. *arXiv preprint arXiv:1110.1527*, 2011.

[42] F. Comets and J. Neveu. The Sherrington-Kirkpatrick model of spin glasses and stochastic calculus: the high temperature case. *Comm. Math. Phys.*, 166(3):549–564, 1995.

[43] Pierre Comon. Independent component analysis, a new concept? *Signal processing*, 36(3):287–314, 1994.

[44] Pierre Comon and Christian Jutten. *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

[45] Edmund A Cornish and Ronald A Fisher. Moments and cumulants in the specification of distributions. *Revue de l'Institut international de Statistique*, pages 307–320, 1938.

[46] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[47] Andrea Crisanti and H-J Sommers. The spherical p-spin interaction spin glass model: the statics. *Zeitschrift für Physik B Condensed Matter*, 87(3):341–354, 1992.

[48] J. R. L. De Almeida and David J. Thouless. Stability of the Sherrington-Kirkpatrick solution of a spin glass model. *Journal of Physics A: Mathematical and General*, 11(5):983, 1978.

[49] J. De Nobel and F. J. Du Chatenier. Specific heats of dilute alloys of manganese in silver and copper at low temperatures and in magnetic fields. *Physica*, 25(7-12):969–979, 1959.

[50] Amir Dembo and Ofer Zeitouni. Matrix optimization under random external fields. *J. Stat. Phys.*, 159(6):1306–1326, 2015.

[51] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[52] Alan Edelman and N Raj Rao. Random matrix theory. *Acta Numerica*, 14:233–297, 2005.

[53] S. F. Edwards and P. W. Anderson. Theory of spin glasses. II. *Journal of Physics F: Metal Physics*, 6:1927, 1976.

[54] Samuel Frederick Edwards and Phil W. Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5(5):965, 1975.

[55] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Spectral statistics of Erdös-Rényi graphs I: Local semicircle law. *Ann. Probab.*, 41(3B):2279–2375, 2013.

[56] Jan Eriksson and Visa Koivunen. Blind identifiability of class of nonlinear instantaneous ICA models. In *2002 11th European Signal Processing Conference*, pages 1–4. IEEE, 2002.

[57] Delphine Féral and Sandrine Péché. The largest eigenvalue of rank one deformation of large Wigner matrices. *Communications in Mathematical Physics*, 272(1):185–228, 2007.

[58] P. J. Forrester. *Log-gases and random matrices*, volume 34 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2010.

[59] Alan Frieze, Mark Jerrum, and Ravi Kannan. Learning linear transformations. In *Proceedings of 37th Conference on Foundations of Computer Science*, pages 359–368. IEEE, 1996.

[60] J. Fröhlich and B. Zegarliński. Some comments on the Sherrington-Kirkpatrick model of spin glasses. *Comm. Math. Phys.*, 112(4):553–566, 1987.

[61] Yan V. Fyodorov and Pierre le Doussal. Topology trivialization and large deviations for the minimum in the simplest random optimization. *J. Stat. Phys.*, 154(1-2):466–490, 2014.

[62] Geoffrey Grimmett, David Stirzaker, et al. *Probability and random processes*. Oxford university press, 2001.

[63] Alice Guionnet, M Maı, et al. A Fourier view on the R-transform and related asymptotics of spherical integrals. *Journal of functional analysis*, 222(2):435–490, 2005.

[64] Simon Haykin and Zhe Chen. The cocktail party problem. *Neural computation*, 17(9):1875–1902, 2005.

[65] Fumio Hiai and Dénes Petz. *The semicircle law, free random variables and entropy*, volume 77 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2000.

[66] Patrik O Hoyer and Aapo Hyvärinen. Independent component analysis applied to feature extraction from colour and stereo images. *Network: computation in neural systems*, 11(3):191–210, 2000.

[67] A. J. Hyvarinen and H. Morioka. Nonlinear ICA of temporally dependent stationary sources. Proceedings of Machine Learning Research, 2017.

[68] Aapo Hyvarinen. A family of fixed-point algorithms for independent component analysis. In *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 3917–3920. IEEE, 1997.

[69] Aapo Hyvarinen. One-unit contrast functions for independent component analysis: A statistical analysis. In *Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop*, pages 388–397. IEEE, 1997.

[70] Aapo Hyvarinen. Fast and robust fixed-point algorithms for Independent Component Analysis. *IEEE transactions on Neural Networks*, 10(3):626–634, 1999.

[71] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent component analysis*, volume 46. John Wiley & Sons, 2004.

[72] Aapo Hyvarinen and Hiroshi Morioka. Unsupervised feature extraction by time-contrastive learning and nonlinear ICA. In *Advances in Neural Information Processing Systems*, pages 3765–3773, 2016.

[73] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000.

[74] Aapo Hyvarinen, Hiroaki Sasaki, and Richard E Turner. Nonlinear ICA using auxiliary variables and generalized contrastive learning. *arXiv preprint arXiv:1805.08651*, 2018.

[75] Pauliina Ilmonen, Klaus Nordhausen, Hannu Oja, and Esa Ollila. A new performance index for ica: properties, computation and asymptotic analysis. In *International Conference on Latent Variable Analysis and Signal Separation*, pages 229–236. Springer, 2010.

[76] Kurt Johansson. On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.*, 91(1):151–204, 1998.

[77] Antti Knowles and Jun Yin. The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.*, 66(11):1663–1750, 2013.

[78] J. M. Kosterlitz, D. J. Thouless, and Raymund C. Jones. Spherical model of a spin-glass. *Phys. Rev. Lett.*, 36(20):1217, 1976.

[79] Jorge Kurchan, Giorgio Parisi, Pierfrancesco Urbani, and Francesco Zamponi. Exact theory of dense amorphous hard spheres in high dimension. ii. the high density regime and the gardner transition. *The Journal of Physical Chemistry B*, 117(42):12979–12994, 2013.

[80] Benjamin Landon and Philippe Sosoe. Fluctuations of the overlap at low temperature in the 2-spin spherical SK model. in preparation.

[81] Daniel D. Lee and H. Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788, 1999.

[82] Te-Won Lee. Independent Component Analysis. In *Independent component analysis*, pages 27–66. Springer, 1998.

[83] A. Lytova and L. Pastur. Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.*, 37(5):1778–1840, 2009.

[84] Camille Male. Traffic distributions and independence: permutation invariant random matrices and the three notions of independence. *arXiv preprint arXiv:1111.4662*, 2011.

[85] W. Marshall. Specific heat of dilute alloys. *Physical Review*, 118(6):1519, 1960.

[86] Madan Lal Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.

[87] Carl D Meyer and Gilbert W Stewart. Derivatives and perturbations of eigenvectors. *SIAM Journal on Numerical Analysis*, 25(3):679–691, 1988.

[88] Marc Mézard, Giorgio Parisi, Nicolas Sourlas, G Toulouse, and Miguel Virasoro. Nature of the spin-glass phase. *Physical review letters*, 52(13):1156, 1984.

[89] Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.

[90] James A Mingo and Roland Speicher. *Free probability and random matrices*, volume 35. Springer, 2017.

[91] Yoshiki Mitsui, Daichi Kitamura, Shinnosuke Takamichi, Nobutaka Ono, and Hiroshi Saruwatari. Blind Source Separation based on independent low-rank matrix analysis with sparse regularization for time-series activity. In *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, pages 21–25. IEEE, 2017.

[92] Patrick Kofod Mogensen and Asbjørn Nilsen Riseth. Optim: A mathematical optimization package for Julia. *Journal of Open Source Software*, 3(24):615, 2018.

[93] J. A. Mydosh. Spin glasses: an experimental introduction. 1993.

[94] Raj Rao Nadakuditi and Hao Wu. lingluanwh/fca.jl: a blind source separation package based on the random matrix theory and free probability, May 2019.

[95] Alexandru Nica and Roland Speicher. *Lectures on the combinatorics of free probability*, volume 13. Cambridge University Press, 2006.

[96] Erkki Oja and Zhijian Yuan. The fastica algorithm revisited: Convergence analysis. *IEEE Transactions on Neural Networks*, 17(6):1370–1381, 2006.

[97] Sean O'Rourke and Alexander Soshnikov. Partial linear eigenvalue statistics for Wigner and sample covariance random matrices. *J. Theoret. Probab.*, 28(2):726–744, 2015.

[98] Pentti Paatero and Unto Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2):111–126, 1994.

[99] Dmitry Panchenko. The parisi ultrametricity conjecture. *arXiv preprint arXiv:1112.1003*, 2011.

[100] Dmitry Panchenko et al. The parisi formula for mixed p-spin models. *The Annals of Probability*, 42(3):946–958, 2014.

[101] Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Physical Review Letters*, 43(23):1754, 1979.

[102] Giorgio Parisi. Toward a mean field theory for spin glasses. *Physics Letters A*, 73(3):203–205, 1979.

[103] Giorgio Parisi. A sequence of approximated solutions to the SK model for spin glasses. *Journal of Physics A: Mathematical and General*, 13(4):L115, 1980.

[104] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[105] Alessandro Pizzo, David Renfrew, and Alexander Soshnikov. On finite rank deformations of Wigner matrices. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(1):64–94, 2013.

[106] David Renfrew and Alexander Soshnikov. On finite rank deformations of Wigner matrices ii: Delocalized perturbations. *Random Matrices: Theory and Applications*, 2(01):1250015, 2013.

[107] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical review letters*, 35(26):1792, 1975.

[108] Peter J Smith. A recursive formulation of the old problem of obtaining moments from cumulants and vice versa. *The American Statistician*, 49(2):217–218, 1995.

[109] Alexander Soshnikov. Universality at the edge of the spectrum in Wigner random matrices. *Communications in mathematical physics*, 207(3):697–733, 1999.

[110] Roland Speicher. Multiplicative functions on the lattice of non-crossing partitions and free convolution. *Mathematische Annalen*, 298(1):611–628, 1994.

[111] Eliran Subag and Ofer Zeitouni. The extremal process of critical points of the pure p-spin spherical spin glass model. *Probability theory and related fields*, 168(3-4):773–820, 2017.

[112] Michel Talagrand. Free energy of the spherical mean field model. *Probab. Theory Related Fields*, 134(3):339–382, 2006.

[113] Michel Talagrand. The Parisi formula. *Annals of Mathematics*, 163(1):221–263, 2006.

[114] Michel Talagrand. *Mean field models for spin glasses: Volume I: Basic examples*, volume 54. Springer Science & Business Media, 2010.

[115] G. Toulouse. Theory of the frustration effect in spin glasses: I. *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*, 9:99, 1987.

[116] Craig A. Tracy and Harold Widom. Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.*, 159(1):151–174, 1994.

[117] Dan Voiculescu. Limit laws for random matrices and free products. *Inventiones mathematicae*, 104(1):201–220, 1991.

[118] Dan Voiculescu. The analogues of entropy and of fisher's information measure in free probability theory, i. *Communications in mathematical physics*, 155(1):71–92, 1993.

[119] Dan Voiculescu. The analogues of entropy and of fisher's information measure in free probability theory, ii. *Inventiones mathematicae*, 118(1):411–440, 1994.

[120] Dan Voiculescu. Operations on certain non-commutative operator-valued random variables, in recent advances in operator algebras. *Astérisque*, 232:243–275, 1995.

[121] Dan Voiculescu. The analogues of entropy and of fisher's information measure in free probability theory, iv: maximum entropy and freeness, in free probability theory. *Fields Inst. Commun.*, 12:293–302, 1997.

[122] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, pages 548–564, 1955.

[123] Hao Wu and Raj Rao Nadakuditi. Free component analysis: Theory, algorithms & applications. *arXiv preprint arXiv:1905.01713*, 2019.

[124] Kei Yosida. Magnetic properties of cu-mn alloys. *Physical Review*, 106(5):893, 1957.

[125] Jo E. Zimmerman. Je zimmerman and fe hoare, j. phys. chem. solids 17, 52 (1960). *J. Phys. Chem. Solids*, 17:52, 1960.