Title
Spectrum Analysis of Correlation Matrices with Weak Factors

Author
Latimer, Van

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Spectrum Analysis of Correlation Matrices with Weak Factors

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Mathematics

by

Van Latimer

2022

ABSTRACT OF THE DISSERTATION

Spectrum Analysis of Correlation Matrices with Weak Factors

by

Van Latimer
Doctor of Philosophy in Mathematics
University of California, Los Angeles, 2022
Professor Jun Yin, Chair

In this work we study the extreme eigenvalues and eigenvectors of sample correlation matrices arising from Johnstone's spiked model, or, a factor model. We make the important assumption that the model has weak factors. Under the assumption of 6 bounded moments and the assumption that the eigenvalues are comparable and not too close to one another, we show that the distribution of the spiked eigenvalues of the sample correlation matrix are close enough to those of the sample covariance matrix to have the same distribution. We show a similar result for the eigenvectors under the additional assumption that the eigenvalues are bounded. We also show that the non-spiked eigenvalues of the sample correlation matrix are close to those of an appropriate random sample covariance matrix, which allows us to establish universal Tracy-Widom statistics if the factors are weak enough. However, we establish a phase transition, whereby the non-spiked eigenvalues may have asymptotically Gaussian distribution if the factors are not weak enough.

The dissertation of Van Latimer is approved.

Terence Chi-Shen Tao

Marek Biskup

Timothy Derek Austin

Jun Yin, Committee Chair

University of California, Los Angeles

2022

To Jesus Christ.

Whom have I in heaven but thee?

And there is nothing upon earth

that I desire beside thee.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

Chapters 4 and 6 of this work are jointly a version of [LLY22], which is in preparation for publication. Chapter 5 is a version of [LY22], which is in preparation for publication.

First, I would like to thank my advisor, Professor Jun Yin. I am grateful to have learned random matrix theory from someone who has occupied such a prominent role in the development of the field.

The questions answered in chapters 4 and 6 were posed to me, though Professor Yin, by Professors Jinchi Lv and Yingying Fan of USC. I am grateful for their collaboration, helpful conversations, and suggestions for making the work useful for a statistical audience. Thank you also to my peer Bertrand Stone for his help in writing section 4.5.

I would like to thank Dr. Ben Robinson at the Air Force Research Laboratory for allowing me to work with him this last summer and for posing to me some questions from signal processing which we were able to solve with RMT. I'm very happy that our paths have crossed and I hope for more fruitful collaboration in the future.

I would like to thank the engineers at Radial R&D for offering to take me on so far in advance of finishing my PhD. I'm excited to join you in Dayton and get to work.

I would like to thank my mathematics and physics Professors at El Camino college: Professors Greg Fry, Perry Hacking, Kristine Numrich, Susana Prieto, and Susan Taylor among others, as well as Mr. Kurashige at West High School. These were all very dedicated and passionate teachers who invested and believed in me, and, at some point while under their instruction, I began to imagine a future pursuing mathematics.

I would like to thank my friends, pastors and family at Redeemed South Bay. As God would have it I had the great opportunity to attend grad school while staying at Redeemed during its formative years. My family and I are eternally grateful for your friendship!

I would like to thank my excellent wife, Nadia, for bearing with me these 5 long years

while giving me three of the most wonderful children in the whole world. I love you more than you know, and I can't wait for the next chapter of our life.

Lastly, I want to thank my God. If I have accomplished anything in my PhD, it is due only to his provision for me. Soli deo gloria!

VITA

2012-2014	A.S. Mathematics, A.S. Physical Sciences, El Camino College
2014	Presidential Scholar of the Physical Sciences Division, El Camino College
2014	Mathematics Faculty Choice Award, El Camino College
2014-2016	Regents Scholar, UCLA
2016	B.S. Mathematics with a Minor in Germanic Languages, Magna Cum Laude, UCLA
2016-2017	Tutor and Curriculum Developer, Merieducation Pasadena
2017-2022	Graduate Student and Teaching Assistant in Mathematics, UCLA
2019	M.A. Mathematics, UCLA
2021	Intern, Autonomous Technology Research Center
2022-	Research Engineer, Radial Research and Development

PUBLICATIONS

Benjamin D. Robinson, Robert Malinas, Van Latimer, Beth Bjorkman Morrison, Alfred Hero. *An Improvement on the Hotelling T^2 Test Using the Ledoit-Wolf Nonlinear Shrinkage Estimator.* European Conference on Signal Processing. Accepted.

CHAPTER 1

Introduction

1.1 The setting and existing results

In this work we investigate correlation matrices arising from the spiked covariance model of Johnstone [Joh01]. First we will describe the motivation for studying covariance matrices in general, then we will introduce the intuition behind the spiked model. We will specialize to the case of spiked models with *weak factors*, where the terminology comes from the discipline of factor analysis where the use of the spiked model is ubiquitous. Finally we will introduce the correlation matrix—the motivation for its use and the difficulty in studying it.

1.1.1 Sample Covariance Matrices

Let M be a large positive integer. Given an M-dimensional real-valued random vector $\mathbf{z} = (z_1, \ldots, z_M)$, a fundamental—even canonical—problem in statistics is to obtain the covariances between the different entries z_i of $\mathbf{z}$, which we call *variables*.

- How do 200 different markers of students' academic performance, socio-economic position, and other quantifiable life circumstances covary with one another?
- How do the prices of 500 different stocks covary?
- How does a customer's rating of one movie on a streaming service covary with his or her rating of another?

We see that the covariances yield valuable understanding in science contexts, which in turn yields monetizable insights in a business context.

In all of these examples, the covariances cannot be known a priori but at best must be estimated from iid samples of the random vector $\mathbf{z}$. The covariance $\mathbb{E}z_i z_j$, being an expectation, may be estimated with the strong law of large numbers: considering N independent measurements $\mathbf{z}^{(\mu)}$ of $\mathbf{z}$, we estimate $\mathbb{E}z_i z_j$ as $N^{-1}\sum_{\mu=1}^{N} z_i^{(\mu)} z_j^{(\mu)}$, which is called the *sample covariance* between z_i and z_j. See that the matrix of sample covariances may be written

$$\left(\frac{1}{N}\sum_{\mu=1}^{N} z_i^{(\mu)} z_j^{(\mu)}\right)_{ij} = \sum_{\mu=1}^{N} \mathbf{z}^{(\mu)} \left(\mathbf{z}^{(\mu)}\right)^*$$

where A^* is the transpose of a matrix or vector.

Now let us further assume, according to the general linear model of statistics, that $\mathbf{z}$ is generated as $S\mathbf{x}$, where S is an $M \times \hat{M}$ deterministic matrix and where $\mathbf{x}$ is a $\hat{M}$-dimensional random vector with independent, centered, variance 1 entries with some tail conditions. This is on the one hand a very restrictive assumption, but on the other hand ubiquitous and necessary for much theoretical analysis. In this case, the *sample covariance* matrix is constructed as

$$\mathcal{V} := \frac{1}{N}\sum_{\mu=1}^{N} \mathbf{z}^{(\mu)}(\mathbf{z}^{(\mu)})^* = \frac{1}{N}\sum_{\mu=1}^{N} S\mathbf{x}^{(\mu)}(\mathbf{x}^{(\mu)})^* S^* = SXX^*S^* \tag{1.1}$$

where the matrix $X_{M\times N}$ is defined through $X_{i\mu} = N^{-1/2}\mathbf{x}_i^{(\mu)}$. This formulation is useful because now all of the randomness is in one quantity X, which has independent, centered, variance N^{-1} entries, and all of the covariance structure is in one quantity S. The true covariance $\mathbb{E}z_i z_j$ is given by $(SS^*)_{ij}$, so that SS^* is the *population covariance matrix.*

Note that only in heuristic discussions about real-world random vectors $\mathbf{z}$ will we use the symbol $\mathbf{x}$ for a random vector in this way. Usually $\mathbf{x}$ and $\mathbf{y}$ will denote generic derterministic vectors.

Letting $N \to \infty$ therefore, we have $(SXX^*S^*)_{i\mu} \to (SS^*)_{i\mu}$ and we may obtain the population covariances at some rate depending on the moment assumptions on X—the bare minimum uniform-in-L^2 bound on $\{N^{1/2}X_{i\mu}\}$ gives almost sure convergence, while a uniform-in-L^4 bound would give a rate of $N^{-1/2}$, and uniform-in-L^p bounds would begin to increase the probability with which the convergence holds. This regime, in which M is fixed and $N \to \infty$, may be referred to as the regime of classical multivariate statistics.

However, modern applications are not adequately handled by classical multivariate statistics. In the age of big data, we may be interested in random vectors whose dimension M is not far exceeded by the number of samples N available. For example, in our financial example above, we may be interested in how the prices of 500 different stocks covary, and, lest we go too far into the past when the covariance structure may very well have been different than it is today, we may only have reliable access to 365 different days on which all these stocks were recorded.

When N is comparable to M, although the sample covariance matrix SXX^*S^* is somewhat close to SS^* in each of its entries, more global quantities calculated from one matrix or the other may vary drastically. The canonical example is when S is the identity (this is called the *null* case, or the uncorrelated case); all of the eigenvalues of SS^* are then 1. But the eigenvalues of SXX^*S^* exhibit a profound difference (see Figure 1.1). Letting $y = M/N$, the non-zero eigenvalues are approximately described by the *Marčenko-Pastur law*, whose density is

$$f(x) = \frac{1}{2\pi} \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{yx} \mathbf{1}_{x \in [\lambda_-, \lambda_+]}, \quad \lambda_\pm = (1 \pm \sqrt{y})^2$$

The regime in which M and N are both large and in which N is comparable to M, or even exceeded by it, is often referred to as the regime of *random matrix theory*. Random matrix theory, or RMT, is most often concerned with the eigenvalues and eigenvectors of such large random matrices.

The bevavior of the eigenvalues of SXX^*S^* for general S is well studied. One relevant form of S is when the eigenvalues of SS^* are "evenly spread out", meaning that, for example, a histogram of the eigenvalues of SS^* looks like a continuous probability density function,

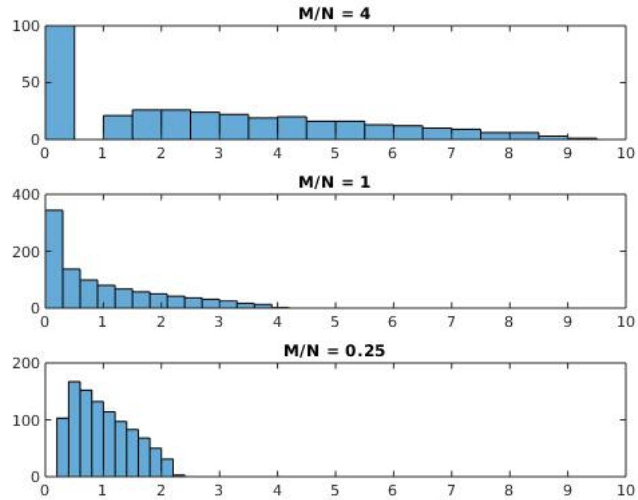

Figure 1.1: The non-zero eigenvalues of XX^* when $M = 1000$ are spread out very predictably according to the Marčenko-Pastur law. Some of the zero eigenvalues in the case $M/N = 4$ are omitted.

or when SS^* has finitely many distinct eigenvalues, and the fraction of eigenvalues taking each distinct value is bounded away from 0 (this is a common toy model; see eg [LP20]). A very strong result for the distribution of the eigenvalues is possible in this setting. Let $\mu = M^{-1}\sum_{i=1}^{M}\delta_{\lambda_i(SXX^*S^*)}$ and $\widehat{\varrho} = M^{-1}\sum_{i=1}^{M}\delta_{\lambda_i(SS^*)}$, where $\lambda_i(A)$ is the ith largest eigenvalue of a symmetric matrix A. We define a measure ϱ through its Stieltjes transform $s_\varrho(z) = \int \frac{\mathrm{d}\varrho(x)}{x-z}$ as follows:

$$\frac{1}{s_\varrho} = -z + y\int \frac{x}{1+s_\varrho x}\mathrm{d}\widehat{\varrho}(x) \tag{1.2}$$

To define the measure ϱ in this way immediately prompts new questions: does a solution to (1.2) exist, is it unique, and is the solution the Stieltjes transform of a measure? There is in fact a unique solution $s_\varrho := s_\varrho(z)$ in $\mathbb{H}$ to this z-dependent equation for $z \in \mathbb{H}$, where $\mathbb{H}$ is the complex upper half plane; the verification of this is found in [SB95] (only note that the measure $\mathcal{A}$ in that paper is δ_0 in our context). That s_ϱ is in fact the Stieltjes transform of a measure is the result of [CS95]. For further discussions of the properties of ϱ, including how to determine its support, and the "square root behavior" of its density in most cases near the edges of its support, see [CS95] and [KY17].

The introduction of the Stieltjes transform, given the technical nature of the proofs of the basic questions above, may seem an unnecessary distraction. This could not be farther from the truth. We will discuss more in Section 1.2 how the Stieltjes transform is a key ingredient in the powerful "resolvent method" of random matrix theory. Using the resolvent method, [KY17] proves that for any interval I,

$$\mu(I) = \varrho(I) + O_{\prec}(N^{-1}) \tag{1.3}$$

where the terminology $O_{\prec}$ will be explained later; it communicates a bound up to factors of N^{ϵ} with very high probability.

1.1.2 The Spiked Model of Johnstone

Another important form of S was introduced by Johnstone in the seminal paper [Joh01]. Rather than assuming the eigenvalues of SS^* are "spread out", we assume that almost all of them are 1, except for a small handful which are away from (usually greater than) 1. That is, SS^* is a bounded rank perturbation of the identity.

Let us introduce a more specialized instance of the spiked model with natural intuitive motivation and which will be the subject of our analysis. Let $\mathbf{s}_1, \ldots, \mathbf{s}_K$ be deterministic vectors in $\mathbb{R}^M$, and assume a random vector $\mathbf{z}$ is of the form

$$\mathbf{z} = \mathbf{s}_1 x_1 + \cdots + \mathbf{s}_K x_K + \mathbf{x}^{(\text{noise})} \tag{1.4}$$

where $\mathbf{x}^{(\text{noise})} \in \mathbb{R}^M$, where $x_1, \ldots, x_K \in \mathbb{R}$, and where $\mathbf{x} = \left(x_1, \ldots, x_K, \left(\mathbf{x}^{(\text{noise})}\right)^*\right)^*$ has independent, centered, variance 1 entries. Thus we are assuming that $\mathbf{z}$ is generated by a K-dimensional "signal" plus "white noise". For example, perhaps $K = 5$, the variables $z_1, \ldots, z_M$ are M different markers of a student's academic performance, and $x_1, \ldots, x_5$ are the values of the student's "Big 5 personality traits": openness, conscientiousness, extroversion, agreeableness, and neuroticism. To say $\mathbf{z}$ is of the form (1.4) is to say that every measure of academic performance has a linear contribution from a student's level of openness, conscientiousness, etc., as well as some unaccounted for randomness in the form of white noise.

The above model for $\mathbf{z}$ is called a *factor model* for the data; the terminology comes from the discipline of factor analysis, which assumes such a model for the data and tries to estimate the vectors $\mathbf{s}_\alpha$. In this context the matrix S is of the form

$$S = \begin{pmatrix} B & I_M \end{pmatrix}, \quad B = \begin{pmatrix} \mathbf{s}_1 & \cdots & \mathbf{s}_K \end{pmatrix}$$

where I_M is the M-dimensional identity matrix, so that

$$SS^* = BB^* + 1_M$$

The rank K of the signal will be considered fixed as M and N go to infinity together. Since $SS^* - I$ has rank K, this is an instance of Johnstone's spiked model. We note at this moment that if SS^* was a finite rank deformation of a diagonal matrix other than the identity, the model would no longer be referred to as Johnstone's spiked model, but rather as a *generalized spiked model*. It would still, however, be appropriate to refer to such a model as a factor model. We note that the technique of factor analysis, though widespread, has a mixed reputation. One of the criticisms leveled against it (see [FMT86]), comparing it to the related technique of principal component analysis, is that it assumes a model for the data rather than only focusing on predictive power (similar to the dichotomy between classical statistics and machine learning, see [Bre01]). This is irrelevant for our random matrix theoretic analysis; for us, the model is the really interesting thing, and we hope that our work will be useful in the very common situation where a factor model is assumed to underlie the data.

There has been a wealth of interest in the spiked model since its introduction; it has been used as a model in many statistical applications, leading to many papers devoted to its theoretical properties and vice versa. The difficulty here is different than in the case when the spectrum of SS^* is "spread out", as discussed above. The eigenvalue interlacing inequality readily yields that SXX^*S^* and XX^* have the same global spectrum, and an optimal local eigenvalue law like equation (1.3) was available in [PY14]. Letting K be the rank of $SS^* - I$, we are now interested in the K eigenvalues which may not fall in with the global spectrum; these are called *outliers*. Important questions include: how large must the eigenvalues of $SS^* - I$ be before SXX^*S^* begins to exhibit outliers? What are the distributions of outlier eigenvalues and their associated eigenvectors? What are the distributions of the leading non-outlier eigenvalues?

In the highly influential paper [BBP05], for complex Gaussian X, the distribution of the leading eigenvalue of SXX^*S^* was computed. For simplicity, assume $SS^* = I + d\mathbf{v}\mathbf{v}^*$ and set $y = \frac{M}{N}$. If d is in a compact subset of $(\sqrt{y}, \infty)$, then $\lambda_1(SXX^*S^*)$ has Gaussian fluctuations around

$$\theta(1+d) = (1+d)(1+yd^{-1}),$$

which is called the *classical location* of $\lambda_1(SXX^*S^*)$, on the order of $N^{-1/2}$, while if d is in a compact subset of $[0, \sqrt{y})$, then $\lambda_1(SXX^*S^*)$ has Tracy-Widom fluctuations around $(1+\sqrt{y})^2$ on the order of $N^{-2/3}$; this is now know as the *BBP phase transition* after the authors Baik, Ben Arous and Péché. This was further refined in [BKY16] for real or complex non-Gaussian X, where d was allowed to converge to the critical point $\sqrt{y}$: if SS^* has boundedly many eigenvalues distinct from 1, and those which differ from 1 satisfy $\left|\lambda_\alpha(SS^*) - \left(1+\sqrt{y}\right)\right| \geq N^{-1/3+\epsilon}$, then

- each $\lambda_\alpha(SXX^*S^*)$ for which $\lambda_\alpha(SS^*)$ is below the critical point $1+\sqrt{y}$ "sticks" very accurately to the extreme eigenvalues of an appropriately chosen sample covariance matrix $\hat{S}XX^*\hat{S}^*$ with $\hat{S}\hat{S}^* = 1$.

- each $\lambda_\alpha(SXX^*S^*)$ for which $1 + d = \lambda_\alpha(SS^*)$ is above $1+\sqrt{y}$ satisfies

$$|\lambda_\alpha(SXX^*S^*) - \theta(1+d)| \prec (d-\sqrt{y})^{1/2} N^{-1/2}$$

Notice that when d is at the critical value $\sqrt{y}$, then $\theta(1+d) = (1+\sqrt{y})^2$ is precisely the maximum of the Marchenko-Pastur distribution.

For the eigenvectors of the extreme eigenvalues, [BKY16] observes that when $\lambda_\alpha(SS^*)$ is beyond (or even slightly below) the critical point $1+\sqrt{y}$, the associated eigenvector is *biased* in the direction of the associated eigenvector $u_\alpha(SS^*)$ and is completely delocalized in all orthogonal directions. All statements about the eigenvectors and about the outlier eigenvalues in [BKY16] are in the form of large deviation estimates, ie, random quantities are bounded away from deterministic quantities in the $O_\prec$ sense. In [BDW20], distributions were gotten for the spiked eigenvalues and eigenvectors.

We note briefly that a large deviation result, that is, a result involving $O_\prec$ bounds between random variables and their means, is distinct in content from a distributional statement, in that neither implies the other. So also are the difficulties in deriving such statements often distinct.

1.1.3 Weak Factors

We will assume the model $S = \begin{pmatrix} B & I \end{pmatrix}$ for the entirety of our work. In addition, we will make the assumption that the factor model SXX^*S^* has *weak factors*—that is, we will assume that the columns $\mathbf{s}_\alpha$ of B satisfy

$$\|\mathbf{s}_\alpha\|_\infty \leq N^{-\epsilon_D}$$

for some $\epsilon_D > 0$. This is to say that in the vector $\mathbf{z}$ in equation (1.4), the signal

$$\mathbf{s}_1 x_1 + \cdots + \mathbf{s}_K x_K$$

is an order of magnitude weaker in every element of $\mathbf{z}$ than the noise. In particular, this is in contrast to another common model in eg [MJM21] and [Pau07], in which each of the vectors $\mathbf{s}_\alpha$ is supported on $\operatorname{span}\{\mathbf{e}_\alpha\}_{\alpha\in\{1,\ldots,K\}} \subseteq \mathbb{R}^M$.

When constructing a factor model for data in real applications, to arrive at a model

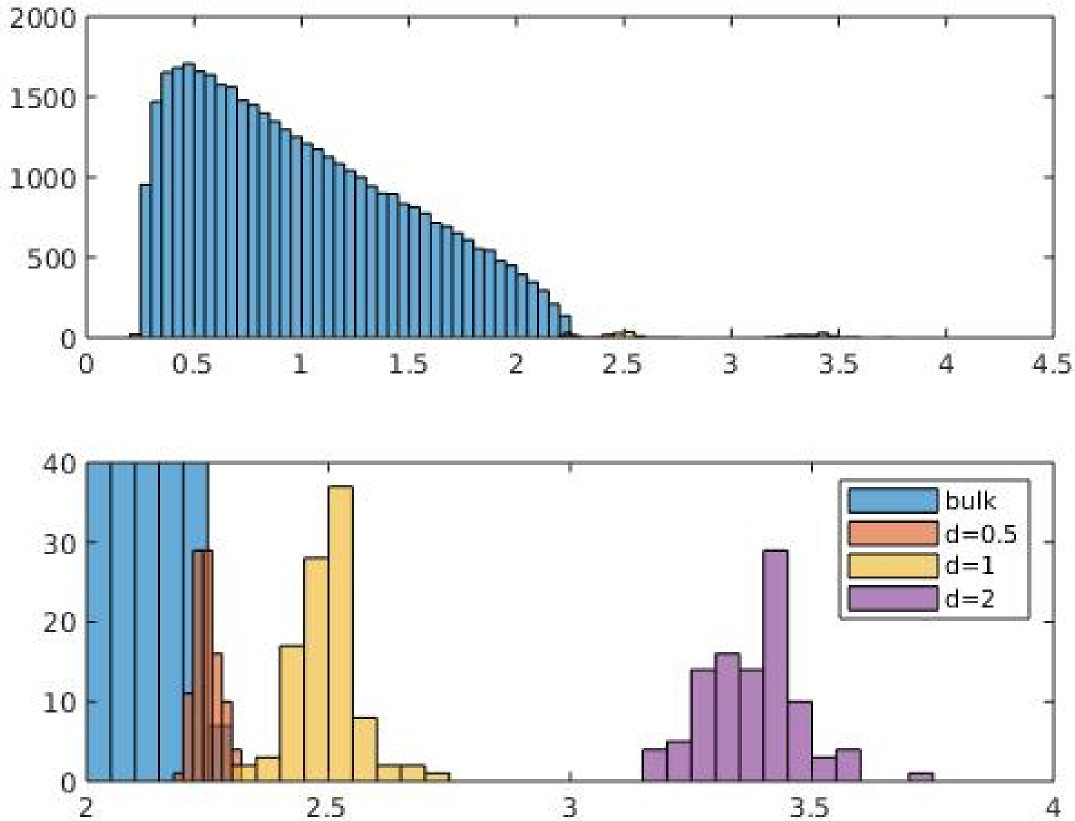

Figure 1.2: The bulk spectrum of SXX^*S^* and the leading eigenvalue when $SS^* = I + d\mathbf{v}\mathbf{v}^*$. We choose $M = N/4 = 1000$ and compute the eigenvalues 100 times. See that when $d = 0.5$, the leading eigenvalue no longer appears to be separated from the bulk spectrum. $0.5 = \sqrt{M/N}$ is the critical point at which this happens.

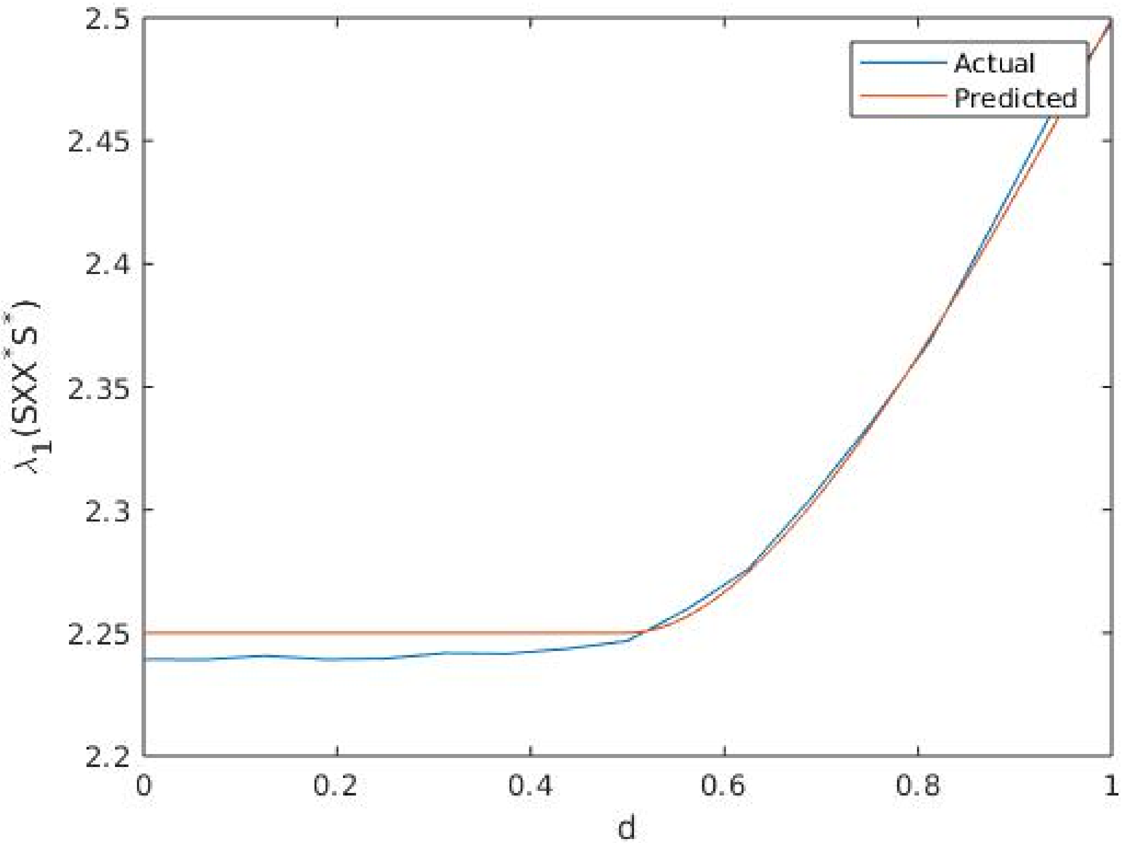

Figure 1.3: Verification of the BBP phase transition for $y = 0.25$. The discrepancy disappears in the large M limit, which is difficult to reproduce on a personal machine. The dicrepancy agrees with the negative expectation of the Tracy-Widom distribution; see [EP18] and [Bor10].

with weak factors would mean that the model would could only explain a small bit of the variance in every observed variable, and the rest would be unaccounted for randomness. This would be an unfortunate situation, but not an unrealistic one. Applications from radar to psychology allow for or even expect the "signal" in the data to be hard to detect because of the presence of noise.

1.1.4 Sample Correlation Matrices

When studying a random vector $\mathbf{z}$ as we have discussed, practitioners oftentimes prefer to replace the samples $z_i^{(\mu)}$ with the L^2 normalized samples $\left(\sum_{\nu=1}^{N}\left|z_i^{(\nu)}\right|^2\right)^{-1/2} z_i^{(\mu)}$. On the level of the sample covariance matrix, this means we study

$$\mathbf{N}(SX)\mathbf{N}(SX)^*$$

rather than SXX^*S^*, where $\mathbf{N}$ is the operator which normalizes the rows of a matrix in L^2. $\mathbf{N}(SX)\mathbf{N}(SX)^*$ is now called the *sample correlation matrix*. There are several reasons for performing this normalization:

- The leading principal component $u_1(SXX^*S^*)$ is supposed to give some explanatory power by providing a linear combination of the entries z_i of $\mathbf{z}$ which has greatest variance. Heuristically, this means it picks out the entries of $\mathbf{z}$ which are "simultaneously large". Entries which are simultaneously large covary strongly with one another, so that the leading principal component reveals a non-trivial covariance or correlation structure. If however one entry z_i of $\mathbf{z}$ tends to be much larger than the others, then $u_1(SXX^*S^*)$ will point very strongly in the direction of z_i, and thus have less potential to describe covariances between many different entries of $\mathbf{z}$. Put another way, that z_1 tends to be large may not be of great interest, but that z_1 and z_2 tend be large simultaneously would be a useful discovery.

- The sizes of the variables z_i may not have any real meaning. On one hand, the variables z_i may each be measured in different units: perhaps in a social science application where our factor model for $\mathbf{z}$ might be used, one may be the income of a person in dollars and another the number of years of the person's education and yet another a standardized test score. On the other hand, even if the variables have the same units, they may be measured with uncalibrated instruments—see e.g. [HRP14].

For these reasons, the preference for the correlation matrix over the covariance matrix when studying factor models is widespread. We remark that whether or not this is advisable is context specific and a matter of some debate (see [DS74]); the fact that it is common and that there is a gap in the mathematical implications of this choice is enough to motivate our investigation.

In contrast to sample covariance matrices, much less is known about sample correlation matrices, which are the main subject of this dissertation. In [El 09], for fairly general S with normalized rows and for X with weak moment assumptions, it was shown that the global spectrum and largest eigenvalue have the same (first-order) limits as the sample covariance matrix. In [PY12], the largest eigenvalue of the correlation matrix was shown to have Tracy-Widom (ie, universal) distribution in the null case. The first time that correlation matrices arising from Johnstone's spiked model were studied was [MJM21]. Under the assumption that the spiked space is spanned by boundedly many standard basis vectors (this is, in some sense, an assumption of very strong factors, as opposed to our assumption of weak factors), they compute the distribution of the spiked eigenvalues and of the projection of the spiked eigenvectors onto the population spike space. The most interesting part of their conclusions is that these distributions are different than they would be for the covariance matrix. We will discuss their work more in Chapter 2 after we have introduced our own results in more detail.

This dissertation is devoted to the study of correlation matrices arising from factor models with weak factors. In Chapters 4 and 6, we extend the results of [MJM21] to the setting of spiked matrices with "weak factors". Ours is arguably the more relevant setting for applications, and our result is qualitatively different: there is no difference between the results for the covariance matrix and the correlation matrix. Moreover, our results deal not only with the distributions of quantities related to the correlation matrix, but rather with the difference between the same quantity computed from the covariance matrix and the correlation matrix of the same data. The results of [MJM21] show that our assumption of weak factors is in some sense necessary for these strong conclusions to hold.

Chapter 5 then treats the non-spiked eigenvalues of spiked correlation matrices. The first order limit of the non-spiked eigenvalues was shown to be $(1+\sqrt{M/N})^2$ in [MJM21], as expected. Just as [MJM21] improves on the first order result of [El 09] by providing distributions for the spiked eigenvalues, we provide distributional results for the non-spiked eigenvalues. Our main result is to bound the difference between the non-spiked eigenvalues of the correlation matrix and the extreme eigenvalues of an appropriate random non-spiked covariance matrix—random meaning of the form $E_M X_M X_M^* E_M^*$, where the matrix E_M is random and independent of X_M; of course, the correlation matrix could almost be viewed as a random covariance matrix, but this would be inappropriate because the population would not be independent of the randomness. The independence between E_M and X_M justifies our speaking of this as a covariance matrix.

1.2 The Resolvent Method

One of the most successful methods that has been used to study the Johnstone spiked model and other finite rank perturbations of random matrices H is to relate the spiked eigenvalues and eigenvectors to a "generalized entry" $\mathbf{x}^* G(x)\mathbf{y}$ of the Green function $G(z) = (H - z)^{-1}$.

Although random matrix theory is considerably broader, the "resolvent method," or the "Green function method," a collection of techniques involving the Green function, has occupied a central role in RMT and has been responsible for much of the field's recent success.

The function $G(z)$ is a matrix-valued meromorphic function. See [BK18] for an insightful overview of some of its useful properties. To list a few of the most relevant,

1. If $H = \sum_{i=1}^{M} \lambda_i \mathbf{v}_i \mathbf{v}_i^*$ is an eigendecomposition, then $G(z) = \sum_{i=1}^{M} \frac{\mathbf{v}_i \mathbf{v}_i^*}{\lambda_i - z}$, ie, $G(z)$ is a matrix-valued meromorphic function with poles at the eigenvalues of H and whose residue at a pole λ_i is the associated spectral projection $\mathbf{v}_i \mathbf{v}_i^*$. This puts the tools of complex analysis at our disposal; we have in mind especially the Helffer-Sjöstrand calculus, see our proof of Lemma 4.2.10.

2. We have that $\underline{G(z)}$ is the Stieltjes transform of the emperical spectral measure $\delta_H = \frac{1}{M_0} \sum \delta_{\lambda_i(H)}$ of H, where $\underline{A}$ is the normalized trace of a square matrix A. In particular, the function $x \mapsto \Im \underline{G(x+i\eta)}$ is equal to the convolution $\delta_H * \theta_\eta$, where $\theta_\eta(x) = \frac{\eta^{-1}}{(x/\eta)^2+1}$ is an approximate δ function for small η. So, by fixing the imaginary part at η, which is then called the "spectral resolution," the imaginary part of the Green function's trace is exactly the spectrum of H "smoothed out" on the scale η. In particular,

   (a) If η is much bigger than the typical separation between the eigenvalues of H, then $\Im \underline{G(x+i\eta)}$ roughly gives the number of eigenvalues of H in an η-wide interval around x. Thus, to get precise control of G is to get precise control of the eigenvalues of H. This has been done to optimal (up to logarithmic factors) precision first in [EYY12] for Wigner matrices and is called a "local eigenvalue law." An optimal local eigenvalue law was the first step in the resolution of the famous Wigner-Dyson-Mehta universality conjecture by Erdős, Schlein, and Yau for the local statistics of random matrices; see [ESY11]. Similar local laws for polynomials in the Green function were very recently used in [CES21b] and [CES21a] to prove the "Eigenstate Thermalization Hypothesis" for Wigner matrices, which says that the eigenvectors of a Wigner matrix are approximately orthonormal with respect to any reasonable deterministic inner product.

   (b) If η is a little smaller than the typical separation of eigenvalues, then $\Im \underline{G(x+i\eta)}$ can be be used to study individual eigenvalues and eigenvectors. This is particularly feasible at the edge of the spectrum of H, and was used by [KY13a] to get universality of the joint eigenvalue-eigenvector statistics at the spectral edge. We use the same idea in Chapter 5 to study the individual extreme non-spiked eigenvalues of spiked sample correlation matrices.

3. The Green function is in some sense more stable than the individual eigenvalues: for derivatives of the Green function with respect to the entries of H we have the simple, easily iterated equation (4.40), whereas the corresponding Hadamard variation formula for the eigenvalues is much more complicated, especially for higher derivatives, and blows up if the eigenvalues are too close together. For these reasons the Green function is easier to work with than individual eigenvalues, although it is possible to get sophisticated results with Hadamard's formulae; see [TV11], which includes an alternative proof of the Wigner-Dyson-Mehta conjecture for ensembles matching the GOE in four moments.

4. The Green function satisfies a collection of identities (see equations (3.13) and (3.14)) which help to quantify the exact dependence of G on certain elements of H (this viewpoint is fundamental to the polynomialization method), and, with the insight that the resolvent G should be close to a deterministic matrix Π (this is called a "local law" for the resolvent), give self-consistent equations that Π would have to satisfy. If H is for example a Wigner matrix, then Π is isotropic, a multiple $m(z)$ of the identity. The function $m(z)$ is itself a Stieltjes transform of a compactly supported measure ϱ, the limiting emperical distribution for the eigenvalues of H. Thus the closeness of $\underline{G}$ to m gives, to the the extent that the Stieltjes transform can be inverted, closeness of δ_H to ϱ, thus demonstrating more fully the link between the Green function and the local eigenvalue laws discussed above.

   Moreover, if rather H is a sample covariance matrix $\Sigma^{1/2}XX^*\Sigma^{1/2}$, then it was observed in [KY17] that if one chooses an approprate alternative form for the Green function,

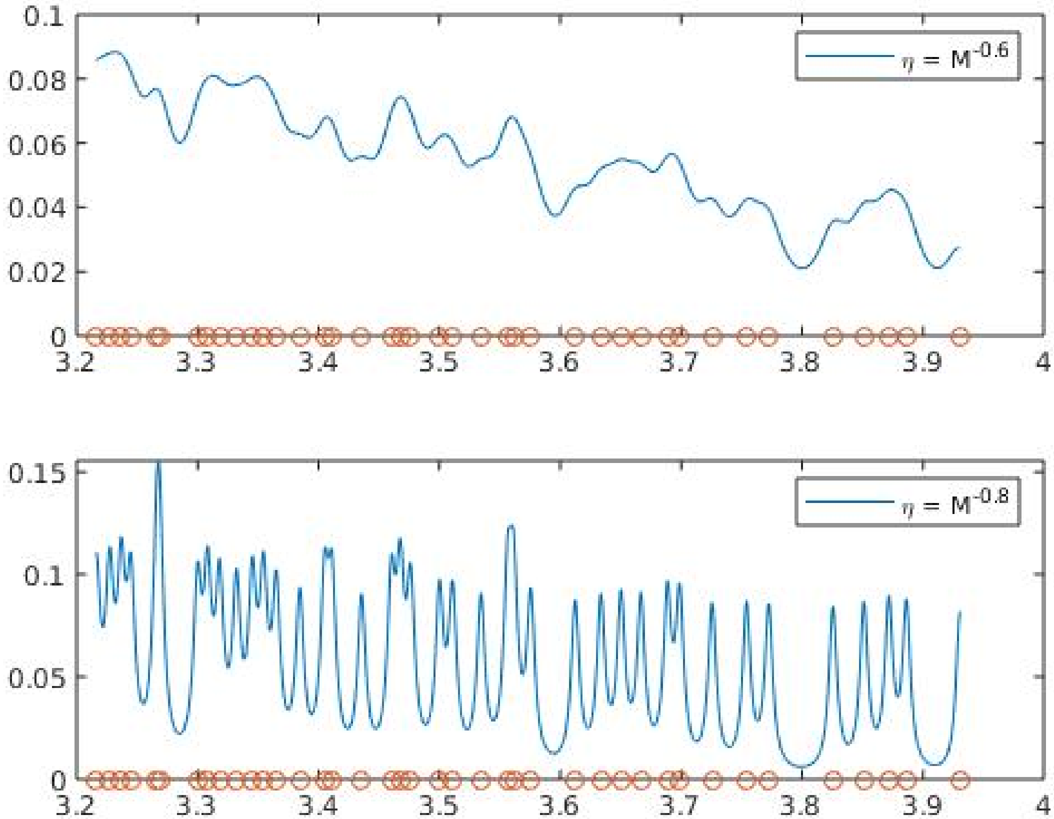

Figure 1.4: The top 40 eigenvalues of XX^* when $M = 1000, y = 1$ and $M^{-1}\Im \operatorname{Tr} G(z)$, where $G(z) = (XX^* - z)^{-1}$. This is the imaginary part of the Stieltjes transform of μ_{XX^*}, which is μ_{XX^*} smoothed out on the scale $\eta = \Im z$.

then the identities satisfied by the Green function, which are of course a function of Σ, immediately yield a self-consistent equation satisfied by the deterministic limit Π. In particular, this gives an elegant derivation of the dependence of ϱ on the matrix Σ which is an alternative to free probablistic derivations. We draw heavily on the work [KY17] and discuss this more precisely in Lemma 3.1.1.

So we are well-justified in phrasing our study of spiked correlation matrices in terms of the Green function. For the non-spiked eigenvalues, we phrase the question as:

Question 1.2.1. How does the Green function of the covariance matrix change upon passing to the correlation matrix?

To answer this, we expand the difference of Green functions in a geometric series which, we find, can be treated with the polynomialization method of [BEK14]. The polynomialization method is a prominent tool from under the umbrella of the resolvent method, and ours is perhaps a new use for it.

For the spiked eigenvalues and eigenvectors we phrase the question instead as:

Question 1.2.2. Can we get a represention of the change in the eigenvalues and eigenvectors in terms of the Green function? How can we bound it?

Once we have gotten a Green function representation for the eigenvalue or eigenvector change, we employ a common technique in RMT: we first treat the case of Gaussian randomness, and then we perform a *Green function comparison*, whereby the change in the distribution of a Green function $(H-z)^{-1}$ is studied as one changes the distribution of H. Green function comparison arguments are another prominent part of the resolvent method, and have historically been done via a Lindeberg replacement strategy, whereby the elements of H are replaced one-by-one. We opt to use a newer and more robust comparison strategy, introduced in [KY17] and based on a continuous interpolation between two different laws for H. This strategy was introduced to handle situations where a Lindeberg strategy fails.

To our knowledge ours is one of the first works since [KY17] to apply this continuous Green function comparison and further demonstrate its utility.

Much more could be said about the Green function and the powerful machinery developed for it recently, and we will say more in Chapter 3; no doubt much more will be said about it in the future. Our work is part of a long line of works using and building on the technology of the Green function to push back the borders of RMT.

CHAPTER 2

Model and Main Result

2.1 Main Model

We consider two large integers N and M which satisfy $M \asymp N$ and also a fixed (small) integer K. We let $y := \frac{M}{N}$. We let $\mathcal{I}_N, \mathcal{I}_M$, and $\mathcal{I}_K$ be index sets which have N, M, and K elements, respectively. We will use these sets to index matrices. Sometimes we will write $\mathcal{I} := \mathcal{I}_K \cup \mathcal{I}_M \cup \mathcal{I}_N$ and also $\mathcal{I}_{K+M} := \mathcal{I}_K \cup \mathcal{I}_M$. We will always use latin letters like i, j for the elements of $\mathcal{I}_M$ (and also sometimes for $\mathcal{I}_K$) and greek letters like μ, ν for the elements of $\mathcal{I}_N$. For example, a matrix A which is $\mathcal{I}_M \times \mathcal{I}_N$ has elements $A_{i\mu}$ for all $i \in \mathcal{I}_M$ and $\mu \in \mathcal{I}_N$.

Unless otherwise noted, all norms $\|A\|$ denote the Euclidean norm if A is a vector and the operator norm induced by the Euclidean norm if A is a matrix. We define the notations A^* for the conjugate transpose of a matrix $A \in \mathbb{C}^{m\times n}$ and $A^{\#} := AA^*$.

We consider the following model: let B be an $\mathcal{I}_M \times \mathcal{I}_K$ matrix. Form the singular value decomposition of B

$$B = \sum_{\alpha=1}^{K} \sqrt{d_\alpha} \mathbf{v}_\alpha \mathbf{w}_\alpha^*. \tag{2.1}$$

The eigenvalues d_α are assumed to satisfy, following [FFH20],

$$d_\alpha \asymp d_{\alpha'} \asymp |d_\alpha - d_{\alpha'}| \asymp \left| d_K - \sqrt{\frac{M}{N}} \right| \asymp N^c \tag{2.2}$$

for any $\alpha \neq \alpha'$ and for some fixed $c \geq 0$. This is to say that the eigenvalues are comparable and not too close to each other or to the limiting spectrum, and they may be allowed to diverge or remain bounded. The condition (2.2) is technical and mostly for convenience and should be removed in future works.

Note that we may also define $d_\alpha = 0$ for $\alpha \in [\![K+1, M]\!]$ and complete $\{\mathbf{v}_1, \ldots, \mathbf{v}_K\}$ to an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_M\}$ of $\mathbb{R}^{\mathcal{I}_M}$, and then we can write $B^{\#} = \sum_{\alpha=1}^{M} d_\alpha \mathbf{v}_\alpha \mathbf{v}_\alpha^*$.

The matrix B is also assumed to satisfy

$$\max_{\alpha \in \mathcal{I}_M} \|e_\alpha^* B\| \leq N^{-\epsilon_D} \tag{2.3}$$

for some fixed constant ϵ_D—in the context of factor analysis, this is means that our model is a factor model with weak factors.

We also define

$$S_{\mathcal{I}_M \times (\mathcal{I}_M \cup \mathcal{I}_K)} = \begin{pmatrix} B & I_{\mathcal{I}_M \times \mathcal{I}_M} \end{pmatrix}, \quad \widetilde{S} := \mathbf{N}(S), \tag{2.4}$$

where $\mathbf{N}$ is the operator which L^2 normalizes the rows of a matrix, i.e., for $A \in \mathbb{C}^{m \times n}$, we have

$$\mathbf{N}(A) := \mathcal{D}_A A, \quad \mathcal{D}_A := \operatorname{diag}\left(\|\mathbf{e}_1^* A\|^{-1}, \ldots, \|\mathbf{e}_M^* A\|^{-1}\right). \tag{2.5}$$

We then introduce the randomness. We define a random matrix $X_{(\mathcal{I}_M \cup \mathcal{I}_K) \times \mathcal{I}_N}$ which we will treat under two different sets of assumptions:

Assumption 1 (High Moment Conditions)**.** *The ratio y satisfies $y \asymp 1$. The elements of X are real-valued, independent random variables and satisfy for all $(i, \mu) \in (\mathcal{I}_M \cup \mathcal{I}_K) \times \mathcal{I}_N$:*

1. $\mathbb{E} X_{i\mu} = 0$.
2. $\mathbb{E}|X_{i\mu}|^2 = N^{-1}$.
3. $\mathbb{E}|X_{i\mu}|^p \leq C_p N^{-p/2}$ *for a universal constant C_p whenever p is a fixed positive integer.*
4. $|X_{i\mu}| \leq N^K$ *for some large fixed K.*

It should be clear that for any matrix satisfying conditions 1-3 of Assumption 1, every quantity at which we will look will change negligibly upon a truncation (and re-standardizing) of the matrix so that it also satisfies condition 4.

Assumption 2 (6 Moment Conditions)**.** *Let* ζ_{atom} *be a real-valued random variable satisfying:*

1. $\mathbb{E}\zeta_{\mathrm{atom}} = 0.$

2. $\mathbb{E}|\zeta_{\mathrm{atom}}|^2 = 1.$

3. $\mathbb{E}|\zeta_{\mathrm{atom}}|^p < \infty$ *for* $p = 1, \ldots, 6.$

For a real-valued random variable Y *and real number* a*, define the truncated and re-standardized random variable*

$$\mathfrak{S}_a(Y) := \frac{(1-\mathbb{E})Y\mathbf{1}_{|Y|\le a}}{\sqrt{\mathbb{E}\left((1-\mathbb{E})Y\mathbf{1}_{|Y|\le a}\right)^2}}.$$

We assume $\{X_{i\mu}\}_{i\in\mathcal{I}_{K+M},\mu\in\mathcal{I}_N}$ *is an iid family of random variables distributed as*

$$N^{-1/2}\mathfrak{S}_{N^{-\epsilon}}\left(N^{-1/2}\zeta_{\mathrm{atom}}\right)$$

for some $\epsilon < 1/6$*, and* y *satisfies* $y \asymp 1$*.*

We record also the adjusted assumption, which is more natural.

Assumption 3. X *satisfies Assumption 2, except* $X_{i\mu}$ *is distributed as* $N^{-1/2}\zeta_{\mathrm{atom}}$ *rather than* $N^{-1/2}\mathfrak{S}_{N^{-\epsilon}}\left(N^{-1/2}\zeta_{\mathrm{atom}}\right)$*.*

We will note in Lemma 4.2.1 that every ensemble satisfying Assumption 3 may without loss of generality be replaced with one satisfying Assumption 2. We call the property $X_{i\mu} \le N^{-\epsilon}$ the *bounded support condition.*

We will sometimes write

$$X_K = X_{\mathcal{I}_K\times\mathcal{I}_N} \quad \text{and} \quad X_M = X_{\mathcal{I}_M\times\mathcal{I}_N}. \tag{2.6}$$

Now we introduce the model: we consider the sample covariance matrices

$$\mathcal{V} := SXX^*S^*, \quad \widetilde{\mathcal{V}} := \widetilde{S}XX^*\widetilde{S}^* \tag{2.7}$$

As described in Chapter 1, $\mathcal{V}$ and $\widetilde{\mathcal{V}}$ are sample covariance matrices of data arising from a factor model. We may describe the matrix $\widetilde{\mathcal{V}}$ as the *sample covariance matrix with normalized variances*, since in the random vector $\mathbf{z}$ of which $\widetilde{\mathcal{V}}$ is the sample covariance matrix, all entries have variance 1.

We remark that the asymptotic spiked eigenvalue distribution of $\widetilde{\mathcal{V}}$ was recently proven in [BJ21] (see our Lemma 4.1.1). Then we define the sample correlation matrix, or, the standardized sample covariance matrix,

$$\mathcal{R} := (\mathbf{N}(SX))^{\#}. \tag{2.8}$$

A first observation, small but important, is that

$$(\mathbf{N}(SX))^{\#} = (\mathbf{N}(\widetilde{S}X))^{\#}$$

The reason for this is that $\mathbf{N}$ is invariant with respect to left-multiplication by diagonal matrices, which effect only a scaling of rows. I.e.,

$$\mathbf{N}(SX) = \mathbf{N}\left(\mathcal{D}_S SX\right) = \mathbf{N}(\widetilde{S}X)$$

if $\mathcal{D}_S$ is as in equation (2.5).

2.2 Main results

Let

$$\widetilde{S} = \widetilde{V} \begin{pmatrix} (\widetilde{D}_1 + 1)^{1/2} & 0 \\ 0 & \widetilde{D}_2^{1/2} \end{pmatrix} \begin{pmatrix} 0_{\mathcal{I}_M \times \mathcal{I}_K} & 1_{\mathcal{I}_M} \end{pmatrix} \widetilde{U}^* \tag{2.9}$$

be a singular value decomposition of $\widetilde{S}$, with

$$\widetilde{D}_1 =: \operatorname{diag}(\widetilde{d_1}, \ldots, \widetilde{d_K}).$$

$\widetilde{D}_1$ is $[\![1, K]\!] \times [\![1, K]\!]$ and $\widetilde{D}_2$ is $[\![K+1, M]\!] \times [\![K+1, M]\!]$, where $[\![1, K]\!]$ and $[\![K+1, M]\!]$ are interepreted as subsets of $\mathcal{I}_M$.

Let $H := H_N$ be the eigenvalue measure of $\widetilde{D}_2$:

$$\frac{1}{M-K}\sum_{i=1}^{M-K}\lambda_i(\widetilde{D}_2) \tag{2.10}$$

Under our assumption (2.3), H_N converges weakly to the measure δ_1 as $N \to \infty$.

Define

$$\phi(x) := \phi_N(x) = x\left(1 + \int \frac{t}{x-t}\mathrm{d}H(t)\right) \tag{2.11}$$

and, for $\alpha \in \{1, \ldots, K\}$,

$$\phi_\alpha := \phi(\widetilde{d}_\alpha + 1), \quad \phi'_\alpha := \phi'(\widetilde{d}_\alpha + 1).$$

The purpose of ϕ_α is that it is the expected location of the α^{th} spiked eigenvalue of $\widetilde{\mathcal{V}}$.

Our main result for the spiked eigenvalues is the following:

Theorem 2.2.1 (Asymptotic Spiked Eigenvalue Distribution of $\mathcal{R}$)**.** *Let* $\alpha \in \{1, \ldots, K\}$*. Letting*

$$\theta_\alpha = \frac{1}{\phi'_\alpha}, \quad \kappa_\alpha = \frac{\phi_\alpha}{\left(\widetilde{d}_\alpha + 1\right)\phi'_\alpha}$$

$$\sigma_\alpha = \sum_{i \in \mathcal{I}_K \times \mathcal{I}_M} \left(\langle \widetilde{\mathbf{u}}_\alpha, \mathbf{e}_i\rangle^4 \mathbb{E}X_{11}^4\right) - 3$$

$$\tau_\alpha = \left(\frac{\widetilde{d}_\alpha + 1}{\phi_\alpha(1 + y(\widetilde{d}_\alpha + 1)^{-1})}\right)^2$$

we have under Assumption 2 and equations (2.2) *and* (2.3) *that*

$$\sqrt{N}\left(\frac{\lambda_\alpha(\mathcal{R})}{\phi_\alpha} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sqrt{2\theta_\alpha + \sigma_\alpha\tau_\alpha}}{\kappa_\alpha}\right), \tag{2.12}$$

We also have the following corresponding large deviations result, although it does not use the full strength of our main technical achievement Theorem 4.1.4 and is somewhat easy to prove.

Theorem 2.2.2 (Large Deviation Bounds for the Spiked Eigenvalues)**.** *Fix* $\alpha \in \{1, \dots, K\}$ *and assume equations* (2.2) *and* (2.3)*. Under Assumption 1, we have*

$$|\lambda_\alpha(\mathcal{R}) - \phi_\alpha| \prec d_1 N^{-1/2}. \tag{2.13}$$

As we will discuss in Chapter 4, our main contribution towards the proof of Theorem 2.2.1, and our main technical achievement for the spiked eigenvalues, is to bound the difference between the spiked eigenvalues of $\mathcal{R}$ and $\widetilde{\mathcal{V}}$, not only to show that the distributions of the spiked eigenvalues of the two matrices are the same. See Theorem 4.1.4. We may get a similar result to Theorem 4.1.4 for the eigenvectors:

Theorem 2.2.3. *Fix* $\alpha \in \{1, \dots, K\}$ *and assume equations* (2.2) *and* (2.3)*. Additionally assume* $\left\|B^{\#}\right\| = O(1)$*. Under Assumption 1, for any fixed deterministic unit vector* $\mathbf{w} \in \mathbb{R}^{\mathcal{I}_M}$*, we have*

$$|\langle \mathbf{w}, u_\alpha(\mathcal{R})\rangle|^2 - \left|\left\langle \mathbf{w}, u_\alpha(\widetilde{\mathcal{V}})\right\rangle\right|^2 = O_\prec\left(|\langle \mathbf{w}, \widetilde{\mathbf{v}}_\alpha\rangle| N^{-1/2-\epsilon} + N^{-1}\right)$$

for some $\epsilon > 0$*. If additionally* $\|\mathbf{w}\|_\infty \le N^{-\epsilon'}$ *for some* $\epsilon' > 0$*, then the error bound may be improved to*

$$O_\prec\left(|\langle \mathbf{w}, \widetilde{\mathbf{v}}_\alpha\rangle| N^{-1/2-\epsilon} + N^{-1-\epsilon}\right)$$

Under rather Assumption 2, all the above holds with $O_\prec$ *replaced with* O_P *in the error bounds.*

We discuss the essential idea for the proof of Theorem 2.2.3 in Chapter 6, but the full proof is omitted. We hope to publish the full proof in a journal soon.

Also let the following be the definition of the $\mathcal{I}_M \times \mathcal{I}_K$ matrix $\widetilde{B}$ and the $\mathcal{I}_M \times \mathcal{I}_M$ matrix $\mathcal{J}$:

$$\widetilde{S} = \begin{pmatrix} \widetilde{B} & \mathcal{J} \end{pmatrix} \tag{2.14}$$

Recalling the definitions of X_M and X_K in equation (2.6), our main result for the non-spiked eigenvalues is the following:

Theorem 2.2.4 (Asymptotic Non-spiked Eigenvalue Distribution of $\mathcal{R}$)**.** *Assume equations* (2.2) *and* (2.3)*, and let Assumption 1 hold. Let*

$$\epsilon = \min\{\epsilon_D/4, 1/12\}.$$

For any fixed $\alpha > K$, we have

$$\mathbb{P}\left(\left|\lambda_\alpha(\mathcal{R}) - \lambda_{\alpha-K}\left((E_M X_M)^{\#}\right)\right| > N^{-2/3-\epsilon}\right) = O(N^{-\phi}) \tag{2.15}$$

for some absolute constant $\phi > 0$, where E_M is a random diagonal matrix, independent of X_M, defined by

$$(E_M)_{ii} := \left(1 + \left\|\mathbf{e}_i^* \widetilde{B} X_K\right\|^2 \mathcal{J}_{ii}^{-2}\right)^{-1/2} \tag{2.16}$$

for $i \in \mathcal{I}_M$.

For any fixed realization of the random matrix E_M, the eigenvalue $\lambda_{\alpha-K}\left((E_M X_M)^{\#}\right)$ is distributed according to the Tracy-Widom law on the scale $N^{-2/3}$, since [LS16] showed edge universality of the sample covariance matrix AXX^*A for diagonal A.[1] So, Theorem 2.2.4 is enough to conclude that $\lambda_\alpha(\mathcal{R})$ and $\lambda_{\alpha-K}\left((E_M X_M)^{\#}\right)$ have the same distribution.

Since the matrices E_M and X_M are independent, the distribution of $\lambda_{\alpha-K}\left((E_M X_M)^{\#}\right)$ has two distinct contributions: the fluctuation of $\lambda_{\alpha-K}\left((E_M X_M)^{\#}\right)$ about its mean for a fixed realization of E_M (which is roughly the same as that of $\lambda_{\alpha-K}\left(X_M^{\#}\right)$), and the fluctuation of the conditional expectation $\mathbb{E}(\lambda_{\alpha-K}(E_M X_M^{\#})|E_M)$. The former has Tracy-Widom fluctuations on the scale $N^{-2/3}$. The latter, if for example $B = B^{(\epsilon_D)} := M^{-\epsilon}(1, \ldots, 1)^*$, has Gaussian fluctuations on the scale $N^{-1/2-2\epsilon_D}$ (although this holds for general B satisfying the assumption (2.3)). Thus Theorem 2.2.4 is sufficient to conclude the distribution of the extreme non-spiked eigenvalues of $\mathcal{R}$, which are Tracy-Widom if $\epsilon_D > 1/12$ and Gaussian if $\epsilon_D < 1/12$.

---

[1] The technical Assupmtion 2.2 in [LS16] readily holds for A equal to the identity, and E_M almost surely differs from the identity in operator norm $O(N^{-1/2+\epsilon})$. Assumption 2.2 is evidently invariant to such small perturbations. Moreover, the results of [LS16] are stated for X with elements having sub-exponential decay, but the results of [KY17], for example, are enough to permit our slightly weaker Assumption 1.

2.3 Comparison with Existing Results

The main competitor to our work on the spiked eigenstructures is the work [MJM21]. We will discuss some differences between our work and this one, as well as the works [PY12] and [BJ21].

In [MJM21], correlation matrices arising from the spiked model were studied for the first time. In that paper, the distribution of the spiked eigenstructures was computed, as we do in this paper (although our full treatment of the eigenvectors is omitted). The difference between our settings is that [MJM21] assumes that the population SS^* is of the form

$$SS^* = \begin{pmatrix} \Gamma & \\ & I \end{pmatrix} \tag{2.17}$$

for a $K \times K$ matrix Γ and bounded K. This corresponds to the situation in which the low-dimensional signal in the random vector $\mathbf{z}$ which is the hallmark of the spiked model is concentrated entirely in the first K entries of $\mathbf{z}$, and the rest of the entries of $\mathbf{z}$ are random noise. This is the same assumption that was made in [BY12], which studied the distribution of the spiked eigenvalues of "generalized" spiked covariance matrices, and stood for nine years as the most general setting for these results until [BJ21]. In this setting [MJM21] shows that the spiked eigenvalues of the correlation matrix have a different distribution than those of the covariance matrix.

We include a brief heuristic note: the behavior of random matrices is oftentimes roughly isotropic. This is to say that no deterministic basis, not even the standard basis, is distinguished from any other. For example, the eigenvectors of XX^* are Haar-distributed if X is Gaussian and still completely de-localized if X is non-Gaussian. When studying the spiked model, the isotropic behavior of XX^* for Gaussian X is enough to reduce all study to that of the model (2.17)—that is, the basis of the spiked eigenvectors can simply be assumed to be the standard basis. Even if X is non-Gaussian, the model (2.17) is only defficient in the loss of some information about the third and fourth moments of the randomness of

X. However, the normalization of rows which produces the sample correlation matrix $\mathcal{R}$ in which we are interested *does* distinguish the standard basis. It is crucial in our work that the population spiked eigenvectors are roughly delocalized with respect to the standard basis as a consequence of the assumption (2.3), so that they do not point too strongly in the direction of any standard basis vector and suffer a great change as a result of the row normalization, which "happens in the standard basis".

In our work, we derive the distribution of the spiked eigenvalues of the correlation matrix under the setting that

$$S = \begin{pmatrix} B & I \end{pmatrix},$$

so that

$$SS^* = BB^* + I,$$

under the additional assumption that the norms of the rows of B satisfy $\|\mathbf{e}_i^* B\| \le N^{-\epsilon_D}$ for some $\epsilon_D > 0$. This corresponds to the situation in which the signal $\mathbf{z}$ from the remarks preceeding equation (1.1) is of the form

$$\mathbf{z} = B\mathbf{x}_{(K)} + \mathbf{x}_{(M)}$$

where $\mathbf{x}_{(K)}$ and $\mathbf{x}_{(M)}$ are independent and are each random vectors with independent, centered, variance 1 entries—that is to say that $\mathbf{z}$ is an additive combination of signal and noise, and the signal in $\mathbf{z}$ is in every entry of $\mathbf{z}$ an order of magnitude smaller than the noise. In the context of factor models, this is to say that our model has "weak factors". Thus our setting and the setting of [MJM21] are completely disjoint (ie, any random vector $\mathbf{z}$ treated by their results is not treated by ours and vice versa), but our setting is arguably more readily applicable to models of real statistical interest.

On the other hand, the moment assumptions of [MJM21] are better than ours—$4 + \epsilon$ as opposed to our 6. The 6 moment assumption in our work is not absolutely vital, and we hope to remove it in future work.

Another interesting aspect of our work in comparison to [MJM21] and [PY12] is that, whereas those works compute the distributions of extreme eigenvalues of spiked and non-spiked correlation matrices, respectively, we bound the *difference* between a spiked eigenvalue of a spiked sample covariance matrix and its associated correlation matrix; in particular we show that the difference is smaller than the scale on which the eigenvalue flucuates. So, whereas the results of [MJM21] offer caution to the practitioner trying to decide between the correlation matrix and the covariance matrix for PCA because the eigenvalues of the correlation matrix fluctuate differently than those of the covariance matrix, our results offer reassurance to the practitioner—given observed data satisfying our structural assumptions (2.2), (2.3), and 2.4 (which, we stress, is not an uncommon situation), one may generate both the covariance matrix $\widetilde{\mathcal{V}}$ (provided one had access to it) and the correlation matrix $\mathcal{R}$ and not see any difference between the spiked eigenvalues.

We will also make here a comparison between our work and the work [BJ21]. One of the first things we do in our paper is reduce the study of the correlation matrix associated to SXX^*S^* to the correlation matrix $\mathcal{R} = (\mathbf{N}(\widetilde{S}X))(\mathbf{N}(\widetilde{S}X))^*$ associated to $\widetilde{\mathcal{V}} = \widetilde{S}XX^*\widetilde{S}^*$, where $\widetilde{S} = \mathbf{N}(S)$. $\widetilde{\mathcal{V}}$ is what [BJ21] calls a "generalized" spiked model, and they derive the distribution of the spiked eigenvalues. Thus our result, which is the distribution of the spiked eigenvalues of $\mathcal{R}$, is as much a novelty compared to [BJ21] as [MJM21] is compared to [BY12], or as [PY12] is compared to previous results on the edge universality of covariance matrices, e.g. [PY14] (we remark that though [PY14] was published later, it seems to have been written before [PY12]).

Thus we situate our work in the context of previous work on correlation matrices and spiked covariance matrices. Crucial to our work is our assumption (2.3) that $\|\mathbf{e}_i^* B\| \leq N^{-\epsilon_D}$. The factors $N^{-1/2-\epsilon}$ in our main results, Lemmas 4.1.7 and 4.1.8, which are necessary to conclude our main result Theorem 4.1.4 that the spiked eigenvalues of the covariance matrix and correlation matrix are asymptotically indistinguishable, would without assumption (2.3) only be $N^{-1/2}$; the results of [MJM21] suggest the assumption (2.3) is actually necessary for such a result.

2.4 Outline of theoretical contributions

In this work, we treat sample correlation matrices corresponding to Johnstone's spiked model under the simple assumption of "weak factors", which is probably a necessary condition for the full strength of our results (given the results of [MJM21]). The setting is very general, and a very reasonable one for statistical applications in diverse disciplines.

Some of the theoretical novelties of our work include:

- For the proof of Lemma 4.1.7, we employ a novel high moment argument to bound the eigenvalue correction in the case of Gaussian X. The technique is essentially an extended investigation of the implications of the orthogonal invariance of real Gaussian random matrices and other related facts, and we hope that it may be useful for proving similar bounds for Gaussian random matrices.

- We extend the conclusion of Lemma 4.1.7 to non-Gaussian matrices with the Green function comparison strategy of [KY17]. The strategy of [KY17] is an flexible alternative to Lindeberg replacement strategies common in RMT, and to our knowledge this is one of the first works subsequent to [KY17] to leverage it.

- In Chapter 5, we use the polynomialization method for a new purpose: to treat polynomials in the resolvent G of a random matrix, ie, expressions of the form

  $$M^{-1} \operatorname{Tr} GAGA \cdots G$$

  for deterministic, or in our case, suitably chosen random A.

CHAPTER 3

Main Tools

We will now discuss some of the important results and techniques in random matrix theory which are important to our analysis. Unless otherwise stated, all notation in this chapter is specific only to this chapter.

3.1 Local Laws for the Green Function

A fundamental part of our analysis, perhaps even the foundation, is a *local law* for the Green function of general sample covariance matrices, discussed very briefly in Chapter 1, in both *isotropic* and *averaged* form.

By a local law we mean a precise estimate of the Green function $(H-z)^{-1} = G(z) = G(x+i\eta)$ for $|\eta| \gg N^{-1}$, or for $|\kappa| + |\eta| \geq N^{-2/3}$, where

$$\kappa := \kappa(z) = \operatorname{dist}(x, \operatorname{supp} \varrho) \tag{3.1}$$

and where ϱ is the deterministic limiting spectrum of H.

For a matrix A and conformable vectors $\mathbf{x}, \mathbf{y}$, we denote the "generalized entry" $A_{\mathbf{x}\mathbf{y}} := \mathbf{x}^* A \mathbf{y}$. We call this a generalized entry because if $\mathbf{x}$ and $\mathbf{y}$ are standard basis vectors, this reduces to the usual definition of a matrix entry.

For this chapter only, define matrices

$$X, \quad T \tag{3.2}$$

and a real number $\tau > 0$. The matrix $X_{M_0 \times N_0}$ is a random matrix with independent entries satisfying $\mathbb{E}X_{i\mu} = 0$, $\mathbb{E}|X_{i\mu}|^2 = N_0^{-1}$ and for any fixed $p \in \mathbb{N}$, $\mathbb{E}|X_{i\mu}|^p \leq C_p N_0^{-p/2}$ for some universal constant C_p. Most often in this paper we will assume X satisfies Assumption 1, which is consistent with this definition for X. We will also sometimes assume X satisfies Assumption 2, which is consistent with this definition except we assume the existence of fewer moments of X. After this chapter we will always state which Assumption, 1 or 2, X satisfies. The matrix $T_{\hat{M}_0 \times M_0}$ is a deterministic matrix with singular values $\sigma_1 \geq \cdots \sigma_{M_0} \geq 0$ satisfying $\sigma_1 \leq \tau^{-1}$ and $\frac{1}{M_0}|\{\alpha : \sigma_\alpha < \tau\}| \leq 1 - \tau$. Let

$$\Sigma = TT^* \tag{3.3}$$

Let $m : \mathbb{H} \to \mathbb{H}$, where $\mathbb{H}$ is the complex upper half-plane, be defined as the unique solution to

$$\frac{1}{m} = -z + y \sum_{\alpha=1}^{M_0} \frac{\sigma_\alpha}{1 + m\sigma_\alpha},$$

where $y := \frac{M_0}{N_0}$ is the dimensional ratio. We always assume that $y \asymp 1$. It is well-known that m is the Stieltjes transform of a compactly supported measure ϱ (see our remarks surrounding equation (1.2)). Usually, m in our paper will correspond to the case $T = 1$, i.e., the null case. In Chapter 4, $\widetilde{m}$ will refer to the case $\Sigma = \widetilde{S}^{\#}$ (or, equivalently for our purposes, the "unspiked" equivalent of $\widetilde{S}^{\#}$).

Define also a region

$$\mathbf{D} := \left\{ z \in \mathbb{H} : |z| \geq \tau, |x| \leq \tau^{-1}, N^{-1+\tau} \leq \eta \leq \tau^{-1} \right\}$$

Now we cite the following collection of results from [KY17] which are crucial to our analysis. Define the control parameter

$$\Psi(z) := \Psi = \sqrt{\frac{\Im m(z)}{N\eta}} + \frac{1}{N\eta}$$

Lemma 3.1.1. *Assume that the measure* $M_0^{-1} \sum_{\alpha=1}^{M_0} \delta_{\sigma_\alpha}$ *satisfies the regularity condition Definition 2.7 of [KY17]. Recall the matrices* X, T *from equation* (3.2). *The following hold for any* $\tau > 0$.

- *(Isotropic local law for the resolvents) For conformable deterministic unit vectors* $\mathbf{x}, \mathbf{y}$, *we have*

$$\left[\Sigma^{-1/2}\left((TXX^*T^* - z)^{-1} - \frac{-1}{z\,(1+mTT^*)}\right)\Sigma^{-1/2}\right]_{\mathbf{xy}} \prec \Psi \tag{3.4}$$

and

$$\left((X^*T^*TX - z)^{-1} - m\right)_{\mathbf{xy}} \prec \Psi \tag{3.5}$$

uniformly for $z \in \mathbf{D}$.

- *(Isotropic local law for the generalized resolvent) Let*

$$G(z) = \begin{pmatrix} -\Sigma^{-1} & X \\ X^* & -zI \end{pmatrix}^{-1}, \quad \Pi = \begin{pmatrix} -\frac{\Sigma}{I+m\Sigma} & \\ & mI \end{pmatrix}, \quad \overline{\Sigma} := \begin{pmatrix} \Sigma & \\ & I \end{pmatrix}$$

For conformable deterministic unit vectors $\mathbf{x}, \mathbf{y}$, *we have*

$$\left(\overline{\Sigma}^{-1}\,(G-\Pi)\,\overline{\Sigma}^{-1}\right)_{\mathbf{xy}} \prec \Psi \tag{3.6}$$

uniformly for $z \in \mathbf{D}$.

- *(Isotropic local law outside the spectrum). Uniformly for* $z \in \mathbf{D}$ *satisfying* $\operatorname{dist}(z, \operatorname{supp}\varrho) \geq N_0^{-2/3+\tau}$, *we have*

$$\left(\overline{\Sigma}^{-1}\,(G-\Pi)\,\overline{\Sigma}^{-1}\right)_{\mathbf{xy}} \prec \frac{N_0^{-1/2}}{(\kappa+\eta)^2 + (\kappa+\eta)^{1/4}} \tag{3.7}$$

The result also holds if the left-hand side of equation (3.7) *is replaced with the left-hand side of either equation* (3.4) *or* (3.5).

- *(Averaged local law) We have*

$$\underline{(XTT^*X^* - z)^{-1}} - m \prec \frac{1}{N_0\eta} \tag{3.8}$$

uniformly for $z \in \mathbf{D}$, *where* $\underline{A}$, *the normalized trace of a square matrix* A, *is defined in equation* (5.14).

On terminology: The equations (3.6) and (3.4) may be referred to either as *isotropic local laws* to denote that there is no preferred basis in which $\mathbf{x}, \mathbf{y}$ must live and to distinguish them from *entrywise local laws* which only allow $\mathbf{x}, \mathbf{y}$ to be standard basis vectors and which historically came first, or as *anisotropic local laws* because the deterministic limit of the resolvent or generalized resolvent is not a multiple of the identity, distinguishing them from, for example, a local law for a Wigner ensemble in which the resolvent is close to a multiple of the identity. The averaged local law is the key ingredient in the proof of local eigenvalue laws discussed briefly in Chapter 1 or, put another way, eigenvalue rigidity.

Lemma 3.1.1 will be especially useful to us when $T = \mathcal{J}$ or $T = E_M$ where these matrices are defined in Chapter 2. Both of the these matrices only have eigenvalues in $[1 - O(N^{-\epsilon_D}), 1 + O(N^{-\epsilon_D})]$, so the regularity condition Definition 2.7 of [KY17] is easily verified.

One of the most important uses of Lemma 3.1.1, and a main reason for the importance of isotropic local laws, is that when studying a finite rank perturbation of a random matrix H (like in the spiked model, which is the subject of this dissertation), the key quantities to control are $\mathbf{x}^*(H - z)^{-1}\mathbf{y}$, where $\mathbf{x}$ and $\mathbf{y}$ are eigenvectors of the perturbation.

3.2 Green Function Comparison Arguments

3.2.1 Lindeberg replacement strategies

A common and very general technique in RMT for proving a theorem about a random matrix H, say, $f(H) \leq a$ in very general terms, is to prove it first for some "nice" distribution of the elements of H and then to bound the change in $f(H)$ as one changes the distribution of H. Two most notable examples of a "nice" distribution for H are

- A Gaussian distribution, which can mean different things in different settings. In the context of Wigner matrices, the relevant Gaussian distribution is the Gaussian unitary ensemble (GUE) in the complex case or the Gaussian orthogonal ensemble (GOE) in the real case, which have iid Gaussian elements up to the symmetry constraint (the diagonal elements also have a difference variance than the other elements). In the context of sample covariance matrices, this could mean TXX^*T^* for a deterministic matrix T and a matrix X of iid Gaussian elements.

- A distribution with uniformly bounded moments.

By way of example, local eigenvalue statistics (that is, the distribution of the point process of the eigenvalues of a random matrix landing in a small interval with about the same width as the typical eigenvalue spacing) were first gotten for Gaussian matrices like the GUE and GOE, and the prevailing storyline of RMT for many years after was the pursuit of *universality*, or, the idea that the local eigenvalue statistics are the same for matrices of any distribution (within reason; there are obvious counter examples if the matrix elements for example do not have finite L^1 or L^2 norm).

One of the most widely used strategies for universality is a Lindeberg replacement strategy, in which the elements are swapped out one-by-one. [TV11] applied this replacement strategy to the individual eigenvalues to obtain 4-moment universality of the local eigenvalue statistics of Wigner matrices in the bulk spectrum. [KY13a] applied the same strategy to the Green function (when applied to the Green function, arguments like this are called Green function comparison arguments) to prove 2-moment universality of the joint eigenvalue and eigenvector statistics of generalized Wigner matrices at the edge of the spectrum (and their work actually provides an alternative proof to the result of [TV11]). Both these works require sub-exponential decay of the atom distributions, or at least the existence of sufficiently high moments.

Another question related to universality is how singular the atom distribution may be, or, how many moments it must have, in order for universality to hold. For example, [KY13a] showed that only 2-moment matching is needed for eigenvalue universality at the edge, but this is false if the atom distribution of either ensemble does not have finite absolute third moment. Using again a Lindeberg strategy applied to the Green functions, [LY12] showed that a necessary and sufficient condition for edge universality of Wigner matrices with iid elements distributed as $N^{-1/2}\xi$ is

$$\lim_{s\to\infty} s^4\mathbb{P}(|\xi| > s) < \infty$$

which is barely weaker than a fourth moment assumption. The difficulty in this proof is different than the difficulty in, say [KY13a]: for the result [KY13a], one can rely on the nice behavior of the Green function thanks to the high moment assumptions on the randomness and then focus on pushing the required number of matching moments as low as possible, while for the result of [LY12], the difficulty is the bad behavior of the Green function when the randomness does not have high moments, but one can match as many moments as one assumes esists.

In Chapter 4 we prove "2 moment universality" of a quantity we construct to represent the change in the eigenstructures upon passing from the covariance matrix to the correlation matrix for general X satisfying assumption 1 in Sections 4.5 and 4.6, and then we extend to distributions satisfying Assumption 2, that is, only having 6 finite moments, by matching 5 moments of the ensembles, which actually makes the argument quite easy. That is, we answer a question about universality like the one answered in [KY13a] and also a question like the one answered in [LY12].

3.2.2 Continuous Interpolation Strategy

After the initial introduction of the Lindeberg replacement strategy in [TV11], a different strategy was introduced in [KY17] which, instead of replacing matrix elements one by one, deforms every matrix element simultaneously from one law to another. This is the strategy that we use in this dissertation, but it is more a strategy than an easily citable theorem and there does not exist to the author's knowledge a detailed description of it, so let us describe it in some detail.

Suppose for instance that one has a collection of probability spaces, each with a choice of two different probability measures: (Ω_i, P_i^0) and (Ω_i, P_i^1), for $i = 1, \dots, n$. Form the product space with the product measure for each choice of P_i to get (Ω, P^0), and (Ω, P^1), where $\Omega = \Omega_1 \times \cdots \times \Omega_n$, and where P^0 is the product measure constructed from the measures P_i^0, and likewise for P^1. Thus P^0 and P^1 represent the joint probability distribution of n independent (but not necessarily identically distributed) random variables $X_1, \dots, X_n$, with X_i having law either P_i^0 or P_i^1.

We may smoothly perturb from P^0 to P^1 by taking convex combinations: define

$$P^\theta = \theta P^1 + (1 - \theta) P^0, \tag{3.9}$$

and likewise define P_i^θ by a convex combination.

A calculus exercise yields that for any sufficiently smooth function F on Ω,

$$\frac{d}{d\theta} \int F dP^\theta = \sum_{i=1}^{n} \int F dP_1^\theta \cdots (\mathrm{d}P_i^1 - \mathrm{d}P_i^0) \cdots \mathrm{d}P_n^\theta; \tag{3.10}$$

that is, the rate of change of the expectation of F as P^0 is deformed to P^1 is described by a sum over the factors in the product space: using the interpolating measure in every factor but the i^{th}, one subtracts the expectation with respect to using P_i^0 in the i^{th} coordinate from the expectation with respect to using P_i^1 in the i^{th} coordinate, repeating for every value of i and summing in i.

To each summand of the expression (3.10), Taylor's Theorem may now be applied. At this point, we refer to our Lemma 4.5.1, where we get the formula

$$\cdot \frac{\partial}{\partial\theta} \mathbb{E} F\left(X^\theta\right) = \sum_{m=1}^{\overline{m}} \sum_{i=1}^{n} \left(K_m(P_i^1, P_i^\theta) - K_m(P_i^0, P_i^\theta)\right) \int \partial_i^m F \mathrm{d}P^\theta + \mathcal{E} \tag{3.11}$$

where $K_m(Q, \widetilde{Q})$ is a function of probability measures $Q, \widetilde{Q}$ and which depends only on their first m moments, and where $\mathcal{E}$ is an error term whose computation is context-specific but generally easy if the moments of P_i^ι decay at all (in our context the mth moment decays as $N_0^{-m/2}$).

The beauty of this is that now, if one has two random matrix ensembles which have the same first m_0 moments in all of their entries, and which both have all their entries living on the scale $N^{-1/2}$ (so that moments of the entries decay roughly as powers of $N^{-1/2}$), then one can get an expression for the difference, in expectation, of any smooth function evaluated on the different ensembles. If the ensembles agree in 3 moments, then three terms in the Taylor polynomials vanish, and then the K_m expressions that are left are all bounded by N^{-2}. Notice that this N^{-2} is already enough, for example, to cancel the factor of N^2 that comes from the sum over every matrix element that arises from our formula (3.10).

This formulation is very useful, because in every term of the sum over i, it involves expressions which are all computed from the same ensemble X^θ, as opposed to in the Lindeberg strategy, where a slightly different matrix appears at every step of the interpolation. This allows for self-consistent comparison arguments, as well as for a finer use of structure in the sum over i, so that one may obtain more precise bounds than one would if one had to bound every element of the sum individually.

3.3 The polynomialization method

The polynomialization method was developed in [BEK14] for proving the isotropic local law for sample covariance matrices (with null covariance) from the entrywise law; we will give some heuristics for the argument of [BEK14] as a way of introducing the method.

When [BEK14] was written, the isotropic local law for sample covariance matrices equation (3.4) was not available, but only the entrywise law. A naive attempt to obtain the isotropic law from the entrywise law would be as follows: let X be a matrix as in equation (3.2) and define $G = \begin{pmatrix} -I & X \\ X^* & -zI \end{pmatrix}^{-1}$, and let $\mathbf{x}$ be a conformable deterministic unit vector. By polarization and linearity, it suffices to consider a "diagonal" generalized entry: the generalized entry $\mathbf{x}^*G\mathbf{x}$ may be written

$$\mathbf{x}^*G\mathbf{x} - m = \sum_{ij} x_i(G_{ij} - \delta_{ij}m)x_j, \tag{3.12}$$

where x_i is the i^{th} entry of $\mathbf{x}$. Each term $G_{ij} - \delta_{ij}m$ has the bound $O_\prec(\Psi)$ by the entrywise local law. Because $\mathbf{x}$ is L^2 normalized, not L^1 normalized, the obvious bound on (3.12) is

$$\begin{aligned}|\mathbf{x}^*G\mathbf{x} - m| &= \sum_i x_i^2(G_{ii} - m) + \sum_{i\neq j} x_iG_{ij}x_j \prec \Psi + \sum_{i\neq j}\Psi|x_ix_j| \lesssim \Psi M_0\sqrt{\sum_{i\neq j}|x_i|^2|x_j|^2} \\ &\lesssim N_0\Psi\end{aligned}$$

by Cauchy-Schwarz, which is clearly a much worse bound than Ψ, which is possible if $\mathbf{x}$ is a standard basis vector. The bound Ψ is in fact possible, and the reason the analysis above was not able to obtain it is that it did not exploit any cancellation, or any independence, between G_{ij} for different values of i and j. The polynomialization method is a way to extract this independence. The essential idea, which is reminiscent of the moment method in RMT, is to look at high moments of the sum and use the expectation to reduce the combinatorics of the sum.

The machinery that makes this work is the following two-fold application of resolvent identities. First, let $A \subseteq \{1, \dots, M_0\}$ and let $G^{(A)}$ be the resolvent $\big(X^{[A]}(X^{[A]})^* - z\big)^{-1}$, where $X^{[A]}$ is the matrix X with the rows indexed by T removed. The following identities show how to remove the dependence of a resolvent entry G_{ab} on the c^{th} row of X:

$$G_{ab} = G_{ab}^{(c)} + \frac{G_{ac}G_{cb}}{G_{cc}}, \quad \frac{1}{G_{aa}} = \frac{1}{G_{aa}^{(c)}} - \frac{G_{ac}G_{ca}}{G_{aa}G_{aa}^{(c)}G_{cc}} \tag{3.13}$$

Note that this identity expresses a resolvent entry as a leading term which is independent of the c^{th} row of X and an error term which is smaller by a factor of Ψ.

Second, we may express a resolvent entry G_{ab} in terms of its dependence on the ath and bth rows of X:

$$G_{ab} = G_{aa}G_{bb}^{(a)}(XG^{(ab)}X^*)_{ab}, \quad \frac{1}{G_{aa}} = (XG^{(a)}X^*)_{aa} \tag{3.14}$$

We can use the identities (3.13) and (3.14) to improve our bound on $\sum_{i\neq j} x_i G_{ij} x_j$. Looking for example at the second moment

$$\mathbb{E}\left|\sum_{i\neq j} x_i G_{ij} x_j\right|^2 = \mathbb{E}\sum_{i_1\neq j_1, i_2\neq j_2} G_{i_1j_1}G_{i_2j_2}x_{i_1}x_{i_2}x_{i_1}x_{i_2}$$

(ignoring the complex conjugates, which are irrelevant to the analysis). We may separate the sum according to coincidences among entries i_1, i_2, j_1, j_2, as in the moment method. For example, we consider

$$\mathbb{E}\sum_{i_1,j_1,j_2}^{*} G_{i_1j_1}G_{i_1j_2}x_{i_1}^2 x_{j_1}x_{j_2}, \tag{3.15}$$

where $\sum_{i_1,j_1,j_2}^{*}$ denotes a sum over distinct values of i_1, j_1, j_2. A problem now is that if x_{j_1} and x_{j_2} here, if summed over j_1 and j_2, each yield a factor of $N_0^{1/2}$. Therefore we must find two factors of $N_0^{-1/2}$ to compensate. Recursively apply the identity (3.13) to all resolvent entries to make them independent of every row of X which is not one of the indices of the resolvent entry. If every application of (3.13) yielded only the leading order term, we would arrive at

$$\mathbb{E}\sum_{i_1,j_1,j_2}^{*} G_{i_1j_1}^{(j_2)}G_{i_1j_2}^{(j_1)}x_{i_1}^2 x_{j_1}x_{j_2}$$

and then applying equation (3.14) would yield

$$\mathbb{E}\sum_{i_1,j_1,j_2}^{*}\mathcal{M}\left(\sum_{\mu_1\nu_1} X_{i_1\mu_1}G_{\mu_1\nu_1}^{(i_1j_1j_2)}X_{j_1\nu_1}\right)\left(\sum_{\mu_2\nu_2} X_{i_1\mu_2}G_{\mu_2\nu_2}^{(i_1j_1j_1)}X_{j_2\nu_2}\right)x_{i_1}^2 x_{j_1}x_{j_2}$$

where $\mathcal{M}$ is a product of diagonal resolvent entries. If we could ignore the error in approximating the diagonal entries by their limit $\frac{-1}{1+m}$, then the above would yield 0 because $X_{j_1\nu_1}$ and $X_{j_2\nu_2}$ are centered and independent of every other factor. If however some of the applications of the identity (3.13) yielded error terms, we might arrive at an epxression like

$$\mathbb{E}\sum_{i_1,j_1,j_2}^{*}\mathcal{M}G_{i_1j_1}^{(j_2)}G_{i_1j_2}^{(j_1)}\left(G_{j_1j_2}^{(i_1)}\right)^2 x_{i_1}^2 x_{j_1}x_{j_2}$$

where again $\mathcal{M}$ is a product of diagonal resolvent entries and their inverses. Again ignoring the error in replacing each one with its limit $\frac{-1}{1+m}$, the identity (3.14) would then yield

$$\mathbb{E} \sum_{i_1,j_1,j_2}^{*} \mathcal{M} \left(\sum_{\mu_1\nu_1} X_{i_1\mu_1} G^{(i_1j_1j_2)}_{\mu_1\nu_1} X_{j_1\nu_1} \right) \left(\sum_{\mu_2\nu_2} X_{i_1\mu_2} G^{(i_1j_1j_2)}_{\mu_2\nu_2} X_{j_2\nu_2} \right)$$
$$\cdot \left(\sum_{\mu_3\nu_3} X_{j_1\mu_3} G^{(i_1j_1j_2)}_{\mu_3\nu_3} X_{j_2\nu_3} \right) \left(\sum_{\mu_4\nu_4} X_{j_1\mu_4} G^{(i_1j_1j_2)}_{\mu_4\nu_4} X_{j_2\nu_4} \right) x_{i_1}^2 x_{j_1} x_{j_2}$$

Now, the eight X entries yield a total of N_0^{-4}, but we have eight new greek summation indices μ_* and ν_*. Because of the independence of the X and G entries, the expectation will be 0 unless the X entries "pair up"; i.e., if any index μ_*—say, μ_1—is distinct from all the other greek indices, then $\mathbb{E}X_{i_1\mu_1}$ will factor out and yield 0. Thus the greek indices must at least pairwise identify, so that the new summation indices yield an N_0^4 to cancel with the N_0^{-4} from the new factors of X, so that we so far see no overall improvement from our application of these identities.

However, two of the indices, j_1 and j_2, appear in three X entries. Looking at the greek indices which appear with j_1 and j_2 as indices of X entries, we see that actually ν_1, μ_3 and μ_4 must all identify, and also ν_2, ν_3 and ν_4 must all identify lest the expression be 0, so that the new summation indices actually only contribute a N_0^3 rather than N_0^4. This improvement of N_0^{-1} is exactly what we claimed we needed, and it came precisely from the fact that the indices j_1 and j_2 appear as indices of resolvent entries an odd number of times in the expression (3.15), which corresponds exactly to the fact that x_{j_1} and x_{j_2} sum in j_1 and j_2 to $N_0^{1/2}$.

This is a faithful representation of the heart of the polynomialization method, except for our approximating all diagonal resolvent entries by their limit. This is not actually permissible, and the diagonal entries must be carried along in the analysis and undergo applications of equations (3.13) and (3.14) along with the off-diagonal edges, but it turns out that this never affects the parity of the number of times j_1 and j_2 appears as an index in a resolvent entry, so the argument still holds.

We apply this same basic idea and expand on it in Chapter 5. It is also a significant part of the Green function comparison argument in Section 4.6.

3.4 The Helffer-Sjöstrand calculus

We will now state a complex analytic result often used in RMT and whose utility stems the fact that the central object in our story, the Green function, is intimately related to the Stieltjes transform, which is inherently a complex-analytic construct. Our statement is taken from [BK18].

Lemma 3.4.1 (Proposition C.1 of [BK18])**.** *Let $n \in \mathbb{N}$ and $f \in \mathcal{C}^{n+1}(\mathbb{R})$. We define the* almost analytic extension of f of degree n *through*

$$\widetilde{f}_n(x + \mathrm{i}y) := \sum_{k=0}^{n} \frac{1}{k!} (\mathrm{i}y)^k f^{(k)}(x)$$

Let $\chi \in \mathcal{C}_c^\infty(\mathbb{C}; [0, 1])$ be a smooth cutoff function. Then for any $\lambda \in \mathbb{R}$ satisfying $\chi(\lambda) = 1$, we have

$$f(\lambda) = \frac{1}{\pi} \int_{\mathbb{C}} \frac{\overline{\partial} \left(\widetilde{f}_n(z) \chi(z) \right)}{\lambda - z} \mathrm{d}^2 z$$

where $\mathrm{d}^2 z$ denotes the Lebesgure measure on $\mathbb{C}$ and $\overline{\partial} := \frac{1}{2}(\partial_x + i\partial_y)$ is the antiholomorphic derivative.

This lemma is useful for deriving a local eigenvalue law from an averaged law for the resolvent. Another utility of the lemma is for proving the right decay in z for local laws. In other words, we may prove a local law for small z, and then use the fact that the relevant resolvent quantities may be written as Stieltjes transforms of measures to show that the local law must necessarily have stronger bounds the further z is from the support of the measure; we do this in the proof of Lemma 4.2.10.

CHAPTER 4

Spiked Eigenvalues

4.1 Proof Elements for Theorem 2.2.1

The proof of Theorem 2.2.1 has two parts. The first is the derivation of the spiked eigenvalue distribution of $\widetilde{\mathcal{V}}$, which is an instance of the generalized spiked model, recently done in [BJ21]. First, define the functions m and $\widetilde{m}$ through

$$\frac{1}{m} = -z + y\frac{1}{1+m} \quad \text{and} \quad \frac{1}{\widetilde{m}} = -z + y\int \frac{x}{1+\widetilde{m}x}\mathrm{d}H(t),$$

where H is defined in equation (2.10). Both m and $\widetilde{m}$ are Stieltjes transforms of probability measures; see [SB95] and [CS95]. The function m is the Stieltjes transform of the usual Marčenko-Pastur law with ratio y, while $\widetilde{m}$ is the Stieltjes transform of the limiting bulk spectrum of $\widetilde{\mathcal{V}}$.

Lemma 4.1.1. *Using all the same notation and assumptions as in Theorem 2.2.1, we have*

$$\sqrt{N}\left(\frac{\lambda_\alpha(\widetilde{\mathcal{V}})}{\phi_{N,\alpha}} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sqrt{2\theta_\alpha + \sigma_\alpha\tau_\alpha}}{\kappa_\alpha}\right). \tag{4.1}$$

Proof. Our form for $\theta_\alpha, \kappa_\alpha, \sigma_\alpha$ and τ_α differ slightly from that of [BJ21]. Applying their

Theorem 3.1, Corollary 3.1, and Remark 3.1 directly would yield

$$\begin{aligned}
\theta_\alpha &= (1+\widetilde{d}_\alpha)^2 \widetilde{m}'(\phi_\alpha) \\
\kappa_\alpha &= 1 + \phi_\alpha(\widetilde{d}_\alpha + 1)\widetilde{m}'(\phi_\alpha) + (\widetilde{d}_\alpha + 1)\widetilde{m}(\phi_\alpha) \\
\sigma_\alpha &= \sum_{i \in \mathcal{I}_K \times \mathcal{I}_M} \left(\langle \widetilde{\mathbf{u}}_\alpha, \mathbf{e}_i \rangle^4 \mathbb{E} X_{11}^4 \right) - 3 \\
\tau_\alpha &= \left(\frac{\widetilde{d}_\alpha + 1}{\phi_\alpha(1 + y m(\phi_\alpha))} \right)^2 + O\left(N^{-\epsilon_D}\right),
\end{aligned}$$

where the expression for τ_α follows from their Remark 3.1 and that $\left\| \widetilde{D}_2 - I \right\| = O(N^{-\epsilon_D})$. To derive our expression for τ_α, again using $\left\| \widetilde{D}_2 - I \right\| = O(N^{-\epsilon_D})$, we write

$$\tau_\alpha = \left(\frac{\widetilde{d}_\alpha + 1}{\phi_\alpha(1 + y\widetilde{m}(\phi_\alpha))} \right)^2 + O\left(N^{-\epsilon_D}\right)$$

since $m(\phi_\alpha) = \widetilde{m}(\phi_\alpha) + O(N^{-\epsilon_D})$ because $\phi_\alpha - \max \operatorname{supp} \varrho \asymp 1$. Now, using the self-consistent equation

$$z = -\frac{1}{\widetilde{m}} + y \int \frac{t}{1 + t\widetilde{m}} \mathrm{d}H(t),$$

one may verify that

$$x^{-1} + 1 = \frac{1}{1 + \widetilde{m}(\phi(x+1))}, \tag{4.2}$$

which yields the result. □

The second part of the proof of Theorem 2.2.1, and the main technical achievement of this chapter, is to compare the spiked eigenvalues of $\mathcal{R}$ with those of $\widetilde{\mathcal{V}}$. First, we define the following notion of size for random variables, introduced in [EKY13], which has proven very helpful for formulating results in RMT.

Definition 4.1.2 (Stochastic Domination)**.** Given two sequences of families of random variables $X := \{X_{N,\omega}\}_{N \in \mathbb{N}, \omega \in A}$ and $Y := \{Y_{N,\omega}\}_{N \in \mathbb{N}, \omega \in A}$ for some index set A, we say that Y *stochastically dominates* X, or that $X \prec Y$, if for any (small) $\epsilon > 0$, (large) $C > 0$, and sufficiently large N, we have

$$P(|X| > N^\epsilon |Y|) < N^{-C} \tag{4.3}$$

uniformly in ω. We also say $X = O_\prec(Y)$ if $X \prec Y$.

We also define the following weaker notion of a probablistic bound:

Definition 4.1.3 (Bound in Probability)**.** Given two sequences of families of random variables $X := \{X_{N,\omega}\}_{N\in\mathbb{N},\omega\in A}$ and $Y := \{Y_{N,\omega}\}_{N\in\mathbb{N},\omega\in A}$ for some index set A, we say that Y *bounds* X *in probability*, or that $X = O_P(Y)$, if for any $\epsilon > 0$,

$$P(|X| > N^\epsilon |Y|) \to 0 \text{ as } N \to \infty \tag{4.4}$$

uniformly in ω.

Theorem 4.1.4. *Fix* $\alpha \in \{1, \ldots, K\}$ *and assume equations* (2.2) *and* (2.3)*. Under Assumption 2, we have*

$$\left|\lambda_\alpha(\mathcal{R}) - \lambda_\alpha(\widetilde{\mathcal{V}})\right| = O_P\left(d_\alpha N^{-1/2-\epsilon}\right) \tag{4.5}$$

for some $\epsilon > 0$*. Under Assumption 1, we have*

$$\left|\lambda_\alpha(\mathcal{R}) - \lambda_\alpha(\widetilde{\mathcal{V}})\right| = O_\prec\left(d_\alpha N^{-1/2-\epsilon}\right). \tag{4.6}$$

Remark 4.1.5. The bound under Assumption 1 is in fact $O_\prec(N^{-1/2-\epsilon_D})$, at least in the case of X with third moment 0; but the proof of this requires an additional high-moment calculation in Section 4.5 which we omit.

This theorem does not apply to the $(K+1)^{\text{th}}$ eigenvalue of $\mathcal{R}$, the reason being that in the proof of Lemma 4.1.6, we analyze the difference $\lambda_\alpha(\mathcal{R}) - \lambda_\alpha(\widetilde{\mathcal{V}})$ through Hadamard's variation formulae, which are hard to control if the eigenvalue $\lambda_\alpha(\widetilde{\mathcal{V}})$ is close to other eigenvalues.

Proof of Theorem 2.2.1. By Lemma 4.1.1, the normalized spiked eigenvalue $\frac{\lambda_\alpha(\widetilde{\mathcal{V}})}{\phi_\alpha} - 1$ has the asymptotically normal distribution claimed in Theorem 2.2.1; the standard deviation of $\frac{\lambda_\alpha(\widetilde{\mathcal{V}})}{\phi_\alpha} - 1$ in particular is $\asymp N^{-1/2}$. Theorem 4.1.4 shows that $\left|\frac{\lambda_\alpha(\widetilde{\mathcal{V}})}{\phi_\alpha} - \frac{\lambda_\alpha(\mathcal{R})}{\phi_\alpha}\right| = O_P(N^{-1/2-\epsilon})$, so that $\frac{\lambda_\alpha(\mathcal{R})}{\phi_\alpha} - 1$ and $\frac{\lambda_\alpha(\widetilde{\mathcal{V}})}{\phi_\alpha} - 1$ have the same asymptotic distribution. □

The proof of Theorem 4.1.4 requires three main sublemmas. First, we require more notation. We define $D \in \mathbb{R}^{\mathcal{I}_M \times \mathcal{I}_M}$ through

$$D_{ij} = \delta_{ij} \left\| \mathbf{e}_i^* \widetilde{S} X \right\|^2. \tag{4.7}$$

D should not be confused with the matrices $\widetilde{D}_1, \widetilde{D}_2$ from the singular value decomposition of $\widetilde{S}$.

Define a $\mathcal{I}_{K+M} \times \mathcal{I}_{K+M}$ orthgonal matrix U through

$$S = (S^{\#})^{1/2} \begin{pmatrix} 0 & I_{\mathcal{I}_M} \end{pmatrix} U^*.$$

Define the *generalized resolvent* and its deterministic limit

$$G := G(z) := \begin{pmatrix} -I & X \\ X^* & -zI \end{pmatrix}^{-1}, \quad \Pi := \Pi(z) := \begin{pmatrix} -(I + m(z)I)^{-1} & \\ & m(z)I \end{pmatrix}.$$

We will for the remainder of the chapter usually consider $\alpha \in \{1, \ldots, K\}$ to be fixed. We define the low-rank matrix $\mathbf{M}$ through $\mathbf{M} := \mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3$ and

$$\begin{aligned} \mathbf{M}_1 &:= \frac{d_\alpha^2 - y}{d_\alpha(d_\alpha + y)} \mathbf{v}_\alpha \mathbf{v}_\alpha^* \\ \mathbf{M}_2 &:= \theta(d_\alpha)^{-1} 2 \frac{f(d_\alpha)}{\sqrt{1 + d_\alpha}} U^*(G - \Pi) U \mathbf{v}_\alpha \mathbf{v}_\alpha^* \\ \mathbf{M}_3 &:= \theta(d_\alpha)^{-2} g(d_\alpha) U^*(G - \Pi) U \mathbf{v}_\alpha \mathbf{v}_\alpha^* U^*(G - \Pi) U, \end{aligned} \tag{4.8}$$

where

$$f(d) := \frac{1}{d}(d+1)(d^2 - y) = O(d^2), \quad g(d) := f(d)(d+y) = O(d^3) \tag{4.9}$$

and where the argument of each matrix $G - \Pi$ is ϕ_α (defined in equation (2.11)). So, $\mathbf{M}$ implicitly depends on α.

For any index set $\mathcal{K}$ and a square $(\mathcal{I}_M \cup \mathcal{K}) \times (\mathcal{I}_M \cup \mathcal{K})$ matrix A, we define

$$\operatorname{Tr}_{(\mathcal{I}_M)} A := \sum_{i \in \mathcal{I}_M} \langle \mathbf{e}_i, A \mathbf{e}_i \rangle.$$

Here is the first of our main sublemmas:

Lemma 4.1.6. *Recall $\alpha \in \{1, \ldots, K\}$ is implicit in the definition of $\mathbf{M}$ and fixed, and assume equations* (2.2) *and* (2.3)*. Under Assumption 1, we have*

$$\lambda_\alpha(\mathcal{R}) - \lambda_\alpha(\widetilde{\mathcal{V}}) = -\lambda_\alpha(\widetilde{\mathcal{V}}) \operatorname{Tr}_{(\mathcal{I}_M)} \mathbf{M}(D - I) + O_\prec \left(d_\alpha N^{-1}\right). \tag{4.10}$$

Under Assumption 2, the error term should be replaced with $O_P(d_\alpha N^{-1/2-\epsilon})$.

This lemma is proven in Section 4.3. The first step is to write $\mathcal{R}$ and $\widetilde{\mathcal{V}}$ as the terminal and inital matrices of a perturbation and use Hadamard's variation formula $\lambda_\alpha(\mathcal{R}) \approx \lambda_\alpha(\widetilde{\mathcal{V}}) + u_\alpha(\widetilde{\mathcal{V}})^* \left(\mathcal{R} - \widetilde{\mathcal{V}}\right) u_\alpha(\widetilde{\mathcal{V}})$ to study the change in the leading eigenvalues. Then using a formula from [BDW20] for the components of $u_\alpha(\widetilde{\mathcal{V}})$, we get a representation of the eigenvalue correction $\lambda_\alpha(\mathcal{R}) - \lambda_\alpha(\widetilde{\mathcal{V}})$ in terms of the generalized resolvent G. This "Green function representation" puts all the useful properties of and powerful results (see Section 1.2) for G, most recently from [KY17], at our disposal.

Lemma 4.1.7. *Assume equations* (2.2) *and* (2.3)*. If X satisfies Assumption 1 and is in addition Gaussian, then for any fixed positive even integer p and large enough N, there is a constant C_p depending only on p such that*

$$\mathbb{E}\left|\operatorname{Tr}_{(\mathcal{I}_M)} \mathbf{M}(D - I)\right|^p \leq C_p \left(N^{-1/2-\epsilon_D}\right)^p. \tag{4.11}$$

This is our desired main result, but only for Gaussian X. We prove it in Section 4.4 by exploiting a degree of independence between the elements of D, which arises from equation (2.3) and the Gaussianity of X, and also that $D - I$ "cares about" only the lengths of the rows of $\widetilde{S}X$, which are very close to 1, whereas $\mathbf{M}$ "cares about" the rows of $\widetilde{S}X$ more holistically. This suggests a degree of independence between $\mathbf{M}$ and $(D - I)$, which would in turn yield an improvement in the naive bound $\operatorname{Tr}_{(\mathcal{I}_M)} \mathbf{M}(D - I) \prec \|D - I\| \prec N^{-1/2}$, but care must be taken to exploit this.

Lastly we show that the smallness of $\operatorname{Tr}_{(\mathcal{I}_M)} \mathbf{M}(D - I)$ is universal in the distribution of X.

Lemma 4.1.8. *Assume equations* (2.2) *and* (2.3)*. Let X^0 satisfy Assumption 1.*

1. *If X^1 satisfies Assumption 1 (so that X^0 and X^1 agree in their first two moments), then there exists $\epsilon > 0$ such that for any fixed even integer p,*

$$\left|\mathbb{E}^{X^1}\left|\mathrm{Tr}_{(\mathcal{I}_M)}\mathbf{M}(D-I)\right|^p - \mathbb{E}^{X^0}\left|\mathrm{Tr}_{(\mathcal{I}_M)}\mathbf{M}(D-I)\right|^p\right| \le C_p N^{-p\left(\frac{1}{2}-\epsilon\right)}. \tag{4.12}$$

Here, $\mathbb{E}^{X^1}$ for instance denotes expectation with respect to the law for X^1.

2. *If X^1 satisfies Assumption 2 and in addition, for all i, μ, N, and $a \in \{1, \dots, 5\}$,*

$$\mathbb{E}\left(X^0_{i\mu}\right)^a = \mathbb{E}\left(X^1_{i\mu}\right)^a,$$

then there exists $\epsilon > 0$ such that

$$\left|\mathbb{E}^{X^1}\left|\mathrm{Tr}_{(\mathcal{I}_M)}\mathbf{M}(D-I)\right|^2 - \mathbb{E}^{X^0}\left|\mathrm{Tr}_{(\mathcal{I}_M)}\mathbf{M}(D-I)\right|^2\right| \le C_2 N^{-1-\epsilon}. \tag{4.13}$$

We prove Lemma 4.1.8 by following the universality strategy of [KY17].

Proof of Theorem 4.1.4. Let X^0 and X^1 be a Gaussian law and a general law for the matrix X satisfying Assumption 1. We have by Lemma 4.1.6 that

$$\lambda_\alpha(\mathcal{R}) - \lambda_\alpha(\widetilde{\mathcal{V}}) = -\,\lambda_\alpha(\widetilde{\mathcal{V}})\,\mathrm{Tr}_{(\mathcal{I}_M)}\mathbf{M}(D-I) + O_\prec\left(d_1 N^{-1/2-\epsilon_D}\right). \tag{4.14}$$

Thus, for any $\delta, C > 0$, we may choose $p \ge \frac{2C}{\delta} + 1$, and then by Markov's inequality, Lemma 4.1.7, and Lemma 4.1.8 we see for large enough N that, using $\lambda_\alpha(\widetilde{\mathcal{V}}) \prec d_1$,

$$\begin{aligned}
&\mathbb{P}^{X^1}\left(\left|\lambda_\alpha(\widetilde{\mathcal{V}})\,\mathrm{Tr}_{(\mathcal{I}_M)}\mathbf{M}(D-I)\right| > d_1 N^{-1/2-\epsilon+\delta}\right)\\
&\le \mathbb{P}^{X^1}\left(\left|\mathrm{Tr}_{(\mathcal{I}_M)}\mathbf{M}(D-I)\right| > N^{-1/2-\epsilon+\delta/2}\right) + \frac{1}{2}N^{-C}\\
&\le \left(N^{1/2+\epsilon-\delta/2}\right)^p \mathbb{E}^{X^1}\left|\mathrm{Tr}_{(\mathcal{I}_M)}\mathbf{M}(D-I)\right|^p + \frac{1}{2}N^{-C}\\
&\le \left(N^{1/2+\epsilon-\delta/2}\right)^p \mathbb{E}^{X^0}\left|\mathrm{Tr}_{(\mathcal{I}_M)}\mathbf{M}(D-I)\right|^p\\
&\qquad + C_p\left(N^{1/2+\epsilon-\delta/2}\right)^p\left(N^{-1/2-\epsilon}\right)^p + \frac{1}{2}N^{-C}\\
&\le 2C_p\left(N^{1/2+\epsilon-\delta/2}\right)^p\left(N^{-1/2-\epsilon}\right)^p + \frac{1}{2}N^{-C}\\
&\le N^{-C},
\end{aligned} \tag{4.15}$$

thus verifying the definition of stochastic domination $\prec$.

At this point, we have in particular proven

$$\mathbb{E}^X \left|\operatorname{Tr}_{(\mathcal{I}_M)} \mathbf{M}(D-I)\right|^2 \le C_2 N^{-1-\epsilon} \tag{4.16}$$

for general X satisfying Assumption 1.

Now, we let X^1 be a general law for X satisfying Assumption 2 and X^0 a law satisfying Assumption 1 which, by Lemma 4.8.2, may be chosen to agree with X^1 in 5 moments. Note that $\mathbb{P}^{X^1}(\lambda_\alpha(\widetilde{\mathcal{V}}) > 3d_1) = o(1)$ by Lemma 4.1.1. The proof may now be easily adjusted, beginning again by Lemma 4.1.6 using equation (4.16) in the last line,

$$\begin{aligned}
&\mathbb{P}^{X^1}\left(\left|\lambda_\alpha(\widetilde{\mathcal{V}}) \operatorname{Tr}_{(\mathcal{I}_M)} \mathbf{M}(D-I)\right| > d_1 N^{-1/2-\epsilon+\delta}\right) \\
&\le \mathbb{P}^{X^1}\left(\left|\operatorname{Tr}_{(\mathcal{I}_M)} \mathbf{M}(D-I)\right| > N^{-1/2-\epsilon+\delta/2}\right) + o(1) \\
&\le \left(N^{1/2+\epsilon-\delta/2}\right)^2 \mathbb{E}^{X^1}\left|\operatorname{Tr}_{(\mathcal{I}_M)} \mathbf{M}(D-I)\right|^2 + o(1) \\
&\le \left(N^{1/2+\epsilon-\delta/2}\right)^2 \mathbb{E}^{X^0}\left|\operatorname{Tr}_{(\mathcal{I}_M)} \mathbf{M}(D-I)\right|^2 \\
&\qquad + C_2\left(N^{1/2+\epsilon-\delta/2}\right)^2\left(N^{-1/2-\epsilon}\right)^2 + o(1) \\
&\le 2C_2\left(N^{1/2+\epsilon-\delta/2}\right)^2\left(N^{-1/2-\epsilon}\right)^2 + o(1) \\
&= o(1),
\end{aligned} \tag{4.17}$$

and we conclude the proof of Theorem 4.1.4. □

4.1.1 Discussion

The intepretation of $\mathcal{V}$, defined in equation (2.7), is that it is the sample covariance matrix constructed from N samples of a random vector which itself generated as a random linear combination of the columns of B, thought of as the "signal", plus white noise. This is known in statistics as a factor model (see [FFH20]). Because of the assumption (2.3), the signal is in every one of the M observed variables an order of magnitude smaller than the noise, so that we speak of this model as having *weak factors*[1]. Note also that $\mathcal{V}$ is an instance of Johnstone's spiked model, introduced in [Joh01].

The correlation matrix $\mathcal{R}$, which is similar to the covariance matrix $\mathcal{V}$ except in that the data are L^2-normalized before forming the sample covariance matrix, is the main object of our interest:

$$\mathcal{R} := (\mathbf{N}(SX))^{\#}. \tag{4.18}$$

The goal of this chapter is to study the spiked eigenvalues of $\mathcal{R}$, ie, its eigenvalues which are separated away from its bulk spectrum—this is a well-defined notion: very coarsely, since the rows of SX have norm $1 + o_N(1)$, we have that $\left\|\mathcal{R} - \widetilde{\mathcal{V}}\right\| = o(1)$, so we have by Weyl's inequality for each $\alpha = 1, \ldots, M$ that $\lambda_\alpha(\mathcal{R}) = \lambda_\alpha(\widetilde{\mathcal{V}}) + o(1)$. Thus, pairing our assumption (2.2) with existing results on the spiked model (see eg [BKY16]), we see that $\mathcal{R}$ has "spiked eigenvalues" as well.

In what follows, we will need to exercise much more delicacy, however, than we did in the above paragraph, since we will derive in this paper both distributional results and large deviation bounds on the spiked eigenvalues of $\mathcal{R}$, which will require us to work on and below the scale $N^{-1/2}$.

As observed in [El 09], most of the change in the spiked eigenvalues from $\mathcal{V}$ to $\mathcal{R}$ can be accounted for by a *deterministic* modification (by contrast, the application of $\mathbf{N}$ to SX is a random, and moreover nonlinear, modification of $\mathcal{V}$). To explain this, we must introduce the auxiliary matrix

$$\widetilde{S} := \mathbf{N}(S) = \begin{pmatrix} \widetilde{B} & \mathcal{J} \end{pmatrix},$$

where the above equation also constitutes the definition of the matrices $\widetilde{B}$ and $\mathcal{J}$, and the auxiliary matrix

$$\widetilde{\mathcal{V}} := (\widetilde{S}X)^{\#}$$

[1] This is a common assumption in, eg, signal processing applications. We hope that this work will find some application in the setting of [HRP14]

Note that while by definition $\mathcal{R} = (\mathbf{N}(SX))^{\#}$, we also have $\mathcal{R} = (\mathbf{N}(\widetilde{S}X))^{\#}$: to see this, observe that $\widetilde{S}$ differs from S, and consequently $\widetilde{S}X$ from SX, only in a scaling of rows, which is then washed out by an application of $\mathbf{N}$.

The matrix $\widetilde{\mathcal{V}}$ is an example of what has been called a *generalized spiked model* by [BY12]. It differs from the typical Johnstone spiked model in that its population covariance matrix $\widetilde{S}^{\#}$ differs not from the identity matrix but from some more general matrix by bounded rank (albeit in our setting the "more general matrix" $\mathcal{J}^{\#}$ is a diagonal matrix which is only $O(N^{-\epsilon_D})$ in norm away from identity).

The generalized spiked model is a very natural thing to study, given both the ubiquity of the spiked model in statistical practice and the somewhat strict assumption that any data should have a covariance structure which is a finite rank deformation of exactly the identity matrix. Despite this, until very recently, the most satisfactory treatment of this new model was in [BY12], which had to assume a strong independence condition between the spiked eigenvalues and the bulk eigenvalues in the form of a block diagonal structure of the population covariance matrix. Now however, this has been relaxed in the recent work [BJ21].

So, the first idea is to see that

$$\mathbf{N}(SX)^{\#} = \mathbf{N}(\widetilde{S}X)^{\#},$$

so that when studying correlation matrices with general populations, it suffices to consider only populations with unit variances. The second idea and the bulk of this paper is showing that, since the rows of $\widetilde{S}X$ are already $1 + O_P(N^{-1/2})$ by the central limit theorem, the normalization when we pass from $(\widetilde{S}X)^{\#}$ to $\mathbf{N}(\widetilde{S}X)^{\#}$ does not affect the spiked eigenvalues to leading order for spiked populations S. A key difficulty is that the error $O_P(N^{-1/2})$ by which the lengths of the rows of SX differ from 1 would appear only to yield a $O_P(N^{-1/2})$ change in the eigenvalues by Weyl's inequality, but we must improve this to $O_P(N^{-1/2-\epsilon})$. Since $(\widetilde{S}X)^{\#}$ is now a "generalized spiked model", we conclude with the main result of

[BJ21].

Remark 4.1.9. Lemma 4.1.1 is actually more general than this, and can treat the possibly that several eigenvalues of $\widetilde{S}^{\#}$ are equal; in that case there arise a number of sample eigenvalues equal to the multiplicity of the population eigenvalue, and their joint distribution is given. The moment assumptions on X are also much weaker, to the point of optimality.

Moreover, though Lemma 4.1.1 is only stated in [BJ21] for square $U_{\widetilde{S}}$, in the course of the paper it becomes clear that the dimensions of $U_{\widetilde{S}}$ may differ by a fixed bounded constant as they do in our setting.

Remark 4.1.10. The condition (2.2) may be too strong for some statistical applications. Provided one assumes existence of enough high moments of the randomness X, then one may weaken (2.2) significantly, allowing

$$|d_\alpha - d_\beta| \gg (1 - \delta_{\alpha\beta}) N^{-\epsilon_D}, \quad (d_K - \sqrt{M/N})^{-1} = O(1), \quad d_\alpha \asymp d_\beta \tag{4.19}$$

for all $\alpha, \beta \in \{1, \ldots, K\}$.

Remark 4.1.11. Variants of Assumption 1 are very common in random matrix theory (see [BKY16]). Assumption 2 is in the spirit of [LY12]. We will prove stronger results under the first assumption and correspondingly weaker results under the second.

Moreover, in Assumption 2, we may weaken the requirement that all entries of X have the same distribution to the requirement that the entries of X have boundedly many different distributions. One may verify that in Lemmas 4.2.1 and 4.8.2 (the only place where the iid rquirement is used), the proof can be easily adjusted.

4.2 Tools

In equation (2.3), D stands for "delocalization" because ϵ_D controls how delocalized the eigenvectors $\mathbf{v}_\alpha$ are:

$$N^{-2\epsilon_D} \geq \|e_\alpha^* B\|^2 = e_\alpha^* B B^* e_\alpha = \sum_{\alpha=1}^{K} d_\alpha |\langle \mathbf{v}_\alpha, e_\alpha \rangle|^2 \tag{4.20}$$

so that

$$|\langle \mathbf{v}_\alpha, e_i \rangle| \leq d_\alpha^{-1/2} N^{-\epsilon_D} \tag{4.21}$$

for all $i \in \mathcal{I}_M$ and $\alpha \in [\![1, K]\!]$.

A second consequence of (2.2) is that $d_1 \leq N^{1-2\epsilon}$ since for any $\mathbf{x} \in \mathbb{R}^{\mathcal{I}_M}$,

$$\|\mathbf{x}^* B\| \leq \sum_{i \in \mathcal{I}_M} \langle \mathbf{x}, \mathbf{e}_i \rangle \|\mathbf{e}_i^* B\| \leq N^{1/2} N^{-\epsilon} \|\mathbf{x}\|, \tag{4.22}$$

where in the first inequality we used the triangle inequality and in the second we used Cauchy-Schwarz.

The following lemma shows that for the purposes of this paper, Assumption 2 with the bounded support condition is no less general than if the bounded support condition is removed.

Lemma 4.2.1. *Let X satisfy assumption 3. Then for some $\epsilon \in (0, 1/6)$, there exists another matrix $\widetilde{X}$ satisfying Assumption 2 and*

$$\left\| (\widetilde{S}\widetilde{X})^\# - (\widetilde{S}X)^\# \right\| = O_P\left(d_1 N^{-1/2-\epsilon'} \right) \tag{4.23}$$

and

$$\left\| \mathbf{N}(\widetilde{S}\widetilde{X})^\# - \mathbf{N}(\widetilde{S}X)^\# \right\| = O_P\left(d_1 N^{-1/2-\epsilon'} \right) \tag{4.24}$$

for some $\epsilon' > 0$.

By Weyl's inequality, the truncation affects the eigenvalues of $\widetilde{\mathcal{V}}$ and $\mathcal{R}$ by $O_P(N^{-1/2-\epsilon'})$, which is an error term according to Theorem 4.1.4.

Definition 4.2.2 (The spectral region $\mathbf{S}^O$)**.** We define the spectral region

$$\mathbf{S}^O := \{z = E + i\eta \in \mathbb{C} : \frac{1}{2}(\max \operatorname{supp} \varrho + d_K) < E < N, |\eta| < 1\}. \tag{4.25}$$

All resolvent-like matrices, like $\mathcal{G}_1$ and G in Definition 4.2.5, in this paper will be evaluated at spectral parameters $z \in \mathbf{S}^O$.

We need to define the following high-probability event in order to control our error terms under Assumption 2.

Definition 4.2.3 (The event Ω_L)**.** We define the event

$$\Omega_L = \{\lambda_1(\widetilde{\mathcal{V}}) \notin \mathbf{S}^O\}$$

which holds with high probability (ie, $\mathbf{1}_{\Omega_L} \prec 0$) under either Assumption 1 or Assumption 2 as a consequence of the Furedi-Komlos argument.

Definition 4.2.4 (Matrix Multiplication)**.** We adopt the same matrix multiplication convention as in [KY17], that is, if

$$\mathcal{J}, \mathcal{K}_1, \mathcal{K}_2, \mathcal{L} \tag{4.26}$$

are index sets like $\mathcal{I}_N$, etc, then if A is a $\mathcal{J} \times \mathcal{K}_1$ matrix and B is a $\mathcal{K}_2 \times \mathcal{L}$ matrix, then for any $j \in \mathcal{J}$ and $l \in \mathcal{L}$, we define

$$(AB)_{jl} = \sum_{k \in \mathcal{K}_1 \cap \mathcal{K}_2} A_{jk} B_{kl}. \tag{4.27}$$

This is of course the usual matrix multiplication, except that we allow the multiplication of matrices of seemingly incompatible dimensions; if the index sets of the matrices still intersect in a meaningful way, we still get a meaningful matrix product.

Define $m : \mathbb{C}_+ \to \mathbb{C}_+$ as the unique solution to

$$\frac{1}{m} = -z + y\frac{1}{1+m},$$

so that m is the Stieltjes transform of a measure ϱ, the Marčenko-Pastur law.

Definition 4.2.5 (Matrices $\mathcal{G}_1$ and G)**.** Recall the defintion of the matrix U. We define

$$\mathcal{G}_1 := \mathcal{G}_1(z) = \left(\left(\begin{pmatrix}0 & I_M\end{pmatrix} U^* X\right)^{\#} - zI\right)^{-1}. \tag{4.28}$$

We also define

$$G := G(z) = \begin{pmatrix} -I & X \\ X^* & -z \end{pmatrix}^{-1} \tag{4.29}$$

Also define

$$\Pi := \Pi(z) = \begin{pmatrix} -\frac{1}{I+m(z)I} & 0 \\ 0 & m(z)I \end{pmatrix}. \tag{4.30}$$

Proposition 4.2.5.1 (Form of U)**.** *The matrix U from Definition 4.2.5 has the form*

$$U = \begin{pmatrix} (U_K)_{\mathcal{I}_K \times (\mathcal{I}_K \cup \mathcal{I}_M)} \\ (U_M)_{\mathcal{I}_M \times (\mathcal{I}_K \cup \mathcal{I}_M)} \end{pmatrix}^* \tag{4.31}$$

where

$$U_M = \begin{pmatrix} \sum_{\alpha=1}^K \sqrt{\frac{d_\alpha}{d_\alpha+1}} \mathbf{v}_\alpha \mathbf{w}_\alpha^* & \sum_{\alpha=1}^M \sqrt{\frac{1}{d_\alpha+1}} \mathbf{v}_\alpha \mathbf{v}_\alpha^* \end{pmatrix}^*, \quad \sum_{\alpha=1}^K \sqrt{d_\alpha} \mathbf{v}_\alpha \mathbf{w}_\alpha^* =: B. \tag{4.32}$$

The columns of U_K all belong to the set

$$\mathbb{R}^{\mathcal{I}_K} \oplus \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_K\}. \tag{4.33}$$

Proof. That U_M is of the claimed form follows from the definitions of U. The columns of U_K then must be orthogonal to the columns of U_M. The span of the columns of U_M is the same as the span of the columns of $\begin{pmatrix} \mathbf{v}_1^* \\ \vdots \\ \mathbf{v}_M^* \end{pmatrix} U_M$, which contains the vectors $\begin{pmatrix}0 & \mathbf{v}_{K+1}^*\end{pmatrix}, \dots, \begin{pmatrix}0 & \mathbf{v}_M^*\end{pmatrix}$. To be orthogonal to these $M-K$ columns is precisely to be in the space (4.33). □

The resolvent G is bears less relation to the matrix $\mathcal{V}$ than does $\mathcal{G}_1$, but it is an easier object to work with. We may transition between the two resolvents with the following lemma.

Lemma 4.2.6. *For deterministic unit vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_M}$ and $z \in \mathbf{S}^O$, we have*

$$\langle \mathbf{x}, \mathcal{G}_1 \mathbf{y} \rangle = \left\langle \mathbf{x}, z^{-1} U^* G U \mathbf{y} \right\rangle + \mathcal{E}_G, \tag{4.34}$$

where $\mathcal{E}_G = O_\prec(N^{-1}|z|^{-2})$ under Assumption 1 and $\sqrt[4]{\mathbb{E}\mathbf{1}_{\Omega_L}|\mathcal{E}_G|^4} = O(N^{-1}|z|^{-2})$ as well as $\mathcal{E}_2 = O_\prec(|z|^{-2}N^{-\epsilon})$ under Assumption 2 with $|X_{i\mu}| \le N^{-\epsilon}$ in the bounded support condition. The multiplication of $(\mathcal{I}_K \cup \mathcal{I}_M \cup \mathcal{I}_N) \times (\mathcal{I}_K \cup \mathcal{I}_M \cup \mathcal{I}_N)$ matrix by a vector $\mathbf{x} \in \mathbb{R}^{\mathcal{I}_M}$ is defined through the definition of matrix multiplication above, or equivalently, the cannonical embedding of $\mathbb{R}^{\mathcal{I}_M} \subseteq \mathbb{R}^{\mathcal{I}_K \cup \mathcal{I}_M \cup \mathcal{I}_N}$ by padding with zeros.

We state the local law for G proven in [KY17]:

Lemma 4.2.7. *We have the averaged local law:*

$$\left| N^{-1} \sum_{\mu \in \mathcal{I}_N} G_{\mu\mu} - m \right| \prec \frac{N^{-1}}{(\kappa + \eta)^2}.$$

We also have the isotropic local law: for any deterministic $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}}$ and $z \in \mathbf{S}^O$, we have

$$\langle \mathbf{x}, (G - \Pi)\mathbf{y} \rangle \prec \frac{N^{-1/2}}{(\kappa + \eta)} \tag{4.35}$$

and the stronger bound

$$\langle \mathbf{x}, (G - \Pi)\mathbf{y} \rangle \prec \frac{N^{-1/2}}{(\kappa + \eta)^2} \tag{4.36}$$

if one of $\mathbf{x}, \mathbf{y}$ is $\in \mathbb{R}^{\mathcal{I}_N} \subseteq \mathbb{R}^{\mathcal{I}}$. Lastly, we have $\|G - \Pi\| = O(z^{-1})$ as well as $\left\|G_{(\mathcal{I}_K \cup \mathcal{I}_M) \times \mathcal{I}_N}\right\| + \left\|G_{\mathcal{I}_N \times (\mathcal{I}_K \cup \mathcal{I}_M)}\right\| + \|G_{\mathcal{I}_N \times \mathcal{I}_N}\| = O(z^{-1})$.

The parts of Theorem 4.2.7 regarding $z \in \mathbf{S}^O$ are not stated as such in [KY17], but may routinely verified with the Helffer-Sjöstrand argument; see, eg, the proof of Lemma 4.2.10 in Section 4.8.

We also recall the following result from [DY18] which extends the local law of Theorem 4.2.7 to matrices with larger fluctuations but bounded supoort.

Theorem 4.2.8 (Theorem 3.11 of [DY18]). *If X is a matrix satisfying Assumption 2 with $|X_{i\mu}| \leq N^{-\epsilon}$ in the bounded support condition, then uniformly for $z \in \mathbf{S}^O$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_K \cup \mathcal{I}_M}$, we have*

$$\langle \mathbf{x}, (G - \Pi)\mathbf{y} \rangle \prec \frac{N^{-\epsilon}}{(\kappa + \eta)}. \tag{4.37}$$

Remark 4.2.9. This version of the local law outside the spectrum is not stated as such in [DY18], but follows from the local law inside the spectrum just as the usual local law for sample covariance or Wigner matrices does.

We also have the following result complementary result, which also provides some bounds on low moments of entries of $G - \Pi$; it is the main technical ingredient in the proof of Lemma 4.1.8 under the low moment assumption.

Lemma 4.2.10. *If X satisfies Assumption 2 with $\epsilon = 1/6 - \delta$ for some sufficiently small $\delta > 0$ in the bounded support condition, and if $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_M}$ are deterministic, then uniformly in $z \in \mathbf{S}^O$,*

$$\mathbb{E}\mathbf{1}_{\Omega_L} |(G - \Pi)_{\mathbf{x}\mathbf{y}}|^4 \leq C \left(\frac{N^{-1/2}}{(\kappa + \eta)} \right)^4. \tag{4.38}$$

One of the most important consequences of the 6-moment condition of assumption 2, once paired with the bounded support condition, is the following lemma, whose proof we postpone to Section 4.7.

Lemma 4.2.11. *If X satisfies Assumption 2 and the bounded support condition, we have for any even $p \geq 4$ that*

$$\mathbb{E}(D_{ii} - 1)^p \leq C_p N^{-2}. \tag{4.39}$$

We may easily differentiate G in the entries of X: we denote by $\partial_{i\mu}$ the derivative with respect to the i, μ element of X and find that

$$\partial_{i\mu} G_{st} = -G_{is} G_{\mu t} - G_{i\mu} G_{it}. \tag{4.40}$$

Or more generally, if $\partial_t (G^{-1}) = \Delta$ for a more general matrix Δ, then $\partial_t G = -G \Delta G$.

We have the following rough bound on $\mathbf{M}_{ij}$ for $i, j \in \mathcal{I}_M$.

Proposition 4.2.11.1. *For $i \in \mathcal{I}_M$ and $\alpha \in \mathcal{I}_K$, we have*

$$\mathbf{e}_i^* U \mathbf{v}_\alpha \le C N^{-\epsilon_D}. \tag{4.41}$$

Proof. By the definition of matrix multiplication and referring to Proposition 4.2.5.1, this is

$$\mathbf{e}_i^* U \begin{pmatrix} 0 \\ \mathbf{v}_\alpha \end{pmatrix} = \mathbf{e}_i^* U_M \begin{pmatrix} 0 \\ \mathbf{v}_\alpha \end{pmatrix} = \sqrt{\frac{1}{d_\alpha + 1}} \mathbf{e}_i^* \mathbf{v}_\alpha \le N^{-\epsilon_D} \tag{4.42}$$

by equation (4.21). □

4.3 Proof of Lemma 4.1.6

Proof of Lemma 4.1.6. Lemma 4.1.6 is implied by the following two lemmas:

Lemma 4.3.1. *Fix $\alpha \in \{1, \dots, K\}$. Under Assumption 1*

$$\lambda_\alpha(\mathcal{R}) - \lambda_\alpha(\widetilde{\mathcal{V}}) = -\lambda_\alpha(\widetilde{\mathcal{V}}) u_\alpha(\mathcal{V})^* (D - I) \mathbf{u}_\alpha(\mathcal{V}) + O_\prec\left(d_\alpha N^{-1}\right). \tag{4.43}$$

Under Assumption 2, the error term should be replaced with $O_P(d_\alpha N^{-1/2-\epsilon})$.

Lemma 4.3.2. *Under Assumption 1, we have*

$$u_\alpha(\mathcal{V})^* (D - I) \mathbf{u}_\alpha(\mathcal{V}) = \operatorname{Tr}_{(\mathcal{I}_M)} \mathbf{M}(D - I) + O_\prec(N^{-1}). \tag{4.44}$$

Under rather Assumption 2, the error term should be replaced with $O_P(N^{-1/2-\epsilon})$.

□

Note the presence of the matrix $\mathcal{V}$ in equation (4.44). This is one of the only places that we will really use $\mathcal{V}$. The reason is that $\mathcal{V}$ is an instance of the Johnstone spiked model rather than of the generalized spiked model, so that we can carry over the results of [BDW20] more directly of $u_\alpha(\mathcal{V})$ than we can for $u_\alpha(\widetilde{\mathcal{V}})$.

For the proof of Lemma 4.3.1, we will require the following additional lemmas:

Lemma 4.3.3. *If* $\mathbf{w} \in \mathbb{R}^{\mathcal{I}_M}$ *with* $\|\mathbf{w}\| \lesssim 1$ *is either deterministic or satisfies* $\|\mathbf{w}\|_\infty \leq N^{-\epsilon_1}$ *for some* $\epsilon_1 > 0$*, and if* X *satisfies Assumption 2, then*

$$\left|\mathbf{w}^*(D-I)^2\mathbf{w}\right| + \left|\sqrt{\mathbf{w}^*(D-I)^4\mathbf{w}}\right| = O_P(N^{-1/2-\epsilon_2}) \tag{4.45}$$

for some $\epsilon_2 > 0$*.*

Lemma 4.3.4. *If* Δ *is a matrix of conformable dimension with* $\|\Delta\| = O_P(N^{-\epsilon}d_1^{1/2})$*, then uniformly in* Δ*,*

$$\left|\lambda_\alpha((\widetilde{S}X+\Delta)^{\#}) - \lambda_\alpha((\widetilde{S}X)^{\#})\right| = O_P\left(d_1^{1/2}\|\Delta\|\right) \tag{4.46}$$

and

$$\left\|u_\alpha((\widetilde{S}X+\Delta)^{\#}) - u_\alpha((\widetilde{S}X)^{\#})\right\| = O_P\left(\frac{2}{d_1}\|\Delta\|\right). \tag{4.47}$$

Proof of Lemma 4.3.1. We will prove the conclusion under Assumption 2. The reader may verify that under Assumption 1, the proof is easier, and only requires the additional input that $\|D-I\| \prec N^{-1/2}$.

Define the matrix flows

$$A(t) = \left(\widetilde{S}X - \frac{t}{2}(D-1)\widetilde{S}X\right)^{\#} \tag{4.48}$$

so that $A(0) = \widetilde{\mathcal{V}}$, and

$$B(t) = \left(\widetilde{S}X - \frac{1-t}{2}(D-1)\widetilde{S}X + t(D^{-1/2}-I)\widetilde{S}X\right)^{\#} \tag{4.49}$$

so that $B(1) = \mathcal{R}$, and moreover, $A(1) = B(0)$. Thus, for $\alpha \in \{1, \ldots, K\}$,

$$\begin{aligned}\lambda_\alpha(\mathcal{R}) - \lambda_\alpha(\widetilde{\mathcal{V}}) &= \lambda_\alpha(B(1)) - \lambda_\alpha(A(0)) \\ &= (\lambda_\alpha(B(1)) - \lambda_\alpha(B(0))) + (\lambda_\alpha(A(1)) - \lambda_\alpha(A(0))) \\ &= \left(\partial_t\lambda_\alpha(B(0)) + \frac{1}{2}\partial_t^2\lambda_\alpha(B(t^B))\right) + \left(\partial_t\lambda_\alpha(A(0)) + \frac{1}{2}\partial_t^2\lambda_\alpha(A(t^A))\right)\end{aligned} \tag{4.50}$$

for some $t^A, t^B \in [0,1]$. Now by Hadamard's first variation formula,

$$\begin{aligned}\partial_t\lambda_\alpha(A(t))|_{t=0} &= -\frac{1}{2}u_\alpha^*\big((D-I)A(0) + A(0)(D-I)\big)u_\alpha \\ &= -\lambda_\alpha u_\alpha^*(D-I)u_\alpha,\end{aligned} \tag{4.51}$$

where the implied argument of every eigenvector and eigenvalue is $A(0) = \widetilde{\mathcal{V}}$, so that this is almost the right-hand side of equation (4.43), except that we need to replace $u_\alpha(\widetilde{\mathcal{V}})$ with $u_\alpha(\mathcal{V})$; we will do this at the end of the proof. We now bound the other terms.

We define the diagonal matrix D_{diff} by

$$(D_{\text{diff}})_{ii} := (D_{ii}^{-1/2} - 1) + \frac{1}{2}(D_{ii} - 1) = O\left((D_{ii} - 1)^2 D_{ii}^{-5/2}\right), \tag{4.52}$$

where the bound holds entry by entry. Thus, we have for some matrix E of norm $O_P(d_1)$,

$$\begin{aligned} \partial_t \lambda_\alpha(B(0)) &= u_\alpha^* \partial_t B(0) u_\alpha \\ &= u_\alpha^* \left(E D_{\text{diff}} + D_{\text{diff}} E\right) u_\alpha \\ &= O\left(\|E\| \cdot \left\|(D - I)^2 u_\alpha\right\|\right) \\ &= O_P(d_1 N^{-1/2-\epsilon}), \end{aligned} \tag{4.53}$$

where the implied argument of every eigenvector and eigenvalue is $B(0)$, by Lemma 4.3.3—to bound $\|u_\alpha(B(0))\|_\infty$, we use Lemma 4.3.5 and equation (4.21) to bound $\|u_\alpha(A(0))\|_\infty$ and Lemma 4.3.4 to bound $\|u_\alpha(B(0)) - u_\alpha(A(0))\| \geq \|u_\alpha(B(0)) - u_\alpha(A(0))\|_\infty$, and then the triangle inequality.

Now we bound the second derivatives, using Hadamard's second variation formula. It is a consequence of Lemma 4.3.4, Assumption 2.2 and $\|D - I\| = O_P(N^{-1/10})$ that

$$\min_{\alpha \neq \beta \in \{1,\ldots,K\}} \min_{t \in [0,1]} |\lambda_\alpha(A(t)) - \lambda_\beta(A(t))| \geq C d_1 \tag{4.54}$$

for some absolute positive constant $C > 0$ a.a.s., so that, recalling t^A from equation (4.50) and using $O(L)$ to denote a conformable matrix with operator norm $O(L)$,

$$\begin{aligned}
&\partial_t^2 \lambda_\alpha(A(t^A)) \\
&= u_\alpha^* \partial_t^2 A(t^A) u_\alpha + 2 \sum_{\beta \neq \alpha} \frac{\left| u_\beta^* \partial_t A(t^A) u_\alpha \right|^2}{\lambda_\beta - \lambda_\alpha} \\
&= u_\alpha^* \cdot (D - I)(\widetilde{S}X)^\# (D - I) u_\alpha \\
&\quad + 2 \sum_{\beta \neq \alpha} \frac{\left| u_\beta^* \cdot \left((D - I)A(t^A) + A(t^A)(D - I) + O(d_1^{1/2} \|D - I\|^2) \right) u_\alpha \right|^2}{\lambda_\beta - \lambda_\alpha} \\
&= O_P\left(d_1 \|(D - I)u_\alpha\|^2 \right) + O_P\left(\frac{d_1^2 \|(D - I)u_\alpha\|^2 + d_1 \|(D - I)^2\|^2}{d_1} \right) \\
&= O_P\left(d_1 \|(D - I)u_\alpha\|^2 \right) + O_P\left(\|D - I\|^4 \right) \\
&= O_P(d_1 N^{-1/2 - \epsilon}),
\end{aligned} \tag{4.55}$$

where the implied argument of every eigenvector and eigenvalue is $A(t^A)$; the last line follows from Lemma 4.3.3 (using Lemmas 4.3.5 and 4.3.4 to bound $\|u_\alpha\|_\infty$)for the first term, and from $\|D - I\|^4 = O_P(N^{-1+\epsilon})$ for any $\epsilon > 0$ by a union bound, Markov's inequality, and Lemma 4.2.11 for the second term.

Now, we may bound $\frac{1}{2}\partial_t^2 \lambda_\alpha(B(t^B))$ as we did $\frac{1}{2}\partial_t^2 \lambda_\alpha(A(t^A))$,

$$\begin{aligned}
&\partial_t^2 \lambda_\alpha(B(t^B)) \\
&= u_\alpha^* \partial_t^2 B(t^B) u_\alpha + 2 \sum_{\beta \neq \alpha} \frac{\left| u_\beta^* \partial_t B(t^B) u_\alpha \right|^2}{\lambda_\beta - \lambda_\alpha} \\
&= u_\alpha^* D_{\text{diff}} \cdot (\widetilde{S}X - \frac{1}{2}(D - I)\widetilde{S}X)^\# D_{\text{diff}} u_\alpha \\
&\quad + 2 \sum_{\beta \neq \alpha} \frac{\left| u_\beta^* \cdot \left(D_{\text{diff}} B(t^B) + B(t^B) D_{\text{diff}} + O(d_1^{1/2} \|D_{\text{diff}}\|^2) \right) u_\alpha \right|^2}{\lambda_\beta - \lambda_\alpha}
\end{aligned} \tag{4.56}$$

$$= O\left(d_1\|D_{\text{diff}}u_\alpha\|^2\right) + O\left(\frac{d_1^2\|D_{\text{diff}}u_\alpha\|^2 + d_1\|D_{\text{diff}}^2\|^2}{d_1}\right)$$
$$= O\left(d_1\|D_{\text{diff}}u_\alpha\|^2\right) + O\left(\|D_{\text{diff}}\|^4\right)$$
$$= O_P(d_1\|D-I\|^4 + d_1\|D-I\|^8)$$
$$= O_P(N^{-1-\epsilon})$$

as before.

Finally, we show how to bound

$$\left|u_\alpha(\widetilde{\mathcal{V}})^*(D-I)u_\alpha(\widetilde{\mathcal{V}}) - u_\alpha(\mathcal{V})^*(D-I)u_\alpha(\mathcal{V})\right| = O_P(N^{-1/2-\epsilon}). \tag{4.57}$$

We first bound $\left|u_\alpha(\widetilde{\mathcal{V}})^*(D-I)(u_\alpha(\widetilde{\mathcal{V}}) - u_\alpha(\mathcal{V}))\right| \leq \left\|u_\alpha(\widetilde{\mathcal{V}})^*(D-I)\right\|\left\|(u_\alpha(\widetilde{\mathcal{V}}) - u_\alpha(\mathcal{V}))\right\|$; the other term arising from the triangle inequality is bounded in the same way. The latter factor is $O_P(N^{-\epsilon_D})$ by simple pertrubation theory. The former factor is

$$\begin{aligned}\left\|u_\alpha(\widetilde{\mathcal{V}})(D-I)\right\| &= \sqrt{\sum_{i\in\mathcal{I}_M}\left\langle u_\alpha(\widetilde{\mathcal{V}}), \mathbf{e}_i\right\rangle^2 (D_{ii}-1)^2} \\ &= \sqrt{\sum_{i\in\mathcal{I}_M}\left(\left\langle u_\alpha(\widetilde{\mathcal{V}}), \mathbf{e}_i\right\rangle^2 - C_{i,\alpha}\right)(D_{ii}-1)^2 + \sum_{i\in\mathcal{I}_M} C_{i,\alpha}(D_{ii}-1)^2}.\end{aligned} \tag{4.58}$$

The second term under the radical is, using that a random variable X satisfies $X = O_P(\mathbb{E}|X|)$,

$$O_P\left(\sum_{i\in\mathcal{I}_M} C_{i,\alpha}\mathbb{E}|D_{ii}-1|^2\right) = O(N^{-1}) \tag{4.59}$$

as desired. The first term under the radical is

$$\begin{aligned}&O_P\left(\sum_{i\in\mathcal{I}_M}\mathbb{E}\left|\left(\left\langle u_\alpha(\widetilde{\mathcal{V}}), \mathbf{e}_i\right\rangle^2 - C_{i,\alpha}\right)(D_{ii}-1)^2\right|\right) \\ &= O_P\left(\sum_{i\in\mathcal{I}_M}\sqrt{\mathbb{E}\left|\left\langle u_\alpha(\widetilde{\mathcal{V}}), \mathbf{e}_i\right\rangle^2 - C_{i,\alpha}\right|^2}\sqrt{\mathbb{E}|D_{ii}-1|^4}\right) \\ &= O_P(N^{-1})\end{aligned} \tag{4.60}$$

by Lemma 4.2.11 and equation 4.64. This concludes the proof of Lemma 4.3.1.

□

Now we prove Lemma 4.3.2. We will now cite the following theorem from [BDW20], which makes our analysis possible. Our statement of the theorem is not exactly as it is in [BDW20], and we will provide an abridged proof in Section 4.8.

Theorem 4.3.5. *Fix $\alpha \in \{1, \ldots, K\}$. Let $\mathbf{w}$ be a deterministic unit vector in $\mathbb{R}^M$. Let*

$$R := \sum_{\beta=1}^{M} \frac{d_\alpha \sqrt{d_\beta + 1}}{d_\alpha - d_\beta} \mathbf{v}_\beta \mathbf{v}_\beta^*, \tag{4.61}$$

and $\Xi := \Xi(z) = \mathcal{G}_1(z) - z^{-1}\Pi(z)$. We have that under the condition (2.2),

$$\begin{aligned}
\langle \mathbf{w}, u_\alpha(\mathcal{V})\rangle^2 = & \frac{d_\alpha^2 - y}{d_\alpha(d_\alpha + y)} |\mathbf{v}_\alpha^* \mathbf{w}|^2 \\
& - 2d_\alpha(d_\alpha+1)\mathbf{w}^*\mathbf{v}_\alpha\mathbf{v}_\alpha^*\Xi\mathbf{v}_\alpha\mathbf{v}_\alpha^*\mathbf{w} - 2\frac{f(d_\alpha)}{\sqrt{1+d_\alpha}}\mathbf{w}^*R^*\Xi\mathbf{v}_\alpha\mathbf{v}_\alpha^*\mathbf{w} \\
& - \frac{f(d_\alpha)^2}{1+d_\alpha}\mathbf{w}^*\mathbf{v}_\alpha\mathbf{v}_\alpha^*\Xi'\mathbf{v}_\alpha\mathbf{v}_\alpha^*\mathbf{w} + g(d_\alpha)\left(\mathbf{v}_\alpha^*\Xi R\mathbf{w}\right)^2 \\
& - d_\alpha(1+d_\alpha)g(d_\alpha) \sum_{\{1,\ldots,K\}\ni\beta\neq\alpha} \frac{d_\beta}{(d_\alpha - d_\beta)^2}\left(\mathbf{v}_\beta^*\Xi\mathbf{v}_\alpha\mathbf{v}_\alpha^*\mathbf{w}\right)^2 \\
& + \mathcal{E}_1(\mathbf{w}) + \mathcal{E}_2(\mathbf{w}),
\end{aligned} \tag{4.62}$$

where every instance of Ξ or Ξ' has argument $\phi_{N,\alpha}$ (defined in Theorem 2.2.1), and where

$$f(d) := \frac{1}{d}(d+1)(d^2 - y) = O(d^2), \quad g(d) := f(d)(d+y) = O(d^3). \tag{4.63}$$

Under Assumption 1 we have $\mathcal{E}_1(\mathbf{w}) \prec N^{-1}\left|\mathrm{proj}_{\mathbf{v}_1,\ldots,\mathbf{v}_K}\mathbf{w}\right|$ and $\mathcal{E}_2(\mathbf{w}) \prec N^{-3/2}$. Under Assumption 2 we rather have $\sqrt{\mathbb{E}\mathbf{1}_{\Omega_L}|\mathcal{E}_1(\mathbf{w})|^2} = O\left(N^{-1}\left|\mathrm{proj}_{\mathbf{v}_1,\ldots,\mathbf{v}_K}\mathbf{w}\right|\right)$ and $\sqrt{\mathbb{E}\mathbf{1}_{\Omega_L}|\mathcal{E}_2(\mathbf{w})|^2} = O(N^{-1-\epsilon})$

Lastly, we have the following weaker statement for the matrix $\widetilde{\mathcal{V}}$: under Assumption 2, we have

$$\sum_{i\in\mathcal{I}_M} \sqrt{\mathbb{E}\mathbf{1}_{\Omega_L}\left|\left\langle u_\alpha(\widetilde{\mathcal{V}}), \mathbf{e}_i\right\rangle^2 - C_{i,\alpha}\right|^2} = O(1), \tag{4.64}$$

where $C_{i,\alpha}$ are deterministic constants such that $\sum_{i\in\mathcal{I}_M}|C_{i,\alpha}| = O(1)$ uniformly in i, α.

Remark 4.3.6. We will only use the equation (4.64) once in the paper; to convert the expression $u_\alpha(\widetilde{\mathcal{V}})(D-I)^*u_\alpha(\widetilde{\mathcal{V}})$ in Lemma 4.1.6 to the better expression $u_\alpha(\mathcal{V})(D-I)^*u_\alpha(\mathcal{V})$.

The point of equation (4.64) is to ensure that $u_\alpha(\widetilde{\mathcal{V}})$ is the sum of a deterministic part and a very small part (so small that its components are absolutely summable).

Proof of Lemma 4.3.2. In this lemma we merely ensure that all but three of the terms in (4.62) are errors, drop the matrix R, and transition from the resolvent $\mathcal{G}_1$ to the resolvent G.

We will do the proof under Assumption 2; under Assumption 1 it is much easier and is omitted. We write

$$u_\alpha(\mathcal{V})^*(D-I)u_\alpha(\mathcal{V}) = \sum_{i\in\mathcal{I}_M} \langle \mathbf{e}_i, u_\alpha(\mathcal{V})\rangle^2 (D_{ii}-1) \tag{4.65}$$

and then apply lemma Theorem 4.3.5 with $\mathbf{w} = \mathbf{e}_i$. We temporarily define

$$\mathbf{M}_2^{(1)} = \frac{f(d_\alpha)}{\sqrt{1+d_\alpha}} R^*\Xi\mathbf{v}_\alpha\mathbf{v}_\alpha^*, \quad \mathbf{M}_3^{(1)} = g(d_\alpha)R^*\Xi\mathbf{v}_\alpha\mathbf{v}_\alpha^*\Xi R. \tag{4.66}$$

We first establish

$$\left| u_\alpha(\mathcal{V})^*(D-I)u_\alpha(\mathcal{V}) - \sum_{i\in\mathcal{I}_M} \left((\mathbf{M}_1 + \mathbf{M}_2^{(1)} + \mathbf{M}_3^{(1)})(D-1) \right)_{ii} \right| = O_P\left(N^{-1/2-\epsilon}\right) \tag{4.67}$$

by bounding the expectation on the event $\mathbf{1}_{\Omega_L}$ of the left-hand side by $O(N^{-1/2-\epsilon})$, which has contributions from the second, fourth, sixth, seventh and eighth terms of the right-hand side of equation (4.62) ($(\mathbf{M}_1 + \mathbf{M}_2^{(1)} + \mathbf{M}_3^{(1)})_{ii}$ are the first, third and fifth terms). Here we have used that a positive random variable X satisfies $X = O_P(\mathbb{E}X)$. We bound the second term's contribution

$$\begin{aligned} &\left| \mathbb{E}\mathbf{1}_{\Omega_L} \sum_{i\in\mathcal{I}_M} 2d_\alpha(d_\alpha+1)\mathbf{e}_i^*\mathbf{v}_\alpha\mathbf{v}_\alpha^*\Xi\mathbf{v}_\alpha\mathbf{v}_\alpha^*\mathbf{e}_i(D_{ii}-1) \right| \\ &\leq \sum_{i\in\mathcal{I}_M} 2d_\alpha(d_\alpha+1)|\mathbf{v}_\alpha^*\mathbf{e}_i|^2 \mathbb{E}\mathbf{1}_{\Omega_L}|\mathbf{v}_\alpha^*\Xi\mathbf{v}_\alpha(D_{ii}-1)|. \end{aligned} \tag{4.68}$$

Cauchy Schwarz on the expectation then yields

$$\begin{aligned} &\leq \sum_{i\in\mathcal{I}_M} 2d_\alpha(d_\alpha+1)|\mathbf{v}_\alpha^*\mathbf{e}_i|^2 \sqrt{\mathbb{E}\mathbf{1}_{\Omega_L}|\mathbf{v}_\alpha^*\Xi\mathbf{v}_\alpha|^2}\sqrt{\mathbb{E}(D_{ii}-1)^2} \\ &\leq C\sum_{i\in\mathcal{I}_M} 2d_\alpha(d_\alpha+1)|\mathbf{v}_\alpha^*\mathbf{e}_i|^2 d_\alpha^{-2}N^{-1/2}N^{-1/2} = O(N^{-1}) \end{aligned} \tag{4.69}$$

where we bounded $\sqrt{\mathbb{E}\mathbf{1}_{\Omega_L}|\mathbf{v}_\alpha^*\Xi\mathbf{v}_\alpha|^2}$ by Lemmas 4.2.6 and 4.2.10. The contribution of the fourth term is bounded in the same way, using $f(x) = O(x^2)$ and $\sqrt{\mathbb{E}\mathbf{1}_{\Omega_L}|\mathbf{v}_\alpha^*\Xi'\mathbf{v}_\alpha|^2} \leq d_\alpha^{-3}N^{-1/2}$; to see why the derivative has an extra factor of d_α^{-1}, let γ be a circular contour centered at $\phi_{N,\alpha}$ and having radius $\asymp d_\alpha$ and use the Cauchy integral formula to write

$$\sqrt{\mathbb{E}\mathbf{1}_{\Omega_L}|\mathbf{v}_\alpha^*\Xi'\mathbf{v}_\alpha|^2} = \sqrt{\mathbb{E}\mathbf{1}_{\Omega_L}\left|\int_\gamma \frac{\mathbf{v}_\alpha^*\Xi(z)\mathbf{v}_\alpha}{(z-\phi_{N,\alpha})^2}\mathrm{d}z\right|^2} \leq \int_\gamma \frac{\sqrt{\mathbb{E}|\mathbf{v}_\alpha^*\Xi(z)\mathbf{v}_\alpha|^2}}{(z-\phi_{N,\alpha})^2}\mathrm{d}z$$
$$= O(d_\alpha N^{-1/2}d_\alpha^{-4})$$

by Minkowski's integral inequality and the triangle inequality for integrals. The sixth term's contribution is also done in a very similar way, using in addition that $\sqrt{\mathbb{E}|\mathbf{v}_\beta\Xi\mathbf{v}_\alpha|^4} \leq d_\alpha^{-4}N^{-1}$ by Lemma 4.2.6 and Theorem 4.2.7. The contribution of the seventh term, $\mathcal{E}_1(\mathbf{e}_i)$, is

$$\begin{aligned}\mathbb{E}\mathbf{1}_{\Omega_L}\mathbf{1}_{\Omega_L}\sum_{i\in\mathcal{I}_M}|\mathcal{E}_1(\mathbf{e}_i)(D_{ii}-1)| &\leq \sum_{i\in\mathcal{I}_M}\sqrt{\mathbb{E}\mathbf{1}_{\Omega_L}|\mathcal{E}_1(\mathbf{e}_i)|^2}\sqrt{\mathbb{E}|D_{ii}-1|^2} \\ &\leq \sum_{i\in\mathcal{I}_M} N^{-1}\left|\mathrm{proj}_{\mathbf{v}_1,\ldots,\mathbf{v}_K}\mathbf{e}_i\right|N^{-1/2} \leq N^{1/2}N^{-1}N^{1/2} = N^{-1}\end{aligned} \tag{4.70}$$

by Cauchy-Schwarz on the sum. Lastly, the contribution of the eighth term $\mathcal{E}_2(\mathbf{e}_i)$ is

$$\begin{aligned}\mathbb{E}\mathbf{1}_{\Omega_L}\sum_{i\in\mathcal{I}_M}|\mathcal{E}_2(\mathbf{e}_i)(D_{ii}-1)| &\leq \sum_{i\in\mathcal{I}_M}\sqrt{\mathbb{E}\mathbf{1}_{\Omega_L}|\mathcal{E}_2(\mathbf{e}_i)|^2}\sqrt{\mathbb{E}|D_{ii}-1|^2} \\ &\leq \sum_{i\in\mathcal{I}_M} N^{-1-\epsilon}N^{-1/2} \leq N^{1/2}N^{-1}N^{1/2} = N^{-1/2-\epsilon}\end{aligned} \tag{4.71}$$

as desired.

Nextly, temporarily setting

$$\mathbf{M}_2^{(2)} = \frac{f(d_\alpha)}{\sqrt{1+d_\alpha}}\Xi\mathbf{v}_\alpha\mathbf{v}_\alpha^*, \quad \mathbf{M}_3^{(2)} = g(d_\alpha)\Xi\mathbf{v}_\alpha\mathbf{v}_\alpha^*\Xi,$$

which only differ from $\mathbf{M}_2^{(1)}$ and $\mathbf{M}_3^{(1)}$ in the absence of the matrix R, we establish

$$\sum_i\left|\mathbf{e}_i^*\left(\mathbf{M}_2^{(2)}-\mathbf{M}_2^{(1)}\right)\mathbf{e}_i(D_{ii}-1)\right| + \left|\mathbf{e}_i^*\left(\mathbf{M}_3^{(2)}-\mathbf{M}_3^{(1)}\right)\mathbf{e}_i(D_{ii}-1)\right| = O_P(N^{-1/2-\epsilon}).$$

Since $R-I$ has rank K, this step is done just like the previous step, and we omit the details.

To conclude the proof, we establish

$$\sum_{i\in\mathcal{I}_M}\left(\left|\mathbf{e}_i^*(\mathbf{M}_2-\mathbf{M}_2^{(2)})\mathbf{e}_i(D_{ii}-1)\right|+\left|\mathbf{e}_i^*(\mathbf{M}_3-\mathbf{M}_3^{(2)})\mathbf{e}_i(D_{ii}-1)\right|\right)=O_P(N^{-1/2-\epsilon}). \quad (4.72)$$

Note that $\mathbf{M}_\iota$ and $\mathbf{M}_\iota^{(1)}$ only differ in replacing Ξ with $\theta(d_\alpha)^{-1}U^*(G-\Pi)U$. Again it suffices to bound the expectation. We use Lemma 4.2.6. First,

$$\begin{aligned}
&\mathbb{E}\mathbf{1}_{\Omega_L}\sum_{i\in\mathcal{I}_M}\left|\mathbf{e}_i^*(\mathbf{M}_2-\mathbf{M}_2^{(1)})\mathbf{e}_i(D_{ii}-1)\right|\\
&=O\left(\sum_{i\in\mathcal{I}_M}d_\alpha^2|\mathbf{e}_i^*\mathbf{v}_\alpha|\sqrt{\mathbb{E}\mathbf{1}_{\Omega_L}\left|\mathbf{v}^*\big(\Xi-\theta(d_\alpha)^{-1}U^*(G-\Pi)U\big)\mathbf{e}_i\right|^2}\sqrt{\mathbb{E}|D_{ii}-1|^2}\right)\\
&=O\left(\sum_{i\in\mathcal{I}_M}d_\alpha^2|\mathbf{e}_i^*\mathbf{v}_\alpha|d_\alpha^{-2}N^{-1}N^{-1/2}\right)\\
&=O(N^{-1}).
\end{aligned} \quad (4.73)$$

Second, noting that, by difference of squares and $|(\Xi+z^{-1}U^*(G-\Pi)U)_{\mathbf{xy}}|\prec z^{-2}N^{-\epsilon}$ by Theorem 4.2.8 and Lemma 4.2.6, we have

$$\begin{aligned}
&\mathbb{E}\mathbf{1}_{\Omega_L}\left|\Xi^2_{\mathbf{v}_\alpha,\mathbf{e}_i}-\theta(d_\alpha)^{-2}(U^*(G-\Pi)U)^2_{\mathbf{v}_\alpha,\mathbf{e}_i}\right|^2\\
&=\mathbb{E}\mathbf{1}_{\Omega_L}\left|\Xi_{\mathbf{v}_\alpha,\mathbf{e}_i}-\theta(d_\alpha)^{-1}(U^*(G-\Pi)U)_{\mathbf{v}_\alpha,\mathbf{e}_i}\right|^2\\
&\qquad\cdot\left|\Xi_{\mathbf{v}_\alpha,\mathbf{e}_i}+\theta(d_\alpha)^{-1}(U^*(G-\Pi)U)_{\mathbf{v}_\alpha,\mathbf{e}_i}\right|^2\\
&\le\theta(d_\alpha)^{-4}N^{-2\epsilon}\mathbb{E}\mathbf{1}_{\Omega_L}\left|\Xi_{\mathbf{v}_\alpha,\mathbf{e}_i}-\theta(d_\alpha)^{-1}(U^*(G-\Pi)U)_{\mathbf{v}_\alpha,\mathbf{e}_i}\right|^2
\end{aligned} \quad (4.74)$$

and then the above is

$$\le\theta(d_\alpha)^{-8}N^{-2-2\epsilon} \quad (4.75)$$

by Lemma 4.2.6. So, we have

$$\begin{aligned}
&\mathbb{E}\mathbf{1}_{\Omega_L}\sum_{i\in\mathcal{I}_M}\left|\mathbf{e}_i^*(\mathbf{M}_3-\mathbf{M}_3^{(1)})\mathbf{e}_i(D_{ii}-1)\right|\\
&=O\left(\sum_{i\in\mathcal{I}_M}d_\alpha^4\sqrt{\mathbb{E}\mathbf{1}_{\Omega_L}\left|\Xi^2_{\mathbf{v}_\alpha,\mathbf{e}_i}-\theta(d_\alpha)^{-2}(U^*(G-\Pi)U)^2_{\mathbf{v}_\alpha,\mathbf{e}_i}\right|^2}\sqrt{\mathbb{E}|D_{ii}-1|^2}\right)\\
&=O\left(\sum_{i\in\mathcal{I}_M}d_\alpha^4d_\alpha^{-4}N^{-1-\epsilon}N^{-1/2}\right)\\
&=O(N^{-1/2-\epsilon}).
\end{aligned} \quad (4.76)$$

This concludes the proof of Lemma 4.3.2. ☐

Proof of Lemma 4.3.3. We treat the case that $\mathbf{w}$ has $\|\mathbf{w}\|_\infty \le N^{-\epsilon_1}$; the case of deterministic $\mathbf{w}$ is easier. Observe that by Lemma 4.2.11 and Markov's inequality, we have

$$P(|D_{ii} - 1| > N^{-a}) = O(N^{-2+4a}). \tag{4.77}$$

This yields on one hand, by a union bound, the bound

$$\begin{aligned} P(\|D - I\|^2 > N^{-1/2+\epsilon}) &= P(\exists i : |D_{ii} - 1|^2 > N^{-1/2+\epsilon}) \\ &= O(NN^{-1-2\epsilon}) \\ &= O(N^{-4\epsilon}). \end{aligned} \tag{4.78}$$

On the other hand, we have for any $\epsilon > 0$,

$$\begin{aligned} \mathbb{E}\#\{i : |D_{ii} - 1|^2 > N^{-1/2-\epsilon}\} &= \mathbb{E} \sum_{i \in \mathcal{I}_M} \mathbf{1}_{|D_{ii}-1|^2 > N^{-1/2-\epsilon}} \\ &= \sum_{\mathcal{I}_M} P(|D_{ii} - 1|^2 > N^{-1/2-\epsilon}) \\ &\le N^{2\epsilon}. \end{aligned} \tag{4.79}$$

Together, equations (4.78) and (4.79) say that for any $\epsilon > 0$, only a few (about $N^{2\epsilon}$) entries of $(D - I)^2$ are expected to exceed $N^{-1/2-\epsilon}$, and those that do are all a.a.s. bounded by $N^{-1/2+\epsilon}$.

Therefore, we bound

$$
\begin{aligned}
\mathbf{w}^*(D-I)^2\mathbf{w} &= \sum_i \langle \mathbf{w}, \mathbf{e}_i\rangle^2 (D_{ii}-1)^2 \\
&= \sum_{i:|D_{ii}-1|^2 > N^{-1/2-\frac{1}{2}\epsilon_{\mathbf{w}}}} \langle \mathbf{w}, \mathbf{e}_i\rangle^2 (D_{ii}-1)^2 \\
&\quad + \sum_{i:|D_{ii}-1|^2 \le N^{-1/2-\frac{1}{2}\epsilon_{\mathbf{w}}}} \langle \mathbf{w}, \mathbf{e}_i\rangle^2 (D_{ii}-1)^2 \\
&\le \sum_{i:|D_{ii}-1|^2 > N^{-1/2-\frac{1}{2}\epsilon_{\mathbf{w}}}} N^{-2\epsilon_{\mathbf{w}}} \left\|(D-I)^2\right\| \\
&\quad + \sum_{i:|D_{ii}-1|^2 \le N^{-1/2-\frac{1}{2}\epsilon_{\mathbf{w}}}} \langle \mathbf{w}, \mathbf{e}_i\rangle^2 N^{-1/2-\frac{1}{2}\epsilon_{\mathbf{w}}} \\
&\le \#\{i : |D_{ii}-1|^2 > N^{-1/2-\frac{1}{2}\epsilon_{\mathbf{w}}}\} N^{-2\epsilon_{\mathbf{w}}} \left\|(D-I)^2\right\| \\
&\quad + N^{-1/2-\frac{1}{2}\epsilon_{\mathbf{w}}} \\
&= O_P\left((N^{\frac{1}{4}\epsilon_{\mathbf{w}}} N^{\epsilon_{\mathbf{w}}}) N^{-2\epsilon_{\mathbf{w}}} \left\|(D-I)^2\right\| + N^{-1/2-\frac{1}{2}\epsilon_{\mathbf{w}}} \right) \\
&= O_P\left(N^{-1/2-\frac{1}{2}\epsilon_{\mathbf{w}}} \right),
\end{aligned}
\tag{4.80}
$$

where in the second to last line we used that the random variable $Y = \#\{i : |D_{ii}-1|^2 > N^{-1/2-\frac{1}{2}\epsilon_{\mathbf{w}}}\}$ is bounded by $O_P(N^{\frac{1}{4}\epsilon_{\mathbf{w}}}\mathbb{E}Y)$ (this is a general fact about positive random variables that follows from Markov's inequality), and in the last line that $\|(D-I)^2\| \le N^{-1/2+\frac{1}{4}\epsilon_{\mathbf{w}}}$ a.a.s.

Lastly, the bound on $\sqrt{\mathbf{w}^*(D-I)^4\mathbf{w}} \prec N^{-1/2-\epsilon}$ follows in amost exactly the same way; we omit the details. ∎

Proof of Lemma 4.3.4. The first inequality is an immediate consequence of Weyl's inequality. The second follows from the formula

$$
\partial_t u_\alpha((\widetilde{S}X + t\Delta)^{\#}) = \sum_{\beta \ne \alpha} \frac{u_\beta u_\beta^* \left((\widetilde{S}X + t\Delta)\Delta^* + \Delta(\widetilde{S}X + t\Delta)^* \right) u_\alpha}{\lambda_\beta - \lambda_\alpha} \tag{4.81}
$$

where the implied argument of every eigenvalue and eigenvector on the right-hand side is

$(\widetilde{S}X + t\Delta)^{\#}$. Thus we have

$$\begin{aligned}
\left\|\partial_t u_\alpha((\widetilde{S}X + t\Delta)^{\#})\right\|^2 &= \left\|\sum_{\beta\neq\alpha} \frac{u_\beta u_\beta^* \left((\widetilde{S}X + t\Delta)\Delta^* + \Delta(\widetilde{S}X + t\Delta)^*\right) u_\alpha}{\lambda_\beta - \lambda_\alpha}\right\|^2 \\
&= \sum_{\beta\neq\alpha} \left\|\frac{u_\beta u_\beta^* \left((\widetilde{S}X + t\Delta)\Delta^* + \Delta(\widetilde{S}X + t\Delta)^*\right) u_\alpha}{\lambda_\beta - \lambda_\alpha}\right\|^2 \\
&= O_P\left(4\sum_{\beta\neq\alpha} \left\|\frac{u_\beta u_\beta^* \left((\widetilde{S}X + t\Delta)\Delta^* + \Delta(\widetilde{S}X + t\Delta)^*\right) u_\alpha}{d_1}\right\|^2\right) \qquad (4.82) \\
&= O_P\left(\frac{4}{d_1^2}\left\|\left((\widetilde{S}X + t\Delta)\Delta^* + \Delta(\widetilde{S}X + t\Delta)^*\right) u_\alpha\right\|^2\right) \\
&= O_P\left(\frac{4}{d_1}\|\Delta\|^2\right)
\end{aligned}$$

as desired. □

4.4 Proof of Lemma 4.1.7

In this section , we will use without proof the following essential and well-known fact regarding Gaussian matrices:

Lemma 4.4.1. *For an $(\mathcal{I}_M \cup \mathcal{I}_K) \times \mathcal{I}_N$ iid Gaussian matrix X and deterministic orthogonal vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{I}_M$, the $2n$ quantities*

$$\frac{1}{\|\mathbf{x}_1^* X\|}\mathbf{x}_1^* X, \ldots, \frac{1}{\|\mathbf{x}_n^* X\|}\mathbf{x}_n^* X, \quad \|\mathbf{x}_1^* X\|, \ldots, \|\mathbf{x}_n^* X\| \qquad (4.83)$$

are mutually independent.

We recall our formula for the eigenvalue correction from Lemma 4.3.2

$$\mathbf{u}^*(D - I)\mathbf{u} = \sum_{i\in\mathcal{I}_M} \langle \mathbf{e}_i, \mathbf{M}\mathbf{e}_i\rangle (D_{ii} - 1) + O_\prec(N^{-1/2-\epsilon}). \qquad (4.84)$$

Recall also that $\mathbf{M}$ is a rank one matrix defined from G, and $D_{ii} = \left\|\mathbf{e}_i^* \widetilde{S}X\right\|^2$.

We look at high moments of the correction (4.84)

$$\begin{aligned}&\mathbb{E}\left|\sum_{i\in\mathcal{I}_M}\langle\mathbf{e}_i,\mathbf{M}\mathbf{e}_i\rangle(D_{ii}-1)\right|^p\\&=\sum_{i_1,\dots,i_p=1}^{M}\mathbb{E}\langle\mathbf{e}_{i_1},\mathbf{M}\mathbf{e}_{i_1}\rangle\cdots\langle\mathbf{e}_{i_p},\mathbf{M}\mathbf{e}_{i_p}\rangle(D_{i_1i_1}-1)\cdots(D_{i_pi_p}-1).\end{aligned}\tag{4.85}$$

Note that $\mathbf{M}$ is a real matrix, being defined from Green functions evaluated at real spectral arguments.

Right away, we will fix a partition

$$2^{\{1,\dots,p\}}\supseteq\mathcal{P}=\{A_1,\dots,A_{p'}\}$$

of the set $\{1,\dots,p\}$. The sets A_r are called blocks of $\mathcal{P}$. We will say that a tuple $\boldsymbol{i}:=(i_1,\dots,i_p)$ *satisfies* $\mathcal{P}$ and write $\boldsymbol{i}\vdash\mathcal{P}$ when $i_r=i_s$ if and only if r and s are in the same block of the partition. We will henceforth only consider

$$\sum_{\boldsymbol{i}\vdash\mathcal{P}}\mathbb{E}\langle\mathbf{e}_{i_1},\mathbf{M}\mathbf{e}_{i_1}\rangle\cdots\langle\mathbf{e}_{i_p},\mathbf{M}\mathbf{e}_{i_p}\rangle(D_{i_1i_1}-1)\cdots(D_{i_pi_p}-1),\tag{4.86}$$

i.e., the sum over all values of the indices $i_1,\dots,i_p$ for which $\boldsymbol{i}=(i_1,\dots,i_p)\vdash\mathcal{P}$. Because the number of partitions of a p-element set is a constant C_p (ie, it does not depend on N), this does not affect our desired $O_\prec$ bound. Denote by ℓ_{i} the number of singletons in $\mathcal{P}$—the indices $i_1,\dots,i_{\ell_{\mathrm{i}}}$ are "lone indices". We will only treat the case that the singletons of $\mathcal{P}$ are $\{1\},\dots,\{\ell_{\mathrm{ii}}\}$ and that the numbers $1,\dots,p'$ all belong to different blocks of the partition; ie, $r\in A_r$ for $r\le p'$. Every other partition is isomorphic to one like this via a permutation of the set $\{1,\dots,p\}$, and these may be treated in the same way.

Note that, by equation (2.3), the unit vectors

$$\widetilde{S}^*\mathbf{e}_{i_1},\dots,\widetilde{S}^*\mathbf{e}_{i_{p'}},\tag{4.87}$$

are nearly orthogonal, each having dot product $N^{-\epsilon_D}$ with each other. We apply Gram-Schmidt to get an orthonormalized set

$$\mathcal{S}_{i_1},\dots,\mathcal{S}_{i_{p'}}\tag{4.88}$$

satisfying $\|\mathcal{S}_{i_r}\| = 1$ for each $r = 1, \ldots, p'$ and $\mathcal{S}_{i_r} \perp \mathcal{S}_{i_s}$ for each $i_r \neq i_s$ and $\widetilde{S}^* \mathbf{e}_{i_r} = \sum_{s=1}^{p'} a_{r,s} \mathcal{S}_{i_s}$. Note that $a_{r,s} = \delta_{rs} + O(N^{-\epsilon_D})$ and that $\sum_{s=1}^{p'} a_{rs}^2 = 1$.

We define now, for each $i = 1, \ldots, p'$ the random vectors

$$\mathbf{y}_r := \mathbf{y}_r(i_1, \ldots, i_p, \mathcal{P}) = \mathcal{S}_{i_r}^* X,$$

so that the vectors $\mathbf{y}_1, \ldots, \mathbf{y}_{p'}$ are iid N-dimensional Gaussian random vectors with covariance $N^{-1} I_{\mathcal{I}_N}$.

The main heuristic of this section is that the centered random variables $(D_{i_1 i_1} - 1) \cdots (D_{i_p i_p} - 1)$, being functions of the lengths of $\mathbf{e}_{i_1}^* \widetilde{S} X, \ldots, \mathbf{e}_{i_{p'}}^* \widetilde{S} X$, are by Lemma 4.4.1 and (2.3) nearly independent. Since the lengths of the vectors $\mathbf{e}_{i_1}^* \widetilde{S} X, \ldots, \mathbf{e}_{i_{p'}}^* \widetilde{S} X$ are all $1 + O_{\prec}(N^{-1/2})$ while their directions are completely delocalized, we expect $\mathbf{M}$ to be largely independent of $(D_{i_1 i_1} - 1), \cdots, (D_{i_p i_p} - 1)$. This near-independence makes for additional smallness in the expression (4.85). The main technical difficulties of this section are first to modify the random variables $(D_{i_1 i_1} - 1), \cdots, (D_{i_p i_p} - 1)$ so as to make them exactly independent of one another (and quantify the error in doing so), leading to the notion of what we will call *lone factors*, and second to perturb the matrix $\mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3$ so as to make it exactly independent of $(D_{i_1 i_1} - 1), \cdots, (D_{i_p i_p} - 1)$ (and quantify the error associated to this perturbation).

The first thing we must do is rewrite each $D_{i_r i_r} - 1 = \left\| \mathbf{e}_{i_r}^* \widetilde{S} X \right\|^2 - 1$ in such a way as rather to involve the vectors $\mathbf{y}_r$. We may write

$$
\begin{aligned}
D_{i_r i_r} - 1 &= \left\langle \sum_{s=1}^{p'} a_{r,s}\mathbf{y}_s, \sum_{s=1}^{p'} a_{r,s}\mathbf{y}_s \right\rangle - 1 \\
&= \sum_{s=1}^{p'} a_{r,s}^2(\|\mathbf{y}_s\|^2 - 1) + \sum_{s\neq t} a_{r,s}\cdot a_{r,t}\langle \mathbf{y}_s, \mathbf{y}_t\rangle \\
&= \sum_{s=1}^{p'} a_{r,s}^2(\|\mathbf{y}_s\|^2 - 1) + \sum_{s\neq t} a_{r,s}\cdot a_{r,t}\langle \mathbf{N}(\mathbf{y}_s), \mathbf{N}(\mathbf{y}_t)\rangle \\
&\quad + \sum_{s\neq t} 2a_{r,s}\cdot a_{r,t}\langle (1-\mathbf{N})(\mathbf{y}_s), \mathbf{N}(\mathbf{y}_t)\rangle \\
&\quad + \sum_{s\neq t} 2a_{r,s}\cdot a_{r,t}\langle (1-\mathbf{N})(\mathbf{y}_s), (1-\mathbf{N})(\mathbf{y}_t)\rangle .
\end{aligned} \tag{4.89}
$$

Here we used that $\sum_{s=1}^{p} a_{r,s}^2 = 1$.

We now have that equation (4.86) can be written as a sum of boundedly many terms of the form

$$
\sum_{i_1,\ldots,i_p \vdash \mathcal{P}} \mathbb{E}\langle \mathbf{e}_{i_1}, \mathbf{M}\mathbf{e}_{i_1}\rangle \cdots \left\langle \mathbf{e}_{i_p}, \mathbf{M}\mathbf{e}_{i_p}\right\rangle \mathcal{L}_{1s_1t_1}^{(\iota_1)} \cdots \mathcal{L}_{ps_pt_p}^{(\iota_p)}, \tag{4.90}
$$

where the factors $\mathcal{L}^{(\iota_r)} := \mathcal{L}_{rst}^{(\iota_r)}$ for $\iota = 1, \ldots, 5$ are defined as follows

- $\mathcal{L}_{rst}^{(1)} = a_{r,r}^2 \left(\left\| \mathcal{S}_{i_r}^* X \right\|^2 - 1 \right)$. Note that $a_{r,r}$ has size $1 + O(N^{-\epsilon})$. Terms of this form are the leading term from the expansion (4.162) and have the ($O_\prec$) bound $N^{-1/2}$. The subscripts s, t are superfluous but we include them to keep the notation consistent.

- $\mathcal{L}_{rst}^{(2)} = \sum_{s\neq r} a_{r,s}^2 \left(\left\| \mathcal{S}_{i_r}^* X \right\|^2 - 1 \right)$ for $s \neq r$. These have the bound $N^{-2\epsilon_D} N^{-1/2}$. The subscript t is again superfluous.

- $\mathcal{L}_{rst}^{(3)} = a_{r,s}a_{r,t}\left\langle \mathbf{N}(\mathcal{S}_{i_s}^* X), \mathbf{N}(\mathcal{S}_{i_t}^* X)\right\rangle$ for $s \neq t$. These have the bound $N^{-\epsilon_D} N^{-1/2}$.

- $\mathcal{L}_{rst}^{(4)} = 2a_{r,s}a_{r,t}\left\langle \mathbf{N}(\mathcal{S}_{i_s}^* X), (1-\mathbf{N})(\mathcal{S}_{i_t}^* X)\right\rangle$ for $s \neq t$. These have the bound $N^{-\epsilon_D} N^{-1} \leq N^{-2\epsilon_D} N^{-1/2}$, since we have a dot product of nearly orthogonal vectors, one of whose lengths is in addition small.

- $\mathcal{L}^{(5)}_{rst} = a_{r,s}a_{r,t}\big\langle(1-\mathbf{N})(\mathcal{S}^*_{i_s}X), (1-\mathbf{N})(\mathcal{S}^*_{i_t}X)\big\rangle$ for $s \neq t$. These have the bound $N^{-\epsilon_D}N^{-3/2} \leq N^{-3\epsilon_D}N^{-1/2}$.

Note that $\mathcal{L}^{(1)}_{rst}$ has the same bound $O_\prec(N^{-1/2})$ as $D_{i_r i_r} - 1$, while $\mathcal{L}^{(2)}_{rst}, \dots, \mathcal{L}^{(5)}_{rst}$ also contain extra factors of $N^{-\epsilon_D}$.

Recall that the expression (4.90) is considered as fixed. Now we call a factor $\mathcal{L}^{(\iota_r)}_{rst}$ *lone* if $\iota = 1$ and if in the expression (4.90) *none* of the following are satisfied:

1. There is another factor $\mathcal{L}^{(1)}_{r's't'}$ for $r' = r$.
2. There is a factor $\mathcal{L}^{(2)}_{r's't'}$ for $s' = r$.
3. There is a factor $\mathcal{L}^{(4)}_{r's't'}$ for $t' = r$.
4. There is a factor $\mathcal{L}^{(5)}_{r's't'}$ for $s' = r$ or $t' = r$.

The rationale behind this definition is that a lone factor $\mathcal{L}^{(\iota_r)}_{rst}$ is independent of all the other factors $\mathcal{L}^{(\iota_{r'})}_{r's't'}$ for $r' \neq r$ by Lemma 4.4.1, because it is a function of the length of $\mathbf{y}_r$, whereas $\mathcal{L}^{(\iota_{r'})}_{r's't'}$ is a function of the direction of $\mathbf{y}_r$ and of the vectors $\mathbf{y}_{r'}$ for $r' \neq r$. The reason there is no reference to a factor of type (3) in the definition of a lone $\mathcal{L}^{(\iota_r)}_{rst}$ factor is that a factor of type (3) is independent of the lengths of all vectors $\mathbf{y}_1, \dots, \mathbf{y}_p$. We will see very shortly the utility in this definition of lone factors.

Recall our partition $\mathcal{P}$. See that

$$\ell_{\mathrm{i}} + 2(p - |\mathcal{P}|) \geq p, \tag{4.91}$$

which follows from the equation

$$2 \cdot \#(\text{non-lone indices}) + \#(\text{lone indices}) \leq p.$$

Let also ℓ_{f} be the number of lone factors in the expression (4.90).

Lemma 4.4.2. *We have*

$$\left|\mathcal{L}^{(\iota_1)}_{1s_1t_1}\cdots\mathcal{L}^{(\iota_p)}_{1s_pt_p}\right| \prec N^{-(\ell_{\mathrm{i}}-\ell_{\mathrm{f}})\epsilon_D}N^{-p/2}.$$

Proof. The way this lemma should be thought about is that the fewer lone factors there are, the more extra factors of $N^{-\epsilon_D}$ you get. Say first that in the expression (4.90), $\iota_r = 1$ for $r = 1, \ldots, p$. Then ℓ_{f} is equal to ℓ_{i}. r for which $\iota_r = 2$, we accumulate a deterministic factor $N^{-2\epsilon_D}$ and ℓ_{f} decreases by as much as 2, since the factor $\mathcal{L}^{(\iota_r)}_{rst}$ is no longer of the form (1)—therefore $\mathcal{L}^{(\iota_r)}_{rst}$ cannot be lone—and it has the potential to force some other $\mathcal{L}^{(\iota_{r'})}_{r's't'}$ to cease to be lone. For example, if i_1 and i_2 are both lone indices, and if $D_{i_1i_1} - 1$ yields $a_{12}^2(\|\mathbf{y}_1\|^2 - 1)$, then $\mathcal{L}^{(\iota_1)}_{1s_1t_1}$ is not lone, and moreover, regardless of what $D_{i_2i_2} - 1$ yields upon expansion according to (4.162), $\mathcal{L}^{(\iota_2)}_{2s_st_s}$ may not be lone. Similarly, if any r is such that ι_r is $3, 4$ or 5, then ℓ_{f} is reduced by as much as 1, 2, or 3, respectively, and we simultaneously accumulate a factor of $N^{-\epsilon_D}$, $N^{-2\epsilon_D}$, or $N^{-3\epsilon_D}$, respectively. This concludes the proof.

□

See at this point that equation (4.90) has the naive bound (naive in that we do not use the expectation or exploit any independence)

$$\begin{aligned}
&\prec N^{-p/2}N^{-(\ell_{\mathrm{i}}-\ell_{\mathrm{f}})\epsilon_D}\left|\sum_{i_1,\ldots,i_p\vdash\mathcal{P}}\mathbb{E}\langle\mathbf{e}_{i_1},\mathbf{M}\mathbf{e}_{i_1}\rangle\cdots\left\langle\mathbf{e}_{i_p},\mathbf{M}\mathbf{e}_{i_p}\right\rangle\right| \\
&\prec N^{-p/2}N^{-(\ell_{\mathrm{i}}-\ell_{\mathrm{f}})\epsilon_D}N^{-2(p-|\mathcal{P}|)\epsilon_D}\sum_{\substack{i_1,\ldots,i_{p'}\\ \text{distinct}}}\mathbb{E}\left|\langle\mathbf{e}_{i_1},\mathbf{M}\mathbf{e}_{i_1}\rangle\cdots\left\langle\mathbf{e}_{i_{p'}},\mathbf{M}\mathbf{e}_{i_{p'}}\right\rangle\right| \qquad (4.92)\\
&\prec N^{-p/2}N^{-(\ell_{\mathrm{i}}-\ell_{\mathrm{f}})\epsilon_D}N^{-2(p-|\mathcal{P}|)\epsilon_D}\cdot 1
\end{aligned}$$

since $\langle\mathbf{e}_{i_1},\mathbf{M}\mathbf{e}_{i_1}\rangle \prec N^{-2\epsilon_D}$ by Lemma 4.2.11.1, and $\mathrm{Tr}|\mathbf{M}| \prec 1$. Combining equations (4.91) and (4.92), we see that we only need to find ℓ_{f} more factors of $-\epsilon_D$.

If $\mathcal{L}^{(\iota_r)}_{rst}$ is lone, then it is centered and independent of all other factors $\mathcal{L}^{(\iota_{r'})}_{r's't'}$ for $r' \neq r$, so if we can perturb the factors $\langle\mathbf{e}_{i_1},\mathbf{M}\mathbf{e}_{i_1}\rangle\cdots\left\langle\mathbf{e}_{i_p},\mathbf{M}\mathbf{e}_{i_p}\right\rangle$ to make them also independent of $\mathcal{L}^{(\iota_r)}_{rst}$, then the expression (4.90) will be 0 because of the expecation in it. The error terms associated to this perturbation will each have a form similar to (4.90), but with an additional factor of at least $N^{-\epsilon_D}$. To put it symbolically,

$$(4.90) = 0 + N^{-\epsilon_D}\,(\text{something resembling (4.90)})\,.$$

If $\mathcal{L}_2$ is also lone, one may perturb again to an expression with expectation 0; the error terms associated to this perturbation will be smaller by an additional factor of $N^{-\epsilon_D}$, and so on, one perturbation for each lone factor. In this way to we hope to add an additional factor of $N^{-\ell_{\mathrm{f}}\epsilon_D}$ to the bound (4.92). Constructing pertrubations with these properties is the object of Lemmas 4.4.3 and 4.4.5.

Now we describe how to use independence to get these ℓ_{f} factors of $-\epsilon_D$. Assume without loss of generality and for notational simplicity that the lone factors are $\mathcal{L}^{(\iota_1)}_{1s_1t_1},\dots,\mathcal{L}^{(\iota_{\ell_{\mathrm{f}}})}_{\ell_{\mathrm{f}}s_{\ell_{\mathrm{f}}}t_{\ell_{\mathrm{f}}}}$.

For a vector $\mathbf{x}\in\mathbb{R}^{\mathcal{I}}$, we define $P_{\mathbf{x}}$ to be the linear operator which projects onto the span of $\mathbf{x}$.

Lemma 4.4.3. *Fix a bounded number n and orthonormal vectors $\mathbf{x}_1,\dots,\mathbf{x}_n\in\mathbb{R}^{\mathcal{I}_K\cup\mathcal{I}_M}$. Define a matrix flow $\boldsymbol{t}:=(t_1,\dots,t_{\ell_{\mathrm{f}}})\mapsto G^{\boldsymbol{t}}_{\mathbf{x}_1,\dots,\mathbf{x}_n}$ through*

where

$$G^{\boldsymbol{t}}_{\mathbf{x}_1,\dots,\mathbf{x}_n} := \left(H^{\boldsymbol{t}}_{\mathbf{x}_1,\dots,\mathbf{x}_n}\right)^{-1},$$
$$H^{\boldsymbol{t}}_{\mathbf{x}_1,\dots,\mathbf{x}_n} := \begin{pmatrix} -1 & X^{\boldsymbol{t}}_{\mathbf{x}_1,\dots,\mathbf{x}_n} \\ \left(X^{\boldsymbol{t}}_{\mathbf{x}_1,\dots,\mathbf{x}_n}\right)^* & -z \end{pmatrix}$$

and

$$X^{\boldsymbol{t}}_{\mathbf{x}_1,\dots,\mathbf{x}_n} := X + \sum_{r=1}^{n} t_r \left(\frac{1}{\|\mathbf{x}_r^* X\|} - 1\right) P_{\mathbf{x}_r} X,$$

so that $X^{1,\dots,1}_{\mathbf{x}_1,\dots,\mathbf{x}_n}$ is X with each of its $\mathbf{x}_r$ components normalized, and $X^{\boldsymbol{t}}_{\mathbf{x}_1,\dots,\mathbf{x}_n}$ interpolates between $X = X^{0,\dots,0}_{\mathbf{x}_1,\dots,\mathbf{x}_n}$ and $X^{1,\dots,1}_{\mathbf{x}_1,\dots,\mathbf{x}_n}$. Then $H^{\boldsymbol{t}}$ is independent of $\|\mathbf{x}_r^* X\|$ for each $r\in\{1,\dots,n\}$ for which $t_r = 1$.

We will prove Lemma 4.4.3 at the end of this section. Now we will use it to complete the proof of Theorem 4.1.7.

For any choice of indices $\boldsymbol{i} := (i_1, \ldots, i_p)$, we set $n = \ell_\mathrm{f}$ and $\mathbf{x}_r = \mathcal{S}_{i_r}$ in Lemma 4.4.3 and define the functions

$$g_{i_r}(\boldsymbol{i}, \boldsymbol{t}) = \Big\langle \mathbf{e}_{i_r}, \mathbf{M}^{\boldsymbol{t}}_{\mathcal{S}_{i_1},\ldots,\mathcal{S}_{i_{\ell_\mathrm{f}}}} \mathbf{e}_{i_r} \Big\rangle.$$

Definition 4.4.4. [Difference Operators] For any disjoint subsets $\{r_1 \neq \cdots \neq r_m\}$ and $\{s_1 \neq \cdots \neq s_n\}$ of $\{1, \ldots, \ell_\mathrm{f}\}$, we define the *difference operators*.

$$\Delta^{r_1,\ldots,r_m} : C([0,1]^{\{s_1,\ldots,s_n,r_1,\ldots,r_m\}}) \to C([0,1]^{\{s_1,\ldots,s_n\}})$$

inductively through

$$\Delta^{r_1,\ldots,r_m} = \Delta^{r_1} \circ \Delta^{r_2,\ldots,r_m}$$

and

$$(\Delta^{r_1} f)(t_{s_1}, \ldots, t_{s_n}) = f(t_{s_1}, \ldots, 0, \ldots, t_{s_n}) - f(t_{s_1}, \ldots, 1, \ldots, t_{s_n})$$

where the 1 and 0 are in the position corresponding to r_1. Note the ambiguity in the domain of $\Delta^{r_1,\ldots,r_m}$; it acts on $C([0,1]^{\{s_1,\ldots,s_n,r_1,\ldots,r_m\}})$ in a well-defined way for any $s_1 \neq \ldots \neq s_n \in \{1, \ldots, \ell_\mathrm{f}\} \setminus \{r_1, \ldots, r_m\}$.

We may therefore write equation (4.90) as, for instance,

$$\begin{aligned} &N^{-(\ell_\mathrm{i}-\ell_\mathrm{f})\epsilon_D} \sum_{\boldsymbol{i}\vdash\mathcal{P}} \mathbb{E} g_{i_1}(\mathbf{0}) \cdots g_{i_p}(\mathbf{0}) \widetilde{\mathcal{L}}_{i_1} \cdots \widetilde{\mathcal{L}}_{i_p} \\ &= N^{-(\ell_\mathrm{i}-\ell_\mathrm{f})\epsilon_D} \sum_{\boldsymbol{i}\vdash\mathcal{P}} \mathbb{E}\left(g_{i_1}(1, 0, \ldots, 0) + \Delta^1 g_{i_1}(\mathbf{0}) \right) \\ &\qquad \cdots \left(g_{i_p}(1, 0, \ldots, 0) + \Delta^1 g_{i_p}(\mathbf{0}) \right) \widetilde{\mathcal{L}}_{i_1} \cdots \widetilde{\mathcal{L}}_{i_p}. \end{aligned} \tag{4.93}$$

We can thus split each g_i into two pieces according to each $r = 1, \ldots, \ell_\mathrm{f}$. We define, for any subset $U \subseteq \{1, \ldots, \ell_\mathrm{f}\}$,

$$g_i^U(\boldsymbol{i}) := \left(\prod_{r\in U} \Delta^r \right) g(\boldsymbol{i}, 1, \ldots, 1) = \Delta^U g(\boldsymbol{i}, 1, \ldots, 1)$$

($\left(\prod_{r\in\{1,2,3\}} \Delta^r\right)$ is the composition $\Delta^1 \circ \Delta^2 \circ \Delta^3$; one can verify this composition is commutative). Implicitly, the domain of $\Delta^U g(\boldsymbol{i}, \cdot)$ is $[0,1]^{\{1,\ldots,\ell_\mathrm{f}\}\setminus U}$. Then, we may write equation

(4.93) as

$$\begin{aligned}&\sum_{\boldsymbol{i}\vdash\mathcal{P}}\mathbb{E}g_1(\boldsymbol{i},\boldsymbol{0})\cdots g_p(\boldsymbol{i},\boldsymbol{0})\widetilde{\mathcal{L}}_{i_1}\cdots\widetilde{\mathcal{L}}_{i_p}\\&=\sum_{\substack{U_1,\ldots,U_p\\\subseteq\{1,\ldots,\ell_{\mathrm{f}}\}}}\sum_{\boldsymbol{i}\vdash\mathcal{P}}\mathbb{E}g_1^{U_1}(\boldsymbol{i})\cdots g_p^{U_p}(\boldsymbol{i})\widetilde{\mathcal{L}}_{i_1}\cdots\widetilde{\mathcal{L}}_{i_p}.\end{aligned}\tag{4.94}$$

For any choice of $U_1,\ldots,U_p$ for which $\sum_{a=1}^p|U_a|<\ell_{\mathrm{f}}$, we have necessarily for some r that $r\notin U_a$ for each $a\in\{1,\ldots,\ell_{\mathrm{f}}\}$. Then $g_{i_a}^{U_a}(\boldsymbol{i})$ is independent of $\mathcal{S}_{i_r}$ for every $a\in\{1,\ldots,\ell_{\mathrm{f}}\}$, so that

$$\sum_{\boldsymbol{i}\vdash\mathcal{P}}\mathbb{E}g_{i_1}^{s_1}\cdots g_{i_p}^{s_p}\widetilde{\mathcal{L}}_{i_1}\cdots\widetilde{\mathcal{L}}_{i_p}=0\tag{4.95}$$

since $\widetilde{\mathcal{L}}_{i_r}$ is lone. Therefore we may restrict our attention to choices of $U_1,\ldots,U_p\subseteq\{1,\ldots,\ell_{\mathrm{f}}\}$ for which $\sum_{a=1}^p|U_a|\geq\ell_{\mathrm{f}}$.

Now, for any choice of $U_1,\ldots,U_p$ for which $\sum_{a=1}^p|U_a|\geq\ell_{\mathrm{f}}$, we have

$$\begin{aligned}&N^{-(\ell_{\mathrm{i}}-\ell_{\mathrm{f}})\epsilon_D}\sum_{\boldsymbol{i}\vdash\mathcal{P}}\Big|g_1^{U_1}(\boldsymbol{i})\cdots g_p^{U_p}(\boldsymbol{i})\widetilde{\mathcal{L}}_{i_1}\cdots\widetilde{\mathcal{L}}_{i_p}\Big|\\&\prec N^{-(\ell_{\mathrm{i}}-\ell_{\mathrm{f}})\epsilon_D}N^{-p/2}\cdot\sum_{\boldsymbol{i}\vdash\mathcal{P}}\big|g_1^{U_1}(\boldsymbol{i})\big|\cdots\big|g_p^{U_p}(\boldsymbol{i})\big|\\&\prec N^{-(\ell_{\mathrm{i}}-\ell_{\mathrm{f}})\epsilon_D}N^{-\frac{1}{2}p}N^{-2(p-p')\epsilon_D}\cdot\sum_{\substack{i_1,\ldots,i_{p'}\\\text{distinct}}}\Big|g_1^{U_1}(\boldsymbol{i})\Big|\cdots\Big|g_{p'}^{U_{p'}}(\boldsymbol{i})\Big|.\end{aligned}\tag{4.96}$$

We have the following lemma about the functions $g_1^{U_1}(\boldsymbol{i})$, proven at the end of this section.

Lemma 4.4.5. *We have, for $r=1,\ldots,p'$,*

$$\sum_{i_r}\max_{i_1,\ldots,i_{r-1},i_{r+1},\ldots,i_{p'}}\big|g_r^{U_r}(\boldsymbol{i})\big|\prec N^{-|U|_r\epsilon_D}.$$

Thus, the above may be bounded by

$$\begin{aligned}&\prec N^{-(\ell_{\mathrm{i}}-\ell_{\mathrm{f}})\epsilon_D}N^{-\epsilon_D\sum_{a=1}^p\|s_a\|}N^{-\frac{1}{2}p}N^{-2(p-p')\epsilon_D}\cdot 1\\&\leq N^{-(1/2-\epsilon_D)p}.\end{aligned}$$

Thus we conclude the proof of Lemma 4.1.7.

Proof of Lemma 4.4.3. For notational simplicity assume that $\boldsymbol{t} = (1, \ldots, 1, t_{r_0+1}, \ldots, t_{\ell_\mathrm{f}})$, and we must show that $A(\boldsymbol{t})$ is independent of $\left\|\mathcal{S}_{i_1}^* X\right\|, \ldots, \left\|\mathcal{S}_{i_{r_0}}^* X\right\|$ (here we used that $\widetilde{\mathcal{L}}_{i_r}$ is a function of $\left\|\mathcal{S}_{i_r}^* X\right\|$). By Lemma 4.4.1, since the matrix $P_{\mathcal{S}_{i_r}} X$ is a function of $\mathcal{S}_{i_r}^* X$, it suffices to show that

$$X + \sum_{r=1}^{r_0} \left(\frac{1}{\|\mathcal{S}_{i_r} X\|} - 1 \right) P_{\mathcal{S}_{i_r}} X \tag{4.97}$$

is independent of $\left\|\mathcal{S}_{i_s}^* X\right\|$ for any $s = 1, \ldots, r_0$. See that equation (4.97) may be written as

$$\begin{aligned} &X + \sum_{r=1}^{r_0} \left(\frac{1}{\|\mathcal{S}_{i_r} X\|} - 1 \right) P_{\mathcal{S}_{i_r}} X \\ &= \left(I - \sum_{r=1}^{r_0} P_{\mathcal{S}_{i_r}} \right) X + \sum_{r=1}^{r_0} \frac{1}{\|\mathcal{S}_{i_r} X\|} P_{\mathcal{S}_{i_r}} X. \end{aligned} \tag{4.98}$$

Since $\frac{1}{\|\mathcal{S}_{i_r} X\|} P_{\mathcal{S}_{i_r}} X = \frac{1}{\|\mathcal{S}_{i_r} X\|} \mathcal{S}_{i_r} \mathcal{S}_{i_r}^* X$ is a deterministic function of $\frac{1}{\|\mathcal{S}_{i_r} X\|} \mathcal{S}_{i_r}^* X$, again by Lemma 4.4.1 it suffices to show that $\left(I - \sum_{r=1}^{r_0} P_{\mathcal{S}_{i_r}}\right) X$ is independendent of $\left\|\mathcal{S}_{i_s}^* X\right\|$. Because the rows of

$$I - \sum_{r=1}^{r_0} P_{\mathcal{S}_{i_r}} = P_{(\mathcal{S}_{i_1}, \ldots, \mathcal{S}_{i_{r_0}})^\perp}$$

are all orthogonal to each of the vectors $\mathcal{S}_{i_1}, \ldots, \mathcal{S}_{i_{r_0}}$, this too follows by Lemma 4.4.1, and we conclude the proof of Lemma 4.4.3. □

Proof of Lemma 4.4.5. Without loss of generality, we show the proof when $r = 1$. $|U_1|$ applications of the intermediate value theorem show that it suffices to bound, letting $U_1 = \{r_1, \ldots, r_s\}$ and $[\ell_\mathrm{f}] = \{1, \ldots, \ell_\mathrm{f}\}$,

$$\sum_{i_1} \max_{i_2, \ldots, i_{p'}} \max_{\substack{\boldsymbol{t} \in [0,1]^{[\ell_\mathrm{f}]}: \\ t_a = 1 \forall a \notin U}} |\partial_{r_1} \cdots \partial_{r_s} g_1(\boldsymbol{i}, \boldsymbol{t})| \prec N^{-s\epsilon_D}. \tag{4.99}$$

A problem with the expression (4.99) is that for each i_1, the values of $t_1, \ldots, t_2$ may be different. To get around this, we use the formula, for $f : \mathbb{R}^{[\ell_\mathrm{f}]} \to \mathbb{R}$,

$$f(\boldsymbol{t}) = f(0) + (\nabla f(0))^* \boldsymbol{t} + \boldsymbol{t}^* (\nabla^2 f(\boldsymbol{t}_1)) \boldsymbol{t}$$

for some $\boldsymbol{t}_1$ between 0 and $\boldsymbol{t}$. Because of this, the left-hand side of equation (4.99) may be written

$$\begin{aligned} &\sum_{i_1} \max_{i_2,\ldots,i_{p'}} |\partial_{r_1} \cdots \partial_{r_s} g_1(\boldsymbol{i}, \mathbf{0})| \\ &+ \sum_{i_1} \max_{i_2,\ldots,i_{p'}} \sum_{a=1}^{\ell_{\mathbf{f}}} O(1) |\partial_a \partial_{r_1} \cdots \partial_{r_s} g_1(\boldsymbol{i}, \mathbf{0})| \\ &+ O\left(\sum_{i_1} \max_{i_2,\ldots,i_{p'}} \max_{\boldsymbol{t} \in [0,1]^{[\ell_{\mathrm{f}}]}} \sum_{a=1}^{\ell_{\mathbf{f}}} \sum_{b=1}^{\ell_{\mathbf{f}}} O(1) |\partial_a \partial_b \partial_{r_1} \cdots \partial_{r_s} g_1(\boldsymbol{i}, \boldsymbol{t})| \right). \end{aligned} \tag{4.100}$$

We claim the following lemma:

Lemma 4.4.6. *For any fixed number $n > 0$ and $r_1, \ldots, r_n \in [\ell_{\mathrm{f}}]$, we have, uniformly in $\boldsymbol{i}$ and $\boldsymbol{t}$,*

$$\partial_{r_1} \cdots \partial_{r_n} g_1(\boldsymbol{i}, \boldsymbol{t}) = \begin{cases} O_{\prec}(N^{-\frac{1}{2} - \epsilon_D}) \left(d_1 |\mathbf{v}_\alpha^* U^* \left(G^{\boldsymbol{t}} - \Pi \right) U \mathbf{e}_{i_1}| + |\mathbf{v}_\alpha^* \mathbf{e}_{i_1}| \right) & n = 1 \\ O_{\prec}(N^{-\frac{n}{2} - 2\epsilon_D}) & n \geq 2. \end{cases}$$

With Lemma 4.4.6, the third line of equation (4.100) may be bounded by

$$N O_{\prec}(N^{-\frac{s}{2} - 1 - 2\epsilon_D}) = O_{\prec}(N^{-s\epsilon_D}).$$

The first line of equation (4.100) may be bounded if $s = 0$ by

$$\sum_{i_1} \max_{i_2,\ldots,i_{p'}} g_1(\boldsymbol{i}, \mathbf{0}) = \sum_{i_1} \mathbf{e}_{i_1}^* \mathbf{M} \mathbf{e}_{i_1} = O_{\prec}(1)$$

as desired, if $s = 1$ by

$$\begin{aligned} \sum_{i_1} \max_{i_2,\ldots,i_{p'}} \partial_{r_1} g_1(\boldsymbol{i}, \mathbf{0}) &= O_{\prec}(N^{-1/2 - \epsilon_D}) \sum_{i_1} \left(d_1 \left| \mathbf{v}_\alpha^* U^* \left(G^{\boldsymbol{t}} - \Pi \right) U \mathbf{e}_{i_1} \right| + |\mathbf{v}_\alpha^* \mathbf{e}_{i_1}| \right) \\ &= O_{\prec}(N^{-\epsilon_D}) \end{aligned}$$

by Cauchy-Schwarz, since perturbing $\boldsymbol{t}$ shows that $\|G - \Pi\| \prec 1$ easily, and if $s \geq 2$ by

$$\sum_{i_1} O_{\prec}(N^{-\frac{s}{2} - 2\epsilon_D}) = O_{\prec}(N^{\frac{s-2}{2} - 2\epsilon_D}) = O_{\prec}(N^{-s\epsilon_D})$$

because $\epsilon_D \leq 1/2$. The second line of equation (4.100) may be bounded in the same way. Thus we conclude the proof of Lemma 4.4.5. $\square$

Proof of Lemma 4.4.6. We demonstrate for the case $d_1 \asymp 1$ only; it is easy to check that if $d_1 \gg 1$, the resolvent entries in the derivatives of $\mathbf{M}$ always provide enough factors of d_1^{-1} to cancel the factors of d_1 in the definition of $\mathbf{M}$. The basic ingredient is the derivative of the resolvent

$$\partial_r G^{\boldsymbol{t}}_{\mathcal{S}_{i_1},\ldots,\mathcal{S}_{i_{\ell_{\mathrm{f}}}}} = \left(\frac{1}{\left\|\mathcal{S}_{i_r}^* X\right\|} - 1\right) G^{\boldsymbol{t}}\left(\mathcal{S}_{i_r}\mathcal{S}_{i_r}^* X + X^*\mathcal{S}_{i_r}\mathcal{S}_{i_r}^*\right) G^{\boldsymbol{t}},$$

which follows from equation (4.40). Note that $\left(\frac{1}{\left\|\mathcal{S}_{i_r}^* X\right\|} - 1\right) = O_\prec(N^{-1/2})$. It is easy to see then that every term in $\partial_{i_r}\langle \mathbf{e}_{i_r}, \mathbf{M}\mathbf{e}_{i_r}\rangle$ has a factor in

$$\left\{\mathcal{S}_{i_r}^* G^{\boldsymbol{t}} U \mathbf{v}_\alpha, \mathcal{S}_{i_r}^* X G^{\boldsymbol{t}} U \mathbf{v}_\alpha\right\}$$

and a factor in

$$\left\{\mathbf{v}_\alpha^* \mathbf{e}_{i_r}, \mathbf{v}_\alpha^* U^*(G^{\boldsymbol{t}} - \Pi) U \mathbf{e}_{i_r}\right\},$$

whereas every second and higher derivative $\partial_{i_{r_1}} \cdots \partial_{i_{r_n}} \langle \mathbf{e}_{i_r}, \mathbf{M}\mathbf{e}_{i_r}\rangle$ has two factors in

$$\left\{\mathcal{S}_{i_r}^* G^{\boldsymbol{t}} U \mathbf{v}_\alpha, \mathcal{S}_{i_r}^* X G^{\boldsymbol{t}} U \mathbf{v}_\alpha\right\}.$$

It is easy to see, by perturbing in $\boldsymbol{t}$, that $\|G^{\boldsymbol{t}}\| = \|G^{\boldsymbol{t}}(\phi_{N,\alpha})\| \prec 1$. To conclude the proof, it only remains to demonstrate that

$$\left|\mathcal{S}_{i_r}^* G^{\boldsymbol{t}} U \mathbf{v}_\alpha\right| + \left|\mathcal{S}_{i_r}^* X G^{\boldsymbol{t}} U \mathbf{v}_\alpha\right| \prec N^{-\epsilon_D}.$$

Now

$$\begin{aligned} \mathbf{x} G^{\boldsymbol{t}} U \mathbf{v}_\alpha &= \mathbf{x}^* G^{\mathbf{0}} U \mathbf{v}_\alpha + \int_{c=0}^1 \mathbf{x}^* \frac{\partial}{\partial c} G^{c\boldsymbol{t}} U \mathbf{v}_\alpha \\ &= \mathbf{x}^* G^{\mathbf{0}} U \mathbf{v}_\alpha + O_\prec\left(N^{-1/2} \max_{c\in[0,1]} \left\|G^{c\boldsymbol{t}}\right\|^2\right) \end{aligned}$$

for any (possibly random) $\mathbf{x}$ (in particular, we care about $\mathbf{x} = \mathcal{S}_{i_r}$ or $X^*\mathcal{S}_{i_r}$) and by Weyl's inequality, $\max_{c\in[0,1]}\|G^{c\boldsymbol{t}}\| \prec 1$, so that it only remains to consider $\boldsymbol{t} = \mathbf{0}$.

It is then a consequence of Lemma 4.4(i) of [KY17] that

$$\mathcal{S}_{i_r}^* X G^{\mathbf{0}} U \mathbf{v}_\alpha = \mathcal{S}_{i_r}^*\left(G^{\mathbf{0}} + I\right) U \mathbf{v}_\alpha.$$

Since $\mathcal{S}_{i_r}^* = \widetilde{S}^* \mathbf{e}_{i_r} + O(N^{-\epsilon_D}) = \mathbf{e}_{i_r} + O(N^{-\epsilon_D})$ and $U\mathbf{v}_\alpha = \sqrt{\frac{1}{d_\alpha}}\mathbf{v}_\alpha + \mathbf{x}$ for a deterministic vector $\mathbf{x} \in \mathbb{R}^{\mathcal{I}_K}$. Then result then follows from Lemma 4.2.7 and equation (4.21). $\mathcal{S}_{i_r}^* G^{\mathbf{0}} U \mathbf{v}_\alpha$ is bounded similarly. □

4.5 Proof of Lemma 4.1.8 under Assumption 1: vanishing third moment

In this section we complete part of the proof of Lemma 4.1.8 (1). The work we do in this section is actually sufficient for Lemma 4.1.8 (1) under the additional assumption that $\mathbb{E}\left(X_{i\mu}^1\right)^3 = 0$. The additional work which is necessary to remove the third moment condition is postponed until the next section. The proof relies on the following two lemmas:

Lemma 4.5.1 (Lemma 7.9 of [KY17])**.** *Given two matrix ensembles X^0 and X^1 with independent entries, let $\rho_{i\mu}^\iota$ be the law of $X_{i\mu}^\iota$ for $\iota = 0, 1$. Then define a law $\rho_{i\mu}^\theta = \theta\rho_{i\mu}^1 + (1-\theta)\rho_{i\mu}^0$ for $\theta \in [0, 1]$, and let X^θ be a new ensemble with $X_{i\mu}^\theta$ distributed according to $\rho_{i\mu}^\theta$. Then for any smooth function F on $\mathbb{R}^{\mathcal{I}}$,*

$$\frac{\partial}{\partial\theta}\mathbb{E}F\left(X^\theta\right) = \sum_{\substack{i \in \mathcal{I}_M \cup \mathcal{I}_K \\ \mu \in \mathcal{I}_N}} \left[\mathbb{E}F\left(X_{(i\mu)}^{\theta, X_{i\mu}^1}\right) - \mathbb{E}F\left(X_{(i\mu)}^{\theta, X_{i\mu}^0}\right)\right] \tag{4.101}$$

where $X_{(i\mu)}^{\theta,y}$ is an ensemble with independent entries whose law in every entry except the entry $i\mu$ is given by the law of X^θ in that entry, but whose law in the $i\mu$ entry is that of a random variable y.

We now particularly let $F : \mathbb{R}^{(\mathcal{I}_K \cup \mathcal{I}_M) \times \mathcal{I}_N} \to \mathbb{C}$ be defined by

$$F(X) = \left(\sum_{i \in \mathcal{I}_\mathcal{M}} \langle \mathbf{e}_i, \mathbf{M}(D - I)\mathbf{e}_i \rangle\right)^p. \tag{4.102}$$

For a function $\widetilde{F} : \mathrm{Mat}_{(\mathcal{I}_K \cup \mathcal{I}_M) \times \mathcal{I}_N} \to \mathbb{C}$, we also define the partial derivative $\partial_{i\mu}$ to the be derivative with respect to the $(i, \mu)^{\mathrm{th}}$ entry of the argument of $\widetilde{F}$.

Lemma 4.5.2. *For any integer $m \geq 4$, we have*

$$N^{-m/2} \sum_{\substack{i \in \mathcal{I}_M \\ \mu \in \mathcal{I}_N}} \left| \mathbb{E} \partial_{i\mu}^m F(X_{i\mu}^\theta) \right| = O_\prec \left(N^{(-1/2-\epsilon_D)p} \right) + C \mathbb{E} F(X^\theta), \tag{4.103}$$

and for any integer $m \geq 3$,

$$N^{-m/2} \sum_{\substack{i \in \mathcal{I}_K \\ \mu \in \mathcal{I}_N}} \left| \mathbb{E} \partial_{i\mu}^m F(X_{i\mu}^\theta) \right| = O_\prec \left(N^{(-1/2-\epsilon)p} \right) + C \mathbb{E} F(X^\theta) \tag{4.104}$$

for some $\epsilon > 0$.

Proof of Lemma 4.1.8. For an entry $i\mu$ of X, we henceforth fix $\theta \in [0, 1]$ and let

$$f_{(i\mu)} : \mathbb{R} \to \mathbb{R}, \quad f_{(i\mu)}(y) = F\left(X_{i\mu}^{\theta, y} \right). \tag{4.105}$$

We will rewrite the difference in equation (4.101), namely $\mathbb{E} f_{(i\mu)}(X_{i\mu}^1) - f_{(i\mu)}(X_{i\mu}^0)$, as the difference

$$\mathbb{E} f_{(i\mu)}(X_{i\mu}^1) - f_{(i\mu)}(X_{i\mu}^0) = \mathbb{E}[f_{(i\mu)}(X_{i\mu}^1) - f_{(i\mu)}(0)] - \mathbb{E}[f_{(i\mu)}(X_{i\mu}^0) - f_{(i\mu)}(0)] \tag{4.106}$$

and then rewrite $\mathbb{E}[f_{(i\mu)}(X_{i\mu}^1) - f_{(i\mu)}(0)]$ (and similarly the other term) using Taylor's formula. This looks like

$$\mathbb{E}(f_{(i\mu)}(X_{i\mu}^1) - f_{(i\mu)}(0)) = \sum_{m=1}^{\bar{m}} K_m(X_{i\mu}^1, X_{i\mu}^\theta) \mathbb{E} f_{(i\mu)}^{(m)}(X_{i\mu}^\theta) + \mathcal{E}_{i\mu}, \tag{4.107}$$

where

$$K_m(\zeta, \xi) := \sum_{q \geq 0} (-1)^q \sum_{n, k_1, \ldots, k_q \geq 1} \mathbf{1}\left(n + k_1 + \cdots + k_q = m \right) \frac{\mathbb{E}\zeta^n}{n!} \prod_{j=1}^q \frac{\mathbb{E}\xi^{k_j}}{k_j!} \tag{4.108}$$

and where

$$|\mathcal{E}|_{i\mu} \leq C_{\overline{m}+1} \max_{|y| \leq \left|X_{i\mu}^\theta\right| + \left|X_{i\mu}^1\right|} f_{(i\mu)}^{(\overline{m}+1)}(y) \sum_{c, d \in \mathbb{N}: c+d = \overline{m}+1} \mathbb{E}(X_{i\mu}^1)^c \mathbb{E}(X_{i\mu}^\theta)^d, \tag{4.109}$$

where $C_{\overline{m}}$ is a constant depending only on $\overline{m}$. What has happened in equation (4.107) is two-fold: we have expanded out $\mathbb{E}(f(X_{i\mu}^1) - f(0))$ as a Taylor polynomial centered at 0, and then we wrote all the derivatives at 0 as themselves Taylor polynomials centered at $X_{i\mu}^{\theta}$. The advantage of this double application of Taylor's formula is that we never, except in the error term, use any ensemble but X^{θ}; i.e., except in the error term, we only evaluate the functions of ensembles whose Green functions we are very confident that we can control, and when we sum over $i\mu$, we will have many expressions which are all in terms of the same matrix X^{θ} so that there will be some simplification that happens in the sum.

Combining equations (4.107) and (4.101) gives us

$$\begin{aligned}\frac{\partial}{\partial\theta}\mathbb{E}F\left(X^{\theta}\right) =& \sum_{m=1}^{\overline{m}} \sum_{\substack{i\in\mathcal{I}_M\cup\mathcal{I}_K\\ \mu\in\mathcal{I}_N}} \left(K_m(X_{i\mu}^1, X_{i\mu}^{\theta}) - K_m(X_{i\mu}^0, X_{i\mu}^{\theta})\right)\mathbb{E}f_{(i\mu)}^{(m)}(X_{i\mu}^{\theta}) + \mathcal{E}\\ =& \sum_{m=4}^{\overline{m}} \sum_{\substack{i\in\mathcal{I}_M\\ \mu\in\mathcal{I}_N}} \left(K_m(X_{i\mu}^1, X_{i\mu}^{\theta}) - K_m(X_{i\mu}^0, X_{i\mu}^{\theta})\right)\mathbb{E}f_{(i\mu)}^{(m)}(X_{i\mu}^{\theta})\\ &+ \sum_{m=3}^{\overline{m}} \sum_{\substack{i\in\mathcal{I}_K\\ \mu\in\mathcal{I}_N}} \left(K_m(X_{i\mu}^1, X_{i\mu}^{\theta}) - K_m(X_{i\mu}^0, X_{i\mu}^{\theta})\right)\mathbb{E}f_{(i\mu)}^{(m)}(X_{i\mu}^{\theta}) + \mathcal{E}.\end{aligned} \tag{4.110}$$

Here we have split the sum in i into according to whether i belongs to $\mathcal{I}_M$ or $\mathcal{I}_K$ and then used that

$$\left(K_m(X_{i\mu}^1, X_{i\mu}^{\theta}) - K_m(X_{i\mu}^0, X_{i\mu}^{\theta})\right) = 0$$

for $m = 1, 2$, and additionally for $m = 3$ if $i \in \mathcal{I}_M$. The first term above is treated by Proposition 4.6.0.1 in Section 4.6. Setting $\overline{m} = 6p$ and using the bound $K_m(X_{i\mu}^1, X_{i\mu}^0) \prec N^{-m/2}$ for any fixed positive integer m, the result now follows by Grönwall's inequality, Lemma 4.5.2 and Proposition 4.5.2.1, which is below.

Proposition 4.5.2.1. *If $\overline{m} = 6p$, then $\mathcal{E} = \sum_{i\mu} \mathcal{E}_{i\mu} \prec N^{-p}$.*

□

Proof of propsition 4.5.2.1. Changing an entry of X by $O_{\prec}(N^{-1/2})$ changes the operator norm of G by $O_\prec(N^{-1/2})$, so that such a perturbed G still has operator norm $O_\prec(1)$. Thus

$$\max_{|y|\le|X^\theta_{i\mu}|+|X^1_{i\mu}|} f^{(\bar{m}+1)}_{(i\mu)}(y) \prec 1$$

uniformly in i,μ. Then $\mathbb{E}|X^\iota_{i\mu}|^p \le C_p N^{-p/2}$ allows us to conclude. □

Remark 4.5.3. The error can just as easily be bounded under Assumption 2 with bounded support condition $|X_{i\mu}| \le N^{-\epsilon}$, replacing $6p$ with, say, $12p/\epsilon$.

4.5.1 Proof of Lemma 4.5.2

Write

$$F := F(X) = h(X)^p, \quad h := h(X) = \sum_{j\in\mathcal{I}_M} \langle \mathbf{e}_j, (\mathbf{M}_1+\mathbf{M}_2+\mathbf{M}_3)\mathbf{e}_j\rangle (D_{jj}-1).$$

We state two propositions:

Proposition 4.5.3.1. *Under Assumption 1 and for $i\in\mathcal{I}_M$, the following bounds hold:*

1. *We have $\partial_{i\mu} h \prec N^{-1/2-\epsilon_D}$.*
2. *For $m\ge 2$, we have $\partial^m_{i\mu} h \prec N^{-2\epsilon_D}$.*
3. *For $m \ge 2$, we have $\partial^m_{i\mu} h = h^{(\mathcal{I}_M)}_{m,i\mu} + h^{(\mathcal{I}_N)}_{m,i\mu}$, where $\sum_{i\in\mathcal{I}_M} \left|h^{(\mathcal{I}_M)}_{m,i\mu}\right|^2 \prec N^{-2\epsilon_D}$ and $\sum_\mu \left|h^{(\mathcal{I}_N)}_{m,i\mu}\right|^2 \prec N^{-2\epsilon_D}$.*

Proposition 4.5.3.2. *Under Assumption 1 and for $i\in\mathcal{I}_K$, we have $\partial_{i\mu}h \prec N^{-1/2}$ and $\partial^m_{i\mu} h \prec N^{-\epsilon_D}$ for $m\ge 2$.*

Proof of Lemma 4.5.2. First we demonstrate equation (4.103). Our goal is to show

$$N^{-m/2}\sum_{i\mu}\left|\mathbb{E}\partial^m_{i\mu}\left(h^p\right)\right| = O_\prec(N^{(-1/2-\epsilon_D)p}). \tag{4.111}$$

Since m and p are fixed integers, we see that the left-hand side of equation (4.111) may be written as a sum of boundedly many terms of the form

$$N^{-m/2}\sum_{i\mu}\left|\mathbb{E}\left(\partial_{i\mu}^{m_1}h\right)\cdots\left(\partial_{i\mu}^{m_q}h\right)\right|\left|h^{p-q}\right|, \tag{4.112}$$

where $m_1+\cdots+m_q=m$ and $m_1\leq\cdots\leq m_q$. Now let q' be the number of $m_1,\ldots,m_q$ which are equal to 1. If $q-q'=0$, then we have by Young's inequality

$$(4.112)\leq CN^{-m/2}\sum_{i\mu}\left(\left|\mathbb{E}\left(\partial_{i\mu}^{m_1}h\right)\cdots\left(\partial_{i\mu}^{m_q}h\right)\right|^{p/q}+\mathbb{E}|h^p|\right)\leq N^{p(-1/2-\epsilon_D)}+\mathbb{E}|h^p|$$

since $m\geq 4$. If $q-q'=1$,

$$\begin{aligned}(4.112)&\prec N^{(-1/2-\epsilon_D)(q-1)}N^{-m/2}|h|^{p-q}\sum_{i\mu}\partial_{i\mu}^{m_q}h\\&\prec N^{(-1/2-\epsilon_D)(q-1)}N^{-m/2}|h|^{p-q}\left(\sum_{\mu}\sum_{i}h_{m,i\mu}^{(\mathcal{I}_M)}+\sum_{i}\sum_{\mu}h_{m,i\mu}^{(\mathcal{I}_N)}\right)\\&\prec N^{(-1/2-\epsilon_D)(q-1)}N^{-1/2-\epsilon_D}|h|^{p-q},\end{aligned} \tag{4.113}$$

where the last line follows from Proposition 4.5.3.1 and Cauchy-Schwarz, since $m\geq 4$. We conclude by Young's inequality again.

We have now reduced to $q-q'\geq 2$. Now we have equation (4.112) may be bounded by

$$\begin{aligned}&\prec N^{(-1/2-\epsilon_D)q'}N^{-m/2}|h|^{p-q}\sum_{i\mu}\partial_{i\mu}^{m_{q'+1}}h\cdots\partial_{i\mu}^{m_q}h\\&\prec N^{(-1/2-\epsilon_D)q'}N^{-m/2}|h|^{p-q}(N^{-2\epsilon_D})^{q-q'-2}\sum_{i\mu}\partial_{i\mu}^{m_{q-1}}h\,\partial_{i\mu}^{m_q}h\\&\prec N^{(-1/2-\epsilon_D)q'}N^{-m/2}(4.112)(N^{-2\epsilon_D})^{q-q'-2}\left(\sum_{i\mu}(\partial_{i\mu}^{m_{q-1}}h)^2\sum_{i\mu}(\partial_{i\mu}^{m_q}h)^2\right)^{\frac{1}{2}}.\end{aligned}$$

Note that

$$\left(\sum_{i\mu}(\partial_{i\mu}^{m_q}h)^2\right)^{\frac{1}{2}}\prec\left(\sum_{i\mu}\left|h_{m,i\mu}^{(\mathcal{I}_M)}\right|^2+\sum_{i\mu}\left|h_{m,i\mu}^{(\mathcal{I}_N)}\right|^2+2\sum_{i\mu}h_{m,i\mu}^{(\mathcal{I}_M)}h_{m,i\mu}^{(\mathcal{I}_N)}\right)^{\frac{1}{2}}\prec N^{1/2-\epsilon_D}, \tag{4.114}$$

where we applied Proposition 4.5.3.1 and used Cauchy-Schwarz again to bound the cross term. Thus, we may bound equation (4.112) by

$$\begin{aligned}
&\prec N^{(-1/2-\epsilon_D)q'} N^{-m/2} |h|^{p-q} (N^{-2\epsilon_D})^{q-q'-2} N^{1-2\epsilon_D} \\
&\prec N^{(-1/2-\epsilon_D)q'} N^{-m/2} |h|^{p-q} N^{-\epsilon_D(q-q')} N \\
&= N^{-(1/2)q'} N^{-m/2+1} |h|^{p-q} N^{-\epsilon_D q} \\
&\leq N^{-(1/2)q} N^{-\epsilon_D q} |h|^{p-q},
\end{aligned} \tag{4.115}$$

where the second line follows from the fact that $q - q' \geq 2$ and the last line follows $m \geq q - q' + 2$, which is a consequence of $q - q' \geq 2$ (ie, at least 2 of $m_{q'+1}, \ldots, m_q$ are at least 2). We conclude again by Young's inequality.

Now we demonstrate equation (4.104). Just as in the proof of equation (4.103), it suffices to bound

$$(4.112) \prec N^{(-1/2-\epsilon)p}. \tag{4.116}$$

We will omit all the applications of Young's inequality, which are done the same here as they were previously. Let again q' be the number of $m_1, \ldots, m_q$ which are equal to 1. Consider first $q \geq 3$. We have then that equation (4.116) may be bounded by, using that $|\mathcal{I}_K| = O(1)$, and that $m \geq q' + 2(q - q')$,

$$N^{-\frac{1}{2}q'} N^{-m/2+1} \prec N^{-q+1}. \tag{4.117}$$

Now $-q + 1 \leq \left(-\frac{1}{2} - \epsilon\right) q$ for $\epsilon := \min\{\epsilon_D, 1/6\}$ as long as $q \geq 3$, so that

$$(4.112) \prec N^{-\left(-\frac{1}{2}-\epsilon\right)q} |h|^{p-q}.$$

So it only remains to consider the cases $q = 1$ and $q = 2$. If $q = 1$, then by Proposition 4.5.3.2,

$$(4.112) \prec N^{-3/2} N N^{-\epsilon_D} |h|^{p-q} = N^{-1/2-\epsilon_D} |h|^{p-q} \tag{4.118}$$

as desired. Finally, if $q = 2$, then either exactly one of m_1, m_2 is 1, in which case equation (4.112) may be bounded by

$$N^{-3/2} N N^{-1/2} N^{-\epsilon_D} |h|^{p-q} = N^{-1-\epsilon_D} |h|^{p-q} = N^{2\left(-\frac{1}{2}-\frac{\epsilon_D}{2}\right)} |h|^{p-q} \tag{4.119}$$

as desired, or $m_1 \geq 2$ and $m_2 \geq 2$, so that $m \geq 4$, and equation (4.112) may be bounded by

$$N^{-2} N N^{-2\epsilon_D} = N^{2\left(-\frac{1}{2}-\epsilon_D\right)} |h|^{p-q} \tag{4.120}$$

as desired. This completes the bound of equation (4.104) and Lemma 4.5.2. □

It remains to prove Propositions 4.5.3.1 and 4.5.3.2. First, we will find a formula for the derivatives of $\mathbf{M}_{jj}$. Suppose $m \geq 1$. For convenience, we define

$$\xi = G - \Pi. \tag{4.121}$$

In the following discussion, we fix $z = z_\alpha = \phi_\alpha + i0^+$, as before. Iterating the derivative formula equation (4.40) gives

$$\partial_{i\mu}^m G = \sum_{(\mathbf{i},\mathbf{j}) \in \mathcal{B}} c_{\mathbf{ij}} G \mathbf{e}_{i_1} \mathbf{e}_{j_1}^* G \cdots \mathbf{e}_{i_m} \mathbf{e}_{j_m}^* G, \tag{4.122}$$

where

$$\mathcal{B} = \{((i_1, \ldots, i_m), (j_1, \ldots, j_m)) : i_l, j_l \in \{i, \mu\}, i_l \neq j_l\} \tag{4.123}$$

and the $c_{\mathbf{ij}}$ are constants.

Observe first that $\partial_{i\mu}^m (\mathbf{M}_1)_{jj} = 0$. For $\mathbf{M}_2$, using the above derivative formula directly yields

$$\begin{aligned} \partial_{i\mu}^m (\mathbf{M}_2)_{jj} &= O(d_1^{1/2})\, \mathbf{e}_j^* \partial_{i\mu}^m (U^* \xi U) \mathbf{v}_\alpha \mathbf{v}_\alpha^* \mathbf{e}_j \\ &= O(d_1^{1/2}) (U\mathbf{e}_j)^* (\partial_{i\mu}^m \xi) U \mathbf{v}_\alpha \mathbf{v}_\alpha^* \mathbf{e}_j \\ &= O(d_1^{1/2}) \cdot \sum_{(\mathbf{s},\mathbf{t}) \in \mathcal{B}} c_{\mathbf{st}} \langle U\mathbf{e}_j, G\mathbf{e}_{s_1} \rangle \langle \mathbf{e}_{t_1}, G\mathbf{e}_{s_2} \rangle \cdots \langle \mathbf{e}_{t_m}, GU\mathbf{v}_\alpha \rangle \langle \mathbf{v}_\alpha, \mathbf{e}_j \rangle. \end{aligned} \tag{4.124}$$

For $\mathbf{M}_3$, we obtain

$$\begin{aligned}
\partial_{i\mu}^m(\mathbf{M}_3)_{jj} &= O(d_1)\sum_{k=0}^{m}\binom{m}{k}\mathbf{e}_j^*\partial_{i\mu}^k(U^*\xi^*U)\mathbf{v}_\alpha\mathbf{v}_\alpha^*\partial_{i\mu}^{m-k}(U^*\xi U)\mathbf{e}_j \\
&= O(d_1)\sum_{k=0}^{m}\binom{m}{k}(U\mathbf{e}_j)^*(\partial_{i\mu}^k\xi^*)U\mathbf{v}_\alpha(U\mathbf{v}_\alpha)^*(\partial_{i\mu}^{m-k}\xi)U\mathbf{e}_j \\
&= O(d_1)\cdot\sum_{(\mathbf{s},\mathbf{t})\in\mathcal{B}} c_{\mathbf{st}}\Bigg(\langle U\mathbf{e}_j,\xi^*U\mathbf{v}_\alpha\rangle\langle U\mathbf{v}_\alpha,G\mathbf{e}_{s_1}\rangle\langle\mathbf{e}_{t_1},G\mathbf{e}_{s_2}\rangle\cdots\langle\mathbf{e}_{t_m},GU\mathbf{e}_j\rangle \\
&\quad+\sum_{k=1}^{m-1}\binom{m}{k}\langle U\mathbf{e}_j,G^*\mathbf{e}_{s_1}\rangle\langle\mathbf{e}_{t_1},G^*\mathbf{e}_{s_2}\rangle\cdots\langle\mathbf{e}_{t_k},G^*U\mathbf{v}_\alpha\rangle\langle U\mathbf{v}_\alpha,G\mathbf{e}_{s_{k+1}}\rangle\cdots\langle\mathbf{e}_{t_m},GU\mathbf{e}_j\rangle \\
&\quad+\langle U\mathbf{e}_j,G^*\mathbf{e}_{s_1}\rangle\langle\mathbf{e}_{t_1},G^*\mathbf{e}_{s_2}\rangle\cdots\langle\mathbf{e}_{t_m},G^*U\mathbf{v}_\alpha\rangle\langle U\mathbf{v}_\alpha,\xi U\mathbf{e}_j\rangle\Bigg).
\end{aligned} \tag{4.125}$$

We now use the fact that $G=\xi+\Pi$ to rewrite all the terms above that contain G or G^* in terms of ξ or Π. For instance,

$$\langle\mathbf{e}_s,G\mathbf{v}\rangle=\langle\mathbf{e}_s,\xi\mathbf{v}\rangle+\langle\mathbf{e}_s,\Pi\mathbf{v}\rangle$$

so we obtain two terms, one in terms of ξ and one in terms of Π. We are now ready to prove Proposition 4.5.3.1.

Proof of Proposition 4.5.3.1. First let us bound the first derivative. We recall

$$h=\sum_{j\in\mathcal{I}_M}\langle\mathbf{e}_j,(\mathbf{M}_1+\mathbf{M}_2+\mathbf{M}_3)\mathbf{e}_j\rangle(D_{jj}-1). \tag{4.126}$$

We have then that

$$\begin{aligned}
\partial_{i\mu}h &= \sum_{j\in\mathcal{I}_M}\langle\mathbf{e}_j,(\partial_{i\mu}\mathbf{M}_1+\partial_{i\mu}\mathbf{M}_2+\partial_{i\mu}\mathbf{M}_3)\mathbf{e}_j\rangle(D_{jj}-1) \\
&\quad+\sum_{j\in\mathcal{I}_M}\langle\mathbf{e}_j,(\mathbf{M}_1+\mathbf{M}_2+\mathbf{M}_3)\mathbf{e}_j\rangle\partial_{i\mu}(D_{jj}-1) \\
&=:\mathbf{T}_1+\mathbf{T}_2.
\end{aligned} \tag{4.127}$$

Note that $\mathbf{T}_1$ and $\mathbf{T}_2$ implicitly depend on i, μ. We calculate

$$\begin{aligned}\partial_{i\mu} D_{jj} = \partial_{i\mu} \left\| \mathbf{e}_j^* \widetilde{S} X \right\|^2 = \partial_{i\mu} (\widetilde{S} X X^* \widetilde{S}^*)_{jj} = (\widetilde{S} \partial_{i\mu} (X X^*) \widetilde{S}^*)_{jj} \\ = (\widetilde{S} (X \mathbf{e}_\mu \mathbf{e}_i^* + \mathbf{e}_i \mathbf{e}_\mu^* X^*) \widetilde{S}^*)_{jj} = 2 \widetilde{S}_{ji} (\widetilde{S} X)_{j\mu}.\end{aligned} \tag{4.128}$$

For $i \in \mathcal{I}_M$, we have that $\widetilde{S}_{ij} = (1 - o(1))\delta_{ij}$. Thus, we see that

$$\mathbf{T}_2 = 2\langle \mathbf{e}_i, (\mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3)\mathbf{e}_i \rangle \widetilde{S}_{ii} (\widetilde{S} X)_{i\mu}. \tag{4.129}$$

Since $\left\| \mathbf{e}_\alpha^* \widetilde{S} \right\| = 1$, we have $(\widetilde{S} X)_{i\mu} \prec N^{-1/2}$. By Theorem 4.2.7 and equation (4.21), we see that

$$\begin{aligned}\langle \mathbf{e}_i, \mathbf{M} \mathbf{e}_i \rangle &= \langle \mathbf{e}_i, \mathbf{v} \rangle \langle \mathbf{v}, \mathbf{e}_i \rangle + O(d_1^{1/2}) \langle \mathbf{e}_i, R^* U^* \xi U \mathbf{v} \rangle \langle \mathbf{v}, \mathbf{e}_i \rangle + O(d_1) \langle \mathbf{e}_i, R^* U^* \xi^* U \mathbf{v} \rangle \langle \mathbf{v}, U^* \xi U R \mathbf{e}_i \rangle \\ &\prec N^{-2\epsilon_D} + d_1^{1/2} (d_1^{-1} N^{-1/2}) d_1^{-1/2} N^{-\epsilon_D} + d_1 (d_1^{-1} N^{-1/2})^2 \\ &\prec N^{-\epsilon_D},\end{aligned} \tag{4.130}$$

so that $\mathbf{T}_2 \prec N^{-1/2-\epsilon_D}$, as desired.

We now bound $\mathbf{T}_1$. Since $D_{jj} - 1 = O_\prec(N^{-1/2})$, it is sufficient to show that

$$\sum_{j \in \mathcal{I}_M} |\langle \mathbf{e}_j, \partial_{i\mu} \mathbf{M} \mathbf{e}_j \rangle| = O_\prec(N^{-\epsilon_D}). \tag{4.131}$$

First, using equations (4.124) and (4.125), we see that $\langle \mathbf{e}_j, \partial_{i\mu} \mathbf{M} \mathbf{e}_j \rangle$ is a sum of boundedly many terms, each of which has one of the following three forms:

$$O(d_1^{1/2}) \langle U \mathbf{e}_j, \mathbf{W}_1 \mathbf{e}_s \rangle \langle \mathbf{e}_t, \mathbf{W}_2 U \mathbf{v}_\alpha \rangle \langle \mathbf{v}_\alpha, \mathbf{e}_j \rangle \tag{4.132}$$

$$O(d_1) \langle U \mathbf{e}_j, \mathbf{W}_1 U \mathbf{v}_\alpha \rangle \langle U \mathbf{v}_\alpha, \mathbf{W}_2 \mathbf{e}_s \rangle \langle \mathbf{e}_t, \mathbf{W}_3 U \mathbf{e}_j \rangle \tag{4.133}$$

$$O(d_1) \langle U \mathbf{e}_j, \mathbf{W}_1 \mathbf{e}_s \rangle \langle \mathbf{e}_t, \mathbf{W}_2 U \mathbf{v}_\alpha \rangle \langle U \mathbf{v}_\alpha, \mathbf{W}_3 U \mathbf{e}_j \rangle, \tag{4.134}$$

where $\mathbf{W}_k \in \{\xi, \xi^*, \Pi, \Pi^*\}$ and $s, t \in \{i, \mu\}$ with $s \neq t$. The bounds are similar for all three forms, so we will consider equation (4.133). The idea is to apply Cauchy-Schwarz to the outer two factors in order to bound the sum in j, and to obtain a factor of $N^{-\epsilon_D}$ from the middle term. We must also find a factor of d_1^{-1} to offset the $O(d_1)$ in front.

Fix two deterministic vectors $\mathbf{x}, \mathbf{y}$. Then by the local law (Theorem 4.2.7), we have

$$\langle \mathbf{x}, \xi(z_\alpha)\mathbf{y} \rangle \prec d_1^{-1} N^{-1/2} \|\mathbf{x}\| \|\mathbf{y}\|. \tag{4.135}$$

We also have

$$\langle \mathbf{x}, \Pi(z_\alpha)\mathbf{y} \rangle \prec \|\mathbf{x}\| \|\mathbf{y}\|, \tag{4.136}$$

using the definition of Π. Moreover, if $\mathbf{y} \in \mathbb{R}^{\mathcal{I}_N}$, we have

$$\langle \mathbf{x}, \Pi(z_\alpha)\mathbf{y} \rangle = m(z_\alpha)\langle \mathbf{x}, \mathbf{y} \rangle \prec d_1^{-1} \|\mathbf{x}\| \|\mathbf{y}\| \tag{4.137}$$

(and similarly for $\mathbf{x} \in \mathbb{R}^{\mathcal{I}_N}$, using Π^*).

Suppose first that $s = i, t = \mu$. Then using equations (4.135) and (4.136) and the delocalization bound equation (4.21), we see that $\langle U\mathbf{v}_\alpha, \mathbf{W}_2\mathbf{e}_i \rangle \prec N^{-\epsilon_D}$, so that by Cauchy-Schwarz,

$$\begin{aligned} \sum_j O(d_1) \, |\langle U\mathbf{e}_j, \mathbf{W}_1 U\mathbf{v}_\alpha \rangle \langle U\mathbf{v}_\alpha, \mathbf{W}_2\mathbf{e}_i \rangle \langle \mathbf{e}_\mu, \mathbf{W}_3 U\mathbf{e}_j \rangle| \\ \prec d_1 N^{-\epsilon_D} \|U^*\mathbf{W}_1 U\mathbf{v}_\alpha\| \|(\mathbf{W}_3 U)^*\mathbf{e}_\mu\|. \end{aligned} \tag{4.138}$$

Now since $\mathbf{e}_\mu \in \mathbb{R}^{\mathcal{I}_N}$, we can use either equation (4.137) or (4.135), depending on whether $\mathbf{W}_3$ is one of Π, Π^* or ξ, ξ^*, to obtain

$$\|(\mathbf{W}_3 U)^*\mathbf{e}_\mu\| = \sup_{\|\mathbf{x}\| \le 1} |\langle \mathbf{x}, (\mathbf{W}_3 U)^*\mathbf{e}_\mu \rangle| \le \sup_{\|\mathbf{x}\| \le 1} d_1^{-1} \|U\mathbf{x}\| \|\mathbf{e}_\mu\|. \tag{4.139}$$

It follows that

$$\begin{aligned} \sum_j O(d_1) \, |\langle U\mathbf{e}_j, \mathbf{W}_1 U\mathbf{v}_\alpha \rangle \langle U\mathbf{v}_\alpha, \mathbf{W}_2\mathbf{e}_i \rangle \langle \mathbf{e}_\mu, \mathbf{W}_3 U\mathbf{e}_j \rangle| \\ \prec N^{-\epsilon_D} \|U^*\mathbf{W}_1 U\mathbf{v}_\alpha\| \|U\|_{\text{op}} \prec N^{-\epsilon_D}. \end{aligned} \tag{4.140}$$

Now suppose that $s = \mu, t = i$. Since $\mathbf{e}_\mu \in \mathbb{R}^{\mathcal{I}_N}$, if $\mathbf{W}_2 \in \{\Pi, \Pi^*\}$ we can use equation (4.137) along with equation (4.21) to obtain $\langle U\mathbf{v}_\alpha, \mathbf{W}_3\mathbf{e}_\mu \rangle \prec d_1^{-1} N^{-\epsilon_D}$. If $\mathbf{W}_2 \in \{\xi, \xi^*\}$ instead, we can use equation (4.135) to obtain $\langle U\mathbf{v}_\alpha, \mathbf{W}_3\mathbf{e}_\mu \rangle \prec d_1^{-1} N^{-1/2}$. We thus have

$$\begin{aligned} \sum_j O(d_1)\,|\langle U^*\mathbf{e}_j, \mathbf{W}_1 U^*\mathbf{v}_\alpha\rangle\langle U\mathbf{v}_\alpha, \mathbf{W}_2\mathbf{e}_\mu\rangle\langle \mathbf{e}_i, \mathbf{W}_3 U\mathbf{e}_j\rangle| \\ \prec N^{-\epsilon_D}\|U\mathbf{W}_1U^*\mathbf{v}_\alpha\|\|(\mathbf{W}_3U)^*\mathbf{e}_i\| \\ \prec N^{-\epsilon_D}. \end{aligned} \tag{4.141}$$

Now let us bound the higher derivatives, $m \geq 2$. We note that

$$\partial_{i\mu}^2(D_{jj}-1) = 2\partial_{i\mu}\widetilde{S}_{ji}(\widetilde{S}X)_{j\mu} = 2\widetilde{S}_{ji}^2 \tag{4.142}$$

and $\partial_{i\mu}^m(D_{jj}-1) = 0$ for $m \geq 3$. We first show part (2). We have

$$\begin{aligned} \partial_{i\mu}^m h &= \sum_j \langle \mathbf{e}_j, \partial_{i\mu}^m \mathbf{M}\mathbf{e}_j\rangle (D_{jj}-1) \\ &\quad + 2\sum_j \langle \mathbf{e}_j, \partial_{i\mu}^{m-1}\mathbf{M}\mathbf{e}_j\rangle(\widetilde{S}X)_{j\mu}\widetilde{S}_{ji} + 2\sum_j \langle \mathbf{e}_j, \partial_{i\mu}^{m-2}\mathbf{M}\mathbf{e}_j\rangle \widetilde{S}_{ji}^2 \\ &=: \mathbf{T}_1^m + \mathbf{T}_2^m + \mathbf{T}_3^m \\ &\prec \sum_j |\langle \mathbf{e}_j, \partial_{i\mu}^m\mathbf{M}\mathbf{e}_j\rangle|\, O_\prec(N^{-1/2}) \\ &\quad + |\langle \mathbf{e}_i, \partial_{i\mu}^{m-1}\mathbf{M}\mathbf{e}_i\rangle|\, O_\prec(N^{-1/2}) + |\langle \mathbf{e}_i, \partial_{i\mu}^{m-2}\mathbf{M}\mathbf{e}_i\rangle|. \end{aligned} \tag{4.143}$$

We can bound $\mathbf{T}_1^m$ in exactly the same way as we did for $\mathbf{T}_1$ above, using equations (4.124) and (4.125). We also see easily, similarly to the computation for $m = 1$, that $\mathbf{T}_2^m = O_\prec(N^{-1/2-\epsilon_D})$ and $\mathbf{T}_3^m = O_\prec(N^{-2\epsilon_D})$.

We now show part (3). First note that

$$\sum_i (\mathbf{T}_1^m)^2 \prec N(N^{-1/2-\epsilon_D})^2 \prec N^{-2\epsilon_D}, \tag{4.144}$$

and similarly for $\mathbf{T}_2^m$. It thus suffices to consider $\mathbf{T}_3^m$.

Using equations (4.124) and (4.125) with $j = i$, we see that $\mathbf{T}_3^m$ consists of boundedly many terms, each of which is bounded above (in the sense of $\prec$) by one of the following four expressions:

$$O(d_1^{1/2})|\langle \mathbf{e}_t, \mathbf{W}_1 U \mathbf{v}_\alpha\rangle \langle \mathbf{v}_\alpha, \mathbf{e}_i\rangle| \tag{4.145}$$

$$O(d_1)|\langle \mathbf{x}, \mathbf{W}_1 U \mathbf{v}_\alpha\rangle \langle U\mathbf{v}_\alpha, \mathbf{W}_2\mathbf{e}_s\rangle \langle \mathbf{e}_t, \mathbf{W}_3\mathbf{e}_u\rangle \langle \mathbf{e}_v, \mathbf{W}_4\mathbf{y}\rangle| \tag{4.146}$$

$$O(d_1)|\langle \mathbf{x}, \mathbf{W}_1 \mathbf{e}_s\rangle \langle \mathbf{e}_t, \mathbf{W}_2 U\mathbf{v}_\alpha\rangle \langle U\mathbf{v}_\alpha, \mathbf{W}_3\mathbf{e}_u\rangle \langle \mathbf{e}_v, \mathbf{W}_4\mathbf{y}\rangle| \tag{4.147}$$

$$O(d_1)|\langle \mathbf{x}, \mathbf{W}_1 \mathbf{e}_s\rangle \langle \mathbf{e}_t, \mathbf{W}_2 \mathbf{e}_u\rangle \langle \mathbf{e}_v, \mathbf{W}_3 U\mathbf{v}_\alpha\rangle \langle U\mathbf{v}_\alpha, \mathbf{W}_4\mathbf{y}\rangle| \tag{4.148}$$

where $\mathbf{W}_k \in \{\xi, \xi^*, \Pi, \Pi^*\}$, $s, t, u, v \in \{i, \mu\}$, $s \neq t$, $u \neq v$, and $\mathbf{x}, \mathbf{y} \in \{\mathbf{e}_i, \mathbf{e}_\mu, U\mathbf{e}_i, U^*\mathbf{e}_i\}$.

The desired bound for terms of type (4.145) follows immediately from equation (4.21). The bounds for the other three types are somewhat similar to one another, so we will show the bound only for (4.147).

There are two cases: if $t = \mu$, then $\mathbf{e}_\mu \in \mathbb{R}^{\mathcal{I}_N}$ so we can use equations (4.135), (4.137), and (4.21) to bound $\langle \mathbf{e}_\mu, \mathbf{W}_2 U\mathbf{v}_\alpha\rangle \prec d_1^{-1} N^{-\epsilon_D}$. We then square sum over the third factor, either in i or μ, according to whether u is i or μ. We bound the other two factors trivially by $O_\prec(1)$. This gives

$$\begin{aligned}\sum_u O(d_1^2)|\langle \mathbf{x}, \mathbf{W}_1 \mathbf{e}_s\rangle \langle \mathbf{e}_t, \mathbf{W}_2 U\mathbf{v}_\alpha\rangle \langle U\mathbf{v}_\alpha, \mathbf{W}_3\mathbf{e}_u\rangle \langle \mathbf{e}_v, \mathbf{W}_4\mathbf{y}\rangle|^2 &\prec N^{-2\epsilon_D}\|(\mathbf{W}_3)^* U\mathbf{v}_\alpha\|^2 \\ &\prec N^{-2\epsilon_D}.\end{aligned} \tag{4.149}$$

If $t = i$, then we instead obtain $O_\prec(N^{-\epsilon_D})$ from the second factor, obtain $O(d_1^{-1})$ from either the third or the fourth factor (whichever one contains $\mathbf{e}_\mu$: here we use equations (4.135) and (4.137)) and square sum over the first factor in μ. This gives a bound of $O_\prec(N^{-2\epsilon_D})$, and completes the proof of part (3).

□

Proof of Proposition 4.5.3.2. We have

$$\partial_{i\mu} h = \sum_{j \in \mathcal{I}_M} (\partial_{i\mu}\mathbf{M}_{jj})(D_{jj} - 1) + \sum_{j \in \mathcal{I}_M} \mathbf{M}_{jj}(\partial_{i\mu}(D_{jj} - 1)). \tag{4.150}$$

By our previous manipulations, it is easy to see that $\sum_{j\in\mathcal{I}_M}|\partial_{i\mu}\mathbf{M}_{jj}|$ and $\sum_{j\in\mathcal{I}_M}|\mathbf{M}_{jj}|$ are $O_\prec(1)$. Also, $D_{jj}-1$ and $\partial_{i\mu}(D_{jj}-1)$ are $O_\prec(N^{-1/2})$. This completes what was was to be proven for $\partial_{i\mu}h$.

Similar, we have seen earlier in the section that $\sum_{j\in\mathcal{I}_M}\left|\partial_{i\mu}^m\mathbf{M}_{jj}\right| = O_\prec(1)$, and because $i\in\mathcal{I}_K$, we also have by equations (4.128) and (4.142) that $\partial_{i\mu}^m(D_{jj}-1) = O_\prec(N^{-\epsilon_D})$. This completes the proof of Proposition 4.5.3.2.

□

4.6 General Third Moment Condition

In order to establish universality for random matrix ensembles with general third moments, recalling our work in the previous two sections, it only remains to establish:

Proposition 4.6.0.1. *Fix $a\in\{1,2,3\}$ and let $h=\sum_j(\mathbf{M}_a)_{jj}(D_{jj}-1)$. Then,*

$$\mathbb{E}\sum_{\substack{i\in\mathcal{I}_N\\ \mu\in\mathcal{I}_M}} K_{3,i\mu}^{\theta}\left(\partial_{i\mu}^{s_1}h\right)\cdots\left(\partial_{i\mu}^{s_p}h\right) \prec N^{-p(1/2-\epsilon_D)}+\mathbb{E}|h|^p, \tag{4.151}$$

where $\sum_{a=1}^q s_a=3$, $s_a\geq 0$, and $K_{3,i\mu}^\theta := K_3(X_{i\mu}^1, X_{i\mu}^\theta)-K_3(X_{i\mu}^0, X_{i\mu}^\theta)$.

This is very close to what we already bounded in Section 4.5, except that we must extract an extra $N^{-1/2}$. Moreover, we have defined h in terms of a single term $\mathbf{M}_a$ rather than in terms of $\mathbf{M}$; this change is not consequential since we allows $a=1,2$ or 3 in the proof; it is merely to solve a small notational difficulty.

In this section, the fixed value of $\alpha\in\{1,\dots,K\}$ will not be mentioned at all; we abbreviate $\mathbf{v}:=\mathbf{v}_\alpha$ be the αth eigenvector of $B^\#$, which appears in the definition of $\mathbf{M}$. We will instead use α as an index for general summations.

We will only treat the case of $d_1\asymp 1$. The only difference between this and the case of $d_1\to\infty$ is that derivatives of h have powers of d_1^{-1} which cancel out the powers of d_1 in the definition of $\mathbf{M}$, but this is very easy to see and a nuisance to keep track of.

Proof of Proposition 4.6.0.1. We replace, as in Section 4.5, each derivative $\partial_{i\mu}^{s_1} h$ with a sum of boundedly many terms of the following form:

$$\sum_{j \in \mathcal{I}_M} \partial_{i\mu}^{r_1} \langle \mathbf{e}_j, (\mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3)\mathbf{e}_j \rangle \partial_{i\mu}^{s_1 - r_1}(D_{jj} - 1). \tag{4.152}$$

According to the equations (4.124) and (4.125), we write, abbreviating $K_3 := K_3(X_{i\mu}^1, X_{i\mu}^\theta) - K_3(X_{i\mu}^0, X_{i\mu}^\theta)$ (suppressing the dependence on $i\mu$, which is not consequential),

$$\begin{aligned} &\mathbb{E} \sum_{i\mu} K_3 \sum_{j_1,\ldots,j_q} \prod_{\alpha=1}^{p} \partial_{i\mu}^{r_\alpha} \langle \mathbf{e}_{j_\alpha}, (\mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3)\mathbf{e}_{j_\alpha} \rangle \partial_{i\mu}^{s_\alpha - r_\alpha}(D_{j_\alpha j_\alpha} - 1) \\ &= \sum_{m_1,\ldots,m_q} \mathbb{E} \sum_{i\mu} K_3 \sum_{j_1,\ldots,j_q} \prod_{\alpha=1}^{p} C_{m_\alpha} \prod_{n_\alpha=1}^{r_\alpha+2} \mathbf{p}_{\alpha,m_\alpha,n_\alpha} \partial_{i\mu}^{s_\alpha - r_\alpha}(D_{j_\alpha j_\alpha} - 1), \end{aligned} \tag{4.153}$$

where

$$\mathbf{p}_{\alpha,m_\alpha,n_\alpha} = \langle \mathbf{x}_{\alpha,m_\alpha,2n_\alpha-1}, \mathbf{W}_{\alpha,m_\alpha,n_\alpha} \mathbf{x}_{m_\alpha,2n_\alpha} \rangle, \tag{4.154}$$

where $\mathbf{x}_{\alpha,m_\alpha,1} = \mathbf{x}_{\alpha,m_\alpha,2r_\alpha+4} \in \{\mathbf{e}_{j_\alpha}\}$ for each $\alpha = 1, \ldots, q$, and where

$$\mathbf{W}_{\alpha,m_\alpha,n_\alpha} \in \{U\Pi U^*, G, UG, GU^*, UGU^*, I\}.$$

From this point on, we consider the choice of $\boldsymbol{m} = (m_1, \ldots, m_p)$ to be fixed, and we denote $C_{\boldsymbol{m}} := \prod_{\alpha=1}^{p} C_{m_\alpha}$. We will now drop the bounded constant $C_{\boldsymbol{m}}$ and the sum over the sum over the boundedly many values of $\boldsymbol{m}$ and not mention them again, and we will drop all m_α subscipts from our notation.

Let $q_0, q_1, q_2 \in \mathbb{N}^{\geq 0}$ be such that $q_0 + q_1 + q_2 = q$, and without loss of generality let $\alpha \in [\![1, q_0]\!]$ satisfy $s_\alpha > 0, s_\alpha - r_\alpha = 0$, let $\alpha \in [\![q_0+1, q_0+q_1]\!]$ satisfy $s_\alpha > 0, s_\alpha - r_\alpha = 1$, let $\alpha \in [\![q_0+q_1+1, q_0+q_1+q_2]\!]$ satisfy $s_\alpha > 0, s_\alpha - r_\alpha = 2$, and let $\alpha \in [\![q_0+q_1+q_2+1, p]\!] = [\![q+1, p]\!]$ satisfy $s_\alpha = 0$. Recall that $s_\alpha - r_\alpha > 2$ for any α makes (4.153) equal 0, so we do not consider this possibility. Because $\partial_{i\mu}^{s}(D_{jj} - 1) = O(\delta_{ij} + \delta_{i \in \mathcal{I}_K})$ for $s \geq 1$, and since i in this section only ranges over $\mathcal{I}_M$, we may write the equation (4.153),

$$\mathbb{E} \sum_{i\mu} K_3 \sum_{j_1,\ldots,j_{q_0}} \prod_{\alpha=1}^{p} \prod_{n_\alpha=1}^{r_\alpha+2} \mathbf{p}_{\alpha,n_\alpha} \partial_{i\mu}^{s_\alpha - r_\alpha}(D_{j_\alpha j_\alpha} - 1), \tag{4.155}$$

where $j_\alpha := i$ for $\alpha > q_0$.

The work of proving Proposition 4.6.0.1 is then divided as follows:

Proposition 4.6.0.2. *If* $q_1 \geq 1$, *equation* (4.155) *may be bounded by* $N^{(-1/2-\epsilon_D)p} + C\mathbb{E}|h|^p$.

Proposition 4.6.0.3. *If* $q_1 = 0$, *equation* (4.155) *may be bounded by* $N^{(-1/2-\epsilon_D)p} + C\mathbb{E}|h|^p$.

□

Let us conclude this subsection by first treating the case that $q_1 \geq 1$.

Proof of Proposition 4.6.0.2. The heuristic of this proof is simple: for each $\alpha \in [\![1, p - q_h]\!]$, we can find a factor with a $\mathbf{v}$ in it to contribute a $N^{-\epsilon_D}$. For $\alpha \in [\![1, q_0]\!]$, we may also find two factors with j_α to cancel the sum over j_α, as well as a factor of $N^{-1/2}$ from the $\partial^0_{i\mu}(D_{j_\alpha j_\alpha} - 1)$, while for $\alpha \in [\![q_0 + q_1 + 1, q_0 + q_1 + q_2]\!]$, we have no j_α over which to sum, but we have a factor containing $\mathbf{e}_i$, which functions as a factor of $N^{-1/2}$ by reducing the size of the sum over i. For $\alpha \in [\![q_0 + 1, q_0 + q_1]\!]$, however, we have both a factor of $N^{-1/2}$ from the $\partial^1_{i\mu}(D_{j_\alpha j_\alpha} - 1)$ and a factor containing i (and no j_α over which to sum). So every $\alpha \in [\![1, q_0 + q_1 + q_2]\!]$ contributes a $N^{-1/2-\epsilon_D}$, and $\alpha \in [\![q_0 + 1, q_0 + q_1]\!]$ contribute additional factors of $N^{-1/2}$.

The assumption of this proposition is that $q_1 \geq 1$. It follows from equations (4.124) and (4.125) that for each $\alpha \in [\![1, q_0 + q_1 + q_2]\!]$, there is an $n^{(1)}_\alpha$ such that $\mathbf{v} \in \{\mathbf{x}_{\alpha, 2n^{(1)}_\alpha - 1}, \mathbf{x}_{\alpha, 2n^{(1)}_\alpha}\}$; and there is moreover, for each $\alpha \in [\![1, q_0]\!]$, values $n^{(3)}_\alpha \neq n^{(4)}_\alpha$, distinct from $n^{(1)}_\alpha$, such that $j_\alpha \in \{\mathbf{x}_{\alpha, 2n^{(3)}_\alpha - 1}, \mathbf{x}_{\alpha, 2n^{(3)}_\alpha}\} \cap \{\mathbf{x}_{\alpha, 2n^{(4)}_\alpha - 1}, \mathbf{x}_{\alpha, 2n^{(4)}_\alpha}\}$; and there is again, for each $\alpha \in [\![q_0 + 1, p]\!]$, a value $n^{(2)}_\alpha$, distinct from $n^{(1)}_\alpha$, for which $\mathbf{e}_i \in \{\mathbf{x}_{\alpha, 2n^{(2)}_\alpha - 1}, \mathbf{x}_{\alpha, 2n^{(2)}_\alpha}\}$.

Since

$$\sum_{j_\alpha} \left| \mathbf{p}_{\alpha, n^{(3)}_\alpha} \mathbf{p}_{\alpha, n^{(4)}_\alpha} \right| \prec 1,$$

since $\left| \mathbf{p}_{\alpha, n^{(1)}_\alpha} \right| \prec N^{-\epsilon_D}$, and since $\partial^{s_\alpha - r_\alpha}_{i\mu}(D_{j_\alpha j_\alpha} - 1) = O_\prec(N^{-1/2})$ for each $\alpha \in [\![1, \ldots, q_0 + q_1]\!]$, equation (4.155) may be bounded in absolute value by, using $K_3 = O_\prec(N^{-3/2})$,

$$\mathbb{E}N^{-(q_0+q_1)/2}N^{-q\epsilon_D}N^{-3/2}\sum_{i\mu}\prod_{\alpha=p_0+1}^{q}\left|\mathbf{p}_{\alpha,n_\alpha^{(2)}}\right||h|^{p-q}. \tag{4.156}$$

Now we proceed by cases: if $q_2 = 1$, then because $q_1 \geq 1$ and $3 = m \geq q_0 + q_1 + 2q_2$, we necessarily have $q_0 = 0$ and $q_1 = 1$, and thus $q = 2$. Note that

$$\sqrt{\sum_i \left|\mathbf{p}_{\alpha,n_\alpha^{(2)}}\right|^2} \prec 1 \tag{4.157}$$

for $\alpha \geq q_0+1$ and that the product $\prod_{\alpha=q_0+1}^{q}$ has 2 factors in it. Thus, equation (4.156) may be bounded by, using Cauchy-Schwarz,

$$\mathbb{E}N^{-1/2}N^{-q\epsilon_D}N^{-3/2}N|h|^{p-q} = N^{-2(1/2+\epsilon_D)}\mathbb{E}|h|^{p-2} \tag{4.158}$$

and then Young's inequality produces the desired result. If rather $q_2 = 0$ (since $m = 3$, it is impossible for $q_2 > 1$), then equation (4.156) is bounded by

$$\mathbb{E}N^{-q/2}N^{-q\epsilon_D}N^{-3/2}N^{3/2}|h|^{p-q} = N^{-q(1/2+\epsilon_D)}\mathbb{E}|h|^{p-q} \tag{4.159}$$

since the product $\prod_{\alpha=q_0+1}^{q}$ has at least 1 factor in it. The result then follows again by Young's inequality. This concludes the proof of Proposition 4.6.0.2. □

Now, restricting to $q_1 = 0$, we may write equation (4.155) as

$$\mathbb{E}\sum_{i\mu}K_3\sum_{j_1,\ldots,j_{p-q_2}}\prod_{\alpha=1}^{p-q_2}\prod_{n_\alpha=1}^{r_\alpha+2}\mathbf{p}_{\alpha,n_\alpha}(D_{j_\alpha j_\alpha}-1)\left(\prod_{n_p=1}^{r_p+2}\mathbf{p}_{p,n_p}\right)^{q_2}, \tag{4.160}$$

where we recall that q_2 may only by 0 or 1, and we have that if $q_2 = 1$, recalling equation (4.154),

$$\mathbf{x}_{p,1} = \mathbf{x}_{p,2r_p+4} = \mathbf{e}_i.$$

4.6.1 Identities for characterising dependence on $X\mathbf{e}_\mu$

In order to prove Proposition 4.6.0.3, which is done in the next subsection, we must find an extra factor of $N^{-1/2}$. Consider the following identities, which we have adapted from [KY17] wherein the technique of this section was developed: for any vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_K \cup \mathcal{I}_M}$, we have

$$\begin{aligned} G_{\mathbf{x}\mu} &= -G_{\mu\mu}(G^{(\mu)}X)_{\mathbf{w}\mu} \\ G_{\mathbf{x}\mathbf{y}} &= G^{(\mu)}_{\mathbf{x}\mathbf{y}} + G_{\mu\mu}(G^{(\mu)}X)_{\mathbf{x}\mu}(X^*G^{(\mu)})_{\mu\mathbf{y}} \\ G_{\mu\mu} &= \sum_{\ell=0}^{L} \mathcal{Y}_{\mu\ell}(X^*G^{(\mu)}X)^{\ell}_{\mu\mu} + O_{\prec}(N^{-p}), \end{aligned} \tag{4.161}$$

where $\mathcal{Y}_{\mu\ell}$ are uniformly $O_{\prec}$ bounded random variables independent of $X\mathbf{e}_\mu$ and L is a fixed constant (dependending on p, which in this argument is considered fixed). In what follows, $\mathcal{Y}$ will always denote a random variable which is implicity indexed by all indices present and which is $O_{\prec}$ bounded and indpenedent of $X\mathbf{e}_\mu$; thus we treat $\mathcal{Y}$ somewhat like a constant, and the particular value of $\mathcal{Y}$ may change from one line to the next. The first two identities follow from basic resolvent identities, and the third may be derived as follows: using the identity

$$\frac{1}{G_{\mu\mu}} = -z - (X^*G^{(\mu)}X)_{\mu\mu}$$

from [KY17], lemma 4.4, we get

$$G_{\mu\mu} = (-z - Y_\mu - Z_\mu)^{-1} = \sum_{\ell=0}^{L}(-z - Y_\mu)^{-\ell-1}Z_\mu^\ell + O_{\prec}(N^{-p}),$$

where

$$Z_\mu := (X^*G^{(\mu)}X)_{\mu\mu} - Y_\mu, \quad Y_\mu := \mathbb{E}\left[(X^*G^{(\mu)}X)_{\mu\mu}|X_\mu\right] = \frac{1}{N}\sum_{j\in\mathcal{I}_M} G^{(\mu)}_{jj}$$

and the truncation is permissible for sufficiently large L by Lemma 4.2.7.

The purpose of the three identities (4.161) is to break up all the resolvent entries which we will encounter in our proof into a piece independent of $X\mathbf{e}_\mu$ and a piece formed from $X\mathbf{e}_\mu$. Note that the parity of the number of entries of X in each resolvent identity for $G_{\mathbf{w}_1\mathbf{w}_2}$ is equal to the parity of the number of $\{\mathbf{w}_1, \mathbf{w}_2\}$ which are μ. This is crucial.

We need a similar, though simpler, identity for the factors $D_{jj} - 1$: we have

$$D_{jj} - 1 = \left(D^{(\mu)}_{jj} - 1\right) + (\widetilde{S}X)^2_{j\mu}, \quad D^{(\mu)}_{jj} := \sum_{j'\in\mathcal{I}_M\setminus\{j\}} (\widetilde{S}X)^2_{j'\mu}. \tag{4.162}$$

With the identities (4.161), we adopt the following notation for $\mathbf{p}_{\alpha,n_\alpha} = G_{\mathbf{x},\mu}, G_{\mathbf{x},\mathbf{y}}$, or $G_{\mu\mu}$, where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_K \cup \mathcal{I}_\mathcal{M}}$ and $\mu \in \mathcal{I}_N$. If $\mathbf{p}_{\alpha,n_\alpha} = G_{\mu\mu}$, then we define $\mathbf{p}^{(-1)}_{\alpha,n_\alpha} = 0$ and

$$\mathbf{p}^{(\ell)}_{\alpha,n_\alpha} := \left(\mathbf{p}^{(\ell)}_{\alpha,n_\alpha}\right)_{k_1,\ldots,k_{2\ell}} := \mathcal{Y}\prod_{\beta=1}^{\ell} G^{(\mu)}_{k_{2\beta-1},k_{2\beta}}$$

so that, choosing L large enough here and in the following equations,

$$\mathbf{p}_{\alpha,n_\alpha} = \sum_{\ell=-1}^{L} \sum_{\substack{k_1,\ldots,k_{2\ell} \\ \in \mathcal{I}_K \cup \mathcal{I}_M}} \mathbf{p}^{(\ell)}_{\alpha,n_\alpha} \prod_{\beta=1}^{2\ell} X_{k_\beta \mu} + O_\prec(N^{-p}).$$

If $\mathbf{p}_{\alpha,n_\alpha} = G_{\mathbf{x}\mu}$ we define $\mathbf{p}^{(-1)}_{\alpha,n_\alpha} = 0$ and

$$\mathbf{p}^{(\ell)}_{\alpha,n_\alpha} := \left(\mathbf{p}^{(\ell)}_{\alpha,n_\alpha}\right)_{k_1,\ldots,k_{2\ell+1}} = \mathcal{Y} G^{(\mu)}_{\mathbf{x}k_1} \prod_{\beta=1}^{\ell} G^{(\mu)}_{k_{2\beta}k_{2\beta+1}}$$

for $\ell \geq 0$, so that

$$\mathbf{p}_{\alpha,n_\alpha} = \sum_{\ell=-1}^{L} \sum_{\substack{k_1,\ldots,k_{2\ell+1} \\ \in \mathcal{I}_K \cup \mathcal{I}_M}} \mathbf{p}^{(\ell)}_{\alpha,n_\alpha} \prod_{\beta=1}^{2\ell+1} X_{k_\beta \mu} + O_\prec(N^{-p}).$$

If $\mathbf{p}_{\alpha,n_\alpha} = G_{\mathbf{x}\mathbf{y}}$, we define $\mathbf{p}^{(-1)}_{\alpha,n_\alpha} = G^{(\mu)}_{\mathbf{x}\mathbf{y}}$ and

$$\mathbf{p}^{(\ell)}_{\alpha,n_\alpha} := \mathcal{Y} G_{\mathbf{x}k_1} G_{k_2\mathbf{y}} \prod_{\beta=1}^{\ell} G^{(\mu)}_{k_{2\ell+1}k_{2\ell+2}}$$

for $\ell \geq 0$, so that

$$\mathbf{p}_{\alpha,n_\alpha} = \sum_{\ell=-1}^{L} \sum_{\substack{k_1,\ldots,k_{2\ell+2} \\ \in \mathcal{I}_K \cup \mathcal{I}_M}} \mathbf{p}^{(\ell)}_{\alpha,n_\alpha} \prod_{\beta=1}^{2\ell+2} X_{k_\beta \mu} + O_\prec(N^{-p}).$$

Lastly, if $\mathbf{W}_{\alpha,n_\alpha} \in \{U\Pi U^*, I\}$, we define $\mathbf{p}^{(\ell)}_{\alpha,n_\alpha} := \mathbf{p}_{\alpha,n_\alpha}$ for $\ell = 0$ and $\mathbf{p}^{(\ell)}_{\alpha,n_\alpha} := 0$ otherwise. In this case we also observe that $\mathbf{x}_{\alpha,1}, \ldots, \mathbf{x}_{\alpha,2n_\alpha+4} \neq \mathbf{e}_\mu$.

4.6.2 Proof of Proposition 4.6.0.3

Thus we may write equation (4.160), reordering labels if necessary,

$$\begin{aligned}
&\sum_{\boldsymbol{\ell}} \mathbb{E} \sum_{k_1,\dots,k_{n_X}} \sum_{i\mu} K_3 \left(\sum_{j_1,\dots,j_{n_\Delta}} \mathbf{P}_1 \right) \left(\sum_{j_{n_\Delta+1},\dots,j_{p-q_2}} \mathbf{P}_2 \right) \mathbf{P}_3 + O_\prec(N^{-p}) \\
&\mathbf{P}_1 := (\mathbf{P}_1)_{i,\mu,\boldsymbol{k},\boldsymbol{j}} := \sum_{j_1,\dots,j_{n_\Delta}} \prod_{\alpha=1}^{n_\Delta} \prod_{n_\alpha=1}^{r_\alpha+2} \mathbf{p}_{\alpha,n_\alpha}^{(\ell_{\alpha,n_\alpha})} \left(\mathbb{E} \prod_{\alpha=1}^{n_X} X_{k_\alpha \mu} \prod_{\alpha=1}^{n_\Delta} (\widetilde{S}X)^2_{j_\alpha \mu} \right) \\
&\mathbf{P}_2 := (\mathbf{P}_2)_{i,\mu,\boldsymbol{k},\boldsymbol{j}} := \prod_{\alpha=n_\Delta+1}^{p-q_2} \prod_{n_\alpha=1}^{r_\alpha+2} \mathbf{p}_{\alpha,n_\alpha}^{(\ell_{\alpha,n_\alpha})} (D^{(\mu)}_{j_\alpha j_\alpha} - 1) \\
&\mathbf{P}_3 := (\mathbf{P}_3)_{i,\mu,\boldsymbol{k}} := \left(\prod_{n_p=1}^{r_p+2} \mathbf{p}_{p,n_p}^{(\ell_{p,n_p})} \right)^{q_2},
\end{aligned} \tag{4.163}$$

where

$$\boldsymbol{\ell} = (\ell_{1,1}, \dots, \ell_{1,r_1+2}, \dots, \ell_{p,1}, \dots, \ell_{p,r_p+2}) \in [\![1, L]\!]^{\sum_{\alpha=1}^{p}(r_\alpha+2)}.$$

We now fix a choice of $\boldsymbol{\ell}$ and drop the sum over $\boldsymbol{\ell}$. We are able to add the expectation into $\mathbf{P}_1$ like we have done because every other factor in $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$ is independent of $X\mathbf{e}_\mu$.

Note that only $\mathbf{P}_1$ depends on $j_1, \dots, j_{n_\Delta}$ while $\mathbf{P}_2, \mathbf{P}_3$ do not. Similarly only $\mathbf{P}_2$ depends on $j_{n_\Delta+1}, \dots, j_{p-q_2}$.

Now consider a tuple

$$(k_1, \dots, k_{n_X}, j_1, \dots, j_{n_\Delta}) = (\boldsymbol{k}, \boldsymbol{j}) \in (\mathcal{I}_K \cup \mathcal{I}_M)^{n_X} \times \mathcal{I}_M^{n_\Delta}.$$

Just to allow us to make the following necessary definition consicely, define, for $\alpha \in 1, \dots, n_\Delta$, the alias $k_{n_X+\alpha} = j_\alpha$. Now we define a very slight generalization of a partition of the indices $\{1, \dots, n_X + n_\Delta\}$ which we will call a *signature*. A signature P of the symbols $\{1, \dots, n_X + n_\Delta\}$ is a pair

$$P = (A_1, \{A_2, \dots, A_n\})$$

such that $\{A_1, A_2, \dots, A_n\}$ is a partition of $\{1, \dots, n_X + n_\Delta\}$—that is, a signature is a partition with one distinguished block. We call $A_1, \dots, A_n$ *blocks* of the signature. We say that a tuple $(\boldsymbol{k}, \boldsymbol{j})$ *satisfies* P and write $(\boldsymbol{k}, \boldsymbol{j}) \vdash P$ if

1. $\alpha \in A_1$ if and only if $k_\alpha \in \mathcal{I}_K$. Note that since $j_\alpha \notin \mathcal{I}_K$, necessarily $\alpha \in \{1, \dots, n_X\}$ in this case.

2. α, α' are in the same element of $\{A_2, \dots, A_n\}$ if and only if $k_\alpha, k_{\alpha'} \notin \mathcal{I}_K$ and $k_\alpha = k_{\alpha'}$.

Let us now fix a signature P for the remainder of the proof. Also define $\epsilon = \min\{\epsilon_D, 1/4\}$.

Let P' be the partition $\{A_2, \dots, A_n\}$ of $\{1, \dots, n_X + n_\Delta\} \setminus A_1$. We form a choice set from P', taking one element α_a from each block A_a for each $a = 2, \dots, n$, in such a way that if A_a contains any index in $\{n_X + 1, \dots, n_X + n_\Delta\}$, then $\alpha_a \in \{n_X + 1, \dots, n_X + n_\Delta\}$. For notational simplicity and without loss of generality, we assume that

$$(\alpha_2, \dots, \alpha_n) = (1, \dots, n'_X, n_X + 1, \dots, n_X + n'_\Delta)$$

and that

$$A_1 = \{n'_X + 1, \dots, n'_X + n_K\}.$$

This is all to say that we can write

$$\sum_{k_1, \dots, k_{n_X}} \sum_{j_1, \dots, j_{n_\Delta}} \mathbf{1}_{(\boldsymbol{k}, \boldsymbol{j}) \vdash P}(\cdots)$$

as

$$\sum_{\substack{k_{n'_x+1}, \dots, k_{n'_X+n_K} \\ \in \mathcal{I}_K}} \sum_{\substack{k_1, \dots, k_{n'_X} \\ \in \mathcal{I}_M}} \sum_{j_1, \dots, j_{n'_\Delta}} \mathbf{1}_{(\boldsymbol{k}, \boldsymbol{j}) \vdash P}(\cdots) \tag{4.164}$$

for any expression $(\cdots)$, and in the above sum no index $k_1, \dots, k_{n'_X}$ is identified with any index $j_1, \dots, j_{n'_\Delta}$. Let now $\widetilde{n}_{\text{diag}}$ be the number of pairs

$$(\alpha, b) \in \{(1, 1), (1, r_1 + 2)), \dots, (n'_\Delta, 1), (n'_\Delta, r_{n'_\Delta} + 2)\}$$

for which $\ell_{\alpha,b} \geq 0$. Also let $n_{\text{diag}} \leq \widetilde{n}_{\text{diag}}$ be the number of those pairs among the $\widetilde{n}_{\text{diag}}$ for which α and $\beta(\alpha, b)$ are identified by P. The number n_{diag} is important because the identification of j_α with $k_{\beta(\alpha,b)}$ means that $\sum_{j_\alpha} \mathbf{p}_{\alpha,b}^{(\ell_{\alpha,b})} = O(N)$ rather than $O(N^{1/2})$, leading to a potentially worse estimate for the sum (4.164); however, the identification of j_α with $k_{\beta(\alpha,b)}$ also reduces the number of indices in the sum, and this will offset the potential loss.

Define also the number

$$n_K := n_K(P) = |A_1|.$$

For brevity, define

$$\boldsymbol{X}_{\boldsymbol{k},\boldsymbol{j}} := \mathbb{E}\prod_{\alpha=1}^{n_X} X_{k_\alpha\mu}\prod_{\alpha=1}^{n_\Delta}(\widetilde{S}X)^2_{j_\alpha\mu},$$

which has the bound $N^{-n_X/2-n_\Delta}$.

Proof of Proposition 4.6.0.3. We state the following three lemmas, which are proven after the conclusion of the current proof.

Lemma 4.6.1. *We have*

$$\sum_{j_1,\dots,n_\Delta}\mathbf{1}_{(\mathbf{k},\mathbf{j})\vdash P}|\mathbf{P}_1| \prec \left(N^{-1/2}+|\langle\mathbf{e}_i,U^*\Pi\mathbf{v}\rangle|+\sum_{\alpha=1}^{n_X}|\langle\mathbf{e}_{k_\alpha},U^*\Pi\mathbf{v}\rangle|\right)^{\mathbf{1}_{\widetilde{n}_{\mathrm{diag}}\neq 0}} \cdot N^{n_{\mathrm{diag}}/2-n_X/2-n_\Delta-\epsilon\mathbf{1}_{n_K=1}} \tag{4.165}$$

uniformly in all indices $i,\mu,k_1,\dots,k_{n_X},j_{n_\Delta+1},\dots,j_{p-q_2}$.

Lemma 4.6.2. *Let* $c=p-q_2-n_\Delta$. *We have that if* $(\boldsymbol{k},\boldsymbol{j})\vdash P$, *then* $\mathbf{P}_2$ *satisfies*

$$\sum_{j_{n_\Delta+1},\dots,j_{p-q_2}}|\mathbf{P}_2| \prec \left(\mathbb{E}|h|^c+N^{-(1/2+\epsilon)c}\right)N^{\epsilon n_K}$$

uniformly in all indices $i,\mu,k_1,\dots,k_{n_X},j_1,\dots,j_{n_\Delta}$.

Lemma 4.6.3. *We have*

$$N^{-3/2}\sum_{i\mu}|\mathbf{P}_3|\left(N^{-1/2}+|\langle\mathbf{e}_i,U^*\Pi\mathbf{v}\rangle|\right) = \begin{cases}1 & q_2=0;\\ O(N^{-1/2-\epsilon_D}) & q_2=1\end{cases}$$

uniformly in all indices $k_1,\dots,k_{n_X},j_1,\dots,j_{p-q_2}$.

The way to intepret the complicated exponentiated factor in Lemma 4.6.1 is that in the case that $\widetilde{n}_{\mathrm{diag}}\neq 0$, we get an additional factor which sums to $N^{1/2}$ rather than N over i or over some k_α. We will need this improvement of a factor of $N^{-1/2}$ in the case that $\widetilde{n}_{\mathrm{diag}}\neq 0$ since in that case the parity argument which we would hope to help us actually fails.

By Lemmas 4.6.1 and 4.6.2, we bound

$$\left|\sum_{i\mu} K_3 \left(\sum_{j_1,\dots,j_{n_\Delta}} \mathbf{P}_1\right)\left(\sum_{j_{n_\Delta+1},\dots,j_{p-q_2}} \mathbf{P}_2\right)\mathbf{P}_3\right|$$
$$\prec \sum_{i\mu} K_3|\mathbf{P}_3|\left(N^{-1/2} + |\langle \mathbf{e}_i, U^*\Pi\mathbf{v}\rangle| + \sum_{\alpha=1}^{n_X}|\langle \mathbf{e}_{k_\alpha}, U^*\Pi\mathbf{v}\rangle|\right)^{\widetilde{n}_{\text{diag}}}$$
$$\cdot N^{n_{\text{diag}}/2-n_X/2-n_\Delta}N^{-\epsilon\mathbf{1}_{n_K=1}}N^{\epsilon n_K}\mathbb{E}\left(N^{-(1/2+\epsilon)(p-q_2-n_\Delta)} + |h|^{p-q_2-n_\Delta}\right),$$

By Lemma 4.6.3, this is

$$\leq \mathcal{A}(\widetilde{n}_{\text{diag}})N^{(-1/2-\epsilon_D)q_2}N^{n_{\text{diag}}/2-n_X/2-n_\Delta}N^{-\epsilon\mathbf{1}_{n_K=1}}N^{\epsilon n_K}$$
$$\cdot\,\mathbb{E}\left(|h|^{p-q_2-n_\Delta} + N^{-(1/2+\epsilon)(p-q_2-n_\Delta)}\right),$$

where

$$\mathcal{A}(n) = \begin{cases} \left(1 + N^{1/2}\sum_{\alpha=1}^{n_X}|\langle \mathbf{e}_{k_\alpha}, U^*\Pi\mathbf{v}\rangle|\right) & n \neq 0 \\ N^{1/2} & n = 0. \end{cases}$$

Using $\epsilon_D < 1/2$, we see that this is

$$\leq \mathcal{A}(\widetilde{n}_{\text{diag}})N^{n_{\text{diag}}/2-n_X/2}N^{-\epsilon\mathbf{1}_{n_K=1}}N^{\epsilon n_K}$$
$$\cdot\,\mathbb{E}\left(N^{(-1/2-\epsilon)(n_\Delta+q_2)}|h|^{p-q_2-n_\Delta} + N^{-(1/2+\epsilon)p}\right).$$

Now we prepare to sum over $k_1,\dots,k_{n_X}$. We find that $n'_X \leq \lfloor\frac{n_X-n_{\text{diag}}-n_K}{2}\rfloor$ lest $\boldsymbol{X}_{\boldsymbol{k},\boldsymbol{j}} = 0$—indeed, $n_{\text{diag}} + n_K$ are already in $\mathcal{I}_K$ or paired with an index j_α which has already been summed over; the remaining indices must at least pairwise identify, lest there be an index k_{α_0} which is distinct from all other indices, and then $\boldsymbol{X}_{\boldsymbol{k},\boldsymbol{j}}$ would be 0 by the independence of the entries of X.

Note now that

$$\sum_{k_1,\dots,k_{n'_X}} \mathcal{A}(\widetilde{n}_{\text{diag}}) \leq N^{\frac{n_X-n_{\text{diag}}-n_K}{2}+\frac{1}{2}\mathbf{1}_{n_K\neq 0}},$$

which follows from Cauchy-Schwarz if $\widetilde{n}_{\text{diag}} \neq 0$, and otherwise is trivial if $n_K = 0$ but follows from the fact that n_X is odd if $n_K = 0$.

Therefore equation (4.163) may be bounded by

$$\begin{aligned}&\prec N^{n_X/2-n_{\mathrm{diag}}/2-n_K/2+\frac{1}{2}\mathbf{1}_{n_K\neq 0}}N^{n_{\mathrm{diag}}/2-n_X/2}N^{-\epsilon\mathbf{1}_{n_K=1}}N^{\epsilon n_K}\\&\quad\cdot\mathbb{E}\left(N^{(-1/2-\epsilon)(n_\Delta+q_2)}|h|^{p-q_2-n_\Delta}+N^{-(1/2+\epsilon)p}\right)\\&\leq 1\cdot\mathbb{E}\left(N^{(-1/2-\epsilon)(n_\Delta+q_2)}|h|^{p-q_2-n_\Delta}+N^{-(1/2+\epsilon)p}\right),\end{aligned}$$

which follows from $-\frac{n_K}{2}+\frac{1}{2}\mathbf{1}_{n_K\neq 0}-\epsilon\mathbf{1}_{n_K=1}+\epsilon n_K\geq 0$. We conclude by Young's inequality and by summing over the boundedly many signatures P. □

Proof of Lemma 4.6.1. Notice that for each pair (α,b) among the number $\widetilde{n}_{\mathrm{diag}}$, we have

$$\mathbf{p}_{\alpha,b}^{(\ell_{\alpha,b})}=\left\langle\mathbf{e}_{j_\alpha},\mathbf{W}\mathbf{e}_{k_{\beta(\alpha,b)}}\right\rangle O_\prec(1) \tag{4.166}$$

for an index $\beta(\alpha,b)\in\{1,\ldots,n_X\}$ and a matrix $\mathbf{W}$ of operator norm $O_\prec(1)$.

Recalling the number n_K, we proceed by cases on n_K. First, if $n_K\neq 1$, we may bound equation (4.165), using the reduction (4.164) and recalling the definition of n_{diag}, by

$$\begin{aligned}&\prec\prod_{\alpha=1}^{n'_\Delta}\left(\sum_{j_\alpha}\left|\mathbf{p}_{\alpha,1}^{(\ell_{\alpha,1})}\mathbf{p}_{\alpha,r_\alpha+2}^{(\ell_{\alpha,r_\alpha+2})}\right|\right)\mathbf{1}_{(\boldsymbol{k},\boldsymbol{j})\vdash P}\boldsymbol{X}_{\boldsymbol{k},\boldsymbol{j}}\\&\prec N^{n_{\mathrm{diag}}/2}N^{-n_X/2-n_\Delta},\end{aligned}\tag{4.167}$$

which follows by Cauchy-Schwarz.

Next, if $n_K=1$, we note that we can actually get a stronger bound on $\boldsymbol{X}_{\boldsymbol{k},\boldsymbol{j}}$: by writing

$$(\widetilde{S}X)_{j_\alpha\mu}=\widetilde{S}_{j_\alpha j_\alpha}X_{j_\alpha\mu}+\sum_{\beta\in\mathcal{I}_K}\widetilde{S}_{j_\alpha\beta}X_{\beta\mu}=\widetilde{S}_{j_\alpha j_\alpha}X_{j_\alpha\mu}+O(N^{-\epsilon_D-1/2}),$$

we see that

$$\begin{aligned}\boldsymbol{X}_{\boldsymbol{k},\boldsymbol{j}}&=\mathbb{E}\prod_{\alpha=1}^{n_X}X_{k_\alpha\mu}\prod_{\alpha=1}^{n_\Delta}(\widetilde{S}_{j_\alpha j_\alpha}X_{j_\alpha\mu})^2+O_\prec\left(N^{-n_X/2-n_\Delta-\epsilon_D}\right)\\&=0+O_\prec\left(N^{-n_X/2-n_\Delta-\epsilon_D}\right)\end{aligned}$$

by independence. Then, as before, we can bound equation (4.164) by

$$\prec N^{n_{\mathrm{diag}}/2}N^{-n_X/2-n_\Delta-\epsilon_D}$$

as desired.

Now we must account for the strange exponentiated factor. If $\widetilde{n}_{\text{diag}} = 0$, we are done. Otherwise, we merely note that $\widetilde{n}_{\text{diag}} \neq 0$ implies that, improving on equation (4.166),

$$\mathbf{p}_{\alpha,b}^{(\ell_{\alpha,b})} = \left\langle \mathbf{e}_{j_\alpha}, \mathbf{W}\mathbf{e}_{k_{\beta(\alpha,b)}} \right\rangle \left(\langle \mathbf{e}_\mu, U^* G\mathbf{v} \rangle + \langle \mathbf{e}_i, U^* G\mathbf{v} \rangle + \sum_{\alpha=1}^{n_X} \langle \mathbf{e}_{k_\alpha}, U^* G\mathbf{v} \rangle \right) O_\prec(1),$$

at which point Lemma 4.2.7 allows us to conclude. □

Proof of Lemma 4.6.2. We can factor

$$\sum_{j_{n_\Delta+1},\dots,j_{p-q_2}} \mathbf{P}_2 = \prod_{\alpha=n_\Delta+1}^{p-q_2} \sum_{j_\alpha} \prod_{n_\alpha=1}^{r_\alpha+2} \mathbf{p}_{\alpha,n_\alpha}^{(\ell_{\alpha,n_\alpha})}. \tag{4.168}$$

Now, similarly to in the proof of Proposition 4.6.0.2, if $r_\alpha \geq 1$, then

$$\sum_{j_\alpha} \left| \mathbf{p}_{\alpha,1}^{(\ell_{\alpha,1})} \mathbf{p}_{\alpha,r_\alpha+2}^{(\ell_{\alpha,r_\alpha+2})} \right| = O_\prec(1),$$

while for some $1 < n_\alpha < r_\alpha + 2$ we have $\mathbf{p}_{\alpha,n_\alpha}^{(\ell_{\alpha,n_\alpha})} = \langle \mathbf{v}, UG\mathbf{y} \rangle$ for $\mathbf{y} \in \{\mathbf{e}_i, \mathbf{e}_{k_\beta}\}$ for some $\beta \in \{1, \dots, n_X\}$; thus $\left| \mathbf{p}_{\alpha,n_\alpha}^{(\ell_{\alpha,n_\alpha})} \right| \prec N^{-\epsilon_D}$ if $\mathbf{y} = \mathbf{e}_i$ or if $k_\beta \in \mathcal{I}_K$ and $\prec 1$ otherwise.

Similarly, if $r_\alpha = 0$ and $\ell_{\alpha,1}, \ell_{\alpha,2}$ are not both -1, say $\ell_{\alpha,2} \neq -1$, then

$$\left| \mathbf{p}_{\alpha,1}^{(\ell_{\alpha,1})} \mathbf{p}_{\alpha,2}^{(\ell_{\alpha,2})} \right| \prec |\langle \mathbf{e}_{j_\alpha}, \mathbf{W}_1 \mathbf{y}_1 \rangle \langle \mathbf{v}, \mathbf{W}_2 \mathbf{y}_2 \rangle \langle \mathbf{y}_3, \mathbf{W}_3 \mathbf{e}_{j_\alpha} \rangle|,$$

where $\mathbf{W}_1, \dots, \mathbf{W}_3 \in \{\Pi, G, UG, GU^*, UGU^*, I\}$ and $\mathbf{y}_1, \dots, \mathbf{y}_3$ are each either $\mathbf{e}_{k_\beta}$ or $\mathbf{v}$; thus we may bound

$$\sum_{j_\alpha} \prod_{n_\alpha=1}^{r_\alpha+2} \left| \mathbf{p}_{\alpha,n_\alpha}^{(\ell_{\alpha,n_\alpha})} \right| \prec N^{-\epsilon_D}$$

if $k_\beta \in \mathcal{I}_M$ and $\prec 1$ otherwise.

Finally, if rather $r_\alpha = 0$ and $\ell_{\alpha,1} = \ell_{\alpha,2} = -1$, corresponding to the leading order term in the application of the identities (4.161) to an undifferentiated term of $\mathbf{M}$, we have

$$\mathbf{p}_{\alpha,1}^{(-1)} \mathbf{p}_{\alpha,2}^{(-1)} = \left(\mathbf{M}_a^{(\mu)} \right)_{j_\alpha j_\alpha},$$

where we define $\mathbf{M}^{(\mu)} = \mathbf{M}_1^{(\mu)} + \mathbf{M}_2^{(\mu)} + \mathbf{M}_3^{(\mu)}$ through the definition of $\mathbf{M}$, replacing the matrix G with the matrix $G^{(\mu)}$.

Assuming now for notational simplicity that $\alpha = p'+1, \ldots, p-q_2$ satisfy $r_\alpha = 0$, $\ell_{\alpha,1} = \ell_{\alpha,2}$ and that $\alpha = n_\Delta + 1, \ldots, p'$ do not, we may now, defining,

$$h^{(\mu)} = \sum_j (\mathbf{M}_a^{(\mu)})_{jj}(D_{jj} - 1),$$

write equation (4.168) as

$$O_\prec(N^{\epsilon_D n_K}) \prod_{\alpha=n_\Delta+1}^{p'} \mathbb{E} O_\prec(N^{-1/2-\epsilon_D}) \prod_{\alpha=p'}^{p-q_2} h^{(\mu)},$$

which may then, by Young's inequality, be bounded by

$$O_\prec\left(N^{\epsilon_D n_K}\right)\left(O_\prec\left(N^{(-1/2-\epsilon_D)(p-q_2-n_\Delta)}\right) + \mathbb{E}\left|h^{(\mu)}\right|^{p-q_2-n_\Delta}\right).$$

We conclude the proof of Lemma 4.6.2 with the help of the following lemma, proven at the end of this section.

Lemma 4.6.4. *For any fixed number b, we have*

$$\mathbb{E}\left|h^{(\mu)}\right|^b \le C\mathbb{E}|h|^b + O_\prec\left(N^{(-1/2-\epsilon_D)b}\right).$$

□

Proof of Lemma 4.6.3. For ease of illustration, recalling the definition of a in Lemma 4.6.0.1, we show the proof for $a = 2$. The proof is easier otherwise. If $q_2 = 0$, the lemma is trivial. Otherwise, if $r_p = 0$, then

$$\sum_i \left| \prod_{n_p=1}^{r_p+2} \mathbf{p}_{p,n_p} \right| = \sum_i |\mathbf{e}_i^* U(G - \Pi)U^* \mathbf{v}\mathbf{v}^* \mathbf{e}_i| \prec 1$$

because of the presence of two factors $\langle \mathbf{e}_i, \mathbf{W}\mathbf{x}\rangle$ for a fixed vector $\mathbf{x}$ and a matrix $\mathbf{W}$ of bounded operator norm. It is easy to see that this same reasoning holds for

$$\prod_{n_p=1}^{r_p+2} \mathbf{p}_{p,n_p}^{(\ell_{p,n_p})},$$

except that the fixed vector $\mathbf{x}$ might be k_α. Lemma 4.6.3 then follows since $|\langle \mathbf{e}_i, U^*\Pi\mathbf{e}_i\rangle| = O(N^{-\epsilon_D})$ by Proposition 4.2.11.1.

If rather $r_p = 1$ (this is the only other option, since we recall $s_p - r_p = 2$ and $\sum_\alpha s_\alpha = 3$), then

$$\sum_i \left| \prod_{n_p=1}^{r_p+2} \mathbf{p}_{p,n_p} \right| \leq \sum_i \left(\left|2\mathbf{e}_i^* UG\mathbf{e}_i\mathbf{e}_\mu^* GU^*\mathbf{v}\mathbf{v}^*\mathbf{e}_i\right| + \left|2\mathbf{e}_i^* UG\mathbf{e}_\mu\mathbf{e}_i^* GU^*\mathbf{v}\mathbf{v}^*\mathbf{e}_i\right| \right) \prec 1,$$

since we can bound the second term by the same reasoning as above and the first term by Cauchy-Schwarz, since we have one factor $|\langle \mathbf{v}, \mathbf{W}\mathbf{x}\rangle|$ and one factor $O_\prec(N^{-1/2})$, by Lemma 4.2.7 since $\langle \mathbf{e}_\mu, \Pi U^*\mathbf{v}\rangle = 0$. □

Proof of Lemma 4.6.4. By Young's inequality, it suffices to show that $|h - h_\mu| \prec N^{-1/2-\epsilon_D}$.

Whereas the manipulations in Section 4.6.1 for removing dependence on $X\mathbf{e}_\mu$ were somewhat lengthy, the identities for adding it back in are short. The following is a well-known resolvent identity, see eg [KY17] lemma 4.4: for vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_K \cup \mathcal{I}_M}$,

$$G^{(\mu)}_{\mathbf{x}\mathbf{y}} = G_{\mathbf{x}\mathbf{y}} - \frac{G_{\mathbf{x}\mu}G_{\mu\mathbf{y}}}{G_{\mu\mu}}.$$

For ease of illustration, we show this for the case $a = 2$, where a was defined in the statement of Proposition 4.6.0.1. The cases of $a = 1$ and $a = 3$ are respectively trivial and similar. See that, recalling that we are only treating the case $d_1 \asymp 1$,

$$\begin{aligned} h - h^{(\mu)} = O(1) &\sum_j \mathbf{e}_j^* UG\mathbf{e}_\mu\mathbf{e}_\mu^* GU^*\mathbf{v}\mathbf{v}^*\mathbf{e}_j(D_{jj} - 1) \\ &+ O(1) \sum_j \mathbf{e}_j^* U(G - \Pi)U^*\mathbf{v}\mathbf{v}^*\mathbf{e}_j(\widetilde{S}X)^2_{j\mu}, \end{aligned}$$

and by Lemma 4.2.7 this is, since $U^*\mathbf{e}_j$ and $U^*\mathbf{v}$ are orthogonal to $\mathbf{e}_\mu$ and Π is a multiple of the identity,

$$\prec N^{-3/2} \sum_j \mathbf{v}^*\mathbf{e}_j \leq \sqrt{M} N^{-3/2},$$

and we conclude because $M \asymp N$. □

4.7 Proof of Lemma 4.1.8 under Assumption 2

In this section we prove:

Lemma 4.7.1. *If X^0 satisfies assumption 1, X^1 sastisfies Assumption 2 and X^0 and X^1 agree in their first 5 moments, then*

$$\left|\mathbb{E}^{X^0}\left|\mathrm{Tr}_{(\mathcal{I}_M)}(\mathbf{M}(D-I))\right|^2 - \mathbb{E}^{X^1}\left|\mathrm{Tr}_{(\mathcal{I}_M)}(\mathbf{M}(D-I))\right|^2\right| \le CN^{-1-\epsilon}. \tag{4.169}$$

With this, we may prove Lemma 4.1.8 part (2).

Proof of Lemma 4.1.8 part (2). By Lemma 4.8.2, we may choose a matrix X^0 satisfying assumption 1 and which agrees with X^1 in its first 5 moments. Then Lemma 4.169 and Lemma 4.1.8 part (1) allow us to conclude. □

Proof of Lemma 4.7.1. By the same reasoning as in the proof of Lemma 4.1.8 part (1), it suffices to bound, for any integer $m \ge 6$, using the terminology of Section 4.5,

$$N^{-3}\sum_{i\mu}\left|\mathbb{E}\partial_{i\mu}^{m_1}h\partial_{i\mu}^{m_2}h\right| \prec N^{-1-\epsilon} \tag{4.170}$$

for some $\epsilon > 0$, where i ranges over $\mathcal{I}_K \cup \mathcal{I}_M$. Here we used that Assumption 2 ensures that what was called $K_m(X^0_{i\mu}, X^1_{i\mu})$ in Section 4.5 is $O(N^{-3})$ for $m \ge 6$. By Cauchy-Schwarz, it suffices to consider $m_1 = m_2$ (although at that point, it is not necessary that $m_1 + m_2 = 2$); then by the product rule, it suffices to bound, for $m_3 + m_4 = m_1$,

$$N^{-3}\sum_{i\mu}\mathbb{E}\left|\sum_{j\in\mathcal{I}_M}\partial_{i\mu}^{m_3}\mathbf{M}_{jj}\partial_{i\mu}^{m_4}(D_{jj}-1)\right|^2 \prec N^{-1-\epsilon}, \tag{4.171}$$

and this constitutes the remainder of the proof.

First consider $m_4 = 0$. We show the stronger (in that it is uniform in i, μ) statement

$$\mathbb{E}\left|\sum_{j\in\mathcal{I}_M}\partial_{i\mu}^{m_3}\mathbf{M}_{jj}\partial_{i\mu}^{m_4}(D_{jj}-1)\right|^2 \prec N^{-\epsilon}.$$

We obtain

$$\begin{aligned}&\left|\mathbb{E}\sum_j \partial_{i\mu}^{m_3}\mathbf{M}_{jj}\partial_{i\mu}^{m_4}(D_{jj}-1)\right|^2\\ \leq&\sum_{j_1j_2}\sqrt{\mathbb{E}\left(\partial_{i\mu}^{m_4}\mathbf{M}_{j_1j_1}\right)^2\left(\partial_{i\mu}^{m_4}\mathbf{M}_{j_2j_2}\right)^2}\sqrt{\mathbb{E}(D_{j_1j_1}-1)^2(D_{j_2j_2}-1)^2}.\end{aligned}$$

By Lemma 4.2.11, the second factor is $O(N^{-1})$. Referring to equations (4.124) and (4.124) and using Young's inequality, we bound the above by

$$\begin{aligned}&\leq\sum_{j_1j_2}\sqrt{|\langle\mathbf{v},\mathbf{e}_{j_1}\rangle|^4|\langle\mathbf{v},\mathbf{e}_{j_2}\rangle|^4+\mathbb{E}(G-\Pi)^4_{U\mathbf{v}U\mathbf{e}_{j_1}}(G-\Pi)^4_{U\mathbf{v}U\mathbf{e}_{j_2}}}N^{-1}\\ &\leq 2N^{-1}\sum_{j_1j_2}\sqrt{\mathbb{E}(G-\Pi)^4_{U\mathbf{v}U\mathbf{e}_{j_1}}(G-\Pi)^4_{U\mathbf{v}U\mathbf{e}_{j_2}}}+2N^{-1}\sum_{j_1j_2}|\langle\mathbf{v},\mathbf{e}_{j_1}\rangle|^2|\langle\mathbf{v},\mathbf{e}_{j_2}\rangle|^2 \qquad (4.172)\\ &=2N^{-1}\sum_{j_1j_2}\sqrt{\mathbb{E}(G-\Pi)^4_{U\mathbf{v}U\mathbf{e}_{j_1}}(G-\Pi)^4_{U\mathbf{v}U\mathbf{e}_{j_2}}}+2N^{-1}.\end{aligned}$$

Then Lemma 4.2.8 followed by Lemma 4.2.10 gives

$$\begin{aligned}&\leq 2N^{-1-\epsilon}\sum_{j_1j_2}\sqrt{\mathbb{E}(G-\Pi)^4_{U\mathbf{v}U\mathbf{e}_{j_1}}}+2N^{-1}\\ &\leq 2CN^{-2-\epsilon}\sum_{j_1j_2}1+2N^{-1}\\ &\leq 2CN^{-\epsilon}\end{aligned}$$

as desired.

Let us now treat the case $m_4\geq 1$. If $i\in\mathcal{I}_M$, we claim

$$\mathbb{E}\left|\sum_j \partial_{i\mu}^{m_3}\mathbf{M}_{jj}\partial_{i\mu}^{m_4}(D_{jj}-1)\right|^2\prec N^{-\epsilon}. \qquad (4.173)$$

Indeed, using equations (4.128) and (4.142) together with $\left|\widetilde{S}_{ij}\right|=O(\delta_{ij})$ implies $\sum_j \partial_{i\mu}^{m_3}\mathbf{M}_{jj}\partial_{i\mu}^{m_4}(D_{jj}-1)=\partial_{i\mu}^{m_3}\mathbf{M}_{ii}$, and then, referring to equations (4.124) and (4.125),

$$\begin{aligned}\mathbb{E}\left|\partial_{i\mu}^{m_3}\mathbf{M}_{ii}\right|^2&\prec\mathbb{E}|\langle\mathbf{v},UG\mathbf{e}_i\rangle|^2+\mathbb{E}|\langle\mathbf{v},UG\mathbf{e}_\mu\rangle|^2\\ &\prec\mathbb{E}|\langle\mathbf{v},U(G-\Pi)\mathbf{e}_i\rangle|^2+\mathbb{E}|\langle\mathbf{v},U(G-\Pi)\mathbf{e}_\mu\rangle|^2+|\langle\mathbf{v},U\Pi\mathbf{e}_i\rangle|^2\\ &\prec N^{-1}+|\langle\mathbf{v},U\Pi\mathbf{e}_i\rangle|^2\prec N^{-2\epsilon_D},\end{aligned}$$

where in the last line we used Proposition 4.2.11.1.

If rather $m_4 \geq 1$ and $i \in \mathcal{I}_K$, we may repeat the bound in equation (4.172), except that

$$\sqrt{\mathbb{E}(D_{j_1 j_1} - 1)^2 (D_{j_2 j_2} - 1)^2} \prec N^{-1}$$

is replaced with

$$\sqrt{\mathbb{E}4\widetilde{S}_{j_1 i}^2 \widetilde{S}_{j_2 i}^2} \prec N^{-\epsilon} \text{ or } \sqrt{\mathbb{E}4\widetilde{S}_{j_1 i}(\widetilde{S}X)_{j_1 \mu}\widetilde{S}_{j_2 i}(\widetilde{S}X)_{j_2 \mu}} \prec N^{-\epsilon},$$

which follows by equations (4.124) and (4.125) and $\widetilde{S}_{ji} = O(N^{-\epsilon_D})$. Thus as opposed to the bound of $N^{-\epsilon}$ in equation (4.172), we now have

$$\mathbb{E}\left|\sum_{j \in \mathcal{I}_M} \partial_{i\mu}^{m_3} \mathbf{M}_{jj} \partial_{i\mu}^{m_4} (D_{jj} - 1)\right|^2 \prec N^{1-\epsilon} \tag{4.174}$$

for $i \in \mathcal{I}_K$. Equations (4.173) and (4.174) together imply equation (4.171). □

4.8 Proofs of lemmas from Section 4.2

Proof of Lemma 4.2.1. Let $X^0 := X$ satisfy assumption 2 and $\epsilon \in (0, 1/6)$. Define the matrix X^1 by

$$X_{i\mu}^1 = X_{i\mu}^0 1\left(\left|X_{i\mu}^0\right| \leq N^{-\epsilon}\right) \tag{4.175}$$

for some yet to be determined $\epsilon > 0$. See that by Markov's inequality, which we will use a few more times,

$$P\left(\left|X_{i\mu}^0\right| > N^{-\epsilon}\right) \leq N^{6\epsilon}\mathbb{E}\left|X_{i\mu}^0\right|^6 \leq CN^{6\epsilon-3}. \tag{4.176}$$

We therefore have that, doing a union bound on the $O(N^2)$ entries of X^0, that

$$P(X^0 \neq X^1) = O\left(N^{6\epsilon-1}\right). \tag{4.177}$$

The exponent is negative, so that X^0 and X^1 agree asymptotically almost sturely.

X^1 now almost satisfies the assumptions Assumption 2, except that its entries are not quite centered, its variances not quite N^{-1} and its third moments not quite 0. Therefore define $\boldsymbol{\mu} = \mathbb{E}X_{11}^1$ and $\boldsymbol{\sigma} = \left(\mathbb{E}\left(X_{11}^1 - \boldsymbol{\mu}\right)^2\right)^{1/2}$. Using the centeredness of X_{11}^0 and Markov's inequality again,

$$\begin{aligned}
\boldsymbol{\mu} = \mathbb{E}X_{11}^1 &= \int_0^{N^{-\epsilon_1}} \left(P(X_{11}^0 > s) - P(X_{11}^0 < -s)\right) \mathrm{d}s \\
&= \int_{N^{-\epsilon_1}}^{\infty} \left(-P(X_{11}^0 > s) + P(X_{11}^0 < -s)\right) \mathrm{d}s \\
&\leq \int_{N^{-\epsilon_1}}^{\infty} P(\left|X_{11}^0\right| > s)\mathrm{d}s \\
&\leq \int_{N^{-\epsilon_1}}^{\infty} s^{-6}N^{-3}\mathrm{d}s \\
&= O\left(N^{5\epsilon-3}\right).
\end{aligned} \tag{4.178}$$

Similarly, we may compute an estimate of $\boldsymbol{\sigma}$:

$$\begin{aligned}
\boldsymbol{\sigma}^2 &= \mathbb{E}(X_{11}^1)^2 - \boldsymbol{\mu}_{11}^2 \\
&= \int_0^{N^{-\epsilon_1}} sP(|X_{11}| > s)\mathrm{d}s + O(N^{5\epsilon-3}) \\
&= N^{-1} - \int_{N^{-\epsilon_1}}^{\infty} sP(|X_{11}| > s)\mathrm{d}s + O(N^{5\epsilon-3}) \\
&= N^{-1} + O\left(N^{4\epsilon-3}\right) + O(N^{5\epsilon-3}) \\
&= N^{-1} + O(N^{5\epsilon-3}).
\end{aligned} \tag{4.179}$$

Therefore,

$$\boldsymbol{\sigma}^{-1}N^{-1/2} = \sqrt{\frac{N^{-1} + O(N^{5\epsilon-3})}{N^{-1}}} = \sqrt{1 + O(N^{5\epsilon-2})} = 1 + O(N^{5\epsilon-2}). \tag{4.180}$$

Now we define $X_{i\mu}^2 := X_{i\mu}^1 - \boldsymbol{\mu}$, and $X^3 := \frac{N^{-1/2}}{\boldsymbol{\sigma}}X^2$. To demonstrate equation (4.23), we first see that the matrix whose every entry is $\boldsymbol{\mu}$ has operator norm $O(N^{-5\epsilon-2})$, so that $\|X^2 - X^1\| = O(N^{5\epsilon-2})$, and that $\|X\| = O(1)$ (see Definition 4.2.3), so that

$$\left\|X^3 - X^2\right\| = \frac{N^{-1/2}}{\boldsymbol{\sigma}}\left\|X^2\right\| = O(N^{5\epsilon-2})$$

with high probability. That $\left\|\widetilde{S}\right\| = O(d_1)$, that $5\epsilon - 2 < -\frac{1}{2} - \epsilon$, and equation (4.177) allow us then conclude equation (4.23).

To demonstrate equation (4.24), we write, using the notation from the definition of $\mathbf{N}$,

$$\begin{aligned}
&\left\|\mathbf{N}(\widetilde{S}X^3)^{\#} - \mathbf{N}(\widetilde{S}X^1)^{\#}\right\| = \left\|(\mathcal{D}_{\widetilde{S}X^3}\widetilde{S}X^3)^{\#} - (\mathcal{D}_{\widetilde{S}X^1}\widetilde{S}X^1)^{\#}\right\| \\
&\leq \left\|(\mathcal{D}_{\widetilde{S}X^3}\widetilde{S}X^3)^{\#} - (\mathcal{D}_{\widetilde{S}X^3}\widetilde{S}X^1)^{\#}\right\| + \left\|(\mathcal{D}_{\widetilde{S}X^3}\widetilde{S}X^1)^{\#} - (\mathcal{D}_{\widetilde{S}X^1}\widetilde{S}X^1)^{\#}\right\|.
\end{aligned}$$

The first term's bound follows from equation (4.23), since $\left\|\mathcal{D}_{\widetilde{S}X^3}\right\| = \max_i\left\|\mathbf{e}_i^*\widetilde{S}X^3\right\| \leq \|X^3\| = O(1)$ with high probability. The second term's bound, using that $\left\|\mathcal{D}_{\widetilde{S}X^3}\right\| + \left\|\mathcal{D}_{\widetilde{S}X^1}\right\| = O(1)$ with high probability and skipping some routine steps,

$$\leq C\left\|\mathcal{D}_{\widetilde{S}X^3} - \mathcal{D}_{\widetilde{S}X^1}\right\|\left\|\widetilde{S}\right\| \leq Cd_1 \max_i \left(\left\|\mathbf{e}_i^*\widetilde{S}X^3\right\|^{-1} - \left\|\mathbf{e}_i^*\widetilde{S}X^1\right\|^{-1}\right).$$

using $\left\|\mathbf{e}_i^*\widetilde{S}\right\| = 1$ in the first and third inequalities allows us to adjust the above to, perhaps adjusting the constant C,

$$\begin{aligned}
&\leq Cd_1 \max_i \left(\left\|\mathbf{e}_i^*\widetilde{S}X^3\right\| - \left\|\mathbf{e}_i^*\widetilde{S}X^1\right\|\right) \leq Cd_1 \max_i\left\|\mathbf{e}_i^*\widetilde{S}(X^3 - X^1)\right\| \\
&\leq Cd_1 \max_i\left\|X^3 - X^1\right\| = O(d_1)N^{5\epsilon-2},
\end{aligned}$$

and we conclude the proof of Lemma 4.2.1. □

Let us now prove Lemma 4.2.11 from Section 4.2.

Proof. We let i and even p be considered as fixed. We write

$$\begin{aligned}
D_{ii} - 1 = \left\|\mathbf{e}_\alpha^*\widetilde{S}X\right\|^2 - 1 &= \sum_{\mu\in\mathcal{I}_M}\left(\sum_{j\in\mathcal{I}_K\cup\{i\}}\widetilde{S}_{ij}X_{j\mu}\right)^2 - 1 \\
&= \sum_{j\in\mathcal{I}_M}\sum_{j_1,j_2\in\mathcal{I}_K\cup\{i\}}\widetilde{S}_{ij_1}\widetilde{S}_{ij_2}(X_{j_1\mu}X_{j_2\mu} - \mathbf{1}_{j_1=j_2}N^{-1}) \qquad (4.181)\\
&:= \sum_{\substack{j_1,j_2\in\mathcal{I}_K\cup\{i\}\\ \mu\in\mathcal{I}_N}} w_{j_1,j_2,\mu},
\end{aligned}$$

where we used that $\sum_{j\in\mathcal{I}_K\cup\{i\}}\widetilde{S}_{ij}^2 = 1$. Now we can bound the pth moment:

$$\mathbb{E}(D_{ii}-1)^p = \sum_{j_1,\dots,j_{2p}}\sum_{\mu_1,\dots,\mu_p}\mathbb{E}w_{j_1,j_2,\mu_1}\cdots w_{j_{2p-1},j_{2p},\mu_p}, \qquad (4.182)$$

where $j_1, \dots j_{2p}$ each range over $\mathcal{I}_K \cup \{i\}$ and $\mu_1, \dots, \mu_p$ range over $\mathcal{I}_N$. We fix now a partition

$$2^{\{1,\dots,p\}} \ni \mathcal{P} = \{A_1, \dots, A_{|\mathcal{P}|}\}$$

of the set $\{1, \dots p\}$ and consider the sum only over indices $\mu_1, \dots, \mu_p$ which satisfy $\mu_a = \mu_b$ if and only if a and b are identified by $\mathcal{P}$. We say such a partition *satisfies* $\mathcal{P}$ and write $(\mu_1, \dots, \mu_p) \vdash \mathcal{P}$. Since there are only boundedly many such $\mathcal{P}$, this is sufficient. Now if $\mu_1 \neq \mu_2$, then w_{j_1,j_2,μ_1} and w_{j_3,j_4,μ_2} are independent (the converse does not hold, so we are getting a coarser bound than is possible, but only by a bounded factor). Therefore the expectation is 0 if $\mathcal{P}$ contains any singletons. We get

$$\begin{aligned} \mathbb{E}(D_{ii} - 1)^p &= \sum_{j_1,\dots,j_{2p}} \sum_{(\mu_1,\dots,\mu_p)\vdash\mathcal{P}} \mathbb{E} w_{j_1,j_2,\mu_1} \cdots w_{j_{2p-1},j_{2p},\mu_p} \\ &\leq C(K+1)^{2p} N^{|\mathcal{P}|} N^{-\min\{|A_1|,3\}} \cdots N^{-\min\{|A_{\mathcal{P}}|,3\}} \\ &= C(K+1)^{2p} N^{\sum_{a=1}^{|\mathcal{P}|}(1-\min\{|A_a|,3\})} \\ &= C(K+1)^{2p} N^{-|\{a:|A_a|=2\}|-2|\{a:|A_a|\geq 3\}|}. \end{aligned} \tag{4.183}$$

One may observe that since $p \geq 4$, the exponent $-|\{a : |A_a| = 2\}| - 2|\{a : |A_a| \geq 3\}|$ must be less or equal to -2. ☐

Proof of Lemma 4.2.6. Define the auxilliary matrices

$$G_U = \begin{pmatrix} -I & U^*X \\ X^*U & -z \end{pmatrix}^{-1}, \quad G_{U_M} = \begin{pmatrix} -I & U_M^*X \\ X^*U_M & -z \end{pmatrix}^{-1}, \tag{4.184}$$

where $U_M = U \begin{pmatrix} 0 & I_{\mathcal{I}_M} \end{pmatrix}^*$. Note now the identity that for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_K \cup \mathcal{I}_M}$,

$$\mathbf{x}^* G_U \mathbf{y} = \mathbf{x} U^* G U \mathbf{y}$$

where G is as defined in Definition 4.2.5; this may be established through the Schur complement formula. By Theorem 4.2.7, this establishes

$$|\mathbf{x}^* G_U \mathbf{y} - \mathbf{x}^* \Pi \mathbf{y}| \prec |z|^{-1} N^{-1/2}. \tag{4.185}$$

Consider the identity

$$\left(A^{(r)}\right)^{-1}_{st} = A^{-1}_{st} - \frac{A^{-1}_{sr}A^{-1}_{rt}}{A^{-1}_{rr}} \tag{4.186}$$

for $s, t \neq r$, for any invertible matrix A and the minor $A^{(r)}$ of A gotten by removing the rth row and column. Using that G_{U_M} is the minor of G_U gotten by removing the rows and columns indexed by $\mathcal{I}_K$, we may repeatedly apply the identity (4.186) to get that, for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_M}$,

$$\mathbf{x}^* G_{U_M} \mathbf{y} = \mathbf{x}^* G_U \mathbf{y} + O_{\prec}\left(\frac{\max_{\alpha \in \mathcal{I}_K} |\mathbf{x}^* G_U \mathbf{e}_\alpha|^2}{\min_\alpha |\mathbf{e}^*_\alpha G_U \mathbf{e}_\alpha|}\right), \tag{4.187}$$

where we may use equation (4.185) to control the size of the denominators in all resulting fractions. Schur's complement formula also shows that

$$\mathbf{x}^* \mathcal{G}_1 \mathbf{y} = z^{-1} \mathbf{x}^* G_{U_M} \mathbf{y},$$

and then result then follows by equations (4.185) and (4.187). □

Proof of Lemma 4.2.10. First, we prove that

$$\mathbb{E}|(G - \Pi)_{\mathbf{xy}}|^4 \leq CN^{-2}. \tag{4.188}$$

We will then we able to exract an additional factor of $(\kappa + \eta)^{-1}$ in the bound with the Helffer-Sjöstrand argument as follows. Actually we show that

$$\mathbb{E}\left|z^{-1}(G - \Pi)_{\mathbf{xy}}\right|^4 \leq C\left|\frac{N^{-1/2}}{z^2}\right|^4. \tag{4.189}$$

We outline this application of the Helffer-Sjöstrand argument and then return to prove (4.188).

By polarization and linearity, it suffices to treat the case of $\mathbf{x} = \mathbf{y}$. We now require a lemma.

Lemma 4.8.1. $z^{-1}(G - \Pi)_{\mathbf{xx}} := m^{\Delta}(z)$ *is the Stieltjes transform of a compactly supported measure* ρ^{Δ}.

Proof. Noting the form of Π, and using a spectral decomposition, it suffices to show that

$$\frac{1}{-z(1+\lambda S_\mu(z))} \tag{4.190}$$

is the Stieltjes transform of a unit measure for any $\lambda > 0$ whenever S_μ is the Stieltjes transform of a compactly measure μ (we need the result for $S_\mu = m$). We outline the sketch of this proof only. If μ is a linear combination of Dirac masses, it is easy to see that equation (4.190) is a rational function, particularly of the form $\sum_{a=1}^{L} \frac{w_a}{z-x_a}$, and the residue theorem ensures $\sum_{a=1}^{L} w_a = 1$, which concludes the result for such μ. This can be staightforwardly applied to general μ by taking a limit of linear combinations of Dirac masses and applying the results of [GH03]. □

From this point we follow section 3.3 of [BKY16]. We choose a C^∞ bump function which is 1 in an ϵ-neighborhood of the limiting spectrum $[(1-\sqrt{M/N})^2, (1+\sqrt{M/N})^2] \times \{0\} \subseteq \mathbb{C}$ of XX^* and 0 outside a 2ϵ-neighborhood of the limiting spectrum. Letting $f_z(x) = \frac{1}{x-z} - \frac{1}{x_0-z}$, where $x_0 := (1+\sqrt{y})^2$, we may write

$$m^\Delta(z) = \frac{1}{\pi}\int_{\mathbb{C}} f_z(w)\partial_{\bar{w}}\chi(w)m^\Delta(w)\mathrm{d}w \tag{4.191}$$

with high probability. Note that $|f_z(w)| \leq C|z|^{-2}$. We have then by Jensen's inequality and Fubini's theorem

$$\begin{aligned}
\mathbb{E}\left|m^\Delta(z)\right|^4 &= \mathbb{E}\left|\frac{1}{\pi}\int_{\mathbb{C}} f_z(w)\partial_{\bar{w}}\chi(w)m^\Delta(w)\mathrm{d}w\right|^4 \\
&\leq C\mathbb{E}\frac{1}{\pi}\int_{\mathbb{C}}\left|f_z(w)\partial_{\bar{w}}\chi(w)m^\Delta(w)\mathrm{d}w\right|^4 \\
&= C\frac{1}{\pi}\int_{\mathbb{C}}\mathbb{E}\left|f_z(w)\partial_{\bar{w}}\chi(w)m^\Delta(w)\mathrm{d}w\right|^4 \\
&\leq C\max|f_z(w)\partial_{\bar{w}}\chi(w)|\int_{\mathbb{C}}\mathbb{E}\left|m^\Delta(w)\right|^4\mathrm{d}w \\
&\leq C|z|^{-2}N^{-2}
\end{aligned} \tag{4.192}$$

as desired.

Now we proceed with the proof. We will require the following lemma:

Lemma 4.8.2. *Let X^0 satisfy Assumption 2. There is another matrix X^1 which agrees with X^0 in its first five moments and which satisfies $\left|N^{1/2}X^1_{i\mu}\right| \le C$ for a universal constant C.*

We use the same interpolation strategy as in the last section. Without repeating ourselves too much, we may let X^θ interpolate linearly between the laws of X^0 satisfying assumption 1 and X^1 satisfying assumption 2. By Lemma 4.8.2, we may assume that X^0 and X^1 agree in their first 5 moments. As in Section 4.5, it suffices to bound, letting K_m abbreviate $K_m(X^1_{i\mu}, X^\theta_{i\mu}) - K_m(X^0_{i\mu}, X^\theta_{i\mu})$,

$$\sum_{m=6}^{\overline{m}} \sum_{\substack{i\in\mathcal{I}_M\cup\mathcal{I}_K\\ \mu\in\mathcal{I}_N}} K_m \mathbb{E}\left(\partial^{m_1}_{i\mu} G_{\mathbf{xy}}\right)\cdots\left(\partial^{m_q}_{i\mu} G_{\mathbf{xy}}\right)\left(G^{4-q}_{\mathbf{xy}}\right) \prec N^{-2},$$

where $q \le 4$ and $m_1 \le \cdots \le m_q$ and $m_1 + \cdots m_q = m$.

Now see that $\partial_{i\mu} G_{\mathbf{xy}} = -G_{\mathbf{x}i}G_{\mu\mathbf{y}} - G_{\mathbf{x}\mu}G_{i\mathbf{y}}$. Since $G := G(z)$ has operator norm $\prec 1$, we see that the vectors $(G_{\mathbf{z}i} : i \in \mathcal{I}_M)$ and $(G_{\mathbf{z}\mu} : \mu \in \mathcal{I}_N)$ have L^2 norm $\prec 1$; it follows by Cauchy Schwarz that $\sum_{i\mu} \partial^{m_1}_{i\mu} G_{\mathbf{xy}} \prec N$ for $m_1 \ge 1$, since it is a sum of monomials which each contain two factors from the set $\{G_{\mathbf{x}i}, G_{\mathbf{x}\mu}, G_{i\mathbf{y}}, G_{\mu\mathbf{y}}\}$. Thus, using that $K_m = O(N^{-3})$, we get that

$$\begin{aligned}
&\sum_{\substack{i\in\mathcal{I}_M\cup\mathcal{I}_K\\ \mu\in\mathcal{I}_N}} K_m(X^0_{i\mu}, X^1_{i\mu}) \mathbb{E}\left(\partial^{m_1}_{i\mu} G_{\mathbf{xy}}\right)\cdots\left(\partial^{m_q}_{i\mu} G_{\mathbf{xy}}\right)\left(G^{4-q}_{\mathbf{xy}}\right)\\
&\quad = \mathbb{E}\left(O(N^{-3})\sum_{i\mu}\left(\partial^{m_1}_{i\mu} G_{\mathbf{xy}}\right)\cdots\left(\partial^{m_q}_{i\mu} G_{\mathbf{xy}}\right)\right)\left(G^{4-q}_{\mathbf{xy}}\right) \qquad (4.193)\\
&\quad = O(N^{-2})\mathbb{E}\left|G^{4-q}_{\mathbf{xy}}\right|.
\end{aligned}$$

Lemma 4.2.8 yields $\mathbb{E}|G_{\mathbf{xy}}|^p \prec 1$ for any p, and we conclude. □

Proof of Lemma 4.8.2. We refer to the results of [CF91]. The idea is that all distributions may be matched in 5 moments by an at most 3-atomic distribution, and that the support of this distribution is a continuous function of the first 5 moments.

Let $\widetilde{X}$ be the matrix satisfying Assumption 3 from which X^0 results after standardization. If the first 5 moments $\widetilde{\gamma}_1, \ldots, \widetilde{\gamma}_5$ of the distribution of $N^{1/2}\widetilde{X}$ lead to a singular Hankel matrix, then by their Theorem 3.8 the distribution is finitely atomic and in particular bounded, so that the truncation of Lemma 4.2.1 does not change $\widetilde{X}$ for sufficiently large N, ie $\widetilde{X} = X^0$, and we may then also take $X^1 = X^0$.

If rather the 5 moments lead to a non-singular Hankel matrix, then by Theorem 3.1, there exists a 3-atomic distribution $\widetilde{P}_*$ whose first 5 moments are $\widetilde{\gamma}_1, \ldots, \widetilde{\gamma}_5$ and whose atoms are the roots of a cubic polynomial $\widetilde{p}$ which is a continuous function of $\widetilde{\gamma}_1, \ldots, \widetilde{\gamma}_5$. Since the atom distribution of X^0 converges to that of $\widetilde{X}$, the Hankel matrix of X^0 is also non-singular for large N, and the polynomial p associated to X^0 converges to $\widetilde{p}$, and thus the roots of p, which are the atoms of an atomic distribution matching X^0 to 5 moments, converge to the roots of $\widetilde{p}$, which concludes the proof.

□

Proof of Theorem 4.3.5. We outline only the key steps of how to adapt out version of the Theorem from the version in [BDW20]. First we prove equation (4.62). Note that because of our strong assumption (2.2), all the contour integrals in [BDW20] may be easily estimated by the triangle inequality for contour integrals rather than the residue theorem, since the contours may be all drawn a distance $\asymp 1$ from any singularities of the matrix they call $L(z)$. The proof begins by writing

$$|\langle \mathbf{w}, u_\alpha(\mathcal{V})\rangle|^2 = S_1 + S_2 + S_3 + R_0 \tag{4.194}$$

(we have called R_0 what they have called R to avoid a clash of notation), which is equation (5.14) of [BDW20]. $S_1 + S_2$ is identically equal to the first four terms of equation (4.62). Equation (5.38) of [BDW20] gives

$$S_3 = \widetilde{S}_3 + O_\prec(\mathcal{R}_3), \tag{4.195}$$

where $\widetilde{S}_3$ is identically the last two non-error terms of equation (4.62) and where equation (5.38) of [BDW20] reads that $\mathcal{R}_3$ is of the form

$$\mathcal{R}_3 = O_\prec(\mathrm{proj}_{\mathbf{v}_1,\ldots,\mathbf{v}_K}\mathbf{w})N^{-1}. \tag{4.196}$$

It may easily verified that the factor of N^{-1} comes directly from two factors of $\Xi_{\mathbf{xy}}$ for deterministic vectors of bounded length $\mathbf{x}, \mathbf{y}$. Thus under either Assumption 1 or 2, using Lemma 4.2.10, we see that $\mathcal{R}_3$ may be absorbed into $\mathcal{E}_1$.

Regarding the term R_0, equation (5.45) of [BDW20] shows that every term of R_0 has three factors of Ξ. Thus, using Lemma 4.2.10, we may absorb all of R_0 into the error term $\mathcal{E}_2$.

For the proof of equation (4.64), we look at equation (5.15) of [BDW20], which is easily seen to hold for $\Sigma = \widetilde{S}^{\#}$ with minor necessary modifications, with $\widetilde{\boldsymbol{w}} := (\widetilde{S}^{\#})^{-1/2}\mathbf{e}_i$. The term S_1 is deterministic and we may call it $C_{i,\alpha}$. The terms S_2, S_3, and R_0 are all easily seen to be $O(\Xi_{\mathbf{xy}} \|\widetilde{\boldsymbol{w}}^* V\|)$, for some deterministic vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_M}$ and where V is a deterministic $\mathcal{I}_M \times \mathcal{I}_K$ matrix; from this it is easy to see, using Lemmas 4.2.10 and 4.2.6, that $\sum_{i\in\mathcal{I}_M} \sqrt{\mathbb{E}|S_2 + S_3 + R_0|^2} = O(1)$.

The proof of Theorem 4.3.5 is concluded. □

4.9 Large Deviation Bounds

So far we have mainly been concerned with the distribution of the spiked eigenvalues of $\mathcal{R}$, although we have proven that, under assumption 1,

$$\left|\lambda_\alpha(\mathcal{R}) - \lambda_\alpha(\widetilde{\mathcal{V}})\right| \prec d_1^{-1} N^{-1/2-\epsilon}.$$

This was enough to establish the distribution of $\lambda_\alpha(\mathcal{R})$, which is on the scale $N^{-1/2}$, and we also relied crucially on the result of [BJ21], which established the distribution for $\lambda_\alpha(\widetilde{\mathcal{V}})$.

We now want a large deviation bound.

Lemma 4.9.1. *The matrix $\mathcal{R}$ satisfies*

$$|\lambda_\alpha(\mathcal{R}) - \phi_{N,\alpha}| \prec N^{-1/2}$$

for $\alpha = 1, \ldots, K$.

Proof. By Theorem 4.1.4, the remaining necessary ingredient is a corresponding result for $\widetilde{\mathcal{V}}$, ie,

$$\left|\lambda_\alpha(\widetilde{\mathcal{V}}) - x_\alpha\right| \prec N^{-1/2}$$

for some deterministic real number x_α, which we establish in this section (it will then follow that x_α can be taken as ϕ_α by Lemma 4.1.1).

First note that if we let $\mathcal{J}$ be a diagonal matrix defined through $\mathcal{J}_{ii} = \|\mathbf{e}_i^* S\|^{-1}$, then

$$\widetilde{S} = \mathcal{J}\begin{pmatrix} B & I \end{pmatrix} = \begin{pmatrix} \mathcal{J}B & \mathcal{J} \end{pmatrix} = \begin{pmatrix} \mathcal{J}B & I \end{pmatrix}\begin{pmatrix} I & \\ & \mathcal{J} \end{pmatrix}.$$

Define $\Sigma = \begin{pmatrix} \mathcal{J}B & I \end{pmatrix}^{\#}$. Pick a matrix O with orthonormal rows so that $\begin{pmatrix} \mathcal{J}B & I \end{pmatrix} = \Sigma^{1/2}O$. We have now that $\Sigma = I + (\mathcal{J}B)^{\#}$ and

$$\widetilde{\mathcal{V}} = \widetilde{S}XX^*\widetilde{S}^* = \Sigma^{1/2}\left(O\begin{pmatrix} I & \\ & \mathcal{J} \end{pmatrix}X\right)^{\#}\Sigma^{1/2}.$$

Lemma 3.10 of [BKY16] then shows that we can characterize $\lambda_1(\widetilde{\mathcal{V}}), \ldots, \lambda_K(\widetilde{\mathcal{V}})$ as solutions to

$$\det\left([d]^{-1} + W(x)\right) = 0,$$

where

$$\begin{aligned}
W(x) &:= \sqrt{y}V^*\left(I + zG^{\mathcal{J}}(z)\right)V \\
G^{\mathcal{J}}(z) &:= \left(H^{\mathcal{J}} - zI\right)^{-1}, \quad H^{\mathcal{J}} := \left(O\begin{pmatrix} I & \\ & \mathcal{J} \end{pmatrix}X\right)^{\#} \\
[d] &:= \begin{pmatrix} \lambda_1\left((\mathcal{J}B)^{\#}\right) & & \\ & \ddots & \\ & & \lambda_K\left((\mathcal{J}B)^{\#}\right) \end{pmatrix} \\
V &:= \begin{pmatrix} | & & | \\ \mathbf{v}_1\left((\mathcal{J}B)^{\#}\right) & \cdots & \mathbf{v}_K\left((\mathcal{J}B)^{\#}\right) \\ | & & | \end{pmatrix},
\end{aligned}$$

where $\mathbf{v}_\alpha\left((\mathcal{J}B)^\#\right)$ is the eigenvector of $(\mathcal{J}B)^\#$ corresponding to the α^{th} largest eigenvalue, provided that the solution x is not also an eigenvalue of $H^{\mathcal{J}}$.

Now the main results of [KY17] apply to the matrix $H^{\mathcal{J}}$ (with the same extra argument as in the proof of Lemma 4.2.10 to get the right decay in z), that is, $G^{\mathcal{J}}$ has a deterministic limit $\Pi^{\mathcal{J}}$ and for any determinisitic unit vectors $\mathbf{x}, \mathbf{y} \in \mathcal{I}_M$, we have

$$\left(G^{\mathcal{J}} - \Pi^{\mathcal{J}}\right)_{\mathbf{xy}} \prec |z|^{-2} N^{-1/2}$$

uniformly for $z \geq \operatorname{supp} \varrho^{\mathcal{J}} + C$ for any fixed constant C, where $\varrho^{\mathcal{J}}$ is the limiting measure for the emperical spectral measure of $G^{\mathcal{J}}$.

Now, because $\|\mathcal{J} - I\| = O(N^{-\epsilon_D})$ and because the nonzero eigenvalues of $(\mathcal{J}B)^\#$ are well-separated by equation (2.2), it is not hard to see that

$$\det\left([d]^{-1} + W_{\det}(x)\right),$$

where $W_{\det}$ is defined through the definition of W, replacing the matrix $G^{\mathcal{J}}$ with $\Pi^{\mathcal{J}}$, has K zeroes $x_1, \ldots, x_K$ which are each $\geq \operatorname{supp} \varrho^{\mathcal{J}} + C$.

Fix now $\alpha \in \{1, \ldots, K\}$ and $\epsilon > 0$ and consider the contour

$$\gamma = \{z : |z - x_\alpha| = x_\alpha N^{-1/2+\epsilon}\}.$$

It is also not hard to see that on γ,

$$W_{\det} \asymp x_\alpha N^{-1/2+\epsilon} W'_{\det}(x_\alpha) \asymp x_\alpha^{-1} N^{-1/2+\epsilon}.$$

By Lemma 4.2.7, we have that on γ, with high probability,

$$|W(z) - W_{\det}(z)| \leq x_\alpha^{-1} N^{-1/2+\epsilon/2},$$

so that by Rouche's theorem, $W_{\det}$ has a zero λ_α in $\{|z - x_\alpha| \leq N^{-1/2+\epsilon}\}$ with high probability. Since λ_α is an eigenvalue of the deterministic matrix $\widetilde{S}XX^*\widetilde{S}^*$, it is real. By Cauchy interlacing, $\widetilde{S}XX^*\widetilde{S}$ may only have $\leq K$ eigenvalues greater that $\operatorname{supp} \varrho^{\mathcal{J}}$, so that we have proven that each eigenvalue λ of $\widetilde{S}XX^*\widetilde{S}^*$ is within $d_1 N^{-1/2+\epsilon}$ of x_α with high probability. Since ϵ was arbitrary, we may now conclude. □

CHAPTER 5

Non-Spiked Eigenvalues

5.1 Proof Strategy for Theorem 2.2.4

Definition 5.1.1 (The spectral region $\mathbf{S}^e$)**.** Fix $\epsilon_0 = \min\{1/6, \epsilon_D/2\}$. We let

$$\mathbf{S}^e := \left\{z = E + i\eta_0 : \kappa(z) < N^{-2/3+\epsilon_0}\right\}, \quad , \eta_0 := N^{-2/3-\epsilon_0}.$$

We will use, in the course of the proof, three different forms of the Green function of a matrix, which will each have their different utilities. A matrix G^A without a subscript will represent a resolvent of the form $(AXX^*A^* - z)^{-1}$. A matrix G_1^A will represent a (generalized) resolvent of the form $\begin{pmatrix} -I & AX \\ X^*A^* & -z \end{pmatrix}^{-1}$. Finally a matrix G_2^A will represent a resolvent of the form $\begin{pmatrix} -(AA^*)^{-1} & X \\ X^* & -z \end{pmatrix}^{-1}$. We will write "generalized entries" of general matrices A as $A_{\mathbf{xy}} := \mathbf{x}^*A\mathbf{y}$, and generalized entries of resolvents with subscipts as $\mathbf{x}^*G_1^A\mathbf{y} =: G_{1,\mathbf{xy}}^A =: G_{1,\mathbf{x},\mathbf{y}}^A$. When appearing as an index of a resolvent entry, indices $s \in \mathcal{I}$ (whose definition we recall below) will now represent the standard basis vector $\mathbf{e}_s$, so that for example $G_{1,s\mathbf{y}}^A = G_{1,\mathbf{e}_s\mathbf{y}}^A$.

Recall the index sets

$$\mathcal{I}_K, \quad \mathcal{I}_M, \quad \mathcal{I}_{K+M} := \mathcal{I}_K \cup \mathcal{I}_M, \quad \mathcal{I}_N := \mathcal{I}_{K+M} \cup \mathcal{I}_N.$$

Let us record a quick lemma, which follows from Schur's complement formula, regarding how to transition between the different resolvents.

Lemma 5.1.2. *For any conformable, invertible, positive definite matrix A and $i, j \in \mathcal{I}_{K+M}$, we have*

$$z^{-1}\left(A^{-1}G_2^A A^{-1}\right)_{ij} = z^{-1}\left(G_1^A\right)_{ij} = G_{ij}^A.$$

Recall the matrices $B, \widetilde{S}, \widetilde{B}$ and $\mathcal{J}$ from equations (2.4) and (2.14) and just above equation (2.1). Begin by noting that

$$\mathbf{N}\left(\widetilde{S}X\right) = \begin{pmatrix} D_M B & I \end{pmatrix} DX,$$

where $D = \begin{pmatrix} I_{\mathcal{I}_K} & \\ & D_M \end{pmatrix}$ is a $(\mathcal{I}_K \cup \mathcal{I}_M) \times (\mathcal{I}_K \cup \mathcal{I}_M)$ diagonal matrix, and D_M is defined through

$$(D_M)_{ii} := \left\|\mathbf{e}_i^* \widetilde{S} X\right\|^{-1} \mathcal{J}_{ii} \text{ for } i \in \mathcal{I}_M.$$

This definition of D is related to but independent of our definition in Chapters 4 and 6. We also define

$$E := \begin{pmatrix} I & \\ & E_M \end{pmatrix},$$

$$E_M := \left(I_M + \mathcal{J}^{-2} \operatorname{diag}\left(\left\|\mathbf{e}_1^* \widetilde{B} X\right\|^2, \cdots, \left\|\mathbf{e}_M^* \widetilde{B} X\right\|^2\right)\right)^{-1/2},$$

$$\text{and } \Lambda := D^{-2} - E^{-2}.$$

E the leading part of D and is a "better" matrix than D because it only depends on the randomness in the first $\mathcal{I}_K$ rows of X, so that as opposed to the matrix product DX, the matrix product EX has much smaller overlap between the source of randomness on which E depends and the source of randomness on which X depends. Our argument would not work if E were deterministic; the dependence of E on X_K, and hence the dependence of the leading eigenvalues of $\mathcal{R}$ on X_K, is a crucial part of this work.

Define the function $m_E(z)$ as in [KY17] through the equation

$$\frac{1}{m_E} = -z + y \int \frac{x}{1 + m_E x} \pi(\mathrm{d}x) \tag{5.1}$$

where π is the eigenvalue measure for E, and we recall y is the dimensional ratio $\frac{M}{N}$. The function m_E is the Stieltjes transform of a measure ϱ_E.

We define $\Pi_2 := \Pi_2(z)$ through

$$\Pi_2 := \begin{pmatrix} -E(1+m_E E)^{-1} & 0 \\ 0 & m_E \end{pmatrix}$$

as well as, independently of our previous definition of Π in chapter 4,

$$\Pi := -z^{-1}(1+m_E E)^{-1}, \quad \Pi_1 := \begin{pmatrix} -(I+m_E E)^{-1} & 0 \\ 0 & m_E I \end{pmatrix}.$$

Π_1, Π_2, and Π satisfy the same conclusion as G_1, G_2, and G in Lemma 5.1.2, that is,

$$z^{-1}\left(E^{-1}\Pi_2 E^{-1}\right)_{ij} = z^{-1}\left(\Pi_1\right)_{ij} = \Pi_{ij}.$$

Unusually for RMT literature, the "deterministic limits" m_E and Π are actually random, depending on X_K. For much of the reasoning in this paper, however, X_K can be thought of as fixed and deterministic. Moreover, note

$$\Pi_2 = \begin{pmatrix} -1/(1+m_E) & \\ & m_E \end{pmatrix} + O_\prec(N^{-1/2}),$$

where the error term is to be taken in an operator norm sense, so that $\Pi_{\mathcal{I}_{K+M}\times\mathcal{I}_{K+M}}$ and $\Pi_{\mathcal{I}_N\times\mathcal{I}_N}$ are nearly isotopic . Π is close enough to isotropy to avoid most potential difficulties arising from anisotropy, but far enough away to cause the non-universality phenomenon described in the remarks following Theorem 2.2.4.

Using notation independent of that of chapters 4, and 6, we also define the left singular vectors $\widetilde{\mathbf{v}}_1, \ldots, \widetilde{\mathbf{v}}_K \in \mathbb{R}^{\mathcal{I}_M}$ of $D_M B$, which may then be completed to an orthonormal basis $\widetilde{\mathbf{v}}_1, \ldots, \widetilde{\mathbf{v}}_M$ of $\mathbb{R}^{\mathcal{I}_M}$), the right singular vectors $\widetilde{\mathbf{w}}_1, \ldots, \widetilde{\mathbf{w}}_K \in \mathbb{R}^{\mathcal{I}_K}$, and the squared singular values $\widetilde{d}_a, \ldots, \widetilde{d}_K$. We define a $(\mathcal{I}_K \cup \mathcal{I}_M) \times \mathcal{I}_M$ matrix with orthonormal columns

$$U_M := \begin{pmatrix} \sum_{\alpha=1}^K \sqrt{\frac{\widetilde{d}_\alpha}{\widetilde{d}_\alpha+1}}\widetilde{\mathbf{v}}_\alpha \widetilde{\mathbf{w}}_\alpha^* & \sum_{\alpha=1}^M \sqrt{\frac{1}{\widetilde{d}_\alpha+1}}\widetilde{\mathbf{v}}_\alpha \widetilde{\mathbf{v}}_\alpha^* \end{pmatrix}^*.$$

Also define an $\mathcal{I}_{K+M} \times \mathcal{I}_{K+M}$ orthogonal matrix U so that $U_M^* = \begin{pmatrix} 0 & I_{\mathcal{I}_{K+M}} \end{pmatrix} U^*$.

Four main lemmas, which compare the eigenvalues of $(E_M X_M)^\#$ to those of $(EX)^\#$, and then to those of $(DX)^\#$ (which are equal to those of $(UDX)^\#$), and then to those of $(U_M DX)^\#$, and then finally to those of $\mathcal{R}$, naturally comprise the steps of the proof. We collect the necessary assumptions as follows:

Assumption 4. *B satisfies equations* (2.2) *and* (2.3) *and X satisfies Assumption 1.*

Lemma 5.1.3. *Fix $\alpha \geq 1$ and let Assumption 4 hold. We have*

$$\left|\lambda_\alpha\left((E_M X_M)^\#\right) - \lambda_\alpha\left((EX)^\#\right)\right| \leq N^{-2/3-\epsilon_0/2} \tag{5.2}$$

with probability $1 - O(N^{-\phi})$ *for some* $\phi > 0$.

Lemma 5.1.4. *Fix $\alpha \geq 1$ and let Assumption 4 hold. We have*

$$\left|\lambda_\alpha\left((EX)^\#\right) - \lambda_\alpha\left((DX)^\#\right)\right| \leq N^{-2/3-\epsilon_0/2} \tag{5.3}$$

with probability $1 - O(N^{-\phi})$ *for some* $\phi > 0$.

Lemma 5.1.5. *Fix $\alpha \geq 1$ and let Assumption 4 hold. We have*

$$\left|\lambda_\alpha\left((U^* DX)^\#\right) - \lambda_\alpha\left((U_M^* DX)^\#\right)\right| \leq N^{-2/3-\epsilon_0/2} \tag{5.4}$$

with probability $1 - O(N^{-\phi})$ *for some* $\phi > 0$.

Lemma 5.1.6. *Fix $\alpha > K$ and let Assumption 4 hold. We have*

$$\left|\lambda_\alpha\left((U_M^* DX)^\#\right) - \lambda_{\alpha-K}\left(\mathcal{R}\right)\right| \leq N^{-2/3-\epsilon_0/2} \tag{5.5}$$

with probability $1 - O(N^{-\phi})$ *for some* $\phi > 0$.

Proof of Theorem 2.2.4. Theorem 2.2.4 follows immediately from Lemmas 5.1.3, 5.1.4, 5.1.5, and 5.1.6. □

5.1.1 Heuristic argument and additional technical results

We will give the heuristic argument in the reverse of the logical order in which the main result must be proven; ie, we start with the matrix $\mathcal{R}$ in which we are interested, and we reason towards the more understandable matrix $(E_M X_M)^{\#}$.

The first step in understanding the eigenvalues of $\mathcal{R}$ is to view it as a bounded rank perturbation of a "non-spiked matrix", leading us to perform an argument very similar to that of [BKY16]; indeed, we have

$$\mathcal{R} = \left(\sqrt{1 + (D_M B)^{\#}} U_M^* D X\right)^{\#},$$

which is the reason for the definition of U_M. One difficulty here is that because of the normalization $\mathbf{N}$ that we apply in passing to $\mathcal{R}$, the bounded rank perturbation $1 + (D_M B)^{\#}$ is random and not independent of the randomness X, and moreover, the non-spiked matrix $(U_M^* D X)^{\#}$ does not have trivial covariance structure, but rather the matrix $U_M^* D$ is random and not independent of X. Other than this, the argument is very similar to the one in [BKY16]—the crucial ingredient there, an isotropic local law for the resolvent of the non-spiked matrix, is replaced in our setting with a necessarily weaker isotropic local law for *random* generalized entries. The main ingredient for this is the isotropic law for $(DX)^{\#}$:

Lemma 5.1.7. *Fix deterministic unit vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_{K+M}}$. Under Assumption 4 we have, uniformly for $z \in \mathbf{S}^e$,*

$$\left|\mathbf{x}^*(G^D - \Pi)\mathbf{y}\right| \prec \psi_{(\mathbf{xy})}. \tag{5.6}$$

The control parameter $\psi_{(\mathbf{xy})}$ is defined in equation (5.13).

Having reduced our study to eigenvalues of the matrix $(U_M^* D X)^{\#}$, we wish to further reduce to the eigenvalues of the matrix $(U^* D X)^{\#}$, because at this point the relevance of U will disappear—conjugation by an orthgonal matrix does not change a matrix's eigenvalues, so we will really at this point have reduced to the eigenvalues of $(DX)^{\#}$. We do this now by comparing the matrices' resolvents; we will see in Section 5.3 how closeness of resolvents translates to closeness of eigenvalues at the spectral edge. Since $U_M^* D$ is just $U^* D$ with a few rows removed, we may compare the resolvents $G_1^{U_M^* D}$ and $G_1^{U^* D}$ with standard resolvent identities (see Lemma 5.2.4). The main technical ingredient for this is again Lemma 5.1.7.

Now we must bound the difference between the eigenvalues of $(DX)^\#$ and $(EX)^\#$, and again we do this by comparing the resolvents. A resolvent expansion yields that

$$G_2^D = G_2^E + G_2^E(D^{-2} - E^{-2})G_2^E + G_2^E(D^{-2} - E^{-2})G_2^E(D^{-2} - E^{-2})G_2^E + \cdots .$$

One of our main contributions (the results described in this paragraph constitute most of the technical novelty of this chapter) is to show that the polynomialization method of [BEK14] can treat such polynomials of resolvents quite efficiently; the randomness of the matrix $(D^{-2} - E^{-2})$ is an added difficulty (as opposed to if it was deterministic, which would place us close to the setting of [CES21b]), but we are still able to treat it with suitable adaptations of the polynomialization method. One technical ingredient to make this work is the following isotropic local law for $(EX)^\#$:

Lemma 5.1.8. *Fix deterministic unit vectors in* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_{K+M}}$*. Under Assumption 4 we have, uniformly for* $z \in \mathbf{S}^e$*,*

$$\left| G_{2,\mathbf{xy}}^E - \Pi_{2,\mathbf{xy}} \right| \prec \psi_{\mathbf{xy}}. \tag{5.7}$$

As we described earlier, $(EX)^\#$ is a much nicer object to work with than $(DX)^\#$ since it is much more decoupled. We can completely decouple the randomness in the population from the randomness in X by finally comparing the eigenvalues of $(EX)^\#$ with those of $(E_M X_M)^\#$, which is done in much the same way as the comparison of $G_1^{U_M D}$ to G_1^{UD}, using Lemma 5.1.8 as input. We can describe the extreme eigenvalues of this matrix $(E_M X_M)^\#$ thanks to the results of [LS16].

5.2 Tools

Eigenvalues of random matrices tend to exhibit a phenomenon called *level repulsion*, whereby only with low probability can eigenvalues be much closer than the average spacing between eigenvalues.

We will require a sort of level repulsion condition as a technical input in order to translate statements about closeness of resolvents into statements about closeness of eigenvalues. It also makes the proof of Lemma 5.1.6 easier, although it is not a requirement (see section 6 of [KY13b])

Definition 5.2.1 (Weak level repulsion at the edge)**.** We say that a random matrix ensemble A satisfies *weak level repulsion at the edge* if for any fixed integer $L > 0$, the eigenvalues $\lambda_\alpha(A)$ satisfy

$$\mathbb{P}\left(\min_{\alpha \le L} |\lambda_\alpha(A) - \lambda_{\alpha+1}(A)| \le CN^{-2/3-\epsilon_0/4}\right) \le N^{-\phi} \tag{5.8}$$

for some absolute constant C and $\phi > 0$.

The matrix G^{E_M}, since the matrix E_M is independent of the randomness X_M, is well-understood. We collect certain facts about it in the following lemma.

Lemma 5.2.2. *Define the high probability event Ω_{reg} on which $|E_{ii} - 1| \le N^{-1/10}$. On Ω_{reg}, the following hold:*

1. *(Isotropic local law) We have for deterministic unit vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}}$ and $z \in \mathbf{S}^e$,*

$$\left|\mathbf{x}^* \left(G^{E_M} - \Pi\right) \mathbf{y}\right| \prec \psi(z) \tag{5.9}$$

   where

$$\psi := \psi(z) := \sqrt{\frac{\Im m(z)}{N\eta_0}} + \frac{1}{N\eta_0}. \tag{5.10}$$

2. *(Weak level repulsion) $(E_M X)^{\#}$ satisfies the weak level repulsion condition 5.2.1.*

3. *(Local Eigenvalue Law) For any interval I, we have*

$$\frac{1}{M}\#\left\{\alpha : \lambda_\alpha\left((E_M X_M)^\#\right) \in I\right\} = \varrho(I) + O_\prec\left(N^{-1}\right) \tag{5.11}$$

Recall that the classical location γ_α of λ_α is the unique value such that $\int_{\gamma_\alpha}^\infty \mathrm{d}\varrho_E = \frac{\alpha}{M}$.

Proof. Let $[a_2, a_1]$ be the right-most connected component of ϱ_E. It is easy to verify that on Ω_{reg}, the bulk component $[a_2, a_1]$ is *regular* in the sense of [KY17] definition 2.7, as is the edge a_1.

1. This is Theorem 3.14 of [KY17].

2. The result for $X^\#$ rather than $(E_M X)^\#$ is Proposition 6.3 of [BKY16]. The corresponding result for $(E_M X)^\#$ follows from the result for $X^\#$ in the same way as Proposition 2.4 of [KY13a]; the key ingredient for the proof of Proposition 2.4 of [KY13a], Lemma 2.6 in [KY13a], is instead given by Proposition 4.1 of [LS16] (whose equation (4.4) is a clear parallel to equation (2.13) of [KY13a]; the corresponding parallel to equation 2.14 of [KY13a] is proven in the same way as equation (4.4) of [LS16] and is in fact easier).

3. This follows from equation (3.11) of [KY17] in the same way that local eigenvalue laws usually follow from averaged local laws for the resolvent; see e.g. [BK18].

□

Note that for $z \in \mathbf{S}^e$, we have

$$\psi \lesssim \sqrt{\frac{\sqrt{\kappa + \eta_0}}{N\eta_0}} \lesssim \sqrt{\frac{N^{-1/3+\epsilon_0}}{NN^{-2/3-\epsilon_0}}} = N^{-1/3+\epsilon_0} \tag{5.12}$$

Let $\delta := \min\left\{\epsilon_D, \frac{1}{24}\right\}$. In analogy to the control parameter ψ, which is $O_\prec(N^{-1/3+\epsilon_0})$ on

$\mathbf{S}^e$, we define a new family of control parameters:

$$\psi_{(\mathbf{xy})} := \begin{cases} N^{-1/3+\epsilon_0} & \mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_M \cup \mathcal{I}_N} \\ N^{-1/6-\delta} & \text{one of } \mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_M \cup \mathcal{I}_N} \\ N^{-3\delta} & \mathbf{x}, \mathbf{y} \notin \mathbb{R}^{\mathcal{I}_M \cup \mathcal{I}_N}. \end{cases} \tag{5.13}$$

The role of $\psi_{(\mathbf{xy})}$ is to replace ψ in isotropic local law statements like equation (5.9) for, for example, G_2^E. For such a resolvent we will have weaker isotropic bounds when $\mathbf{x}, \mathbf{y}$ have components in $\mathbb{R}^{\mathcal{I}_K}$.

Also define $\psi_{(st)} = \psi_{(\mathbf{e}_s \mathbf{e}_t)}$ for $s, t \in \mathcal{I}$, and also define $\psi_{(KM)} := \psi_{(st)}$ for $s \in \mathcal{I}_K, t \in \mathcal{I}_M$, and similarly $\psi_{(KK)}, \psi_{(sM)}, \psi_{(\mathbf{x},M)}$, etc., in the same way.

Fundamental to the analysis of this paper are the following identities for the resolvents G_1^A and G_2^A.

Definition 5.2.3. If $\mathcal{K} \subseteq \mathcal{I}_{K+M} \cup \mathcal{I}_N$ is an index set, H is a $\mathcal{K} \times \mathcal{K}$ matrix, and if $T \subseteq \mathcal{K}$, then we define

$$H^{(T)} := (H_{st})_{s,t \in \mathcal{K} \setminus T}$$

and we define, for $\iota = 1, 2$,

$$G_\iota^{A^{(T)}} = \left(\left(\left(G_\iota^A \right)^{-1} \right)^{(T)} \right)^{-1}.$$

We use, for $T \subseteq \mathcal{K}$ and $r_1, \ldots, r_L \in \mathcal{K} \setminus T$, the shorthand $G^{A^{(\{r_1, \ldots, r_L\})}} =: G^{A^{(r_1 \cdots r_L)}}$ and $G^{A^{(T \cup \{r_1\})}} =: G^{A^{(T r_1)}}$.

Lemma 5.2.4. *For $\iota = 1, 2$, and for $r, s, t \in \mathcal{K}$, $s, t \neq r$,*

$$G_{\iota, st}^{A^{(r)}} = G_{\iota, st}^A - \frac{G_{\iota, sr}^A G_{\iota, rt}^A}{G_{\iota, rr}^A}.$$

If $\mu, \nu \in \mathcal{I}_N$, then

$$G_{2, \mu\nu}^A = G_{2, \mu\mu}^A G_{2, \nu\nu}^{A^{(\mu)}} \left(X^* G_2^{A^{(\mu\nu)}} X \right)_{\mu\nu}.$$

If A is diagonal and $i, j \in \mathcal{I}_{K+M}$, then

$$G_{2, ij}^A = G_{2, ii}^A G_{2, jj}^{A^{(i)}} \left(X G_2^{A^{(ij)}} X^* \right)_{ij}.$$

If A is diagonal and $i \in \mathcal{I}_{K+M}, \mu \in \mathcal{I}_N$, then

$$G^A_{2,i\mu} = G^A_{2,ii} G^{A^{(i)}}_{2,\mu\mu} \left(-X_{i\mu} + \left(X G_2^{A^{(i\mu)}} X \right)_{i\mu} \right).$$

Proof. This is a consequence of lemma 4.4 of [KY17]. □

Let $\sigma(H)$ denote the spectrum of a matrix H. Also let

$$\underline{A} := M_0^{-1} \operatorname{Tr} A \tag{5.14}$$

denote the normalized of an $M_0 \times M_0$ matrix A.

Recall that for a $M_0 \times M_0$ Hermitian matrix H, the function $x \mapsto \Im \underline{(H - (x + i\eta))^{-1}}$ is equal to $\delta_H * \theta_\eta$, where $\delta_H = \frac{1}{M_0} \sum \delta_{\lambda_i(H)}$ and $\theta_\eta(x) = \frac{\eta^{-1}}{(x/\eta)^2+1}$ is an approximate δ function for small η. In order to infer statements about the eigenvalues of H from $\Im \underline{(H - (x + i\eta))^{-1}}$, we would like to have, roughly, that if η is smaller than the typical separation of eigenvalues around x, then the graph of $\underline{G^H}$ looks essentially flat with sharp, defined peaks at each eigenvalue. Put another way, $\delta_H * \theta_\eta(x)$ is always large when $x \in \sigma(H)$, but we want the converse to hold as well. The following definition and lemma establish the validity of this heuristic at the edge for $(E_M X)^\#$. It is, in addition to the weak level repulsion condition, the second technical ingredient we will need for translating closeness of resolvents to closeness of eigenvalues.

Definition 5.2.5 (The valley condition)**.** Let $\epsilon > 0$ and $\mathcal{K} \subseteq \mathcal{I}_{K+M}$ have $|\mathcal{I}_{K+M} \setminus \mathcal{K}|$ bounded. We say that a $\mathcal{K} \times \mathcal{K}$ matrix H satisfies the *ϵ-valley condition* if, for all $z = x + i\eta_0 \in \mathbf{S}^e$, we have that $\operatorname{dist}(x, \sigma(H)) > M^{-2/3-\epsilon}$ implies

$$\left| \Im \underline{(H - z)^{-1}} \right| \prec \frac{M^{-\epsilon'}}{M\eta_0}$$

and if $\lambda \in \sigma(H), \operatorname{dist}(\lambda, \sigma(H) \setminus \{\lambda\}) > 2M^{-2/3-\epsilon}$ and $|x - \lambda| < M^{-2/3-\epsilon}$ imply

$$\left| \Im \underline{(H - z)^{-1}} - \frac{1}{M\eta_0} \frac{1}{((x - \lambda)/\eta_0)^2 + 1} \right| \prec \frac{M^{-\epsilon'}}{M\eta_0}$$

for some $\epsilon' > 0$.

Lemma 5.2.6. *$(E_M X)^{\#}$ satisfies the ϵ-valley condition at z for any $z \in \mathbf{S}^e$ and any $\epsilon < \epsilon_0$.*

Proof. Let $z = x + i\eta_0 \in \mathbf{S}^e$ be fixed. Denote $\lambda_\alpha = \lambda_\alpha((E_M X)^{\#})$. Define, for a positive integer $n > 0$,

$$J_n = \left\{\alpha : |x - \lambda_\alpha| \in [M^{-2/3+n\epsilon}, M^{-2/3+(n+1)\epsilon})\right\}$$

Because $\lambda_\alpha \geq 0$, we see that $J_n = \emptyset$ for $n > \overline{n} := \lceil \epsilon \rceil^{-1}$. It follows from equation (5.11) that

$$|J_n| \leq \sum_{n'=1}^{n} |J_{n'}| \prec M^{\frac{3}{2}(n+1)\epsilon} \tag{5.15}$$

We write

$$\begin{aligned}
\Im \underline{G^{E_M}(z)} &= \frac{1}{M} \sum_{\alpha=1}^{M} \frac{\eta_0^{-1}}{((x-\lambda_\alpha)/\eta_0)^2 + 1} \\
&= \frac{1}{M} \sum_{n=-1}^{\overline{n}} \sum_{\alpha \in J_n} \frac{\eta_0^{-1}}{((x-\lambda_\alpha)/\eta_0)^2 + 1} \leq \frac{1}{M\eta_0} \sum_{n=-1}^{\overline{n}} |J_n| \frac{1}{(M^{\epsilon_0 + n\epsilon})^2} \\
&\prec \frac{1}{M\eta_0} \sum_{n=-1}^{\overline{n}} M^{\frac{3}{2}(n+1)\epsilon} M^{-2\epsilon_0 - 2n\epsilon} \leq \frac{1}{M\eta_0} (\overline{n}+1) M^{-2\epsilon_0 + 2\epsilon}
\end{aligned}$$

which is as desired, since $\overline{n}$ is constant. For the second statement, this follows from applying the same computation as above to

$$\Im \underline{G^{E_M}(z)} - \frac{1}{M\eta_0} \frac{1}{((x-\lambda_\beta)/\eta_0)^2 + 1} = \frac{1}{M} \sum_{\alpha \neq \beta} \frac{\eta_0^{-1}}{((x-\lambda_\alpha)/\eta_0)^2 + 1}$$

Thus we conclude the proof of Lemma 5.2.6. □

Here is the essential linear algebra formula which allows us to study the eigenvalues of a finite rank deformation of a random matrix.

Lemma 5.2.7 (Lemma 3.10 of [BKY16])**.** *If $\Sigma = I + VCV^*$ is a positive definite matrix with its eigen-decomposition and H is another matrix of which $x \in \mathbb{R}$ is not an eigenvalue, then $\Sigma^{1/2} H \Sigma^{1/2}$ has an eigenvalue at x if and only if*

$$\det\left(C^{-1} + V^*(1 + xG(x))V\right) = 0 \tag{5.16}$$

where $G(z) = \left(H^{-1} - zI\right)^{-1}$ is the resolvent.

We will also need to pair the following two facts. The first is a consequence of Theorem 2.2.2 and the second is well-known.

Lemma 5.2.8. *With high probability, the spectrum of $\mathcal{R}$ has K outlier eigenvalues. That is,*

$$\lambda_1(\mathcal{R}) > \cdots > \lambda_K(\mathcal{R}) > \operatorname{supp} \varrho_E + C_0$$

for a fixed constant C_0.

Lemma 5.2.9 (Cauchy eigenvalue interlacing)**.** *For a square matrix H and a finite rank deformation $\widetilde{H} = \Sigma^{1/2} H \Sigma^{1/2}$ with $\Sigma = I + VCV^*$ and $\Sigma > -1$, we have*

$$\lambda_i(\widetilde{H}) \in [\lambda_{i+K}(H), \lambda_{i-K}(H)]$$

Let $\kappa := \kappa(z)$ for $z = x + i\eta_0$ be defined as $\operatorname{dist}(x, \{\min \operatorname{supp} \varrho_E, \max \operatorname{supp} \varrho_E\})$. Regarding m_E we have

Lemma 5.2.10. *For $\mathbf{z} \in \mathbf{S}^e$, the function m_E satisfies*

$$\Im m_E \asymp \begin{cases} \sqrt{\kappa + \eta_0} & E \in \operatorname{supp} \varrho_E \\ \frac{\eta_0}{\kappa + \eta_0} & E \notin \operatorname{supp} \varrho_E \end{cases}$$

Proof. This is a exactly as lemma A.4 of [KY17], and follows from the regularity condition 2.7 of [KY17]. □

We also need the following lemma and corollary, which follow from the analogous statements for m, the usual Marčenko-Pastur law, since $E = I + O_\prec(N^{-1/2}$.

Lemma 5.2.11. *The measure ϱ_E is supported on $[(1-\sqrt{y})^2 + O_\prec(N^{-1/2}), (1+\sqrt{y})^2 + O_\prec(N^{-1/2})]$. Moreover,*

$$m_E\left((1+\sqrt{y})^2\right) = -\frac{1}{1+\sqrt{y}} + O_\prec(N^{-\delta})$$

Corollary 5.2.11.1. *When $z = (1+\sqrt{y})^2$, we have*

$$1 + z\Pi = -\frac{1}{\sqrt{y}} + O_\prec(N^{-\delta})$$

5.2.1 Properties of U_M

The purpose of U_M is that

$$\begin{pmatrix} D_M B & 1 \end{pmatrix} = \sqrt{(D_M B)^{\#} + I} U_M^*$$

which will be important for the proof of Lemma 5.1.6. Indeed, we know such a U_M exists because of the singular value decomposition, and if existing then it must satisfy

$$U_M^* = \left((D_M B)^{\#} + I\right)^{-1/2} \begin{pmatrix} D_M B & I \end{pmatrix} = \left(\sum_{\alpha=1}^{K} (d_\alpha + 1) \widetilde{\mathbf{v}}_\alpha \widetilde{\mathbf{v}}_\alpha\right)^{-1/2} \begin{pmatrix} \sum_{\alpha=1}^{K} \sqrt{\widetilde{d}_\alpha} \widetilde{\mathbf{v}}_\alpha \widetilde{\mathbf{w}}_\alpha^* & I \end{pmatrix}$$

at which point we see the desired form for U_M. This is sufficient to conclude that the columns of U_M columns are orthonormal, but it is helpful to observe this manually. Consider the $\mathcal{I}_K \times \mathcal{I}_K$ orthogonal matrix $\widetilde{V}$ which has $\widetilde{\mathbf{v}}_\alpha$ for its columns. Thus U_M has orthnormal columns if and only if $U_M^* U_M = I$, which is if and only if $(U_M \widetilde{V})^* (U_M \widetilde{V}) = I$, which is if and only if $U_M \widetilde{V}$ has orthonormal columns, and indeed,

$$U_M V = U_M \sum_{\alpha=1}^{M} \widetilde{\mathbf{v}}_\alpha \mathbf{e}_\alpha^* = \begin{pmatrix} \sum_{\alpha=1}^{K} \sqrt{\frac{\widetilde{d}_\alpha}{\widetilde{d}_\alpha + 1}} \widetilde{\mathbf{w}}_\alpha \mathbf{e}_\alpha^* \\ \sum_{\alpha=1}^{M} \sqrt{\frac{1}{\widetilde{d}_\alpha + 1}} \widetilde{\mathbf{v}}_\alpha \mathbf{e}_\alpha^* \end{pmatrix}$$

whose columns are $\begin{pmatrix} \sqrt{\frac{\widetilde{d}_\alpha}{\widetilde{d}_\alpha+1}} \widetilde{\mathbf{w}}_\alpha^* & \sqrt{\frac{1}{\widetilde{d}_\alpha+1}} \widetilde{\mathbf{v}}_\alpha^* \end{pmatrix}^*$, interpreting $\widetilde{\mathbf{w}}_\alpha = 0$ for $\alpha > K$, which are easily seen to be orthogonal and also unit length by the Pythagorean theorem.

U_M may be written as $U \begin{pmatrix} 0 & I_M \end{pmatrix}$ for an $(\mathcal{I}_K \cup \mathcal{I}_M) \times (\mathcal{I}_K \cup \mathcal{I}_M)$ orthogonal matrix U. We do not need an exact formula for the columns of U which do not also belong to U_M, but it may be verified that

$$U \mathbf{e}_\alpha \in \mathbf{R}^{\mathcal{I}_K} \oplus \operatorname{span}\{\widetilde{\mathbf{v}}_\alpha : \alpha \in \mathcal{I}_K\}$$

for $\alpha \in \mathcal{I}_K$.

Of course,

$$U_M := \begin{pmatrix} 0_{\mathcal{I}_M \times \mathcal{I}_K} & 1_{\mathcal{I}_M} \end{pmatrix}^* + \begin{pmatrix} \sum_{\alpha=1}^{K} \sqrt{\frac{\widetilde{d}_\alpha}{\widetilde{d}_\alpha+1}} \widetilde{\mathbf{v}}_\alpha \widetilde{\mathbf{w}}_\alpha^* & \sum_{\alpha=1}^{K} \left(\sqrt{\frac{1}{\widetilde{d}_\alpha+1}} - 1\right) \widetilde{\mathbf{v}}_\alpha \widetilde{\mathbf{v}}_\alpha^* \end{pmatrix}^*$$

so that U differs from the identity by bounded rank.

Recall the right singular vectors $\mathbf{v}_\alpha$ of B and note $\|\mathbf{v}_\alpha - \widetilde{\mathbf{v}}_\alpha\| \prec N^{-1/2}$ by Hadamard's variation formula and Weyl's inequality. All together, we have the lemma,

Lemma 5.2.12. *For deterministic* $\mathbf{x} \in \mathbb{R}^{\mathcal{I}_M}$*, we have*

$$U\mathbf{x} = \mathbf{x} + \sum_{\alpha \in \mathcal{I}_K} a_\alpha \mathbf{e}_\alpha + \sum_{\alpha \in \{1,\ldots,K\}} b_\alpha \mathbf{v}_\alpha + O_\prec(N^{-1/2})$$

for some bounded, random numbers $\{a_\alpha\}, \{b_\alpha\}$*.*

5.3 The Proof of Lemma 5.1.3

The main technical ingredient for the proof of Lemma 5.1.3 is the following:

Lemma 5.3.1. *The resolvents of* $(EX)^\#$ *and* $(E_M X_M)^\# = (E_M X)^\#$ *are close at the spectral edge: for* $z \in \mathbf{S}^e$*, we have*

$$\left|\underline{G^E} - \underline{G^{E_M}}\right| \prec \frac{N^{-\delta}}{N\eta_0} \tag{5.17}$$

Proof of Lemma 5.1.3. Define $\lambda_i := \lambda_i\left((E_M X)^\#\right), \widetilde{\lambda}_i := \lambda_i\left((EX)^\#\right)$. Define also the points $x_i^\pm = \lambda_i \pm N^{-2/3-\epsilon_0/2}$. We let $\epsilon := \epsilon_0/2$ in the lemma. Since $(E_M X)^\#$ is a contraction of $(EX)^\#$, the standard interlacing inequality yields that

$$\widetilde{\lambda}_i \le \lambda_i \tag{5.18}$$

First note that because $(E_M X)^\#$ satisfies the $\epsilon_0/2$-valley condition (Lemma 5.2.6) and by Lemma 5.3.1, we have (also using that E_M satisfies the weak level repulsion condition)

$$\left|\Im \underline{G^{E_M}(x_1^-)}\right| \prec \frac{M^{-\epsilon'}}{M\eta_0} \text{ so that } \left|\Im \underline{G^{E}(x_1^-)}\right| \prec \frac{M^{-\epsilon'}}{M\eta_0}. \tag{5.19}$$

Similarly,

$$\left|\Im \underline{G^{E_M}(\lambda_1)}\right| \succ \frac{1}{M\eta_0} \text{ so that } \left|\Im \underline{G^{E}(\lambda_1)}\right| \succ \frac{1}{M\eta_0}. \tag{5.20}$$

Now see that $\left|\left\{i : \widetilde{\lambda}_i \in [x_i^-, x_i^+]\right\}\right| \geq 1$. Indeed, if this were not the case, then $\widetilde{\lambda}_1 < x_1^-$ by equation (5.18), so that for $x \geq x_i^-$, $x \mapsto \Im \underline{G^E(x)}$ is decreasing, which is a contradiction to equations (5.19) and (5.20).

Now we show that $\left|\left\{i : \widetilde{\lambda}_i \in [x_1^-, x_1^+]\right\}\right| \leq 1$. Otherwise, we see that

$$\int_{x_i^- - N^{-2/3-\epsilon_0/2}}^{x_i^+ + N^{-2/3-\epsilon_0/2}} \Im \underline{G^E(x)} \mathrm{d}x \geq (2 - \epsilon_0) \frac{1}{M\eta_0}$$

so that by Lemma 5.1.3,

$$\int_{x_i^- - N^{-2/3-\epsilon_0/2}}^{x_i^+ + N^{-2/3-\epsilon_0/2}} \Im \underline{G^{E_M}(x)} \mathrm{d}x \geq (2 - 2\epsilon_0) \frac{1}{M\eta_0}$$

which is a contradiction to the valley condition and weak level repulsion condition for $(E_M X)^\#$.

We conclude that $\left|\left\{i : \widetilde{\lambda}_i \in [x_1^-, x_1^+]\right\}\right| = 1$. The valley condition for E_M and Lemma 5.1.3 are enough to conclude $\left|\left\{i : \widetilde{\lambda}_i \in [x_2^+, x_1^-]\right\}\right| = 0$, lest $\Im \underline{G^{E_M}(x)} \geq \frac{1}{M\eta_0}$ for $x \in [x_2^+, x_1^-]$.

To establish that $\left|\left\{i : \widetilde{\lambda}_i \in [x_2^-, x_2^+]\right\}\right| = 1$, we apply all the same argument to the functions (slightly abusing notation)

$$\begin{aligned} \underline{G^{E_M}} - \frac{M^{-1}}{\lambda_1 - z} &= \underline{G^{E_M}} + O_\prec \left(\frac{M^{-\epsilon}}{M\eta_0} + \mathbf{1}_{\Re z > x_2^+} \right) \\ \text{and } \underline{G^E} - \frac{M^{-1}}{\widetilde{\lambda}_1 - z} &= \underline{G^E} + O_\prec \left(\frac{M^{-\epsilon}}{M\eta_0} + \mathbf{1}_{\Re z > x_2^+} \right) \end{aligned}$$

Repeating L times, we conclude the proof of Lemma 5.1.3. □

Observe the following immediate corollary to Lemma 5.1.3 which we will use later: $(EX)^\#$ satisfies the $\epsilon_0/2$-valley and weak level repulsion conditions.

Proof of Lemma 5.3.1. The main technical ingredient in the proof of Lemma 5.3.1 is Lemma 5.1.8, whose proof will constitute the majority of this section.

We write

$$\underline{G^E} = (M+K)^{-1} \sum_{i \in \mathcal{I}_M \cup \mathcal{I}_\mathcal{K}} G^E_{ii}$$
$$= M^{-1} \sum_{i \in \mathcal{I}_M} G^E_{ii} + (M+K)^{-1} \sum_{i \in \mathcal{I}_\mathcal{K}} G^E_{ii} + \frac{-K}{M(M+K)} \sum_{i \in \mathcal{I}_M} G^E_{ii}$$

and

$$\underline{G^{E_M}} = M^{-1} \sum_{i \in \mathcal{I}_M} G^{E_M}_{ii}.$$

Since $M^{-1} = O\left(\frac{N^{-\delta}}{N\eta_0}\right)$, it suffices to show that

$$\max_{i \in \mathcal{I}_M} \left| G^E_{ii} - G^{E_M}_{ii} \right| \prec \frac{N^{-\delta}}{N\eta_0} \text{ and } \max_{i \in \mathcal{I}_M \cup \mathcal{I}_K} \left| G^E_{ii} \right| \prec 1. \tag{5.21}$$

We begin by establishing

$$G^{E_M}_{1,ii} = G^E_{1,ii} + O_\prec\left(\max_\alpha \left| G^E_{1,i\alpha} \right|^2\right) \tag{5.22}$$

for $i \in \mathcal{I}_M$. Note that $G^{E_M}_1 = G^{E^{(\mathcal{I}_K)}}_1$. Let $\mathcal{I}_K = \{\alpha_1, \ldots, \alpha_K\}$. We use Lemma 5.2.4 to see that

$$G^{E^{(\alpha_1 \cdots \alpha_K)}}_{1,ii} = G^{E^{(\alpha_1 \cdots \alpha_{K-1})}}_{1,ii} - \frac{G^{E^{(\alpha_1 \cdots \alpha_{K-1})}}_{1,i\alpha_K} G^{E^{(\alpha_1 \cdots \alpha_{K-1})}}_{1,\alpha_K,i}}{G^{E^{(\alpha_1 \cdots \alpha_{K-1})}}_{1,\alpha_K\alpha_K}}.$$

Inductively we may assume that

$$G^{E^{(\alpha_1 \cdots \alpha_{K-1})}}_{1,\mathbf{e}_j\mathbf{e}_k} = G^E_{1,\mathbf{e}_j\mathbf{e}_k} + O_\prec\left(\max_{\alpha \in \mathcal{I}_K \setminus \{\alpha_K\}} \left| G^E_{1,j,\alpha} \right| \left| G^E_{1,k\alpha} \right|\right)$$

for $j, k \in \mathcal{I}_M \cup \{\alpha_K\}$. Equation (5.22) then follows by

$$\left| G^E_{1,i\alpha_K} \right| + \left| G^E_{1,\alpha_K i} \right| + \left| G^E_{1,\alpha_K\alpha_K} \right|^{-1}$$
$$= \left| zG^E_{i\alpha_K} \right| + \left| zG^E_{\alpha_K i} \right| + \left| zG^E_{\alpha_K\alpha_K} \right|^{-1} \prec 1$$

which is a consequence of Lemma 5.1.8.

Now we may write for $i \in \mathcal{I}_M$,

$$G^E_{ii} = z^{-1} \mathbf{e}_i^* E^{-1} G^E_2 E^{-1} \mathbf{e}_i = z^{-1} E^{-1}_{ii} G^E_{2,ii} E^{-1}_{ii}$$
$$= z^{-1} (E_M)^{-1}_{ii} G^{E_M}_{2,ii} (E_M)^{-1}_{ii} + O_\prec\left(\max_{\alpha \in \mathcal{I}_K} \left| G^E_{i\alpha} \right|^2\right) \tag{5.23}$$
$$= G^{E_M}_{ii} + O_\prec\left(\max_{\alpha \in \mathcal{I}_K} \left| G^E_{i\alpha} \right|^2\right)$$

where we used that $(E_M)_{ii} = E_{ii}$ for $i \in \mathcal{I}_M$. For $i \in \mathcal{I}_M$, Lemma 5.1.8 gives

$$\left|G^E_{i\alpha}\right| \prec N^{-1/6-\delta}$$

Since $N^{-1/3-2\delta} \leq \frac{N^{-\delta}}{N\eta_0}$, equation (5.23) and Lemma 5.1.8 establish the first bound of (5.21). Lemma 5.1.8 also implies the second bound of (5.21). This concludes the proof of Lemma 5.1.3. □

Now we may begin the proof of Lemma 5.1.8. The polynomialization method was developed in [BEK14] to prove the iostropic local law from the entrywise law. In analogy to this, to prove the isotropic law Lemma 5.1.8, we will need an entrywise law for G^E. This and a closely related fact are collected in the following lemma.

Lemma 5.3.2. *Let $s, t \in \mathcal{I}_K \cup \mathcal{I}_M \cup \mathcal{I}_N$. Fix a subset $T \subseteq \mathcal{I}_K \cup \mathcal{I}_M \cup \mathcal{I}_N \setminus \{s, t\}$ with $|T|$ bounded. We have the following bound:*

$$\left|G^{E(T)}_{st} - \Pi_{st}\right| \prec \psi_{(st)}. \tag{5.24}$$

Moreover, if $s, t \in \mathcal{I}_N$, then

$$\left|\sum_{\mu\nu} X_{s\mu} G^{E(Tst)}_{\mu\nu} X_{t\nu} - \delta_{st} m\right| \prec \psi_{(st)}. \tag{5.25}$$

The proof of Lemma 5.3.2 is less interesting than that of 5.1.8 and is postponed until Section 5.3.4. Notice that Lemma 5.3.2 would actually have been sufficient to conclude the proof of the main Lemma 5.1.3. But, we will require the full strength of Lemma 5.1.8 later in the paper.

5.3.1 Introduction of Graphs

Now we outline the terminology and techniques from [BEK14] that we will use throughout the rest of the paper.

We prove Lemma 5.1.8 in the next section; here we will introduce the structures and terminologies necessary for its proof. The bulk of the work in proving that proposition is bounding

$$\left|G^E_{2,\alpha\mathbf{x}}\right| \prec N^{-1/6-\delta} \tag{5.26}$$

for $\alpha \in \mathcal{I}_K, \mathbf{x} \in \mathbb{R}^{\mathcal{I}_M}$. Note that of course $\langle \mathbf{e}_\alpha, \mathbf{x}\rangle = 0$. This fact is why we do not have an analog to section 5.2 "Reduction to off-diagonal entries" in [BEK14]. We seek to establish the sufficient high moment bound

$$\mathbb{E}\left|G^E_{2,\alpha\mathbf{x}}\right|^p \prec \psi^p_{KM} \tag{5.27}$$

We use the polynomialization technique of [BEK14]. Define a (edge-colored, directed, multi-) graph Δ^{pre} with vertices $0, 1, \ldots, p$ and edges $\{(0,1), \ldots, (0, p/2)\}$ with color G and $\{(0, p/2+1), \ldots, (0, p)\}$ with color $\overline{G}$. We let $\mathfrak{P}$ be the set of partitions of $V(\Delta)$ which contain the singleton block $i_0 := \{0\}$ and for each $P \in \mathfrak{P}$ let $\Delta := \Delta(P)$ be the quotient graph Δ^{pre}/P. We enumerate the vertice $i_0, \ldots, i_{|P|-1}$ of Δ. We write the set of vertices of Δ as $V(\Delta)$ or as $V_b(\Delta)$. As we will explain in the next paragraph, every vertex of Δ will be assigned a value in $\mathcal{I}_K \cup \mathcal{I}_M$, and be called "black" vertices. Later we will see graphs which also have "white" vertices which get assigned values in $\mathcal{I}_N$, and the white vertices of a graph Γ will be denote $V_w(\Gamma)$. So currently we have $V_w(\Delta) = \emptyset$ and $V_b(\Delta) = V(\Delta)$. For each edge $e \in E(\Delta)$, we denote by $\alpha(e)$ and $\beta(e)$ the initial and terminal vertices of the edge e. It should always be clear from context whether α refers to the vertex $\alpha \in \mathcal{I}_K$ which is the first index of the generalized resolvent entry in equation (5.26) or the function $e \mapsto \alpha(e)$.

For each of the vertices $i_1, \ldots, i_{|P|}$ of Δ, we associate a value $a_{i_1}, \ldots, a_{i_{|P|}} \in \mathcal{I}_M$. For i_0 we associate the value $a_{i_0} = \alpha$. We say that the vertices $i_1, \ldots, i_{|P|}$ "land in" $\mathcal{I}_M$, while i_0 "lands in" $\mathcal{I}_K$. We write the tuple of vertex values $\mathbf{a}_b = (a_{i_0}, a_{i_1}, \ldots, a_{i_{|P|}})$.

We define the subset $V_b^* \subseteq V(\Delta)$ as the set of vertices with odd degree. Since p is even and by our defition of $\mathfrak{P}$, $i_0 \notin V_b^*$.

We arrive at

$$\mathbb{E}\left|G^E_{2,\alpha\mathbf{x}}\right|^p = \sum_{P\in\mathfrak{P}_p} Y(\Delta(P)),$$

where

$$Y(\Delta) = \sum_{\mathbf{a}_b}^{*} w_{\mathbf{a}_b}(\Delta)\mathbb{E}\mathcal{A}_{\mathbf{a}_b}(\Delta),$$

where $\sum^*_{\mathbf{a}_b}$ denotes the sum over all values $\mathbf{a}_b$ of the vertices of Δ subject to the constraint that distinct vertices get distinct values, and

$$w_{\mathbf{a}_b}(\Delta) = \prod_{e\in E(\Delta)} (\mathbf{e}_\alpha)_{\alpha(e)}\mathbf{x}_{a_{\beta(e)}} = \prod_{e\in E(\Delta)} \mathbf{x}_{a_{\beta(e)}}$$

$$\mathcal{A}_{\mathbf{a}_b}(\Delta) = \prod_{e:\xi(E)=G} G_{2,a_{\alpha(e)}a_{\beta(e)}} \prod_{e:\xi(E)=\overline{G}} \overline{G_{2,a_{\alpha(e)}a_{\beta(e)}}},$$

where $\overline{z}$ denotes a complex conjugate for $z \in \mathbb{C}$, and we have called the colors of edges G and $\overline{G}$ rather than G and G^* as in [BEK14]. In our setting, $\alpha(e) = i_0$ for every $e \in E(\Delta)$, and we only consider $\mathbf{a}_b$ with $a_{i_0} = \alpha$, hence our form for the weight $w_{\mathbf{a}_b}(\Delta)$. For the rest of this section we consider the partition P as fixed.

Now let us more fully define the graphs used in [BEK14].

Definition 5.3.3. [Graph] A *graph* is a finite, directed, edge-colored multigraph

$$\Gamma = (V(\Gamma), E(\Gamma), \xi(\Gamma)),$$

where V is a finite set of vertices, E is a finite set of edges, and ξ is a coloring of the edges. We also let $V(\Gamma) = V_b(\Gamma) \cup V_w(\Gamma)$.

Up to now we have only seen graphs with $V_w(\Gamma) = \emptyset$. But for general graphs Γ, we will associate a tuple $\mathbf{a}_w = (a_i)_{i\in V_w(\Gamma)}$ of values in $\mathcal{I}_N$.

Remark 5.3.4. We only ever consider tuples $\mathbf{a}_b$ with distinct entries, but until we introduce a partition on the white vertices, we will allow $\mathbf{a}_w$ to have non-distinct entries. Once we introduce a partition we will restrict $\mathbf{a}_w$ to have distinct entries.

As a set of colors, we choose

$$\{\xi = (\xi_1, \xi_2, \xi_3) : \xi_1 \in \{G, \overline{G}, X, \overline{X}\}, \xi_2 \in \{+, -\}, \xi_3 \in V_b(\Gamma)\}.$$

We will denote $E_G(\Gamma) := \{e \in E(\Gamma) : \xi_1(e) \in \{G, \overline{G}\}$, $E_o(\Gamma) := \{e \in E_G(\Gamma) : \alpha(e) \neq \beta(e)\}$ and $E_X(\Gamma) := \{e \in E(\Gamma) : \xi_1(e) \in \{X, \overline{X}\}$. We call edges in the first set G edges, in the second set off-diagonal (G) edges, and in the third set X edges. Edges e in a graph Γ with color $\xi_1(e) \in \{X, \overline{X}\}$ will always have $\alpha(e) \in V_b(\Gamma)$ and $\beta(e) \in V_w(\Gamma)$.

For tuple $\mathbf{a}_b = (a_i)_{i \in V_b(\Gamma)}$ of values in $\mathcal{I}_K \cup \mathcal{I}_M$ of the vertices $V_b(\Gamma)$, we define $\mathbf{a}_{\xi_3} = \{a_i\}_{i \in \xi_3}$.

For a graph Γ and $e \in E(\Gamma)$ with $\xi(e) = (\xi_1, \xi_2, \xi_3)$, we define the evaluation

$$\mathcal{A}_{\mathbf{a}_b}(e, \Gamma) = \begin{cases} G^{E^{(\mathbf{a}_{\xi_3})}}_{2, a_{\alpha(e)} a_{\beta(e)}} & \xi_1 = G \text{ and } \xi_2 = + \\ 1/G^{E^{(\mathbf{a}_{\xi_3})}}_{2, a_{\alpha(e)} a_{\beta(e)}} & \xi_1 = G \text{ and } \xi_2 = - \\ \overline{G^{E}}^{(\mathbf{a}_{\xi_3})}_{2, a_{\alpha(e)} a_{\beta(e)}} & \xi_1 = \overline{G} \text{ and } \xi_2 = + \\ 1/\overline{G^{E}}^{(\mathbf{a}_{\xi_3})}_{2, a_{\alpha(e)} a_{\beta(e)}} & \xi_1 = \overline{G} \text{ and } \xi_2 = - \\ X_{a_{\alpha(e)} a_{\beta(e)}} & \xi_1(e) = X \\ \overline{X}_{a_{\alpha(e)} a_{\beta(e)}} & \xi_1(e) = \overline{X}. \end{cases} \tag{5.28}$$

We say that an edge e is maximally expanded if $\{\alpha(e), \beta(e)\} \cup \xi_3(e) = V_b(\Gamma)$. We now define the evaluation of a general graph, subsuming our previous definition for the graph Δ,

$$\mathcal{A}_{\mathbf{a}_b}(\Gamma) = \prod_{e \in E(\Gamma)} \mathcal{A}_{\mathbf{a}_b}(e, \Gamma).$$

Actually, though, we adjust slightly and define

$$\mathcal{A}_{\mathbf{a}_b}(\Gamma) = u(\Gamma) \prod_{e \in E(\Gamma)} \mathcal{A}_{\mathbf{a}_b}(e, \Gamma)$$

where $u(\Gamma)$ is a bounded deterministic prefactor. Its presence allow us to disregard factors of -1 or z that arise in later steps by absorbing them into $u(\Gamma)$.

Let us now outline the expansion procedure, which relies on the identities, for any $a, b, c \in V_b(\Gamma)$ with $c \notin \{a, b\}$ and $T \subseteq V_b(\Gamma) \setminus \{a, b, c\}$, abbreviating (Tc) for the superscript $(T \cup \{c\})$,

$$G_{2,ab}^{E(T)} = G_{2,ab}^{E(Tc)} + \frac{G_{2,ac}^{E(T)} G_{2,cb}^{E(T)}}{G_{2,cc}^{E(T)}}, \quad \frac{1}{G_{2,aa}^{E(T)}} = \frac{1}{G_{2,aa}^{E(Tc)}} - \frac{G_{2,ac}^{E(T)} G_{2,ca}^{E(T)}}{G_{2,aa}^{E(T)} G_{2,aa}^{E(Tc)} G_{2,cc}^{E(T)}}. \tag{5.29}$$

Using this identity and given a graph Γ, we define 2 graphs $\tau_0(\Gamma)$ and $\tau_1(\Gamma)$ constructed from Γ as follows. Considering an arbitrary ordering on the set $\{G_{2,ab}^{E(T)} : a, b \notin T\}$ of all resolvent entries and the vertex set $V_b(\Gamma)$, we take the first edge e in Γ which is not maximally expanded. $\tau_0(\Gamma)$ is identical to Γ except that we add the first $c \in \mathbf{a}_b \setminus \big(\mathbf{a}_{\xi_3(e)} \cup \{a_\alpha(e), a_\beta(e)\}\big)$ to $\xi_3(e)$. $\tau_1(\Gamma)$ also differs from Γ only locally; the edge e is replaced with 3 edges if $\xi_2(e) = +$: $(\alpha(e), c)$, $(c, \beta(e))$ with color $(\xi_1(e), +, \xi_3(e))$ and an edge (c, c) with color $(\xi_1(e), -, \xi_3(e))$. If rather $\xi_2(e) = -$, a similar definition holds according to the identity (5.29), so that

$$\mathcal{A}_{\mathbf{a}_b}(\Gamma) = \mathcal{A}_{\mathbf{a}_b}(\tau_0 \Gamma) + \mathcal{A}_{\mathbf{a}_b}(\tau_1 \Gamma).$$

Within the graph $\tau_1 \Gamma$, we say that the vertex c has been "pulled to", because, besides the other diagonal edges which are produces, the essential effect in passing from Γ to $\tau_1 \Gamma$ is that an edge of Γ not already incident on c is re-routed so as to pass through c on its way from its initial vertex to its terminal vertex; or, if the vertices are thought of as pins on a board and the edges as strings from pin to pin, then some string's midpoint is attached to the pin at c.

Now, for graphs in which every off-diagonal edge is maximally expanded, we define another operation ρ, which also follows from a resolvent identity:

$$G_{2,ab}^{E(T)} = z G_{2,aa}^{E(T)} G_{2,bb}^{E(Ta)} \sum_{\mu, \nu \in \mathcal{I}_N} X_{a\mu} G_{2,ab}^{E(T)} \overline{X}_{b\nu}. \tag{5.30}$$

The graph $\rho(\Gamma)$ is defined as the graph encoding the monomial $\mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(\Gamma)$ after replacing every maximally expanded off-diagonal resolvent entry according to the identity (5.30). On the level of graphs, every off-diagonal edge e in the graph Γ is replaced by a chain of edges: an X edge from $\alpha(e)$ to j, a G edge from j to j', and an X edge from $\beta(e)$ to j', where j

and j' are two new white vertices added to $V(\Gamma)$; so after the first application of ρ, we now consider graphs with $V_w(\Gamma) \neq \emptyset$. We call this chain of edges an R-group (this terminology is an artifact; in [BEK14], there are two matrices: G indexed by $(\mathcal{I}_K \cup \mathcal{I}_M) \times (\mathcal{I}_K \cup \mathcal{I}_M)$ and R indexed by $\mathcal{I}_N \times \mathcal{I}_N$, whereas we use one matrix, the "linearizing block matrix" G_2^E). As we defined $\mathbf{a}_b$, now we define $\mathbf{a}_w = (a_i)_{i \in V_w(\Gamma)} \in \mathcal{I}_N^{|V_w(\Gamma)|}$. Thus by definition we have

$$\sum_{\mathbf{a}_w} \mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(\Gamma) = \sum_{\mathbf{a}_w} \mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(\rho\Gamma).$$

As opposed to the sum $\sum^*_{\mathbf{a}_b}$ which was restricted to distinct values for distinct vertices, we let the sum $\sum_{\mathbf{a}_w}$ be totally unrestricted subject to $a_i \in \mathcal{I}_N$ for each $i \in V_w(\Gamma)$.

We now recursively apply the operations τ and ρ to the graph Δ, defining a new family of graphs Θ_σ where σ ranges over the vertices of a finite, complete binary tree, or equivalently, over a set of finite binary strings. We define for a binary string σ

$$\Theta_\emptyset = \Delta, \quad \Theta_{0\sigma} = \rho(\tau_0(\Theta_\sigma)), \quad \Theta_{1\sigma} = \rho(\tau_1(\Theta_\sigma)),$$

where $i\sigma$ is the binary string σ with i appended on the left. We define the binary tree $\mathcal{T}$, whose vertices are finite binary strings, which will index the expansion as follows: starting with $\mathcal{T} =$ the trivial tree with the single vertex $\emptyset$, for every leaf σ of $\mathcal{T}$ for which Θ_σ does not satisfy the stopping rule below, we add two children 0σ and 1σ to the leaf σ. Notice that every vertex σ of the tree $\mathcal{T}$ satisfies $V_b(\Theta_\sigma) = V_b(\Delta)$.

Definition 5.3.5 (Stopping Rule)**.** We say a graph Γ satisfies the stopping rule if either all off-diagonal edges of Γ are maximally expanded or if Γ has $\geq L := (p + |V_b(\Delta)|)/\delta$ off-diagonal G edges.

The stopping rule ensures that for every leaf σ of $\mathcal{T}$, the graph Θ_σ has only maximally expanded off-diagonal edges (we call these leaves the non-trivial leaves) or else satisfies, by Lemma 5.3.2,

$$\left| \sum_{\mathbf{a}_b, \mathbf{a}_w} \mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(\Theta_\sigma) \right| \prec N^{-p}$$

and thus is an error term (we call these leaves the trivial leaves). It is not hard to see and is explained in [BEK14] why $\mathcal{T}$ has bounded depth—every leaf has either one more off-diagonal edge than its parent or is closer to being maximally expanded than its parent. We arrive at

$$\begin{aligned}\mathcal{A}_{\mathbf{a}_b}(\Delta) &= \sum_{\sigma \in L(\mathcal{T})} \sum_{\mathbf{a}_w} \mathcal{A}_{\mathbf{a}_b \mathbf{a}_w}(\Theta_\sigma) \\ &= \sum_{\sigma \in L_1(\mathcal{T})} \sum_{\mathbf{a}_w} \mathcal{A}_{\mathbf{a}_b \mathbf{a}_w}(\Theta_\sigma) + \sum_{\sigma \in L_2(\mathcal{T})} \sum_{\mathbf{a}_w} \mathcal{A}_{\mathbf{a}_b \mathbf{a}_w}(\Theta_\sigma),\end{aligned} \tag{5.31}$$

where $L_1(\mathcal{T})$ is the set of non-trivial leaves and $L_2(\mathcal{T})$ is the set of trivial leaves. The trivial leaves already having the desired bound $O_\prec(N^{-p})$, we now move to bound the non-trivial leaves, that is, the ones in which every off-diagonal G edge has been replaced with an R-group, and every remaining diagonal edge is maximally expanded. Very similary to the previous operations τ and ρ and as explained more fully in [BEK14], for every non-trivial leaf Θ_σ, we replace each maximally expanded diagonal edge e with $\xi_2(e) = -$ according to the identity

$$1/G_{2,aa}^{E^{(T)}} = -z - z \sum_{\mu\nu} X_{a\mu} G_{2,\mu\nu}^{E^{(Ta)}} \overline{X}_{b\nu}, \tag{5.32}$$

resulting in two new graphs for each such edge, one in which the edge e is removed, and one in which the edge e is replaced with a diagonal R-group, ie, two X edges having one vertex at $\alpha(e) = \beta(e)$ and having their other vertices joined by a G edge (in both new graphs $-z$ is absorbed into the determistic prefactor $u(\Gamma)$). To treat diagonal edges e with $\xi_2(e) = +$, we do standard manipulations on the identity (5.32): first write the above as

$$\begin{aligned}1/G_{2,aa}^{E^{(T)}} &= -z - zm - z\left(-m + \sum_{\mu\nu} X_{a\mu} G_{2,\mu\nu}^{E^{(Ta)}} \overline{X}_{b\nu}\right) \\ &= (-z - zm)\left(1 - \frac{1}{-z - zm}\left(-m + \sum_{\mu\nu} X_{a\mu} G_{2,\mu\nu}^{E^{(Ta)}} \overline{X}_{b\nu}\right)\right)\end{aligned}$$

and then do a geometric expansion

$$G_{2,aa}^{E^{(T)}} = (-z - zm) \sum_{k=0}^{L} \left(\frac{1}{-z - zm}\left(-m + \sum_{\mu\nu} X_{a\mu} G_{2,\mu\nu}^{E^{(Ta)}} \overline{X}_{b\nu}\right)\right)^k + O_\prec(N^{-p}) \tag{5.33}$$

with the same choice of L as in the stopping rule; Lemma 5.3.2 is what allows the truncation. Applying equation (5.32) or (5.33) to every diagonal edge of Θ_σ for every non-trivial leaf σ of $\mathcal{T}$ and multiplying everything out, we get

$$\sum_{\mathbf{a}_w} \mathcal{A}_{\mathbf{a}_b \mathbf{a}_w}(\Theta_\sigma) = \sum_{\Gamma \in \mathfrak{G}} \sum_{\mathbf{a}_w} \mathcal{A}_{\mathbf{a}_b \mathbf{a}_w}(\Gamma) + O_\prec(N^{-p}) \tag{5.34}$$

for a family $\mathfrak{G}$ of graphs. Note that for $\Gamma \in \mathfrak{G}$, every edge $e \in E_G(\Gamma)$ joins two white vertices, ie, it encodes a resolvent entry $G_{2,\mu\nu}^{E^{(\mathbf{a}_b)}}$ for $\mu, \nu \in \mathcal{I}_N$. The big upshot of this construction is that every G-edge in one of these graphs Γ is independent of every X edge in Γ, *except* that all resolvent entries $G_{2,\mu\nu}^{E^{(\mathbf{a}_b)}}$ retain some dependence on X edges whose initial vertices land in $\mathcal{I}_K$; this is because the matrix E is itself dependent on E and represents a main difference between our setting and [BEK14].

Note that white vertices only appear in the graphs $\Gamma \in \mathfrak{G}$ as parts of diagonal or off-diagonal R-groups. As such every $j \in V_w(\Gamma)$ is connected by an X edge to a unique $i := \pi(j) \in V_b(\Gamma)$.

5.3.2 Proof of Lemma 5.1.8

Proof of Lemma 5.1.8. The proof of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_M}$ is identical to the proof of the main result of [BEK14], using only the additional input $G_{2,\mu\nu}^{E^{(T)}} \prec \delta_{\mu\nu}\psi$ for $T \subseteq \mathcal{I}_M \cup \mathcal{I}_K$, which is a consequence of Lemma 5.3.2.

The case of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_K}$ is a consequence of Lemma 5.3.2.

It remains to bound

$$\left|G_{2,\alpha\mathbf{x}}^{E}\right| \prec N^{-1/6-\delta}.$$

We consider the high moment $\mathbb{E}\left|G_{2,\alpha\mathbf{x}}^{E}\right|^p$ and, expanding it as explained in the previous section, it suffices to consider

$$\mathbb{E} \sum_{\mathbf{a}_b}^{*} \sum_{\mathbf{a}_w} \mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(\Gamma) \tag{5.35}$$

for $\Gamma \in \mathfrak{G}$.

Now we fix a partition ζ of the white vertices of Γ (here, we diverge from [BEK14], where they fix a partition of the white vertices adjacent to a black vertex i separately for every i).

From $E(\Gamma)$, choose a collection $\widetilde{E_o}(\Gamma) \subset E_o(\Gamma)$ of p edges connecting $j \in \pi^{-1}(i_0)$ to $j \in \pi^{-1}(i)$ for $i \neq i_0$; note that this is always possible: the graph Δ has p off-diagonal edges incident on i_0, the operations τ_0 and τ_1 do no decrease the number of such edges, and each such edge yields after the expansion of section 5.3.1 of [BEK14] one edge for $\widetilde{E_o}(\Gamma)$.

Consider now the following lemma, proven later in this section. Define the restriction of the partition ζ to $\pi^{-1}(i)$ as ζ_i.

Lemma 5.3.6. *Let $i \in V_b \setminus \{i_0\}$. Then either*

$$|\zeta_i| \leq \begin{cases} \frac{\left|\pi^{-1}(i)\right|-1}{2} & i \in V_b^* \\ \frac{\left|\pi^{-1}(i)\right|}{2} & i \in V_b \setminus V_b^* \end{cases} \tag{5.36}$$

or

$$\mathbb{E} \prod_{e \in E(\Gamma_\zeta)} \mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e, \Gamma_\zeta) = 0. \tag{5.37}$$

Lemma 5.3.6 says that the set of white vertices adjacent to one black vertex must at least pairwise identify. One important aspect of this lemma is that it does not treat black vertices landing in $\mathcal{I}_K$. This is because for $\alpha \in \mathcal{I}_K, \mu \in \mathcal{I}_N$, maximally expanded edges are not actually independent of $X_{\alpha\mu}$. Again, this is one of the main differences between this paper and [BEK14].

Now identify all the white vertices according to ζ, and let Γ_ζ be the resulting graph. $\mathbf{a}_w$ will now refer to a tuple of values for the vertices of Γ_ζ. With the introduction of a partition, we will now only consider $\mathbf{a}_w$ with distinct values. Note that the function π, which maps a white vertex to its black neighbor, is no longer a function after identifying some white vertices, but it may still be regarded as a relation or a multi-valued function, so that $\pi^{-1}(i)$ is still defined.

Now let ζ_{b} be the partition which is maximally unrestricted subject to the constraint that if ζ identifies two endpoints of some $e \in \widetilde{E_o}(\Gamma)$, then so does ζ_{b}. So ζ_{b} has only blocks of size 1 or 2, and ζ_{b} is a strictly coarser partition that ζ. Let $\Gamma_{\zeta_{\mathrm{b}}}$ be the graph which results after identifying vertices of Γ according to ζ_{b}. Using the cannonical mapping of edges from Γ to $\Gamma_{\zeta_{\mathrm{b}}}$ and defining r to be the number of blocks of ζ_b of size 2, we have that $p-r$ of the edges $e \in \widetilde{E_o}(\Gamma_{\zeta_{\mathrm{b}}})$ are still off-diagonal. Here b stands for "bad", because the identications between vertices induced by ζ_{b} cause edges to cease to be off-diagonal, and we need a certain number of off-diagonal G entries, which satisfy $|G_{\mu\nu}| \prec \psi_{(NN)} = \psi$ for $\mu \neq \nu \in \mathcal{I}_N$ in order to establish (5.27).

Define the subset $V_w^*(\Gamma_\zeta) \subseteq V_w(\Gamma_\zeta)$ of vertices j which are contained in trivial (singleton) blocks of ζ, and define $\ell = |V_w(\Gamma_\zeta)|$. Define also $\widetilde{V_w^*}(\Gamma_\zeta) \subseteq V_w^*(\Gamma_\zeta)$ as the set of those vertices j such that j is one end of an edge $e \in \widetilde{E_o}(\Gamma)$, and let $\widetilde{\ell} = \left|\widetilde{V_w^*}(\Gamma_\zeta)\right|$, so that $\widetilde{\ell} \leq p$. The purpose of recording the number $\widetilde{\ell}$ is that all vertices $j \in V_w^*(\Gamma_\zeta)$, by virtue of being unidentified with any other white vertices, are the endpoint of some off-diagonal G edge; the off-diagonal edges incident on vertices in $\widetilde{V_w^*}(\Gamma_\zeta)$ are however in $\widetilde{E_o}(\Gamma_\zeta)$ and thus already accounted for, so we must be sure not to double count them. We state the following lemma regarding the number of white vertices in Γ_ζ, which is proven after the conclusion of the current lemma.

Lemma 5.3.7. *For any graph* Γ *and any partition* ζ *such that*

$$\mathbb{E} \prod_{e \in E(\Gamma_\zeta)} \mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(e, \Gamma_\zeta) \neq 0,$$

we have

$$|V_w(\Gamma_\zeta)| \leq \frac{|E_X(\Gamma)|}{2} - \frac{r}{2} + \frac{\ell}{2} - \frac{|V_b^*|}{2}.$$

Now we need to find some off-diagonal edges in Γ_ζ which are not present among $\widetilde{E_o}(\Gamma_\zeta)$. Note that e with $\alpha(e), \beta(e) \in \pi^{-1}(i_0)$ may be included in $E_o(\Gamma_\zeta)$ if, for instance, each $\alpha(e)$ or $\beta(e)$ is $\in V_w^*(\Gamma_\zeta)$. In this way, every $j \in V_w^*(\Gamma_\zeta) \setminus \widetilde{V_w^*}(\Gamma_\zeta)$ is one endpoint of some

$e \in E_o(\Gamma_\zeta) \setminus \widetilde{E_o}(\Gamma_\zeta)$. This implies that

$$\left| E_o(\Gamma_\zeta) \setminus \widetilde{E_o}(\Gamma_\zeta) \right| \geq \frac{1}{2}\left(\ell - \widetilde{\ell}\right) \tag{5.38}$$

Now, we will need a secondary sort of resolvent expansion to account for the fact that the population matrix E depends on X_K. Analogously to the definitions of τ_0 and τ_1, we define operations

$$\omega_0, \omega_1$$

such that, just as for τ_0, τ_1,

$$\mathbf{A}_{\mathbf{a}_b,\mathbf{a}_w}(\Gamma') = \mathbf{A}_{\mathbf{a}_b,\mathbf{a}_w}(\omega_0\Gamma') + \mathbf{A}_{\mathbf{a}_b,\mathbf{a}_w}(\omega_1\Gamma')$$

for any graph Γ'. To maintain the flow of the argument, we will define these operations ω_0, ω_1 after we conclude the proof of Lemma 5.1.8. The important property of the operations ω_0, ω_1 is summarized by the following lemma

Lemma 5.3.8. *Let Γ_ζ be as defined above. Let $n_E = |E_G(\Gamma_\zeta)|$, and let $\sigma \in \{0,1\}^{n_E}$ be a binary string. Define the composition*

$$\omega_\sigma = \omega_{\sigma_1} \cdots \omega_{\sigma_{n_E}}.$$

We have

$$\left| \prod_{e \in E_G(\Gamma_\zeta)} \mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e, \omega_\sigma \Gamma_\zeta) \right| \prec \psi^{\left|E_o(\Gamma_\zeta)\right| + \sum \sigma} N^{-2\epsilon_D \sum \sigma} \tag{5.39}$$

(where $\sum \sigma := \sum_{r=1}^{n_E} \sigma_r$ is the number of 1s in σ), and moreover, If $\sum \sigma < \ell$, then

$$\mathbb{E}\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(\omega_\sigma \Gamma_\zeta) = 0.$$

This is to say that a sequence of expansions ω_σ only produce graphs whose expected evaluation are 0 or whose evaluations are less by a factor of $\psi^\ell N^{-2\epsilon_D \ell}$ than the number of off-diagonal edges in Γ would suggest.

Now let σ and ζ be such that the expression (5.41) below is nonzero. Lemma 5.3.8 and equation (5.38) tell us that

$$\left|\prod_{e\in E(\Gamma_\zeta)} \mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\omega_\sigma\Gamma_\zeta)\right| \prec \psi^{\left|E_o(\Gamma_\zeta)\right|+\ell} \le \psi^{p-r+\frac{1}{2}(\ell-\widetilde{\ell})+\ell} N^{-\epsilon_D \ell} \tag{5.40}$$

where we have used that $\left|G_{2,\mu\nu}^{E(T)}\right|(1-\delta_{\mu\nu}) \prec \psi$ by Lemma 5.3.2. Equation (5.40) then yields, also using $\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\omega_\sigma\Gamma_\zeta) \prec N^{-1/2}$ for $e \in E_X(\Gamma) = E_X(\omega_\sigma\Gamma_\zeta)$, we get

$$\begin{aligned} &\mathbb{E}\sum_{\mathbf{a}_w} \prod_{e\in E(\omega_\sigma\Gamma_\zeta)} \mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\omega_\sigma\Gamma_\zeta) \\ &\quad = \mathbb{E}\sum_{\mathbf{a}_w} \prod_{e\in E_G(\omega_\sigma\Gamma_\zeta)} \mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\omega_\sigma\Gamma_\zeta) \cdot O_\prec\left(N^{-\frac{1}{2}|E_X(\Gamma)|}\right). \end{aligned}$$

Using Lemma 5.3.7 to estimate $|V_w(\Gamma_\zeta)|$ which is the number of indices in $\mathbf{a}_w$, we get

$$\begin{aligned} &= \mathbb{E}\left(\sum_{\mathbf{a}_w} 1\right) \cdot O_\prec\left(\psi^{p-r}\psi^{\ell+\frac{1}{2}(\ell-\widetilde{\ell})} N^{-2\ell\epsilon_D}\right) \cdot O_\prec\left(N^{-\frac{1}{2}|E_X(\Gamma)|}\right) \\ &\prec N^{\frac{1}{2}|E_X(\Gamma)|-\frac{r}{2}+\frac{\ell}{2}} \cdot \left(\psi^{p-r}\psi^{\ell+\frac{1}{2}(\ell-\widetilde{\ell})} N^{-2\ell\epsilon_D}\right) \cdot \left(N^{-\frac{1}{2}|E_X(\Gamma)|}\right) \cdot N^{-\frac{1}{2}\left|V_b^*\right|} \\ &= \left(N^{-\frac{r}{2}}\psi^{-r}\right)\left(\psi^{\frac{3}{2}\ell} N^{\frac{\ell}{2}} N^{-2\ell\epsilon_D}\right)\left(\psi^{p-\frac{1}{2}\widetilde{\ell}}\right) \cdot N^{-\frac{1}{2}\left|V_b^*\right|}. \end{aligned} \tag{5.41}$$

Now we use that $N^{-\frac{1}{2}} \lesssim \psi$ for any $\eta_0 \le 1$, that $\psi^3 N^{-2\epsilon_D} N \le N^{-8\delta}$, that $\widetilde{\ell} \le p$, and that $\psi \le N^{-8\delta}$ to bound the above by

$$\lesssim \psi^{p/2}\left(N^{-4\delta}\right)^\ell \left(\sqrt{\psi}\right)^{p-\widetilde{\ell}} \cdot N^{-\frac{1}{2}\left|V_b^*\right|} \le \psi^{p/2}\left(N^{-4\delta}\right)^\ell \left(N^{-4\delta}\right)^{p-\widetilde{\ell}} \cdot N^{-\frac{1}{2}\left|V_b^*\right|}$$

and then $\widetilde{\ell} \le \ell$ and $\sqrt{\psi}N^{-4\delta} \le \psi_{(KM)}$ to bound the above by

$$(\sqrt{\psi}N^{-4\delta})^p N^{-\frac{1}{2}\left|V_b^*\right|} \le \psi_{(KM)}^p N^{-\frac{1}{2}\left|V_b^*\right|}.$$

Now, we may bound equation (5.35), using the definition of $w_{\mathbf{a}_b}$ and that $\|\mathbf{x}\| = 1$, by

$$\begin{aligned} &\left|\mathbb{E}\sum_{\mathbf{a}_b}^* \sum_{\mathbf{a}_w} \mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(\Gamma)\right| \prec \psi_{(KM)}^p N^{-\frac{1}{2}\left|V_b^*\right|} \sum_{\mathbf{a}_b}^* |w_{\mathbf{a}_b}| \\ &\le \psi_{(KM)}^p N^{-\frac{1}{2}\left|V_b^*\right|} \sum_{i_1,\ldots,i_{\left|V_b\setminus V_b^*\right|}} \prod_{\iota=1}^{\left|V_b\setminus V_b^*\right|} \left|\mathbf{x}_{i_\iota}\right|^2 \sum_{j_1,\ldots,j_{\left|V_b^*\right|}} \prod_{\iota=1}^{\left|V_b^*\right|} |\mathbf{x}_{j_\iota}| \\ &= \psi_{(KM)}^p N^{-\frac{1}{2}\left|V_b^*\right|} O_\prec(N^{1/2})^{\left|V_b^*\right|} \end{aligned}$$

by Cauchy-Schwarz, and we conclude the proof of Lemma 5.1.8. □

Proof of Lemma 5.3.6. As in [BEK14], this follows from the fact that if equation (5.36) is not satisfied, then there must be a block of ζ containing only one element j of $\pi^{-1}(i)$, so that, letting e_0 be the edge connecting i and j, equation (5.37) factors as

$$\begin{aligned}\mathbb{E}\prod_{e\in E(\Gamma_\zeta)}\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\Gamma_\zeta) &= \mathbb{E}X_{a_ia_j}\prod_{e\in E(\Gamma_\zeta)\setminus\{e_0\}}\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\Gamma_\zeta)\\ &= \mathbb{E}X_{a_ia_j}\mathbb{E}\prod_{e\in E(\Gamma_\zeta)\setminus\{e_0\}}\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\Gamma_\zeta)=0,\end{aligned}$$

where the second equality follows from the facts that for every edge $e \in E_G(\Gamma_\zeta)$, we have $\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\Gamma_\zeta)$ independent of X_{a_i,a_j}, and that our assumption in this lemma implies for each $e \in E_X(\Gamma_\zeta)\setminus\{e_0\}$ that $\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\Gamma_\zeta)$ is independent of X_{a_i,a_j}. □

Proof of Lemma 5.3.7. Because every edge in $\widetilde{E_o}(\Gamma)$ joins $j \in \pi^{-1}(i_0)$ and $j \in \pi^{-1}(i)$ for $i \in V_b \setminus \{i_0\}$, and because the partition ζ_{b} only makes identifications between vertices joined by $e \in \widetilde{E_o}(\Gamma)$, it follows that the collection of sets $\{\pi^{-1}(i) : i \in V_b^*\}$ is disjoint—this is slightly nontrivial, because the preimages of π, though being disjoint before identification of vertices by ζ_{b}, may no longer be disjoint after some vertices are identified; one may simply regard $\pi^{-1}(i) \subseteq V_w(\Gamma_{\zeta_{\mathrm{b}}})$ as the set of all neighbors of the vertex i in the graph $\Gamma_{\zeta_{\mathrm{b}}}$, and then it becomes clear that the sets $\pi^{-1}(i)$ are disjoint for distinct $i \in V_b \setminus \{i_0\}$.

Introduce one more partition ζ_{int} on $V_w(\Gamma_{\zeta_{\mathrm{b}}})$, which is the least restrictive partition on $V_w(\Gamma_{\zeta_{\mathrm{b}}}) \setminus V_w^*(\Gamma_\zeta)$ which identifies two vertices j_1, j_2 whenever ζ identifies j_1 and j_2 and $j_1, j_2 \in \pi^{-1}(i)$ for the same $i \in V_b$. Here "int" stands for *internal.*

Note that the graph $\Gamma_{\zeta_{\mathrm{b}}}$ has $|V_w(\Gamma) - r|$ vertices. By Lemma 5.3.6, every block ζ_{int} has size at least 2, but for each $i \in V_b^*$, there is a distinct block of ζ_{int} with size at least 3.

Therefore, we see that

$$
\begin{aligned}
|V_w(\Gamma_\zeta)| &\le \ell + |V_w(\Gamma_{\zeta_{\mathrm{int}}})| \le \ell + \frac{|V_w(\Gamma_{\zeta_{\mathrm{b}}}) \setminus V_w^*(\Gamma_{\zeta_{\mathrm{b}}})| - |V_b^*|}{2} \\
&\le \ell + \frac{|V_w(\Gamma)| - \ell - r - |V_b^*|}{2} \\
&= \frac{|E_X(\Gamma)|}{2} - \frac{r}{2} + \frac{\ell}{2} - \frac{|V_b^*|}{2}.
\end{aligned}
$$

□

Remark 5.3.9. The fact that the preimages $\pi^{-1}(i)$ are disjoint for distinct $i \neq i_0$ in the above proof makes the proof somewhat easy to get the $-\frac{|V_b^*|}{2}$ term. In section 5.4, we will encounter graphs where this is not the case and we will have to work a litle harder.

5.3.3 The operations ω_0 and ω_1

The definitions of ω_0, ω_1 and the proof of Lemma 5.3.8. Recall that E is a $(\mathcal{I}_K \cup \mathcal{I}_M) \times (\mathcal{I}_K \cup \mathcal{I}_M)$ matrix, and that for $T \subseteq \mathcal{I}_K$, we have defined $E^{(T)}$ as the minor of E by removing the rows and columns indexed by T.

Now, for $\nu \in \mathcal{I}_N$, we will define the diagonal matrices $E^{(\nu)}$ and $\Phi^{(\nu)}$ through

$$
E_{ii}^{(\nu)} := \left(1 + \mathcal{J}_{ii}^{-2} \sum_{\mu \in \mathcal{I}_N \setminus \{\nu\}} (\widetilde{B}X)_{i\mu}^2 \right)^{-1/2}, \quad \Phi_{ii}^{(\nu)} := \mathcal{J}_{ii}^{-2} (\widetilde{B}X)_{i\nu}^2.
$$

Notice that $\left\| \Phi^{(\nu)} \right\| \prec N^{-1-2\epsilon_D}$. See that

$$
\begin{aligned}
(E^{-2})_{ii} &= 1 + \mathcal{J}_{ii}^{-2} \left\| \mathbf{e}_i^* \widetilde{B} X \right\|^2 = 1 + \mathcal{J}^{-2} \sum_{\mu \in \mathcal{I}_N} (\widetilde{B}X)_{i\mu}^2 \\
&= \left(E_{ii}^{(\nu)} \right)^{-2} + \Phi_{ii}^{(\nu)}.
\end{aligned} \tag{5.42}
$$

Therefore we are able by a resolvent expansion, for any index set $T \subseteq \mathcal{I}_K \cup \mathcal{I}_M$, to write

$$
G^{E^{(T)}} = G^{E^{(T\nu)}} + G^{E^{(T\nu)}} \Phi^{(\nu)} G^{E^{(T)}}. \tag{5.43}
$$

We similarly define $E^{(\nu_1 \cdots \nu_n)}$ for $\nu_1, \ldots, \nu_n \in \mathcal{I}_N$, and also, for $T \subseteq \mathcal{I}_K$, we define $E^{(T\nu_1 \cdots \nu_n)}$ to be the obvious minor of $E^{(\nu_1 \cdots \nu_n)}$.

At this point, fix a graph Γ_ζ. We now confront the difference between our context and that of [BEK14]: the partition ζ may have singletons, and still $\mathbb{E}W_{\mathbf{a}_b,\mathbf{a}_w}(\Gamma_\zeta) \neq 0$, the reason being that all the resolvent entries $G_{2,\mu_0,\mu_1}^{E^{(\mathbf{a}_b)}}$ appearing in $E_G(\Gamma_\zeta)$ are, despite being maximally expanded, not independent of $\mathbf{e}_\alpha^* X$, because the population matrix $E^{(\mathbf{a}_b)}$ depends on $\mathbf{e}_\alpha^* X$. This leads us to perform this new binary-tree-indexed expansion. Fix an ordering $j_1, \ldots, j_\ell$ of $V_w^*(\Gamma_\zeta)$, $\ell := |V_w^*(\Gamma_\zeta)|$ (recall $V_w^*(\Gamma_\zeta)$ are the vertices which are singletons of ζ). We define, analogously to the operations τ_0 and τ_1, two operations ω_0 and ω_1. The first $|E_G(\Gamma_\zeta)|$ applications of ω_0 and ω_1 are as follows: $\omega_0(\Gamma_\zeta)$ replaces the first edge in $E_G(\Gamma_\zeta)$, whose evaluation is, say, $G_{2,\mu_0,\mu_1}^{E^{(\mathbf{a}_b)}}$, with an edge with evaluation $G_{2,\mu_0,\mu_1}^{E^{(\mathbf{a}_b a_{j_1})}}$—note that this edge is now in fact independent of $X_{\alpha,a_{j_1}}$—while $\omega_1(\Gamma_\zeta)$ replaces it with an edge, which we will still call a G edge, with evaluation

$$\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e, \omega_1(\Gamma_\zeta)) = \left(G_2^{E^{(\mathbf{a}_b a_{j_1})}} \Phi^{(a_{j_1})} G_2^{E^{(\mathbf{a}_b)}} \right)_{\mu_0,\mu_1}.$$

Note that this edge e has $\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e, \omega_1(\Gamma_\zeta)) = O_\prec(\psi^2 N^{-2\epsilon_D})$, an improvement over $O_\prec(\psi)$. Indeed, by Lemma 5.3.2,

$$\begin{aligned} \left| \left(G_2^{E^{(\mathbf{a}_b a_{j_1})}} \Phi^{(a_{j_1})} G_2^{E^{(\mathbf{a}_b)}} \right)_{\mu_0,\mu_1} \right| &= \left| \sum_{i \in \mathcal{I}_M} G_{2,\mu_0 i}^{E^{(\mathbf{a}_b a_{j_1})}} \Phi_{ii}^{(a_{j_1})} G_{2,i\mu_1}^{E^{(\mathbf{a}_b)}} \right| \\ &\prec m_E \psi N^{-1-2\epsilon_D} \psi \lesssim \psi^2 N^{-2\epsilon_D}. \end{aligned} \tag{5.44}$$

Remark 5.3.10. Here we need a bound on $G_{2,i\mu}^E$, which is the reason for the treatment of such an off-diagonal entry in Lemma 5.3.2.

The second application of ω_1 or ω_2 does the same thing for the second edge of $E_G(\Gamma_\zeta)$, and so forth, $|E_G(\Gamma_\zeta)|$ times. Note that $|E_G(\Gamma')| = |E_G(\omega_\iota(\Gamma'))|$ for any graph Γ'. Before we describe the next applications of ω_0 and ω_1, note that if the first $|E_G(\Gamma)|$ applications $\omega_{\iota_1}, \ldots, \omega_{\iota_{|E_G(\Gamma)|}}$ of ω_ι are ω_1, then

$$\mathbb{E}\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(\omega_{\iota_1} \cdots \omega_{\iota_{|E_G(\Gamma_\zeta)|}} \Gamma_\zeta) = \mathbb{E}\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(\omega_1^{|E_G(\Gamma)|} \Gamma_\zeta) = 0$$

for the same reason that singletons in ζ_i led to 0 in Lemma 5.3.6, or in [BEK14]: we now have $X_{i_0 a_{j_1}}$ independent of $\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e, \omega_0^{|E_G(\Gamma_\zeta)|}\Gamma_\zeta)$ for all $e \in E_G(\omega_0^{|E_G(\Gamma_\zeta)|}\Gamma_\zeta)$.

The rest of the applications ω_0 and ω_1 are as follows: say $c|E_G(\Gamma_\zeta)|$ applications of ω_0, ω_1 have already been performed, $1 \leq c \leq \ell - 1$, $c \in \mathbb{N}$. The next $|E_G(\Gamma_\zeta)|$ applications of ω_0 replace each of the $|E_G(\Gamma)|$ edges with evaluation

$$\left(G_2^{E^{(\mathbf{a}_b T_1)}} \Phi^{(\nu_1)} G_2^{E^{(\mathbf{a}_b T_2)}} \Phi^{(\nu_2)} \cdots \Phi^{(\nu_{d-1})} G_2^{E^{(\mathbf{a}_b T_d)}}\right)_{\mu_0 \mu_1} := \mathcal{A}, \tag{5.45}$$

where $T_1, \ldots, T_d \subseteq \{a_{j_1}, \ldots, a_{j_c}\}$ and $d \leq c$, with edges with evaluation

$$\begin{aligned}\big(G_2^{E^{(\mathbf{a}_b T_1 a_{j_{c+1}})}} \Phi^{(\nu_1)} G_2^{E^{(\mathbf{a}_b T_2 a_{j_{c+1}})}} \Phi^{(\nu_2)} \cdots \\ \cdots \Phi^{(\nu_{d-1})} G_2^{E^{(\mathbf{a}_b T_d a_{j_{c+1}})}}\big)_{\mu_0 \mu_1} := \mathcal{A}'.\end{aligned}$$

An application of ω_1 replaces the edge $\mathcal{A}$ with $\mathcal{A}' - \mathcal{A}$, so that by equation 5.43, one sees that $|\mathcal{A}| \prec \psi^d N^{-2(d-1)\epsilon_D}$ and $|\mathcal{A} - \mathcal{A}'| \prec \psi^{d+1} N^{-2d\epsilon_D}$ (likewise, an edge with evaluation $\mathcal{A}_1 + \cdots + \mathcal{A}_n$, each $\mathcal{A}_\iota$ having the form (5.45) for the same value of d_0 of d, is replaced by ω_0 with an edge with evaluation $\mathcal{A}'_1 + \cdots + \mathcal{A}'_n$, and by ω_1 with an edge with evaluation $(\mathcal{A}_1 - \mathcal{A}'_1) + \cdots + (\mathcal{A} - \mathcal{A}'_n)$, which a sum of many terms of the form (5.45), each with the same value $d_0 + 1$ of d). At this point, it is clear that every application of ω_1 improves the evaluation of some G edge by a factor of $\psi N^{-2\epsilon_D}$, which allows us to conclude equation (5.39).

Again, if each of the applications $c|E_G(\Gamma_\zeta)| + 1, \ldots, (c+1)|E_G(\Gamma_\zeta)|$ of ω_ι are ω_0, then

$$\mathbb{E}\mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(\omega_{\iota_1} \cdots \omega_{\iota_{(c+1)|E_G(\Gamma_\zeta)|}} \Gamma_\zeta) = 0$$

since we now have $X_{i_0 a_{j_{c+1}}}$ independent of $\mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}\left(e, \omega_{\iota_1} \cdots \omega_{\iota_{(c+1)|E_G(\Gamma_\zeta)|}} \Gamma_\zeta\right)$ for all $e \in E_G\left(\omega_{\iota_1} \cdots \omega_{\iota_{(c+1)|E_G(\Gamma_\zeta)|}} \Gamma_\zeta\right)$.

Applying this argument for each $c = 0, \ldots, \ell - 1$, we see that

$$\mathbb{E}\mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(\omega_\sigma \Gamma_\zeta) = 0$$

unless σ has ℓ entries equal to 1; that is, unless ℓ of the applications of ω_ι were ω_1, which concludes the definition of ω_0, ω_1 and the proof of Lemma 5.3.8. $\square$

5.3.4 Proof of Lemma 5.3.2

Proof of Lemma 5.3.2. We prove by a bootstrapping argument common in RMT; the proof is easier than many such proofs, though, because we already have a local law for G^{E_M}. Because $\mathbf{e}_s$ and $\mathbf{e}_t$ are standard basis vectors, ie, they have L^1 norm 1, we do not have to find any independence between different resolvent entries, and therefore we do not have to treat high moments.

Fix $z = E + i\eta_0 \in \mathbf{S}^e$ and a constant C. We define, for $L = 1, \dots, N^{4\lceil\delta^{-1}\rceil}$, $z_L = z + i\eta_L$ and $\eta_L = \eta_0 + LN^{-4\delta^{-1}}$ and let

$$\mathbf{A}_L = \left\{ \max_{r_1,r_2,T} \left| G_{2,r_1r_2}^{E(T\setminus\{r_1r_2\})}(z_L) - \Pi_{2,r_1r_2}(z_L) \right| \le N^{-\delta/10} \right\}$$

$$\mathbf{B}_L = \left\{ \max_{r_1,r_2,T} \left| G_{2,r_1r_2}^{E(T\setminus\{r_1r_2\})}(z_L) - \Pi_{2,r_1r_2}(z_L) \right| \le \psi_{(r_1r_2)} \right\}$$

where the max over T ranges over all subsets T of $\mathbf{a}_b$ of size $|T| \le C$.

The main technical result needed for the proof of Lemma 5.3.2 is the following:

Lemma 5.3.11 ($\mathbf{B}_L$ holds with high probability on $\mathbf{A}_L$). *For any constant $D > 0$,*

$$P(\mathbf{A}_L \cap \mathbf{B}_L^C) \le N^{-D}.$$

We may now prove Lemma 5.3.2. First observe, for a general matrix H of bounded operator norm,

$$\left\| (H - z_L I)^{-1} - (H + H_\Delta - z_L I)^{-1} \right\| \lesssim \|H_\Delta\| \eta_L^{-2} \tag{5.46}$$

To start the bootstrapping, we see that for $L = N^{4\lceil\delta^{-1}\rceil}$, equation (5.46) yields

$$\left\| G_2^E(z_L) - G_2^I(z_L) \right\| \prec N^{-1/2}$$

(also note $\left\| \Pi_2^E - \Pi_2^I \right\| \prec N^{-1/2}$). Since $G_2^I - \Pi_2^I$ has the usual local law, we conclude that B_L holds with high probability for $L = N^{4\lceil\delta^{-1}\rceil}$.

For $L \ge 1$, equation (5.46) also shows that

$$P(\mathbf{B}_L \cap \mathbf{A}_{L-1}^C) \le N^{-D}. \tag{5.47}$$

Lemma (5.3.11) and equation (5.47) yield then

$$\begin{aligned}
P(B_1^C) &= P(A_1^C \cap B_1^C) + P(A_1 \cap B_1^C) \\
&\leq P(A_1^C) + N^{-D} \\
&= P(B_2^C \cap A_1^C) + P(B_2 \cap A_1^C) + N^{-D} \\
&\leq P(B_2^C) + 2N^{-D} \\
&\quad\vdots \\
&\leq P(B_{N^4}^C) + 2N^4 N^{-D} \\
&\leq 3N^{4-D}.
\end{aligned}$$

Since D is arbitrary, we are done. □

Proof of Lemma 5.3.11. We let Δ be the graph with $K + |\{s,t\} \setminus \mathcal{I}_K|$ black vertices $V_b(\Delta) = \mathcal{I}_K \cup \{s,t\}$. Even if either of s, t are in $\mathcal{I}_N$, we still include them in V_b; this is inconvenient notation, since we otherwise think of black vertices as taking values in $\mathcal{I}_{K+M}$ and white vertice in $\mathcal{I}_N$, but we will not need it for long. K of the vertices are labeled $1, \ldots, K$, and for each $\alpha = 1, \ldots, K$, we write $a_\alpha = \alpha$, and the $|\{s,t\} \setminus \mathcal{I}_K|$ other vertices shall be labeled s and t and shall satisfy $a_s = s$, $a_t = t$. Δ has one edge connecting s and t. Here we are trying to preserve some continuity with 5.3.1, wherein there must be a distinction between the vertices of the graph i, and the values a_i in $\mathcal{I}_K \cup \mathcal{I}_M$ that the vertices take—we are working with standard basis vectors now, so every vertex takes only one value and this distinction is not really necessary.

The main difference between this graph and the graph of section 5.3.1 is that it can (if s or $t \in \mathcal{I}_N$) contain G edges initiating and/or terminating on vertices landing in $\mathcal{I}_N$. This does not actually affect the construction of section 5.3.1; the identity (5.29) still holds, and in addition to the identity (5.30) we have

$$G_{i\mu} = G_{ii} G_{\mu\mu}^{(i)} \left(-X_{i\mu} + \left(X G^{(i\mu)} X \right)_{i\mu} \right)$$

The edges $-X_{i\mu}$ and $\left(XG^{(i\mu)}X\right)_{i\mu}$ above will both be refered to as R-groups because of the role that they play. Now, we construct the tree of operations just like in subsection 5.3.1 and write

$$Y(\Delta) = \mathbb{E} \sum_{\sigma \in L(\mathcal{T})} \sum_{\mathbf{a}_w} \mathcal{A}_{\mathbf{a}_w}(\Theta_\sigma),$$

where we have removed $\mathbf{a}_b$ from the subscript of $\mathcal{A}$ since $\mathbf{a}_b$ is now a constant; also, $w_{\mathbf{a}_b}(\Delta) = 1$ identically.

Now to estimate the trivial leaves $\sigma \in L(\mathcal{T})$, we need a bound on off-diagonal edges

$$\left| G_{2,r_1 r_2}^{E^{(T)}} \right| \prec N^{-\epsilon_0} \tag{5.48}$$

for some $\epsilon_0 > 0$—which is given by the bootstrapping assumption $\mathbf{A}_L$—and a bound on the R-groups, which we collect in the following lemma:

Lemma 5.3.12. *We have the following bounds:*

1. *For* $r_1, r_2 \in \mathcal{I}_{K+M}$,
$$\left| \left(X G_2^{E^{(\mathbf{a}_b)}} X^* \right)_{r_1 r_2} - \delta_{r_1 r_2} m \right| \prec \psi_{(r_1 r_2)}.$$
2. *For* $r_1 \in \mathcal{I}_{K+M}, r_2 \in \mathcal{I}_N$,
$$\left| \left(X G_2^{E^{(\mathbf{a}_b)}} X \right)_{r_1 r_2} \right| \prec \psi_{(r_1 r_2)}.$$
3. *For* $r_1, r_2 \in \mathcal{I}_N$,
$$\left| \left(X^* G_2^{E^{(\mathbf{a}_b)}} X \right)_{r_1 r_2} - \delta_{r_1 r_2} N^{-1} \sum_{i \in \mathcal{I}_M} \Pi_{2,ii} \right| \prec \psi_{(r_1 r_2)}.$$

Now the bound on the trivial leaves follows just as it did in Section 5.3.1, and we can now treat the non-trivial leaves, in which all off-diagonal edges are replaced with R-groups and all diagonal edges are maximally expanded. Because again we are not dealing with the fully general situation of high moments of generalized resolvent entries, we do not need to exploit any independence, and we can simply bound the R-groups individually rather than going through the process of identifying white vertices as in Lemma 5.1.8.

First note that every self-loop e has

$$|\mathcal{A}_{\mathbf{a}_w}(e,\Theta_\sigma)| \prec 1$$

since it may be expanded by operatrion (c) of [BEK14] just as in that paper; the bounds necessary to perform the truncation in this operation are provided by Lemma 5.3.12.

We proceed by cases on r_1, r_2. If $r_1, r_2 \in \mathcal{I}_M$, then there is one non-trivial leaf σ such that Θ_σ has a single R-group

$$R_* = \sum_{\mu\nu} X_{r_1\mu} G_{2,\mu\nu}^{E^{(\mathbf{a}_b)}} X_{r_2\nu}$$

(where for brevity we are conflating an edge with its evaluation) which satisfies $|R_* - \delta_{r_1r_2} m| \prec \psi_{(r_1r_2)}$ by Lemma 5.3.12; every other non-trivial leaf has at least two R-groups R_1 and R_2 of the form

$$R_\iota = \left| \sum_{\mu\nu} X_{r_\iota\mu} G_{2,\mu\nu}^{E^{(\mathbf{a}_b)}} X_{\alpha_\iota\nu} \right|$$

for some $\alpha_1, \alpha_2 \in \mathcal{I}_K$. Every other self-edge and R-group is bounded by $O_\prec(1)$. Since $R_1R_2 \prec \psi_{(MK)}^2 \leq \psi_{(r_1r_2)}$, we are done with the case $r_1, r_2 \in \mathcal{I}_M$.

The case of $r_1 \in \mathcal{I}_M$, $r_2 \in \mathcal{I}_N$, the symmetric case, and the case of $r_1, r_2 \in \mathcal{I}_N$ are treated almost identically, using the identity (5.48) and Lemma 5.3.12, with only the additional input that $X_{r_1r_2} \prec N^{-1/2}$.

If $r_1 \in \mathcal{I}_K$ and $r_2 \in \mathcal{I}_M$, then every non-trivial leaf σ has an R-group of the form

$$\left| \sum_{\mu\nu} X_{r_2\mu} G_{2,\mu\nu}^{E^{(\mathbf{a}_b)}} X_{\alpha\nu} \right| \prec \psi_{(r_1r_2)}$$

for some $\alpha \in \mathcal{I}_K$, so that again we are done.

The case of $r_1 \in \mathcal{I}_K, r_2 \in \mathcal{I}_N$ is treated similarly, with the same adjustments as for the case $r_1 \in \mathcal{I}_M, r_2 \in \mathcal{I}_N$.

Finally if $r_1, r_2 \in \mathcal{I}_K$, exactly one non-trivial leaf has exactly one R-group

$$R_* = \sum_{\mu\nu} X_{r_1\mu} G_{2,\mu\nu}^{E^{(\mathbf{a}_b)}} X_{r_2\nu}$$

which satisfies $\left|R_* - \delta_{r_1 r_2} m \prec N^{-\delta}\right| \prec \psi_{(KK)}$ and every other non-trivial leaf has at least one off-diagonal R-group

$$R_* = \left|\sum_{\mu\nu} X_{\alpha\mu} G_{2,\mu\nu}^{E^{(\mathbf{a}_b)}} X_{\beta\nu}\right| \prec \psi_{(NN)}$$

for some $\alpha \neq \beta \in \mathcal{I}_K$.

□

Proof of Lemma 5.3.12. We separate into cases:

1. $r_1, r_2 \in \mathcal{I}_M$,
2. $r_1, r_2 \in \mathcal{I}_N$,
3. $r_1 \in \mathcal{I}_M, r_2 \in \mathcal{I}_N$,
4. $r_1 \in \mathcal{I}_K, r_2 \in \mathcal{I}_M$,
5. $r_1 \in \mathcal{I}_K, r_2 \in \mathcal{I}_N$,
6. $r_1, r_2 \in \mathcal{I}_K$.

We prove each item separately.

1. Since $G_2^{E^{(\mathbf{a}_b)}}$ is a matrix for which the usual local law holds ($\mathcal{I}_K \subseteq \mathbf{a}_b$, so that the population matrix $E^{(\mathbf{a}_b)}$ is independent of the randomness), the usual proof holds, which we outline. Using the large deviation bounds (see lemma 3.6 of [BK18]) for real-valued, independent, centered, variance 1 random variables $(x_i)_{i\in\mathcal{K}}, (y_i)_{i\in\mathcal{K}}$ having $x_i \prec 1$ and $y_i \prec 1$ and deterministic complex constants $(b_{ij})_{i,j\in\mathcal{K}}$ for some finite index

set $\mathcal{K}$,

$$\begin{aligned}
\sum_i x_i b_{ii} &\prec \left(\sum_i |b_{ii}|^2\right)^{1/2} \\
\sum_{i\neq j} x_i b_{ij} x_j &\prec \left(\sum_{i\neq j} |b_{ij}|^2\right)^{1/2} \\
\sum_{ij} x_i b_{ij} y_j &\prec \left(\sum_{ij} |b_{ij}|^2\right)^{1/2}
\end{aligned} \tag{5.49}$$

and also using that $N^{1/2}X_{r_\iota \mu_{\iota'}}$ is centered, variance 1, and $O_\prec(1)$, we obtain

$$\begin{aligned}
&\sum_{\mu_1,\mu_2} X_{r_1\mu_1} G^{E^{(\mathbf{a}_b)}}_{2,\mu_1\mu_2} X_{r_2\mu_2} - \delta_{r_1 r_2} m_E \\
=& \sum_{\mu_1\neq\mu_2} X_{r_1\mu_1} G^{E^{(\mathbf{a}_b)}}_{2,\mu_1\mu_2} X_{r_2\mu_2} + \left(\sum_{\mu_1} X_{r_1\mu_1}\left(G^{E^{(\mathbf{a}_b)}}_{2,\mu_1\mu_1} - m\right) X_{r_2\mu_1}\right) \\
&\quad + \sum_{\mu_1} m(X_{r_1\mu_1}X_{r_2\mu_1} - \delta_{r_1r_2}N^{-1}) \\
=& N^{-1}O_\prec\left(\left(\sum_{\mu_1\neq\mu_2}\left|G^{E^{(\mathbf{a}_b)}}_{2,\mu_1\mu_2}\right|^2\right)^{1/2}\right) + N^{-1}O_\prec\left(\left(\sum_{\mu_1}\left|G^{E^{(\mathbf{a}_b)}}_{2,\mu_1\mu_1} - m\right|^2\right)^{1/2}\right) \\
&\quad + N^{-1}O_\prec\left(\left(\sum_{\mu_1}|m|^2\right)^{1/2}\right) \\
=& O_\prec(\psi_{(NN)}) + O_\prec(N^{-1/2}\psi_{(NN)}) + O_\prec(N^{-1/2}) \\
=& O_\prec(\psi_{(NN)})
\end{aligned}$$

by Lemma 5.2.2.

2. This bound is proven just like item 1, using that $\left|G^{E^{(\mathbf{a}_b)}}_{2,i_1i_2} - \Pi_{2,i_1i_2}\right| \prec \psi_{(MM)}$ (which is also a consequence of Lemma 5.2.2) in place of $\left|G^{E^{(\mathbf{a}_b)}}_{2,\mu_1\mu_2} - \Pi_{2,\mu_1\mu_2}\right| \prec \psi_{(NN)}$.

3. This proof is again like items 1 and 2, using $\left|G^{E^{(\mathbf{a}_b)}}_{2,i\mu} - \Pi_{2,i\mu}\right| \prec \psi_{(MN)}$.

4. The reason this differs from items 1, 2 and 3 is that $G_2^{(\mathbf{a}_b\cup\mathcal{I}_K)}$ is not idependent of $X_{\alpha\mu}$

for $\alpha \in \mathcal{I}_K$. We write

$$\begin{aligned}
\left|\left(XG_2^{E^{(\mathbf{a}_b)}}X^*\right)_{r_1r_2}\right| &= \left|\sum_{\mu_1,\mu_2} X_{r_1\mu_1}G_{2,\mu_1\mu_2}^{E^{(\mathbf{a}_b)}}X_{r_2\mu_2}\right| \\
&\leq \left|\sum_{\mu_1,\mu_2} X_{r_1\mu_1}G_{2,\mu_1\mu_2}^{E^{(\mathbf{a}_b\mu_2)}}X_{r_2\mu_2}\right| \\
&\quad + \left|\sum_{\mu_1,\mu_2} X_{r_1\mu_1}\left(G_2^{E^{(\mathbf{a}_b\mu_2)}}\Phi^{(\mu_2)}G_2^{E^{(\mathbf{a}_b)}}\right)_{\mu_1\mu_2}X_{r_2\mu_2}\right|.
\end{aligned} \tag{5.50}$$

In the first term, the resolvent is independent of both X factors, so we treat it as we did before, getting the bound $\psi_{(NN)} \leq \psi_{(KM)}$. In the second, we do not have independence with respect to $X_{r_2\mu_2}$, so we use a weaker large deviation bound: note that, using equation (5.44), we may crudely bound

$$Y_{\mu_1} := \sum_{\mu_2}\left(G_2^{E^{(\mathbf{a}_b\mu_2)}}\Phi^{(\mu_2)}G_2^{E^{(\mathbf{a}_b)}}\right)_{\mu_1\mu_2}X_{r_2\mu_2} \prec N^{1/2}\psi_{(NN)}^2 N^{-2\epsilon_D} \leq \psi_{(KM)}.$$

Also see that Y_{μ_1} is independent of $X_{r_1\mu_1}$. Thus, we bound the second term of (5.50):

$$\left|\sum_{\mu_1} X_{r_1\mu_1}Y_{\mu_1}\right| \prec N^{-1/2}\left(\sum_{\mu_1}|Y_{\mu_1}|^2\right)^{1/2} \prec \psi_{(KM)} \tag{5.51}$$

as desired.

5. This proof is identical to the proof of item 4, noting that the implicit range of i in the sum excludes $\mathcal{I}_K$ so that the factor X_{r_2i} is independent of $G_{2,\mu i}^{E^{(\mathbf{a}_b)}}$.

6. We may write, using parentheses to demonstrate the provenance of the terms,

$$\begin{aligned}
G^{E^{(\mathbf{a}_b)}} &= \left(G^{E^{(\mathbf{a}_b\mu_1)}}\right) + \left(G^{E^{(\mathbf{a}_b\mu_1)}}\Phi^{(\mu_1)}G^{E^{(\mathbf{a}_b)}}\right) \\
&= \left(G^{E^{(\mathbf{a}_b\mu_1\mu_2)}} + G^{E^{(\mathbf{a}_b\mu_1\mu_2)}}\Phi^{(\mu_2)}G^{E^{(\mathbf{a}_b\mu_1)}}\right) \\
&\quad + \left(G^{E^{(\mathbf{a}_b\mu_1\mu_1)}}\Phi^{(\mu_2)}G^{E^{(\mathbf{a}_b\mu_2)}} + O_\prec(N^{-1+3\epsilon_0-4\epsilon_D})\right) \\
:&= A_1 + A_2 + A_3 + O_\prec(N^{-1+3\epsilon_0-4\epsilon_D})
\end{aligned} \tag{5.52}$$

where in the last line follows from

$$G^{(T_1)}\Phi^{(\mu_2)}G^{(T_2)}\Phi^{(\mu_1)}G^{(T_3)} = O_\prec(\psi^3 N^{-4\epsilon_D}) = O(N^{-1+3\epsilon_0-4\epsilon_D})$$

which is proven similarly to equation (5.44). Thus, to establish this last item of Lemma 5.3.12, we must establish

$$\left|\sum_{\mu_1,\mu_2} X_{r_1\mu_1} A_\iota X_{r_2\mu_2}\right| \prec \psi_{(KK)} \tag{5.53}$$

for $\iota = 1,2,3$—each of these bounds is treated just as one of the previous items—and

$$\left|\sum_{\mu_1,\mu_2} X_{r_1\mu_1} O_\prec(N^{-1+3\epsilon_0-4\epsilon_D}) X_{r_2\mu_2}\right| \tag{5.54}$$

but the above is $O_\prec(N^{3\epsilon_0-4\epsilon_D}) \leq \psi_{(KK)}$ somewhat naively, so that we are done.

□

5.4 The Proof of Lemma 5.1.4

We have the following lemma:

Lemma 5.4.1. *The resolvents of $(DX)^\#$ and $(EX)^\#$ are close at the spectral edge: for $z \in \mathbf{S}^e$, we have*

$$\left|\underline{G^E} - \underline{G^D}\right| \prec \frac{N^{-\delta}}{N\eta_0}. \tag{5.55}$$

Proof of Lemma 5.1.4. An immediate corollary of Lemma 5.1.3 is that $(EX)^\#$ satisfies the weak level repulsion and $\epsilon_0/2$-valley conditions. The proof is then concluded in exactly the same way that the proof of Lemma 5.1.3 was. □

Proof of Lemma 5.4.1. Use a Taylor expansion to obtain

$$\begin{aligned} D &= E + \sum_{a=1}^{5} \Lambda^a \mathbf{C}_a + O_\prec(N^{-3}) \\ &= \sum_{a=0}^{5} \Lambda^a \mathbf{C}_a + O_\prec(N^{-3}) \end{aligned} \tag{5.56}$$

where

$$\mathbf{C}_a = \frac{(-1)^a}{a!} \frac{(2a-1)!!}{2^a} E^{2a+1}.$$

Only note that $\mathbf{C}_a$ is power of E (up to bounded deterministic factors). We moreover have by a resolvent expansion

$$G_2^D = G_2^E + \sum_{k=1}^{\overline{k}} (G_2^E \Lambda)^k G_2^E + (G_2^E \Lambda)^{\overline{k}} G_2^D \tag{5.57}$$

Consider the following lemma, proven momentarily:

Lemma 5.4.2. *We have for any integer $k \geq 1$ and any $O_\prec$-bounded random variables $\mathcal{C}_i$ which are independent of X_M and have $\mathcal{C}_\alpha = 1$ for $\alpha \in \mathcal{I}_K$,*

$$N^{-1} \sum_{i \in \mathcal{I}_K \cup \mathcal{I}_M} \mathcal{C}_i \mathbf{e}_i^* G (\Lambda G)^{k-1} \mathbf{e}_i \prec \psi^{(k+1)/2}.$$

It is then a consequence of Lemma 5.4.2 that

$$\left\| (G_2^E \Lambda)^{\overline{k}-1} G_2^E \right\| \prec N (N^{-1/3+\epsilon_0})^{\overline{k}/2}$$

so that if $\overline{k} = 100$, we may truncate equation (5.57) as

$$G_2^D = G_2^E + \sum_{k=1}^{\overline{k}} (G_2^E \Lambda)^k G_2^E + O_\prec \left(N^{-3} \right) \tag{5.58}$$

where $O_\prec(N^{-3})$ is a matrix of operator norm $O_\prec(N^{-3})$. Moreover, using Lemma 5.4.2 to bound the operator norm by the trace, see that

$$\left\| G_2^E \right\| + \sum_{k=1}^{\overline{k}} \left\| (G_2^E \Lambda)^k G_2^E \right\| \prec N\psi \lesssim \eta_0^{-1}.$$

Consider also the following lemma:

Lemma 5.4.3. *For any integers $a, b \geq 0$ and $k \geq 1$ satisfying $a + b + k \geq 2$, we have*

$$N^{-1} \sum_{i \in \mathcal{I}_K \cup \mathcal{I}_M} \mathbf{e}_i^* \mathbf{C}_a \Lambda^a (G (\Lambda G)^{k-1}) \Lambda^b \mathbf{C}_b \mathbf{e}_i \prec \frac{N^{-\delta}}{N\eta_0}.$$

Thus we may write $\underline{G^E} - \underline{G^D}$ as

$$\begin{aligned}(M+K)^{-1} \sum_{i \in \mathcal{I}_K \cup \mathcal{I}_M} (G^E - G^D)_{ii} &= z(M+K)^{-1} \sum_{i \in \mathcal{I}_K \cup \mathcal{I}_M} \left(E G_2^E E - D G_2^D D \right)_{ii} \\ &= z(M+K)^{-1} \sum_{i \in \mathcal{I}_K \cup \mathcal{I}_M} \sum_{a=0}^{5} \sum_{b=0}^{5} \sum_{k=0}^{\overline{k}} \mathbf{1}_{a+b+k \geq 1} \left(\mathbf{C}_a \Lambda^a (G_2^E \Lambda)^k G_2^E \Lambda^b \mathbf{C}_b \right)_{ii} \\ &\quad + O_\prec \left(N^{-3} \eta_0^{-2} \right).\end{aligned}$$

Since $N^{-3}\eta_0^{-2} \prec \frac{N^{-\delta}}{N\eta_0}$, we may conclude. □

The rest of this section is devoted to the proofs of Lemmas 5.4.3 and 5.4.2. We break their proof into four sublemmas:

Lemma 5.4.4. *If $a, b \geq 0$ and $k \geq 1$ are integers with $a + b + k \geq 2$, then*

$$N^{-1}\left|\sum_{i\in\mathcal{I}_M} \mathbf{e}_i^* \mathbf{C}_a \Lambda^a (G(\Lambda G)^{k-1})\Lambda^b \mathbf{C}_b \mathbf{e}_i\right| \prec \frac{N^{-\delta}}{N\eta_0}. \tag{5.59}$$

Lemma 5.4.5. *We have for any integer $k \geq 3$,*

$$N^{-1}\left|\sum_{i\in\mathcal{I}_M} \mathcal{C}_i \mathbf{e}_i^* G(\Lambda G)^{k-1}\mathbf{e}_i\right| \prec \psi^{\frac{k+1}{2}}. \tag{5.60}$$

Lemma 5.4.6. *We have for $k = 2$,*

$$N^{-1}\left|\sum_{i\in\mathcal{I}_M} \mathcal{C}_i \mathbf{e}_i^* G(\Lambda G)^{k-1}\mathbf{e}_i\right| \prec \psi^{\frac{k+1}{2}}. \tag{5.61}$$

Lemma 5.4.7. *We have for any integer $k \geq 2$ and $\alpha \in \mathcal{I}_K$,*

$$\left|\mathbf{e}_\alpha^* G(\Lambda G)^{k-1}\mathbf{e}_\alpha\right| \prec \psi^{\frac{k-1}{2}} \tag{5.62}$$

In the above $\mathcal{C}_i$ is as it was in the statement of Lemma 5.4.2. We only ever actually need $\mathcal{C}_i = E_{ii}$ or $\mathcal{C}_i = 1$. .

Proof of Lemma 5.4.3. Lemma 5.4.3 is a consequence of Lemma 5.4.4 as well as the following result: if $a, b \geq 0$ and $k \geq 1$ are integers with $a + b + k \geq 2$ and $\alpha \in \mathcal{I}_K$, then

$$N^{-1}\left|\mathbf{e}_\alpha^* \mathbf{C}_a \Lambda^a (G(\Lambda G)^{k-1})\Lambda^b \mathbf{C}_b \mathbf{e}_\alpha\right| \prec \frac{N^{-\delta}}{N\eta_0} \tag{5.63}$$

which we prove now. The case of $a = b = 0$ is treated by Lemma 5.4.7. The case of $a \geq 0$ or $b \geq 0$ is trivial, since $\Lambda \mathbf{e}_\alpha = 0$ for $\alpha \in \mathcal{I}_K$. □

Proof of Lemma 5.4.2. This follows from Lemmas 5.4.5, 5.4.6, and 5.4.7. Indeed,

$$\begin{aligned}
&N^{-1}\left|\sum_{i\in\mathcal{I}_K\cup\mathcal{I}_M}\mathcal{C}_i\mathbf{e}_i^*G(\Lambda G)^{k-1}\mathbf{e}_i\right| \\
&\leq N^{-1}\left|\sum_{i\in\mathcal{I}_M}\mathcal{C}_i\mathbf{e}_i^*G(\Lambda G)^{k-1}\mathbf{e}_i\right| + N^{-1}\left|\sum_{\alpha\in\mathcal{I}_K}\mathbf{e}_\alpha^*G(\Lambda G)^{k-1}\mathbf{e}_\alpha\right| \\
&\prec \psi^{\frac{k+1}{2}} + KN^{-1}\psi^{\frac{k-1}{2}} \prec \psi^{\frac{k+1}{2}}
\end{aligned}$$

since K is bounded and $N^{-1}\leq\psi^2$. □

In order to prove Lemmas 5.4.4, 5.4.5, 5.4.6, and 5.4.7, we must first introduce some new graphs, similar to the ones introduced in section 5.3.1.

5.4.1 Graphs with X^2 edges, and the proof of Lemma 5.4.3

Throughout this section, we will work with graphs like the ones introduced in section 5.3.1. The difference now is that the graphs in this section will have a new sort of X edge, and whereas in section 5.3.1 we start with a graph Δ of only black vertices and G edges and then introduce white vertices and X edges through applications of resolvent identities, in this section the new sorts of X edges will be present in the original graph Δ.

Recalling our definition 5.3.3 of a graph, we now adjust it to allow one more possibility for the color of an edge: we add the colors

$$X^2, \overline{X^2}$$

to the range of ξ_1. We extend the definition (5.28) of the evaluation of an edge now:

$$\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\Gamma) = \begin{cases} \left(X_{i\mu}^2 - N^{-1}\right) + 2\left(\mathcal{J}_{ii}^{-1}\sum_{\alpha\in\mathcal{I}_K}\widetilde{B}_{i\alpha}X_{\alpha\mu}\right)X_{i\mu} & \xi_1(e) = X^2 \\ \overline{\left(X_{i\mu}^2 - N^{-1}\right) + 2\left(\mathcal{J}_{ii}^{-1}\sum_{\alpha\in\mathcal{I}_K}\widetilde{B}_{i\alpha}X_{\alpha\mu}\right)X_{i\mu}} & \xi_1(e) = \overline{X^2} \end{cases}$$

where $i := a_{\alpha(e)}, \mu := a_{\beta(e)}$. The motivation for this definition is that

$$\begin{aligned}\Lambda_{ii} &= \left(\|\mathbf{e}_i^* X\|^2 - 1\right) + 2\mathcal{J}_{ii}^{-1}\left\langle \mathbf{e}_i^* \widetilde{B} X, \mathbf{e}_i^* X\right\rangle \\ &= \sum_{\mu \in \mathcal{I}_N} \left(\left(X_{i\mu}^2 - N^{-1}\right) + 2\left(\mathcal{J}_{ii}^{-1} \sum_{\alpha \in \mathcal{I}_K} \widetilde{B}_{i\alpha} X_{\alpha\mu}\right) X_{i\mu}\right).\end{aligned} \tag{5.64}$$

As with edges with color $\xi_1(e) \in \{X, \overline{X}\}$, we assume that edges with color $\xi_1(e) \in \{X^2, \overline{X^2}\}$ in graph Γ have $\alpha(e) \in V_b(\Gamma)$, $\beta(e) \in V_w(\Gamma)$.

Crucially, these new X edges (we will now refer to an edge of any color $\{X, \overline{X}, X^2, \overline{X^2}\}$ as an X edge) have a very similar independence property to before:

Lemma 5.4.8. *Let Γ be a graph and $\mathbf{a}_b$ and $\mathbf{a}_w$ indices for its black and white vertices. Let Γ have X^1 edges $e_1, \ldots, e_{n_K}$ and $f_1, \ldots, f_{n_M^1}$ with $a_{\alpha(e_\iota)} \in \mathcal{I}_K$ and $a_{\alpha(f_\iota)} \in \mathcal{I}_M$, and X^2 edges $g_1, \ldots, g_{n_M^2}$ having $a_{\alpha(g_i)} \in \mathcal{I}_M$. Then, unless*

- *every edge $e \in \{f_\iota\} \cup \{g_\iota\}$ has $a_{\alpha(e)} = a_{\alpha(e')}$ and $a_{\beta(e)} = a_{\beta(e')}$ for some other $e' \neq e \in \{f_\iota\} \cup \{g_\iota\}$* and
- *every $e \in \{e_\iota\}$ has either*
  - *$a_{\alpha(e)} = a_{\alpha(e')}$ and $a_{\beta(e)} = a_{\beta(e')}$ for some $e' \neq e \in \{e_\iota\}$*, or
  - *$a_{\beta(e)} = a_{\beta(e')}$ for some $e' \neq e \in \{g_\iota\}$.*

we have

$$\mathbb{E} \prod_{\iota=1}^{n_K} \mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(e_\iota, \Gamma) \prod_{\iota=1}^{n_M^1} \mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(f_\iota, \Gamma) \prod_{\iota=1}^{n_M^2} \mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(g_\iota, \Gamma) = 0.$$

In other words, in order for a graph's X-edges to have nonzero evaluation, every edge must share both endpoints with a different edge just as in section 5.3, with the exception of X edges whose initial vertex lies in $\mathcal{I}_K$—these edges may be allowed to share only their terminal vertex with another edge (specifically an X^2 edge). The purpose of this lemma and the parallel phenomenon in Section 5.3 is to reduce the number of vertices which may range freely over $\mathcal{I}_M$ or $\mathcal{I}_N$. Since $\mathcal{I}_K$ is of bounded size, the relaxation of the requirement for edges e with $a_{\alpha(e)} \in \mathcal{I}_K$ does not change the argument.

Proof. The essential reason this lemma holds is that the evaluation of an edge $e \in \{e_\iota\} \cup \{f_\iota\}$ depends only on $X_{a_{\alpha(e)},a_{\beta(e)}}$, while the evaluation of edge g_ι depends on $X_{a_{\alpha(g_\iota)},a_{\beta(g_\iota)}}$ as well as $X_{\alpha_1,a_{\beta(g_\iota)}}, \dots, X_{\alpha_K,a_{\beta(g_\iota)}}$, where $\alpha_1, \dots, \alpha_K$ is an enumeration of the elements of $\mathcal{I}_K$.

If the first item does not hold, let this be witnessed by an edge f_ι or g_ι. If the witnessing edge is f_ι (which we assume for brevity to have $\xi_1(f_\iota) = X$), then

$$Z := \prod_{\iota=1}^{n_K} \mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e_\iota, \Gamma) \prod_{\iota=1}^{n_M^1} \mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(f_\iota, \Gamma) \prod_{\iota=1}^{n_M^2} \mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(g_\iota, \Gamma)$$

factors as

$$Z = X_{a_{\alpha(f_\iota)},a_{\beta(f_\iota)}} \widetilde{Z},$$

where $\widetilde{Z}$ is independent of $X_{a_{\alpha(f_\iota)},a_{\beta(f_\iota)}}$ because of the independence of the entries of X. We conclude by the centeredness of the entries of X. If the witnessing edge is rather g_ι (which we assume for brevity to have $\xi_1(f_\iota) = X^2$, then Z factors as

$$\begin{aligned} Z &= \left(\left(X_{i\mu}^2 - N^{-1} \right) + 2 \left(\mathcal{J}_{ii}^{-1} \sum_{\alpha \in \mathcal{I}_K} \widetilde{B}_{i\alpha} X_{\alpha\mu} \right) X_{i\mu} \right) \widetilde{Z} \\ &= \left(X_{i\mu}^2 - N^{-1} \right) \widetilde{Z} + 2 X_{i\mu} \left(\mathcal{J}_{ii}^{-1} \sum_{\alpha \in \mathcal{I}_K} \widetilde{B}_{i\alpha} X_{\alpha\mu} \right) \widetilde{Z} \\ &= \left(X_{i\mu}^2 - N^{-1} \right) \widetilde{Z}_1 + 2 X_{i\mu} \widetilde{Z}_2, \end{aligned}$$

where $\widetilde{Z}_1, \widetilde{Z}_2$ are again independent of $X_{a_{\alpha(f_\iota)},a_{\beta(f_\iota)}}$, and we conclude as before.

If rather the second item does not hold, let this be witnessed by an edge e_ι (which we assume for brevity to have $\xi_1(e_\iota) = X$), which therefore does not share its terminal vertex with any other edge, except perhaps some $f_{\iota'}$, but e_ι and $f_{\iota'}$ may not also share their intial vertices (one lies in $\mathcal{I}_K$ and the other in $\mathcal{I}_M$). Then Z factors as

$$Z = X_{a_{\alpha(f_\iota)},a_{\beta(f_\iota)}} \widetilde{Z}$$

where $\widetilde{Z}$ is independent of $X_{a_{\alpha(f_\iota)},a_{\beta(f_\iota)}}$, and we conclude as before. □

Now, we will construct a graph to model the expressions that we will need to bound for the proof of Lemma 5.4.2. Fix values $a, b \geq 0, k \geq 1$. Similarly to before and to in [BEK14], we begin with the following graph Δ^{pre} whose vertices are partitioned into *black vertices* $V_b(\Delta^{\text{pre}})$ and *white vertices* $V_w(\Delta^{\text{pre}})$:

$$\begin{aligned} V_b(\Delta^{\text{pre}}) :& \{1, \ldots, k\} \times \{1, \ldots, p\} \\ V_w(\Delta^{\text{pre}}) :& \{\bar{1}_1, \ldots, \bar{1}_{a+b}, \bar{2}, \ldots, \bar{k}\} \times \{1, \ldots, p\} \\ E(\Delta^{\text{pre}}) :& \{((k', p'), (k'+1, p')) \text{ for every } (k', p') \in \{1, \ldots, k\} \times \{1, \ldots, p\}\} \\ & \cup \left\{((k', p'), (\bar{k'}, p')) \text{ for every } (k', p') \in \{2, \ldots, k\} \times \{1, \ldots, p\}\right\} \\ & \cup \{((1, p'), (\bar{1}_c, p')) \text{ for every } p' \in \{1, \ldots, p\}, c \in \{1, \ldots, a+b\}\} \end{aligned} \tag{5.65}$$

where in the above the addition $k+1$ is taken modulo k. When $a = b = 0$, we refer to the vertices $(1, p')$ as *weightless vertices*.

In Δ^{pre}, for $p' \leq p/2$, the color of every edge connecting (k', p') to $(k'+1, p')$ is G, and the color of every edge connecting $(1, p')$ to $(\bar{1}_s, p')$ or (k', p') to $(\bar{k'}, p')$ is X^2. For $p' > p/2$ the colors are $\overline{G}$ and $\overline{X^2}$ respectively.

Proof of Lemma 5.4.4. The case of $a = b = 0$ and $k \geq 2$ is treated by Lemma 5.4.2, since $\psi^{3/2} = (N^{-1/3+\epsilon_0})^{3/2} = N^{-1/2+\frac{3}{2}\epsilon_0} \leq N^{-1/3+\epsilon_0-\delta} = \frac{N^{-\delta}}{N\eta_0}$. Therefore it remains to treat the case of $a + b \geq 1$.

We show that

$$\left| M^{-1} \sum_{i \in \mathcal{I}_M} \mathbf{e}_i^* \mathbf{C}_a \Lambda^a (G(\Lambda G)^{k-1}) \Lambda^b \mathbf{C}_b \mathbf{e}_i \right| \prec N^{-1}. \tag{5.66}$$

We accomplish this using the technique of Section 5.3.2, except we start with the graph Δ^{pre} introduced in this section.

Let $\mathfrak{P}$ be the set of all partitions of $V_b(\Delta^{\text{pre}})$, and for $P \in \mathfrak{P}$ let $\Delta(P)$ be the quotient graph of Δ^{pre} by P. This section contains only the first of several different very closely related graphs with which we will be working later in the paper, and then we will let $\mathfrak{P}$ take slightly different definitions.

Let $p>0$ be even, $\mathbf{a}_b=(a_i)_{i\in V_b(\Gamma)}\in\mathcal{I}_M^{|V_b(\Gamma)|}$ (notice that we do not have black vertices landing in $\mathcal{I}_K$ in this proof) and $\mathbf{a}_w=(a_i)_{i\in V_w(\Gamma)}\in\mathcal{I}_N^{|V_b(\Gamma)|}$. Just as in Section 5.3.1 we get

$$\mathbb{E}\left|\sum_{i\in\mathcal{I}_M}\mathbf{e}_i^*\mathbf{C}_a\Lambda^a(G(\Lambda G)^{k-1})\Lambda^b\mathbf{C}_b\mathbf{e}_i\right|^p=\sum_{P\in\mathfrak{P}}\sum_{\mathbf{a}_b}^{*}w_{\mathbf{a}_b}(\Delta(P))\sum_{\mathbf{a}_w}\mathcal{A}(\Delta(P))$$

where $w_{\mathbf{a}_b}(\Delta(P))=\prod(\mathbf{C}_a\mathbf{C}_b)_{a_ia_i}^{\deg_{\Delta(P)}(i)/2}M^{-p}$, the product ranging over the weightless vertices, which is justified just as in Section 5.3.1, together with equation (5.64). Now we fix P and let $\Delta:=\Delta(P)$. Throughout the rest of this proof, the black vertices of every graph that appears are the same as the black vertices of Δ, so that we may write $V_b:=V_b(\Delta)$ for the set of black vertices of every graph in this proof without ambiguity. Exactly as in Section 5.3.1,

$$\sum_{\mathbf{a}_w}\mathcal{A}(\Delta)=\sum_{\sigma\in L_1(\mathcal{T})}\sum_{\mathbf{a}_w}\mathcal{A}(\Theta_\sigma)+O_\prec(N^{-p})$$

where we recall $L_1(\mathcal{T})$ is the set of non-trivial leaves of the tree $\mathcal{T}$, and then for each of the boundedly many $\sigma\in L_1(\mathcal{T})$,

$$\sum_{\mathbf{a}_w}\mathcal{A}(\Theta_\sigma)=\sum_{\Gamma\in\mathfrak{G}}\sum_{\mathbf{a}_w}\mathcal{A}(\Gamma)+O_\prec(N^{-p}).$$

We show that

$$\sum_{\mathbf{a}_b}^{*}w_{\mathbf{a}_b}(\Delta)\sum_{\mathbf{a}_w}\mathbb{E}\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(\Gamma)\prec N^{-p},$$

which is sufficient because $|\mathfrak{G}|$ is bounded. We fix a partition ζ, and define for $i\in V_b$ the partitions ζ_i on $\pi^{-1}(i)$ as the restriction of ζ to $\pi^{-1}(i)$, and we define $n_\zeta(i):=|\zeta_i|$. Define also

$$W_{\mathbf{a}_b,\mathbf{a}_w}(\Gamma_\zeta)=\prod_{e\in E_X(\Gamma_\zeta)}\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\Gamma_\zeta).$$

We have the estimate

$$|W_{\mathbf{a}_b,\mathbf{a}_w}(\Gamma_\zeta)|\prec\left(\prod_{e\in E_X(\Gamma_\zeta)}N^{-1/2}\right)N^{-\frac{1}{2}(a+b+k-1)p}\tag{5.67}$$

since every X-edge e yields a factor of $N^{-1/2}$, except $(a+b+k-1)p$ of them (the ones with color $\xi_1(e) \in \{X^2, \overline{X^2}\}$), which yield an additional $N^{-1/2}$. We then note that

$$\prod_{e \in E_X(\Gamma_\zeta)} N^{-1/2} = \prod_{i \in V_b} N^{-\frac{1}{2}\deg_{\Gamma_\zeta}(i)}. \tag{5.68}$$

By Lemma 5.4.8, since $\mathbf{w}_{\mathbf{a}_b}(\Delta)$ is independent of X_M, we may assume that each partition ζ_i for $i \in V_b$ has no singletons.

We proceed from the simple estimate:

$$\sum_{\mathbf{a}_b}^{*} w_{\mathbf{a}_b}(\Delta) \sum_{\mathbf{a}_w} \mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(\Gamma) \prec \left(\prod_{i \in V_b} N^{n_\zeta(i)+1} \right) \max_{\mathbf{a}_b,\mathbf{a}_w} |W_{\mathbf{a}_b,\mathbf{a}_w}(\Gamma_\zeta)| N^{-p}$$

so that recalling equations (5.67) and (5.68), it now suffices to show that

$$\left(\prod_{i \in V_b} N^{n_\zeta(i)+1} N^{-\frac{1}{2}\deg_{\Gamma_\zeta}(i)} \right) N^{-\frac{1}{2}(a+b+k-1)p} \leq 1. \tag{5.69}$$

Since ζ_i has no singletons for any i, we have

$$n_\zeta(i) \leq \begin{cases} \frac{\deg_{\Gamma_\zeta}(i)}{2} & \text{if } \deg_{\Gamma_\zeta} \text{ is even} \\ \frac{\deg_{\Gamma_\zeta}(i)-1}{2} & \text{if } \deg_{\Gamma_\zeta} \text{ is odd.} \end{cases}$$

Bounding the left-hand side of equation (5.69), we have

$$\begin{aligned} &\left(\prod_{i \in V_b} N^{n_\zeta(i)+1} N^{-\frac{1}{2}\deg_{\Gamma_\zeta}(i)} \right) N^{-\frac{1}{2}(a+b+k-1)p} \\ &\leq N^{|V_b|-\frac{1}{2}\left|V_b^*\right|} N^{-\frac{1}{2}(a+b+k-1)p} \end{aligned} \tag{5.70}$$

where V_b^* is the set all $i \in V_b$ for which $\deg_{\Gamma_\zeta}(i)$ is odd. Recall the partition P which defined $\Delta := \Delta(P)$.

Now first assume $a+b=1$. One may see that if no black vertices are identified by P, then every black vertex i has $\deg_{\Gamma_\zeta}(i)$ odd; in this way, $|V_b^*|$ is at least the number of black vertices which are in a singleton of P. Letting ℓ be the number of singletons in P, we therefore have

$$|V_b| - \frac{1}{2}|V_b^*| \leq |V_b| - \frac{1}{2}\ell \leq \ell + \frac{1}{2}(kp - \ell) - \frac{1}{2}\ell = \frac{1}{2}kp. \tag{5.71}$$

Referring to equation (5.70) and using that $a+b=1$, we are done with the case $a+b=1$.

For the case $a+b \geq 2$, we have now that whereas previously $|V_b^*| \geq \ell$, now $|V_b^*| \geq \ell - p$, since p of the vertices $i \in V_b$ (the ones with the factor Λ^{a+b}), may be unidentified by P and yet have $\deg_{\Gamma_\varsigma}(i)$ even (if $a+b$ is even). Thus, equation (5.71) instead yields

$$|V_b| - \frac{1}{2}|V_b^*| \leq \frac{1}{2}kp + \frac{1}{2}p \tag{5.72}$$

but since $a+b \geq 2$, equation (5.70) still yields the result. We conclude the proof of Lemma 5.4.4 by summing over the boundedly many partitions $P \in \mathfrak{P}$. □

5.4.2 The proof of Lemmas 5.4.5, 5.4.6, and 5.4.7

We prove each of the three lemmas whose proof constitutes this section with the use of two or three sublemmas, prove the sublemmas, and then move on to the next one.

A common theme in each of the three lemmas is the counting of the number of vertices in the graph, the number of off-diagonal edges in the graph, and the number of special vertices i, collected in a set designated V_b^* which is chosen differently for the proof of each lemma, which have $\deg_X(i)$ odd and thus lead to reductions in the combinatorics, thanks to the independence of the entries of X.

Proof of Lemma 5.4.5. We define $\mathfrak{P}$ and construct the graph Δ^{pre} as in section 5.4.1 for $a=b=0$ and $k \geq 2$, and fix a partition $P \in \mathfrak{P}$. We let V_b be the set of the resulting black vertices and V_b^* be the set of non-weightless black vertices which correspond to singletons in the partition P. The following lemma counts the relationship between V_b, V_b^*, and the number of off-diagonal G-edges in Δ.

Lemma 5.4.9. *We have*

$$\left(-\frac{(k+1)p}{2} + |V_b| - \frac{1}{2}|V_b^*|\right) \leq 0 \tag{5.73}$$

and

$$-2\left(-\frac{(k+1)p}{2}+|V_b|-\frac{1}{2}|V_b^*|\right)+|E_o(\Delta)|\geq\frac{(k+1)p}{2}. \tag{5.74}$$

Equality in equation (5.73) is easily achieved, and holds in the worst case scenario for equation (5.74) that we describe in the next paragraph.

We may understand equation (5.74) heuristically as follows: off-diagonal edges and V_b^* black vertices are good, while black vertices generally speaking are bad ("good" means "more of them leads to a better bound" while "bad" means the opposite). The worst case scenario for equation (5.74) is if all $(k-1)p$ vertices from V_b^* should pairwise identity. This is the worst case because each pairwise identification of V_b^* vertices (1) decreases V_b^* by two and V_b by one, so that the difference $|V_b|-\frac{1}{2}|V_b^*|$ is not changed, but also (2) may also cause one off-diagonal edge to cease to be off-diagonal. Since $\frac{(k-1)p}{2}$ such identifications are possible, we lose $\frac{(k-1)p}{2}$ off-diagonal edges in this case and are left with $\frac{(k+1)p}{2}$ of them. Any additional identifications of black vertices then strictly decrease the difference $|V_b|-\frac{1}{2}|V_b^*|$, which compensates for any additional loss of off-diagonal edges.

Lemma 5.4.9 is proven very shortly. Now we construct the tree $\mathcal{T}$ from Δ just as in section 5.3.1 and we get

$$\mathbb{E}Y(\Delta)=\mathbb{E}\sum_{\sigma\in L(\mathcal{T})}\sum_{\mathbf{a}_b}^{*}w_{\mathbf{a}_b}(\Delta)\sum_{\mathbf{a}_w}\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(\Theta_\sigma), \tag{5.75}$$

where $w_{\mathbf{a}_b}(\Delta)=\prod(\mathbf{C}_a\mathbf{C}_b)_{a_ia_i}^{\deg_{\Delta(P)}(i)/2}M^{-p}$, the product ranging over the weightless vertices. We now fix a leaf $\sigma\in L(\mathcal{T})$. The trivial leaves are treated just as in section 5.3.1, so we assume σ is a non-trivial leaf. We complete the operations of Section 5.3.1, and fix a $\Gamma\in\mathfrak{G}(\Theta_\sigma)$. We now fix a partition ζ on the white vertices of Γ and let Γ_ζ be the quotient graph after identifying white vertices in Γ according to ζ.

We must now count $|V_w(\Gamma_\zeta)|$ in relation to the number of off-diagonal G edges $E_o(\Gamma_\zeta)$ in Γ_ζ:

Lemma 5.4.10. *Consider a fixed graph Γ and partition ζ which is such that*

$$\mathbb{E} \prod_{e \in E(\Gamma_\zeta)} \mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(e, \Gamma_\zeta) \neq 0 \tag{5.76}$$

Let the following be the definition of the numbers c_1, c_2:

$$|V_w(\Gamma_\zeta)| = \frac{|E_X(\Gamma)|}{2} - \frac{|V_b^*|}{2} - c_1 \tag{5.77}$$

and

$$|E_o(\Gamma_\zeta)| = |E_o(\Delta)| - c_2. \tag{5.78}$$

Then, $c_1 \geq 0$ and $2c_1 \geq c_2$.

Lemma 5.4.10 is proven very shortly. We have then, fixing Γ and ζ such that equation (5.76) holds,

$$\begin{aligned}
&\left| \mathbb{E} \sum_{\mathbf{a}_b}^{*} w_{\mathbf{a}_b}(\Delta) \sum_{\mathbf{a}_w} \prod_{e \in E(\Gamma_\zeta)} \mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(e, \Gamma_\zeta) \right| \\
&\lesssim N^{-p} \mathbb{E} N^{|V_b|} N^{\left|V_w(\Gamma_\zeta)\right|} \left| \prod_{e \in E_X(\Gamma_\zeta)} \mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(e, \Gamma_\zeta) \right| \left| \prod_{e \in E_G(\Gamma_\zeta)} \mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(e, \Gamma_\zeta) \right| \\
&\prec N^{-p} N^{|V_b|} N^{\left|V_w(\Gamma_\zeta)\right|} N^{-\frac{|E_X(\Gamma)|}{2}} N^{-\frac{(k-1)p}{2}} \psi^{\left|E_o(\Gamma_\zeta)\right|}
\end{aligned}$$

and then using Lemma 5.4.10 and that $N^{-1/2} \leq \psi$,

$$\begin{aligned}
&\leq N^{-p} N^{|V_b|} N^{\frac{|E_X(\Gamma)|}{2} - \frac{|V_b^*|}{2}} N^{-\frac{|E_X(\Gamma)|}{2}} N^{-\frac{(k-1)p}{2}} \psi^{|E_o(\Delta)|} \\
&= N^{-\frac{1}{2}(k+1)p + |V_b| - \frac{1}{2}\left|V_b^*\right|} \psi^{|E_o(\Delta)|}
\end{aligned}$$

Then, using Lemma 5.4.9, we get

$$\begin{aligned}
&\leq \psi^{2\left(-\frac{1}{2}(k+1)p + |V_b| - \frac{1}{2}\left|V_b^*\right|\right)} \psi^{|E_o(\Delta)|} \\
&\leq \psi^{\frac{(k+1)p}{2}}
\end{aligned}$$

thus completing the proof of Lemma 5.4.5. □

Proof of Lemma 5.4.9. We recall the partition P which induces the graph Δ. We may think of the operation of producing the quotient graph $\Delta = (\{1,\dots,k\} \times \{1,\dots,p\})/P$ as a sequence of identifications of vertices of $\{1,\dots,k\} \times \{1,\dots,p\}$. We may perform them in this order:

1. First perform identifications between two non-weightless vertices. Let P_1 be the partition on Δ vertices described by these identifications; the blocks of the partition are identified with the vertices of the resulting graph.

2. Then perform all identifications between a weightless vertex and a non-weightless vertex which is a singleton of P_1.

3. Then perform all other identifications between a weightless vertex and a non-weightless vertex.

4. Finally perform the remaining identifications between two weightless vertices.

Let s_1, s_2, s_3 and s_4 respectively be the number of each sort of identification. We may reason simply by counting how these identifcations affect $|V_b|$, $|V_b^*|$, and $|E_o(\Delta)|$, collecting our observations as follows:

1. The number of vertices $|V_b(\Delta)|$ is precisely
$$|V_b(\Delta)| = kp - s_1 - s_2 - s_3 - s_4.$$

2. $|V_b^*|$ is at least
$$|V_b^*(\Delta)| \geq ((k-1)p - 2s_1 - s_2)^+ .$$

3. The number of off-diagonal edges is at least
$$|E_o(\Delta)| \geq kp - s_1 - s_2 - 2s_3.$$

   The reason for this is that the subgraph of Δ^{pre} induced by the subset $\{2,\dots,k\} \times \{1,\dots,p\} \subseteq V_b(\Delta^{\text{pre}})$ has no cycles, so that each of the s_1 identifications between the vertices $(k', p') \in \{2, \ldots, k\} \times \{1, \ldots, p\}$, which may be assumed to be performed in order of increasing k' and then increasing p', cannot cause more than one edge of to cease to be off-diagonal. Then, identifications among the s_2 can only cause 1 edge to cease to be off-diagonal—this is because if a weightless vertex has two edges joining it a non-weightless vertex, then because $k \geq 3$, the non-weightless vertex must have already been identified by one of the first s_1 identifications—while identifications among the s_3 can cause 2 edges to cease to be off-diagonal—ie, if $(2, k')$ and (p, k') have already been identified, then identifying $(1, k')$ to $(2, k')$ causes two edges to cease to be off-diagonal. Identifications among the s_4 do not alter the number of off-diagonal edges.

The proof now easily follows:

$$
\begin{aligned}
& -\frac{(k+1)p}{2} + |V_b| - \frac{1}{2}|V_b^*| \\
& \leq -\frac{(k+1)p}{2} + kp - s_1 - s_2 - s_3 - s_4 - \frac{1}{2}\left((k-1)p - 2s_1 - s_2\right)^+ \\
& = \begin{cases} -\frac{s_2}{2} - s_3 - s_4 & \text{if } 2s_1 + s_2 \leq (k-1)p \\ \frac{(k-1)p}{2} - s_1 - s_2 - s_3 - s_4 & \text{if } 2s_1 + s_2 \geq (k-1)p \end{cases} \\
& \leq 0,
\end{aligned}
$$

i.e., the last inequality holds regardless of the size of s_1. And then,

$$
\begin{aligned}
& -2\left(-\frac{(k+1)p}{2} + |V_b| - \frac{1}{2}|V_b^*|\right) + |E_o(\Delta)| \\
& \geq \begin{cases} -2\left(-\frac{s_2}{2} - s_3 - s_4\right) + kp - s_1 - s_2 - 2s_3 & \text{if } 2s_1 + s_2 \leq (k-1)p \\ -2\left(\frac{(k-1)p}{2} - s_1 - s_2 - s_3 - s_4\right) + kp - s_1 - s_2 - 2s_3 & \text{if } 2s_1 + s_2 \geq (k-1)p \end{cases} \\
& = \begin{cases} kp - s_1 + 2s_4 & \text{if } 2s_1 + s_2 \leq (k-1)p \\ p + s_1 + s_2 + 2s_4 & \text{if } 2s_1 + s_2 \geq (k-1)p \end{cases} \\
& \geq \frac{(k+1)p}{2}
\end{aligned}
$$

as desired; this completes the proof of Lemma 5.4.9. □

Proof of Lemma 5.4.10. Let ζ_{b} be defined as the least-restrictive partition of $V_w(\Gamma_\zeta)$ which identifies two vertices $a_i \neq a_j \in V_w(\Gamma_\zeta)$ if a_i and a_j are identified by ζ and if $\{a_i, a_j\} = \{\alpha(e), \beta(e)\}$ for some $e \in E_o(\Gamma_\zeta)$. Let $r = |V_w(\Gamma)| - |V_w(\Gamma_{\zeta_{\mathrm{b}}})|$, the number of off-diagonal edges whose endpoints are identified by ζ. Here, b stands for "bad", because ζ_{b} is the part of ζ which induces *bad* identifications between $j \in V_w(\Gamma)$, ie, those which cause off-diagonal edges to cease to be off-diagonal, and we are trying to preserve off-diagonal edges. The subscript b for "black" should not be confused with the subscript b. For $i_1, i_2 \in V_b$, we say i_1 and i_2 are "part of a bad identification" if for some $j_1 \in \pi^{-1}(i_1), j_2 \in \pi^{-1}(i_2)$, we have $\{j_1, j_2\} \in \zeta_{\mathrm{b}}$.

Let also ζ_{int} be least restrictive partition which identifies two vertices $j_1 \neq j_2 \in V_w(\Gamma)$ if j_1 and j_2 are identified by ζ_{b}, or if j_1 and j_2 are both in $\pi^{-1}(i)$ for the same $i \in V_b(\Gamma)$ and j_1 and j_2 are identified by ζ. Here, "int" is short for "internal". ζ_{int} induces a partition on each set $\pi^{-1}(i)$ which we may assume has no singletons, lest by the same reasoning as lemma 5.13 of [BEK14] we get a trivial expression. Note also that $|\pi^{-1}(i)|$ is odd.

Let $V_{\mathrm{pb}} \subseteq V_b^*$ be the vertices which are pulled to (see the remarks following equation (5.29)) in the construction of Γ and which are part of a bad identification. Let also $V_{\mathrm{rest}} := V_b^* \setminus V_{\mathrm{pb}}$.

First perform the identifications entailed by ζ_{b}: the number of vertices in $V_w(\Gamma_{\zeta_{\mathrm{b}}})$ is $|V_w(\Gamma)| - r$. Now by construction, for each $i \in V_{\mathrm{rest}}$, one of two cases holds:

1. there are exactly 2 distinct $j \in \pi^{-1}(i)$ which are endpoints of an off-diagonal edge $e \in \Gamma_{\zeta_{\mathrm{b}}}(\Gamma)$, and hence only 2 vertices $j \in \pi^{-1}(i)$ which can be part of a bad identification.
2. no vertex $j \in \pi^{-1}(i)$ is part of a bad identification.

Therefore, for each $i \in V_{\mathrm{rest}}$ there is one vertex $:= j(i) \in \pi^{-1}(i)$ which is part of a block of ζ_{int} of size ≥ 3 and which is not part of a bad identification. Similarly, for each $i \in V_{\mathrm{pb}}$, there is a vertex $j(i) \in \pi^{-1}(i)$ which is part of a block of ζ_{int} of size ≥ 3, but in this case, however, because i may be part of a bad identification, $j(i)$ may be equal under ζ_{b} to $j(i')$ for some $i' \neq i$.

We conclude that $\{j(i) : i \in V_{\mathrm{rest}}\}$ and $\{j(i) : i \in V_{\mathrm{pb}}\}$ are disjoint and have sizes $|V_{\mathrm{rest}}|$, and $\geq \frac{1}{2}|V_{\mathrm{pb}}|$, respectively. Since each $j(i)$ for $i \in V_{\mathrm{rest}} \cup V_{\mathrm{pb}} = V_b^*$ is part of a block of ζ_{int} of size ≥ 3, and every block of ζ_{int} has size at least 2, we conclude that

$$\begin{aligned} |V_w(\Gamma_\zeta)| \leq |V_w(\Gamma_{\zeta_{\mathrm{int}}})| &\leq \frac{|E_X(\Gamma)| - r - |V_{\mathrm{rest}}| - \frac{|V_{\mathrm{pb}}|}{2}}{2} \\ &= \frac{|E_X(\Gamma)|}{2} - \frac{|V_b^*(\Gamma)|}{2} - \frac{r}{2} + \frac{|V_{\mathrm{pb}}|}{4}. \end{aligned}$$

Therefore, c_1 in Lemma 5.4.10 is at least $\frac{r}{2} - \frac{|V_{\mathrm{pb}}|}{4}$. This is a non-negative number, as desired, because every one of the r identifications performed by ζ_{b} adds at most 2 vertices to V_{pb}.

The number of off-diagonal edges $|E_o(\Gamma_{\zeta_{\mathrm{int}}})| = |E_o(\Gamma_\zeta)|$ is at least $|E_o(\Delta)| + |V_{\mathrm{pb}}| - r$. Therefore c_2 in Lemma 5.4.10 is at most $r - |V_{\mathrm{pb}}|$. This concludes the proof of Lemma 5.4.10. □

Proof of Lemma 5.4.6. This proof is very similar to the previous proof. We construct the graph Δ^{pre} for $k = 2$ and follow the same process as in the last proof to arrive at a graph Γ. Now that $k = 2$, every non-weightless vertex in Δ^{pre} has a unique weightless neighbor in Δ^{pre}.

In this proof, let $V_b^* \subseteq V_b$ be those black vertices whose preimage under the quotienting by P is either a single non-weightless vertex, or a non-weightless vertex together with its unique weightless neighbor only.

Lemma 5.4.11. *With our new definition of V_b^*, the same statement as Lemma 5.4.9 holds.*

Lemma 5.4.12. *With our new definition of V_b^*, the same statement as Lemma 5.4.10 holds.*

Given that, with our new definition of V_b^*, exact analogs of Lemmas 5.4.9 and 5.4.10 hold, we conclude the proof the proof of Lemma 5.4.6 just as we concluded Lemma 5.4.5. □

Proof of Lemma 5.4.11. This proof is similar to the proof of Lemma 5.4.9, and we give an abridged version. Recall P from which $\Delta = \Delta(P)$ is formed. We let s_1 be the number of identifications entailed by P which identify a non-weightless vertex with its unique weightless neighbor, and let s_2 be the number of all other identifications. Then we have

$$|V_b| = kp - s_1 - s_2$$

and

$$|V_b^*| \geq ((k-1)p - 2s_2)^+$$

and

$$|E_o(\Delta)| = kp - 2s_1,$$

so that

$$\begin{aligned}
& -\frac{(k+1)p}{2} + |V_b| - \frac{1}{2}|V_b^*| \\
& \leq -\frac{(k+1)p}{2} + kp - s_1 - s_2 - \frac{1}{2}\left((k-1)p - 2s_2\right)^+ \\
& = \begin{cases} -s_1 & \text{if } 2s_2 \leq (k-1)p \\ \frac{(k-1)p}{2} - s_1 - s_2 & \text{if } 2s_2 \geq (k-1)p \end{cases} \\
& \leq 0
\end{aligned}$$

and

$$\begin{aligned}
& -2\left(-\frac{(k+1)p}{2} + |V_b| - \frac{1}{2}|V_b^*|\right) + |E_o(\Delta)| \\
& \geq \begin{cases} -2\left(-s_1\right) + kp - 2s_1 & \text{if } 2s_2 \leq (k-1)p \\ -2\left(\frac{(k-1)p}{2} - s_1 - s_2\right) + kp - 2s_1 & \text{if } 2s_2 \geq (k-1)p \end{cases} \\
& = \begin{cases} kp & \text{if } 2s_2 \leq (k-1)p \\ p + 2s_2 & \text{if } 2s_2 \geq (k-1)p \end{cases} \\
& \geq kp,
\end{aligned}$$

and we conclude. □

Proof of Lemma 5.4.12. This proof is a verbatim repeat of the proof of Lemma 5.4.10, using only the additional insight that each $i \in V_b^*$ has at most 2 off-diagonal edges incident on it in Δ, unless it is pulled to in the construction of Γ. $\square$

Proof of Lemma 5.4.7. We define the graph Δ^{pre} according to section 5.4.1 for $a = b = 0$. We fix a partition P of the $V_b(\Delta^{\mathrm{pre}})$ which contains the block $i_0 := \{(1,1), \ldots, (1,p)\}$, and let $\Delta = \Delta^{\mathrm{pre}}/P$. The reason for this insistence on the form of partition is that p different vertices in the graphs of the previous proofs which previously each could range over $\mathcal{I}_M$ are now all constrained to equal $\alpha \in \mathcal{I}_K$.

We write the expression which we desire to bound as

$$\begin{aligned}\mathbb{E}\big|\mathbf{e}_\alpha^*(G(\Lambda G)^{k-1})\mathbf{e}_\alpha\big|^p &= \mathbb{E}\sum_P Y(\Delta(P)) \\ &= \mathbb{E}\sum_P \sum_{\mathbf{a}_b}^{*} \sum_{\mathbf{a}_w} \mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(\Delta(P))\end{aligned} \tag{5.79}$$

as before, where a_{i_0} is now only allowed to take the value $\alpha \in \mathcal{I}_K$.

The proof is now very similar to that of 5.4.5, except that we have fewer black vertices over which to sum (since a_{i_0} takes a single value rather than ranging over $\mathcal{I}_M$). We also must account for the possibility of un-identified white vertices like in section 5.3.3.

As we have done several times before, follow the procedure of section 5.3.1 and fix one of the resulting graphs Γ.

We first adjust Lemma 5.4.9 to count only the black vertices which land in $\mathcal{I}_M$. Notice that the conclusion of this lemma is better by $-p$ than Lemma 5.4.9, owing to the fact that we have p fewer vertices over which to sum.

Lemma 5.4.13. *Let $V_b^s = V_b \setminus \{i_0\}$ be the set of vertices $i \in V(\Delta)$ such that $a_i \in \mathcal{I}_M$ ("s" stands for "summing").*

Then,

$$\left(-\frac{(k-1)p}{2} + |V_b^s| - \frac{1}{2}|V_b^*|\right) \leq 0$$

and

$$-2\left(-\frac{(k-1)p}{2}+|V_b^s|-\frac{1}{2}|V_b^*|\right)+|E_o(\Delta)|\geq\frac{(k+1)p}{2}.$$

Lemma 5.4.13 is proven later in the section. The following lemma is in analogy to Lemma 5.4.10.

Lemma 5.4.14. *Fix a partition ζ on $V_w(\Gamma)$ such that*

$$\mathbb{E}\prod_{e\in E(\Gamma_\zeta)}\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\Gamma_\zeta)\neq 0 \tag{5.80}$$

and let ℓ be the number of white vertices $i\in\pi^{-1}(i_0)$ which are in a block of ζ of size 1. Let the following be the definitions of the numbers c_1, c_2:

$$|V_w(\Gamma_\zeta)|=\frac{|E_X(\Gamma)|}{2}-\frac{|V_b^*|}{2}+\frac{\ell}{2}-c_1 \tag{5.81}$$

and

$$|E_o(\Gamma_\zeta)|=|E_o(\Gamma)|-c_2. \tag{5.82}$$

Then, $c_1\geq 0$ and $2c_1\geq c_2$.

Lemma 5.4.14 is proven later in the section. We now have a lemma which says that lone vertices always lead to more factors of ψ.

Lemma 5.4.15. *Fix $\sigma\in\{0,1\}^\ell$ such that*

$$\left|\mathbb{E}\sum_{\mathbf{a}_b}^{*}\sum_{\mathbf{a}_w}\prod_{e\in E(\omega_\sigma\Gamma_\zeta)}\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\omega_\sigma\Gamma_\zeta)\right|\neq 0.$$

Let n_{p0} be the number of times in the construction of Γ that an edge is "pulled" to i_0. Then,

$$\left|\prod_{e\in E_G(\omega_\sigma\Gamma_\zeta)}\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\omega_\sigma\Gamma_\zeta)\right|\prec\psi^{\left|E_o(\Gamma_\zeta)\right|+\ell+\frac{1}{2}[\ell-2p-2n_{p0}]^+}N^{-2\ell\epsilon_D}.$$

Lemma 5.4.15 is proven later in the section. Now we may prove Lemma 5.4.7. We have by Lemma 5.4.15

$$
\begin{aligned}
&\left|\mathbb{E}\sum_{\mathbf{a}_b}^{*}\sum_{\mathbf{a}_w}\prod_{e\in E(\Gamma_\zeta)}\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\Gamma_\zeta)\right| \\
&= \left|\sum_{\sigma}\mathbb{E}\sum_{\mathbf{a}_b}^{*}\sum_{\mathbf{a}_w}\prod_{e\in E(\omega_\sigma\Gamma_\zeta)}\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\omega_\sigma\Gamma_\zeta)\right| \\
&\prec \left(\sum_{\mathbf{a}_b}^{*}1\right)\left(\sum_{\mathbf{a}_w}1\right)\left(\prod_{e\in E_X(\omega_\sigma\Gamma_\zeta)}\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\omega_\sigma\Gamma_\zeta)\right)\left(\prod_{e\in E_G(\omega_\sigma\Gamma_\zeta)}\mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e,\omega_\sigma\Gamma_\zeta)\right) \\
&\prec N^{\left|V_b^s\right|}N^{\left|V_w(\Gamma_\zeta)\right|}N^{-\frac{1}{2}|E_X(\Gamma)|}N^{-\frac{(k-1)p}{2}}\psi^{\left|E_o(\Gamma_\zeta)\right|+\ell+\frac{1}{2}[\ell-2p-2n_{p0}]^+}N^{-2\ell\epsilon_D}.
\end{aligned}
$$

Lemma 5.4.14 gives

$$
\begin{aligned}
&\le N^{\left|V_b^s\right|}N^{\frac{1}{2}|E_X(\Gamma)|-\frac{1}{2}\left|V_b^*\right|+\frac{\ell}{2}}N^{-\frac{1}{2}|E_X(\Gamma)|}N^{-\frac{(k-1)p}{2}}\psi^{|E_o(\Gamma)|+\ell+\frac{1}{2}[\ell-2p-2n_{p0}]^+}N^{-2\ell\epsilon_D} \\
&= N^{-\frac{1}{2}(k-1)p+\left|V_b^s\right|-\frac{1}{2}\left|V_b^*\right|+\frac{\ell}{2}}\psi^{|E_o(\Gamma)|+\ell+\frac{1}{2}[\ell-2p-2n_{p0}]^+}N^{-2\ell\epsilon_D}.
\end{aligned}
$$

Using $|E_o(\Gamma)| \ge |E_o(\Delta)| + n_{p0}$, followed by the fact that $\psi^3 N^{-4\epsilon_D} \le N^{-1}$, yields

$$
\begin{aligned}
&\le N^{-\frac{1}{2}(k-1)p+|V_b|-\frac{1}{2}\left|V_b^*\right|+\frac{\ell}{2}}\psi^{|E_o(\Delta)|+\frac{3}{2}\ell-p}N^{-2\ell\epsilon_D} \\
&\le N^{-\frac{1}{2}(k-1)p+|V_b|-\frac{1}{2}\left|V_b^*\right|}\psi^{|E_o(\Delta)|-p}.
\end{aligned}
$$

Lemma 5.4.13 and $N^{-1/2} \le \psi$ then yield

$$
\le \psi^{\frac{k+1}{2}p}\psi^{-p} = \psi^{\frac{k-1}{2}p}
$$

which concludes the proof of Lemma 5.4.7. □

Proof of Lemma 5.4.13. The proof is identical to the proof of Lemma 5.4.9, except that only the vertices in $\{2,\dots,p\}\times\{1,\dots,k\}$ need to be summed over, that is, p fewer vertices. □

Proof of Lemma 5.4.14. The proof is almost identical to that of Lemma 5.4.10; the only difference is that ℓ vertices $i \in V_w(\Gamma_{\zeta_\mathrm{b}})$ can be blocks of ζ_{int} of size 1. Using the notation of the proof of Lemma 5.4.10, we have

$$\begin{aligned}|V_w(\Gamma_\zeta)| \le |V_w(\Gamma_{\zeta_{\text{int}}})| &\le \frac{|E_X(\Gamma)| - r - |V_{\text{rest}}| - \frac{|V_{\text{pb}}|}{2} - \ell}{2} + \ell \\ &= \frac{|E_X(\Gamma)|}{2} - \frac{|V_b^*(\Gamma)|}{2} + \frac{\ell}{2} - \frac{r}{2} + \frac{|V_{\text{pb}}|}{4}\end{aligned}$$

and $|E_o(\Gamma_{\zeta_{\text{int}}})| = |E_o(\Gamma_\zeta)|$ is at least $|E_o(\Gamma)| + |V_{\text{pb}}| - r$ as in the proof of Lemma 5.4.10, and we conclude as in that lemma. □

Proof of Lemma 5.4.15. Just as in section 5.3.3, each lone vertex yields an extra factor of $\psi N^{-2\epsilon_D}$, and moreover, each lone vertex, except at most $2p + 2n_{p0}$ of them (noting that there are exactly $2p+2n_{p0}$ off-diagonal edges in Δ incident on i_0), is one endpoint of an edge $e \in E_o(\Gamma_\zeta)$ but not already listed among $\widetilde{E_o}(\Gamma_\zeta)$, which concludes the proof. □

5.4.3 Isotropic local law for G_2^D.

Our goal for this section is the proof of Lemma 5.1.7

Lemma 5.4.16.

Proof of Lemma 5.1.7. Recalling Lemma 5.1.8, it suffices to prove: for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_K \cup \mathcal{I}_M}$, we have

$$\left|G^D_{2,\mathbf{xy}} - G^E_{2,\mathbf{xy}}\right| \prec \psi_{(\mathbf{xy})}. \tag{5.83}$$

Using the same reasoning as in the proof of Lemma 5.1.4, it suffices to prove that, recalling $G := G_2^E$ in this section,

$$\left|\mathbf{x}^* \mathbf{C}_a \Lambda^a G (\Lambda G)^{k-1} \Lambda^b \mathbf{C}_b \mathbf{y}\right| \prec \psi_{(\mathbf{xy})}$$

for $a, b \ge 0$ and $k \ge 1$ satisfying $a + b + k \ge 2$.

A short lemma will save us some work here:

Lemma 5.4.17. *If $\mathcal{G}$ is a conformable random matrix satisfying*

$$|\mathbf{x}^* \mathcal{G} \mathbf{y}| \prec \psi_{(\mathbf{xy})}$$

for deterministic $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_{K+M}}$*, then if* $\mathbf{x}_{\text{rand}}, \mathbf{y}_{\text{rand}} \in \mathbb{R}^{\mathcal{I}_{K+M}}$ *are random vectors with norm* $O_\prec(N^{-1/2})$*, we have*

$$|\mathbf{x}^* \mathcal{G} \mathbf{y}_{\text{rand}}| + |\mathbf{x}_{\text{rand}}^* \mathcal{G} \mathbf{y}| + |\mathbf{x}_{\text{rand}}^* \mathcal{G} \mathbf{y}_{\text{rand}}| \prec \psi_{(\mathbf{xy})}, \tag{5.84}$$

where $O_\prec(N^{-1/2})$ *denotes a random vector in* $\mathbb{R}^{\mathcal{I}_{K+M}}$ *of norm* $O_\prec(N^{-1/2})$*.*

Proof. Equation (5.84) may be bounded by

$$O_\prec(N^{-1/2}) \|\mathbf{x}^* \mathcal{G}\|_{\mathbb{R}^{\mathcal{I}_{K+M}}} + O_\prec(N^{-1/2}) \|\mathcal{G}\mathbf{y}\|_{\mathbb{R}^{\mathcal{I}_{K+M}}} + O_\prec(N^{-1}) \|\mathcal{G}\|_{\mathbb{R}^{\mathcal{I}_{K+M}}}$$

where $\|\cdot\|_{\mathbb{R}^{\mathcal{I}_{K+M}}}$ denotes the norm of a vector's projection onto $\mathbb{R}^{\mathcal{I}_{K+M}}$ or the operator norm of a matrix restricted to this space. The condition $|\mathbf{x}^* \mathcal{G} \mathbf{y}| \prec \psi_{(\mathbf{xy})}$ ensures

$$\|\mathcal{G}\mathbf{y}\|_{\mathbb{R}^{\mathcal{I}_{K+M}}} = \sqrt{\sum_{i \in \mathcal{I}_{K+M}} |\mathbf{e}_i^* \mathcal{G} \mathbf{y}|^2} \prec \sqrt{\psi_{(K\mathbf{y})}^2 + N \psi_{(M\mathbf{y})}^2} \lesssim N^{1/2} \psi_{(M\mathbf{y})}$$

as well as

$$\|\mathcal{G}\|_{\mathbb{R}^{\mathcal{I}_{K+M}}} \le \sum_{i \in \mathcal{I}_{K+M}} \mathbf{e}_i^* \mathcal{G} \mathbf{e}_i \prec N\psi$$

so that we may conclude. □

The following lemmas, whose proofs constitute most of the rest of this section, treat the case of $a = b = 0$, so that we may restrict to $k \ge 2$. Note the omission of $\mathbf{C}_0 = E$, which is permissible by Lemma 5.4.17 since $\|E - 1\| \prec N^{-1/2}$.

Lemma 5.4.18. *We have for* $\alpha, \beta \in \mathcal{I}_K$ *and* $k \ge 2$

$$\left|\mathbf{e}_\alpha^* G (\Lambda G)^{k-1} \mathbf{e}_\beta\right| \prec \psi_{(KK)}.$$

Lemma 5.4.19. *We have for* $\alpha \in \mathcal{I}_K, \mathbf{x} \in \mathbb{R}^{\mathcal{I}_M}$ *and* $k \ge 2$

$$\left|\mathbf{e}_\alpha^* G (\Lambda G)^{k-1} \mathbf{x}\right| \prec \psi_{(KM)}.$$

Lemma 5.4.20. *We have for* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_M}$ *and* $k \ge 2$

$$\left|\mathbf{x}^* G (\Lambda G)^{k-1} \mathbf{y}\right| \prec \psi_{(MM)}.$$

Now we treat the case of $a+b \geq 1$. If $k=1$, we write

$$\left|\mathbf{x}^*\mathbf{C}_a\Lambda^a G\Lambda^b\mathbf{C}_b\mathbf{y}\right| = \left|\mathbf{x}^*\mathbf{C}_a\Lambda^a(G-\Pi_2)\Lambda^b\mathbf{C}_b\mathbf{y}\right| + \left|\mathbf{x}^*\mathbf{C}_a\Lambda^a\Pi_2\Lambda^b\mathbf{C}_b\mathbf{y}\right|.$$

Using that $\Lambda^b\mathbf{C}_b\mathbf{y} = \mathbf{1}_{b=0}\mathbf{y} + O_\prec(N^{-1/2})$, the first term is bounded by $O_\prec(\psi_{(\mathbf{xy})})$ by Lemmas 5.4.17 and 5.1.8. The second is trivially $O_\prec(N^{-1/2})$. If rather $k \geq 2$, we conclude by Lemmas 5.4.17 and 5.4.18, 5.4.19, and 5.4.20. □

Proof of Lemma 5.4.18. For $k \geq 2$, this proof is exactly the same as that of Lemma 5.4.7; the only difference is that since we are working in this lemma with potentially $\alpha \neq \beta$, the single black vertex i_0 with $a_{i_0} \in \mathcal{I}_K$ in that proof is replaced with two such vertices, but it is easy to see that the proof is not affected by this. □

Proof of Lemma 5.4.19. By polarization and linearity, it suffices to consider $\mathbf{x} = \mathbf{y}$. We make one more reduction, similar to Section 5.2 "Reduction to off-diagonal entries" in [BEK14]. Writing

$$\mathbf{x}^* G_2^E(\Lambda G_2^E)^{k-1}\mathbf{x} = \sum_{i,j\in\mathcal{I}_M} \overline{\mathbf{x}_i}\left(G_2^E(\Lambda G_2^E)^{k-1}\right)_{ij}\mathbf{x}_j,$$

the case $i=j$ is actually treated by our Lemma 5.4.5. The only difference is that what in that proof were called "weightless" vertices and had weight identically N^{-1}, in this case correspond to the vertex $i=j$ has a weight $\mathbf{x}_i^2$ that only sums to 1. One may observe that the proof of 5.4.5 is not sensitive to this (a way the proof could conceivably be sensitive to this difference is if we needed to get an improvement in the order of the sum by identifying two vertices with weight $\mathbf{x}_i^2$, but we never actually utilize such an improvement).

The proof is now very similar to the earlier proofs in this section, and we only outline the differences. Very similarly to in section 5.4.1, we define a graph Δ^{pre} through

$$\begin{aligned} V_b(\Delta^{\text{pre}}) :& \{1,\dots,k+1\}\times\{1,\dots,p\} \\ V_w(\Delta^{\text{pre}}) :& \{\bar{2}_1,\dots,\bar{k}_1\}\times\{1,\dots,p\} \\ E(\Delta^{\text{pre}}) :& \{((k',p'),(k'+1,p')) : (k',p')\in\{1,\dots,k\}\times\{1,\dots,p\}\} \\ & \cup\left\{((k',p'),(\bar{k}'_1,p')) : (k',p')\in\{2,\dots,k\}\times\{1,\dots,p\}\right\}. \end{aligned} \tag{5.85}$$

Edges in the first set in the definition of $E(\Delta^{\text{pre}})$ have color G or $\overline{G}$ depending on whether $p' \le p/2$ or not, and similarly edges in the second set have color X^2 or $\overline{X^2}$. Also let $V_b^{\text{end}}(\Delta^{\text{pre}}) := \{1, k+1\} \times \{1, \dots, p\}$.

We fix a partition P which does not identify $(1, p')$ and $(k+1, p')$ for any $p' = 1, \dots, p$ (thanks to our reduction to off-diagonal entries) on $V_b(\Delta^{\text{pre}})$ and let us consider graphs of the form $\Delta := \Delta(P) = \Delta^{\text{pre}}/P$. This quotient induces a set $V_b^{\text{end}}(\Delta)$ in the obvious way.

Using all the same notation as before, we get

$$\mathbb{E}\left|\sum_{i \ne j \in \mathcal{I}_M} \overline{\mathbf{x}_i} \left(G_2^E (\Lambda G_2^E)^{k-1}\right)_{ij} \mathbf{x}_j\right|^p = \mathbb{E} \sum_P \sum_{\mathbf{a}_b}^{*} \sum_{\mathbf{a}_w} \mathbb{E} w_{\mathbf{a}_b}(\Delta(P)) \mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(\Delta(P)),$$

where $w_{\mathbf{a}_b}(\Delta) = \prod_{i \in V_b^{\text{end}}(\Delta)} \boldsymbol{c}_i(\mathbf{x}_{a_i})$ and $\boldsymbol{c}_i(z)$ is either z or $\overline{z}$ depending on the value of i; the complex conjugate is not important and we will not elaborate on how the function $\boldsymbol{c}_i$ depends on i.

The first lemma is analogous to Lemma 5.4.13. Its proof is very close to that of Lemma 5.4.13 and is omitted.

For $i \in V_b(\Delta)$, recalling that i arises as a block of the partition P, let $\deg_{\mathbf{x}}(i) = \left|\left\{ i \cap \bigcup_{p'=1}^{p} \{(1, p'), (k+1, p')\} \right\}\right|$.

Lemma 5.4.21. *Let $V_b^0 \subseteq V_b$ be the set of black vertices i in Γ which have $\deg_{\mathbf{x}}(i) = 0$, and let $V_b^1 \subseteq V_b$ be those with $\deg_{\mathbf{x}}(i) = 1$. Let $V_b^* \subseteq V_b$ be the vertices which correspond to singletons of P. Then,*

$$\left|V_b^0\right| + \frac{1}{2}\left|V_b^1\right| - \frac{1}{2}|V_b^*| \le \frac{(k-1)p}{2}.$$

Note that V_b^* is not necessarily disjoint with V_b^0 or V_b^1. Also note that this statement is simpler than the statement of of Lemma 5.4.9; we do not need to count the number of off-diagonal edges in the graph $\Delta(P)$ since we only consider partitions P in this section such that $\Delta(P)$ includes at least p off-diagonal edges.

The graph Δ has p off-diagonal edges incident on vertices i with $\deg_{\mathbf{x}}(i) \ge 1$. Indeed, after performing all pullings, there are at least p such edges (thanks to the defintion of the partition P), and the collection of them induces a collection of edges $:= \widetilde{E_o}(\Gamma) \subseteq E_o(\Gamma)$. The following lemma is proven just like Lemma 5.4.10.

Lemma 5.4.22. *Fix a partition ζ on $V_w(\Gamma)$ such that*

$$\mathbb{E} \prod_{e \in E(\Gamma_\zeta)} \mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e, \Gamma_\zeta) \neq 0. \tag{5.86}$$

Let the following be the definitions of the numbers c_1, c_2:

$$|V_w(\Gamma_\zeta)| = \frac{|E_X(\Gamma)|}{2} - \frac{|V_b^*|}{2} + c_1 \tag{5.87}$$

and

$$\left|E_o^T(\Gamma_\zeta)\right| = |E_o(\Gamma)| - c_2. \tag{5.88}$$

Then, $c_1 \geq 0$ and $2c_1 \geq c_2$.

We may conclude the proof now. Just as in earlier sections, the thing we desire to bound may be written

$$\mathbb{E} \sum_P \sum_{\mathbf{a}_b}^{*} \sum_{\mathbf{a}_w} \mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(\Delta(P)). \tag{5.89}$$

Then,

$$\begin{aligned}
&\left| \mathbb{E} \sum_{\mathbf{a}_b}^{*} w_{\mathbf{a}_b} \sum_{\mathbf{a}_w} \prod_{e \in E(\Gamma_\zeta)} \mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e, \Gamma_\zeta) \right| \\
&\leq \left(\sum_{\mathbf{a}_b}^{*} w_{\mathbf{a}_b}(\Delta) \right) N^{\left|V_w(\Gamma_\zeta)\right|} \left| \prod_{e \in E_X(\Gamma_\zeta)} \mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e, \Gamma_\zeta) \right| \left| \prod_{e \in E_o(\Gamma_\zeta)} \mathcal{A}_{\mathbf{a}_b,\mathbf{a}_w}(e, \Gamma_\zeta) \right| \\
&\prec N^{\left|V_b^0\right| + \frac{1}{2}\left|V_b^1\right|} \cdot N^{\left|V_w(\Gamma_\zeta)\right|} \cdot N^{-\frac{|E_X(\Gamma)|}{2} - \frac{(k-1)p}{2}} \cdot \psi^{\left|E_o(\Gamma_\zeta)\right|} \\
&\prec N^{\left|V_b^0\right| + \frac{1}{2}\left|V_b^1\right|} \cdot N^{\frac{|E_X(\Gamma)|}{2} - \frac{|V_b^*|}{2}} \cdot N^{-\frac{|E_X(\Gamma)|}{2} - \frac{(k-1)p}{2}} \cdot \psi^{|E_o(\Gamma)|} \\
&\leq \psi^{|E_o(\Gamma)|} = \psi^p
\end{aligned}$$

as desired. □

Proof of Lemma 5.4.20. This is a strict generalization of what we did in Lemma 5.1.8. It is very similar to what we did in that proof, together with some adjustments which we have already used in this section. We define the graph Δ^{pre} through

$$\begin{aligned} V_b(\Delta^{\text{pre}}) :=& \{1\} \cup \{2, \dots, k+1\} \times \{2, \dots, p\} \\ V_w(\Delta^{\text{pre}}) :=& \{\bar{2}_1, \dots, \bar{k}_1\} \times \{1, \dots, p\} \\ E(\Delta^{\text{pre}}) :=& \{(1, (2, p')) : p' \in \{1, \dots, p\}\} \\ & \cup \{((k', p'), (k'+1, p')) : (k', p') \in \{2, \dots, k\} \times \{1, \dots, p\}\} \\ & \cup \left\{((k', p'), (\bar{k}'_1, p')) : (k', p') \in \{2, \dots, k\} \times \{1, \dots, p\}\right\} \end{aligned} \tag{5.90}$$

and we let now $\mathfrak{P}$ be the set of partitions of $V_b(\Delta^{\text{pre}})$ which include the singleton $i_0 := \{1\}$, so that we have

$$\mathbb{E}\left|\left(G_2^E(\Lambda G_2^E)^{k-1}\right)_{2,\alpha\mathbf{x}}\right|^p = \sum_{P\in\mathfrak{P}} \sum_{\mathbf{a}_b} w_{\mathbf{a}_b}(\Delta(P)) \mathcal{A}_{\mathbf{a}_b}(\Delta(P)),$$

where the sum $\sum_{\mathbf{a}_b}$ ranges over $\mathbf{a}_b$ with $a_{i_0} = \alpha$. Similarly to the case $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_M}$, we have the following lemma.

Lemma 5.4.23. *Let $V_b^0 \subseteq V_b \backslash \{i_0\}$ be the set of black vertices i in Γ which have $\deg_{\mathbf{x}}(i) = 0$, and let $V_b^1 \subseteq V_b \backslash \{i_0\}$ be those with $\deg_{\mathbf{x}}(i) = 1$. Let $V_b^* \subseteq V_b$ be the vertices which correspond to singletons of P. Then,*

$$\left|V_b^0\right| + \frac{1}{2}\left|V_b^1\right| - \frac{1}{2}|V_b^*| \leq \frac{(k-1)p}{2} \tag{5.91}$$

We fix a partition ζ of $V_w(\Gamma)$ as before and define the quotient graph Γ_ζ whose white vertices are identified according to ζ. We extract a subset $\widetilde{E_o}(\Gamma_\zeta)$ of p off-diagonal edges incident on i_0. As in the proof of Lemma 5.1.8, we define the collection $V_w^*(\Gamma)$ of lone white vertices of set ℓ and the number $\widetilde{\ell}$ of them which are one end-point of an edge in $\widetilde{E_o}(\Gamma)$. We have as in equation (5.38), that

$$\left|E_o(\Gamma_\zeta) \setminus \widetilde{E_o}(\Gamma_\zeta)\right| \geq \frac{1}{2}\left(\ell - \widetilde{\ell}\right). \tag{5.92}$$

The following lemma is just like Lemma 5.3.7, letting ζ_{b} be the bad identifications and r be the number of blocks of size 2 in ζ_{b}.

Lemma 5.4.24. *For any graph Γ and any partition ζ such that*

$$\mathbb{E} \prod_{e \in E(\Gamma_\zeta)} \mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(e, \Gamma_\zeta) \neq 0,$$

we have

$$|V_w(\Gamma_\zeta)| \leq \frac{|E_X(\Gamma)|}{2} - \frac{r}{2} + \frac{\ell}{2} - \frac{|V_b^*|}{2}.$$

Lemma 5.3.8 holds verbatim. We conclude:

$$\begin{aligned}
&\mathbb{E} \sum_{\mathbf{a}_b} w_{\mathbf{a}_b} \sum_{\mathbf{a}_w} \prod_{e \in E(\omega_\sigma \Gamma_\zeta)} \mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(e, \omega_\sigma \Gamma_\zeta) \\
&\lesssim \left(\sum_{\mathbf{a}_b} w_{\mathbf{a}_b} \right) \sum_{\mathbf{a}_w} \prod_{e \in E_G(\omega_\sigma \Gamma_\zeta)} \mathcal{A}_{\mathbf{a}_b, \mathbf{a}_w}(e, \omega_\sigma \Gamma_\zeta) \cdot O_\prec \left(N^{-\frac{1}{2}|E_X(\Gamma)|} N^{-(k-1)p/2} \right) \\
&\prec \left(\sum_{\mathbf{a}_b} w_{\mathbf{a}_b} \right) \left(\sum_{\mathbf{a}_w} 1 \right) \cdot \left(\psi^{p-r} \psi^{\ell + \frac{1}{2}(\ell - \widetilde{\ell})} N^{-2\ell\epsilon_D} \right) \left(N^{-\frac{1}{2}|E_X(\Gamma)|} N^{-(k-1)p/2} \right) \\
&\prec N^{\left|V_b^0\right| + \frac{1}{2}\left|V_b^1\right|} N^{\frac{1}{2}|E_X(\Gamma)| - \frac{r}{2} + \frac{\ell}{2}} \cdot \left(\psi^{p-r} \psi^{\ell + \frac{1}{2}(\ell - \widetilde{\ell})} N^{-2\ell\epsilon_D} \right) \cdot \left(N^{-\frac{1}{2}|E_X(\Gamma)|} \right) \cdot N^{-(k-1)p/2} N^{-\frac{1}{2}\left|V_b^*\right|} \\
&= \left(N^{-\frac{r}{2}} \psi^{-r} \right) \left(\psi^{\frac{3}{2}\ell} N^{\frac{\ell}{2}} N^{-2\ell\epsilon_D} \right) \left(\psi^{p - \frac{1}{2}\widetilde{\ell}} \right) \cdot N^{\left|V_b^0\right| + \frac{1}{2}\left|V_b^1\right| - \frac{1}{2}\left|V_b^*\right| - \frac{1}{2}(k-1)p}.
\end{aligned}$$

Now we use that $N^{-\frac{1}{2}} \lesssim \psi$, that $\psi^3 N^{-2\epsilon_D} \leq N^{-1}$, that $\widetilde{\ell} \leq p$, and equation (5.91) to bound the above by $\psi^{p/2}$, and we conclude.

□

5.5 The proof of Lemma 5.1.5

Lemma 5.5.1. *The resolvents of $(U_M^* DX)^\#$ and $(U^* DX)^\#$ are close at the spectral edge: for $z \in \mathbf{S}^e$, we have*

$$\left| \underline{G^{U^* D}} - \underline{G^{U_M^* D}} \right| \prec \frac{N^{-\delta}}{N\eta_0}. \tag{5.93}$$

Proof of Lemma 5.1.5. An immediate corollary of Lemma 5.1.4 is that $(DX)^{\#}$ satisfies the $\epsilon_0/2$-valley and weak level repulsion conditions, so the result follows from Lemma 5.5.1 in the same way that Lemma 5.1.3 was proven. $\square$

Proof of Lemma 5.5.1. We will need the following consequence of Lemmas 5.1.7, 5.2.12 and 5.4.17.

Lemma 5.5.2. *For deterministic unit vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}_{K+M}}$, we have*

$$\left|\left(G^D - \Pi\right)_{\mathbf{xy}}\right| \prec \psi_{(KK)}.$$

We begin by establishing

$$G_{1,ii}^{U_M^* D} = G_{1,ii}^{U^* D} + O_{\prec}\left(\max_{\alpha}\left|G_{1,i\alpha}^{U^* D}\right|^2\right) \tag{5.94}$$

for $i \in \mathcal{I}_M$. Note that $G_1^{U_M^* D} = G_1^{UD^{(\mathcal{I}_K)}}$. Let $\mathcal{I}_K = \{\alpha_1, \ldots, \alpha_K\}$. We use Lemma 5.2.4 to see that

$$G_{1,ii}^{U^* D^{(\alpha_1 \cdots \alpha_K)}} = G_{1,ii}^{U^* D^{(\alpha_1 \cdots \alpha_{K-1})}} - \frac{G_{1,i\alpha_K}^{U^* D^{(\alpha_1 \cdots \alpha_{K-1})}} G_{1,\alpha_K i}^{U^* D^{(\alpha_1 \cdots \alpha_{K-1})}}}{G_{1,\alpha_K \alpha_K}^{U^* D^{(\alpha_1 \cdots \alpha_{K-1})}}}.$$

Inductively we may assume that

$$G_{1,jk}^{U^* D^{(\alpha_1 \cdots \alpha_{K-1})}} = G_{1,jk}^{U^* D} + O_{\prec}\left(\max_{\alpha \in \mathcal{I}_K \setminus \{\alpha_K\}} \left|G_{j,\alpha}^{U^* D}\right|\left|G_{k,\alpha}^{U^* D}\right|\right).$$

For $j, k \in \mathcal{I}_M \cup \{\alpha_K\}$. Equation (5.94) then follows by

$$\begin{aligned} &\left|G_{1,i\alpha_K}^{U^* D}\right| + \left|G_{1,\alpha_K i}^{U^* D}\right| + \left|G_{1,\alpha_K \alpha_K}^{U^* D}\right|^{-1} \\ &= \left|zG_{i\alpha_K}^{U^* D}\right| + \left|zG_{\alpha_K i}^{U^* D}\right| + \left|zG_{\alpha_K \alpha_K}^{U^* D}\right|^{-1} \prec 1, \end{aligned}$$

which is a consequence of Lemma 5.5.2.

Equation (5.94) then yields

$$\begin{aligned} &\underline{G^{U^* D}} - \underline{G^{U_M^* D}} \\ &= \frac{1}{M}\sum_{i \in \mathcal{I}_M}\left(G_{ii}^{U^* D} - G_{ii}^{U_M^* D}\right) + \frac{1}{M+K}\sum_{i \in \mathcal{I}_K} G_{ii}^{U^* D} + \frac{K}{M(M+K)}\sum_{i \in \mathcal{I}_M} G_{ii}^{U^* D} \\ &\prec \max_{\alpha \in \mathcal{I}_K} \frac{1}{M}\sum_{i \in \mathcal{I}_M}\left|G_{i,\alpha}^{U^* D}\right|^2 + M^{-1} \end{aligned}$$

by Lemma 5.5.2, the first term of which we may then write as

$$\frac{1}{N}\sum_{i\in\mathcal{I}_K\cup\mathcal{I}_M}\left|G^{U^*D}_{1,i\alpha}\right|^2=\frac{1}{N}\left\|G_1^DU^*\mathbf{e}_\alpha\right\|^2_{\mathbb{R}^{\mathcal{I}_{K+M}}}=\frac{1}{N}\sum_{i\in\mathcal{I}_K\cup\mathcal{I}_M}\left|\mathbf{e}_i^*G_1^DU\mathbf{e}_\alpha\right|^2. \tag{5.95}$$

Lemma 5.2.12, Lemma 5.1.7, and Lemma 5.4.17 then yield together

$$\prec\frac{1}{N}\sum_{i\in\mathcal{I}_K\cup\mathcal{I}_M}\left|\psi_{(iK)}\right|^2\lesssim\left|\psi_{(MK)}\right|^2\leq\frac{N^{-\delta}}{N\eta_0}$$

as desired. □

5.6 Proofs of Lemma 5.1.6

Recalling the singular values and vectors $d_\alpha,\mathbf{v}_\alpha,\mathbf{w}_\alpha$ of B, we define the random singular values and vectors $\widetilde{d}_\alpha,\widetilde{\mathbf{v}}_\alpha,\widetilde{\mathbf{w}}_\alpha$ of D_MB. Since $\|D-I\|\prec N^{-1/2}$, simple perturbation theory yields the estimates

$$\max_\alpha\left(\left|\widetilde{d}_\alpha-d_\alpha\right|+\|\widetilde{\mathbf{v}}_\alpha-\mathbf{v}_\alpha\|+\|\widetilde{\mathbf{w}}_\alpha-\mathbf{w}_\alpha\|\right)\prec N^{-1/2} \tag{5.96}$$

We define the matrices

$$[\widetilde{d}]:=\operatorname{diag}(\widetilde{d}_1,\ldots,\widetilde{d}_K),\quad [\widetilde{\mathbf{v}}]:=\begin{pmatrix}| & & |\\ \widetilde{\mathbf{v}}_1 & \cdots & \widetilde{\mathbf{v}}_K\\ | & & |\end{pmatrix}\quad [\widetilde{\mathbf{w}}]:=\begin{pmatrix}| & & |\\ \widetilde{\mathbf{w}}_1 & \cdots & \widetilde{\mathbf{w}}_K\\ | & & |\end{pmatrix}.$$

We will need to following, somewhat crude, result for the proof of Lemma 5.1.6.

Lemma 5.6.1. *For $\alpha,\beta\in\{1,\ldots,K\}$, we have*

$$\left|\left(G^{U_M^*D}-\Pi\right)_{\widetilde{\mathbf{v}}_\alpha\widetilde{\mathbf{v}}_\beta}\right|\prec N^{-\delta}.$$

A particular consequence of corollary 5.6.1 is that for a fixed integer $L>0$, letting $v_1,\ldots,v_N$ be the eigenvectors of $(U_M^*DX)^\#$

$$|\widetilde{\mathbf{v}}_\alpha^*v_i|\prec N^{-1/3-\delta-\epsilon_0} \tag{5.97}$$

for $i = 1, \ldots, L$ and $\alpha \in \mathcal{I}_K$. Indeed, letting $\mathbf{S}^e \ni z = \lambda_i + i\eta_0$,

$$\begin{aligned} N^{-\delta} &\succ \Im \Pi_{\widetilde{\mathbf{v}}_\alpha \widetilde{\mathbf{v}}_\alpha}(z) + O_\prec(\psi_{(KK)}) = \Im G^{U_M^* D}_{\widetilde{\mathbf{v}}_\alpha \widetilde{\mathbf{v}}_\alpha}(z) \\ &= \sum_j \frac{\eta_0}{(\lambda_i - \lambda_j)^2 + \eta_0^2} |\widetilde{\mathbf{v}}_\alpha^* v_j|^2 \geq \frac{1}{\eta_0} |\widetilde{\mathbf{v}}_\alpha^* v_i|^2. \end{aligned} \tag{5.98}$$

Proof of Lemma 5.1.6. This proof is similar to the proof of lemma 6.8 in [KY13b]. However, because in our setting we are able to assume the weak level repulsion condition, the proof is easier (the authors note in the beginning of the proof that the proof would in fact be easier under this assumption).

We must add a little independent randomness to take care of a technical detail later. Rather than considering the eigenvalues of $[\mathbf{N}(SX)]^\# = \left[\begin{pmatrix} D_M B & I \end{pmatrix} DX\right]^\#$, we consider the eigenvalues of

$$\widetilde{H}_\gamma := \begin{pmatrix} D_M B & I \end{pmatrix} \left([DX]^\# + \gamma \mathcal{X}\right) \begin{pmatrix} D_M B & I \end{pmatrix}^*$$

for some very small $\gamma > 0$, where $\mathcal{X}$ is an independent conformable Wishart matrix. Sending $\gamma \to 0$ and using the Lipschitz continuity of the eigenvalues with respect to γ on the high probability event $\left\{\left\|\mathbf{N}(SX)^\#\right\| \leq 10\right\}$, thanks to Weyl's inequality, we may conclude.

In order to use Lemma 5.2.7 to study the eigenvalues of $\widetilde{H}_\gamma$, we write

$$\begin{pmatrix} D_M B & I \end{pmatrix} = \left(\begin{pmatrix} D_M B & I \end{pmatrix}^\#\right)^{1/2} U_M^*$$

which is the reason for the definition of U_M^*. Therefore, since

$$\begin{pmatrix} D_M B & I \end{pmatrix}^\# = I + [\widetilde{\mathbf{v}}]^* [\widetilde{d}] [\widetilde{\mathbf{v}}],$$

Lemma 5.2.7 shows us that $\widetilde{H}_\gamma$ has an eigenvalue at $x \notin \sigma(H_\gamma)$, where $H_\gamma := U_M^* \left([DX]^\# + \gamma \mathcal{X}\right) U_M$ if and only if

$$M(x) := [\widetilde{d}]^{-1} + [\widetilde{\mathbf{v}}]^* (I + x (H_\gamma - x)^{-1}) [\widetilde{\mathbf{v}}]$$

is singular. Fix an integer $L > 0$. Let $\Omega_{\mathrm{LR}}^{U_M^* D}$ be the event that $(U_M^* DX)^\#$ satisfies the weak level repulsion condition; by choosing γ very small we may assume H_γ does as well. For

$i, j \in \{1, \ldots, M\}$, define

$$\widetilde{\lambda}_i = \lambda_i\left(\widetilde{H}_\gamma\right), \quad \lambda_i = \lambda_i\left(H_\gamma\right)$$

We first establish the claim that for each $i \le L$, on the event $\Omega_{\mathrm{LR}}^{U_M^* D}$,

$$n_j := \left|\left\{j : \widetilde{\lambda}_j \in [x_i^-, x_i^+]\right\}\right| \geq 1 \tag{5.99}$$

where $x_i^\pm = \lambda_i \pm N^{-2/3-\gamma/2}$. Let us demonstrate why the claim (5.99) is sufficient before proving it. If $n_1 \geq 1$, then by Lemma 5.2.8 and the weak level repulsion condition, we see that

$$\left|\left\{j : \widetilde{\lambda}_j > x_2^+\right\}\right| \geq K + 1$$

so that by Lemma 5.2.9,

$$\left|\left\{j : \widetilde{\lambda}_j > x_2^+\right\}\right| = K + 1$$

so that $n_j = 1$ and

$$\left|\{j : \widetilde{\lambda}_j \in [x_2^+, \operatorname{supp} \varrho_E + C_0/2] \setminus [x_1^-, x_1^+]\right| = 0$$

We may repeat this argument for $n_2, \ldots, n_L$, establishing that $n_\alpha = 1$ for $\alpha = 1, \ldots, L$ and

$$\left|\{j : \widetilde{\lambda}_j \in [x_{L+1}^+, \operatorname{supp} \varrho_E + C_0/2] \setminus [x_1^-, x_1^+] \setminus \cdots \setminus [x_L^-, x_L^+]\right| = 0$$

Now let us establish the claim (5.99). It suffices, by Lemma 5.2.7, to establish that $M(x)$ is singular for at least 1 value x in each of the (centerless) intervals $[x_i^-, x_i^+] \setminus \{\lambda_i\}$ for $i = 1, \ldots, L$. Fix such an i and let $x \in [x_i^-, x_i^+]$. Write

$$M(x) = [\widetilde{d}]^{-1} + [\widetilde{\mathbf{v}}]^* \left(x \frac{v_i v_i^*}{\lambda_i - x}\right) [\widetilde{\mathbf{v}}] + [\widetilde{\mathbf{v}}]^* \left(1 + x \sum_{j \neq i} \frac{v_j v_j^*}{\lambda_j - x}\right) [\widetilde{\mathbf{v}}] \tag{5.100}$$

where $\sum_j \lambda_j v_j v_j^*$ is an eigendecomposition of H_γ. Since

$$\operatorname{dist}\left(x, \sigma\left(H_\gamma\right) \setminus \{\lambda_i\}\right) > \eta_0$$

because of the weak level repulsion condition, we may write

$$\begin{aligned}
&\left\| [\widetilde{\mathbf{v}}]^* \left(1 + x\sum_{j\neq i} \frac{v_j v_j^*}{\lambda_j - x}\right)[\widetilde{\mathbf{v}}] - [\widetilde{\mathbf{v}}]^*\left(1 + x\sum_j \frac{v_j v_j^*}{\lambda_j - (x+i\eta_0)}\right)[\widetilde{\mathbf{v}}]\right\| \\
&\le C \max_{\alpha,\beta\in\mathcal{I}_K}\left(\left|\Im\widetilde{\mathbf{v}}_\alpha^*\left(x\sum_j \frac{v_j v_j^*}{\lambda_\beta - (x+i\eta_0)}\right)\widetilde{\mathbf{v}}_\beta\right| + \left|\widetilde{\mathbf{v}}_\alpha^*\left(x\frac{v_i v_i^*}{\lambda_i - (x+i\eta_0)}\right)\widetilde{\mathbf{v}}_\beta\right|\right) \qquad (5.101)\\
&= C \max_{\alpha,\beta\in\mathcal{I}_K}\left(\left|\Im([DX]^\# + \gamma\mathcal{X} - (x+i\eta_0))^{-1}_{\widetilde{\mathbf{v}}_\alpha\widetilde{\mathbf{v}}_\beta}\right| + O_\prec(N^{-1/3-\delta-\epsilon_0}\eta_0^{-1})\right) \\
&= O_\prec(N^{-\delta}),
\end{aligned}$$

where the second to last equality follows from equation (5.97) and the last from Lemma 5.2.2. Then we may write

$$M(x) = [\widetilde{d}]^{-1} + [\widetilde{\mathbf{v}}]^*\left(x\frac{v_i v_i^*}{\lambda_i - x}\right)[\widetilde{\mathbf{v}}] + [\widetilde{\mathbf{v}}]^*\left(1 + xG^{U_M^* D}(x+i\eta_0)\right)[\widetilde{\mathbf{v}}] + O_\prec(N^{-\delta}).$$

By Lemma 5.6.1,

$$\left\| [\widetilde{\mathbf{v}}]^*\left(1 + xG^{U_M^* D}(x+i\eta_0)\right)[\widetilde{\mathbf{v}}] - [\widetilde{\mathbf{v}}]^*\left(1 + x\Pi(x+i\eta_0)\right)[\widetilde{\mathbf{v}}]\right\| \prec N^{-\delta}$$

and, using Lemma 5.2.11, we have

$$[\widetilde{\mathbf{v}}]^*\left(1 + x\Pi(x+i\eta_0)\right)[\widetilde{\mathbf{v}}] = -\frac{1}{\sqrt{y}} + O_\prec(N^{-\delta}).$$

We therefore get

$$M(x) = \left([\widetilde{d}]^{-1} - \frac{1}{\sqrt{y}}\right) + [\widetilde{\mathbf{v}}]^* U_M^*\left(x\frac{v_i v_i^*}{\lambda_i - x}\right)U_M[\widetilde{\mathbf{v}}] + O_\prec(N^{-\delta}).$$

By equation (5.97), $\left\| [\widetilde{\mathbf{v}}]^* U_M^*\left(x\frac{v_i v_i^*}{\lambda_i - x}\right)U_M[\widetilde{\mathbf{v}}]\right\| \prec N^{-\delta/2}$ if $x \in [x_i^-, x_i^+], x < \lambda_i - \eta_0 N^{\delta/2}$. Because of assumption $d_\alpha \ge y + C$, the eigenvalues of $\left([\widetilde{d}]^{-1} - \frac{1}{\sqrt{y}}\right)$ are all $< -c_1$ for some small positive constant c_1. Thus $M(x_i^-)$ has only negative eigenvalues. Because $\mathcal{X}$ is independent of X and has absolutely continuous distribution, we see that

$$\mathbb{P}\left([v_i^* U_M \widetilde{\mathbf{v}}] = 0\right) = 0.$$

So we see that as $x \to \lambda_i$, $M(x)$ has only positive eigenvalues. It is then easy by the intermediate value theorem to conclude that $M(x)$ is singular for some $x \in (x_i^-, \lambda_i)$, and we conclude the proof of Lemma 5.1.6. □

Proof of Lemma 5.6.1. Note first that $\Pi = \frac{-1}{z(1+m(z))} I + O_{\prec}(N^{-1/2})$, ie, Π is nearly isotropic.

We write, just as in the proof of equation (5.22),

$$\begin{aligned} G^{U_M^* D}_{\widetilde{\mathbf{v}}_\alpha, \widetilde{\mathbf{v}}_\beta} &= z^{-1} G^{U_M^* D}_{1, \widetilde{\mathbf{v}}_\alpha, \widetilde{\mathbf{v}}_\beta} \\ &= z^{-1} G^{U^* D}_{1, \widetilde{\mathbf{v}}_\alpha, \widetilde{\mathbf{v}}_\beta} + O_{\prec}\left(\max_{\substack{\gamma \in \{1,\ldots,K\} \\ \delta \in \mathcal{I}_K}} \left| G^{U^* D}_{1, \widetilde{\mathbf{v}}_\gamma, \mathbf{e}_\delta} \right|^2 \right) \\ &= G^{U^* D}_{\widetilde{\mathbf{v}}_\alpha, \widetilde{\mathbf{v}}_\beta} + O_{\prec}\left(\max_{\substack{\gamma \in \{1,\ldots,K\} \\ \delta \in \mathcal{I}_K}} \left| G^{U^* D}_{\widetilde{\mathbf{v}}_\gamma, \mathbf{e}_\delta} \right|^2 \right) \\ &= G^{D}_{U\widetilde{\mathbf{v}}_\alpha, U\widetilde{\mathbf{v}}_\beta} + O_{\prec}\left(\max_{\substack{\gamma \in \{1,\ldots,K\} \\ \delta \in \mathcal{I}_K}} \left| G^{D}_{U\widetilde{\mathbf{v}}_\gamma, U\mathbf{e}_\delta} \right|^2 \right). \end{aligned} \tag{5.102}$$

Now it follows by the definition of U that $U\widetilde{\mathbf{v}}_\alpha, U\mathbf{e}_\beta \in \mathbb{R}^{\mathcal{I}_K} \oplus \operatorname{span}\{\widetilde{\mathbf{v}}_\alpha : \alpha \in \mathcal{I}_K\}$ for $\alpha = 1, \ldots, K$ and $\beta \in \mathcal{I}_K$. Therefore, for $\mathbf{x}, \mathbf{y} \in \{U\widetilde{\mathbf{v}}_\alpha\} \cup \{U\mathbf{e}_\beta\}$, we may write

$$\begin{aligned} &\mathbf{x}^*(G^D - \Pi)\mathbf{y} \\ &= \left(\sum_{\alpha \in \{1,\ldots,K\}} a_\alpha \mathbf{e}_\alpha + \sum_{\alpha \in \mathcal{I}_K} b_\alpha \widetilde{\mathbf{v}}_\alpha \right)^* (G^D - \Pi) \left(\sum_{\alpha \in \{1,\ldots,K\}} a'_\alpha \mathbf{e}_\alpha + \sum_{\alpha \in \mathcal{I}_K} b'_\alpha \widetilde{\mathbf{v}}_\alpha \right)^* \end{aligned}$$

for possibly random $a_\alpha, b_\alpha, a'_\alpha, b'_\alpha$. Expanding this out into $(2K)^2$ terms and applying Lemmas 5.1.7 and 5.4.17 (because $\widetilde{\mathbf{v}}_\alpha$ is deterministic $+O_{\prec}(N^{-1/2})$) to each term bounds the above by $O_{\prec}(\psi_{(KK)})$. Therefore, using equation (5.102) and also that $U^*\Pi U = \Pi + O_{\prec}(N^{-1/2})$, we obtain

$$\left(G^{U_M^* D} - \Pi\right)_{\widetilde{\mathbf{v}}_\alpha \widetilde{\mathbf{v}}_\beta} = (G^D - \Pi)_{U\widetilde{\mathbf{v}}_\alpha, U\widetilde{\mathbf{v}}_\beta} + O_{\prec}(\psi_{(KK)}) = O_{\prec}(\psi_{(KK)}).$$

To obtain the error term in the first equality above, note that $\widetilde{\mathbf{v}}_\alpha$ and $\mathbf{e}_\beta$ are orthogonal.

□

CHAPTER 6

Future Work

Using the technique of [BDW20] to represent the eigenvector components as residues of the Green function, the key ingredient for Theorem 2.2.3 is as follows:

Lemma 6.0.1. *Fix $\alpha \leq K$ and assume $d_1 = 0(1)$. For any fixed non-negative integers n_1, n_2 and deterministic unit vectors $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^{\mathcal{I}_M}$ satisfying $\|\mathbf{w}_\iota\|_\infty \leq N^{-\epsilon'}$ for some ϵ', and letting*

$$\begin{aligned}
\mathbf{M}^{(n_1,n_2)} &:= \mathbf{M}_1^{(n_1,n_2)} + \mathbf{M}_2^{(n_1,n_2)} + \mathbf{M}_3^{(n_1,n_2)} \\
\mathbf{M}_1^{(n_1,n_2)} &:= \mathbf{w}_1\mathbf{w}_2^* \\
\mathbf{M}_2^{(n_1,n_2)} &:= \mathbf{w}_1\mathbf{w}_2^*\partial_z^{(n_2)}(G - \Pi) \\
\mathbf{M}_3^{(n_1,n_2)} &:= \partial_z^{(n_1)}(G - \Pi)\mathbf{w}_1\mathbf{w}_2^*\partial_z^{(n_2)}(G - \Pi)
\end{aligned}$$

where again the omitted spectral argument of G and Π is $\phi_\alpha \asymp d_1$, we have under Assumptions 1 and equations (2.2) *and* (2.3) *that*

$$\operatorname{Tr}_{(\mathcal{I}_M)} \mathbf{M}^{(n_1,n_2)}(D - I) = O_\prec\left(N^{-1/2-\epsilon}\right)$$

for some $\epsilon > 0$. Under Assumption 2 the error term should be replace with $O_P\left(N^{-1/2-\epsilon}\right)$.

Lemma 6.0.1 is proven in exactly the same way as Lemmas 4.1.7 and 4.1.8. It is fairly routine to verify that $\partial_z^{(n_1)}G$ satisfies the same sort of local laws, Lemmas 4.2.7, 4.2.8 and 4.2.10, as G for $z \in \mathbf{S}^O$, and also the same derivative rule (4.40). The replacement of $\mathbf{v}_\alpha$ in the definition of $\mathbf{M}$ with $\mathbf{w}_\iota$ now is also inconsequential: the only property of $\mathbf{v}_\alpha$ that we use is its delocalization, which we assume now for $\mathbf{w}$.

We postpone further justification of Lemma 6.0.1 and an explanation of how it implies Theorem 2.2.3 to future work.

CHAPTER 7

Summary

In this work we discussed the spiked eigenvalues and eigenvectors of sample covariance matrices—our work is only the second on the topic, and extends the scope of the existing work [MJM21] to the statistically relevant setting of spiked matrices with weak factors. Moreover, ours is the first work to discuss the non-spiked eigenvalues of spiked correlation matrices.

A defining characteristic of our work has been the treatment of eigenstructures of correlation matrices by *a deterministic comparison* between the Green function of the correlation matrix and suitable covariance matrix; that is, we show that for a fixed realization of the randomness X, the correlation matrix and the covariance matrix have very similar Green functions, yielding similarity of their eigenstructures. This is as opposed to a distributional result which would only establish that the eigenstructures of the correlation matrix have the same *distribution* as those of the covariance matrix. In this way, we have alleviated some of the practictioner's concern surrounding the use of correlation matrices: for spiked matrices with sufficiently weak factors (see the remark following Theorem 2.2.4), the practitioner may confidently use the sample correlation matrix as a substitute for the sample covariance matrix with normalized variances. In other words, for factor models with weak factors, our work has answered in the affirmative the important question of whether normalizing data by the sample variances, thus forming the sample correlation matrix, is an acceptable substitute for normalizing by the true variances. This is especially useful in the most common scenario that the sample covariance matrix with normalized variances is unavailable, i.e., in the case of unknown calibrations or different units of the many variables being measured.

However, our work also uncovered a surprising feature of the sample correlation matrix of a factor model: if the factors are not weak enough (as manifested by eigenvalues of the model exceeding $N^{5/6}$, or equivalently, the parameter ϵ_D controlling the weakness of the factors satisfying $\epsilon_D < 1/12$), then the fluctuations of the extreme non-spiked eigenvalues will be driven by the "signal part" of the randomness X_K and have Gaussian fluctuations on the scale $N^{-1/2-2\epsilon_D}$ rather than the typical "random matrix behavior," i.e., Tracy-Widom fluctuations on the scale $N^{-2/3}$. Thus we have uncovered a phase-transition phenomenon not entirely unlike the BBP phase transition (although one should be careful not to push the analogy too far).

Our establishment of deterministic bounds on the difference between the eigenstructures of sample correlation and sample covariance matrices has obvious statistical implications. Beyond this, we note that such deterministic comparisons of Green functions, which are a proxy for the eigenstructures, are somewhat unusual in the random matrix theory literature, with distributional statements usually being preferred. We hope that the techniques we developed for such deterministic comparisons, in particular our novel use of the polynomialization method in Chapter 5, may be useful for future work in RMT.

REFERENCES

[BBP05] Jinho Baik, Gerard Ben Arous, and Sandrine Péché. "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices." *Ann. Probab.*, **33**:1643–1697, 2005.

[BDW20] Z. Bao, X. Ding, K. Wang, and K. Wang. *Statistical Inference for Principal Components of Spiked Covariance Matrices.* [math.ST], 2020.

[BEK14] Alex Bloemendal, László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. "Isotropic local laws for sample covariance and generalized Wigner matrices." *Electron. J. Probab.*, **19**:1–53, 2014.

[BJ21] Zhidong Bai and Dandan Jiang. "Generalized four moment theorem and an application to CLT for spiked eigenvalues of high-dimensional covariance matrices." *Bernoulli*, **27**(1):274–294, February 2021.

[BK18] Florent Benaych-Georges and Antti Knowles. "Lectures on the local semicircle law for Wigner matrices." In *Panorama et Synthèses*, volume 53. Société Mathématique de France, 2018.

[BKY16] Alex Bloemendal, Antti Knowles, Horng-Tzer Yau, and Jun Yin. "On the principal components of sample covariance matrices." *Probab. Theory Relat. Fields*, **164**:459–552, 2016.

[Bor10] F. Bornemann. "On the numerical evaluation of distributions in random matrix theory: A review with an invitation to experimental mathematics." *Markov Processes and Related Fields*, **16 (4)**:803–866, 2010.

[Bre01] Leo Breiman. "Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author)." *Statistical Science*, **16**(3):199 – 231, 2001.

[BY12] Z. Bai and J. Yao. "On sample eigenvalues in a generalized spiked population model." *Journal of Multivariate Analysis*, **106**:167–177, 2012.

[CES21a] Giorgio Cipolloni, László Erdoős, and Dominik Schr"oder. "Eigenstate Thermalization for Wigner Matrices." 2021.

[CES21b] Giorgio Cipolloni, László Erdoős, and Dominik Schröder. "Thermalisation for Wigner matrices." 2021.

[CF91] R. Curto and L. Fialkow. "Recursiveness, positivity, and truncated moment problems." *Houston J. Math*, **17**(4):603–635, 1991.

[CS95] S. I. Choi and J. W. Silverstein. "Analysis of the limiting spectral distribution of large dimensional random matrices." *J. Multivariate Anal*, **54**:295–309, 1995.

[DS74] C. D. Dziuban and E. C. Shirkey. “When is a correlation matrix appropriate for factor analysis? Some decision rules.” *Psychological Bulletin*, **81(6)**:358–361, 1974.

[DY18] X. Ding and F. Yang. “A Necessary and Sufficient Condition for Edge Universality at the Largest Singular Value of Sample Covariance Matrices.” *Annals of Applied Probability*, **28**:1679–1738, 2018.

[EKY13] László Erdős, Antti Knowles, and Horng-Tzer Yau. “Averaging fluctuations in resolvents of random band matrices.” *Annales Henri Poincare*, **8**:1837–1926, 2013.

[El 09] Noureddine El Karoui. “Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond.” *Ann. Appl. Probab.*, **19**(6):2362–2405, December 2009.

[EP18] Alan Edelman and Per-Olof Persson. “Numerical Methods for Eigenvalues Distributions of Random Matrices.” 2018.

[ESY11] L. Erdős, B. Schlein, and HT. Yau. “Universality of random matrices and local relaxation flow.” *Invent. math.*, **185**:75–119, 2011.

[EYY12] László Erdoős, Horng-Tzer Yau, and Jun Yin. “Rigidity of eigenvalues of generalized Wigner matrices.” *Advances in Mathematics*, **229**(3):1435–1515, 2012.

[FFH20] Jianqing Fan, Yingying Fan, Xiao Han, and Jinchi Lv. *Asymptotic Theory of Eigenvectors for Random Matrices With Diverging Spikes.* Journal of the American Statistical Association, 2020.

[FMT86] J. Ford, ROBERT MacCALLUM, and MARIANNE TAIT. “The Application of Exploratory Factor Analysis in Applied Psychology: A Critical Review and Analysis.” *Personnel Psychology*, **39**:291 – 314, 06 1986.

[GH03] J. Geronimo and T. P. Hill. “Necessary and Sufficient Condition that the Limit of Stieljes Transforms is a Stieljes Transform.” *Journ. Approx. Theory*, **121**:54–60, 2003.

[HRP14] Daniel E. Hack, Carl W. Rossler, and Lee K. Patton. “Multichannel Detection of an Unknown Rank-N Signal Using Uncalibrated Receivers.” *IEEE Signal Process. Lett.*, **21**(8):998–1002, 2014.

[Joh01] I. M. Johnstone. “On the distribution of the largest eigenvalue in principal components analysis.” *Ann. Statist.*, **29**:295–327, 2001.

[KY13a] Antti Knowles and Jun Yin. “Eigenvector Distribution of Wigner Matrices.” *Probab. Theory Relat. Fields*, **155**:543–582, 2013.

[KY13b] Antti Knowles and Jun Yin. "The Isotropic Semicircle Law and Deformation of Wigner Matrices." *Commun. Pur. Appl. Math*, **66**:1663–1749, 2013.

[KY17] Antti Knowles and Jun Yin. "Anisotropic Local Laws for Random Matrices." *Probab. Theory Relat. Fields*, **169**:257–352, 2017.

[LLY22] Van Latimer, Jinchi Lv, and Jun Yin. "Asymptotic Theory of Spiked Eigenvalues and Eigenvectors for Correlation Matrices with Weak Factors." In preparation, 2022.

[LP20] Olivier Ledoit and Sandrine Péché. "Analytical nonlinear shrinkage of large-dimensinal sample covariance matrices." *Annals of Statistics*, **48(5)**:3043–3065, 2020.

[LS16] Ji Oon Lee and Kevin Schnelli. "Tracy–Widom distribution for the largest eigenvalue of real sample covariance matrices with general population." *Ann. Appl. Probab*, **26(6)**:3786–3839, 2016.

[LY12] J. O. Lee and Jun Yin. "A Necessary and Sufficient Condition for Edge Universality of Wigner Matrices." *Duke Mathematical Journal*, **163**(10):27094–24147, 2012.

[LY22] Van Latimer and Jun Yin. "Extreme Non-Spiked Eigenvalues of Spiked Sample Correlation Matrices." In preparation, 2022.

[MJM21] David Morales-Jimenez, Iain Johnstone, Matthew R. McKay, and Jeha Yang. "Asymptotics of eigenstructure of sample correlation matrices for high-dimensional spiked models." *Statistica Sinica*, **31**:571–601, 2021.

[Pau07] Debashis Paul. "Asymptotics of sample eigenstructure for a large-dimensional spiked covariance model." *Statistica Sinica*, **17**:1617–1642, 2007.

[PY12] Natesh Pillai and Jun Yin. "Edge universality of correlation matrices." *Ann. Statist.*, **40**(3):1737–1763, June 2012.

[PY14] Natesh Pillai and Jun Yin. "Universality of Covariance Matrices." *Ann. Appl. Probab.*, **24**:935–1001, 2014.

[SB95] J. W. Silverstei and Z. D. Bai. "On the empirical distribution of eigenvalues of a class of large dimensional random matrices." *J. Multivar. Anal.*, **54**:175–192, 1995.

[TV11] Terence Tao and Van Vu. "Random matrices: Universality of local eigenvalue statistics." *Acta Mathematica*, **206**(1):127 – 204, 2011.