Problems at the interface of high dimensional probability and random matrix theory

By

Xiaoyu Xie

A dissertation submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in the Division of Applied Mathematics at Brown University

PROVIDENCE, RHODE ISLAND

May 2024

This dissertation by Xiaoyu Xie is accepted in its present form by the Division of Applied Mathematics as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Date ______________ ______________________________

Kavita Ramanan, Advisor

Recommended to the Graduate Council

Date ______________ ______________________________

Kavita Ramanan, Reader

Date ______________ ______________________________

Paul Dupuis, Reader

Date ______________ ______________________________

Grigorios Paouris, Reader

Approved by the Graduate Council

Date ______________ ______________________________

Thomas A. Lewis, Dean of the Graduate School

Curriculum Vitae

Education

2019–2024 Ph.D., Applied Mathematics
Brown University (Providence, RI, USA)

2017-2019
M.Phil., Mathematics
The Chinese University of Hong Kong (Hong Kong, Hong Kong SAR)

2013-2017
Honors B.Sc., Mathematics
The Chinese University of Hong Kong (Hong Kong, Hong Kong SAR)

Publications

Papers

Xie, X., Goel, A., and Lopatto, P. (2023). Central limit theorem for the complex eigenvalues of Gaussian random matrices. arXiv preprint arXiv:2306.10243 (2023).

Xie, X., Benigni, L., Chen, N., and Lopatto, P. (2023). Fluctuations in Quantum Unique Ergodicity at the Spectral Edge. arXiv preprint arXiv:2303.11142 (2023).

Xie, X., Lopatto, P., and Ramanan, K. (2023). Quenched large deviation principles for random projections of ℓ_p^n balls. arXiv preprint arXiv:2308.00649 (2023).

Xie, X., and Ramanan, K. (2024). The ℓ_r-Levy-Grothendieck problem and $r \to p$ norms of Levy matrices arXiv preprint arXiv:2404.18299 (2024).

Teaching

Feb 2021 - Jun 2021
Mentor, *Directed Research Program*, (Brown University, Providence, RI)

Jan 2021 - May 2021
Teaching assistant, *APMA 1860/1941B Graphs and Networks*, (Brown University, Providence, RI)

Sep 2020 - Dec 2020
Teaching assistant, *APMA 2630 Theory of Probability I*, (Brown University, Providence, RI)

Awards

Sigma Xi Award, Brown University, (2022)

Graduate Fellowship, Brown University, (2019)

Dean's List, The Chinese University of Hong Kong, (2014-2016)

HKSAR Government Scholarship, The Chinese University of Hong Kong, (2016)

Undergraduate Mathematics Scholorship, The Chinese University of Hong Kong, (2015)

C.N.Yang Scholarship, The Chinese University of Hong Kong, (2014)

Acknowledgments

I would like to express my deepest gratitude to all those who provided me the possibility to complete this thesis. A heartfelt thank you to my advisor, Professor Kavita Ramanan, whose encouragement and insightful guidance were invaluable throughout my research journey.

I am also deeply thankful to my thesis committee members, Professor Paul Dupuis and Professor Grigorios Paouris, for their critical insights and expert feedback.

Special thanks are due to my collaborators Patrick Lopatto, Nixia Chen, Lucas Benigni, and Advay Goel, whose contributions were essential to my research. The discussions and support from fellow students I-Hsun Chen, Kevin Hu, Juniper Cocomello, Yanqiu Guo, and Yin-Ting Liao have been incredibly enriching and beneficial.

I must acknowledge the unwavering support of my family, particularly my father and mother, who provided more than I could ever ask for during this challenging journey.

I am also grateful to my friends Xin Lian, Sicheng Liu, Mengjie Liu, Enrui Zhang, Qian Zhang, Zhen Zhang, and many more, who have been there for me through thick and thin. Their support and encouragement have kept my spirits high.

Finally, I would like to express my gratitude to everyone who was involved in any manner and supported me throughout this process.

Contents

List of Figures

CHAPTER 1

INTRODUCTION

In the era of big data, the complexity of models in fields like statistical physics, bioinformatics, finance, and social network analysis has underscored the importance of understanding intricate structures within high-dimensional spaces. High-dimensional probability, which delves into the behavior of objects such as random vectors and random matrices in such contexts, has proven to be instrumental in uncovering key insights and sometimes surprising phenomena. By shedding light on the underlying patterns and behaviors, this branch of probability theory plays a pivotal role in navigating and making sense of the complexities inherent in high-dimensional data.

One of the fundamental challenges in understanding data residing in high-dimensional spaces is the 'curse of dimensionality'. This phenomenon refers to the exponential growth of the volume of the underlying space, which necessitates a corresponding increase in data, computational resources, and time to achieve reliable results. To address this challenge, researchers often turn to the study of asymptotic behavior, where the focus is on understanding the problem as the dimension tends toward infinity. This approach aims to uncover underlying patterns and structures that are otherwise obscured in finite-dimensional settings.

In this thesis, we delve into three distinct problems within the realm of high-dimensional probability, each associated with different ensembles of random matrices. Our exploration is conducted through the lens of asymptotic analysis, providing insights into the different aspects of these matrices and related objects as the dimensionality grows. Specifically, we focus on the following three ensembles:

1. heavy-tailed random matrices, characterized by their probability distributions with slowly decaying tails, may have infinite variance and/or infinite expectation;
2. light-tailed random matrices, known for their exponentially decaying tails;
3. and random matrices uniformly sampled from the Stiefel manifold, which is the set of matrices with orthonormal columns equipped with a natural manifold structure.

These ensembles, each with its unique properties and the questions that motivate their study, will be discussed in detail in the following sections.

1.1 Heavy-tailed random matrices, Grothendieck ℓ_r-problem, and the $r \to p$ operator norm

In recent years, heavy-tailed random variables have garnered significant attention due to their ability to model real-world phenomena, such as financial markets and power-law distributions in complex networks. In Chapter 2, we investigate problems associated with symmetric random matrices whose entries are independent heavy-tailed random variables. Specifically, we consider a symmetric random matrix $A = (a_{ij})_{1 \le i,j \le n}$ with entries in the domain of attraction of an α-stable law, where $\alpha \in (0, 2)$. This means that there exists a slowly varying function L (defined in Definition 2.2.1) such that

$$\mathbb{P}\left(|a_{ij}| \ge x\right) = L(x)x^{-\alpha}.$$

It is important to note that, as evident from the definition, the random variables a_{ij} do not possess a finite variance. Furthermore, when $\alpha \le 1$, a_{ij} does not even have a finite expectation. These properties distinguish heavy-tailed random variables from their light-tailed counterparts and require specialized techniques for their analysis.

The spectral statistics of heavy-tailed random matrices have been extensively studied over the past several decades. In [156], the distribution of the largest eigenvalues was investigated and shown to converge to a Poisson point process. Furthermore, it was proved in [14] that the limiting spectrum of symmetric heavy-tailed random matrices is supported on the entire real line. These findings stand in stark contrast to the behavior observed in light-tailed random matrices, where it is well known that the largest eigenvalue follows the Tracy-Widom distribution, while the limiting spectrum is supported on $[-2, 2]$ and follows semi-circle law (for a comprehensive overview of the spectral statistics of light-tailed random matrices, see the survey paper [85]).

In Chapter 2, we delve into the study of two optimization problems involving heavy-tailed random matrices. Specifically, we consider an $n \times n$ matrix A and examine the following:

1. The ℓ_r-Grothendieck problem, which is concerned with the quantity

$$M_r(A) := \max_{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_r \leq 1} \boldsymbol{x}^\top A \boldsymbol{x}, \quad \text{for } r \geq 1;$$

2. The $r \to p$ operator norm, defined as

$$\|A\|_{r \to p} := \sup_{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_r \leq 1} \|A\boldsymbol{x}\|_p.$$

Note that when $r = p = 2$, this corresponds to the spectral norm and is related to the largest eigenvalue problem mentioned above. Additionally The motivation for studying these problems is as follows:

First, both $M_r(A)$ and $\|A\|_{r \to p}$ are connected to a wide array of key quantities in various fields. For instance, when $r = 2$ and $r = \infty$, the ℓ_r-Grothendieck problem is linked to spectral partitioning and correlation clustering, respectively, and the computation of $\|A\|_{2\to 4}$ is known to be related to intriguing problems in theoretical computer science, such as the small-set expansion problem and the approximability of the injective tensor norm. For a more comprehensive list of quantities related to these two optimization problems, we refer the reader to Section 2.1.1.

Secondly, both quantities are known to be challenging to compute in general. Indeed, accurately computing $M_r(A)$ and $\|A\|_{r\to p}$ is NP-hard except for some special choices of r and p, such as $r = 2$ for $M_r(A)$. We discuss this in more depth in Section 2.1.1. Given this computational complexity, it is natural to investigate these problems for typical matrices, i.e., matrices with independent and identically distributed entries drawn from certain distributions.

In this work, we characterize the asymptotic distribution of $M_r(A)$ and $\|A\|_{r\to p}$ in terms of the parameters r, p and the heavy-tail index α. Our results elucidate the relationship between these quantities and the tail behavior of the matrix entries, providing insights with potential applications in various fields. In the non-spectral and heavy-tailed settings, new techniques have to be developed to analyze the desired asymptotics.

1.2 Wigner matrices and quantum unique ergodicity

Since its inception, a foundational principle of random matrix theory has been the notion that the asymptotic spectral properties of random matrices do not depend on the details of the distributions of their entries. This concept is known as universality. In Chapter 3, we explore the universality of the eigenvector, focusing on the Wigner matrix $H = (h_{ij})_{1\le i,j\le N}$, a symmetric random matrix with real entries having zero mean and variance $(1+\delta_{ij})/N$. We impose moment conditions on h_{ij}, requiring that for every $p \in \mathbb{N}$, there exists $\mu_p < \infty$ such that $\mathbb{E}\left[\left|\sqrt{N}h_{ij}\right|^p\right] \leqslant \mu_p$. Under this settings, we study the asymptotic fluctuations of the following quantity:

$$\sum_{\alpha\in\mathcal{I}}\langle \boldsymbol{q}_\alpha, \boldsymbol{u}\rangle^2$$

where $\mathcal{I} \in [N]$ and $\{\boldsymbol{q}_\alpha\}_{\alpha\in[N]}$ is an orthonormal basis of $\mathbb{R}^N$.

Our investigation is related to the Quantum Unique Ergodicity (QUE) conjecture [147], which roughly states that the L^2-normalized eigenfunctions of Laplacian, which corresponds to the solution to the time-independent Schrödinger equation in absence of potential energy, on compact negatively curved manifold becomes equidistributed in the large eigenvalue limit (or in physics jargon, the semiclassical limit).

The first step towards QUE conjecture is established by [150, 168, 58] in the 1980s through the quantum ergodicity theorem, which states that most eigenfunctions of a quantum system with classically ergodic dynamics become equidistributed in the semiclassical limit. It is important to note, however, that the quantum ergodicity theorem does not preclude the existence of a subsequence of eigenfunctions that fail to converge to uniform distribution as eigenvalues tends to infinity. While Quantum Ergodicity has been widely proven, QUE remains a challenging conjecture, proven only in specific cases such as arithmetic hyperbolic surfaces [131]. The exploration of QUE continues to be a vibrant area of research in understanding quantum chaos.

In the context of Wigner random matrix, an analogue of QUE is the delocalization phenomenon of eigenvectors. A quantitative formulation of delocalization is articulated as follows: Consider a deterministic index set $\mathcal{I} \subset [N]$ and an orthonormal basis $\{\boldsymbol{q}_\alpha\}_{\alpha\in[N]}$ of $\mathbb{R}^N$. For any unit eigenvector $\boldsymbol{u}$ of H, we have the high-probability bound

$$\left|\sum_{\alpha\in\mathcal{I}}\langle \boldsymbol{q}_\alpha, \boldsymbol{u}\rangle^2 - \frac{|\mathcal{I}|}{N}\right| \leqslant \frac{N^\varepsilon\sqrt{|\mathcal{I}|}}{N}.$$

This result is proved using a dynamical method [39, 24, 52]. Beyond the high probability bound, one can also ask the fluctuations of $\sum_{\alpha\in\mathcal{I}}\langle \boldsymbol{q}_\alpha, \boldsymbol{u}\rangle^2$ around the leading order term.

Label the eigenvalues of H in increasing order, $\lambda_1 \leq \ldots \leq \lambda_N$, it is established in [52] that

$$\sqrt{\frac{N^3}{2|\mathcal{I}|(N-|\mathcal{I}|)}}\left(\sum_{\alpha\in\mathcal{I}}\langle \boldsymbol{q}_\alpha, \boldsymbol{u}\rangle^2 - \frac{|\mathcal{I}|}{N}\right) \to \mathcal{N}(0,1), \tag{1.2.1}$$

whenever $\frac{|\mathcal{I}|(N-|\mathcal{I}|)}{N^2} \geq \delta$ for some small $\delta > 0$, and $\boldsymbol{u}$ is an eigenvector of H corresponds to λ_i with $\min(i, N-i) > cN$ for some small $c > 0$.

In this part of the thesis, our objective is to complement the fluctuation result of $\sum_{\alpha\in\mathcal{I}}\langle \boldsymbol{q}_\alpha, \boldsymbol{u}\rangle^2$ in [52] by establishing (1.2.1) for eigenvector $\boldsymbol{u}$ corresponding to λ_i with $\min(i, N-i) \leq N^{1-\tau}$ for a small constant $\tau > 0$. By establishing this result, and in conjunction with prior findings [52, 24], we provide a complete characterization of the fluctuations of $\sum_{\alpha\in\mathcal{I}}\langle \boldsymbol{q}_\alpha, \boldsymbol{u}\rangle^2$ for Wigner matrices across the entire spectrum.

1.3 Non-Hermitian random matrices and eigenvalue counting

The Ginibre ensemble, introduced by Jean Ginibre in 1965 [99], is a fundamental class of non-Hermitian random matrices in random matrix theory. This ensemble encompasses three types of matrices: real, complex, and quaternionic. After appropriate scaling, the limiting spectrum of these ensembles converges in distribution to the circular law, characterized by a uniform distribution over the unit disk.

Other quantities of interest are the eigenvalue counts, which refers to the number of eigenvalues in any region of the plane. The fluctuation of eigenvalue counting is well understood for the complex Ginibre ensemble, as a consequence of more general CLT for determinantal point process (see [155, Section 2]). However, the real Ginibre ensemble presents additional challenges: first, there is a non-zero probability that some eigenvalues are real; second, the underlying point correlation function associated with the eigenvalues is Pfaffian, rather than determinantal. Both factors contribute to the complexity of the analysis.

In Chapter 4, we establish a Central Limit Theorem (CLT) for eigenvalue counts in the real Ginibre Ensemble (GinOE). We consider the random matrix $G = (g_{ij})_{1\le i,j\le n}$, where the entries g_{ij}, $1 \le i, j \le n$, are mutually independent standard normal random variables. We characterize the fluctuation of the number of eigenvalues within any reasonable region in the upper half disk. More specifically, given region A in the upper half disk, and let N_A be the number of eigenvalues in A. Then we show that

$$\lim_{N\to\infty} \frac{N_A - \mathbb{E}[N_A]}{n^{1/4}} = \mathcal{N}\left(0, \frac{\ell(\partial A)}{2\pi^{3/2}}\right),$$

where $\ell(\partial A)$ is the length of boundary of A. Given that the eigenvalues of GinOE appear in conjugate pairs, this result extends to characterizing the fluctuations in the lower half disk as well.

1.4 Stiefel manifold, random projection, and large deviation principle

The Stiefel manifold, denoted as $\mathbb{V}_{n,k}$, is the set of all k-dimensional orthonormal frames in $\mathbb{R}^n$. In other words, it consists of all $n \times k$ matrices V with real entries such that $V^\top V = I_k$, where I_k is the $k \times k$ identity matrix. In Chapter 5, we consider the random matrix ensemble $A_{n,k}$ distributed according to the Haar measure (unique orthogonally invariant measure) on $\mathbb{V}_{n,k}$. Given a random vector $X^{(n)}$ in $\mathbb{R}^n$, we are interested in the large deviation behavior or $A_{n,k_n}^\top X^{(n)}$ for suitably sequences $\{k_n\}_{n\in\mathbb{N}}$.

The problem we explore here is motivated by the analysis of high-dimensional data through their lower-dimensional projections, for instance, using algorithms such as Projection Pursuit. This technique, initially introduced by [94] and further examined in [68, 106], aims to identify directions along which the projections are significantly distinct from a normal distribution. However, the quest for such projection directions faces a challenge due to Klartag's celebrated CLT for convex sets [117, 118]. This theorem, a cornerstone in the realm of universality results in convex geometry, asserts that in high-dimensional settings, projections of vectors uniformly distributed on a convex body along most directions closely resemble a normal distribution, rendering non-normal projection directions exceedingly rare.

One potential way to address the challenges presented by Klartag's CLT is to shift the focus beyond the scale of fluctuations and investigate the Large Deviation Principle (LDP) of the projection. The LDP, which focuses on the asymptotic behavior of the tails of sequences of probability distributions, is recognized for its ability to reveal the unique characteristics of the underlying probability distribution. This is achieved through the LDP rate function, which provides a detailed description of the exponential decay rates of the tail probabilities. By examining the LDP rate function, we can potentially uncover insights that extend beyond the universal behaviors captured by the CLT, thereby gaining a more nuanced understanding of the projection's behavior in high-dimensional settings. Motivated by this, the authors of [96] asked the following question: given a sequence of projection directions $\{\theta^{(n)}\}_{n\in\mathbb{N}}$, which are uniformly chosen from unit spheres of increasing dimension n, and consider the projection of random vectors $\{X_p^{(n)}\}_{n\in\mathbb{N}}$ onto $\theta^{(n)}$, where $X_p^{(n)}$ are independent of $\theta^{(n)}$ and are uniformly distributed on the so-called n-dimensional ℓ_p-balls $\mathbb{B}_p^n$, that is,

$$\mathbb{B}_p^n := \left\{ \boldsymbol{x} \in \mathbb{R}^n : \sum_{i=1}^n |x_i|^p \le n \right\},$$

can one identify the LDP rate function of the sequence $\{n^{-1/2}\langle \theta^n, X_p^{(n)}\rangle\}$ in a p-dependent way? They provided an affirmative answer to this question and further demonstrated that this rate function exhibits insensitivity to almost all sequences of projection directions $\theta^{(n)}$. Subsequent research by [116] expanded the scope of the one-dimensional projection discussed in [96] to encompass projections of any finite dimension. A key component of the proof involves analysis of the random matrix from the Stiefel manifold, which are pre-fixed independent of the n-dimensional random vectors $X^{(n)}$.

In this thesis, we extend the projection LDP from a finite dimensional projection to projection onto sequence of subspaces with an increasing dimension k_n, provided that k_n is asymptotically negligible compared to n. In the course of establishing this generalization, we addressed both technical and conceptual hurdles. The main technical challenge involved proving that the empirical measures of scalar products, between rows of a random matrix drawn from the Stiefel manifold and a deterministic vector of dimension k_n, converge in a suitable Wasserstein sense to a normal distribution, uniformly over e^{k_n} deterministic vectors. And the conceptual challenge stems from the absence of a canonical space to accommodate projected vectors with increasing dimensions.

CHAPTER 2

The ℓ_r-Levy-Grothendieck problem and $r \to p$ norms of Levy matrices

2.1 Introduction

2.1.1 Motivation and Context

Given an $n \times n$ matrix A, consider the following two optimization problems.

1. For $1 \le r \le \infty$, the ℓ_r-Grothendieck problem concerns the quantity

$$M_r(A) \coloneqq \max_{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_r \le 1} \boldsymbol{x}^\top A \boldsymbol{x}, \qquad \text{OPT 1}$$

   where $\|\boldsymbol{x}\|_r \coloneqq \left(\sum_{i=1}^n |x_i|^r\right)^{1/r}$ denotes the usual ℓ_r-norm of the vector $\boldsymbol{x}$.

2. For $1 \le r, p \le \infty$, the $r \to p$ operator norm of A is defined as

$$\|A\|_{r \to p} \coloneqq \sup_{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_r \le 1} \|A\boldsymbol{x}\|_p. \qquad \text{OPT 2}$$

For different choices of r and p, $M_r(A)$ and $\|A\|_{r \to p}$ correspond to key quantities of interest arising in different fields. Specifically, $M_2(A)$ represents the largest eigenvalue of the symmetric matrix $(A + A^\top)/2$, where $A^\top$ denotes the transpose of A. In the context of clustering, the ℓ_r-Grothendieck problem for $r = 2$ and $r = \infty$ is related to spectral partitioning [69, 86] and correlation clustering [45], respectively. Moreover, for general $r \in (2, \infty)$, the ℓ_r-Grothendieck problem can be viewed as a smooth interpolation between these two clustering measures. When A is diagonal free, and has off-diagonal entries that are independent and identically distributed (i.i.d.) standard Gaussian random variables, $M_r(A)$ is equivalent to the ground state energy of Sherrinton-Kirkpatrick spin glass model with spins coming from the ℓ_r-sphere and couplings given by the entries of A (see the discussion in [47] after (1.2) and also the discussion in Section 2.2 before Corollary 2.2.9).

Turning to the $r \to p$ operator norm, $\|A\|_{2\to 2}$ is equivalent to the largest singular value of A. In the scenario where $r < p$ and A represents the normalized adjacency matrix of a graph (or, more broadly, of a Markov operator), upper bounds on $\|A\|_{r\to p}$ are recognized as hypercontractive inequalities, which are instrumental in demonstrating rapid mixing properties of associated random walks (see [98, Section 2.2]). Additionally, the computation of $\|A\|_{2\to 4}$ is closely related to the small-set expansion and injective tensor norm, both of which are important in theoretical computer science [18, 42]. Other application domains where the $\|A\|_{r\to p}$ norm arises include the development of oblivious routing schemes in transportation networks for the ℓ_p norm [28, 74, 103, 144], and data dimensionality sketching of $r \to p$ norms, with applications to data stream model and robust regression [123]. The special case $\|A\|_{p\to p}$ also appears in the estimation of matrix condition numbers [105].

It is worth pointing out that the two optimization problems are connected; the quantity $M_r(A)$ of OPT 1 is closely related to $\|A\|_{r\to p}$ in OPT 2 in the special case $p = r^* \coloneqq r/(r-1)$. Indeed, given any $\boldsymbol{x} \in \mathbb{R}^n$ with $\|\boldsymbol{x}\|_r \le 1$, using the well-known duality representation $\|A\boldsymbol{x}\|_{r^*} = \sup_{y\in\mathbb{R}^n:\|\boldsymbol{y}\|_r\le 1} \boldsymbol{y}^\top A\boldsymbol{x}$ (see for example [88, Proposition 6.13]) it follows that

$$\|A\|_{r\to r^*} = \sup_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n:\|\boldsymbol{x}\|_r\le 1,\|\boldsymbol{y}\|_r\le 1} \boldsymbol{y}^\top A\boldsymbol{x}. \tag{2.1.1}$$

Thus, comparing OPT 1 with (2.1.1), it follows that $\|A\|_{r\to r^*} \geq M_r(A)$. In certain specific instances, such as when $r \geq 2$, A has nonnegative entries, and $A^\top A$ is irreducible, $M_r(A)$ is equal to $\|A\|_{r\to r^*}$ (see [67, Proposition 2.14]). However, in general, these two quantities need not coincide.

The computational complexity of both optimization problems, $M_r(A)$ and $\|A\|_{r\to p}$, have attracted immense attention in the theoretical computer science community. For $r \in (2,\infty)$, there exists a polynomial time algorithm to approximate $M_r(A)$ within a factor of ξ_r^2, where $\xi_r := \mathbb{E}\left[|z|^r\right]^{1/r}$ denotes the ℓ_r-norm of the standard Gaussian random variable [114]. However, no polynomial-time algorithm can approximate $M_r(A)$ within a factor strictly less than ξ_r^2 unless P=NP [114]. For the $r \to p$ operator norm, when A is a nonnegative matrix and $r \geq p$, efficient algorithms, such as the nonlinear power method by Boyd [41], can compute $\|A\|_{r\to p}$ in polynomial time. However, for general A, when $r \geq 2 \geq p \geq 1$, it is known that $\|A\|_{r\to p}$ is NP-hard to approximate within any factor smaller than $1/(\xi_p \xi_{r/(r-1)})$ [29]. And when either $r \geq p$ and $2 \notin [p,r]$, or $r < p$, approximating $\|A\|_{r\to p}$ within any constant factor is NP-hard [29]. Given the computational challenges associated with $M_r(A)$ and $\|A\|_{r\to p}$ for generic matrices A, a natural question arises: is the approximation of these quantities also hard for typical matrices, such as random matrices with i.i.d. entries, modulo symmetry consideration?

2.1.2 Discussion of Related Prior Work

While the majority of classical work has focused on spectral norms, early work on operator norms includes that of Bennett [27], which obtains an upper bound for the expected $2 \to p$ norm for matrices with independent, mean-zero entries lying in $[-1,1]$ when $p > 2$, and [26], which extrapolated the latter results to obtain bounds on $r \to p$ norms in the full range $1 \leq r, p \leq \infty$. Recent years have witnessed a resurgence in interest in operator norms of random matrices. On the one hand, one line of inquiry focuses on non-asymptotic bounds. Latała and Strzelecka [124, 125] established matching two-sided nonasymptotic bounds for the expected $r \to p$ norm, with $r, p \in [1, \infty]$, for matrices with i.i.d. entries drawn from a large class of light-tailed distribution. On the other hand, for matrices with independent (but not necessarily identical) mean-zero Gaussian entries, two-sided bounds were derived in [1] that match up to a logarithmic factor depending on the dimension of matrix. It is conjectured in the same paper that this logarithmic factor could be removed to yield a dimension-free estimation of $r \to p$ operator norm of inhomogeneous Gaussian random matrices.

On the other hand, another line of recent work focuses on high-dimensional asymptotic limits of $\|A\|_{r\to p}$ and $M_r(A)$. For $r \geq p$ and symmetric, nonnegative random matrices A with i.i.d. light-tailed upper triangular entries, both the almost sure limit and the fluctuations of suitably normalized $\|A\|_{r\to p}$ were analyzed in [67]. As mentioned earlier, for this class of matrices, namely matrices with nonnegative entries, efficient algorithms, such as the (nonlinear) power method by Boyd [41], are available to compute $\|A\|_{r\to p}$ when $r \geq p$. When A is an $n \times n$ matrix composed of n^2 i.i.d. standard Gaussian random variables, Chen and Sen [47] use Chevet's inequality to examine the almost sure limit of an appropriately scaled version of $M_r(A)$. They discovered that the scaling behavior of $M_r(A)$ undergoes a phase transition at $r = 2$: for $r \in (1, 2)$, it scales as $n^{(r-1)/r}$, and for $r \in [2, \infty]$, it scales as $n^{3/2-2/r}$. Moreover, they also characterized the limit of $M_r(A)$ for all $r \geq 1$, showing in the sub-regime $r > 2$ that it can be characterized in terms of a variant of the Parisi formula for the free energy of the Sherrington-Kirkpatrick spin glass model. It is worth mentioning that at the critical value $r = 2$, $M_2(A)$ is equal to the largest eigenvalue of $\frac{A+A^\top}{2}$, and the fluctuations of $M_2(A)$ is well known to be governed by the celebrated Tracy-Widom distribution [161]. It remains an interesting open problem to study the asymptotic fluctuations of $M_r(A)$ in the Gaussian setting when $r \neq 2$.

In contrast to their light-tailed counterparts, random matrices with heavy tails exhibit very different properties. For instance, the limiting empirical spectral distribution and distribution of the largest eigenvalue (as the dimension of matrix tends to infinity) of symmetric light-tailed random matrices, such as those in the Gaussian Orthogonal Ensemble, are known to follow the Wigner semicircle distribution and Tracy-Widom distributions, respectively (see the survey paper [85] for more details). However, in the heavy-tailed regime, the limiting empirical spectral distribution of symmetric random matrix $A = (a_{ij})_{i,j\in[n]}$ is supported on the whole real line when a_{ij} is heavy-tailed with index $\alpha \in (0, 2)$ (see Definition 2.2.3) [14] and the largest eigenvalue of A follows the Fréchet distribution when a_{ij} is heavy-tailed with index $\alpha \in (0, 4)$ [156, 12].

2.1.3 Summary of Results

Both the non-asymptotic and asymptotic results for general operator norms mentioned in Section 2.1.2 have been concerned with random matrices with light-tailed entries. Our work takes a first step towards understanding the asymptotics of non-spectral norms in the heavy-tailed setting. We consider both the asymptotic almost sure limit and fluctuations of $M_r(A_n)$ and $\|A_n\|_{r\to p}$, as $n \to \infty$, in the case when each A_n is a symmetric $n \times n$ random matrix with independent upper-triangular entries with a common *heavy-tailed* distribution. Specifically, let $A_n = (a_{ij})_{i,j\in[n]}$ be a symmetric $n \times n$ matrix consisting of i.i.d. heavy-tailed entries with a tail index $\alpha \in (0, 2)$; see Definition 2.2.3. Our main results can be broadly categorized into two distinct regimes:

1. Suppose $1 \leq r \leq 2$ for $M_r(A_n)$ and $1 \leq r \leq p$ for $\|A_n\|_{r\to p}$, and $\alpha \in (0, 2)$. Then we show that the quantities $M_r(A_n)$ and $\|A_n\|_{r\to p}$ are both solely determined by the entries with the largest absolute value among all entries of A_n (there are two largest entries due to symmetry), and suitably scaled versions of both quantities converge, as $n \to \infty$, to a Frechét distribution (see Theorem 2.2.5 and Theorem 2.2.10).

2. Suppose $2 < r \leq \infty$ for $M_r(A_n)$ and $1 \leq p < r \leq \infty$ for $\|A_n\|_{r\to p}$, and α lies in $(0, \alpha_*)$ for a certain explicit constant $\alpha_* = \alpha_*(r)$ or $\alpha_* = \alpha_*(r, p)$ lying in $(1, 2)$. Then the high-dimensional asymptotics of $M_r(A)n)$ and $\|A_n\|_{r\to p}$ are dictated by the largest entries of A_n amongst all those whose absolute values exceed a certain (n-dependent)

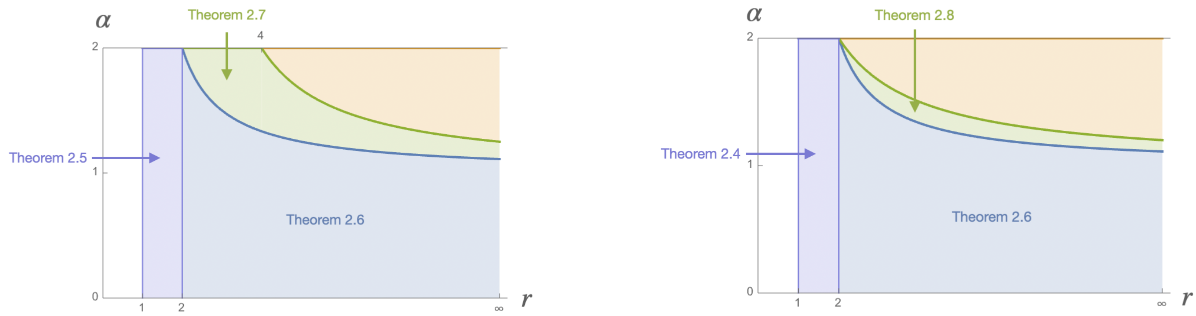
threshold, and both quantities converge to corresponding stable distributions after suitable scaling (see Theorem 2.2.6 and Theorem 2.2.11).

Furthermore, we also establish the following results to show that in a sense the choice of α_* is optimal:

3. Consider the case where either $2 < r < \infty$ for $M_r(A_n)$ or $1 \leq r < p \leq \infty$ for $\|A_n\|_{r\to p}$, with the corresponding constant $\alpha_* \in (1,2)$ as in Case 2. Then we show that there exist (explicit) constants $\bar{\alpha}_{*,1}$ and $\bar{\alpha}_{*,2}$, depending on r or both r and p (as relevant), that both lie in the interval $(\alpha_*, 2)$ and are such that the following properties are satisfied:
   - If the entries of A have zero mean and $\alpha \in (\alpha_*, \bar{\alpha}_{*,1})$, then the asymptotic fluctuations of suitably scaled versions of $M_r(A)$ and $\|A\|_{r\to p}$ converge to corresponding stable distributions. The scaling is the same as in Case 2 (see Theorem 2.2.7 and Theorem 2.2.12).
   - If the entries of A have positive mean for $M_r(A)$, or nonzero mean for $\|A\|_{r\to p}$, and $\alpha \in (\alpha_*, \bar{\alpha}_{*,2})$, then the asymptotic fluctuations of suitably scaled and centered versions of $M_r(A)$ and $\|A\|_{r\to p}$ again converge to the stable distributions. The limiting stable distributions are the same as in Case 2; however, the scaling is different and an additional centering of $M_r(A)$ and $\|A\|_{r\to p}$ isalso necessary (see Theorem 2.2.8 and Theorem 2.2.13). The additional centering becomes necessary because in the regime $\alpha > \alpha_*$, the fluctuations contributed by the largest entries are asymptotically negligible compared to the effect of shift in centering of individual entries.

Also, as an immediate consequence of our results we characterize the ground state of the Levy spin glass model with from statistical physics when $\alpha \in (0,1)$ (see Corollary 2.2.9).

Our results on the ℓ_r-Levy-Grothendieck problem in the heavy-tailed setting are summarized in Figure 2.1(a) and (b) for the case of centered and non-centered entries, respectively. An interesting open problem is to identify the asymptotic fluctuations in the orange regime in these two figures. This corresponds to the less heavy-tailed regime (depending on r), where we expect the fluctuations to be determined by the small entries instead of the largest ones. The influence of the small entries on fluctuation already starts to manifest itself in the third set of results described above, which can be seen from the difference in scaling factors. The challenge we met in the orange regime is also consistent with the fact (mentioned in Section 2.1.2) that even for centered Gaussian i.i.d. entries, the asymptotic fluctuations of $M_r(A_n)$ are not well understood; only the almost sure limit has been characterized. It is also worth mentioning that, the change in scalings for different α regimes and the difference for centered and non-centered entries are not observed in the spectral case $r = 2$, where, as mentioned at the end of Section 2.1.2, properly scaled $M_2(A_n)$ converges to Fréchet distribution as $n \to \infty$.



(a) Entries with zero mean

(b) Entries with nonzero mean

Figure 2.1: Centered and noncentered cases

2.1.4 Common Notation

Let $[n]$ denote the set $\{1, \ldots, n\}$. For a sequence of random variables $\{Y_n\}_{n \in \mathbb{N}}$ and a random variable Y, we denote the convergence in distribution of Y_n to Y by $Y_n \xrightarrow{d} Y$. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the events $E_n \in \mathcal{F}, n \in \mathbb{N}$ are said to occur *with high probability* (whp) if $\mathbb{P}(E_n) \to 1$ as $n \to \infty$. For a sequence of random variables $(X_n)_{n \in \mathbb{N}}$ and positive deterministic constants $(k_n)_{n \in \mathbb{N}}$:

1. We say $X_n = O(k_n)$ whp if there exists a constant $C > 0$ such that the events

$\{|X_n| \le Ck_n\}_{n\in\mathbb{N}}$ occur whp. We say $X_n = O_\delta(k_n)$ whp if for any $\delta > 0$ the events $\{|X_n| \le n^\delta k_n\}_{n\in\mathbb{N}}$ occur whp.

2. We say $X_n = o(k_n)$ whp if $k_n^{-1}X_n$ converges to 0 in probability.

3. We say $X_n = \Theta_\delta(k_n)$ whp if for any $\delta > 0$ the events $\{n^{-\delta}k_n \le |X_n| \le n^\delta k_n\}_{n\in\mathbb{N}}$ occur whp.

For a sequence of deterministic quantities $\{D_n\}_{n\in\mathbb{N}}$ and positive deterministic constants $(k_n)_{n\in\mathbb{N}}$:

1. We say $D_n = O(k_n)$ if there exists a constant $C > 0$ such that $\limsup_{n\to\infty} k_n^{-1}|D_n| \le C$. We say $D_n = O_\delta(k_n)$ if for any $\delta > 0$, $\lim_{n\to\infty} n^{-\delta}k_n^{-1}D_n = 0$.

2. We say $D_n = o(k_n)$ if $\lim_{n\to\infty} k_n^{-1}D_n = 0$.

3. We say $D_n = \Theta_\delta(k_n)$ if for any $\delta > 0$, $\lim_{n\to\infty} n^{-\delta}k_n^{-1}D_n = 0$ and $\lim_{n\to\infty} n^\delta k_n^{-1}|D_n| = \infty$.

Given an $n \times n$ matrix A, we denote its transpose by $A^\top$ and the matrix obtained by taking the absolute value of each entry in A by $|A|$. Given a vector $\boldsymbol{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$ with positive entries and for $p \in \mathbb{R}$, we define $\boldsymbol{v}^p := (v_1^p, \ldots, v_n^p)$. Given nonnegative functions f, g defined on $[0, \infty)$, we say $f \ll g$ (respectively, $f \gg g$) if $f(x)/g(x) \to 0$ (respectively, $g(x)/f(x) \to 0$) as $x \to \infty$. The p-norm of a vector $\boldsymbol{x} = (x_i)_{i\in[n]} \in \mathbb{R}^n$ is defined as

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

For $p \in (1, \infty)$, we denote its Hölder conjugate by $p^* := p/(p-1)$. Additionally, we write $1^* = \infty$ and $\infty^* = 1$. We use $\mathrm{sgn} : \mathbb{R} \to \mathbb{R}$ to denote the sign function, where $\mathrm{sgn}(t) = -1, 1$ and 0 for $t < 0, t > 0$ and $t = 0$, respectively. For $q > 1$, $t \in \mathbb{R}$ and

$\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we define $\Psi_q : \mathbb{R}^n \to \mathbb{R}^n$ as follows: for $i \in [n]$,

$$(\Psi_q(\boldsymbol{x}))_i := \psi_q(x_i), \quad \text{where} \quad \psi_q(t) := |t|^{q-1} \operatorname{sgn}(t). \tag{2.1.2}$$

We adopt the convention $0^0 = 0$.

2.2 Main results

In Section 2.2.1, we state our basic assumption on A_n after introducing some basic definitions related to heavy-tailed distributions. In Section 2.2.2 and Section 2.2.3, we present our main results concerning OPT 1 and OPT 2 respectively. A rough outline of the proofs of the main results is given in Section 2.2.4, and Section 2.2.5 summarizes the rest of the paper.

2.2.1 Assumption on the random matrix

To state our main assumption, we first introduce some definitions.

Definition 2.2.1 (Slowly Varying Function). *A function $L : (0, \infty) \to (0, \infty)$ is said to be a* slowly varying function *if it satisfies*

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1, \quad \forall\, t > 0.$$

The next lemma (see [139, Lemma 2.7]) states that slowly varying functions grow (or decay) asymptotically slower than any polynomial.

Lemma 2.2.2. *If the function $L : (0, \infty) \to (0, \infty)$ is slowly varying, then*

$$x^{-\delta} \ll L(x) \ll x^{\delta}, \quad \forall\, \delta > 0. \tag{2.2.1}$$

Definition 2.2.3 (Heavy-tailed random variable). *A random variable X is said to be* heavy-tailed *with index* $\alpha > 0$ *if it satisfies*

$$\bar{F}(x) \equiv \bar{F}_\alpha(x) := \mathbb{P}(|X| > x) = x^{-\alpha}L(x), \quad x > 0, \tag{2.2.2}$$

where L *is a slowly varying function.*

Definition 2.2.4 (Stable random variable)**.** *For* $\alpha \in (0,2)$*, a nondegenerate random variable* X *is said to be* α*-stable if* $X = aZ + b$*, where* $a \neq 0$*,* $b \in \mathbb{R}$*, and* Z *is a random variable whose characteristic function takes the following form, for some* $\beta \in [-1,1]$*:*

$$\mathbb{E}\left[\exp(itZ)\right] = \begin{cases} \exp\left(-|t|^\alpha \left[1 - i\beta \tan\frac{\pi\alpha}{2}(\operatorname{sgn} t)\right]\right), & \alpha \neq 1, \\ \exp\left(-|t| \left[1 + i\beta\frac{2}{\pi}(\operatorname{sgn} t)\log|t|\right]\right), & \alpha = 1. \end{cases}$$

We consider the following class of random matrices.

Assumption 1. $A = (a_{ij})_{i,j\in[n]}$ is a symmetric random matrix, whose upper-triangular entries $(a_{ij})_{1\le i\le j\le n}$ are i.i.d. copies of heavy-tailed random variables with index α.

Fix $\alpha \in (0,2)$. Given the complementary cumulative distribution function $\bar{F}_\alpha$ of $|X|$ in (2.2.2), define

$$b_n \equiv b_{n,\alpha} := \inf\left\{x \ge 0 : \bar{F}_\alpha(x) \le \frac{2}{n(n+1)}\right\}, \quad n \in \mathbb{N}. \tag{2.2.3}$$

It can be shown that (see [70, (3.8.6)])

$$\bar{F}_\alpha(b_{n,\alpha}) \cdot \frac{n(n+1)}{2} \to 1, \text{ as } n \to \infty,$$

which implies that there exists another slowly varying function L_o such that

$$b_{n,\alpha} = L_o(n)n^{2/\alpha}, \tag{2.2.4}$$

For notational simplicity, we often drop the α subscript whenever it is clear from the context.

2.2.2 The Levy-Grothendieck ℓ_r problem and the ground state of Levy spin glasses

Our first set of main results concerns the ℓ_r-Grothendieck problem (OPT 1) for the class of random matrices satisfying Assumption 1. The proof of Theorem 2.2.5, Theorem 2.2.6, Theorem 2.2.8 and Corollary 2.2.9 are deferred to Section 2.6.1, Section 2.6.2, Section 2.6.3 and Section 2.6.4, respectively.

Theorem 2.2.5. *Fix $1 \leq r \leq 2$. Suppose $A_n = (a_{ij})_{i,j\in[n]}$ satisfies Assumption 1 with $\alpha \in (0,2)$. Then with b_n as defined in (2.2.3), we have*

$$\lim_{n\to\infty} \mathbb{P}\left(b_n^{-1} 2^{\frac{2}{r}-1} M_r(A_n) \leq x\right) = \exp\left(-x^{-\alpha}\right).$$

The proof of Theorem 2.2.5 is given in Section 2.6.1. As we show there, in the case $1 \leq r \leq 2$, $M_r(A_n)$ is completely determined by the largest (in absolute value) entry of A_n. Thus the limiting distribution of the properly normalized $M_r(A_n)$ resembles that of the limiting distribution of the largest entry, which is a Fréchet distribution with index α.

Theorem 2.2.6. *Fix $2 < r \leq \infty$ and let $\alpha_* \equiv \alpha_*(r) := \frac{r}{r-1}$. Suppose $A_n = (a_{ij})_{i,j\in[n]}$ satisfies Assumption 1 with index α, such that*

$$0 < \alpha < \alpha_*(r). \tag{2.2.5}$$

Then with b_n as defined in (2.2.3), we have

$$b_n^{-1} M_r(A_n) \xrightarrow{d} \left(Y_{\frac{\alpha(r-2)}{r}}\right)^{\frac{r-2}{r}},$$

where $Y_{\frac{\alpha(r-2)}{r}}$ is an $\frac{\alpha(r-2)}{r}$-stable random variable.

The upper bound $\alpha_*(r) = \frac{r}{r-1}$ is tight for the result to hold in such generality. Beyond α_*, things become more subtle, with scaling and asymptotics depending on whether or not the entries are centered, as demonstrated in the following two results.

Theorem 2.2.7. *Fix $2 < r \leq \infty$, $\alpha_* \equiv \alpha_*(r) := \frac{r}{r-1}$ and $\bar{\alpha}_{*,1} \equiv \bar{\alpha}_{*,1}(r) := \min\{2, \frac{r}{r-2}\}$. Suppose $A_n = (a_{ij})_{i,j\in[n]}$ satisfies Assumption 1 with index α, such that*

$$(a_{ij})_{i,j\in[n]} \text{ are centered and } \quad \alpha_*(r) \leq \alpha < \bar{\alpha}_{*,1}(r). \tag{2.2.6}$$

Then with b_n as defined in (2.2.3), we have

$$b_n^{-1} M_r(A_n) \xrightarrow{d} \left(Y_{\frac{\alpha(r-2)}{r}}\right)^{\frac{r-2}{r}},$$

where $Y_{\frac{\alpha(r-2)}{r}}$ is an $\frac{\alpha(r-2)}{r}$-stable random variable.

The proof of Theorem 2.2.6 and Theorem 2.2.7 are deferred to Section 2.6.2. Both theorems share the same conclusions, but their conditions differ, which can be understood as follows. When α is small, that is, when (2.2.5) holds, the entries have "extreme" heavy tails, making the impact of the mean (or median for $\alpha \in (0, 1]$) of the distribution, whether zero or not, asymptotically negligible when compared to the the largest, sparse entries. On the other hand, when $\alpha \geq \alpha_*$ with α_* in Theorem 2.2.6, the assumption of zero mean is necessary to mitigate the influence of smaller entries, thereby exposing the sparse structure of largest entries. When $\alpha > \bar{\alpha}_{*,1}$ with $\bar{\alpha}_{*,1}$ in Theorem 2.2.7, we expect the fluctuations of $M_r(A_n)$ to be dominated by the small, dense matrix terms. Consequently, the approach of leveraging the sparsity of large terms in heavy-tailed entries, as employed in this work, becomes ineffective.

Theorem 2.2.8. *Fix $2 < r < \infty$, $\mu > 0$, $\alpha_* \equiv \alpha_*(r) := \frac{r}{r-1}$ and $\bar{\alpha}_{*,2} \equiv \alpha_{*,2}(r) := \frac{r+2}{r}$.*

Suppose $A_n = (a_{ij})_{i,j\in[n]}$ satisfies Assumption 1 with index α, such that

$$(a_{ij})_{i,j\in[n]} \text{ are centered and } \quad \alpha_*(r) < \alpha < \bar{\alpha}_{*,2}(r). \tag{2.2.7}$$

Define

$$A_{\mu,n} \coloneqq \mu\mathbf{1}\mathbf{1}^\top + A_n = (\mu + a_{ij})_{i,j\in[n]}.$$

Then with b_n as defined in (2.2.3), we have

$$b_n^{-\frac{r}{r-2}} n^{\frac{r}{r-2}-\frac{r-2}{r}} \left(M_r(A_{\mu,n}) - n^{2-\frac{2}{r}}\mu\right) \xrightarrow{d} Y_{\frac{\alpha(r-2)}{r}},$$

where $Y_{\frac{\alpha(r-2)}{r}}$ is an $\frac{\alpha(r-2)}{r}$-stable random variable.

Theorem 2.2.8 and its corollary below are proved in Section 2.6.3 and Section 2.6.4, respectively. Comparing with Theorem 2.2.6 and Theorem 2.2.7, $M_r(A_{\mu,n})$ requires a centering and a different scaling. The appearance of the centering can be seen from the definition of $M_r(A_{\mu,n})$,

$$M_r(A_{\mu,n}) = \max_{\boldsymbol{x}\in\mathbb{R}^n:\|\boldsymbol{x}\|_r\le 1} \left[\mu\left(\mathbf{1}^\top\boldsymbol{x}\right)^2 + \boldsymbol{x}^\top A_n \boldsymbol{x}\right]. \tag{2.2.8}$$

Indeed, for $\alpha \in (\alpha_*, \bar{\alpha}_{*,2})$, $\max_{\|\boldsymbol{x}\|_r\le 1} \boldsymbol{x}^\top A_n\boldsymbol{x} = O_\delta(b_n)$ whp by Theorem 2.2.7 and $\max_{\|\boldsymbol{x}\|_r\le 1} \mu(\mathbf{1}^\top\boldsymbol{x})^2 = n^{2-\frac{2}{r}}\mu$ by Hölder's inequality. When $\alpha > \alpha_*$ and $\mu > 0$, the condition $b_n \ll n^{2-\frac{2}{r}}$ implies that the dominant term in $M_r(A_n)$ is the first term in (2.2.8), necessitating centering. Conversely, when $\alpha < \alpha_*$ as in the setting of Theorem 2.2.6 or $\mu = 0$ as in the setting of Theorem 2.2.7, the second term in (2.2.8), which governs the fluctuation, outweighs the first term, rendering centering unnecessary. The scaling in Theorem 2.2.8 is more subtle and not completely obvious. Unlike the light-tailed case discussed in [67], where the scaling factor is $n^{-\frac{r-2}{r}}\sigma^{-1}$ with σ being the variance of the matrix entries, the variance is not well-defined in the heavy-tailed setting. Instead, an additional factor $(b_n^{-1}n)^{\frac{r}{r-2}}$ appears, reflecting the different scaling behavior in the presence of heavy tails.

As a corollary of Theorem 2.2.6, we obtain asymptotics of the ground state energy of Levy spin glass model, which is a result of independent interest. Spin glasses are a type of disordered magnetic system characterized by the presence of both ferromagnetic and antiferromagnetic interactions between the spins of the atoms in the material. Mathematically, a spin glass can be described as a random system where the state space consists of n spins taking values from ± 1. The probability distribution of the state space is given by the Gibbs measure, which is defined as:

$$P(\boldsymbol{\sigma}) \propto \exp\left(-\beta \sum_{i,j\in[n]} h_{ij}\sigma_i\sigma_j\right), \quad \forall\, \boldsymbol{\sigma} \in \{+1,-1\}^n,$$

where $\beta > 0$ is the inverse temperature, $(h_{ij})_{i,j\in[n]}$ is the interaction matrix and $H(\boldsymbol{\sigma}) = -\beta \sum_{i,j\in[n]} h_{ij}\sigma_i\sigma_j$ is the Hamiltonian of the system. The Levy spin glass model, which features interaction matrix with independent heavy-tailed entries, was introduced in [57] to understand surprising experimental results on dilute spin glasses with dipolar interactions. The ground states of the spin glass refer to the limiting states as $\beta \to \infty$, which are the minimizers of the corresponding Hamiltonian, and the associated energy $\sup_{\sigma\in\{\pm 1\}^n} \sum_{i,j\in[n]} h_{ij}\sigma_i\sigma_j$ is called the ground state energy.

Corollary 2.2.9. *Suppose $A_n = (a_{ij})_{i,j\in[n]}$ satisfies Assumption 1 with index $\alpha \in (0,1)$. Then with b_n as defined in (2.2.3), we have*

$$b_n^{-1} \sup_{\boldsymbol{x}\in\{-1,1\}^n} \boldsymbol{x}^\top A_n \boldsymbol{x} \xrightarrow{d} Y_\alpha,$$

where Y_α is an α-stable random variable.

2.2.3 The $r \to p$ operator norm

We next turn to our results on asymptotics of the $r \to p$ operator norm problem (OPT 2) for heavy-tailed random matrices. At this juncture, we recall the dual formulation of the $r \to p$ operator norm to emphasize its connection with OPT 1 and for subsequent reference. Given $\boldsymbol{x} \in \mathbb{R}^n$, using the well-known dual representation $\|A\boldsymbol{x}\|_p = \sup_{\boldsymbol{y}\in\mathbb{R}^n:\|\boldsymbol{y}\|_{p^*}\leq 1} \boldsymbol{x}^\top A\boldsymbol{y}$ (see [88, Proposition 6.13]), it follows that

$$\|A_n\|_{r\to p} := \sup_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n:\|\boldsymbol{x}\|_r\leq 1,\|\boldsymbol{y}\|_{p^*}\leq 1} \boldsymbol{y}^\top A_n\boldsymbol{x}. \tag{2.2.9}$$

Our results for the $r \to p$ operator norm parallel those for the Levy-Grothendieck ℓ_r problem.

Theorem 2.2.10. *Fix $1 \leq r \leq p \leq \infty$. Suppose $A_n = (a_{ij})_{i,j\in[n]}$ satisfy Assumption 1 with index $\alpha \in (0,2)$. Then with b_n as defined in (2.2.3), we have*

$$\lim_{n\to\infty} \mathbb{P}\left(b_n^{-1}\|A_n\|_{r\to p} \leq x\right) = \exp\left(-x^{-\alpha}\right).$$

Theorem 2.2.10, whose proof is postponed to Section 2.3, is an analogue of Theorem 2.2.5. The additional factor $2^{\frac{2}{r}-1}$ in Theorem 2.2.5 comes from the facts that the largest entries of A_n are unlikely to reside on the diagonal (see Lemma 2.3.3) and that when $r \leq 2$, $M_r(A)$ is fully determined by the largest entries of A (recall that we have two largest entries due to symmetry). Hence, we cannot localize all the mass in one entry in the optimizer of OPT 1. In contrast, in the dual formulation (2.1.1) of OPT 2, the optimal vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ can be chosen separately, allowing mass to be concentrated in single, yet distinct entries in $\boldsymbol{x}$ and $\boldsymbol{y}$.

Theorem 2.2.11 and Theorem 2.2.12 below, which parallel Theorem 2.2.6 and Theorem 2.2.7 respectively, are proved in Section 2.4.

Theorem 2.2.11. *Fix $1 \leq p < r \leq \infty$, denote $\gamma := \frac{1}{p} - \frac{1}{r}$ and define $\alpha_* \equiv \alpha_*(\gamma) := \frac{2}{1+\gamma}$.*

Suppose $A_n = (a_{ij})_{i,j\in[n]}$ *satisfy Assumption 1 with index* α, *such that*

$$0 < \alpha < \alpha_*(\gamma). \tag{2.2.10}$$

Then with b_n *defined in* (2.2.3), *we have*

$$b_n^{-1}\|A_n\|_{r\to p} \xrightarrow{d} (Y_{\alpha\gamma})^{\gamma},$$

where $Y_{\alpha\gamma}$ *is an* $\alpha\gamma$*-stable random variable.*

Theorem 2.2.12. *Fix* $1 \le p < r \le \infty$, *denote* $\gamma := \frac{1}{p} - \frac{1}{r}$, *define* $\alpha_* \equiv \alpha_*(\gamma) := \frac{2}{1+\gamma}$ *and* $\bar{\alpha}_{*,1} \equiv \bar{\alpha}_{*,1}(\gamma) := \min\{2, \frac{1}{\gamma}\}$. *Suppose* $A_n = (a_{ij})_{i,j\in[n]}$ *satisfy Assumption 1 with* α, *such that*

$$(a_{ij})_{i,j\in[n]} \text{ are centered and } \quad \alpha_*(\gamma) \le \alpha < \bar{\alpha}_{*,1}(\gamma). \tag{2.2.11}$$

Then with b_n *defined in* (2.2.3), *we have*

$$b_n^{-1}\|A_n\|_{r\to p} \xrightarrow{d} (Y_{\alpha\gamma})^{\gamma},$$

where $Y_{\alpha\gamma}$ *is an* $\alpha\gamma$*-stable random variable.*

Theorem 2.2.13. *Fix* $1 < p < r < \infty$, $\mu \ne 0$ *and denote* $\gamma := \frac{1}{p} - \frac{1}{r}$ *and* $\gamma' := \frac{1}{\min\{2,p\}} - \frac{1}{\max\{2,r\}}$. *Define* $\alpha_* \equiv \alpha_*(\gamma) := \frac{2}{1+\gamma}$ *and* $\bar{\alpha}_{*,2} \equiv \bar{\alpha}_{*,2}(\gamma,\gamma') := \frac{2-\gamma}{\gamma(\gamma'-\gamma)+1}$. *Suppose* $A_n = (a_{ij})_{i,j\in[n]}$ *satisfy Assumption 1 with* α, *such that*

$$\alpha_*(\gamma) < \alpha < \bar{\alpha}_{*,2}(\gamma,\gamma'). \tag{2.2.12}$$

Define

$$A_{\mu,n} := \mu\mathbf{1}\mathbf{1}^{\top} + A_n = (\mu + a_{ij})_{i,j\in[n]}.$$

Then with b_n as defined in (2.2.3),

$$b_n^{-\frac{1}{\gamma}} n^{\frac{1}{\gamma}-\gamma} \left(\|A_{\mu,n}\|_{r\to p} - n^{1+\gamma}|\mu| \right) \to Y_{\alpha\gamma},$$

where $Y_{\alpha\gamma}$ is an $\alpha\gamma$-stable random variable.

Theorem 2.2.13, whose proof is postponed to Section 2.5, is the analogue of Theorem 2.2.8.

2.2.4 Rough proof outline

We now briefly discuss the main strategy used in this work. To deal with the heavy-tailed natures of entries, we decompose A based on the magnitude of the entries, in the following manner:

$$A = \mu \mathbf{1}\mathbf{1}^T + \tilde{A} + \hat{A} + \bar{A},$$

where $\mu\mathbf{1}\mathbf{1}^T$ represents the centering of A. The matrices $\hat{A}$ collects the largest entries in A which can be shown to be 'sparse' throughout the matrix, $\tilde{A}$ collects entries of moderate magnitude in A which can be shown to be 'sparse' within each row and column and the remaining matrix consisting of relatively small and centered entries.

For the proof of Theorem 2.2.11, in the first step we exploit the cancellation coming from $\bar{A}$, utilizing concentration of measure and random matrix theories as main tools. The next step exploits the sparsity and heavy-tailed nature of the matrix $\tilde{A}$ to argue that the impact of $\tilde{A}$ is negligible compared to $\hat{A}$. Finally, we show that the 'extreme' sparsity of $\hat{A}$ simplifies its structure, making it amenable to an explicit optimization.

For Theorem 2.2.13, the estimate on $\|\tilde{A}\|$ used in Theorem 2.2.11 proves inadequate. We instead employ a 'sandwich' approach, where we construct an upper bound and a lower bound that match at the fluctuation level.

2.2.5 Organization of the rest of the paper

In view of (2.1.1), the $r \to r^*$ operator norm provides a natural upper bound for $M_r(A)$. For this reason, we first present the proofs of the $r \to p$ operator norm, and then complete proofs of the ℓ_r-Grothendieck problem. The proof of Theorem 2.2.10, Theorem 2.2.11 and Theorem 2.2.13 are presented in Section 2.3, Section 2.4 and Section 2.5, respectively. Then in Section 2.6 we provide the proof of results relating to the Levy-Grothendieck ℓ_r problem.

2.3 Proof of Theorem 2.2.10

The proof of Theorem 2.2.10 proceeds by showing that the distribution of $\|A_n\|_{r\to p}$ coincides with the distribution of the largest entry of A_n. For brevity, in this and subsequent sections, we suppress the subscript n of A_n.

We start with the limiting distribution of the largest matrix entries (after proper normalization), which was identified in [156, (24)].

Lemma 2.3.1. *Fix $\alpha \in (0,2)$. Suppose $A = (a_{ij})_{i,j\in[n]}$ satisfies Assumption 1, and let $(b_n)_{n\in\mathbb{N}}$ be as in (2.2.3). Then the following holds:*

$$\lim_{n\to\infty} \mathbb{P}\left(b_n^{-1} \max_{i,j\in[n]} |a_{ij}| \le x\right) = \exp\left(-x^{-\alpha}\right). \tag{2.3.1}$$

Next, we identify the condition that should be satisfied by the maximizing vector of OPT 2 when $1 < r, p < \infty$. The following lemma is proved in [67, (4.4)].

Lemma 2.3.2. *Fix $1 < r, p < \infty$. Let $\boldsymbol{v} \equiv \boldsymbol{v}_n = (v_i)_{i\in[n]}$ be a maximizing vector of OPT 2 with $\|\boldsymbol{v}\|_r = 1$. Then the following relation holds:*

$$A^\top \Psi_p(A\boldsymbol{v}) = \|A\|_{r\to p}^p \Psi_r(\boldsymbol{v}).$$

We will make repeated use of the following result (see [156, Lemma 4]), which describes the typical structure of heavy-tailed random matrix A.

Lemma 2.3.3. *Suppose $A_n = (a_{ij})_{i,j\in[n]}$ satisfies Assumption 1 with $\alpha \in (0,2)$. Then the following hold:*

(a) whp, $\left\{i \in [n] :\ a_{ii} \geq b_n^{11/20}\right\} = \varnothing$;

(b) $\forall\, \delta > 0$, whp, $\forall\, i \in [n]$, $\#\left\{j \in [n] :\ |a_{ij}| \geq b_n^{3/4+\delta}\right\} \leq 1$;

(c) $\exists\, \delta(\alpha) > 0$, such that, whp, $\forall\, i \in [n]$, $\sum_{j\in[n]} |a_{ij}|\, \mathbb{1}\left\{|a_{ij}| \leq b_n^{3/4+\delta(\alpha)}\right\} \leq b_n^{3/4+\frac{\alpha}{8}}$.

The following inequality is a direct consequence of Lemma 2.3.3.

Corollary 2.3.4. *Suppose $A_n = (a_{ij})_{i,j\in[n]}$ satisfies Assumption 1 with $\alpha \in (0,2)$. Let $a_* := \max_{i,j\in[n]} |a_{ij}|$. Then whp, we have*

$$\max_{i\in[n]} \sum_{j\in[n]} |a_{ij}| \leq a_*(1 + o(1)) \quad \textit{and} \quad \max_{j\in[n]} \sum_{i\in[n]} |a_{ij}| \leq a_* \left(1 + o(1)\right). \tag{2.3.2}$$

Proof. Due to the symmetry of A, we only prove the first inequality.

Note that a_* is of order $O(b_n)$ according to (2.3.1). Fix $i \in [n]$ and consider the event $\mathcal{E}_i$ that a_* appears in row i or there exists one entry in row i that is at least $b_n^{3/4+\delta(\alpha)}$. Lemma 2.3.3 (b) implies that, whp on $\mathcal{E}_i$, all the remaining entries are less than $b_n^{3/4+\delta(\alpha)}$. Then by Lemma 2.3.3 (c), whp on $\mathcal{E}_i$, the sum of the remaining terms is of order $o(a_*)$.

On the other hand, on the event $\mathcal{E}_i^c$, that is when a_* is not in row i and every entry in row i is less than $b_n^{3/4+\delta(\alpha)}$, Lemma 2.3.3 (c) implies that, whp the sum of this row is $o(a_*)$. This proves the corollary. □

Proof of Theorem 2.2.10. Let $\boldsymbol{v} = (v_i)_{i\in[n]}$ be a maximizing vector of OPT 2 with $\|\boldsymbol{v}\|_r = 1$. Let $v_* := \max_{i\in[n]} |v_i|$ and $a_* := \max_{i,j\in[n]} |a_{ij}|$.

We first show that, for $1 \le r \le p \le \infty$,

$$\|A\|_{r \to p} \le a_* \left(1 + o(1)\right). \tag{2.3.3}$$

This is proved separately for three parameter regimes.

1. Suppose $1 < r \le p < \infty$. Using Lemma 2.3.2 and the definition of Ψ_p, Ψ_r in (2.1.2) in the first line, the definition of v_* and (2.3.2) in the second and third lines, respectively, and again using (2.3.2) in the last line below, we conclude that whp, for all $i \in [n]$,

$$\begin{aligned} \|A\|_{r \to p}^p |v_i|^{r-1} &= \left| \sum_{j=1}^n a_{ij} \left| \sum_{k=1}^n a_{jk} v_k \right|^{p-1} \operatorname{sgn}\left(\sum_{k=1}^n a_{jk} v_k \right) \right| \\ &\le \sum_{j=1}^n |a_{ij}| \left(\sum_{k=1}^n |a_{jk}| \right)^{p-1} v_*^{p-1} \\ &\le \sum_{j=1}^n |a_{ij}| \; a_*^{p-1} (1 + o(1)) v_*^{p-1} \\ &\le a_*^p v_*^{p-1} (1 + o(1)). \end{aligned} \tag{2.3.4}$$

   Choosing i to be the (random) index such that $|v_i| = v_* \le 1$ in (2.3.4), and recalling $r \le p$, we have whp,

$$\|A\|_{r \to p} \le a_* (1 + o(1)).$$

2. Suppose $r = 1$. We have

$$\|A\|_{1 \to p} = \sup_{\|\boldsymbol{x}\|_1, \|\boldsymbol{y}\|_{p^*} \le 1} \boldsymbol{x}^\top A \boldsymbol{y} \le \sup_{\|\boldsymbol{x}\|_1, \|\boldsymbol{y}\|_\infty \le 1} \boldsymbol{x}^\top A \boldsymbol{y} = \|A\|_{1 \to 1} = \sup_{\|\boldsymbol{x}\|_1 = 1} \sum_{i \in [n]} |(A\boldsymbol{x})_i|. \tag{2.3.5}$$

   Since $\boldsymbol{x} \mapsto \sum_{i \in [n]} |(A\boldsymbol{x})_i|$ is convex, the maximal value must be obtained at extremal points of the unit ℓ_1-sphere, i.e. $\boldsymbol{x} \in \{\boldsymbol{e}_j\}_{j\in[n]}$. Hence, (2.3.2) and (2.3.5) imply

$$\|A\|_{1\to p} \le \max_{j\in[n]} \sum_{i\in[n]} |a_{ij}| \le a_* \left(1 + o(1)\right).$$

3. Suppose $p = \infty$. We have

$$\|A\|_{r\to\infty} = \sup_{\|\boldsymbol{x}\|_r, \|\boldsymbol{y}\|_1 \le 1} \boldsymbol{x}^\top A\boldsymbol{y} \le \sup_{\|\boldsymbol{x}\|_\infty, \|\boldsymbol{y}\|_1 \le 1} \boldsymbol{x}^\top A\boldsymbol{y} = \|A\|_{\infty\to\infty} = \sup_{\|\boldsymbol{x}\|_\infty = 1} \max_{i\in[n]} |(A\boldsymbol{x})_i|. \tag{2.3.6}$$

Since $\boldsymbol{x} \mapsto \max_{i\in[n]}$ is convex, the maximal value must be obtained at extremal points of the unit ℓ_∞-sphere, i.e. $\boldsymbol{x} \in \{1, -1\}^n$. Hence (2.3.2) and (2.3.6) imply

$$\|A\|_{r\to\infty} \le \max_{i\in[n]} \sum_{j\in[n]} |a_{ij}| \le a_* \left(1 + o(1)\right).$$

On the other hand, let (i_*, j_*) be a (random) pair of indices such that $a_* = a_{i_* j_*}$. Then we have

$$\|A\|_{r\to p} \ge \|A\boldsymbol{e}_{i_*}\|_p \ge a_*. \tag{2.3.7}$$

Combining (2.3.3) and (2.3.7), we conclude that

$$\|A\|_{r\to p} = a_*(1 + o(1)).$$

Hence, Lemma 2.3.1 implies that $b_n^{-1}\|A\|_{r\to p}$ converges to the desired distribution. $\square$

2.4 Proof of Theorem 2.2.11 and Theorem 2.2.12

Fix $\alpha \in (0, \min\{2, \frac{rp}{r-p}\})$ and $1 \le p < r \le \infty$. Define

$$\begin{aligned} \eta &:= \frac{1}{2}\left(1 - \frac{\alpha(r-p)}{rp}\right), \\ \zeta &:= \min\left\{\frac{1}{4}\left(1 - \frac{\alpha(r-p)}{rp}\right), 1 - \frac{\alpha}{2}\right\}. \end{aligned} \tag{2.4.1}$$

Under the assumptions of Theorem 2.2.11 or Theorem 2.2.12, we have $0 < \zeta < \eta$. The main strategy of the proof is to carefully decompose the heavy-tailed matrix A based on the absolute values of entries, and to control the effect of each of these separately. To this end, define

$$\begin{aligned}
\bar{A} &\equiv (\bar{a}_{ij}) := \left(a_{ij}\mathbb{1}\left(|a_{ij}| \leq n^{\frac{1+\eta}{\alpha}}\right)\right), \\
\hat{A} &\equiv (\hat{a}_{ij}) := \left(a_{ij}\mathbb{1}\left(|a_{ij}| > n^{\frac{2-\zeta}{\alpha}}\right)\right), \\
\tilde{A} &\equiv (\tilde{a}_{ij}) := A - \bar{A} - \hat{A}.
\end{aligned} \tag{2.4.2}$$

In Section 2.4.1 and Section 2.4.2, we show that both $\|\bar{A}\|_{r\to p}$ and $\|\tilde{A}\|_{r\to p}$ are of order $o(b_n)$ whp, where b_n is the constant defined in (2.2.3). Then in Section 2.4.3, we analyze the fluctuations of $b_n^{-1}\|\hat{A}\|_{r\to p}$ and use that to conclude the proof of Theorem 2.2.11.

2.4.1 Estimating the small terms

The goal of this section is to prove the following estimate.

Proposition 2.4.1. *Recall b_n in (2.2.3). Under assumptions of Theorem 2.2.11 or Theorem 2.2.12, whp,*

$$\|\bar{A}\|_{r\to p} = o\left(b_n\right).$$

To prove the result, we start by establishing a general upper bound for the $r \to p$ norm of an arbitrary $n \times n$ matrix.

Lemma 2.4.2. *Fix $1 \leq p < r \leq \infty$. For any symmetric matrix $V = (v_{ij})_{i,j\in[n]} \in \mathbb{R}^n$, we have*

$$\|V\|_{r\to p} \leq \left(\sum_{i\in[n]}\left(\sum_{j\in[n]}|v_{ij}|\right)^{\frac{rp}{r-p}}\right)^{\frac{r-p}{rp}}.$$

Proof. Fix any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ such that $\|\boldsymbol{x}\|_r = \|\boldsymbol{y}\|_{p^*} = 1$. Since $r > p$, applying Hölder's inequality twice we obtain

$$
\begin{aligned}
\sum_{i,j\in[n]} v_{ij} x_i y_j &\le \sum_{i,j\in[n]} |v_{ij}|^{\frac{p}{rp-r+p}} |x_i| \cdot |v_{ij}|^{\frac{rp-r}{rp-r+p}} |y_j| \\
&\le \left(\sum_{i,j\in[n]} |v_{ij}| |x_i|^{\frac{rp-r+p}{p}} \right)^{\frac{p}{rp-r+p}} \left(\sum_{i,j\in[n]} |v_{ij}| |y_j|^{\frac{rp-r+p}{rp-r}} \right)^{\frac{rp-r}{rp-r+p}} \\
&\le \left(\sum_{i\in[n]} |x_i|^r \right)^{\frac{1}{r}} \left(\sum_{i\in[n]} \left(\sum_{j\in[n]} |v_{ij}| \right)^{\frac{rp}{r-p}} \right)^{\frac{r-p}{r(rp-r+p)}} \\
&\quad \times \left(\sum_{j\in[n]} |y_j|^{p^*} \right)^{\frac{1}{p^*}} \left(\sum_{j\in[n]} \left(\sum_{i\in[n]} |v_{ij}| \right)^{\frac{rp}{r-p}} \right)^{\frac{r-p}{p^*(rp-r+p)}} \\
&= \left(\sum_{i\in[n]} \left(\sum_{j\in[n]} |v_{ij}| \right)^{\frac{rp}{r-p}} \right)^{\frac{r-p}{rp}},
\end{aligned}
\tag{2.4.3}
$$

where we use $\|\boldsymbol{x}\|_r = \|\boldsymbol{y}\|_{p^*} = 1$ and the symmetry of V for the last equality. In view of the corresponding dual formulation (2.2.9) of $\|V\|_{r\to p}$, one can take the supremum over $\boldsymbol{x}, \boldsymbol{y}$ with $\|\boldsymbol{x}\|_r = \|\boldsymbol{y}\|_{p^*} = 1$ on the left-hand side of (2.4.3) to complete the proof. $\square$

We now bound the row sums of $\bar{A}$. Note that in conjunction with Lemma 2.4.2, this will yield an upper bound for $\|\bar{A}\|_{r\to p}$. To this end, we need the following standard concentration result [163, Theorem 2.8.4].

Lemma 2.4.3 (Bernstein's inequality)**.** *Let $X_1, \ldots, X_n$ be independent, mean zero random variables, such that $|X_k| \le K$ for all $k \in [n]$. Then, for every $t \ge 0$, we have*

$$
\mathbb{P}\left\{ \left| \sum_{k=1}^{n} X_k \right| \ge t \right\} \le 2\exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3} \right),
$$

where $\sigma^2 = \sum_{k=1}^{n} \mathbb{E}[X_k^2]$ is the variance of the sum.

Lemma 2.4.4. *Under the assumptions of Theorem 2.2.11 or Theorem 2.2.12, whp, we* *have for all $i \in [n]$,*

$$\left|\sum_{i=1}^{n} \bar{a}_{ij}\right| \leq \sum_{j=1}^{n} |\bar{a}_{ij}| = O_\delta\left(n^{\max\left\{\frac{1+\eta}{\alpha},1\right\}}\right).$$

Proof. First, we calculate the second moment of $\bar{a}_{ij}$ using $\bar{F}$ from (2.2.2) and (2.2.1) to obtain

$$\mathbb{E}\left[\bar{a}_{ij}^2\right] \leq \int_0^{n^{2(1+\eta)/\alpha}} \left(\bar{F}\left(x^{\frac{1}{2}}\right) - \bar{F}\left(n^{\frac{1+\eta}{\alpha}}\right)\right) dx = O_\delta\left(n^{\left(\frac{2}{\alpha}-1\right)(1+\eta)}\right).$$

This implies that

$$\operatorname{Var}\left(\sum_{j=1}^{n} |\bar{a}_{ij}|\right) \leq n\mathbb{E}\left[\bar{a}_{11}^2\right] = O_\delta\left(n^{\frac{2}{\alpha}+\eta\left(\frac{2}{\alpha}-1\right)}\right). \tag{2.4.4}$$

We now consider three cases. First, suppose $\alpha < 1$. Then

$$\mathbb{E}\left[\sum_{j=1}^{n} |\bar{a}_{ij}|\right] = n\int_0^{n^{(1+\eta)/\alpha}} \left(\bar{F}(x) - \bar{F}\left(n^{\frac{1+\eta}{\alpha}}\right)\right) dx = O_\delta\left(n^{\frac{1}{\alpha}+\eta\left(\frac{1}{\alpha}-1\right)}\right),$$

where we again used (2.2.1) in the last equality. Next suppose $\alpha = 1$. Then similarly, we have

$$\mathbb{E}\left[\sum_{j=1}^{n} |\bar{a}_{ij}|\right] = n\int_0^{n^{(1+\eta)/\alpha}} \left(\bar{F}(x) - \bar{F}\left(n^{\frac{1+\eta}{\alpha}}\right)\right) dx = O_\delta(n\log n) = O_\delta(n).$$

Finally, in the case $\alpha \in \left(1, \min\{2, \frac{rp}{r-p}\}\right)$, $\mathbb{E}\left[|a_{11}|\right] < \infty$, hence

$$\mathbb{E}\left[\sum_{j=1}^{n} |\bar{a}_{ij}|\right] \leq n\mathbb{E}\left[|a_{11}|\right] = O_\delta(n).$$

Since $|\bar{a}_{ij}| \le n^{\frac{1+\eta}{\alpha}}$ by (2.4.2), Lemma 2.4.3 and (2.4.4) imply that for any $\delta > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{j=1}^{n} |\bar{a}_{ij}| - \mathbb{E}\left[\sum_{j=1}^{n} |\bar{a}_{ij}|\right]\right| \ge n^{\frac{1+\eta}{\alpha}+\delta}\right) \\
&\le 2\exp\left(-\frac{cn^{\frac{2(1+\eta)}{\alpha}+2\delta}}{\operatorname{Var}\left(\sum_{j=1}^{n} |\bar{a}_{ij}|\right) + n^{\frac{1+\eta}{\alpha}} n^{\frac{1+\eta}{\alpha}+\delta}}\right) \\
&= 2\exp\left(-\frac{cn^{\delta}}{O_\delta\left(n^{-\eta}\right) + 1}\right) \\
&\le 2\exp\left(-cn^{\delta}\right).
\end{aligned}$$

To complete the proof, combine the last four displays with the union bound. □

We now establish three additional lemmas that hold under the assumptions of Theorem 2.2.12. The first one bounds the mean of $\bar{a}_{ij}$ as defined in (2.4.2), the second lemma concerns the spectral norm of $\bar{A}$ and the third one relates the spectral norm to the $r \to p$ operator norm.

Lemma 2.4.5. *Fix $1 \le p < r \le \infty$ and suppose A satisfies Assumption 1 with index α and (2.2.11). Then, for $i, j \in [n]$, we have*

$$|\mathbb{E}\left[\bar{a}_{ij}\right]| = O\left(n^{\frac{1}{\alpha}-1}\right).$$

Proof. Fix $i, j \in [n]$. By the mean zero condition on a_{ij}, we have

$$\int_0^\infty \mathbb{P}\left(a_{ij} > x\right) dx = \int_0^\infty \mathbb{P}\left(a_{ij} < -x\right) dx =: K_0.$$

By the definition of $\bar{a}_{ij}$ in (2.4.2), it follows that

$$\begin{aligned}\mathbb{E}\left[\bar{a}_{ij}\right] &= \int_0^{n^{\frac{1+\eta}{\alpha}}} \mathbb{P}\left(x < a_{ij} \le n^{\frac{1+\eta}{\alpha}}\right) dx - \int_0^{n^{\frac{1+\eta}{\alpha}}} \mathbb{P}\left(-n^{\frac{1+\eta}{\alpha}} \le a_{ij} < -x\right) dx \\ &= \left(K_0 - \int_{n^{\frac{1+\eta}{\alpha}}}^{\infty} \mathbb{P}\left(a_{ij} > x\right) dx - n^{\frac{1+\eta}{\alpha}} \mathbb{P}\left(a_{ij} > n^{\frac{1+\eta}{\alpha}}\right)\right) - \\ &\quad \left(K_0 - \int_{n^{\frac{1+\eta}{\alpha}}}^{\infty} \mathbb{P}\left(a_{ij} < -x\right) dx - n^{\frac{1+\eta}{\alpha}} \mathbb{P}\left(a_{ij} < -n^{\frac{1+\eta}{\alpha}}\right)\right) \\ &= \int_{n^{\frac{1+\eta}{\alpha}}}^{\infty} \mathbb{P}\left(a_{ij} < -x\right) dx - \int_{n^{\frac{1+\eta}{\alpha}}}^{\infty} \mathbb{P}\left(a_{ij} > x\right) dx \\ &\quad + n^{\frac{1+\eta}{\alpha}} \mathbb{P}\left(a_{ij} < -n^{\frac{1+\eta}{\alpha}}\right) - n^{\frac{1+\eta}{\alpha}} \mathbb{P}\left(a_{ij} > n^{\frac{1+\eta}{\alpha}}\right).\end{aligned}$$

This implies that

$$\left|\mathbb{E}\left[\bar{a}_{ij}\right]\right| \le \int_{n^{\frac{1+\eta}{\alpha}}}^{\infty} \mathbb{P}\left(|a_{ij}| > x\right) dx + n^{\frac{1+\eta}{\alpha}} \mathbb{P}\left(|a_{ij}| > n^{\frac{1+\eta}{\alpha}}\right) = O_\delta\left(n^{\frac{1}{\alpha}-1+\frac{\eta(1-\alpha)}{\alpha}}\right) = O\left(n^{\frac{1}{\alpha}-1}\right),$$

where we used (2.2.2) in the first equality and $\alpha > 1$ from (2.2.11) in the last equality. □

Lemma 2.4.6. *Fix $1 \le p < r \le \infty$ and suppose A satisfies Assumption 1 with index α and (2.2.11). Let b_n and $\bar{A}$ be as defined in (2.2.3) and (2.4.2), respectively. Then whp,*

$$\|\bar{A}\|_{2\to 2} = O_\delta(n^{\frac{1+\eta}{\alpha}}). \tag{2.4.5}$$

Proof. Let $\mu := \mathbb{E}\left[\bar{a}_{11}\right]$. Using Lemma 2.4.5 in the last step, we have

$$\left|\|\bar{A}\|_{2\to 2} - \|\bar{A} - \mu \mathbf{1}\mathbf{1}^\top\|_{2\to 2}\right| \le \|\mu \mathbf{1}\mathbf{1}^\top\|_{2\to 2} = \mu n = O\left(n^{\frac{1}{\alpha}}\right).$$

Hence, to show (2.4.5), we may assume without loss of generality that $\{\bar{a}_{ij} : i, j \in [n]\}$ are centered. Furthermore, by a standard symmetrization argument (see [159, (2.60)]), in the rest of the proof we may assume $\bar{a}_{ij} \overset{d}{=} -\bar{a}_{ij}$.

We first show that

$$\mathbb{E}\left[\|\bar{A}\|_{2\to 2}\right] = O_\delta\left(n^{\frac{1+\eta}{\alpha}}\right). \tag{2.4.6}$$

To prove (2.4.6), define $\sigma_{ij} := \left(\mathbb{E}\left[\bar{a}_{ij}^2\right]\right)^{1/2}$ and $\xi_{ij} := \left(\mathbb{E}\left[\bar{a}_{ij}^2\right]\right)^{-1/2} \bar{a}_{ij}$, and $\sigma := \max_{i\in[n]} \left(\sum_{j\in[n]} \sigma_{ij}^2\right)^{1/2}$. Then it is easy to verify that

$$\sigma = \left(n\mathbb{E}\left[\bar{a}_{11}^2\right]\right)^{\frac{1}{2}} = O_\delta\left(n^{\frac{1+\eta}{\alpha}}\right),$$
$$\max_{i,j} \mathbb{E}\left[|\sigma_{ij}\xi_{ij}|^{2\lceil 3\log n\rceil}\right]^{\frac{1}{2\lceil 3\log n\rceil}} = O_\delta\left(n^{\frac{1+\eta}{\alpha}}\right).$$

Then [16, Corollary 3.6], together with the assumption $\bar{a}_{ij} \overset{d}{=} -\bar{a}_{ij}$, yields the bound

$$\mathbb{E}\left[\|\bar{A}\|_{2\to 2}\right] \le e^{2/3}\left[2\sigma + 52\max_{i,j}\mathbb{E}\left[|\sigma_{ij}\xi_{ij}|^{2\lceil 3\log n\rceil}\right]^{\frac{1}{2\lceil 3\log n\rceil}}\sqrt{\log n}\right] = O_\delta\left(n^{\frac{1+\eta}{\alpha}}\right).$$

This proves the claim (2.4.6).

Next, define the map $F : \mathbb{R}^{(n+1)n/2} \to \mathbb{R}$ as follows: given $\boldsymbol{x} = (x_{ij})_{1\le i\le j\le n}$, $F(\boldsymbol{x})$ is the $2 \to 2$ norm of the $n \times n$ symmetric matrix with upper triangular part equal to $(x_{ij})_{1\le i\le j\le n}$. Since F is convex with Lipschitz constant 1 and $|\bar{a}_{ij}| \le n^{\frac{1+\eta}{\alpha}}$ by (2.4.2), Talagrand's concentration inequality [159, Theorem 2.1.13] implies that

$$\mathbb{P}\left(\|\bar{A}\|_{2\to 2} - \mathbb{E}\|\bar{A}\|_{2\to 2} \ge \lambda n^{\frac{1+\eta}{\alpha}}\right) \le C\exp\left(-c\lambda^2\right).$$

When combined with (2.4.6), this completes the proof of the lemma. □

Lemma 2.4.7. *Fix $1 \le p < r \le \infty$ and suppose A satisfies Assumption 1 with index α and (2.2.11). Let b_n and $\bar{A}$ be as defined in (2.2.3) and (2.4.2), respectively. Then whp,*

$$\|\bar{A}\|_{r\to p} = O_\delta\left(n^{\frac{1}{\min\{2,p\}} - \frac{1}{\max\{2,r\}} + \frac{1+\eta}{\alpha}}\right). \tag{2.4.7}$$

Proof. We prove this by considering three cases:

1. Suppose $2 \le p < r \le \infty$. First note that, for any $\boldsymbol{x} \in \mathbb{R}^n$,

$$\|\boldsymbol{x}\|_p = \sup_{\|\boldsymbol{y}\|_{p^*}\le 1} \langle \boldsymbol{x}, \boldsymbol{y}\rangle \le \sup_{\|\boldsymbol{y}\|_2\le 1} \langle \boldsymbol{x}, \boldsymbol{y}\rangle = \|\boldsymbol{x}\|_2, \tag{2.4.8}$$

where we use $p^* \leq 2$ in the inequality. Using (2.4.8), Hölder's inequality and Lemma 2.4.6, we have whp,

$$\|\bar{A}\|_{r\to p} = \sup_{\boldsymbol{x}\neq 0} \frac{\|\bar{A}\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_r} \leq \sup_{\boldsymbol{x}\neq 0} \frac{\|\bar{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_r} \leq \sup_{\boldsymbol{x}\neq 0} \frac{\|\bar{A}\boldsymbol{x}\|_2}{n^{\frac{1}{r}-\frac{1}{2}}\|\boldsymbol{x}\|_2} = n^{\frac{1}{2}-\frac{1}{r}}\|\bar{A}\|_{2\to 2} = O_\delta(n^{\frac{1}{2}-\frac{1}{r}+\frac{1+\eta}{\alpha}}).$$

2. Suppose $1 \leq p < r \leq 2$, notice that $\|\bar{A}\|_{r\to p} = \|\bar{A}\|_{p^*\to r^*}$ and $2 \leq r^* < p^* \leq \infty$. The result in this case follows as in Case 1.

3. Finally, suppose $1 \leq p < 2 < r \leq \infty$. Then Hölder's inequality and Lemma 2.4.6 imply

$$\|\bar{A}\|_{r\to p} = \sup_{x\neq 0} \frac{\|\bar{A}x\|_p}{\|x\|_r} \leq \sup_{x\neq 0} \frac{n^{\frac{1}{p}-\frac{1}{2}}\|\bar{A}x\|_2}{n^{\frac{1}{r}-\frac{1}{2}}\|x\|_2} = n^{\frac{1}{p}-\frac{1}{r}}\|\bar{A}\|_{2\to 2} = O_\delta(n^{\frac{1}{p}-\frac{1}{r}+\frac{1+\eta}{\alpha}}).$$

Together, the three cases imply (2.4.7). $\square$

Proof of Proposition 2.4.1. Suppose $1 \leq p < r \leq \infty$ and (2.2.10) holds. Invoking first Lemma 2.4.2, then Lemma 2.4.4 and finally using (2.2.10), (2.4.1) for the last equality, we obtain

$$\|\bar{A}\|_{r\to p} \leq \left(\sum_{i\in[n]}\left(\sum_{j\in[n]}|\bar{a}_{ij}|\right)^{\frac{rp}{r-p}}\right)^{\frac{r-p}{rp}} = O_\delta\left(n^{\max\{1,\frac{1+\eta}{\alpha}\}+\frac{r-p}{rp}}\right) = o(b_n).$$

This completes the proof under the assumptions of Theorem 2.2.11.

The conclusion under assumptions of Theorem 2.2.12 follows by combining (2.2.11), (2.4.1) and (2.4.7). $\square$

2.4.2 Estimating terms of intermediate magnitude

Note from (2.4.2) that

$$\tilde{a}_{ij} \coloneqq a_{ij}\mathbb{1}\left(n^{\frac{1+\eta}{\alpha}} < |a_{ij}| \leq n^{\frac{2-\zeta}{\alpha}}\right). \tag{2.4.9}$$

The goal of this section is the following estimate.

Proposition 2.4.8. *Recall b_n in (2.2.3). Under assumptions of Theorem 2.2.11 or Theorem 2.2.12, whp,*

$$\|\tilde{A}\|_{r\to p} = o\left(b_n\right).$$

The proof of this result, which is postponed to the end of this section, relies on two auxiliary results. The first one, Lemma 2.4.9, establishes the sparsity of $\tilde{A}$.

Lemma 2.4.9. *Under assumptions of Theorem 2.2.11 or Theorem 2.2.12, there exists $\kappa \equiv \kappa(\eta) \in \mathbb{N}$ independent of n such that, whp, there exist at most κ nonzero entries in any row or column of $\tilde{A}$.*

Proof. Define

$$\kappa \equiv \kappa(\eta) := \left\lceil \frac{1}{\eta} \right\rceil + 1. \tag{2.4.10}$$

Fix $n \in \mathbb{N}$, define the event $\tilde{\Omega} \equiv \tilde{\Omega}_n$ as follows:

$$\tilde{\Omega} := \{\exists\, i \in [n] \text{ such that } \exists\, \kappa \text{ nonzero entries on row (or column) } i \text{ of } A\}.$$

By (2.2.2), (2.4.9) and (2.4.10), it follows that

$$\mathbb{P}\left(\tilde{\Omega}\right) \leq \mathbb{P}\left(|a_{ij}| \geq n^{\frac{1+\eta}{\alpha}}\right)^{\kappa} \binom{n}{\kappa} n = O_{\delta}\left(n^{1-\kappa\eta}\right) = O_{\delta}(n^{-1}),$$

which converges to zero as $n \to \infty$. This completes the proof of the lemma. $\square$

Remark 2.4.10. As a simple consequence, with κ as in Lemma 2.4.9, for any $s, t > 0$, we have whp, for all $i \in [n]$

$$\left(\sum_{j\in[n]} |\tilde{a}_{ij}|\right)^s \le \kappa^s \max_{j\in[n]} |\tilde{a}_{ij}|^s \le \kappa^s \sum_{j\in[n]} |\tilde{a}_{ij}|^s, \tag{2.4.11}$$

and, using the rearrangement inequality [104, Theorem 368], we also have

$$\sum_{j\in[n]} |\tilde{a}_{ij}|^s \sum_{k\in[n]} |\tilde{a}_{ik}|^s \le \kappa \sum_{j\in[n]} |\tilde{a}_{ij}|^{s+t}.$$

We now state the second auxiliary result, which exploits the heavy-tailed nature of $(a_{ij})_{i,j\in[n]}$.

Lemma 2.4.11. *Let $(X_\ell)_{\ell\in[n(n+1)/2]}$ be i.i.d. copies of a nonnegative heavy-tailed random variable with index $\alpha \in (0,2)$. Fix $q > \alpha$ and $\beta < \frac{2}{\alpha}$. Define $\tilde{X}_\ell := X_\ell \mathbb{1}\left\{X_\ell \le n^\beta\right\}$. Then whp,*

$$\sum_{\ell\in[n(n+1)/2]} \tilde{X}_\ell^q = O_\delta\left(n^{q\beta-\alpha\beta+2}\right). \tag{2.4.12}$$

Proof. We prove (2.4.12) by directly calculating the second moment of the left-hand side of (2.4.12). Indeed, note that by Definition 2.2.3 and Lemma 2.2.2, it follows that

$$\begin{aligned}
\mathbb{E}\left[\tilde{X}_1^{2q}\right] &\le O_\delta(1) \int_0^{n^{2q\beta}} x^{-\frac{\alpha}{2q}} dx = O_\delta\left(n^{2q\beta-\beta\alpha}\right), \\
\mathbb{E}\left[\tilde{X}_1^{q}\right] &\le O_\delta(1) \int_0^{n^{q\beta}} x^{-\frac{\alpha}{q}} dx = O_\delta\left(n^{q\beta-\beta\alpha}\right).
\end{aligned}$$

Therefore, by the assumption $\beta < \frac{2}{\alpha}$, it follows that

$$\mathbb{E}\left[\left(\sum_{i\in[n(n+1)/2]} X_i^q\right)^2\right] \le n^2 \mathbb{E}\left[\tilde{X}_1^{2q}\right] + n^4\left(\mathbb{E}\left[\tilde{X}_1^q\right]\right)^2 = O_\delta\left(n^{2q\beta-2\beta\alpha+4}\right).$$

Together with Markov's inequality, this proves (2.4.12). □

Proof of Proposition 2.4.8. Fix an arbitrary $\boldsymbol{x} = (x_i)_{i\in[n]} \in \mathbb{R}^n$ such that $\|\boldsymbol{x}\|_r = 1$. Let

κ be as in Lemma 2.4.9. By (2.4.11), it follows that, whp, for each $i \in [n]$,

$$\left| \sum_{j\in[n]} \tilde{a}_{ij} x_j \right|^p \le \kappa^p \sum_{j\in[n]} |\tilde{a}_{ij} x_j|^p . \tag{2.4.13}$$

Let M be the (random) number of nonzero entries in $\tilde{A}$, and for each $i \in [M]$, let $\{\tilde{a}_{*,\ell}\}_{\ell\in[M]}$ be a non-increasing rearrangement of the nonzero entries of $\tilde{A}$:

$$\{\tilde{a}_{*,\ell} : \ell \in [M]\} = \{\tilde{a}_{ij} : \tilde{a}_{ij} \neq 0,\, i, j \in [n]\} \quad \text{and} \quad |\tilde{a}_{*,\ell}| \ge |\tilde{a}_{*,\ell+1}| \quad \forall\, \ell \in [M]. \tag{2.4.14}$$

Consider the random function $\varphi : [M] \to [n]$ associated to this rearrangement that maps each $\ell \in [M]$ to the column index of the term corresponding to $\tilde{a}_{*,\ell}$ in $\tilde{A}$. Then by Hölder's inequality, we have

$$\sum_{i\in[n]} \sum_{j\in[n]} |\tilde{a}_{ij} x_j|^p = \sum_{\ell\in[M]} |\tilde{a}_{*,\ell} x_{\varphi(\ell)}|^p \le \left(\sum_{\ell\in[M]} |\tilde{a}_{*,\ell}|^{\frac{rp}{r-p}} \right)^{\frac{r-p}{r}} \left(\sum_{\ell\in[M]} |x_{\varphi(\ell)}|^r \right)^{\frac{p}{r}} . \tag{2.4.15}$$

By definition, for $j \in [n]$, $\varphi^{-1}(\{j\})$ denotes the number of nonzero entries in the j-th column of $\tilde{A}$, which is at most κ whp by Lemma 2.4.9. Therefore, since $\|\boldsymbol{x}\|_r = 1$, we obtain whp,

$$\sum_{\ell\in[M]} |x_{\varphi(\ell)}|^r \le \sum_{j\in[n]} |\varphi^{-1}(\{j\})| |x_j|^r \le \kappa \sum_{j\in[n]} |x_j|^r = \kappa. \tag{2.4.16}$$

Combining (2.4.13), (2.4.15) and (2.4.16), we have whp,

$$\|\tilde{A}\|_{r\to p}^p = \sup_{\boldsymbol{x}:\|\boldsymbol{x}\|_r=1} \sum_{i\in[n]} \left| \sum_{j\in[n]} \tilde{a}_{ij} x_j \right|^p \le \kappa^{p+\frac{p}{r}} \left(\sum_{\ell=1}^{m} |\tilde{a}_{*,\ell}|^{\frac{rp}{r-p}} \right)^{\frac{r-p}{r}} .$$

By (2.4.9) and the definition of $\tilde{a}_{*,\ell}$ in (2.4.14), we have

$$\sum_{\ell\in[M]} |\tilde{a}_{*,\ell}|^{\frac{rp}{r-p}} \le \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^{\frac{rp}{r-p}} \mathbb{1}\left(|a_{ij}| \le n^{\frac{2-\zeta}{\alpha}} \right) \le 2 \sum_{1\le i\le j\le n} |a_{ij}|^{\frac{rp}{r-p}} \mathbb{1}\left(|a_{ij}| \le n^{\frac{2-\zeta}{\alpha}} \right) .$$

Under assumptions of Theorem 2.2.11 or Theorem 2.2.12, we have $\alpha < \frac{rp}{r-p}$. Applying Lemma 2.4.11 with $\{X_\ell\}_{\ell\in[n(n+1)/2]} = \{|a_{ij}|\}_{1\le i\le j\le n}$, $q = \frac{rp}{r-p}$ and $\beta = \frac{2-\zeta}{\alpha}$, and with b_n as in (2.2.3), we have whp,

$$\sum_{1\le i\le j\le n} |a_{ij}|^{\frac{rp}{r-p}} \mathbb{1}\left\{|a_{ij}| \le n^{\frac{2-\zeta}{\alpha}}\right\} = O_\delta\left(n^{\frac{rp}{r-p}\frac{2}{\alpha}-\zeta\left(\frac{1}{\alpha}\frac{rp}{r-p}-1\right)}\right) = o\left(b_n^{\frac{rp}{r-p}}\right).$$

Together, the last three displays prove the proposition. □

2.4.3 Characterizing the fluctuation of large terms

We now turn our attention to the matrix $\hat{A}$ in (2.4.2), starting with the following lemma that counts the number of nonzero entries in $\hat{A}$.

Lemma 2.4.12. *Let $(X_\ell)_{\ell\in[n(n+1)/2]}$ be i.i.d. copies of a heavy-tailed random variable with index $\alpha \in (0,2)$. Fix any $\theta \in (0,1)$, and let $N_n = \#\left\{X_\ell : |X_\ell| > n^{\frac{2\theta}{\alpha}}\right\}$. Then whp,*

$$N_n = \Theta_\delta\left(n^{2-2\theta}\right).$$

Proof. Letting $\bar{F}$ denote complement cumulative distribution function of X_ℓ, note that $N_n \sim \mathrm{Ber}\left(\frac{n(n+1)}{2}, \bar{F}\left(n^{\frac{2\theta}{\alpha}}\right)\right)$. Thus, by (2.2.2), we have

$$\begin{aligned}\mathbb{E}[N_n] &= \frac{n(n+1)}{2}\cdot \bar{F}\left(n^{\frac{2\theta}{\alpha}}\right)\\ &= \frac{n(n+1)}{2} L\left(n^{\frac{2\theta}{\alpha}}\right) n^{-2\theta} \\ &= \Theta_\delta\left(n^{2-2\theta}\right),\end{aligned} \tag{2.4.17}$$

where the last equality uses (2.2.1). An application of Chernoff's bound [137, Theorem 4.4 and Theorem 4.5] then implies that

$$\mathbb{P}\left(|N_n - \mathbb{E}[N_n]| > \frac{1}{2}\mathbb{E}[N_n]\right) \le 2\exp\left(-\Theta_\delta\left(n^{2-2\theta}\right)\right).$$

Together with (2.4.17), the assumption $\theta \in (0,1)$ and the union bound, this proves the lemma. $\square$

We now state the second lemma, which simplifies the structure of $\hat{A}$ while preserving its $r \to p$ norm (and also the optimal value of OPT 1 which is going to be used in Section 2.6.2).

Lemma 2.4.13. *Let V be a symmetric $n \times n$ matrix that satisfies the following three properties:*

(1) zero diagonal entries, that is, $V_{ii} = 0, \forall i \in [n]$;

(2) there is at most one non-zero entry in each row and each column;

(3) $\exists\, m \in \mathbb{N}$ such that $m < \frac{n}{2}$ and $\#\{(i,j) \in [n] \times [n] : V_{ij} \neq 0\} = 2m$.

Then there exists a symmetric $n \times n$ matrix $\check{V}$ with $\left\{\check{V}_{ij} : i,j \in [n]\right\} = \{V_{ij} : i,j \in [n]\}$ that satisfies properties (1) $\sim$ (3) above and in addition, also satisfies three additional properties:

(4) $\check{V}_{ij} = 0$ if $\max\{i,j\} > 2m$;

(5) For $k \in [m]$, $\check{V}_{2k-1,2k} = \check{V}_{2k,2k-1}$ has a magnitude that is k-th largest in $\{|V_{ij}| : 1 \leq i < j \leq n\}$;

(6) $\|\check{V}\|_{r \to p} = \|V\|_{r \to p}$ and $M_r(\check{V}) = M_r(V)$, where we recall that $M_r(V), M_r(\check{V})$ are the optimal values of the corresponding ℓ_r-Grothendieck problems defined in OPT 1.

Moreover, one can ensure that the map from V to $\check{V}$ is measurable.

Remark 2.4.14. The output matrix $\check{V}$ of Lemma 2.4.13 is illustrated in Figure 2.2, where blue region represents region of zero entries and the red squares represent nonzero entries. Note that each row or column of $\check{V}$ has at most one red square and all red squares are off-diagonal.

Proof. The result is intuitive, but we provide a proof for completeness.

Given $i,j \in [n]$, consider the mapping $\Pi_{i,j} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n \times n}$ (respectively, $\Pi^{i,j} : \mathbb{R}^{n\times n} \to$

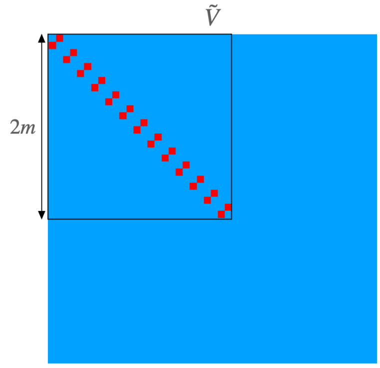


Figure 2.2: Structure of $\check{V}$

$\mathbb{R}^{n\times n}$) that transposes the i-th and j-th rows (respectively, columns) of the matrix. More precisely, given $V \in \mathbb{R}^{n\times n}$, $V' = \Pi_{i,j}V$ satisfies $V'_{ik} = V_{jk}$, $V'_{jk} = V_{ik}$ and $V'_{\ell k} = V_{\ell k}$ for all $k \in [n]$ and $\ell \in [n]\backslash\{i,j\}$, and $\Pi^{i,j}V = \left(\Pi_{i,j}V^\top\right)^\top$. Then it is easy to verify that $\|\Pi V\|_{r\to p} = \|V\|_{r\to p}$ for $\Pi = \Pi_{i,j}$ or $\Pi = \Pi^{i,j}$ and $M_r(\Pi_{i,j}\Pi^{i,j}V) = M_r(V)$. It is also immediate that if V satisfies properties (2) and(3), then so does $\Pi_{i,j}V$ and $\Pi^{i,j}V$.

Now, set $V^{(0)} := V$, which satisfies properties (1) – (3), and recursively, for $k \in [m]$, define $i(k), j(k) \in [n]$ to be such that $j(k) > i(k)$ and $\left|V^{(2k-2)}_{i(k),j(k)}\right| = \mathrm{argmax}_{i>2k-2}\left|V^{(2k-2)}_{ij}\right|$, and then set $V^{(2k-1)} := \Pi_{2k-1,i(k)}\Pi^{2k-1,i(k)}V^{(2k-2)}$ and $V^{(2k)} := \Pi_{2k,j(k)}\Pi^{2k,j(k)}V^{(2k-1)}$. Then define $\check{V} := V^{(2m)}$. By the properties of $\Pi_{i,j}$ and $\Pi^{i,j}$ mentioned above, it is immediate that $\check{V}$ satisfies properties (2), (3) and (6), and by construction, it is clear that $\check{V}$ is symmetric (since V is) and satisfies (1) and (5). Finally, the fact that $\check{V}$ satisfies (5) and V satisfies (3) together imply that $\check{V}$ satisfies (4). $\square$

The following result characterizes the fluctuations of $\|\hat{A}\|_{r\to p}$ and $M_r(\hat{A})$. The latter will be used in Section 2.6.2.

Proposition 2.4.15. *Suppose* $1 \le p < r \le \infty$ *and* $\alpha \in \left(0, \min\left\{2, \frac{1}{\gamma}\right\}\right)$, *where* $\gamma := \frac{1}{p} - \frac{1}{r}$. *Let* $A_n \in \mathbb{R}^{n\times n}$ *satisfy Assumption 1 with index* α, $\hat{A}_n$ *as in* (2.4.2) *and* b_n *as in* (2.2.3).

1. *There exists an* $\alpha\gamma$*-stable random variable such that*

$$b_n^{-1}\|\hat{A}_n\|_{r\to p} \xrightarrow{(d)} Y^{\gamma}_{\alpha\gamma}.$$

2. *Suppose further that $r > 2$ and $p = r^*$. Then there exists an $\alpha\gamma$-stable random variable such that*

$$b_n^{-1} M_r(\hat{A}_n) \overset{(d)}{\to} Y_{\alpha\gamma}^{\gamma}.$$

Proof. In this proof, we drop the subscript n of A_n and $\hat{A}_n$ for brevity.

Since A is symmetric, (2.4.2) ensures $\hat{A}$ is also symmetric and by Lemma 2.3.3, $\hat{A}$ satisfies properties (1) and (3) of Lemma 2.4.13. Moreover, if $2m(n)$ denotes the number of nonzero entries of $\hat{A} = \hat{A}_n$, then an application of Lemma 2.4.12 with $\theta = 1 - \frac{\zeta}{2}$, together with the choice of ζ in (2.4.1), shows that $m(n) = o(n)$ whp. Thus, Lemma 2.4.13 shows that whp, there exists another symmetric matrix $\check{A}$ that satisfies properties (1) – (6) of Lemma 2.4.13.

1. Fix $1 \le p < r \le \infty$. By the structure of $\check{A}$ and the corresponding dual formulation (2.2.9) of $\|\check{A}\|_{r\to p}$, we see that,

$$\|\check{A}\|_{r\to p} = \max_{\|\boldsymbol{x}\|_r = \|\boldsymbol{y}\|_{p^*}=1} \sum_{k\in[m]} \left(\check{A}_{2k-1,2k} x_{2k} y_{2k-1} + \check{A}_{2k,2k-1} x_{2k-1} y_{2k}\right). \tag{2.4.18}$$

For all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ with $\|\boldsymbol{x}\|_r = \|\boldsymbol{y}\|_{p^*} = 1$, using Hölder's inequality twice, we have

$$\begin{aligned}
&\sum_{k\in[m(n)]} \left(\check{A}_{2k-1,2k} x_{2k} y_{2k-1} + \check{A}_{2k,2k-1} x_{2k-1} y_{2k}\right) \\
&\le \left(2\sum_{k\in[m(n)]} \left|\check{A}_{2k-1,2k}\right|^{\frac{1}{\gamma}}\right)^{\gamma} \left[\sum_{k\in[m(n)]} \left(|x_{2k} y_{2k-1}|^{\frac{rp}{rp-r+p}} + |x_{2k-1} y_{2k}|^{\frac{rp}{rp-r+p}}\right)\right]^{\frac{rp-r+p}{rp}} \\
&\le \left(2\sum_{k\in[m(n)]} \left|\check{A}_{2k-1,2k}\right|^{\frac{1}{\gamma}}\right)^{\gamma} \left(\sum_{k\in[2m(n)]} |x_k|^r\right)^{\frac{1}{r}} \left(\sum_{k\in[2m(n)]} |y_k|^{p^*}\right)^{\frac{1}{p^*}} \\
&\le \left(2\sum_{k\in[m(n)]} \left|\check{A}_{2k-1,2k}\right|^{\frac{1}{\gamma}}\right)^{\gamma},
\end{aligned} \tag{2.4.19}$$

and the equalities can be achieved simultaneously by choosing $\boldsymbol{x} = (x_1, \dots, x_n), \boldsymbol{y} =$

$(y_1, \ldots, y_n)$ such that $\|\boldsymbol{x}\|_r = \|\boldsymbol{y}\|_{p^*} = 1$, and satisfy

$$x_i \propto \begin{cases} \psi_{\frac{r}{r-p}}\left(\breve{A}_{2k-1,2k}\right), & i = 2k-1, 1 \leq k \leq m(n) \\ \left|\breve{A}_{2k-1,2k}\right|^{\frac{p}{r-p}}, & i = 2k, 1 \leq k \leq m(n) \\ 0, & 2m(n)+1 \leq i \leq n, \end{cases} \tag{2.4.20}$$

and

$$y_i \propto \begin{cases} \psi_{\frac{rp-p}{r-p}}\left(\breve{A}_{2k-1,2k}\right), & i = 2k-1, 1 \leq k \leq m(n) \\ \left|\breve{A}_{2k-1,2k}\right|^{\frac{rp-r}{r-p}}, & i = 2k, 1 \leq k \leq m(n) \\ 0, & 2m(n)+1 \leq i \leq n. \end{cases} \tag{2.4.21}$$

Due to the fact that $\left\{\breve{A}_{ij} : i,j \in [n]\right\} = \left\{\hat{A}_{ij} : i,j \in [n]\right\}$, properties (1) — (3) of Lemma 2.4.13, and the definition of $\hat{A}$ in (2.4.2), we have whp,

$$2\sum_{k\in[m(n)]}\left|\breve{A}_{2k-1,2k}\right|^{\frac{1}{\gamma}} = \sum_{i,j\in[n]}\left|\breve{A}_{ij}\right|^{\frac{1}{\gamma}} = \sum_{i,j\in[n]}\left|\hat{A}_{ij}\right|^{\frac{1}{\gamma}} = 2\sum_{1\leq i<j\leq n}|a_{ij}|^{\frac{1}{\gamma}}\mathbb{1}\left(|a_{ij}| \geq n^{\frac{2-\zeta}{\alpha}}\right). \tag{2.4.22}$$

Since $\alpha < \frac{1}{\gamma}$, applying Lemma 2.4.11 with $\{X_\ell\}_{\ell\in[n(n+1)/2]} = \{|a_{ij}|\}_{1\leq i\leq j\leq n}$, $q = \frac{1}{\gamma}$ and $\beta = \frac{2-\zeta}{\alpha}$, and with b_n as in (2.2.3), we have whp,

$$\sum_{1\leq i\leq j\leq n}|a_{ij}|^{\frac{1}{\gamma}}\mathbb{1}\left\{|a_{ij}| \leq n^{\frac{2-\zeta}{\alpha}}\right\} = O_\delta\left(n^{\frac{2}{\alpha\gamma}-\zeta\left(\frac{1}{\alpha\gamma}-1\right)}\right) = o\left(b_n^{\frac{1}{\gamma}}\right). \tag{2.4.23}$$

Moreover, the generalized central limit theorem [70, Theorem 3.8.2] implies

$$b_n^{-\frac{1}{\gamma}} \cdot 2\sum_{1\leq i<j\leq n}|a_{ij}|^{\frac{1}{\gamma}} \to Y_{\alpha\gamma} \tag{2.4.24}$$

in distribution, where $Y_{\alpha\gamma}$ is an $\alpha\gamma$-stable random variable.

Combining (2.4.18), (2.4.19), (2.4.22), (2.4.23) and (2.4.24) and the continuous mapping theorem, we conclude that

$$b_n^{-1}\|\hat{A}\|_{r\to p} \to Y^{\gamma}_{\alpha,r,p}.$$

This completes the proof.

2. Fix $2 < r \le \infty$. On the one hand, comparing OPT 1 and (2.1.1), it follows that

$$M_r(\check{A}) \le \|\check{A}\|_{r\to r^*}. \tag{2.4.25}$$

On the other hand, note that with $p = r^*$, the optimizers $\boldsymbol{x}, \boldsymbol{y}$ for $\|\check{A}\|_{r\to r^*}$, satisfying $\|\boldsymbol{x}\|_r = \|\boldsymbol{y}\|_{p^*} = 1$, (2.4.20) and (2.4.21), coincide. Hence,

$$\|\check{A}\|_{r\to r^*} = \boldsymbol{x}^\top \check{A}\boldsymbol{x} \le M_r(\check{A}). \tag{2.4.26}$$

Together, (2.4.25) and (2.4.26) imply $M_r(\check{A}) = \|\check{A}\|_{r\to r^*}$, and the conclusion follows from the first part and property (6) of Lemma 2.4.13. □

We are now ready to present the proof of Theorem 2.2.11.

Proof of Theorem 2.2.11 and Theorem 2.2.12. Recall the decomposition of $A = \bar{A}+\tilde{A}+\hat{A}$ in (2.4.2). By the triangle inequality, Proposition 2.4.1 and Proposition 2.4.8, under assumptions of Theorem 2.2.11 or Theorem 2.2.12, we have whp,

$$\Big|\|A\|_{r\to p} - \|\hat{A}\|_{r\to p}\Big| \le \|\bar{A}\|_{r\to p} + \|\tilde{A}\|_{r\to p} = o(b_n).$$

Together with Proposition 2.4.15, this completes the proof. □

2.5 Proof of Theorem 2.2.13

Throughout this section, denote by $\gamma = \frac{r-p}{rp}$ as in the statement of Theorem 2.2.13. Fix $\alpha \in \left(0, \min\left\{2, \frac{1}{\gamma}\right\}\right)$ and $1 < p < r < \infty$, define

$$\begin{aligned} \eta &\coloneqq \frac{1}{2}(1 - \alpha\gamma), \\ \zeta &\coloneqq \min\left\{\frac{1}{4}(1 - \alpha\gamma), \left(1 - \frac{\alpha}{2}\right)\frac{p-1}{r}\right\}. \end{aligned} \tag{2.5.1}$$

With these η and ζ, we adopt the same decomposition scheme of A as in (2.4.2). Also note that under the assumptions of Theorem 2.2.13, $0 < \zeta < \eta$.

The section is organized as follow. Section 2.5.1 contains some auxiliary results used in the derivation of upper and lower bounds on fluctuation of $r \to p$ norm of $\mu\mathbf{1}\mathbf{1}^\top + \hat{A} + \tilde{A}$, which are presented in Section 2.5.2 and Section 2.5.3 respectively. Then in Section 2.5.4 we completes the proof of Theorem 2.2.13 by showing that these two bounds match asymptotically and that adding back $\bar{A}$ leads to a negligible change in asymptotic fluctuation.

2.5.1 Auxiliary results

We start with four auxiliary lemmas. The first two concern the structures of matrix $\tilde{A} + \hat{A}$ and $\hat{A}$.

Lemma 2.5.1. *Suppose the assumptions of Theorem 2.2.13 hold and* $\hat{A} = (\hat{a}_{ij})_{i,j\in[n]}, \tilde{A} = (\tilde{a}_{ij})_{i,j\in[n]}$ *are as in* (2.4.2)*. Then whp, for all* $i, j, k \in [n]$,

$$\hat{a}_{ij}\tilde{a}_{ik} = 0.$$

Proof. By (2.5.1), (2.4.2), the assumption that (a_{ij}) are heavy-tailed with index α and Definition 2.2.3, the probability that there exists $i, j, k \in [n]$ such that $\hat{a}_{ij}\tilde{a}_{ik} \neq 0$ is upper bounded by

$$n\binom{n}{2}L\left(n^{\frac{2-\zeta}{\alpha}}\right)L\left(n^{\frac{1+\eta}{\alpha}}\right)\left(n^{\frac{2-\zeta}{\alpha}}\right)^{-\alpha}\left(n^{\frac{1+\eta}{\alpha}}\right)^{-\alpha}\le n^{(-\eta+\zeta)/2}\to 0.$$

The proof is complete. $\square$

Lemma 2.5.2. *Suppose the assumptions of Theorem 2.2.13 hold and $\hat{A}=(\hat{a}_{ij})_{i,j\in[n]}$ is as in (2.4.2). Fix $q\ge 0$. Then whp, for all $i\in[n]$,*

$$\left|(\hat{A}\mathbf{1})_i\right|^q=\sum_{j\in[n]}|\hat{a}_{ij}|^q. \tag{2.5.2}$$

Proof. Recalling the choice of ζ in (2.5.1), definition of $\hat{A}$ in (2.4.2) and (2.2.4), Lemma 2.3.3 (b) implies that whp, there is at most one nonzero entry on each row or column of $\hat{A}$. Say the nonzero entry on the i-th row of $\hat{A}$ is on the i^*-column. We have whp, for all $i\in[n]$,

$$\left|\left(\hat{A}\mathbf{1}\right)_i\right|^q=|a_{ii^*}|^q=\sum_{j\in[n]}|\hat{a}_{ij}|^q.$$

This completes the proof. $\square$

The third result identifies a typical property of heavy-tailed entries.

Lemma 2.5.3. *Let $(X_i)_{i\in[n(n+1)/2]}$ be i.i.d. copies of heavy-tailed random variable with index $\alpha\in(0,2)$. Fix $q\ge 0$ and $\beta\in(0,\frac{2}{\alpha})$. Denote by $\tilde{X}_i=X_i\mathbb{1}\left\{|X_i|>n^\beta\right\}$. Then whp,*

$$\sum_{i\in[n(n+1)/2]}\left|\tilde{X}_i\right|^q=O_\delta\left(n^{\max\left\{q\beta-\alpha\beta+2,\frac{2q}{\alpha}\right\}}\right).$$

Proof. By considering $|X_i|$, we may, without loss of generality, assume that X_i's are nonnegative. Fix any $\varepsilon>0$ and θ such that

$$\max\left\{\frac{\alpha\beta}{2},1-\varepsilon\right\}<\theta<1. \tag{2.5.3}$$

Consider the following decomposition of $\tilde{X}_i$:

$$\tilde{X}_i = \bar{X}_i + \hat{X}_i,$$

where

$$\bar{X}_i \coloneqq \tilde{X}_i \mathbb{1}\left\{\tilde{X}_i \leq n^{\frac{2\theta}{\alpha}}\right\}, \quad \hat{X}_i \coloneqq \tilde{X}_i \mathbb{1}\left\{\tilde{X}_i > n^{\frac{2\theta}{\alpha}}\right\}.$$

By Lemma 2.4.12, whp, there are $\Theta_\delta\left(n^{2-2\theta}\right)$ nonzero $\hat{X}_i$. By (2.2.2) and Lemma 2.2.2, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\max_{i\in[n(n+1)/2]} \hat{X}_i \geq n^{\frac{2}{\alpha}+\varepsilon}\right) \leq \mathbb{P}\left(\max_{i\in[n(n+1)/2]} X_i \geq n^{\frac{2}{\alpha}+\varepsilon}\right) = O_\delta\left(n^2 \cdot n^{-2-\alpha\varepsilon}\right) = o(1),$$

and hence $\max_{i\in[n(n+1)/2]} \hat{X}_i = O_\delta\left(n^{\frac{2}{\alpha}}\right)$. Therefore whp, for any $q \geq 0$,

$$\sum_{i\in[n(n+1)/2]} \hat{X}_i^q = O_\delta\left(n^{2-2\theta+\frac{2q}{\alpha}}\right) = O_\delta\left(n^{2\varepsilon+\frac{2q}{\alpha}}\right), \tag{2.5.4}$$

where we used (2.5.3) in the last step.

Recalling the convention $0^0 = 0$, we have

$$\mathbb{E}\left[\bar{X}_1^0\right] = \mathbb{P}\left(\bar{X}_1 \neq 0\right) = \mathbb{P}\left(n^\beta < \tilde{X}_1 \leq n^{\frac{2\theta}{\alpha}}\right) = O_\delta\left(n^{-\alpha\beta}\right).$$

This implies that

$$\mathbb{E}\left[\sum_{i\in[n(n+1)/2]} \bar{X}_i^0\right] = O_\delta\left(n^{2-\alpha\beta}\right), \quad \mathrm{Var}\left(\sum_{i\in[n(n+1)/2]} \bar{X}_i^0\right) \leq \frac{n(n+1)}{2}\mathbb{E}\left[\bar{X}_1^0\right] = O_\delta\left(n^{2-\alpha\beta}\right). \tag{2.5.5}$$

On the other hand, for $q > 0$, we have by (2.5.3),

$$\mathbb{E}\left[\sum_{i\in[n(n+1)/2]} \bar{X}_i^q\right] = \frac{n(n+1)}{2}\mathbb{E}\left[\bar{X}_1^q\right] \quad \le n^{\varepsilon/2+2}\int_{n^{\beta q}}^{n^{2\theta q/\alpha}} x^{-\frac{\alpha}{q}}\,dx = n^{2+\varepsilon+\max\left\{\beta(q-\alpha),\frac{2\theta}{\alpha}(q-\alpha)\right\}},$$

$$\mathrm{Var}\left(\sum_{i\in[n(n+1)/2]} \bar{X}_i^q\right) \le \frac{n(n+1)}{2}\mathbb{E}\left[\bar{X}_1^{2q}\right] \quad = n^{\varepsilon/2+2}\int_{n^{2q\beta}}^{n^{4\theta q/\alpha}} x^{-\frac{\alpha}{2q}}\,dx = n^{2+\varepsilon+\max\left\{\beta(2q-\alpha),\frac{2\theta}{\alpha}(2q-\alpha)\right\}}. \tag{2.5.6}$$

Combining (2.5.5), (2.5.6) with Markov's inequality and (2.5.3), we have whp, for all $q \ge 0$,

$$\sum_{i\in[n(n+1)/2]} \bar{X}_i^q = O_\delta\left(n^{2+\varepsilon+\max\left\{\beta(q-\alpha),\frac{2\theta}{\alpha}(q-\alpha)\right\}}\right). \tag{2.5.7}$$

Combining $\theta < 1$ from (2.5.3), (2.5.4) and (2.5.7), we have whp, for any $q \ge 0$,

$$\sum_{i\in[n(n+1)/2]} \tilde{X}_i^q = O_\delta\left(n^{\max\left\{2\varepsilon+\frac{2q}{\alpha},2+\varepsilon+\beta(q-\alpha)\right\}}\right).$$

Recalling that $\varepsilon > 0$ is arbitrary, the proof is complete. □

The fourth lemma, which we will use multiple times in the rest of the section, is an immediate consequence of Lemma 2.4.11 and Lemma 2.5.3,

Lemma 2.5.4. *Suppose the assumptions of Theorem 2.2.13 hold and recall that $\gamma = \frac{r-p}{rp}$. With b_n defined in (2.2.3) and ζ in (2.5.1), we have whp,*

$$\sum_{i,j\in[n]} |a_{ij}|^{\frac{1}{\gamma}}\mathbb{1}\left(|a_{ij}| \le n^{\frac{2-\zeta}{\alpha}}\right) = o\left(b_n^{\frac{1}{\gamma}}\right), \tag{2.5.8}$$

$$\sum_{i,j\in[n]} |a_{ij}|^{\frac{1}{\gamma}} = o\left(n^{1+\frac{1}{\gamma}}\right). \tag{2.5.9}$$

Proof. Note that (2.2.12) implies $\alpha < \frac{1}{\gamma}$. Hence, Lemma 2.4.11 with $\{X_i : i \in [n(n+1)/2]\} = \{a_{ij} : 1 \le i \le j \le n\}$, $q = \frac{1}{\gamma}$ and $\beta = \frac{2-\zeta}{\alpha}$ implies that whp,

$$\sum_{i,j\in[n]} |a_{ij}|^{\frac{1}{\gamma}}\mathbb{1}\left(|a_{ij}| \le n^{\frac{2-\zeta}{\alpha}}\right) = O_\delta\left(n^{\frac{2}{\alpha\gamma}-\left(\frac{1}{\alpha\gamma}-1\right)\zeta}\right) = o\left(b_n^{\frac{1}{\gamma}}\right),$$

where we use (2.2.4) in the last equality. This proves (2.5.8).

By Lemma 2.5.3, with $\{X_i : i \in [n(n+1)/2]\} = \{a_{ij} : 1 \le i \le j \le n\}$, $q = \frac{1}{\gamma} > \alpha$ and $\beta = \frac{2-\zeta}{\alpha}$, we have whp,

$$\sum_{i,j\in[n]} |a_{ij}|^{\frac{1}{\gamma}} \mathbb{1}\left(|a_{ij}| > n^{\frac{2-\zeta}{\alpha}}\right) = O_\delta\left(n^{\max\left\{\frac{2}{\alpha\gamma} - \left(\frac{1}{\alpha\gamma}-1\right)\zeta,\, \frac{2}{\alpha\gamma}\right\}}\right) = O_\delta\left(n^{\frac{2}{\alpha\gamma}}\right) = o\left(n^{1+\frac{1}{\gamma}}\right),$$

where we use (2.2.12) in the second and third equalities. Together with (2.2.4) and (2.5.8), this completes the proof of (2.5.9). □

2.5.2 Upper bound

The goal of this section is the following lemma.

Proposition 2.5.5. *Under the assumptions of Theorem 2.2.13, recall that $\gamma = \frac{r-p}{rp}$, we have whp*

$$\|\mu\mathbf{1}\mathbf{1}^\top + \hat{A} + \widetilde{A}\|_{r\to p} \le n^{1+\gamma}|\mu| + \gamma|\mu|^{1-\frac{1}{\gamma}} n^{\gamma-\frac{1}{\gamma}} \sum_{i,j=1}^{n} |a_{ij}|^{\frac{1}{\gamma}} + o\left(n^{\gamma-\frac{1}{\gamma}} b_n^{\frac{1}{\gamma}}\right).$$

Proof. In this proof, we will use C to denote a positive constant that may depend on α, r, p, and μ and whose value may vary from one occurrence to another.

Since $\|\mu\mathbf{1}\mathbf{1}^\top + \tilde{A} + \hat{A}\|_{r\to p} \le \||\mu|\mathbf{1}\mathbf{1}^\top + |\hat{A}| + |\tilde{A}|\|_{r\to p}$, it suffices to prove the lemma in the case that $\mu > 0$ and that $\hat{A}, \tilde{A}$ contain only nonnegative entries.

By Lemma 2.4.2, we have

$$\|\mu\mathbf{1}\mathbf{1}^\top + \tilde{A} + \hat{A}\|_{r\to p} \le \mathcal{U}(A) := \left(\mathbf{1}^\top\left((\mu\mathbf{1}\mathbf{1}^\top + \tilde{A} + \hat{A})\mathbf{1}\right)^{\frac{1}{\gamma}}\right)^{\gamma}. \tag{2.5.10}$$

Hence it suffices to show that whp,

$$\mathcal{U}(A) \le n^{1+\gamma}|\mu| + \gamma|\mu|^{1-\frac{1}{\gamma}} n^{\gamma-\frac{1}{\gamma}} \sum_{i,j=1}^{n} |a_{ij}|^{\frac{1}{\gamma}} + o\left(n^{\gamma-\frac{1}{\gamma}} b_n^{\frac{1}{\gamma}}\right). \tag{2.5.11}$$

To show this, we first claim that, we have whp,

$$\left(\left(\mu\mathbf{1}\mathbf{1}^\top+\tilde{A}+\hat{A}\right)\mathbf{1}\right)^{\frac{rp}{r-p}}\le(n\mu)^{\frac{1}{\gamma}}\mathbf{1}+\left(\hat{A}\mathbf{1}\right)^{\frac{1}{\gamma}}+Cn\left(\hat{A}\mathbf{1}\right)^{\frac{1}{\gamma}-1}+Cn^{\frac{1}{\gamma}-1}\tilde{A}\mathbf{1}+C\left(\tilde{A}\mathbf{1}\right)^{\frac{1}{\gamma}},\tag{2.5.12}$$

where the inequality stands for entrywise comparison. Let $i\in[n]$ be fixed. We prove (2.5.12) by considering two cases.

(1) Suppose $(\hat{A}\mathbf{1})_i\neq 0$. Then by Lemma 2.5.1, we have whp, $(\tilde{A}\mathbf{1})_i=0$. By the definition of $\hat{A}$ in (2.4.2) and the choice of ζ in (2.5.1), it follows that

$$(\hat{A}\mathbf{1})_i\ge n^{\frac{2-\zeta}{\alpha}}\gg n.$$

Using the elementary inequality $(x+y)^q\le x^q+Cx^{q-1}y$, $\forall x,q>0$, $0\le y\le\frac{x}{2}$, we have whp,

$$\left(\left(\mu\mathbf{1}\mathbf{1}^\top+\tilde{A}+\hat{A}\right)\mathbf{1}\right)_i^{\frac{1}{\gamma}}=\left(n\mu+\left(\hat{A}\mathbf{1}\right)_i\right)^{\frac{1}{\gamma}}\le\left(\hat{A}\mathbf{1}\right)_i^{\frac{1}{\gamma}}+Cn\left(\hat{A}\mathbf{1}\right)^{\frac{1}{\gamma}-1}.\tag{2.5.13}$$

(2) Suppose $(\hat{A}\mathbf{1})_i=0$. Then by the elementary inequality $(x+y)^q\le x^q+Cx^{q-1}y+Cy^q$, $\forall\, x,q>0,y\ge 0$, we have

$$\begin{aligned}\left(\left(\mu\mathbf{1}\mathbf{1}^\top+\tilde{A}+\hat{A}\right)\mathbf{1}\right)_i^{\frac{1}{\gamma}}&=\left(n\mu+\left(\tilde{A}\mathbf{1}\right)_i\right)^{\frac{1}{\gamma}}\\&\le(n\mu)^{\frac{1}{\gamma}}+Cn^{\frac{1}{\gamma}-1}\left(\tilde{A}\mathbf{1}\right)_i+C\left(\tilde{A}\mathbf{1}\right)_i^{\frac{1}{\gamma}}.\end{aligned}\tag{2.5.14}$$

The claim (2.5.12) is complete upon combining (2.5.13) and (2.5.14).

By (2.4.11), (2.5.2) and (2.5.12), we have whp,

$$\mathbf{1}^\top\left(\left(\mu\mathbf{1}\mathbf{1}^\top+\tilde{A}+\hat{A}\right)\mathbf{1}\right)^{\frac{1}{\gamma}}\le n^{\frac{1}{\gamma}+1}\mu^{\frac{1}{\gamma}}+\sum_{i,j\in[n]}\hat{a}_{ij}^{\frac{1}{\gamma}}+\sum_{i=1}^{3}R_i,\tag{2.5.15}$$

where

$$R_1 := Cn \sum_{i,j\in[n]} \hat{a}_{ij}^{\frac{1}{\gamma}-1},$$

$$R_2 := Cn^{\frac{1}{\gamma}-1} \sum_{i,j\in[n]} \tilde{a}_{ij},$$

$$R_3 := C \sum_{i,j\in[n]} \tilde{a}_{ij}^{\frac{1}{\gamma}}.$$

We next claim that whp,

$$R_i = o\left(b_n^{\frac{1}{\gamma}}\right), \quad i = 1, 2, 3. \tag{2.5.16}$$

The verification is as follow.

(1) By Lemma 2.5.3 with $\{X_i : i \in [n(n+1)/2]\} = \{a_{ij} : 1 \leq i \leq j \leq n\}$, $q = \frac{1}{\gamma} - 1$ and $\beta = \frac{2-\zeta}{\alpha}$, we have whp,

$$\begin{aligned}\sum_{i,j\in[n]} \hat{a}_{ij}^{\frac{1}{\gamma}-1} &= \sum_{i,j\in[n]} a_{ij}^{\frac{1}{\gamma}-1} \mathbb{1}\left(a_{ij} \geq n^{\frac{2-\zeta}{\alpha}}\right) \\ &= O_\delta\left(n^{\max\left\{\left(\frac{1}{\gamma}-1\right)\frac{2-\zeta}{\alpha}+\zeta,\, \left(\frac{1}{\gamma}-1\right)\frac{2}{\alpha}\right\}}\right) \\ &= O_\delta\left(n^{\left(\frac{1}{\gamma}-1\right)\frac{2}{\alpha}+\zeta}\right).\end{aligned}$$

Together with the choice of ζ in (2.5.1), this implies that whp,

$$R_1 = O_\delta\left(n^{\frac{2}{\alpha\gamma}+1-\frac{2}{\alpha}+\zeta}\right) = o\left(b_n^{\frac{1}{\gamma}}\right).$$

(2) By Lemma 2.5.3 with $\{X_i : i \in [n(n+1)/2]\} = \{a_{ij} : 1 \leq i \leq j \leq n\}$, $q = 1$ and

$\beta = \frac{1+\eta}{\alpha}$, we have whp,

$$\begin{aligned}\sum_{i,j\in[n]} \tilde{a}_{ij} &\leq \sum_{i,j\in[n]} a_{ij}\mathbb{1}\left(a_{ij} \geq n^{\frac{1+\eta}{\alpha}}\right) \\ &= O_\delta\left(n^{\max\left\{\frac{1}{\alpha}+1+\eta\left(\frac{1}{\alpha}-1\right),\,\frac{2}{\alpha}\right\}}\right) \\ &= O_\delta\left(n^{\frac{1}{\alpha}+1}\right),\end{aligned}$$

where we use (2.2.12) in the last equality. Using (2.2.12) again, this implies that whp,

$$R_2 = O_\delta\left(n^{\frac{1}{\gamma}+\frac{1}{\alpha}}\right) = o\left(b_n^{\frac{1}{\gamma}}\right).$$

(3) By (2.5.8), we have whp

$$R_3 = C\sum_{i,j\in[n]} a_{ij}\mathbb{1}\left(a_{ij} \leq n^{\frac{2-\zeta}{\alpha}}\right) = o\left(b_n^{\frac{1}{\gamma}}\right).$$

Combining (2.5.15), (2.5.16), (2.4.2) and (2.5.8), we have whp,

$$\begin{aligned}\mathbf{1}^\top\left(\left(\mu\mathbf{1}\mathbf{1}^\top + \tilde{A} + \hat{A}\right)\mathbf{1}\right)^{\frac{1}{\gamma}} &\leq n^{\frac{1}{\gamma}+1}\mu^{\frac{1}{\gamma}} + \sum_{i,j\in[n]} \hat{a}_{ij}^{\frac{1}{\gamma}} + o\left(b_n^{\frac{1}{\gamma}}\right) \\ &= n^{\frac{1}{\gamma}+1}\mu^{\frac{1}{\gamma}} + \sum_{i,j\in[n]} a_{ij}^{\frac{1}{\gamma}} + o\left(b_n^{\frac{1}{\gamma}}\right).\end{aligned}$$

Together with (2.5.9), the definition of $\mathcal{U}(A)$ in (2.5.10), the fact that $\gamma = \frac{r-p}{rp} \in (0,1]$, and the elementary inequality $(x+y)^q \leq x^q + qx^{q-1}y$ for all $x > 0$, $q \in (0,1]$ and $y \geq 0$, this implies that whp,

$$\begin{aligned}\mathcal{U}(A) &\leq \left(n^{\frac{1}{\gamma}+1}\mu^{\frac{1}{\gamma}} + \sum_{i,j\in[n]} a_{ij}^{\frac{1}{\gamma}} + o\left(b_n^{\frac{1}{\gamma}}\right)\right)^\gamma \\ &\leq n^{1+\gamma}\mu + \gamma n^{\gamma-\frac{1}{\gamma}}\mu^{1-\frac{1}{\gamma}}\sum_{i,j\in[n]} a_{ij}^{\frac{1}{\gamma}} + o\left(n^{\gamma-\frac{1}{\gamma}}b_n^{\frac{1}{\gamma}}\right).\end{aligned}$$

This establishes (2.5.11), hence completes the proof of the lemma. □

2.5.3 Lower bound

The goal of this section is the following proposition.

Proposition 2.5.6. *Under the assumptions of Theorem 2.2.13, recall that* $\gamma = \frac{r-p}{rp}$*, we have whp*

$$\|\mu\mathbf{1}\mathbf{1}^\top + \tilde{A} + \hat{A}\|_{r\to p} \geq n^{1+\gamma}|\mu| + \gamma|\mu|^{1-\frac{1}{\gamma}}n^{\gamma-\frac{1}{\gamma}}\sum_{i,j=1}^{n}|a_{ij}|^{\frac{1}{\gamma}} + o\left(n^{\gamma-\frac{1}{\gamma}}b_n^{\frac{1}{\gamma}}\right).$$

We start by providing some intuition on the choice of $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}$ in (2.5.21) below. Although the full object of interest is $\mu\mathbf{1}\mathbf{1}^\top + \tilde{A} + \hat{A}$, for intuition consider the modified matrix $\hat{A}_{\mu,n} := |\mu|\mathbf{1}\mathbf{1}^\top + |\hat{A}|$, where we only focus on the largest entries and the centering.

Recall the definitions of ψ_q in (2.1.2), and define the operator S by

$$S\boldsymbol{x} := \Psi_{r^*}(\hat{A}_{\mu,n}^\top\Psi_p(\hat{A}_{\mu,n}\boldsymbol{x})).$$

By the nonlinear power method, if $\boldsymbol{v}^{(0)}$ has all positive entries, and we define $\boldsymbol{v}^{(k+1)} := S\boldsymbol{v}^{(k)}$, then it is known that (see [67, Section 4.1])

$$\lim_{k\to\infty}\frac{\left\|\hat{A}_{\mu,n}\boldsymbol{v}^{(k)}\right\|_p}{\|\boldsymbol{v}^{(k)}\|_r} = \left\|\hat{A}_{\mu,n}\right\|_{r\to p}. \tag{2.5.17}$$

We make two simple observations. First, by the definition of $\hat{A}$ in (2.4.2) and the choice of ζ in (2.5.1), we have the following property,

$$\text{for each } i \in [n] \text{, we either have } (|\hat{A}|\mathbf{1})_i = 0 \text{ or } \mu n \ll (|\hat{A}|\mathbf{1})_i. \tag{2.5.18}$$

Secondly, by the symmetry of $\hat{A}$ and the fact that, whp, each row of $\hat{A}$ has at most one nonzero entry (as a consequence of Lemma 2.3.3 (b)), for all $q > 0$, we have

$$|\hat{A}|(|\hat{A}|\mathbf{1})^q = (|\hat{A}|\mathbf{1})^{q+1}. \tag{2.5.19}$$

Starting from an initial condition of the form $\boldsymbol{v}^{[k]} = (\mu n)^k + \left(|\hat{A}|\mathbf{1}\right)^k$ for some $k \geq 0$, and then using (2.5.18) and (2.5.19) repeatedly, we have

$$S\boldsymbol{v}^{[k]} \approx (n|\mu|)^{\frac{p}{r-1}+\frac{(p-1)k}{r-1}} + \left(|\hat{A}|\mathbf{1}\right)^{\frac{p}{r-1}+\frac{(p-1)k}{r-1}}.$$

Therefore, the operation of S on this class of vectors can be approximately regarded as an operator on the exponent $k \mapsto \frac{p}{r-1} + \frac{(p-1)k}{r-1}$. It is easy to see that the iteration of such a map converges to the unique fixed point of the form $\boldsymbol{v}^{[k_*]}$, where $k_* = \frac{p}{r-p}$. Hence, as indicated in (2.5.17), it is natural to choose the near optimizer to be the limiting vector of the nonlinear power method, that is

$$\hat{\boldsymbol{x}} := (n|\mu|)^{\frac{p}{r-p}} + \left(|\hat{A}|\mathbf{1}\right)^{\frac{p}{r-p}}.$$

The duality formulation suggests that $\hat{\boldsymbol{y}}$ should be set equal to

$$\operatorname*{argmax}_{\boldsymbol{y}} \left\langle \boldsymbol{y}, \hat{A}_{\mu,n}\hat{\boldsymbol{x}} \right\rangle, \tag{2.5.20}$$

which can be explicitly determined using the equality condition of Hölder's inequality. But in fact the $\hat{\boldsymbol{y}}$ we choose in (2.5.21) is a suitable perturbation of the actual optimizer in (2.5.20) that facilitates subsequent calculations and can be shown to be close to the actual optimizer using (2.5.18).

Motivated by the above intuition, we define $\hat{\boldsymbol{x}} = (\hat{x}_1, \ldots, \hat{x}_n), \hat{\boldsymbol{y}} = (\hat{y}_1, \ldots, \hat{y}_n)$ such that, for $i \in [n]$,

$$\begin{aligned}\hat{x}_i &:= (n|\mu|)^{\frac{p}{r-p}} \operatorname{sgn}(\mu) + \sum_{j:\, j<i} |\hat{a}_{ij}|^{\frac{p}{r-p}} + \sum_{j:\, j>i} \psi_{\frac{r}{r-p}}(\hat{a}_{ij}), \\ \hat{y}_i &:= (n|\mu|)^{\frac{r(p-1)}{r-p}} + \sum_{j:\, j<i} |\hat{a}_{ij}|^{\frac{r(p-1)}{r-p}} + \sum_{j:\, j>i} \psi_{\frac{p(r-1)}{r-p}}(\hat{a}_{ij}),\end{aligned} \tag{2.5.21}$$

where $\psi_{\frac{r}{r-p}}, \psi_{\frac{p(r-1)}{r-p}}$ are the functionals defined in (2.1.2).

Consider

$$\mathcal{L}(A) := \frac{\mathcal{L}_1(A)}{\mathcal{L}_2(A)}, \tag{2.5.22}$$

where

$$\begin{aligned}\mathcal{L}_1(A) &:= \hat{\boldsymbol{y}}^\top \left(\mu \mathbf{1}\mathbf{1}^\top + \tilde{A} + \hat{A}\right) \hat{\boldsymbol{x}}, \\ \mathcal{L}_2(A) &:= \|\hat{\boldsymbol{y}}\|_{p^*} \|\hat{\boldsymbol{x}}\|_r.\end{aligned} \tag{2.5.23}$$

By the dual formulation (2.2.9) of the $r \to p$ norm, we have

$$\mathcal{L}(A) \le \|\mu \mathbf{1}\mathbf{1}^\top + \tilde{A} + \hat{A}\|_{r\to p}.$$

To prove Proposition 2.5.6, it suffices to establish the following lemma.

Lemma 2.5.7. *Under the assumptions of Theorem 2.2.13, recall that $\gamma = \frac{r-p}{rp}$, we have whp,*

$$\mathcal{L}(A) \ge n^{1+\gamma}|\mu| + \gamma |\mu|^{1-\frac{1}{\gamma}} n^{\gamma - \frac{1}{\gamma}} \sum_{i,j=1}^{n} |a_{ij}|^{\frac{1}{\gamma}} + o\left(n^{\gamma-\frac{1}{\gamma}} b_n^{\frac{1}{\gamma}}\right).$$

Proof. In this proof, we will use C to denote a positive constant that may depend on α, r, p, and μ and whose value may vary from one occurrence to another.

Note that since $\|\mu \mathbf{1}\mathbf{1}^\top + \tilde{A} + \hat{A}\|_{r\to p} = \| - \mu \mathbf{1}\mathbf{1}^\top - \tilde{A} - \hat{A}\|_{r\to p}$, we may assume $\mu > 0$ for the rest of the proof. We proceed in three steps

Step 1: Estimation of $\mathcal{L}_1(A)$. Expanding the definition of $\mathcal{L}_1$ in (2.5.23), we see that

$$\mathcal{L}_1(A) = \mu \mathbf{1}^\top \hat{\boldsymbol{x}} \hat{\boldsymbol{y}}^\top \mathbf{1} + \hat{\boldsymbol{y}}^\top \tilde{A} \hat{\boldsymbol{x}} + \hat{\boldsymbol{y}}^\top \hat{A} \hat{\boldsymbol{x}}. \tag{2.5.24}$$

By the definitions of $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}$ in (2.5.21), we have whp,

$$\begin{aligned} \mu \mathbf{1}^\top \hat{\boldsymbol{x}} \hat{\boldsymbol{y}}^\top \mathbf{1} \geq n^{\frac{rp}{r-p}+1} \mu^{\frac{rp}{r-p}} - Cn^{\frac{r}{r-p}} \sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{rp-r}{r-p}} - Cn^{\frac{rp-p}{r-p}} \sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{p}{r-p}} \\ - C \sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{rp-r}{r-p}} \sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{p}{r-p}}. \end{aligned} \tag{2.5.25}$$

Furthermore, (2.5.21) and Lemma 2.5.1 imply that whp,

$$\hat{\boldsymbol{y}}^\top \tilde{A} \hat{\boldsymbol{x}} = (n|\mu|)^{\frac{rp}{r-p}-1} \operatorname{sgn}(\mu) \mathbf{1}^\top \tilde{A} \mathbf{1} \geq -Cn^{\frac{rp}{r-p}-1} \sum_{i,j\in[n]} |\tilde{a}_{ij}|. \tag{2.5.26}$$

For the last term in (2.5.24), we first write it as

$$\hat{\boldsymbol{y}}^\top \hat{A} \hat{\boldsymbol{x}} = \sum_{i,j\in[n]} \hat{a}_{ij} \hat{x}_i \hat{y}_j,$$

and analyze the contribution of each summand by considering the following two cases:

1. When $\hat{a}_{ij} = 0$, the contribution is trivially 0.

2. Suppose $\hat{a}_{ij} \neq 0$. By Lemma 2.3.3 item (a), (2.5.1) and (2.4.2), we have whp, $i \neq j$. We only consider the case $i < j$, the case $j < i$ can be analyzed in the same way with the same conclusion. By Lemma 2.3.3 item (b), $\hat{a}_{ij}$ is the only nonzero entry on i-th row and j-th column. Together with the symmetry of $\hat{A}$, we have whp,

   $$\hat{x}_i = (n\,|\mu|)^{\frac{p}{r-p}} \operatorname{sgn}(\mu) + |\hat{a}_{ij}|^{\frac{p}{r-p}} \operatorname{sgn}(\hat{a}_{ij}), \qquad \hat{y}_j = (n\,|\mu|)^{\frac{r(p-1)}{r-p}} + |\hat{a}_{ij}|^{\frac{r(p-1)}{r-p}}.$$

   Therefore,

   $$\hat{a}_{ij} x_i y_j \geq |\hat{a}_{ij}|^{\frac{rp}{r-p}} - Cn^{\frac{p}{r-p}} |\hat{a}_{ij}|^{\frac{p(r-1)}{r-p}} - Cn^{\frac{r(p-1)}{r-p}} |\hat{a}_{ij}|^{\frac{r}{r-p}} - Cn^{\frac{rp}{r-p}-1} |\hat{a}_{ij}|.$$

Combining the above two cases, we have whp,

$$\begin{aligned}\hat{\boldsymbol{y}}^\top \hat{A}\hat{\boldsymbol{x}} \geq \sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{rp}{r-p}} - Cn^{\frac{p}{r-p}} \sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{rp-p}{r-p}} - Cn^{\frac{rp-r}{r-p}} \sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{r}{r-p}} \\ - Cn^{\frac{rp}{r-p}-1} \sum_{i,j\in[n]} |\hat{a}_{ij}|.\end{aligned} \tag{2.5.27}$$

Inserting (2.5.25), (2.5.26) and (2.5.27) into (2.5.24) and recall that $\gamma = \frac{r-p}{rp}$, we have whp,

$$\mathcal{L}_1(A) \geq n^{\frac{1}{\gamma}+1}\mu^{\frac{1}{\gamma}} + \sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{1}{\gamma}} - \sum_{i=1}^{7} R_i', \tag{2.5.28}$$

where

$$\begin{aligned}
R_1' &:= Cn^{\frac{r}{r-p}} \sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{rp-r}{r-p}}, \quad R_2' := Cn^{\frac{rp-p}{r-p}} \sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{p}{r-p}}, \\
R_3' &:= Cn^{\frac{p}{r-p}} \sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{rp-p}{r-p}}, \quad R_4' := Cn^{\frac{rp-r}{r-p}} \sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{r}{r-p}}, \\
R_5' &:= (n\mu)^{\frac{rp}{r-p}-1} \sum_{i,j\in[n]} |\hat{a}_{ij}|, \quad R_6' := (n\mu)^{\frac{rp}{r-p}-1} \sum_{i,j\in[n]} |\tilde{a}_{ij}|,
\end{aligned}$$

and

$$R_7' := C \sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{rp-r}{r-p}} \sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{p}{r-p}}.$$

We now claim that whp,

$$R_i' = o\left(b_n^{\frac{1}{\gamma}}\right), \quad 1 \leq i \leq 7. \tag{2.5.29}$$

This can be verified as follows:

(1) By the definition of $\hat{a}_{ij}$ in (2.4.2) and Lemma 2.5.3 with $q = \frac{rp-r}{r-p}$ and $\beta = \frac{2-\zeta}{\alpha}$, we

have whp

$$\sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{rp-r}{r-p}} = \sum_{i,j\in[n]} |a_{ij}|^{\frac{rp-r}{r-p}} \mathbb{1}\left\{|a_{ij}| > n^{\frac{2-\zeta}{\alpha}}\right\} = O_\delta\left(n^{\frac{2}{\alpha}\frac{rp-r}{r-p}+\zeta}\right).$$

Hence, the choice of ζ in (2.5.1), $\alpha < 2$ and (2.2.4) imply that whp,

$$R_1' = O_\delta\left(n^{\frac{2}{\alpha}\frac{rp-r}{r-p}+\frac{r}{r-p}+\zeta}\right) = O_\delta\left(n^{\frac{2}{\alpha\gamma}+\left(2-\frac{2}{\alpha}-\frac{\alpha}{2}\right)\frac{p-1}{r}}\right) = o\left(b_n^{\frac{1}{\gamma}}\right).$$

The estimation of R_2', R_3', R_4', R_5' and R_7' are similar.

(2) By the definition of $\tilde{a}_{ij}$ in (2.4.2) and Lemma 2.5.3 with $q = 1$ and $\beta = \frac{1+\eta}{\alpha}$, we have whp,

$$\sum_{i,j\in[n]} |\tilde{a}_{ij}| = \sum_{i,j\in[n]} |a_{ij}| \mathbb{1}\left\{|a_{ij}| \geq n^{\frac{1+\eta}{\alpha}}\right\} = O_\delta\left(n^{\frac{1}{\alpha}+1}\right).$$

Hence, (2.2.12), together with (2.2.4), implies that whp,

$$R_6' = O_\delta\left(n^{\frac{1}{\alpha}+1}\right) = o\left(b_n^{\frac{1}{\gamma}}\right).$$

Combining (2.5.8), (2.5.28) and (2.5.29), we have established that whp,

$$\mathcal{L}_1(A) \geq n^{\frac{1}{\gamma}+1}\mu^{\frac{1}{\gamma}} + \sum_{i,j\in[n]} |a_{ij}|^{\frac{1}{\gamma}} + o\left(b_n^{\frac{1}{\gamma}}\right). \tag{2.5.30}$$

Step 2: Estimation of $\mathcal{L}_2(A)$. Fix $i \in [n]$. We estimate the r-th power of $\hat{x}_i$, considering two cases:

1. Suppose there exists $i^* \in [n]$ such that $\hat{a}_{ii^*} \neq 0$. Then by (2.4.2) and the choice of ζ in (2.5.1), we have whp,

$$|\hat{a}_{ii^*}|^{\frac{p}{r-p}} \geq (n\mu)^{\frac{p}{r-p}}. \tag{2.5.31}$$

By Lemma 2.3.3 (b), (2.5.1) and (2.4.2), we have whp, $\hat{a}_{ij} = 0$ for all $j \neq i^*$. Together with definition of $\hat{x}$ in (2.5.21), (2.5.31) and the elementary inequality $(x+y)^q \leq x^q + Cx^{q-1}y$ for all $x, q > 0$ and $0 \leq y \leq x$, this implies that whp,

$$|\hat{x}_i|^r \leq \sum_{j\in[n]} |\hat{a}_{ij}|^{\frac{1}{\gamma}} + Cn^{\frac{p}{r-p}} \sum_{j\in[n]} |\hat{a}_{ij}|^{\frac{p(r-1)}{r-p}}. \tag{2.5.32}$$

2. Suppose $\hat{a}_{ij} = 0$ for all $j \in [n]$. Then by definition of $\hat{x}$ in (2.5.21), we have

$$|\hat{x}_i|^r = (n\mu)^{\frac{1}{\gamma}}. \tag{2.5.33}$$

Combining (2.5.32) and (2.5.33), we have whp,

$$\|\hat{\boldsymbol{x}}\|_r^r \leq n^{\frac{1}{\gamma}+1}\mu^{\frac{1}{\gamma}} + \sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{1}{\gamma}} + Cn^{\frac{p}{r-p}} \sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{p(r-1)}{r-p}}. \tag{2.5.34}$$

By Lemma 2.5.3 with $q = \frac{p(r-1)}{r-p}$ and $\beta = \frac{2-\zeta}{\alpha}$, the choice of ζ in (2.5.1) and (2.2.4), we have whp,

$$\begin{aligned}
n^{\frac{p}{r-p}} \sum_{i,j\in[n]} |\hat{a}_{ij}|^{\frac{p(r-1)}{r-p}} &= n^{\frac{p}{r-p}} \sum_{i,j\in[n]} |a_{ij}|^{\frac{p(r-1)}{r-p}} \mathbb{1}\left\{|a_{ij}| > n^{\frac{2-\zeta}{\alpha}}\right\} \\
&= O_\delta\left(n^{\frac{2}{\alpha}\frac{p(r-1)}{r-p}+\frac{p}{r-p}+\zeta}\right) \\
&= O_\delta\left(n^{\frac{2}{\alpha\gamma}+\left(2-\frac{2}{\alpha}-\frac{\alpha}{2}\right)\frac{p-1}{r}}\right) \\
&= o\left(b_n^{\frac{1}{\gamma}}\right).
\end{aligned}$$

Together with (2.5.8) and (2.5.34), this implies that whp,

$$\|\hat{\boldsymbol{x}}\|_r^r \leq n^{\frac{1}{\gamma}+1}\mu^{\frac{1}{\gamma}} + \sum_{i,j\in[n]} |a_{ij}|^{\frac{1}{\gamma}} + o\left(b_n^{\frac{1}{\gamma}}\right).$$

Similarly, we have whp,

$$\|\hat{\boldsymbol{y}}\|_{p^*}^{p^*} \leq n^{\frac{1}{\gamma}+1}\mu^{\frac{1}{\gamma}} + \sum_{i,j\in[n]} |a_{ij}|^{\frac{1}{\gamma}} + o\left(b_n^{\frac{1}{\gamma}}\right).$$

Together with (2.5.9), the definition of $\mathcal{L}_2(A)$ in (2.5.23), the elementary inequality $(x+y)^q \geq x^q + qx^{q-1}y$ for all $x > 0, q \leq 0$ and $y \geq 0$, and the fact that $-\frac{1}{r} - \frac{1}{p^*} = \gamma - 1 \leq 0$, the last two displays imply that whp, we have

$$\begin{aligned} \frac{1}{\mathcal{L}_2(A)} &\geq \left(n^{\frac{1}{\gamma}+1}\mu^{\frac{1}{\gamma}} + \sum_{i,j\in[n]} |a_{ij}|^{\frac{1}{\gamma}} + o\left(b_n^{\frac{1}{\gamma}}\right)\right)^{\gamma-1} \\ &\geq n^{-\frac{1}{\gamma}+\gamma}\mu^{-\frac{1}{\gamma}+1} + (\gamma-1)\, n^{-\frac{2}{\gamma}+\gamma-1}\mu^{-\frac{2}{\gamma}+1} \sum_{i,j\in[n]} |a_{ij}|^{\frac{1}{\gamma}} + o\left(n^{-\frac{2}{\gamma}+\gamma-1} b_n^{\frac{1}{\gamma}}\right). \end{aligned} \tag{2.5.35}$$

The proof is complete on combining (2.5.9), (2.5.30) and (5.1.1) with the definition of $\mathcal{L}(A)$ in (2.5.22). □

2.5.4 Asymptotic equivalence of upper and lower bounds

Proof of Theorem 2.2.13. Comparing Proposition 2.5.5 and Proposition 2.5.6, we have shown that

$$\|\mu\mathbf{1}\mathbf{1}^\top + \tilde{A} + \hat{A}\|_{r\to p} = n^{1+\gamma}|\mu| + \gamma n^{\gamma-\frac{1}{\gamma}}|\mu|^{1-\frac{1}{\gamma}} \sum_{i,j\in[n]} |a_{ij}|^{\frac{1}{\gamma}} + o\left(n^{\gamma-\frac{1}{\gamma}} b_n^{\frac{1}{\gamma}}\right). \tag{2.5.36}$$

Note that condition (2.2.12) implies (2.2.11), and in particular, (2.4.7) holds whp. Combining (2.4.7) and (2.2.12), we have whp,

$$\|\bar{A}\|_{r\to p} = o\left(n^{\gamma-\frac{1}{\gamma}} b_n^{\frac{1}{\gamma}}\right). \tag{2.5.37}$$

Combining (2.5.36) and (2.5.37), we have whp

$$\|\mu\mathbf{1}\mathbf{1}^\top + A\|_{r\to p} = n^{1+\gamma}|\mu| + \gamma n^{\gamma-\frac{1}{\gamma}}|\mu|^{1-\frac{1}{\gamma}} \sum_{i,j\in[n]} |a_{ij}|^{\frac{1}{\gamma}} + o\left(n^{\gamma-\frac{1}{\gamma}} b_n^{\frac{1}{\gamma}}\right). \tag{2.5.38}$$

Finally, by [70, Theorem 3.8.2] and (2.2.12),

$$b_n^{-\frac{1}{\gamma}} \sum_{i,j\in[n]} |a_{ij}|^{\frac{1}{\gamma}} \xrightarrow{d} Y_{\alpha\gamma}, \tag{2.5.39}$$

where $Y_{\alpha\gamma}$ is an $\alpha\gamma$-stable random variable. The proof is complete by combining (2.5.38) and (2.5.39). $\square$

2.6 Proofs related to the ℓ_r-Levy-Grothendieck problem

2.6.1 Proof of Theorem 2.2.5

Proof of Theorem 2.2.5. Let $a_* := \max_{i,j} |a_{i,j}|$ and let $i_*, j_* \in [n]$ be the (random) location of the maximal entry of A, that is $a_* = |a_{i_*,j_*}|$. From (2.3.1), we have that $a_* = \Theta_\delta(b_n)$ whp. By Lemma 2.3.3 (a), this implies $i_* \neq j_*$ whp. Next, set $\boldsymbol{x} := 2^{-1/r}(\boldsymbol{e}_{i_*} + \operatorname{sgn}(a_*)\, \boldsymbol{e}_{j_*})$. Then by the symmetry of A, we have whp,

$$M_r(A) \geq \boldsymbol{x}^\top A \boldsymbol{x} = 2^{1-\frac{2}{r}} a_*. \tag{2.6.1}$$

We now consider two cases.

Case 1. Suppose $r \in (1, 2]$. Let $\boldsymbol{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$ with $\|\boldsymbol{v}\|_r \leq 1$ denote a maximizing vector for the optimization problem OPT 1 associated with $M_r(A)$, and let $v_* := \max_{i\in[n]} |v_i|$. By the method of Langrange multipliers, it is easy to see that the following relationship must hold:

$$M_r(A)|v_i|^{r-1} = \left| \sum_{j=1}^{n} a_{ij} v_j \right|. \tag{2.6.2}$$

We claim that whp,

$$v_* \leq 2^{-1/r}(1+o(1)). \tag{2.6.3}$$

To see why the claim is true, let $k_* \in [n]$ satisfy $v_* = |v_{k_*}|$. By (2.6.2) we have

$$M_r(A)v_*^{r-1} = M_r(A)|v_{k_*}|^{r-1} = \left|\sum_{j\in[n]} a_{k_*j}v_j\right|. \tag{2.6.4}$$

Let $\delta(\alpha)$ be as defined in Lemma 2.3.3 (c). Suppose $\max_{j\in[n]}|a_{k_*j}| \leq b_n^{3/4+\delta(\alpha)}$. Then, since $\|\boldsymbol{v}\|_r \leq 1$ implies $\|\boldsymbol{v}\|_\infty \leq 1$, Lemma 2.3.3 (c) combined with (2.6.4) yields

$$M_r(A)v_*^{r-1} \leq \sum_{j=1}^{n} |a_{k_*j}| = o(b_n). \tag{2.6.5}$$

On the other hand, suppose there exists $k' \in [n]$ such that $|a_{k_*k'}| > b_n^{3/4+\delta(\alpha)}$. Then by properties (a) and (b) of Lemma 2.3.3, whp, there exists a unique such k', and further $k' \neq k_*$. Moreover, Lemma 2.3.3 (c) combined with (2.6.4) and the constraint $\|\boldsymbol{v}\|_r \leq 1$, yields whp

$$M_r(A)v_*^{r-1} \leq |a_{k_*k'}||v_{k'}| + \sum_{j:\,|a_{k_*j}|\leq b_n^{3/4+\delta(\alpha)}} |a_{k_*j}| \leq (1-v_*^r)^{\frac{1}{r}}\, a_*\,(1+o(1)) + o(b_n) \tag{2.6.6}$$

Combining (2.6.1), (2.6.5) and (2.6.6), we have, whp

$$2^{1-\frac{2}{r}}a_*v_*^{r-1} \leq M_r(A)v_*^{r-1} \leq (1-v_*^r)^{\frac{1}{r}}\, a_*\,(1+o(1)) + o(b_n). \tag{2.6.7}$$

Suppose that whp, $v_* = 1-o(1)$. Then (2.6.7) and the relation $a_* = \Theta_\delta(b_n)$, which holds by (2.3.1), imply that whp, $v_* = o(1)$, leading to a contradiction. So $v^* \neq 1-o(1)$. Together with (2.6.7) and $a_* = \Theta_\delta\,(b_n)$, this implies that whp,

$$2^{1-\frac{2}{r}}v_*^{r-1} \leq (1-v_*^r)^{\frac{1}{r}}\,(1+o(1))\,,$$

which is equivalent to

$$\frac{v_*^{r-1}}{(1-v_*^r)^{\frac{1}{r}}} \leq 2^{\frac{2}{r}-1}\left(1+o(1)\right). \tag{2.6.8}$$

Note that the map $\varphi : x \mapsto \frac{x^{r-1}}{(1-x^r)^{1/r}}$ is increasing for $x \in (0,1)$, and $\varphi(2^{-1/r}) = 2^{2/r-1}$. Together with (2.6.8), this completes the proof of claim (2.6.3).

Combining (2.6.4) and (2.3.2), we have whp,

$$M_r(A)v_*^{r-1} \leq v_* \sum_{j=1}^{n} |a_{k_*j}| \leq v_* a_* \left(1+o(1)\right).$$

Along with (2.6.3) and the fact that $r \leq 2$, this implies that whp,

$$M_r(A) \leq v_*^{2-r} a_* \left(1+o(1)\right) \leq 2^{1-\frac{2}{r}} a_*(1+o(1)). \tag{2.6.9}$$

To complete the proof of $r \in (1,2]$, combine (2.6.1), (2.6.9) and (2.3.1).

Case 2. Suppose $r = 1$. On the one hand, note that the lower bound (2.6.1) continues to hold. On the other hand, by definition $M_1(A) \leq \inf_{r\in(1,2]} M_r(A)$. Together with (2.6.9), this implies that whp,

$$M_1(A) \leq \inf_{r\in(1,2]} M_r(A) = 2^{-1}a_*(1+o(1)). \tag{2.6.10}$$

The proof is complete on combining (2.6.1), (2.6.10) and (2.3.1). □

2.6.2 Proof of Theorem 2.2.6 and Theorem 2.2.7

Proof of Theorem 2.2.6 and Theorem 2.2.7. Fix $r \in (2, \infty]$. Recall the decomposition $A = \bar{A} + \tilde{A} + \hat{A}$ in (2.4.2). Using the triangle inequality and (2.2.9), we have

$$\begin{aligned}\left|M_r(A) - M_r(\hat{A})\right| &\leq \sup_{\|\boldsymbol{x}\|_r=1}\left(\left|\boldsymbol{x}^\top\tilde{A}\boldsymbol{x}\right| + \left|\boldsymbol{x}^\top\bar{A}\boldsymbol{x}\right|\right)\\ &\leq \sup_{\|\boldsymbol{x}\|_r=\|\boldsymbol{y}\|_r=1}\left(\boldsymbol{y}^\top\tilde{A}\boldsymbol{x} + \boldsymbol{y}^\top\bar{A}\boldsymbol{x}\right)\\ &= \|\bar{A}\|_{r\to r^*} + \|\tilde{A}\|_{r\to r^*}.\end{aligned}$$

Now set $p := \frac{r}{r-1} = r^*$ and $\gamma := \frac{r-2}{r} = \frac{1}{p} - \frac{1}{r}$. Then $p^* = r > p \geq 1$, and conditions (2.2.5) and (2.2.6) imply that γ satisfies conditions (2.2.10) and (2.2.11), respectively. Thus, the results in Section 2.4 are valid for r and $p = r^*$. In particular, together with the last display, Proposition 2.4.1 and Proposition 2.4.8 imply that whp,

$$\left|M_r(A) - M_r(\hat{A})\right| = o(b_n),$$

under assumptions of Theorem 2.2.6 or Theorem 2.2.7. Together with Proposition 2.4.15, this completes the proof. □

2.6.3 Proof of Theorem 2.2.8

Proof of Theorem 2.2.8. Fix $r \in (2, \infty), \mu > 0$ and set $p = r^*$. Recall the decomposition of $A = \bar{A}+\tilde{A}+\hat{A}$ in (2.4.2) and let $\hat{A}_\mu := \mu\mathbf{1}\mathbf{1}^\top + \tilde{A} + \hat{A}$. Note that (2.2.7) implies (2.2.11) with $\gamma = \frac{1}{p} - \frac{1}{r} = \frac{r-2}{r}$. Hence, (2.4.7) of Lemma 2.4.7 and the condition $r > 2$ imply that

$$\|\bar{A}\|_{r\to r^*} = O_\delta\left(n^{\frac{1}{r^*}-\frac{1}{r}+\frac{1+\eta}{\alpha}}\right) = o\left(n^{\frac{r-2}{r}-\frac{r}{r-2}}b_n^{\frac{r}{r-2}}\right), \tag{2.6.11}$$

where the last equality uses (2.2.7).

With $A_\mu \coloneqq \mu \mathbf{1}\mathbf{1}^\top + A$, by the duality formula (2.2.9) and (2.6.11), we have whp,

$$\left|M_r(A_\mu) - M_r\left(\hat{A}_\mu\right)\right| \le \sup_{\|\boldsymbol{x}\|_r=1} \left|\boldsymbol{x}^\top \bar{A}\boldsymbol{x}\right| \le \sup_{\|\boldsymbol{x}\|_r=\|\boldsymbol{y}\|_r=1} \boldsymbol{y}^\top \bar{A}\boldsymbol{x} = \|\bar{A}\|_{r\to p} = o\left(n^{\frac{r-2}{r}-\frac{r}{r-2}} b_n^{\frac{r}{r-2}}\right). \tag{2.6.12}$$

Using (2.2.9) again, we have

$$M_r\left(\hat{A}_\mu\right) \le \sup_{\|\boldsymbol{x}\|_r=\|\boldsymbol{y}\|_r=1} \boldsymbol{y}^\top \hat{A}_\mu \boldsymbol{x} = \|\hat{A}_\mu\|_{r\to r^*}. \tag{2.6.13}$$

Since (2.2.7) implies (2.2.12) with $\gamma = \frac{1}{p} - \frac{1}{r} = \frac{r-2}{r}$, by Proposition 2.5.5 and (2.6.13), we have whp,

$$M_r\left(\hat{A}_\mu\right) \le n^{2-\frac{2}{r}}\mu + \frac{r-2}{r}\mu^{-\frac{2}{r-2}} n^{\frac{r-2}{r}-\frac{r}{r-2}} \sum_{i,j=1}^{n} |a_{ij}|^{\frac{r}{r-2}} + o\left(n^{\frac{r-2}{r}-\frac{r}{r-2}} b_n^{\frac{r}{r-2}}\right). \tag{2.6.14}$$

Recall the definition of $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}$ in (2.5.21). When $p = r^*$, we have $\hat{\boldsymbol{x}} = \hat{\boldsymbol{y}}$, and hence,

$$M_r\left(\hat{A}_\mu\right) \ge \frac{\hat{\boldsymbol{x}}^\top \hat{A}_\mu \hat{\boldsymbol{x}}}{\|\hat{\boldsymbol{x}}\|_r^2}. \tag{2.6.15}$$

Combining (2.6.15), the definition of $\mathcal{L}(A)$ in (2.5.22) and Lemma 2.5.7, we have whp,

$$M_r\left(\hat{A}_\mu\right) \ge n^{2-\frac{2}{r}}\mu + \frac{r-2}{r}\mu^{-\frac{2}{r-2}} n^{\frac{r-2}{r}-\frac{r}{r-2}} \sum_{i,j=1}^{n} |a_{ij}|^{\frac{r}{r-2}} + o\left(n^{\frac{r-2}{r}-\frac{r}{r-2}} b_n^{\frac{r}{r-2}}\right). \tag{2.6.16}$$

Combining (2.6.12), (2.6.14) and (2.6.16), we have whp,

$$M_r(A_\mu) = n^{2-\frac{2}{r}}\mu + \frac{r-2}{r}\mu^{-\frac{2}{r-2}} n^{\frac{r-2}{r}-\frac{r}{r-2}} \sum_{i,j=1}^{n} |a_{ij}|^{\frac{r}{r-2}} + o\left(n^{\frac{r-2}{r}-\frac{r}{r-2}} b_n^{\frac{r}{r-2}}\right) \tag{2.6.17}$$

Finally, [70, Theorem 3.8.2] implies that

$$\sum_{i,j\in[n]} |a_{ij}|^{\frac{r}{r-2}} \xrightarrow{d} Y_{\frac{\alpha(r-2)}{r}},$$

where $Y_{\frac{\alpha(r-2)}{r}}$ is an $\frac{\alpha(r-2)}{r}$-stable random variable. Together with (2.6.17), this completes the proof. $\square$

2.6.4 Proof of Corollary 2.2.9

Proof of Corollary 2.2.9. Fix $\alpha \in (0,1)$. Then condition (2.2.5) of Theorem 2.2.6 is satisfied and hence, by the theorem we have By Theorem 2.2.6 with the first condition, we have

$$b_n^{-1} \sup_{\|\boldsymbol{x}\|_\infty \le 1} \boldsymbol{x}^\top A_n \boldsymbol{x} \xrightarrow{(d)} Y_\alpha, \tag{2.6.18}$$

where Y_α is an α-stable random variable. Let $\Lambda_n \in \mathbb{R}^{n\times n}$ denote the diagonal of A_n. Then we have whp,

$$\|\Lambda\|_{\infty\to 1} = \sup_{\|\boldsymbol{x}\|_\infty = \|\boldsymbol{y}\|_\infty = 1} \sum_{i\in[n]} a_{ii} x_i y_i \le \sum_{i\in[n]} |a_{ii}| = O_\delta\left(n^{\frac{1}{\alpha}}\right), \tag{2.6.19}$$

where the last equality uses [70, Theorem 3.8.2]. Define $\mathring{A}_n := A_n - \Lambda_n$. Combining (2.6.18), (2.6.19) and (2.2.4), we have

$$b_n^{-1} \sup_{\|\boldsymbol{x}\|_\infty \le 1} \boldsymbol{x}^\top \mathring{A}_n \boldsymbol{x} \xrightarrow{(d)} Y. \tag{2.6.20}$$

Now we argue that the supremum in (2.6.20) must be obtained at the extremal points $\{-1,1\}^n$ of the n-dimensional hypercube. To see this, pick a point $\boldsymbol{x} \in \mathbb{R}^n$ with $\|\boldsymbol{x}\|_\infty \le 1$. Since $\mathring{A}_n$ is diagonal free, the map $\boldsymbol{x} \mapsto \boldsymbol{x}^\top \mathring{A}_n \boldsymbol{x}$ is linear in x_i for each $i \in [n]$. Hence, we can iteratively replace x_i by either $+1$ or -1 without decreasing $\boldsymbol{x}^\top \mathring{A}_n \boldsymbol{x}$ and the terminal point of this iterative procedure would be a point in $\{-1,1\}^n$. Given the above argument, (2.6.20) implies that

$$b_n^{-1} \sup_{\boldsymbol{x}\in\{-1,1\}^n} \boldsymbol{x}^\top \mathring{A}_n \boldsymbol{x} \xrightarrow{(d)} Y.$$

Once again invoking (2.6.19) and (2.2.4), we conclude that

$$b_n^{-1} \sup_{\boldsymbol{x} \in \{-1,1\}^n} \boldsymbol{x}^\top A_n \boldsymbol{x} \xrightarrow{(d)} Y.$$

This completes the proof. $\square$

CHAPTER 3

Fluctuations in Quantum Unique Ergodicity at the Spectral Edge

3.1 Introduction

Quantum Unique Ergodicity (QUE) refers to the observation that for the quantization of a chaotic dynamical system, the eigenstates of the Hamiltonian become uniformly distributed in phase space in the high-energy limit. This phenomenon has been intensely studied by both physicists and mathematicians, and we refer the reader to [148] for a survey. Recently, a number of works have investigated QUE, and other closely related principles, in the context of Wigner random matrices [39, 50, 2, 17, 48, 24, 23, 53, 52]. Because such matrices are the simplest class of chaotic quantum Hamiltonians, they form a natural testbed for the study of these ideas.

We recall that a Wigner matrix is a symmetric matrix $H = \{h_{ij}\}_{1\le i,j\le N}$ of real random variables with mean zero and variance N^{-1}, such that the upper triangular elements $\{h_{ij}\}_{1\le i,j,\le N}$ are independent. The eigenvectors of Wigner matrices are *delocalized*, meaning that their mass is spread approximately uniformly among their entries. The simplest manifestation of delocalization is the high-probability bound

$$\sup_{\alpha \in [\![1,N]\!]} N \langle \boldsymbol{q}_\alpha, \boldsymbol{u} \rangle^2 \le N^\varepsilon, \tag{3.1.1}$$

which holds for any $\varepsilon > 0$, eigenvector $\boldsymbol{u}$, and orthonormal basis $(\boldsymbol{q}_\alpha)_{\alpha=1}^N$, for sufficiently large N [32].[1]

QUE for Wigner matrices asserts a more refined form of delocalization, concerning the equidistribution of the eigenvector coordinates. Let $\mathcal{I} \subset [\![1,N]\!]$ be any determinstic subset of indices. Then for any eigenvector $\boldsymbol{u}$, we have the high-probability bound

$$\left| \sum_{\alpha \in \mathrm{I}} \langle \boldsymbol{q}_\alpha, \boldsymbol{u} \rangle^2 - \frac{|\mathrm{I}|}{N} \right| \le \frac{N^\varepsilon \sqrt{|\mathcal{I}|}}{N}. \tag{3.1.2}$$

A weaker version of this claim was first established in [39], and the optimal error term stated in (3.1.2) was shown in [24, 52].

In this article, we consider the fluctuations around the leading-order term identified in (3.1.2). Based on explicit calculations with Gaussian random matrices [141, Theorem 2.4], we expect that

$$\sqrt{\frac{N^3}{2|\mathrm{I}|\big(N - |\mathrm{I}|\big)}} \left(\sum_{\alpha \in \mathrm{I}} \langle \boldsymbol{q}_\alpha, \boldsymbol{u} \rangle^2 - \frac{|\mathrm{I}|}{N} \right) \to \mathcal{N}(0,1), \tag{3.1.3}$$

with convergence in distribution, for all Wigner matrices, whenever $|\mathcal{I}| \gg 1$. We observe that when $|\mathrm{I}| \ll N$, the summands act as independent Gaussians, while correlations arising from the condition that $\|\boldsymbol{u}\|_2 = 1$ are present when when $|\mathcal{I}|$ is of order N. It has been shown in the recent work [52] that (3.1.3) is true for eigenvectors $\boldsymbol{u}$ corresponding to eigenvalues in the bulk of the spectrum in the following sense. Label the eigenvalues of H in increasing order, $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_N$, and let $i = i_N$ be a sequence of indices such that $\min(i, N-i) > cN$, for some constant $c > 0$ and all $N \in \mathbb{N}$. Then (3.1.3) holds for

[1] All eigenvectors in this work are normalized so that $\|\boldsymbol{u}\|_2 = 1$.

the eigenvectors $\boldsymbol{u}_i$.

At the edge of the spectrum, previous results are less complete. In [24], it was shown that (3.1.3) holds for *any* eigenvector $\boldsymbol{u}$, if $N^{\tau} \leq |\mathcal{I}| \leq N^{1-\tau}$ for some $\tau > 0$. However, this leaves open the case where $|\mathcal{I}|$ is proportional to N where correlations between eigenvector entries arise. This case is of particular interest since it parallels the original QUE conjecture, which concerned the mass of eigenstates on subsets containing a constant fraction of phase space. In this article, we address this regime and show that (3.1.3) holds for any $\mathcal{I}$ such that $|\mathcal{I}| \geq N^{1-c}$ and any eigenvector $\boldsymbol{u}_i$ such that $\min(i, N-i) \leq N^{1-\tau}$, where $\tau > 0$ is an arbitrary constant, and $c(\tau) > 0$ is a small constant depending on τ. This completes the characterization of fluctuations in QUE for Wigner matrices at the spectral edge.

3.1.1 Main Results

We first define Wigner matrices.

Definition 3.1.1 (Wigner matrix)**.** *A Wigner matrix $H = H_N = \{h_{ij}\}_{1\leq i,j\leq N}$ is a real symmetric or complex Hermitian $N \times N$ matrix whose upper triangular elements $\{h_{ij}\}_{1\leq i\leqslant j\leq N}$ are independent random variables that satisfy*

$$\mathbb{E}[h_{ij}] = 0 \qquad \mathbb{E}\Big[|h_{ij}|^2\Big] = \frac{1+\delta_{ij}}{N}\,. \tag{3.1.4}$$

Further, we suppose that the normalized entries have finite moments, uniformly in N, i, and j, in the sense that for all $p \in \mathbb{N}$ there exists a constant μ_p such that

$$\mathbb{E}\left[\left|\sqrt{N}h_{ij}\right|^p\right] \leqslant \mu_p \tag{3.1.5}$$

for all N, i, and j.

Remark 3.1.2. The second condition in (3.1.4) could be relaxed to requiring only that

$$\mathbb{E}\left[|h_{ij}|^2\right] = \frac{1+(d-1)\delta_{ij}}{N}$$

for some constant $d > 0$. The modifications to the proofs in this case are straightforward, and we omit them for brevity.

Our main theorem is the following central limit theorem for Wigner matrix eigenvectors. We let

$$\mathbb{S}^{N-1} = \{\boldsymbol{v} \in \mathbb{R}^N : \|\boldsymbol{v}\|_2 = 1\}.$$

Theorem 3.1.3 (Central Limit Theorem)**.** *Let H be a Wigner matrix and fix $\tau \in (0,1)$. Then there exists $\delta(\tau) > 0$ such that the following holds. Let $\mathrm{I} = \mathrm{I}_N$ be a deterministic sequence of subsets such that $|\mathcal{I}| \in [\![N^{1-\delta}, N]\!]$, let $\ell = \ell_N \in [\![1, N^{1-\tau}]\!] \cup [\![N - N^{1-\tau}, N]\!]$ be a sequence of indices, and let $\boldsymbol{u} = \boldsymbol{u}_{\ell_N}^{(N)} = (u(1), \dots, u(N))$ be the corresponding sequence of ℓ^2-normalized eigenvectors of H. Let $(\boldsymbol{q}_\alpha^{(N)})_{\alpha\in\mathrm{I}} = (\boldsymbol{q}_\alpha)_{\alpha\in\mathrm{I}}$ be a deterministic sequence of sets of orthogonal vectors in $\mathbb{S}^{N-1}$. Then*

$$\sqrt{\frac{\beta N^3}{2|\mathrm{I}|(N-|\mathrm{I}|)}}\left(\sum_{\alpha\in\mathrm{I}} \langle \boldsymbol{q}_\alpha, \boldsymbol{u}_{\ell_N}\rangle^2 - \frac{|\mathrm{I}|}{N}\right) \to \mathcal{N}(0,1), \tag{3.1.6}$$

with convergence in the sense of moments. Here $\mathcal{N}(0,1)$ is a standard real Gaussian random variable; we take $\beta = 1$ if H is real symmetric, or $\beta = 2$ if it is complex Hermitian.

Because the Gaussian distribution is determined by its moments, the above theorem immediately implies that (3.1.6) also holds for convergence in distribution [30, Theorem 30.2].

3.1.2 Related Works

Delocalization estimates have received significant attention from the random matrix community over the past decade and a half. The estimate (3.1.1) has a long history, and increasingly strong version of this statement were proved in [80, 79, 81, 158, 160, 83, 84, 165, 3, 102, 101]. The optimal high-probability upper bound of $\sqrt{(2+\varepsilon)\log N}$ was recently established in [25]. Going beyond Wigner matrices, similar estimates have been shown for band matrices [166, 40, 76], heavy-tailed random matrices [34, 35, 6, 4], and adjacency matrices of sparse random graphs [19, 77]. Fluctuations of individual eigenvector entries of Wigner matrices were first studied in [39], where they were shown to be Gaussian (see also [25, Corollary B.18] and [22]). Arbitrary finite collections of bulk eigenvector entries were shown to be jointly Gaussian in [133].

As noted above, the first QUE estimate for Wigner matrices was shown in [39]. Estimates of the form (3.1.2) have also been shown for deformed Wigner matrices, [21], band matrices [166, 40], sparse random matrices [19, 38, 10, 9, 8], and heavy-tailed random matrices [5]. Further, a more general version of (3.1.2), known as *eigenvector thermalization*, has appeared recently (motivated by the phenomena surveyed in [157, 64, 66]). Let A be a deterministic $N \times N$ matrix such that $\|A\| \leq 1$, where $\|A\|$ denotes the spectral norm of A. Then for any eigenvector $\boldsymbol{u}$ of a Wigner matrix, we have the high-probability bound

$$\left| \langle \boldsymbol{u}, A\boldsymbol{u} \rangle - \frac{1}{N} \operatorname{Tr} A \right| \leq \frac{N^{\varepsilon}}{\sqrt{N}}, \tag{3.1.7}$$

for any $\varepsilon > 0$ and sufficiently large N [49]. Subsequently, fluctuations around the leading order term in (3.1.7) were identified in [50], and an optimal-order error term was established in [52].

We expect that our proof strategy can be extended straightforwardly in two directions. One is to determine the fluctuations for the more general observables $\langle \boldsymbol{u}, A\boldsymbol{u} \rangle$ when $\boldsymbol{u}$ is an edge eigenvector and A is essentially full-rank, i.e. $\frac{1}{N} \operatorname{Tr}(A^2) \geq N^{-\delta}\|A\|^2$ for some small $\delta > 0$. The other is to yield the joint fluctuations for any finite edge eigenvectors, i.e.

$$\sqrt{\frac{N^3}{2|\mathrm{I}|(N-|\mathrm{I}|)}} \left(\sum_{\alpha \in \mathrm{I}} \langle \boldsymbol{q}_\alpha, \boldsymbol{u}_{\ell_1} \rangle^2 - \frac{|\mathrm{I}|}{N}, \cdots, \sum_{\alpha \in \mathrm{I}} \langle \boldsymbol{q}_\alpha, \boldsymbol{u}_{\ell_k} \rangle^2 - \frac{|\mathrm{I}|}{N} \right) \to (Z_1, \cdots, Z_k),$$

with convergence in distribution, where $\ell_1 < \cdots < \ell_k \leq N^{1-\tau}$ and $Z_1, \ldots, Z_k$ are independent Gaussian random variables with zero mean and unit variance. However, we consider only the case (3.1.3) for brevity.

Our work does not address the intermediate spectral regime where i/N tends to 0 slower than any negative power of N. We expect that this regime can be handled by a straightforward modification of the arguments in [52]. However, we leave this as an open question for future work.

3.1.3 Proof Strategy

Previous works determining the fluctuations in QUE have all followed the dynamical approach to random matrix universality (surveyed in [82]). This approach uses the following three steps.

1. Establish various *a priori* estimates on the eigenvalues and eigenvectors of Wigner matrices, such as (3.1.1), which are used as input in the following steps.

2. Determine the fluctuations in QUE for random matrices of the form $H+\sqrt{t}W$, where H is an arbitrary Wigner matrix, W is a Gaussian Wigner matrix, and $t \approx N^{-c}$ for some $c > 0$. This is done by recognizing $H + \sqrt{t}W$ as the evolution of a matrix Brownian motion with initial data H until time t. Under this stochastic process, the moments of the QUE observable in (3.1.3) evolve according to a parabolic differential equation known as the *eigenvector moment flow*. A detailed analysis of this evolution shows that these moment observables converge to their equilibrium states, the Gaussian moments, after time t. This convergence in moments establishes (3.1.3) for the matrix $H + \sqrt{t}W$.

3. Transfer the conclusion from the previous step to all Wigner matrices. Given an arbitrary Wigner matrix H, there exists a Wigner matrix H' such that the first three moments of H and $H' + \sqrt{t}W$ match exactly, and the difference of the fourth

moments is order t. By a moment matching argument similar to the one used in Lindeberg's proof of the central limit theorem (see [20, Section 11]), one can show that this moment condition is enough to establish that H has the same fluctuations in QUE as $H' + \sqrt{t}W$, completing the proof.

Thus far, obstacles related to Step 2 of the dynamical approach have blocked a proof of (3.1.3) for $|\mathcal{I}|$ proportional to N at the spectral edge. The analysis of the eigenvector moment flow in [24] was applicable throughout the entire spectrum, but was only effective for index sets $\mathcal{I}$ with cardinality $|\mathcal{I}| \ll N$. The works [50, 52] analyzed a variation of the eigenvector moment flow introduced in [133], called the *colored eigenvector moment flow*, which allow them to access $\mathcal{I}$ with $|\mathcal{I}|$ proportional to N. However, these works depend on an intricate analysis of the colored evolution dynamics presented in [133], which was only given in the bulk. In principle, such an analysis could also be carried out at the edge. However, given the length and sophistication of [133], and additional complications that arise at the edge due to the curvature of the spectral density (the semicircle law, given in (3.2.1) below), this extension seems far from straightforward.

Instead, we adopt an argument that has no dynamical component, and uses only moment matching. We draw inspiration from [119] and [33], which characterize the joint eigenvector–eigenvalue distribution of Wigner matrices at the edge (see [33, Remark 8.5]). Specifically, the authors show that given any finite collection of edge eigenvalues and entries of the corresponding eigenvectors, their joint distribution is asymptotically the same with the one for a Gaussian ensemble. In particular, any finite collection of such eigenvector entries is asymptotically distributed as independent Gaussians. The proof proceeds by a "two moment matching" argument, which shows that two random matrix ensembles who entries are independent, centered, and have the same variance matrix also have the same eigenvector–eigenvalue statistics at the edge. As an immediate consequence, the edge statistics of any Wigner matrix match those of a Gaussian Wigner matrix, which may be computed explicitly. The decay of spectral density at the edge is crucial to the proof, and renders it inapplicable to the bulk, where the full dynamical approach is necessary.

We now give an overview of our proof. Let H be a Wigner matrix. Our first step is to regularize the QUE observable in (3.1.3). Let $f_n(H)$ denote a smooth approximation to n-th moment of the observable on the left side of (3.1.3), corresponding to some eigenvector $\boldsymbol{u}$, which is differentiable in the matrix entries.[2] We wish to proceed as follows. Fix indices $a, b \in [\![1, N]\!]$. Let W denote the matrix such that $w_{ij} = h_{ij}$ for all $i, j \in [\![1, N]\!]$ such that $(i, j) \notin \{(a, b), (b, a)\}$, and such that $w_{ab} = w_{ba} = g$, where g is a Gaussian variable with mean 0 and variance N^{-1}. We observe that the first two moments of g match those of h_{ab}. Finally, let Q denote the matrix such that $q_{ij} = h_{ij}$ for all $i, j \in [\![1, N]\!]$ such that $(i, j) \notin \{(a, b), (b, a)\}$, where $q_{ab} = q_{ba} = 0$. Then by Taylor expansion, we have

$$f_n(H) = f_n(Q) + \partial_{ab} f_n(Q) h_{ab} + \frac{1}{2}\partial_{ab}^2 f_n(Q) h_{ab}^2 + \frac{1}{6}\partial_{ab}^3 f_n(Q) h_{ab}^3 + \frac{1}{24}\partial_{ab}^4 f_n(Q) h_{ab}^4 + X_H,$$

where X_H is the error term in expansion. Subtracting the analogous expansion for $f_n(W)$, and taking expectations, we obtain

$$\begin{aligned}
\mathbb{E}\Big[f_n(H) - f_n(W)\Big] &= \mathbb{E}\Big[\partial_{ab} f_n(Q)(h_{ab} - w_{ab})\Big] + \frac{1}{2}\mathbb{E}\Big[\partial_{ab}^2 f_n(Q)(h_{ab}^2 - w_{ab}^2)\Big] \\
&+ \frac{1}{6}\mathbb{E}\Big[\partial_{ab}^3 f_n(Q)(h_{ab}^3 - w_{ab}^3)\Big] + \frac{1}{24}\mathbb{E}\left[\partial_{ab}^4 f_n(Q)(h_{ab}^4 - w_{ab}^4)\right] + \mathbb{E}\Big[(X_H - X_W)\Big] \\
&= \frac{1}{6}\mathbb{E}\Big[\partial_{ab}^3 f_n(Q)\Big]\mathbb{E}\Big[h_{ab}^3 - w_{ab}^3\Big] + \frac{1}{24}\mathbb{E}\left[\partial_{ab}^4 f_n(Q)\right]\mathbb{E}\left[h_{ab}^4 - w_{ab}^4\right] + \mathbb{E}\Big[(X_H - X_W)\Big].
\end{aligned}$$

In the previous equation, we observed that Q is independent from h_{ab} and w_{ab}, and used $\mathbb{E}[\partial_{ab} f_n(Q)(h_{ab} - w_{ab})] = 0$, which follows from $\mathbb{E}[h_{ab}] = \mathbb{E}[w_{ab}]$. We also used the analogous reasoning for the second-moment term.

We consider the third-moment and fourth-moment terms, and neglect the error term for now. From the definition of a Wigner matrix, we have $\mathbb{E}\Big[h_{ab}^3 - w_{ab}^3\Big] = O(N^{-3/2})$ and

[2]The observable itself is not differentiable in the matrix entries, which necessitates the smoothing.

$\mathbb{E}\left[h_{ab}^4 - w_{ab}^4\right] = O(N^{-2})$. If we had the estimates

$$\mathbb{E}\Big[\partial_{ab}^3 f_n(Q)\Big] \ll N^{-1/2}, \quad \mathbb{E}\Big[\partial_{ab}^4 f_n(Q)\Big] \ll 1, \tag{3.1.8}$$

then $\mathbb{E}\Big[f_n(H) - f_n(W)\Big] \ll N^{-2}$. This estimates the error accrued when exchanging one entry of W for a Gaussian. Since we need to exchange $O(N^2)$ entries, the total error will be $o(1)$, and the moments $\mathbb{E}[f_n(H)]$ will match those of a Gaussian random matrix in the large N limit. Because (3.1.3) can directly be established for Gaussian matrices, this would complete the proof.

The crux of the problem is then to produce a suitable regularization f_n and demonstrate that its derivatives decay suitably in N near the edge of the spectrum. While regularizations of (3.1.3) have appeared before, the necessary decay at the edge has not been established. For example, the regularized QUE observable in [24] was only shown to satisfy $\mathbb{E}\Big[\partial_{ab}^3 f_n(Q)\Big] = O(1)$. To illustrate how our regularization works, and how we achieve the additional gain at the edge, we begin by describing the regularization of a single eigenvector entry, as accomplished in [119, 33].

Let $\boldsymbol{u}_\ell$ be an eigenvector with associated eigenvalue λ_ℓ, and let $\eta > 0$ be chosen so that $\eta \ll \Delta_\ell$, where Δ_ℓ is the typical size of the eigenvalue gap $\lambda_{\ell+1} - \lambda_\ell$. Recall the Poisson kernel identity

$$\frac{\eta}{\pi}\int_{\mathbb{R}} \frac{\mathrm{d}E}{(E-\lambda_\ell)^2 + \eta^2} = 1.$$

Fix $k \in [\![1, N]\!]$. We have the high probability estimate

$$\boldsymbol{u}_\ell(k)^2 = \frac{\eta}{\pi}\int_{\mathbb{R}} \frac{\boldsymbol{u}_\ell(k)^2\,\mathrm{d}E}{(E-\lambda_\ell)^2 + \eta^2} \approx \frac{\eta}{\pi}\int_{I} \frac{\boldsymbol{u}_\ell(k)^2\,\mathrm{d}E}{(E-\lambda_\ell)^2 + \eta^2} \approx \frac{\eta}{\pi}\int_{I} \sum_{i=1}^{N} \frac{\boldsymbol{u}_i(k)^2\,\mathrm{d}E}{(E-\lambda_i)^2 + \eta^2},$$

where I is any interval centered at λ_ℓ such that $\eta \ll |I| \ll \Delta_\ell$, and we used

$$\max(\lambda_{\ell+1} - \lambda_\ell, \lambda_\ell - \lambda_{\ell-1}) \gg \eta$$

to neglect the terms with $i \neq \ell$ in the sum. Letting $G = (H - E - \mathrm{i}\eta)^{-1}$ denote the resolvent of H, we have by the spectral theorem that

$$\frac{\eta}{\pi} \int_I \sum_{i=1}^{N} \frac{\boldsymbol{u}_i(k)^2 \, \mathrm{d}E}{(E - \lambda_i)^2 + \eta^2} = \frac{\eta}{\pi} \int_I (G\bar{G})_{kk} \, \mathrm{d}E. \tag{3.1.9}$$

For illustrative purpose, let us treat I as a deterministic interval. Then, we see that

$$f_n(H) := \left(\frac{N\eta}{\pi} \int_I (G\bar{G})_{kk} \mathrm{d}E \right)^n \approx \left(\sqrt{N} \boldsymbol{u}_\ell(k) \right)^{2n},$$

is a smooth function of the matrix entries. We multiplied $\boldsymbol{u}_\ell(k)$ by $\sqrt{N}$ to make it an $O(1)$ quantity.

Letting $R = (Q - E - \mathrm{i}\eta)^{-1}$ denote the resolvent of Q and taking derivatives, one readily finds that[3]

$$\partial_{ab}^m f_n(Q) \approx n(n-1) \cdots (n-m+1) (\sqrt{N} \boldsymbol{u}_\ell(k))^{2n-m} \int_I N \tilde{P}_m \, \mathrm{d}E + \cdots , \tag{3.1.10}$$

where $\tilde{P}_m$ is a polynomial with constant number of terms, and each term consists of one $(R\bar{R})_{**}$ factor and m R_{**}'s or $\bar{R}_{**}$'s. Here $* \in \{a, b, k\}$ and different $*$ may take different values.

So far, we have not used the fact that λ_ℓ is an edge eigenvalue. The crucial use of this fact is that we are able to choose the spectral parameter η such that $1 \ll \Delta_\ell \eta^{-1} \ll (N\eta)^{1/4}$. Indeed, if λ_ℓ were in the bulk, we would have $\Delta_\ell = O\left(N^{-1}\right)$, and such choice would not be possible. Combined with standard local law for resolvents of Wigner matrices (see (3.4.8) below), for any $\varepsilon > 0$, we have

$$(R\bar{R})_{ij} \leq N^{\varepsilon} (N\eta)^{-2}$$

[3] There are multiple terms as a result of applying product rule, so we focus on one representative term for clarity.

with high probability for all $i, j \in [\![1, N]\!]$. Therefore, we have the bound

$$\int_I N\tilde{P}_m \leq |I| N^{1+\varepsilon} (N\eta)^{-2} \leq (N\eta)^{-1/2} \ll 1, \tag{3.1.11}$$

by the choice of I and η. Inserting this into (3.1.10), we have

$$\left|\partial_{ab}^m f_n(Q)\right| \ll 1,$$

with high probability. Upon taking expectation, this establishes the second inequality in (3.1.8).

For the first inequality in (3.1.8), we need to exploit an additional cancellation that is introduced when taking expectation. To uncover this cancellation, we use the *polynomialization* technique, which first appeared systematically in [75], and was further developed in [167, 33]. The main idea is to write $\tilde{P}_m$ in the form

$$\tilde{P}_m \approx \sum_{i_1,\ldots,i_d \neq a} \tilde{P}^{(a)}_{m,i_1,\ldots,i_d} h_{ai_1} \cdots h_{ai_d},$$

where $\tilde{P}^{(a)}_{m,i_1,\ldots,i_d}$ is independent of a-th row and column of Q. When d is an odd number, we have

$$\left|\mathbb{E}\Big[\tilde{P}_m\Big]\right| \lesssim N^{-1/2} \sqrt{\mathbb{E}\Big[|\tilde{P}_m|^2\Big]}. \tag{3.1.12}$$

To see why (3.1.12) is true, consider a simple case, where $\mathcal{P} = \sum_{i_1,i_2,i_3 \neq a} h_{ai_1} h_{ai_2} h_{ai_3}$. Taking expectation forces i_1, i_2, i_3 to coincide, and therefore

$$\Big|\mathbb{E}[\mathcal{P}]\Big| = \left|\mathbb{E}\left[\sum_i h_{ai}^3\right]\right| \lesssim N^{-1/2} = N^{-1/2} \sqrt{\mathbb{E}\left[\sum_{i_1,i_2,i_3} h_{ai_1}^2 h_{ai_2}^2 h_{ai_3}^2\right]} \leq N^{-1/2} \sqrt{\mathbb{E}\left[|\mathcal{P}|^2\right]}.$$

Moreover, it can be shown that, $\tilde{P}_m$ can be approximated by an odd polynomial as long as m is odd and a, b, k are distinct indices. Combining (3.1.12) with (3.1.10) and (3.1.11), we obtain the first inequality in (3.1.8) for all indices a, b, expect for $O(n)$ pairs such that $a = b, a = k$ or $b = k$, which is sufficient for our purpose.

Regularizing the QUE observable in (3.1.3) can be accomplished similarly by replacing each term $\langle \boldsymbol{q}_\alpha, \boldsymbol{u}_\ell \rangle^2$ appearing there by the regularization given in (3.1.9). However, to appropriately control the size of the resulting moments, we need to detect additional cancellations in the sum; it is not enough to bound each term individually. For this, we use *multi-resolvent local laws*, which bound the quantities $(GA\overline{G})_{cd}$ and $(GA\overline{G}G)_{cd}$ for any choice of $c, d \in [\![1, N]\!]$ and deterministic $N \times N$ matrix A such that $\|A\| \leq 1$ and $\operatorname{Tr} A = 0$; see Lemma 3.4.2 below. While such laws have been established previously [52, 51, 49], we prove a new version with improved error terms at the spectral edge. These improved estimates allow us to obtain sharper bounds in the moment matching argument, which are necessary to complete the proof.

3.1.4 Outline

Section 3.2 introduces our notational conventions, and states several preliminary results from previous works that are used throughout this paper. Section 3.3 defines the smoothed QUE observables needed for our moment matching argument. Section 3.4 proves our main result, Theorem 3.1.3, assuming two preliminary lemmas, Lemma 3.4.2, and Lemma 3.4.4. Lemma 3.4.4 is proved in Section 3.5. Lemma 3.4.2 is proved in Appendix 3.6.

3.2 Preliminaries

3.2.1 Conventions

For the remainder of the paper, we fix arbitrary constants $\tau \in (0, 1)$, a sequence of deterministic subsets $\mathrm{I} = \mathrm{I}_N \subset [\![N^{1-\delta}, N]\!]$, with $\delta \equiv \delta(\tau) = 10^{-5}\tau$, and a sequence of deterministic indices $\ell = \ell_N \in [\![1, N^{1-\tau}]\!] \cup [\![N - N^{1-\tau}, N]\!]$. We also fix a deterministic sequence $(\boldsymbol{q}_\alpha^{(N)})_{\alpha \in \mathrm{I}} = (\boldsymbol{q}_\alpha)_{\alpha \in \mathrm{I}}$ of sets of orthogonal vectors in $\mathbb{S}^{N-1}$, and a sequence of positive reals $(\mu_p)_{p=1}^{\infty}$. We assume that all Wigner matrices mentioned below satisfy Definition 3.1.1 with this sequence of constants. Our claims hold for any choices of these parameters.

We also define the spectral domain

$$\boldsymbol{S} = \boldsymbol{S}(N) = \left\{ z = E + \mathrm{i}\eta \in \mathbb{C} : |E| \leqslant \frac{10}{\tau}, N^{-1+\tau/10} \leqslant |\eta| \leqslant \frac{10}{\tau} \right\}.$$

Throughout this article, we typically suppress the dependence of various constants in our results on the choices of τ. These dependencies do not affect our arguments in any substantial way. Additionally, we focus on the case of real symmetric Wigner matrices in our proof of Theorem 3.1.3. The details for the complex Hermitian case are nearly identical, and hence omitted.

3.2.2 Notations and Definitions

Let Mat_N be the set of $N \times N$ real symmetric matrices and $\{\boldsymbol{e}_i\}_{i=1}^N$ be the standard basis of $\mathbb{R}^N$. Let $\|M\|$ denote the spectral norm of M. We index the eigenvalues of matrices $M \in \mathrm{Mat}_N$ in increasing order, and denote them $\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_N$. For $z \in \mathbb{C}\backslash\mathbb{R}$, the resolvent of $M \in \mathrm{Mat}_N$ is given by $G(z) = (M - z)^{-1}$. The Stieltjes transform of M is

$$m_N(z) = \operatorname{Tr} G(z) = \frac{1}{N} \sum_i \frac{1}{\lambda_i - z}.$$

The resolvent has the spectral decomposition

$$G(z) = \sum_{i=1}^{N} \frac{\boldsymbol{u}_i \boldsymbol{u}_i^*}{\lambda_i - z},$$

where we let $\boldsymbol{u}_i$ denote the eigenvector corresponding to the eigenvalue λ_i of M such that $\|\boldsymbol{u}_i\|_2 = 1$. We fix the sign of $\boldsymbol{u}_i$ arbitrarily by demanding that $\boldsymbol{u}_i(1) \geqslant 0$. For deterministic vectors $\boldsymbol{x}, \boldsymbol{y}$, we abbreviate $\langle \boldsymbol{x}, M\boldsymbol{y} \rangle$ by $M_{\boldsymbol{xy}}$ and we abbreviate $M_{\boldsymbol{xy}}$ further by $M_{i\boldsymbol{y}}$ or $M_{\boldsymbol{x}j}$ if $\boldsymbol{x} = \boldsymbol{e}_i$ or $\boldsymbol{y} = \boldsymbol{e}_j$ respectively.

The semicircle density and its Stieltjes transform are

$$\rho_{\rm sc}(E) = \frac{\sqrt{(4-E^2)_+}}{2\pi}\,\mathrm{d}E, \qquad m_{\rm sc}(z) = \int_{\mathbb{R}} \frac{\rho_{\rm sc}(x)}{x-z}\mathrm{d}x = \frac{-z+\sqrt{z^2-4}}{2}, \tag{3.2.1}$$

for $E \in \mathbb{R}$ and $z \in \mathbb{C}\backslash\mathbb{R}$. The square root in $\sqrt{z^2-4}$ is defined with a branch cut in $[-2,2]$, so that $\Im m_{\rm sc}(z) > 0$ for $\Im z > 0$.

For $i \in [\![1, N]\!]$, we denote the i-th N-quantile of the semicircle distribution by γ_i and define it implicitly by

$$\frac{i}{N} = \int_{-2}^{\gamma_i} \rho_{\rm sc}(x)\,\mathrm{d}x.$$

We will often differentiate functions of a matrix $M \in \mathrm{Mat}_N$ with respect to some entry m_{ab} of M. For example, we will consider quantities such as $\partial_{ab} f(M)$, where ∂_{ab} means that we consider M as a function of its upper-triangular elements $\{m_{ij}\}_{1\le i\le j\le N}$ and differentiate with respect to m_{ab} when $a \le b$, or with respect to m_{ba} when $b \le a$. Mostly commonly, we take f to be the resolvent $f(M) = (M-z)^{-1}$, or some product of resolvents.

Finally, we adopt the convention that $\mathbb{N} = \{1, 2, 3, \dots\}$.

3.2.3 Local law for resolvent and multi-resolvent

We require the isotropic local law proved in [32] and multi-resolvent local law proved in [51]. We begin by recalling the notion of stochastic domination (which was introduced in [76]).

Definition 3.2.1 (Stochastic domination). *Let*

$$X = \left(X^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)}\right), \quad Y = \left(Y^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)}\right)$$

be two families of nonnegative random variables, where $U^{(N)}$ is a possibly N-dependent parameter set. We say that X is stochastically dominated by Y, uniformly in u, if for all *(small)* $\varepsilon > 0$ *and (large)* $D > 0$ *we have*

$$\sup_{u \in U^{(N)}} \mathbb{P}\left[X^{(N)}(u) > N^{\varepsilon} Y^{(N)}(u)\right] \leqslant N^{-D}$$

for large enough $N \geqslant N_0(\varepsilon, D)$. *Unless stated otherwise, throughout this paper the stochastic domination will always be uniform in all parameters apart from* δ, τ, *and the constants* μ_p *(which were fixed in Section 3.2.1); thus,* $N_0(\varepsilon, D)$ *also depends on* μ_p, τ, δ. *If* X *is stochastically dominated by* Y, *uniformly in* u, *we use the notation* $X \prec Y$. *Moreover, if for some complex family* X *we have* $|X| \prec Y$ *we also write* $X = O_{\prec}(Y)$

We first introduce the isotropic local law for a single resolvent.

Theorem 3.2.2 (Isotropic local law)**.** *Let* H *be a Wigner matrix, let* $G = (H - z)^{-1}$ *be its resolvent.*

$$\sup_{z \in \boldsymbol{S}} \left|\langle \boldsymbol{x}, G(z)\boldsymbol{y}\rangle - \langle \boldsymbol{x}, \boldsymbol{y}\rangle m_{\mathrm{sc}}(z)\right| \prec \sqrt{\frac{|\operatorname{Im} m_{\mathrm{sc}}(z)|}{N\eta}} + \frac{1}{N\eta} \tag{3.2.2}$$

for any choice of deterministic vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{S}^{N-1}$, *where* $\eta = |\Im z|$.

Remark 3.2.3. In this and all following local laws, the high probability bound may be strengthened to hold simultaneously for all z in the specified domain. For instance, (3.2.2) may be strengthened to

$$\mathbb{P}\left[\bigcap_{z \in \boldsymbol{S}} \left\{|\langle \boldsymbol{x}, G(z)\boldsymbol{y}\rangle - m_{\mathrm{sc}}(z)\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \leqslant N^{\varepsilon}\left(\sqrt{\frac{\operatorname{Im} m_{\mathrm{sc}}(z)}{N\eta}} + \frac{1}{N\eta}\right)\right\}\right] \geqslant 1 - N^{-D},$$

for all $\varepsilon > 0, D > 0$ and $N \geq N_0(\varepsilon, D)$. It follows from a simple lattice argument combined with the Lipschitz continuity of G, m_{sc} on $\boldsymbol{S}$. See [20, Remark 2.7] for details.

We next recall the definitions of a non-crossing partition and the Kreweras complement of a partition. We follow the notation of [100]; see also [121].

Definition 3.2.4. *For* $k \in \mathbb{N}$, *let* $[k]$ *denote the set* $\{1, 2, \ldots, k\}$. *A* set partition *of* $[k]$ *is* *a set π of disjoint subsets of $[k]$ whose union is $[k]$. The elements of π are called* blocks. *Given a set partition π, a* bump *is an ordered pair (i_1, i_2) such that i_1 and i_2 lie in the same block of π, $i_1 < i_2$, and there is no j in the same block with $i_1 < j < i_2$. We say that π is a* noncrossing partition *if for every pair of bumps (i_1, i_2) and (j_1, j_2) in π, it is not the case that $i_1 < j_1 < i_2 < j_2$. We let $\mathrm{NC}[k]$ denote the set of non-crossing partitions of $[k]$.*

Definition 3.2.5. *Arrange the points in $[k]$ equidistantly on the boundary of the unit disk $\mathbb{D}$, with labels increasing counterclockwise. Label the arcs between adjacent points so that arc i connects point i to its neighbor in the counterclockwise direction. Given $\pi \in \mathrm{NC}[k]$, we define the* Kreweras complement $K(\pi) \in \mathrm{NC}[k]$ *of π to be the partition such that two points $x, y \in [k]$ belong to the same block of $K(\pi)$ if and only if the arcs x, y are in the same connected component of $\mathbb{D} \setminus \cup_{B\in\pi} P_B$, where P_B denotes the convex hull of the vertices in the block B.*

Moreover, for $\pi \in \mathrm{NC}[k]$, and matrices $A_1, \ldots, A_{k-1} \in \mathrm{Mat}_N$, we define the partial trace pTr_π associated to partition π to be

$$\mathrm{pTr}_\pi (A_1, \ldots, A_{k-1}) = \frac{1}{N} \prod_{B\in\pi\setminus B(k)} \mathrm{Tr}\left(\prod_{j\in B} A_j\right) \prod_{j\in B(k)\setminus\{k\}} A_j,$$

where $B(k) \in \pi$ denotes the unique block containing k. We recall that by convention, an empty product is equal to 1.

For any subset $B \subset [k]$ we define

$$m[B] \coloneqq m_{\mathrm{sc}}\left[\{z_i \mid i \in B\}\right] = \int_{-2}^{2} \rho_{\mathrm{sc}}(x) \prod_{i\in B} \frac{1}{x - z_i}\,\mathrm{d}x.$$

For every $k \in \mathbb{N}$, let $m_\circ[\cdot] : 2^{[k]} \to \mathbb{C}$ denote the free-cumulant transform of $m[\cdot]$, which is defined implicitly by requiring that the relation

$$m[B] = \sum_{\pi \in \mathrm{NC}(B)} \prod_{B' \in \pi} m_\circ [B'] , \quad \forall B \subset [k].$$

holds for all k. For example, when $k = 1$, we have $m_\circ[i] = m[i]$, and for $k = 2$ we have $m_\circ[i, j] = m[i, j] - m[i]m[j]$. For further details, see the discussion following [53, Definition 2.3].

Definition 3.2.6. *For arbitrary deterministic matrices $A_1, \ldots, A_{k-1} \in \mathrm{Mat}_N$ and spectral parameters $z_1, \ldots, z_k \in \mathbb{C}\backslash\mathbb{R}$, define*

$$M(z_1, A_1, \ldots, A_{k-1}, z_k) := \sum_{\pi \in \mathrm{NC}[k]} \mathrm{pTr}_{K(\pi)} (A_1, \ldots, A_{k-1}) \prod_{B \in \pi} m_\circ[B]. \tag{3.2.3}$$

We are now ready to state the multi-resolvent local laws necessary for our work.

Lemma 3.2.7 ([51, Lemma 2.4])**.** *Fix $k, m \in \mathbb{N}$ with $m \le k$ and a constant $C_0 > 0$. Let $A_1, \ldots, A_k \in \mathrm{Mat}_N$ be deterministic matrices such that $\|A_i\| \le C_0$ for all $1 \le i \le k$, and suppose that $\mathrm{Tr}\, A_j = 0$ holds for at least m distinct indices j. Then there exists a constant $C = C(C_0, k) > 0$ such that*

$$|\mathrm{Tr}\left(M\left(z_1, A_1, \ldots, z_{k-1}, A_{k-1}, z_k\right) A_k\right)| \le \begin{cases} \frac{CN}{\eta^{k-1-\lceil m/2 \rceil}} & d \le 1 \\ \frac{CN}{d^k} & d \ge 1 \end{cases}$$

and

$$\|M\left(z_1, A_1 \ldots, z_k, A_k, z_{k+1}\right)\| \le \begin{cases} \frac{C}{\eta^{k-\lceil m/2 \rceil}} & d \le 1 \\ \frac{C}{d^{k+1}} & d \ge 1, \end{cases} \tag{3.2.4}$$

where $\eta := \min_j |\Im z_j|$ and $d := \min_j \mathrm{dist}\left(z_j, [-2, 2]\right)$.

Theorem 3.2.8 ([51, Theorem 2.5])**.** *Let H be a $N \times N$ Wigner matrix and let $G = (H - z)^{-1}$ be its resolvent. Fix $m, k \in \mathbb{N}$ with $m \le k$ and $z_1, \ldots, z_{k+1} \in \boldsymbol{S}$. Fix $C_0 > 0$,* *and let $A_1, \dots, A_k$ be deterministic matrices such that $\|A_j\| \le C_0$ for all $1 \le j \le k$, and $\operatorname{Tr} A_j = 0$ for at least m distinct indices j. Then*

$$\left|\operatorname{Tr}\left(G_1 A_1 \cdots G_k A_k - M\left(z_1, A_1, \dots, A_{k-1}, z_k\right) A_k\right)\right| \prec \begin{cases} \frac{1}{\eta^{k-m/2}} & d \le 1 \\ \frac{1}{d^{k+1}} & d \ge 1, \end{cases} \tag{3.2.5}$$

and for any deterministic vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ such that $\|\boldsymbol{x}\| + \|\boldsymbol{y}\| \le C_0$, we have

$$\left|\left\langle \boldsymbol{x}, \left(G_1 A_1 \cdots G_k A_k G_{k+1} - M\left(z_1, A_1, \dots, A_k, z_{k+1}\right)\right) \boldsymbol{y} \right\rangle\right| \prec \begin{cases} \frac{1}{\sqrt{N}\eta^{k-m/2+1/2}} & d \le 1 \\ \frac{1}{\sqrt{N} d^{k+2}} & d \ge 1. \end{cases} \tag{3.2.6}$$

Here $G_j := G\left(z_j\right)$, $\eta := \min_j |\Im z_j|$, and $d := \min_j \operatorname{dist}\left(z_j, [-2,2]\right)$. In (3.2.5) and (3.2.6), the numbers N_0 in the definition of the $\prec$ notation (recall Definition 3.2.1) may depend on k and C_0.

3.2.4 Central limit theorem for GOE

We recall the following result from [141, Theorems 2.3 and 2.4].

Theorem 3.2.9 (Central Limit Theorem for GOE)**.** *Let H be drawn from Gaussian Orthogonal Ensemble and fix $\delta \in (0,1)$. For $N \ge 2$, let $\mathrm{I} = \mathrm{I}_N \in [\![N^{1-\delta}, N]\!]$ be a deterministic sequence of subsets, let $\ell = \ell_N \in [\![1, N]\!]$ be a sequence of indices, and let $\boldsymbol{u} = \boldsymbol{u}_{\ell_N}^{(N)}$ be the corresponding sequence of ℓ^2-normalized eigenvectors of H. Let $\left(\boldsymbol{q}_\alpha^{(N)}\right)_{\alpha \in \mathrm{I}} = \left(\boldsymbol{q}_\alpha\right)_{\alpha \in \mathrm{I}}$ be a deterministic sequence of sets of orthogonal vectors in $\mathbb{S}^{N-1}$. Then*

$$\sqrt{\frac{N^3}{2|\mathcal{I}|(N - |\mathcal{I}|)}} \left(\sum_{\alpha \in \mathcal{I}} \langle \boldsymbol{q}_\alpha, \boldsymbol{u} \rangle^2 - \frac{|\mathcal{I}|}{N} \right) \to \mathcal{N}(0,1),$$

with convergence in the sense of moments.

3.2.5 Quantum Unique Ergodicity

We also recall the following quantum unique ergodicity result from [24, Theorem 1.4].

Theorem 3.2.10. *Let H be a Wigner matrix. Let $\mathrm{I} = \mathrm{I}_N \subset [\![1, N]\!]$ be a deterministic sequence of subsets, let $\ell = \ell_N \subset [\![1, N]\!]$ be a sequence of indices, and let $\boldsymbol{u} = \boldsymbol{u}_{\ell_N}^{(N)}$ be the corresponding sequence of eigenvectors of H. Let $(\boldsymbol{q}_\alpha)_{\alpha \in I} = \left(\boldsymbol{q}_\alpha^{(N)}\right)_{\alpha \in \mathrm{I}}$ be a deterministic sequence of sets of orthogonal vectors in $\mathbb{S}^{N-1}$. Then*

$$\left| \sum_{\alpha \in \mathrm{I}} \langle \boldsymbol{q}_\alpha, \boldsymbol{u} \rangle^2 - \frac{|\mathrm{I}|}{N} \right| \prec \frac{\sqrt{|\mathrm{I}|}}{N}. \tag{3.2.7}$$

3.3 Regularized observables

3.3.1 Definitions

We begin by defining notation for the self-overlaps and typical eigenvalue spacings.

Definition 3.3.1 (Self-overlaps and spacings)**.** *Let H be a $N \times N$ Wigner matrix. Define the self-overlap of the eigenvector $\boldsymbol{u}_k$ by*

$$p_k = p_k(\mathrm{I}) \coloneqq \sum_{\alpha \in \mathcal{I}} \langle \boldsymbol{q}_\alpha, \boldsymbol{u}_k \rangle^2 - \frac{|\mathcal{I}|}{N}.$$

Denote the normalized p_k as

$$\widehat{p}_k \coloneqq \sqrt{\frac{N^3}{2|\mathcal{I}|(N - |\mathcal{I}|)}} p_k.$$

Denote the typical size of the k-th eigenvalue gap by

$$\Delta_k = N^{-2/3} k^{-1/3}.$$

We next define v_ℓ, which serves as the regularization of a scaled version of p_ℓ.

Definition 3.3.2 (Smoothed indicator function)**.** *For any* $E_1, E_2 \in \mathbb{R}$ *with* $E_1 < E_2$*, and* $\eta > 0$*, let* $f_{E_1,E_2,\eta}$ *denote a function such that* $f_{E_1,E_2,\eta} = 1$ *on* $[E_1, E_2]$*,* $f_{E_1,E_2,\eta} = 0$ *on* $\mathbb{R}\backslash[E_1 - \eta, E_2 + \eta]$*, and* $|f'_{E_1,E_2,\eta}| \leq C\eta^{-1}, |f''_{E_1,E_2,\eta}| \leq C\eta^{-2}$ *on* $\mathbb{R}$*.*

Definition 3.3.3 (Regularized self-overlap)**.** *Let* $\delta_i > 0$ *for* $1 \leq i \leq 5$ *be parameters, and let* H *be a Wigner matrix. Define*

$$\begin{gathered}\eta_\ell = \Delta_\ell N^{-\delta_1}, \quad I_\ell = \left[\gamma_\ell - \Delta_\ell N^{\delta_2}, \gamma_\ell + \Delta_\ell N^{\delta_2}\right], \quad E^{\pm} = E \pm \Delta_\ell N^{-\delta_3}, \\ f_E = f_{-3,E^+,\Delta_\ell N^{-\delta_4}}, \quad \tilde{\eta}_\ell = \Delta_\ell N^{-\delta_5}.\end{gathered} \tag{3.3.1}$$

Let $\tilde{f} = f_{-\frac{\kappa}{2},\frac{\kappa}{2},\frac{\kappa}{2}}$*, where* $\kappa = \left(\frac{\ell}{N}\right)^{2/3}$*, and set* $q = f_{\ell-\frac{1}{3},\ell+\frac{1}{3},\frac{1}{3}}$*. Let* $\tilde{f} = f_{-\frac{\kappa}{2},\frac{\kappa}{2},\frac{\kappa}{2}}$*, where* $\kappa = \left(\frac{\ell}{N}\right)^{2/3}$*, and set* $q = f_{\ell-\frac{1}{3},\ell+\frac{1}{3},\frac{1}{3}}$*.*

Define

$$\begin{aligned}x(E) \equiv x_\ell(E) &= \frac{\eta_\ell}{\pi}\sum_i \frac{\hat{p}_i}{(\lambda_i - E)^2 + \eta_\ell^2} \\ &= \frac{\eta_\ell}{\pi}\sqrt{\frac{N^3}{2|\mathrm{I}|(N-|\mathrm{I}|)}}\left[\sum_{\alpha\in\mathrm{I}}\sum_i G_{\boldsymbol{q}_\alpha i}\bar{G}_{i\boldsymbol{q}_\alpha} - \sum_\beta\sum_i G_{\boldsymbol{q}_\beta i}\bar{G}_{i\boldsymbol{q}_\beta}\right] \\ &= \frac{\eta_\ell}{\pi}\sqrt{\frac{N^3}{2|\mathrm{I}|(N-|\mathrm{I}|)}}\left[\sum_{\alpha\in\mathrm{I}}\left(1 - \frac{|\mathrm{I}|}{N}\right)(G\bar{G})_{\boldsymbol{q}_\alpha\boldsymbol{q}_\alpha} - \sum_{\alpha\notin\mathrm{I}}\frac{|\mathrm{I}|}{N}(G\bar{G})_{\boldsymbol{q}_\alpha\boldsymbol{q}_\alpha}\right],\end{aligned} \tag{3.3.2}$$

and

$$\begin{aligned}y(E) \equiv y_\ell(E) = &\frac{1}{2\pi}\int_{\mathbb{R}^2} \mathrm{i}\sigma f''_E(e)\tilde{f}(\sigma)\operatorname{Tr} G(e+\mathrm{i}\sigma)\mathbf{1}\left(|\sigma| > \tilde{\eta}_\ell\right)\mathrm{d}e\ \mathrm{d}\sigma \\ &+ \frac{1}{2\pi}\int_{\mathbb{R}^2}\left(\mathrm{i}f_E(e)\tilde{f}'(\sigma) - \sigma f'_E(e)\tilde{f}'(\sigma)\right)\operatorname{Tr} G(e+\mathrm{i}\sigma)\,\mathrm{d}e\ \mathrm{d}\sigma,\end{aligned} \tag{3.3.3}$$

Finally, set $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_5)$ *and define the regularized observable*

$$v_\ell = v_\ell(\boldsymbol{\delta}, \mathcal{I}) = \int_{I_\ell} x(E) q\left(y_E\right)\mathrm{d}E. \tag{3.3.4}$$

Before stating the main lemma in this section, we need the following several results.

Theorem 3.3.4 (Eigenvalue Rigidity [84, Theorem 2.2]). *Let H be a Wigner matrix. For all $i \in [\![1, N]\!]$, we have*

$$|\lambda_i - \gamma_i| \prec \Delta_i. \tag{3.3.5}$$

Proposition 3.3.5 (Level repulsion at the edge [25, Proposition 5.7]). *Let H be a Wigner matrix. Then there exists $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0)$, there exists constants $C = C(\varepsilon)$ such that for any $i \in [\![1, \lfloor N/2 \rfloor]\!]$,*

$$\mathbb{P}\left(\lambda_{i+1} - \lambda_i < N^{-2/3-\varepsilon} i^{-1/3}\right) \leq CN^{-\varepsilon}. \tag{3.3.6}$$

Lemma 3.3.6 ([25, Lemma 4.9]). *With the definitions in Definition 3.3.3, for any $\varepsilon > 0$, there exists $C = C(\varepsilon)$ such that*[4]

$$\sum_{i:|i-\ell|\geqslant N^{\varepsilon}} \frac{1}{(\lambda_i - \lambda_\ell)^2} \prec N^{4/3-\varepsilon} \ell^{2/3}. \tag{3.3.7}$$

Lemma 3.3.7. *With I_ℓ defined in (3.3.1) and $x(E)$ defined in (3.3.2), we have*

$$\int_{I_\ell} |x(E)| \chi(E) \, \mathrm{d}E \prec N^{\delta_2}, \tag{3.3.8}$$

where $\chi(E) = \mathbf{1}(\lambda_\ell \leq E^+ \leq \lambda_{\ell+1})$.

Proof. By the QUE bound (3.2.7), it suffices to show

$$\sum_i \int_{I_\ell} \frac{\eta_\ell}{\pi} \frac{1}{(\lambda_i - E)^2 + \eta_\ell^2} \chi(E) \, \mathrm{d}E \prec N^{\delta_2}. \tag{3.3.9}$$

[4] There is a misprint in [25, Lemma 4.9]. The sign of the ω on the right hand side of the inequality should be negative.

We break the integral (3.3.9) into two parts, and find that it equals

$$\sum_{i:|i-\ell|<N^{\delta_2}} \int_{I_\ell} \frac{\eta_\ell}{\pi} \frac{1}{(\lambda_i - E)^2 + \eta_\ell^2} \chi(E)\, \mathrm{d}E + \sum_{i:|i-\ell|\geq N^{\delta_2}} \int_{I_\ell} \frac{\eta_\ell}{\pi} \frac{1}{(\lambda_i - E)^2 + \eta_\ell^2} \chi(E)\, \mathrm{d}E. \tag{3.3.10}$$

For the first term in (3.3.10), using the integral

$$\int \frac{\eta_\ell}{E^2 + \eta_\ell^2} \mathrm{d}E = \pi,$$

we can bound it by

$$\sum_{i:|i-\ell|<N^{\delta_2}} \int_{I_\ell} \frac{\eta_\ell}{\pi} \frac{1}{(\lambda_i - E)^2 + \eta_\ell^2} \chi(E) \mathrm{d}E < 2N^{\delta_2}. \tag{3.3.11}$$

For the second term in (3.3.10), using (3.3.7), it follows that

$$\sum_{i:|i-\ell|\geq N^{\delta_2}} \int_{I_\ell} \frac{\eta_\ell}{\pi} \frac{1}{(\lambda_i - E)^2 + \eta_\ell^2} \chi(E) \mathrm{d}E \prec N^{-\delta_2-\delta_1}. \tag{3.3.12}$$

Combining (3.3.11) and (3.3.12), the proof is complete. □

The following lemma is our main comparison result for the smoothed observable v_ℓ.

Lemma 3.3.8. *Let H be a Wigner matrix and suppose that λ_ℓ satisfies level repulsion estimate* (3.3.6) *with*

$$\varepsilon = \varepsilon_1 := \min\{\varepsilon_0/2, \tau/10000\}.$$

Fix parameters

$$\delta_1 = 2\varepsilon_1,\ \delta_2 = \frac{\varepsilon_1}{\max\{C_0+1, 100\}},\ \delta_3 = \frac{\varepsilon_1}{2},\ \delta_4 = 6\varepsilon_1,\ \delta_5 = 8\varepsilon_1.$$

Fix $C_0 > 0$, and let $g : \mathbb{R} \to \mathbb{R}$ be a smooth function satisfying

$$|g'(x)| \leq C_0(1+|x|)^{C_0}. \tag{3.3.13}$$

Then there exists a constant $c(\tau) > 0$ *such that*

$$\left|\mathbb{E}\left[g\left(\hat{p}_\ell(\mathcal{I})\right)\right] - \mathbb{E}\Big[g\Big(v_\ell(\boldsymbol{\delta}, \mathcal{I})\Big)\Big]\right| \le c^{-1}N^{-c}.$$

To prove Lemma 3.3.8, it suffices to establish Lemmas 3.3.9, 3.3.10 and 3.3.11.

Lemma 3.3.9. *Maintain the notation and assumptions of Lemma 3.3.8. Recalling Definition 3.3.3, we have*

$$\mathbb{E}[g(\widehat{p}_\ell)] - \mathbb{E}\left[g\left(\int_{I_\ell} x(E)\chi(E)\right)\right] = O(N^{-\varepsilon_1/4}),$$

where $\chi(E) := \mathbf{1}\left(\lambda_\ell \le E^+ \le \lambda_{\ell+1}\right)$*.*

Proof. We first write

$$\widehat{p}_\ell = \frac{\eta_\ell}{\pi}\int \frac{\widehat{p}_\ell}{(E-\lambda_\ell)^2 + \eta_\ell^2}\, \mathrm{d}E.$$

By the assumption (3.3.13) on g, rigidity (3.3.5), and quantum unique ergodicity (3.2.7), we write

$$\begin{aligned}\mathbb{E}\left[g(\hat{p}_\ell)\right] =& \mathbb{E}\left[g\left(\frac{\eta_\ell}{\pi}\int_{E_1}^{E_2} \frac{\hat{p}_\ell}{(E-\lambda_\ell)^2+\eta_\ell^2}\mathrm{d}E\right)\right] + O_\prec(N^{-\delta_1+\delta_3})\\ =& \mathbb{E}\left[g\left(\frac{\eta_\ell}{\pi}\int_{E_1}^{E_2} \frac{\hat{p}_\ell}{(E-\lambda_\ell)^2+\eta_\ell^2}\mathrm{d}E\right)\right] + O(N^{-\varepsilon_1/2}),\end{aligned}$$

where

$$E_1 = \lambda_\ell^-, \quad E_2 = \max\left\{\lambda_\ell^+, \lambda_{\ell+1}^-\right\}.$$

We now show that the integral over $[E_1, E_2]$ can be approximated by integrating over $[\lambda_\ell^-, \lambda_{\ell+1}^-]$. From (3.3.6) and the parameter choice $\delta_3 = \varepsilon_1/2$, we have

$$\mathbb{P}\left(\lambda_{\ell+1}^- \le \lambda_\ell^+\right) \le N^{-\varepsilon_1/2}.$$

Decomposing the integral

$$\int_{E_1}^{E_2} = \int_{\lambda_\ell^-}^{\lambda_{\ell+1}^-} + \mathbf{1}\left(\lambda_{\ell+1}^- \le \lambda_\ell^+\right) \int_{\lambda_{\ell+1}^-}^{\lambda_\ell^+}, \tag{3.3.14}$$

we have

$$\begin{aligned}\mathbb{E}\Big[g(\hat{p}_\ell)\Big] =& \mathbb{E}\left[g\left(\frac{\eta_\ell}{\pi}\int_{\lambda_\ell^-}^{\lambda_{\ell+1}^-} \frac{\hat{p}_\ell}{(E-\lambda_\ell)^2+\eta_\ell^2}\mathrm{d}E\right)\right] + O_\prec(\mathbb{P}\left(\lambda_{\ell+1}^- \le \lambda_\ell^+\right)) + O(N^{-\varepsilon_1/2}) \\ =& \mathbb{E}\left[g\left(\frac{\eta_\ell}{\pi}\int_{\lambda_\ell^-}^{\lambda_{\ell+1}^-} \frac{\hat{p}_\ell}{(E-\lambda_\ell)^2+\eta_\ell^2}\mathrm{d}E\right)\right] + O(N^{-\varepsilon_1/4})\end{aligned}$$

where we used the polynomial growth of g (3.3.13) and the quantum unique ergodicity (3.2.7) in the first step and (3.3.14) in the second step. By rigidity (3.3.5),

$$|\lambda_\ell - \gamma_\ell| \prec \Delta_\ell, \quad |\lambda_{\ell+1} - \gamma_\ell| \prec \Delta_\ell,$$

we have

$$\mathbb{E}\left[g(\hat{p}_\ell)\right] = \mathbb{E}\left[g\left(\frac{\eta_\ell}{\pi}\int_{I_\ell} \frac{\hat{p}_\ell}{(E-\lambda_\ell)^2+\eta_\ell^2}\chi(E)\mathrm{d}E\right)\right] + O(N^{-\varepsilon_1/4}). \tag{3.3.15}$$

Our next goal is to replace the first term on the right hand side of (3.3.15) by the following.

$$\mathbb{E}g\left[\int_{I_\ell} x(E)\chi(E)\right] = \mathbb{E}g\left[\frac{\eta_\ell}{\pi}\int_{I_\ell} \frac{\hat{p}_\ell}{(E-\lambda_\ell)^2+\eta_\ell^2}\chi(E)\,\mathrm{d}E + \frac{\eta_\ell}{\pi}\sum_{i\neq\ell}\int_{I_\ell} \frac{\hat{p}_i}{(\lambda_i - E)^2+\eta_\ell^2}\chi(E)\,\mathrm{d}E\right].$$

Using mean value theorem, (3.3.8) and (3.3.15) in the first step, and (3.3.7) and (3.2.7) in the second step, we have

$$
\begin{aligned}
&\left|\mathbb{E}\left[g(\hat{p}_\ell)\right]-\mathbb{E}\left[g\left(\int_{I_\ell} x(E)\chi(E)\mathrm{d}E\right)\right]\right| \\
\leq & O_\prec(N^{C_0\delta_2})\left|\mathbb{E}\left[\sum_{i\neq\ell}\int_{I_\ell}\frac{\eta_\ell}{\pi}\frac{\hat{p}_i}{(\lambda_i-E)^2+\eta_\ell^2}\chi(E)\mathrm{d}E\right]\right|+O(N^{-\varepsilon_1/4}) \\
= & O_\prec(N^{C_0\delta_2})\mathbb{E}\left[\sum_{i:1\leq|i-\ell|<N^{\delta_2}}\int_{I_\ell}\frac{\eta_\ell}{\pi}\frac{1}{(\lambda_i-E)^2+\eta_\ell^2}\chi(E)\mathrm{d}E\right]+O_\prec(N^{C_0\delta_2-\delta_1})+O(N^{-\varepsilon_1/4}) \\
= & O_\prec(N^{C_0\delta_2})\mathbb{E}\left[\sum_{i:1\leq|i-\ell|<N^{\delta_2}}\int_{I_\ell}\frac{\eta_\ell}{\pi}\frac{1}{(\lambda_i-E)^2+\eta_\ell^2}\chi(E)\mathrm{d}E\right]+O(N^{-\varepsilon_1/4}).
\end{aligned}
\tag{3.3.16}
$$

Next we would like to bound the expectation term in the right hand side of (3.3.16). We decompose it into two parts.

Firstly, for $i>\ell$, we have

$$
\mathbb{E}\left[\sum_{i:1\leq i-\ell<N^{\delta_2}}\int_{I_\ell}\frac{\eta_\ell}{\pi}\frac{1}{(\lambda_i-E)^2+\eta_\ell^2}\chi(E)\mathrm{d}E\right]\leq N^{\delta_2}\mathbb{E}\left[\int_{-\infty}^{\lambda_{\ell+1}^-}\frac{\eta_\ell}{\pi}\frac{1}{(\lambda_{\ell+1}-E)^2+\eta_\ell^2}\mathrm{d}E\right]
$$
$$
\prec N^{\delta_2-\delta_1+\delta_3}. \tag{3.3.17}
$$

Suppose now that $i<\ell$. On the event $\mathcal{B}:=\mathbf{1}(\lambda_\ell-\lambda_{\ell-1}>4\Delta_\ell N^{-\delta_3})$, we have

$$
\chi(E)(E-\lambda_i)^2\geq(\lambda_\ell-\lambda_i)^2-2\Delta_\ell N^{-\delta_3}(\lambda_\ell-\lambda_i)\geq\frac{1}{2}(\lambda_\ell-\lambda_i)^2\geq\frac{1}{2}(\lambda_\ell-\lambda_{\ell-1})^2.
$$

Therefore,

$$
\mathbf{1}(\mathcal{B})\chi(E)\frac{1}{(E-\lambda_i)^2+\eta_\ell^2}\leq\mathbf{1}(\mathcal{B})\chi(E)\frac{2}{(\lambda_\ell-\lambda_{\ell-1})^2+\eta_\ell^2}\leq\chi(E)\frac{N^{2\delta_3}}{8\Delta_\ell^2}.
$$

Now we have,

$$
\mathbb{E}\left[\sum_{i:1\leq\ell-i<N^{\delta_2}}\int_{I_\ell}\frac{\eta_\ell}{\pi}\frac{1}{(\lambda_i-E)^2+\eta_\ell^2}\chi(E)\mathrm{d}E\right]\prec N^{\delta_2}\mathbb{P}\left(\mathcal{B}^c\right)+N^{\delta_2}\eta_\ell\Delta_\ell\frac{N^{2\delta_3}}{\Delta_\ell^2}\leq N^{\delta_2-\frac{2}{3}\delta_3},
\tag{3.3.18}
$$

where we used $\lambda_{\ell+1}-\lambda_\ell \prec \Delta_\ell$ in the first inequality and (3.3.6) in the second inequality. Combining (3.3.17) and (3.3.18), we have

$$\mathbb{E}\left[\sum_{i:1\le|i-\ell|<N^{\delta_2}}\int_{I_\ell}\frac{\eta_\ell}{\pi}\frac{1}{(\lambda_i-E)^2+\eta_\ell^2}\chi(E)\mathrm{d}E\right]\prec N^{\delta_2-\frac{2}{3}\delta_3}. \tag{3.3.19}$$

Inserting (3.3.19) into (3.3.16), we have

$$\left|\mathbb{E}\left[g(\hat{p}_\ell)\right]-\mathbb{E}\left[g\left(\int_{I_\ell}x(E)\chi(E)\mathrm{d}E\right)\right]\right| = O(N^{-\varepsilon_1/4}).$$

$\square$

Lemma 3.3.10. *Maintain the same assumptions as in Lemma 3.3.8. Using Definition 3.3.3, the following holds*

$$\mathbb{E}\left[g\left(\int_{I_\ell}x(E)\chi(E)\,\mathrm{d}E\right)\right]-\mathbb{E}\left[g\left(\int_{I_\ell}x(E)q\left(\operatorname{Tr}f_E(H)\right)\,\mathrm{d}E\right)\right]=O(N^{-\varepsilon_1/2}),$$

where $\chi(E):=\mathbf{1}\left(\lambda_\ell\le E^+\le\lambda_{\ell+1}\right)$.

Proof. Let $\theta:=\mathbf{1}_{[-3,E^+]}$ and $\mathcal{B}$ denote the event $\{\lambda_1\ge -3\}$. By the definition of $\chi(E)$, the first term becomes

$$\begin{aligned}\mathbf{1}(\mathcal{B})\int_{I_\ell}x(E)\,\mathbf{1}(\lambda_\ell\le E^+\le\lambda_{\ell+1})\,\mathrm{d}E &= \mathbf{1}(\mathcal{B})\int_{I_\ell}x(E)\,\mathbf{1}(\mathcal{N}(-\infty,E^+)=\ell)\,\mathrm{d}E\\ &= \mathbf{1}(\mathcal{B})\int_{I_\ell}x(E)\,q(\operatorname{Tr}\theta(H))\,\mathrm{d}E,\end{aligned}$$

where $\mathcal{N}(E_1,E_2)$ denotes the number of eigenvalues in $[E_1,E_2]$.

By the definition of $f_E=f_{-3,E^+,\Delta_\ell N^{-\delta_4}}$, we know

$$\mathbf{1}(\mathcal{B})|\operatorname{Tr}\theta(H)-\operatorname{Tr}f_E(H)|\le\mathcal{N}(E^+,E^++\Delta_\ell N^{-\delta_4})=\sum_i\mathbf{1}(|\lambda_i-E^+|\le\Delta_\ell N^{-\delta_4}).$$

Hence we have

$$
\begin{aligned}
&\mathbf{1}(\mathcal{B})\left|\int_{I_\ell} x(E)\chi(E)\,\mathrm{d}E-\int_{I_\ell} x(E)\,q(\operatorname{Tr} f_E(H))\,\mathrm{d}E\right|\\
=&\mathbf{1}(\mathcal{B})\left|\int_{I_\ell} x(E)\,[q(\operatorname{Tr}\theta(H))-q(\operatorname{Tr} f_E(H))]\,\mathrm{d}E\right|\\
\leq& C\mathbf{1}(\mathcal{B})\int_{I_\ell}|x(E)|\,|\operatorname{Tr}\theta(H)-\operatorname{Tr} f_E(H)|\,\mathrm{d}E\\
\leq& C\sum_i\int_{I_\ell}|x(E)|\mathbf{1}(|\lambda_i-E^+|\leq\Delta_\ell N^{-\delta_4})\mathrm{d}E\\
\prec& N^{-\delta_4/2}\Delta_\ell\sup_{E\in I_\ell}|x(E)|,
\end{aligned}
\tag{3.3.20}
$$

where we use the rigidity (3.3.5) in the final step. By (3.2.7) and (3.3.7) with $\varepsilon=2\delta_2$, we have

$$
\sup_{E\in I_\ell}|x(E)|\prec\Delta_\ell^{-1}N^{\delta_1+2\delta_2}. \tag{3.3.21}
$$

Combining (3.3.20) and (3.3.21), we obtain the desired bound on $\mathcal{B}$. On $\mathcal{B}^c$, we simply use the rigidity (3.3.5) and (3.3.21). The proof is complete. $\square$

Lemma 3.3.11. *Maintain the same assumptions as in Lemma 3.3.8. Using Definition 3.3.3, the following holds*

$$
\mathbb{E}\left[g\left(\int_{I_\ell} x(E)q(\operatorname{Tr} f_E(H))\right)\right]-\mathbb{E}\left[g\left(\int_{I_\ell} x(E)q(y_E)\right)\right]=O(N^{-\varepsilon_1/2}).
$$

Proof. We first express $f_E(H)$ in terms of Green functions using Helffer-Sjöstrand functional calculus (see Equation (B.12) of [78]):

$$
f_E(\lambda)\;=\;\frac{1}{2\pi}\int_{\mathbb{R}^2}\frac{\mathrm{i}\sigma f_E''(e)\tilde{f}(\sigma)+\mathrm{i}f_E(e)\tilde{f}'(\sigma)-\sigma f_E'(e)\tilde{f}'(\sigma)}{\lambda-e-\mathrm{i}\sigma}\,\mathrm{d}e\,\mathrm{d}\sigma.
$$

Let $G(z) = (H - z)^{-1}$ and recall $\tilde{\eta}_\ell = \Delta_\ell N^{-\delta_5}$, then we have:

$$
\begin{aligned}
\operatorname{Tr} f_E(H) &= \frac{1}{2\pi}\int_{\mathbb{R}^2}\Big(\mathrm{i}\sigma f_E''(e)\tilde{f}(\sigma) + \mathrm{i} f_E(e)\tilde{f}'(\sigma) - \sigma f_E'(e)\tilde{f}'(\sigma)\Big)\operatorname{Tr} G(e+\mathrm{i}\sigma)\,\mathrm{d}e\,\mathrm{d}\sigma \\
&= \frac{1}{2\pi}\int_{\mathbb{R}^2}\Big(\mathrm{i} f_E(e)\tilde{f}'(\sigma) - \sigma f_E'(e)\tilde{f}'(\sigma)\Big)\operatorname{Tr} G(e+\mathrm{i}\sigma)\,\mathrm{d}e\,\mathrm{d}\sigma \\
&+ \frac{1}{2}\int_{|\sigma|>\tilde{\eta}_\ell}\int \mathrm{i}\sigma f_E''(e)\tilde{f}(\sigma)\operatorname{Tr} G(e+\mathrm{i}\sigma)\,\mathrm{d}e\,\mathrm{d}\sigma \\
&+ \frac{1}{2\pi}\int_{|\sigma|<\tilde{\eta}_\ell}\int \mathrm{i}\sigma f_E''(e)\tilde{f}(\sigma)\operatorname{Tr} G(e+\mathrm{i}\sigma)\,\mathrm{d}e\,\mathrm{d}\sigma.
\end{aligned}
$$

We show that the last term is negligible. From [119, Lemma 5.1], we know $\sigma \operatorname{Tr} G(e+\mathrm{i}\sigma) = O_\prec(1)$. Since $\int |f_E''(e)| = O(\Delta_\ell^{-1}N^{\delta_4})$, the last term is bounded by:

$$
\left|\frac{1}{2\pi}\int_{|\sigma|<\tilde{\eta}_\ell}\int \mathrm{i}\sigma f_E''(e)\tilde{f}(\sigma)\operatorname{Tr} G(e+\mathrm{i}\sigma)\,\mathrm{d}e\,\mathrm{d}\sigma\right| = O_\prec(N^{\delta_4-\delta_5}).
$$

Hence by the mean value theorem on q and definition of y_E (see (3.3.3)), we know $q(\operatorname{Tr} f_E(H)) - q(y_E) = O_\prec(N^{\delta_4-\delta_5})$. Now by mean value theorem on g and (3.3.8), we have

$$
\mathbb{E}\left[g\left(\int_{I_\ell} x(E) q\left(\operatorname{Tr} f_E(H)\right)\right)\right] - \mathbb{E}\left[g\left(\int_{I_\ell} x(E) q\left(y_E\right)\right)\right] = O_\prec\left(N^{(C_0+1)\delta_2+\delta_4-\delta_5}\right),
$$

which, by the choice of parameters, completes the proof. □

3.4 Two moment comparison and proof of the main theorem

We retain the conventions stated in Section 3.2.1 and the choice of parameters made in the statement of Lemma 3.3.8.

3.4.1 Preliminary Lemmas

For $z \in \mathbb{C}\backslash\mathbb{R}$, we define the control parameter

$$\Psi(z) = \sqrt{\frac{\operatorname{Im} m_{\mathrm{sc}}(z)}{N \operatorname{Im} z}} + \frac{1}{N|\operatorname{Im} z|}. \tag{3.4.1}$$

Let H be a $N \times N$ Wigner matrix. Fix $a, b \in [\![1, N]\!]$, and define

$$Q = Q(a,b) = \{q_{ij}\}_{1 \le i,j \le N} \in \operatorname{Mat}_N$$

as follows. Set $q_{ij} = h_{ij}$ if $(i,j) \notin \{(a,b),(b,a)\}$, and set $q_{ij} = 0$ otherwise. In other words, Q is the matrix obtained by starting with H and replacing entries h_{ab} and h_{ba} by zeros. Given $z \in \mathbb{C} \setminus \mathbb{R}$, set

$$G = (H - z)^{-1}, \qquad R = (Q - z)^{-1}. \tag{3.4.2}$$

Let x^G, x^R, y^G, y^R be the quantities in (3.3.2) and (3.3.3) defined using the resolvents G or R, as indicated by the superscript. Finally, let $U = H - Q$.

We summarize some preliminary bounds in the following lemma.

Lemma 3.4.1. *Let $H = \{h_{ij}\}_{i,j=1}^N, G, R, \Psi$ be as defined above. We have*

$$|h_{ij}| \prec N^{-1/2}, \tag{3.4.3}$$

$$\|G(z)\| \le \frac{1}{\eta}, \tag{3.4.4}$$

$$\|R(z)\| \le \frac{1}{\eta}, \tag{3.4.5}$$

$$C^{-1}\tau^{1/4}N^{-1/2} \le \Psi(z) \le C\tau^{-1/4}N^{-\tau/20}, \quad \forall\, z \in \boldsymbol{S}, \tag{3.4.6}$$

for some constant $C > 0$, where $\eta = |\Im z|$. Moreover, when $z = E + \mathrm{i}\eta_\ell$, $z \in I_\ell$ (see

Definition 3.3.3), there exists constant $C > 0$ such that

$$\Psi(z) \leq \frac{C}{N\eta_\ell}. \tag{3.4.7}$$

Proof. Fix any $\varepsilon > 0, D > 0$. By Markov's inequality and the moment assumption (3.1.5) on h_{ij}, we have

$$\mathbb{P}\left(|\sqrt{N}h_{ij}| > N^\varepsilon\right) = \mathbb{P}\left(|\sqrt{N}h_{ij}|^p > N^{p\varepsilon}\right) \leq \mu_p N^{-p\varepsilon} \leq N^{-D},$$

for large enough p and $N > N_0(\varepsilon, D)$. This proves (3.4.3).

By spectral decomposition, we have

$$G(z) = \sum_i \frac{\boldsymbol{u}_i \boldsymbol{u}_i^\top}{\lambda_i - z},$$

where $\{\lambda_i\}_{i=1}^N$ and $\{\boldsymbol{u}_i\}_{i=1}^N$ are the corresponding eigenvalues and eigenvectors of H. Observe that

$$\left|\frac{1}{\lambda_i - z}\right| \leq \frac{1}{\eta}.$$

Pick any unit vector $\boldsymbol{x}$. We have

$$|\boldsymbol{x}^* G \boldsymbol{x}| \leq \frac{1}{\eta} \boldsymbol{x}^* \sum_i \boldsymbol{u}_i \boldsymbol{u}_i^\top \boldsymbol{x} = \frac{1}{\eta}.$$

This proves (3.4.4). The inequality (3.4.5) follows similarly.

To obtain the lower bound of $\Psi(z)$, we use the following result (see [20, Lemma 3.3]):

$$\begin{aligned} c\sqrt{\kappa + \eta} \leq |\operatorname{Im} m_{\mathrm{sc}}(z)| &\leq c^{-1}\sqrt{\kappa + \eta}, && \text{if } |E| \leq 2, \\ \frac{c\eta}{\sqrt{\kappa + \eta}} \leq |\Im m_{\mathrm{sc}}(z)| &\leq \frac{c^{-1}\eta}{\sqrt{\kappa + \eta}}, && \text{if } |E| \geq 2, \end{aligned}$$

for some constant $c > 0$, where $E = \Re z$ and $\kappa \equiv \kappa(E) := ||E| - 2|$.

For $|E| \le 2$, we have

$$\Psi(z) \ge \sqrt{\frac{c\sqrt{\kappa+\eta}}{N\eta}} + \frac{1}{N\eta} \ge CN^{-1/2}\eta^{-1/4} \ge C\tau^{1/4}N^{-1/2},$$

$$\Psi(z) \le \sqrt{\frac{c^{-1}\sqrt{\kappa+\eta}}{N\eta}} + \frac{1}{N\eta} \le C\sqrt{\frac{\tau^{-1/2}}{NN^{-1+\tau/10}}} + \frac{1}{NN^{-1+\tau/10}} \le C\tau^{-1/4}N^{-\tau/20},$$

where we used $z \in \boldsymbol{S}$ in the last step of the first line and in the second inequality of the second line.

For $|E| \ge 2$, we have

$$\Psi(z) \ge \sqrt{\frac{c}{\sqrt{\kappa+\eta}N}} + \frac{1}{N\eta} \ge \left\{\begin{matrix} CN^{-1/2}\eta^{-1/4} \text{ if } \kappa \le \eta \\ CN^{-1/2}\kappa^{-1/4} \text{ if } \kappa \ge \eta \end{matrix}\right\} \ge C\tau^{1/4}N^{-1/2},$$

$$\Psi(z) \le \sqrt{\frac{c^{-1}}{\sqrt{\kappa+\eta}N}} + \frac{1}{N\eta} \le C\sqrt{\frac{1}{N^{-1/2+\tau/20}N}} + \frac{1}{NN^{-1+\tau/10}} \le CN^{-1/2-\tau/20},$$

where we used $z \in \boldsymbol{S}$ in the last step of the first line and in the second inequality of the second line. This completes the proof of (3.4.6).

Finally, for $z = E + \mathrm{i}\eta_\ell$, $E \in I_\ell$. We have

$$\Psi(z) \le C\sqrt{\frac{\sqrt{\kappa+\eta_\ell}}{N\eta}} + \frac{1}{N\eta_\ell}.$$

Note that $\kappa \le C(\ell/N)^{2/3} + N^{\delta_2}\Delta_\ell$. It is not hard to check that $\kappa + \eta_\ell \le C(N\eta_\ell)^{-2}$. This completes the proof of (3.4.7). $\square$

In the next lemma, we collect several local laws, which will be used frequently for the current section and Section 3.5.

Lemma 3.4.2. *Let H be a Wigner matrix, and let S be either G or R, as defined in* (3.4.2)*.*

1. *For all $z \in \boldsymbol{S}$, we have*

$$\left|\langle \boldsymbol{x}, S\boldsymbol{y}\rangle - \langle \boldsymbol{x}, \boldsymbol{y}\rangle m_{\mathrm{sc}}\right| \prec \Psi, \quad |(SS)_{\boldsymbol{x}\boldsymbol{y}}| \prec N\Psi^2, \quad |(S\bar{S})_{\boldsymbol{x}\boldsymbol{y}}| \prec N\Psi^2. \tag{3.4.8}$$

uniformly over all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{S}^{N-1}$.

2. *Further, if $z = E + \mathrm{i}\eta \in \boldsymbol{S}$ satisfies*

$$E \in I_\ell, \quad \eta = \eta_\ell.$$

Then for any deterministic $A \in \mathrm{Mat}_N$ such that $\|A\| \leq 1$ and $\operatorname{Tr} A = 0$, we have

$$|(SA\bar{S})_{cd}| \prec N^{1/2}\Psi, \tag{3.4.9}$$

$$|(SA\bar{S}S)_{cd}| \prec N^{3/2}\Psi^{9/4}, \tag{3.4.10}$$

uniformly over all $c, d \in [\![1, N]\!]$.

The proof of Lemma 3.4.2 is postponed to Section 3.6.

The resolvent expansion formulae

$$\begin{aligned} G &= R - RUR + RURUR - RURURUR + (RU)^4 G, \\ R &= G - GUG + GUGUG - GUGUGUG + (GU)^4 R, \end{aligned} \tag{3.4.11}$$

follow immediately from the definitions of G, R, and U. They imply

$$\begin{aligned} G\bar{G} - R\bar{R} = &- R\overline{RUR} - RUR\bar{R} \\ &+ R\overline{RURUR} + RUR\overline{RUR} + RURUR\bar{R} \\ &- R\overline{RURURUR} - RUR\overline{RURUR} - RURUR\overline{RUR} - RURURUR\bar{R} \\ &+ R\overline{(RU)^4 G} + RUR\overline{(RU)^3 R} + (RU)^2 R\overline{(RU)^2 R} + (RU)^3 R\overline{RUR} + (RU)^4 G\bar{R}. \end{aligned} \tag{3.4.12}$$

These identities facilitate resolvent expansions for the terms x^G and y^G, which are stated in the following lemma. We recall that ℓ was fixed earlier in Section 3.2.1, and that I_ℓ was defined in (3.3.1).

Lemma 3.4.3. *Let H be a $N \times N$ Wigner matrix.*

1. *We have*

$$x^G - x^R = \sum_{r=1}^{3} x_r h_{ab}^r + x_{\mathrm{err}},$$

where

$$|x_i(E)| \prec N^{1+\delta/2}\Psi^{5/4}, \quad i = 1,2,3, \quad |x_{\mathrm{err}}(E)| \prec N^{-1+\delta/2}\Psi^{5/4},$$

uniformly over $E \in I_\ell$. Moreover

$$|x^R(E)| \prec N^{1+\delta/2}\Psi^{1/2},$$

uniformly over $E \in I_\ell$.

2. *We have*

$$\operatorname{Tr} G - \operatorname{Tr} R = \sum_{r=1}^{3} J_r h_{ab}^r + J_{err}$$

where

$$|J_i(z)| \prec N\Psi^2, \quad i = 1,2,3, \quad |J_{\mathrm{err}}(z)| \prec N^{-2}\Psi^2,$$

uniformly for $z \in \boldsymbol{S}$.

3. *We have*

$$y^G - y^R = \sum_{r=1}^{3} y_r h_{ab}^r + y_{\mathrm{err}}$$

where

$$|y_i(E)| \prec \Psi, \quad i = 1,2,3, \quad |y_{\mathrm{err}}(E)| \prec N^{-2}\Psi,$$

uniformly over $E \in I_\ell$.

Proof. We first present the proofs for the claims about x_1 and x^R. The others are similar.

By (3.3.2) and (3.4.12), we have

$$x_1 = \frac{\eta_\ell}{\pi}\sqrt{\frac{N^3}{2|\mathrm{I}|(N-|\mathrm{I}|)}}\frac{1}{1+\delta_{ab}}[(RA\bar{R}R)_{ab} + (RA\bar{R}R)_{ba}] + [C],$$

where $[C]$ denotes the complex conjugate of the previous terms, and

$$A = \sum_{\alpha \in \mathrm{I}}\left(1 - \frac{|\mathrm{I}|}{N}\right)\boldsymbol{q}_\alpha \boldsymbol{q}_\alpha^\top - \sum_{\alpha \notin \mathrm{I}} \frac{|\mathrm{I}|}{N}\boldsymbol{q}_\alpha \boldsymbol{q}_\alpha^\top. \tag{3.4.13}$$

Note that $\|A\| = \max\{1 - |\mathrm{I}|/N, |\mathrm{I}|/N\} \le 1$ and $\operatorname{Tr} A = 0$. By Lemma 3.4.2, we have that $|x_1| \prec N^{1+\delta/2}\Psi^{5/4}$, where we used $|\mathrm{I}| \ge N^{1-\delta}$. For x^R we have,

$$x^R = \frac{\eta_\ell}{\pi}\sqrt{\frac{N^3}{2|\mathrm{I}|(N-|\mathrm{I}|)}}\operatorname{Tr}(RA\bar{R}).$$

By definition, we have $M(z, A, \bar{z}) = |m_{\mathrm{sc}}(z)|^2 A$ which is traceless and from (3.2.5), we have $|x^R| \prec N^{1+\delta/2}\Psi^{1/2}$.

By resolvent expansion, we have

$$\operatorname{Tr} S - \operatorname{Tr} R = -\operatorname{Tr} RVR + \operatorname{Tr} RVRVR - \operatorname{Tr} RVRVRVR + \operatorname{Tr}(RV)^4 G =: \sum_{i=1}^{3} J_i + J_{\mathrm{err}}.$$

Using the high probability bound (3.4.8) we have

$$|J_i| \prec N\Psi^2, \quad i = 1, 2, 3, \quad \text{and} \quad |J_{\mathrm{err}}| \prec N^{-1}\Psi^2.$$

Using the bound on J_i from the second part, and notice that

$$\begin{aligned}
y_i =& \frac{1}{2\pi}\int_{\mathbb{R}^2} \mathrm{i}\sigma f_E''(e)\widetilde{f}(\sigma)J_i(e+\mathrm{i}\sigma)\mathbf{1}\left(|\sigma|>\widetilde{\eta}_\ell\right)\mathrm{d}e\,\mathrm{d}\sigma\\
&+\frac{1}{2\pi}\int_{\mathbb{R}^2}\left(\mathrm{i}f_E(e)\widetilde{f}'(\sigma)-\sigma f_E'(e)\widetilde{f}'(\sigma)\right)J_i(e+\mathrm{i}\sigma)\mathrm{d}e\,\mathrm{d}\sigma,\\
y_{\mathrm{err}} =& \frac{1}{2\pi}\int_{\mathbb{R}^2} \mathrm{i}\sigma f_E''(e)\widetilde{f}(\sigma)J_{\mathrm{err}}(e+\mathrm{i}\sigma)\mathbf{1}\left(|\sigma|>\widetilde{\eta}_\ell\right)\mathrm{d}e\,\mathrm{d}\sigma\\
&+\frac{1}{2\pi}\int_{\mathbb{R}^2}\left(\mathrm{i}f_E(e)\widetilde{f}'(\sigma)-\sigma f_E'(e)\widetilde{f}'(\sigma)\right)J_{\mathrm{err}}(e+\mathrm{i}\sigma)\mathrm{d}e\,\mathrm{d}\sigma,
\end{aligned}$$

the proof of the bound on y_i and y_{err} follows from the proof of [33, Lemma 7.7]. $\square$

Using Lemma 3.4.3, and Taylor expansion of $q(y^G)$ around y^R up to order 3, we have

$$\begin{aligned}
&\int_{I_\ell} x^G q(y^G)\mathrm{d}E-\int_{I_\ell} x^R q(y^R)\mathrm{d}E\\
=&\int_{I_\ell}\left(x^R+\sum_{r=1}^3 x_r h_{ab}^r + x_{\mathrm{err}}\right)\cdot\left(q(y^R)+\sum_{k=1}^3\frac{q^{(k)}(y^R)}{k!}\left(\sum_{r=1}^3 y_r h_{ab}^r+y_{\mathrm{err}}\right)^k+O_\prec(N^{-2}\Psi)\right)\mathrm{d}E\\
&-\int_{I_\ell} x^R q(y^R)\mathrm{d}E\\
=&\sum_{\boldsymbol{l}\in\mathcal{L}} P_{\boldsymbol{l}} h_{ab}^{|\boldsymbol{l}|}+O_\prec(N^{-2-c}),
\end{aligned} \tag{3.4.14}$$

where we define

$$\mathcal{L}:=\left\{\boldsymbol{l}=(l_0,\ldots,l_k)\in[\![0,3]\!]\times[\![1,3]\!]^k : k\in\mathbb{N}, 1\leqslant|\boldsymbol{l}|\leqslant 3\right\},\quad |\boldsymbol{l}|:=\sum_{i=0}^k l_i,$$
$$P_{\boldsymbol{l}}:=\int_{I_\ell}\frac{q^{(k)}\left(y^R\right)}{k!}x_{l_0}y_{l_1}\cdots y_{l_k}\,\mathrm{d}E,$$

and $c>0$ depends only on τ. Here we denote $x_0:=x^R$.

Using (3.4.14), for g satisfying (3.3.13), we have

$$\begin{aligned}
\mathbb{E}\left[g\left(\int_{I_\ell}x^G q\left(y^G\right)\mathrm{d}E\right)\right]-\mathbb{E}\left[g\left(\int_{I_\ell}x^R q\left(y^R\right)\mathrm{d}E\right)\right]=\mathbb{E}\left[\mathcal{A}\right]+O_\prec\left(N^{-2-c}\right)\\
+\mathbb{E}\left[h_{ab}^3\right]\mathbb{E}\left[\sum_{k=1}^3\frac{1}{k!}g^{(k)}\left(P_{\mathbf{0}}\right)\sum_{\boldsymbol{l}_1,\ldots,\boldsymbol{l}_k\in\mathcal{L}}\mathbf{1}\left(\sum_{i=1}^k|\boldsymbol{l}_i|=3\right)\prod_{i=1}^k P_{\boldsymbol{l}_i}\right],
\end{aligned} \tag{3.4.15}$$

where $P_{\mathbf{0}} := \int_{I_\ell} x^R q(y^R)\, \mathrm{d}E$, and $\mathbb{E}[\mathcal{A}]$ depends only on R and the first two moments of h_{ab}.

We need the following lemma:

Lemma 3.4.4. *Fix indices a, b such that $a \neq b$. Let Y denote any of the following terms:*

$$\begin{aligned}
&g^{(3)}\left(P_{\mathbf{0}}\right) P_{(1)}^m P_{(0,1)}^n \quad (m+n=3), \\
&g^{(2)}\left(P_{\mathbf{0}}\right)\left(P_{(2)}+P_{(0,2)}+P_{(1,1)}+P_{(0,1,1)}\right)\left(P_{(1)}+P_{(0,1)}\right), \\
&g^{(1)}\left(P_{\mathbf{0}}\right)\left(P_{(3)}+P_{(0,3)}+P_{(1,2)}+P_{(2,1)}+P_{(0,1,2)}+P_{(1,1,1)}+P_{(0,1,1,1)}\right).
\end{aligned} \tag{3.4.16}$$

Then there is a constant $c > 0$ such that

$$\left|\mathbb{E}[Y]\right| \prec N^{-1/2-c}.$$

The lemma is proved in Section 3.5.

3.4.2 Resolvent Comparison

Given this lemma, we can conclude by standard Green's comparison argument.

Proof of Theorem 3.1.3. Let W be drawn from the Gaussian Orthogonal Ensemble, and let H be a Wigner matrix. By Theorem 3.2.9 and Lemma 3.3.8, it suffices to show

$$\left|\mathbb{E}\left[g(v_\ell^W) - g(v_\ell^H)\right]\right| \leq c^{-1} N^{-c} \tag{3.4.17}$$

for some constant $c > 0$, where v_ℓ^W and v_ℓ^H are corresponding regularized observables (see (3.3.4)) for W and H respectively.

To this end, we fix a bijection

$$\psi : \{(i,j) : 1 \leqslant i \leqslant j \leqslant N\} \to [\![1, \xi_N]\!],$$

where $\xi_N = N(N+1)/2$, and define the interpolating matrices $H^0, H^1, H^2, \ldots, H^{\xi_N}$ by

$$h_{ij}^{\xi} = \begin{cases} h_{ij} & \text{if } \psi(i,j) > \xi, \\ w_{ij} & \text{if } \psi(i,j) \leq \xi, \end{cases}$$

for $i \leq j$. Therefore, $H^0 = H$ and $H^{\xi_N} = W$. We may rewrite (3.4.17) as a telescopic summation,

$$\left|\mathbb{E}\left[g(v_\ell^W) - g(v_\ell^H)\right]\right| \leq \sum_{\xi=1}^{\xi_N} \left|\mathbb{E}\left[g\left(v_\ell^{H^\xi}\right) - g\left(v_\ell^{H^{\xi-1}}\right)\right]\right| \tag{3.4.18}$$

Fix some $\xi \in [\![1, \xi_N]\!]$ and consider the indices (a,b) such that $\psi(a,b) = \xi$. Let Q^ξ be the matrix obtained from H^ξ by setting h_{ab}^ξ and h_{ba}^ξ to zero. Note that Q^ξ can also be obtained from $H^{\xi-1}$ by setting $h_{ab}^{\xi-1}$ and $h_{ba}^{\xi-1}$ to zero. We consider the following two cases. First, suppose $a = b$. Lemma 3.3.1 and Lemma 3.4.3 imply that, with Y denoting any term from (3.4.16), $|Y| \prec N^{-\tau/30}$, where we use the upper bound of $\Psi(z)$ (see (3.4.7)). Combining (3.4.15), we have

$$\left|\mathbb{E}\left[g(v_\ell^{H^\xi}) - g(v_\ell^{H^{\xi-1}})\right]\right| \leq \left|\mathbb{E}\left[g(v_\ell^{H^\xi}) - g(v_\ell^{Q^\xi})\right] + \left[g(v_\ell^{Q^\xi}) - g(v_\ell^{H^{\xi-1}})\right]\right| \leq N^{-3/2-c},$$

where we use the fact that the first two moments of Wigner matrices H^ξ and $H^{\xi-1}$ are the same, and therefore $\mathbb{E}[\mathcal{A}]$ in (3.4.15) is the same for both cases.

Now, if $a \neq b$. Combining Lemma 3.4.4 and (3.4.15), we have

$$\left|\mathbb{E}\left[g(v_\ell^{H^\xi}) - g(v_\ell^{H^{\xi-1}})\right]\right| \leq \left|\mathbb{E}\left[g(v_\ell^{H^\xi}) - g(v_\ell^{Q^\xi})\right] + \left[g(v_\ell^{Q^\xi}) - g(v_\ell^{H^{\xi-1}})\right]\right| \leq N^{-2-c}$$

for some $c > 0$. These two estimates conclude the proof in view of (3.4.18). $\square$

3.5 Proof of Lemma 3.4.4

We now fix indices $a, b \in [\![1, N]\!]$ such that $a \neq b$, and carry this choice throughout the current section. All of the bounds stated below are uniform in the choice of a and b.

Recall from (3.2.1) that m_{sc} denotes the Stieltjes transform of the semicircle law, which is deterministic and satisfies

$$m_{\mathrm{sc}}(z) + z + \frac{1}{m_{\mathrm{sc}}(z)} = 0. \tag{3.5.1}$$

Recall from the discussion below (3.4.1) that Q is the matrix obtained by setting h_{ab}, h_{ba} in H to 0, and R is the resolvent of Q. Let $Q^{(a)}$ be the matrix obtained by setting a-th row and column of H to 0 and let $R^{(a)}$ be the resolvent of $Q^{(a)}$. Then it follows from the Schur complement formula that

$$\left(R^{(a)}\right)_{ij} = \begin{cases} 0, & \text{if exactly one of } i, j \text{ is } 0, \\ -z^{-1}, & \text{if } i = j = 0, \\ R_{ij}, & \text{otherwise.} \end{cases} \tag{3.5.2}$$

We state the following local laws for $R^{(a)}$, whose proof follows directly from the corresponding local laws in Lemma 3.4.2 and (3.5.2), and is therefore omitted.

Lemma 3.5.1. *Let H be a Wigner matrix.*

1. *For all $z \in \boldsymbol{S}$, we have*

$$\left|R^{(a)}_{\boldsymbol{x}\boldsymbol{y}} - \langle \boldsymbol{x}, \boldsymbol{y}\rangle m_{\mathrm{sc}}\right| \prec \Psi, \quad \left|(R^{(a)}R^{(a)})_{cd}\right| \prec N\Psi^2, \quad \left|(R^{(a)}\bar{R}^{(a)})_{cd}\right| \prec N\Psi^2, \tag{3.5.3}$$

uniformly over $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{S}^{N-1}$ such that at least one of $\boldsymbol{x}, \boldsymbol{y}$ is $\boldsymbol{e}_s$ with some $s \neq a$, and $c, d \in [\![1, N]\!]$ such that $c, d \neq a$.

2. Furthermore, if $z = E + \mathrm{i}\eta \in \boldsymbol{S}$ satisfies

$$E \in I_\ell, \quad \eta = \eta_\ell,$$

then for any deterministic $A \in \mathrm{Mat}_N$ such that $\|A\| \leq 1$ and $\operatorname{Tr} A = 0$, we have

$$\begin{aligned} \left|(R^{(a)} A \bar{R}^{(a)})_{cd}\right| &\prec N^{1/2}\Psi, \\ \left|(R^{(a)} A \bar{R}^{(a)} R^{(a)})_{cd}\right| &\prec N^{3/2}\Psi^{9/4}, \end{aligned} \tag{3.5.4}$$

uniformly over all $c, d \in [\![1, N]\!]$ such that $c, d \neq a$.

Finally, for conciseness, we let A stand for the matrix defined in (3.4.13), which satisfies

$$\|A\| \leq 1, \qquad \operatorname{Tr}(A) = 0.$$

The main goal of this section is to rewrite the resolvent expansion terms x_i and y_i from Lemma 3.4.3 into the polynomial form as introduced in [33, Section 7]. For instance, we want to rewrite

$$x_1 \approx \left(\sum_{i_1,\ldots,i_d \neq a} V_{i_1,\ldots,i_d} h_{i_1 a} \cdots h_{i_d a} \right) \cdot \left(\prod_{j=d+1}^{d+m} \sum_{i_j=1}^{N} V_{i_j} \left(h_{i_j a}^2 - \frac{1}{N} \right) \right),$$

where the V terms are $Q^{(a)}$-measurable. The reason to write it into this form is that, whenever d is odd, we gain a factor of $N^{-1/2}$ upon taking the expectation (see Lemma 3.5.8 for the precise statement), which is essential in the proof of Lemma 3.4.4.

Fixing a universal constant $C_0 > 0$ and following the set up in [33], we make the following definitions.

Definition 3.5.2 (Admissible weights). *Let $\varrho = (\varrho_i : i \in [\![1, N]\!])$ be a sequence of deter-*

ministic nonnegative real numbers. We say that ϱ is an admissible weight *if*

$$\frac{1}{N^{1/2}}\left(\sum_i \varrho_i^2\right)^{1/2} \leqslant 1, \quad \frac{1}{N^{1/2}}\left(\sum_i \varrho_i^3\right)^{1/3} \leqslant N^{-1/6}.$$

Definition 3.5.3 ($O_{\prec,d}(\cdot)$)**.** *For a given degree $d \in \mathbb{N}$ let*

$$\mathcal{P} = \sum_{\substack{1\leq i_1,\dots,i_d\leq N \\ i_1,\dots,i_d\neq a}} V_{i_1\cdots i_d} h_{ai_1}\cdots h_{ai_d}$$

be a polynomial in entries of the a-th row of Q. We write $\mathcal{P} = O_{\prec,d}(K)$ if the following conditions are satisfied.

1. *K is deterministic and $V_{i_1\cdots i_d}$ is $Q^{(a)}$-measurable.*
2. *There exist admissible weights $\varrho^{(1)},\dots,\varrho^{(d)}$ such that*

$$|V_{i_1\cdots i_d}| \prec K\varrho_{i_1}^{(1)}\cdots\varrho_{i_d}^{(d)}.$$

3. *We have the deterministic bound $|V_{i_1\cdots i_d}| \leqslant N^{C_0}$.*

The above definition also extends to $d = 0$, where $\mathcal{P} = V$ is $Q^{(a)}$-measurable.

Definition 3.5.4 ($O_{\prec,\diamond}(\cdot)$)**.** *Let $\mathcal{P}$ be a polynomial of the form*

$$\mathcal{P} = \sum_{i=1}^N V_i\left(h_{ai}^2 - \frac{1}{N}\right).$$

We write $\mathcal{P} = O_{\prec,\diamond}(K)$ if V_i is $Q^{(a)}$-measurable, $|V_i| \leqslant N^{C_0}$, and $|V_i| \prec K$ for some deterministic K.

Definition 3.5.5 (Graded polynomials)**.** *We write $\mathcal{P} = O_{\prec,\mathsf{even}}(K)$ if $\mathcal{P}$ is a sum of at most C_0 terms of the form*

$$K\mathcal{P}_0\prod_{s=1}^n \mathcal{P}_i, \quad \mathcal{P}_0 = O_{\prec,2d}(1), \quad \mathcal{P}_i = O_{\prec,\diamond}(1)$$

where $d, n \leqslant C_0$ and K is deterministic. Moreover, we write $\mathcal{P} = O_{\prec,\mathsf{odd}}(K)$ if $\mathcal{P} = \widehat{\mathcal{P}}\mathcal{P}_{\mathsf{even}}$, where $\widehat{\mathcal{P}} = O_{\prec,1}(1)$ and $\mathcal{P}_{\mathsf{even}} = O_{\prec,\mathsf{even}}(K)$.

Remark 3.5.6. The graded polynomials satisfy simple algebraic rules by definition, which we state without proof:

$$
\begin{gathered}
O_{\prec,*}(K_1) + O_{\prec,*}(K_2) = O_{\prec,*}(K_1 + K_2),\\
O_{\prec,*}(K_1)O_{\prec,*}(K_2) = O_{\prec,\mathsf{even}}(K_1K_2),\\
O_{\prec,\mathsf{odd}}(K_1)O_{\prec,\mathsf{even}}(K_2) = O_{\prec,\mathsf{odd}}(K_1K_2),
\end{gathered}
$$

after possibly increasing C_0. Here $*$ represents either odd or even. It should be noted that all of these operations can be done for an arbitrary, but finite, number of times (independent of N).

Remark 3.5.7. Definitions 3.5.3-3.5.5 refine the stochastic domination notation 3.2.1. More precisely, it means that

$$
\mathcal{P} = O_{\prec,*}(K) \Longrightarrow \mathcal{P} = O_{\prec}(K), \tag{3.5.5}
$$

where $*$ can represent $d, \diamond, \mathsf{even}$ or odd. See lines under [33, Equation (7.56)] for details.

We now state the following lemma proved in [33, Lemma 7.13].

Lemma 3.5.8. *Let $\mathcal{P} = O_{\prec,\mathsf{odd}}(K)$ for some deterministic $K \leqslant N^{C_0}$. Then for any fixed $D > 0$ we have*

$$
\left|\mathbb{E}[\mathcal{P}]\right| \prec N^{-1/2}K + N^{-D}.
$$

The following resolvent identities are standard (see [20, Lemma 3.5] and [20, Equation (4.1)]):

$$
R_{aa} = \frac{1}{-z - \sum_{r,s\neq a} R_{rs}^{(a)} h_{ar}h_{as}}, \tag{3.5.6}
$$

$$R_{ai} = -R_{aa} \sum_{r \neq a} R_{ir}^{(a)} h_{ar}, \tag{3.5.7}$$

$$R_{ij} = R_{ij}^{(a)} + R_{aa} \sum_{r \neq a} R_{ir}^{(a)} h_{ar} \cdot \sum_{r \neq a} R_{jr}^{(a)} h_{ar}. \tag{3.5.8}$$

In the next several lemmas, we write resolvents and multi-resolvents in terms of graded polynomials.

Lemma 3.5.9. *Fix $D > 0$. For spectral parameter $z \in \boldsymbol{S}$, we have*

$$R_{aa} = O_{\prec,\mathsf{even}}(1) + O_{\prec}\left(N^{-D}\right), \tag{3.5.9}$$

$$R_{ab} = O_{\prec,\mathsf{odd}}(\Psi) + O_{\prec}(N^{-D}),$$

$$R_{bb} = O_{\prec,\mathsf{even}}(1) + O_{\prec}(N^{-D}),$$

$$\operatorname{Tr}(R) = \sum_{i \neq a} R_{ii}^{(a)} + O_{\prec,\mathsf{even}}(N\Psi^2) + O_{\prec}(N^{-D}). \tag{3.5.10}$$

Proof. We begin with some preliminary claims. Using (3.5.3) and the definition of graded polynomial, we have

$$\begin{aligned}
\sum_{r,s \neq a} R_{rs}^{(a)} h_{ar} h_{as} - m_{\mathrm{sc}} &= \sum_{r \neq a} R_{rr}^{(a)} \left(h_{ar}^2 - \frac{1}{N}\right) + \left(\frac{1}{N} \sum_{r \neq a} R_{rr}^{(a)} - m_{\mathrm{sc}}\right) + \sum_{\substack{r \neq s \\ r,s \neq a}} R_{rs}^{(a)} h_{ar} h_{as} \\
&= O_{\prec,0}(1) + O_{\prec,0}(\Psi) + O_{\prec,2}(\Psi) \\
&= O_{\prec,\mathsf{even}}(1),
\end{aligned} \tag{3.5.11}$$

and

$$\begin{aligned}
\sum_{r,s \neq a} R_{rs}^{(a)} h_{ar} h_{as} - m_{\mathrm{sc}} &= \sum_{r \neq a} R_{rr}^{(a)} \left(h_{ar}^2 - \frac{1}{N}\right) + \left(\frac{1}{N} \sum_{r \neq a} R_{rr}^{(a)} - m_{\mathrm{sc}}\right) + \sum_{\substack{r \neq s \\ r,s \neq a}} R_{rs}^{(a)} h_{ar} h_{as} \\
&= O_{\prec}\left(N^{-1/2}\right) + O_{\prec}(\Psi) + O_{\prec}(\Psi) \\
&= O_{\prec}(\Psi),
\end{aligned} \tag{3.5.12}$$

where we used the definition of graded polynomials and (3.5.5) in the second step and

(3.4.6) in the last step.

Note also that

$$\sum_{r \neq a} R_{ir}^{(a)} h_{ra} = R_{ii}^{(a)} h_{ia} + O_{\prec,\mathsf{odd}}(\Psi).$$

For any $n \in \mathbb{N}$, we have by (3.5.6) that

$$\begin{aligned} R_{aa} = \frac{1}{-z - \sum_{r,s\neq a} R_{rs}^{(a)} h_{ar} h_{as}} &= \frac{1}{-z - m_{\mathsf{sc}} + \left(m_{\mathsf{sc}} - \sum_{r,s\neq a} R_{rs}^{(a)} h_{ar} h_{as}\right)} \\ &= \frac{1}{1/m_{\mathsf{sc}} + \left(m_{\mathsf{sc}} - \sum_{r,s\neq a} R_{rs}^{(a)} h_{ar} h_{as}\right)} \\ &= \frac{m_{\mathsf{sc}}}{1 + m_{\mathsf{sc}}\left(m_{\mathsf{sc}} - \sum_{r,s\neq a} R_{rs}^{(a)} h_{ar} h_{as}\right)} \\ &= \sum_{j=0}^{n} m_{\mathsf{sc}}^{i+1} \left(\sum_{r,s\neq a} R_{rs}^{(a)} h_{ar} h_{as} - m_{\mathsf{sc}} \right)^{j} + O_{\prec}\left(m_{\mathsf{sc}}^{n} \Psi^{n}\right) \\ &= O_{\prec,\mathsf{even}}(1) + O_{\prec}\left(m_{\mathsf{sc}}^{n} \Psi^{n}\right), \end{aligned} \tag{3.5.13}$$

where we used (3.5.12) in the second-to-last line and (3.5.11) in the last step.

By (3.5.7) and (3.5.13), we have

$$R_{ab} = R_{aa} \sum_{r \neq a} R_{rb}^{(a)} h_{ar} = O_{\prec,\mathsf{odd}}(\Psi) + O_{\prec}(m_{\mathsf{sc}}^{n} \Psi^{n+1}). \tag{3.5.14}$$

By (3.5.8) and (3.5.13), we have

$$R_{bb} = R_{bb}^{(a)} + R_{aa} \left(\sum_{r \neq a} R_{br}^{(a)} h_{ar} \right)^2 = O_{\prec,\mathsf{even}}(1) + O_{\prec}(m_{\mathsf{sc}}^{n} \Psi^{n+1}). \tag{3.5.15}$$

The lemma follows from (3.5.13), (3.5.14), and (3.5.15) by choosing a sufficiently large $n \equiv n(\tau, D)$. □

Lemma 3.5.10. *Fix $D > 0$. For a spectral parameter $z \in \boldsymbol{S}$, we have*

$$(R\bar{R})_{aa} = O_{\prec,\mathsf{even}}(N\Psi^2) + O_{\prec}(N^{-D}), \tag{3.5.16}$$

$$
\begin{aligned}
(R\bar{R})_{ab} &= O_{\prec,\mathsf{odd}}(N\Psi^2) + O_{\prec}(N^{-D}),\\
(R\bar{R})_{bb} &= O_{\prec,\mathsf{even}}(N\Psi^2) + O_{\prec}\left(N^{-D}\right),\\
\left(R^2\right)_{aa} &= O_{\prec,\mathsf{even}}(N\Psi^2) + O_{\prec}(N^{-D}),\\
\left(R^2\right)_{ab} &= O_{\prec,\mathsf{odd}}(N\Psi^2) + O_{\prec}(N^{-D}),\\
\left(R^2\right)_{bb} &= O_{\prec,\mathsf{even}}(N\Psi^2) + O_{\prec}(N^{-D}).
\end{aligned}
$$

Proof. We present only the proof of (3.5.16); the others can be done similarly. By the resolvent identity (3.5.7),

$$
\begin{aligned}
(R\bar{R})_{aa} = \sum_{i\neq a}\left(R_{ai}\bar{R}_{ia}\right) + |R_{aa}|^2 &= |R_{aa}|^2 \sum_{i\neq a}\left(\sum_{r\neq a} R^{(a)}_{ri} h_{ar}\right)\left(\sum_{r\neq a}\bar{R}^{(a)}_{ir} h_{ar}\right) + |R_{aa}|^2 \\
&= |R_{aa}|^2\left(\sum_{r,s\neq a}\left(R^{(a)}\bar{R}^{(a)}\right)_{rs} h_{ar}h_{as} + 1\right).
\end{aligned}
\tag{3.5.17}
$$

Combining (3.4.6), (3.5.3), (3.5.9), and (3.5.17), we deduce (3.5.16). □

Lemma 3.5.11. *There exists an orthonormal basis $\{\boldsymbol{w}_i\}_{i=1}^N$ of $\mathbb{R}^N$ such that*

$$
\boldsymbol{w}_i(j) \prec N^{-1/2} \text{ and } \left|\sum_k \boldsymbol{e}_j^\top A \boldsymbol{w}_k\right| \prec 1, \tag{3.5.18}
$$

uniformly over $i, j \in [\![1, N]\!]$.

Proof. Let the matrix $W = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N]$ be uniformly distributed on the orthogonal group $O(n)$ with respect to Haar measure. Then the marginal distribution of each column vector $\boldsymbol{w}_i$ is uniform on the unit sphere $\mathbb{S}^{N-1}$. By a standard concentration result (see [162, Theorem 3.4.6]), we have

$$
\boldsymbol{w}_i(j) \prec N^{-1/2}. \tag{3.5.19}
$$

Consider the function on $N \times N$ matrices defined by

$$\varphi(M) := \boldsymbol{e}_j^\top AW\mathbf{1},$$

where $\mathbf{1}$ has all entries equal to 1. We claim that φ is Lipshitz with respect to the Frobenius norm with Lipshitz constant at most $N^{1/2}$. Indeed, with $\boldsymbol{u} := A\boldsymbol{e}_j$,

$$\begin{aligned}|\varphi(W) - \varphi(\tilde{W})| \le \left|\sum_{i,k} u_i(w_{ik} - \tilde{w}_{ik})\right| &\le \sum_i |u_i| \left(\sum_k (w_{ik} - \tilde{w}_{ik})^2\right)^{1/2} N^{1/2} \\ &\le N^{1/2}\|\boldsymbol{u}\|\|W - \tilde{W}\|_F \\ &\le N^{1/2}\|W - \tilde{W}\|_F,\end{aligned}$$

where we used the Cauchy–Schwarz inequality in the second and third inequalities, and we used $\|A\| \le 1$ in the last inequality. By a standard concentration result for the orthogonal group (see [162, Theorem 5.2.7]), we have

$$|\varphi(W)| \prec 1. \tag{3.5.20}$$

By (3.5.19) and (3.5.20), there exists an orthonormal basis satisfying (3.5.18). □

Lemma 3.5.12. *Fix $D > 0$. For a spectral parameter $z = E + \mathrm{i}\eta$ satisfying*

$$E \in \mathrm{I}_\ell, \quad \eta = \eta_\ell,$$

we have

$$\begin{aligned}(RA\bar{R})_{aa} &= O_{\prec,\mathsf{even}}\left(N^{1/2}\Psi\right) + O_{\prec,\mathsf{odd}}(\Psi) + O_\prec(N^{-D}), &(3.5.21)\\ (RA\bar{R})_{ab} &= O_{\prec,\mathsf{odd}}(N^{1/2}\Psi) + O_{\prec,\mathsf{even}}(\Psi) + O_\prec(N^{-D}),\\ (RA\bar{R})_{bb} &= O_{\prec,\mathsf{even}}(N^{1/2}\Psi) + O_{\prec,\mathsf{odd}}(\Psi^2) + O_\prec(N^{-D}),\end{aligned}$$

$$\operatorname{Tr}\left(RA\bar{R}\right) = \operatorname{Tr}\left(R^{(a)}A\bar{R}^{(a)}\right) + O_{\prec,\mathsf{even}}(N^{3/2}\Psi^{9/4}) + O_{\prec,\mathsf{odd}}(N\Psi^2) + O_\prec(N^{-D}) \tag{3.5.22}$$

$$
\begin{aligned}
(RA\bar{R}R)_{aa} &= O_{\prec,\mathsf{even}}(N^{3/2}\Psi^{9/4}) + O_{\prec,\mathsf{odd}}(N\Psi^2) + O_{\prec}\left(N^{-D}\right), \\
(RA\bar{R}R)_{ab} &= O_{\prec,\mathsf{odd}}(N^{3/2}\Psi^{9/4}) + O_{\prec,\mathsf{even}}(N\Psi^2) + O_{\prec}\left(N^{-D}\right), \qquad (3.5.23)\\
(RA\bar{R}R)_{bb} &= O_{\prec,\mathsf{even}}(N^{3/2}\Psi^{9/4}) + O_{\prec,\mathsf{odd}}(N\Psi^3) + O_{\prec}\left(N^{-D}\right), \\
(RA\bar{R}R)_{ba} &= O_{\prec,\mathsf{odd}}(N^{3/2}\Psi^{9/4}) + O_{\prec,\mathsf{even}}(N\Psi^3) + O_{\prec}(N^{-D}). \qquad (3.5.24)
\end{aligned}
$$

Proof. We present the proof of (3.5.21). The others can be shown similarly.

By the resolvent identity (3.5.7), we have

$$
\begin{aligned}
(RA\bar{R})_{aa} = {} & |R_{aa}|^2 \sum_{r,s\neq a} \left(R^{(a)}A\bar{R}^{(a)}\right)_{rs} h_{ar}h_{as} \\
& + |R_{aa}|^2 \left(\sum_{s\neq a} \left(A\bar{R}^{(a)}\right)_{as} h_{as} + \sum_{r\neq a} \left(R^{(a)}A\right)_{ra} h_{ar} + A_{aa} \right). \qquad (3.5.25)
\end{aligned}
$$

By the definition of graded polynomial (see Definition 3.5.5), (3.5.9) and (3.5.4), the first line of (3.5.25) is

$$
|R_{aa}|^2 \sum_{r,s\neq a} \left(R^{(a)}A\bar{R}^{(a)}\right)_{rs} h_{ar}h_{as} = O_{\prec,\mathsf{even}}(N^{1/2}\Psi) + O_{\prec}(N^{-D}). \qquad (3.5.26)
$$

Select an orthonormal basis $\{\boldsymbol{w}_i\}_{i=1}^N$ of $\mathbb{R}^N$ according to Lemma 3.5.11, we have

$$
\begin{aligned}
\left|\left(A\bar{R}^{(a)}\right)_{as}\right| = \left|\sum_i \boldsymbol{e}_a^\top A\boldsymbol{w}_i\boldsymbol{w}_i^\top \bar{R}^{(a)}\boldsymbol{e}_s\right| &\leq \left|\sum_i \boldsymbol{e}_a^\top A\boldsymbol{w}_i \left(\langle \boldsymbol{w}_i, \boldsymbol{e}_s\rangle m_{\mathrm{sc}}(z) + \Psi\right)\right| \\
&\prec \left|\Psi \sum_i \boldsymbol{e}_a^\top A\boldsymbol{w}_i\right| \prec \Psi,
\end{aligned}
$$

where we used (3.2.2) in the first inequality, and $\Psi(z) > C\tau^{1/4}N^{-1/2}$ for $z \in \boldsymbol{S}$ in the second inequality.

Similarly, we have $\left|\left(R^{(a)}A\right)_{ra}\right| \prec \Psi$. Moreover, $|A_{aa}| \leq 1$ as a consequence of $\|A\| \leq 1$.

Therefore, by definition of graded polynomial, the second line of (3.5.25) is

$$|R_{aa}|^2 \left(\sum_{s \neq a} \left(A\bar{R}^{(a)} \right)_{as} h_{as} + \sum_{r \neq a} \left(R^{(a)} A \right)_{ra} h_{ar} + A_{aa} \right) = O_{\prec,\mathsf{odd}}(\Psi) + O_{\prec,\mathsf{even}}(1) + O_{\prec}(N^{-D}). \tag{3.5.27}$$

(3.5.21) now follows from (3.5.26) and (3.5.27). □

Remark 3.5.13. We note that the even graded polynomials in (3.5.23) and (3.5.24) are different by an order of Ψ. This should not be surprising, since we are expanding with respect to the a-th row and column in Q, but a does not take a "symmetric" position in the multi-resolvents $(RA\bar{R}R)_{ab}, (RA\bar{R}R)_{ba}$. This point will be exploited in Lemma 3.5.15.

Using Lemma 3.5.9, Lemma 3.5.10, Lemma 3.5.12 and the definitions of x_i, y_i in Lemma 3.4.3, we have the following lemma. We omit the proof, since it is straightforward.

Lemma 3.5.14. *Let $x_i, i = 1, 2$ and $y_i, i = 1, 2, 3$ be the resolvent expansion as in Lemma 3.4.3. For spectral parameter $z = E + \mathrm{i}\eta$ satisfying*

$$E \in I_\ell, \quad \eta = \eta_\ell,$$

we have

$$\begin{aligned} x_1 &= O_{\prec,\mathsf{odd}}(N^{1+\delta/2}\Psi^{5/4}) + O_{\prec,\mathsf{even}}(N^{1/2+\delta/2}\Psi) + O_{\prec}(N^{-D}), \\ x^R, x_2 &= O_{\prec,\mathsf{even}}(N^{1+\delta/2}\Psi^{5/4}) + O_{\prec,\mathsf{odd}}(N^{1/2+\delta/2}\Psi) + O_{\prec}(N^{-D}), \\ y_1, y_3 &= O_{\prec,\mathsf{odd}}(\Psi) + O_{\prec}(N^{-D}), \\ y_2 &= O_{\prec,\mathsf{even}}(\Psi) + O_{\prec}(N^{-D}). \end{aligned}$$

To bound x_3, we need an extra lemma.

Lemma 3.5.15. *Denote by $O_{\prec,\mathsf{odd,b}}(K)$ the graded polynomial with expansion in b-th row and column of Q instead of a-th row and column. Then for spectral parameter $z = E + \mathrm{i}\eta$*

satisfying

$$E \in I_\ell, \quad \eta = \eta_\ell,$$

we have

$$x_3 = O_{\prec,\mathsf{odd}}(N^{1+\delta/2}\Psi^{5/4}) + O_{\prec,\mathsf{odd,b}}(N^{1+\delta/2}\Psi^{5/4}) + O_{\prec,\mathsf{even}}(N^{1/2+\delta/2}\Psi^2) + O_{\prec,\mathsf{even,b}}(N^{1/2+\delta/2}\Psi^2) + O_\prec(N^{-D}). \quad (3.5.28)$$

Proof. By definition,

$$\begin{aligned}
&\sqrt{\frac{|\mathrm{I}|(N-|\mathrm{I}|)}{N^3\eta_\ell^2}}x_3 \\
&= -(RA\bar{R}R)_{ab}R_{aa}R_{bb} - (RA\bar{R}R)_{ba}R_{aa}R_{bb} - (RA\bar{R})_{aa}(R\bar{R})_{ab}R_{bb} - \left(RA\bar{R}\right)_{bb}(R\bar{R})_{ba}R_{aa} \\
&- (RA\bar{R})_{ab}(R\bar{R})_{aa}R_{bb} - (RA\bar{R})_{ba}(R\bar{R})_{bb}R_{aa} + [C] + [L],
\end{aligned} \quad (3.5.29)$$

where $[C]$ denotes the conjugate of the previous terms and $[L]$ refers to terms with smaller order for either parity (even or odd).

For the first term in (3.5.29), using resolvent identities with respect to the b-th row and column as in the proof of Lemma 3.5.9 and Lemma 3.5.12, we have

$$(RA\bar{R}R)_{ab} = O_{\prec,\mathsf{odd,b}}(N\Psi^{9/4}) + O_{\prec,\mathsf{even,b}}(N^{1/2}\Psi^3) + O_\prec(N^{-D}),$$
$$R_{aa} = O_{\prec,\mathsf{even,b}}(1) + O_\prec(N^{-D}), \quad R_{bb} = O_{\prec,\mathsf{even,b}}(1) + O_\prec(N^{-D}).$$

The conclusion follows by applying Lemma 3.5.9, Lemma 3.5.10 and Lemma 3.5.12 to the remaining terms of (3.5.29). □

Remark 3.5.16. Note that it is still legitimate to apply Lemma 3.5.8 to (3.5.28). Since by linearity of expectation, we may apply Lemma 3.5.8 to $O_{\prec,\mathsf{odd}}(K)$ and $O_{\prec,\mathsf{odd,b}}(K)$

separately.

By (3.5.10), (3.5.22), Lemma 3.3.1 and Taylor expansion, we have

$$g^{(k)}\left(\int_{I_\ell} x^R q(y^R)\mathrm{d}E\right) = O_{\prec,\mathsf{even}}(N^{C_0\delta_2}) + O_{\prec,\mathsf{odd}}(N^{-1/2+\delta/2+C_0\delta_2}), \quad k=1,2,3, \tag{3.5.30}$$
$$q^{(k)}(y^R) = O_{\prec,\mathsf{even}}(1), \quad k=1,2,3$$

for some constant $C>0$ which depends only on g. The same graded polynomial expansion holds for the expansion respect to b-th row and column.

We are now ready for the proof of Lemma 3.4.4.

Proof of Lemma 3.4.4. Define

$$\Psi_\ell = (N\Delta_\ell)^{-1} \le N^{-\tau/3}, \quad \text{and} \quad \hat{N} = N^{2\delta_1+\delta_2+\delta/2} \ll N^{\tau/1000}.$$

Combining Lemma 3.5.14, Lemma 3.5.15, definition of P_l, Remark 3.2.3 and (3.4.7), we have

$$P_{(1)}, P_{(0,1)} = O_{\prec,\mathsf{odd}}(\hat{N}\Psi_\ell^{1/4}) + O_{\prec,\mathsf{even}}(\hat{N}N^{-1/2}) + O_\prec(N^{-D}),$$
$$P_{(2)}, P_{(0,2)}, P_{(1,1)}, P_{(0,1,1)} = O_{\prec,\mathsf{even}}(\hat{N}\Psi_\ell^{1/4}) + O_{\prec,\mathsf{odd}}(\hat{N}N^{-1/2}) + O_\prec(N^{-D}),$$

as well as

$$\begin{aligned}&P_{(0,3)}, P_{(1,2)}, P_{(2,1)}, P_{(0,1,2)}, P_{(1,1,1)}, P_{(0,1,1,1)} \\ &\qquad = O_{\prec,\mathsf{odd}}(\hat{N}\Psi_\ell^{5/4}) + O_{\prec,\mathsf{even}}(\hat{N}N^{-1/2}\Psi_\ell) + O_\prec(N^{-D}),\end{aligned}$$

and

$$\begin{aligned}P_{(3)} &= O_{\prec,\mathsf{odd}}(\hat{N}\Psi_\ell^{1/4}) + O_{\prec,\mathsf{odd,b}}(\hat{N}\Psi_\ell^{1/4}) \\ &\qquad + O_{\prec,\mathsf{even}}(\hat{N}N^{-1/2}\Psi_\ell) + O_{\prec,\mathsf{even,b}}(\hat{N}N^{-1/2}\Psi_\ell) + O_\prec(N^{-D}).\end{aligned}$$

Combining (3.5.30) and Lemma 3.5.8, we conclude the existence of $c \equiv c(\tau) > 0$, such that

$$\mathbb{E}[Y] \prec N^{-1/2-c} + N^{-D} \leq N^{-1/2-c},$$

when Y represents any term in (3.4.16). $\square$

3.6 Proof of Lemma 3.4.2

The proof is based on the cumulant expansion which can be found in [126, Lemma 3.2].

Lemma 3.6.1 (Cumulant expansion). *Fix $T \in \mathbb{N}$ and let $F : \mathbb{R} \to \mathbb{C}^+$ be $T+1$ times continuously differentiable. Let Y be a mean zero random variable with finite moments to order $T+2$. Then there exists a function $\Omega_T : \mathbb{C} \to \mathbb{C}$ such that*

$$\mathbb{E}\Big[YF(Y)\Big] = \sum_{r=1}^{T} \frac{\kappa^{(r+1)}(Y)}{r!}\mathbb{E}\left[F^{(r)}(Y)\right] + \mathbb{E}\Big[\Omega_T\Big(YF(Y)\Big)\Big], \tag{3.6.1}$$

where $\mathbb{E}$ denotes the expectation with respect to Y, $\kappa^{(r+1)}(Y)$ denotes the $(r+1)$-th cumulant of Y, and $F^{(r)}$ denotes the r-th derivative of the function F. Further, there exists a constant $C > 0$ (not depending on F or Y) such that

$$\Big|\mathbb{E}\Big[\Omega_T\Big(YF(Y)\Big)\Big]\Big| \leqslant \frac{(CT)^T}{T!} \cdot \mathbb{E}\left[|Y|^{T+2}\right] \sup_{t\in\mathbb{R}} \Big|F^{(T+1)}(t)\Big|.$$

Proof of Lemma 3.4.2. First, we claim that it suffices to prove the local laws for the resolvent G, because the local laws for R are straightforward consequences of the ones for G. We illustrate the procedure of deducing a local law for R from the corresponding local law for G for the first inequality in (3.4.8). The other deductions are similar.

Pick any $z = E + \mathrm{i}\eta \in \boldsymbol{S}$. By (3.4.11), we have

$$R_{\boldsymbol{x}\boldsymbol{y}} = G_{\boldsymbol{x}\boldsymbol{y}} - (GUG)_{\boldsymbol{x}\boldsymbol{y}} + (GUGUG)_{\boldsymbol{x}\boldsymbol{y}} - ((GU)^3G)_{\boldsymbol{x}\boldsymbol{y}} + ((GU)^4R)_{\boldsymbol{x}\boldsymbol{y}}.$$

Combining the estimates $h_{ab} \prec N^{-1/2}$, $\|R\| \le \eta^{-1}$, $\Psi \ge C\tau^{1/4}N^{-1/2}$ from Lemma 3.4.1 and the isotropic local law (3.2.2) for G, the isotropic local law for R follows.

In light of the previous discussion, we focus only on the local laws for G. By Theorem 3.2.2, we have

$$\begin{aligned}(G\bar{G})_{\boldsymbol{x}\boldsymbol{y}} = \sum_k G_{\boldsymbol{x}k}\bar{G}_{k\boldsymbol{y}} &\prec \left|\sum_k \langle \boldsymbol{x}, \boldsymbol{e}_k\rangle\langle \boldsymbol{e}_k, \boldsymbol{y}\rangle\right| + \left|\sum_k \langle \boldsymbol{x}, \boldsymbol{e}_k\rangle\right|\Psi + |\langle \boldsymbol{e}_k, \boldsymbol{y}\rangle|\,\Psi + N\Psi^2 \\ &\le |\langle \boldsymbol{x}, \boldsymbol{y}\rangle| + 2N^{1/2}\Psi + N\Psi^2 \prec N\Psi^2,\end{aligned} \tag{3.6.2}$$

where we use Cauchy inequality in the second last inequality and $\Psi \ge C\tau^{1/4}N^{-1/2}$ on $\boldsymbol{S}$ in the last step. Similarly we have

$$(G^2)_{cd} \prec N\Psi^2. \tag{3.6.3}$$

Then (3.4.8) follows from (3.2.2), (3.6.2), and (3.6.3).

Next, the expression (3.4.9) is a direct consequence of (3.2.4) and (3.2.6) with two resolvents and one traceless matrix. It remains to prove (3.4.10).

We proceed by computing the moments of the quantity $\left(GA\bar{G}G\right)_{cd}$. For the rest of the proof, we assume the spectral parameter $z = E + \mathrm{i}\eta \in \boldsymbol{S}$ satisfies $E \in I_\ell$ and $\eta = \eta_\ell$ as in the statement of Lemma 3.4.2.

Fix $D > 1$, we have

$$\begin{aligned}\mathbb{E}\left[z\left|(GA\bar{G}G)_{cd}\right|^{2D}\right] =&\mathbb{E}\left[z(GA\bar{G}G)_{cd}(GA\bar{G}G)_{cd}^{D-1}(\bar{G}AG\bar{G})_{cd}^{D}\right]\\ =&\mathbb{E}\left[\sum_k h_{ck}(GA\bar{G}G)_{kd}(GA\bar{G}G)_{cd}^{D-1}(\bar{G}AG\bar{G})_{cd}^{D}\right]\\ &-\mathbb{E}\left[(A\bar{G}G)_{cd}(GA\bar{G}G)_{cd}^{D-1}(\bar{G}AG\bar{G})_{cd}^{D}\right],\end{aligned} \tag{3.6.4}$$

where we used the identity $zG = HG - I$ in the last equality.

Let $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N$ be an orthonormal basis of $\mathbb{R}^N$, such that $\boldsymbol{e}_c^\top A\boldsymbol{w}_1 \le 1$ and $\boldsymbol{e}_c^\top A\boldsymbol{w}_i = 0$ for $i = 2, 3, \cdots, N$. We have the following high probability bound

$$\left|(A\bar{G}G)_{cd}\right| = \left|\sum_i \boldsymbol{e}_c^\top A\boldsymbol{w}_i\boldsymbol{w}_i^\top \bar{G}G\boldsymbol{e}_d\right| \le (G\bar{G})_{\boldsymbol{w}_1 d} \prec N\Psi^2,$$

where we used (3.4.8) and (3.4.6) in the last step.

Using Young's inequality with powers $p = 2D$ and $q = (2D)/(2D-1)$, the last line in (3.6.4) can be bounded by

$$\left|\mathbb{E}\left[(A\bar{G}G)_{cd}(GA\bar{G}G)_{cd}^{D-1}(\bar{G}AG\bar{G})_{cd}^{D}\right]\right| \leqslant N^{2D\varepsilon}(N\Psi^2)^{2D} + N^{-\frac{2D\varepsilon}{2D-1}}\mathbb{E}\left[\left|(GA\bar{G}G)_{cd}\right|^{2D}\right], \tag{3.6.5}$$

for any small $\varepsilon > 0$.

Applying Lemma 3.6.1 to the third line in (3.6.4), with $T = 12D$, we have

$$\begin{aligned}&\mathbb{E}\left[\sum_k h_{ck}(GA\bar{G}G)_{kd}(GA\bar{G}G)_{cd}^{D-1}(\bar{G}AG\bar{G})_{cd}^{D}\right]\\ =&\sum_{r=1}^{12D}\frac{1}{r!N^{(r+1)/2}}\sum_k \kappa_{r+1}^{c,k}\mathbb{E}\left[\partial_{ck}^{r}(GA\bar{G}G)_{kd}(GA\bar{G}G)_{cd}^{D-1}(\bar{G}AG\bar{G})_{cd}^{D}\right]+\Omega,\end{aligned} \tag{3.6.6}$$

where $\kappa_r^{c,k}$ is the r-th cumulant of $\sqrt{N}h_{ck}$ and Ω denotes the error term in (3.6.1).

We split the sum in (3.6.6) into two cases and handle the error term Ω at the end.

1. When $r=1$, the summand is

$$\sum_k \frac{1+\delta_{ck}}{N}\mathbb{E}\left[\partial_{ck}(GA\bar{G}G)_{kd}(GA\bar{G}G)_{cd}^{D-1}(\bar{G}AG\bar{G})_{cd}^{D}\right]. \tag{3.6.7}$$

Note that

$$(1+\delta_{ck})\partial_{ck}G_{ij} = -G_{ic}G_{kj} - G_{ik}G_{cj}.$$

We have

$$\begin{aligned}
&(1+\delta_{ck})\partial_{ck}\left[(GA\bar{G}G)_{kd}\left(GA\bar{G}G\right)_{cd}^{D-1}\left(\bar{G}AG\bar{G}\right)_{cd}^{D}\right]\\
=&-[G_{kk}(GA\bar{G}G)_{cd}+G_{kc}(GA\bar{G}G)_{kd}+(GA\bar{G})_{kc}(\bar{G}G)_{kd}+(GA\bar{G})_{kk}(\bar{G}G)_{cd}\\
&\qquad+(GA\bar{G}G)_{kc}G_{kd}+(GA\bar{G}G)_{kk}G_{cd}]\cdot\left(GA\bar{G}G\right)_{cd}^{D-1}\left(\bar{G}AG\bar{G}\right)_{cd}^{D}\\
&-(D-1)\cdot[G_{cc}(GA\bar{G}G)_{kd}+G_{ck}(GA\bar{G}G)_{cd}+\left(GA\bar{G}\right)_{cc}\left(\bar{G}G\right)_{kd}+\left(GA\bar{G}\right)_{ck}\left(\bar{G}G\right)_{cd}\\
&\qquad+\left(GA\bar{G}G\right)_{cc}G_{kd}+\left(GA\bar{G}G\right)_{ck}G_{cd}]\cdot\left(GA\bar{G}G\right)_{kd}\left(GA\bar{G}G\right)_{cd}^{D-2}\left(\bar{G}AG\bar{G}\right)_{cd}^{D}\\
&-D\cdot[\bar{G}_{cc}\left(\bar{G}AG\bar{G}\right)_{kd}+\bar{G}_{ck}\left(\bar{G}AG\bar{G}\right)_{cd}+\left(\bar{G}AG\right)_{cc}\left(G\bar{G}\right)_{kd}+\left(\bar{G}AG\right)_{ck}\left(G\bar{G}\right)_{cd}\\
&\qquad+\left(\bar{G}AG\bar{G}\right)_{cc}\bar{G}_{kd}+\left(\bar{G}AG\bar{G}\right)_{ck}\bar{G}_{cd}]\cdot(GA\bar{G}G)_{kd}\left(GA\bar{G}G\right)_{cd}^{D-1}\left(\bar{G}AG\bar{G}\right)_{cd}^{D-1}.
\end{aligned} \tag{3.6.8}$$

Inserting (3.6.8) into (3.6.7), we have

$$\begin{aligned}
&\sum_k \frac{1+\delta_{ck}}{N}\partial_{ck}(GA\bar{G}G)_{kd}(GA\bar{G}G)_{cd}^{D-1}(\bar{G}AG\bar{G})_{cd}^{D}\\
=&-m\left|(GA\bar{G}G)_{cd}\right|^{2D}-\left(\sum_{i=1}^{5}\alpha_i\right)(GA\bar{G}G)_{cd}^{D-1}(\bar{G}AG\bar{G})_{cd}^{D}\\
&-(D-1)\left(\sum_{i=1}^{6}\beta_i\right)(GA\bar{G}G)_{cd}^{D-2}(\bar{G}AG\bar{G})_{cd}^{D}-D\left(\sum_{i=1}^{6}\bar{\beta}_i\right)(GA\bar{G}G)_{cd}^{D-1}(\bar{G}AG\bar{G})_{cd}^{D-1},
\end{aligned} \tag{3.6.9}$$

where

$$\alpha_1 = \frac{1}{N}(GGA\bar{G}G)_{cd}, \quad \alpha_2 = \frac{1}{N}(\bar{G}AG\bar{G}G)_{cd}, \quad \alpha_3 = \frac{1}{N}(G\bar{G}AGG)_{cd},$$
$$\alpha_4 = \frac{1}{N}\operatorname{Tr}(GA\bar{G})(G\bar{G})_{cd}, \quad \alpha_5 = \frac{1}{N}G_{cd}\operatorname{Tr}(GA\bar{G}G),$$
$$\beta_1 = \frac{1}{N}G_{cc}(G\bar{G}AGGA\bar{G}G)_{dd}, \quad \beta_2 = \frac{1}{N}(GA\bar{G}G)_{cd}(GGA\bar{G}G)_{cd},$$
$$\beta_3 = \frac{1}{N}(GA\bar{G})_{cc}(G\bar{G}GA\bar{G}G)_{dd}, \quad \beta_4 = \frac{1}{N}(\bar{G}G)_{cd}(GA\bar{G}GA\bar{G}G)_{cd},$$
$$\beta_5 = \frac{1}{N}(GA\bar{G}G)_{cc}(GGA\bar{G}G)_{dd}, \quad \beta_6 = \frac{1}{N}G_{cd}(GA\bar{G}GGA\bar{G}G)_{cd}.$$

We used the fact that G is symmetric, with $G_{ij} = G_{ji}$ for all $i, j \in [\![1, N]\!]$, in the above calculations. We now start to estimate α_i, β_i.

- By (3.2.4) and (3.2.6), we have

$$|\alpha_1| = \left|\frac{1}{N}(GGA\overline{G}G)_{cd}\right| \prec N^{-3/2}\eta^{-3} \leq N^{3/2}\Psi^3,$$

  where we used $1/(N\eta) \leq \Psi$ in the last inequality.

- By (3.2.4) and (3.2.6), we have

$$|\alpha_2| = \left|\frac{1}{N}(\overline{G}AG\overline{G}G)_{cd}\right| \prec N^{-3/2}\eta^{-3} \leq N^{3/2}\Psi^3.$$

- By the definition of M in (3.2.3), we have

$$\operatorname{Tr}(M(z, A, \overline{z})I) = \operatorname{Tr}(A|m_{\mathrm{sc}}(z)|^2) = 0.$$

  Combining the previous equation with (3.2.5), we have

$$\left|\frac{1}{N}\operatorname{Tr}(GA\overline{G})\right| \prec N^{-1}\eta^{-3/2} \leq N^{1/2}\Psi^{3/2}.$$

Therefore

$$|\alpha_3| = \left|\frac{1}{N}\operatorname{Tr}(GA\overline{G})(G\overline{G})_{cd}\right| \prec N^{3/2}\Psi^{7/2},$$

where we used $\left(G\overline{G}\right)_{cd} \prec N\Psi^2$ (from (3.4.8)).

- By the same arguments used to bound α_1 and α_2, we have

$$|\alpha_4| \prec N^{-3/2}\eta^{-3} \leq N^{3/2}\Psi^3.$$

- By definition of M in (3.2.3), we have

$$\begin{aligned}\operatorname{Tr}(M(z, A, \overline{z}, I, z)I) &= \frac{1}{2\eta}\operatorname{Tr}(M(z, A, \overline{z})I - M(z, A, z)I)\\ &= \frac{1}{2\eta}\operatorname{Tr}(A|m_{\mathrm{sc}}(z)|^2 - Am_{\mathrm{sc}}(z)^2) = 0.\end{aligned}$$

Combining this equation with (3.2.5), we have

$$\left|\frac{1}{N}\operatorname{Tr}(GA\overline{G}G)\right| \prec N^{-1}\eta^{-5/2} \leq N^{3/2}\Psi^{5/2}.$$

Therefore

$$|\alpha_5| = \left|\frac{1}{N}G_{cd}\operatorname{Tr}(GA\overline{G}G)\right| \prec N^{3/2}\Psi^{5/2}.$$

- For β_1, we cannot directly use (3.2.6) since the operator norm of the deterministic part

$$M(z, I, \overline{z}, A, z, I, z, A, \overline{z}, I, z)$$

is not small compared to its fluctuations.

By [51, Equation (3.12)], we have

$$M(\bar{z}, A, z, I, z, A, \bar{z}, I, z) = \frac{1}{2\eta}\Big(M(\bar{z}, A, z, I, z, A, \bar{z}) - M(\bar{z}, A, z, I, z, A, z)\Big).$$

By definition, we have

$$\begin{aligned}
&M(\bar{z}, A, z, I, z, A, \bar{z})\\
=&\frac{1}{N} m_\circ[\bar{z}, z] m_\circ[z, \bar{z}] \operatorname{Tr}(A^2) I + A^2 m_\circ[\bar{z}, z] m_\circ[z] m_\circ[\bar{z}] + A^2 m_\circ^2[z] m_\circ^2[\bar{z}]\\
=&\frac{1}{N} \operatorname{Tr}(A^2) I \left(m_{\mathrm{sc}}[\bar{z}, z]^2 - 2 m_{\mathrm{sc}}[z, \bar{z}] |m_{\mathrm{sc}}(z)|^2 + |m_{\mathrm{sc}}(z)|^4\right)\\
&+ A^2 m_{\mathrm{sc}}[\bar{z}, z] |m_{\mathrm{sc}}(z)|^2,
\end{aligned}$$

and similarly,

$$\begin{aligned}
&M(\bar{z}, A, z, I, z, A, z)\\
=&\frac{1}{N} m_\circ[\bar{z}, z] m_\circ[z, z] \operatorname{Tr}(A^2) I + A^2 m_\circ[\bar{z}, z] m_\circ^2[z] + A^2 m_\circ^3[z] m_\circ[\bar{z}]\\
=&\frac{1}{N} \operatorname{Tr}(A^2) I \left(m_{\mathrm{sc}}[\bar{z}, z] m_{\mathrm{sc}}[z, z] - m_{\mathrm{sc}}[z, \bar{z}] m_{\mathrm{sc}}^2(z) - m_{\mathrm{sc}}[z, z] |m_{\mathrm{sc}}(z)|^2 + m_{\mathrm{sc}}^3(z) m_{\mathrm{sc}}(\bar{z})\right)\\
&+ A^2 m_{\mathrm{sc}}[\bar{z}, z] m_{\mathrm{sc}}(z)^2.
\end{aligned}$$

By [51, Equation (A.1)], there exists a constant $C > 0$ such that

$$\left|m_{\mathrm{sc}}(z)\right| \leq C, \quad \left|m_{\mathrm{sc}}[\bar{z}, z]\right| \leq C\eta^{-1}, \quad \left|m_{\mathrm{sc}}[z, \bar{z}]\right| \leq C\eta^{-1}, \quad \left|m_{\mathrm{sc}}[z, z]\right| \leq C\eta^{-1}.$$

Since $\operatorname{Tr}(A^2)I$ is diagonal, the off-diagonal entries of $\overline{G}AGGA\overline{G}G$ are bounded by $N^{-1/2}\eta^{-7/2} \leq N^3\Psi^{7/2}$ with high probability by (3.2.6). Therefore, by the local law (3.2.2) and (3.4.7), we have

$$\left|\left(G\overline{G}AGGA\overline{G}G\right)_{dd}\right| \leq \sum_{k=1}^{N} |G_{dk}| \left|\left(\overline{G}AGGA\overline{G}G\right)\right|_{kd} \prec N^4\Psi^{9/2}.$$

Thus we have

$$|\beta_1| = \left|\frac{1}{N} G_{cc}(G\bar{G}AGGA\bar{G}G)_{dd}\right| \prec N^3\Psi^{9/2}.$$

- By (3.2.4) and (3.2.6), we have

$$|\beta_2| = \left|\frac{1}{N}(GA\bar{G}G)_{cd}(GGA\bar{G}G)_{cd}\right| \prec \frac{1}{N}(N^{-1/2}\eta^{-2})(N^{-1/2}\eta^{-3}) \leq N^3\Psi^5.$$

- By (3.2.4) and (3.2.6), we have

$$|\beta_3| = \left|\frac{1}{N}(GA\bar{G})_{cc}(G\bar{G}GA\bar{G}G)_{dd}\right| \prec \frac{1}{N}(N^{-1/2}\eta^{-1})(N^{-1/2}\eta^{-4}) \leq N^3\Psi^5.$$

- By (3.2.4), (3.2.6) and (3.4.7), we have

$$|\beta_4| = \left|\frac{1}{N}(\bar{G}G)_{cd}(GA\bar{G}G\bar{G}AG)_{cd}\right| \prec \frac{1}{N}(N\Psi^2)(\eta^{-3}) \leq N^3\Psi^5.$$

- By the same argument that is used to bound β_2, we have

$$|\beta_5| = \left|\frac{1}{N}(GA\bar{G}G)_{cc}(GGA\bar{G}G)_{dd}\right| \prec N^3\Psi^5.$$

- By the same argument that is used to bound β_1, we have

$$|\beta_6| = \left|\frac{1}{N} G_{cd}(GA\bar{G}GGA\bar{G}G)_{cd}\right| \prec N^3\Psi^{9/2}.$$

let $\varepsilon > 0$ be a parameter. In (3.6.9), we use Young's inequality twice in the second inequality, with powers $p = 2D$ and $q = (2D)/(2D-1)$ for terms containing α_i's and powers $p = D$ and $q = D/(D-1)$ for terms containing β_i's, to show that

$$\begin{aligned}
&\left|\mathbb{E}\left[m|(GA\bar{G}G)_{cd}|^{2D} + \sum_k \frac{1+\delta_{ck}}{N}\partial_{ck}(GA\bar{G}G)_{kd}(GA\bar{G}G)_{cd}^{D-1}(\bar{G}AG\bar{G})_{cd}^{D}\right]\right| \\
&\leqslant \sum_{i=1}^{5}\mathbb{E}\left[|\alpha_i|\left|(GA\bar{G}G)_{cd}\right|^{2D-1}\right] + 2\sum_{i=1}^{6}\mathbb{E}\left[|\beta_i|\left|(GA\bar{G}G)_{cd}\right|^{2D-2}\right] \\
&\leqslant \sum_{i=1}^{5} N^{2D\varepsilon}\mathbb{E}\left[|\alpha_i|^{2D}\right] + 2\sum_{i=1}^{6} N^{D\varepsilon}\mathbb{E}\left[|\beta_i|^{D}\right] + \left(5N^{-\frac{2D\varepsilon}{2D-1}} + 12N^{-\frac{D\varepsilon}{D-1}}\right)\mathbb{E}\left[\left|(GA\bar{G}G)_{cd}\right|^{2D}\right] \\
&\prec 20N^{2D\varepsilon}N^{3D}\Psi^{9D/2} + 20N^{-\frac{2D\varepsilon}{2D-1}}\mathbb{E}\left[\left|(GA\bar{G}G)_{cd}\right|^{2D}\right].
\end{aligned} \tag{3.6.10}$$

2. Now we fix $r \geq 2$ in the cumulant expansion. In this case, we use the following relaxation of (3.2.4) and (3.2.6). Since it is a direct consequence of these inequalities, we omit the proof.

   Corollary 3.6.2 ("Coarser" isotropic local law)**.** *Fix $k \in \mathbb{N}$, and $z_1, \ldots, z_{k+1} \in \boldsymbol{S}$. Let $A_1, \ldots, A_k$ be deterministic matrices such that $\|A_j\| \leq 1$, and m of them satisfy* $\operatorname{Tr} A_j = 0$ *for some $0 \leqslant m \leqslant k$. Suppose that* $\min_j \operatorname{dist}(z_j, [-2,2]) \leqslant 1$. *Then for arbitrary deterministic vectors $\boldsymbol{x}, \boldsymbol{y}$ such that $\|\boldsymbol{x}\| + \|\boldsymbol{y}\| \leq 2$, we have*

$$|\langle \boldsymbol{x}, (G_1A_1\cdots G_kA_kG_{k+1})\,\boldsymbol{y}\rangle| \prec N^{k-\frac{m}{2}} \tag{3.6.11}$$

   with $G_j := G(z_j)$.

   We call a term of the form $(G_1B_1\cdots G_{s+1})_{ij}$ a block. The effect of one differentiation operator ∂_{ak} on a product of blocks is to increase the number of blocks and the number of G 's by exactly one each, while keeping the number of A unchanged. And from (3.6.11), we know the effect of each traceless matrix A is a contribution of a factor of $N^{-1/2}$.

   - Suppose $r < 2D-1$. Note that when using the product rule, ∂_{ck}^r can at most operate on r different blocks, there are at least $2D-r-1$ blocks of $(GA\bar{G}G)_{cd}$

or $(\bar{G}AG\bar{G})_{cd}$ which are unaffected. In view of this, we have

$$\begin{aligned}
&\left|\partial_{ck}^{r}\left[\left(GA\bar{G}G\right)_{kd}(GA\bar{G}G)_{cd}^{D-1}(\bar{G}AG\bar{G})_{cd}^{D}\right]\right| \\
&\le \sum_{\substack{r_1+r_2=r\\ r_1\le D-1\\ r_2\le D}}\binom{D}{r_1}\binom{D-1}{r_2}\left|\partial_{ck}^{r}\left[\left(GA\bar{G}G\right)_{kd}(GA\bar{G}G)_{cd}^{r_1}(\bar{G}AG\bar{G})_{cd}^{r_2}\right]\right|\left|\left(GA\bar{G}G\right)_{cd}\right|^{2D-1-r} \\
&\le D^{r}\sum_{\substack{r_1+r_2=r\\ r_1\le D-1\\ r_2\le D}}\left|\partial_{ck}^{r}\left[\left(GA\bar{G}G\right)_{kd}(GA\bar{G}G)_{cd}^{r_1}(\bar{G}AG\bar{G})_{cd}^{r_2}\right]\right|\cdot\left|\left(GA\bar{G}G\right)_{cd}\right|^{2D-1-r}.
\end{aligned} \tag{3.6.12}$$

There are $1+r$ blocks, $3(r+1)$ G's, and $1+r$ A's in

$$\left(GA\bar{G}G\right)_{kd}(GA\bar{G}G)_{cd}^{r_1}(\bar{G}AG\bar{G})_{cd}^{r_2}.$$

and after the operation of ∂_{ck}^{r}, there are $2r+1$ blocks, $4r+3$ G's and $1+r$ A's. By Corollary 3.6.2, we know

$$\left|\partial_{ck}^{r}\left(GA\bar{G}G\right)_{kd}(GA\bar{G}G)_{cd}^{r_1}(\bar{G}AG\bar{G})_{cd}^{r_2}\right| \prec N^{(4r+2)-(2r)-\frac{1+r}{2}} = N^{3(r+1)/2}. \tag{3.6.13}$$

Combining (3.6.12) and (3.6.13), we deduce that a summand in (3.6.6) with $2\le r<2D-1$ can be bounded by

$$\begin{aligned}
&\left|\frac{\kappa_{r+1}}{r!N^{(r+1)/2}}\sum_{k}\mathbb{E}\left[\partial_{ck}^{r}(GA\bar{G}G)_{kd}(GA\bar{G}G)_{cd}^{D-1}(\bar{G}AG\bar{G})_{cd}^{D}\right]\right| \\
&\prec D^{r+1}N^{r+2}\mathbb{E}\left[\left|\left(GA\bar{G}G\right)_{cd}\right|^{2D-1-r}\right] \\
&\le D^{r+1}N^{\frac{2D\varepsilon}{r+1}+\frac{2D(r+2)}{r+1}}+D^{r+1}N^{-\frac{2D\varepsilon}{2D-r-1}}\mathbb{E}\left[\left|\left(GA\bar{G}G\right)_{cd}\right|^{2D}\right],
\end{aligned} \tag{3.6.14}$$

for any small $\varepsilon>0$, where we used Young's inequality in the last step.

- Suppose $r\ge 2D-1$. There are initially $2D$ blocks, $6D$ G's and $2D$ A's in

$$(GA\bar{G}G)_{kd}(GA\bar{G}G)_{cd}^{D-1}(\bar{G}AG\bar{G})_{cd}^{D},$$

and after the operation of ∂_{ck}^r, there are $2D+r$ blocks, $6D+r$ G's and $2D$ A's. Then by Corollary 3.6.2, we have

$$\left|\partial_{ck}^r (GA\bar{G}G)_{kd}(GA\bar{G}G)_{cd}^{D-1}(\bar{G}AG\bar{G})_{cd}^D\right| \prec N^{3D}.$$

Therefore, a summand in (3.6.6) with $r \geq 2D-1$ can be bounded by

$$\begin{aligned} \left|\frac{\kappa_{r+1}}{r!N^{(r+1)/2}} \sum_k \mathbb{E}\left[\partial_{ck}^r (GA\bar{G}G)_{kd}(GA\bar{G}G)_{cd}^{D-1}(\bar{G}AG\bar{G})_{cd}^D\right]\right| &\prec C(D)N^{3D-\frac{r+1}{2}+1} \\ &\leq C(D)N^{D+1}. \end{aligned} \tag{3.6.15}$$

3. Finally, we consider the error term Ω in (3.6.6). Define $H_t^{(k)}$ by

$$\left(H_t^{(k)}\right)_{ij} = \begin{cases} t, & \text{if } i=c, j=k, \\ h_{ij}, & \text{otherwise.} \end{cases}$$

By Lemma 3.6.1,

$$\Omega \leq C_{12D} \sum_k \mathbb{E}\left[|h_{ck}|^{12D+2}\right] \mathbb{E}\left[\left|\partial_{ck}^{12D+1}(G_t^{(k)}A\bar{G}_t^{(k)}G_t^{(k)})_{kd}(G_t^{(k)}A\bar{G}_t^{(k)}G_t^{(k)})_{cd}^{D-1}(\bar{G}_t^{(k)}AG_t^{(k)}\bar{G}_t^{(k)})_{cd}^D\right|\right], \tag{3.6.16}$$

where $G_t^{(k)}$ is the resolvent of $H_t^{(k)}$.

Originally, (3.6.16) has $2D$ blocks and $6D$ G's, after taking $(12D+1)$-th derivatives, there are $14D+1$ blocks and $18D+1$ G's. Using the moment assumption (3.1.5) and the trivial bound $\|G_t^{(k)}\| \leq \eta^{-1} \leq N$ from (3.4.4), we have

$$\Omega \leq C(D)NN^{-6D-1}N^{(18D+1)-(14D+1)} = C(D)N^{-2D}. \tag{3.6.17}$$

Combining (3.6.4), (3.6.5), (3.6.6), (3.6.10), (3.6.14), (3.6.15) and (3.6.17), we have

$$\begin{aligned}&\left|\mathbb{E}\left[(z+m)|(GA\bar{G}G)_{cd}|^{2D}\right]\right|\\ \prec\,&C(D)N^{2D\varepsilon}N^{3D}\Psi^{9D/2}+C(D)N^{-\frac{2D\varepsilon}{2D-1}}\mathbb{E}\left[\left|(GA\bar{G}G)_{cd}\right|^{2D}\right].\end{aligned}$$

Using the local law $|m-m_{\mathrm{sc}}|\prec\Psi$ [20, Theorem 2.6], we have

$$\begin{aligned}&(z+m_{\mathrm{sc}})\left|\mathbb{E}\left[|(GA\bar{G}G)_{cd}|^{2D}\right]\right|\\ \prec\,&C(D)N^{2D\varepsilon}N^{3D}\Psi^{9D/2}+C(D)N^{-\frac{2D\varepsilon}{2D-1}}\mathbb{E}\left[\left|(GA\bar{G}G)_{cd}\right|^{2D}\right]+\Psi\cdot\mathbb{E}\left[|(GA\bar{G}G)_{cd}|^{2D}\right]\end{aligned}$$

Since $z+m_{\mathrm{sc}}=1/m_{\mathrm{sc}}$ (recall (3.5.1)), which is bounded away from 0 (by [20, Equation (3.2)]), and $\varepsilon>0$ is arbitrary, we have

$$\mathbb{E}\left[\left|(GA\bar{G}G)_{cd}\right|^{2D}\right]\prec N^{3D}\Psi^{9D/2}.$$

Since $D>1$ is arbitrary, we have

$$|(GA\bar{G}G)_{cd}|\prec N^{3/2}\Psi^{9/4}.$$

This completes the proof of (3.4.10). □

CHAPTER 4

Central limit theorem for the complex eigenvalues of Gaussian random matrices

4.1 Introduction

4.1.1 Main Result

We prove a central limit theorem (CLT) for the eigenvalue counting function of a matrix of real Gaussian random variables in regions of the complex plane. While such a result is well known for matrices of complex Gaussians (see [43, Section 3.1] for a survey), to the best of our knowledge, the analogous statement for real Gaussian matrices has not previously been addressed.

We begin by defining the random matrix ensemble of interest in this work.

Definition 4.1.1. *For all $N \in \mathbb{N}$, let $G_N = (g_{ij})_{1\leq i,j\leq N}$ be a random matrix whose entries are mutually independent Gaussian random variables with mean zero and variance one. We call G_N the* real Ginibre matrix *(GinOE) of dimension N. We also denote*

$W_N = N^{-1/2} G_N$.

In the limit as N goes to infinity, it is known that the empirical spectral distribution of W_N tends to the uniform measure on the unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ [36]. We note that the eigenvalues of W_N come in conjugate pairs, since W_N is real; if $\lambda \in \mathbb{C}$ is an eigenvalue, then so is $\bar{\lambda}$. It is therefore natural when studying the fluctuations of the eigenvalues of W_N to restrict attention to the upper half disk $\mathbb{D}^+ = \{z \in \mathbb{C} : |z| < 1, \Im z > 0\}$. We recall that a domain is defined as a non-empty connected open subset of $\mathbb{C}$.

Definition 4.1.2. *We say that a domain A is* admissible *if $\overline{A} \subset \mathbb{D}^+$.*

This condition is slightly stronger than requiring $A \subset \mathbb{D}^+$, since it enforces a separation between A and the boundary of $\mathbb{D}^+$. We also recall that a domain is said to be Lipschitz if its boundary is locally the graph of a Lipschitz continuous function; see [127, Definition 12.9]. Given an admissible Lipschitz domain A, we let $\ell(\partial A)$ denote the length of its boundary.

Denote the eigenvalues of W_N by $\lambda_1, \ldots, \lambda_N$, in an arbitrary order. Given an admissible domain A, we define $f_A : \mathbb{C} \to \mathbb{R}$ by $f_A(z) = \mathbb{1}_A(z)$, and define the (N-dependent) random variable

$$X_A = \sum_{i=1}^{N} f_A(\lambda_i) - \mathbb{E}\left[\sum_{i=1}^{N} f_A(\lambda_i)\right]. \tag{4.1.1}$$

The following theorem is our main result. We let $\mathcal{N}(0, c)$ denote a Gaussian random variable with mean zero and variance $c > 0$.

Theorem 4.1.3. *Let A be an admissible Lipschitz domain. Then we have the weak convergence*

$$\lim_{N\to\infty} \frac{X_A}{N^{1/4}} = \mathcal{N}\left(0, \frac{\ell(\partial A)}{2\pi^{3/2}}\right). \tag{4.1.2}$$

The variance of X_A is of order $N^{1/2}$, which is smaller than the variance of order N seen in sums of independent random variables. This is due to the strong correlations between the eigenvalues of W_N [135]. Further, the variance of the Gaussian in (4.1.2) is identical to the one in the analogous theorem for complex Gaussian matrices [43, (3.9)].

4.1.2 Background

The analogue of Theorem 5.2.5 for a complex Ginibre matrix (GinUE) is known. It is a consequence of a theorem that provides a CLT for a broad class of determinantal point processes proved in [155, Section 2] (see also [63]), together with the the explicit computation of the asymptotic variance in [130, Corollary 1.2.1].[1] See [46, Corollary 1.7] for an alternative proof in the case where A has a smooth boundary. Further, a local CLT for the counting function of the GinUE eigenvalues was derived in [90].

All of these works crucially rely on the fact that the eigenvalues of the GinUE form a determinantal point process. While this determinantal structure enables a precise analysis of many aspects of the GinUE, it is absent in the GinOE. Instead, the eigenvalues of the GinOE form a Pfaffian point process, and consequently they are more difficult to study [44].

Previous work on linear statistics of the GinOE has considered smooth test functions of the complex eigenvalues [120, 142], differentiable functions of the real eigenvalues [151, 120, 87], general functions of the real eigenvalues [87], and the number of real eigenvalues [87, 151, 89, 72, 71, 112]. There have also been a few recent articles proving CLTs for linear statistics of matrices of general i.i.d. random variables when the test function has at least two derivatives [56, 55, 54]. Proving a CLT for the eigenvalue counting function in this more general setting remains an open problem.

4.1.3 Outline

In Section 4.2, we collect several preliminary lemmas, and show that the Pfaffian correlations of the GinOE eigenvalues may be quantitatively approximated by determinantal

[1] While [63] gives details only for certain Gaussian matrices, the authors note (in a remark attributed to H. Widom) that their method works in much greater generality, as later demonstrated in [155].

correlations. In Section 4.3, we compute the variance and higher cumulants of X_A, and show that they match those of the desired Gaussian distribution, concluding the proof of Theorem 5.2.5. Using the results of [130], it is straightforward to extend Theorem 5.2.5 to all domains with finite perimeter (so-called Caccioppoli sets) and certain domains with fractal boundaries. We briefly discuss this point in Remark 4.3.8.

4.1.4 Acknowledgments

The authors thank P. Bourgade for suggesting the problem, C. D. Sinclair for helpful comments on the references [37, 152], and P. J. Forrester for comments on a preliminary draft. They are grateful to the referees for several useful suggestions.

P. L. was supported by the NSF postdoctoral fellowship DMS-2202891. X. X. was supported by NSF grant DMS-1954351.

4.2 Preliminary Results

Set $\mathbb{C}^* = \mathbb{C}\backslash\mathbb{R}$. We recall that for all $k \in \mathbb{N}$, the complex–complex correlation functions $\rho_k^{(N)} \colon (\mathbb{C}^*)^k \to \mathbb{R}$ for G_N are defined by the following property [37, (5.1)]. For every compactly supported, bounded Borel-measurable function $f \colon (\mathbb{C}^*)^k \to \mathbb{R}$, we have

$$\int_{(\mathbb{C}^*)^k} f(z_1, \dots, z_k)\rho_k^{(N)}(z_1, \dots, z_k)\, dz_1 \dots dz_k = \mathbb{E}\left[\sum_{(i_1,\dots,i_k)\in\mathcal{I}_k} f(w_{i_1}, \dots, w_{i_k})\right], \quad (4.2.1)$$

where $\mathcal{I}_k \subset \{1, \dots, N\}^N$ is the set of pairwise distinct k-tuples of indices, $\{w_i\}_{i=1}^N$ are the eigenvalues of G_N, and we use dz_i to denote the Lebesgue measure on $\mathbb{C}$. We typically write ρ_k instead of $\rho_k^{(N)}$, since the value of N will be clear from context. We also recall that if $M = (M_{ij})_{i,j=1}^{2n}$ is a $2n \times 2n$ skew-symmetric matrix, its Pfaffian is defined as

$$\mathrm{Pf}(M) = \frac{1}{2^n n!} \sum_{\sigma\in S_{2n}} \mathrm{sgn}(\sigma) \prod_{i=1}^{n} M_{\sigma(2i-1),\sigma(2i)}, \quad (4.2.2)$$

where S_{2n} is the symmetric group of degree $2n$.

The following lemma, taken from [136, Appendix B.3], identifies the correlation functions ρ_k explicitly.

Lemma 4.2.1. *The k-point complex–complex correlation functions of the N-dimensional real Ginibre ensemble G_N are given by*

$$\rho_k(z_1, \dots, z_k) = \mathrm{Pf}(K(z_i, z_j))_{1 \le i,j \le k}, \tag{4.2.3}$$

where $(K(z_i, z_j))_{1 \le i,j \le k}$ is a $2k \times 2k$ matrix composed of the 2×2 blocks

$$K(z_i, z_j) = \begin{pmatrix} D_N(z_i, z_j) & S_N(z_i, z_j) \\ -S_N(z_j, z_i) & I_N(z_i, z_j) \end{pmatrix},$$

and D_N, I_N, and S_N are defined by

$$\begin{aligned} S_N(z, w) &= \frac{\mathrm{i}e^{-(1/2)(z-\bar{w})^2}}{\sqrt{2\pi}}(\bar{w} - z)G(z, w)s_N(z\bar{w}), \\ D_N(z, w) &= \frac{e^{-(1/2)(z-w)^2}}{\sqrt{2\pi}}(w - z)G(z, w)s_N(zw), \\ I_N(z, w) &= \frac{e^{-(1/2)(\bar{z}-\bar{w})^2}}{\sqrt{2\pi}}(\bar{z} - \bar{w})G(z, w)s_N(\bar{z}\bar{w}), \end{aligned}$$

where $z, w \in \mathbb{C}^*$ and

$$G(z, w) = \sqrt{\operatorname{erfc}(\sqrt{2}\Im(z)) \operatorname{erfc}(\sqrt{2}\Im(w))}, \qquad \operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-t^2)\, dt,$$
$$s_N(z) = e^{-z} \sum_{j=0}^{N-1} \frac{z^j}{j!}.$$

Remark 4.2.2. The functions ρ_k were first determined explicitly in [92]. The Pfaffian form in (4.2.3) was derived in the case of even N in [37]. Subsequently, a variety of methods have been used to recover this form for all N [91, 152, 153] (see also [136, Section

4.6]).

By a change of variable it is straightforward to see that the k-th correlation function for the complex eigenvalues of W_N is $N^k \rho_k(\sqrt{N} z_1, \ldots \sqrt{N} z_k)$. The following lemma is useful for controlling these functions and is proved in Section 4.4. We let $d_A = \inf\{|z - w| : z \in A, w \in \partial \mathbb{D}^+\}$ denote the distance between A and the boundary of $\mathbb{D}^+$, and use the standard "big O" notation $O(\cdot)$ for estimates that hold in the limit $N \to \infty$.

Lemma 4.2.3. *Let A be an admissible domain. Then there exists a constant $c(d_A) > 0$ such that*

$$\sup_{z,w \in A} D_N(\sqrt{N} z, \sqrt{N} w) = O(e^{-cN}), \qquad \sup_{z,w \in A} I_N(\sqrt{N} z, \sqrt{N} w) = O(e^{-cN}),$$
$$\sup_{z,w \in A} S_N(\sqrt{N} z, \sqrt{N} w) = O(1),$$

where the implicit constants in the asymptotic notation depend only on d_A.

We next state a useful lemma about Pfaffians, proved in [97, Appendix B].[2]

Lemma 4.2.4. *Let $M = (M_{ij})_{i,j=1}^{2n}$ be a skew-symmetric $2n \times 2n$ matrix such that $M_{ij} = 0$ when $i \equiv j \bmod 2$. Let $\tilde{M} = (\tilde{M})_{i,j=1}^{n}$ be the $n \times n$ matrix formed by setting $\tilde{M}_{ij} = M_{2i-1,2j}$. Then $\mathrm{Pf}(M) = \det(\tilde{M})$.*

Finally, we require the following integral formula from [130, Corollary 3.1.4].

Lemma 4.2.5. *Let $J : \mathbb{C} \to \mathbb{R}$ be radially symmetric (meaning $J(z) = J(|z|)$) and nonnegative. Suppose further that $\int_{\mathbb{C}} J(z) \cdot |z| \, dz = 1$. Then for any admissible Lipschitz region A,*

$$\lim_{N \to \infty} N^{3/2} \int_A \int_{A^c} J\big(\sqrt{N}(z - w)\big) \, dz \, dw = \frac{4}{\pi} \cdot \ell(\partial A).$$

4.3 Proof of Theorem 5.2.5

[2] The statement has appeared earlier in the literature, for example in [93].

4.3.1 Variance Calculation

The following lemma follows from results proved in [120]. We sketch the proof for completeness.

Lemma 4.3.1. *For any admissible domain A,*

$$\operatorname{Var}[X_A] = \frac{N}{\pi}\operatorname{area}(A) - \frac{N^2}{\pi^2}\int_A\int_A \exp(-N|z-w|^2)\,dz\,dw + O(N^{-1}), \tag{4.3.1}$$

where the implicit constant in the asymptotic notation depends only on d_A.

Proof. From the definition (4.1.1) of X_A, we compute

$$\operatorname{Var}[X_A] = \mathbb{E}\left[\sum_{i=1}^N f_A(\lambda_i)\right] + \mathbb{E}\left[\sum_{i\neq j} f_A(\lambda_i)f_A(\lambda_j)\right] - \mathbb{E}\left[\sum_{i=1}^N f_A(\lambda_i)\right]^2.$$

Writing this expression in terms of correlation functions using (4.2.1) and (4.2.3), we obtain

$$\begin{aligned}\operatorname{Var}[X_A] =& N\int_A S_N(\sqrt{N}z,\sqrt{N}z)\,dz - N^2\int_{A^2} S_N(\sqrt{N}z,\sqrt{N}w)^2\,dz\,dw \\ &- N^2\int_{A^2} D_N(\sqrt{N}z,\sqrt{N}w)I_N(\sqrt{N}z,\sqrt{N}w)\,dz\,dw.\end{aligned} \tag{4.3.2}$$

The last term in (4.3.2) vanishes exponentially, by Lemma 4.2.3. The first term is computed in [120, Lemma 7] and equals

$$\frac{N}{\pi}\operatorname{area}(A) - \frac{1}{4\pi}\int_A \frac{dz}{\Im(z)^2} + O(N^{-1}). \tag{4.3.3}$$

The second term is computed in the proof of [120, Lemma 9] and equals

$$-\frac{N^2}{\pi^2}\int_A\int_A \exp(-N|z-w|^2)\,dz\,dw + \frac{1}{4\pi}\int_A \frac{dz}{\Im(z)^2} + O(N^{-1}). \tag{4.3.4}$$

Inserting (4.3.3) and (4.3.4) into (4.3.2) completes the proof.[3] We observe that the asymptotic bounds in the proofs of the cited lemmas rely only on Lemma 4.2.3 and the estimates Lemma 4.4.1 and (4.3.15) stated below, whose error terms depend on A only through d_A. This justifies the claim that the implicit constant in (4.3.1) depends only on d_A, even though this dependence was not made explicit in [120]. □

Lemma 4.3.2. *For any admissible Lipschitz domain A,*

$$\lim_{N\to\infty} \frac{\mathrm{Var}[X_A]}{N^{1/2}} = \frac{1}{2\pi^{3/2}} \cdot \ell(\partial A).$$

Proof. We write

$$\begin{aligned} \int_A \int_A \exp(-N|z-w|^2)\, dz\, dw = \int_A \int_{\mathbb{C}} \exp(-N|z-w|^2)\, dz\, dw \\ - \int_A \int_{A^c} \exp(-N|z-w|^2)\, dz\, dw \end{aligned} \tag{4.3.5}$$

By a change of variable and the Gaussian integral formula $\int_{\mathbb{R}} e^{-x^2} = \sqrt{\pi}$, we have

$$\int_{\mathbb{C}} \exp(-N|z-w|^2)\, dz = \int_{\mathbb{C}} \exp(-N|z|^2)\, dz = \frac{\pi}{N}. \tag{4.3.6}$$

Combining (4.3.1), (4.3.5), and (4.3.6), we obtain

$$\mathrm{Var}[X_A] = \frac{N^2}{\pi^2} \int_A \int_{A^c} \exp(-N|z-w|^2)\, dz\, dw + O(N^{-1}). \tag{4.3.7}$$

By Lemma 4.2.5 applied to the radially-symmetric kernel function $J\colon \mathbb{R}^2 \to \mathbb{R}$ given by $J(r) = 2\pi^{-3/2} \exp(-2r^2)$, we find

$$\lim_{N\to\infty} N^{3/2} \int_A \int_{A^c} \exp(-N|z-w|^2)\, dz\, dw = \frac{\sqrt{\pi}}{2} \cdot \ell(\partial A). \tag{4.3.8}$$

[3] While the main results of [120] require the test function to be smooth, these calculations do not. Further, they were given for even N in [120] using the statement of (4.2.3) for even N in [37]. Their extension to odd N requires only notational changes, given that (4.2.3) is now known for all N.

We conclude by combining (4.3.7) and (4.3.8). $\square$

4.3.2 Higher Cumulants

We recall that given a random variable X, its cumulants $\{\kappa_n\}_{n=1}^{\infty}$ are defined by

$$\log \mathbb{E}[e^{\mathrm{i}tX}] = \sum_{n=1}^{\infty} \kappa_n \frac{(\mathrm{i}t)^n}{n!},$$

and that for every $n \in \mathbb{N}$, there exists a degree n polynomial L_n (independent of the choice of X) such that $\kappa_n = L_n\Big(\mathbb{E}[X], \mathbb{E}[X^2], \dots, \mathbb{E}[X^n]\Big)$.

Let Y be a point process on a subset $\mathcal{D} \subset \mathbb{C}$ with N particles $\{y_i\}_{i=1}^{N}$ and correlation functions $\tau_k \colon \mathcal{D}^k \to \mathbb{R}$ such that

$$\int_{\mathcal{D}^k} f(z_1, \dots, z_k) \tau_k(z_1, \dots, z_k)\, dz_1 \dots dz_k = \mathbb{E}\left[\sum_{(i_1,\dots,i_k)\in\mathcal{I}_k} f(y_{i_1}, \dots, y_{i_k}) \right] \tag{4.3.9}$$

for all $k \in \mathbb{N}$ and all compactly supported, bounded Borel-measurable functions $f \colon \mathcal{D}^k \to \mathbb{R}$. For every domain $A \subset \mathbb{C}$, let $N_A = \sum_{i=1}^{N} \mathbb{1}_A(y_i)$ denote the counting function for A. We insert the test function $f(z_1, \dots, z_k) = \mathbb{1}_A(z_1) \cdots \mathbb{1}_A(z_k)$ into (4.3.9), and note that the number of elements $(i_1, \dots, i_k) \in \mathcal{I}_k$ such that $f(y_i, \dots, y_{i_k}) = 1$ is equal to $N_A(N_A - 1) \cdots (N_A - k + 1)$, since there are N_A choices for i_1, and then $N_A - 1$ choices remaining for i_2, and so on until i_k. This implies the well-known identity

$$\mathbb{E}\Big[N_A(N_A - 1) \cdots (N_A - k + 1)\Big] = \int_{A^k} \tau_k(z_1, \dots, z_k)\, dz_1 \dots dz_k. \tag{4.3.10}$$

Let J_k denote the integral on the right-hand side of (4.3.10). Then (4.3.10) implies that for all $n \in \mathbb{N}$, the moment $\mathbb{E}[N_A^n]$ is equal to a linear combination of the terms $J_1, \dots, J_n$, with universal coefficients (independent of Y). Recalling the definition of the polynomial

L_n, we conclude that for every $n \in \mathbb{N}$, there exists a universal polynomial H_n such that

$$\kappa_n(N_A) = H_n(J_1, \ldots, J_n). \tag{4.3.11}$$

The following lemma is a consequence of Lemma 4.2.3 and Lemma 4.2.4. Let A be an admissible domain, and for all $k \in \mathbb{N}$, set $Q^{(k)}(z_1, \ldots, z_k) = (S_N(z_i, z_j))_{1 \le i,j \le k}$. We define

$$T_k = N^k \int_{A^k} \det Q^{(k)}(\sqrt{N} z_1, \ldots, \sqrt{N} z_k)\, dz_1 \ldots dz_k. \tag{4.3.12}$$

Lemma 4.3.3. *For any admissible domain A, there exists a constant $c(d_A) > 0$ such that $\kappa_n(X_A) = H_n(T_1, \ldots, T_n) + O(e^{-cN})$ for all $n \in \mathbb{N}$. The implicit constant depends only on n and d_A.*

Proof. We begin by computing $\rho_k(\sqrt{N} z_1, \ldots, \sqrt{N} z_k)$ using the definition of ρ_k in (4.2.3) and the definition of a Pfaffian in (4.2.2). By Lemma 4.2.3, all terms in the defining sum (4.2.2) containing a factor of D_N or I_N are exponentially small. We conclude that

$$\sup_{z_1,\ldots,z_k \in \Omega} \left| N^k \rho_k(\sqrt{N} z_1, \ldots, \sqrt{N} z_k) - N^k \operatorname{Pf}\left(\tilde{K}(\sqrt{N} z_i, \sqrt{N} z_j) \right)_{1 \le i,j \le k} \right| \le c^{-1} e^{-cN}, \tag{4.3.13}$$

where where $(\tilde{K}(z_i, z_j))_{1 \le i,j \le k}$ is a $2k \times 2k$ matrix composed of the 2×2 blocks

$$\tilde{K}(z_i, z_j) = \begin{pmatrix} 0 & S_N(z_i, z_j) \\ -S_N(z_j, z_i) & 0 \end{pmatrix}.$$

Lemma 4.2.4 implies

$$\operatorname{Pf}\left(\tilde{K}(\sqrt{N} z_i, \sqrt{N} z_j) \right)_{1 \le i,j \le k} = \det Q^{(k)}(z_1, \ldots, z_k). \tag{4.3.14}$$

Combining (4.3.10), (4.3.13), (4.3.14), and the definition of T_k in (4.3.12), we find

$$\begin{aligned} T_k &= N^k \int_{A^k} \rho_k(\sqrt{N}z_1, \dots, \sqrt{N}z_k)\, dz_1 \dots dz_k + O(e^{-cN}) \\ &= \mathbb{E}\Big[X_A(X_A - 1)\cdots(X_A - k + 1)\Big] + O(e^{-cN}), \end{aligned}$$

since $N^k \rho_k(\sqrt{N}z_1, \dots, \sqrt{N}z_k)$ is the k-th correlation function for the complex eigenvalues of W_N. The conclusion follows after recalling the definition of H_n from (4.3.11) and using the trivial inequality $|T_k| \le 2N^k$. $\square$

Lemma 4.3.3 motivates the next definition.

Definition 4.3.4. *We define the* pseudo-cumulants *of X_A by $\tilde{\kappa}_n = H_n(T_1, \dots, T_n)$ for all $n \in \mathbb{N}$.*

The following cumulant identity is known for determinantal processes [154, (2.6)]. The proof in [154] works for the pseudo-cumulants without modification, since they are defined in terms of a determinantal kernel.[4]

Lemma 4.3.5. *For all $n \in \mathbb{N}$, we have*

$$\begin{aligned} \tilde{\kappa}_n &= \sum_{m=1}^{n} \frac{(-1)^{m-1}}{m} \sum_{\substack{n_1+\cdots+n_m=n \\ n_1,\dots,n_m>0}} \frac{n!}{n_1! \cdots n_m!} \cdot R_m, \\ R_m &= N^m \int_{A^m} \prod_{i=1}^{m} S_N(\sqrt{N}z_i, \sqrt{N}z_{i+1})\, dz_i, \end{aligned}$$

with the convention that $z_{m+1} = z_1$.

The next lemma follows from the previous one by induction; see [155, Lemma 1] for the statement in the case of determinantal processes.

Lemma 4.3.6. *For all $n \in \mathbb{N}$, there exist constants $(\alpha_{nj})_{j=2}^{n-1}$ (independent of A and N)*

[4] They are precisely the cumulants of the determinantal point process defined by the kernel $S_N(\sqrt{N}z, \sqrt{N}w)$, if such a process exists. We do not address the question of existence here, since this claim is not needed.

such that

$$\tilde{\kappa}_n = (-1)^n (n-1)!(R_1 - R_n) + \sum_{j=2}^{n-1} \alpha_{nj} \tilde{\kappa}_j.$$

In light of the previous lemma, we now aim to calculate the terms R_n.

Lemma 4.3.7. *Fix $\delta \in (0, 1/2)$. For $k \geq 2$, we have*

$$R_1 = \frac{N}{\pi} \operatorname{area}(A) + O(1), \quad R_k = \frac{N}{\pi} \operatorname{area}(A) + O(N^{1/2+\delta}),$$

where the implicit constant in the asymptotic notation depends only on d_A, k, and δ.

To prepare for the proof, we recall the standard error function asymptotic

$$\operatorname{erfc}(x) = \frac{e^{-x^2}}{\sqrt{\pi} x} \left(1 + O(x^{-2})\right). \tag{4.3.15}$$

Proof of Lemma 4.3.7. The case $k = 1$ is [120, Lemma 7], so we suppose that $k \geq 2$. Then by (4.3.15) and Lemma 4.4.1, for all $z, w \in A$, we have the asymptotic expansion

$$S_N(\sqrt{N} z, \sqrt{N} w) = U(z, w) \left(1 - \frac{e^{-2(1 - z\bar{w})}}{\sqrt{2\pi N}(1 - z\bar{w})} e^{N(1 - z\bar{w})} (z\bar{w})^N\right) \left(1 + O(N^{-1})\right) \tag{4.3.16}$$

where

$$U(z, w) = \frac{\mathrm{i} e^{(-N/2)(z - \bar{w})^2 - N(\Im(z)^2 + \Im(w)^2)}}{2\pi \sqrt{\Im(z)\Im(w)}} (\bar{w} - z).$$

We claim that the leading order term in R_k is $N^k \int_{A^k} \prod_{i=1}^k U(z_i, z_{i+1})\, dz_i$. To show this, we begin by illustrating how to bound one of the other terms in R_k coming from (4.3.16). We note that there exists a constant $C(d_A, k) > 0$ such that

$$\begin{aligned}
&N^k \left| \int_{A^k} \prod_{i=1}^k U(z_i, z_{i+1}) \frac{e^{-2(1 - z_i \bar{z}_{i+1})}}{\sqrt{2N} \pi z_i \bar{z}_{i+1}} e^{N(1 - z_i \bar{z}_{i+1})} (z_i \bar{z}_{i+1})^N \, dz_i \right| \\
&\leq C N^k \int_{A^k} \prod_{i=1}^k e^{(-N/2)\Re((z_i - \bar{z}_{i+1})^2) - N(\Im(z_i)^2 + \Im(z_{i+1})^2)} e^{N(1 - \Re(z_i \bar{z}_{i+1}) + \ln|z_i \bar{z}_{i+1}|)} \, dz_i \\
&= C N^k \int_{A^k} \prod_{i=1}^k e^{(-N/2)(|z_i|^2 + |z_{i+1}|^2 - 2 - \ln|z_i|^2 - \ln|z_{i+1}|^2)} \, dz_i. \qquad (4.3.17)
\end{aligned}$$

We now observe that the integral in (4.3.17) decays exponentially in z, since $|z|^2-1-\ln|z|^2$ is positive and bounded away from zero for $z \in A$ (since A is admissible). The other error terms can be treated similarly; each has an integrand that decays exponentially.

Introducing the notation

$$g(z,w) = \Re(z)\Im(w) - \Re(w)\Im(z),$$

using (4.3.16), and bounding the error terms as indicated in (4.3.17), we obtain (after observing some cancellation in the exponent) that

$$R_k = \big(1+O(N^{-1})\big)N^k \int_{A^k} \exp\left(-\frac{N}{2}\sum_{i=1}^{k}|z_i - z_{i+1}|^2 + \mathrm{i}N\sum_{i=1}^{N} g(z_i,z_{i+1})\right) \times \prod_{i=1}^{k} \frac{\mathrm{i}(\bar{z}_{i+1}-z_i)}{2\pi\Im(z_i)}\,dz_i + O(e^{-cN}), \tag{4.3.18}$$

for some constant $c(d_A,k)>0$. We now decompose

$$\prod_{i=1}^{k}(\bar{z}_{i+1}-z_i) = \prod_{i=1}^{k}(z_{i+1}-z_i-2\mathrm{i}\Im(z_{i+1})) = (-2i)^k\prod_{i=1}^{k}\Im(z_i) + \varepsilon(z_1,\dots,z_k),$$

where $\varepsilon(z_1,\dots,z_k)$ is the sum of terms containing at least one copy of $(z_i - z_{i+1})$. We claim that all integrals arising from $\varepsilon(z_1,\dots,z_k)$ are negligible. The following computation demonstrates this for terms containing exactly one copy of $(z_i - z_{i+1})$; the other terms are bounded similarly (and are lower order). We have

$$\begin{aligned}
&N^k\left|\int_{A^k}\exp\left(-\frac{N}{2}\sum_{i=1}^{k}|z_i-z_{i+1}|^2 + \mathrm{i}N\sum_{i=1}^{N}g(z_i,z_{i+1})\right)(z_1-z_2)\,dz_1\dots dz_k\right| \\
&\quad\le N^k\int_{A^k}\exp\left(-\frac{N}{2}\sum_{i=1}^{k}|z_i-z_{i+1}|^2\right)|z_1-z_2|\,dz_1\dots dz_k \\
&\quad\le N^k\int_{A^k\cap\{|z_1-z_2|\le N^{-1/2+\delta}\}}\exp\left(-\frac{N}{2}\sum_{i=1}^{k}|z_i-z_{i+1}|^2\right)|z_1-z_2|\,dz_1\dots dz_k + O(e^{-cN}),
\end{aligned}$$

due to the exponential decay of the integrand on the set $A^k\cap\{|z_1-z_2|>N^{-1/2+\delta}\}$. We have

$$
\begin{aligned}
&N^k \int_{A^k \cap \{|z_1 - z_2| \leq N^{-1/2+\delta}\}} \exp\left(-\frac{N}{2}\sum_{i=1}^{k} |z_i - z_{i+1}|^2\right) |z_1 - z_2|\, dz_1 \dots dz_k \\
&\leq N^{k-1/2+\delta} \int_{\mathbb{C}^{k-1}\times A} \exp\left(-\frac{N}{2}\sum_{i=1}^{k-1} |z_i - z_{i+1}|^2\right) dz_1 \dots dz_k = O(N^{1/2+\delta}),
\end{aligned}
$$

where the last inequality follows by directly evaluating the integrals in the variables z_1 through z_{k-1}, then using the fact that $\mathrm{area}(A) \leq 2$.

After bounding these lower-order terms, (4.3.18) becomes

$$
R_k = \pi^{-k} N^k \int_{A^k} \exp\left(-\frac{N}{2}\sum_{i=1}^{k} |z_i - z_{i+1}|^2 + \mathrm{i}N \sum_{i=1}^{k} g(z_i, z_{i+1})\right) dz_1 \dots dz_k + O(N^{1/2+\delta}).
$$

We write R_k as

$$
R_k = \pi^{-k} N^k \left(I_0 - \sum_{j=1}^{k-1} I_j\right) + O(N^{1/2+\delta}), \tag{4.3.19}
$$

where

$$
I_0 = \int_A \int_{\mathbb{C}^{k-1}} \exp\left(-\frac{N}{2}\sum_{i=1}^{k} |z_i - z_{i+1}|^2 + \mathrm{i}N \sum_{i=1}^{k} g(z_i, z_{i+1})\right) dz_1 \dots dz_k,
$$

and I_j for $j \geq 1$ is defined similarly to I_0, with the integral over $A \times \mathbb{C}^{k-1}$ replaced by one over $A^j \times A^c \times \mathbb{C}^{k-j-1}$. I_0 is the leading-order term, and may be computed explicitly. After the change variables by $z_i \mapsto z_i + z_k$ for $i < k$, the variable z_k disappears from the exponent and may be integrated directly. After some simplification, we obtain

$$
I_0 = \mathrm{area}(A) \int_{\mathbb{C}^{k-1}} \exp\left(-N \sum_{i=1}^{k-1} |z_i|^2 + N \sum_{i=1}^{k-2} \overline{z}_i z_{i+1}\right) dz_1 \dots dz_{k-1}
$$

Changing variables to polar coordinates by setting $z_i = r_i e^{\mathrm{i}\theta_i}$, and using the identity

$$
\int_0^{2\pi} \exp(\alpha e^{\mathrm{i}\theta})\, d\theta = \oint \exp(\alpha z) \frac{dz}{\mathrm{i}z} = 2\pi,
$$

valid for any $\alpha \in \mathbb{C}$, to integrate out the θ_i variables, we obtain[5]

$$I_0 = \pi^{k-1} N^{1-k} \operatorname{area}(A). \tag{4.3.20}$$

Next, we note that for every j such that $1 \le j \le k-1$, we have

$$|I_j| \le \int_{A \times A^c \times \mathbb{C}^{k-2}} \exp\left(-\frac{N}{2} \sum_{i=1}^{k} |z_i - z_{i+1}|^2\right) dz_1 \dots dz_k.$$

Recalling (4.3.6), have

$$\int_{\mathbb{C}} \exp\left(-\frac{N}{2}|z_{k-1} - z_k|^2 - \frac{N}{2}|z_k - z_1|^2\right) dz_k \le \int_{\mathbb{C}} \exp\left(-\frac{N}{2}|z_k - z_1|^2\right) dz_k = \frac{\pi}{N}.$$

The variables $z_{k-1}, \dots, z_3$ can then be integrated directly using (4.3.6). By Lemma 4.2.5,

$$\int_{A \times A^c} \exp\left(-\frac{N}{2}|z_1 - z_2|^2\right) dz_1\, dz_2 = O(N^{-3/2}).$$

We conclude that for $j \ge 1$,

$$I_j = O(N^{-k+1/2}) \tag{4.3.21}$$

Inserting (4.3.20) and (4.3.21) into (4.3.19) completes the proof. □

4.3.3 Conclusion

Proof of Theorem 5.2.5. By Lemma 4.3.3, for every $n \in \mathbb{N}$ the cumulant $\kappa_n(X_A)$ is equal to the pseudo-cumulant $\tilde{\kappa}_n$ plus an exponentially small error term. Then by Lemma 4.3.2, Lemma 4.3.6, Lemma 4.3.7, and induction, we have for every $n \ge 3$ that

$$\lim_{N\to\infty} \kappa_2(N^{-1/4} X_A) = \frac{\ell(\partial A)}{2\pi^{3/2}}, \qquad \lim_{N\to\infty} \kappa_n(N^{-1/4} X_A) = 0.$$

[5]We learned of this integration method from [73], which derives a general formula for integrals of exponentials of complex quadratic forms.

We conclude that the limiting cumulants of $N^{-1/4}X_A$ are the same as the cumulants of a Gaussian random variable with variance $2^{-1}\pi^{-3/2}\ell(\partial A)$. Since the cumulants of a random variable determine its moments, the limiting moments also match those of this Gaussian. Because the Gaussian distribution is uniquely determined by its moments [31, Theorem 30.1], this implies the desired weak convergence [31, Theorem 30.2]. □

Remark 4.3.8. The Lipschitz hypothesis in Theorem 5.2.5 was used only to compute the variance in Lemma 4.3.2. To relax this hypothesis, one only needs to compute the integral (4.3.1) for more general domains (with rougher boundaries). This can be done for Caccioppoli sets using [130, Corollary 3.1.4] and for the Koch snowflake using [130, Theorem 3.3.2]. We note that the proof technique for the latter result is applicable to many other domains with self-similar boundaries.

4.4 Technical Estimates

The following estimate improves [37, Lemma 9.2] by establishing a quantitative error term. It is implicit in [122, Remark 3.4]; we provide a short proof here for completeness.

Lemma 4.4.1. *Let A be an admissible domain. Define $\tilde{d}_A = \inf\{|z-1| : z \in A\}$. Then there exists a constant $C(\tilde{d}_A) > 0$ such that for all $z \in A$,*

$$s_N(Nz) = 1 - \frac{1}{\sqrt{2\pi N}} \frac{(ze^{1-z})^N}{1-z}\Big(1 + R(z;N)\Big), \qquad \Big|R(z;N)\Big| \le CN^{-1}.$$

Proof. By repeated integration by parts, we have

$$s_N(Nz) = 1 - \frac{1}{(N-1)!}\int_0^{Nz} \zeta^{N-1}e^{-\zeta}d\zeta = 1 - \frac{N^N}{(N-1)!}\int_0^{z} \zeta^{N-1}e^{-N\zeta}d\zeta. \qquad (4.4.1)$$

The integral $\int_0^z$ can be taken along any curve connecting 0 and z since the integrand is analytic. Inserting Stirling's formula [134]

$$N! = \sqrt{2\pi N}\left(\frac{N}{e}\right)^N \tau_N, \qquad \tau_N = 1 + \frac{1}{12N} + O(N^{-2}) \tag{4.4.2}$$

into (4.4.1), we have

$$e^{-Nz} s_N(Nz) = 1 - \frac{1}{\tau_N}\sqrt{\frac{N}{2\pi}} \int_0^z \zeta^{N-1} e^{N(1-\zeta)} d\zeta. \tag{4.4.3}$$

Consider the map $\varphi\colon \zeta \mapsto \zeta e^{1-\zeta}$ on $\mathbb{C}$. We recall that the function $W(z)$ solving the equation $-e\varphi(-W(z)) = z$ is known as the Lambert W function. Standard facts about this function imply that there is a multi-valued inverse of φ with a (single-valued) principal branch defined on $\mathbb{C} \setminus [1, \infty)$ (see [62, Section 4]). It is given by $\psi(z) = -W(-z/e)$. Applying the change of variable $\zeta = \psi((1-t)ze^{1-z})$ to (4.4.3), we have

$$\begin{aligned} s_N(Nz) =& 1 - \frac{1}{\tau_N}\sqrt{\frac{N}{2\pi}} \int_0^z e^{1-\zeta} \varphi(\zeta)^{N-1} d\zeta \\ =& 1 - \frac{1}{\tau_N}\sqrt{\frac{N}{2\pi}} \left(ze^{1-z}\right)^N \int_0^1 \frac{(1-t)^{N-1}}{1-\psi((1-t)ze^{1-z})} dt. \end{aligned} \tag{4.4.4}$$

Define $f(z;t) = (1-\psi((1-t)ze^{1-z}))^{-1}$. Note that $f(z;\cdot)$ is infinitely differentiable in a neighborhood of $t = 1$ since ψ is analytic in a neighborhood of ze^{1-z} for $|z| < 1$. Direct differentiation shows that

$$f(z;t) = \frac{1}{1-z} + r(z;t),$$

where

$$r(z;t) = \int_0^t \frac{-\psi((1-\tau)ze^{1-z})}{(1-\tau)(1-\psi((1-\tau)ze^{1-z}))^3} d\tau.$$

By the continuity of $(\tau, z) \mapsto \psi((1-\tau)ze^{1-z})$ over $[0,1] \times A$, together with $\psi(ze^{1-z}) = z$, there exists $t_A > 0$ such that for all $0 \le \tau \le t_A$ and all $z \in A$, $|1 - \psi((1-\tau)ze^{1-z})| \le \frac{1}{2}\tilde{d}_A$. Therefore, there exists $c_A > 0$ depending on A (only through $\tilde{d}_A$) such that $|r(z;t)| \le c_A t$.

A standard application of Laplace's method (see [169, Section 19.2.4, Theorem 1(a)]) implies that there exists a constant $C > 0$ depending only on c_A, and consequently only on $\tilde{d}_A$, such that

$$\int_0^1 (1-t)^{N-1} f(z;t)\, dt = \frac{1}{N(1-z)}(1 + R(z;N)), \tag{4.4.5}$$

where $|R(z;N)| \leq CN^{-1}$. Combining (4.4.4) and (4.4.5) and recalling the definition of τ_N in (4.4.2) completes the proof. $\square$

Proof of Lemma 4.2.3. Using (4.3.15), we obtain

$$G(\sqrt{N}z, \sqrt{N}w) = \frac{e^{-N(\Im(z)^2+\Im(w)^2)}}{\sqrt{2N\pi}|\Im(z)\Im(w)|}\left(1 + O\left(\frac{1}{N\min(|z|,|w|)^4}\right)\right).$$

Combining this estimate with Lemma 4.4.1 and $|(2\pi)^{-1}(w-z)| \leq 1$, we get

$$\begin{aligned}\left|D_N(\sqrt{N}z, \sqrt{N}w)\right| \leq & e^{-(N/2)\Re(z-w)^2}\frac{e^{-N(\Im(z)^2+\Im(w)^2)}}{\sqrt{N}|\Im(z)\Im(w)|}\left(1 + O\left(\frac{1}{N\min(|z|,|w|)^4}\right)\right) \\ & \times \left(1 + \left|\frac{e^{-2(1-zw)}}{\sqrt{2\pi N}(1-zw)}e^{N(1-zw)}(zw)^N\left(1 + O(N^{-1})\right)\right|\right). \end{aligned} \tag{4.4.6}$$

We observe that

$$-\frac{N}{2}\Re(z-w)^2 - N(\Im(z)^2 + \Im(w)^2) \leq -N\Im(z)\Im(w). \tag{4.4.7}$$

We also note that

$$\begin{aligned} &|e^{-(N/2)\Re(z-w)^2}e^{-N(\Im(z)^2+\Im(w)^2)}e^{N(1-zw)}(zw)^N| \\ &\leq \exp\left(-\frac{N}{2}(|z|^2 - \ln|z|^2 - 1) - \frac{N}{2}(|w|^2 - \ln|w|^2 - 1)\right) \\ &\leq \exp\left(-\frac{N}{8}(|z|-1)^2 - \frac{N}{8}(|w|-1)^2\right). \end{aligned} \tag{4.4.8}$$

Inserting (4.4.7) and (4.4.8) into (4.4.6) completes the proof of the bound on D_N. The proof for I_N is similar, so we omit the details.

For S_N, we have

$$\begin{aligned}\left|S_N(\sqrt{N}z,\sqrt{N}w)\right| \leq & e^{-(N/2)\Re(z-\overline{w})^2}\frac{e^{-N(\Im(z)^2+\Im(w)^2)}}{\sqrt{N}|\Im(z)\Im(w)|}\left(1+O\left(\frac{1}{N\min(|z|,|w|)^4}\right)\right) \\ & \times\left(1+\left|\frac{e^{-2(1-z\overline{w})}}{\sqrt{2\pi N}\pi(1-z\overline{w})}e^{N(1-z\overline{w})}(z\overline{w})^N\Big(1+O(N^{-1})\Big)\right|\right)\end{aligned}$$

We note that

$$e^{-(N/2)\Re(z-\overline{w})^2}e^{-N(\Im(z)^2+\Im(w)^2)}=e^{-(N/2)|z-w|^2}\leq 1,$$

and

$$\begin{aligned}&|e^{-(N/2)\Re(z-\overline{w})^2}e^{-N(\Im(z)^2+\Im(w)^2)}e^{N(1-z\overline{w})}(z\overline{w})^N| \\ &\leq e^{(N/2)(-|z-\overline{w}|^2+2+2\Re(z\overline{w})+2\ln|z\overline{w}|)}=e^{-(N/2)(|z|^2+|w|^2-2-\ln|z|^2-\ln|w|^2)}\leq 1,\end{aligned}$$

where the last inequality follows from $|z|^2-1-\ln|z|^2\geq 0$ for $z\in A$. This completes the proof of the bound on S_N. $\square$

CHAPTER 5

Quenched Large Deviation Principles for Random Projections of ℓ_p^n balls

5.1 Introduction

5.1.1 Background and Motivation

High-dimensional measures are ubiquitous in mathematics, and are often profitably studied through their lower-dimensional projections. This approach has been successfully applied to problems in numerous fields, including statistics [68, 94, 132], asymptotic functional analysis [110] and convex geometry [13, 117]. In the case of the uniform measure on a high-dimensional convex body (a compact convex set with non-empty interior), low-dimensional projections are known to satisfy a central limit theorem. This theorem states that most k-dimensional projections of an n-dimensional isotropic convex body are approximately Gaussian in total variation norm, if n is sufficiently large and k is sufficiently small relative to n [13, 118]. The typical behavior of a low-dimensional projection is therefore uninformative about the high-dimensional convex body from which it originated. However, it was recently discovered that the tail behavior of random orthogonal projections retains interesting information about the original measure [96, 95].

This tail behavior is quantified through large deviation principles (LDPs), of which there are two main types: *quenched* LDPs, which provide almost-sure statements with respect to the projection or sequence of projections (which are independent of the high-dimensional measure), and *annealed* LDPs, which consider an average over the measure from which the projection directions or projection matrices are sampled. In this article, we focus on quenched LDPs for random projections of random vectors from a class of distributions that includes the uniform distribution on the unit ball in ℓ_p^n, one of the most fundamental examples of a convex body. Denoting this ball by $\mathbb{B}_p^n$, we have

$$\mathbb{B}_p^n := \left\{ \boldsymbol{x} \in \mathbb{R}^n : \sum_{i=1}^n |x_i|^p \leq n \right\}.$$

We consider projection directions chosen uniformly from $\mathbb{S}^{n-1}$, the unit sphere in $\mathbb{R}^n$ (in the one-dimensional case), or projection matrices chosen from the Haar measure on the Stiefel manifold $\mathbb{V}_{n,k}$ of orthonormal k-frames in $\mathbb{R}^n$ (in the multi-dimensional case).

We begin by reviewing previous work on LDPs for projections of $\mathbb{B}_p^n$, starting with the annealed case. Annealed LDPs are typically easier to analyze than quenched LDPs, since averaging over the randomness of the projection renders the entries of the projected vector exchangeable. For $\mathbb{B}_p^n$, annealed LDPs for one-dimensional projections were first established in [95, 96] for all $p \in [1, \infty]$. Later, annealed LDPs for the ℓ_2 norm of k_n-dimensional projections were established in [7] for all $p \in [1, \infty]$ when $\lim_{n\to\infty} k_n/n = \lambda \in [0, 1]$, with the additional requirement that $\lambda > 0$ when $p \leq 2$. The restriction that $\lambda > 0$ was removed in [115], which also proved a phase transition in the speed of the LDP with respect to the growth rate of k_n (see [115, Remark 3.6]). Further, the article [115] established LDPs for the empirical measures of the coordinates of the projections, in addition to the ℓ_2 norm. Additionally, [115] went beyond the balls $\mathbb{B}_p^n$ and established results for general high-dimensional measures satisfying an asymptotic thin shell condition, including uniform measures on Orlicz balls and certain Gibbs measures. A more refined version of this condition, and corresponding annealed sharp large deviation estimates for projection were also subsequently obtained in [128].

Compared to the annealed case, much less is known about quenched LDPs. Previous works have focused exclusively on k-dimensional projections for k independent of n. For one-dimensional projections of the unit cube $\mathbb{B}_\infty^n$ and more general product measures, quenched LDPs were established in [95]. For one-dimensional projections of $\mathbb{B}_p^n$, quenched LDPs were established for all $p \in [1, \infty)$ in [96]. In the case $p \in (1, \infty]$, the speed and rate function of the LDP are insensitive to the choice of projection matrix, except for a measure zero set of so-called *atypical* sequences of projection matrices. (In the case $p = 1$, the LDP is more subtle and depends on the sequence; see [96, Theorem 2.6]). Later, *sharp* LDPs that identify the precise prefactor were obtained for one-dimensional projections of $\mathbb{B}_p^n$ in [129]. Quenched LDPs for k-dimensional projections were established for $p \geq 2$ in [116], for any fixed positive integer k. Similar to the one-dimensional case, the speed and rate function are almost surely independent of the choice of sequence of projection matrices. Quenched LDPs for projections of various radially symmetric measures on $\mathbb{B}_p^n$ were also recently studied in [113]. We also remark that LDPs for the ℓ_q norm of a random element of $\mathbb{B}_p^n$ (with $q \neq p$) were established in [111].

In this work, we prove a quenched LDP for sequences of k_n-dimensional projections of ℓ_p^n balls when $p \in [2, \infty)$ and k_n grows sublinearly in n, that is, $\lim_{n\to\infty} k_n = \infty$ and $\lim_{n\to\infty} k_n/n = 0$ (see Theorem 5.1.2). As in previous work, the speed and rate function are almost surely insensitive to the sequence of projection matrices. As corollaries, we also obtain LDPs for the sequences of ℓ_2 norms and ℓ_∞ norms of the projections (see Corollaries 5.1.3 and 5.1.4). The extension to LDPs for the ℓ_q norms of the projections for all $q \in [1, \infty)$ is also possible using our methods; see Remark 5.5.1 below. Our results also generalize to projections of a larger class of random vectors, including those uniformly distributed on the ℓ_p^n sphere and a broad class of product measures with sub-Gaussian marginals (see Theorem 5.2.5). In addition to asymptotic convex geometry, LDP results of this kind are also relevant to problems arising in statistical mechanics. Indeed, establishing the LDP for ℓ_2-norms of projections is equivalent to identifying the scaled logarithmic asymptotics for expectations of exponential functionals of the sequence of ℓ_2 norms of the projections. The latter has close parallels with the study of the quenched log-partition function for statistical mechanical models with random disorder, such as the Hopfield model of neural networks [61], in which the projection matrix is replaced by an $(n \times k_n)$-matrix of i.i.d. entries and $Y^{(n)}$ is sampled from a product distribution with bounded support. Similar logarithmic asymptotics also appear in the study of the normalizing constant of the posterior distribution in high-dimensional linear regression [138, 143], where the distribution of $Y^{(n)}$ corresponds to the prior, which is taken to be a product distribution in [138] and uniform on $\mathbb{S}^{n-1}$ in [143].

5.1.2 Proof Techniques.

There are two primary technical challenges involved in establishing our results. First, any LDP needs to be defined relative to a sequence of measures on a single probability space, but each element of a sequence of k_n-dimensional projections has a different codomain. To remedy this problem, each projection must be embedded into a suitable parent space. In the annealed case, the empirical measure of the coordinates completely determines the distribution of the projected vector, due to the exchangeability of the coordinates. Previous work on annealed LDPs for projections of growing dimension identified the projections with the empirical measures of their coordinates, which are probability measures on $\mathbb{R}$, and established LDPs for the latter [115]. However, in the quenched case, the lack of exchangeability of the coordinates renders the empirical measure an unsuitable state variable. Instead, we take inspiration from the work of Comets and Dembo on large deviations for mean-field spin glass models [61], and embed our projections into a space consisting of infinite sequences whose coordinates appear in descending order (in absolute value), paired with their ℓ_2 norms. This approach, while leading to some technicalities, also captures a great deal of information about the sequence of projections, allowing us to also prove the additional LDPs for the ℓ_∞ and ℓ_2 norms mentioned previously. We remark that the previous works [115, 116] on multidimensional projections in the annealed setting did not address LDPs with the projection dimension k_n growing to infinity for the ℓ_∞ norm, since the latter is not a sufficiently nice function of the empirical coordinate measure.

However, the paper [61] considers the setting where $Y^{(n)}$ is a product measure with a bounded distribution and the projection matrix is replaced with a matrix with i.i.d. entries. In contrast, no similar boundedness hypothesis can be made in our work, which creates several additional technical complications. Further, the second main technical obstacle in our setting is that, in the course of our proof, we must show that the entries of the projection matrix sampled uniformly from the Haar measure on $\mathbb{V}_{n,k_n}$ can be well approximated by independent Gaussians for averages of a function of the rows. This claim is made precise in Lemma 5.3.6 below, and may be of independent interest, as it generalizes several existing approximation results [116, 108]. The result [116, Corollary 2.11] establishes convergence of the empirical measure of rows of the projection matrix to a Gaussian in the space of Borel probability measures on $\mathbb{R}^k$ (equipped with suitable Wasserstein topologies), only for fixed k. However, such a statement does not address the case when k_n grows in n, in which case such empirical measures would live on different spaces. Instead, in Lemma 5.3.6 we establish convergence of the empirical measures of scalar products of rows of the projection matrix with a k_n-dimensional deterministic vector towards a 1-dimensional Gaussian distribution in a suitable Wasserstein topology (see Remark 5.3.7) on the space of probability measures on $\mathbb{R}$, uniformly over any collection of up to e^{k_n} deterministic vectors. Prior results on Gaussian approximations for entries of Haar-distributed random matrices from the Stiefel manifold, such as those obtained in [109, 108], appear inadequate for our purposes, thus necessitating our alternative approach to such Gaussian approximations using Weingarten calculus; we expand on this point in Remark 5.3.8.

5.1.3 Definitions

For $k \in \mathbb{N}$, let I_k be the $k \times k$ identity matrix, and for $n \geq k$, set

$$\mathbb{V}_{n,k} := \left\{ A \in \mathbb{R}^{n \times k} : A^\top A = I_k \right\}.$$

The set $\mathbb{V}_{n,k}$ is called the Stiefel manifold of k-frames in $\mathbb{R}^n$ and consists of k-dimensional orthonormal bases in $\mathbb{R}^n$. We also use $O_n := \mathbb{V}_{n,n}$ to denote the set of $n \times n$ orthogonal matrices.

Let ℓ^2 denote the set of infinite sequences $\boldsymbol{x} = (x_1, x_2, \dots) \in \mathbb{R}^\infty$ of real numbers such that the norm

$$\|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^{\infty} x_i^2 \right)^{1/2} \tag{5.1.1}$$

is finite. Given $\boldsymbol{x}, \boldsymbol{y} \in \ell^2$, we define the inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{i=1}^{\infty} x_i y_i$. We say that $\boldsymbol{x} \in \ell^2$ is ordered if $|x_k| \geq |x_{k+1}|$, and $x_k \geq x_{k+1}$ if $|x_k| = |x_{k+1}|$, for all $k \in \mathbb{N}$. Given a sequence $\boldsymbol{x} \in \ell^2$ with a finite number of nonzero entries, let

$$\mathfrak{n}(\boldsymbol{x}) := \Big| \{ i \in \mathbb{N} : x_i \neq 0 \} \Big|$$

denote the number of such entries. Let $[\boldsymbol{x}] \in \ell^2$ denote the sequence whose first $\mathfrak{n}(\boldsymbol{x})$ coordinates $[x]_1, \dots, [x]_{\mathfrak{n}(\boldsymbol{x})}$ are equal to the nonzero entries of $\boldsymbol{x}$ arranged so that $[\boldsymbol{x}]$ is ordered; this implies that $[x]_i = 0$ for $i \geq \mathfrak{n}(\boldsymbol{x}) + 1$. Given $k \in \mathbb{N}$ and $\boldsymbol{x} \in \mathbb{R}^k$, we define $[\boldsymbol{x}]$ by first defining $\tilde{\boldsymbol{x}} \in \ell^2$ as follows: $\tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i$ for $i \leq k$ and $\boldsymbol{x}_i = 0$ for $i > k$, then setting $[\boldsymbol{x}] := [\tilde{\boldsymbol{x}}]$. We define $\mathcal{X}$ as the set of pairs

$$\mathcal{X} := \{ (\boldsymbol{x}, r) : \boldsymbol{x} \in \ell^2, \, r \in [0, \infty), \, \boldsymbol{x} \text{ is ordered}, \, \|\boldsymbol{x}\|_2 \leq r \}. \tag{5.1.2}$$

We equip $\mathcal{X}$ with the distance

$$d\Big((\boldsymbol{x}, r), (\boldsymbol{y}, s) \Big) := \|\boldsymbol{x} - \boldsymbol{y}\|_\infty + |r - s|. \tag{5.1.3}$$

For $\boldsymbol{x} \in \mathbb{R}^k$, let

$$\Pi(\boldsymbol{x}) := ([\boldsymbol{x}], \|\boldsymbol{x}\|_2) \in \mathcal{X} \tag{5.1.4}$$

denote the corresponding representative in $\mathcal{X}$.

The function $\log(x)$ denotes the natural logarithm. We let $\mathbb{R}_+$ denote the non-negative reals. We define the unit ℓ_n^p ball as the set

$$\mathbb{B}_p^n := \left\{ \boldsymbol{x} \in \mathbb{R}^n : \sum_{i=1}^n |x_i|^p \leq n \right\}$$

and the unit ℓ_n^p sphere by

$$\mathbb{S}_p^n := \left\{ \boldsymbol{x} \in \mathbb{R}^n : \sum_{i=1}^n |x_i|^p = n \right\}.$$

Given a sequence $\{k_n\}_{n\in\mathbb{N}}$ of positive integers, let σ_n be the Haar measure on $\mathbb{V}_{n,k_n}$, and let σ be a probability measure on $\mathbb{V} = \otimes_n \mathbb{V}_{n,k_n}$ whose n-th marginal is σ_n for all $n \in \mathbb{N}$. All statements about probabilities on $\mathbb{V}_{n,k_n}$ and $\mathbb{V}$ in this work are made with respect to the measures σ_n and σ. For brevity, we do not explicitly denote the dependence of $\mathbb{V}$, σ_n, and σ on the sequence $\{k_n\}$. We say that the sequence $\{k_n\}_{n\in\mathbb{N}}$ is increasing if $k_{n+1} \geq k_n$ for all $n \in \mathbb{N}$.

Finally, we recall the definition of an LDP.

Definition 5.1.1 ([65, Section 1.2])**.** *Let $\mathcal{T}$ be a topological space with Borel σ-algebra $\mathcal{B}$. A sequence $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ of probability measures on $\mathcal{T}$ satisfies a large deviation principle (LDP) with speed $s_n : \mathbb{N} \to \mathbb{R}_+$ and rate function $I \colon \mathcal{T} \to [0, \infty]$ if for all $B \in \mathcal{B}$,*

$$-\inf_{x\in B^\circ} I(x) \leq \liminf_{n\to\infty} \frac{1}{s_n} \log \mathbb{P}_n(B) \leq \limsup_{n\to\infty} \frac{1}{s_n} \log \mathbb{P}_n(B) \leq -\inf_{x\in\overline{B}} I(x),$$

where B° and $\overline{B}$ denote the interior and closure of B, respectively. We say that $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ satisfies a weak LDP when these inequalities hold for all $B \in \mathcal{B}$ such that $\overline{B}$ is compact.

A sequence of $\mathcal{T}$-valued random variables $\{Z^{(n)}\}_{n\in\mathbb{N}}$ is said to satisfy an LDP with speed s_n and rate function I if and only if the corresponding sequence of image measures $\{\mathbb{P}(Z^{(n)} \in \cdot)\}_{n\in\mathbb{N}}$ satisfies an LDP with that speed and rate function. When the speed is not mentioned explicitly, we use the default $s_n = n$. The function I is said to be a good rate function if it has compact level sets $\Psi_I(\alpha) = \{x \in \mathcal{T} : I(x) \leq \alpha\}$ for all $\alpha \in [0, \infty)$.

5.1.4 Main Result: A Specific Setting

We begin by introducing some general notation, which will be used throughout the paper. Let Z be a mean zero random variable and let $\eta\colon \mathbb{R} \to \mathbb{R}_+$ be a continuous function. We define $\overline{\Lambda}\colon \mathbb{R}^2 \to \mathbb{R} \cup \{\infty\}$ to be the log-moment generating function of $(Z, \eta(Z))$:

$$\overline{\Lambda}(t_1, t_2) = \overline{\Lambda}_{Z,\eta}(t_1, t_2) \coloneqq \log \mathbb{E}\left[\exp\left(t_1 Z + t_2 \eta(Z)\right)\right]. \tag{5.1.5}$$

Further, let $\{g_i\}_{i=0}^{\infty}$ be an infinite sequence of independent Gaussian random variables with mean zero and variance one, denote $\boldsymbol{g} = (g_1, g_2, \dots)$, and define $\Lambda : \ell^2 \times \mathbb{R}^2 \to \mathbb{R} \cup \{\infty\}$ as[1]

$$\Lambda(\boldsymbol{u}, b, c) \coloneqq \mathbb{E}\Big[\overline{\Lambda}\Big(\langle \boldsymbol{u}, \boldsymbol{g}\rangle + b g_0, c\Big)\Big]. \tag{5.1.6}$$

The Legendre transform Λ^* of Λ is defined by

$$\Lambda^*(\boldsymbol{w}, r, s) \coloneqq \sup_{(\boldsymbol{u},b,c)\in \ell^2\times\mathbb{R}\times\mathbb{R}} \left\{\langle \boldsymbol{u}, \boldsymbol{w}\rangle + br + cs - \Lambda(\boldsymbol{u}, b, c)\right\}. \tag{5.1.7}$$

We also set

$$I(\boldsymbol{w}, r, s) \coloneqq \Lambda^*\left(\boldsymbol{w}, \sqrt{r - \|\boldsymbol{w}\|_2^2}, s\right) \tag{5.1.8}$$

for $(\boldsymbol{w}, r, s) \in \mathcal{X} \times \mathbb{R}_+$. Let $\rho\colon \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function. For $(\boldsymbol{w}, r) \in \mathcal{X}$, we

[1] The function Λ represents a suitable average of the log-moment generating function $\overline{\Lambda}$ over the random "environment" (i.e. projection matrix), with the Gaussian vector $(g_0, \boldsymbol{g})$ arising as an approximation to the typical row of a Haar-distributed element of $\mathbb{V}_{n,k}$ in the sublinear regime. See Lemma 5.3.6.

define

$$\mathcal{I}(\boldsymbol{w}, r) = \inf_{s>0} I\left(\frac{\boldsymbol{w}}{\rho(s)}, \frac{r}{\rho(s)}, s\right). \tag{5.1.9}$$

We now specialize the previous definitions for the purpose of stating our main results, concerning $\mathbb{B}_p^n$. We fix $p \in [1,\infty)$, and let $f_p : \mathbb{R} \to \mathbb{R}$ be the probability density function of the p-generalized normal distribution:

$$f_p(x) \coloneqq \frac{1}{2p^{1/p}\Gamma(1+1/p)} \exp\left(-\frac{|x|^p}{p}\right).$$

Let Z_p be a random variable with density f_p, let $\eta_2(x) = x^2$, and define $\overline{\Lambda}_p \coloneqq \overline{\Lambda}_{Z_p,\eta_2}$, as in (5.1.5). We let Λ_p denote the functional in (5.1.6) when $\overline{\Lambda}$ is replaced by $\overline{\Lambda}_p$. Also, the let definitions (5.1.7), (5.1.8), and (5.1.9) hold when Λ^*, I, and $\mathcal{I}$ are replaced with Λ_p^*, I_p, and $\mathcal{I}_p$, respectively, and $\rho(x) = x^{1/p}$. We also set

$$\hat{\mathcal{I}}_p(\boldsymbol{w}, r) \coloneqq \inf_{s>0} I_p\left(\boldsymbol{w}, r, s\right).$$

The following theorem is our primary result. It establishes a quenched LDP for multidimensional projections of ℓ_p^n balls when the projection dimension grows sublinearly. It is proved in Section 5.5. We emphasize that in this theorem and its corollaries, $\boldsymbol{a} \in \mathbb{V}$ always denotes a fixed, deterministic sequence of projection matrices.

Theorem 5.1.2. *Fix $p \in [2,\infty)$, and let $\{k_n\}_{n\in\mathbb{N}}$ be an increasing sequence of positive integers such that $\lim_{n\to\infty} k_n = \infty$ and $\lim_{n\to\infty} k_n/n = 0$. For each $n \in \mathbb{N}$, let $Y^{(n)}$ be uniformly distributed on $\mathbb{B}_p^n$. Then for σ-a.e. $\boldsymbol{a} = \{\boldsymbol{a}^{(n)}\}_{n\in\mathbb{N}} \in \mathbb{V}$, the sequence*

$$\left\{\Pi\left(n^{-1/2}(\boldsymbol{a}^{(n)})^\top Y^{(n)}\right)\right\}_{n\in\mathbb{N}},$$

with Π defined as in (5.1.4), satisfies an LDP in $\mathcal{X}$ with speed n and good rate function $\mathcal{I}_p$.

From Theorem 5.1.2, we can deduce LDPs for the sequence of Euclidean norms and maximal coordinates of the random projection $n^{-1/2}(\boldsymbol{a}^{(n)})^\top Y^{(n)}$. We state these in the following corollaries, which are also proved in Section 5.5.

Corollary 5.1.3. *Let p, $\{k_n\}_{n\in\mathbb{N}}$, and $Y^{(n)}$ be as in Theorem 5.1.2. Then for σ-a.e. $\boldsymbol{a} = \left\{\boldsymbol{a}^{(n)}\right\}_{n\in\mathbb{N}} \in \mathbb{V}$, the sequence*

$$\left\{\left\|n^{-1/2}\left(\boldsymbol{a}^{(n)}\right)^\top Y^{(n)}\right\|_2\right\}_{n\in\mathbb{N}}$$

satisfies an LDP $\mathbb{R}_+$ with speed n and good rate function $\mathbb{I}_p$ given by

$$\mathbb{I}_p\left(r\right) := \sup_{t_1,t_2\in\mathbb{R}}\left\{t_1 r + t_2 - \mathbb{E}\left[\overline{\Lambda}_p\left(t_1 g_0, t_2\right)\right]\right\},$$

where $\overline{\Lambda}_p$ is defined in (5.1.5). Furthermore, $\mathbb{I}_p$ is convex.

Corollary 5.1.4. *Let p, $\{k_n\}_{n\in\mathbb{N}}$, and $Y^{(n)}$ be as in Theorem 5.1.2. Then for σ-a.e. $\boldsymbol{a} = \{\boldsymbol{a}^{(n)}\}_{n\in\mathbb{N}} \in \mathbb{V}$, the sequence*

$$\left\{\left\|n^{-1/2}(\boldsymbol{a}^{(n)})^\top Y^{(n)}\right\|_\infty\right\}_{n\in\mathbb{N}} \tag{5.1.10}$$

satisfies an LDP in $\mathbb{R}$ with speed n and good rate function

$$\mathcal{I}_{\max}(r) := \mathcal{I}_p((r,0,0\ldots),r).$$

Remark 5.1.5. The conclusions of Theorem 5.1.2, Corollary 5.1.3, and Corollary 5.1.4 also hold when $Y^{(n)}$ is uniformly distributed on $\mathbb{S}_p^n$. The proofs of these results also contain, as intermediate steps, proofs of the analogous claims for $\mathbb{S}_p^n$.

Outline

Theorem 5.1.2 is a consequence of a more general result, which provides LDPs for the projections of a large class of sequences of random vectors. This general result in stated in Section 5.2 below as Theorem 5.2.5, along with the necessary assumptions and associated corollaries. The proof of Theorem 5.2.5 follows on combining an upper bound and a lower bound, given in the next section as Proposition 5.2.3 and Proposition 5.2.4, respectively. In Section 5.3, the upper bound Proposition 5.2.3 is proved assuming certain preliminary lemmas compiled in Section 5.3.1 and proved in Section 5.6. The complementary lower bound is established in Section 5.4, building on preparatory results stated in Section 5.4.1, whose proofs are deferred to Section 5.7. Section 5.8 contains auxiliary computations for p-Gaussian random variables. Section 5.9 recalls the Weingarten calculus and proves an auxiliary lemma.

5.1.5 Open Problems.

This work gives rise to several open problems.

1. *The case $p \in (1,2)$.* It would be of interest to study the quenched LDP for the projection or its norm when $Y^{(n)}$ is uniformly distributed on $\mathbb{B}_p^n$ with $p \in (1,2)$. The sub-Gaussianity assumption $p \geq 2$ is used in several places in the proof, including in the proof of the exponential tightness result in Lemma 5.3.4 and the concentration result in Lemma 5.7.3. We believe that at least a weak (quenched) LDP at speed n will hold for sufficiently slowly growing k_n. But we expect a phase transition in the growth rate of k_n, wherein the speed of the LDP would change in a p-dependent way.

2. *The linear setting $k_n = \lambda n$ for $\lambda \in (0,1)$.* It would be of interest to study the case when k_n grows linearly in n, whis would likely require a different approach since the condition $k_n = o(n)$ is currently being used in several places in the proof, including the Gaussian approximation lemma, as well as to show that the tail of the ordered

projection vector has negligible ℓ_2-norm, and associated concentration results.

3. *A broad class of high-dimensional vectors.* It would be desirable to identify a broad sequence of random vectors $\{Y^{(n)}\}_{n\in\mathbb{N}}$ for which the corresponding (quenched) random projections satisfy an LDP. For example, does there exist a sufficient condition for LDPs of quenched projections, analogous to the asymptotic thin shell condition in the annealed setting?

The fact that all the above questions are fully understood in the annealed setting [96, 7, 115] further points to the additional subtleties present in the quenched setting when compared to the annealed one.

Acknowledgments

P.L. is supported by NSF postdoctoral fellowship DMS-220289. K.R. is supported by NSF grant DMS-1954351 and Vannevar Bush Faculty Fellowship N0014-21-1-2887. X.X. is supported by NSF grant DMS-1954351. The second author would like to thank Amir Dembo for bringing to her attention the reference [61], which suggested the convenient topological space in which the LDP in this paper is established.

5.2 Main Technical Results

In this section, we list the general versions of our main technical results and the assumptions they require. We first recall the following concept from convex analysis.

Definition 5.2.1 (Essential Smoothness)**.** *Let $f : \mathbb{R}^d \to (-\infty, \infty]$ be a convex function and let $\mathcal{D}_f \coloneqq \{x \in \mathbb{R}^d : f(x) < \infty\}$. Then f is said to be essentially smooth if*

1. *$\mathcal{D}_f^\circ$ is non-empty.*

2. *$f(\cdot)$ is differentiable throughout $\mathcal{D}_f^\circ$.*

3. *$f(\cdot)$ satisfies $\lim_{n\to\infty} |\nabla f(\lambda_n)| = \infty$ whenever $\{\lambda_n\}_{n\in\mathbb{N}}$ is a sequence in $\mathcal{D}_f^\circ$ converg-*

ing to a boundary point of $\mathcal{D}_f^\circ$.

Given a sequence $\{Y^{(n)}\}_{n\in\mathbb{N}}$ of random vectors $Y^{(n)} = (Y_1, Y_2, \dots, Y_n) \in \mathbb{R}^n$, we consider the following assumptions.

Assumption 2. There exists a sequence of non-constant and independent, identically distributed (i.i.d.) real-valued random variables $\{X_j\}_{j\in\mathbb{N}}$, a non-constant continuous function $\eta\colon \mathbb{R} \to \mathbb{R}_+$, and a continuous function $\rho\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that we have the distributional equality

$$Y_j^{(n)} \overset{d}{=} X_j \cdot \rho\left(\frac{1}{n}\sum_{i=1}^{n}\eta(X_i)\right) \tag{5.2.1}$$

for all $n \in \mathbb{N}$ and $j \in \{1, \dots, n\}$. We further suppose that $\rho(x) > 0$ for all $x > 0$.

Assumption 3. The variable X_1 has mean zero and there exists a constant $C_2 > 0$ such that for all $s \geq 0$

$$\mathbb{P}\left(|X_1| \geq s\right) \leq 2\exp(-s^2/C_2^2).$$

Assumption 4. Set $\overline{\Lambda}(t_1, t_2) = \overline{\Lambda}_{X_1,\eta}$, as defined in (5.1.5) and using the choice of η from Assumption 2. There exists some $0 < T \leq \infty$ such that $\overline{\Lambda}(t_1, t_2)$ is finite for all $(t_1, t_2) \in \mathbb{R} \times (-\infty, T)$ and $\overline{\Lambda}$ is essentially smooth. Further, the derivatives $\partial_1^\alpha \partial_2^\beta \overline{\Lambda}(t_1, t_2)$ exist for all $(t_1, t_2) \in \mathbb{R} \times (-\infty, T)$ and all integers $\alpha, \beta \geq 0$ such that $\alpha + \beta \leq 2$.

Assumption 5. There exists a continuous function $\tilde{C}\colon (-\infty, T) \to \mathbb{R}_+$ such that

$$\left|\partial_1^\alpha \partial_2^\beta \overline{\Lambda}\,(t_1, t_2)\right| \leq \tilde{C}(t_2)\left(1 + |t_1|^{2-\alpha}\right), \tag{5.2.2}$$

for all $(t_1, t_2) \in \mathbb{R} \times (-\infty, T)$ and all integers $\alpha, \beta \geq 0$ such that $\alpha + \beta \leq 2$.

Remark 5.2.2. In addition to the uniform measure on the ℓ^p balls considered in Theorem 5.1.2, the framework above allows us to deal with product measures whose marginals are symmetric and sub-Gaussian. Indeed, by taking $\rho(x) = 1$ for all $x \in \mathbb{R}_+$, we see that these assumptions are satisfied when $Y^{(n)} = (Y_1, \dots, Y_n)$ is constructed from a sequence of i.i.d. random variables $\{Y_j\}_{j\in\mathbb{N}}$ satisfying Assumptions 2, 3, and 4.

Additionally, we often consider increasing sequences $\{k_n\}_{n\in\mathbb{N}}$ of positive integers such that

$$\lim_{n\to\infty} k_n = \infty \qquad \text{and} \qquad \lim_{n\to\infty} \frac{k_n}{n} = 0. \tag{5.2.3}$$

The following large deviation upper bound is proved in Section 5.3.2.

Proposition 5.2.3. *Let $\{k_n\}_{n\in\mathbb{N}}$ be an increasing sequence of positive integers satisfying (5.2.3). Suppose that $\{X_j\}_{j\in\mathbb{N}}$ is a sequence of i.i.d. random variables satisfying Assumptions 2, 3, and 4. Let $X^{(n)} = (X_1,\dots,X_n) \in \mathbb{R}^n$ for all $n \geq 1$. Then for σ-a.e. $\boldsymbol{a} = \{\boldsymbol{a}^{(n)}\}_{n\in\mathbb{N}} \in \mathbb{V}$, we have*

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{P}\left(\left(\Pi(n^{-1/2}\left(\boldsymbol{a}^{(n)}\right)^{\top} X^{(n)}), \frac{1}{n}\sum_{i=1}^{n}\eta(X_i)\right) \in \mathcal{S}\right) \leq -\inf_{(\boldsymbol{w},r,s)\in\mathcal{S}} I(\boldsymbol{w},r,s) \tag{5.2.4}$$

for all closed sets $\mathcal{S} \subset \mathcal{X} \times \mathbb{R}_+$.

The following large deviations lower bound is proved in Section 5.4.2.

Proposition 5.2.4. *Let $\{k_n\}_{n\in\mathbb{N}}$ be an increasing sequence of positive integers satisfying (5.2.3). Suppose that $\{X_j\}_{j\in\mathbb{N}}$ is a sequence of i.i.d. random variables satisfying Assumptions 2, 3, and 4. Let $X^{(n)} = (X_1,\dots,X_n) \in \mathbb{R}^n$ for all $n \geq 1$. Then for σ-a.e. $\boldsymbol{a} = \{\boldsymbol{a}^{(n)}\}_{n\in\mathbb{N}} \in \mathbb{V}$, we have*

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{P}\left(\left(\Pi(n^{-1/2}\left(\boldsymbol{a}^{(n)}\right)^{\top} X^{(n)}), \frac{1}{n}\sum_{i=1}^{n}\eta(X_i)\right) \in \mathcal{O}\right) \geq -\inf_{(\boldsymbol{w},r,s)\in\mathcal{O}} I(\boldsymbol{w},r,s)$$

for all open sets $\mathcal{O} \subset \mathcal{X} \times \mathbb{R}_+$.

The following theorem is proved in Section 5.5 using the previous two lemmas, where it is then used to prove the main results stated in the previous section.

Theorem 5.2.5. *Let $\{k_n\}_{n\in\mathbb{N}}$ be an increasing sequence of positive integers satifying (5.2.3). Let $\{Y^{(n)}\}_{n\in\mathbb{N}}$ be a sequence of random vectors $Y^{(n)} = (Y_1, Y_2, \dots, Y_n)$ satisfying*

Assumptions 1–4. Then for σ-a.e. $\boldsymbol{a} = \{\boldsymbol{a}^{(n)}\}_{n\in\mathbb{N}} \in \mathbb{V}$, the sequence

$$\left\{\Pi\left(n^{-1/2}(\boldsymbol{a}^{(n)})^{\top}Y^{(n)}\right)\right\}_{n\in\mathbb{N}}$$

satisfies an LDP in $\mathcal{X}$ with good rate function $\mathcal{I}$.

5.3 Upper Bound

5.3.1 Preliminary Lemmas

In this section, we state some results required for the proof of Proposition 5.2.4. The first lemma, which collects several topological properties of $\mathcal{X}$, is proved in Section 5.6.1.

Lemma 5.3.1. *1. The topology on $\mathcal{X}$ is equivalent to the product topology on $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}_+$, where $\mathbb{R}^{\mathbb{N}}$ is itself equipped with the product topology.*

2. The topology on $\mathcal{X}$ is equivalent to the product topology on $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}_+$, where $\mathbb{R}^{\mathbb{N}}$ is itself equipped with the weak-ℓ^2 topology.

3. For any fixed $A < \infty$, the set $\{(\boldsymbol{w}, r) \in \mathcal{X} : r \leq A\}$ is compact.

Our second lemma shows that the rate function in Proposition 5.2.3 is a good rate function (as defined in Definition 5.1.1). Its proof is deferred to Section 5.6.2.

Lemma 5.3.2. *The function I defined in (5.1.8) is a good rate function.*

To state the following lemmas, we recall the definition of exponential tightness.

Definition 5.3.3. *A family of measures $\{\mu_n\}_{n\in\mathbb{N}}$ on a topological space space $\mathcal{T}$ is exponentially tight with speed $s_n : \mathbb{N} \to \mathbb{R}_+$ if for every $\alpha < \infty$, there exists a compact set $K_\alpha \subset \mathcal{T}$ such that*

$$\limsup_{s_n\to\infty} \frac{1}{n} \log \mu_\varepsilon(K_\alpha^c) \leq -\alpha.$$

We say that a sequences of random variables $\{X_n\}_{n\in\mathbb{N}}$ is exponentially tight with speed s_n if the sequence of measures $\{\mu_n\}_{n\in\mathbb{N}}$ defined by $\mu_n(A) = \mathbb{P}(X_n \in A)$ also has this property.

We default to $s_n = n$ when no speed is explicitly stated.

The next two lemmas address exponential tightness of the sequence of the norms of the projections $n^{-1/2}(\boldsymbol{a}^{(n)})^\top X^{(n)}$ and the sums $\frac{1}{n}\sum_{i=1}^n \eta(X_i)$. They are proved in Section 5.6.3.

Lemma 5.3.4. *Let $\{X_j\}_{j\in\mathbb{N}}$ be random variables satisfying Assumption 3, and let $X^{(n)} = (X_1, \ldots, X_n)$. Let $\{k_n\}_{n\in\mathbb{N}}$ be an increasing sequence of positive integers satisfying (5.2.3). Then there exists a constant $\gamma > 0$, depending only on the sequence $\{k_n\}_{n\in\mathbb{N}}$ and the constant C_2 from Assumption 3, such that for every $n \in \mathbb{N}$, $\boldsymbol{a}^{(n)} \in \mathbb{V}_{n,k_n}$, and $t \geq C_2^2$,*

$$\frac{1}{n}\log \mathbb{P}\left(\|n^{-1/2}(\boldsymbol{a}^{(n)})^\top X^{(n)}\|_2^2 \geq t+1\right) \leq -\gamma t.$$

Lemma 5.3.5. *Let $\{X_j\}_{j=1}^\infty$ be i.i.d. random variables satisfying Assumptions 2 and 3. Then the sequence of random variables $\{Z_n\}_{n\in\mathbb{N}}$ given by $Z_n := \frac{1}{n}\sum_{i=1}^n \eta(X_i)$ is exponentially tight.*

For $\boldsymbol{u}^{(n)} \in \mathbb{R}^{k_n}$, $\boldsymbol{a}^{(n)} \in \mathbb{V}_{n,k_n}$, and $c \in \mathbb{R}$, we let $\boldsymbol{a}_i^{(n)}$ denote the i-th row of $\boldsymbol{a}^{(n)}$, and define

$$F_n(\boldsymbol{u}^{(n)}, \boldsymbol{a}^{(n)}, c) := \frac{1}{n}\sum_{i=1}^n \overline{\Lambda}\left(\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\rangle, c\right). \tag{5.3.1}$$

Recall that g denotes a Gaussian random variable with zero mean and unit variance. The next lemma shows that $F_n(\boldsymbol{u}^{(n)}, \boldsymbol{a}^{(n)}, c)$ is closely approximated by $\mathbb{E}\left[\bar{\Lambda}\left(\left\|\boldsymbol{u}^{(n)}\right\|_2 g, c\right)\right]$; this holds uniformly for any subexponential collection of vectors in $\mathbb{R}^{k_n}$. It is proved in Section 5.6.5.

Lemma 5.3.6. *Fix a constant $D \in (0,\infty)$, a deterministic sequence $\{d_n\}_{n\in\mathbb{N}}$ of positive integers such that $\limsup_{n\to\infty} n^{-1}\log(d_n) = 0$, and i.i.d. random variables $\{X_j\}_{j=1}^\infty$ satisfying Assumptions 2–4. For every $n \in \mathbb{N}$, let*

$$\mathcal{W}_n = \{\boldsymbol{v}^{(n,1)}, \boldsymbol{v}^{(n,2)}, \ldots, \boldsymbol{v}^{(n,d_n)}\} \subset \mathbb{R}^{k_n}$$

be a finite collection of vectors such that $\|\boldsymbol{v}^{(n,j)}\|_2 = D$ *for all* $n \in \mathbb{N}$ *and* $j \in \{1, \ldots, d_n\}$. *Fix* $c \in (-\infty, T)$, *where* T *is the constant from Assumption 4. Then for* σ*-a.e.* $\boldsymbol{a} = \{\boldsymbol{a}^{(n)}\}_{n\in\mathbb{N}} \in \mathbb{V}$,

$$\lim_{n\to\infty} \sup_{1\le j\le d_n} \left| F_n(\boldsymbol{v}^{(n,j)}, \boldsymbol{a}^{(n)}, c) - \mathbb{E}\left[\bar{\Lambda}\left(Dg, c\right)\right] \right| = 0,$$

where g *is a Gaussian random variable with zero mean and unit variance.*

Remark 5.3.7. From the proof of Lemma 5.3.6, it is clear that the result continues to hold if in the definition of F_n, $\bar{\Lambda}(\cdot, c)$ is replaced by any differentiable function $H\colon \mathbb{R} \to \mathbb{R}$ that satisfies, for some constant $C \in \mathbb{R}_+$, $|H(t)|^2 \le C(1+t^2)$ and $|H'(t)|^2 \le C(1+|t|)$.

Remark 5.3.8. As mentioned in the introduction, Jiang has proved a Gaussian approximation result for Haar-distributed elements of the Stiefel manifold in [108, Theorem 5] by coupling elements of the Haar-distributed matrix to matrices of independent Gaussians through Gram-Schmidt orthogonalization. However, as noted in [109, Theorem 3], this coupling is effective only when $k_n = o(\log n/n)$. Further, in order for this coupling to be strong enough to prove the almost-sure convergence in Lemma 5.3.6 (rather than the convergence in probability considered in [109]), it appears that additional restrictions on k_n and d_n are required. Hence, [108, Theorem 5] is not sufficient to study the entire sublinear regime, which motivates the alternative approach to Gaussian approximation via the Weingarten calculus taken in our proof of Lemma 5.3.6 in Section 5.6.5.

For $\boldsymbol{u} = (u_1, u_2, \ldots) \in \ell^2$ and a given $m \in \mathbb{N}$, we let $\boldsymbol{u}_\le = \boldsymbol{u}_{\le m} \in \mathbb{R}^m$ be the vector containing the first m coordinates of $\boldsymbol{u}$,

$$\boldsymbol{u}_\le := (u_1, u_2, \ldots, u_m, 0, 0, \ldots)^\top, \tag{5.3.2}$$

and let $\boldsymbol{u}_> = \boldsymbol{u}_{>m} \in \ell^2$ denote the vector defined by $\boldsymbol{u}_> = \boldsymbol{u} - \boldsymbol{u}_\le$, so that

$$\boldsymbol{u}_> := (0, \ldots, 0, u_{m+1}, u_{m+2}, \ldots)^\top.$$

For $\hat{\boldsymbol{u}} \in \mathbb{R}^m$ and $b, c \in \mathbb{R}$, and Λ defined as in (5.1.6), set

$$\Lambda_m(\hat{\boldsymbol{u}}, b, c) := \Lambda((\hat{\boldsymbol{u}}^\top, 0, 0, \ldots), b, c). \tag{5.3.3}$$

For all $\boldsymbol{u} \in \ell^2$ and $b, c \in \mathbb{R}$, we observe by Jensen's inequality that[2]

$$\Lambda(\boldsymbol{u}, b, c) \geq \Lambda_m(\boldsymbol{u}_{\leq}, b, c).$$

Additionally, using the dominated convergence theorem and (5.2.2), we have the monotonic limit

$$\lim_{m\to\infty} \Lambda_m(\boldsymbol{u}_{\leq m}, b, c) = \Lambda(\boldsymbol{u}, b, c). \tag{5.3.4}$$

5.3.2 Proof of the Upper Bound

We first reduce the proof of the upper bound to a verification of the upper bound for a simpler subclass of closed sets.

Lemma 5.3.9. *Retain the notation and hypotheses of Proposition 5.2.3. Suppose that for every closed set $\mathcal{S} \subset \mathcal{X} \times \mathbb{R}_+$ and constant $A \in (0, \infty)$, the upper bound (5.2.4) holds for the set $\mathcal{S} \cap \mathcal{K}_A$, where*

$$\mathcal{K}_A = \{(\boldsymbol{w}, r, s) : 0 \leq r \leq A, 0 \leq s \leq A\}, \tag{5.3.5}$$

for a set $\Omega_{\mathcal{S},A}$ of sequences $\boldsymbol{a} = \left(\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}, \ldots\right) \in \mathbb{V}$ such that $\sigma(\Omega_{\mathcal{S},A}) = 1$. Then for almost every sequence $\boldsymbol{a} \in \mathbb{V}$, the upper bound (5.2.4) holds for all closed sets $\mathcal{S}$.

Proof. We assume the hypotheses of the lemma, and for any $A \in (0, \infty)$, we let $\mathcal{K}_A \subset \mathcal{X} \times \mathbb{R}_+$ denote the set defined in (5.3.5).

Step 1. We first fix an arbitrary closed set $\mathcal{S} \subset \mathcal{X} \times \mathbb{R}_+$ and establish the claim in

[2]Note that for any $t_2 \geq 0$, the function $t_1 \mapsto \overline{\Lambda}(t_1, t_2)$ is convex by Hölder inequality.

(5.2.4) for $\mathcal{S}$; namely,

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\left(\Pi(n^{-1/2}(\boldsymbol{a}^{(n)})^\top X^{(n)}),\frac{1}{n}\sum_{i=1}^n\eta(X_i)\right)\in\mathcal{S}\right)\le-\inf_{(\boldsymbol{w},r,s)\in\mathcal{S}}I(\boldsymbol{w},r,s)$$

holds for σ-almost every sequence $\boldsymbol{a}\in\mathbb{V}$.

For any $A\in(0,\infty)$ let $\Omega_{\mathcal{S},A}\subset\mathbb{V}$ be as in the lemma statement, and set $\Omega_{\mathcal{S}}=\cap_{m=1}^\infty\Omega_{\mathcal{S},m}$. Then clearly $\sigma(\Omega_{\mathcal{S}})=1$. We next observe that we have $I(\boldsymbol{w},r,s)\ge 0$ for all choices of $\boldsymbol{w},r,s$; this follows on taking $\boldsymbol{u}=(0,0,\dots)$ and $b=c=0$ in (5.1.7). Let $\alpha\in[0,\infty)$ be a parameter such that $\mathcal{S}\subset\{I\ge\alpha\}$. By Lemma 5.3.4 and Lemma 5.3.5, there exists some $m\in\mathbb{N}$ such that for every sequence $\boldsymbol{a}\in\mathbb{V}$,

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\left(\Pi(n^{-1/2}(\boldsymbol{a}^{(n)})^\top X^{(n)}),\frac{1}{n}\sum_{i=1}^n\eta(X_i)\right)\in\mathcal{K}_m^c\right)\le-\alpha.$$

By hypothesis, the upper bound (5.2.4) holds for $\mathcal{S}\cap\mathcal{K}_m$ and all $\boldsymbol{a}\in\Omega_{\mathcal{S}}$. Since $\mathcal{S}\subset\{I\ge\alpha\}$, this implies that for every sequence $\boldsymbol{a}\in\Omega_{\mathcal{S}}$,

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\left(\Pi(n^{-1/2}(\boldsymbol{a}^{(n)})^\top X^{(n)}),\frac{1}{n}\sum_{i=1}^n\eta(X_i)\right)\in\mathcal{S}\cap\mathcal{K}_m\right)\le-\alpha.$$

Together, the previous two displays imply that for all $\boldsymbol{a}\in\Omega_{\mathcal{S}}$,

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\left(\Pi(n^{-1/2}(\boldsymbol{a}^{(n)})^\top X^{(n)}),\frac{1}{n}\sum_{i=1}^n\eta(X_i)\right)\in\mathcal{S}\right)\le-\alpha.$$

We can set $\alpha:=\inf_{(\boldsymbol{w},r,s)\in\mathcal{S}}I(\boldsymbol{w},r,s)$, since we required only that $\mathcal{S}\subset\{I\ge\alpha\}$, and hence the claim of Step 1 follows.

Step 2. We next show that for σ-almost every sequence $\boldsymbol{a}=\left(\boldsymbol{a}^{(1)},\boldsymbol{a}^{(2)},\dots\right)\in\mathbb{V}$, the upper bound (5.2.4) holds for all closed sets $\mathcal{S}$.

Note that the space $\mathcal{X}\times\mathbb{R}_+$ possesses a countable basis $\{\mathcal{O}_i\}_{i\in\mathbb{N}}$ since it is a separable metric space. Consider the countable family of closed sets $\mathfrak{C}=\{\cap_{i\in\mathcal{F}}\mathcal{O}_i^c$:

$\mathcal{F}$ is finite subset of $\mathbb{N}\}$ and define $\widehat{\Omega} = \cap_{\mathcal{S}\in\mathfrak{C}}\Omega_{\mathcal{S}}$. Then $\mathbb{P}(\widehat{\Omega}) = 1$. For an arbitrary closed set $\mathcal{S}$, we have $\mathcal{S} = \cap_{i\in\mathcal{F}'}\mathcal{O}_i^c$ for some $\mathcal{F}' \subset \mathbb{N}$. For any $\varepsilon > 0$, there exists a finite set $\mathcal{F} \subset \mathcal{F}' \subset \mathbb{N}$, such that for $\mathcal{C} = \cap_{i\in\mathcal{F}}\mathcal{O}_i^c$, we have

$$\inf_{(\boldsymbol{w},r,s)\in\mathcal{C}} I(\boldsymbol{w}, r, s) > \inf_{(\boldsymbol{w},r,s)\in\mathcal{S}} I(\boldsymbol{w}, r, s) - \varepsilon.$$

Therefore, for $\boldsymbol{a} \in \widehat{\Omega}$, using $\mathcal{S} \subset \mathcal{C}$ and the previous step, we have

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left(\Pi(n^{-1/2}(\boldsymbol{a}^{(n)})^\top X^{(n)}), \frac{1}{n}\sum_{i=1}^n \eta(X_i)\right) \in \mathcal{S}\right) \\
\leq\ &\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left(\Pi(n^{-1/2}(\boldsymbol{a}^{(n)})^\top X^{(n)}), \frac{1}{n}\sum_{i=1}^n \eta(X_i)\right) \in \mathcal{C}\right) \\
\leq\ & -\inf_{(\boldsymbol{w},r,s)\in\mathcal{C}} I(\boldsymbol{w}, r, s) \\
<\ & -\inf_{(\boldsymbol{w},r,s)\in\mathcal{S}} I(\boldsymbol{w}, r, s) + \varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, for every $\boldsymbol{a} \in \widehat{\Omega}$ the upper bound (5.2.4) holds for all closed sets $\mathcal{S}$. The proof is complete. □

Proof of Proposition 5.2.3. We begin by introducing some useful notation. Given positive integers $k \geq m$, let $\mathcal{J}_{k,m}$ denote the set of injective mappings from $\{1, \ldots, m\}$ to $\{1, \ldots, k\}$. For $\boldsymbol{x} \in \mathbb{R}^k$ and $\tau \in \mathcal{J}_{k,m}$, we define $\tau(\boldsymbol{x}) \in \mathbb{R}^k$ as the vector

$$\tau(\boldsymbol{x}) = (x_{\tau(1)}, x_{\tau(2)}, \ldots, x_{\tau(m)}, x_{m+1}, x_{m+2}, \ldots) \tag{5.3.6}$$

whose first m coordinates are $x_{\tau(1)}, x_{\tau(2)}, \ldots, x_{\tau(m)}$, with the remaining coordinates taken in their original order.

By Lemma 5.3.1(3) and Lemma 5.3.9, to prove the upper bound (5.2.4) for σ-a.e. sequence $\boldsymbol{a} \in \mathbb{V}$, it suffices to prove (5.2.4) for compact sets. For the remainder of this proof, we fix a compact set $\mathcal{C}$ and a corresponding constant A such that $\max(r, s) \leq A$ for all $(\boldsymbol{w}, r, s) \in \mathcal{C}$. The proof consists of two further steps.

Step 1: Reduction to a finite number of coordinates. We start by establishing some general properties of the set $\mathcal{C}$. Set

$$\alpha = \alpha(\mathbb{C}) := \inf_{(\boldsymbol{w},r,s)\in\mathbb{C}} I(\boldsymbol{w}, r, s). \tag{5.3.7}$$

Let $\delta > 0$ be a parameter and define

$$\alpha_\delta = \min(\alpha, \delta^{-1}) - \delta. \tag{5.3.8}$$

Then from the definition of α in (5.1.8), and (5.1.7), we see that for every $(\boldsymbol{w}, r, s) \in \mathbb{C}$, there exists a triple

$$(\boldsymbol{u}, b, c) = \Big(\boldsymbol{u}(\boldsymbol{w}, r, s, \delta), b(\boldsymbol{w}, r, s, \delta), c(\boldsymbol{w}, r, s, \delta\Big) \in \ell^2 \times \mathbb{R}_+ \times \mathbb{R}_+ \tag{5.3.9}$$

such that

$$\langle \boldsymbol{u}, \boldsymbol{w}\rangle + bt + cs - \Lambda(\boldsymbol{u}, b, c) > \alpha_\delta, \tag{5.3.10}$$

with $t = \sqrt{r^2 - \|\boldsymbol{w}\|_2^2}$. By (5.3.4), there exists $m = m(\boldsymbol{w}, r, s, \delta) \in \mathbb{N}$ such that

$$\langle \boldsymbol{u}_\le, \boldsymbol{w}_\le\rangle + bt_m + cs - \Lambda_m(\boldsymbol{u}_\le, b, c) > \alpha_\delta,$$

where $t_m = \sqrt{r_1^2 - \|\boldsymbol{w}_\le\|_2^2}$, and $\boldsymbol{u}_\le = \boldsymbol{u}_{\le m}$.

Consider the open set of near-optimizers

$$B_{\boldsymbol{w},r,s} = \left\{(\hat{\boldsymbol{w}}, t, s) \in \mathbb{R}^m \times \mathbb{R}_+ \times \mathbb{R}_+ : \boldsymbol{u}_\le \cdot \hat{\boldsymbol{w}} + bt + cs - \Lambda_m(\boldsymbol{u}_\le, b, c) > \alpha_\delta\right\}, \tag{5.3.11}$$

where $(\boldsymbol{u}, b, c)$ are chosen as in (5.3.9), and for $(\boldsymbol{w}, r, s) \in \mathcal{X} \times \mathbb{R}_+$, define the set

$$\mathcal{W}_{\boldsymbol{w},r,s} = \left\{(\boldsymbol{w}, r, s) \in \mathcal{X} \times \mathbb{R}_+ : \left(\boldsymbol{w}_\le, \sqrt{r^2 - |\boldsymbol{w}_\le|^2}, s\right) \in B_{\boldsymbol{w},r,s}, r < A, s < A\right\}.$$

This is an open set that contains $(\boldsymbol{w}, r, s)$, by (5.3.10). Since $\mathbb{C}$ is compact, there exists a collection of finitely many open sets $\mathcal{W}^{(j)} = \mathcal{W}_{\boldsymbol{w}_j, r_j, s_j}$, $j = 1, \ldots, m_0$, with $(\boldsymbol{w}_j, r_j, s_j) \in \mathcal{X} \times \mathbb{R}_+$, such that

$$\mathbb{C} \subset \bigcup_{j=1}^{m_0} \mathcal{W}^{(j)}. \tag{5.3.12}$$

Let $B^{(j)} = B_{\boldsymbol{w}_j, r_j, s_j}$, and let $(\boldsymbol{u}^{(j)}, b^{(j)}, c^{(j)})$ be the triple in (5.3.9) associated with $(\boldsymbol{w}_j, r_j, s_j)$.

Next, we write $k = k_n$, suppressing the dependence on n, and recall the definition of $\mathcal{J}_{k,m}$ made before (5.3.6). Given $\boldsymbol{w} \in \mathbb{R}^k$, we consider the map $\tau^{\boldsymbol{w}} \in \mathcal{J}_{k,m}$ such that the corresponding induced map $\tau^{\boldsymbol{w}} \colon \mathbb{R}^k \to \mathbb{R}^m$ defined in (5.3.6) satisfies $\tau^{\boldsymbol{w}}(\boldsymbol{w})_{\leq m} = [\boldsymbol{w}]_{\leq m}$. Then we see that for any $j = 1, \ldots, m_0$,

$$\begin{aligned}
&\left\{(\boldsymbol{w}, s) \in \mathbb{R}^k \times \mathbb{R}_+ : ([\boldsymbol{w}], \|\boldsymbol{w}\|_2, s) \in \mathcal{W}^{(j)}\right\} \\
&\subset \left\{(\boldsymbol{w}, s) \in \mathbb{R}^k \times \mathbb{R}_+ : ([\boldsymbol{w}]_{\leq}, \|[\boldsymbol{w}]_{>}\|_2, s) \in B^{(j)}, \|\boldsymbol{w}\|_2 \leq A\right\} \\
&= \left\{(\boldsymbol{w}, s) \in \mathbb{R}^k \times \mathbb{R}_+ : (\tau^{\boldsymbol{w}}(\boldsymbol{w})_{\leq}, \|\tau^{\boldsymbol{w}}(\boldsymbol{w})_{>}\|_2, s) \in B^{(j)}, \|\boldsymbol{w}\|_2 \leq A\right\}.
\end{aligned} \tag{5.3.13}$$

Now, for $j = 1, \ldots, m_0$, define, for any $\tau \in \mathcal{J}_{k,m}$,

$$\widehat{B}^{(j)}_{k,\tau} := \left\{(\boldsymbol{w}, s) \in \mathbb{R}^k \times \mathbb{R}_+ : (\tau(\boldsymbol{w})_{\leq}, \|\tau(\boldsymbol{w})_{>}\|_2, s) \in B, s \leq A, \|\boldsymbol{w}\|_2 \leq A\right\},$$

and note that (5.3.13) implies

$$\{(\boldsymbol{w}, s) \in \mathbb{R}^k \times \mathbb{R}_+ : ([\boldsymbol{w}], \|\boldsymbol{w}\|_2, s) \in \mathcal{W}\} \subset \widehat{B}^{(j)}_k := \bigcup_{\tau \in \mathcal{J}_{k,m}} \widehat{B}^{(j)}_{k,\tau}. \tag{5.3.14}$$

To apply these properties of $\mathcal{C}$, we introduce the notation

$$W = W^{(n)} := n^{-1/2} (\boldsymbol{a}^{(n)})^\top X^{(n)}, \qquad L = L^{(n)} := \frac{1}{n} \sum_{i=1}^{n} \eta(X_i). \tag{5.3.15}$$

By (5.3.13) and (5.3.14), if $([W], \|W\|_2, L) \in \mathcal{W}^{(j)}$, then $(W, L) \in \widehat{B}_k^{(j)}$. Combining this observation with (5.3.12) and a union bound, we find

$$\mathbb{P}\left((\Pi(W), L) \in \mathcal{C}\right) \leq \sum_{j=1}^{m_0} \mathbb{P}\left((\Pi(W), L) \in \mathcal{V}^{(j)}\right) \leq \sum_{j=1}^{m_0} \mathbb{P}\left((W, L) \in \widehat{B}_k^{(j)}\right). \tag{5.3.16}$$

It therefore suffices to bound the probabilities $\mathbb{P}((W, L) \in \widehat{B}_k^{(j)})$. For concreteness, we focus on $\mathbb{P}((W, L) \in \widehat{B}_k^{(1)})$. The argument for the other terms is analogous.

Step 2: Probability bound for $\widehat{B}_k^{(1)}$. Set $\widehat{B}_k = \widehat{B}_k^{(1)}$ and $(\boldsymbol{u}, b, c) = (\boldsymbol{u}^{(1)}, b^{(1)}, c^{(1)})$. From (5.3.14), we have

$$\mathbb{1}_{\{(W,L)\in\widehat{B}_k\}} \leq \sum_{\tau\in\mathcal{J}_{k,m}} \mathbb{1}_{\{(\tau(W)_\leq, \|\tau(W)_>\|_2, L)\in B\}} \mathbb{1}_{\{\|W\|_2\leq A\}} \mathbb{1}_{\{L\leq A\}}. \tag{5.3.17}$$

Using the quantity

$$H_{\boldsymbol{u},b,c}(B) \coloneqq \inf_{(\hat{\boldsymbol{w}},t,s)\in B} \{\langle \boldsymbol{u}_\leq, \hat{\boldsymbol{w}}\rangle + bt + cs\} \tag{5.3.18}$$

to bound the first indicator in the summand in (5.3.17), we obtain

$$\begin{aligned} \mathbb{1}_{\{(W,L)\in\widehat{B}_k\}} &\leq \exp\Big(-nH_{\boldsymbol{u},b,c}(B)\Big) \\ &\qquad \times \sum_{\tau\in\mathcal{J}_{k,m}} \exp\Big(n(\boldsymbol{u}_\leq \cdot \tau(W)_\leq + b\|\tau(W)_>\|_2 + cL)\Big) \mathbb{1}_{\{\|W\|_2\leq A\}} \mathbb{1}_{\{L\leq A\}}. \end{aligned} \tag{5.3.19}$$

With a view toward bounding the right-hand side of (5.3.19), we first let $\kappa > 0$ be a parameter. We define

$$\mathcal{G} \coloneqq \left\{\boldsymbol{v}' \in \mathbb{R}^{k-m} : \|\boldsymbol{v}'\|_2 = b\right\},$$

and let $\mathrm{Cov}(\mathcal{G})$ denote a minimal cover of the set $\mathcal{G}$ by open ℓ^2 balls of radius $b\kappa$ with centers in $\mathcal{G}$. We let $\mathcal{U}'$ denote the set of the centers of balls in $\mathrm{Cov}(\mathcal{G})$, and define

$$\mathcal{U} := \bigcup_{\tau \in \mathcal{J}_{k,m}} \Big\{ \boldsymbol{v} \in \mathbb{R}^k : \tau(\boldsymbol{v})_{\leq} = \boldsymbol{u}_{\leq}, \tau(\boldsymbol{v})_{>} \in \mathcal{U}' \cup \{0\} \Big\}. \tag{5.3.20}$$

A standard volume estimate (see [164, Proposition 4.2.12]) shows that

$$|\mathcal{U}'| \leq (1 + 2/\kappa)^{k-m} \leq (1 + 2/\kappa)^k.$$

Together with the bound $|\mathcal{J}_{k,m}| \leq k^m$, this implies that there exist $\hat{\varepsilon} > 0$ and $n_0 = n_0(\{k_n\}_{n\in\mathbb{N}}, \kappa) \in \mathbb{N}$ such that for all $n \geq n_0$,

$$|\mathcal{U}| \leq \exp\Big((1 - \hat{\varepsilon})n\Big). \tag{5.3.21}$$

We now show that we can approximate the sum in (5.3.19) by a sum over elements in $\mathcal{U}$. Fix $\tau \in \mathcal{J}_{k,m}$. We first consider the event where $\|W\|_2 \leq A$ and $\|\tau(W)_{>}\|_2 > \kappa$. Since

$$U := b\frac{\tau(W)_{>}}{\|\tau(W)_{>}\|_2}$$

lies in $\mathcal{G}$, there exists $\boldsymbol{v} \in \mathcal{U}$ such that $\tau(\boldsymbol{v})_{\leq} = \boldsymbol{u}_{\leq}$ and $\|U - \tau(\boldsymbol{v})_{>}\|_2 < b\kappa$. Then

$$\begin{aligned}
&\boldsymbol{u}_{\leq} \cdot \tau(W)_{\leq} + b\|\tau(W)_{>}\|_2 \\
&= \tau(\boldsymbol{v})_{\leq} \cdot \tau(W)_{\leq} + U \cdot \tau(W)_{>} \\
&= \tau(\boldsymbol{v})_{\leq} \cdot \tau(W)_{\leq} + \tau(\boldsymbol{v})_{>} \cdot \tau(W)_{>} + (U - \tau(\boldsymbol{v})_{>}) \cdot \tau(W)_{>} \\
&\leq \boldsymbol{v} \cdot W + Ab\kappa,
\end{aligned} \tag{5.3.22}$$

where to obtain the last inequality, we used the Cauchy–Schwarz inequality, and the estimates $\|\tau(W)_{>}\|_2 \leq \|W\|_2 \leq A$ and $\|U - \tau(\boldsymbol{v})_{>}\|_2 \leq b\kappa$. On the other hand, on the event $\|W\|_2 \leq A$ and $\|\tau(W)_{>}\|_2 \leq \kappa$, a direct bound shows that for any $\boldsymbol{v}$ such that

$\tau(\boldsymbol{v})_{\leq} = \boldsymbol{u}_{\leq}$ and $\tau(\boldsymbol{v})_{>} = 0$ (which in particular satisfies $\boldsymbol{v} \in \mathcal{U}$), we have

$$\boldsymbol{u}_{\leq} \cdot \tau(W)_{\leq} + b\|\tau(W)_{>}\|_2 = \tau(\boldsymbol{v})_{\leq} \cdot \tau(W)_{\leq} + b\|\tau(W)_{>}\|_2 \leq \boldsymbol{v} \cdot W + b\kappa. \tag{5.3.23}$$

The upper bounds (5.3.22) and (5.3.23) together with (5.3.19) show that

$$\mathbb{P}((W, L) \in \widehat{B}_k) \leq \exp(-n(H_{\boldsymbol{u},b,c}(B) - (A+1)b\kappa)) \, |\mathcal{J}_{k,m}| \sum_{\boldsymbol{v} \in \mathcal{U}} \mathbb{E}\left[\exp(n(\boldsymbol{v} \cdot W + cL))\right].$$

Taking logarithms, we find

$$\begin{aligned} n^{-1} \log \mathbb{P}((W, L) \in \widehat{B}_k) \leq & -(H_{\boldsymbol{u},b,c}(B) - (A+1)b\kappa) + n^{-1} \log |\mathcal{J}_{k,m}| \, |\mathcal{U}| \\ & + \max_{\boldsymbol{v} \in \mathcal{U}} n^{-1} \log \mathbb{E}\left[\exp(n(\boldsymbol{v} \cdot W + cL))\right]. \end{aligned} \tag{5.3.24}$$

Recalling the definitions of W and L in (5.3.15), the fact that the variables $\{X_i\}_{i \in \mathbb{N}}$ are i.i.d., the definition of $\overline{\Lambda}$ in (5.1.5), and the definition of F_n in (5.3.1), we have

$$\begin{aligned} & n^{-1} \log \mathbb{E}\left[\exp(n(\boldsymbol{v} \cdot W + cL))\right] \\ & \quad = n^{-1} \log \mathbb{E}\left[\exp\left(n^{1/2} \sum_{i=1}^{n} X_i \left(\sum_{j=1}^{k_n} v_j a_{ji}^{(n)}\right) + c \sum_{i=1}^{n} \eta(X_i)\right)\right] \\ & \quad = n^{-1} \sum_{i=1}^{n} \log \mathbb{E}\left[\exp\left(X_i \langle \boldsymbol{v}, \sqrt{n} \boldsymbol{a}_i^{(n)} \rangle + c\eta(X_i)\right)\right] \\ & \quad = F_n(\boldsymbol{v}, \boldsymbol{a}^{(n)}, c). \end{aligned}$$

Further, by (5.3.20), each $\boldsymbol{v} \in \mathcal{U}$ satisfies either $\|\boldsymbol{v}\|^2 = \|\boldsymbol{u}_{\leq}\|_2^2$ or $\|\boldsymbol{v}\|^2 = \|\boldsymbol{u}_{\leq}\|_2^2 + b^2$; that is, the ℓ^2 norms of all vectors $\boldsymbol{v} \in \mathcal{U}$ take one of two values. Recalling the previous estimate on $|\mathcal{U}|$ from (5.3.21), taking the limit superior of both sides of (5.3.24), using the last display and Lemma 5.3.6 to control the maximum therein, and recalling $k_n/n \to 0$ from the assumption (5.2.3), we obtain

$$\limsup n^{-1} \log \mathbb{P}\Big((W, L) \in \widehat{B}_k\Big) \leq -H_{\boldsymbol{u},b,c}(B) + (A+1)b\kappa + \max(\Lambda_m(\boldsymbol{u}_{\leq}, b, c), \Lambda_m(\boldsymbol{u}_{\leq}, 0, c))$$

$$\le -H_{\boldsymbol{u},b,c}(B) + (A+1)b\kappa + \Lambda_m(\boldsymbol{u}_{\le}, b, c).$$

In the last line, we used Jensen's inequality (as in (5.3.4)) to simplify the maximum. By the definitions of the sets $H_{\boldsymbol{u},u,c}$ and B in (5.3.18) and (5.3.11), respectively, this becomes

$$\begin{aligned}\limsup n^{-1} \log \mathbb{P}\Big((W, L) \in \widehat{B}_k\Big) &\le -\alpha_\delta - \Lambda_m(\boldsymbol{u}_{\le}, b, c) + (A+1)b\kappa + \Lambda_m(\boldsymbol{u}_{\le}, b, c) \\ &\le -\alpha_\delta + (A+1)b\kappa.\end{aligned}$$

Since this bound holds for all $\kappa > 0$, it implies that

$$\limsup n^{-1} \log \mathbb{P}\Big((W, L) \in \widehat{B}_k\Big) \le -\alpha_\delta.$$

Finally, since this holds for all $\delta > 0$, we have by (5.3.7) and (5.3.8) that

$$\limsup n^{-1} \log \mathbb{P}\Big((W, L) \in \widehat{B}_k\Big) \le -\alpha = \inf_{(\boldsymbol{w},r,s)\in\mathcal{C}} I(\boldsymbol{w}, r, w).$$

Inserting this and the analogous bounds for all $B_k^{(j)}$, $j = 1, \ldots, m_0$, into (5.3.16) completes the proof. □

5.4 Lower Bound

5.4.1 Preliminary Lemmas

We begin by stating some preliminary results. Given a metric space $(\mathcal{T}, d)$, $t_0 \in \mathcal{T}$ and $r \ge 0$, set

$$B(t_0, r) := \{t \in \mathcal{T} : d(t_0, t) < r\}. \tag{5.4.1}$$

Given $m \in \mathbb{N}$ and $\boldsymbol{v} \in \ell^2$, we recall the quantities $\boldsymbol{v}_{>m}$ and $\boldsymbol{v}_{\leq m}$ defined in (5.3.2). Recalling the constant T from Assumption 4, we set

$$\mathcal{D}_m \coloneqq \{(\boldsymbol{v}, b, c) : (\boldsymbol{v}, b) \in \mathbb{R}^m \times \mathbb{R}, c \in (-\infty, T)\}.$$

Let Λ_m^* denote the Legendre transform of the function Λ_m from (5.3.3), defined by

$$\Lambda_m^*(\hat{\boldsymbol{w}}, t, s) \coloneqq \sup_{(\hat{\boldsymbol{u}}, b, c) \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}} \left\{ \langle \hat{\boldsymbol{u}}, \hat{\boldsymbol{w}} \rangle + bt + cs - \Lambda_m(\hat{\boldsymbol{u}}, b, c) \right\}. \tag{5.4.2}$$

We denote the domain of Λ_m^* by

$$\mathcal{D}_m^* \coloneqq \left\{ (\hat{\boldsymbol{w}}, r, s) : \Lambda_m^*(\hat{\boldsymbol{w}}, r, s) < \infty \right\}. \tag{5.4.3}$$

Finally, let $\nabla \Lambda_m \in \mathbb{R}^{m+2}$ denote the gradient of Λ_m:

$$\begin{aligned} &\nabla \Lambda_m(\boldsymbol{v}, b, c) = \\ &\qquad \left(\partial_{v_1} \Lambda_m(\boldsymbol{v}, b, c), \partial_{v_2} \Lambda_m(\boldsymbol{v}, b, c), \dots, \partial_{v_m} \Lambda_m(\boldsymbol{v}, b, c), \partial_b \Lambda_m(\boldsymbol{v}, b, c), \partial_c \Lambda_m(\boldsymbol{v}, b, c) \right). \end{aligned}$$

The first result pertains to the existence of a point with useful properties in any open set $\mathcal{O} \subset \mathcal{X} \times \mathbb{R}$. Its proof is relegated to Section 5.7.2.

Lemma 5.4.1. *Fix $\varepsilon > 0$, an open set $\mathcal{O} \subset \mathcal{X} \times \mathbb{R}$, and a point $(\boldsymbol{w}, r, s) \in \mathcal{O}$ such that $I(\boldsymbol{w}, r, s) < \infty$. Then there exist $\bar{\kappa} > 0$, $m_0 \in \mathbb{N}$ and $(\overline{\boldsymbol{w}}, \bar{t}, \bar{s}) \in \mathbb{R}^{m_0} \times \mathbb{R}_+ \times \mathbb{R}_+$ such that the following claims hold:*

1. *$(\bar{\boldsymbol{w}}, \bar{t}, \bar{s}) = \nabla \Lambda_{m_0}(\bar{\boldsymbol{v}}, \bar{b}, \bar{c})$ for some $(\bar{\boldsymbol{v}}, \bar{b}, \bar{c}) \in \mathcal{D}_{m_0}$;*
2. *$|\bar{\boldsymbol{w}}_j| > |\bar{\boldsymbol{w}}_{j+1}| > 0$, for all $1 \leq j \leq m_0 - 1$;*
3. *$\Lambda_{m_0}^* \left(\bar{\boldsymbol{w}}, \bar{t}, \bar{s} \right) \leq I(\boldsymbol{w}, r, s) + \varepsilon$;*

4. For all $(\boldsymbol{w}', r', s') \in \mathcal{X} \times \mathbb{R}_+$, *if*

$$\left(\boldsymbol{w}'_{\leq m_0}, \sqrt{r'^2 - \|\boldsymbol{w}'_{\leq m_0}\|_2^2}, s'\right) \in B\left((\overline{\boldsymbol{w}}, \bar{t}, \bar{s}), \bar{\kappa}/2\right), \tag{5.4.4}$$

then $(\boldsymbol{w}', r', s') \in \mathcal{O}$.

We now present the second result, whose proof is deferred to Section 5.7.3. Recall the definitions of $W^{(n)}$ and $L^{(n)}$ from (5.3.15).

Lemma 5.4.2. *Fix* $m \in \mathbb{N}$, $\kappa > 0$, *and* $(\hat{\boldsymbol{u}}, b, c) \in \mathcal{D}_m$. *Define* $(\hat{\boldsymbol{w}}, t, s) := \nabla\Lambda_m(\hat{\boldsymbol{u}}, b, c)$ *and* $B := B\left((\hat{\boldsymbol{w}}, t, s), \kappa\right)$. *Then, for* σ*-a.e.* $\boldsymbol{a} = (\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}, \cdots)$, *the sequence* $\{W^{(n)}, L^{(n)}\}$, $n \in \mathbb{N}$, *satisfies*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left(W^{(n)}_{\leq m}, \left\|W^{(n)}_{>m}\right\|_2, L^{(n)}\right) \in B, \left\|W^{(n)}_{>m}\right\|_\infty \leq 2m^{-1/2}\right) \geq -\Lambda_m^*(\hat{\boldsymbol{w}}, t, s). \tag{5.4.5}$$

5.4.2 Proof of the Lower Bound.

Proof of Proposition 5.2.4. Recall the definitions of $W^{(n)}, L^{(n)}$ in (5.3.15). We first establish the following claim.

Claim 1. For every $\varepsilon > 0$, open set $\mathcal{O} \subset \mathcal{X} \times \mathbb{R}_+$, and point $(\boldsymbol{w}, r, s) \in \mathcal{O}$, there exists a set $\Omega_{\mathcal{O},\varepsilon,(\boldsymbol{w},r,s)} \subset \mathbb{V}$ of sequences $\boldsymbol{a} = \{\boldsymbol{a}^{(n)}\}_{n\in\mathbb{N}}$ with $\sigma(\Omega_{\mathcal{O},\varepsilon,(\boldsymbol{w},r,s)}) = 1$ and such that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left([W^{(n)}], \|W^{(n)}\|_2, L^{(n)}\right) \in \mathcal{O}\right) \geq -I(\boldsymbol{w}, r, s) - \varepsilon. \tag{5.4.6}$$

We may assume that $I(\boldsymbol{w}, r, s) < \infty$, since otherwise the claim is trivial. In this case, there exist $\bar{\kappa} > 0$, $m_0 \in \mathbb{N}$, and $(\bar{\boldsymbol{w}}, \bar{t}, \bar{s}) \in \mathbb{R}^{m_0} \times \mathbb{R}_+ \times \mathbb{R}_+$ that satisfy the properties of Lemma 5.4.1.

Fix $m > m_0$ such that $3m^{-1/2} < |\bar{\boldsymbol{w}}_{m_0}|$ (such an m exists since $|\bar{\boldsymbol{w}}_{m_0}| > 0$ by property

(2) of Lemma 5.4.1) and set

$$\kappa \coloneqq \min\left\{\frac{\bar{\kappa}}{2}, m^{-1/2}, \frac{1}{3}\min_{1\le j\le m_0-1}\left(|\bar{\boldsymbol{w}}_j| - |\bar{\boldsymbol{w}}_{j+1}|\right)\right\}.$$

Note that property (2) of Lemma 5.4.1 ensures $\kappa > 0$. Set

$$\widehat{\boldsymbol{w}} \coloneqq (\bar{\boldsymbol{w}}, 0, \dots, 0) \in \mathbb{R}^m, \quad B_1 \coloneqq B\left((\widehat{\boldsymbol{w}}, \bar{t}, \bar{s}), \kappa\right), \quad B_2 \coloneqq B\left((\overline{\boldsymbol{w}}, \bar{t}, \bar{s}), \bar{\kappa}/2\right).$$

Property (1) of Lemma 5.4.1 implies that there exists $\hat{\boldsymbol{u}} \coloneqq (\bar{\boldsymbol{v}}, 0, \dots, 0) \in \mathbb{R}^m$ such that

$$(\hat{\boldsymbol{u}}, \bar{b}, \bar{c}) \in \mathcal{D}_{m_0}, \text{ and } \nabla\Lambda_m(\hat{\boldsymbol{u}}, \bar{b}, \bar{c}) = (\hat{\boldsymbol{w}}, \bar{t}, \bar{c}). \tag{5.4.7}$$

Given $(\boldsymbol{w}', r', s') \in \ell^2 \times \mathbb{R}_+ \times \mathbb{R}_+$ with $r' \ge \|\boldsymbol{w}'\|_2$, we write

$$t'_{\boldsymbol{w}',r'} \coloneqq \sqrt{(r')^2 - \|\boldsymbol{w}'_{\le m}\|_2^2}, \quad t''_{\boldsymbol{w}',r'} \coloneqq \sqrt{(r')^2 - \|\boldsymbol{w}_{\le m_0}\|_2^2}.$$

Consider $(\boldsymbol{w}', r', s') \in \ell^2 \times \mathbb{R}_+ \times \mathbb{R}_+$ with $(\boldsymbol{w}'_{\le m}, t'_{\boldsymbol{w}',r'}, s') \in B_1$. By the definition of B_1, it follows that $\|\boldsymbol{w}'_{\le m} - \hat{\boldsymbol{w}}\|_\infty \le \kappa$. The definition of κ, m and the strict ordering in the first m_0 coordinates of $\hat{\boldsymbol{w}}$ imply that $\boldsymbol{w}'$ satisfy the following properties:

1. $|\boldsymbol{w}'_j| > |\boldsymbol{w}'_{j+1}| > 2m^{-1/2}$ for all $1 \le j \le m_0 - 1$;
2. $|\boldsymbol{w}'_j| \le m^{-1/2}$ for all $m_0 + 1 \le j \le m$.

If we further require that $\|\boldsymbol{w}'_{>m}\|_\infty \le 2m^{-1/2}$, then it follows that $[\boldsymbol{w}']_{\le m_0} = \boldsymbol{w}'_{\le m_0}$.

The above discussion, and the definitions of $t'_{\boldsymbol{w}',r'}$ and $t''_{\boldsymbol{w}',r'}$, imply that

$$\begin{aligned}&\{(\boldsymbol{w}', r', s') \in \ell^2 \times \mathbb{R}_+ \times \mathbb{R}_+ : \left(\boldsymbol{w}'_{\le m}, t'_{\boldsymbol{w}',r'}, s'\right) \in B_1 \text{ and } \|\boldsymbol{w}'_{>m}\|_\infty \le 2m^{-1/2}\} \\ &\subset \{(\boldsymbol{w}', r', s') \in \ell^2 \times \mathbb{R}_+ \times \mathbb{R}_+ : [\boldsymbol{w}']_{\le m_0} = \boldsymbol{w}'_{\le m_0} \text{ and } \left(\boldsymbol{w}'_{\le m_0}, t''_{\boldsymbol{w}',r'}, s'\right) \in B_2\}.\end{aligned} \tag{5.4.8}$$

Thus, for σ-a.e. $\boldsymbol{a} = (\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}, \cdots)$, we see that

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(([W^{(n)}], \|W^{(n)}\|_2, L^{(n)}) \in \mathcal{O}\right) \\
&\quad \geq \liminf_{n\to\infty} \ \frac{1}{n} \log \mathbb{P}\Big(\left([W^{(n)}]_{\leq m_0}, \|W^{(n)}_{>m_0}\|_2, L^{(n)}\right) \in B_2\Big) \\
&\quad \geq \liminf_{n\to\infty} \ \frac{1}{n} \log \mathbb{P}\left(\left(W^{(n)}_{\leq m}, \|W^{(n)}_{>m}\|_2, L^{(n)}\right) \in B_1, \|W^{(n)}_{>m}\|_\infty \leq 2m^{-1/2}\right) \\
&\quad \geq - \Lambda^*_m(\widehat{\boldsymbol{w}}, \bar{t}, \bar{s}) \\
&\quad \geq - I(\boldsymbol{w}, r, s) - \varepsilon,
\end{aligned}$$

where we used (5.4.4) in the first step, (5.4.8) in the second step, (5.4.7) and Lemma 5.4.2 (with $\delta = \kappa$) in the third step and property (3) of Lemma 5.4.1 in the last step. This proves the claim (5.4.6).

Next let $\{\boldsymbol{w}_l, r_l, s_l\}_{l=1}^\infty$ be a sequence such that

$$\lim_{l\to\infty} I(\boldsymbol{w}_l, r_l, s_l) = \inf_{(\boldsymbol{w},r,s)\in\mathcal{O}} I(\boldsymbol{w}, r, s),$$

and define $\Omega_{\mathcal{O}} := \cap_{l=1}^\infty \Omega_{\mathcal{O},\frac{1}{l},(\boldsymbol{w}_l,r_l,s_l)}$, where $\Omega_{\mathcal{O},\frac{1}{\ell},(\boldsymbol{w}_\ell,r_\ell,s_\ell)}$ is the set in from Claim 1. Then $\sigma(\Omega_{\mathcal{O}}) = 1$ and

$$\liminf_{n\to\infty} n^{-1} \log \mathbb{P}_n \left(\left([W^{(n)}], \|W^{(n)}\|_2, L^{(n)}\right) \in \mathcal{O}\right) \geq - \inf_{(\boldsymbol{w},r,s)\in\mathcal{O}} I(\boldsymbol{w}, r, s). \tag{5.4.9}$$

Next, observe that since $\mathcal{X} \times \mathbb{R}_+$ is a separable metric space, it possesses a countable basis $\{\mathcal{O}_i\}_{i\in\mathbb{N}}$. Setting $\widehat{\Omega} := \cap_{i\in\mathbb{N}} \Omega_{\mathcal{O}_i}$, we have $\sigma(\widehat{\Omega}) = 1$. Now consider an arbitrary open set $\mathcal{O} \subset \mathcal{X} \times \mathbb{R}_+$ and any $(\boldsymbol{w}, r, s) \in \mathcal{O}$. By the definition of a basis, there exists a basis element $\mathcal{O}_i$ such that $(\boldsymbol{w}, r, s) \in \mathcal{O}_i \subset \mathcal{O}$. For sequence $\boldsymbol{a} \in \widehat{\Omega} \subset \Omega_{\mathcal{O}_i}$, the corresponding sequence $W^{(n)} := W^{\boldsymbol{a}^{(n)}}$ satisfies

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left([W^{(n)}], \|W^{(n)}\|_2, L^{(n)}\right) \in \mathcal{O}\right) \\
&\geq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left([W^{(n)}], \|W^{(n)}\|_2, L^{(n)}\right) \in \mathcal{O}_i\right) \\
&\geq -I(\boldsymbol{w}, r, s),
\end{aligned} \tag{5.4.10}$$

where the latter follows from (5.4.9) applied with $\mathcal{O}$ replaced with $\mathcal{O}_i$. Since $(\boldsymbol{w}, r, s) \in \mathcal{O}$ is arbitrary, taking the supremum of the right-hand side of (5.4.10) over $(\boldsymbol{w}, r, s) \in \mathcal{O}$ yields (5.2.4). □

5.5 Proof of the Main Result

Proof of Theorem 5.2.5. Combining Proposition 5.2.3 and Proposition 5.2.4, we conclude that, for σ-a.e. $\boldsymbol{a} = (\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}, \dots) \in \mathbb{V}$, the sequence of random variables

$$\left(\Pi\left(n^{-1/2}\left(\boldsymbol{a}^{(n)}\right)^\top X^{(n)}\right), \frac{1}{n}\sum_{i=1}^n \eta(X_i)\right), \quad n \in \mathbb{N},$$

satisfies an LDP with speed n and good rate function $I(\boldsymbol{w}, r, s) = \Lambda^*(\boldsymbol{x}, \sqrt{r_1^2 - \|\boldsymbol{x}\|_2^2}, r_2)$. By Assumption 2 and the definition of Π in (5.1.4),

$$\Pi\left(n^{-1/2}\left(\boldsymbol{a}^{(n)}\right)^\top Y^{(n)}\right) \stackrel{d}{=} \Pi\left(n^{-1/2}\left(\boldsymbol{a}^{(n)}\right)^\top X^{(n)}\right) \times \rho\left(\frac{1}{n}\sum_{i=1}^n \eta(X_i)\right),$$

where $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous. By applying the contraction principle (see [65, Theorem 4.2.1]) to the continuous map $f : \mathcal{X} \times \mathbb{R}_+ \to \mathcal{X}$ given by $f(\boldsymbol{w}, r, s) := (\rho(r_2)\boldsymbol{x}, \rho(r_2)r_1)$, we obtain an LDP for $\{\Pi(n^{-1/2}\left(\boldsymbol{a}^{(n)}\right)^\top Y^{(n)})\}_{n\in\mathbb{N}}$ with speed n and good rate function

$$\mathcal{I}(\boldsymbol{y}, r) = \inf\{I(\boldsymbol{w}, r, s) : \boldsymbol{y} = \rho(r_2)\boldsymbol{x},\ \rho(r_2)r_1 = r\} = \inf_{r_2\in\mathbb{R}_+} I\left(\frac{\boldsymbol{y}}{\rho(r_2)}, \frac{r}{\rho(r_2)}, r_2\right).$$

□

Given Theorem 5.2.5, the proof of Theorem 5.1.2 closely follows [116, Section 5]. We include the details for completeness.

Proof of Theorem 5.1.2. We now verify that $Y^{(n)}$ satisfies Assumptions 1–4 when uniformly distributed on the ℓ_n^p sphere for some $p \in [2, \infty)$.

Assumption 1: By [149, Lemma 1], the relation (5.2.1) holds with $\{X_i\}_{i\in\mathbb{N}}$ being the i.i.d. sequence with common distribution equal to the p-Gaussian distribution (the probability measure on $\mathbb{R}$ with density proportional to $e^{-|y|^p/p}$), $\eta(x) = |x|^p$, and $\rho(y) = y^{-1/p}$.

Assumption 2: This is a direct consequence of the fact that X_1 is a p-Gaussian distribution, and hence sub-Gaussian for $p \in [2, \infty)$.

Assumption 3: It is straightforward to see that the domain of $\overline{\Lambda}_p$ is $\mathbb{R} \times (-\infty, \frac{1}{p})$. The essential smoothness is established in [96, Lemma 5.8].

Assumption 4: The growth conditions on $\overline{\Lambda}_p$ and the derivatives of $\overline{\Lambda}_p$ are established in Lemma 5.8.2.

By Theorem 5.2.5, this proves the LDP when $Y^{(n)}$ is uniformly distributed on $\S_n^p$, and the corresponding rate function is $\mathcal{I}_p$ defined in (5.1.9).

Next, we consider the case where $Y^{(n)}$ is uniformly distributed on $\mathbb{B}_p^n$. Let U be a uniform random variable on $[0, 1]$. We recall that if the random vector $X^{(n)} \in \mathbb{R}^n$ is uniformly distributed on the sphere $\mathbb{S}_p^n$ and independent from U, then $U^{1/n}X^{(n)}$ is uniform on $\mathbb{B}_p^n$ [149, Lemma 1]. It then suffices to prove the LDP for $Y^{(n)} = U^{1/n}X^{(n)}$. By [96, Lemma 3.3], the sequence $\{U^{1/n}\}_{n\in\mathbb{N}}$ satisfies an LDP with good rate function I_U given by $I_U(u) = -\log u$ for $u \in (0, 1]$, and $I_U(u) = +\infty$ otherwise. Recalling that U and $X^{(n)}$ are independent, and using Theorem 5.2.5, we find that $\{(\Pi\left(n^{-1/2}(\boldsymbol{a}^{(n)})^\top X^{(n)}\right), U^{1/n})\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{X} \times \mathbb{R}_+$ with good rate function $I_U(u) + \mathcal{I}_p(\boldsymbol{x}, r)$. By the contraction principle applied to the continuous map $\mathcal{X} \times \mathbb{R}_+ \ni (\boldsymbol{x}, r, u) \mapsto (u\boldsymbol{x}, ur) \in \mathcal{X}$, and the positive homogeneity of the map Π defined in (5.1.4), it follows that $\left\{\Pi\left(U^{1/n}n^{-1/2}(\boldsymbol{a}^{(n)})^\top X^{(n)}\right)\right\}_{n\in\mathbb{N}}$ satisfies an

LDP on $\mathcal{X}$ with rate function

$$\begin{aligned}\widehat{I}_p(\boldsymbol{x},r) &= \inf_{(\boldsymbol{y},s)\in\mathcal{X},u\in\mathbb{R}} \{I_U(u)+\mathcal{I}_p(\boldsymbol{y},s) : (\boldsymbol{x},r)=(u\boldsymbol{y},us)\} \\ &= \inf_{u\in(0,1]} \left\{-\log u + \mathcal{I}_p\left(\frac{\boldsymbol{x}}{u},\frac{r}{u}\right)\right\}.\end{aligned} \tag{5.5.1}$$

It is clear from the definition (5.1.9) that the map $t\mapsto \mathcal{I}_p(t\boldsymbol{x},tr)$ is increasing for $t\in\mathbb{R}_+$. Since $u\mapsto u^{-1}$ is monotonically decreasing, we deduce that $u\mapsto \mathcal{I}_p(u^{-1}\boldsymbol{x},u^{-1}r)$ is decreasing. Combined with the fact that $-\log u$ is decreasing, we find that the infimum in (5.5.1) is attained at $u=1$. Thus

$$\widehat{I}_p(\boldsymbol{x},r)=\mathcal{I}_p(\boldsymbol{x},r), \quad (\boldsymbol{x},r)\in\mathcal{X},$$

as claimed. □

Proof of Corollary 5.1.3. It was shown in the proof of Theorem 5.1.2 that $Y^{(n)}$ satisfies the assumptions of Theorem 5.2.5 when uniformly distributed on $\mathbb{S}_n^p$ for some $p\in[2,\infty)$. By applying the contraction principle to Theorem 5.2.5 with the continuous projection map $\mathcal{X}\ni(\boldsymbol{y},r)\mapsto r\in\mathbb{R}_+$, we conclude that $\{\|n^{-1/2}\left(\boldsymbol{a}^{(n)}\right)^{\top} Y^{(n)}\|_2\}_{n\in\mathbb{N}}$ satisfies an LDP with speed n and good rate function

$$\mathbb{I}_p(r)=\inf\left\{\mathcal{I}_p(\boldsymbol{y},r) : (\boldsymbol{y},r)\in\mathcal{X}\right\}.$$

Substituting the definition of $\mathcal{I}_p$ from (5.1.9) and recalling the definition of I_p from (5.1.8), we have

$$\begin{aligned}\mathbb{I}_p(r) &= \inf\left\{I_p\left(\boldsymbol{w},r,s\right) : (\boldsymbol{w},r,s)\in\mathcal{X}\times\mathbb{R}_+,\ r_1|r_2|^{-1/p}=r\right\} \\ &= \inf\left\{I_p\left(\boldsymbol{x},r_1,\left(\frac{r_1}{r}\right)^p\right) : (\boldsymbol{x},r_1)\in\mathcal{X}\right\} \\ &= \inf\left\{\Lambda_p^*\left(\boldsymbol{x},\sqrt{r_1^2-\|\boldsymbol{x}\|_2^2},\left(\frac{r_1}{r}\right)^p\right) : (\boldsymbol{x},r_1)\in\mathcal{X}\right\}.\end{aligned} \tag{5.5.2}$$

By definitions of Λ_p and Λ_p^* from (5.1.6) and (5.1.7), respectively, and an elementary property of Gaussian variables, we see that

$$\begin{aligned}
&\Lambda_p^*\left(\boldsymbol{x}, \sqrt{r_1^2-\|\boldsymbol{x}\|_2^2}, \left(\frac{r_1}{r}\right)^p\right) \\
&= \sup_{(\boldsymbol{u},b,c)\in\ell^2\times\mathbb{R}\times\mathbb{R}} \left\{\langle \boldsymbol{u},\boldsymbol{x}\rangle + b\sqrt{r_1^2-\|\boldsymbol{x}\|_2^2} + c\left(\frac{r_1}{r}\right)^p - \Lambda_p\left(\boldsymbol{u},b,c\right)\right\} \\
&= \sup_{(\boldsymbol{u},b,c)\in\ell^2\times\mathbb{R}\times\mathbb{R}} \left\{\langle \boldsymbol{u},\boldsymbol{x}\rangle + b\sqrt{r_1^2-\|\boldsymbol{x}\|_2^2} + c\left(\frac{r_1}{r}\right)^p - \mathbb{E}\left[\overline{\Lambda}_p\left(\|\left(\boldsymbol{u},b\right)\|_2 g, c\right)\right]\right\}
\end{aligned} \tag{5.5.3}$$

By the Cauchy–Schwarz inequality and the fact that $\|\boldsymbol{x}\|_2 \le r_1$, we have $\langle \boldsymbol{u},\boldsymbol{x}\rangle + b\sqrt{r_1^2-\|\boldsymbol{x}\|_2^2} \le \|(\boldsymbol{u},b)\|_2 r_1$. Further, if $\boldsymbol{x}$ and r_1 are given, there exists a pair $(\boldsymbol{u},b)\in\ell^2\times\mathbb{R}$ such that equality is attained. Hence, (5.5.3) implies

$$\begin{aligned}
&\Lambda_p^*\left(\boldsymbol{x}, \sqrt{r_1^2-\|\boldsymbol{x}\|_2^2}, \left(\frac{r_1}{r}\right)^p\right) \\
&= \sup_{(\boldsymbol{u},b,c)\in\ell^2\times\mathbb{R}\times\mathbb{R}} \left\{\|(\boldsymbol{u},b)\|_2\, r_1 + c\left(\frac{r_1}{r}\right)^p - \mathbb{E}\left[\overline{\Lambda}_p\left(\|\left(\boldsymbol{u},b\right)\|_2 g, c\right)\right]\right\} \\
&= \sup_{(v,c)\in\mathbb{R}_+\times\mathbb{R}} \left\{v r_1 + c\left(\frac{r_1}{r}\right)^p - \mathbb{E}\left[\overline{\Lambda}_p\left(vg,c\right)\right]\right\} \\
&= \sup_{(v,c)\in\mathbb{R}\times\mathbb{R}} \left\{v r_1 + c\left(\frac{r_1}{r}\right)^p - \mathbb{E}\left[\overline{\Lambda}_p\left(vg,c\right)\right]\right\},
\end{aligned} \tag{5.5.4}$$

where the last equality follows from the observation that, since $r_1 \ge 0$ and g is symmetric, replacing v by $|v|$ increases the quantity we are taking the supremum over. Combining (5.5.2) and (5.5.4), we have

$$\mathbb{I}_p(r) = \inf_{\tau\in\mathbb{R}_+} \sup_{(v,c)\in\mathbb{R}\times\mathbb{R}} \left\{v\tau + c\left(\frac{\tau}{r}\right)^p - \mathbb{E}\left[\Lambda\left(vg,c\right)\right]\right\}.$$

By [129, Lemma 2.1], this implies

$$\mathbb{I}_p(r) = \sup_{(v,c)\in\mathbb{R}\times\mathbb{R}} \left\{vr + c - \mathbb{E}\left[\Lambda\left(vg,c\right)\right]\right\}.$$

The convexity of $\mathbb{I}$ now follows from a standard result in convex analysis (see [146, Theorem 11.1]). $\square$

Remark 5.5.1. The space $\mathcal{X}$ is defined in (5.1.2) using the ℓ_2 norm in one coordinate. This allowed us to deduce Corollary 5.1.3 from Theorem 5.1.2 by applying the contraction principle in the previous proof. The ℓ_2 norm in this definition could be replaced by ℓ_q for any $1 \leq q < \infty$, and a parallel argument would give the analogue of Corollary 5.1.3 for the ℓ_q norm of the projections. For brevity, we do not take this up here.

Proof of Corollary 5.1.4. Consider the map $\pi \colon \mathcal{X} \to \mathbb{R}$ given by $\pi(\boldsymbol{w}, r) = w_1$, the first element of the ordered sequence $\boldsymbol{w}$. By Theorem 5.2.5 and the contraction principle, the sequence (5.1.10) satisfies an LDP with good rate function

$$\mathcal{I}_{\max}(r) = \inf_{(\boldsymbol{w},s)\in\mathcal{X}} \{\mathcal{I}_p(\boldsymbol{w}, s) : r = \pi(\boldsymbol{w}, s)\} = \mathcal{I}_p((r, 0, 0 \ldots), r),$$

where the last inequality follows from the definition (5.1.9) and the fact that Λ is even in its first two arguments.

The proof for $\mathbb{B}_p^n$ is nearly identical, so we omit it. $\square$

5.6 Preliminary Lemmas: Upper Bound

5.6.1 Topological properties of $\mathcal{X}$

Proof of Lemma 5.3.1. Due to the ordering of the first component in $\mathcal{X}$, each $(\boldsymbol{w}, r) \in \mathcal{X}$ satisfies

$$|w_m| \leq m^{-1/2}\|\boldsymbol{w}\|_2 \leq m^{-1/2} r. \tag{5.6.1}$$

Proof of Claim (1). Suppose $(\boldsymbol{w}_n, r_n) \to (\boldsymbol{w}, r)$ in $\mathcal{X}$. Then, by the definition of the metric in (5.1.3), $r_n \to r$, and $\boldsymbol{w}_n \to \boldsymbol{w}$ in ℓ_∞. The latter implies that for any coordinate projection map $p_i : \boldsymbol{y} \mapsto y_i$, we have $p_i(\boldsymbol{w}_n - \boldsymbol{w}) \to 0$ as $n \to \infty$. Thus $(\boldsymbol{w}_n, r_n) \to (\boldsymbol{w}, r)$ in $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}_+$ when $\mathbb{R}^{\mathbb{N}}$ is equipped with the product topology.

On the other hand, suppose $(\boldsymbol{w}_n, r_n) \to (\boldsymbol{w}, r)$ in $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}_+$ when $\mathbb{R}^{\mathbb{N}}$ is equipped with the product topology. To show that $(\boldsymbol{w}_n, r_n) \to (\boldsymbol{w}, r)$ in $\mathcal{X}$, by (5.1.3) it suffices to show that $\|\boldsymbol{w}_n - \boldsymbol{w}\|_\infty \to 0$ as $n \to \infty$. Pick any $\varepsilon > 0$. Then there exists $N \in \mathbb{N}$ such that $\|(\boldsymbol{w}_n)_{\leq N} - \boldsymbol{w}_{\leq N}\|_\infty < \varepsilon$ and $N^{-1/2} < \varepsilon/(2r+1)$, and for all $n > N$, $r_n < r + 1$. Let $n > N$ be arbitrary. By (5.6.1), we have

$$\begin{aligned}
\|\boldsymbol{w}_n - \boldsymbol{w}\|_\infty &= \max\left(\|(\boldsymbol{w}_n)_{\leq N} - \boldsymbol{w}_{\leq N}\|_\infty, \|(\boldsymbol{w}_n)_{>N} - \boldsymbol{w}_{>N}\|_\infty\right) \\
&\leq \max\left(\|(\boldsymbol{w}_n)_{\leq N} - \boldsymbol{w}_{\leq N}\|_\infty, N^{-1/2} r_n + N^{-1/2} r\right) \\
&\leq \varepsilon.
\end{aligned}$$

Sending first $n \to \infty$, and then $\varepsilon \to 0$, we obtain the desired conclusion.

Proof of Claim (2). Suppose $(\boldsymbol{w}_n, r_n) \to (\boldsymbol{w}, r)$ in $\mathcal{X}$. Pick any $\boldsymbol{u} \in \ell^2$. It suffices to show that $\langle \boldsymbol{u}, \boldsymbol{w}_n - \boldsymbol{w}\rangle \to 0$. Without loss of generality, assume $\boldsymbol{u} \neq 0$. Fix $\varepsilon > 0$. Since $\boldsymbol{u} \in \ell^2$, there exists $m \in \mathbb{N}$ such that $\|\boldsymbol{u}_{>m}\|_2 < \varepsilon/(2r+1)$. Let $N = N(\varepsilon, m) \in \mathbb{N}$ be large enough so that $\|\boldsymbol{w}_n - \boldsymbol{w}\|_\infty < \varepsilon m^{-1/2}/\|\boldsymbol{u}\|_2$ and for all $n > N$, $r_n < r + 1$. Then applying the Cauchy–Schwarz and triangle inequalities, and invoking the identities $\|w_n\| = r$ and $\|w\|_2 = r$ (which follow from the definition of $\mathcal{X}$ in (5.1.2)), we conclude that for any $n > N$,

$$\begin{aligned}
\langle \boldsymbol{u}, \boldsymbol{w}_n - \boldsymbol{w}\rangle &= \left\langle \boldsymbol{u}_{\leq m}, (\boldsymbol{w}_n - \boldsymbol{w})_{\leq m}\right\rangle + \left\langle \boldsymbol{u}_{>m}, (\boldsymbol{w}_n - \boldsymbol{w})_{>m}\right\rangle \\
&\leq \|\boldsymbol{u}\|_2 \|\left(\boldsymbol{w}_n - \boldsymbol{w}\right)_{\leq m}\|_2 + \|\boldsymbol{u}_{>m}\|_2 \|\left(\boldsymbol{w}_n - \boldsymbol{w}\right)\|_2 \\
&\leq \|\boldsymbol{u}\|_2 \|\left(\boldsymbol{w}_n - \boldsymbol{w}\right)_{\leq m}\|_\infty m^{1/2} + \|\boldsymbol{u}_{>m}\|_2 \left(r_n + r\right) \\
&\leq 2\varepsilon.
\end{aligned}$$

Since ε is arbitrary, this shows $\langle \boldsymbol{u}, \boldsymbol{w}_n - \boldsymbol{w}\rangle = 0$, as desired.

On the other hand, suppose $(\boldsymbol{w}_n, r_n) \to (\boldsymbol{w}, r)$ in $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}_+$ when $\mathbb{R}^{\mathbb{N}}$ is equipped with the weak-ℓ^2 topology. Pick any $\varepsilon > 0$. Choose $m \in \mathbb{N}$ large enough such that $m^{-1/2} < \varepsilon/(2r+1)$. Choose $N \in \mathbb{N}$ large enough such that for all $n > N$, $\|(\boldsymbol{w}_n)_{\leq m} - (\boldsymbol{w})_{\leq m}\|_2 < \varepsilon$

and $r_n < r+1$. Together with (5.6.1), this implies that for any $n > N$,

$$\begin{aligned}\|\boldsymbol{w}_n - \boldsymbol{w}\|_\infty &= \max\left(\left\|(\boldsymbol{w}_n)_{\le m} - \boldsymbol{w}_{\le m}\right\|_\infty, \left\|(\boldsymbol{w}_n)_{>m} - \boldsymbol{w}_{>m}\right\|_\infty\right)\\ &\le \max\left(\left\|(\boldsymbol{w}_n)_{\le m} - \boldsymbol{w}_{\le m}\right\|_2, m^{-1/2}\left(r_n + r\right)\right)\\ &< 2\varepsilon.\end{aligned}$$

Since ε is arbitrary, by (5.1.3) this implies $(\boldsymbol{w}_n, r_n) \to (\boldsymbol{w}, r)$ in $\mathcal{X}$.

Proof of Claim (3). By definition, for all $(\boldsymbol{w}, r) \in \mathcal{X}$, $|w_i| \le \|\boldsymbol{w}\|_2 \le r$. Therefore, $\{(\boldsymbol{w}, r) \in \mathcal{X} : r \le A\}$ is a closed subset of $\prod_{i\in\mathbb{N}}[-A, A]\times[0, A]$. By item (1), the topology of $\mathcal{X}$ is equivalent to $\mathbb{R}^{\mathbb{N}}\times\mathbb{R}_+$ where $\mathbb{R}^{\mathbb{N}}$ is equipped with the product topology, $\prod_{i\in\mathbb{N}}[-A, A]\times[0, A]$ is compact by Tychonoff's Theorem. Thus, as a closed subset of $\prod_{i\in\mathbb{N}}[-A, A]\times[0, A]$, the set $\{(\boldsymbol{w}, r) \in \mathcal{X} : r \le A\}$ is compact. □

5.6.2 Rate Function

Proof of Lemma 5.3.2. Step 1. Fix $m \in \mathbb{N}$, and recall from Section 5.4.1 that Λ_m^* denotes the Legendre transform of the function Λ_m defined in (5.3.3). We first show that the following inequality holds for all $(\boldsymbol{w}, r, s) \in \mathcal{X}$:

$$I(\boldsymbol{w}, r, s) \ge \Lambda_m^*\left(\boldsymbol{w}_{\le}, \sqrt{r^2 - \|\boldsymbol{w}_{\le}\|_2^2}, s\right). \tag{5.6.2}$$

If $\boldsymbol{w}_> = \mathbf{0}$, then $\Lambda(\boldsymbol{w}, r, s) = \Lambda_m(\boldsymbol{w}_{\le}, r, s)$, and for any $\boldsymbol{u} \in \ell^2$, $\langle \boldsymbol{u}, \boldsymbol{w}\rangle = \beta$ if and only if $\langle \boldsymbol{u}_{\le}, \boldsymbol{w}_{\le}\rangle = \beta$. It follows that $\Lambda^*(\boldsymbol{w}, b, c) = \Lambda_m^*(\boldsymbol{w}_{\le}, b, c)$ for all $b, c \in \mathbb{R}_+$. Hence, the definition of (5.1.8) of I shows that the inequality in (5.6.2) becomes an equality if $\boldsymbol{w}_> = \mathbf{0}$.

We henceforth suppose that $\|w_>\|_2 > 0$, which then implies $r > \|\boldsymbol{w}_{\le}\|_2$. Pick any $\hat{\boldsymbol{u}} \in \mathbb{R}^m$ and $b \ge 0$. Let $\boldsymbol{u} := \left(\hat{\boldsymbol{u}}, b\boldsymbol{w}_>/\sqrt{r^2 - \|\boldsymbol{w}_{\le}\|_2^2}\right)$. Then by the definition of Λ^* in

(5.1.7), we have for any $c \in \mathbb{R}$ that

$$\begin{aligned}
I\left(\boldsymbol{w}, r, s\right) &= \Lambda^*\left(\boldsymbol{w}, \sqrt{r^2 - \|\boldsymbol{w}\|_2^2}, s\right) \\
&\geq \langle \boldsymbol{u}, \boldsymbol{w}\rangle + \left(\frac{b\sqrt{r^2 - \|\boldsymbol{w}\|_2^2}}{\sqrt{r^2 - \|\boldsymbol{w}_{\leq}\|_2^2}}\right)\sqrt{r^2 - \|\boldsymbol{w}\|_2^2} + cs - \Lambda\left(\boldsymbol{u}, \frac{b\sqrt{r^2 - \|\boldsymbol{w}\|_2^2}}{\sqrt{r^2 - \|\boldsymbol{w}_{\leq}\|_2^2}}, c\right) \\
&= \langle \hat{\boldsymbol{u}}, \boldsymbol{w}_{\leq}\rangle + \frac{b\|\boldsymbol{w}_{>}\|_2^2}{\sqrt{r^2 - \|\boldsymbol{w}_{\leq}\|_2^2}} + \left(\frac{b\sqrt{r^2 - \|\boldsymbol{w}\|_2^2}}{\sqrt{r^2 - \|\boldsymbol{w}_{\leq}\|_2^2}}\right)\sqrt{r^2 - \|\boldsymbol{w}\|_2^2} + cs \\
&\quad - \Lambda\left(\boldsymbol{u}, \frac{b\sqrt{r^2 - \|\boldsymbol{w}\|_2^2}}{\sqrt{r^2 - \|\boldsymbol{w}_{\leq}\|_2^2}}, c\right) \\
&= \langle \hat{\boldsymbol{u}}, \boldsymbol{w}_{\leq}\rangle + \frac{b(\|\boldsymbol{w}_{>}\|_2^2 + r^2 - \|\boldsymbol{w}\|_2^2)}{\sqrt{r^2 - \|\boldsymbol{w}_{\leq}\|_2^2}} + cs - \Lambda\left(\boldsymbol{u}, \frac{b\sqrt{r^2 - \|\boldsymbol{w}\|_2^2}}{\sqrt{r^2 - \|\boldsymbol{w}_{\leq}\|_2^2}}, c\right) \\
&= \langle \hat{\boldsymbol{u}}, \boldsymbol{w}_{\leq}\rangle + b\sqrt{r^2 - \|\boldsymbol{w}_{\leq}\|_2^2} + cs - \Lambda_m\left(\hat{\boldsymbol{u}}, b, c\right),
\end{aligned}$$

where we used $\Lambda(\boldsymbol{u}, b', c) = \Lambda_m(\boldsymbol{u}_{\leq}, \sqrt{(b')^2 + \|\boldsymbol{u}_{>}\|_2^2}, c)$ in the last step. Taking the supremum of the right-hand side over $\boldsymbol{w} \in \mathbb{R}^m$ and $b \in \mathbb{R}_+$, we obtain (5.6.2).

Step 2. We next claim that I is lower semicontinuous.

First note that Λ_m^* is lower semicontinuous, since it is the supremum of a collection of continuous functions. By claim (2) of Lemma 5.3.1, the topology on $\mathcal{X}$ is equivalent to the topology on $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}_+$, where $\mathbb{R}^{\mathbb{N}}$ is equipped with the weak-ℓ^2 topology. Therefore Λ^*, which is equal to the supremum of a collection of continuous functions, is lower semicontinuous. Let $(\boldsymbol{w}^{(n)}, r^{(n)}, s^{(n)})_{n\in\mathbb{N}}$ be a sequence converging to $(\boldsymbol{w}, r, s)$ in $\mathcal{X} \times \mathbb{R}_+$. Applying (5.6.2), the lower semicontinuity of Λ_m^*, and (5.3.3), we obtain

$$\begin{aligned}
\liminf_{n\to\infty} I\left(\boldsymbol{w}^{(n)}, r^{(n)}, s^{(n)}\right) &\geq \liminf_{n\to\infty} \Lambda_m^*\left(\boldsymbol{w}_{\leq}^{(n)}, \sqrt{\left(r^{(n)}\right)^2 - \left\|\boldsymbol{w}_{\leq}^{(n)}\right\|_2^2}, s^{(n)}\right) \\
&\geq \Lambda_m^*\left(\boldsymbol{w}_{\leq}, \sqrt{r^2 - \|\boldsymbol{w}_{\leq}\|_2^2}, s\right) \\
&= \Lambda^*\left(\left(\boldsymbol{w}_{\leq}, \boldsymbol{0}\right), \sqrt{r^2 - \|\boldsymbol{w}_{\leq}\|_2^2}, s\right).
\end{aligned}$$

Recalling that $\boldsymbol{w}_{\leq}$ is the abbreviation for $\boldsymbol{w}_{\leq m}$, passing to the $m \to \infty$ limit in the last display, and using the lower semicontinuity of Λ^*, the lower semicontinuity of I follows.

Step 3. We claim $\{I \leq \alpha\}$ is compact for any $\alpha \geq 0$.

Since I is lower semicontinuous, by Step 2, $\{I \leq \alpha\}$ is closed. We will show that

$$\{(\boldsymbol{w}, r, s) \in \mathcal{X} \times \mathbb{R} : I(\boldsymbol{w}, r, s) \leq \alpha\} \subset \{(\boldsymbol{w}, r) \in \mathcal{X} : r \leq A\} \times [0, B]. \tag{5.6.3}$$

for some $A, B > 0$. This suffices to prove the claim, because the latter set is compact by Lemma 5.3.1, and closed subsets of compact sets are compact. To show (5.6.3), we argue by contradiction. Fix $\alpha \geq 0$, and suppose that there exists a sequence

$$(\boldsymbol{w}^{(n)}, r^{(n)}, s^{(n)}) \in \{I \leq \alpha\}, \qquad n \in \mathbb{N} \tag{5.6.4}$$

such that $r^{(n)} \to \infty$ or $s^{(n)} \to \infty$. Fix $0 < c < T$, where $\mathbb{R} \times (-\infty, T)$ is the domain of Λ from Assumption 4. By Assumption 5 on $\overline{\Lambda}$, and the definition of Λ_m in (5.3.3) and (5.1.6), it follows that there exists an absolute constnat $C \in (0, \infty)$ such that for all $\hat{\boldsymbol{u}} = (u_1, \ldots, u_m) \in \mathbb{R}^m$ and $b \geq 0$,

$$\Lambda_m\left(\hat{\boldsymbol{u}}, b, 0\right) = \mathbb{E}\left[\overline{\Lambda}\left(\sum_{i=1}^m u_i g_i + b g_0, 0\right)\right] \leq C\left(1 + \|(\hat{\boldsymbol{u}}, b)\|_2^2\right). \tag{5.6.5}$$

Suppose first that $r^{(n)} \to \infty$. By (5.6.2) and (5.6.5), it follows that

$$\begin{aligned}
I\left(\boldsymbol{w}^{(n)}, r^{(n)}, s^{(n)}\right) &\geq \Lambda_m^*\left(\boldsymbol{w}_{\leq}^{(n)}, \sqrt{\left(r^{(n)}\right)^2 - \|\boldsymbol{w}_{\leq}^{(n)}\|_2^2}, s^{(n)}\right) \\
&= \sup_{(\hat{\boldsymbol{u}}, b, c) \in \mathcal{X} \times \mathbb{R}_+} \left\{\langle \hat{\boldsymbol{u}}, \boldsymbol{w}_{\leq}^{(n)}\rangle + b\sqrt{\left(r^{(n)}\right)^2 - \|\boldsymbol{w}_{\leq}^{(n)}\|_2^2} + c s^{(n)} - \Lambda_m\left(\hat{\boldsymbol{u}}, b, c\right)\right\} \\
&\geq \sup_{(\hat{\boldsymbol{u}}, b) \in \mathcal{X}} \left\{\langle \hat{\boldsymbol{u}}, \boldsymbol{w}_{\leq}^{(n)}\rangle + b\sqrt{\left(r^{(n)}\right)^2 - \|\boldsymbol{w}_{\leq}^{(n)}\|_2^2} - C\left(1 + \|(\hat{\boldsymbol{u}}, b)\|_2^2\right)\right\}.
\end{aligned}$$

Using the Cauchy–Schwarz inequality, we note that

$$\langle \hat{\boldsymbol{u}}, \boldsymbol{w}_{\leq}^{(n)}\rangle + b\sqrt{\left(r^{(n)}\right)^2 - \|\boldsymbol{w}_{\leq}^{(n)}\|_2^2} \leq r^{(n)}\|(\hat{\boldsymbol{u}}, b)\|_2,$$

and equality holds when $(\hat{\boldsymbol{u}}, b) = \alpha(\boldsymbol{w}^{(n)}_{\leq}, \sqrt{\left(r^{(n)}\right)^2 - \|\boldsymbol{w}^{(n)}_{\leq}\|_2^2})$ for some $\alpha \in \mathbb{R}$. Therefore

$$\begin{aligned}
&\sup_{(\hat{\boldsymbol{u}},b)\in\mathcal{X}} \left\{ \langle \hat{\boldsymbol{u}}, \boldsymbol{w}^{(n)}_{\leq} \rangle + b\sqrt{\left(r^{(n)}\right)^2 - \|\boldsymbol{w}^{(n)}_{\leq}\|_2^2} - C\left(1 + \|(\hat{\boldsymbol{u}}, b)\|_2^2\right) \right\} \\
&= \sup_{(\hat{\boldsymbol{u}},b)\in\mathcal{X}} \left\{ r^{(n)} \|(\hat{\boldsymbol{u}}, b)\|_2 - C\|(\hat{\boldsymbol{u}}, b)\|_2^2 - C \right\} = \frac{\left(r^{(n)}\right)^2}{4C} - C,
\end{aligned}$$

where to get the last equality, we compute the supremum directly as a function of $\|(\hat{\boldsymbol{u}}, b)\|_2$. The last three displays together show that

$$I(\boldsymbol{w}^{(n)}, r^{(n)}, s^{(n)}) \geq \frac{(r^{(n)})^2}{4C} - C,$$

and this lower bound goes to infinity as $n \to \infty$, due to the assumption that $r^{(n)} \to \infty$. However, this contradicts the assumption (5.6.4).

Next, suppose $s^{(n)} \to \infty$. By (5.6.2) and (5.3.3),

$$I\left(\boldsymbol{w}^{(n)}, r^{(n)}, s^{(n)}\right) \geq \Lambda^*_m \left(\boldsymbol{w}^{(n)}_{\leq}, \sqrt{\left(r^{(n)}\right)^2 - \|\boldsymbol{w}^{(n)}_{\leq}\|_2^2}, s^{(n)}\right) \geq \sup_{t\in\mathbb{R}} \left\{ ts^{(n)} - \Lambda(\mathbf{0}, 0, t) \right\}.$$

The supremum on the right side of the previous display is infinite, by (5.1.6), (5.1.5), and (5). Again, this contradicts the assumption that $(\boldsymbol{w}^{(n)}, r^{(n)}, s^{(n)}) \in \{I \leq \alpha\}$. The proof is complete. $\square$

5.6.3 Exponential Tightness

Our proof will appeal to the Hanson–Wright concentration inequality for quadratic forms. We recall the following definition from [164, Section 2.5].

Definition 5.6.1. *A random variable X is said to be sub-Gaussian if there exists a constant $C > 0$ such that*

$$\mathbb{P}(|X| \geq x) \leq 2\exp\left(-\frac{x^2}{C^2}\right).$$

In this case, the sub-Gaussian norm of X is defined as

$$\|X\|_{\psi_2} \coloneqq \inf\Big\{t > 0 : \mathbb{E}\Big[\exp(X^2/t^2)\Big] \le 2\Big\}.$$

Moreover, a random vector $X \in \mathbb{R}^n$ is called sub-Gaussian if all one dimensional marginals $\langle X, \boldsymbol{x}\rangle$, $\boldsymbol{x} \in \mathbb{R}^n$, are sub-Gaussian random variables. The sub-Gaussian norm of X is defined as

$$\|X\|_{\psi_2} \coloneqq \sup_{\boldsymbol{x}\in\mathbb{S}^{n-1}} \|\langle X, \boldsymbol{x}\rangle\|_{\psi_2}. \tag{5.6.6}$$

Theorem 5.6.2 ([164, Theorem 6.2.1])**.** *There exists a constant $\gamma_0 > 0$ such that the following holds. Let $X = (X_1, \ldots, X_n)$ be a random vector with i.i.d., mean zero, sub-Gaussian entries. Let A be an $n \times n$ matrix. Then for every $t \ge 0$,*

$$\mathbb{P}\left(\left|X^\top AX - \mathbb{E}[X^\top AX]\right| \ge t\right) \le 2\exp\left[-\gamma_0 \min\left(\frac{t^2}{R^4\|A\|_F^2}, \frac{t}{R^2\|A\|_2}\right)\right],$$

where $R \coloneqq \|X\|_{\psi_2}$ is the sub-Gaussian norm of X_1 defined in Definition 5.6.1, and $\|A\|_2$ denotes the Frobenius norm of A.

Proof of Lemma 5.3.4. For $n \in \mathbb{N}$, we set

$$A^{(n)} = \frac{1}{n}\boldsymbol{a}^{(n)}(\boldsymbol{a}^{(n)})^\top, \quad Q^{(n)} = \frac{1}{n}(X^{(n)})^\top \boldsymbol{a}^{(n)}(\boldsymbol{a}^{(n)})^\top X^{(n)} = \frac{1}{n}\sum_{i,j=1}^{n} A_{ij}^{(n)} X_i X_j. \tag{5.6.7}$$

For the remainder of this proof, we suppress the dependence on n in the notation by omitting all superscripts. Observe that

$$\mathbb{E}\left[X^\top AX\right] = \frac{1}{n}\sum_{i=1}^{n} A_{ii}\mathbb{E}[X_i^2] = \frac{1}{n}\sum_{\ell}^{k_n}\sum_{i=1}^{n} a_{i\ell}^2 = \frac{k_n}{n}. \tag{5.6.8}$$

We write the matrix A as a sum of rank one matrices, $A = n^{-1}\sum_{\ell=1}^{k_n} Y_\ell Y_\ell^\top$, where $Y_\ell = [a_{1\ell}, \ldots, a_{n\ell}]^\top$. Since $\boldsymbol{a} \in \mathbb{V}_{n,k_n}$, the vectors $\{Y_\ell\}_{1\le\ell\le k_n}$ are orthonormal. This implies

$$\begin{aligned}
\|A\|_F^2 &= \operatorname{Tr} AA^\top \\
&= n^{-2} \operatorname{Tr} \left(\sum_{\ell=1}^{k_n} Y_\ell Y_\ell^\top \right) \left(\sum_{\ell=1}^{k_n} Y_\ell Y_\ell^\top \right)^\top \\
&= n^{-2} \operatorname{Tr} \sum_{\ell=1}^{k_n} Y_\ell Y_\ell^\top \\
&= n^{-2} \operatorname{Tr} A \\
&= n^{-2} k_n.
\end{aligned} \tag{5.6.9}$$

Further, we also have

$$\begin{aligned}
\|A\|_2^2 &= \max_{\|v\|=1} \|Av\|_2^2 \\
&= \max_{\|v\|=1} v^\top A^\top A v \\
&= \max_{\|v\|=1} n^{-2} \sum_{\ell=1}^{k_n} v^\top Y_\ell Y_\ell^\top v \\
&= \max_{\|v\|=1} n^{-2} \langle Y_\ell, v \rangle^2.
\end{aligned} \tag{5.6.10}$$

The last quantity is equal to n^{-2} because the Y_ℓ are unit vectors. Also, note that by Definition 5.6.1, $\|X\|_{\psi_2} = C_2$, where C_2 is the constant in Assumption 2. By applying Theorem 5.6.2 to the quantity Q from (5.6.7) (and hence with $A = A^{(n)}$ and $R = C_2$), and using (5.6.8), (5.6.9), and (5.6.10), we find that for any $\boldsymbol{a} \in \mathbb{V}_{n,k_n}$,

$$\mathbb{P}\left(\left| \|n^{-1/2} \boldsymbol{a}^\top X^{(n)}\|_2^2 - \frac{k_n}{n} \right| \geq t \right) \leq 2 \exp\left[-\gamma_0 \min\left(k_n^{-1} C_2^{-4} (nt)^2, C_2^{-2} nt \right) \right],$$

where $\gamma_0 > 0$ is the constant from Theorem 5.6.2. This implies the conclusion after noting that since $k_n/n \leq 1$ for $n \in \mathbb{N}$, for $t \geq C^2$, we have

$$\min\left(k_n^{-1} (C_2^{-2} nt)^2, C_2^{-2} nt \right) = C_2^{-2} nt.$$

$\square$

Proof of Lemma 5.3.5. Pick any $M > 0$ and fix $\lambda > 0$ such that $\overline{\Lambda}(0, \lambda) < \infty$. We note that such a λ exists by Assumption 4. Fix $\alpha > 0$ satisfying $\alpha > \big(\overline{\Lambda}(0, \lambda) + M\big)/\lambda$. By Markov's inequality, we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\eta\left(X_i\right) \geq \alpha\right) \leq e^{-n\lambda\alpha}\left(\mathbb{E}\Big[\exp\Big(\lambda\eta(X_1)\Big)\Big]\right)^n.$$

This implies that

$$\frac{1}{n}\log\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\eta\left(X_i\right) \geq \alpha\right) \leq -\lambda\alpha + \mathbb{E}\Big[\exp\Big(\lambda\eta(X_1)\Big)\Big] < -M,$$

which proves the exponential tightness of $(\frac{1}{n}\sum_{i=1}^{n}\eta(X_i))_{n\in\mathbb{N}}$. $\square$

5.6.4 Preliminary Lemmas for Gaussian Approximation

Let μ be the standard Gaussian measure on $\mathbb{R}$, corresponding to a Gaussian random variable with mean zero and variance one. Fix $k, n \in \mathbb{N}$ with $k \leq n$, and $\boldsymbol{a} = \boldsymbol{a}^{(n)} \in \mathbb{V}_{n,k}$. Let $\boldsymbol{a}_i$ be the i-th row of $\boldsymbol{a}$. For any choice of $\boldsymbol{u} \in \mathbb{R}^k$, let $\mu_{n,\boldsymbol{u}}$ be the measure on $\mathbb{R}$ given by

$$\mu_n \equiv \mu_{n,\boldsymbol{u}} := \frac{1}{n}\sum_{i=1}^{n}\delta_{\langle \boldsymbol{u}, \sqrt{n}\boldsymbol{a}_i\rangle}. \tag{5.6.11}$$

Lemma 5.6.3. *Fix $k, n \in \mathbb{N}$ with $k \leq n$, and $\boldsymbol{u} \in \mathbb{R}^k$. Let $\boldsymbol{a}$ be a random element of $\mathbb{V}_{n,k}$ distributed according to the Haar measure σ_n. Then for every $m \in \mathbb{N}$, the quantity μ_n from (5.6.11) satisfies*

$$\mathbb{E}\left[\int x^m \, d\mu_{n,\boldsymbol{u}}\right] = \|\boldsymbol{u}\|_2^m\left(\int x^m \, d\mu\right)(1 + T_{m,n}), \qquad |T_{m,n}| \leq \frac{C_m}{n},$$

for some $T_{m,n} \in \mathbb{R}$ and $C_m > 0$, where C_m depends only on m (and not k or n).

Proof. We just treat the even m case. The odd m case is clear by symmetry, as the moments on the left and right side of the equality both vanish. By exchangeability of the rows of $\boldsymbol{a}$,

$$\mathbb{E}\left[\int x^m \, d\mu_{n,\boldsymbol{u}}\right] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}\left(\sum_{j=1}^{k} u_j(\sqrt{n}a_{ij}^{(n)})\right)^m\right] = \mathbb{E}\left[\left(\sum_{j=1}^{k} u_j(\sqrt{n}a_{1j})\right)^m\right].$$

We recognize this quantity as

$$\mathbb{E}\left[\langle \boldsymbol{u}, \sqrt{n}\boldsymbol{a}_1\rangle^m\right] = \|\boldsymbol{u}\|_2^m \cdot \mathbb{E}\left[(\sqrt{n}\boldsymbol{a}_{11})^m\right].$$

where the equality is justified by the orthogonal invariance of μ_n. By Wick's theorem (Theorem 5.9.5) and Lemma 5.9.6,

$$\mathbb{E}\left[(\sqrt{n}\boldsymbol{a}_{11})^m\right] = \left(\int x^m \, d\mu\right)(1 + T_{m,n}), \qquad |T_{m,n}| \leq \frac{C_m}{n}.$$

This completes the proof.

□

The next lemma is a minor adaptation of a modified version of Gromov's concentration inequality for Haar measure on the special orthogonal group $SO_n := \{\boldsymbol{a}^{(n)} \in O_n : \det(\boldsymbol{a}^{(n)}) = 1\}$ (presented in [11, Corollary 4.4.28]), tailored to the Haar measure $\sigma_n = \sigma_{n,k}$ on the Stiefel manifold $\mathbb{V}_{n,k}$.

Lemma 5.6.4. *Fix $k, n \in \mathbb{N}$ with $k < n$. Let $f\colon \mathbb{V}_{n,k} \to \mathbb{R}$ be a Lipshitz function with Lipschitz constant L in the Hilbert–Schmidt norm, meaning*

$$|f(\boldsymbol{a}^{(n)}) - f(\boldsymbol{b}^{(n)})| \leq L\|\boldsymbol{a}^{(n)} - \boldsymbol{b}^{(n)}\|_{\mathrm{HS}}$$

for all $\boldsymbol{a}^{(n)}, \boldsymbol{b}^{(n)} \in \mathbb{V}_{n,k}$. Then there is a constant $C > 0$ such that for all $\delta \geq 0$,

$$\sigma_n\left(|f(\boldsymbol{a}^{(n)}) - \mathbb{E}_n[f(\boldsymbol{a}^{(n)})] \geq \delta\right) \leq 2\exp\left(-\frac{C\delta^2 n}{L^2}\right).$$

Proof. We suppress the superscript (n) in the proof for brevity. Fix $k \in \mathbb{N}$, and let $\pi = \pi_k : O_n \to \mathbb{V}_{n,k}$ be the canonical projection map, i.e. $\pi : \boldsymbol{a} = (a_{ij})_{1\leq i,j\leq n} \mapsto (a_{ij})_{1\leq i\leq n, 1\leq j\leq k}$, for all $\boldsymbol{a} \in O_n$, let $\bar{\sigma}_n$ be the Haar measure on O_n and let $\bar{\mathbb{E}}_n$ be the corresponding expectation. Then $\sigma_n = \sigma_{n,k}$ is the pushforward of $\bar{\sigma}_n$ under π. Finally, set $\bar{f} = f \circ \pi$ and let $H_n \subset O_n$ be the cyclic group generated by

$$\boldsymbol{h} = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & -1 \end{pmatrix},$$

that is, H_n consists of the identity matrix and $\boldsymbol{h}$. Observe that $O_n = \{\boldsymbol{b}_1\boldsymbol{b}_2 : \boldsymbol{b}_1 \in SO_n, \boldsymbol{b}_2 \in H_n\}$. Fix $\boldsymbol{c} \in O_n$. Then there exists $\boldsymbol{c}_1 \in SO_n$ and $\boldsymbol{c}_2 \in H_n$ such that $\boldsymbol{c} = \boldsymbol{c}_1\boldsymbol{c}_2$. Let $\boldsymbol{a}$ be uniformly distributed on O_n with respect to Haar measure. Using right-invariance of Haar measure in the second equality and combining the definition of $\bar{f}$, the fact that $k < n$ and $\boldsymbol{c}_2 \in H_n$ in the third equality below, we have

$$\bar{\mathbb{E}}_n\left[\bar{f}(\boldsymbol{ac})\right] = \bar{\mathbb{E}}_n\left[\bar{f}\left(\boldsymbol{ac}_1\boldsymbol{c}_2\right)\right] = \bar{\mathbb{E}}_n\left[\bar{f}\left(\boldsymbol{ac}_2\right)\right] = \bar{\mathbb{E}}_n\left[\bar{f}\left(\boldsymbol{a}\right)\right].$$

This implies that the function $O_n \ni \boldsymbol{c} \mapsto \bar{\mathbb{E}}_n[\bar{f}(\boldsymbol{ac})]$ is a constant over O_n equal to $\bar{\mathbb{E}}_n[\bar{f}(\boldsymbol{a})]$. This observation, together with [11, Corollary 4.4.28] and the definitions of $\bar{\sigma}_n$ and $\bar{f}$, completes the proof. □

5.6.5 Proof of Gaussian Approximation

Proof of Lemma 5.3.6. Fix $c \in (-\infty, T)$ and a sequence $\{d_n\}_{n\in\mathbb{N}}$ as in the lemma statement. To rewrite F_n in (5.3.1) more succinctly, set $\Lambda_1(t) = \overline{\Lambda}(t, c)$ with $\overline{\Lambda}$ as in (5.1.5), and define the maps $f^{(n,j)}:\mathbb{V}_{n,k_n}\to\mathbb{R}$

$$f^{(n,j)}(\boldsymbol{a}^{(n)}) := \frac{1}{n}\sum_{i=1}^{n}\Lambda_1\left(\left\langle \boldsymbol{v}^{(n,j)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\right\rangle\right), \quad j=1,2,\dots,d_n. \tag{5.6.12}$$

A direct calculation shows that, for $i=1,\dots,n$ and $l=1,\dots,k_n$,

$$\partial_{il}f^{(n,j)}(\boldsymbol{a}^{(n)}) = \frac{1}{\sqrt{n}}\Lambda_1'\left(\left\langle \boldsymbol{v}^{(n,j)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\right\rangle\right)v_l^{(n,j)},$$

where $\partial_{il}f^{(n,j)}(\boldsymbol{a}^{(n)})$ represents the derivative of $f^{(n,j)}$ with respect to $\boldsymbol{a}_{il}^{(n)}$. Hence, by Assumption 5, there exists $\overline{C}\equiv\overline{C}(c)<\infty$ such that

$$\begin{aligned}
\left\|\nabla f^{(n,j)}(\boldsymbol{a}^{(n)})\right\|_2^2 =& \sum_{i=1}^{n}\sum_{l=1}^{k_n}\left(\partial_{il}f^{(n,j)}(\boldsymbol{a}^{(n)})\right)^2 \\
=& \frac{1}{n}\sum_{i=1}^{n}\left(\Lambda_1'\left(\left\langle \boldsymbol{v}^{(n,j)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\right\rangle\right)\right)^2\sum_{l=1}^{k_n}\left(v_l^{(n,j)}\right)^2 \\
\le& \frac{\left\|\boldsymbol{v}^{(n,j)}\right\|_2^2}{n}\sum_{i=1}^{n}C\left(1+\left\langle \boldsymbol{v}^{(n,j)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\right\rangle^2\right) \\
=& \overline{C}\|\boldsymbol{v}^{(n,j)}\|_2^2\left(1+\left\|\boldsymbol{v}^{(n,j)}\right\|_2^2\right) \\
=& \overline{C}D^2\left(1+D^2\right).
\end{aligned} \tag{5.6.13}$$

Equation (5.6.13) implies that $f^{(n,j)}$ is Lipshitz continuous in the Hilbert–Schmidt topology with Lipshitz constant $\sqrt{CD^2\left(1+D^2\right)}$. By Lemma 5.6.4 and union bound, it follows that there exists a constant $C'>0$ such that for any $\varepsilon>0$,

$$\sigma_n\left(\max_{1\le j\le d_n}\left|f^{(n,j)}(\boldsymbol{a}^{(n)})-\mathbb{E}_n\left[f^{(n,j)}(\boldsymbol{a}^{(n)})\right]\right|>\varepsilon\right)\le 2d_n\exp\left(-\frac{C'\varepsilon^2 n}{D^2(1+D^2)}\right) \tag{5.6.14}$$

Recalling $d_n=\exp(o(n))$, the inequality in (5.6.14) and the Borel–Cantelli lemma together imply that for σ-a.e. $\boldsymbol{a}=\left(\boldsymbol{a}^{(1)},\boldsymbol{a}^{(2)},\dots\right)\in\mathbb{V}$,

$$\lim_{n\to\infty}\max_{1\le j\le d_n}\left|f^{(n,j)}(\boldsymbol{a}^{(n)})-\mathbb{E}_n\left[f^{(n,j)}(\boldsymbol{a}^{(n)})\right]\right|=0. \tag{5.6.15}$$

Next, consider the random measures

$$\mu^{(n,j)} := \frac{1}{n}\sum_{i=1}^{n} \delta_{\langle \boldsymbol{v}^{(n,j)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\rangle}, \qquad j = 1, \dots, d_n.$$

By the definition of $f^{(n,j)}$ in (5.6.12), we have

$$f^{(n,j)}\left(\boldsymbol{a}^{(n)}\right) = \int_{\mathbb{R}} \Lambda_1(t)\, d\mu^{(n,j)}(t). \tag{5.6.16}$$

By Lemma 5.6.3, the sequence

$$\left\{\mathbb{E}_1\left[\mu^{(1,1)}\right], \dots, \mathbb{E}_1\left[\mu^{(1,d_1)}\right], \mathbb{E}_2\left[\mu^{(2,1)}\right], \dots, \mathbb{E}_2\left[\mu^{(2,d_2)}\right], \dots, \mathbb{E}_n\left[\mu^{(n,1)}\right], \dots, \mathbb{E}_n\left[\mu^{(n,d_n)}\right], \dots\right\}$$

converges in the sense of moments to a Gaussian measure, with mean zero and variance D^2. Combined with [15, Lemma B.1] and [140, Theorem 1.7], this implies that for any $f : \mathbb{R} \to \mathbb{R}$ with at most polynomial growth,

$$\lim_{n\to\infty} \max_{1\le j\le d_n} \left| \mathbb{E}\left[\int f d\mu^{(n,j)}\right] - \mathbb{E}\left[f(Dg)\right] \right| = 0, \tag{5.6.17}$$

where g is a Gaussian random variable with zero mean and unit variance. By (5.6.16), (5.6.17) and the fact that Λ_1 can be bounded by a quadratic polynomial due to Assumption 5, it follows that

$$\lim_{n\to\infty} \max_{1\le j\le d_n} \left| \mathbb{E}\left[f^{(n,j)}\left(\boldsymbol{a}^{(n)}\right)\right] - \mathbb{E}\left[\Lambda_1(Dg)\right] \right| = 0. \tag{5.6.18}$$

To complete the proof, combine (5.6.15) and (5.6.18), and recall the definition of F_n in (5.3.1). $\square$

5.7 Preliminary Lemmas: Lower Bound

5.7.1 Notations and Conventions

We adopt the following notation for this section. Let T be the constant from Assumption 4 and let $m \in \mathbb{N}$ and $(\hat{\boldsymbol{u}}, b, c) \in \mathbb{R}^m \times \mathbb{R} \times (-\infty, T)$ be parameters. Recall that Λ_m was defined in (5.3.3), and set

$$(\hat{\boldsymbol{w}}, t, s) := \nabla \Lambda_m(\hat{\boldsymbol{u}}, b, c). \tag{5.7.1}$$

Let $\{k_n\}_{n\in\mathbb{N}}$ be an increasing sequence of positive integers such that $\lim_{n\to\infty} k_n = \infty$. Recall that $\sigma = \sigma_n$ denotes the Haar measure on $\mathbb{V}_{n,k_n}$. Define $m \in \mathbb{N}$ and $\boldsymbol{b} \in \mathbb{R}_+^M$ by

$$M = M(m) := \left\lceil (t+1)^2 m \right\rceil, \quad \boldsymbol{b} := \underbrace{\left(bM^{-1/2}, \dots, bM^{-1/2}\right)}_{M \text{ times}}, \tag{5.7.2}$$

and also define a sequence of vectors $\{\boldsymbol{u}^{(n)}\}_{n>m+M} \in \mathbb{R}^{k_n}$ as follows:

$$\boldsymbol{u}^{(n)} := (\hat{\boldsymbol{u}}, \boldsymbol{b}, \underbrace{0, 0, \dots, 0}_{k_n - m - M \text{ times}}). \tag{5.7.3}$$

For $j \in \{1, \dots, m+M\}$, define[3]

$$\nu_j^{(n)} := \frac{1}{n} \sum_{i=1}^{n} \delta\Big(\sqrt{n} a_{ij}^{(n)}, \langle \boldsymbol{u}^{(n)}, \sqrt{n} \boldsymbol{a}_i^{(n)} \rangle\Big). \tag{5.7.4}$$

Finally, define

$$\nu_j := \operatorname{Law}\left(g_j, \sum_{l=1}^{m+M} u_l g_l\right), \tag{5.7.5}$$

where $g_0, \dots, g_{m+M}$ are independent, mean zero, variance one Gaussian random variables.

[3] We sometimes write $\delta(x)$ instead of δ_x in this subsection for legibility. The difference is only notational.

5.7.2 Proof of Lemma 5.4.1

We begin with an auxiliary result that allows us to analyze open neighborhoods of a point in $\mathcal{X} \times \mathbb{R}$ by looking at open neighborhoods around the truncated version of the point in $\mathbb{R}^m \times \mathbb{R}_+^2$, for sufficiently large m. Recall our notation for a metric ball given in (5.4.1).

Lemma 5.7.1. *Fix an open set $O \subset \mathcal{X} \times \mathbb{R}_+$ and a point $(\boldsymbol{w}, r, s) \in O$. Then there exists $m_0 \in \mathbb{N}$ and $\rho' > 0$ such that the following holds for all $m \geq m_0$: If $(\boldsymbol{w}', r', s') \in \mathcal{X} \times \mathbb{R}_+$ and*

$$\left(\left(\boldsymbol{w}'_{\leq m}, \sqrt{(r')^2 - \|\boldsymbol{w}'_{\leq m}\|_2^2}\right), s'\right) \in B\left(\left(\boldsymbol{w}_{\leq m}, \sqrt{r^2 - \|\boldsymbol{w}\|_2^2}, s\right), \rho'\right), \tag{5.7.6}$$

then $(\boldsymbol{w}', r', s') \in O$.

Proof. Fix $(\boldsymbol{w}, r, s) \in O$. By Lemma 5.3.1(1), the topology on $\mathcal{X}$ is equivalent to the product topology on $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}_+ \times \mathbb{R}_+$. Therefore, using the definition of the product topology, there exists $m_0 \in \mathbb{N}$ and $\delta_1, \delta_2, \delta_3 > 0$ such that for every $m > m_0$, the following claim holds:

$$\begin{aligned} &\text{If } (\tilde{\boldsymbol{w}}, \tilde{r}, \tilde{s}) \in \mathcal{X} \times \mathbb{R}_+ \text{ and } (\tilde{\boldsymbol{w}}_{\leq m}, \tilde{r}, \tilde{s}) \in B(\boldsymbol{w}_{\leq m}, \delta_1) \times B(r, \delta_2) \times B(s, \delta_3), \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\text{then } (\boldsymbol{w}', r', s') \in O. \end{aligned} \tag{5.7.7}$$

By increasing m_0 if necessary, we may also suppose that for all $m \geq m_0$, we have

$$\|\boldsymbol{w}_{>m}\|_2 \leq \frac{\delta_2 r}{2}. \tag{5.7.8}$$

We set $t := \sqrt{r^2 - \|\boldsymbol{w}\|_2^2}$ and $\rho' := \min(\delta_1, \delta_2/4, \delta_3)$.

Now, let $(\boldsymbol{w}', r', s') \in \mathcal{X} \times \mathbb{R}_+$ be such that (5.7.6) holds for some $m \geq m_0$ and our choice of ρ', and let this choice of m be fixed for the remainder of the proof. Then clearly,

$$\boldsymbol{w}'_{\leq m} \in B\left(\boldsymbol{w}_{\leq m}, \rho'\right) \subset B\left(\boldsymbol{w}_{\leq m}, \delta_1\right), \tag{5.7.9}$$

$$s' \in B\left(s, \rho'\right) \subset B\left(s, \delta_3\right), \tag{5.7.10}$$

and

$$\rho' > \left| \sqrt{r'^2 - \|\boldsymbol{w}'_{\leq m}\|_2^2} - \sqrt{r^2 - \|\boldsymbol{w}\|_2^2} \right|.$$

Multiplying both sides by $\sqrt{r'^2 - \|\boldsymbol{w}'_{\leq m}\|_2^2} + \sqrt{r^2 - \|\boldsymbol{w}\|_2^2}$ we have

$$\rho'\left(\sqrt{r'^2 - \|\boldsymbol{w}'_{\leq m}\|_2^2} + \sqrt{r^2 - \|\boldsymbol{w}\|_2^2}\right) > \left| r^2 - (r')^2 - \|\boldsymbol{w}_{\leq m}\|_2^2 + \|\boldsymbol{w}'_{\leq m}\|_2^2 - \|\boldsymbol{w}_{>m}\|_2^2 \right|.$$

Note because that $(\boldsymbol{w}, r, s)$ and $(\boldsymbol{w}', r', s')$ lie in $\mathcal{X} \times \mathbb{R}_+$, we have $\|\boldsymbol{w}\|_2 \leq r$ and $\|\boldsymbol{w}'\|_2 \leq r'$. Combining this fact with the triangle inequality, (5.7.8), (5.7.9), and the previous display, we have

$$\begin{aligned}
|r^2 - r'^2| &< \left| \|\boldsymbol{w}_{\leq m_0}\|_2^2 - \left\|\boldsymbol{w}'_{\leq m}\right\|_2^2 \right| + \|\boldsymbol{w}_{>m}\|^2 + \rho'(r + r') \\
&\leq \left\|\boldsymbol{w}_{\leq m} - \boldsymbol{w}'_{\leq m}\right\|_2 \left(\|\boldsymbol{w}_{\leq m}\|_2 + \|\boldsymbol{w}'_{\leq m}\|_2\right) + \delta_2 r/2 + \rho'(r + r') \\
&\leq 2\rho'(r + r') + \delta_2 r/2.
\end{aligned}$$

Dividing both sides by $(r + r')$ and recalling $2\rho' \leq \delta_2/2$, this implies

$$|r - r'| < 2\rho' + \delta_2/2 \leq \delta_2. \tag{5.7.11}$$

When combined, (5.7.9), (5.7.10), and (5.7.11) show that the supposition in (5.7.7) holds with $(\tilde{\boldsymbol{w}}, \tilde{r}, \tilde{s}) = (\boldsymbol{w}', r', s')$. Thus (5.7.7) implies $(\boldsymbol{w}', r', s') \in O$, which proves the lemma. □

In the proof of the next lemma, we use the following concept from convex analysis.

Definition 5.7.2 (Relative interior)**.** *For every non-empty convex set C, the relative interior of C, denoted* $\mathrm{ri}(C)$, *is defined as the set*

$$\mathrm{ri}(C) = \{x \in C : \textit{for all } y \in C, \textit{ there exists some } \mu > 1 \textit{ such that } \mu x + (1-\mu)y \in C\}.$$

Proof of Lemma 5.4.1. We first show that Λ_m, defined in (5.3.3), is essentially smooth for all $m \in \mathbb{N}$ (see Definition 5.2.1). The finiteness of Λ_m on $\mathbb{R}^{m_0} \times \mathbb{R} \times (-\infty, T)$ follows from the finiteness of $\bar{\Lambda}$ on $\mathbb{R} \times (-\infty, T)$. The differentiability of Λ_{m_0} at $(\boldsymbol{v}', b', c') \in \mathbb{R}^{m_0} \times \mathbb{R} \times (-\infty, T)$ is a direct consequence of the differentiability of $\bar{\Lambda}$ on $\mathbb{R} \times (-\infty, T)$, the bounds in Assumption 5, and the dominated convergence theorem. To check the third condition in Definition 5.2.1, first note that for $t_2 \in (-\infty, T)$,

$$\partial_2 \bar{\Lambda}\left(t_1, t_2\right) = e^{-\bar{\Lambda}(t_1,t_2)} \mathbb{E}\left[\eta(X_1) e^{t_1 X_1 + t_2 \eta(X_1)}\right] \geq 0. \tag{5.7.12}$$

Hence, we have

$$\|\nabla \Lambda_{m_0}(\boldsymbol{v}', b', c')\|_2 \geq \mathbb{E}\left[\partial_2 \bar{\Lambda}\left(\sum_{l=1}^{m_0} v_l' g_l + b' g_0, c'\right)\right] \geq 0.$$

Taking the limit infimum over $(\boldsymbol{v}', b', c') \in \mathbb{R}^{m_0} \times \mathbb{R} \times (-\infty, T)$ above and using Fatou's lemma, one obtains

$$\liminf_{\boldsymbol{v}', b', c' \to \boldsymbol{v}_0, b_0, T} \|\nabla \Lambda_{m_0}(\boldsymbol{v}', b', c')\|_2 \geq \mathbb{E}\left[\liminf_{\boldsymbol{v}', b', c' \to \boldsymbol{v}_0, b_0, T} \partial_2 \bar{\Lambda}\left(\sum_{l=1}^{m_0} v_l' g_l + b' g_0, c'\right)\right] = \infty,$$

where the last equality follows from the essential smoothness of $\bar{\Lambda}$, which in turn follows from Assumption 4. This establishes the essential smoothness of Λ_m.

Fix an open set $\mathcal{O}$ and a point $(\boldsymbol{w}, r, s) \in \mathcal{O}$ such that $I(\boldsymbol{w}, r, s) < \infty$. Let m_0 be as in Lemma 5.7.1 and recall the definition of Λ_{m_0} from (5.3.3), and the definitions of $\Lambda^*_{m_0}$ and $\mathcal{D}^*_{m_0}$ from (5.4.2) and (5.4.3). Since Λ_{m_0} is essentially smooth, by [145, Corollary 26.4.1] we have

$$\nabla\Lambda_{m_0}(\mathcal{D}_{m_0}) \subset \mathcal{D}^*_{m_0}. \tag{5.7.13}$$

Note that $\Lambda_{m_0}(0,0,0) = 0$ and the cumulant generating function $\bar{\Lambda}(s_1, s_2) = \log \mathbb{E}[e^{s_1 X_1 + s_2 \eta(X_1)}]$ satisfies $\bar{\Lambda}(s_1, 0) \geq \mathbb{E}[s_1 X_1] = 0$. Hence, the mapping $(\boldsymbol{v}', b') \mapsto \Lambda_{m_0}(\boldsymbol{v}', b', 0)$ achieves its minimum of 0 at $(\mathbf{0}, 0)$. Combining this with (5.7.12), we have $\nabla\Lambda_{m_0}(\mathbf{0}, 0, 0) = (\mathbf{0}, 0, s_0)$ for some $s_0 \in \mathbb{R}_+$. By differentiating Λ_{m_0} twice, we have

$$\nabla^2\Lambda_{m_0}(\mathbf{0}, 0, 0) = \begin{pmatrix} \partial_1^2\bar{\Lambda}(\mathbf{0}, 0) I_{m_0+1} & \mathbf{0} \\ \mathbf{0}^\top & \partial_2^2\bar{\Lambda}(\mathbf{0}, 0) \end{pmatrix} = \begin{pmatrix} \mathrm{Var}(X_1) I_{m_0+1} & \mathbf{0} \\ \mathbf{0}^\top & \mathrm{Var}(\eta(X_1)) \end{pmatrix},$$

which is positive definite (since, by Assumption 2, X_1 is not degenerate and η is not a constant function). Thus, the inverse function theorem implies that $\nabla\Lambda_{m_0}$ is locally a diffeomorphism, which, along with (5.7.13), implies there exists $\varepsilon_0 > 0$ such that

$$B_{\varepsilon_0}(0, 0, s_0) \subset \mathrm{ri}(\mathcal{D}^*_{m_0}).$$

Now, pick $(\tilde{\boldsymbol{w}}, \tilde{t}, \tilde{s}) \in B_{\varepsilon_0}(0, 0, s_0)$ such that $\tilde{\boldsymbol{w}}$ is strictly ordered in the sense that $|\tilde{\boldsymbol{w}}_1| > \ldots > |\tilde{\boldsymbol{w}}_{m_0}|$. Set $t = \sqrt{r^2 - \|\boldsymbol{w}\|_2^2}$. By Jensen's inequality, $\Lambda_{m_0}(\boldsymbol{w}_{\leq m_0}, t, s) \leq I(\boldsymbol{w}, r, s) < \infty$ by (5.4.2), and hence, $(\boldsymbol{w}_{\leq m_0}, t, s) \in \mathcal{D}_{m_0}$. Since $(\tilde{\boldsymbol{w}}, \tilde{t}, \tilde{s}) \in \mathrm{ri}(\mathcal{D}^*_{m_0})$, [145, Theorem 6.1] and [145, Corollary 26.4.1] imply that, for all $\lambda \in [0, 1)$,

$$\lambda(\boldsymbol{w}_{\leq m_0}, t, s) + (1-\lambda)(\tilde{\boldsymbol{w}}, \tilde{t}, \tilde{s}) \in \mathrm{ri}(\mathcal{D}^*_{m_0}) \subset \{\nabla\Lambda_{m_0}(y) : y \in \mathcal{D}_{m_0}\}.$$

By the definitions of $(\boldsymbol{w}_{\leq m_0}, t, s)$ and $(\tilde{\boldsymbol{w}}, \tilde{t}, \tilde{s})$, note that the points on the line segment

$$\ell := \{\lambda(\boldsymbol{w}_{\leq m_0}, t, s) + (1-\lambda)(\tilde{\boldsymbol{w}}, \tilde{t}, \tilde{s}) : \lambda \in (0, 1)\}$$

are strictly ordered in the first component. Since $\Lambda^*_{m_0}$ is a convex function (by definition of the Legendre transform), it is continuous along $\ell \subset \mathcal{D}^*_{m_0}$. Together with Lemma 5.7.1 (to establish (5.4.4)), this implies that there exists a point $(\bar{\boldsymbol{w}}, \bar{t}, \bar{s}) \in \ell$ that satisfies all conditions of the lemma. □

5.7.3 Proof of the modified lower bound

The proof of Lemma 5.4.2, which is given in Section 5.7.3, uses a change of measure.

Throughout this section, we fix μ to be the law of a random variable satisfying Assumptions 2, 3 and 4, with associated constants $C_2, T > 0$ and function $\tilde{C}(\cdot)$. For $s_1 \in \mathbb{R}, s_2 < T$, recalling $\bar{\Lambda}$ defined in (5.1.5), we define the exponentially tilted measure $\tilde{\mu}_{s_1,s_2}$ by

$$\frac{d\tilde{\mu}_{s_1,s_2}}{d\mu}(x) := \exp(s_1 x + s_2 \eta(x) - \overline{\Lambda}\,(s_1, s_2)), \quad x \in \mathbb{R}. \tag{5.7.14}$$

Preparatory results

The first lemma claims that $\{\tilde{\mu}_{s_1,s_2}\}_{(s_1,s_2)\in\mathbb{R}\times(-\infty,T)}$ is uniformly sub-Gaussian in s_1.

Lemma 5.7.3. *Fix $(s_1, s_2) \in \mathbb{R} \times (-\infty, T)$. Let $\tilde{X}_1$ be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P}_{s_1,s_2})$ that has law $\tilde{\mu}_{s_1,s_2}$, and let $\tilde{\mathbb{E}}_{s_1,s_2}$ denote expectation with respect to $\tilde{\mathbb{P}}_{s_1,s_2}$. Then there exists a constant $C = C(s_2) > 0$ such that*

$$\tilde{\mathbb{P}}_{s_1,s_2}\left(|\tilde{X}_1 - \tilde{\mathbb{E}}_{s_1,s_2}[\tilde{X}_1]| > t\right) \leq 2\exp(-t^2/C^2).$$

Proof of Lemma 5.7.3. Fix $(s_1, s_2) \in \mathbb{R} \times (-\infty, T)$. By the definition of $\bar{\Lambda}$ in (5.1.5) and the definition of $\tilde{\mu}_{s_1,s_2}$ in (5.7.14), we have

$$\tilde{\mathbb{E}}_{s_1,s_2}\left[\tilde{X}_1\right] = \int x e^{s_1 x + s_2 \eta(x) - \bar{\Lambda}(s_1,s_2)} d\mu(x) = \partial_1 \bar{\Lambda}\,(s_1, s_2)\,.$$

Now Assumption 5 implies there exists a continuous function $\tilde{C}_0\colon (-\infty, T) \to \mathbb{R}$ such that

$$\begin{aligned}\tilde{\mathbb{E}}_{s_1,s_2}[\exp(\lambda(\tilde{X}_1 - \tilde{\mathbb{E}}_{s_1,s_2}[\tilde{X}_1]))] &= e^{-\lambda\tilde{\mathbb{E}}_{s_1,s_2}[\tilde{X}_1]-\Lambda(s_1,s_2)} \int \exp(\lambda x)\exp(s_1 x + s_2\eta(x))\,d\mu(x) \\ &= \exp(\bar{\Lambda}(s_1+\lambda, s_2) - \bar{\Lambda}(s_1,s_2) - \lambda\partial_1\bar{\Lambda}(s_1,s_2)) \\ &\leq \exp\left(\tilde{C}_0(s_2)^2\lambda^2\right).\end{aligned}$$

The lemma then follows from a standard argument using Chernoff's bound (see [164, Proposition 2.5.2]). □

Next, we state a Gaussian approximation result for the Haar measure on $\mathbb{V}_{n,k_n}$ in Lemma 5.7.4, and establish an asymptotic decorrelation result in Lemma 5.7.6. The proofs of these lemmas are deferred to Section 5.10.

Recall from Section 5.7.1 that T denotes the constant from Assumption 4, and that M was defined in (5.7.2) in terms of m and a point $(\hat{\boldsymbol{u}}, b, c) \in \mathbb{R}^m \times \mathbb{R} \times (-\infty, T)$. Recall from (5.7.5) that $g_1, g_2, \ldots, g_{m+M}$ denote i.i.d. Gaussian random variables with zero mean and unit variance. Additionally, given any $n \in \mathbb{N}$ and $j \in \{1, \ldots, k_n\}$, and any function $h : \mathbb{R} \to \mathbb{R}$, we define a map $f_j^{(n)}\colon \mathbb{V}_{n,k_n} \to \mathbb{R}$ as follows:

$$f_j^{(n)}\left(\boldsymbol{a}^{(n)}\right) := \frac{1}{n}\sum_{i=1}^{n} \sqrt{n}a_{ij}^{(n)} h\left(\left\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\right\rangle\right) = \int xh(y)\,d\nu_j^{(n)}(x,y), \qquad (5.7.15)$$

where we recall that $\boldsymbol{u}^{(n)} = (u_1, \ldots, u_{k_n})$ and $\nu_j^{(n)}$ were specified in (5.7.3) and (5.7.4) respectively.

Lemma 5.7.4. *Fix $C_0 \in (0, \infty)$ and a function $h : \mathbb{R} \to \mathbb{R}$ satisfying*

$$\left|\frac{d^\alpha}{dx^\alpha}h(x)\right| \leq C_0(1 + |x|^{1-\alpha}), \qquad \alpha = 0, 1. \qquad (5.7.16)$$

Fix $m \in \mathbb{N}$ and $(\hat{\boldsymbol{u}}, b, c) \in \mathbb{R}^m \times \mathbb{R} \times (-\infty, T)$. Then for σ-a.e. $\boldsymbol{a} = (\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}, \ldots) \in \mathbb{V}$, *we have for all $j \in \{1, \ldots, m+M\}$ that*

$$\lim_{n\to\infty} f_j^{(n)}\left(\boldsymbol{a}^{(n)}\right) = \mathbb{E}\left[g_j h\left(\sum_{l=1}^{m+M} u_l g_l\right)\right].$$

Remark 5.7.5. For any fixed $x \in (-\infty, T)$, by Assumption 5, $\partial_1 \bar{\Lambda}(\cdot, x)$ satisfies the condition imposed on h in Lemma 5.7.4.

Lemma 5.7.6. *Fix $m \in \mathbb{N}$ and $(\hat{\boldsymbol{u}}, b, c) \in \mathbb{R}^m \times \mathbb{R} \times (-\infty, T)$, and retain the assumption (5.7.16). Then for σ-a.e. $\boldsymbol{a} = (\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}, \ldots) \in \mathbb{V}$,*

$$\lim_{n\to\infty} \sum_{\ell=m+M+1}^{k_n} \left(f_\ell^{(n)}\left(\boldsymbol{a}^{(n)}\right)\right)^2 = 0.$$

Proof of Lemma 5.4.2

We now show how the preparatory results from Section 5.7.3 can be combined to establish Lemma 5.4.2. Let $m \in \mathbb{N}$, $(\hat{\boldsymbol{u}}, b, c) \in \mathcal{D}_m$ and $(\hat{\boldsymbol{w}}, t, s) = \nabla \Lambda_m(\hat{\boldsymbol{u}}, b, c)$ be as in the statement of the lemma. Let M be defined in terms of t and m as in (5.7.2) and let $\boldsymbol{u}^{(n)} \in \mathbb{R}^{k_n}$ be defined in terms of $\hat{\boldsymbol{u}}$ and b as in (5.7.3). Let $(X_1, \ldots, X_n)$ be random variables defined on a measure space $(\Omega, \mathcal{F})$ equipped with the probability measures $\mathbb{P}^{(n)}$ and $\tilde{\mathbb{P}}^{(n)}$. Suppose that under $\mathbb{P}^{(n)}$, the variables $(X_1, \ldots, X_n)$ are i.i.d. with a common law μ that satisfies Assumptions 1–4, and let $\mu^{(n)} := \mu^{\otimes n}$. Let $\tilde{\mu}_{s_1,s_2}$ be defined in terms of μ as in (5.7.14), and let $\tilde{\mathbb{P}}^{(n)}$ be a probability measure on $(\Omega, \mathcal{F})$ such that under $\tilde{\mathbb{P}}^{(n)}$, $(X_1, \ldots, X_n)$ has law $\tilde{\mu}^{(n)}$, where

$$\tilde{\mu}^{(n)} := \otimes_{i=1}^n \tilde{\mu}_{\lambda_i, c}, \quad \lambda_i \equiv \lambda_i^{(n)} := \left\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\right\rangle, \quad i = 1, \ldots, n. \tag{5.7.17}$$

Let $(\tilde{X}_1, \ldots, \tilde{X}_n)$ be an $\mathbb{R}^n$-valued random vector with law $\tilde{\mu}^{(n)}$. Let $\mathbb{E}^{(n)}$ and $\tilde{\mathbb{E}}^{(n)}$ denote expectations with respect to $\mathbb{P}^{(n)}$ and $\tilde{\mathbb{P}}^{(n)}$, respectively. Note that the Radon–Nikodym derivative of $\tilde{\mu}^{(n)}$ with respect to $\mu^{(n)}$ can be rewritten in terms of the function F_n from

(5.3.1) as follows:

$$\frac{d\tilde{\mu}^{(n)}}{d\mu^{(n)}}(\boldsymbol{x}) = e^{n\left\langle \boldsymbol{u}^{(n)}, \frac{1}{\sqrt{n}}\left(\boldsymbol{a}^{(n)}\right)^\top \boldsymbol{x}\right\rangle + c\sum_{i=1}^n \eta(x_i) - nF_n\left(\boldsymbol{u}^{(n)}, \boldsymbol{a}^{(n)}, c\right)}, \ \forall \boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n. \tag{5.7.18}$$

Next, recall that $\kappa > 0$ and $B = B((\hat{\boldsymbol{w}}, t, s), \kappa)$ were fixed in the statement of Lemma 5.4.2, and for n such that $k_n \geq M + m$, let

$$\boldsymbol{w}^{(n)} = (w_1, \ldots, w_{k_n}) := (\hat{\boldsymbol{w}}, \underbrace{tM^{-1/2}, \ldots, tM^{-1/2}}_{M \text{ times}}, \underbrace{0, \ldots, 0}_{k_n - m - M \text{ times}}) \in \mathbb{R}^{k_n}. \tag{5.7.19}$$

Fix $\delta \in (0, \min(\kappa, m^{-1/2}))$ and define

$$E_n := \left\{ \boldsymbol{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n : \left(\frac{1}{\sqrt{n}} (\boldsymbol{a}^{(n)})^\top \boldsymbol{x}, \ \frac{1}{n}\sum_{i=1}^n \eta(x_i) \right) \in B\left((\boldsymbol{w}^{(n)}, s), \delta \right) \right\}.$$

Let $W^{(n)}$ and $L^{(n)}$ be defined in terms of $(X_1, \ldots, X_n)$ as in (5.3.15). Then using the definitions of $\mu^{(n)}$, $\tilde{\mu}^{(n)}$, and E_n, and (5.7.18), we have

$$\begin{aligned}
&\mathbb{P}^{(n)}\left((W^{(n)}_{\leq m}, \|W^{(n)}_{>m}\|_2, L^{(n)}) \in B, \|W^{(n)}_{>m}\|_\infty \leq 2m^{-1/2} \right) \\
\geq &\mathbb{P}^{(n)}\left(\left(W^{(n)}, L^{(n)}\right) \in B\left((\boldsymbol{w}^{(n)}, s), \delta \right) \right) \\
= &\int_{E_n} d\mu^{(n)}(\boldsymbol{x}) \\
= &\int_{E_n} e^{-n\left(\left\langle \boldsymbol{u}^{(n)}, \frac{1}{\sqrt{n}}\left(\boldsymbol{a}^{(n)}\right)^\top \boldsymbol{x}\right\rangle + \frac{c}{n}\sum_{i=1}^n \eta(x_i)\right) + nF_n(\boldsymbol{u}^{(n)}, \boldsymbol{a}^{(n)}, c)} d\tilde{\mu}^{(n)}(\boldsymbol{x}) \\
\geq &\tilde{\mathbb{P}}^{(n)}\left(\left(W^{(n)}, L^{(n)}\right) \in B\left((\boldsymbol{w}^{(n)}, s), \delta \right) \right) e^{-n\left(\langle \hat{\boldsymbol{u}}, \hat{\boldsymbol{w}} \rangle + bt + cs - F_n(\boldsymbol{u}^{(n)}, \boldsymbol{a}^{(n)}, c) + \delta \|(\hat{\boldsymbol{u}}, b, c)\|_2\right)}.
\end{aligned} \tag{5.7.20}$$

We show below that the following three claims hold for σ-a.e. $\boldsymbol{a} \in \mathbb{V}$ and any $\varepsilon > 0$:

$$\lim_{n\to\infty} \tilde{\mathbb{P}}^{(n)}\left(|W^{(n)}_j - w_j| \geq \varepsilon \right) = 0, \ j = 1, \ldots, m + M, \tag{5.7.21}$$

$$\lim_{n\to\infty} \tilde{\mathbb{P}}^{(n)}\left(\sum_{l=m+M+1}^{k_n} (W^{(n)}_l)^2 \geq \varepsilon \right) = 0, \tag{5.7.22}$$

$$\lim_{n\to\infty} \tilde{\mathbb{P}}^{(n)}\left(\left| L^{(n)} - s \right| \geq \varepsilon \right) = 0. \tag{5.7.23}$$

In essence, the limits (5.7.21) and (5.7.23) are weak laws of large numbers for $W_j^{(n)}, j = 1, \ldots, m$, and $L^{(n)}$ respectively, under the tilted measure, and (5.7.22) states that the ℓ^2-norms of $W_{>m+M}^{(n)}$ are asymptotically negligible under the tilted measure.

Postponing the proofs of (5.7.21)–(5.7.23), we first note that together with (5.7.18), they imply that for σ-a.e. $\boldsymbol{a} \in \mathbb{V}$,

$$\liminf_{n\to\infty} \tilde{\mathbb{P}}^{(n)} \left(\left(W^{(n)}, L^{(n)} \right) \in B \left((\boldsymbol{w}^{(n)}, s), \delta \right) \right) = 1. \tag{5.7.24}$$

Moreover, $F_n(\boldsymbol{u}^{(n)}, \boldsymbol{a}^{(n)}, c) \to \Lambda_m(\hat{\boldsymbol{u}}, b, c)$ for σ-a.e. $\boldsymbol{a} \in \mathbb{V}$ by Lemma 5.3.6, with $d_n = 1$. Thus, combining (5.7.20) and (5.7.24), and recalling the defining of Λ_m^* in (5.4.2), we conclude that for σ-a.e. $\boldsymbol{a} \in \mathbb{V}$,

$$\begin{aligned}
\liminf_{n\to\infty} n^{-1} \log \mathbb{P}^{(n)} &\left((W_{\le m}^{(n)}, \|W_{>m}^{(n)}\|_2, L) \in B, \|W_{>m}^{(n)}\|_\infty \le 2m^{-1/2} \right) \\
&\ge - \left(\langle \hat{\boldsymbol{u}}, \hat{\boldsymbol{w}} \rangle + bt + cs - \Lambda_m \left(\hat{\boldsymbol{u}}, b, c \right) + \delta \| (\hat{\boldsymbol{u}}, b, c) \|_2 \right) \\
&\ge - \Lambda_m^*(\hat{\boldsymbol{w}}, t, s) - \delta \| (\hat{\boldsymbol{u}}, b, c) \|_2.
\end{aligned}$$

Taking $\delta \searrow 0$ completes the proof of (5.4.5), given (5.7.21)–(5.7.23).

We now turn to the proofs of (5.7.21)–(5.7.23).

Proof of (5.7.21). By the definition of $\tilde{\mu}^{(n)}$ in (5.7.18) and Lemma 5.7.3 (with $s_1 = \lambda_i$ from (5.7.17), and $s_2 = c$, where c is from the statement of the lemma), we have, for each $i = 1, \ldots, n$,

$$\tilde{\mathbb{P}}^{(n)} \left(|X_i - \tilde{\mathbb{E}}^{(n)} \left[X_i \right] | > t \right) \le 2 \exp(-t^2/C^2),$$

for some constant $C > 0$ depending only on c. Also, the definition of $W^{(n)}$ in (5.3.15), Hoeffding's inequality for sub-Gaussian random variables (see [164, Theorem 2.6.2]), and a union bound imply that for all $\boldsymbol{a} \in \mathbb{V}$,

$$\begin{aligned}
&\tilde{\mathbb{P}}^{(n)}\left(\max_{1\le j\le m+M}\left|W_j^{(n)} - \tilde{\mathbb{E}}^{(n)}[W_j^{(n)}]\right| \ge \varepsilon/3\right) \\
&\le (m+M)\max_{1\le j\le m+M}\tilde{\mathbb{P}}^{(n)}\left(\left|\sum_{i=1}^{n}\frac{a_{ij}^{(n)}}{\sqrt{n}}(X_i - \tilde{\mathbb{E}}^{(n)}[X_i])\right| \ge \varepsilon/3\right) \\
&\le 2(m+M)\exp(-Cn\varepsilon^2),
\end{aligned} \tag{5.7.25}$$

which decays to zero as $n \to \infty$. Also, by the definitions of $\bar{\Lambda}$ and $\tilde{\mu}^{(n)}$ in (5.1.5) and (5.7.18), respectively, we have,

$$\tilde{\mathbb{E}}^{(n)}[X_i] = \partial_1\bar{\Lambda}\left(\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\rangle, c\right), \quad i = 1, \dots, n. \tag{5.7.26}$$

By the definition of $W^{(n)}$ in (5.3.15) and the relation (5.7.18), it follows that

$$\tilde{\mathbb{E}}^{(n)}[W_j^{(n)}] = \frac{1}{n}\sum_{i=1}^{n}\sqrt{n}a_{ij}^{(n)}\partial_1\bar{\Lambda}\left(\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\rangle, c\right), \quad j = 1, \dots, n.$$

Then Lemma 5.7.4 with $h = \partial_1\Lambda(\cdot, c)$ and Remark 5.7.5 imply that for σ-a.e. $\boldsymbol{a}$ and each $j = 1, \dots, m+M$,

$$\lim_{n\to\infty}\tilde{\mathbb{E}}^{(n)}[W_j^{(n)}] = \mathbb{E}\left[g_j\partial_1\bar{\Lambda}\left(\sum_{l=1}^{m+M}u_l^{(n)}g_l, c\right)\right]. \tag{5.7.27}$$

Let g_0 be a Gaussian random variable independent of $g_1, \dots, g_m$. Recalling from (5.7.3) that $u_\ell^{(n)} = \hat{u}_\ell$ for $\ell = 1, \dots, m$ and $u_{m+1}^{(n)} = \cdots = u_{m+M}^{(n)} = bM^{-1/2}$, we have $\sum_{l=m+1}^{m+M}u_l^{(n)}g_l \overset{d}{=} bg_0$ and $(g_0, g_1, \dots, g_m) \overset{d}{=} (M^{-1/2}\sum_{l=m+1}^{m+M}g_l, g_1, \dots, g_m)$. Together with (5.3.3) and (5.7.1), for $j = 1, 2, \dots, m$, we have

$$\mathbb{E}\left[g_j\partial_1\bar{\Lambda}\left(\sum_{l=1}^{m+M}u_lg_l, c\right)\right] = \mathbb{E}\left[g_j\partial_1\bar{\Lambda}\left(\sum_{l=1}^{m}u_lg_l + bg_0, c\right)\right] = \partial_j\Lambda_m(\hat{\boldsymbol{u}}, b, c) = u_j. \tag{5.7.28}$$

and for $j = m+1, \dots, m+M$, also using (5.7.19), we have

$$\begin{aligned}\mathbb{E}\left[g_j \partial_1 \bar{\Lambda}\left(\sum_{l=1}^{m+M} u_l^{(n)} g_l, c\right)\right] =&\mathbb{E}\left[\frac{\sum_{l=m+1}^{m+M} g_l}{M^{1/2}} \partial_1 \bar{\Lambda}\left(\sum_{l=1}^{m} \hat{u}_l g_l + bM^{-1/2} \sum_{l=m+1}^{m+M} g_l, c\right)\right] \\ =&M^{-1/2}\mathbb{E}\left[g_0 \partial_1 \bar{\Lambda}\left(\sum_{l=1}^{m} \hat{u}_l g_l + bg_0, c\right)\right] \\ =&M^{-1/2}\partial_2 \Lambda_m\left(\hat{\boldsymbol{u}}, b, c\right) \\ =&tM^{-1/2} \\ =&w_j^{(n)}.\end{aligned} \tag{5.7.29}$$

Combining (5.7.25), (5.7.27), (5.7.28) and (5.7.29), (5.7.21) follows. □

Proof of (5.7.22). By Markov's inequality and the independence of $\{X_i,\, i = 1, \dots, n\}$ under the tilted measure $\tilde{\mathbb{P}}^{(n)}$ (see (5.7.17)), we have

$$\begin{aligned}\tilde{\mathbb{P}}^{(n)}\left(\sum_{l=m+M+1}^{k_n} (W_l^{(n)})^2 \geq \varepsilon\right) \leq&\varepsilon^{-1}\mathbb{E}^{(n)}\left[\sum_{l=m+M+1}^{k_n} (W_l^{(n)})^2\right] \\ =&\varepsilon^{-1} n^{-1} \sum_{i,j=1}^{n} \tilde{\mathbb{E}}^{(n)}\left[X_i X_j\right] \sum_{l=m+M+1}^{k_n} a_{il}^{(n)} a_{jl}^{(n)} \\ =&\varepsilon^{-1} n^{-1}\left(\sum_{i,j=1}^{n} \tilde{\mathbb{E}}^{(n)}[X_i]\tilde{\mathbb{E}}^{(n)}[X_j] \sum_{l=m+M+1}^{k_n} a_{il}^{(n)} a_{jl}^{(n)} + \right. \\ &\left. \sum_{i=1}^{n} \widetilde{\mathrm{Var}}\left(X_i\right) \sum_{l=m+M+1}^{k_n} (a_{il}^{(n)})^2\right),\end{aligned}$$

where $\widetilde{\mathrm{Var}}(X_i)$ denotes the variance of X_i under the tilted measure $\tilde{\mathbb{P}}^{(n)}$. By a calculation similar to the one carried out in (5.7.26), we have $\widetilde{\mathrm{Var}}(X_i) = \partial_1^2 \bar{\Lambda}\left(\left\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\right\rangle, c\right)$, which is bounded uniformly in $n \in \mathbb{N}$ and $i = 1, \dots, n$ by Assumption 5. Additionally, $\sum_{l=m+M+1}^{k_n} (a_{il}^{(n)})^2$ is uniformly bounded in $n \in \mathbb{N}$ and $i = 1, \dots, n$. Therefore,

$$\tilde{\mathbb{P}}^{(n)}\left(\sum_{l=m+M+1}^{k_n} W_l^2 \geq \varepsilon\right) \leq \varepsilon^{-1} n^{-1} \sum_{i,j=1}^{n} \tilde{\mathbb{E}}^{(n)}[X_i]\tilde{\mathbb{E}}^{(n)}[X_j] \sum_{l=m+M+1}^{k_n} a_{il}^{(n)} a_{jl}^{(n)} + \varepsilon^{-1} O\left(\frac{k_n}{n}\right).$$

Hence, it suffices to show that for σ-a.e. $\boldsymbol{a} \in \mathbb{V}$,

$$\lim_{n\to\infty} n^{-1} \sum_{i,j=1}^{n} \tilde{\mathbb{E}}^{(n)}[X_i]\tilde{\mathbb{E}}^{(n)}[X_j] \sum_{l=m+M+1}^{k_n} a_{il}^{(n)} a_{jl}^{(n)} = 0.$$

Recalling $\tilde{\mathbb{E}}^{(n)}[X_i] = \partial_1 \bar{\Lambda}\left(\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\rangle, c\right)$ from (5.7.26), the above convergence follows directly from Lemma 5.7.6, with $h = \bar{\Lambda}(\cdot, c)$, and Remark 5.7.5. $\square$

Proof of (5.7.23). Let $\lambda > 0$ be a small parameter, to be chosen later. Note that $\tilde{\mathbb{E}}^{(n)}[\eta(X_i)] = \partial_2\bar{\Lambda}(\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\rangle, c)$, by a calculation similar to (5.7.26) (using (5.1.5) and (5.7.18)). Then by Chebyshev's inequality, the mutual independence of $(X_1, \dots, X_n)$ under $\tilde{\mathbb{P}}^{(n)}$, the definition of $\overline{\Lambda}$ in (5.1.5), and calculations similar to those in the proof of Lemma 5.7.3, we have

$$\begin{aligned}
&\tilde{\mathbb{P}}^{(n)}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\eta(X_i) - \tilde{\mathbb{E}}^{(n)}[\eta(X_i)]\right) > \frac{\varepsilon}{2}\right) \\
&\leq e^{-\lambda n\varepsilon/2} \prod_{i=1}^{n} \tilde{\mathbb{E}}^{(n)}\left[\exp\left(\lambda\eta(X_i) - \lambda\tilde{\mathbb{E}}^{(n)}[\eta(X_i)]\right)\right] \\
&= e^{-\lambda n\varepsilon/2} \prod_{i=1}^{n} \exp\left(\bar{\Lambda}\left(\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\rangle, c+\lambda\right) - \bar{\Lambda}\left(\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\rangle, c\right) - \lambda\partial_2\bar{\Lambda}\left(\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\rangle, c\right)\right) \\
&= \exp\left(\frac{\lambda^2}{2}\sum_{i=1}^{n}\partial_2^2\bar{\Lambda}\left(\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\rangle, \xi_i\right) - \lambda n\frac{\varepsilon}{2}\right),
\end{aligned}$$

where $\xi_i \in (c, c+\lambda)$, $i = 1, \dots, n$. Together with Assumption 5, we have

$$\tilde{\mathbb{P}}^{(n)}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\eta(X_i) - \tilde{\mathbb{E}}^{(n)}[\eta(X_i)]\right) > \frac{\varepsilon}{2}\right) \leq \exp\left(\left(-\frac{\lambda}{2}\varepsilon + C(1 + \|\boldsymbol{u}\|_2^2)\lambda^2\right)n\right) \to 0$$

for all $\boldsymbol{a} \in \mathbb{V}$, if λ is chosen small enough. To prove (5.7.23), it suffices to show that, for σ-a.e. $\boldsymbol{a} = \left(\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}, \dots\right) \in \mathbb{V}$,

$$\frac{1}{n}\sum_{i=1}^{n}\partial_2\bar{\Lambda}(\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\rangle, c) \to \partial_3\Lambda_m(\hat{\boldsymbol{u}}, b, c) = s. \tag{5.7.30}$$

Since $H = \partial_2\bar{\Lambda}$ satisfies the conditions in (5.3.7) due to Assumption 5 and (5.7.2), and because $\|\boldsymbol{u}^{(n)}\|_2 = D := \sqrt{\|\hat{\boldsymbol{u}}\|_2^2 + |b|^2}$ for all $n \in \mathbb{N}$, it follows from Lemma 5.3.6 and Remark 5.3.7 that (5.7.30) holds with the right-hand side equal to $\mathbb{E}[\partial_2\bar{\Lambda}(Dg, c)]$. Then the fact that $\mathbb{E}[\partial_2\bar{\Lambda}(Dg, c)] = \partial_3\Lambda_m(\hat{\boldsymbol{u}}, b, c)$ follows after recalling (5.1.6), (5.3.3), and $\langle \boldsymbol{u}^{(n)}, g\rangle + bg_0 \overset{d}{=} Dg$ (by (5.7.2)). This completes the proof. □

Since all three claims (5.7.21)–(5.7.23) have been established, this completes the proof of Lemma 5.4.2.

5.8 Properties of p-Gaussian Variables

Lemma 5.8.1. *Fix $p \geq 2$, let X be a random variable with probability density function equal to $\bar{c}_p \exp(-p^{-1}|x|^p)$ for a suitable normalization constant $\bar{c}_p$, and define*

$$\widehat{\Lambda}(t) := \log \mathbb{E}\Big[\exp(tX)\Big], \quad t \in \mathbb{R}.$$

Then there exists a constant $C \in (0, \infty)$ such that the following properties hold for all $t \in \mathbb{R}$:

1. $0 \leq \widehat{\Lambda}(t) \leq C\,(1 + t^2)$.
2. $|\widehat{\Lambda}'(t)| \leq C\,(1 + |t|)$.
3. $\widehat{\Lambda}''(t) \leq C$.

Proof. Since the distribution of X is symmetric by definition, the log moment gerating function Λ satisfies $\Lambda(t) = \Lambda(-t)$ for all $t \in \mathbb{R}$, so we may assume that $t \geq 0$. Let

$$M(t) = \mathbb{E}\Big[\exp(tX)\Big] = \bar{c}_p \int_{\mathbb{R}} \exp\left(tx - \frac{|x|^p}{p}\right) dx.$$

By the change of variables $x = t^{1/(p-1)}y$, we have

$$M(t) = \bar{c}_p t^{\frac{1}{p-1}} \int_{\mathbb{R}} \exp\left(t^{\frac{p}{p-1}}\left(y - \frac{|y|^p}{p}\right)\right) dy.$$

Define $S(y) \coloneqq y - \frac{|y|^p}{p}$ for $y \in \mathbb{R}$. Note that $y_0 = 1$ is the unique maximizer of S, S is smooth in a neighborhood of $y = 1$, and $S''(1) = 1 \neq 0$. An application of Laplace's method (see [169, Section 19.2.5, Theorem 2(b) & Remark 4]) shows that, as $t \to \infty$,

$$M(t) = \bar{c}_p t^{\frac{1}{p-1}} \sqrt{\frac{2\pi}{p-1}} \exp\left(t^{\frac{p}{p-1}} \frac{p-1}{p}\right) t^{-\frac{p}{2(p-1)}} \left(1 + O\left(t^{-\frac{p}{p-1}}\right)\right),$$

where in each case, the implicit constant in the big O notation depends only on p. Similarly, as $t \to \infty$,

$$M'(t) = \bar{c}_p t^{\frac{2}{p-1}} \sqrt{\frac{2\pi}{p-1}} \exp\left(t^{\frac{p}{p-1}} \frac{p-1}{p}\right) t^{-\frac{p}{2(p-1)}} \left(1 + O\left(t^{-\frac{p}{p-1}}\right)\right),$$

$$M''(t) = \bar{c}_p t^{\frac{3}{p-1}} \sqrt{\frac{2\pi}{p-1}} \exp\left(t^{\frac{p}{p-1}} \frac{p-1}{p}\right) t^{-\frac{p}{2(p-1)}} \left(1 + O\left(t^{-\frac{p}{p-1}}\right)\right).$$

Therefor, as $t \to \infty$, we also have

$$\begin{aligned}
\widehat{\Lambda}(t) &= \log\left(M(t)\right) = O\left(t^{\frac{p}{p-1}}\right), \\
|\widehat{\Lambda}'(t)| &= \left|\frac{M'(t)}{M(t)}\right| = t^{\frac{1}{p-1}} \left(1 + O\left(t^{-\frac{p}{p-1}}\right)\right), \\
\widehat{\Lambda}''(t) &= \frac{M''(t)M(t) - \left(M'(t)\right)^2}{\left(M(t)\right)^2} = O\left(t^{\frac{2-p}{p-1}}\right).
\end{aligned}$$

Since $p \geq 2$, the conclusions of the lemma follow from the previous displays and $p \geq 2$. $\square$

Lemma 5.8.2. *Fix $p \geq 2$, let X be as in Lemma 5.8.1, and define*

$$\widehat{\Lambda}\left(t_1, t_2\right) \coloneqq \mathbb{E}\Big[\exp(t_1 X + t_2 |X|^p)\Big], \quad (t_1, t_2) \in \mathbb{R}^2.$$

Then for every pair of integers $\alpha, \beta \geq 0$ with $\alpha + \beta \leq 2$, there exists a continuous map

$(0, p^{-1}) \ni t_2 \mapsto C_{t_2,\alpha,\beta} \in (0, \infty)$ *such that*

$$\partial_1^{\alpha} \partial_2^{\beta} \widehat{\Lambda}\,(t_1, t_2) \leq C_{t_2,\alpha,\beta} \left(1 + |t_1|^{2-\alpha}\right), \quad t_1 \in \mathbb{R}.$$

Proof. Let $M(t) = \mathbb{E}[\exp(tX)]$. It was shown in [96, Lemma 5.7] that for $(t_1, t_2) \in \mathbb{R} \times (-\infty, p^{-1})$,

$$\widehat{\Lambda}\,(t_1, t_2) = -\frac{1}{p} \log\left(1 - pt_2\right) + \log M \left(\frac{t_1}{\left(1 - pt_2\right)^{1/p}}\right).$$

The lemma follows from this representation and Lemma 5.8.1. $\square$

5.9 Weingarten Calculus

This appendix contains some preliminary remarks on the Weingarten calculus, and then a lemma necessary for the Gaussian approximation results proved in Section 5.6.4 and Section 5.6.5. We recall the following definitions from [59]. Fix $d \in \mathbb{N}$ and let $\mathcal{M}(2d)$ be the set of pair partitions m_{sc} of $\{1, 2, \ldots, 2d\}$; these are partitions where each block of the element has exactly two elements. They have a canonical form $\{(m_{\text{sc}}(1), m_{\text{sc}}(2)), \ldots, (m_{\text{sc}}(2d-1), m_{\text{sc}}(2d))\}$ for $m_{\text{sc}}(2i-1) \leq m_{\text{sc}}(2i)$ and $m_{\text{sc}}(1) < m_{\text{sc}}(3) < \cdots m_{\text{sc}}(2d-1)$.

Given pair partitions $m_{\text{sc}}, \mathfrak{n} \in \mathcal{M}(2d)$, we let $\Gamma(m_{\text{sc}}, \mathfrak{n})$ denote the graph that has vertices $\{1, 2, \ldots, 2d\}$, and edges $\{(m_{\text{sc}}(2i-1), m_{\text{sc}}(2i)), (\mathfrak{n}(2i-1), \mathfrak{n}(2i))\}_{i=1}^{d}$. The Gram matrix $G_d^{(n)} = [G^{(n)}(m_{\text{sc}}, \mathfrak{n})]_{m_{\text{sc}}, \mathfrak{n} \in \mathcal{M}(2d)}$ is defined through its entries

$$G^{(n)}(m_{\text{sc}}, \mathfrak{n}) = n^{\text{loop}(m_{\text{sc}}, \mathfrak{n})},$$

where $\text{loop}(m_{\text{sc}}, \mathfrak{n})$ is defined as the number of connected components of $\Gamma(m_{\text{sc}}, \mathfrak{n})$. The matrix $\mathrm{Wg}^{(n)} = \mathrm{Wg}_d^{(n)}$ is defined as the pseudo-inverse of $G_d^{(n)}$. We denote its entries by

$$\mathrm{Wg}^{(n)} = [\mathrm{Wg}^{(n)}(m_{\mathrm{sc}}, \mathfrak{n})]_{m_{\mathrm{sc}}, \mathfrak{n} \in \mathcal{M}(2d)}.$$

We recall the following theorem, which was originally proved in [60].

Theorem 5.9.1 ([59, Theorem 2.1])**.** *Given $i_1, \ldots i_{2d}$ and $j_1, \ldots, j_{2d}$ in $\{1, 2, \ldots, n\}$,*

$$\begin{aligned}\int_{g \in O_n} & g_{i_1 j_1} \cdots g_{i_{2d} j_{2d}}\, dg \\ &= \sum_{m_{\mathrm{sc}}, \mathfrak{n} \in \mathcal{M}(2d)} \mathrm{Wg}^{(n)}(m_{\mathrm{sc}}, \mathfrak{n}) \prod_{k=1}^{d} \delta(i_{m_{\mathrm{sc}}(2k-1)}, i_{m_{\mathrm{sc}}(2k)}) \delta(j_{m_{\mathrm{sc}}(2k-1)}, j_{m_{\mathrm{sc}}(2k)}),\end{aligned} \tag{5.9.1}$$

where $\delta(i, j)$ denotes the Kronecker delta function satisfying $\delta(i, j) = 1$ if $i = j$ and $\delta(i, j) = 0$ otherwise.

Note that the sum on the right side of (5.9.1) is over all choices of pair partitions that pair indices with the same value. In particular, Theorem 5.9.1 shows that any moment $g_{i_1 j_1} \cdots g_{i_{2d} j_{2d}}$ without an even number of entries in each row and column vanishes.

Example 5.9.2. *In the expectation of $g_{11}^2 g_{22}^2$ computed according to (5.9.1), the only possibility for a nonzero term in the sum is that both partitions are $m_{\mathrm{sc}} = \mathfrak{n} = \{(1, 2), (3, 4)\}$.*

Considering m_{sc} and $\mathfrak{n}$ as members of the symmetric group S_{2d} (products of transpositions), we note that the value of $\mathrm{Wg}^{(n)}(m_{\mathrm{sc}}, \mathfrak{n})$ depends only on the value of $\sigma = m_{\mathrm{sc}}^{-1}\mathfrak{n}$ [59, Theorem 3.1]. Consider the graph $\overline{\Gamma}(\sigma)$ whose vertex set is $\{1, 2, \ldots, 2d\}$ with edges $\{2i - 1, 2i\}$ and $\{\sigma(2i - 1), \sigma(2i)\}$ for $1 \leq i \leq d$. Then $\Gamma(\sigma)$ has connected components of even sizes $2\rho_1 \geq 2\rho_2 \geq \cdots$, which determine a partition $\rho = (\rho_1, \rho_2, \ldots)$ of d (in the number-theoretic sense). The length $\ell(\rho)$ of ρ is defined to be the number of elements ρ_i that it contains. We let $\mathrm{Wg}^{(n)}(\rho)$ equal the value of any $\mathrm{Wg}^{(n)}(m_{\mathrm{sc}}, \mathfrak{n})$ such that $\sigma = m_{\mathrm{sc}}^{-1}\mathfrak{n}$ has the corresponding partition ρ of d.

Theorem 5.9.3 ([60, Corollary 2.7])**.** *Fix $d \in \mathbb{N}$ and a partition ρ of d. As $n \to \infty$,*

$$\mathrm{Wg}^{(n)}(\rho) = \left(\prod_{i\geq 1} c_{\rho_i - 1}\right) n^{-2d+\ell(\rho)}(1 + O(n^{-1})).$$

Here $c_k = \frac{(2k)!}{(k+1)!k!}$ denotes the k-th Catalan number, and the implicit constant in the asymptotic notation depends on d and ρ, but not n.

Example 5.9.4. *If $m_{\mathrm{sc}} = \mathfrak{n} = \{(1,2),(3,4)\}$ as in the previous example, then $\sigma = m_{\mathrm{sc}}^{-1}\mathfrak{n}$ is the identity permutation. The graph $\Gamma(\sigma)$ has the edges $\{1,2\}$ and $\{3,4\}$, so $\ell(\rho) = 2$, and we conclude the leading order term in the asymptotic is n^{-2}, as in the Gaussian case. The coefficient of this term is 1 because $c_0 = 1$.*

We next recall Wick's theorem on the expectations of products of centered jointly normal random variables.

Theorem 5.9.5 ([107, Theorem 1.28])**.** *Let $g_1, ..., g_n$ be independent normal random variables with mean zero and variance one. Then*

$$\mathbb{E}[g_1 \cdot \dots \cdot g_n] = \begin{cases} 0, & \textit{if } n \textit{ is odd,} \\ \sum_{m_{\mathrm{sc}} \in \mathcal{M}(n)} \prod_{i=1}^{n/2} \mathbb{E}\left[g_{m_{\mathrm{sc}}(2i-1)} g_{m_{\mathrm{sc}}(2i)}\right] & \textit{otherwise.} \end{cases}$$

Lemma 5.9.6. *Fix $k, n \in \mathbb{N}$, and suppose $\boldsymbol{a} \in \mathbb{V}_{n,k}$ is a random matrix distributed according to the Haar measure on $\mathbb{V}_{n,k}$. Then for all $m \in \mathbb{N}$,*

$$\mathbb{E}\left[\prod_{i=1}^{m} \sqrt{n} a_{1j_i}\right] = \mathbb{E}\left[\prod_{i=1}^{m} g_{j_i}\right]\left(1 + O\left(n^{-1}\right)\right). \tag{5.9.2}$$

where the g_i are independent standard Gaussian random variables, and the implicit constant depends only on m.

Proof. By symmetry of Haar measure and Gaussian random variables, both sides of (5.9.2) are zero when m is odd. It remains to consider the case $m = 2d$ for some $d \in \mathbb{N}$.

Let $\Xi(\sigma)$ denote the coset type of $\sigma \in S_{2d}$ and set

$$\begin{pmatrix} \mathfrak{m} \\ i_1 \ldots i_{2d} \end{pmatrix} = \prod_{l=1}^{d} \delta_{i_{\mathfrak{m}(2l-1),\mathfrak{m}(2l)}}.$$

Let $\mathbf{1}_{2d} \in S_{2d}$ be the identity permutation. With these notations, we have

$$\begin{aligned}
&\mathbb{E}\left[\prod_{i=1}^{2d} \sqrt{n} a_{1 j_i}\right] \\
=& n^d \sum_{\mathfrak{m},\mathfrak{n} \in \mathcal{M}(2d)} \mathrm{Wg}^{(n)}(\mathfrak{m},\mathfrak{n}) \begin{pmatrix} \mathfrak{n} \\ j_1 \ldots j_{2d} \end{pmatrix} \\
=& n^d \sum_{\rho \vdash d} \mathrm{Wg}^{(n)}(\rho) \sum_{\Xi(\mathfrak{m}^{-1}\mathfrak{n}) = \rho} \begin{pmatrix} \mathfrak{n} \\ j_1 \ldots j_{2d} \end{pmatrix} \\
=& n^d \, \mathrm{Wg}^{(n)}\left((\mathbf{1}_{2d}), n\right) \mathbb{E}\left[\prod_{i=1}^{2d} g_{j_i}\right] + n^d \sum_{\substack{\rho \vdash d \\ \rho \neq \mathbf{1}_{2d}}} \mathrm{Wg}^{(n)}(\rho) \sum_{\Xi(\mathfrak{m}^{-1}\mathfrak{n}) = \rho} \begin{pmatrix} \mathfrak{n} \\ j_1 \ldots j_{2d} \end{pmatrix},
\end{aligned} \tag{5.9.3}$$

where the first equality follows from Theorem 5.9.1, the second equality follows from the fact that $\mathrm{Wg}^{(n)}(\mathfrak{m},\mathfrak{n})$ depends only on the coset type of $m_{\mathrm{sc}}^{-1}\mathfrak{n}$ (see [59, Theorem 3.1]), and the third equality follows from Wick's Theorem, Theorem 5.9.5.

By Theorem 5.9.3, the first term in (5.9.3) is

$$\mathbb{E}\left[\prod_{i=1}^{2d} g_{j_i}\right] \left(1 + O\left(n^{-1}\right)\right), \tag{5.9.4}$$

and the absolute value of the second term of (5.9.3) is upper bounded by

$$n^d \mathbb{E}\left[\prod_{i=1}^{2d} g_{j_i}\right] \sum_{\substack{\rho \vdash d \\ \rho \neq \mathbf{1}_{2d}}} |\,\mathrm{Wg}^{(n)}(\rho)| \leq C_d \mathbb{E}\left[\prod_{i=1}^{2d} g_{j_i}\right] n^{-1} \tag{5.9.5}$$

for n sufficiently large, where $C_d = \max_{\rho \vdash d} \prod_{i \geq 1} c_{\rho_i - 1} + 1 < \infty$.

Combining (5.9.3), (5.9.4) and (5.9.5), we obtain the desired conclusion. $\square$

5.10 Proof of Auxiliary Lemmas in Section 5.7

Maintain the notations in Section 5.7.1. The proof of Lemma 5.7.4 is split into several parts, which we first briefly summarize:

1. Establish the concentration of the quantity of interest around the expectation taken over the projection direction $\boldsymbol{a}$. This step requires a truncation argument and good control of the truncated part , which is the subject of Section 5.10.1 (see Lemma 5.10.2).

2. Prove the Gaussian approximation using the result of the first step. This second step is achieved through moment calculations, which are the subject of Section 5.10.2 (see Lemma 5.10.3 and Lemma 5.10.4). Lemma 5.10.3 uses the Weingarten calculus (see Section 5.9) and Lemma 5.10.4 is a multi-dimensional moment convergence theorem, which we include for completeness.

3. Finally, Section 5.10.3 contains the proof of the asymptotic decorrelation result of Lemma 5.7.6.

5.10.1 Preliminary estimate

Remark 5.10.1. In this appendix, we will use the fact that the rows $\{\boldsymbol{a}_i^{(n)}\}_{i=1}^n$ of the matrix $\boldsymbol{a}^{(n)} \in \mathbb{V}_{n,k_n}$ are exchangeable. This follows from the fact that $\boldsymbol{a}^{(n)}$ has the same distribution as the matrix obtained by taking a Haar-distributed element of O_n and removing the last $n - k_n$ columns.

Lemma 5.10.2. *Fix $C, R \in (0, \infty)$, and let $h_R : \mathbb{R} \to \mathbb{R}$ be a function satisfying $|h_R(y)| \leq CR^{-1}y^2$ for all $y \in \mathbb{R}$. Then the measure $\nu_j^{(n)}$ defined in (5.7.4) satisfies*

$$\left|\mathbb{E}\left[\int_{\mathbb{R}^2} x h_R(y)\, d\nu_j^{(n)}(x, y)\right]\right| \leq \frac{C_1 \|\boldsymbol{u}^{(n)}\|_2^2}{R}, \qquad \textit{for } j = 1, \ldots, m + M, \tag{5.10.1}$$

where the constant $C_1 > 0$ depends only on C. We also have the deterministic bounds

$$\int_{\{|x|\geq R\}\times\mathbb{R}} |x|\, d\nu_j^{(n)}(x,y) \leq \frac{1}{R}, \text{ for } j = 1,\dots,m+M. \tag{5.10.2}$$

Proof. For $j = 1,\dots,m$, applying first Jensen's inequality, next the Cauchy-Schwartz inequality and the fact that $(\boldsymbol{a}^{(n),\top})_j$, the j-th column of $\boldsymbol{a}^{(n)}$, lies in $\mathbb{S}^{n-1}$, then Remark 5.10.1, the rotational invariance of the Haar measure σ_n, and the fact that $\sqrt{n}a_{11}$ is sub-Gaussian with $\|\sqrt{n}a_{11}\|_{\psi_2}$ bounded by a constant independent of n [164, Theorem 3.4.6], we have

$$\begin{aligned}
\left(\mathbb{E}\left[\sum_{i=1}^n a_{ij}^{(n)} h_R\left(\left\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\right\rangle\right)\right]\right)^2 &\leq \mathbb{E}\left[\left(\sum_{i=1}^n a_{ij}^{(n)} h_R\left(\left\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\right\rangle\right)\right)^2\right] \\
&\leq \mathbb{E}\left[\sum_{i=1}^n h_R\left(\left\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\right\rangle\right)^2\right] \\
&= n\mathbb{E}\left[h_R\left(\left\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_1\right\rangle\right)^2\right] \\
&= n\mathbb{E}\left[h_R\left(\|\boldsymbol{u}^{(n)}\|_2\sqrt{n}a_{11}\right)^2\right] \\
&\leq C^2 n\|\boldsymbol{u}^{(n)}\|_2^4\mathbb{E}\left[\frac{(\sqrt{n}a_{11})^4}{R^2}\right] \\
&\leq \frac{C_1 n\|\boldsymbol{u}^{(n)}\|_2^4}{R^2}.
\end{aligned}$$

Inequality (5.10.1) now follows by the definition of $\nu_n^{(j)}$. The inequality (5.10.2) can be proved by noticing that

$$\int_{\{|x|\geq R\}\times\mathbb{R}} |x|\, d\nu_j^{(n)}(x,y) \leq \frac{1}{R}\int_{\mathbb{R}^2} |x|^2 d\nu_j^{(n)}(x,y) = \frac{1}{R}\sum_{i=1}^n \left(a_{ij}^{(n)}\right)^2 = \frac{1}{R},$$

where the last inequality uses the fact that $(\boldsymbol{a}^{(n),\top})_j \in \S^{n-1}$. $\square$

5.10.2 Proofs of Gaussian approximations

Lemma 5.10.3. *For any fixed $\alpha, \beta \in \mathbb{N} \cup \{0\}$, we have*

$$\mathbb{E}\left[\int x^\alpha y^\beta d\nu_j^{(n)}\right] = \left(\int x^\alpha y^\beta d\nu_j\right)\left(1 + O\left(\frac{1}{n}\right)\right),$$

where the implicit constant depends only on α, β.

Proof. Using first (5.7.4), then Remark 5.10.1, and then Lemma 5.9.6 with $m = \alpha + \beta$, we have

$$\begin{aligned}
\mathbb{E}_n\left[\int x^\alpha y^\beta d\nu_j^{(n)}\right] &= \frac{1}{n}\mathbb{E}_n\left[\sum_{i=1}^{n}\left(\sqrt{n}a_{ij}^{(n)}\right)^\alpha \cdot \left\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\right\rangle^\beta\right] \\
&= \mathbb{E}_n\left[\left(\sqrt{n}a_{1j}^{(n)}\right)^\alpha \cdot \left(\sum_{\ell=1}^{k_n} u_\ell \sqrt{n}a_{1\ell}^{(n)}\right)^\beta\right] \\
&= \sum_{i_1,\cdots,i_\beta=1}^{k_n}\prod_{r=1}^{\beta} u_{i_r}^{(n)} \cdot \mathbb{E}_n\left[\left(\sqrt{n}a_{1j}^{(n)}\right)^\alpha \cdot \prod_{r=1}^{\beta}\sqrt{n}a_{1i_r}^{(n)}\right] \qquad (5.10.3) \\
&= \sum_{i_1,\cdots,i_\beta=1}^{k_n}\prod_{r=1}^{\beta} u_{i_r}^{(n)} \cdot \mathbb{E}\left[(g_j)^\alpha \cdot \prod_{r=1}^{\beta} g_{i_r}\right]\left(1 + O\left(\frac{1}{n}\right)\right) \\
&= \mathbb{E}\left[(g_j)^\alpha \cdot \left(\sum_{i=1}^{k_n} u_i^{(n)} g_i\right)^\beta\right]\left(1 + O\left(\frac{1}{n}\right)\right),
\end{aligned}$$

where the implicit constant depends only on α, β. By the definition of ν_j in (5.7.5), we have

$$\mathbb{E}\left[(g_j)^\alpha \cdot \left(\sum_{i=1}^{k_n} u_i^{(n)} g_i\right)^\beta\right] = \int x^\alpha y^\beta d\nu_j.$$

Together with (5.10.3), this proves the lemma. □

Lemma 5.10.4. *A sequence of probability measures $\{\mu_n\}$ on $\mathbb{R}^2$ converges weakly to a probability measure μ if the following conditions are satisfied:*

1. *All moments of μ are finite.*

2. *All moments of μ_n are finite and*

$$\int x^{\alpha} y^{\beta} d\mu_n(x,y) \to \gamma_{\alpha,\beta}, \quad \forall\, \alpha, \beta \in \mathbb{N} \cup \{0\},$$

where

$$\gamma_{\alpha,\beta} \coloneqq \int x^{\alpha} y^{\beta} d\mu_n(x,y), \quad \forall\, \alpha, \beta \in \mathbb{N} \cup \{0\}.$$

3. *μ is uniquely determined by $\{\gamma_{\alpha,\beta}\}_{\alpha,\beta\in\mathbb{N}\cup\{0\}}$.*

Proof. By property (1) and (2), for each polynomial P, we have

$$C_p \coloneqq \sup_{n\geq 1} \int P d\mu_n < \infty. \tag{5.10.4}$$

Let $B_R \coloneqq \{\boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x}\|_2 \leq R\}$. Then Markov's inequality implies that

$$\mu_n\left((B_R)^c\right) \leq \frac{C_{x^2+y^2}}{R^2}. \tag{5.10.5}$$

Therefore, μ_n is tight. By Prokhorov's Theorem, it suffices to show that if $\mu_{n_k} \to \nu$ weakly, then $\nu = \mu$.

Pick any polynomial P. For any $R \in (0,\infty)$ define φ_R to be a nonnegative continuous function such that $\mathbf{1}_{B_R} \leq \varphi_R \leq \mathbf{1}_{B_{R+1}}$ and φ_R is monotonically increasing in R. We write

$$\int P d\mu_{n_k} = \int \varphi_R P d\mu_{n_k} + \int (1-\varphi_R) P d\mu_{n_k}. \tag{5.10.6}$$

For the first term the right-hand side of (5.10.6), since μ_{n_k} weakly converges to ν, and $\varphi_R P$ is bounded and continuous, it follows that

$$\lim_{k\to\infty} \int \varphi_R P d\mu_{n_k} = \int \varphi_R P d\nu.$$

For the second term, by the Cauchy–Schwarz inequality, (5.10.4), and (5.10.5), we have

$$\left|\int (1-\varphi_R) P d\mu_{n_k}\right|^2 \leq \mu_{n_k}\left((B_R)^c\right) \int P^2 d\mu_{n_k} \leq \frac{C_{x^2+y^2} C_{P^2}}{R^2}.$$

Moreover, properties (1) and (2) imply that

$$\lim_{k\to\infty} \int P\, d\mu_{n_k} = \int P\, d\mu < \infty. \tag{5.10.7}$$

Combining (5.10.6)-(5.10.7), we have

$$\lim_{R\to\infty} \int \varphi_R P\, d\nu = \int P\, d\mu. \tag{5.10.8}$$

Since P is arbitrary, replacing P by P^2 in the above display and applying the monotone convergence theorem implies that

$$\int P^2 d\nu = \lim_{R\to\infty} \int \varphi_R P^2 d\nu = \int P^2\, d\mu < \infty.$$

Therefore, $P \in L^2(\nu) \subset L^1(\nu)$. An application of the dominated convergence theorem shows that the left-hand side of (5.10.8) is equal to $\int P\, d\nu$ and hence that

$$\int P\, d\nu = \int P\, d\mu.$$

Again, since P is arbitrary, setting $P = x^\alpha y^\beta\, (\alpha, \beta \in \mathbb{N} \cup \{0\})$, and invoking property (3), we have $\mu = \nu$. This completes the proof. □

Proof of Lemma 5.7.4. In this proof, we use C to denote a constant that may depend on C_0, whose value might change from line to line.

A straightforward differentiation of the function $f_j^{(n)}$ defined in (5.7.15) shows that for

$i = 1, \ldots, n$ and $l = 1, 2, \ldots, k_n$,

$$\partial_{il} f_j^{(n)}\left(\boldsymbol{a}^{(n)}\right) = \begin{cases} \frac{1}{\sqrt{n}} h\left(\left\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\right\rangle\right) + a_{ij}^{(n)} h'\left(\left\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\right\rangle\right) u_j^{(n)}, & l = j, \\ a_{ij}^{(n)} h'\left(\left\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\right\rangle\right) u_l^{(n)}, & l \neq j. \end{cases}$$

Hence, by (5.7.16) and the definition of $\boldsymbol{u}^{(n)}$ in (5.7.3), it follows that for $j = 1, \ldots, n$ that

$$\begin{aligned} \left\|\nabla f_j^{(n)}\left(\boldsymbol{a}^{(n)}\right)\right\|_2^2 =& \sum_{i=1}^{n} \sum_{l=1}^{k_n} \left(\partial_{il} f_j^{(n)}\left(\boldsymbol{a}^{(n)}\right)\right)^2 \\ \leq& \frac{2}{n} \sum_{i=1}^{n} \left(h\left(\left\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\right\rangle\right)\right)^2 \\ &+ 2 \sum_{i=1}^{n} \left(a_{ij}^{(n)} h'\left(\left\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)}\right\rangle\right)\right)^2 \sum_{l=1}^{k_n} \left(u_j^{(n)}\right)^2 \\ \leq& \frac{C}{n} \sum_{i=1}^{n} \left(1 + n\left\langle \boldsymbol{u}^{(n)}, \boldsymbol{a}_i^{(n)}\right\rangle^2\right) + C \sum_{i=1}^{n} \left(a_{ij}^{(n)}\right)^2 \sum_{l=1}^{k_n} \left(u_j^{(n)}\right)^2 \\ \leq& C\left(1 + \left\|\boldsymbol{u}^{(n)}\right\|_2^2\right) \\ =& C\left(1 + \|\hat{\mathbf{u}}\|_2^2 + b^2\right) \end{aligned}$$

This implies that $f_j^{(n)}$ is Lipshitz continuous with Lipshitz constant $\sqrt{C\left(1 + \|\hat{\mathbf{u}}\|_2^2 + b^2\right)}$. By Lemma 5.6.4, there exists $C' > 0$ such that for any $\varepsilon > 0$

$$\sigma_n\left(\left|f_j^{(n)}\left(\boldsymbol{a}^{(n)}\right) - \mathbb{E}_n\left[f_j^{(n)}\left(\boldsymbol{a}^{(n)}\right)\right]\right| > \varepsilon\right) \leq \exp\left(-\frac{C'\varepsilon^2 n}{1 + \|\hat{\mathbf{u}}\|_2^2 + b^2}\right).$$

Together with the Borel–Cantelli lemma, this implies that, for σ-a.e. $\boldsymbol{a} = \left(\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}, \ldots\right) \in \mathbb{V}$,

$$\lim_{n\to\infty} \left|f_j^{(n)}\left(\boldsymbol{a}^{(n)}\right) - \mathbb{E}_n\left[f_j^{(n)}\left(\boldsymbol{a}^{(n)}\right)\right]\right| = 0. \tag{5.10.9}$$

Recall the definition of $\nu_j^{(n)}$ in (5.7.4), we may rewrite

$$f_j^{(n)}\left(\boldsymbol{a}^{(n)}\right) = \int_{\mathbb{R}^2} x\partial_1\bar{\Lambda}\left(y, c\right) d\nu_j^{(n)}(x, y). \tag{5.10.10}$$

By Lemma 5.10.3, Lemma 5.10.4, and [140, Theorem 1.7], it follows that $\mathbb{E}\left[\int f d\nu_j^{(n)}\right] \to \int f d\nu_j$ for all $f : \mathbb{R}^2 \to \mathbb{R}$ with polynomial growth. By Assumption 5, $f(x, y) \coloneqq x\partial_1\bar{\Lambda}(y, c)$ can be bounded by a quadratic polynomial. Hence, by (5.10.10)

$$\lim_{n\to\infty} \left| \mathbb{E}_n \left[f_j^{(n)} \left(\boldsymbol{a}^{(n)} \right) \right] - \int_{\mathbb{R}^2} x\partial_1\bar{\Lambda}(y, c) d\nu_j \right| = 0. \tag{5.10.11}$$

Combining (5.10.9), (5.10.11) and recalling the definition of ν_j in (5.7.5), the proof of the lemma is complete. □

5.10.3 Proof of asymptotic decorrelation

The proof of Lemma 5.7.6 relies on the sub-Gaussianity of a random vector chosen uniformly from a "large enough" portion of the scaled unit sphere, which is the content of the next lemma. We recall that the sub-Gaussian norm $\|\cdot\|_{\psi_2}$ was defined in (5.6.6).

Lemma 5.10.5. *There exists a universal constant $K \in (0, \infty)$ such that the following holds. Fix $m, n \in \mathbb{N}$ such that $m < n/2$, and a collection $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m \in \mathbb{R}^n$ of random vectors that are almost surely orthonormal. Let $\boldsymbol{v}$ be a random vector whose distribution conditional on $\{\boldsymbol{v}_i\}_{i=1}^m$ is uniform on the subset of $\mathbb{S}^{n-1}$ orthogonal to the subspace generated by $\{\boldsymbol{v}_i\}_{i=1}^m$. Then*

$$\|\sqrt{n}\boldsymbol{v}\|_{\psi_2} \leq K.$$

Proof. Let $O \in O_n$ be a random matrix such such that $O\boldsymbol{v}_i = \boldsymbol{e}_i$ for all $i = 1, \ldots, m$. By the assumption that $\boldsymbol{v}$ is perpendicular to each $\boldsymbol{v}_i$, it follows that $(O\boldsymbol{v})_i = 0$ for all $i = 1, \ldots, m$ and $\hat{\boldsymbol{v}} \coloneqq ((O\boldsymbol{v})_i)_{i=m+1}^n$ is distributed uniformly on $\mathbb{S}^{n-m}$. Hence, by the definition of the sub-Gaussian norm in (5.6.6), which shows it is invariant under orthogonal transformations,

$$\|\boldsymbol{v}\|_{\psi_2} = \|O\boldsymbol{v}\|_{\psi_2} = \sup_{\boldsymbol{x}\in\mathbb{S}^{n-1}} \|\langle O\boldsymbol{v}, \boldsymbol{x}\rangle\|_{\psi_2} = \sup_{\substack{\boldsymbol{x}\in\mathbb{S}^{n-1} \\ x_i=0,\, 1\leq i\leq m}} \|\langle O\boldsymbol{v}, \boldsymbol{x}\rangle\|_{\psi_2} = \|\hat{\boldsymbol{v}}\|_{\psi_2} \tag{5.10.12}$$

Since $\hat{\boldsymbol{v}}$ is uniformly distributed on $\mathbb{S}^{n-m}$, by [164, Theorem 3.4.6] there exists a universal constant $K \in (0, \infty)$ such that

$$\|\sqrt{(n-m)}\hat{\boldsymbol{v}}\|_{\psi_2} \leq \frac{K}{\sqrt{2}}. \tag{5.10.13}$$

Combining (5.10.12), (5.10.13), and using the hypothesis that $m < n/2$, the proof is complete. □

Proof of Lemma 5.7.6. Let

$$S_n(\boldsymbol{a}^{(n)}) \coloneqq \sum_{\ell=m+M+1}^{k_n} \left(f_\ell^{(n)}\left(\boldsymbol{a}^{(n)}\right)\right)^2.$$

Fix $\lambda > 0$, and let $\boldsymbol{Z} \coloneqq (Z_{m+M+1}, \ldots, Z_{k_n})^\top$ be a random vector of i.i.d. standard Gaussian random variables independent of $\boldsymbol{a}^{(n)}$. Since $\mathbb{E}\left[\exp\left(xZ_\ell\right)\right] = \exp(x^2/2)$ for any $x \in \mathbb{R}$ and $\ell = m+M+1, \ldots, k_n$, we have, recalling the definition of $f_\ell(\boldsymbol{a}^{(n)})$ in (5.7.15),

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\frac{\lambda^2}{2}S_n(\boldsymbol{a}^{(n)})\right)\right] &= \mathbb{E}\left[\prod_{\ell=m+M+1}^{k_n} \mathbb{E}\left[\exp\left(\frac{1}{2}\left(\lambda f_\ell^{(n)}(\boldsymbol{a}^{(n)})\right)^2\right) \middle| \boldsymbol{a}^{(n)}\right]\right] \\
&= \mathbb{E}\left[\prod_{\ell=m+M+1}^{k_n} \mathbb{E}\left[\exp\left(\lambda f_\ell^{(n)}(\boldsymbol{a}^{(n)})Z_\ell\right) \middle| \boldsymbol{a}^{(n)}\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\exp\left(\lambda \sum_{\ell=m+M+1}^{k_n} f_\ell^{(n)}(\boldsymbol{a}^{(n)})Z_\ell\right) \middle| \boldsymbol{a}^{(n)}\right]\right] \\
&= \mathbb{E}\left[\exp\left(\lambda \sum_{\ell=m+M+1}^{k_n} f_\ell^{(n)}(\boldsymbol{a}^{(n)})Z_\ell\right)\right].
\end{aligned} \tag{5.10.14}$$

Next, note that since $\{Z_\ell\}_{\ell=m+M+1}^{k_n}$ are symmetric and independent of $\boldsymbol{a}^{(n)}$, we have

$$\mathbb{E}\left[\lambda \sum_{\ell=m+M+1}^{k_n} Z_\ell f_\ell(\boldsymbol{a}^{(n)})\right] = \lambda \sum_{\ell=m+M+1}^{k_n} \mathbb{E}\left[Z_\ell\right]\mathbb{E}\left[f_\ell(\boldsymbol{a}^{(n)})\right] = 0. \tag{5.10.15}$$

Now, for $i = 1, \dots, n$, set

$$Y_i \equiv Y_i^{(n)}(\boldsymbol{a}^{(n)}) := \frac{1}{n} h \left(\sum_{j=1}^{m+M} \sqrt{n} u_j a_{ij}^{(n)} \right), \qquad \boldsymbol{Y} \equiv \boldsymbol{Y}^{(n)}(\boldsymbol{a}^{(n)}) := (Y_1, \dots, Y_n)^\top, \tag{5.10.16}$$

$$\hat{A} \equiv \hat{A}^{(n)} := (\sqrt{n} a_{i\ell}^{(n)})_{1 \le i \le n, m+M+1 \le \ell \le k_n}.$$

Since $\boldsymbol{Z}$ is a vector of standard Gaussians (and in particular has a rotationally symmetric distribution), there exists a $\mathbb{R}^{k_n - m - M} \times \mathbb{R}^{k_n - m - M}$-valued, Haar-distributed random orthogonal matrix O such that $O^\top \boldsymbol{Z} = \|\boldsymbol{Z}\|_2 \boldsymbol{e}_1$, where $\boldsymbol{e}_1 \in \mathbb{R}^{k_n - m - M}$. By the identity $OO^\top = I$, it follows that

$$\sum_{\ell=m+M+1}^{k_n} Z_\ell f_\ell^{(n)}(\boldsymbol{a}^{(n)}) = \boldsymbol{Y}^\top \hat{A} \boldsymbol{Z} \overset{d}{=} \|\boldsymbol{Z}\|_2 \boldsymbol{Y}^\top \hat{A} O \boldsymbol{e}_1, \tag{5.10.17}$$

Note that $\hat{A} O \boldsymbol{e}_1$ has norm 1 and lies in the subset of $\mathbb{S}^{n-1}$ orthogonal to the first $m+M$ columns of $\boldsymbol{a}^{(n)}$ (that is, $(a_{i\ell}^{(n)})_{1 \le i \le n, 1 \le \ell \le m+M}$). Moreover, conditional on the first $m+M$ columns of $\boldsymbol{a}^{(n)}$, it is uniformly distributed on this subset (since O is Haar-distributed). By Lemma 5.10.5, with m in that lemma statement equal to the quantity $m+M$ in this proof, there exists a constant $K \in (0, \infty)$ such that for all sufficiently large n (depending only on the choice of m and $(\hat{\boldsymbol{u}}, b, c)$), we have $\left\| \hat{A} O \boldsymbol{e}_1 \right\|_{\psi_2} \le K$, where $\| \cdot \|_{\psi_2}$ denotes the sub-Gaussian norm defined in Definition 5.6.1. Then it is immediate from (5.6.6) that the sub-Gaussian norm of $\boldsymbol{Y}^\top \hat{A} O \boldsymbol{e}_1$ satisfies the estimate

$$\|\boldsymbol{Y}^\top \hat{A} O \boldsymbol{e}_1\|_{\psi_2} \le K \|\boldsymbol{Y}\|_2. \tag{5.10.18}$$

Combining (5.10.15), (5.10.17), (5.10.18), (5.6.1), and the exponential moment estimate [164, Proposition 2.5.2(v)], we have

$$\mathbb{E}\left[\exp\left(\lambda \sum_{\ell=m+M+1}^{k_n} Z_\ell f_\ell(\boldsymbol{a}^{(n)}) \right) \right] \le \exp\left(K^2 \lambda^2 \|\boldsymbol{Y}\|_2^2 \|\boldsymbol{Z}\|_2^2 \right). \tag{5.10.19}$$

By (5.10.16) and the assumption (5.7.16) on h, there exists a constant $C \in (0, \infty)$ such that

$$\|\boldsymbol{Y}\|_2^2 \le \frac{C}{n^2} \sum_{i=1}^n \left(1 + \left\langle \boldsymbol{u}^{(n)}, \sqrt{n}\boldsymbol{a}_i^{(n)} \right\rangle^2\right) \le \frac{C}{n}(1 + \|\boldsymbol{u}^{(n)}\|_2^2). \tag{5.10.20}$$

Set $\lambda \coloneqq \sqrt{\frac{n}{8CK^2(1+\|\boldsymbol{u}^{(n)}\|_2^2)}}$. Combining (5.10.14), (5.10.19) and (5.10.20), with Markov's inequality, and recalling the moment generating function for a Gaussian, we have

$$\mathbb{P}\left(S_n > \varepsilon\right) \le e^{-\frac{\lambda^2}{2}\varepsilon}\mathbb{E}\left[e^{\frac{\lambda^2}{2}S_n}\right] \le e^{-\frac{\lambda^2}{2}\varepsilon} \prod_{\ell=m+M+1}^{k_n} \mathbb{E}\left[e^{\frac{Z_\ell^2}{8}}\right] = e^{-\frac{\varepsilon n}{16CK^2(1+\|\boldsymbol{u}^{(n)}\|_2^2)}} \left(\frac{4}{3}\right)^{k_n/2}.$$

Since $k_n = o(n)$ by assumption, by the Borel–Cantelli lemma, we have $S_n \to 0$ σ-a.s. $\square$

References

[1] Radosław Adamczak, Joscha Prochno, Marta Strzelecka, and Michał Strzelecki. Norms of structured random matrices. *Mathematische Annalen*, pages 1–65, 2023.

[2] A. Adhikari, S. Dubova, C. Xu, and J. Yin. Eigenstate thermalization hypothesis for generalized Wigner matrices. *arXiv preprint arXiv:2302.00157*, 2023.

[3] A. Aggarwal. Bulk universality for generalized Wigner matrices with few moments. *Probab. Theory Related Fields*, 173(1-2):375–432, 2019.

[4] A. Aggarwal, C. Bordenave, and P. Lopatto. Mobility edge of Lévy matrices. *arXiv preprint arXiv:2210.09458*, 2022.

[5] A. Aggarwal, P. Lopatto, and J. Marcinek. Eigenvector statistics of Lévy matrices. *Ann. Probab.*, 49(4):1778–1846, 2021.

[6] A. Aggarwal, P. Lopatto, and H.-T. Yau. GOE statistics for Lévy matrices. *J. Eur. Math. Soc.*, 2021.

[7] David Alonso-Gutiérrez, Joscha Prochno, and Christoph Thäle. Large deviations for high-dimensional random projections of ℓ_p^n-balls. *Advances in Applied Mathematics*, 99:1–35, 2018.

[8] N. Anantharaman. Entropy and the localization of eigenfunctions. *Ann. of Math. (2)*, 168:435–475, 2008.

[9] N. Anantharaman and E. Le Masson. Quantum ergodicity on large regular graphs. *Duke Math. J.*, 164(4):723–765, 2015.

[10] N. Anantharaman and M. Sabri. Quantum ergodicity on graphs: from spectral to spatial delocalization. *Ann. of Math. (2)*, 189(3):753–835, 2019.

[11] Greg Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*. Number 118. Cambridge University Press, 2010.

[12] Gérald Ben Arous Antonio Auffinger and Sandrine Péché. Poisson convergence for the largest eigenvalues of heavy tailed random matrices. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*, 45(3):589–610, 2009.

[13] Milla Anttila, Keith Ball, and Irini Perissinaki. The central limit problem for convex bodies. *Transactions of the American Mathematical Society*, 355(12):4723–4735, 2003.

[14] Gérard Ben Arous and Alice Guionnet. The spectrum of heavy tailed random matrices. *Communications in Mathematical Physics*, 278(3):715–751, 2008.

[15] Zhidong Bai and Jack Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. Springer, 2010.

[16] Afonso S Bandeira and Ramon van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability*, pages 2479–2506, 2016.

[17] Z. Bao, L. Erdős, and K. Schnelli. Equipartition principle for Wigner matrices. *Forum Math. Sigma*, 9:1–21, 2021.

[18] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 307–326, 2012.

[19] R. Bauerschmidt, J. Huang, and H.-T. Yau. Local Kesten–Mckay law for random regular graphs. *Comm. Math. Phys.*, pages 1–114, 2016.

[20] F. Benaych-Georges and A. Knowles. Local semicircle law for Wigner matrices. In *Advanced topics in random matrices*, volume 53 of *Panor. Synthèses*, pages 1–90. 2017.

[21] L. Benigni. Eigenvectors distribution and quantum unique ergodicity for deformed Wigner matrices. *Ann. Inst. Henri Poincar ́e Probab. Stat.*, 56(4):2822–2867, 2020.

[22] L. Benigni. Fermionic eigenvector moment flow. *Probab. Theory Related Fields*, 179:733–775, 2021.

[23] L. Benigni and G. Cipolloni. Fluctuations of eigenvector overlaps and the berry conjecture for Wigner matrices. *arXiv preprint arXiv:2212.10694*, 2022.

[24] L. Benigni and P. Lopatto. Fluctuations in local quantum unique ergodicity for generalized Wigner matrices. *Comm. Math. Phys.*, 391(2):401–454, 2022.

[25] L. Benigni and P. Lopatto. Optimal delocalization for generalized Wigner matrices. *Adv. Math.*, 396:108109, 2022.

[26] Grahame Bennett. Schur multipliers. *Duke Math. J.*, 44:603–639, 1977.

[27] Grahame Bennett, Victor Goodman, and Charles Newman. Norms of random matrices. *Pacific Journal of Mathematics*, 59(2):359–365, 1975.

[28] Aditya Bhaskara and Aravindan Vijayaraghavan. Approximating matrix p-norms. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete algorithms*, pages 497–511, 2011.

[29] Vijay Bhattiprolu, Mrinalkanti Ghosh, Venkatesan Guruswami, Euiwoong Lee, and Madhur Tulsiani. Inapproximability of matrix $p \to q$ norms. *arXiv preprint arXiv:1802.07425*, 2018.

[30] P. Billingsley. *Probability and Measure.* John Wiley & Sons, 2008.

[31] Patrick Billingsley. *Probability and Measure.* John Wiley & Sons, 2017.

[32] A. Bloemendal, L. Erdos, A. Knowles, H.-T. Yau, and J. Yin. Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab*, 19(33):1–53, 2014.

[33] A. Bloemendal, A. Knowles, H.-T. Yau, and J. Yin. On the principal components of sample covariance matrices. *Probab. Theory Related Fields*, 164(1):459–552, 2016.

[34] C. Bordenave and A. Guionnet. Localization and delocalization of eigenvectors for heavy-tailed random matrices. *Probab. Theory Related Fields*, 157(3-4):885–953, 2013.

[35] C. Bordenave and A. Guionnet. Delocalization at small energy for heavy-tailed random matrices. *Comm. Math. Phys.*, 354:115–159, 2017.

[36] Charles Bordenave and Djalil Chafaï. Around the circular law. *Probability Surveys*, 9:1–89, 2012.

[37] Alexei Borodin and Christopher D. Sinclair. The Ginibre ensemble of real random matrices and its scaling limits. *Communications in Mathematical Physics*, 291:177–224, 2009.

[38] P. Bourgade, J. Huang, and H.-T. Yau. Eigenvector statistics of sparse random matrices. *Electron. J. Probab.*, 22:1–64, 2017.

[39] P. Bourgade and H.-T. Yau. The eigenvector moment flow and local quantum unique ergodicity. *Comm. Math. Phys.*, 350:231–278, 2017.

[40] P. Bourgade, H.-T. Yau, and J. Yin. Random band matrices in the delocalized phase I: Quantum unique ergodicity and universality. *Comm. Pure Appl. Math.*, 73(7):1526–1596, 2020.

[41] David W Boyd. The power method for ℓ_p norms. *Linear Algebra and its Applications*, 9:95–101, 1974.

[42] Fernando GSL Brandao and Aram W Harrow. Estimating operator norms using covering nets. *arXiv preprint arXiv:1509.05065*, 2015.

[43] Sung-Soo Byun and Peter J. Forrester. Progress on the study of the Ginibre ensembles I: GinUE. *arXiv preprint arXiv:2211.16223*, 2022.

[44] Sung-Soo Byun and Peter J. Forrester. Progress on the study of the Ginibre ensembles II: GinOE and GinSE. *arXiv preprint arXiv:2301.05022*, 2023.

[45] M. Charikar and A. Wirth. Maximizing quadratic programs: extending Grothendieck's inequality. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 54–60, 2004.

[46] Laurent Charles and Benoit Estienne. Entanglement entropy and Berezin–Toeplitz operators. *Communications in Mathematical Physics*, 376(1):521–554, 2020.

[47] Wei-Kuo Chen and Arnab Sen. On ℓ_p-Gaussian–Grothendieck problem. *International Mathematics Research Notices*, 2020.

[48] G. Cipolloni, L. Erdős, J. Henheik, and O. Kolupaiev. Gaussian fluctuations in the equipartition principle for Wigner matrices. *arXiv preprint arXiv:2301.05181*, 2023.

[49] G. Cipolloni, L. Erdős, and D. Schröder. Eigenstate thermalization hypothesis for Wigner matrices. *Comm. Math. Phys.*, 388:1005–1048, 2021.

[50] G. Cipolloni, L. Erdős, and D. Schröder. Normal fluctuation in quantum ergodicity for Wigner matrices. *Ann. Probab.*, 50(3):984–1012, 2022.

[51] G. Cipolloni, L. Erdős, and D. Schröder. Optimal multi-resolvent local laws for Wigner matrices. *Electron. J. Probab.*, 27:1–38, 2022.

[52] G. Cipolloni, L. Erdős, and D. Schröder. Rank-uniform local law for Wigner matrices. *Forum Math. Sigma*, 10:1–43, 2022.

[53] G. Cipolloni, L. Erdős, and D. Schröder. Thermalisation for Wigner matrices. *J. Funct. Anal.*, 282(8):109394, 2022.

[54] Giorgio Cipolloni, László Erdős, and Dominik Schröder. Central limit theorem for linear eigenvalue statistics of non-Hermitian random matrices. *arXiv preprint arXiv:1912.04100*, 2019.

[55] Giorgio Cipolloni, László Erdős, and Dominik Schröder. Fluctuation around the circular law for random matrices with real entries. *Electronic Journal of Probability*, 2021.

[56] Giorgio Cipolloni, László Erdős, and Dominik Schröder. Mesoscopic central limit theorem for non-Hermitian random matrices. *arXiv preprint arXiv:2210.12060*, 2022.

[57] Pierre Cizeau and Jean-Philippe Bouchaud. Mean field theory of dilute spin-glasses with power-law interactions. *Journal of Physics A: Mathematical and General*, 26(5):L187, 1993.

[58] Y Colin de Verdière. Ergodicité et fonctions propres du laplacien. *Communications in Mathematical Physics*, 102:497–502, 1985.

[59] Benoît Collins and Sho Matsumoto. On some properties of orthogonal Weingarten functions. *Journal of Mathematical Physics*, 50(11):113516, 2009.

[60] Benoît Collins and Piotr Śniady. Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Communications in Mathematical Physics*, 264(3):773–795, 2006.

[61] Francis Comets and Amir Dembo. Ordered overlaps in disordered mean-field models. *Probability Theory and Related Fields*, 121(1):1–29, 2001.

[62] Robert M Corless, Gaston H Gonnet, David EG Hare, David J Jeffrey, and Donald E Knuth. On the lambert w function. *Advances in Computational mathematics*, 5:329–359, 1996.

[63] Ovidiu Costin and Joel L. Lebowitz. Gaussian fluctuation in random matrices. *Physical Review Letters*, 75(1):69, 1995.

[64] L. D'Alessio, Y. Kafri, A. Polkovnikov, and M. Rigol. From quantum chaos and

eigenstate thermalization to statistical mechanics and thermodynamics. *Adv. Phys.*, 65(3):239–362, 2016.

[65] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*. Springer, 2nd edition, 2009.

[66] J. Deutsch. Quantum statistical mechanics in a closed system. *Phys. Rev. A*, 43(4):2046, 1991.

[67] Souvik Dhara, Debankur Mukherjee, and Kavita Ramanan. The r-to-p norm of non-negative random matrices: asymptotic normality and entry-wise bounds. *Annals of Applied Probability, to appear, arXiv preprint arXiv:2005.14056*, 2020.

[68] Persi Diaconis and David Freedman. Asymptotics of graphical projection pursuit. *The Annals of Statistics*, pages 793–815, 1984.

[69] William E Donath and Alan J Hoffman. Lower bounds for the partitioning of graphs. *IBM Journal of Research and Development*, 17(5):420–425, 1973.

[70] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[71] Alan Edelman. The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law. *Journal of Multivariate Analysis*, 60:203–232, 1997.

[72] Alan Edelman, Eric Kostlan, and Michael Shub. How many eigenvalues of a random matrix are real? *Journal of the American Mathematical Society*, 7(1):247–267, 1994.

[73] Michele Elia and Giorgio Taricco. Integration of the exponential function of a complex quadratic form. *Applied Mathematics E-Notes*, 3:95–98, 2003.

[74] Matthias Englert and Harald Räcke. Oblivious routing for the lp-norm. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 32–40. IEEE, 2009.

[75] L. Erdős, A. Knowles, and H.-T. Yau. Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré*, 14(8):1837–1926, 2013.

[76] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. The local semicircle law for a general class of random matrices. *Electron. J. Probab*, 18(59):1–58, 2013.

[77] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdős–Rényi graphs I: Local semicircle law. *Ann. Probab*, 41(3B):2279–2375, 2013.

[78] L. Erdős, J. Ramírez, B. Schlein, and H.-T. Yau. Universality of sine-kernel for Wigner matrices with a small Gaussian perturbation. *Electron. J. Probab.*, 15:526–604, 2010.

[79] L. Erdős, B. Schlein, and H.-T. Yau. Local semicircle law and complete delocalization for Wigner random matrices. *Comm. Math. Phys.*, 287(2):641–655, 2009.

[80] L. Erdős, B. Schlein, and H.-T. Yau. Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.*, 37(3):815–852, 2009.

[81] L. Erdős, B. Schlein, and H.-T. Yau. Wegner estimate and level repulsion for Wigner random matrices. *Int. Math. Res. Not. IMRN*, 2010(3):436–479, 2010.

[82] L. Erdős and H.-T. Yau. *A dynamical approach to random matrix theory*, volume 28. American Mathematical Soc., 2017.

[83] L. Erdős, H.-T. Yau, and J. Yin. Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields*, 154(1-2):341–407, 2012.

[84] L. Erdős, H.-T. Yau, and J. Yin. Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math*, 229(3):1435–1515, 2012.

[85] László Erdős. Universality of wigner random matrices: a survey of recent results. *Russian Mathematical Surveys*, 66(3):507, 2011.

[86] Miroslav Fiedler. Algebraic connectivity of graphs. *Czechoslovak mathematical journal*, 23(2):298–305, 1973.

[87] Will FitzGerald and Nick Simm. Fluctuations and correlations for products of real asymmetric random matrices. *arXiv preprint arXiv:2109.00322*, 2021.

[88] Gerald B Folland. *Real analysis: modern techniques and their applications*, volume 40. John Wiley & Sons, 1999.

[89] Peter J. Forrester. Local central limit theorem for real eigenvalue fluctuations of elliptic GinOE matrices. *arXiv preprint arXiv:2305.09124*, 2023.

[90] Peter J Forrester and Joel L Lebowitz. Local central limit theorem for determinantal point processes. *Journal of Statistical Physics*, 157(1):60–69, 2014.

[91] Peter J. Forrester and Anthony Mays. A method to calculate correlation functions for $\beta = 1$ random matrices of odd size. *Journal of Statistical Physics*, 134(3):443–462, 2009.

[92] Peter J Forrester and Taro Nagao. Eigenvalue statistics of the real Ginibre ensemble. *Physical Review Letters*, 99(5):050603, 2007.

[93] Peter J Forrester and Taro Nagao. Skew orthogonal polynomials and the partly symmetric real Ginibre ensemble. *Journal of Physics A: Mathematical and Theoretical*, 41(37):375003, 2008.

[94] Jerome Friedman and John Tukey. A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*, 100(9):881–890, 1974.

[95] Nina Gantert, Steven Soojin Kim, and Kavita Ramanan. Cramér's theorem is atypical. In *Advances in the mathematical sciences*, pages 253–270. Springer, 2016.

[96] Nina Gantert, Steven Soojin Kim, and Kavita Ramanan. Large deviations for random projections of ℓ^p balls. *The Annals of Probability*, 45(6B):4419–4476, 2017.

[97] Martin Gebert and Mihail Poplavskyi. On pure complex spectrum for truncations of random orthogonal matrices and Kac polynomials. *arXiv preprint arXiv:1905.03154*, 2019.

[98] Evarist Giné, Geoffrey R Grimmett, Laurent Saloff-Coste, and Laurent Saloff-Coste. Lectures on finite markov chains. *Lectures on probability theory and statistics: École d'été de Probabilités de Saint-Flour XXVI-1996*, pages 301–413, 1997.

[99] Jean Ginibre. Statistical ensembles of complex, quaternion, and real matrices. *Journal of Mathematical Physics*, 6(3):440–449, 1965.

[100] T. Gobet and N. Williams. Noncrossing partitions and Bruhat order. *European J. Combin.*, 53:8–34, 2016.

[101] F. Götze, A. Naumov, and A. Tikhomirov. Local semicircle law under fourth moment condition. *J. Theoret. Probab.*, 33(3):1327–1362, 2020.

[102] F. Götze, A. Naumov, A. Tikhomirov, and D. Timushev. On the local semicircular law for Wigner ensembles. *Bernoulli*, 24(3):2358–2400, 2018.

[103] Anupam Gupta, Mohammad T Hajiaghayi, and Harald Räcke. Oblivious network design. In *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 970–979, 2006.

[104] Godfrey Harold Hardy, John Edensor Littlewood, and George Pólya. *Inequalities*. Cambridge university press, 1952.

[105] Nicholas J Higham. A survey of condition number estimation for triangular matrices. *SIAM Review*, 29(4):575–596, 1987.

[106] Peter J Huber. Projection pursuit. *The annals of Statistics*, pages 435–475, 1985.

[107] Svante Janson. *Gaussian Hilbert Spaces*. Number 129. Cambridge University Press, 1997.

[108] Tiafeng Jiang. Maxima of entries of Haar distributed matrices. *Probability Theory and Related Fields*, 131:121–144, 2005.

[109] Tiafeng Jiang. How many entries of a typical orthogonal matrix can be approximated by independent normals? *The Annals of Probability*, 34(4):1497 – 1529, 2006.

[110] William B Johnson. Extensions of Lipschitz mappings into a Hilbert space. *Contemp. Math.*, 26:189–206, 1984.

[111] Zakhar Kabluchko, Joscha Prochno, and Christoph Thäle. High-dimensional limit theorems for random vectors in ℓ_p^n-balls. *Communications in Contemporary Mathematics*, 23(03):1950073, 2021.

[112] Eugene Kanzieper and Gernot Akemann. Statistics of real eigenvalues in Ginibre's ensemble of random real matrices. *Physical Review Letters*, 95(23):230201, 2005.

[113] Tom Kaufmann, Holger Sambale, and Christoph Thäle. Large deviations for uniform projections of p-radial distributions on ℓ_p^n-balls. *arXiv preprint arXiv:2203.00476*, pages 1–11, 2022.

[114] Subhash Khot and Assaf Naor. Grothendieck-type inequalities in combinatorial optimization. *Communications on Pure and Applied Mathematics*, 65:0992–1035, 2012.

[115] Steven Soojin Kim, Yin-Ting Liao, and Kavita Ramanan. An asymptotic thin shell condition and large deviations for random multidimensional projections. *Advances in Applied Mathematics*, 134:102306, 2022.

[116] Steven Soojin Kim and Kavita Ramanan. Large deviation principles induced by the Stiefel manifold, and random multi-dimensional projections. *arXiv preprint arXiv:2105.04685*, pages 1–18, 2021.

[117] Bo'az Klartag. A central limit theorem for convex sets. *Inventiones mathematicae*, 168(1):91–131, 2007.

[118] Bo'az Klartag. Power-law estimates for the central limit theorem for convex sets. *Journal of Functional Analysis*, 245(1):284–310, 2007.

[119] A. Knowles and J. Yin. Eigenvector distribution of Wigner matrices. *Probab. Theory Related Fields*, 155(3):543–582, 2013.

[120] Phil Kopel. Linear statistics of non-Hermitian matrices matching the real or complex Ginibre ensemble to four moments. *arXiv preprint arXiv:1510.02987*, 2015.

[121] G. Kreweras. Sur les partitions non croisées d'un cycle. *Discrete Math.*, 1(4):333–350, 1972.

[122] T Kriecherbauer, ABJ Kuijlaars, KD McLaughlin, and PD Miller. Locating the zeros of partial sums of $\exp(z)$ with riemann-hilbert methods. *arXiv preprint arXiv:0709.1213*, 2007.

[123] Aditya Krishnan, Sidhanth Mohanty, and David P Woodruff. On sketching the q to p norms. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 2018.

[124] Rafał Latała and Marta Strzelecka. Chevet-type inequalities for subexponential weibull matrices. *arXiv preprint arXiv:2309.04214*, 2023.

[125] Rafał Latała and Marta Strzelecka. Operator $\ell_p \to \ell_q$ norms of random matrices with iid entries. *arXiv preprint arXiv:2401.09814*, 2024.

[126] J. O. Lee and K. Schnelli. Local law and Tracy–Widom limit for sparse random matrices. *Probab. Theory Related Fields*, 171:543–616, 2018.

[127] Giovanni Leoni. *A first course in Sobolev spaces.* American Mathematical Soc., 2017.

[128] Yin-Ting Liao. *Sharp large deviation estimates and their applications to asymptotic convex geometry.* PhD thesis, Brown University, 2022.

[129] Yin-Ting Liao and Kavita Ramanan. Geometric sharp large deviations for random projections of ℓ_p^n spheres and balls. *arXiv preprint arXiv:2001.04053*, pages 1–38, 2020.

[130] Zhengjiang Lin. Nonlocal energy functionals and determinantal point processes on non-smooth domains. *arXiv preprint arXiv:2304.00118*, 2023.

[131] Elon Lindenstrauss. Invariant measures and arithmetic quantum unique ergodicity. *Annals of Mathematics*, pages 165–219, 2006.

[132] Odalric Maillard and Rémi Munos. Compressed least-squares regression. *Advances in neural information processing systems*, 22:1–9, 2009.

[133] J. Marcinek and H.-T. Yau. High dimensional normality of noisy eigenvectors. *Comm. Math. Phys.*, 395(3):1007–1096, 2022.

[134] George Marsaglia and John CW Marsaglia. A new derivation of stirling's approximation to n! *The American Mathematical Monthly*, 97(9):826–829, 1990.

[135] Ph. A. Martin and T. Yalcin. The charge fluctuations in classical Coulomb systems. *Journal of Statistical Physics*, 22:435–463, 1980.

[136] Anthony Mays. A geometrical triumvirate of real random matrices. *arXiv preprint arXiv:1202.1218*, 2012.

[137] Michael Mitzenmacher and Eli Upfal. *Probability and computing: randomization and probabilistic techniques in algorithms and data analysis.* Cambridge university press, 2017.

[138] Sumit Mukherjee and Subhabrata Sen. Variational inference in high-dimensional linear regression. *The Journal of Machine Learning Research*, 23(1):13703–13758, 2022.

[139] Jayakrishnan Nair, Adam Wierman, and Bert Zwart. The fundamentals of heavy-tails: Properties, emergence, and identification. In *Proceedings of the ACM SIG-*

METRICS/international conference on Measurement and modeling of computer systems, pages 387–388, 2013.

[140] Yann Ollivier, Hervé Pajot, and Cédric Villani. *Optimal Transport: Theory and Applications*, volume 413. Cambridge University Press, 2014.

[141] S. O'Rourke, V. Vu, and K. Wang. Eigenvectors of random matrices: a survey. *J. Combin. Theory Ser. A*, 144:361–442, 2016.

[142] Sean O'Rourke and Noah Williams. Partial linear eigenvalue statistics for non-Hermitian random matrices. *Theory of Probability & Its Applications*, 67(4):613–632, 2023.

[143] Jiaze Qiu and Subhabrata Sen. The TAP free energy for high-dimensional linear regression. *The Annals of Applied Probability*, 33(4):2643–2680, 2023.

[144] Harald Räcke. Optimal hierarchical decompositions for congestion minimization in networks. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 255–264, 2008.

[145] Ralph Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 2015.

[146] Ralph Tyrrell Rockafellar and Roger Wets. *Variational Analysis*, volume 317. Springer Science & Business Media, 2009.

[147] Zeév Rudnick and Peter Sarnak. The behaviour of eigenstates of arithmetic hyperbolic manifolds. *Communications in Mathematical Physics*, 161:195–213, 1994.

[148] P. Sarnak. Recent progress on the quantum unique ergodicity conjecture. *Bull. Amer. Math. Soc*, 48:211–228, 2012.

[149] Gideon Schechtman and Joel Zinn. On the volume of the intersection of two L_p^n balls. *Proceedings of the American Mathematical Society*, 110(1):217–224, 1990.

[150] Alexander I Shnirelman. Ergodic properties of eigenfunctions. *Uspekhi Matematicheskikh Nauk*, 29(6):181–182, 1974.

[151] Nick J. Simm. Central limit theorems for the real eigenvalues of large Gaussian random matrices. *Random Matrices: Theory and Applications*, 6(01):1750002, 2017.

[152] Christopher D. Sinclair. Correlation functions for $\beta = 1$ ensembles of matrices of odd size. *Journal of Statistical Physics*, 136(1):17–33, 2009.

[153] Hans-Jürgen Sommers and Waldemar Wieczorek. General eigenvalue correlations for the real Ginibre ensemble. *Journal of Physics A: Mathematical and Theoretical*, 41(40):405003, 2008.

[154] Alexander Soshnikov. The central limit theorem for local linear statistics in classical compact groups and related combinatorial identities. *Annals of Probability*, 28(3):1353–1370, 2000.

[155] Alexander Soshnikov. Gaussian fluctuation for the number of particles in Airy, Bessel, sine, and other determinantal random point fields. *Journal of Statistical Physics*, 100:491–522, 2000.

[156] Alexander Soshnikov. Poisson statistics for the largest eigenvalues of Wigner random matrices with heavy tails. *Electronic Communications in Probability*, 9:82–91, 2004.

[157] M. Srednicki. Chaos and quantum thermalization. *Phys. Rev. E*, 50(2):888, 1994.

[158] T. Tao and V. Vu. Random matrices: universality of local eigenvalue statistics. *Acta Math.*, 206(1):127–204, 2011.

[159] Terence Tao. *Topics in random matrix theory*, volume 132. American Mathematical Society, 2023.

[160] Terence Tao and Van Vu. Random matrices: Universality of local eigenvalue statistics up to the edge. *Communications in Mathematical Physics*, 298(2):549–572, 2010.

[161] Craig A Tracy and Harold Widom. On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics*, 177:727–754, 1996.

[162] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[163] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[164] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018.

[165] V. Vu and K. Wang. Random weighted projections, random quadratic forms and random eigenvectors. *Random Structures Algorithms*, 47(4):792–821, 2015.

[166] C. Xu, F. Yang, H.-T. Yau, and J. Yin. Bulk universality and quantum unique ergodicity for random band matrices in high dimensions. *arXiv preprint arXiv:2207.14533*, 2022.

[167] J. Yin. The local circular law III: general case. *Probab. Theory Related Fields*, 160:679–732, 2014.

[168] Steven Zelditch. Uniform distribution of eigenfunctions on compact hyperbolic surfaces. *Duke Mathematical Journal*, 55(4):919, 1987.

[169] Vladimir Antonovich Zorich and Octavio Paniagua. *Mathematical analysis II*, volume 220. Springer, 2016.