The extreme eigenvalues of two types of random matrices

Zhang, Zhixiang

2021

Zhang, Z. (2021). The extreme eigenvalues of two types of random matrices. Doctoral thesis, Nanyang Technological University, Singapore. <https://hdl.handle.net/10356/150804>

<https://hdl.handle.net/10356/150804>

<https://doi.org/10.32657/10356/150804>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 20:38:23 SGT

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**The Extreme Eigenvalues of Two Types
of Random Matrices**

ZHANG ZHIXIANG

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

2021

The Extreme Eigenvalues of Two Types of Random Matrices

ZHANG ZHIXIANG

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

A thesis submitted to the Nanyang Technological

University in partial fulfilment of the requirement for the

degree of Doctor of Philosophy

2021

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me except where otherwise stated in this thesis. The thesis work has not been submitted for a degree or professional qualification to any other university or institution. I declare that this thesis is written by myself and is free of plagiarism and of sufficient grammatical clarity to be examined. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

.....Jan. 14, 2021.....

Date

....

ZHANG ZHIXIANG

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it of sufficient grammatical clarity to be examined. To the best of my knowledge, the thesis is free of plagiarism and the research and writing are those of the candidate's except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

.....Jan. 14, 2021.....

Date

.....

Prof. Pan Guangming

Authorship Attribution Statement

(A) This thesis does not contain any materials from papers published in peer-reviewed journals or from papers accepted at conferences in which I am listed as an author.

..... Jan 14, 2021

Date

.....

ZHANG ZHIXIANG

Abstract

The fluctuations of extreme eigenvalues of a large random matrix model is a central topic in random matrix theory, motivated by applications in principle component analysis, factor analysis, or signal detection problems. This thesis establishes asymptotic distributions for the largest eigenvalues of two types of matrices under the high dimensional setting where the dimension goes to infinity proportionally with the sample size.

The first type is the spiked sample covariance matrix. We prove that the spiked eigenvalues converge in distribution to Gaussian distributions without the conventional assumption of the block-diagonal structure on the population covariance matrices. We also show that the sample spiked eigenvalues and linear spectral statistics are asymptotically independent for such a spiked sample covariance model. With these theoretical results, we propose a statistic by combining the largest eigenvalues and the linear spectral statistics together to test the equality of two population covariance matrices.

The second type is the signal-plus-noise matrix. To be more specific, let $\mathbf{S} = \mathbf{R} + \mathbf{X}$ where $\mathbf{R}$ is the signal matrix and $\mathbf{X}$ is the noise matrix contains *i.i.d.* standardized entries. The signal matrix $\mathbf{R}$ is allowed to be full rank, which is rarely studied in literature compared with the low rank cases. Under a regularity condition of $\mathbf{R}$ that assures the square root behaviour of the spectral density near the edge, we prove that the largest eigenvalue of $\mathbf{S}\mathbf{S}^*$ has Tracy-Widom distribution under a tail condition on the entries of $\mathbf{X}$. Moreover, such a condition is proved to be necessary and sufficient to assure the Tracy-Widom law.

Acknowledgements

I would like to express my deep and sincere attitude to my supervisor, Professor Pan Guangming, for his invaluable guidance and unfailing encouragement throughout my Ph.D. study. He led me to the scientific research, raised many interesting questions to me, and helped me a lot with his wide knowledge. I feel very fortunate for having the opportunity to work with him.

I would like to thank Division of Mathematical Sciences of Nanyang Technological university for providing me the scholarships for pursuing the research study. I am thankful to my thesis committee members, Prof. Xiang Liming and Prof. Feng Qu for their help and guidance. I also appreciate many teachers, Prof. Wu Guohua, Prof. Li Yi, Prof. Bei Xiaohui, who have given me instructions in my first year study.

I would like to owe my gratitude to all the members of our research group, Liu Yiming, Zhang Bo, Han Xiao, Yang Qing, and Lin Zeqin. In particular, I want to thank Liu Yiming, who introduced me many ideas and finally led to some interesting results through our collaborate works. I want to thank those who have participated in our weekly seminars, including Wang Shaochen, Yang Wenzhi, Zhang Lingyue, Qian Fang, Zhang Yangchun and many others. It has been a great pleasure learning and discussing with them.

I am grateful to all my friends for their company during the good and hard times and helps in the life. Thanks for Sun Xin, Zhou Yongli, Chen Yonghao, Chen Shengyang, and Cui Dong, who have been enjoying playing football together with me these years.

Last but not the least, I would express great love and gratitude to my parents for supporting me throughout my Ph.D. study. Finally I owe my loving thanks to my girlfriend, Liu Xiaoyu, for all her warm encouragement and consistent support.

Contents

Abstract	vi
Acknowledgements	vii
List of Figures	xii
List of Tables	xiv
1 Introduction	1
2 Asymptotic joint distribution of spiked eigenvalues and LSS of sample covariance matrices	7
2.1 Introduction	7
2.2 The main results	8
2.2.1 Limiting laws for spiked eigenvalues	9
2.2.2 Asymptotic joint distribution of sample spiked eigenvalues and linear spectral statistics	13
2.2.3 Estimating the population eigenvectors associated with the spiked eigenvalues	15
2.2.4 Testing the equality of two spiked covariance matrices	16
2.3 Simulations	21
2.3.1 Approximation accuracy	22
2.3.2 Power discussion	22
2.4 Proof of Theorem 2.2.1	25
2.5 Proof of Theorems 2.2.2 and 2.2.3	39
2.6 Proof of Theorem 2.2.4	45
2.7 Proof of Theorem 2.2.5	48
2.8 Appendix	52
2.8.1 Truncation and Centralization	52

2.8.2	Proof of Lemma 2.2.6	56
2.8.3	Proof of Lemma 2.4.2	58
3	Tracy-Widom law for the largest eigenvalue of signal-plus-noise matrices	61
3.1	Introduction	61
3.2	The main result	64
3.2.1	The main result	64
3.2.2	Applications	69
3.2.2.1	Detection of interference signals	69
3.2.2.2	Data privacy	71
3.3	Notations and Tools	74
3.4	Tracy-Widom distribution for Gaussian $\mathbf{X}$	81
3.4.1	Rescaling of the model	82
3.4.2	Green function comparison	84
3.4.3	Green function flow	85
3.4.4	Proof of Proposition 3.4.2	87
3.4.5	Derivatives of Green functions	88
3.5	Entrywise and average local laws for Gaussian $\mathbf{X}$	95
3.5.1	Weak local law	96
3.5.1.1	Basic tools	96
3.5.1.2	A prior estimates on domain $\Omega(z)$	99
3.5.1.3	Proof of Lemma 3.5.2	108
3.5.2	Fluctuation lemma and entrywise local law	110
3.5.2.1	Fluctuation Lemma	110
3.5.2.2	Proof of Lemma 3.5.1	114
3.6	Anisotropic local law for general $\mathbf{X}$ with support $q = N^{-\delta}$, $0 < \delta < 1/2$	116
3.6.1	Anisotropic local law for Gaussian $\mathbf{X}$	116
3.6.2	Anisotropic local law for general $\mathbf{X}$ with support $q = N^{-1/2}$ and $\mathbb{E}x_{ij}^3 = 0$	119
3.6.3	Anisotropic and averaged local law for general $\mathbf{X}$ with sup- port $N^{-1/2}$	125
3.6.3.1	Anisotropic local law in the domain $\mathcal{D}(\tilde{c}, \tau, c_0)$	126
3.6.3.2	Averaged local law in the domain $\mathcal{D}(\tilde{c}, \tau, c_0)$	128
3.6.4	Anisotropic local law for $\mathbf{X}$ with support $q = N^{-\delta}$ ($0 < \delta < 1/2$).	132

3.6.5	Green function comparison for two matrices with support $N^{-1/2}$	133
3.7	Proof of Theorem 3.2.1 and Theorem 3.3.3	134
3.7.1	Averaged local law (3.3.15) and (3.3.16)	134
3.7.2	Proof of Theorem 3.2.1	142
3.8	Appendix	147
3.8.1	Preliminaries	147
3.8.2	Optical theorem from X_{22}	153
3.8.2.1	Expansion of B_1	153
3.8.2.2	Expansion of B_2	164
3.8.2.3	Expansion of B_3	166
3.8.2.4	Optical theorem from X_{22}	169
3.8.3	Optical theorem from X_{33}	171
3.8.4	Optical theorem from X_{32}	174
3.8.5	Optical theorem from $m_N X_{22}$	175
3.8.6	Proof of Lemma 3.8.3 and Lemma 3.8.4	176
4	Discussion and Future Research	179
	Bibliography	181

List of Figures

3.1	The histogram of centralized and rescaled largest eigenvalues with $M = N = 500$ and 10000 matrices.	69
3.2	Left: The first three eigenvalues of $\mathbf{Q}_1(t)$. Right: The first three eigenvalues of $\mathbf{Q}(t)$. The largest, second largest and third largest eigenvalues are represented by red triangle, black circle and green plus respectively.	73

List of Tables

2.1	Empirical size for testing $H_0 : \Sigma_1 = \Sigma_2 = \Sigma^{(i)}$ for data generated from Model i ($i = 1, 2, 3, 4, 5$) with $N(0,1)$ entries. The sample size is 100 for both samples.	23
2.2	Empirical size for testing $H_0 : \Sigma_1 = \Sigma_2 = \Sigma^{(i)}$ for data generated from Model i ($i = 1, 2, 3, 4, 5$) with $t_{10}/\sqrt{5/4}$ entries. The sample size is 100 for both samples.	24
2.3	Empirical power for testing $H_0 : \Sigma_1 = \Sigma_2$ where $\Sigma_1 = \Sigma^{(1)}$ and $\Sigma_2 = \Sigma^{(2)}$ with two types of data entries: $N(0,1)$ and $t_{10}/\sqrt{5/4}$. The sample size is 100 for both samples.	25
2.4	Empirical power for testing $H_0 : \Sigma_1 = \Sigma_2$ where $\Sigma_1 = \Sigma^{(3)}$ and $\Sigma_2 = \Sigma^{(4)}$ with two types of data entries: $N(0,1)$ and $t_{10}/\sqrt{5/4}$. The sample size is 100 for both samples.	25
2.5	Empirical power for testing $H_0 : \Sigma_1 = \Sigma_2$ where $\Sigma_1 = \Sigma^{(2)}$ and $\Sigma_2 = \Sigma^{(5)}$ with two types of data entries: $N(0,1)$ and $t_{10}/\sqrt{5/4}$. The sample size is 100 for both samples.	26
2.6	Empirical averaged relative error (η) and standard deviation (sd) of the proposed estimators for $\sum_{i=1}^p u_{1i}^4$, with true value 0.5427 in Model 4. The simulation replication is 500.	26
3.1	Size under Models 1 and 2	70
3.2	Power under Models 1 and 2 for $N(0,1)$	71
3.3	Power under Models 1 and 2 for t_{10}	72

Chapter 1

Introduction

The extreme eigenvalues of random matrices play an important role in many statistical problems such as principal component analysis (PCA), factor analysis, MANOVA, or classification. With the increasing volume of data collected in many scientific researches such as genomics, image processing or finance, the data usually can be modelled by a matrix with large size of dimensions and measurements. Random matrix theory turns out to be a powerful tool for studying asymptotic properties of largest eigenvalues of large random matrices. One can refer to the monograph of [Bai and Silverstein \(2010\)](#) or the review paper of [Paul and Aue \(2014\)](#) for a comprehensive review. The present thesis is devoted to studying the asymptotic distributions of largest eigenvalues for two types of large random matrices.

The first type of matrices we consider are the sample covariance matrices. Sample covariance matrices play a fundamental role in traditional multivariate statistics, see [Anderson \(1962\)](#). In the last two decades, there has been a lot of work in studying the asymptotic properties of the extreme eigenvalues of the sample covariance matrix under the high-dimensional setting, where the data dimension p grows with the sample size n proportionally. [Johnstone \(2001\)](#) showed that after suitable centralizing and scaling, the largest eigenvalues of a real null Wishart matrix converge to Tracy-Widom law. [Baik et al. \(2005\)](#) studied the phase transition behaviour of largest eigenvalues of a sample covariance matrix with Gaussian samples. The phase transition of sample eigenvalues depends on the magnitude of population eigenvalues. To be more specific, on one hand, when the largest population eigenvalues are smaller than a critical value, the corresponding largest sample eigenvalues adhere to the bulk spectrum and have Tracy-Widom distributions. On

the other hand, if the population eigenvalues are greater than that critical value, the largest sample eigenvalues are like outliers that located away from the bulk spectrum and converge to Gaussian distributions. There have been a large amount of following works on two types of behaviours of extreme eigenvalues for more general sample covariance models. We focus on those extreme eigenvalues that behave like outliers, also referred to as spiked eigenvalues in the random matrix theory literature.

The spiked covariance model was first named by [Johnstone \(2001\)](#), where the population covariance matrix is expressed as a $p \times p$ diagonal matrix with the form

$$\Sigma = \text{diag}(\alpha_1, \dots, \alpha_K, 1, \dots, 1), \quad (1.1)$$

where $\alpha_1 \geq \dots \geq \alpha_K > 1$ are referred to as spikes and $K < \infty$ is the number of spikes. [Baik and Silverstein \(2006\)](#) investigated the almost sure limits of the extreme eigenvalues which depend on the critical value $1 + \sqrt{\gamma}$ when $p/n \rightarrow \gamma > 0$. [Paul \(2007\)](#) established the central limit theorem (CLT) for the spiked sample eigenvalues under the Gaussian assumption on the data. [Bai et al. \(2008\)](#) extended Paul's results by removing the Gaussian assumption, but assumed a block diagonal structure on the population covariance matrix. [Bai and Yao \(2012\)](#) further generalized the spiked model by considering an arbitrary non-spiked part of the population covariance matrix instead of identity (but still a block diagonal structure). They classified the spikes into distant spikes and close spikes, and discussed the almost sure limits for two types of spikes and established a central limit theorem for distant spikes.

The study of the CLT for the largest eigenvalues of spiked sample covariance matrices mentioned above relies heavily on the block diagonal structure which provides a simple representation of the largest eigenvalue statistics. The asymptotic normality of spiked eigenvalues is expected to be true for a more general population covariance matrix model. [Jiang and Bai \(2018\)](#) obtained the CLT for the spiked eigenvalues without assuming a block diagonal structure on the population covariance matrix but the finite 4th moment condition on the entries of the data matrix is relaxed to a tail probability decay. However they had to pay a price that the largest entries of the population eigenvectors corresponding to the spikes tend to zero or the fourth moment of the underlying random variables must match with that of Gaussian distribution due to their moment matching method. In addition to the above literature about the bounded spikes, we would also like to mention

that there is some literature about the unbounded spikes and one may see [Cai et al. \(2017\)](#) and the reference therein.

Besides considering the spiked eigenvalues for sample covariance matrices, we also pay attention to the joint distribution of spiked eigenvalues and linear spectral statistics (LSS) which has not been well understood.

The study of LSS of sample covariance matrices is another important topic in statistics and random matrix theory. The most influential work is [Bai and Silverstein \(2004\)](#). They showed that the LSS of sample covariance matrices converge to the normal distribution under some moment restrictions. Further refinements were carried out under different relaxed settings. [Pan and Zhou \(2008\)](#) improved Theorem 1.1 in [Bai and Silverstein \(2004\)](#) by removing the constraint on the fourth moment of the underlying random variables. [Pan \(2014\)](#) showed the CLT of LSS for non-centered sample covariance matrices and discussed the difference between the centered and non-centered sample covariance matrices. [Zheng et al. \(2015\)](#) provided similar results for centralized and noncentralized sample covariance matrices in a unified framework. Furthermore [Najim and Yao \(2016\)](#) also provided CLT in terms of vanishing Levy-Prohorov distance between the LSS distribution and a Gaussian probability distribution.

There are few works considering the joint distribution of largest eigenvalue and LSS for random matrix ensembles. [Baik et al. \(2018\)](#) obtained the joint normal distribution of the largest eigenvalue and LSS for a spiked Wigner matrix. They showed that the asymptotic joint distribution of the largest eigenvalue and LSS converge to a bivariate normal distribution with the covariance dependent on the third moment of entries of the Wigner matrix. In this case the spiked eigenvalues and LSS are asymptotically independent if the third moment is zero. Recently, [Li et al. \(2020\)](#) established that the extreme eigenvalues and the trace of sample covariance matrices are jointly asymptotically normal and independent for a block diagonal population covariance matrix.

The second type of matrices we consider are the signal-plus-noise matrices. Specifically speaking, we consider a signal-plus-noise matrix with the form of

$$\mathbf{S} = \mathbf{R} + \mathbf{X}, \tag{1.2}$$

where $\mathbf{R}$ is the signal matrix and $\mathbf{X}$ is the noise matrix. This model is popular in many fields such as machine learning ([Yang et al., 2016](#)), matrix denoising ([Nadakuditi, 2014](#)) or signal processing ([Vallet et al., 2012](#)). When the signal

matrix $\mathbf{R}$ is assumed to be finite rank, there has been a huge amount of work on eigenvalues and eigenvectors for such signal-plus-noise type matrices. To name a few, [Loubaton et al. \(2011\)](#) derived the almost localization of eigenvalues, [Ding \(2020\)](#) obtained the convergent limits and rates of the leading eigenvalues and eigenvectors, [Bao et al. \(2018\)](#) showed the distributions of the principal singular vectors and singular subspaces. More works can be found in literature ([Hachem et al., 2007](#); [Benaych-Georges and Nadakuditi, 2011](#); [Vallet et al., 2012](#); [Knowles and Yin, 2013](#); [Yang et al., 2016](#); [Cape et al., 2019](#)).

However, it may be the case that the signals are not low rank matrices in practice, such as the direction of arrival (DOA) estimation ([Mani and Bose, 2010](#)), or the models appeared in signal detection for hyperspectral images ([Chang and Du, 1999](#)) or massive MIMO system ([Vinogradova et al., 2016](#)). A signal-plus-noise model without a finite rank restriction on $\mathbf{R}$ has been investigated in several papers. For the case when $\mathbf{X}$ in (1.2) contains *i.i.d.* entries, [Dozier and Silverstein \(2007b\)](#) studied the limiting spectral distribution(LSD) of $\mathbf{S}\mathbf{S}^*$ and [Dozier and Silverstein \(2007a\)](#) provided analytic properties of the stieltjes transform of LSD. The exact separation property for Gaussian noise was characterized in [Loubaton et al. \(2011\)](#). [Banna et al. \(2020\)](#) established a central limit theorem for the linear spectral statistics of $\mathbf{S}\mathbf{S}^*$. For the case when $\mathbf{X}$ has a complicated variance structure, the research works include [Hachem et al. \(2007\)](#) where the determinant equivalent of the resolvent of $\mathbf{S}\mathbf{S}^*$ was derived, and [Hachem et al. \(2013\)](#) where the limit of random bilinear forms of the resolvent of $\mathbf{S}\mathbf{S}^*$ was obtained. Most of these works focus on the full spectrum which is in the sense of global regime in random matrix theory. However, the study on the largest eigenvalues which belongs to the scope of the local regime is rare.

The Tracy-Widom distribution has been widely established for the largest eigenvalues of different types of random matrices. [Tracy and Widom \(1994, 1996\)](#) studied the most basic random matrix models known as Gaussian Orthogonal Ensemble (GOE) and Gaussian Unitary Ensemble (GUE) that are self-adjoint matrices contains independent standard real and complex Gaussian entries, and derived the distribution of the extreme eigenvalues, named it Tracy-Widom distribution, for GOE and GUE. [Johansson \(2000\)](#) and [Johnstone \(2001\)](#) established the Tracy-Widom distribution for the standard Wishart matrices with real and complex elements. These works utilize the joint distribution of eigenvalues of studying model

and this priori knowledge relies on Gaussian assumptions. In the last decade, significant development has been made on universality property of random matrices theory. Universality property usually refers to the phenomenon that the eigenvalues converge to the same distribution regardless of exact distributions of entries. When one cares about the extreme eigenvalues, i.e., the edge of spectrum, we refer to such a property as the edge universality property. [Tao and Vu \(2010\)](#) showed the edge universality for two Wigner matrices with first four moments matching. [Erdős et al. \(2012\)](#) considered generalized Wigner matrices and showed the Tracy-Widom law holds with first two moments matching with the GOE. A necessary and sufficient condition to guarantee the Tracy-Widom distribution is provided in [Lee and Yin \(2014\)](#) for Wigner matrices and [Ding and Yang \(2018\)](#) for sample covariance matrices. Extensive researches ([Lee and Schnelli, 2015](#); [Han et al., 2016](#); [Fan and Johnstone, 2017](#); [Bao et al., 2019](#)) have signified that the limiting distributions of extreme eigenvalues of a large group of random matrices are Tracy-Widom distributions. In addition, the Tracy-Widom distribution has been used frequently in hypothesis testing, such as testing the number of factors ([Onatski, 2009](#)), testing the covariance structure of covariance ([Han et al., 2016](#)), testing linear independence between high dimensional vectors ([Bao et al., 2019](#)).

Our considered model $\mathbf{SS}^*$ is a deformed rectangular model from a random matrix point of view. There are some related results on the largest eigenvalues of deformed Hermitian models. [Lee and Schnelli \(2015\)](#) established the Tracy-Widom distributions of extreme eigenvalues of deformed Wigner matrices. [Landon and Yau \(2017\)](#) considered edge statistics of Dyson Brownian motion with deterministic initial data after short time scale. This model can be regarded as a deformed GOE where the GOE is in smaller scale compared with the deterministic initial data. It was proved that the edge statistics have the Tracy-Widom distributions. Very recently, [Ding and Yang \(2020a\)](#) considered the deformed rectangular model $\mathbf{SS}^*$ where the variance of $\mathbf{X}$ tends to 0. Such model can be regarded as the rectangular version of that considered in [Landon and Yau \(2017\)](#). However, to the best of the author's knowledge, there is no result on the Tracy-Widom distribution of $\mathbf{SS}^*$ when the $\mathbf{X}$ is in the same scale as $\mathbf{R}$.

In view of the literature mentioned above, the main objects of this thesis are to develop the asymptotic normality of spiked eigenvalues for a general sample covariance matrix model, and develop the asymptotic Tracy-Widom law for the largest eigenvalue of a signal-plus-noise model. For the sample covariance matrix

model, we also consider the asymptotic joint distribution of the spiked eigenvalues and LSS.

The main content of the thesis is organized as follows.

- In Chapter 2, we show the asymptotic normality of spiked eigenvalues of sample covariance matrices. We also prove that the spiked eigenvalues are asymptotically independent of linear spectral statistics. We propose a statistic combining the largest eigenvalue and linear spectral statistics to test equality of two population covariance matrices. Simulation results demonstrate satisfied performance of the proposed statistic. More details of this chapter can be found in the submitted paper (Zhang et al., 2020).
- In Chapter 3, we consider the model $\mathbf{SS}^*$ where $\mathbf{S}$ is defined by (1.2). Under some regularity condition on $\mathbf{R}$, we show that the largest eigenvalue have Tracy-Widom distribution under a tail condition on the entries of $\mathbf{X}$. Moreover, the tail condition is proved to be necessary and sufficient to assure the Tracy-Widom law.
- In Chapter 4, we conclude the thesis and have a discussion about future research directions.

Throughout the thesis, we use C to denote generic large constant and its value may change from line to line. We use $O(\cdot)$ and $o(\cdot)$ for the standard big O and little o notation, and occasionally, we use $\ll$ to stand for $o(\cdot)$, i.e. $a \ll b$ represents $a = o(b)$. We say $a \sim b$ if there exists C , such that $C^{-1}b \leq a \leq Cb$. For any function f we use $f^{(n)}$ to denote the n -th derivative of f . We use $I(\mathcal{A})$ to denote an indicator function of an event $\mathcal{A}$. The intersection of events $\mathcal{A}$ and $\mathcal{B}$ is denoted by $\mathcal{A} \cap \mathcal{B}$, or abbreviated by $\mathcal{AB}$. The spectral norm of a matrix M is denoted by $\|M\|$.

Chapter 2

Asymptotic joint distribution of spiked eigenvalues and LSS of sample covariance matrices

2.1 Introduction

In this chapter, we consider the spiked covariance matrix models which appear naturally in the area of wireless communication, speech recognition, genetics and finance, etc. Specifically speaking, consider a population covariance matrix

$$\Sigma = \mathbf{V} \begin{pmatrix} \Lambda_S & 0 \\ 0 & \Lambda_P \end{pmatrix} \mathbf{V}^\top \quad (2.1.1)$$

where $\mathbf{V}$ is an orthogonal matrix, Λ_S is a diagonal matrix consisting of the bounded and descending spiked eigenvalues, and Λ_P is the diagonal matrix of non-spiked eigenvalues.

Our main contributions are summarized as follows. For the first time we establish CLT for the leading spiked eigenvalues of the sample covariance matrices with the general spiked covariance matrices Σ in (2.1.1). We need neither the block diagonal structure unlike [Bai et al. \(2008\)](#); [Bai and Yao \(2012\)](#) nor the maximum absolute value of the eigenvector of the corresponding spikes tending to zero nor requiring the match of the 4th moment with the standard Gaussian distribution (i.e., the 4-th moment is 3) unlike [Jiang and Bai \(2018\)](#). We also show that the extreme eigenvalues and LSS of large sample covariance matrices are asymptotically independent. A direct consequence of above results are the asymptotic joint

distributions of the spiked eigenvalues and LSS. Moreover consistent estimators of the L_4 norm of population eigenvectors associated with the leading sample spikes are proposed.

The remaining sections are organized as follows. Section 2.2 presents the main results about the asymptotic distribution of the largest sample spikes, the asymptotic independence between the largest sample spikes and the linear spectral statistics and the estimator of the population eigenvectors corresponding the largest spikes. We also explore an application of our main results in the two sample hypothesis testing about covariances in Section 2.2. The simulation is reported in Section 2.3. The proofs are given from Section 2.4 to Section 2.8.

2.2 The main results

Consider the data matrix $\mathbf{\Gamma}\mathbf{X}$, where $\mathbf{\Gamma}$ is a $p \times p$ deterministic matrix with $\mathbf{\Gamma}\mathbf{\Gamma}^\top = \mathbf{\Sigma}$ and $\mathbf{X} = (x_{ij})$ is a $p \times n$ random matrix with entries $x_{ij} = n^{-1/2}q_{ij}$ where q_{ij} are independent random variables satisfying Assumption 1 below. The sample covariance matrix has the form

$$\mathbf{S}_n = \mathbf{\Gamma}\mathbf{X}\mathbf{X}^\top\mathbf{\Gamma}^\top.$$

Order the eigenvalues of $\mathbf{S}_n$ as $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Denote the singular value decomposition of $\mathbf{\Gamma}$ by

$$\mathbf{\Gamma} = \mathbf{V} \begin{pmatrix} \mathbf{\Lambda}_S^{1/2} & 0 \\ 0 & \mathbf{\Lambda}_P^{1/2} \end{pmatrix} \mathbf{U}^\top \quad (2.2.1)$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices, $\mathbf{\Lambda}_S$ is a diagonal matrix consisting of the spiked eigenvalues in descending order and $\mathbf{\Lambda}_P$ is the diagonal matrix of the non-spiked eigenvalues. To be more specific, denote the eigenvalues of the spiked part $\mathbf{\Lambda}_S$ as $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_K$, and eigenvalues of the non-spiked part as $\alpha_{K+1} \geq \alpha_{K+2} \dots \geq \alpha_p$. Partition $\mathbf{U}$ as $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2)$, where $\mathbf{U}_1$ is a $p \times K$ submatrix of $\mathbf{U}$. Let $\mathbf{u}_i = (u_{i1}, \dots, u_{ip})^\top$ be the i -th column of $\mathbf{U}_1$. Define

$$\mathbf{\Sigma}_{1P} = \mathbf{U}_2\mathbf{\Lambda}_P\mathbf{U}_2^\top. \quad (2.2.2)$$

2.2.1 Limiting laws for spiked eigenvalues

We first specify the assumptions for establishing CLTs of the leading sample spiked eigenvalues.

Assumption 1. The double array $\{q_{ij} : i = 1, \dots, p, j = 1, \dots, n\}$ consists of independent and identically distributed random variables, with $\mathbb{E}q_{11} = 0, \mathbb{E}|q_{11}|^2 = 1$ and $\mathbb{E}|q_{11}|^4 = \gamma_4$.

Assumption 2. $p/n = c_n \rightarrow c > 0$ as $n \rightarrow \infty$.

Assumption 3. The $p \times p$ matrix $\Sigma = \Gamma\Gamma^\top$ has a bounded spectral norm. Furthermore, denote the empirical spectral distribution (ESD) of Σ by H_n , which tends to a nonrandom limiting distribution H as $p \rightarrow \infty$.

For the next assumption, we denote by Γ_μ the support for any measure μ on $\mathbb{R}$. For $\alpha \notin \Gamma_H$ and $\alpha \neq 0$, define

$$\psi(\alpha) := \alpha + c\alpha \int \frac{t}{\alpha - t} dH(t). \quad (2.2.3)$$

Its derivative is

$$\psi'(\alpha) = 1 - c \int \frac{t^2}{(\alpha - t)^2} dH(t). \quad (2.2.4)$$

Define $\psi_n(\alpha)$ from (2.2.3) with H, c replaced by H_n, c_n .

Assumption 4. Suppose that the population covariance matrix Σ has K spiked eigenvalues: $\alpha_1 > \dots > \alpha_K$, lying outside the support of H , and satisfying $\psi'(\alpha_k) > 0$ for $1 \leq k \leq K$.

Bai and Yao (2012) provided a complete characterization of sample spikes according to the sign of $\psi'(\alpha)$. If $\psi'(\alpha) > 0$ then the corresponding sample spiked eigenvalues have limits outside the support of $F^{c,H}$, the limit of the empirical spectral distribution of $\mathbf{S}_n$. They called them distant spikes in this case. Here we need to clarify that although Bai and Yao (2012) assumed that the population covariance matrices are block diagonal but this assumption is not essential and can be removed. This is because their method of deriving almost sure convergence relies on their Propositions 3.1 and 3.2 and these two results regarding the exact separation were first appeared in Bai et al. (1998); Bai and Silverstein (1999) without a block diagonal structure of the population covariance matrices.

Assumption 5. Assume that for $i = 1, \dots, K$ the following limits exist:

$$\sigma_i^2 := (\gamma_4 - 3) \frac{\alpha_i^2 \{\psi'(\alpha_i)\}^2}{\psi^2(\alpha_i)} \lim_{p \rightarrow \infty} \sum_{j=1}^p u_{ij}^4 + 2 \frac{\alpha_i^2 \psi'(\alpha_i)}{\psi^2(\alpha_i)} \text{ and}$$

$$\sigma_{ij} := (\gamma_4 - 3) \frac{\alpha_i \alpha_j \psi'(\alpha_i) \psi'(\alpha_j)}{\psi(\alpha_i) \psi(\alpha_j)} \lim_{p \rightarrow \infty} \sum_{k=1}^p u_{ik}^2 u_{jk}^2.$$

We will show that the sample spiked eigenvalues λ_i ($i = 1, \dots, K$) of $\mathbf{S}_n$ are associated with a random quadratic form given by the following equation (see the details given in the proof of Theorem 2.2.2):

$$\det\{\Lambda_S^{-1} - \mathbf{U}_1 \mathbf{X} (\lambda_i \mathbf{I} - \mathbf{X}^\top \Sigma_{1P} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{U}_1\} = 0. \quad (2.2.5)$$

Thus, our results rely on a new technique tool, a CLT for a type of random quadratic forms. The result in Theorem 2.2.1 is crucial to removing the block diagonal structure of the population covariance matrices (hence the proof of Theorems 2.2.2 and 2.2.3 below). It can be of independent interest.

Theorem 2.2.1. *Suppose that Assumptions 1-3 hold. Moreover, suppose that the non-random orthogonal unit vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ satisfy $\mathbf{w}_1^\top \mathbf{U}_2 = \mathbf{w}_2^\top \mathbf{U}_2 = 0$ and $\mathbf{w}_1^\top \mathbf{w}_2 = 0$, and α satisfies $\psi'(\alpha) > 0$. Then*

$$\frac{\sqrt{n}}{\tilde{\sigma}_1} \left(\mathbf{w}_1^\top \mathbf{X} \left(\mathbf{I} - \frac{1}{\psi_n(\alpha)} \mathbf{X}^\top \Sigma_{1P} \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{w}_1 - \frac{\psi_n(\alpha)}{\alpha} \right) \xrightarrow{D} N(0, 1) \quad (2.2.6)$$

and

$$\frac{\sqrt{n}}{\tilde{\sigma}_{12}} \mathbf{w}_1^\top \mathbf{X} \left(\mathbf{I} - \frac{1}{\psi_n(\alpha)} \mathbf{X}^\top \Sigma_{1P} \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{w}_2 \xrightarrow{D} N(0, 1) \quad (2.2.7)$$

where $\tilde{\sigma}_1^2 := \psi^2(\alpha) \{(\gamma_4 - 3) \sum_{i=1}^p w_{1i}^4 + 2/\psi'(\alpha)\} / \alpha^2$, $\tilde{\sigma}_{12}^2 := \{(\gamma_4 - 3) \psi^2(\alpha)\} \sum_{i=1}^p w_{1i}^2 w_{2i}^2 / \alpha^2$ and w_{ij} is the j -th element of $\mathbf{w}_i$, $i = 1, 2$.

We are ready to provide the central limit theorem for the sample spiked eigenvalues. We consider the case when the eigenvalues of Λ_S are all simple first.

Theorem 2.2.2. *Let $\theta_i = \psi_n(\alpha_i)$, $i = 1, \dots, K$. Suppose that Assumptions 1-5 hold. Then for all $i = 1, 2, \dots, K$,*

$$\sqrt{n} \frac{\lambda_i - \theta_i}{\theta_i} \xrightarrow{D} N(0, \sigma_i^2). \quad (2.2.8)$$

Moreover, for any fixed $1 \leq r \leq K$,

$$\left(\sqrt{n} \frac{\lambda_1 - \theta_1}{\theta_1}, \dots, \sqrt{n} \frac{\lambda_r - \theta_r}{\theta_r} \right) \xrightarrow{D} N(0, \Sigma_{\lambda r}), \quad (2.2.9)$$

where $\Sigma_{\lambda r} = (\Sigma_{\lambda r, ij})$ with

$$\Sigma_{\lambda r, ij} := \begin{cases} \sigma_i^2, & i = j \\ \sigma_{ij}, & i \neq j \end{cases}.$$

Remark 1. Since the convergence rate of $c_n \rightarrow c$ and $H_n \rightarrow H$ can be arbitrarily slow, $\theta_i = \psi_n(\alpha_i)$ is used in the CLT, rather than $\psi(\alpha_i)$, which is the almost sure limit of λ_i .

Remark 2. If one cares about the asymptotic distribution for an individual sample spiked eigenvalue, the existence of limit defined in Assumption 5 (the σ_i) is not necessary since $\sqrt{n}(\lambda_i - \theta_i)/\theta_i$ can be normalized by $[(\gamma_4 - 3)\alpha_i^2\{\psi'(\alpha_i)\}^2 \sum_{j=1}^p u_{ij}^4 + 2\alpha_i^2\psi'(\alpha_i)]/\psi^2(\alpha_i)$. Moreover Assumption 5 is not necessary if we restate the asymptotic convergence results using Lévy–Prokhorov metric, as (2.2.10) stated below. To be more concrete, recall Lévy–Prokhorov metric π which characterize the distance of two probability measure μ and ν defined by

$$\pi(\mu, \nu) := \inf \{ \varepsilon > 0 \mid \mu(A) \leq \nu(A^\varepsilon) + \varepsilon \text{ and } \nu(A) \leq \mu(A^\varepsilon) + \varepsilon \text{ for all } A \in \mathcal{B}(M) \}.$$

Define

$$\Lambda_r := \left(\sqrt{n} \frac{\lambda_1 - \theta_1}{\theta_1}, \dots, \sqrt{n} \frac{\lambda_r - \theta_r}{\theta_r} \right).$$

Let $\Sigma_{p, \lambda r} = ((\Sigma_{p, \lambda r})_{ij})$ be $r \times r$ matrices with

$$(\Sigma_{p, \lambda r})_{ij} := \begin{cases} \sigma_{p,i}^2, & i = j \\ \sigma_{p,ij}, & i \neq j \end{cases}.$$

where

$$\begin{aligned} \sigma_{p,i}^2 &:= (\gamma_4 - 3) \frac{\alpha_i^2 \{\psi'(\alpha_i)\}^2}{\psi^2(\alpha_i)} \sum_{j=1}^p u_{ij}^4 + 2 \frac{\alpha_i^2 \psi'(\alpha_i)}{\psi^2(\alpha_i)}, \\ \sigma_{p,ij} &:= (\gamma_4 - 3) \frac{\alpha_i \alpha_j \psi'(\alpha_i) \psi'(\alpha_j)}{\psi(\alpha_i) \psi(\alpha_j)} \sum_{k=1}^p u_{ik}^2 u_{jk}^2. \end{aligned}$$

Then

$$\lim_{n \rightarrow \infty} \pi(\Lambda_r, N(0, \Sigma_{p, \lambda r})) \rightarrow 0. \quad (2.2.10)$$

To conclude this it is equivalent to show that for any subsequence $\{\pi(\Lambda r, N(0, \Sigma_{p,\lambda r}))\}_{n \geq 1}$, there is a further subsequence that converges to 0. Since the entries of $\Sigma_{p,\lambda r}$ are bounded and $r < \infty$, without loss of generality, we can assume that the limit of $\Sigma_{p,\lambda r}$ exists. Therefore by our conclusion of Theorem 2.2.2, $\lim_n \pi(\Lambda r, N(0, \Sigma_{p,\lambda r})) \rightarrow 0$.

Remark 3. Compared with earlier asymptotic results on spiked eigenvalues of sample covariance matrices obtained by Bai et al. (2008); Bai and Yao (2012) and Li et al. (2020), we do not assume a block diagonal structure on population covariance matrices. Moreover Bai et al. (2008); Bai and Yao (2012) and Jiang and Bai (2018) did not consider the joint distribution of the different leading sample spiked eigenvalues corresponding to the different population eigenvalues. Instead they considered the joint distribution of the different leading sample spiked eigenvalues corresponding to the same population eigenvalues.

We next consider the case when the multiplicity of the spiked eigenvalues of Λ_S are more than one.

Assumption 6. Suppose that the population covariance matrix Σ has K spiked eigenvalues: $\alpha_1 > \dots > \alpha_{\mathcal{L}}$ with respective multiplicities $m_1, \dots, m_{\mathcal{L}}$, laying outside the support of H , and satisfying $\psi'(\alpha_k) > 0$ for $1 \leq k \leq \mathcal{L}$. Furthermore, we assume that the following limits exist for $i = 1, \dots, \mathcal{L}$:

$$\begin{aligned} g(r_i, k_1, l_1, k_2, l_2) := & (\gamma_4 - 3) \frac{\alpha_i^2 \{\psi'(\alpha_i)\}^2}{\psi^2(\alpha_i)} \lim_{p \rightarrow \infty} \sum_{j=1}^p u_{r_i+k_1,j} u_{r_i+l_1,j} u_{r_i+k_2,j} u_{r_i+l_2,j} \\ & + \frac{\alpha_i^2 \psi'(\alpha_i)}{\psi^2(\alpha_i)} \{(\mathbf{u}_{r_i+k_1}^\top \mathbf{u}_{r_i+k_2})(\mathbf{u}_{r_i+l_1}^\top \mathbf{u}_{r_i+l_2}) + (\mathbf{u}_{r_i+k_1}^\top \mathbf{u}_{r_i+l_2})(\mathbf{u}_{r_i+k_2}^\top \mathbf{u}_{r_i+l_1})\}, \end{aligned} \quad (2.2.11)$$

where $r_i := \sum_{j=0}^{i-1} m_j$, $m_0 = 0$ and $1 \leq k_1, l_1, k_2, l_2 \leq m_i$.

Theorem 2.2.3. Suppose that Assumptions 1, 2, 3 and 6 hold. Then

$$\left(\sqrt{n} \frac{\lambda_{r_i+1} - \theta_i}{\theta_i}, \dots, \sqrt{n} \frac{\lambda_{r_i+m_i} - \theta_i}{\theta_i} \right) \quad (2.2.12)$$

converges weakly to the joint distribution of the eigenvalues of $m_i \times m_i$ Gaussian random matrix $\mathbf{G}_i$ with $\mathbb{E} \mathbf{G}_i = 0$ and covariance of $(\mathbf{G}_i)_{k_1, l_1}$ and $(\mathbf{G}_i)_{k_2, l_2}$ being $g(r_i, k_1, l_1, k_2, l_2)$ defined in (2.2.11).

Remark 4. This result is similar to those in Bai et al. (2008); Bai and Yao (2012), Theorem 3.1 and Corollary 3.1 in Jiang and Bai (2018). However we neither need

a block diagonal population covariance structure as in [Bai et al. \(2008\)](#); [Bai and Yao \(2012\)](#) nor the maximum absolute value of the eigenvector of the corresponding spikes tending to zero (i.e., $\max_{1 \leq i \leq K, 1 \leq j \leq K} |u_{ij}| \rightarrow 0$) nor requiring the match of the 4th moment with Gaussian distribution (i.e., $\gamma_4 = 3$) as in [Jiang and Bai \(2018\)](#). The assumption [D] about the population eigenvectors in [Jiang and Bai \(2018\)](#) excludes all the diagonal population covariance matrices when $\max_{1 \leq i \leq K, 1 \leq j \leq K} |u_{ij}| \rightarrow 0$. Under their assumption [D] we have

$$g(r_i, k_1, l_1, k_2, l_2) = \begin{cases} 2\alpha_i^2 \psi'(\alpha_i) / \psi^2(\alpha_i) & k_1 = k_2 = l_1 = l_2 \\ \alpha_i^2 \psi'(\alpha_i) / \psi^2(\alpha_i) & k_1 = k_2 \text{ and } l_1 = l_2 \text{ or } k_1 = l_2 \text{ and } l_1 = k_2 \\ 0 & \text{otherwise,} \end{cases} \quad (2.2.13)$$

which is consistent with theirs.

2.2.2 Asymptotic joint distribution of sample spiked eigenvalues and linear spectral statistics

We now turn to the asymptotic joint distribution of sample spiked eigenvalues and linear spectral statistics of sample covariance matrices. Some notations are introduced first. Let $F^{\mathbf{S}_n}$ be the ESD of the sample covariance matrix $\mathbf{S}_n$. It is well known that $F^{\mathbf{S}_n}$ under some mild assumptions converges weakly to a non-random distribution $F^{c,H}$ with probability one, whose Stieltjes transform is the unique solution in $\mathbb{C}^+$ to the equation

$$m = \int \frac{1}{t(1 - c - czm) - z} dH(t) \quad (2.2.14)$$

for $z \in \mathbb{C}^+$ (see [Bai and Silverstein \(2010\)](#)). We also need the non-asymptotic version of $F^{c,H}$ whose stieltjes transform solves the above equation by replacing c and H with c_n and H_n respectively, and we denote it by F^{c_n, H_n} . Define the conjugate matrix of $\mathbf{S}_n$ by $\underline{\mathbf{S}}_n := \mathbf{X}^\top \mathbf{\Gamma}^\top \mathbf{\Gamma} \mathbf{X}$ that shares the same non-zero eigenvalues of $\mathbf{S}_n$. The ESD of $\underline{\mathbf{S}}_n$ has an almost sure limit whose Stieltjes transform $\underline{m} \equiv \underline{m}(z) \in \mathbb{C}^+$ is the unique solution for any $z \in \mathbb{C}^+$ to the equation

$$z = -\frac{1}{\underline{m}} + c \int \frac{t}{1 + t\underline{m}} dH(t). \quad (2.2.15)$$

Now we introduce the linear spectral statistics of sample covariance matrices defined by

$$L_p(\varphi) := \sum_{i=1}^p \varphi(\lambda_i) - p \int \varphi(x) dF^{c_n, H_n}(x), \quad (2.2.16)$$

where $\varphi(x)$ is an analytic function on an open interval containing

$$[\liminf_n \lambda_{\min}^{\Sigma} I_{(0,1)}(c)(1 - \sqrt{c})^2, \limsup_n \lambda_{\max}^{\Sigma}(1 + \sqrt{c})^2].$$

Assumption 7. Suppose that

$$\frac{1}{p} \sum_{i=1}^p \mathbf{e}_i^{\top} \mathbf{\Gamma}^{\top} (\underline{m}(z_1) \mathbf{\Gamma} \mathbf{\Gamma}^{\top} + \mathbf{I})^{-1} \mathbf{\Gamma} \mathbf{e}_i \mathbf{e}_i^{\top} \mathbf{\Gamma}^{\top} (\underline{m}(z_2) \mathbf{\Gamma} \mathbf{\Gamma}^{\top} + \mathbf{I})^{-1} \mathbf{\Gamma} \mathbf{e}_i \rightarrow h_1(z_1, z_2) \quad (2.2.17)$$

and

$$\frac{1}{p} \sum_{i=1}^p \mathbf{e}_i^{\top} \mathbf{\Gamma}^{\top} (\underline{m}(z) \mathbf{\Gamma} \mathbf{\Gamma}^{\top} + \mathbf{I})^{-2} \mathbf{\Gamma} \mathbf{e}_i \mathbf{e}_i^{\top} \mathbf{\Gamma}^{\top} (\underline{m}(z) \mathbf{\Gamma} \mathbf{\Gamma}^{\top} + \mathbf{I})^{-1} \mathbf{\Gamma} \mathbf{e}_i \rightarrow h_2(z). \quad (2.2.18)$$

Remark 5. This assumption has been used to establish the CLT for LSS, see Theorem 1.4. in [Pan and Zhou \(2008\)](#). Remark 1.4 therein also tells that such assumption can be removed under our assumptions 1-3 together with either $\gamma_4 = 3$ or $\mathbf{\Gamma}^{\top} \mathbf{\Gamma}$ is diagonal. Actually, if $\gamma_4 = 3$, the term in Assumption 7 will not appear in the variance (covariance). And if $\mathbf{\Gamma}^{\top} \mathbf{\Gamma}$ is diagonal,

$$h_2(z) = \int \frac{t^2 dH(t)}{(\underline{m}(z)t + 1)^3},$$

$$h_1(z_1, z_2) = \int \frac{t^2 dH(t)}{(m(z_1)t + 1)(m(z_2)t + 1)}.$$

Theorem 2.2.4. Suppose that Assumptions 1-5 and 7 hold, then for $1 \leq i \leq K$, $\sqrt{n}(\lambda_i - \theta_i)/\theta_i$ and $L_p(\varphi)$ are asymptotically independent.

Remark 6. This theorem implies that the joint distribution of $\sqrt{n}(\lambda_i - \theta_i)/\theta_i$ and $L_p(\varphi)$ is bivariate normal with asymptotic independent marginal distribution. The result can be generalized to the joint distribution of the multiple spikes and LSS easily with the spiked eigenvalues part and the LSS part still being independent.

For the results regarding to marginal distribution of LSS, one can refer to [Bai and Silverstein \(2004\)](#) and [Pan and Zhou \(2008\)](#).

Remark 7. Compared with Theorem 3.1 in [Li et al. \(2020\)](#), we have two advantages. Firstly, we don't need the block diagonal assumption on the population covariance matrices. Secondly, our LSS is not restricted to the trace of sample covariance matrices.

2.2.3 Estimating the population eigenvectors associated with the spiked eigenvalues

This section is to explore the estimation of the L_4 norm for population spiked eigenvectors associated with the simple spiked eigenvalues $\alpha_1, \dots, \alpha_K$ involved in (2.2.8). The estimators will be used in the next section in our proposed test statistic. Although many studies of the spiked eigenvectors have been carried out, most of them have not provided consistent estimators for the population eigenvectors in terms of certain norm. For example, [Paul \(2007\)](#) established the almost sure limit of $\mathbf{u}_i^\top \hat{\mathbf{u}}_i$ and a CLT for $\hat{\mathbf{u}}_i$ for any $1 \leq i \leq K$ under the assumption that $\mathbf{X}$ is Gaussian and $\mathbf{\Gamma}$ is diagonal with the nonspiked covariance being identity. [Ding \(2017\)](#) further characterized the limit of $\mathbf{u}_i^\top \hat{\mathbf{u}}_i$ for a general spiked model. However these results are not helpful for estimating the population eigenvectors in terms of certain norm. Our following theorem provides a consistent estimator of $\sum_{k=1}^p u_{ik}^4$ inspired by the results in [Mestre \(2008\)](#), which considered an estimator of $\mathbf{s}^\top \mathbf{u}_i$ where $\mathbf{s}$ is any fixed vector with a bounded norm in R^p when the underlying random variables are continuous with finite eighth order moments.

Theorem 2.2.5. *Suppose that the assumptions of Theorem 2.2.2 hold and $\mathbf{\Gamma}$ is symmetric, i.e. the left orthogonal matrix $\mathbf{V}$ in (2.2.1) equals $\mathbf{U}$. Let $\hat{\mathbf{u}}_i$ be eigenvectors of $\mathbf{S}_n$ associated with eigenvalue λ_i and $\hat{u}_{ik}$ be the k -th coordinate of $\hat{\mathbf{u}}_i$. For $1 \leq i \leq K$, $\sum_{k=1}^p u_{ik}^4$ is consistently estimated by $\sum_{j=1}^p \{\sum_{k=1}^p \theta_i(k) \hat{u}_{kj}^2\}^2$, where*

$$\begin{aligned} \theta_i(k) &:= \begin{cases} -\phi_i(k), & k \neq i \\ 1 + \varrho_i(k), & k = i \end{cases}, \\ \phi_i(k) &:= \frac{\lambda_i}{\lambda_k - \lambda_i} - \frac{\nu_i}{\lambda_k - \nu_i}, \\ \varrho_i(k) &:= \sum_{j \neq i}^p \left(\frac{\lambda_j}{\lambda_k - \lambda_j} - \frac{\nu_j}{\lambda_k - \nu_j} \right), \end{aligned} \tag{2.2.19}$$

and where $\nu_1 \geq \nu_2 \geq \dots \geq \nu_p$ are the real valued solutions to the equation in x :

$$\frac{1}{p} \sum_{i=1}^p \frac{\lambda_i}{\lambda_i - x} = \frac{1}{c}. \quad (2.2.20)$$

When $c > 1$, take $\nu_n = \dots = \nu_p = 0$. In the expressions of $\phi_i(k)$ and $\varrho_i(k)$, we use the convention that any term of form $\frac{0}{0}$ is 0.

Remark 8. Table 5 below shows that such an estimator of $\sum_{k=1}^p u_{ik}^4$ is quite accurate.

The L_4 norm of eigenvector is also related to the inverse participation ratio (IPR) (Plerou et al. (2002); Pastor-Satorras and Castellano (2016)) for an unit vector $\mathbf{u} = (u_1, \dots, u_p)^\top$, denoted by

$$\text{IPR}(\mathbf{u}) := \sum_{i=1}^p u_i^4. \quad (2.2.21)$$

The IPR quantifies the reciprocal of number of components that participant significantly to the unit vector. Consider two extreme cases for illustration. If only one component of $\mathbf{u}$ is 1 and others are 0, then IPR equals 1. While if all the components have identical contribution that equal $1/\sqrt{p}$, then IPR equals $1/p$. A large IPR indicates that the vector is sparse or localized. The IPR has been used in Plerou et al. (2002) to quantify how many stocks contribute to the principle eigenvector, and in Pastor-Satorras and Castellano (2016) to measure the localization of networks. Estimating the L_4 norm of spiked eigenvector may lead to more theoretical understanding about related problems.

2.2.4 Testing the equality of two spiked covariance matrices

This subsection is to explore an application of our results. Consider the problem of testing the equality of two spiked covariance matrices Σ_1 and Σ_2 . Let $\{\mathbf{y}_{1i} = \Sigma_1^{1/2} \mathbf{q}_{1i}, i = 1, \dots, n_1\}$ be i.i.d p variate random samples from the population F_1 with mean zero and covariance matrix Σ_1 , and $\{\mathbf{y}_{2i} = \Sigma_2^{1/2} \mathbf{q}_{2i}, i = 1, \dots, n_2\}$ be i.i.d. p variate random samples from the population F_2 with mean zero and covariance matrix Σ_2 . Suppose F_1 and F_2 are independent. Several tests on the hypothesis:

$$H_0 : \Sigma_1 = \Sigma_2 \quad \text{versus} \quad H_1 : \Sigma_1 \neq \Sigma_2 \quad (2.2.22)$$

have been proposed under high-dimensional settings. To name a few, [Li and Chen \(2012\)](#) suggested a test based on an unbiased estimator for $\text{tr}[(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2)^2]$. The test in [Cai et al. \(2013\)](#) is motivated by studying the maximum of standardized differences between entries of two sample covariance matrices to deal with sparse alternatives. [Yang and Pan \(2017\)](#) proposed a weighted statistic that is powerful for dense or sparse alternatives.

Let $\mathbf{Y}_1 = (\mathbf{y}_{11}, \dots, \mathbf{y}_{1n_1})$ and $\mathbf{Y}_2 = (\mathbf{y}_{21}, \dots, \mathbf{y}_{2n_2})$. Denote $\mathbf{x}_{1i} = n_1^{-1/2} \mathbf{y}_{1i}$, $i = 1, \dots, n_1$ and $\mathbf{x}_{2i} = n_2^{-1/2} \mathbf{y}_{2i}$, $i = 1, \dots, n_2$. Let $\mathbf{X}_1 = (\mathbf{x}_{11}, \dots, \mathbf{x}_{1n_1})$ and $\mathbf{X}_2 = (\mathbf{x}_{21}, \dots, \mathbf{x}_{2n_2})$. Denote two sample covariance matrices by

$$\mathbf{S}_1 = \frac{1}{n_1} \mathbf{Y}_1 \mathbf{Y}_1^\top = \mathbf{\Sigma}_1^{\frac{1}{2}} \mathbf{X}_1 \mathbf{X}_1^\top \mathbf{\Sigma}_1^{\frac{1}{2}} \quad \text{and} \quad \mathbf{S}_2 = \frac{1}{n_2} \mathbf{Y}_2 \mathbf{Y}_2^\top = \mathbf{\Sigma}_2^{\frac{1}{2}} \mathbf{X}_2 \mathbf{X}_2^\top \mathbf{\Sigma}_2^{\frac{1}{2}}. \quad (2.2.23)$$

We also assume that the respective largest spike eigenvalues of $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are simple for simplicity. Denote the largest eigenvalues of $\mathbf{S}_1$ and $\mathbf{S}_2$ as $\lambda_1(\mathbf{S}_1)$ and $\lambda_1(\mathbf{S}_2)$ respectively. Denote the largest spiked eigenvalues of $\mathbf{\Sigma}_k$ by α_{k1} , $k = 1, 2$, and the corresponding eigenvector by $\mathbf{u}_{1,k} = (u_{11,k}, \dots, u_{1p,k})^\top$, $k = 1, 2$. Let γ_{4k} , $k = 1, 2$ be the fourth moment of $\{q_{1ij}, j = 1, \dots, p, i = 1, \dots, n_1\}$ and $\{q_{2ij}, j = 1, \dots, p, i = 1, \dots, n_2\}$ respectively. A natural test statistic for (2.2.22) by using the largest eigenvalues and the linear spectral statistics is

$$\left\{ \sqrt{n} \frac{\lambda_1(\mathbf{S}_1) - \lambda_1(\mathbf{S}_2)}{\sigma_{spi}} \right\}^2 + \left\{ \frac{\text{tr}(\mathbf{S}_1) + \text{tr}(\mathbf{S}_1^2) - \text{tr}(\mathbf{S}_2) - \text{tr}(\mathbf{S}_2^2)}{\sigma_{lin}} \right\}^2 \quad (2.2.24)$$

where

$$\sigma_{spi}^2 = \sigma_{spi1}^2 + \sigma_{spi2}^2, \quad \sigma_{lin}^2 = \sigma_{lin1}^2 + \sigma_{lin2}^2, \\ \sigma_{spi1}^2 := (\gamma_{4k} - 3) \alpha_{k1}^2 (\psi'(\alpha_{k1}))^2 \sum_{j=1}^p u_{1j,k}^4 + 2 \alpha_{k1}^2 \psi'(\alpha_{k1}), \quad k = 1, 2 \quad (2.2.25)$$

and

$$\sigma_{lin1}^2 := 8c_k r_{k4} + 16c_k^2 r_{k3} r_{k1} + 8c_k r_{k3} + 8c_k^3 r_{k2} (r_{k1})^2 + 8c_k^2 r_{k2} r_{k1} + 4c_k^2 (r_{k2})^2 + 2c_k r_{k2} \\ + (\gamma_{4k} - 3) [4c_k r_{k4} + 8c_k^2 r_{k3} r_{k1} + 4c_k r_{k3} + 4c_k^3 r_{k2} (r_{k1})^2 + 4c_k^2 r_{k2} r_{k1} + c_k r_{k2}], \quad k = 1, 2 \quad (2.2.26)$$

with $r_{km} := \text{tr}(\mathbf{\Sigma}_k^m)/p$ and $c_k := p/n_k$. The expression (2.2.26) is obtained by calculating the contour integral in (1.20) in [Pan and Zhou \(2008\)](#). This statistic is modified further below.

The statistic in (2.2.24) is asymptotic χ_2^2 under the null hypothesis by Theorem 2.2.4. We next develop the estimators of unknown parameters α_{1k} , $\psi'(\alpha_{1k})$, $\sum_{j=1}^p u_{1j,k}^4$,

γ_{4k} and r_{km} for practical implementation. For notational simplicity, the population index k is omitted and we aim to find estimators of $\alpha_1, \psi'(\alpha_1), \sum_{j=1}^p u_{1j}^4, r_m = \text{tr} \Sigma_1^m / p$ and γ_4 associated with the population F_1 . The similar estimators are applicable to F_2 as well. In the following, we use n to denote the sample size and recall $c_n = p/n$.

The estimator of $\sum_{j=1}^p u_{1j}^4$ is given in Theorem 2.2.5. For the estimation of α_1 , we use the result in Bai and Ding (2012). Note that

$$\underline{m}_n^*(z) := -\frac{1-c_n}{z} + \frac{1}{n} \sum_{j \geq 2} \frac{1}{\lambda_j - z} \xrightarrow{a.s.} \underline{m}(z),$$

and

$$-\frac{1}{\underline{m}_n^*(\lambda_1)} \xrightarrow{a.s.} \alpha_1.$$

Therefore, as proposed by Bai and Ding (2012), α_1 is estimated by

$$\left(\frac{1-c_n}{\lambda_1} + \frac{1}{n} \sum_{j \geq 2} \frac{1}{\lambda_1 - \lambda_j} \right)^{-1}. \quad (2.2.27)$$

Consider an estimator of $\psi'(\alpha_1)$ now. Since $\psi(\cdot)$ is the inverse of the function $\alpha : x \mapsto -1/\underline{m}(x)$, we obtain

$$\psi'(\alpha_1) = \frac{1}{\alpha_1^2 \underline{m}'\{\psi(\alpha_1)\}}. \quad (2.2.28)$$

Thus we can estimate $\underline{m}'\{\psi(\alpha_1)\}$ by taking $z = \lambda_1$ in the expression of $d\underline{m}_n^*(z)/dz$, which is

$$\frac{1-c_n}{\lambda_1^2} + \frac{1}{n} \sum_{j \geq 2} \frac{1}{(\lambda_j - \lambda_1)^2}. \quad (2.2.29)$$

An estimator of $\psi'(\alpha_1)$ follows by replacing α_1 with (2.2.27) and $\underline{m}'\{\psi(\alpha)\}$ with (2.2.29) in (2.2.28).

Let $s_m = \text{tr}(\mathbf{S}_1^m)/p$. According to Lemma 2.16 in Yao et al. (2015) and Theorem 1.4 in Pan and Zhou (2008), we have the following consistent estimators A_m for $r_m, m = 1, 2, 3, 4$,

$$\begin{aligned} A_1 &= s_1, \quad A_2 = s_2 - c_n(A_1)^2, \quad A_3 = s_3 - 3c_n A_1 A_2 - c_n^2(A_1)^3, \\ A_4 &= s_4 - 2c_n(A_2)^2 - 4c_n A_1 A_3 - 6c_n^2(A_1)^2 A_2 - c_n^3(A_1)^4. \end{aligned} \quad (2.2.30)$$

To estimate γ_4 , notice that

$$\mathfrak{M} := \frac{1}{p} \mathbb{E}(\mathbf{y}_{11}^\top \mathbf{y}_{11} - \text{tr} \mathbf{\Sigma}_1)^2 = \frac{\gamma_4 - 3}{p} \sum_{i=1}^p (\mathbf{\Sigma}_{1ii})^2 + 2r_2, \quad (2.2.31)$$

where $\mathbf{\Sigma}_{1ii}$ and $\mathbf{S}_{1ii}$ are, respectively, the i -th diagonal entry of $\mathbf{\Sigma}_1$ and $\mathbf{S}_1$. Since r_2 can be estimated by A_2 above, we just need to find estimators of $\mathfrak{M}$ and $\sum_{i=1}^p (\mathbf{\Sigma}_{1ii})^2/p$. The following Lemma specifies their consistent estimators.

Lemma 2.2.6. *Under Assumptions 1 and 2, and assuming that $\mathbf{\Sigma}_1$ has bounded spectral norm, we have*

$$\frac{1}{p} \sum_{i=1}^p \mathbf{S}_{1ii}^2 - \frac{1}{p} \sum_{i=1}^p (\mathbf{\Sigma}_{1ii})^2 \xrightarrow{i.p.} 0, \quad (2.2.32)$$

and

$$\frac{1}{pn} \sum_{i=1}^n (\mathbf{y}_{1i}^\top \mathbf{y}_{1i} - \text{tr} \mathbf{S}_1)^2 - \mathfrak{M} \xrightarrow{i.p.} 0, \quad (2.2.33)$$

where $\mathbf{y}_{1i}$ denotes the i -th observation from the first population.

We assume that $\sum_{i=1}^p (\mathbf{\Sigma}_{1ii})^2/p$ does not converge to 0, which is a mild assumption for a population covariance matrix (otherwise the variances of the majority of the underlying random variables tend to zero). From (2.2.31) and Lemma 2.2.6, we propose a consistent estimator for γ_4 as follows

$$\hat{\gamma}_4 = \frac{n^{-1} \sum_{i=1}^n (\mathbf{y}_{1i}^\top \mathbf{y}_{1i})^2 - (1 - 2/n)(\text{tr} \mathbf{S}_1)^2 - 2\text{tr} \mathbf{S}_1^2}{\sum_{i=1}^p \mathbf{S}_{1ii}^2} + 3. \quad (2.2.34)$$

Up to now we can get consistent estimators of σ_{spi} and σ_{lin} in (2.2.24) by plugging in the asymptotic estimators of $\sum_{j=1}^p u_{1j,k}^4$, α_{k1} , $\psi(\alpha_{k1})$, γ_{4k} , r_{k1} , r_{k2} , r_{k3} , r_{k4} ($k = 1, 2$) into them, and we denote them by $\widehat{\sigma_{spi}}$ and $\widehat{\sigma_{lin}}$ respectively.

Through our simulations, we find that the largest sample spiked eigenvalue and the full linear spectral statistics have large correlations although they are asymptotic uncorrelated in theory. This is due to the fact that $\text{cov}(\sqrt{n}\lambda_1/\psi(\alpha_1), \lambda_1) = O\{\psi(\alpha)/\sqrt{n}\}$ by Theorem 2.2.2, which is theoretically negligible. However, in practice, it may happen that $\psi(\alpha_1)$ is comparable to $\sqrt{n}$ (for example $\psi(8) = 8 + 7c/8$ for model 1 in the simulation part) which results in significant covariance. Therefore, we correct the statistic in (2.2.24) by removing the largest sample eigenvalue from the linear spectral statistics part. Actually, by using Slutsky's theorem and the fact that the single sample spiked eigenvalues converge to a constant, our proof

of Theorem 2.2.4 also applies to the case when linear spectral statistics do not include the sample spiked eigenvalues.

It then suffices to recalculate the variance of LSS part without the largest sample spiked eigenvalue and estimate it. By taking contour z enclosing all the sample eigenvalues except the largest spiked one, and after analyzing the contour integral, we find that the more accurate variance is just to replace $r_m = \text{tr} \Sigma^m / p$ with $r_m - \alpha_1^m / p$, $m = 1, 2, 3, 4$ in (2.2.26), and denote it by $(\sigma_{lin}^{(1)})^2$. The corresponding estimator for $r_m - \alpha_1^m / p$ is obtained by replacing $s_m = \text{tr} \mathbf{S}_1^m / n$ with $s_m - \lambda_1^m / n$ in (2.2.30). Thus we find an estimator for $(\sigma_{lin}^{(1)})^2$, and denote it by $\widehat{(\sigma_{lin}^{(1)})^2}$. Since λ_1 has an almost sure limit which is bounded, we can easily get that

$$\frac{\widehat{\sigma_{lin}^{(1)}}}{\widehat{\sigma_{lin}}} \xrightarrow{i.p.} 1. \quad (2.2.35)$$

Thus $\widehat{\sigma_{lin}^{(1)}}$ is still a consistent estimator of σ_{lin} .

Let

$$\begin{aligned} M_n &:= \sqrt{n} \frac{\lambda_1(\mathbf{S}_1) - \lambda_1(\mathbf{S}_2)}{\widehat{\sigma_{spi}}}; \\ L_n &:= \frac{\sum_{i=1}^p f\{\lambda_i(\mathbf{S}_1)\} - f\{\lambda_i(\mathbf{S}_2)\}}{\widehat{\sigma_{lin}}}, \text{ where } f(x) = x + x^2; \\ L_n^{(1)} &:= \frac{\sum_{i=2}^p f\{\lambda_i(\mathbf{S}_1)\} - f\{\lambda_i(\mathbf{S}_2)\}}{\widehat{\sigma_{lin}^{(1)}}}, \text{ where } f(x) = x + x^2; \\ T_n &:= M_n^2 + (L_n^{(1)})^2. \end{aligned} \quad (2.2.36)$$

Then we have the following corollary which provides the asymptotic distribution of T_n under the null hypothesis (2.2.22) with both Σ_1 and Σ_2 being spiked population covariance matrices.

Corollary 2.2.7. *Suppose that $\mathbf{S}_1$ and $\mathbf{S}_2$ defined in (2.2.23) satisfies Assumptions 1, 2, 3, 4, 5 and 7. Moreover we assume that $\sum_{i=1}^p (\Sigma_{1ii})^2 / p$ is bounded from below for sufficiently large p . Under the null hypothesis $\Sigma_1 = \Sigma_2$, we have*

$$T_n \xrightarrow{D} \chi_2^2 \quad (2.2.37)$$

Remark 9. The assumption that $\sum_{i=1}^p (\Sigma_{1ii})^2 / p$ is bounded from below assures that the estimator of γ_4 is consistent. See the discussion above (2.2.34).

Proof. From CLT of LSS (Theorem 1.4 of [Pan and Zhou \(2008\)](#)) and the independence between $\mathbf{S}_1$ and $\mathbf{S}_2$, we have

$$\frac{\sum_{i=1}^p f\{\lambda_i(\mathbf{S}_1)\} - f\{\lambda_i(\mathbf{S}_2)\}}{\sigma_{lin}} \xrightarrow{D} N(0, 1). \quad (2.2.38)$$

Hence by Slutsky's theorem and $\widehat{\sigma_{lin}} \xrightarrow{i.p.} \sigma_{lin}$, L_n converge weakly to $N(0, 1)$. By the almost convergence of the spiked eigenvalues we find that

$$\frac{f\{\lambda_1(\mathbf{S}_1)\} - f\{\lambda_1(\mathbf{S}_2)\}}{\sigma_{lin}} \xrightarrow{i.p.} 0. \quad (2.2.39)$$

Then (2.2.38), (2.2.39) and $\widehat{\sigma_{lin}^{(1)}}/\sigma_{lin} \xrightarrow{i.p.} 1$ imply that

$$L_n^{(1)} \xrightarrow{D} N(0, 1). \quad (2.2.40)$$

Similarly, it can be inferred from Theorem 2.2.2 that $M_n \xrightarrow{D} N(0, 1)$. Therefore, using Theorem 2.2.4, we obtain

$$T_n \xrightarrow{D} \chi_2^2. \quad (2.2.41)$$

□

2.3 Simulations

This section is to conduct the simulations to verify the performance of the earlier proposed statistics and the accuracy of the estimator of the population eigenvectors corresponding to the spikes.

We introduce five covariance models to be used in simulations.

- Model 1: $\Sigma^{(1)} = \text{diag}(8, 1, \dots, 1)_{p \times p}$.
- Model 2: $\Sigma^{(2)} = \text{diag}(6, 2, \dots, 2, 1, \dots, 1)_{p \times p}$ where the number of 2 is 10.
- Model 3: $\Sigma^{(3)} = \mathbf{O}_p \text{diag}(12, d_2, \dots, d_p)_{p \times p} \mathbf{O}_p^T$ where $d_i = 3 - 1.5(i - 1)/p$, and

$$\mathbf{O}_p = \begin{bmatrix} \mathbf{O}_1 & 0 \\ 0 & \mathbf{I}_{p-3} \end{bmatrix}$$

where $\mathbf{O}_1$ is a 3×3 orthogonal matrix.

- Model 4: $\Sigma^{(4)} = \mathbf{O}_p \text{diag}(15, d_2, \dots, d_p)_{p \times p} \mathbf{O}_p^T$ where $d_i = 3 - 2(i - 1)/p$ and $\mathbf{O}_p$ is the same as Model 3.

- Model 5: $\Sigma^{(5)} = \text{diag}(12, 2, \dots, 2, 1, \dots, 1)_{p \times p}$ where the number of 2 is 10.

We consider two types of distribution for entries of $\mathbf{X}_1$ and $\mathbf{X}_2$: standard normal distribution, and $t_{10}/\sqrt{(5/4)}$. We investigate the performance of T_n , and compare it with the tests in Li and Chen (2012) and Cai et al. (2013), respectively, denoted as Chen's test and Cai's test. The performance of M_n and L_n are also reported.

2.3.1 Approximation accuracy

In Tables 2.1 and 2.2, we report the empirical sizes of testing $H_0 : \Sigma_1 = \Sigma_2 = \Sigma^{(i)}$ for $\Sigma^{(i)}$ given by the above Model 1-5. The results listed in Table 2.1 are for standard normal distributed entries while Table 2.2 is for normalized t_{10} distributed entries. We run 500 simulation replications for each test of population covariance matrices. The nominal test size is 0.05. From the tables, we can see that the empirical sizes are around 0.05, which indicates that the χ_2^2 approximation is accurate. We would like to point out that although $\lambda_1 \xrightarrow{a.s.} \psi(\alpha_1)$ as n goes to infinity, the approximation is not accurate enough when $n = 100$. The estimating errors in (2.2.29) and (2.2.30) are slightly amplified if λ_1^2 is involved. This accounts for the slightly smaller size for statistics T_n, M_n and L_n in Tables 2.1 and 2.2. As a comparison, $L_n^{(1)}$ does not include the largest eigenvalue, and it has empirical sizes close to 0.05.

In Table 2.6, we record the performance of our estimator of $\sum_{i=1}^p u_{1i}^4$ for Model 4. We run 500 replications for each pair of (p, n) and list the averaged relative error and standard deviation. It can be seen that the estimator performs well.

2.3.2 Power discussion

We consider the power of the test for comparing three pairs of covariances $\Sigma^{(1)}$ vs $\Sigma^{(2)}$, $\Sigma^{(3)}$ vs $\Sigma^{(4)}$ and $\Sigma^{(2)}$ vs $\Sigma^{(5)}$. The empirical power of the above three comparisons are, respectively, summarized in Tables 2.3, 2.4 and 2.5. In Tables 2.3-2.5, we find that T_n always outperforms Cai's test and Chen's test. In Table 2.3, all the three tests i.e., T_n , Chen and Cai's tests have competitive power. However, in Table 2.4, both Cai and Chen's tests lose power while T_n has significant better power and the power increase as p increases. For the test comparing $\Sigma^{(2)}$ and $\Sigma^{(5)}$ in Table 2.5, Cai's test loses power, while T_n and Chen's tests have satisfactory power. The performance of T_n is more stable than Chen's test as p increases, and

TABLE 2.1: Empirical size for testing $H_0 : \Sigma_1 = \Sigma_2 = \Sigma^{(i)}$ for data generated from Model i ($i = 1, 2, 3, 4, 5$) with $N(0,1)$ entries. The sample size is 100 for both samples.

Model	p	40	60	80	100	120	150	240	300
$\Sigma^{(1)}$	T_n	0.042	0.048	0.038	0.050	0.042	0.042	0.052	0.038
	M_n	0.044	0.044	0.054	0.042	0.042	0.042	0.042	0.038
	L_n	0.026	0.030	0.046	0.042	0.038	0.030	0.066	0.044
	$L_n^{(1)}$	0.046	0.044	0.050	0.052	0.040	0.048	0.050	0.044
	Cai	0.036	0.048	0.030	0.058	0.044	0.040	0.044	0.046
	Chen	0.084	0.078	0.072	0.068	0.070	0.060	0.048	0.032
$\Sigma^{(2)}$	T_n	0.036	0.030	0.040	0.052	0.026	0.063	0.042	0.048
	M_n	0.028	0.032	0.052	0.042	0.040	0.042	0.044	0.046
	L_n	0.030	0.030	0.042	0.042	0.036	0.054	0.044	0.044
	$L_n^{(1)}$	0.052	0.060	0.056	0.056	0.036	0.034	0.046	0.034
	Cai	0.044	0.038	0.048	0.058	0.044	0.048	0.044	0.038
	Chen	0.056	0.050	0.066	0.056	0.046	0.053	0.054	0.056
$\Sigma^{(3)}$	T_n	0.024	0.030	0.046	0.034	0.040	0.044	0.042	0.068
	M_n	0.038	0.036	0.044	0.072	0.034	0.034	0.038	0.066
	L_n	0.040	0.032	0.036	0.046	0.050	0.038	0.050	0.070
	$L_n^{(1)}$	0.052	0.058	0.046	0.058	0.046	0.054	0.054	0.054
	Cai	0.040	0.046	0.048	0.036	0.032	0.054	0.048	0.028
	Chen	0.062	0.060	0.050	0.056	0.058	0.050	0.044	0.038
$\Sigma^{(4)}$	T_n	0.056	0.044	0.048	0.034	0.048	0.044	0.030	0.030
	M_n	0.046	0.040	0.043	0.040	0.044	0.044	0.032	0.030
	L_n	0.034	0.040	0.038	0.032	0.038	0.052	0.048	0.036
	$L_n^{(1)}$	0.052	0.052	0.054	0.052	0.046	0.038	0.052	0.054
	Cai	0.046	0.056	0.050	0.040	0.054	0.054	0.026	0.040
	Chen	0.072	0.064	0.086	0.066	0.058	0.058	0.030	0.042
$\Sigma^{(5)}$	T_n	0.028	0.052	0.040	0.046	0.060	0.056	0.046	0.058
	M_n	0.042	0.028	0.046	0.036	0.038	0.042	0.030	0.052
	L_n	0.026	0.024	0.042	0.032	0.042	0.034	0.038	0.040
	$L_n^{(1)}$	0.054	0.058	0.056	0.054	0.036	0.032	0.042	0.042
	Cai	0.046	0.054	0.054	0.044	0.046	0.038	0.028	0.042
	Chen	0.074	0.062	0.090	0.054	0.064	0.052	0.078	0.056

T_n outperforms Chen's test for large enough p . In fact, we can infer from (2.2.3) that the limit of difference of two sample spiked eigenvalues increases as p increases, so it is understandable that M_n has good power for large p cases.

We observe T_n has good power whenever the differences between the two covariances are introduced by either the non-spike eigenvalues (Table 2.3 and 2.4) or the spike eigenvalues (Table 2.5). Specifically, for the tests comparing $\Sigma^{(1)}$ and $\Sigma^{(2)}$ (Table 2.3), the main differences between $\Sigma^{(1)}$ and $\Sigma^{(2)}$ are from non-spike eigenvalues. Thus, M_n has relatively small power in this scenario, and the power of T_n in Table 2.3 are inherited from the difference of LSS excluding the largest

TABLE 2.2: Empirical size for testing $H_0 : \Sigma_1 = \Sigma_2 = \Sigma^{(i)}$ for data generated from Model i ($i = 1, 2, 3, 4, 5$) with $t_{10}/\sqrt{5/4}$ entries. The sample size is 100 for both samples.

Model	p	40	60	80	100	120	150	240	300
$\Sigma^{(1)}$	T_n	0.052	0.034	0.054	0.038	0.034	0.042	0.030	0.040
	M_n	0.048	0.028	0.042	0.036	0.042	0.038	0.036	0.040
	L_n	0.034	0.026	0.042	0.036	0.042	0.046	0.06	0.056
	$L_n^{(1)}$	0.070	0.050	0.050	0.054	0.046	0.042	0.044	0.056
	Cai	0.036	0.040	0.040	0.022	0.022	0.032	0.034	0.020
	Chen	0.098	0.102	0.112	0.094	0.092	0.054	0.048	0.050
$\Sigma^{(2)}$	T_n	0.050	0.026	0.020	0.042	0.058	0.046	0.048	0.038
	M_n	0.048	0.028	0.026	0.034	0.042	0.030	0.034	0.038
	L_n	0.054	0.026	0.042	0.048	0.046	0.052	0.052	0.052
	$L_n^{(1)}$	0.046	0.044	0.050	0.052	0.040	0.048	0.056	0.038
	Cai	0.036	0.036	0.034	0.032	0.046	0.020	0.034	0.026
	Chen	0.086	0.058	0.082	0.080	0.072	0.052	0.056	0.050
$\Sigma^{(3)}$	T_n	0.038	0.030	0.034	0.046	0.042	0.030	0.038	0.048
	M_n	0.044	0.036	0.044	0.042	0.034	0.026	0.040	0.018
	L_n	0.028	0.038	0.036	0.040	0.050	0.042	0.044	0.046
	$L_n^{(1)}$	0.056	0.052	0.058	0.058	0.044	0.034	0.034	0.036
	Cai	0.028	0.032	0.032	0.020	0.030	0.032	0.038	0.028
	Chen	0.084	0.066	0.072	0.054	0.066	0.050	0.038	0.040
$\Sigma^{(4)}$	T_n	0.040	0.038	0.046	0.034	0.046	0.046	0.038	0.040
	M_n	0.042	0.034	0.034	0.054	0.030	0.036	0.038	0.040
	L_n	0.018	0.018	0.032	0.026	0.032	0.020	0.046	0.044
	$L_n^{(1)}$	0.060	0.054	0.060	0.056	0.040	0.036	0.034	0.040
	Cai	0.038	0.042	0.040	0.028	0.022	0.024	0.032	0.018
	Chen	0.108	0.078	0.066	0.090	0.056	0.094	0.058	0.064
$\Sigma^{(5)}$	T_n	0.044	0.042	0.052	0.060	0.030	0.060	0.034	0.042
	M_n	0.022	0.042	0.056	0.032	0.024	0.038	0.036	0.026
	L_n	0.022	0.034	0.028	0.030	0.018	0.038	0.034	0.038
	$L_n^{(1)}$	0.068	0.058	0.060	0.058	0.048	0.046	0.058	0.040
	Cai	0.026	0.038	0.030	0.038	0.034	0.046	0.028	0.032
	Chen	0.100	0.118	0.104	0.086	0.084	0.104	0.070	0.048

population eigenvalue, i.e. the statistic $L_n^{(1)}$. This phenomenon can be also seen by comparing the powers of L_n and $L_n^{(1)}$. Note that L_n does not have good power because the largest eigenvalue in $\Sigma^{(2)}$ is smaller than that of $\Sigma^{(1)}$, which offsets the effect of $L_n^{(1)}$. In Table 2.4, the power of T_n are also mainly contributed by $L_n^{(1)}$, but different from the results in Table 2.3, L_n and $L_n^{(1)}$ both have powers close to 1 for large p . However, in Table 2.5, M_n has significant power but $L_n^{(1)}$ loses power because $\Sigma^{(2)}$ and $\Sigma^{(5)}$ shares the same eigenvalues except the large difference between their largest spiked population eigenvalues. In this scenario, the power of T_n are mainly due to the contribution of M_n .

TABLE 2.3: Empirical power for testing $H_0 : \Sigma_1 = \Sigma_2$ where $\Sigma_1 = \Sigma^{(1)}$ and $\Sigma_2 = \Sigma^{(2)}$ with two types of data entries: $N(0, 1)$ and $t_{10}/\sqrt{5/4}$. The sample size is 100 for both samples.

Data Entries	p	40	60	80	100	120	150	240	300
$N(0, 1)$	T_n	1.000	1.000	1.000	0.994	0.992	0.978	0.894	0.822
	M_n	0.248	0.266	0.224	0.218	0.218	0.246	0.264	0.220
	L_n	0.194	0.188	0.232	0.256	0.286	0.230	0.274	0.270
	$L_n^{(1)}$	1.000	1.000	1.000	1.000	0.996	0.978	0.884	0.844
	Cai	0.796	0.668	0.584	0.530	0.434	0.372	0.260	0.194
	Chen	0.852	0.786	0.722	0.606	0.548	0.448	0.332	0.240
$t_{10}/\sqrt{5/4}$	T_n	1.000	0.998	0.996	0.980	0.942	0.898	0.728	0.600
	M_n	0.148	0.164	0.150	0.160	0.198	0.136	0.194	0.136
	L_n	0.142	0.176	0.182	0.204	0.200	0.204	0.200	0.216
	$L_n^{(1)}$	1.000	0.996	0.994	0.984	0.966	0.924	0.748	0.632
	Cai	0.448	0.326	0.236	0.190	0.162	0.132	0.066	0.062
	Chen	0.814	0.774	0.710	0.586	0.522	0.458	0.346	0.258

TABLE 2.4: Empirical power for testing $H_0 : \Sigma_1 = \Sigma_2$ where $\Sigma_1 = \Sigma^{(3)}$ and $\Sigma_2 = \Sigma^{(4)}$ with two types of data entries: $N(0, 1)$ and $t_{10}/\sqrt{5/4}$. The sample size is 100 for both samples.

Data Entries	p	40	60	80	100	120	150	240	300
$N(0, 1)$	T_n	0.792	0.938	0.976	0.994	1.000	1.000	1.000	1.000
	M_n	0.176	0.176	0.156	0.150	0.142	0.132	0.098	0.084
	L_n	0.040	0.062	0.178	0.318	0.506	0.762	0.994	1.000
	$L_n^{(1)}$	0.818	0.964	0.986	0.998	1.000	1.000	1.000	1.000
	Cai	0.062	0.050	0.092	0.074	0.068	0.076	0.070	0.052
	Chen	0.250	0.236	0.206	0.172	0.174	0.150	0.130	0.160
$t_{10}/\sqrt{5/4}$	T_n	0.636	0.818	0.930	0.954	0.982	0.998	1.000	1.000
	M_n	0.158	0.146	0.104	0.114	0.136	0.116	0.072	0.074
	L_n	0.024	0.054	0.144	0.266	0.374	0.632	0.986	0.996
	$L_n^{(1)}$	0.686	0.862	0.952	0.960	0.992	1.000	1.000	1.000
	Cai	0.032	0.058	0.026	0.052	0.040	0.030	0.032	0.036
	Chen	0.248	0.236	0.212	0.178	0.216	0.184	0.168	0.122

2.4 Proof of Theorem 2.2.1

This section is to give the proof of Theorem 2.2.1. We begin with a list of results.

1. Two matrix formulas:

$$\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{B}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{A}^{-1} \quad (2.4.1)$$

$$\mathbf{A}(\mathbf{I} + \mathbf{BA})^{-1} = (\mathbf{I} + \mathbf{AB})^{-1}\mathbf{A} \quad (2.4.2)$$

2. Let $X = (X_1, \dots, X_n)$, where X_i 's are i.i.d real random variables with

TABLE 2.5: Empirical power for testing $H_0 : \Sigma_1 = \Sigma_2$ where $\Sigma_1 = \Sigma^{(2)}$ and $\Sigma_2 = \Sigma^{(5)}$ with two types of data entries: $N(0, 1)$ and $t_{10}/\sqrt{5/4}$. The sample size is 100 for both samples.

Data Entries	p	40	60	80	100	120	150	240	300
$N(0, 1)$	T_n	0.870	0.858	0.864	0.842	0.842	0.820	0.830	0.844
	M_n	0.910	0.940	0.946	0.926	0.908	0.888	0.900	0.906
	L_n	0.882	0.892	0.878	0.856	0.818	0.772	0.722	0.704
	$L_n^{(1)}$	0.046	0.048	0.046	0.046	0.058	0.056	0.036	0.062
	Cai	0.236	0.140	0.130	0.110	0.098	0.096	0.076	0.058
	Chen	0.908	0.872	0.878	0.854	0.778	0.730	0.630	0.596
$t_{10}/\sqrt{5/4}$	T_n	0.706	0.644	0.670	0.702	0.694	0.672	0.670	0.634
	M_n	0.822	0.748	0.790	0.786	0.804	0.786	0.782	0.770
	L_n	0.756	0.696	0.688	0.682	0.668	0.600	0.546	0.490
	$L_n^{(1)}$	0.054	0.046	0.050	0.040	0.050	0.054	0.044	0.042
	Cai	0.096	0.062	0.058	0.058	0.060	0.050	0.034	0.034
	Chen	0.838	0.784	0.790	0.768	0.778	0.704	0.634	0.588

TABLE 2.6: Empirical averaged relative error (η) and standard deviation (sd) of the proposed estimators for $\sum_{i=1}^p u_{1i}^4$, with true value 0.5427 in Model 4. The simulation replication is 500.

Data Entries	(p, n)	(100,100)	(180,100)	(180,200)	(180,300)	(300,100)	(300,200)	(300,300)	(300,500)
$N(0, 1)$	η	0.092	0.110	0.066	0.052	0.134	0.070	0.061	0.044
	sd	0.062	0.077	0.045	0.036	0.098	0.050	0.044	0.033
$t_{10}/\sqrt{5/4}$	η	0.097	0.120	0.076	0.062	0.144	0.079	0.063	0.054
	sd	0.066	0.082	0.052	0.042	0.100	0.054	0.043	0.042

mean zero and variance one. Let $\mathbf{A} = (a_{ij})_{n \times n}$ and $\mathbf{B} = (b_{ij})_{n \times n}$ be two real or complex matrices. Then we have an identity

$$\mathbb{E}(X^\top \mathbf{A} X - \text{tr} \mathbf{A})(X^\top \mathbf{B} X - \text{tr} \mathbf{B}) = (\mathbb{E}|X_1|^4 - 3) \sum_{i=1}^n a_{ii} b_{ii} + \text{tr} \mathbf{A} \mathbf{B}^\top + \text{tr} \mathbf{A} \mathbf{B}. \quad (2.4.3)$$

Lemma 2.4.1. (Theorem 35.12 of Billingsley (1995)) Suppose that for each n , $Y_{n1}, Y_{n2}, \dots, Y_{nr_n}$ is a real martingale difference sequence with respect to an increasing σ -field $\{\mathcal{F}_{nj}\}$ having second moments. If as $n \rightarrow \infty$,

$$(i) \quad \sum_{j=1}^{r_n} \mathbb{E}(Y_{nj}^2 | \mathcal{F}_{n,j-1}) \xrightarrow{i.p.} \sigma^2,$$

where σ^2 is a positive constant, and for each $\epsilon > 0$,

$$(ii) \quad \sum_{j=1}^{r_n} \mathbb{E}\{Y_{nj}^2 I(|Y_{nj}| > \epsilon)\} \rightarrow 0,$$

then

$$\sum_{j=1}^{r_n} Y_{nj} \xrightarrow{D} N(0, \sigma^2).$$

4. Suppose that entries of $\mathbf{x}$ is truncated at $\eta_n n^{1/4}$ and centralized, i.e. $x_{ij} = \frac{1}{\sqrt{n}} q_{ij}$, where q_{ij} satisfying Assumption 1, are truncated at $\eta_n n^{1/4}$ and centralized. $\mathbf{M}, \mathbf{M}_1$ and $\mathbf{M}_2$ are $p \times p$ non-random matrices (or independent of $\mathbf{x}$). $\mathbf{w}$ is a $p \times 1$ non-random vector with a bounded spectral norm. We conclude the following simple results from Lemma 2.2 in [Bai and Silverstein \(2004\)](#):

$$\mathbb{E}|\mathbf{x}^\top \mathbf{M} \mathbf{x} - \frac{1}{n} \text{tr} \mathbf{M}|^d \leq C \|\mathbf{M}\|^d n^{-d/2}, \quad (2.4.4)$$

$$\mathbb{E}|\mathbf{x}^\top \mathbf{M}_1 \mathbf{w} \mathbf{w}^\top \mathbf{M}_2 \mathbf{x} - \frac{1}{n} \mathbf{w}^\top \mathbf{M}_2 \mathbf{M}_1 \mathbf{w}|^d \leq C \|\mathbf{M}_1\|^d \|\mathbf{M}_2\|^d \eta_n^{2d-4} n^{-d/2-1}, \quad (2.4.5)$$

$$\mathbb{E}|\mathbf{x}^\top \mathbf{M}_1 \mathbf{w} \mathbf{w}^\top \mathbf{M}_2 \mathbf{x}|^d \leq C \|\mathbf{M}_1\|^d \|\mathbf{M}_2\|^d \eta_n^{2d-4} n^{-d/2-1}. \quad (2.4.6)$$

Proof of Theorem 2.2.1. We below only prove (2.2.6) and the proof of (2.2.7) is similar. The overall strategy of the proof is to decompose $\mathbf{w}_1^\top \mathbf{X} (\mathbf{I} - \mathbf{X}^\top \frac{\Sigma_{1P}}{\psi_n(\alpha)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{w}_1$ into summation of martingale differences and then apply Lemma 2.4.1. We assume that X has already been truncated at $\eta_n n^{1/4}$ and centralized according to the argument in the Appendix.

CLT of the random part. Throughout the rest of the chapter, let $\mathbf{x}_k$ be the k -th ($k = 1, \dots, n$) column of $\mathbf{X}$, and $\mathbf{e}_k = (0, \dots, 0, 1, 0, \dots, 0)$ be an n -dimensional vector with the k -th element being 1. We use C to denote constants which may

change from line to line. Introduce notations

$$\begin{aligned}
\theta &= \lim \psi_n(\alpha) = \psi(\alpha), \quad \mathbf{X}_k = \mathbf{X} - \mathbf{x}_k \mathbf{e}_k^\top, \quad \mathbf{X}_{jk} = \mathbf{X} - \mathbf{x}_k \mathbf{e}_k^\top - \mathbf{x}_j \mathbf{e}_j^\top, \\
\tilde{\Sigma}_1 &= \frac{\Sigma_{1P}}{\psi_n(\alpha)}, \quad \mathbf{A} = \mathbf{I}_n - \mathbf{X}^\top \tilde{\Sigma}_1 \mathbf{X}, \quad \mathbf{A}_k = \mathbf{I}_n - \mathbf{X}_k^\top \tilde{\Sigma}_1 \mathbf{X}_k, \\
\mathbf{D}_k &= \mathbf{I}_p - \tilde{\Sigma}_1 \mathbf{X}_k \mathbf{X}_k^\top, \quad \mathbf{D} = \mathbf{I}_p - \tilde{\Sigma}_1 \mathbf{X} \mathbf{X}^\top, \quad \mathbf{D}_{jk} = \mathbf{I}_p - \tilde{\Sigma}_1 \mathbf{X}_{jk} \mathbf{X}_{jk}^\top, \\
\mathbf{B}_k &= \mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top (\mathbf{D}_k^\top)^{-1}, \quad \delta_k = \mathbf{x}_k^\top \mathbf{B}_k \mathbf{x}_k - \frac{1}{n} \text{tr} \mathbf{B}_k \\
\alpha_k &= \frac{1}{1 - \mathbf{x}_k^\top \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1} \mathbf{x}_k}, \quad \alpha_{jk} = \frac{1}{1 - \mathbf{x}_k^\top \tilde{\Sigma}_1 (\mathbf{D}_{jk}^\top)^{-1} \mathbf{x}_k} \\
\bar{\alpha}_k &= \frac{1}{1 - \frac{1}{n} \text{tr} \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1}}, \quad \bar{\alpha}_{jk} = \frac{1}{1 - \frac{1}{n} \text{tr} \tilde{\Sigma}_1 (\mathbf{D}_{jk}^\top)^{-1}}, \\
a_n &= \frac{1}{1 - \frac{1}{n} \mathbb{E} \text{tr} \tilde{\Sigma}_1 (\mathbf{D}_1^\top)^{-1}}, \quad a_{1n} = \frac{1}{1 - \mathbb{E} \frac{1}{n} \text{tr} \tilde{\Sigma}_1 (\mathbf{D}_{12}^\top)^{-1}} \\
\gamma_k &= \mathbf{x}_k^\top \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1} \mathbf{x}_k - \frac{1}{n} \text{tr} \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1}, \quad \gamma_{1k} = \mathbf{x}_k^\top \tilde{\Sigma}_1 (\mathbf{D}_{1k}^\top)^{-1} \mathbf{x}_k - \frac{1}{n} \text{tr} \tilde{\Sigma}_1 (\mathbf{D}_{1k}^\top)^{-1}.
\end{aligned} \tag{2.4.7}$$

With the help of (2.4.2), it is not difficult to conclude the following facts:

$$\mathbf{e}_k^\top \mathbf{X}_k^\top = \mathbf{e}_k^\top \mathbf{A}_k^{-1} \mathbf{X}_k^\top = 0, \tag{2.4.8}$$

and

$$\mathbf{e}_k^\top \mathbf{A}^{-1} \mathbf{X}^\top = \mathbf{e}_k^\top \mathbf{X}^\top \mathbf{D}^{-1} = \mathbf{x}_k^\top \mathbf{D}_k^{-1} \alpha_k. \tag{2.4.9}$$

Note that the quantities defined in (2.4.7) such as α_k , $\bar{\alpha}_k$ and a_n are not always bounded and the matrices such as $\mathbf{A}$, $\mathbf{D}_k$ are not always invertible. So we introduce events

$$\begin{aligned}
\mathcal{B}_1 &= \{|\tilde{\Sigma}_1 \mathbf{X} \mathbf{X}^\top| \leq 1 - \epsilon\}, \quad \mathcal{B}_{1k} = \{|\tilde{\Sigma}_1 \mathbf{X}_k \mathbf{X}_k^\top| \leq 1 - \epsilon\}, \\
\mathcal{B}_{1jk} &= \{|\tilde{\Sigma}_1 \mathbf{X}_{jk} \mathbf{X}_{jk}^\top| \leq 1 - \epsilon\}, \quad \mathcal{B}_{2k} = \{|\mathbf{x}_k^\top \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1} \mathbf{x}_k - (1 + \frac{1}{\theta \underline{m}(\theta)})| < \epsilon\}, \\
\mathcal{B}_{2jk} &= \{|\mathbf{x}_k^\top \tilde{\Sigma}_1 (\mathbf{D}_{jk}^\top)^{-1} \mathbf{x}_k - (1 + \frac{1}{\theta \underline{m}(\theta)})| < \epsilon\}, \\
\mathcal{B}_{3k} &= \{|\frac{1}{n} \text{tr} \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1} - (1 + \frac{1}{\theta \underline{m}(\theta)})| < \epsilon\}, \\
\mathcal{B}_{3jk} &= \{|\frac{1}{n} \text{tr} \tilde{\Sigma}_1 (\mathbf{D}_{jk}^\top)^{-1} - (1 + \frac{1}{\theta \underline{m}(\theta)})| < \epsilon\},
\end{aligned} \tag{2.4.10}$$

where ϵ is a small positive constant. Note that $\mathcal{B}_1 \subseteq \mathcal{B}_{1k} \subseteq \mathcal{B}_{1jk}$. Denote

$$\mathcal{B} = \mathcal{B}_1 \bigcap \left(\bigcap_{i=2,3} \bigcap_{k=1}^n \mathcal{B}_{ik} \right) \bigcap \left(\bigcap_{i=2,3} \bigcap_{1 \leq j \neq k \leq n} \mathcal{B}_{ijk} \right). \quad (2.4.11)$$

Then we have following lemma and the proof is postponed to the Appendix.

Lemma 2.4.2. *The event $\mathcal{B}$ holds with high probability (i.e. $P(\mathcal{B}) = 1 - n^{-l}$ for any large constant l).*

This lemma ensures that it suffices to establish CLT of $\mathbf{w}_1^\top \mathbf{X} \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{w}_1 I(\mathcal{B})$. When the event $\mathcal{B}$ holds the terms α_k , α_{jk} , $\bar{\alpha}_k$ and $\bar{\alpha}_{jk}$ defined in (2.4.7) are bounded. We remark here that the more accurate definition of a_n should be

$$a_n = \frac{1}{1 - \frac{1}{n} \mathbb{E} \text{tr} \tilde{\Sigma}_1 (\mathbf{D}_1^\top)^{-1} I(\mathcal{B}_1)},$$

which is bounded for sufficient large n , see (2.8.24) in the proof of Lemma 2.4.2. The definition of a_n in (2.4.7) is just for notational simplicity. Another important fact of a_n is

$$\lim_{n \rightarrow \infty} a_n \rightarrow \frac{\psi(\alpha)}{\alpha}. \quad (2.4.12)$$

This is because we have $a_n \rightarrow -\theta \underline{m}(\theta)$ as $n \rightarrow \infty$, see (2.8.24). Recall that $\theta = \psi(\alpha)$. By the fact that ψ is the inverse function of $\alpha : x \mapsto -1/\underline{m}(x)$, we have $\underline{m}(\theta) = -1/\alpha$. The above comment for a_n also applies to a_{1n} , and the limit of a_{1n} is also $\theta \underline{m}(\theta)$.

Let $\mathbb{E}_0(\cdot)$ denote expectation, and $\mathbb{E}_k(\cdot)$ denote the conditional expectation with respect to σ -field generated by $\mathbf{x}_1, \dots, \mathbf{x}_k$. We have

$$\begin{aligned}
& \sqrt{n}\{\mathbf{w}_1^\top \mathbf{X} \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{w}_1 I(\mathcal{B}) - \mathbb{E} \mathbf{w}_1^\top \mathbf{X} \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{w}_1 I(\mathcal{B})\} \\
&= \sqrt{n} \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) \{\mathbf{w}_1^\top \mathbf{X} \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{w}_1 I(\mathcal{B})\} \\
&= \sqrt{n} \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) \{\mathbf{w}_1^\top \mathbf{X} \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{w}_1 I(\mathcal{B}) - \mathbf{w}_1^\top \mathbf{X}_k \mathbf{A}_k^{-1} \mathbf{X}_k^\top \mathbf{w}_1 I(\mathcal{B}_{1k})\} \\
&= \sqrt{n} \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) \{\mathbf{w}_1^\top \mathbf{X} \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{w}_1 I(\mathcal{B}) - \mathbf{w}_1^\top \mathbf{X}_k \mathbf{A}_k^{-1} \mathbf{X}_k^\top \mathbf{w}_1 I(\mathcal{B})\} + o_p(1) \\
&= \sqrt{n} \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) \{\alpha_k \mathbf{x}_k^\top B_k \mathbf{x}_k I(\mathcal{B})\} + o_p(1),
\end{aligned} \tag{2.4.13}$$

where the last step uses the fact that by (2.4.8) and (2.4.9)

$$\begin{aligned}
& \mathbf{w}_1^\top \mathbf{X} \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{w}_1 - \mathbf{w}_1^\top \mathbf{X}_k \mathbf{A}_k^{-1} \mathbf{X}_k^\top \mathbf{w}_1 \\
&= \mathbf{w}_1^\top (\mathbf{X} - \mathbf{X}_k) \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{w}_1 + \mathbf{w}_1^\top \mathbf{X}_k (\mathbf{A}^{-1} - \mathbf{A}_k^{-1}) \mathbf{X}^\top \mathbf{w}_1 + \mathbf{w}_1^\top \mathbf{X}_k \mathbf{A}_k^{-1} (\mathbf{X}^\top - \mathbf{X}_k^\top) \mathbf{w}_1 \\
&= \mathbf{w}_1^\top \mathbf{x}_k \mathbf{e}_k^\top \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{w}_1 + \mathbf{w}_1^\top \mathbf{X}_k \mathbf{A}_k^{-1} (\mathbf{X}^\top \tilde{\Sigma}_1 \mathbf{X} - \mathbf{X}_k^\top \tilde{\Sigma}_1 \mathbf{X}_k) \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{w}_1 \\
&= \mathbf{w}_1^\top \mathbf{x}_k \mathbf{x}_k^\top \mathbf{D}_k^{-1} \mathbf{w}_1 \alpha_k + \mathbf{w}_1^\top \mathbf{X}_k \mathbf{A}_k^{-1} (\mathbf{e}_k \mathbf{x}_k^\top \tilde{\Sigma}_1 \mathbf{X} + \mathbf{X}_k^\top \tilde{\Sigma}_1 \mathbf{x}_k \mathbf{e}_k^\top) \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{w}_1 \\
&= \mathbf{w}_1^\top \mathbf{x}_k \mathbf{x}_k^\top \mathbf{D}_k^{-1} \mathbf{w}_1 \alpha_k + \mathbf{w}_1^\top \mathbf{X}_k \mathbf{X}_k^\top \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1} \mathbf{x}_k \mathbf{e}_k^\top \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{w}_1 \\
&= \mathbf{w}_1^\top (\mathbf{D}_k^\top)^{-1} \mathbf{x}_k \mathbf{x}_k^\top \mathbf{D}_k^{-1} \mathbf{w}_1 \alpha_k = \alpha_k \mathbf{x}_k^\top B_k \mathbf{x}_k.
\end{aligned} \tag{2.4.14}$$

Note that $\alpha_k = \bar{\alpha}_k + \bar{\alpha}_k^2 \gamma_k + \bar{\alpha}_k^2 \gamma_k^2 \alpha_k$. It follows that

$$\begin{aligned}
(2.4.13) &= \sqrt{n} \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) [(\bar{\alpha}_k + \bar{\alpha}_k^2 \gamma_k + \bar{\alpha}_k^2 \gamma_k^2 \alpha_k) (\delta_k + \frac{1}{n} \text{tr} B_k) I(\mathcal{B})] \\
&= \sqrt{n} \sum_{k=1}^n \mathbb{E}_k [(\bar{\alpha}_k \delta_k + \frac{1}{n} \bar{\alpha}_k^2 \gamma_k \text{tr} B_k) I(\mathcal{B}_{1k} \mathcal{B}_{3k})] \\
&\quad + \sqrt{n} \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) [(\bar{\alpha}_k^2 \gamma_k \delta_k + \bar{\alpha}_k^2 \gamma_k^2 \alpha_k \mathbf{x}_k^\top B_k \mathbf{x}_k) I(\mathcal{B})] + o_p(1),
\end{aligned} \tag{2.4.15}$$

where in the second equality, we use $\mathbb{E}_{k-1}\{\bar{\alpha}_k \delta_k I(\mathcal{B})\} = \mathbb{E}_{k-1}\{\bar{\alpha}_k \delta_k I(\mathcal{B}_{1k} \mathcal{B}_{3k})\} + o_p(n^{-2}) = o_p(n^{-2})$, and similarly, $\mathbb{E}_{k-1} \frac{1}{n} \bar{\alpha}_k^2 \gamma_k \text{tr} B_k I(\mathcal{B}) = \mathbb{E}_{k-1} \frac{1}{n} \bar{\alpha}_k^2 \gamma_k \text{tr} B_k I(\mathcal{B}_{1k} \mathcal{B}_{3k}) + o_p(n^{-2}) = o_p(n^{-2})$. We below omit the indicator functions such as $I(\mathcal{B}), I(\mathcal{B}_{1k})$ for simplicity, but one should bear in mind that a suitable indicate function of events is needed whenever handling the inverses of random matrices.

Using the Burkholder inequality, (2.4.4) and (2.4.5), we have

$$\mathbb{E}|\sqrt{n} \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) \gamma_k \delta_k|^2 \leq Cn^2 (\mathbb{E}|\gamma_k|^4)^{\frac{1}{2}} (\mathbb{E}|\delta_k|^4)^{\frac{1}{2}} = o(1). \quad (2.4.16)$$

By similar arguments, together with the fact that $\bar{\alpha}_k$ and α_k are bounded, we have

$$\begin{aligned} \sqrt{n} \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) (\bar{\alpha}_k^2 \gamma_k^2 \alpha_k \mathbf{x}_k^\top \mathbf{B}_k \mathbf{x}_k) &= o_p(1), \\ \sqrt{n} \sum_{k=1}^n \mathbb{E}_k \left(\frac{1}{n} \bar{\alpha}_k^2 \gamma_k \text{tr} \mathbf{B}_k \right) &= o_p(1). \end{aligned} \quad (2.4.17)$$

Therefore we only need to consider $\sqrt{n} \sum_{k=1}^n \mathbb{E}_k \bar{\alpha}_k \delta_k = \sqrt{n} \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) (\bar{\alpha}_k \delta_k) = \sqrt{n} \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) [(\bar{\alpha}_k - a_n) \delta_k] + \sqrt{n} a_n \sum_{k=1}^n \mathbb{E}_k \delta_k$. Similarly to (2.4.16), it is easy to get

$$\mathbb{E}|\sqrt{n} \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) [(\bar{\alpha}_k - a_n) \delta_k]|^2 = o(1).$$

Summarizing the above we conclude that

$$\sqrt{n} (\mathbf{w}_1^\top \mathbf{X} \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{w}_1 - \mathbb{E} \mathbf{w}_1^\top \mathbf{X} \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{w}_1) = \sqrt{n} \sum_{k=1}^n a_n \mathbb{E}_k (\delta_k) + o_p(1). \quad (2.4.18)$$

Let $Y_k = \mathbb{E}_k \delta_k = (\mathbb{E}_k - \mathbb{E}_{k-1}) \delta_k$. By the fact that a_n is bounded we obtain

$$\sum_{k=1}^n a_n^2 \mathbb{E} (n Y_k^2 I(|\sqrt{n} Y_k| \geq \epsilon)) \leq \frac{C}{\epsilon^2} \sum_{k=1}^n \mathbb{E} |\sqrt{n} Y_k|^4 \leq \frac{Cn^2}{\epsilon^2} \sum_{k=1}^n \mathbb{E} |\delta_k|^4 = o(1), \quad (2.4.19)$$

where in the last step we use (2.4.5) and $\eta_n \rightarrow 0$. By Lemma 2.4.1 it suffices to verify

$$\sum_{k=1}^n a_n^2 \mathbb{E}_{k-1} (n Y_k^2) \xrightarrow{i.p.} \sigma^2. \quad (2.4.20)$$

It follows from (2.4.3) that

$$n \sum_{k=1}^n \mathbb{E}_{k-1} (Y_k^2) = \frac{1}{n} (\gamma_4 - 3) \sum_{k=1}^n \sum_{i=1}^p (\mathbb{E}_k (\mathbf{B}_k)_{ii})^2 + \frac{2}{n} \sum_{k=1}^n \text{tr} (\mathbb{E}_k \mathbf{B}_k)^2. \quad (2.4.21)$$

It suffices to find the limits of

$$\frac{1}{n} \sum_{k=1}^n \sum_{i=1}^p \left(\mathbf{e}_i^\top \mathbb{E}_k \mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top (\mathbf{D}_k^\top)^{-1} \mathbf{e}_i \right)^2, \quad (2.4.22)$$

and

$$\frac{2}{n} \sum_{k=1}^n \text{tr} \left(\mathbb{E}_k \mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top (\mathbf{D}_k^\top)^{-1} \right)^2. \quad (2.4.23)$$

First, we deal with (2.4.22). Write

$$\begin{aligned} & \mathbf{e}_i^\top (\mathbf{D}_k^{-1} - \mathbb{E} \mathbf{D}_k^{-1}) \mathbf{w}_1 \mathbf{w}_1^\top (\mathbf{D}_k^\top)^{-1} \mathbf{e}_i \\ &= \mathbf{e}_i^\top (\mathbf{D}_k^{-1} - \mathbb{E} \mathbf{D}_k^{-1}) \mathbf{w}_1 \mathbf{w}_1^\top ((\mathbf{D}_k^\top)^{-1} - \mathbb{E} (\mathbf{D}_k^\top)^{-1}) \mathbf{e}_i + \mathbf{e}_i^\top (\mathbf{D}_k^{-1} - \mathbb{E} \mathbf{D}_k^{-1}) \mathbf{w}_1 \mathbf{w}_1^\top \mathbb{E} (\mathbf{D}_k^\top)^{-1} \mathbf{e}_i \\ &= \sum_{j_1, j_2 \neq k} \mathbf{e}_i^\top (\mathbb{E}_{j_1} \mathbf{D}_k^{-1} - \mathbb{E}_{j_1-1} \mathbf{D}_k^{-1}) \mathbf{w}_1 \mathbf{w}_1^\top (\mathbb{E}_{j_2} \mathbf{D}_k^{-1} - \mathbb{E}_{j_2-1} \mathbf{D}_k^{-1}) \mathbf{e}_i \\ & \quad + \sum_{j \neq k} \mathbf{e}_i^\top (\mathbb{E}_j \mathbf{D}_k^{-1} - \mathbb{E}_{j-1} \mathbf{D}_k^{-1}) \mathbf{w}_1 \mathbf{w}_1^\top \mathbb{E} (\mathbf{D}_k^\top)^{-1} \mathbf{e}_i. \end{aligned} \quad (2.4.24)$$

By the Burkholder inequality, (2.4.1), (2.4.6) and (2.4.24), we have

$$\begin{aligned} & \left[\mathbb{E} \left| \sum_{i=1}^p \mathbb{E}_k \mathbf{e}_i^\top (\mathbf{D}_k^{-1} - \mathbb{E} \mathbf{D}_k^{-1}) \mathbf{w}_1 \mathbf{w}_1^\top \mathbf{D}_k^{\top-1} \mathbf{e}_i \times \mathbf{e}_i^\top \mathbb{E}_k \mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \mathbf{D}_k^{\top-1} \mathbf{e}_i \right| \right]^2 \\ & \leq \sum_{i=1}^p \mathbb{E} |\mathbf{e}_i^\top (\mathbf{D}_k^{-1} - \mathbb{E} \mathbf{D}_k^{-1}) \mathbf{w}_1 \mathbf{w}_1^\top \mathbf{D}_k^{\top-1} \mathbf{e}_i|^2 \sum_{i=1}^p \mathbb{E} |\mathbf{e}_i^\top \mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \mathbf{D}_k^{\top-1} \mathbf{e}_i|^2 \\ & \leq C \sum_{i=1}^p \left[\mathbb{E} \left| \sum_{j_1 \neq k} \alpha_{kj_1} \mathbf{e}_i^\top (\mathbb{E}_{j_1} - \mathbb{E}_{j_1-1}) \mathbf{D}_{kj_1}^{-1} \tilde{\Sigma}_1 \mathbf{x}_{j_1} \mathbf{x}_{j_1}^\top \mathbf{D}_{kj_1}^{-1} \mathbf{w}_1 \right|^4 \right]^{1/2} \\ & \quad \times \left[\mathbb{E} \left| \sum_{j_2 \neq k} \alpha_{kj_2} \mathbf{e}_i^\top (\mathbb{E}_{j_2} - \mathbb{E}_{j_2-1}) \mathbf{D}_{kj_2}^{-1} \tilde{\Sigma}_1 \mathbf{x}_{j_2} \mathbf{x}_{j_2}^\top \mathbf{D}_{kj_2}^{-1} \mathbf{w}_1 \right|^4 \right]^{1/2} \\ & \quad + C \sum_{i=1}^p |\mathbf{w}_1^\top \mathbb{E} \mathbf{D}_k^{\top-1} \mathbf{e}_i|^2 \mathbb{E} \left| \sum_{j \neq k} \alpha_{kj} \mathbf{e}_i^\top (\mathbb{E}_j - \mathbb{E}_{j-1}) \mathbf{D}_{jk}^{-1} \tilde{\Sigma}_1 \mathbf{x}_j \mathbf{x}_j^\top \mathbf{D}_{jk}^{-1} \mathbf{w}_1 \right|^2 \\ & = O(n^{-1}), \end{aligned} \quad (2.4.25)$$

where the second inequality uses

$$\sum_{i=1}^n \mathbb{E} |\mathbf{e}_i^\top \mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \mathbf{D}_k^{\top-1} \mathbf{e}_i|^2 \leq \mathbb{E} \left| \sum_{i=1}^n \mathbf{w}_1^\top \mathbf{D}_k^{\top-1} \mathbf{e}_i \mathbf{e}_i^\top \mathbf{D}_k^{-1} \mathbf{w}_1 \right|^2 = \mathbb{E} |\mathbf{w}_1^\top \mathbf{D}_k^{\top-1} \mathbf{D}_k \mathbf{w}_1|^2 = O(1). \quad (2.4.26)$$

(2.4.25) implies

$$\frac{1}{n} \sum_{k=1}^n \sum_{i=1}^p (\mathbb{E}_k \mathbf{e}_i^\top (\mathbf{D}_k^{-1} - \mathbb{E} \mathbf{D}_k^{-1}) \mathbf{w}_1 \mathbf{w}_1^\top \mathbf{D}_k^{\top-1} \mathbf{e}_i) \times \mathbf{e}_i^\top \mathbb{E}_k \mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \mathbf{D}_k^{\top-1} \mathbf{e}_i \xrightarrow{i.p.} 0. \quad (2.4.27)$$

Similarly, we have

$$\frac{1}{n} \sum_{k=1}^n \sum_{i=1}^p \mathbf{e}_i^\top \mathbb{E} \mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \mathbb{E}_k (\mathbf{D}_k^{\top-1} - \mathbb{E} \mathbf{D}_k^{\top-1}) \mathbf{e}_i \times \mathbf{e}_i^\top \mathbb{E}_k \mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \mathbf{D}_k^{\top-1} \mathbf{e}_i \xrightarrow{i.p.} 0.$$

Therefore

$$\frac{1}{n} \sum_{k=1}^n \sum_{i=1}^p (\mathbf{e}_i^\top \mathbb{E}_k \mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \mathbf{D}_k^{\top-1} \mathbf{e}_i - \mathbf{e}_i^\top \mathbb{E} \mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \mathbb{E} \mathbf{D}_k^{\top-1} \mathbf{e}_i) \times \mathbf{e}_i^\top \mathbb{E}_k \mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \mathbf{D}_k^{\top-1} \mathbf{e}_i \xrightarrow{i.p.} 0.$$

By similar arguments, we have

$$\frac{1}{n} \sum_{k=1}^n \sum_{i=1}^p \mathbf{e}_i^\top \mathbb{E} \mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \mathbb{E} \mathbf{D}_k^{\top-1} \mathbf{e}_i \times \mathbf{e}_i^\top (\mathbb{E}_k \mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \mathbf{D}_k^{\top-1} - \mathbb{E} \mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \mathbb{E} \mathbf{D}_k^{\top-1}) \mathbf{e}_i \xrightarrow{i.p.} 0.$$

Thus to find the limit of (2.4.22), it is equivalent to considering the limit of

$$\frac{1}{n} \sum_{k=1}^n \sum_{i=1}^p (\mathbf{e}_i^\top \mathbb{E} \mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \mathbb{E} \mathbf{D}_k^{\top-1} \mathbf{e}_i)^2. \quad (2.4.28)$$

Let

$$\mathbf{T} = \mathbf{I} - \mathbb{E} \frac{1}{n} \sum_{k=2}^n \alpha_{1k} \tilde{\Sigma}_1. \quad (2.4.29)$$

Using (2.8.29) and the dominated convergence theorem, it is easy to verify that

$$\lim_{n \rightarrow \infty} \mathbb{E} \frac{1}{n} \sum_{k=2}^n \alpha_{1k} \rightarrow -\theta \underline{m}(\theta) = \frac{\psi(\alpha)}{\alpha}.$$

Since α has a positive distance to the support of Σ_{1P} , $\mathbf{T}$ is invertible for large n .

Write

$$\begin{aligned}
& \mathbb{E}(\mathbf{e}_i^\top \mathbf{D}_1^{-1} \mathbf{w}_1) - \mathbf{e}_i^\top \mathbf{T}^{-1} \mathbf{w}_1 \\
&= \mathbb{E}[\mathbf{e}_i^\top \mathbf{T}^{-1} (\sum_{j \geq 2} \tilde{\Sigma}_1 \mathbf{x}_j \mathbf{x}_j^\top - \mathbb{E} \frac{1}{n} \sum_{j \geq 2} \alpha_{1j} \tilde{\Sigma}_1) \mathbf{D}_1^{-1} \mathbf{w}_1] \\
&= \sum_{j \geq 2} \mathbb{E}[\alpha_{1j} \mathbf{x}_j^\top \mathbf{D}_{1j}^{-1} \mathbf{w}_1 \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{x}_j - \frac{\mathbb{E} \alpha_{1j}}{n} \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{D}_1^{-1} \mathbf{w}_1] \\
&= A_1 + A_2 + A_3,
\end{aligned} \tag{2.4.30}$$

where

$$\begin{aligned}
A_1 &= \sum_{j \geq 2} \mathbb{E}((\alpha_{1j} - \bar{\alpha}_{1j})(\mathbf{x}_j^\top \mathbf{D}_{1j}^{-1} \mathbf{w}_1 \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{x}_j - \frac{1}{n} \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{1j}^{-1} \mathbf{w}_1)), \\
A_2 &= \sum_{j \geq 2} \mathbb{E}(\frac{1}{n} \alpha_{1j} \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 (\mathbf{D}_{1j}^{-1} - \mathbf{D}_1^{-1}) \mathbf{w}_1), \\
A_3 &= \sum_{j \geq 2} \mathbb{E}(\frac{1}{n} \alpha_{1j} \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 (\mathbf{D}_1^{-1} - \mathbb{E} \mathbf{D}_1^{-1}) \mathbf{w}_1).
\end{aligned} \tag{2.4.31}$$

We prove $A_1 = O(n^{-1})$ first. Using $\alpha_{1j} - \bar{\alpha}_{1j} = \bar{\alpha}_{1j}^2 \gamma_{1j} + \bar{\alpha}_{1j}^2 \gamma_{1j}^2 \alpha_{1j}$, we can write $A_1 = A_{11} + A_{12}$, where

$$\begin{aligned}
A_{11} &= \sum_{j \geq 2} \mathbb{E} \bar{\alpha}_{1j}^2 \gamma_{1j} (\mathbf{x}_j^\top \mathbf{D}_{1j}^{-1} \mathbf{w}_1 \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{x}_j - \frac{1}{n} \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{1j}^{-1} \mathbf{w}_1), \\
A_{12} &= \sum_{j \geq 2} \mathbb{E} \alpha_{1j} \bar{\alpha}_{1j}^2 \gamma_{1j}^2 (\mathbf{x}_j^\top \mathbf{D}_{1j}^{-1} \mathbf{w}_1 \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{x}_j - \frac{1}{n} \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{1j}^{-1} \mathbf{w}_1).
\end{aligned} \tag{2.4.32}$$

Using (2.4.3), we obtain

$$\begin{aligned}
& \mathbb{E}[\bar{\alpha}_{1j}^2 \gamma_{1j} (\mathbf{x}_j^\top \mathbf{D}_{1j}^{-1} \mathbf{w}_1 \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{x}_j - \frac{1}{n} \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{1j}^{-1} \mathbf{w}_1) | \mathbf{X}_j] \\
&\leq \frac{C}{n^2} \sum_{k=1}^p (\tilde{\Sigma}_1 \mathbf{D}_{1j}^{-1})_{kk} (\mathbf{D}_{1j}^{-1} \mathbf{w}_1 \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1)_{kk} + C \frac{\text{tr}(\tilde{\Sigma}_1 \mathbf{D}_{1j}^{-2} \mathbf{w}_1 \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1)}{n^2} \\
&+ \frac{\text{tr}(\tilde{\Sigma}_1 \mathbf{D}_{1j}^{-1} \tilde{\Sigma}_1^\top \mathbf{T}^{-1} \mathbf{e}_i \mathbf{w}_1^\top \mathbf{D}_{1j}^{-1})}{n^2}.
\end{aligned} \tag{2.4.33}$$

The first summation is bounded by

$$\begin{aligned} & \left(\sum_{k=1}^p |\mathbf{e}_k^\top \mathbf{D}_{1j}^{-1} \mathbf{w}_1|^2 \right)^{1/2} \left(\sum_{k=1}^p |\mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{e}_k \mathbf{e}_k^\top \tilde{\Sigma}_1 \mathbf{D}_{1j}^{-1} \mathbf{e}_k|^2 \right)^{1/2} \\ & \leq \|\mathbf{D}_{1j}^{-1}\|^2 \|\tilde{\Sigma}_1\|^2 \|\mathbf{T}^{-1}\| < C. \end{aligned} \quad (2.4.34)$$

Thus the first term in (2.4.33) is $O(n^{-2})$. By similar but easier arguments, the second and third term also have bounds of the same order, so that we can conclude that

$$|A_{11}| = O(n^{-1}). \quad (2.4.35)$$

For A_{12} , using (2.4.4) and (2.4.5), we have

$$\begin{aligned} |A_{12}| & \leq \sum_{j \geq 2} (\mathbb{E} |\alpha_{1j} \bar{\alpha}_{1j}^2 \gamma_{1j}^2|)^{1/2} (\mathbb{E} |\mathbf{x}_j^\top \mathbf{D}_{1j}^{-1} \mathbf{w}_1 \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{x}_j - \frac{1}{n} \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{1j}^{-1} \mathbf{w}_1|^2)^{1/2} \\ & = O(n^{-1}). \end{aligned} \quad (2.4.36)$$

Thus

$$|A_1| = O(n^{-1}). \quad (2.4.37)$$

Consider the terms A_2 and A_3 now in (2.4.31). It follows from (2.4.1), (2.4.5), (2.4.6) and the Burkholder inequality that

$$\begin{aligned} |A_2| & = \left| \sum_{j \geq 2} \mathbb{E} \left(\frac{1}{n} \alpha_{1j}^2 \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{1j}^{-1} \tilde{\Sigma}_1 \mathbf{x}_j \mathbf{x}_j^\top \mathbf{D}_{1j}^{-1} \mathbf{w}_1 \right) \right| = O(n^{-1}) \\ |A_3| & = \left| \sum_{j \geq 2} \mathbb{E} \left[\frac{1}{n} (\alpha_{1j} - a_n) \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 (\mathbf{D}_1^{-1} - \mathbb{E} \mathbf{D}_1^{-1}) \mathbf{w}_1 \right] \right| \\ & = \left| \sum_{j \geq 2} \mathbb{E} \left[\frac{1}{n} (\alpha_{1j} - a_n) \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 \left(\sum_{k=2}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) (\mathbf{D}_1^{-1} - \mathbf{D}_{1k}^{-1}) \right) \mathbf{w}_1 \right] \right| \\ & \leq \frac{1}{n} \sum_{j \geq 2} (\mathbb{E} |\alpha_{1j} - a_n|^2)^{1/2} (\mathbb{E} \left| \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) \mathbf{e}_i^\top \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{1k}^{-1} \tilde{\Sigma}_1 \mathbf{x}_k \mathbf{x}_k^\top \mathbf{D}_{1k}^{-1} \mathbf{w}_1 \right|^2)^{1/2} \\ & = O(n^{-1}). \end{aligned} \quad (2.4.38)$$

From (2.4.30), (2.4.37), (2.4.38) and $\mathbf{w}_1^\top \mathbf{U}_2 = 0$ we have

$$\begin{aligned} \mathbb{E}(\mathbf{e}_i^\top \mathbf{D}_1^{-1} \mathbf{w}_1) &= \mathbf{e}_i^\top \mathbf{T}^{-1} \mathbf{w}_1 + O(n^{-1}) \\ &= \mathbf{e}_i^\top \left(\mathbf{I} + \mathbf{T}^{-1} \mathbb{E} \frac{1}{n} \sum_{k=2}^n \alpha_{1k} \tilde{\Sigma}_1 \right) \mathbf{w}_1 + O(n^{-1}) \\ &= w_{1i} + O(n^{-1}). \end{aligned} \quad (2.4.39)$$

Substituting these back into (2.4.22) we conclude that (2.4.28) asymptotically equals

$$(1 + o(1)) \sum_{i=1}^p w_{1i}^4.$$

For (2.4.23), it has the similar form to equation (4.7) in Bai et al. (2007). Hence rewrite (2.4.23) as

$$\begin{aligned} & \frac{2}{n} \sum_{k=1}^n \mathbb{E}_{k-1} \text{tr} \mathbb{E}_k (\mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \mathbf{D}_k^{\top-1}) \mathbb{E}_k (\mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \mathbf{D}_k^{\top-1}) \\ &= \frac{2}{n} \sum_{k=1}^n \mathbb{E}_{k-1} (\mathbf{w}_1^\top \mathbf{D}_k^{\top-1} \check{\mathbf{D}}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \check{\mathbf{D}}_k^{\top-1} \mathbf{D}_k^{-1} \mathbf{w}_1), \end{aligned} \quad (2.4.40)$$

where $\check{\mathbf{D}}_k^{-1}$ is defined similarly to $\mathbf{D}_k^{-1}$ by $(\mathbf{x}_1, \dots, \mathbf{x}_{k-1}, \check{\mathbf{x}}_{k+1}, \dots, \check{\mathbf{x}}_n)$ and where $\check{\mathbf{x}}_{k+1}, \dots, \check{\mathbf{x}}_n$ are i.i.d copies of $\mathbf{x}_{k+1}, \dots, \mathbf{x}_n$. Following the argument similar to (4.7)-(4.22) of their work, we can obtain

$$\begin{aligned} & \mathbb{E}_{k-1} (\mathbf{w}_1^\top \mathbf{D}_k^{\top-1} \check{\mathbf{D}}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \check{\mathbf{D}}_k^{\top-1} \mathbf{D}_k^{-1} \mathbf{w}_1) \times \left[1 - \frac{k-1}{n} a_{1n}^2 \frac{1}{n} \text{tr} \tilde{\Sigma}_1 \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{T}^{-1} \right] \\ &= (\mathbf{w}_1^\top \mathbf{T}^{-2} \mathbf{w}_1)^2 \left[1 + \frac{k-1}{n} a_{1n}^2 \frac{1}{n} \mathbb{E}_{k-1} \text{tr} \mathbf{D}_k^{-1} \tilde{\Sigma}_1 \check{\mathbf{D}}_k^{-1} \tilde{\Sigma}_1 \right] + o_p(1). \end{aligned} \quad (2.4.41)$$

Note that $\mathbf{w}_1^\top \mathbf{T}^{-2} \mathbf{w}_1 = 1$ as in (2.4.39). Combining (2.4.2) with (2.18) of Bai and Silverstein (2004) we have

$$\mathbb{E}_{k-1} \text{tr} \mathbf{D}_k^{-1} \tilde{\Sigma}_1 \check{\mathbf{D}}_k^{-1} \tilde{\Sigma}_1 = \frac{\text{tr} \tilde{\Sigma}_1 \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{T}^{-1} + o_p(1)}{1 - \frac{k-1}{n^2} \theta^2(\underline{m}(\theta))^2 \text{tr} \tilde{\Sigma}_1 \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{T}^{-1}} \quad (2.4.42)$$

Recall that $a_{1n} \rightarrow -\theta \underline{m}(\theta)$, $\underline{m}(\theta) \rightarrow \underline{m}(\psi(\alpha)) = -\alpha^{-1}$, and $F^{\Sigma_{1P}} \rightarrow H$. Hence

$$\begin{aligned}
 d &:= \lim \frac{a_{1n}^2}{n} \text{tr} \tilde{\Sigma}_1 \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{T}^{-1} \\
 &= \underline{m}^2(\theta) \int \frac{ct^2}{(1 + t\underline{m}(\theta))^2} dH(t) \\
 &= \int \frac{ct^2}{(\alpha - t)^2} dH(t) \\
 &= 1 - \psi'(\alpha).
 \end{aligned} \tag{2.4.43}$$

By (2.4.41) and (2.4.43), and similar argument to (2.4.39), we obtain

$$\begin{aligned}
 (2.4.40) &\rightarrow 2(\mathbf{w}_1^\top \mathbf{T}^{-2} \mathbf{w}_1)^2 \left(\int_0^1 \frac{1}{1 - td} dt + \int_0^1 \frac{td}{(1 - td)^2} dt \right) \\
 &= \frac{2}{1 - d} = \frac{2}{\psi'(\alpha)}.
 \end{aligned} \tag{2.4.44}$$

Consequently, from (2.4.21)-(2.4.44), and $a_n \rightarrow \psi(\alpha)/\alpha$, we conclude that

$$\sum_{k=1}^n a_n^2 \mathbb{E}_{k-1}(nY_k^2) \rightarrow \frac{\psi^2(\alpha)}{\alpha^2} [(\gamma_4 - 3) \sum_{i=1}^p w_{1i}^4 + \frac{2}{\psi'(\alpha)}]. \tag{2.4.45}$$

Calculation of the mean. We next show that

$$\sqrt{n} \left(\mathbb{E} w_1^\top \mathbf{X} \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{w}_1 - \frac{\psi_n(\alpha)}{\alpha} \right) \rightarrow 0 \tag{2.4.46}$$

Let $\mathbf{X}^0 = (\mathbf{x}_1^0, \dots, \mathbf{x}_n^0)$ be a $p \times n$ matrix with entries consisting of *i.i.d.* Gaussian variables with mean 0 and variance $1/n$. Denote $\mathbf{A}^0 = \mathbf{I} - (\mathbf{X}^0)^\top \tilde{\Sigma}_1 \mathbf{X}^0$. We finish (2.4.46) by verifying that

$$\sqrt{n} \mathbb{E} \left(\mathbf{w}_1^\top \mathbf{X} \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{w}_1 - \mathbf{w}_1^\top \mathbf{X}^0 (\mathbf{A}^0)^{-1} (\mathbf{X}^0)^\top \mathbf{w}_1 \right) \rightarrow 0, \tag{2.4.47}$$

and

$$\sqrt{n} \mathbb{E} \left(\mathbf{w}_1^\top \mathbf{X}^0 (\mathbf{A}^0)^{-1} (\mathbf{X}^0)^\top \mathbf{w}_1 - \frac{\psi_n(\alpha)}{\alpha} \right) \rightarrow 0. \tag{2.4.48}$$

Let

$$\begin{aligned}
\mathbf{Z}_k^1 &= \sum_{i=1}^k \mathbf{x}_i \mathbf{e}_i^\top + \sum_{i=k+1}^n \mathbf{x}_i^0 \mathbf{e}_i^\top, \quad \mathbf{Z}_k^0 = \sum_{i=1}^{k-1} \mathbf{x}_i \mathbf{e}_i^\top + \sum_{i=k}^n \mathbf{x}_i^0 \mathbf{e}_i^\top, \\
\mathbf{Z}_k &= \sum_{i=1}^{k-1} \mathbf{x}_i \mathbf{e}_i^\top + \sum_{i=k+1}^n \mathbf{x}_i^0 \mathbf{e}_i^\top, \quad \hat{\mathbf{A}}_k^1 = \mathbf{I} - \frac{1}{n} (\mathbf{Z}_k^1)^\top \tilde{\Sigma}_1 \mathbf{Z}_k^1, \\
\hat{\mathbf{A}}_k^0 &= \mathbf{I} - \frac{1}{n} (\mathbf{Z}_k^0)^\top \tilde{\Sigma}_1 \mathbf{Z}_k^0, \quad \hat{\mathbf{A}}_k = \mathbf{I} - \frac{1}{n} (\mathbf{Z}_k)^\top \tilde{\Sigma}_1 \mathbf{Z}_k, \quad \hat{\mathbf{D}}_k^1 = \mathbf{I} - \tilde{\Sigma}_1 \mathbf{Z}_k^1 (\mathbf{Z}_k^1)^\top, \\
\hat{\mathbf{D}}_k^0 &= \mathbf{I} - \tilde{\Sigma}_1 \mathbf{Z}_k^0 (\mathbf{Z}_k^0)^\top, \quad \hat{\mathbf{D}}_k = \mathbf{I} - \tilde{\Sigma}_1 \mathbf{Z}_k \mathbf{Z}_k^\top, \quad \alpha_k^z = \frac{1}{1 - \mathbf{x}_k^\top \tilde{\Sigma}_1 (\hat{\mathbf{D}}_k^\top)^{-1} \mathbf{x}_k}, \\
\alpha_k^{0z} &= \frac{1}{1 - \mathbf{x}_k^{0\top} \tilde{\Sigma}_1 (\hat{\mathbf{D}}_k^\top)^{-1} \mathbf{x}_k^0}, \quad \bar{\alpha}_k^z = \frac{1}{1 - \frac{1}{n} \text{tr} \tilde{\Sigma}_1 (\hat{\mathbf{D}}_k^\top)^{-1}}.
\end{aligned} \tag{2.4.49}$$

We prove (2.4.47) first. To this end introduce the following notation

$$\begin{aligned}
& \sqrt{n} \mathbb{E} \left[\mathbf{w}_1^\top \mathbf{X} \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{w}_1 - \mathbf{w}_1^\top \mathbf{X}^0 (\mathbf{A}^0)^{-1} (\mathbf{X}^0)^\top \mathbf{w}_1 \right] \\
&= \sqrt{n} \sum_{k=1}^n \mathbb{E} \left[\mathbf{w}_1^\top \mathbf{Z}_k^1 (\hat{\mathbf{A}}_k^1)^{-1} (\mathbf{Z}_k^1)^\top \mathbf{w}_1 - \mathbf{w}_1^\top \mathbf{Z}_k^0 (\hat{\mathbf{A}}_k^0)^{-1} (\mathbf{Z}_k^0)^\top \mathbf{w}_1 \right] \\
&= \sqrt{n} \sum_{k=1}^n \mathbb{E} \left[\mathbf{w}_1^\top \mathbf{Z}_k^1 (\hat{\mathbf{A}}_k^1)^{-1} (\mathbf{Z}_k^1)^\top \mathbf{w}_1 - \mathbf{w}_1^\top \mathbf{Z}_k \hat{\mathbf{A}}_k^{-1} (\mathbf{Z}_k)^\top \mathbf{w}_1 \right] \\
&\quad - \sqrt{n} \sum_{k=1}^n \mathbb{E} \left[\mathbf{w}_1^\top \mathbf{Z}_k^1 (\hat{\mathbf{A}}_k^0)^{-1} (\mathbf{Z}_k^1)^\top \mathbf{w}_1 - \mathbf{w}_1^\top \mathbf{Z}_k \hat{\mathbf{A}}_k^{-1} (\mathbf{Z}_k)^\top \mathbf{w}_1 \right] \\
&= \sqrt{n} \sum_{k=1}^n \mathbb{E} \left[\alpha_k^z \mathbf{x}_k^\top (\hat{\mathbf{D}}_k)^{-1} \mathbf{w}_1 \mathbf{w}_1^\top (\hat{\mathbf{D}}_k^\top)^{-1} \mathbf{x}_k \right] - \sqrt{n} \sum_{k=1}^n \mathbb{E} \left[\alpha_k^{0z} (\mathbf{x}_k^0)^\top (\hat{\mathbf{D}}_k)^{-1} \mathbf{w}_1 \mathbf{w}_1^\top (\hat{\mathbf{D}}_k^\top)^{-1} \mathbf{x}_k^0 \right].
\end{aligned} \tag{2.4.50}$$

For the first summation above, we write

$$\begin{aligned}
\sqrt{n} \sum_{k=1}^n \mathbb{E} \left[\alpha_k^z \mathbf{x}_k^\top (\hat{\mathbf{D}}_k)^{-1} \mathbf{w}_1 \mathbf{w}_1^\top (\hat{\mathbf{D}}_k^\top)^{-1} \mathbf{x}_k \right] &= \sqrt{n} \sum_{k=1}^n \mathbb{E} \left[(\alpha_k^z - \bar{\alpha}_k^z) \mathbf{x}_k^\top (\hat{\mathbf{D}}_k)^{-1} \mathbf{w}_1 \mathbf{w}_1^\top (\hat{\mathbf{D}}_k^\top)^{-1} \mathbf{x}_k \right] \\
&\quad + \sqrt{n} \sum_{k=1}^n \mathbb{E} \left[\bar{\alpha}_k^z \mathbf{x}_k^\top (\hat{\mathbf{D}}_k)^{-1} \mathbf{w}_1 \mathbf{w}_1^\top (\hat{\mathbf{D}}_k^\top)^{-1} \mathbf{x}_k \right].
\end{aligned} \tag{2.4.51}$$

By the same strategy used in estimation of A_1 in (2.4.31), one can get

$$\sqrt{n} \sum_{k=1}^n \mathbb{E} \left[(\alpha_k^z - \bar{\alpha}_k^z) \mathbf{x}_k^\top (\hat{\mathbf{D}}_k)^{-1} \mathbf{w}_1 \mathbf{w}_1^\top (\hat{\mathbf{D}}_k^\top)^{-1} \mathbf{x}_k \right] = O(n^{-1/2}).$$

Thus, we have

$$\sqrt{n} \sum_{k=1}^n \mathbb{E} \left[\alpha_k^z \mathbf{x}_k^T (\hat{\mathbf{D}}_k)^{-1} \mathbf{w}_1 \mathbf{w}_1^T (\hat{\mathbf{D}}_k^T)^{-1} \mathbf{x}_k \right] = \frac{1}{\sqrt{n}} \sum_{k=1}^n \mathbb{E} \left[\bar{\alpha}_k^z \text{tr} \left((\hat{\mathbf{D}}_k)^{-1} \mathbf{w}_1 \mathbf{w}_1^T (\hat{\mathbf{D}}_k^T)^{-1} \right) \right] + O(n^{-1/2}). \quad (2.4.52)$$

Similarly, one can get

$$\sqrt{n} \sum_{k=1}^n \mathbb{E} \left[\alpha_k^{0z} (\mathbf{x}_k^0)^T (\hat{\mathbf{D}}_k)^{-1} \mathbf{w}_1 \mathbf{w}_1^T (\hat{\mathbf{D}}_k^T)^{-1} \mathbf{x}_k^0 \right] = \frac{1}{\sqrt{n}} \sum_{k=1}^n \mathbb{E} \left[\bar{\alpha}_k^z \text{tr} \left((\hat{\mathbf{D}}_k)^{-1} \mathbf{w}_1 \mathbf{w}_1^T (\hat{\mathbf{D}}_k^T)^{-1} \right) \right] + O(n^{-1/2}). \quad (2.4.53)$$

(2.4.47) follows from (2.4.50), (2.4.52) and (2.4.53).

Next we consider (2.4.48). By $\mathbf{w}_1^T \mathbf{U}_2 = 0$, we know that

$$\mathbb{E} \mathbf{w}_1^T \mathbf{X}^0 (\mathbf{A}^0)^{-1} (\mathbf{X}^0)^T \mathbf{w}_1 = \frac{1}{n} \mathbb{E} \text{tr} (\mathbf{A}^0)^{-1}. \quad (2.4.54)$$

Denote by $B_\epsilon(\theta)$ the ball with center θ and radius ϵ in the complex plane. By the strategy used in section 4 of [Bai and Silverstein \(2004\)](#), one can prove that

$$\sup_{z \in B_\epsilon(\theta)} \sqrt{n} |\mathbb{E} m_{F(\mathbf{X}^0)^T \Sigma_{1P} \mathbf{X}^0}(z) - \underline{m}_n(z)| \rightarrow 0.$$

Since $\psi_n(\alpha) \rightarrow \theta$ we have

$$\sqrt{n} \left(\mathbb{E} m_{F(\mathbf{X}^0)^T \Sigma_{1P} \mathbf{X}^0} \{ \psi_n(\alpha) \} - \underline{m}_n \{ \psi_n(\alpha) \} \right) = o(1).$$

This implies

$$\sqrt{n} \left(\frac{1}{n} \mathbb{E} \text{tr} (\mathbf{A}^0)^{-1} - \frac{\psi_n(\alpha)}{\alpha} \right) \rightarrow 0. \quad (2.4.55)$$

We conclude (2.4.48) from (2.4.55) and (2.4.54). □

2.5 Proof of Theorems 2.2.2 and 2.2.3

Proof of Theorem 2.2.2. First we give an outline of the proof, which is similar to the proof of Theorem 2.2 in [Cai et al. \(2017\)](#). As $n \rightarrow \infty$, $\lambda_k \xrightarrow{a.s.} \psi(\alpha_k)$, $1 \leq k \leq K$, which lie outside the support of F^{c_n, H_n} by Theorem 4.1 in [Bai and Yao \(2012\)](#). Moreover, no eigenvalue of $\mathbf{X}^T \Sigma_{1P} \mathbf{X}$ appears in a small interval that lies outside the support of F^{c_n, H_n} and that contains $\psi(\alpha_k)$ almost surely by [Bai et al. \(1998\)](#).

Therefore, $(\lambda_k \mathbf{I} - \mathbf{X}^\top \boldsymbol{\Sigma}_{1P} \mathbf{X})$ is invertible with probability 1. Conditional on this event, λ_k solves the determinant equation

$$\det(\boldsymbol{\Lambda}_S^{-1} - \mathbf{U}_1^\top \mathbf{X}(\lambda_k \mathbf{I} - \mathbf{X}^\top \boldsymbol{\Sigma}_{1P} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{U}_1) = 0. \quad (2.5.1)$$

To prove (2.2.8) we rewrite the normalized largest spike in terms of random quadratic forms (see (2.5.7) below) from the determinant equation. Then its limiting distribution is governed by the fluctuation of the random quadratic forms. The idea of proving (2.2.9) is similar.

We prove (2.2.8) first. Define $\mathbf{B}(x) = x\mathbf{I} - \mathbf{X}^\top \boldsymbol{\Sigma}_{1P} \mathbf{X}$ and $\chi_i = (\lambda_i - \theta_i)/\theta_i$. We only prove the central limit theorem for χ_1 , and the others can be similarly proved.

Using (2.4.1) we write

$$\mathbf{U}_1^\top \mathbf{X} \mathbf{B}^{-1}(\lambda_1) \mathbf{X}^\top \mathbf{U}_1 = \mathbf{U}_1^\top \mathbf{X} \mathbf{B}^{-1}(\theta_1) \mathbf{X}^\top \mathbf{U}_1 - \chi_1 \theta_1 \mathbf{U}_1^\top \mathbf{X} \mathbf{B}^{-1}(\lambda_1) \mathbf{B}^{-1}(\theta_1) \mathbf{X}^\top \mathbf{U}_1. \quad (2.5.2)$$

Denote by δ_{ij} the Kronecker delta function; i.e. $\delta_{ii} = 1$ and $\delta_{ij} = 0$ if $i \neq j$. By Theorem 2.2.1 we have

$$P(\max_{1 \leq i, j \leq K} |\frac{\delta_{ij}}{\alpha_1} - \mathbf{u}_i^\top \mathbf{X} \mathbf{B}^{-1}(\theta_1) \mathbf{X}^\top \mathbf{u}_j| \geq \frac{\epsilon}{\sqrt{n}}) \quad (2.5.3)$$

$$\leq K^2 P\left(\sqrt{n} |\frac{\delta_{ij}}{\alpha_1} - \mathbf{u}_i^\top \mathbf{X} \mathbf{B}^{-1}(\theta_1) \mathbf{X}^\top \mathbf{u}_j| \geq \epsilon\right) = O(1). \quad (2.5.4)$$

It follows that

$$\begin{aligned} & \boldsymbol{\Lambda}_S^{-1} - \mathbf{U}_1^\top \mathbf{X} \mathbf{B}^{-1}(\theta_1) \mathbf{X}^\top \mathbf{U}_1 \\ &= \begin{bmatrix} \hat{S}_n & O_p\left(\frac{1}{\sqrt{n}}\right) & \cdots & O_p\left(\frac{1}{\sqrt{n}}\right) \\ O_p\left(\frac{1}{\sqrt{n}}\right) & O_p(1) & \cdots & O_p\left(\frac{1}{\sqrt{n}}\right) \\ \cdots & \cdots & O_p(1) & O_p\left(\frac{1}{\sqrt{n}}\right) \\ O_p\left(\frac{1}{\sqrt{n}}\right) & \cdots & O_p\left(\frac{1}{\sqrt{n}}\right) & O_p(1) \end{bmatrix} \end{aligned} \quad (2.5.5)$$

where $\hat{S}_n = \frac{1}{\alpha_1} - \mathbf{u}_1^\top \mathbf{X} \mathbf{B}^{-1}(\theta_1) \mathbf{X}^\top \mathbf{u}_1$. Moreover, we have

$$\begin{aligned} & \chi_1 \theta_1 \mathbf{U}_1^\top \mathbf{X} \mathbf{B}^{-1}(\lambda_1) \mathbf{B}^{-1}(\theta_1) \mathbf{X}^\top \mathbf{U}_1 \\ &= \begin{bmatrix} \chi_1 \theta_1 R_1 & O_p(\chi_1) & \cdots & O_p(\chi_1) \\ O_p(\chi_1) & O_p(\chi_1) & \cdots & O_p(\chi_1) \\ \cdots & \cdots & O_p(\chi_1) & O_p(\chi_1) \\ O_p(\chi_1) & \cdots & O_p(\chi_1) & O_p(\chi_1) \end{bmatrix}, \end{aligned} \quad (2.5.6)$$

where $R_1 = \mathbf{u}_1^\top \mathbf{X} \mathbf{B}^{-1}(\lambda_1) \mathbf{B}^{-1}(\theta_1) \mathbf{X}^\top \mathbf{u}_1$. By (2.5.1), (2.5.2), (2.5.5), (2.5.6), the fact that $\chi_1 \xrightarrow{a.s.} 0$, and the Leibniz formula for a determinant, one can show that

$$\hat{S}_n + \chi_1 \theta_1 R_1 = o_p\left(\frac{1}{\sqrt{n}}\right),$$

which implies

$$\sqrt{n} \chi_1 = \frac{-\sqrt{n} \hat{S}_n}{\theta_1 R_1}. \quad (2.5.7)$$

By Theorem 2.2.1, $-\sqrt{n} \hat{S}_n / \theta_1$ converges in distribution to a normal distribution with mean 0 and variance $\{\alpha_1^2 \psi^2(\alpha_1)\}^{-1}[(\gamma_4 - 3) \sum_{i=1}^p u_{1i}^4 + \frac{2}{\psi'(\alpha)}]$. To handle R_1 in (2.5.7) by (2.4.1) we further expand it as

$$R_1 = \mathbf{u}_1^\top \mathbf{X} \mathbf{B}^{-2}(\theta_1) \mathbf{X}^\top \mathbf{u}_1 + (\theta_1 - \lambda_1) \mathbf{u}_1^\top \mathbf{X} \mathbf{B}^{-1}(\lambda_1) \mathbf{B}^{-2}(\theta_1) \mathbf{X}^\top \mathbf{u}_1. \quad (2.5.8)$$

It is easy to obtain

$$(\theta_1 - \lambda_1) \mathbf{u}_1^\top \mathbf{X} \mathbf{B}^{-1}(\lambda_1) \mathbf{B}^{-2}(\theta_1) \mathbf{X}^\top \mathbf{u}_1 \xrightarrow{i.p.} 0,$$

Next, we handle the first term of R_1 , that equals $\theta_1^{-2} \mathbf{u}_1^\top \mathbf{X} \mathbf{A}^{-2}(\theta_1) \mathbf{X}^\top \mathbf{u}_1$. Here $\mathbf{A}(\theta_1)$ is obtained from $\mathbf{A}$ in (2.4.7) with $\psi_n(\alpha)$ replaced by $\theta_1 = \psi_n(\alpha_1)$. Write

$$\begin{aligned} & \sqrt{n}(\mathbf{u}_1^\top \mathbf{X} \mathbf{A}^{-2}(\theta_1) \mathbf{X}^\top \mathbf{u}_1 - \mathbb{E} \mathbf{u}_1^\top \mathbf{X} \mathbf{A}^{-2}(\theta_1) \mathbf{X}^\top \mathbf{u}_1) \\ &= \sqrt{n} \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1})(I_1 + I_2 + I_3 + I_4) \end{aligned} \quad (2.5.9)$$

where

$$\begin{aligned} I_1 &= -\alpha_k \mathbf{x}_k^\top \mathbf{D}_k^{-1} \mathbf{u}_1 \mathbf{u}_1^\top (\mathbf{D}_k^\top)^{-1} \mathbf{x}_k, & I_2 &= \alpha_k \mathbf{x}_k^\top \mathbf{D}_k^{-1} \mathbf{u}_1 \mathbf{u}_1^\top (\mathbf{D}_k^\top)^{-2} \mathbf{x}_k, \\ I_3 &= \alpha_k \mathbf{x}_k^\top \mathbf{D}_k^{-2} \mathbf{u}_1 \mathbf{u}_1^\top (\mathbf{D}_k^\top)^{-1} \mathbf{x}_k, & I_4 &= -\alpha_k \mathbf{x}_k^\top \mathbf{D}_k^{-2} \tilde{\Sigma}_1 \mathbf{x}_k \times I_1. \end{aligned} \quad (2.5.10)$$

Here we should have put an indicator of the event as in the proof of Theorem 2.2.1 to ensure the existence of the expectation involving the inverse of the matrices of interest but we ignore it for simplicity. By the above decomposition as in the proof of Theorem 2.2.1 one can verify that

$$\mathbb{E}|\sqrt{n}(\mathbf{u}_1^\top \mathbf{X} \mathbf{A}^{-2}(\theta_1) \mathbf{X}^\top \mathbf{u}_1 - \mathbb{E} \mathbf{u}_1^\top \mathbf{X} \mathbf{A}^{-2}(\theta_1) \mathbf{X}^\top \mathbf{u}_1)|^2 = O(1). \quad (2.5.11)$$

Moreover by the interpolation method using in the calculating the asymptotic mean in the proof of Theorem 2.2.1 we conclude that

$$\mathbb{E} \mathbf{u}_1^\top \mathbf{X} \mathbf{A}^{-2}(\theta_1) \mathbf{X}^\top \mathbf{u}_1 - \mathbb{E} \mathbf{u}_1^\top \mathbf{X}^0 \{\mathbf{A}^0(\theta_1)\}^{-2} \mathbf{X}^{0\top} \mathbf{u}_1 = o(1). \quad (2.5.12)$$

Note that

$$\begin{aligned} \mathbb{E} \mathbf{u}_1^\top \mathbf{X}^0 \{\mathbf{A}^0(\theta_1)\}^{-2} \mathbf{X}^{0\top} \mathbf{u}_1 &= \frac{1}{n} \mathbb{E} \text{tr} \{\mathbf{A}^0(\theta_1)\}^{-2} = \mathbb{E} \int \frac{\theta_1^2}{(t - \theta_1)^2} dF^{\mathbf{X}^{0\top} \boldsymbol{\Sigma}_{1P} \mathbf{X}^0}(t) \\ &\rightarrow \int \frac{\psi(\alpha_1)^2}{\{t - \psi(\alpha_1)\}^2} dF^{c,H}(t) = \psi(\alpha_1)^2 \underline{m}'\{\psi(\alpha_1)\}. \end{aligned} \quad (2.5.13)$$

By the fact that $\psi(\alpha)$ is the inverse function of $\alpha : x \mapsto -1/\underline{m}(x)$ we have

$$\underline{m}'\{\psi(\alpha_1)\} = \frac{1}{\alpha_1^2 \psi'(\alpha_1)}. \quad (2.5.14)$$

It follows from (2.5.8), (2.5.11), (2.5.12) (2.5.13) and (2.5.14) that

$$R_1 \xrightarrow{i.p.} \frac{1}{\alpha_1^2 \psi'(\alpha_1)}. \quad (2.5.15)$$

One application of Slutsky's theorem implies that $\sqrt{n} \chi_1$ converges in distribution to a normal distribution with mean 0 and variance σ_1^2 .

To finish the proof of Theorem 2.2.2, we prove (2.2.9) next. Without loss of generality, we just deal with the joint distribution of χ_1 and χ_2 .

For $i = 1, 2$, let

$$\begin{aligned}\tilde{\Sigma}_1(\theta_i) &= \frac{\Sigma_{1P}}{\theta_i}, \quad \mathbf{D}_k^{(i)} = \mathbf{I} - \tilde{\Sigma}_1(\theta_i) \mathbf{X}_k \mathbf{X}_k^\top, \quad \mathbf{B}_k^{(i)} = (\mathbf{D}_k^{(i)})^{-1} \mathbf{u}_i \mathbf{u}_i^\top (\mathbf{D}_k^{(i)})^{-1}, \\ a_n(\theta_i) &= \frac{1}{1 - \frac{1}{n} \mathbb{E} \text{tr} \tilde{\Sigma}_1(\theta_i) (\mathbf{D}_k^{(i)})^{-1}}, \quad \delta_k(\theta_i) = \mathbf{x}_k^\top \mathbf{B}_k^{(i)} \mathbf{x}_k - \frac{1}{n} \text{tr} \mathbf{B}_k^{(i)}.\end{aligned}\tag{2.5.16}$$

For any constants c_1 and c_2 , as in (2.4.18), write

$$\begin{aligned}& c_1 \sqrt{n} \left\{ \mathbf{u}_1^\top \mathbf{X} \mathbf{B}^{-1}(\theta_1) \mathbf{X}^\top \mathbf{u}_1 - \mathbb{E} \mathbf{u}_1^\top \mathbf{X} \mathbf{B}^{-1}(\theta_1) \mathbf{X}^\top \mathbf{u}_1 \right\} + \\ & \quad c_2 \sqrt{n} \left\{ \mathbf{u}_2^\top \mathbf{X} \mathbf{B}^{-1}(\theta_2) \mathbf{X}^\top \mathbf{u}_2 - \mathbb{E} \mathbf{u}_2^\top \mathbf{X} \mathbf{B}^{-1}(\theta_2) \mathbf{X}^\top \mathbf{u}_2 \right\} \\ &= \sqrt{n} \sum_{k=1}^n \mathbb{E}_k \left\{ c_1 \theta_1^{-1} a_n(\theta_1) \delta_k(\theta_1) + c_2 \theta_2^{-1} a_n(\theta_2) \delta_k(\theta_2) \right\} + o_p(1).\end{aligned}\tag{2.5.17}$$

Condition (ii) of Lemma 2.4.1 can be verified as in (2.4.19). For condition (i), we write

$$\begin{aligned}& n \sum_{k=1}^n \mathbb{E}_{k-1} [\mathbb{E}_k \{ c_1 \theta_1^{-1} a_n(\theta_1) \delta_k(\theta_1) + c_2 \theta_2^{-1} a_n(\theta_2) \delta_k(\theta_2) \}]^2 \\ &= n \sum_{i=1}^2 c_i^2 \theta_i^{-2} a_n^2(\theta_i) \sum_{k=1}^n \mathbb{E}_{k-1} [\mathbb{E}_k \{ \delta_k(\theta_i) \}]^2 \\ & \quad + n \frac{c_1 c_2 a_n(\theta_1) a_n(\theta_2)}{\theta_1 \theta_2} \sum_{k=1}^n \mathbb{E}_{k-1} [\mathbb{E}_k \{ \delta_k(\theta_1) \} \mathbb{E}_k \{ \delta_k(\theta_2) \}].\end{aligned}\tag{2.5.18}$$

The first term corresponds to marginal variances which have been derived before. The second term above representing the covariance of λ_1 and λ_2 is slightly different from (2.4.21). By (2.4.3) we need to find the limit of

$$\frac{1}{n} (\gamma_4 - 3) \sum_{k=1}^n \sum_{i=1}^p \mathbb{E}_k (\mathbf{B}_k^{(1)})_{ii} \mathbb{E}_k (\mathbf{B}_k^{(2)})_{ii} + \frac{2}{n} \sum_{k=1}^n \text{tr} (\mathbb{E}_k \mathbf{B}_k^{(1)} \mathbb{E}_k \mathbf{B}_k^{(2)}).\tag{2.5.19}$$

Following the arguments similar to those from (2.4.21)-(2.4.39), we conclude that the first term of (2.5.19) equals $(\gamma_4 - 3) \sum_{k=1}^p u_{1k}^2 u_{2k}^2 + o_p(1)$. We claim that the

second term of (2.5.19) is negligible. Indeed the second term can be rewritten as

$$\begin{aligned} & \frac{2}{n} \sum_{k=1}^n \mathbb{E}_{k-1} \text{tr} \mathbb{E}_k ((\mathbf{D}_k^{(1)})^{-1} \mathbf{u}_1 \mathbf{u}_1^\top (\mathbf{D}_k^{(1)\top})^{-1}) \mathbb{E}_k ((\mathbf{D}_k^{(2)})^{-1} \mathbf{u}_2 \mathbf{u}_2^\top (\mathbf{D}_k^{(2)\top})^{-1}) \\ &= \frac{2}{n} \sum_{k=1}^n \mathbb{E}_{k-1} \left(\mathbf{u}_1^\top (\mathbf{D}_k^{(1)\top})^{-1} (\check{\mathbf{D}}_k^{(2)})^{-1} \mathbf{u}_2 \mathbf{u}_2^\top (\check{\mathbf{D}}_k^{(2)\top})^{-1} (\mathbf{D}_k^{(1)})^{-1} \mathbf{u}_1 \right), \end{aligned} \quad (2.5.20)$$

where $(\check{\mathbf{D}}_k^{(i)})^{-1}$ is defined similarly to $(\mathbf{D}_k^{(i)})^{-1}$ by $(\mathbf{x}_1, \dots, \mathbf{x}_{k-1}, \check{\mathbf{x}}_{k+1}, \dots, \check{\mathbf{x}}_n)$ and $\check{\mathbf{x}}_{k+1}, \dots, \check{\mathbf{x}}_n$ are i.i.d copies of $\mathbf{x}_{k+1}, \dots, \mathbf{x}_n$. Define $\mathbf{T}(\theta_i), i = 1, 2$ to be analogues of $\mathbf{T}$ in (2.4.29) with ψ_n replaced by $\theta_i, i = 1, 2$. The argument for (2.5.20) is parallel to (2.4.41)-(2.4.44), so we do not provide all details and list only the differences below. The term $\mathbf{u}_1^\top \mathbf{T}(\theta_1)^{-1} \mathbf{T}(\theta_2)^{-1} \mathbf{u}_2$ replaces $\mathbf{w}_1^\top \mathbf{T}^{-2} \mathbf{w}_1$ in (2.4.44). Since $\mathbf{u}_1^\top \mathbf{u}_2 = 0$, and $\mathbf{u}_1^\top \mathbf{U}_2 = \mathbf{u}_2^\top \mathbf{U}_2 = 0$, we can get $\mathbf{u}_1^\top \mathbf{T}(\theta_1)^{-1} \mathbf{T}(\theta_2)^{-1} \mathbf{u}_2 = 0$ by using (2.4.1). The term replacing $n^{-1} a_{1n}^2 \text{tr} \tilde{\Sigma}_1 \mathbf{T}^{-1} \tilde{\Sigma}_1 \mathbf{T}^{-1}$ in (2.4.41) and (2.4.42) is $n^{-1} a_{1n}(\theta_1) a_{1n}(\theta_2) \text{tr} \tilde{\Sigma}_1(\theta_1) (\mathbf{T}(\theta_1))^{-1} \tilde{\Sigma}_1(\theta_2) (\mathbf{T}(\theta_2))^{-1}$, that equals $1 - (\psi(\alpha_2) - \psi(\alpha_1))/(\alpha_2 - \alpha_1)$ asymptotically. These imply that the second term is $o_p(1)$ via (2.4.41). Together with the fact that for $i = 1$ or 2 , $a_n(\theta_i) \rightarrow \theta_i/\alpha_i$, the limit of the second term of (2.5.18) in probability is

$$c_1 c_2 \frac{\gamma_4 - 3}{\alpha_1 \alpha_2} \sum_{k=1}^p u_{1k}^2 u_{2k}^2. \quad (2.5.21)$$

Therefore, the asymptotic covariance of $\sqrt{n}\chi_1$ and $\sqrt{n}\chi_2$ is given by σ_{12} defined in assumption 5.

Proof of Theorem 2.2.3. We use the first m_1 spiked eigenvalues to illustrate the idea. Denote $(\lambda_j - \theta_1)/\theta_1$ by χ_{1j} for $j = 1, \dots, m_1$. Partition the matrix $\mathbf{U}_1 = (\mathbf{U}_{11}, \dots, \mathbf{U}_{1\mathcal{L}})$. Similar to (2.5.5) and (2.5.6), we write

$$\begin{aligned} & \Lambda_S^{-1} - \mathbf{U}_1^\top \mathbf{X} \mathbf{B}^{-1} (\lambda_j) \mathbf{X}^\top \mathbf{U}_1 \\ &= \begin{bmatrix} \tilde{\Sigma}_n & O_p(\frac{1}{\sqrt{n}}) + O_p(\chi_{1j}) & \cdots & O_p(\frac{1}{\sqrt{n}}) + O_p(\chi_{1j}) \\ O_p(\frac{1}{\sqrt{n}}) + O_p(\chi_{1j}) & O_p(1) & \cdots & \cdots \\ \cdots & \cdots & O_p(1) & O_p(\frac{1}{\sqrt{n}}) + O_p(\chi_{1j}) \\ O_p(\frac{1}{\sqrt{n}}) + O_p(\chi_{1j}) & \cdots & O_p(\frac{1}{\sqrt{n}}) + O_p(\chi_{1j}) & O_p(\frac{1}{\sqrt{n}}) + O_p(\chi_{1j}) \end{bmatrix} \end{aligned} \quad (2.5.22)$$

where

$$\begin{aligned}
 \tilde{S}_n &= \frac{\mathbf{I}}{\alpha_1} - \mathbf{U}_{11}^\top \mathbf{X} \mathbf{B}^{-1}(\lambda_j) \mathbf{X}^\top \mathbf{U}_{11} \\
 &= \frac{\mathbf{I}}{\alpha_1} - \mathbf{U}_{11}^\top \mathbf{X} \mathbf{B}^{-1}(\theta_1) \mathbf{X}^\top \mathbf{U}_{11} + \chi_{1j} \theta_1 \mathbf{U}_{11}^\top \mathbf{X} \mathbf{B}^{-2}(\theta_1) \mathbf{X}^\top \mathbf{U}_{11} + \\
 &\quad \chi_{1j}(\theta_1 - \lambda_j) \mathbf{U}_{11}^\top \mathbf{X} \mathbf{B}^{-1}(\lambda_1) \mathbf{B}^{-1}(\theta_1) \mathbf{X}^\top \mathbf{U}_{11} \\
 &= \frac{\mathbf{I}}{\alpha_1} - \mathbf{U}_{11}^\top \mathbf{X} \mathbf{B}^{-1}(\theta_1) \mathbf{X}^\top \mathbf{U}_{11} + \chi_{1j} \left(\frac{\psi(\alpha_1)}{\alpha_1^2 \psi'(\alpha_1)} \mathbf{I} + o_p(1) \right).
 \end{aligned} \tag{2.5.23}$$

Using arguments similar to the proof of Theorem 2.2.2 we have

$$\sqrt{n} \frac{\alpha_1^2 \psi'(\alpha_1)}{\psi(\alpha_1)} \left(\frac{\mathbf{I}}{\alpha_1} - \mathbf{U}_{11}^\top \mathbf{X} \mathbf{B}^{-1}(\theta_1) \mathbf{X}^\top \mathbf{U}_{11} \right) \xrightarrow{D} \mathbf{G}_1, \tag{2.5.24}$$

where $\mathbf{G}_1$ is a Gaussian random matrix with mean 0 and covariance structure characterized by (2.2.11). Therefore, if we multiply $n^{1/4}$ on the first m_1 rows and multiply $n^{1/4}$ on the first m_1 columns of (2.5.22), by using the Skorokhod strong representation similar to arguments in pages 464-465 of Bai et al. (2008) we see that $\sqrt{n} \chi_{11}, \dots, \sqrt{n} \chi_{1m_1}$ converge weakly to the joint distribution of eigenvalues of $\mathbf{G}_1$. Thus we conclude Theorem 2.2.3.

2.6 Proof of Theorem 2.2.4

Set $\theta_n = \psi_n(\alpha)$, where α is one of $\alpha_1, \dots, \alpha_K$. Let $\varphi_n = \sqrt{n}(\mathbf{w}_1^\top \mathbf{X} \mathbf{A}^{-1}(\theta_n) \mathbf{X}^\top \mathbf{w}_1 - \theta_n/\alpha)$, and $M_n(z) = p(m_{F\mathbf{S}_n}(z) - m_n(z))$ where $m_n(z)$ is the Stieltjes transformation of $F_{c_n, H_n}(x)$. Define the contour $\mathcal{C}_n$ as in page 561 of Bai and Silverstein (2004). We assume that the $M_n(z)$ has been truncated on the contour as in Bai and Silverstein (2004). To prove Theorem 2.2.4 it suffices to establish the following lemma.

Lemma 2.6.1. *Under the assumptions of Theorem 2.2.4, we have for $z \in \mathcal{C}_n$,*

$$(\varphi_n, M_n(z)) \rightarrow (\varphi, M(z)),$$

where φ is normal random variable with mean 0 and variance $(\gamma_4 - 3) \sum_{i=1}^p w_{1i}^4 + 2/\psi'(\alpha)$, independent of Gaussian process $M(z)$ with mean

$$\begin{aligned} \mathbb{E}M(z) = & \frac{c \int \underline{m}^3(z) t^2 dH(t) / (1 + t\underline{m}(z))^3}{1 - c \int \underline{m}^2(z) t^2 dH(t) / (1 + t\underline{m}(z))^2} dz \\ & + (\gamma_4 - 3) \int \frac{c \underline{m}^3(z) h_2(z)}{1 - c \int \underline{m}^2(z) t^2 dH(t) / (1 + t\underline{m}(z))^2} dz, \end{aligned} \quad (2.6.1)$$

and covariance

$$\begin{aligned} \text{Cov}(M(z_1), M(z_2)) = & \frac{2\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2} - \frac{1}{(z_1 - z_2)^2} \\ & + c(\gamma_4 - 3) \frac{d^2}{dz_1 dz_2} [\underline{m}(z_1) \underline{m}(z_2) h_1(z_1, z_2)]. \end{aligned} \quad (2.6.2)$$

Proof. Define $\mathbf{Q}(z) = \mathbf{S}_n - z\mathbf{I}$, $\mathbf{Q}_k(z) = \mathbf{Q}(z) - \Gamma \mathbf{x}_k \mathbf{x}_k^\top \Gamma^\top$, $\mathbf{C}_k = \Gamma^\top \mathbf{Q}_k^{-1}(z) \Gamma$,

$$\epsilon_k(z) = \mathbf{x}_k^\top \mathbf{C}_k \mathbf{x}_k - \frac{1}{n} \text{tr} \mathbf{C}_k, \quad b_n(z) = \frac{1}{1 + \frac{1}{n} \mathbb{E} \text{tr} \mathbf{C}_k}, \quad \beta_k(z) = \frac{1}{1 + \mathbf{x}_k^\top \mathbf{C}_k \mathbf{x}_k}, \quad \bar{\beta}_k = \frac{1}{1 + \frac{1}{n} \text{tr} \mathbf{C}_k}.$$

For any constants c_1 and c_2 , we consider the distribution of $c_1 \varphi_n + c_2 M_n(z)$. Its tightness is from that of $M_n(z)$ which was proved in [Bai and Silverstein \(2004\)](#). We truncate the entries of the matrix $\mathbf{X}$ in the linear spectral statistic at $\eta_n n^{1/2}$ as in [Bai and Silverstein \(2004\)](#) while we truncate the entries of the matrix $\mathbf{X}$ at $\eta_n n^{1/4}$ for the largest spike eigenvalues which is in Appendix. Hence it is equivalent to considering

$$c_1 \sqrt{n} (\mathbf{w}_1^\top \dot{\mathbf{X}} \dot{\mathbf{A}}^{-1} \dot{\mathbf{X}}^\top \mathbf{w}_1 - \frac{\theta_n}{\alpha}) + c_2 p(m_{F\check{\mathbf{s}}_n}(z) - m_n(z)) \quad (2.6.3)$$

where we use ‘.’ to mean that the entries of $\mathbf{X}$ is truncated at $\eta_n n^{1/4}$ and ‘..’ to mean that the entries of $\mathbf{X}$ is truncated at $\eta_n n^{1/2}$. We below use $\ddot{\beta}_k(z)$ to represent $\beta_k(z)$ but the underlying random variables are truncated at $\eta_n n^{1/2}$, and $\ddot{\epsilon}_k(z), \dot{\alpha}_k, \dot{\delta}_k$, etc, are similarly defined. Section 2 of [Bai and Silverstein \(2004\)](#) establishes that

$$p(m_{F\check{\mathbf{s}}_n}(z) - \mathbb{E} m_{F\check{\mathbf{s}}_n}(z)) = - \sum_{k=1}^n \mathbb{E}_k \frac{d}{dz} \ddot{\beta}_k(z) \ddot{\epsilon}_k(z) + o_p(1).$$

Thus, referring to (2.4.18), we have

$$\begin{aligned} & c_1 \sqrt{n} (\mathbf{w}_1^\top \dot{\mathbf{X}} \dot{A}^{-1} \dot{\mathbf{X}}^\top \mathbf{w}_1 - \mathbb{E} \mathbf{w}_1^\top \dot{\mathbf{X}} \dot{A}^{-1} \dot{\mathbf{X}}^\top \mathbf{w}_1) + c_2 p (m_{F\check{s}_n}(z) - \mathbb{E} m_{F\check{s}_n}(z)) \\ &= \sum_{k=1}^n \mathbb{E}_k (c_1 \sqrt{n} \dot{\alpha}_k \dot{\delta}_k - c_2 \frac{d}{dz} \ddot{\beta}_k(z) \ddot{\epsilon}_k(z)) + o_p(1). \end{aligned} \quad (2.6.4)$$

Note that both Theorem 2.2.1 and the main result in Bai and Silverstein (2004) use martingale to prove central limit theorem as in (2.6.4). Hence by Lemma 4.1 it suffices to consider the limit of the cross product

$$\sum_{k=1}^n \mathbb{E}_{k-1} (\mathbb{E}_k \sqrt{n} \dot{\alpha}_k \dot{\delta}_k) (\mathbb{E}_k \ddot{\beta}_k(z) \ddot{\epsilon}_k(z)). \quad (2.6.5)$$

Using (2.1) in Bai and Silverstein (2004), and (2.4.4) we conclude that

$$(2.6.5) = \sqrt{n} \dot{a}_n \ddot{b}_n(z) \sum_{k=1}^n \mathbb{E}_{k-1} [(\mathbb{E}_k \dot{\delta}_k) (\mathbb{E}_k \ddot{\epsilon}_k(z))] + o_p(1). \quad (2.6.6)$$

From (2.4.3), we have

$$|\mathbb{E}_{k-1} [(\mathbb{E}_k \dot{\delta}_k) (\mathbb{E}_k \ddot{\epsilon}_k(z))]| \leq \frac{C}{n^2} \left[\left| \sum_{i=1}^p (\mathbb{E}_k \dot{\mathbf{B}}_k)_{ii} (\mathbb{E}_k \ddot{\mathbf{C}}_k)_{ii} \right| + |tr(\mathbb{E}_k \dot{\mathbf{B}}_k \mathbb{E}_k \ddot{\mathbf{C}}_k)| \right]. \quad (2.6.7)$$

By the fact that $(\dot{\mathbf{B}}_k)_{ii} \geq 0$, and $|(\mathbb{E}_k \ddot{\mathbf{C}}_k)_{ii}| \leq \|\mathbb{E}_k \ddot{\mathbf{C}}_k\|$, we get

$$\left| \sum_{i=1}^p (\mathbb{E}_k \dot{\mathbf{B}}_k)_{ii} (\mathbb{E}_k \ddot{\mathbf{C}}_k)_{ii} \right| \leq \|\mathbb{E}_k \ddot{\mathbf{C}}_k\| |tr \mathbb{E}_k \dot{\mathbf{B}}_k|. \quad (2.6.8)$$

A simple application of Von Neumann's trace inequality yields that

$$|tr(\mathbb{E}_k \dot{\mathbf{B}}_k \mathbb{E}_k \ddot{\mathbf{C}}_k)| \leq \|\mathbb{E}_k \ddot{\mathbf{C}}_k\| |tr \mathbb{E}_k \dot{\mathbf{B}}_k|. \quad (2.6.9)$$

It follows from (2.6.7), (2.6.8) and (2.6.9) that

$$\begin{aligned} & \mathbb{E} |\sqrt{n} \dot{a}_n \ddot{b}_n(z) \sum_{k=1}^n \mathbb{E}_{k-1} (\mathbb{E}_k \dot{\delta}_k) (\mathbb{E}_k \ddot{\epsilon}_k(z))|^2 \\ & \leq C n^{-3} |\dot{a}_n \ddot{b}_n(z)|^2 \mathbb{E} \left| \sum_{k=1}^n \|\mathbb{E}_k \ddot{\mathbf{C}}_k\| |tr \mathbb{E}_k \dot{\mathbf{B}}_k| \right|^2 \\ & \leq C n^{-2} |\dot{a}_n \ddot{b}_n(z)|^2 \sum_{k=1}^n \mathbb{E} |tr \mathbb{E}_k \dot{\mathbf{B}}_k|^2 = O(n^{-1}). \end{aligned} \quad (2.6.10)$$

Then (2.6.6) and (2.6.10) imply (2.6.5) = $o_p(1)$. Lemma 2.3 in Bai and Silverstein (2004) further ensures

$$\sum_{k=1}^n \mathbb{E}_{k-1}(\mathbb{E}_k \sqrt{n} \dot{\alpha}_k \dot{\delta}_k) (\mathbb{E}_k \frac{d}{dz} \ddot{\beta}_k(z) \ddot{\epsilon}_k(z)) = o_p(1).$$

It follows that $\sqrt{n}(\mathbf{w}_1^\top \dot{\mathbf{X}} \dot{\mathbf{A}}^{-1} \dot{\mathbf{X}}^\top \mathbf{w}_1 - \mathbb{E} \mathbf{w}_1^\top \dot{\mathbf{X}} \dot{\mathbf{A}}^{-1} \dot{\mathbf{X}}^\top \mathbf{w}_1)$ and $p(m_{F\hat{\mathbf{s}}_n(z)} - \mathbb{E} m_{F\hat{\mathbf{s}}_n(z)})$ are asymptotically independent from Lemma 2.4.1 and Cramér-Wold's device. Consequently, φ_n and $M_n(z)$ are asymptotically independent. The marginal distribution is from our Theorem 2.2.2 and Theorem 1.4 of Pan and Zhou (2008). $\square$

Proof of Theorem 2.2.4: Without loss of generality, we consider the first spiked eigenvalue λ_1 . Recall that $\chi_1 = (\lambda_1 - \theta_1)/\theta_1$, and R_1 in (2.5.8). Asymptotic independence of $\sqrt{n}\chi_1$ and $L_p(\varphi)$ follows easily from the above lemma, (2.5.7) and (2.5.15).

2.7 Proof of Theorem 2.2.5

Proof. We take the estimation of $\sum_{k=1}^p u_{1k}^4$ as an illustration. Let $\mathbf{s}_1, \mathbf{s}_2$ be two deterministic $p \times 1$ vectors, and $\mathbf{u}_i, i = 1, \dots, p$ be the population eigenvectors of Σ . We need to clarify that although results in Mestre (2008) are given under absolutely continuous random entries with a bounded 8th moment, this assumption can be relaxed to random entries with a bounded 4th moment. Indeed, note that the conclusion (7) therein is a well known result, see Bai and Silverstein (2010), and the conclusion (9) therein is also true by appropriately modifying the proof of Bai et al. (2007) (just change one unit vector there to $\mathbf{s}_2$) assuming that the bounded 4th moment.

Some notations are introduced first:

$$\begin{aligned} \eta_1 &= \mathbf{s}_1^\top \mathbf{u}_1 \mathbf{u}_1^\top \mathbf{s}_2, \quad f_n(z) = \frac{z}{1 - c - cz m_n(z)}, \quad s_n = \mathbf{s}_1^\top (-z \underline{m}_n(z) \Sigma - zI)^{-1} \mathbf{s}_2, \\ g_n(z) &= \sum_{i=1}^p \frac{\mathbf{s}_1^\top \mathbf{u}_i \mathbf{u}_i^\top \mathbf{s}_2}{\mu_i - f_n(z)} f'_n(z) = s_n \frac{1 - c + cz^2 m'_n(z)}{1 - c - cz m_n(z)}, \\ \hat{s}_n(z) &= \mathbf{s}_1^\top (\mathbf{S}_n - zI)^{-1} \mathbf{s}_2, \quad \hat{m}_n(z) = m_{F\hat{\mathbf{s}}_n(z)}, \quad \hat{g}_n(z) = \hat{s}_n(z) \frac{1 - c + cz \hat{m}'_n(z)}{1 - c - cz \hat{m}_n(z)}. \end{aligned} \tag{2.7.1}$$

Let Υ be a negatively oriented contour described by the boundary of a rectangle

$$\{z \in \mathbb{C}, E_1 \leq \operatorname{Re}(z) \leq E_2, |\operatorname{Im}(z)| \leq b\},$$

where b is positive, and $[E_1, E_2]$ encloses $\psi(\alpha_1)$ and no other components of the support of F^{c_n, H_n} . Note that since α_1 is the largest distant spike, the associated support of F^{c_n, H_n} for this spike is separated from other components, and the length of the support decreases to 0 when n goes to infinity. So it is legitimate to choose Υ as above according to our assumptions.

From Proposition 3 in [Mestre \(2008\)](#)

$$\eta_1 = \frac{1}{2\pi i} \oint_{\Upsilon} g_n(z) dz.$$

Furthermore Proposition 4 in [Mestre \(2008\)](#) yields

$$\eta_1 - \hat{\eta}_1 \xrightarrow{a.s.} 0 \tag{2.7.2}$$

where $\hat{\eta}_1 := \frac{1}{2\pi i} \oint_{\Upsilon} \hat{g}_n(z) dz$. We below prove a stronger result in terms of the moment convergence (see [2.7.4](#) below) for our purpose.

Let Ξ be the event such that the contour Υ contains the largest sample eigenvalue but excludes all the remaining sample eigenvalues. The event Ξ holds with high probability as pointed out in the proof of Theorem [2.2.2](#). The next aim is to show that

$$\mathbb{E} \left| (\hat{\eta}_1 - \eta_1) I(\Xi) \right|^4 \leq C n^{-2}, \tag{2.7.3}$$

where the constant C is independent of the choice of the vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ as long as they are bounded in terms of the Euclidean norm. We first show that for any $z \in \Upsilon$

$$\mathbb{E}(|\hat{g}_n(z) - g_n(z)|^4 I(\Xi)) \leq C n^{-2}. \tag{2.7.4}$$

When there is no confusion we below omit the variable z from notations in [\(2.7.1\)](#). Set

$$g_z(u, v, w) = u \frac{1 - c + cz^2 w}{1 - c - czv},$$

which is an analytic function of u , v and w in a region excluding $v \neq (1 - c)/cz$. Due to the structure of the contour Υ by the exact separation arguments in [Bai et al. \(1998\)](#), $\hat{s}_n, \hat{m}_n, \hat{m}'_n$ are uniformly bounded for z on the contour Υ given the

event Ξ . Note that $|\underline{m}(z)|$ is bounded below away from zero on a bounded set by (5.1) in [Bai and Silverstein \(2004\)](#) (one may verify that $\underline{m}_n(z)$ has a similar property) and the end point E_1 involved in the contour Υ is positive. Hence from the relationship between $\underline{m}_n(z)$ and $m_n(z)$ we conclude that the absolute value of the denominator of $g(\hat{s}_n, \hat{m}_n, \hat{m}'_n)$ is bounded below away from zero on the contour. Hence a straightforward calculation indicates that

$$\begin{aligned} & |g_z(\hat{s}_n, \hat{m}_n, \hat{m}'_n) - g_z(s_n, m_n, m'_n)|I(\Xi) \\ & \leq C(|\hat{s}_n - s_n| + |\hat{m}_n - m_n| + |\hat{m}'_n - m'_n|)I(\Xi), \end{aligned} \quad (2.7.5)$$

where the constant C in the above inequality is independent of z . We next consider the 4th moment of the above three terms term by term.

By modifying the proof of Theorem 2 in [Bai and Silverstein \(2004\)](#) we conclude that $\mathbb{E}|\hat{s}_n - s_n|^4 I(\Xi) \leq C_1 n^{-2}$. The proof of Lemma 1.1 in [Bai and Silverstein \(2004\)](#) shows that $\mathbb{E}|\hat{m}_n - m_n|^4 I(\Xi) \leq C_2 n^{-2}$. Note that although [Bai et al. \(2007\)](#) and [Bai and Silverstein \(2004\)](#) established the above conclusions on the two horizontal lines of the contour Υ they still hold on the two vertical lines of the contour because $\min_i |\lambda_i - z| \geq C > 0$ on the two vertical lines given that the respective distances from $\varphi(\alpha_1)$ to the end points E_1 and E_2 involved in the contour can be positive. To handle the last term by using a martingale method it is not difficult to derive

$$\mathbb{E}|\hat{m}'_n - \mathbb{E}\hat{m}'_n|^4 \leq C n^{-2}. \quad (2.7.6)$$

Moreover note that $\sqrt{n}(\mathbb{E}\hat{m}_n - m_n)$ converges uniformly to zero on the contour by [Bai and Silverstein \(2004\)](#). Applying Vitali's convergence theorem, on the contour,

$$\sqrt{n}(\mathbb{E}\hat{m}'_n - m'_n) \rightarrow 0 \quad \text{uniformly for } z. \quad (2.7.7)$$

(2.7.6), together with (2.7.7), implies $\mathbb{E}|\hat{m}'_n - m'_n|^4 \leq C n^{-2}$. Consequently, we have (2.7.4). It follows from Jensen's inequality that

$$\mathbb{E} \left| \oint_{\Upsilon} (\hat{g}_n - g_n) I(\Xi) dz \right|^4 \leq C \oint_{\Upsilon} \mathbb{E}(|\hat{g}_n - g_n|^4 I(\Xi)) |dz| \leq C n^{-2}. \quad (2.7.8)$$

Therefore (2.7.3) is true.

Now we are at a position to show that our estimator for $\sum_{k=1}^p u_{1k}^4$ is weakly consistent. If we take $\mathbf{s}_1 = \mathbf{s}_2 = \mathbf{e}_1$ in the definition η_1 of (2.7.3) then one could

find the estimator $\hat{v}_{1j}$ of u_{1j}^2 from (2.7.2) for $j = 1, \dots, p$. It turns out that

$$\hat{v}_{1j} = \sum_{k=1}^p \theta_1(k) \hat{u}_{kj}^2, \quad (2.7.9)$$

and recall that $\theta_1(k)$ is defined in (2.2.19). One may refer to [Mestre \(2008\)](#) for details of calculating the explicit expression of $\hat{v}_{1j}$.

Using (2.7.3), we have

$$\begin{aligned} P(\max_j |\hat{v}_{1j} - u_{1j}^2| > \epsilon) &\leq P(\max_j |\hat{v}_{1j} - u_{1j}^2| > \epsilon, \Xi) + P(\Xi^c) \\ &\leq \sum_{i=1}^p P(|\hat{v}_{1i} - u_{1i}^2| > \epsilon, \Xi) + P(\Xi^c) = o(1). \end{aligned} \quad (2.7.10)$$

It follows that

$$\sum_{j=1}^p |\hat{v}_{1j}^2 - u_{1j}^4| \leq \max_j |\hat{v}_{1j} - u_{1j}^2| \sum_{j=1}^p (|\hat{v}_{1j}| + u_{1j}^2) = o_p(1),$$

where we use the fact $\sum_{j=1}^p |\hat{v}_{1j}| = O_p(1)$ to be proved below.

We finally verify the fact $\sum_{j=1}^p |\hat{v}_{1j}| = O_p(1)$. We consider the case $p/n < 1$ first. By analyzing equation (2.2.20) we conclude have the following interlacing relationship

$$\lambda_1 > \nu_1 > \lambda_2 > \nu_2 > \dots > \lambda_p > \nu_p > 0. \quad (2.7.11)$$

It follows from (2.2.19) that $\theta_1(1) > 0$ and $\theta_1(k) < 0$ for $k \geq 2$. Thus $|\hat{v}_{1j}| \leq \theta_1(1) \hat{u}_{1j}^2 - \sum_{k \geq 2}^p \theta_1(k) \hat{u}_{kj}^2$ by (2.7.9). Note that $\sum_{k=1}^p \theta_1(k) = 1$ by the definition of $\theta_1(k)$. Furthermore, we claim that $\theta_1(1)$ is bounded with high probability. Indeed, from (2.7.11) we have $\frac{\lambda_j}{\lambda_1 - \lambda_j} < \frac{\nu_{j-1}}{\lambda_1 - \nu_{j-1}}$ for $j = 2, \dots, p$. This further implies that $\theta_1(1) < 1 + \lambda_2/(\lambda_1 - \lambda_2)$ by the definition of $\theta_1(k)$. Since with high probability, λ_1 lies in a small interval containing $\psi(\alpha_1)$ while λ_2 lies outside the interval, we get $\theta_1(1)$ is bounded with high probability, as claimed. Thus $\sum_{j=1}^p |\hat{v}_{1j}| \leq \theta_1(1) - \sum_{k \geq 2}^p \theta_1(k) = 2\theta_1(1) - 1 = O_p(1)$. This is also true for $p/n \geq 1$ through similar arguments. Therefore we conclude that $\sum_{k=1}^p u_{1k}^4$ is consistently estimated by $\sum_{j=1}^p (\sum_{k=1}^p \theta_1(k) \hat{u}_{kj}^2)^2$.

□

2.8 Appendix

This appendix provides justifications for truncation and centralization on the entries of $\mathbf{X}$, and the proof of Lemma 2.4.2, Lemma 2.2.6.

2.8.1 Truncation and Centralization

Recall $x_{ij} = q_{ij}/\sqrt{n}$. By Assumption 1 we can select $\eta_n \rightarrow 0$ such that

$$\eta_n^{-4} \int_{|q_{11}| > \eta_n n^{1/4}} |q_{11}|^4 \rightarrow 0. \quad (2.8.1)$$

Let $\hat{x}_{ij} = \frac{1}{\sqrt{n}} q_{ij} I(|q_{ij}| \leq \eta_n n^{1/4}) - \mathbb{E} \frac{1}{\sqrt{n}} q_{ij} I(|q_{ij}| \leq \eta_n n^{1/4})$, $\tilde{x}_{ij} = x_{ij} - \hat{x}_{ij}$, $\hat{\mathbf{X}} = (\hat{x}_{ij})$, $\tilde{\mathbf{X}} = (\tilde{x}_{ij})$. Let $\hat{\mathbf{X}}_k = \hat{\mathbf{X}} - \hat{\mathbf{x}}_k \mathbf{e}_k^\top$, $\hat{\mathbf{X}}_{jk} = \hat{\mathbf{X}} - \hat{\mathbf{x}}_k \mathbf{e}_k^\top - \hat{\mathbf{x}}_j \mathbf{e}_j^\top$. Recall $\tilde{\Sigma}_1, \mathbf{D}, \mathbf{D}_k, \mathbf{D}_{jk}, \alpha_k, \alpha_{jk}$ defined in (2.4.7) and similarly define their respective analogues $\hat{\mathbf{D}}_k, \hat{\mathbf{D}}, \hat{\mathbf{D}}_{jk}, \hat{\alpha}_k, \hat{\alpha}_{jk}$ corresponding to the truncated version.

The first aim is to prove that

$$\sqrt{n} \left(\mathbf{w}_1^\top \mathbf{X} (\mathbf{I}_p - \mathbf{X}^\top \tilde{\Sigma}_1 \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{w}_1 - \mathbf{w}_1^\top \hat{\mathbf{X}} (\mathbf{I}_p - \hat{\mathbf{X}}^\top \tilde{\Sigma}_1 \hat{\mathbf{X}})^{-1} \hat{\mathbf{X}}^\top \mathbf{w}_1 \right) \xrightarrow{i.p.} 0. \quad (2.8.2)$$

With the help of (2.4.1) and (2.4.2), we write

$$\begin{aligned} & \sqrt{n} (\mathbf{w}_1^\top \mathbf{X} (\mathbf{I}_n - \mathbf{X}^\top \tilde{\Sigma}_1 \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{w}_1 - \mathbf{w}_1^\top \hat{\mathbf{X}} (\mathbf{I}_n - \hat{\mathbf{X}}^\top \tilde{\Sigma}_1 \hat{\mathbf{X}})^{-1} \hat{\mathbf{X}}^\top \mathbf{w}_1) \\ &= \sqrt{n} (\mathbf{w}_1^\top \mathbf{X} \mathbf{X}^\top \mathbf{D}^{-1} \mathbf{w}_1 - \mathbf{w}_1^\top \hat{\mathbf{X}} \hat{\mathbf{X}}^\top \hat{\mathbf{D}}^{-1} \mathbf{w}_1) \\ &= \sqrt{n} \mathbf{w}_1^\top (\mathbf{X} \mathbf{X}^\top - \hat{\mathbf{X}} \hat{\mathbf{X}}^\top) \mathbf{D}^{-1} \mathbf{w}_1 + \sqrt{n} \mathbf{w}_1^\top \hat{\mathbf{X}} \hat{\mathbf{X}}^\top (\mathbf{D}^{-1} - \hat{\mathbf{D}}^{-1}) \mathbf{w}_1 \\ &= \sqrt{n} \mathbf{w}_1^\top (\tilde{\mathbf{X}} \tilde{\mathbf{X}}^\top + \hat{\mathbf{X}} \hat{\mathbf{X}}^\top) \mathbf{D}^{-1} \mathbf{w}_1 + \sqrt{n} \mathbf{w}_1^\top \hat{\mathbf{X}} \hat{\mathbf{X}}^\top \hat{\mathbf{D}}^{-1} \tilde{\Sigma}_1 (\mathbf{X} \mathbf{X}^\top - \hat{\mathbf{X}} \hat{\mathbf{X}}^\top) \mathbf{D}^{-1} \mathbf{w}_1 \\ &= \sqrt{n} \sum_{i=1}^n \alpha_i \mathbf{w}_1^\top \tilde{\mathbf{x}}_i \mathbf{x}_i^\top \mathbf{D}_i^{-1} \mathbf{w}_1 + \sqrt{n} \sum_{i=1}^n \mathbf{w}_1^\top \hat{\mathbf{x}}_i \tilde{\mathbf{x}}_i^\top \mathbf{D}^{-1} \mathbf{w}_1 \\ &\quad + \sqrt{n} \sum_{i=1}^n \alpha_i \mathbf{w}_1^\top \hat{\mathbf{X}} \hat{\mathbf{X}}^\top \hat{\mathbf{D}}^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_i \mathbf{x}_i^\top \mathbf{D}_i^{-1} \mathbf{w}_1 + \sqrt{n} \sum_{i=1}^n \mathbf{w}_1^\top \hat{\mathbf{X}} \hat{\mathbf{X}}^\top \hat{\mathbf{D}}^{-1} \tilde{\Sigma}_1 \hat{\mathbf{x}}_i \tilde{\mathbf{x}}_i^\top \mathbf{D}^{-1} \mathbf{w}_1. \end{aligned} \quad (2.8.3)$$

For the first and second term above they can be handled by the same way as the ω_1 and ω_2 in (7.3) of Bai et al. (2007). For the third and fourth term, they are still similar but more complicated. We below only show the third term is $o_p(1)$ for illustration.

By extracting the i -th column of $\mathbf{X}$ respectively from $\hat{\mathbf{X}}\hat{\mathbf{X}}^\top$ and $\hat{\mathbf{D}}^{-1}$ we split the third term above into summations of the four terms: J_1, J_2, J_3, J_4 , where

$$\begin{aligned} J_1 &= \sqrt{n} \sum_{i=1}^n \alpha_i \mathbf{w}_1^\top \hat{\mathbf{X}}_i \hat{\mathbf{X}}_i^\top \hat{\mathbf{D}}_i^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_i \mathbf{x}_i^\top \mathbf{D}_i^{-1} \mathbf{w}_1, \\ J_2 &= \sqrt{n} \sum_{i=1}^n \alpha_i \mathbf{w}_1^\top \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^\top \hat{\mathbf{D}}_i^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_i \mathbf{x}_i^\top \mathbf{D}_i^{-1} \mathbf{w}_1, \\ J_3 &= \sqrt{n} \sum_{i=1}^n \alpha_i \hat{\alpha}_i \mathbf{w}_1^\top \hat{\mathbf{X}}_i \hat{\mathbf{X}}_i^\top \hat{\mathbf{D}}_i^{-1} \tilde{\Sigma}_1 \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^\top \hat{\mathbf{D}}_i^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_i \mathbf{x}_i^\top \mathbf{D}_i^{-1} \mathbf{w}_1, \\ J_4 &= \sqrt{n} \sum_{i=1}^n \alpha_i \hat{\alpha}_i \mathbf{w}_1^\top \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^\top \hat{\mathbf{D}}_i^{-1} \tilde{\Sigma}_1 \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^\top \hat{\mathbf{D}}_i^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_i \mathbf{x}_i^\top \mathbf{D}_i^{-1} \mathbf{w}_1. \end{aligned} \tag{2.8.4}$$

We only show $J_1 = o_p(1)$ below and others can be handled similarly. We can define a event similar to (2.4.11) which holds with high probability such that α_i and the spectral norm of $\hat{\mathbf{D}}_i^{-1}, \mathbf{D}_i^{-1}$ are bounded on this event. Therefore, we may ignore α_i inside J_1 when calculating the corresponding moment such as (2.8.5) below for simplicity. Moreover, an indicator function of such an event should be added inside the expectation to handle the inverse of the matrices of interest, but for notational simplicity, we omit it as in the earlier section. A direct calculation indicates that

$$\begin{aligned} \mathbb{E} \left| \sqrt{n} \sum_{i=1}^n \mathbf{w}_1^\top \hat{\mathbf{X}}_i \hat{\mathbf{X}}_i^\top \hat{\mathbf{D}}_i^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_i \mathbf{x}_i^\top \mathbf{D}_i^{-1} \mathbf{w}_1 \right|^2 &\leq n \sum_{i=1}^n \mathbb{E} \left| \mathbf{w}_1^\top \hat{\mathbf{X}}_i \hat{\mathbf{X}}_i^\top \hat{\mathbf{D}}_i^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_i \mathbf{x}_i^\top \mathbf{D}_i^{-1} \mathbf{w}_1 \right|^2 \\ &+ n \sum_{i_1 \neq i_2} \mathbb{E} \left(\mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1} \hat{\mathbf{X}}_{i_1}^\top \hat{\mathbf{D}}_{i_1}^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_{i_1} \mathbf{x}_{i_1}^\top \mathbf{D}_{i_1}^{-1} \mathbf{w}_1 \times \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_2} \hat{\mathbf{X}}_{i_2}^\top \hat{\mathbf{D}}_{i_2}^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_{i_2} \mathbf{x}_{i_2}^\top \mathbf{D}_{i_2}^{-1} \mathbf{w}_1 \right). \end{aligned} \tag{2.8.5}$$

It is easy to see the first term above converges to zero by the fact that

$$\mathbb{E} |\mathbf{x}_i^\top \mathbf{D}_i^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \hat{\mathbf{X}}_i \hat{\mathbf{X}}_i^\top \hat{\mathbf{D}}_i^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_i|^2 = o(n^{-2}).$$

For the second term, by further extracting $\mathbf{x}_{i_1}$ from $\hat{\mathbf{X}}_{i_2}$ and $\hat{\mathbf{D}}_{i_2}$ we have

$$\begin{aligned} \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_2} \hat{\mathbf{X}}_{i_2}^\top \hat{\mathbf{D}}_{i_2}^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_{i_2} \mathbf{x}_{i_2}^\top \mathbf{D}_{i_2}^{-1} \mathbf{w}_1 &= \mathbf{w}_1^\top (\hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top + \hat{\mathbf{x}}_{i_1} \hat{\mathbf{x}}_{i_1}^\top) \\ &\times (\hat{\mathbf{D}}_{i_1 i_2}^{-1} + \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \hat{\mathbf{x}}_{i_1} \hat{\mathbf{x}}_{i_1}^\top \hat{\mathbf{D}}_{i_2}^{-1}) \tilde{\Sigma}_1 \tilde{\mathbf{x}}_{i_2} \mathbf{x}_{i_2}^\top (\mathbf{D}_{i_1 i_2}^{-1} + \mathbf{D}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{x}_{i_1} \mathbf{x}_{i_1}^\top \mathbf{D}_{i_2}^{-1}) \mathbf{w}_1. \end{aligned} \tag{2.8.6}$$

Plugging the above expansion into the second term of (2.8.5) we have eight terms. We below consider only one of eight terms

$$\Delta_1 = n \sum_{i_1 \neq i_2} \mathbb{E} \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1} \hat{\mathbf{X}}_{i_1}^\top \hat{\mathbf{D}}_{i_1}^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_{i_1} \mathbf{x}_{i_1}^\top \mathbf{D}_{i_1}^{-1} \mathbf{w}_1 \times \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_{i_2} \mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1, \quad (2.8.7)$$

and the other seven terms can be estimated similarly.

By calculating expectation with respect to $\mathbf{x}_{i_1}$, we find

$$\Delta_1 = n \sum_{i_1 \neq i_2} \mathbb{E} \tilde{x}_{11}^2 \mathbb{E} \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1} \hat{\mathbf{X}}_{i_1}^\top \hat{\mathbf{D}}_{i_1}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1}^{-1} \mathbf{w}_1 \times \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_{i_2} \mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1. \quad (2.8.8)$$

Since we have $\mathbb{E} \tilde{x}_{11}^2 = o(n^{-3/2})$ by (2.8.1) we just need to show that the summation term in (2.8.8) is $o(n^{1/2})$. To prove this, we extract $\mathbf{x}_{i_2}$ from $\mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1} \hat{\mathbf{X}}_{i_1}^\top \hat{\mathbf{D}}_{i_1}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1}^{-1} \mathbf{w}_1$, and decompose Δ_1 into eight terms:

$$\begin{aligned} \Delta_{11} &= \sum_{i_1 \neq i_2} \mathbb{E} \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \times \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_{i_2} \mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1, \\ \Delta_{12} &= \sum_{i_1 \neq i_2} \mathbb{E} \mathbf{w}_1^\top \hat{\mathbf{x}}_{i_2} \hat{\mathbf{x}}_{i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \times \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_{i_2} \mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1, \\ \Delta_{13} &= \sum_{i_1 \neq i_2} \mathbb{E} \hat{\alpha}_{i_1 i_2} \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \hat{\mathbf{x}}_{i_2} \mathbf{x}_{i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_{i_2} \mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1, \\ \Delta_{14} &= \sum_{i_1 \neq i_2} \mathbb{E} \hat{\alpha}_{i_1 i_2} \mathbf{w}_1^\top \hat{\mathbf{x}}_{i_2} \hat{\mathbf{x}}_{i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \hat{\mathbf{x}}_{i_2} \mathbf{x}_{i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \times \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_{i_2} \mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1, \\ \Delta_{15} &= \sum_{i_1 \neq i_2} \mathbb{E} \alpha_{i_1 i_2} \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{x}_{i_2} \mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \\ &\quad \times \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_{i_2} \mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1, \\ \Delta_{16} &= \sum_{i_1 \neq i_2} \mathbb{E} \alpha_{i_1 i_2} \mathbf{w}_1^\top \hat{\mathbf{x}}_{i_2} \hat{\mathbf{x}}_{i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{x}_{i_2} \mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \times \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_{i_2} \mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1, \\ \Delta_{17} &= \sum_{i_1 \neq i_2} \mathbb{E} \hat{\alpha}_{i_1 i_2} \alpha_{i_1 i_2} \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \hat{\mathbf{x}}_{i_2} \hat{\mathbf{x}}_{i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{x}_{i_2} \mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \\ &\quad \times \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_{i_2} \mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1, \\ \Delta_{18} &= \sum_{i_1 \neq i_2} \mathbb{E} \hat{\alpha}_{i_1 i_2} \alpha_{i_1 i_2} \mathbf{w}_1^\top \hat{\mathbf{x}}_{i_2} \hat{\mathbf{x}}_{i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \hat{\mathbf{x}}_{i_2} \hat{\mathbf{x}}_{i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{x}_{i_2} \mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \\ &\quad \times \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_{i_2} \mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1. \end{aligned}$$

Next we consider above eight terms one by one.

For Δ_{11} , by calculating expectation with respect to $\mathbf{x}_{i_2}$, we have

$$|\Delta_{11}| = \left| \sum_{i_1 \neq i_2} \mathbb{E} \tilde{x}_{11}^2 \mathbb{E} \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \times \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \right| = o(n^{1/2}).$$

We conclude that $|\Delta_{12}| = o(1)$ from

$$\begin{aligned} \mathbb{E} |\mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \hat{\mathbf{x}}_{i_2}|^2 &= O(n^{-2}), \\ \mathbb{E} |\hat{\mathbf{x}}_{i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \times \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \hat{\mathbf{x}}_{i_2}|^2 &= o(n^{-2}). \end{aligned} \quad (2.8.9)$$

For Δ_{13} , we find

$$\begin{aligned} |\Delta_{13}| &\leq \sum_{i_1 \neq i_2} (\mathbb{E} |\mathbf{x}_{i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \times \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \hat{\mathbf{x}}_{i_2}|^2)^{1/2} \\ &\quad \times (\mathbb{E} |\mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \hat{\mathbf{x}}_{i_2}|^2)^{1/2} \\ &= o(1). \end{aligned} \quad (2.8.10)$$

It follows from

$$\begin{aligned} \mathbb{E} |\hat{\mathbf{x}}_{i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \hat{\mathbf{x}}_{i_2}|^4 &= O(1), \\ \mathbb{E} |\mathbf{x}_{i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \times \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \hat{\mathbf{x}}_{i_2}|^2 &= o(n^{-2}), \\ \mathbb{E} |\mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \hat{\mathbf{x}}_{i_2}|^4 &= o(n^{-3}). \end{aligned} \quad (2.8.11)$$

that $|\Delta_{14}| = o(n^{1/4})$. By the same argument as (2.8.10) Δ_{15} is of order $o(1)$. For Δ_{16} , one can verify that

$$\mathbb{E} |\hat{\mathbf{x}}_{i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{x}_{i_2}|^4 = O(1). \quad (2.8.12)$$

This, together with (2.8.11), implies

$$\begin{aligned} |\Delta_{16}| &\leq \sum_{i_1 \neq i_2} (\mathbb{E} |\hat{\mathbf{x}}_{i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{x}_{i_2}|^4)^{1/4} \times (\mathbb{E} |\mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \hat{\mathbf{x}}_{i_2}|^2)^{1/2} \\ &\quad \times (\mathbb{E} |\mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \hat{\mathbf{x}}_{i_2}|^4)^{1/4} \\ &= o(n^{1/4}). \end{aligned} \quad (2.8.13)$$

For Δ_{17} and Δ_{18} , we have

$$\begin{aligned}
|\Delta_{17}| &\leq \sum_{i_1 \neq i_2} (\mathbb{E} |\hat{\mathbf{x}}_{i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{x}_{i_2}|^4)^{1/4} \times (\mathbb{E} |\mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \tilde{\mathbf{x}}_{i_2}|^2)^{1/2} \\
&\quad \times (\mathbb{E} |\mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \hat{\mathbf{X}}_{i_1 i_2} \hat{\mathbf{X}}_{i_1 i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \hat{\mathbf{x}}_{i_2}|^4)^{1/4} \\
&= o(n^{1/4}), \\
|\Delta_{18}| &\leq \sum_{i_1 \neq i_2} (\mathbb{E} |\hat{\mathbf{x}}_{i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \hat{\mathbf{x}}_{i_2}|^4)^{1/4} \times (\mathbb{E} |\hat{\mathbf{x}}_{i_2}^\top \hat{\mathbf{D}}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{D}_{i_1 i_2}^{-1} \tilde{\Sigma}_1 \mathbf{x}_{i_2}|^4)^{1/4} \\
&\quad \times (\mathbb{E} |\mathbf{x}_{i_2}^\top \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \tilde{\mathbf{x}}_{i_2}|^4)^{1/4} \times (\mathbb{E} |\mathbf{x}_{i_2}^\top H \mathbf{D}_{i_1 i_2}^{-1} \mathbf{w}_1 \mathbf{w}_1^\top \hat{\mathbf{x}}_{i_2}|^4)^{1/4} \\
&= o(n^{1/4}).
\end{aligned} \tag{2.8.14}$$

Combining above arguments we conclude that $\Delta_1 \rightarrow 0$ and therefore the third term in (2.8.3) is $o_p(1)$.

Let $\check{x}_{ij} = \frac{\hat{x}_{ij}}{\sqrt{n}\sigma_n}$, with $\sigma_n^2 = \mathbb{E}|\hat{x}_{ij}|^2$, $\check{\mathbf{X}} = (\check{x}_{ij})$. We have where we use $1 - n\sigma_n^2 = \mathbb{E}q_{11}^2 I(|q_{11}| > \eta_n n^{1/4}) + \mathbb{E}^2 q_{11} I(|q_{11}| > \eta_n n^{1/4}) = o(n^{-1/2})$.

Up to now, we conclude that truncation at $\eta_n n^{1/4}$ and centralization do not influence the limiting distribution of the statistics in Theorem 2.2.1.

2.8.2 Proof of Lemma 2.2.6

Proof. For simplicity, we omit the index of population, thus denote the population covariance by Σ and the sample covariance by $\mathbf{S} = \frac{1}{n} \mathbf{Y} \mathbf{Y}^\top = \Sigma^{\frac{1}{2}} \mathbf{X} \mathbf{X}^\top \Sigma^{\frac{1}{2}}$. To conclude (2.2.32) it suffices to show the following two facts:

$$\max_{1 \leq i \leq p} |\mathbf{S}_{ii} - \Sigma_{ii}| \xrightarrow{D} 0, \tag{2.8.15}$$

$$\frac{\sum_{i=1}^p |\mathbf{S}_{ii} - \Sigma_{ii}|}{p} = O_p(1). \tag{2.8.16}$$

The proof of (2.8.16) is trivial, since

$$\frac{\sum_{i=1}^p |\mathbf{S}_{ii} - \Sigma_{ii}|}{p} = \frac{\text{tr}(\mathbf{S}) - \text{tr}(\Sigma)}{p} = O_p(1).$$

To derive (2.8.15) we first truncate and centralize $\mathbf{X}$. Select $\eta_n \rightarrow 0$ such that $\eta_n^{-4} \int_{|q_{11}| > \eta_n n^{1/2}} |q_{11}|^4 \rightarrow 0$. Let $\hat{\mathbf{X}}' = (\hat{x}'_{ij})_{p \times n}$ where $\hat{x}'_{ij} = \frac{1}{\sqrt{n}} q_{ij} I(|q_{ij}| \leq \eta_n n^{1/2})$, and $\hat{\mathbf{S}}' = \Sigma^{1/2} \hat{\mathbf{X}}' (\hat{\mathbf{X}}')^\top \Sigma^{1/2}$. Then define the truncated and centralized matrix $\check{\mathbf{X}}' = (\check{x}'_{ij})_{p \times n}$, where $\check{x}'_{ij} = (\hat{x}'_{ij} - \mathbb{E} \hat{x}'_{ij}) / (\sqrt{n} \sigma'_n)$, with $\sigma'_n = \mathbb{E}^{1/2} |\hat{x}'_{ij} - \mathbb{E} \hat{x}'_{ij}|^2$, and

$\check{\mathbf{S}}' = \Sigma^{1/2} \check{\mathbf{X}}' (\check{\mathbf{X}}')^\top \Sigma^{1/2}$. We have

$$P(\mathbf{S} \neq \hat{\mathbf{S}}') \leq pnP(|q_{11}| \geq \eta_n \sqrt{n}) = o(1).$$

It follows that for any $\epsilon > 0$,

$$P(\max_{1 \leq i \leq p} |\mathbf{S}_{ii} - \hat{\mathbf{S}}'_{ii}| > \epsilon) = o(1)$$

For any deterministic unit vector $\mathbf{s}_1$, we have

$$\begin{aligned} & \mathbb{E} |\mathbf{s}_1^\top \hat{\mathbf{S}}' \mathbf{s}_1 - \mathbf{s}_1^\top \check{\mathbf{S}}' \mathbf{s}_1| \\ & \leq (\mathbb{E} |\mathbf{s}_1^\top \Sigma^{1/2} (\hat{\mathbf{X}}' - \check{\mathbf{X}}') (\hat{\mathbf{X}}')^\top \Sigma^{1/2} \mathbf{s}_1| + \mathbb{E} |\mathbf{s}_1^\top \Sigma^{1/2} \check{\mathbf{X}}' (\hat{\mathbf{X}}' - \check{\mathbf{X}}')^\top \Sigma^{1/2} \mathbf{s}_1|) \\ & \leq C \left(\frac{|\sqrt{n}\sigma'_n - 1|}{\sqrt{n}\sigma'_n} + \frac{|\mathbb{E} \hat{x}'_{11}|}{\sigma'_n} \right) = o(n^{-1}) \end{aligned}$$

where the last step uses two facts that $|n\sigma_n'^2 - 1| \leq 2 \int_{\{|q_{11}| \geq \eta_n n^{1/2}\}} |q_{11}|^2 = o(n^{-1})$, and that $|\mathbb{E} \hat{x}'_{11}| \leq \int_{\eta_n n^{1/2}}^\infty P(|q_{11}| \geq t) dt = o(n^{-3/2})$. From above estimates, we obtain for any $\epsilon > 0$,

$$P(\max_{1 \leq i \leq p} |\hat{\mathbf{S}}'_{ii} - \check{\mathbf{S}}'_{ii}| > \epsilon) \leq \sum_{i=1}^p \frac{\mathbb{E} |\mathbf{e}_i^\top \hat{\mathbf{S}}' \mathbf{e}_i - \mathbf{e}_i^\top \check{\mathbf{S}}' \mathbf{e}_i|}{\epsilon} = o(1).$$

Therefore we below assume the underlying variables are truncated at $\eta_n n^{1/2}$ and centralized.

Then for any deterministic unit vector $\mathbf{s}_1$, using Lemma 2.1 in [Bai and Silverstein \(2004\)](#), and the Burkholder's inequality, we have

$$\begin{aligned} & \mathbb{E} |\mathbf{s}_1^\top \Sigma^{1/2} \mathbf{X} \mathbf{X}^\top \Sigma^{1/2} \mathbf{s}_1 - \mathbf{s}_1^\top \Sigma \mathbf{s}_1|^4 \\ & \leq \mathbb{E} \left| \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) \mathbf{s}_1^\top \Sigma^{1/2} \mathbf{X} \mathbf{X}^\top \Sigma^{1/2} \mathbf{s}_1 \right|^4 \\ & \leq \mathbb{E} \left| \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) (\mathbf{s}_1^\top \Sigma^{1/2} \mathbf{x}_k \mathbf{x}_k^\top \Sigma^{1/2} \mathbf{s}_1) \right|^4 \\ & = n \sum_{k=1}^n \mathbb{E} |\mathbf{s}_1^\top \Sigma^{1/2} \mathbf{x}_k \mathbf{x}_k^\top \Sigma^{1/2} \mathbf{s}_1|^4 = o(n^{-1}). \end{aligned} \tag{2.8.17}$$

This implies (2.8.15) and hence (2.2.32).

Next we show that (2.2.33) holds. The law of large numbers ensures that

$$\frac{1}{pn} \sum_{i=1}^n (\mathbf{y}_i^\top \boldsymbol{\Sigma}_1 \mathbf{y}_i - \text{tr} \boldsymbol{\Sigma}_1)^2 - \mathfrak{M} \xrightarrow{i.p.} 0. \quad (2.8.18)$$

Note that

$$\begin{aligned} & \left| \frac{1}{pn} \sum_{i=1}^n (\mathbf{y}_i^\top \boldsymbol{\Sigma}_1 \mathbf{y}_i - \text{tr} \mathbf{S}_1)^2 - \frac{1}{pn} \sum_{i=1}^n (\mathbf{y}_i^\top \boldsymbol{\Sigma}_1 \mathbf{y}_i - \text{tr} \boldsymbol{\Sigma}_1)^2 \right| \\ & \leq \frac{2}{pn} \sum_{i=1}^n |(\mathbf{y}_i^\top \boldsymbol{\Sigma}_1 \mathbf{y}_i - \text{tr} \boldsymbol{\Sigma}_1)(\text{tr} \mathbf{S}_1 - \text{tr} \boldsymbol{\Sigma}_1)| + \frac{1}{p} |\text{tr} \mathbf{S}_1 - \text{tr} \boldsymbol{\Sigma}_1|^2. \end{aligned} \quad (2.8.19)$$

The second term is $o_p(1)$ by the law of large numbers. Also from the law of large numbers,

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{p}} |\mathbf{y}_i^\top \boldsymbol{\Sigma}_1 \mathbf{y}_i - \text{tr} \boldsymbol{\Sigma}_1| - \mathbb{E} \frac{1}{\sqrt{p}} |\mathbf{y}_i^\top \boldsymbol{\Sigma}_1 \mathbf{y}_i - \text{tr} \boldsymbol{\Sigma}_1| \xrightarrow{i.p.} 0. \quad (2.8.20)$$

Moreover (2.4.3) implies that $\mathbb{E} \frac{1}{\sqrt{p}} |\mathbf{y}_i^\top \boldsymbol{\Sigma}_1 \mathbf{y}_i - \text{tr} \boldsymbol{\Sigma}_1|$ is bounded and from Lemma 2.4.1 $\text{tr} \mathbf{S}_1 - \text{tr} \boldsymbol{\Sigma}_1$ is asymptotically normal. These imply that the first term in (2.8.19) is also $o_p(1)$. Thus we have

$$\left| \frac{1}{pn} \sum_{i=1}^n (\mathbf{y}_i^\top \boldsymbol{\Sigma}_1 \mathbf{y}_i - \text{tr} \mathbf{S}_1)^2 - \frac{1}{pn} \sum_{i=1}^n (\mathbf{y}_i^\top \boldsymbol{\Sigma}_1 \mathbf{y}_i - \text{tr} \boldsymbol{\Sigma}_1)^2 \right| \xrightarrow{i.p.} 0. \quad (2.8.21)$$

From (2.8.18) and (2.8.21), we conclude (2.2.33). $\square$

2.8.3 Proof of Lemma 2.4.2

First we show that $\mathcal{B}_1$ holds with high probability. By Lemma 3.1 in Bai and Yao (2012), $\psi(\alpha)$ lies outside the support of $F^{c,H}$ and F^{c_n,H_n} . Thus we can select a small interval $[a, b]$ containing $\psi(\alpha)$ and $\psi_n(\alpha)$ for sufficiently large n , such that this interval lies outside the support of $F^{c,H}$ and F^{c_n,H_n} . By carefully checking on the proofs of Bai et al. (1998) one can show that under the truncation at $\eta_n n^{1/4}$ no eigenvalues of $\mathbf{X}^\top \boldsymbol{\Sigma}_{1P} \mathbf{X}$ lies in $[a, b]$ with high probability. Therefore, $\mathcal{B}_1$ holds with high probability. Actually, we note that proof of Lemma C.3 in Jiang and Bai (2018) also uses this conclusion. Following a similar argument $\mathcal{B}_{1k}$ and $\mathcal{B}_{1jk}$ hold with high probability.

We next show that the event $\mathcal{B}_{2k}$ and $\mathcal{B}_{3k}$ hold with high probability. For any small positive constant ϵ , and large positive constant l , using (2.4.4), we find

$$P \left(\left| \mathbf{x}_k^\top \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1} \mathbf{x}_k I(\mathcal{B}_{1k}) - \frac{1}{n} \text{tr} \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1} I(\mathcal{B}_{1k}) \right| > \epsilon \right) \leq n^{-l}. \quad (2.8.22)$$

By Burkholder's inequality

$$\begin{aligned} & \mathbb{E} \left| \text{tr} \frac{1}{n} \tilde{\Sigma}_1 (\mathbf{D}_1^\top)^{-1} I(\mathcal{B}_{11}) - \mathbb{E} \text{tr} \frac{1}{n} \tilde{\Sigma}_1 (\mathbf{D}_1^\top)^{-1} I(\mathcal{B}_{11}) \right|^p \\ & \leq \frac{1}{n^p} \mathbb{E} \left| \sum_{j=2}^n (\mathbb{E}_j - \mathbb{E}_{j-1}) [\text{tr} \tilde{\Sigma}_1 (\mathbf{D}_1^\top)^{-1} I(\mathcal{B}_{11}) - \text{tr} \tilde{\Sigma}_1 (\mathbf{D}_{1j}^\top)^{-1} I(\mathcal{B}_{11j})] \right|^p \\ & \leq \frac{C_p}{n^p} \mathbb{E} \left(\sum_{j=2}^n |(\mathbb{E}_j - \mathbb{E}_{j-1}) \mathbf{x}_j^\top \tilde{\Sigma}_1 (\mathbf{D}_1^\top)^{-1} \tilde{\Sigma}_1 (\mathbf{D}_{1j}^\top)^{-1} \mathbf{x}_j I(\mathcal{B}_{11} \mathcal{B}_{11j})|^2 \right)^{\frac{p}{2}} \\ & \leq \frac{C_p}{n^{p/2+1}} \sum_{j=2}^n \mathbb{E} |(\mathbb{E}_j - \mathbb{E}_{j-1}) \mathbf{x}_j^\top \tilde{\Sigma}_1 (\mathbf{D}_1^\top)^{-1} \tilde{\Sigma}_1 (\mathbf{D}_{1j}^\top)^{-1} \mathbf{x}_j I(\mathcal{B}_{11} \mathcal{B}_{11j})|^p \\ & \leq \frac{C_p}{n^{p/2}} \mathbb{E} |\mathbf{x}_j^\top \tilde{\Sigma}_1 (\mathbf{D}_1^\top)^{-1} \tilde{\Sigma}_1 (\mathbf{D}_{1j}^\top)^{-1} \mathbf{x}_j I(\mathcal{B}_{11} \mathcal{B}_{11j})|^p \\ & \leq \frac{C_p}{n^{p/2}} \mathbb{E} |\mathbf{x}_j^\top \mathbf{x}_j|^p = O(n^{-\frac{p}{2}}), \end{aligned}$$

where we use the fact that $\mathcal{B}_{11} \subseteq \mathcal{B}_{11j}$. This implies that

$$P \left(\left| \frac{1}{n} \text{tr} \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1} I(\mathcal{B}_{1k}) - \mathbb{E} \frac{1}{n} \text{tr} \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1} I(\mathcal{B}_{1k}) \right| > \epsilon \right) \leq n^{-l}. \quad (2.8.23)$$

We claim that

$$\mathbb{E} \left(\frac{1}{n} \text{tr} \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1} I(\mathcal{B}_{1k}) \right) \rightarrow 1 + \frac{1}{\theta \underline{m}(\theta)}, \text{ as } n \rightarrow \infty, \quad (2.8.24)$$

(to be proved later). We conclude from (2.8.22), (2.8.23) and (2.8.24) that

$$P \left(\left| \mathbf{x}_k^\top \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1} \mathbf{x}_k I(\mathcal{B}_{1k}) - \left(1 + \frac{1}{\theta \underline{m}(\theta)} \right) \right| > \epsilon \right) = o(n^{-l}), \quad (2.8.25)$$

and

$$P \left(\left| \frac{1}{n} \text{tr} \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1} I(\mathcal{B}_{1k}) - \left(1 + \frac{1}{\theta \underline{m}(\theta)} \right) \right| > \epsilon \right) = o(n^{-l}). \quad (2.8.26)$$

Thus the events $\mathcal{B}_{2k}$ and $\mathcal{B}_{3k}$ hold with high probability. Using the similar arguments we conclude that $\mathcal{B}_{2jk}$ and $\mathcal{B}_{3jk}$ hold with high probability. To conclude, the event $\mathcal{B}$ holds with high probability.

Finally we show (2.8.24). Lemma 3.3 in [Bai and Silverstein \(1999\)](#), together with the fact $\mathbf{x}_k^\top \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1} \mathbf{x}_k = \mathbf{x}_k^\top \tilde{\Sigma}_1^{1/2} (\mathbf{I} - \tilde{\Sigma}_1^{1/2} \mathbf{X}_k \mathbf{X}_k^* \tilde{\Sigma}_1^{1/2})^{-1} \tilde{\Sigma}_1^{1/2} \mathbf{x}_k$, ensures that

$$\mathbf{x}_k^\top \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1} \mathbf{x}_k I(\mathcal{B}_{1k}) \xrightarrow{a.s.} 1 + \frac{1}{\theta \underline{m}(\theta)}, \text{ as } n \rightarrow \infty. \quad (2.8.27)$$

We need to be careful here that although Lemma 3.3 in [Bai and Silverstein \(1999\)](#) is derived under the truncation of entries of $\mathbf{X}$ at a constant, we can still get the same conclusion using our truncation. From (2.8.22), we see that

$$|\mathbf{x}_k^\top \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1} \mathbf{x}_k I(\mathcal{B}_{1k}) - \frac{1}{n} \text{tr} \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1} I(\mathcal{B}_{1k})| \xrightarrow{a.s.} 0. \quad (2.8.28)$$

From (2.8.27) and (2.8.28), it follows that

$$\frac{1}{n} \text{tr} \tilde{\Sigma}_1 (\mathbf{D}_k^\top)^{-1} I(\mathcal{B}_{1k}) \xrightarrow{a.s.} 1 + \frac{1}{\theta \underline{m}(\theta)}. \quad (2.8.29)$$

The dominated convergence theorem implies (2.8.24).

Chapter 3

Tracy-Widom law for the largest eigenvalue of signal-plus-noise matrices

3.1 Introduction

Consider a signal-plus-noise data matrix. It takes the form of

$$\mathbf{S} = \mathbf{R} + \mathbf{X}, \tag{3.1.1}$$

where $\mathbf{R}$ is a deterministic signal matrix and $\mathbf{X}$ is the noise matrix consisting of *i.i.d.* standardized entries. There have been a lot of works on the spectral properties of signal-plus-noise type matrices, as was briefly reviewed in the first Chapter. Studying the asymptotic distributions of extreme eigenvalues is crucial to many research tasks such as signal detection. However, detection of some special signals when $\mathbf{R}$ is allowed to be full rank has not been well studied, due to the lack of knowledge on the asymptotic distributions of the extreme eigenvalues.

In this Chapter, we aim to develop the Tracy-Widom distributions of extreme eigenvalues for $\mathbf{S}\mathbf{S}^*$. Since the detailed proof is lengthy we give an outline of the method first. We divide the problem into two main steps. The first main step is to obtain the asymptotic distribution of extreme eigenvalues when $\mathbf{X}$ is Gaussian. The method used in this step follows from [Lee and Schnelli \(2015, 2016\)](#). A dynamic signal matrix is constructed to link the target matrix with a sample covariance matrix, and then we make a continuous Green function comparison over the time flow. The analysis requires a local law as an ingredient, that is, an estimation of the

difference between a Green function of the target matrix and a fixed limit uniformly on a domain of the complex plane. We establish an entrywise local law in this step and track the rate of change of Green functions over time flow accurately with its help. Some complicated cancellation schemes referred as optical theorems are needed. An averaged local law is also necessary to show that the sample eigenvalues are close to the classical locations defined through quantiles of LSD, known as the rigidity property in the random matrix theory literature.

The second main step is to show the universality of largest eigenvalues. To be more specific, we should relax the Gaussian assumption to the general $\mathbf{X}$ that contains entries with a larger support. Then a condition that is slightly weaker than the finiteness of fourth moment is proved to be necessary and sufficient at the end of the proof. The extension of $\mathbf{X}$ from Gaussian to the general case with a large support also requires standard Green function comparison arguments and rigidity of eigenvalues near the edge as two technical inputs, as in the first main step. Thus the local laws shall be extended, and the extension is twofold. First, since the left and right orthogonal matrices of $\mathbf{R}$ are not removable when we study the eigenvalues of $\mathbf{S}\mathbf{S}^*$ for non-Gaussian $\mathbf{X}$, we need the anisotropic local law. Another extension is in the sense of support of entries of $\mathbf{X}$, where we need local laws for variables with the larger support.

To conclude the anisotropic local law for the general $\mathbf{X}$ with a small support, we adopt the continuous interpolation method used in [Knowles and Yin \(2017\)](#). This method has been applied to prove local laws for different random matrix models, see [Ding and Yang \(2020b\)](#), [Han et al. \(2018\)](#), [Han et al. \(2016\)](#), [Yang \(2019\)](#). One of the challenges in the procedure is to remove the condition that the third moment vanishes, which relies heavily on the model structure. This has been completely solved for the models in [Han et al. \(2018\)](#); [Knowles and Yin \(2017\)](#) where a further expansion on the Green function was performed together with a careful parity analysis, but for the separable model studied in [Yang \(2019\)](#), the third moment vanishing condition is not so easy to remove, thus there are some additional conditions put on the third moment together with the spectral domain. We use a slightly different argument here to remove this third moment condition by restricting the spectral domain. The averaged local law is also proved through analogous arguments to the anisotropic local law. We remark that although we do not provide the local laws in the largest permissible domain, our allowed domain is enough when one cares about the edge behaviours.

To obtain the strong average local for the general $\mathbf{X}$ with the large support, we use the method developed in [Lee and Yin \(2014\)](#) where the discrete Lindeberg replacement was applied many times. This method is robust and has been applied in, e.g., [Ding and Yang \(2018\)](#), [Yang \(2019\)](#).

We also notice that a model similar to ours was studied in very recent works ([Ding and Yang, 2020a,b](#)). They considered the case when the variance of the entries of noise matrix tends to zero, which can be regarded as a rectangular Gaussian divisible model if the noise matrix is Gaussian. The local laws and the edge universality were studied in [Ding and Yang \(2020a\)](#). The Tracy-Widom law for Gaussian case was proved in [Ding and Yang \(2020b\)](#) by introducing a Dyson Brownian motion (DBM) of eigenvalues coupled to DBM of a Wishart matrix, followed by a short time range analysis on the difference of eigenvalues locations that obey a discrete parabolic equation, see also the previous work ([Landon and Yau, 2017](#)). In contrast, the method developed in [Lee and Schnelli \(2015\)](#) to prove the Tracy-Widom law for Gaussian case is by a direct Green comparison for long time range, which is kind of extension of [Erdős et al. \(2012\)](#) but involves much more complicated higher order cancellation of Green functions. We adopt the method in [Lee and Schnelli \(2015\)](#) to prove the Tracy-Widom distribution, which is unrelated to the method in ([Ding and Yang, 2020a,b](#)).

The remaining sections are organized as follows. In Section [3.2](#) we state our main results regarding the Tracy-Widom distribution of extreme eigenvalues. Section [3.3](#) includes the local laws for the linearization matrix of the signal plus noise model, which serves as the main tools to prove the Tracy-Widom distribution. Some notations and preliminaries are provided in this section. In Section [3.4](#) we prove the Tracy-Widom distribution for the Gaussian case. In Sections [3.5](#), [3.6](#), [3.7](#), we prove the local laws step by step. We consider the entrywise and averaged local laws for the Gaussian case in Section [3.5](#), extend it to the anisotropic local law for the general $\mathbf{X}$ in Section [3.6](#), and finally conclude the averaged local law for the large support case in Section [3.7](#). The proof of the main result is also finished in Section [3.7](#). Section [3.8](#) includes some “optical theorems” used in Section [3.4](#).

3.2 The main result

3.2.1 The main result

In this section, we introduce the model and state the main result of this chapter. We consider a matrix $\mathbf{S}$ of the form (3.1.1). The signal part $\mathbf{R}$ is an $M \times N$ ($M \leq N$) deterministic matrix that has a singular value decomposition (SVD)

$$\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{V}^*, \quad (3.2.1)$$

where $\mathbf{U}, \mathbf{V}$ are left and right singular vectors respectively and the diagonal matrix $\mathbf{D}$ consists of the singular values $d_1 \geq d_2 \geq \dots \geq d_M \geq 0$, i.e.

$$\mathbf{D} = \begin{pmatrix} d_1 & \dots & 0 & \dots & 0 \\ 0 & \ddots & 0 & \dots & 0 \\ 0 & \dots & d_M & \dots & 0 \end{pmatrix}. \quad (3.2.2)$$

Moreover we need the following assumption on $\mathbf{X}$.

Assumption 1. *We assume that $\mathbf{X} = (x_{ij})$ is an $M \times N$ matrix, whose entries $\{x_{ij} : 1 \leq i \leq M, 1 \leq j \leq N\}$ are independent real random variables satisfying*

$$\mathbb{E}x_{ij} = 0, \quad \mathbb{E}|x_{ij}|^2 = \frac{1}{N}. \quad (3.2.3)$$

Moreover, we assume that

$$\lim_{t \rightarrow \infty} t^4 \mathbb{P}(|\sqrt{N}x_{ij}| > t) \rightarrow 0 \quad (3.2.4)$$

We consider the high dimensional setting where $M/N \rightarrow c \in (0, \infty)$. If $c > 1$, we can exchange the role of $\mathbf{S}$ and $\mathbf{S}^*$. Therefore we assume the following for convenience.

Assumption 2. $c_N := M/N \rightarrow c \leq 1$.

Let $\mathbf{Q} := \mathbf{S}\mathbf{S}^* = (\mathbf{R} + \mathbf{X})(\mathbf{R} + \mathbf{X})^*$. Denote the empirical spectral distribution (ESD) of $\mathbf{R}\mathbf{R}^*$ by

$$\hat{\rho} := \frac{1}{M} \sum_{i=1}^M \delta_{d_i^2},$$

where δ_x is the Dirac measure at point x . Use $\mu_1 \geq \mu_2, \dots, \geq \mu_M$ to denote the ordered eigenvalues of $\mathbf{Q}$. The ESD of $\mathbf{Q}$ is

$$F^{\mathbf{Q}} := \frac{1}{M} \sum_{i=1}^M \delta_{\mu_i}.$$

The Stieltjes transform of $F^{\mathbf{Q}}$ is

$$s_N(z) = s_{\mathbf{Q}}(z) := \int \frac{dF^{\mathbf{Q}}(t)}{t - z}. \quad (3.2.5)$$

If $\hat{\rho} \rightarrow \rho$ and $M/N \rightarrow c$, it was shown in [Dozier and Silverstein \(2007b\)](#) that $F^{\mathbf{Q}}$ converges to a deterministic distribution $F^{c,\rho}$ whose stieltjes transform $s_0 = s_0(z)$ is defined through a self-consistent equation:

$$s_0 = \int \frac{d\rho(t)}{\frac{t}{1+cs_0} - z(1+cs_0) + (1-c)}, \quad s_0 \in \mathcal{C}^+, \Im(zs_0) \geq 0. \quad (3.2.6)$$

In the following, we need the non-asymptotic version of s_0 , denoted by s , which is the unique solution in $\mathcal{C}^+$ satisfying (3.2.6) where we replace c and ρ by c_N and $\hat{\rho}$ respectively. According to Theorem 2.1 in [Dozier and Silverstein \(2007a\)](#), $\lim_{z \rightarrow E, z \in \mathcal{C}_+} s(z)$ exists for $E \in \mathcal{R}/\{0\}$, so the definition of $s(z)$ can be extended to $z \in \mathcal{C}_+ \cup \mathcal{R}/\{0\}$, and we still denote it by s . By the well known inverse formula of the Stieltjes transform, we know that the non asymptotic distribution $F^{c,\rho}$ has a continuous density ρ_0 with

$$\rho_0(E) = \lim_{\eta \rightarrow 0} \frac{1}{\pi} \Im s(E + i\eta) = \frac{1}{\pi} \Im s(E), \quad E \in \mathcal{R}/\{0\}. \quad (3.2.7)$$

For convenience, we also need to introduce the $N \times N$ conjugate matrix $\tilde{\mathbf{Q}} = \mathbf{S}^* \mathbf{S}$ which shares the same non-zero eigenvalues of $\mathbf{Q}$. We have that $s_{\tilde{\mathbf{Q}}} = -\frac{1-c_N}{z} + c_N s_{\mathbf{Q}}$. If we use $\tilde{s}$ to denote the non-asymptotic version of limit of $s_{\tilde{\mathbf{Q}}}$, it is easy to verify that

$$\tilde{s} = -\frac{1-c}{z} + cs. \quad (3.2.8)$$

The support of $F^{c,\rho}$ has been analyzed carefully in [Loubaton et al. \(2011\)](#). We introduce the following functions that were used in [Loubaton et al. \(2011\)](#), but

omit the subscript N for ease of notation:

$$\begin{aligned} f(w) &:= \frac{1}{M} \text{Tr}(\mathbf{R}\mathbf{R}^* - wI_M)^{-1}, \\ \phi(w) &:= w(1 - cf(w))^2 + (1 - c)(1 - cf(w)), \\ w(z) &:= z(1 + cs(z))^2 - (1 - c)(1 + cs(z)). \end{aligned} \tag{3.2.9}$$

From the non-asymptotic version of (3.2.6) we can find

$$\frac{1}{1 + cs} = 1 - cf(w). \tag{3.2.10}$$

Substituting this into $\phi(w)$ we find

$$\phi(w(z)) = z. \tag{3.2.11}$$

Let ξ_r be the largest solution to $\phi^{(1)}(w) = 0$ and satisfy $1 - cf(w) > 0$. Denote

$$\lambda_r := \phi(\xi_r). \tag{3.2.12}$$

By Theorem 3 in Loubaton et al. (2011), λ_r is the rightmost boundary of the support of $F^{c,\rho}$. By Lemma 2 in Loubaton et al. (2011), we also have $\xi_r = \lim_{\eta \rightarrow 0} w(\lambda_r + i\eta) = w(\lambda_r)$. Denote

$$b(z) := 1 + cs(z), \quad \text{and} \quad b \equiv b(\lambda_r) = 1 + cs(\lambda_r). \tag{3.2.13}$$

Then according to the last equation in (3.2.9), we have the relation

$$\xi_r = \lambda_r b^2 - (1 - c)b. \tag{3.2.14}$$

From (3.2.10) and Theorem 2.1 in Dozier and Silverstein (2007a), we can infer that

$$1 - cf(\xi_r) = \frac{1}{b}. \tag{3.2.15}$$

Since

$$\phi^{(1)}(w) = (1 - cf(w))^2 - 2cw(1 - cf(w))f^{(1)}(w) - c(1 - c)f^{(1)}(w), \tag{3.2.16}$$

by the fact that $\phi^{(1)}(\xi_r) = 0$ together with (3.2.13), (3.2.14) and (3.2.15), we have the following crucial relation between quantities defined above:

$$\frac{1}{N} \sum_{i=1}^M \frac{b(\lambda_r b^2 + \xi_r)}{(d_i^2 - \xi_r)^2} = 1. \quad (3.2.17)$$

To ensure that there is a square root behaviour (see Lemma 3.5.4 below) near the edge λ_r , we need the following assumption.

Assumption 3. *There exists a positive constant $\hat{c}$, such that*

$$\limsup_{N \rightarrow \infty} d_1^2 - \xi_r < -\hat{c}. \quad (3.2.18)$$

Remark 10. We consider two examples to see that this assumption is reasonable. Example 1. If $d > 0$ and $d_i = d$ for all i , then by some calculations, we find that $\phi^{(1)}(w) = 0$ is equivalent to $g(w) := -w^3 + 3d^2w^2 - (3d^4 - 2cd^2 - c)w + d^6 - 2cd^4 + (2c^2 - c)d^2 = 0$. Observing that $g(d^2) = 2c^2d^2 > 0$ and $g(+\infty) < 0$, by the mean value theorem, there exists at least one $w > d^2$, s.t. $\phi^{(1)}(w) = 0$. Denote the largest one by ξ_r , and it is easy to verify that $1 - cf(\xi_r) > 0$ since $f(\xi_r) < 0$. Hence Assumption 3 holds.

Example 2. If $c = 1$, and $\mathbf{R}\mathbf{R}^* = \text{diag}(2, 2-1/(M-1), 2-2/(M-1), \dots, 1+1/(M-1), 1)$, then the limiting spectral distribution of $\mathbf{R}\mathbf{R}^*$ is a uniform distribution on the interval $[1, 2]$. We find the largest solution to $\phi^{(1)}(w) = 0$ is approximately 3.89 and $1 - f(3.89) > 0$. The Assumption 3 also holds obviously.

Recall that $b = b(\lambda_r) = 1 + cs(\lambda_r)$. Define γ_0 to be

$$\gamma_0 := \left\{ \left[-\frac{1}{b^3} - \frac{1}{N} \sum_{i=1}^M \frac{(2\lambda_r b - (1-c))^2}{(d_i^2 - \xi_r)^3} - \frac{1}{N} \sum_{i=1}^M \frac{\lambda_r}{(d_i^2 - \xi_r)^2} \right] \left[\frac{1}{N} \sum_{i=1}^M \frac{b^2}{(d_i^2 - \xi_r)^2} \right]^{-1} \right\}^{-1/3}. \quad (3.2.19)$$

We are ready to give the main theorem of this chapter.

Theorem 3.2.1. *1. (Sufficient condition) Suppose that Assumptions 1,2,3 hold. Let μ_1 be the largest eigenvalue of $\mathbf{Q}$. Then there exists γ_0 defined in (3.2.19) such that the distribution of the rescaled eigenvalue of $\mathbf{Q}$ converges to the type-1 Tracy-Widom distribution denoted by F_1 , i.e.,*

$$\lim_{N \rightarrow \infty} \mathbb{P}(\gamma_0 N^{2/3}(\mu_1 - \lambda_r) \leq t) = F_1(t). \quad (3.2.20)$$

2. (Necessary condition) If (3.2.4) does not hold, then μ_1 does not converge to the Tracy-Widom distribution. Actually, we have for any fixed $T > \lambda_r$,

$$\limsup_{N \rightarrow \infty} \mathbb{P}(\mu_1 > T) > 0 \quad (3.2.21)$$

Remark 11. F_1 can be described by Painlevé equation. One may refer to Tracy and Widom (1996) for more characterization of type-1 Tracy-Widom distribution. The result (3.2.20) can be extended to finite centered and scaled extreme eigenvalues. We have the joint distribution of $\gamma_0 N^{2/3}(\mu_1 - \lambda_r), \gamma_0 N^{2/3}(\mu_2 - \lambda_r), \dots, \gamma_0 N^{2/3}(\mu_m - \lambda_r)$ converges weakly to the m -dimensional joint Tracy-Widom distribution for any fixed m . Similar extension has been made on different types of random matrix models, for instance, the sample covariance matrices discussed in Remark 2.6 of Lee and Schnelli (2016). The convergence to joint Tracy-Widom law can be deduced by using (3.3.27) and an extension of Proposition 3.4.2 (see Remark 16).

Remark 12. If one considers $\tilde{\mathbf{Q}} = (\mathbf{R} + \sigma \mathbf{X})(\mathbf{R} + \sigma \mathbf{X})^*$ for constant $\sigma > 0$, the Tracy-Widom distribution still holds by some modification on Assumption 3. Notice that we can write $\tilde{\mathbf{Q}} = \sigma^2(\tilde{\mathbf{R}} + \mathbf{X})(\tilde{\mathbf{R}} + \mathbf{X})^*$ where $\tilde{\mathbf{R}} = \sigma^{-1}\mathbf{R}$. Therefore by replacing $\mathbf{R}$ with $\tilde{\mathbf{R}}$ and redefining $f, \phi, w, \xi_r, \lambda_r, b$ from (3.2.9) to (3.2.13), under an assumption similar to Assumption 3, the Tracy-Widom law still holds for the largest eigenvalue of $\tilde{\mathbf{Q}}$ with a different limit and a scaling constant.

Remark 13. It is easy to see that this theorem also holds if we replace $\mathbf{X}$ by $\mathbf{U}_1 \mathbf{X} \mathbf{V}_1^*$ where $\mathbf{U}_1, \mathbf{V}_1$ are any orthogonal matrices with dimensions M and N respectively.

We use a simple example to illustrate this result. Let $\mathbf{R}$ be a randomly generated orthogonal matrix, and the entries of $\sqrt{N}\mathbf{X}$ be taken from a discrete distribution defined by

$$\frac{1}{9}\delta_{-2} + \frac{4}{9}\delta_{-1/2} + \frac{4}{9}\delta_1, \quad (3.2.22)$$

where δ_a is the Dirac measure at a . Random variables generated by such a distribution have mean 0 and variance 1. We find $\xi_r = 3$, $\lambda_r = 27/4$ and $\gamma_0 = (27^2/16)^{-1/3}$. We plot the histogram of $\gamma_0 N^{2/3}(\mu_1 - \lambda_r)$ for 10000 generated matrices in Figure 3.1, we see that the simulation results fit the Tracy-Widom law very well.

FIGURE 3.1: The histogram of centralized and rescaled largest eigenvalues with $M = N = 500$ and 10000 matrices.

3.2.2 Applications

In this section, we discuss some potential applications of our results in signal processing and data privacy.

3.2.2.1 Detection of interference signals

The PCA and signal detection for hyperspectral images arouse researchers' interests in recent years, for example in [Chang and Du \(1999\)](#) and [Svensson and Jakobsson \(2011\)](#). Assume the observed N independent samples can be modelled as

$$\mathbf{s}_i = \mathbf{f}_i + \mathbf{r}_i + \mathbf{x}_i, \quad (3.2.23)$$

where $\mathbf{f}_i$ and $\mathbf{r}_i$ are vectors of strong interference signal and interested signal, and where $\mathbf{x}_i$ is the noise vector. In a matrix form denote by $\mathbf{S}$ the matrix with i -th column being $\mathbf{s}_i$. We aim to consider the following hypothesis under high dimensional setting:

$$H_0 : \text{there is no strong interference signal among these } N \text{ observations.} \quad (3.2.24)$$

To test H_0 , we use the following pivotal statistic used in [Onatski \(2009\)](#):

$$T_i = \frac{\mu_i - \mu_{i+1}}{\mu_{i+1} - \mu_{i+2}} \text{ for } i = 1, \dots, m, \quad (3.2.25)$$

	Model 1					Model 2				
$N(0, 1)$	T_1	T_2	T_3	T_4	Ben	T_1	T_2	T_3	T_4	Ben
$c = 1/2$	0.052	0.038	0.032	0.052	0.026	0.056	0.046	0.036	0.028	0.052
$c = 3/4$	0.048	0.034	0.046	0.030	0.048	0.046	0.052	0.030	0.038	0.044
$c = 1$	0.038	0.034	0.030	0.044	0.056	0.050	0.048	0.032	0.040	0.046
t_{10}										
$c = 1/2$	0.046	0.036	0.052	0.042	0.052	0.070	0.052	0.040	0.052	0.050
$c = 3/4$	0.048	0.048	0.040	0.040	0.048	0.064	0.036	0.034	0.054	0.034
$c = 1$	0.050	0.038	0.036	0.040	0.044	0.032	0.054	0.032	0.044	0.022

TABLE 3.1: Size under Models 1 and 2

where m is a constant, μ_i is the i -th largest sample eigenvalue of $\mathbf{S}\mathbf{S}^\top$. From Theorem 1, it is easy to prove that (3.2.25) follows a variant of Tracy-Widom law. Hence, we make the final decision as follows: if all T_i for $i = 1, \dots, m$ follows the variant of Tracy-Widom law, we accept H_0 , and if one of T_i exceeds some specific thresholds, i.e., $T_i > \tau_n$, we reject H_0 . In practice, we set $m = \lfloor \log n \rfloor$ and we use Bonferroni correction to this test.

To evaluate the performance of the proposed method, we also conduct the following numerical experiments. Specifically, we consider the following model:

- Model 1: each $\mathbf{r}_i$ follows a uniform distribution on the surface of unit sphere.
- Model 2: $\mathbf{R} = UDV^\top$ in which i -th column is $\mathbf{r}_i$, $D = (\mathbf{I}, \mathbf{0}) \in \mathbb{R}^{M \times N}$ with $\mathbf{0} \in \mathbb{R}^{M \times (N-M)}$.

For both models, we set $N = 100$, and consider the cases with $\mathbf{X}$ follows normal distribution and t_{10} distribution. All the reported results are based on 500 replications. Table 3.1 displays the size under Models 1 and 2 for different c . Note that $\mathcal{H}_0$ is a multiple hypothesis for $i = 1, \dots, m$, and hence we also use Bonferroni correction to this test.

We also consider the power of the proposed statistic. Based on aforementioned two models, we consider the following two cases:

Case 1: Under Model 1, we assume that the first k observations contain strong interference signals with $\mathbf{f}_i = (\sqrt{a+1} - 1) \cdot \mathbf{u}_i$ for $i = 1, \dots, k$, where $a > 0$ and $\mathbf{u}_i$ follows a uniform distribution on the surface of unit sphere.

Case 2: Under Model 2, we set the $\mathbf{f}_i = a \cdot \mathbf{e}_i$ for $i = 1, \dots, k$, where $\mathbf{e}_i$ is a unit vector with i -th coordinate being one.

		Model 1					Model 2				
$a = 4$		T_1	T_2	T_3	T_4	Ben	T_1	T_2	T_3	T_4	Ben
	$c = 1/2$	0.084	0.052	0.050	0.034	0.076	1.000	0.036	0.040	0.036	1.000
$k = 1$	$c = 3/4$	0.208	0.044	0.056	0.042	0.064	1.000	0.052	0.042	0.044	1.000
	$c = 1$	0.172	0.052	0.048	0.050	0.094	1.000	0.046	0.048	0.050	1.000
	$c = 1/2$	0.036	0.078	0.058	0.042	0.072	0.000	1.000	0.068	0.056	1.000
$k = 2$	$c = 3/4$	0.024	0.128	0.050	0.034	0.064	0.000	1.000	0.056	0.050	1.000
	$c = 1$	0.012	0.160	0.042	0.042	0.076	0.000	1.000	0.068	0.052	0.998
<hr/>											
$a = 9$											
	$c = 1/2$	0.832	0.046	0.064	0.024	0.364	1.000	0.046	0.046	0.042	1.000
$k = 1$	$c = 3/4$	0.872	0.052	0.032	0.028	0.488	1.000	0.056	0.032	0.034	1.000
	$c = 1$	0.884	0.052	0.044	0.044	0.502	1.000	0.038	0.034	0.050	1.000
	$c = 1/2$	0.000	0.590	0.062	0.050	0.098	0.000	1.000	0.056	0.060	1.000
$k = 2$	$c = 3/4$	0.000	0.786	0.052	0.050	0.270	0.000	1.000	0.068	0.044	1.000
	$c = 1$	0.000	0.822	0.046	0.044	0.382	0.000	1.000	0.058	0.038	1.000
<hr/>											
$a = 14$											
	$c = 1/2$	0.966	0.036	0.058	0.030	0.712	1.000	0.056	0.044	0.042	1.000
$k = 1$	$c = 3/4$	0.988	0.056	0.044	0.036	0.818	1.000	0.054	0.044	0.046	1.000
	$c = 1$	0.996	0.046	0.044	0.024	0.844	1.000	0.054	0.036	0.050	1.000
	$c = 1/2$	0.000	0.964	0.036	0.046	0.400	0.000	1.000	0.046	0.032	1.000
$k = 2$	$c = 3/4$	0.000	0.992	0.050	0.044	0.790	0.000	1.000	0.052	0.032	1.000
	$c = 1$	0.000	1.000	0.050	0.042	0.834	0.000	1.000	0.054	0.056	1.000

TABLE 3.2: Power under Models 1 and 2 for $N(0, 1)$

Tables 3.2 and 3.3 report the power of above two cases. Here we set $k = 1, 2$ for simplicity.

3.2.2.2 Data privacy

Introducing some random noise to sensitive data can be an efficient way to achieve data privacy. For example, see Chaudhuri et al. (2013) and Yang et al. (2020). Some big companies like Google (Erlingsson et al., 2014) and Apple (Apple, 2017) have adopted systems that are able to learn information on the user community without obtaining the exact information for each user. We consider a scenario that a researcher is interested in the largest eigenvalue of averaged matrix data aggregated from m individuals. The matrix data $\{\mathbf{R}_i\}_{i=1}^m$ can be some medical images or survey results, which are sensitive for each individual in the sense that the largest singular value carries information of some specific features that are private. To hide the information, adding noise matrices $\mathbf{X}_i$ with i.i.d. entries on the raw matrix data is one method. Then a natural question to ask is how to determine the variance of the introduced noise.

		Model 1					Model 2				
$a = 4$		T_1	T_2	T_3	T_4	Ben	T_1	T_2	T_3	T_4	Ben
	$c = 1/2$	0.148	0.038	0.048	0.056	0.082	1.000	0.038	0.042	0.046	1.000
$k = 1$	$c = 3/4$	0.180	0.058	0.044	0.036	0.086	1.000	0.048	0.046	0.050	0.998
	$c = 1$	0.146	0.060	0.036	0.050	0.086	1.000	0.040	0.048	0.046	1.000
	$c = 1/2$	0.164	0.066	0.056	0.044	0.080	0.000	1.000	0.050	0.032	1.000
$k = 2$	$c = 3/4$	0.238	0.088	0.060	0.044	0.062	0.000	1.000	0.058	0.038	1.000
	$c = 1$	0.016	0.176	0.046	0.058	0.060	0.000	1.000	0.046	0.044	1.000
<hr/>											
$a = 9$											
	$c = 1/2$	0.784	0.034	0.048	0.026	0.272	1.000	0.042	0.034	0.034	1.000
$k = 1$	$c = 3/4$	0.886	0.074	0.042	0.392	0.038	1.000	0.054	0.034	0.046	1.000
	$c = 1$	0.876	0.052	0.046	0.048	0.516	1.000	0.052	0.048	0.050	1.000
	$c = 1/2$	0.000	0.738	0.052	0.042	0.386	0.000	1.000	0.048	0.054	1.000
$k = 2$	$c = 3/4$	0.000	0.738	0.05	0.036	0.368	0.000	1.000	0.05	0.028	1.000
	$c = 1$	0.000	0.798	0.062	0.036	0.404	0.000	1.000	0.044	0.050	1.000
<hr/>											
$a = 14$											
	$c = 1/2$	0.968	0.068	0.062	0.040	0.790	1.000	0.078	0.048	0.040	1.000
$k = 1$	$c = 3/4$	0.996	0.070	0.036	0.044	0.782	1.000	0.038	0.044	0.036	1.000
	$c = 1$	0.998	0.046	0.038	0.036	0.516	1.000	0.042	0.048	0.034	1.000
	$c = 1/2$	0.000	0.988	0.050	0.034	0.516	0.000	1.000	0.048	0.032	1.000
$k = 2$	$c = 3/4$	0.000	0.996	0.048	0.042	0.750	0.000	1.000	0.058	0.040	1.000
	$c = 1$	0.000	0.986	0.050	0.058	0.790	0.000	1.000	0.048	0.042	1.000

TABLE 3.3: Power under Models 1 and 2 for t_{10}

On one hand, the noise should hide the information of individual data matrix. In the above scenario, the added noise shall conceal the largest singular value of the raw matrix $\mathbf{R}_i$. [Montanari et al. \(2015\)](#) concluded that for a rank one perturbation of Gaussian tensors (GOE as a special case), when the perturbation is not larger than a specific critical value, eigenvalue based algorithm is unable to detect signals with error probability vanishing. It is reasonable to conjecture that a similar conclusion holds for the signal-plus-noise type of models we are discussing, and more research is needed. On the other hand, the noise cannot be too large since large noise prevents the learning algorithm to extract information from the aggregated data. In other words, the largest singular value of averaged raw matrix data should be significant enough for the researcher to detect or estimate it after adding noise.

We provide a simple simulation to illustrate above. We randomly generate m signal matrices $\{\mathbf{R}_i\}_{i=1}^m$ with SVD being $\mathbf{R}_i = \mathbf{U}\mathbf{D}_i\mathbf{V}^* \in \mathbb{R}^{M \times N}$. Let $m = 100$, $M = N = 200$ and $\mathbf{U}$ and $\mathbf{V}$ be two randomly generated orthogonal matrices. For each i , the largest eigenvalue value of $\mathbf{R}_i\mathbf{R}_i^\top$ is randomly generated from a uniform distribution in $[3, 4]$, and the eigenvalues excluding the largest one are from a uniform distribution in $[1, 2]$. The added noise matrices are $t\mathbf{X}_i$ where $\mathbf{X}_i$

FIGURE 3.2: Left: The first three eigenvalues of $\mathbf{Q}_1(t)$. Right: The first three eigenvalues of $\bar{\mathbf{Q}}(t)$. The largest, second largest and third largest eigenvalues are represented by red triangle, black circle and green plus respectively.

contains i.i.d. variables with normalized Rademacher distribution (the variance is $1/N$), and we set t , the parameter for noise level, ranges from $[0, 3]$ with equal spacing 0.1. Let $\mathbf{Q}_i(t) = (\mathbf{R}_i + t\mathbf{X}_i)(\mathbf{R}_i + t\mathbf{X}_i)^\top$, and $\bar{\mathbf{Q}}(t) = \frac{1}{m} \sum_{i=1}^m \mathbf{Q}_i(t)$. Figure 3.2 depicts the first three eigenvalues of $\mathbf{Q}_1(t)$ and $\bar{\mathbf{Q}}(t)$ respectively.

As noise level increases, the largest eigenvalue is more difficult to be distinguished from the second and third eigenvalues for a particular matrix, demonstrated by the left panel in Figure 3.2. After averaging, the variance of noise is reduced from t^2/N to $t^2/(m^2N)$, thus the largest eigenvalue for the averaged matrices can be distinguished from the other eigenvalues for any $t \in [0, 3]$, demonstrated by the right panel in Figure 3.2. For $\mathbf{Q}_1(t)$, $d_1^2 \approx 9.23$, then the largest sample eigenvalue of $\mathbf{Q}_1(t)$ satisfies Tracy-Widom distribution when t is larger than the critical value approximated to be 1.54. (A quick way to obtain this critical value is to utilize the discussion in Example 2 of Remark 10. If $t = 1$, when $d_1^2 < 3.89$, the largest sample eigenvalue satisfies Tracy-Widom distribution. Thus the critical value for t here is square root of $9.23/3.89$, which approximately equals 1.54). Actually when $t < 1.54$, the largest eigenvalue will be an outlier to the spectrum of $\mathbf{Q}_1(t)$, indicated by the left plot in Figure 3.2. This suggests that if the existence of an outlier eigenvalue is a sensitive information, one should set t at least to be 1.54 to achieve privacy. And choosing any $t \in [1.54, 3]$ will not hinder from learning the existence of the outlier eigenvalue of averaged data.

3.3 Notations and Tools

In this section, we provide the local laws and some related results which serve as basic tools in the proof of Theorem 3.2.1. Since studying the eigenvalues of $\mathbf{S}\mathbf{S}^*$ is equivalent to the eigenvalues of $\mathbf{Y}\mathbf{Y}^*$ where

$$\mathbf{Y} = \mathbf{D} + \mathbf{U}^* \mathbf{X} \mathbf{V}, \quad (3.3.1)$$

we develop the local laws that give estimates of the resolvent of $\mathbf{Y}\mathbf{Y}^*$. This is formulated by introducing linearization matrices

$$H(z) := \begin{pmatrix} -zI_M & \mathbf{Y} \\ \mathbf{Y}^* & -I_N \end{pmatrix}, \quad (3.3.2)$$

and the Green function

$$G(z) := H^{-1}(z) = \begin{pmatrix} G_1 & G_1 \mathbf{Y} \\ \mathbf{Y}^* G_1 & G_2 \end{pmatrix}, \quad (3.3.3)$$

where $G_1(z) = (\mathbf{Y}\mathbf{Y}^* - zI_M)^{-1}$ and $G_2 = (\mathbf{Y}^* \mathbf{Y} / z - I_N)^{-1}$. The averaged Green function of $\mathbf{Q}$ satisfies

$$s_N(z) = \frac{1}{M} \text{Tr} G_1(z). \quad (3.3.4)$$

For sufficiently small positive $\tau, \tilde{c}$, we define the domain of spectral parameter z by

$$\mathcal{D}(\tilde{c}, \tau) \equiv \mathcal{D}(\tilde{c}, \tau, N) := \{z = E + i\eta \in \mathcal{C}^+ : |z| \geq \tau, \lambda_r - \tilde{c} \leq E \leq \tau^{-1}, N^{-1+\tau} \leq \eta \leq \tau^{-1}\}. \quad (3.3.5)$$

We introduce notations to label the entries of H and G . Define index sets

$$\mathcal{I}_M := \llbracket 1, M \rrbracket, \quad \mathcal{I}_N := \llbracket M+1, M+N \rrbracket, \quad \mathcal{I} := \mathcal{I}_M \cup \mathcal{I}_N = \llbracket 1, M+N \rrbracket.$$

We consistently use Latin letters, e.g. i, j , to denote indices in $\mathcal{I}_M$, and Greek letters, e.g. μ, ν , to denote indices in $\mathcal{I}_N$. Then we label the indices of $\mathbf{Y}$ according to

$$\mathbf{Y} = (y_{i\mu} : i \in \mathcal{I}_M, \mu \in \mathcal{I}_N), \quad (3.3.6)$$

and $\mathbf{X}$ is relabelled accordingly. Note that when $\mu = i + M$, $y_{i\mu}$ may have non-zero mean d_i . We use

$$\bar{i} := i + M \quad (3.3.7)$$

to denote such a special Greek letter index corresponding to each $i \in \mathcal{I}_M$. We shall call $(i, \bar{i})$ index pair.

Let $T \subset \mathcal{I}$. We define the minor $H^{(T)} := (H_{st} : s, t \in \mathcal{I} \setminus T)$ by removing all the rows and columns of H indexed by T . Note that we keep the original values of the matrix indices. We also write $G^{(T)} = (H^{(T)})^{-1}$.

We frequently use shorthand notation

$$\sum_i^{(T)} := \sum_{\substack{i=1 \\ i \notin T}}^M, \quad \sum_\mu^{(T)} := \sum_{\substack{\mu=M+1 \\ \mu \notin T}}^{M+N}.$$

If $T = \{i\}$, we abbreviate $(i) = (\{i\})$; similarly, when $T = \{a, b\}$, write $(ab) = (\{a, b\})$. If there are both Latin letter indices and Greek letter indices, we use shorthand notation

$$\sum_{i, \mu, \nu}^{(\bar{i})} := \sum_i \sum_{\mu, \nu}^{(\bar{i})}, \quad \sum_{i, j, \mu, \nu}^{(\bar{i})} := \sum_{i, j} \sum_{\mu, \nu}^{(\bar{i})}$$

We introduce the partial expectation notations $\mathbb{E}_a$ and $\mathbb{E}_\mu$ to indicate that the expectations are taken with respect to variables in a -th row and μ -th column of $\mathbf{X}$ respectively.

In the following definition, we introduce notation used in [Erdős et al. \(2013a\)](#) which provide a simple and systematic way to control the orders of random variables or their expectations, and that will be used throughout this chapter.

Definition 3.3.1. Let

$$X = \{X^{(N)}(u) : N \in \mathcal{N}, u \in U^{(N)}\}, Y = \{Y^{(N)}(u) : N \in \mathcal{N}, u \in U^{(N)}\}$$

be two families of nonnegative random variables and where $U^{(N)}$ is a possibly N -dependent parameter set. We say that X is stochastically dominated by Y ,

uniformly in u , if for all fixed small $\epsilon > 0$, and large $D > 0$,

$$\sup_{u \in U^{(N)}} P(X^{(N)}(u) > N^\epsilon Y^{(N)}(u)) \leq N^{-D},$$

for all large $N \geq N(\epsilon, D)$, and we denote this by $X \prec Y$ or $X = O_{\prec}(Y)$. Moreover, if for some complex valued family X we have $|X| \prec Y$, we can still write $X \prec Y$ or $X = O_{\prec}(Y)$.

We say that an event Ξ holds with high probability if $1 \prec 1(\Xi)$, i.e., $P(\Xi) = 1 - O(N^{-D})$ for any large constant $D > 0$.

Remark 14. We have following arithmetic rules for $\prec$.

- (i) Suppose that $X(u, v) \prec Y(u, v)$ uniformly in $u \in U$ and $v \in V$. If $|V| = O(N^C)$ for some constant C , then $\sum_u X(u, v) \prec \sum_u Y(u, v)$.
- (ii) Suppose that $X_1(u) \prec Y_1(u)$ uniformly in $u \in U$ and $X_2(u) \prec Y_2(u)$ uniformly in $u \in U$, then $X_1(u)X_2(u) \prec Y_1(u)Y_2(u)$ uniformly in $u \in U$. These two properties can be proved by a simple union bound.
- (iii) If $\Gamma := \Gamma(u)$ is deterministic and $\Gamma(u) \geq N^{-C}$ for some $C > 0$, let $X = X(u)$ satisfies $X \prec \Gamma$ and $EX^2 \leq N^{C'}$, then $EX \prec \Gamma$. This follows from an application of Chebyshevis inequality.

Definition 3.3.2. We say that a family of random matrices $\mathbf{X}$ satisfies a bounded support condition with q if for $i \in \mathcal{I}_M$, $\mu \in \mathcal{I}_N$,

$$|x_{i\mu}| \prec q. \quad (3.3.8)$$

Here q may depend on N and usually satisfies $N^{-1/2} \leq q \leq N^{-\delta}$ for some small constant δ . We also say that $\mathbf{X}$ has support q if (3.3.8) holds for all i, μ .

Recall $b(z)$ defined in (3.2.13). Let

$$\tilde{b}(z) := zb(z) - (1 - c), \quad (3.3.9)$$

and define

$$\Pi(z) := \begin{pmatrix} -\tilde{b}(z)\mathbf{I}_M & \mathbf{D} \\ \mathbf{D}^* & -b(z)\mathbf{I}_N \end{pmatrix}^{-1}. \quad (3.3.10)$$

We also need to define the control parameter

$$\Psi(z) := \sqrt{\frac{\Im s(z)}{N\eta}} + \frac{1}{N\eta}. \quad (3.3.11)$$

For $z = E + i\eta$, let $\kappa \equiv \kappa_E = |E - \lambda_r|$ be the distance between the real part of z to the right edge λ_r . We have the following estimates (see Lemma 3.5.4 below)

$$\Im s(z) \sim \begin{cases} \frac{\eta}{\sqrt{\kappa+\eta}}, & \text{if } E \geq \lambda_r + \eta, \\ \sqrt{\kappa+\eta}, & \text{if } E \in [\lambda_r - \delta, \lambda_r + \eta]. \end{cases} \quad (3.3.12)$$

Recall (3.3.5). Define restricted spectral domain

$$\mathcal{D}(\tilde{c}, \tau, c_0) := \mathcal{D}(\tilde{c}, \tau) \cap \{z = E + i\eta : \Psi(z) \leq N^{-1/4-c_0}\}, \quad (3.3.13)$$

We are ready to provide the local law.

Theorem 3.3.3. *Under the Assumptions 1,2,3, and suppose that $\mathbf{X}$ satisfies a bounded support condition with $q = N^{-\delta}$ for any $0 < \delta < 1/2$, for any fixed small positive constant $\tilde{c}, \tau, c_0$, we have for any $z \in \mathcal{D}(\tilde{c}, \tau, c_0)$ and for any deterministic vector $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^{M+N}$ with bounded spectral norms,*

$$|\langle \mathbf{w}_1, (G(z) - \Pi(z))\mathbf{w}_2 \rangle| \prec q + \Psi(z), \quad (3.3.14)$$

and

$$|s_N(z) - s(z)| \prec \frac{1}{N\eta}. \quad (3.3.15)$$

Moreover, for $z \in \mathcal{D}(\tilde{c}, \tau, c_0) \cap \{z = E + i\eta : E \geq \lambda_r, N\eta\sqrt{\kappa+\eta} > N^\tau\}$, we have the following stronger estimate

$$|s_N(z) - s(z)| \prec \frac{N^{-c_0/2}}{N\eta} + \frac{1}{N(\kappa+\eta)} + \frac{1}{(N\eta)^2\sqrt{\kappa+\eta}}. \quad (3.3.16)$$

The above estimates hold uniformly for the spectral parameter z and (3.3.14) is uniform in any set of deterministic vectors with cardinality $N^{O(1)}$.

If the third moment of entries of $\mathbf{X}$ vanishes, we have (3.3.14), (3.3.15) for any $z \in \mathcal{D}(\tilde{c}, \tau)$, and (3.3.16) (the additional term $N^{-c_0/2}/(N\eta)$ can be removed) for $z \in \mathcal{D}(\tilde{c}, \tau) \cap \{z = E + i\eta : E \geq \lambda_r, N\eta\sqrt{\kappa+\eta} > N^\tau\}$.

The proof can be divided into three steps. In the first step, we prove the entrywise local law and the averaged local law for Gaussian $\mathbf{X}$ in Section 3.5. The second step is to extend it to the anisotropic case (3.3.14) for general $\mathbf{X}$ with small support $q = N^{-1/2}$, which is given in Section 3.6. In this step, we start from Gaussian case in Subsection 3.6.1, then extend to general distributions with the third moment of $\mathbf{X}$ vanishing in Subsection 3.6.2, and finally we remove the third moment condition in Subsection 3.6.3. The averaged local laws (3.3.15) and (3.3.16) are also given in Subsection 3.6.3. In the third step, we extend the local law to the case when $\mathbf{X}$ has support $q = N^{-\delta}$ for any $0 < \delta < 1/2$. The anisotropic version is derived in Subsection 3.6.4. The averaged version is given in Corollary 3.7.7.

To make a Green comparison at the edge, we need the rigidity results, which characterize the typical locations of extreme eigenvalues. Besides (3.3.16), another technical input to get the rigidity property is the averaged local law in a domain with imaginary part down to the scale $\eta \geq N^{-1+\epsilon}$ for sufficiently small ϵ . Therefore, we give the following definition and result.

Definition 3.3.4. (Typical locations of eigenvalues) Let γ_i be the typical location of μ_i which is also the i/M quantile of the distribution with p.d.f. ρ_0 , satisfying

$$M \int_{-\infty}^{\gamma_i} \rho_0(x) dx = i. \quad (3.3.17)$$

Theorem 3.3.5. (Rigidity of extreme eigenvalues) Suppose that the assumptions of Theorem 3.3.3 hold. Define spectral domain

$$\mathcal{D}_1(\tilde{c}, \tau, \mathfrak{c}) := \mathcal{D}(\tilde{c}, \tau) \cap \{z = E + i\eta : |E - \lambda_r| \leq N^{-1/2+\mathfrak{c}}\}. \quad (3.3.18)$$

For any $z \in \mathcal{D}_1(\tilde{c}, \tau, \mathfrak{c})$ with sufficiently small $\mathfrak{c}$, we have

$$|s_N(z) - s(z)| \prec \frac{1}{N\eta}. \quad (3.3.19)$$

We have for any $1 \leq j \leq j_0$,

$$|\mu_j - \gamma_j| \prec N^{-2/3} \quad (3.3.20)$$

Proof. First we prove (3.3.19) for any $z \in \mathcal{D}_1(\tilde{c}, \tau, \mathfrak{c})$. Take $\eta_0 = N^{-3/4+3c_0}$. Notice that for any $z = E + i\eta_0 \in \mathcal{D}_1(\tilde{c}, \tau, \mathfrak{c})$ with sufficiently small $\mathfrak{c}$, it also belongs to

$\mathcal{D}(\tilde{c}, \tau, c_0)$. Hence we have

$$s_N^G(E + i\eta_0) - s(E + i\eta_0) \prec 1/(N\eta_0). \quad (3.3.21)$$

Let $\tilde{G}$ be the Green function corresponding to Gaussian $\mathbf{X}$, and denote by $s_N^{\tilde{G}} = M^{-1} \sum_j \tilde{G}_{jj}$. We have for any $0 < \eta_1 \leq \eta_0$,

$$\tilde{s}_N^G(E + i\eta_1) - s(E + i\eta_1) \prec 1/(N\eta_1). \quad (3.3.22)$$

Using $\partial_\eta G(E + i\eta) = iG^2(E + i\eta)$, we can write

$$\begin{aligned} s_N^G(E + i\eta_1) - s_N^{\tilde{G}}(E + i\eta_1) &= s_N^G(E + i\eta_0) - s_N^{\tilde{G}}(E + i\eta_0) \\ &\quad + \frac{i}{M} \int_{\eta_0}^{\eta_1} \sum_j \left((G^2)_{jj}(E + i\eta) - (\tilde{G}^2)_{jj}(E + i\eta) \right). \end{aligned} \quad (3.3.23)$$

Using the estimate

$$\frac{1}{M} \left| \sum_j (G^2)_{jj} \right| \leq \frac{1}{M} \sum_j \Im \langle \mathbf{e}_j, GG^* \mathbf{e}_j \rangle \leq \frac{1}{M\eta} \sum_j \Im G_{jj}, \quad (3.3.24)$$

we can bound the integrand for any $\eta \leq \eta_0$ by

$$\frac{\Im s_N(E + i\eta)}{\eta} \leq \frac{\eta_0}{\eta^2} \Im s_N(E + i\eta_0) \prec \frac{\eta_0 \sqrt{|E - \lambda_r| + \eta_0}}{\eta^2} + \frac{1}{N\eta^2} \prec \frac{1}{N\eta^2}, \quad (3.3.25)$$

where in the first step we use the monotonicity of function $\eta \mapsto \eta \Im G_{jj}(E + i\eta)$, and in the second step we use (3.3.15) since $E + i\eta_0 \in \mathcal{C}(\tilde{c}, \tau, c_0)$. Substituting the bound above into (3.3.23) and taking integration, we find the integral is bounded by $1/(N\eta_1)$. Combining this with (3.3.21) and (3.3.22), we conclude (3.3.19).

Next we prove (3.3.20). This follows from a standard argument by using (3.3.16) and (3.3.19) so we omit the details here. One can refer to the proof of Theorem 2.2 in Erdős et al. (2012) or the proof of Theorem 3.3 in Pillai et al. (2014). $\square$

The following edge universality result is stated for the proof of the main Theorem 3.2.1.

Theorem 3.3.6. *Under the Assumptions 2,3, suppose that both $\mathbf{X}^V$ and $\mathbf{X}^W$ satisfies Assumption 1 and have bounded support q with $n^{-1/2} \leq q \leq n^{-\delta}$ for some*

$\delta > 0$. Then there exist constants $\epsilon_1, \epsilon_2 > 0$, such that for any $s \in \mathbb{R}$, we have

$$\begin{aligned} \mathbb{P}^V(N^{2/3}(\mu_1 - \lambda_r) \leq s - N^{-\epsilon_1}) - N^{-\epsilon_2} &\leq \mathbb{P}^W(N^{2/3}(\mu_1 - \lambda_r) \leq s) \\ &\leq \mathbb{P}^V(N^{2/3}(\mu_1 - \lambda_r) \leq s + N^{-\epsilon_1}) + N^{-\epsilon_2}, \end{aligned} \quad (3.3.26)$$

where $\mathbb{P}^V$ and $\mathbb{P}^W$ stand for the distributions of $\mathbf{X}^V$ and $\mathbf{X}^W$ respectively.

Remark 15. As in Erdős et al. (2012), it is easy to extend the result to finite correlation functions of extreme eigenvalues:

$$\begin{aligned} \mathbb{P}^V(N^{2/3}(\mu_1 - \lambda_r) \leq s_1 - N^{-\epsilon_1}, \dots, N^{2/3}(\mu_m - \lambda_r) \leq s_m - N^{-\epsilon_1}) - N^{-\epsilon_2} \\ \leq \mathbb{P}^W(N^{2/3}(\mu_1 - \lambda_r) \leq s_1, \dots, N^{2/3}(\mu_m - \lambda_r) \leq s_m) \\ \leq \mathbb{P}^V(N^{2/3}(\mu_1 - \lambda_r) \leq s_1 + N^{-\epsilon_1}, \dots, N^{2/3}(\mu_m - \lambda_r) \leq s_m + N^{-\epsilon_1}) + N^{-\epsilon_2}, \end{aligned} \quad (3.3.27)$$

for all $s_1, \dots, s_m \in \mathbb{R}$, where m is fixed and N is sufficiently large.

Proof. The proof follows from standard arguments as in Section 6 of Erdős et al. (2012). The technical inputs are the averaged local (3.3.19), rigidity properties of extreme eigenvalues (3.3.20), and Green function comparison result which is a combination of Proposition 3.6.12 and Proposition 3.7.8. $\square$

We also need the following large deviation estimates taken from Lemma 3.8 in Erdős et al. (2013c).

Lemma 3.3.7. (*Large deviation bounds*). *Let $(x_i), (y_i)$ be independent families of random variables with mean 0, variance $1/N$ and satisfy the bounded support condition with $q = N^{-\epsilon}$ for some constant $\epsilon > 0$. Let (A_i) and (B_{ij}) be deterministic complex numbers. Then we have*

$$\begin{aligned} |\sum_i A_i x_i| &\prec q \max_i |A_i| + \frac{1}{\sqrt{N}} (\sum_i |A_i|^2)^{1/2}, \\ |\sum_{i,j} x_i B_{ij} y_j| &\prec q^2 B_d + q B_o + \frac{1}{N} (\sum_{i \neq j} |B_{ij}|^2)^{1/2}, \\ |\sum \bar{x}_i B_{ii} x_i - \sum_i (\mathbb{E}|x_i|^2) B_{ii}| &\prec q B_d, \\ |\sum_{i \neq j} \bar{x}_i B_{ij} x_j| &\prec q B_o + \frac{1}{N} (\sum_{i \neq j} |B_{ij}|^2)^{1/2}, \end{aligned}$$

where $B_d := \max_i |B_{ii}|$, and $B_o := \max_{i \neq j} |B_{ij}|$.

We record the following Green function identities, see Lemma 4.4 in [Knowles and Yin \(2017\)](#) for instance.

Lemma 3.3.8. *(Green function identities) Let $G \equiv G(z)$ be defined in (3.3.3). We have*

(i) *For any $i \in \mathcal{I}_M$, and $\mu \in \mathcal{I}_N$, we have*

$$\begin{aligned} G_{ii} &= \frac{1}{-z - (\mathbf{Y}G^{(i)}\mathbf{Y}^*)_{ii}}, \\ G_{\mu\mu} &= \frac{1}{-1 - (\mathbf{Y}^*G^{(\mu)}\mathbf{Y})_{\mu\mu}}. \end{aligned} \quad (3.3.28)$$

(ii) *For $i \neq j \in \mathcal{I}_M$*

$$G_{ij} = -G_{ii}(\mathbf{Y}G^{(i)})_{ij} = -G_{jj}(G^{(j)}\mathbf{Y}^*)_{ij} = G_{ii}G_{jj}^{(i)}(\mathbf{Y}G^{(ij)}\mathbf{Y}^*)_{ij}. \quad (3.3.29)$$

For $\mu \neq \nu \in \mathcal{I}_N$

$$G_{\mu\nu} = -G_{\mu\mu}(\mathbf{Y}^*G^{(\mu)})_{\mu\nu} = -G_{\nu\nu}(G^{(\nu)}\mathbf{Y})_{\mu\nu} = G_{\mu\mu}G_{\nu\nu}^{(\mu)}(\mathbf{Y}^*G^{(\mu\nu)}\mathbf{Y})_{\mu\nu}. \quad (3.3.30)$$

(iii) *For $i \in \mathcal{I}_M$ and $\mu \in \mathcal{I}_N$, we have*

$$\begin{aligned} G_{i\mu} &= -G_{\mu\mu}(G^{(\mu)}\mathbf{Y})_{i\mu} = -G_{ii}(\mathbf{Y}G^{(i)})_{i\mu}, \\ G_{\mu i} &= -G_{\mu\mu}(\mathbf{Y}^*G^{(\mu)})_{\mu i} = -G_{ii}(G^{(i)}\mathbf{Y}^*)_{\mu i}. \end{aligned} \quad (3.3.31)$$

(iv) *For $r \in \mathcal{I}$ and $s, t \in \mathcal{I} \setminus \{k\}$, we have*

$$G_{st}^{(r)} = G_{st} - \frac{G_{sr}G_{rt}}{G_{rr}}. \quad (3.3.32)$$

(v) *All of the identities above hold if we replace G by $G^{(T)}$.*

3.4 Tracy-Widom distribution for Gaussian $\mathbf{X}$

In this section, we prove (3.2.20) for Gaussian $\mathbf{X}$. The main technical ingredient is pursuing a continuous Green function comparison near the rightmost edge. This has been successfully used to derive the Tracy Widom law for the deformed Wigner matrix ([Lee and Schnelli, 2015](#)) and the real sample covariance matrix with the general population covariance structure ([Lee and Schnelli, 2016](#)). In Subsection 3.4.1 we rescale $\mathbf{Q}$ and state the entrywise local law, whose proof is postponed to

Section 3.5. In Subsection 3.4.2 we give the Green function comparison result to conclude (3.2.20) for the Gaussian $\mathbf{X}$. In Subsection 3.4.3 we introduce the Green function flow to link the target model with the well studied sample covariance matrix model. And in Subsection 3.4.4 we prove the Green function comparison result in Subsection 3.4.2 by tracking the rate of change of Green functions over the time flow.

3.4.1 Rescaling of the model

We need to rescale the matrix $\mathbf{Q}$ such that the eigenvalue gaps of the renormalized matrix $\hat{\mathbf{Q}} = \gamma\mathbf{Q}$ has same eigenvalue gaps predicted by the Tracy-Widom distribution. The rescaling factor γ for our target matrix is actually defined by γ_0 through equation (3.2.19). We now set

$$\hat{\mathbf{Y}} = \gamma_0^{1/2}(\mathbf{D} + \mathbf{X}), \quad \hat{\mathbf{Q}} := \hat{\mathbf{Y}}\hat{\mathbf{Y}}^*. \quad (3.4.1)$$

The rescaled linearization matrix is defined by

$$\hat{H}(z) := \begin{pmatrix} -zI_M & \hat{\mathbf{Y}} \\ \hat{\mathbf{Y}}^* & -I_N \end{pmatrix} \quad (3.4.2)$$

Let

$$\hat{G}(z) := \hat{H}^{-1}(z) = \begin{pmatrix} \hat{G}_1 & \hat{G}_1\hat{\mathbf{Y}} \\ \hat{\mathbf{Y}}^*\hat{G}_1 & \hat{G}_2 \end{pmatrix}, \quad (3.4.3)$$

where $\hat{G}_1(z) = (\hat{\mathbf{Y}}\hat{\mathbf{Y}}^* - zI_M)^{-1}$ and $\hat{G}_2 = (\hat{\mathbf{Y}}^*\hat{\mathbf{Y}}/z - I_N)^{-1}$. The averaged Green function of $\hat{\mathbf{Q}}$ is

$$m_N(z) := \frac{1}{M} \text{Tr} \hat{G}_1(z). \quad (3.4.4)$$

Denote the stieltjes transform of (non-asymptotic) limiting spectral distribution of $\hat{\mathbf{Q}}$ by $m = m(z)$, which is the unique solution of equation

$$m(z) = \frac{1}{M} \sum_{i=1}^M \frac{1}{\frac{\gamma_0 d_i^2}{1+c\gamma_0 m(z)} - z(1+c\gamma_0 m(z)) + \gamma_0(1-c)}, \quad m \in \mathcal{C}^+. \quad (3.4.5)$$

$m(z)$ can be extended to $\mathbb{R}/\{0\}$. Define

$$\begin{aligned}\hat{f}(\hat{w}) &:= \frac{1}{M} \text{Tr}(\gamma_0 \mathbf{D} \mathbf{D}^* - \hat{w} I_M)^{-1} \\ \hat{\phi}(w) &:= w(1 - c\gamma_0 \hat{f}(\hat{w}))^2 + \gamma_0(1 - c)(1 - c\gamma_0 \hat{f}(\hat{w})) \\ \hat{w}(z) &:= z(1 + c\gamma_0 m(z))^2 - \gamma_0(1 - c)(1 + c\gamma_0 m(z)).\end{aligned}\tag{3.4.6}$$

We denote by ξ the largest solution of $\hat{\phi}'(w) = 0$. Denote the edge for $\hat{Q}$ by E_+ , satisfying $E_+ = \hat{\phi}(\xi)$, which parallels to the relation in (3.2.12). Note that $\gamma_0 m(E_+) = s(\lambda_r)$. Recalling $b = b(\lambda_r)$ defined in (3.2.13), we also have

$$b = 1 + c\gamma_0 m(E_+),\tag{3.4.7}$$

and

$$\xi = E_+ b^2 - \gamma_0(1 - c)b = \gamma_0 \xi_r.\tag{3.4.8}$$

Denote by $\tilde{m}$ the Stieltjes transform of the limiting spectral distribution of $\hat{\mathbf{Y}}^* \hat{\mathbf{Y}}$. Define

$$\tilde{b} := E_+[1 + \gamma_0 \tilde{m}(E_+)].\tag{3.4.9}$$

It is easy to see that $\tilde{b} = E_+ b - \gamma_0(1 - c)$ from the fact that $\tilde{m}(z) = -(1 - c)/z + cm$. It follows easily that

$$\xi = b\tilde{b}.\tag{3.4.10}$$

Now we can rewrite equations (3.2.17) and (3.2.19) as

$$\frac{1}{N} \sum_{i=1}^M \frac{\gamma_0 b(E_+ b^2 + \xi)}{(\gamma_0 d_i^2 - \xi)^2} = 1,\tag{3.4.11}$$

and

$$\frac{1}{N\gamma_0^2} \sum_{i=1}^M \frac{b^2}{(\gamma_0 d_i^2 - \xi)^2} = -\frac{1}{\gamma_0 b^3} - \frac{1}{N} \sum_{i=1}^M \frac{(2E_+ b - \gamma_0(1 - c))^2}{(\gamma_0 d_i^2 - \xi)^3} - \frac{1}{N} \sum_{i=1}^M \frac{E_+}{(\gamma_0 d_i^2 - \xi)^2}.\tag{3.4.12}$$

The equation (3.4.11) can be understood as multiplying γ_0^2 on both the numerator and the denominator of the LHS of (3.2.17).

Next, we collect some estimates on the Green functions of $\hat{\mathbf{Q}}$, which will be used later in Subsection 3.4.5. For $\tau > 0$, and sufficiently small positive constant $\tilde{c}$, define the domain

$$\hat{\mathcal{D}}(\tilde{c}, \tau) \equiv \hat{\mathcal{D}}(\tilde{c}, \tau, N) := \{z \in \mathcal{C}_+, |z| \geq \tau, E_+ - \tilde{c} \leq E \leq \tau^{-1}, N^{-1+\tau} \leq \eta \leq \tau^{-1}\}.$$

Lemma 3.4.1. *With slight abuse of notations, we let $b(z) = 1 + \gamma_0 c m(z)$ and $\tilde{b}(z) = z b(z) - \gamma_0(1 - c)$. Under the Assumptions 2, 3, and suppose that $\mathbf{X}$ contains i.i.d. standard Gaussian variables, we have that*

$$|m_N(z) - m(z)| \prec \frac{1}{N\eta},$$

and

$$\begin{aligned} & \max_{i \in \mathcal{I}_M} \left| \hat{G}_{ii} - \frac{b(z)}{\gamma_0 d_i^2 - \hat{w}(z)} \right| + \max_{i \in \mathcal{I}_M} \left| \hat{G}_{\bar{i}\bar{i}} - \frac{\tilde{b}(z)}{\gamma_0 d_i^2 - \hat{w}(z)} \right| + \max_{i \in \mathcal{I}_M} \left| \hat{G}_{ii}^{(\bar{i})} + \frac{1}{\tilde{b}(z)} \right| \\ & + \max_{i \in \mathcal{I}_M} \left| \hat{G}_{ii}^{(i)} + \frac{1}{b(z)} \right| + \max_{\mu \in [2M+1, M+N]} \left| \hat{G}_{\mu\mu} + \frac{1}{b(z)} \right| + \max_{i \in \mathcal{I}_M} \left| \hat{G}_{i\bar{i}} - \frac{\gamma_0^{1/2} d_i}{\gamma_0 d_i^2 - \hat{w}(z)} \right| \\ & + \max_{s \neq t, (s,t) \text{ is not index pair}} |\hat{G}_{st}| \prec \sqrt{\frac{\Im m(z)}{N\eta}} + \frac{1}{N\eta} \end{aligned}$$

uniformly in $z \in \hat{\mathcal{D}}(\tilde{c}, \tau)$.

Proof. The proof is similar to Lemma 3.5.1 below without essential difference. $\square$

3.4.2 Green function comparison

We introduce rescaled sample covariance matrices

$$\mathbf{W} := c^{1/6} (1 + \sqrt{c})^{-4/3} X X^*, \quad (3.4.13)$$

and let

$$m_{\mathbf{W}}(z) = \frac{1}{N} \text{Tr}(\mathbf{W} - z)^{-1}.$$

Recall the renormalized matrix $\hat{\mathbf{Q}}$ defined in (3.4.1). The following Green function comparison proposition is the key technical input for the proof of Theorem 3.2.1.

Proposition 3.4.2. *Let $\epsilon > 0$ and set $\eta = N^{-2/3-\epsilon}$. Denote the right edge of $\mathbf{W}$ by M_+ . Let E_1 and E_2 satisfy*

$$|E_1|, |E_2| \leq N^{-2/3+\epsilon}.$$

Let $K : \mathbb{R} \rightarrow \mathbb{R}$ be a real function whose derivatives satisfy

$$\sup_x |K^{(k)}(x)| / (|x| + 1)^C \leq C, \quad k = 0, 1, 2, 3, 4.$$

Then there exists a constant $\phi > 0$ such that for any sufficiently large N and for any sufficiently small $\epsilon > 0$, we have

$$\left| EK \left(N \int_{E_1}^{E_2} \Im m_{\hat{\mathbf{Q}}}(x + E_+ + i\eta) dx \right) - EK \left(N \int_{E_1}^{E_2} \Im m_{\mathbf{W}}(x + M_+ + i\eta) dx \right) \right| \leq N^{-\phi}.$$

Remark 16. Proposition 3.4.2 can be easily extended to a general form, see Remark 5.1 in Lee and Schnelli (2015), or Remark 4.2 in Lee and Schnelli (2016).

Proof of (3.2.20) for the Gaussian $\mathbf{X}$. We need to assume the validity of Proposition 3.4.2 which is proved in the next section, and require a similar result to Lemma 3.4 in Lee and Schnelli (2016) using Lemma 3.4.1. Then the proof involves standard arguments developed in Section 6 in Erdős et al. (2012). One may also refer to the proof of Theorem 2.4 in Lee and Schnelli (2016).

3.4.3 Green function flow

In this subsection we prepare for proof of Proposition 3.4.2 by introducing the continuous interpolation Green functions to link $m_{\mathbf{W}}$ with $m_{\hat{\mathbf{Q}}}$. Let the signal matrix $\mathbf{D}(t)$ evolve with time parameter t , defined by

$$\mathbf{D}(t) := \begin{pmatrix} d_1(t) & \cdots & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \cdots & 0 \\ 0 & \cdots & d_M(t) & \cdots & 0 \end{pmatrix}, \quad (3.4.14)$$

where

$$d_i(t) = e^{-t/2} d_i, \quad i = 1, \dots, M. \quad (t \geq 0). \quad (3.4.15)$$

Notice that $\mathbf{D}(0) = \mathbf{D}$, and $\mathbf{D}(\infty)$ is an zero matrix. We can define $\lambda_r(t)$, $b(t)$ and $\xi_r(t)$ according to (3.2.12), (3.2.13) and (3.2.14) respectively. The only difference is that in the current case, $\mathbf{D}(t)$ evolves with the time parameter t . Denote $\gamma(t)$ to be the solution of

$$\begin{aligned} \frac{1}{\gamma(t)^3} \frac{1}{N} \sum_{i=1}^N \frac{b(t)^2}{(d_i(t)^2 - \xi_r(t))^2} &= -\frac{1}{b(t)^3} - \frac{1}{N} \sum_{i=1}^N \frac{(2\lambda_r(t)b(t) - (1-c))^2}{(d_i(t)^2 - \xi_r(t))^3} \\ &\quad - \frac{1}{N} \sum_{i=1}^N \frac{\lambda_r(t)}{(d_i(t)^2 - \xi_r(t))^2}. \end{aligned} \quad (3.4.16)$$

Let

$$\hat{\mathbf{Y}}(t) = \gamma^{1/2}(t)(\mathbf{D}(t) + \mathbf{X}), \quad \hat{\mathbf{Q}}(t) = \hat{\mathbf{Y}}(t)\hat{\mathbf{Y}}(t)^*. \quad (3.4.17)$$

Now we consider the rescaled linearization matrix defined by

$$\hat{H}(z, t) := \begin{pmatrix} -zI_M & \hat{\mathbf{Y}}(t) \\ \hat{\mathbf{Y}}(t)^* & -I_N \end{pmatrix}, \quad (3.4.18)$$

The resolvent

$$\hat{G}(z, t) := \hat{H}(z, t)^{-1} = \begin{pmatrix} \hat{G}_1(z, t) & \hat{G}_1(z, t)\hat{\mathbf{Y}}(t) \\ \hat{\mathbf{Y}}(t)^*\hat{G}_1(z, t) & \hat{G}_2(z, t) \end{pmatrix}, \quad (3.4.19)$$

where $\hat{G}_1(z, t) = [\hat{\mathbf{Y}}(t)\hat{\mathbf{Y}}(t)^* - zI_M]^{-1}$ and $\hat{G}_2(z, t) = [\hat{\mathbf{Y}}(t)^*\hat{\mathbf{Y}}(t)/z - I_N]^{-1}$.

Now the matrix $\hat{\mathbf{Q}}(t)$ evolves with time parameter t , moreover $\hat{\mathbf{Q}}(0) = \hat{\mathbf{Q}}$ where $\hat{\mathbf{Q}}$ is defined in (3.4.1) and $\hat{\mathbf{Q}}(\infty) = \mathbf{W}$. Recall the argument from (3.4.6) to (3.4.8). By replacing d_i and γ_0 with $d_i(t)$ and $\gamma(t)$ in (3.4.6), we can similarly define the edge $E_+(t)$ that varies with t , so like the $b(t)$, $\xi(t)$ and $\tilde{b}(t)$ appeared in (3.4.7), (3.4.8) and (3.4.9). With slight abuse of notations, we still use notation E_+ , b , ξ , $\tilde{b}$ frequently and omit the dependence on t in most cases later for simplicity. Also for ease of notation, we use notion $\mathbf{Y}$, G instead of $\hat{\mathbf{Y}}$, $\hat{G}$ to stand for the rescaled data matrix and the Green function of the rescaled signal plus noise matrix throughout the rest of this section and in Section 3.8. We remark that with these notations, Lemma 3.3.8 still holds for the rescaled model without any difference. Choose the

domain of the spectral parameter z as

$$\mathcal{E}_\epsilon(t) := \{z = E_+(t) + y + i\eta \in \mathcal{C}^+ : y \in [-N^{-2/3+\epsilon}, N^{-2/3+\epsilon}], \eta = N^{-2/3-\epsilon}\}. \quad (3.4.20)$$

The averaged Green function $m_N(t, z)$ (defined in (3.4.4)) for $\hat{\mathbf{Q}}$ is defined by

$$m_N(t, z) := \frac{1}{M} \text{tr}(\hat{\mathbf{Q}}(t) - zI)^{-1}. \quad (3.4.21)$$

Recalling Lemma 3.4.1 and noticing that $\Im m$ has similar property to $\Im s$ in (3.5.13), we have

$$|m_N(t, z) - m(t, z)| \prec \Psi_0,$$

uniformly in $\mathcal{E}_\epsilon(t), t \geq 0$, where

$$\Psi_0 := N^{-1/3+C'\epsilon}, \quad (3.4.22)$$

for some small constant C' independent of N, ϵ and t . Furthermore, we have $m(t, z) - m(E_+) \prec \Psi_0$ for $z \in \mathcal{E}_\epsilon(t)$, see Lemma 3.8.1 below. Therefore for z uniformly in $\mathcal{E}_\epsilon(t), t \geq 0$, we have

$$m_N(t, z) - m(E_+) \prec \Psi_0, \quad (3.4.23)$$

and similar to Lemma 3.4.1, we can obtain

$$\begin{aligned} & \max_{i \in \mathcal{I}_M} |G_{ii} - \frac{b}{\gamma d_i^2 - \xi}| + \max_{i \in \mathcal{I}_M} |G_{\bar{i}\bar{i}} - \frac{\tilde{b}}{\gamma d_i^2 - \xi}| + \max_{i \in \mathcal{I}_M} |G_{ii}^{(\bar{i})} + \frac{1}{\tilde{b}}| + \max_{i \in \mathcal{I}_M} |G_{\bar{i}\bar{i}}^{(i)} + \frac{1}{b}| \\ & + \max_{\mu \in [2M+1, M+N]} |G_{\mu\mu} + \frac{1}{b}| + \max_{i \in \mathcal{I}_M} |G_{i\bar{i}} - \frac{\gamma^{1/2} d_i}{\gamma d_i^2 - \xi}| + \max_{s \neq t, (s,t) \text{ is not index pair}} |G_{st}| \prec \Psi_0. \end{aligned} \quad (3.4.24)$$

3.4.4 Proof of Proposition 3.4.2

With the preparations in the last subsection, we prove Proposition 3.4.2 by using an accurate estimate on the imaginary part of derivatives of the Green function with respect to the time parameter t , see (3.4.25).

Proof of Proposition 3.4.2. Similar to the proof of Proposition 4.1 in Lee and Schnelli (2016), we also only consider the case $K' \equiv 1$ for simplicity. Using Lemma

3.4.3 together with Lemma 3.4.4 below, we find

$$\mathbb{E} \frac{1}{N} \Im \sum_i \frac{\partial G_{ii}}{\partial t} = O(N \Psi_0^5). \quad (3.4.25)$$

Integrating on both sides of (3.4.25) from $t = 0$ to $t = 2 \log N$, we get

$$\left| \mathbb{E} N \int_{E_1}^{E_2} \Im m_N(x + E_+ + i\eta)|_{t=0} dx - \mathbb{E} N \int_{E_1}^{E_2} \Im m_N(x + E_+ + i\eta)|_{t=2 \log N} dx \right| \leq N^{-1/3+C\epsilon}, \quad (3.4.26)$$

for some constant $C > 0$.

From (3.4.15), at $t = \infty$, $d_i = 0$ for $i = 1, \dots, M$. Therefore, according to definition of ξ_r, λ_r, b and γ (see (3.2.9)-(3.2.19)), we have

$$\xi_r(\infty) = \sqrt{c}, \quad \lambda_r = (1 + \sqrt{c})^2, \quad b(\infty) = \frac{1}{1 + \sqrt{c}}, \quad \gamma(\infty) = (1 + \sqrt{c})^{-4/3} c^{1/6}.$$

Let $t_0 = 2 \log N$. It can be easily shown that $\gamma(t_0) = \gamma(\infty) + O(N^{-2})$, and $z(t_0) = z(\infty) + O(N^{-2})$. Using the matrix identity $A^{-1} - B^{-1} = -A^{-1}(A - B)B^{-1}$ and (3.4.24), we have

$$|G_{ii}(t_0) - G_{ii}(\infty)| = \left| - \sum_{a,b} G_{ia}(t_0) \left(\hat{\mathbf{Q}}_{ab}(t_0) - \hat{\mathbf{Q}}_{ab}(\infty) \right) G_{bi} \right| \prec N^{-2/3}.$$

It follows that

$$\left| \mathbb{E} N \int_{E_1}^{E_2} \Im m_N(x + E_+ + i\eta)|_{t=2 \log N} dx - \mathbb{E} N \int \Im m_N(x + E_+ + i\eta)|_{t=\infty} dx \right| \leq N^{-1/3+C\epsilon}, \quad (3.4.27)$$

for some constant $C > 0$. Note that $m_N(x + E_+ + i\eta)|_{t=\infty} = m_W(x + M_+ + i\eta)$.

We conclude the proof from (3.4.26) and (3.4.27).

3.4.5 Derivatives of Green functions

We now consider the derivative of G_{ii} with respect to the time parameter t for $z \in \mathcal{E}_\epsilon(t)$, see (3.4.20). In the following, we denote by $\dot{z}$ the derivative of z with respect to the time parameter t , and similarly denote by $\dot{d}_i$ and $\dot{\gamma}$, see (3.4.15) and

(3.4.16). A direct calculation yields $\dot{d}_i = -\frac{d_i(t)}{2}$. Then we have

$$\begin{aligned}
 \frac{\partial G_{ii}}{\partial t} &= \dot{z} \sum_j G_{ij} G_{ji} - 2 \sum_{j,\mu} G_{ij} \frac{\partial y_{j\mu}}{\partial t} G_{\mu i} \\
 &= \dot{z} \sum_j G_{ij} G_{ji} + \gamma^{1/2} \sum_j G_{ij} d_j G_{\bar{j}i} - \gamma^{-1/2} \dot{\gamma} \sum_{j,\mu} G_{ij} (x_{j\mu} + d_{j\mu}) G_{\mu i} \\
 &= \dot{z} \sum_j G_{ij} G_{ji} + \sum_j G_{i\bar{j}} G_{\bar{j}i} - \gamma^{1/2} \sum_{j,p} G_{ip} x_{p\bar{j}} G_{\bar{j}i} \\
 &\quad - \gamma^{-1} \dot{\gamma} \sum_j G_{i\bar{j}} G_{\bar{j}i} + \gamma^{-1/2} \dot{\gamma} \sum_{j,p} G_{ip} x_{p\bar{j}} G_{\bar{j}i} - \gamma^{-1/2} \dot{\gamma} \sum_{j,\mu} G_{ij} x_{j\mu} G_{\mu i},
 \end{aligned} \tag{3.4.28}$$

where in the last step, we use $G_{i\bar{j}} = \gamma^{1/2} G_{ij} d_j + \gamma^{1/2} \sum_p G_{ip} x_{p\bar{j}}$, which can be obtained from (3.4.3). To obtain the desired order of $\mathbb{E} \frac{1}{N} \Im \sum_i \frac{\partial G_{ii}}{\partial t}$ as in (3.4.25), we start from calculating expectation of the right side of (3.4.28) above and ignore those terms with order $O(\Psi_0^2)$.

We use the Stein's lemma to calculate those terms in (3.4.28) involving entries of $\mathbf{X}$ except G . We obtain

$$\mathbb{E} \sum_{j,p} G_{ip} x_{p\bar{j}} G_{\bar{j}i} = -\frac{\gamma^{1/2}}{N} \sum_{j,p} (2\mathbb{E} G_{ip} G_{p\bar{j}} G_{\bar{j}i} + \mathbb{E} G_{i\bar{j}} G_{pp} G_{\bar{j}i} + \mathbb{E} G_{ip} G_{\bar{j}\bar{j}} G_{pi}). \tag{3.4.29}$$

$$\mathbb{E} \gamma^{-1/2} \sum_{j,\mu} G_{ij} x_{j\mu} G_{\mu i} = -\frac{1}{N} \sum_{j,\mu} (2\mathbb{E} G_{ij} G_{j\mu} G_{\mu i} + \mathbb{E} G_{i\mu} G_{jj} G_{\mu i} + \mathbb{E} G_{ij} G_{\mu\mu} G_{ji}). \tag{3.4.30}$$

Plugging (3.4.29) and (3.4.30) into (3.4.28), we obtain that

$$\begin{aligned}
 \mathbb{E} \frac{\partial G_{ii}}{\partial t} &= \dot{z} \sum_j \mathbb{E} G_{ij} G_{ji} + \sum_j \mathbb{E} G_{i\bar{j}} G_{\bar{j}i} + \frac{\gamma}{N} \sum_{j,p} (2\mathbb{E} G_{ip} G_{p\bar{j}} G_{\bar{j}i} + \mathbb{E} G_{i\bar{j}} G_{pp} G_{\bar{j}i} + \mathbb{E} G_{ip} G_{\bar{j}\bar{j}} G_{pi}) \\
 &\quad - \gamma^{-1} \dot{\gamma} \sum_j \mathbb{E} G_{i\bar{j}} G_{\bar{j}i} - \frac{\dot{\gamma}}{N} \sum_{j,p} (2\mathbb{E} G_{ip} G_{p\bar{j}} G_{\bar{j}i} + \mathbb{E} G_{i\bar{j}} G_{pp} G_{\bar{j}i} + \mathbb{E} G_{ip} G_{\bar{j}\bar{j}} G_{pi}) \\
 &\quad + \frac{\dot{\gamma}}{N} \sum_{j,\mu} (2\mathbb{E} G_{ij} G_{j\mu} G_{\mu i} + \mathbb{E} G_{i\mu} G_{jj} G_{\mu i} + \mathbb{E} G_{ij} G_{\mu\mu} G_{ji}) \\
 &= \dot{z} \sum_j \mathbb{E} G_{ij} G_{ji} + \sum_j \mathbb{E} G_{i\bar{j}} G_{\bar{j}i} + \frac{\gamma}{N} \sum_{j,p} (2\mathbb{E} G_{ip} G_{p\bar{j}} G_{\bar{j}i} + \mathbb{E} G_{i\bar{j}} G_{pp} G_{\bar{j}i} + \mathbb{E} G_{ip} G_{\bar{j}\bar{j}} G_{pi}) \\
 &\quad - \gamma^{-1} \dot{\gamma} \sum_j \mathbb{E} G_{i\bar{j}} G_{\bar{j}i} + \frac{\dot{\gamma}}{N} \sum_{\mu > 2M, j} (2\mathbb{E} G_{ij} G_{j\mu} G_{\mu i} + \mathbb{E} G_{i\mu} G_{jj} G_{\mu i} + \mathbb{E} G_{ij} G_{\mu\mu} G_{ji}).
 \end{aligned} \tag{3.4.31}$$

Actually we have a simpler expression of $\mathbb{E} \frac{1}{N} \mathfrak{S} \sum_i \frac{\partial G_{ii}}{\partial t}$ given in Lemma 3.4.3 below.

We need the following notations. Let

$$\begin{aligned}
X_{22} &= \frac{1}{N} \sum_r G_{ir} G_{ri}, & X_{32} &= (cm(E_+) - cm_N) \frac{1}{N} \sum_r G_{ir} G_{ri}, \\
X_{33} &= \frac{1}{N^2} \sum_{r,s} G_{ir} G_{rs} G_{si}, & X_{42} &= (cm(E_+) - cm_N)^2 \frac{1}{N} \sum_r G_{ir} G_{ri}, \\
X_{43} &= (cm(E_+) - cm_N) \frac{1}{N^2} \sum_{r,s} G_{ir} G_{rs} G_{si}, & X_{44} &= \frac{1}{N^3} \sum_{r,s,t} G_{ir} G_{rs} G_{st} G_{ti}, \\
\tilde{X}_{44} &= \frac{1}{N^3} \sum_{r,s,t} G_{ir} G_{ri} G_{st} G_{ts}.
\end{aligned} \tag{3.4.32}$$

Observe that the above z -dependent random variables satisfy

$$X_{22} = O(\Psi_0^2), \quad X_{32}, X_{33} = O(\Psi_0^3), \quad X_{42}, X_{43}, X_{44}, \tilde{X}_{44} = O(\Psi_0^4).$$

Some variations of above random variables are defined by

$$\begin{aligned}
X'_{22} &= \frac{1}{N} \sum_\alpha G_{i\alpha} G_{\alpha i}, & X'_{32} &= (cm(E_+) - cm_N) \frac{1}{N} \sum_\alpha G_{i\alpha} G_{\alpha i}, \\
X'_{33} &= \frac{1}{N^2} \sum_{\alpha,k} G_{i\alpha} G_{\alpha k} G_{ki}, & X''_{33} &= \frac{1}{N^2} \sum_{\alpha,\beta} G_{i\alpha} G_{\alpha\beta} G_{\beta i}, \\
X'_{42} &= (cm(E_+) - cm_N)^2 \frac{1}{N} \sum_\alpha G_{i\alpha} G_{\alpha i}, & X'_{43} &= (cm(E_+) - cm_N) \frac{1}{N^2} \sum_{\alpha,s} G_{i\alpha} G_{\alpha s} G_{si}, \\
X''_{43} &= (cm(E_+) - cm_N) \frac{1}{N^2} \sum_{\alpha,\beta} G_{i\alpha} G_{\alpha\beta} G_{\beta i}, & X'_{44} &= \frac{1}{N^3} \sum_{\alpha,s,t} G_{i\alpha} G_{\alpha s} G_{st} G_{ti}, \\
X''_{44} &= \frac{1}{N^3} \sum_{\alpha,\beta,t} G_{i\alpha} G_{\alpha\beta} G_{\beta t} G_{ti}, & X'''_{44} &= \frac{1}{N^3} \sum_{\alpha,\beta,\mu} G_{i\alpha} G_{\alpha\beta} G_{\beta\mu} G_{\mu i}.
\end{aligned} \tag{3.4.33}$$

Lemma 3.4.3. *Let*

$$\begin{aligned}
\mathcal{X}_3 &= \sum_i (X_{32} - X_{33}), \quad \mathfrak{X}_4 = (cm - cm_N) + (E_+ - z) \sum_i X_{22}, \\
\mathcal{X}_4 &= \sum_i (3X_{42} - 6X_{43} + 12X_{44} + 3\tilde{X}_{44}).
\end{aligned} \tag{3.4.34}$$

Then we have

$$\mathbb{E} \frac{1}{N} \sum_i \frac{\partial G_{ii}}{\partial t} = C_0 + C_1 \mathbb{E} \mathcal{X}_3 + C_2 \mathbb{E} \mathfrak{X}_4 + C_3 \mathbb{E} \mathcal{X}_4 + O(N \Psi_0^5)$$

where C_0, C_1, C_2 and C_3 are explicitly given by

$$\begin{aligned} C_0 &= \left(\frac{h}{\gamma E_+} - \frac{1}{\gamma}\right)\left(b - \frac{\dot{\gamma}}{\gamma}\right), \quad C_1 = \frac{2\tilde{b}b}{\gamma E_+} - 2\gamma E_+ - \frac{2\tilde{b}\dot{\gamma}}{\gamma^2 E_+}, \\ C_2 &= \frac{2b^2(b\tilde{b} - \gamma^2 E_+^2)}{\gamma^2 E_+ h} - \gamma cm - \frac{\gamma^3(1-c)}{b^3} - \frac{2b(b\tilde{b} - \gamma^2 E_+^2)}{\gamma^3 E_+ h}\dot{\gamma} + \frac{\gamma^2(1-c)}{b^3}\dot{\gamma}, \\ C_3 &= \left(\frac{\gamma^3 h^4 \tilde{b} \varphi_4}{E_+} - \frac{3\tilde{b}}{h} - \frac{3\gamma^2 \tilde{b}}{b^3} - \frac{3\gamma^2 E_+ \tilde{b}}{b^2 h} - \frac{E_+ b^2 + \gamma^2 E_+^2}{bh} + \frac{h\gamma^3(1-c)}{E_+ b^4} - \frac{\gamma^3(1-c)}{b^4}\right)\left(b - \frac{\dot{\gamma}}{\gamma}\right) \\ &\quad + \dot{\gamma} \frac{\gamma^3 cm(1-c)}{b^4}, \end{aligned}$$

and where $h = E_+ b + \tilde{b}$, φ_4 is defined in (3.8.14).

Proof. The first step is to rewrite each term in (3.4.31) as a linear combination of $\mathbb{E}X_{22}, \mathbb{E}X_{32}, \mathbb{E}X_{33}, \mathbb{E}(cm - cm_N), (E_+ - z)\mathbb{E}X_{22}, \mathbb{E}X_{42}, \mathbb{E}X_{43}, \mathbb{E}X_{44}, \mathbb{E}\tilde{X}_{44}$. Although there exist linear relations within these terms, for example, see (3.8.100), we do not pursue the cancellation between them in this step. Notice that the first term is already expressed by $\mathbb{E}X_{22}$. By a simple fact that $\sum_j \mathbb{E}G_{i\bar{j}}G_{\bar{j}i} = N\mathbb{E}X'_{22} - \sum_{\alpha > 2M} EG_{i\alpha}G_{\alpha i}$, the second term can be rewritten according to (3.8.76) and (3.8.80) below. Using (3.8.86) and (3.8.88) below, we find the third term is

$$\frac{\gamma}{N} \sum_{j,p} 2\mathbb{E}G_{ip}G_{p\bar{j}}G_{\bar{j}i} = 2\gamma N\mathbb{E}X'_{33} - \frac{2\gamma}{N} \sum_{\alpha > 2M,p} \mathbb{E}G_{ip}G_{p\alpha}G_{\alpha i} = 2\gamma\left(E_+ - \frac{\gamma(1-c)}{b^2}\right)N\mathbb{E}X_{33}.$$

To handle the fourth term, observe that

$$\frac{\gamma}{N} \sum_{j,p} \mathbb{E}G_{i\bar{j}}G_{pp}G_{\bar{j}i} = \gamma cm \sum_j \mathbb{E}G_{i\bar{j}}G_{\bar{j}i} - \gamma \sum_j \mathbb{E}(cm - cm_N)G_{i\bar{j}}G_{\bar{j}i}. \quad (3.4.35)$$

We have

$$\begin{aligned} \frac{1}{N} \sum_j \mathbb{E}(cm - cm_N)G_{i\bar{j}}G_{\bar{j}i} &= \mathbb{E}X'_{32} - \frac{1}{N} \sum_{\alpha > 2M} \mathbb{E}(cm - cm_N)G_{i\alpha}G_{\alpha i} \\ &= \left(E_+ - \frac{\gamma(1-c)}{b^2}\right)\mathbb{E}X_{32} + \frac{b}{\gamma d_i^2 - \xi} \mathbb{E} \frac{1}{N} (cm - cm_N) + \frac{\gamma^2(1-c)}{b^3} \mathbb{E}(-2X_{42} + 2X_{43} - 2X_{44}) + O(\Psi) \end{aligned}$$

where in the second step we use (3.8.92) and (3.8.96). Next we deal with the fifth term. We have the following relation that can be verified without difficulty,

$$E_+ \tilde{m}(E_+) - \frac{1}{N} \sum G_{\mu\mu} = E_+(cm(E_+) - \frac{1}{N} \sum G_{jj}) + (E_+ - z)cm(E_+) + O(\Psi_0^3). \quad (3.4.36)$$

Using this we find

$$\begin{aligned}
& \frac{\gamma}{N^2} \sum_{p,j} \mathbb{E} G_{ip} G_{\bar{j}\bar{j}} G_{pi} \\
&= -\gamma E_+ \mathbb{E} X_{32} - \gamma(E_+ - z)cm(E_+) \mathbb{E} X_{22} - \frac{\gamma}{N^2} \sum_{\mu > 2M} G_{ip} G_{\mu\mu} G_{pi} + \gamma E_+ \tilde{m}(E_+) \mathbb{E} X_{22} \\
&= -\gamma E_+ \mathbb{E} X_{32} - \gamma(E_+ - z)cm(E_+) \mathbb{E} X_{22} + \gamma E_+ \tilde{m}(E_+) \mathbb{E} X_{22} \\
&+ \frac{\gamma(1-c)}{b} \mathbb{E} X_{22} + \frac{\gamma^2(1-c)}{b^2} \mathbb{E} X_{32} + \frac{\gamma^3(1-c)}{b^3} \mathbb{E}(X_{42} + 4X_{44} + \tilde{X}_{44}) + O(\Psi_0^5),
\end{aligned}$$

where in the second equality we use (3.8.92). The last four terms can be reorganized in a similar manner to above terms, so we do not provide details here.

In the next step, using above expressions of terms in (3.4.31), summing over index i and dividing by N , we summarize the following.

- (1). The terms do not contain expectations of Green entries are $(\frac{h}{\gamma E_+} - \frac{1}{\gamma})(b - \frac{\dot{\gamma}}{\gamma})$.
- (2). With the help of (3.4.47) and (3.8.12), the coefficient of $\sum_i \mathbb{E} X_{22}$ is 0.
- (3). With the help of (3.8.76) and (3.8.80), both the coefficients of $\sum_i \mathbb{E} X_{32}$ and $-\sum_i \mathbb{E} X_{33}$ are

$$\frac{2\tilde{b}b}{\gamma E_+} - 2\gamma E_+ - \frac{2\tilde{b}\dot{\gamma}}{\gamma^2 E_+}.$$

- (4). With the help of (3.8.97) and (3.8.100), both the coefficients of $\mathbb{E}(cm - cm_N)$ and $(E_+ - z) \sum_i \mathbb{E} X_{22}$ are

$$-\frac{2b^2(\gamma^2 E_+^2 - b\tilde{b})}{\gamma^2 E_+ h} - \gamma cm - \frac{\gamma^3(1-c)}{b^3} + \frac{2b(\gamma^2 E_+^2 - b\tilde{b})}{\gamma^3 E_+ h} \dot{\gamma} + \dot{\gamma} \frac{\gamma^2(1-c)}{b^3}$$

- (5). With the help of (3.8.76) and (3.8.80), the coefficient of $\sum_i \mathbb{E}(3X_{42} - 6X_{43} + 12X_{44} + 3\tilde{X}_{44})$ is

$$\begin{aligned}
& \left\{ \frac{\gamma^3 h^4 \tilde{b} \varphi_4}{E_+} - \frac{3\tilde{b}}{h} - \frac{3\gamma^2 \tilde{b}}{b^3} - \frac{3\gamma^2 E_+ \tilde{b}}{b^2 h} - \frac{E_+ b^2 + \gamma^2 E_+^2}{bh} + \frac{h\gamma^3(1-c)}{E_+ b^4} - \frac{\gamma^3(1-c)}{b^4} \right\} (b - \frac{\dot{\gamma}}{\gamma}) \\
& + \dot{\gamma} cm \frac{\gamma^3(1-c)}{b^4}.
\end{aligned}$$

We conclude this lemma from above. $\square$

Lemma 3.4.4.

$$\Im(C_1 \mathbb{E} \mathcal{X}_3 + C_2 \mathbb{E} \mathfrak{X}_4 + C_3 \mathbb{E} \mathcal{X}_4) = O(N \Psi_0^5) \quad (3.4.37)$$

Proof. Subtracting $(C_1\gamma/2)$ -times the imaginary part of (3.8.78) from (3.4.37), we find the coefficient of $\Im\mathbb{E}\mathcal{X}_3$ cancels, the coefficient of $\Im\mathbb{E}\mathfrak{X}_4$ is

$$-\gamma cm - \frac{\gamma^3(1-c)}{b^3} + \frac{2bE_+}{\gamma h}\dot{\gamma} + \frac{\gamma^2(1-c)}{b^3}\dot{\gamma}, \quad (3.4.38)$$

and the coefficient of $\Im\mathbb{E}\mathcal{X}_4$ is $C_3 - \frac{C_1\gamma}{2}\theta_4$, which is explicitly given by

$$\begin{aligned} & \left(\frac{\tilde{b}}{h} + \frac{\gamma^2\tilde{b}}{b^3} + \frac{\gamma^2E_+\tilde{b}}{b^2h} - \frac{E_+b}{h} - \frac{\gamma^2E_+^2}{bh} \right) \left(b - \frac{\dot{\gamma}}{\gamma} \right) - \frac{\gamma^2E_+\tilde{b}}{bh} + \frac{\gamma^3h(1-c)}{E_+b^3} - \frac{\gamma^3(1-c)}{b^3} \\ & + \gamma^5h^4E_+\varphi_4 - \frac{4\gamma^2E_+^2}{h} - \frac{4\gamma^4E_+^2}{b^3} - \frac{3\gamma^4E_+^3}{b^2h} + \frac{\gamma E_+\tilde{b}}{b^2h}\dot{\gamma}. \end{aligned} \quad (3.4.39)$$

From (3.8.97) and (3.8.100), we find

$$\mathbb{E}\mathfrak{X}_4 = -\mathbb{E}\mathcal{X}_4 + \sum_i (-4\mathbb{E}X_{43} + 6\mathbb{E}X_{44} + 2\mathbb{E}\tilde{X}_{44})$$

It follows from (3.8.88) that

$$\mathbb{E}\mathfrak{X}_4 = -\mathbb{E}\mathcal{X}_4 + O(N\Psi_0^5). \quad (3.4.40)$$

To prove (3.4.37), it suffices to show that

$$-(3.4.38) + (3.4.39) = 0. \quad (3.4.41)$$

Replacing $\gamma^5h^4E_+\varphi_4$ in (3.4.39) with

$$\gamma^5h^4E_+\varphi_4 = \frac{\dot{\gamma}}{\gamma} + \gamma^3h^3 \left(\frac{4E_+^2}{\gamma h^4} + \frac{2\gamma E_+^3}{b^2h^4} + \frac{2\gamma E_+^2}{b^3h^3} - \frac{2b\tilde{b}}{\gamma^3h^4} + \frac{\gamma E_+}{h^2b^4} + \frac{1}{\gamma^3h^3} \right)$$

which is obtained from (3.4.43), we find the coefficient of $\dot{\gamma}$ vanishes, and the other terms also cancel. So we conclude this lemma. $\square$

Lemma 3.4.5. *For the b and γ (see (3.4.16) and arguments above it), we have*

$$\dot{b} = \gamma^2E_+ + b(b-1) \quad (3.4.42)$$

$$\dot{\gamma} = \gamma^6h^4E_+\varphi_4 - \gamma^4h^3 \left(\frac{4E_+^2}{\gamma h^4} + \frac{2\gamma E_+^3}{b^2h^4} + \frac{2\gamma E_+^2}{b^3h^3} - \frac{2b\tilde{b}}{\gamma^3h^4} + \frac{\gamma E_+}{h^2b^4} + \frac{1}{\gamma^3h^3} \right) \quad (3.4.43)$$

Proof. Recalling the definition of $\hat{f}$ and E_+ in (3.4.6) and the argument below it, we have

$$E_+ = \xi(1 - c\gamma\hat{f}(\xi))^2 + \gamma(1 - c)(1 - c\gamma\hat{f}(\xi)). \quad (3.4.44)$$

Then using the fact that $1 - c\gamma\hat{f}(\xi) = 1/b$ which can be deduced from (3.4.5), (3.4.7) and (3.4.8), we find

$$\dot{z} = \dot{E}_+ = \dot{\xi} \frac{1}{b^2} - \frac{2\xi(\gamma c \dot{\hat{f}} + c \hat{f} \dot{\gamma})}{b} + \dot{\gamma} \frac{1-c}{b} - \gamma(1-c)c\gamma\dot{\hat{f}} - \gamma(1-c)c\hat{f}\dot{\gamma}. \quad (3.4.45)$$

Actually we can calculate

$$c\dot{\hat{f}}(\xi) = -\frac{\dot{\gamma}\psi_2}{\gamma} + \psi_2 + \varphi_2\dot{\xi} \quad (3.4.46)$$

Plugging (3.4.46) back into (3.4.45), we find that the coefficient for $\dot{\xi}$ is 0, and get

$$\dot{z} = \dot{E}_+ = -\gamma(E_+b + \tilde{b})\psi_2 + \frac{E_+}{\gamma}\dot{\gamma} = -(E_+b + \tilde{b}) + E_+ + \frac{E_+}{\gamma}\dot{\gamma}. \quad (3.4.47)$$

Then we compute the derivative of ξ and h for use later,

$$\dot{\xi} = \frac{b\tilde{b}}{\gamma}\dot{\gamma} + h\dot{b} + E_+b^2 - hb^2, \quad (3.4.48)$$

and

$$\dot{h} = 2\dot{E}_+b + 2E_+\dot{b} - \dot{\gamma}(1-c) = 2E_+b - 2hb + 2E_+\dot{b} + \frac{h}{\gamma}\dot{\gamma}. \quad (3.4.49)$$

We differentiate the equation $\varphi_2 = 1/\gamma b^2 h$ in (3.8.15) with respect to t to get

$$\frac{2}{N} \sum_{i=1} \frac{\dot{\gamma}d_i^2 - \gamma d_i^2 - \dot{\xi}}{(\gamma d_i^2 - \xi)^3} = \frac{\dot{\gamma}}{\gamma^2 b^2 h} + \frac{2\dot{b}}{\gamma b^3 h} + \frac{\dot{h}}{\gamma b^2 h^2}.$$

Plugging (3.4.48) and (3.4.49) into the above we find the coefficient of $\dot{\gamma}$ cancels, and we can obtain that

$$\dot{b} = \gamma^2 E_+ + b(b-1) \quad (3.4.50)$$

Similarly, by differentiating the expression of φ_3 in (3.8.15) with respect to t , we find

$$\begin{aligned} \frac{3}{N} \sum_i \frac{\dot{\gamma} d_i^2 - \gamma d_i^2 - \dot{\xi}}{(\gamma d_i^2 - \xi)^4} + \frac{3\dot{\gamma}}{\gamma^4 h^3} + \frac{3\dot{h}}{\gamma^3 h^4} + \frac{\dot{\gamma}}{\gamma^2 b^3 h^2} + \frac{3\dot{b}}{\gamma b^4 h^2} \\ + \frac{2\dot{h}}{\gamma b^3 h^3} + \frac{E_+ \dot{\gamma}}{\gamma^2 b^2 h^3} + \frac{2E_+ \dot{b}}{\gamma b^3 h^3} + \frac{3E_+ \dot{h}}{\gamma b^2 h^4} - \frac{\dot{E}_+}{\gamma b^2 h^3} = 0. \end{aligned} \quad (3.4.51)$$

Substituting (3.4.47), (3.4.48), (3.4.49) and (3.4.50) into above, we obtain the derivative of $\dot{\gamma}$ given by (3.4.43) through tedious calculations. $\square$

3.5 Entrywise and average local laws for Gaussian $\mathbf{X}$

In this section, we prove the entrywise local law (Lemma 3.5.1) for the Gaussian $\mathbf{X}$ which serves as the first step of proof for Theorem 3.3.3. Since $\mathbf{U}^* \mathbf{X}^{Gauss} \mathbf{V} \stackrel{d}{=} \mathbf{X}^{Gauss}$, it is equivalent to studying G defined in (3.3.3) where $\mathbf{Y} = \mathbf{D} + \mathbf{X}$.

Lemma 3.5.1. *Assume $\mathbf{X}$ contains i.i.d. standard Gaussian variables. Under the assumptions 2,3, we have the estimates*

$$|s_N(z) - s(z)| \prec \frac{1}{N\eta}, \quad \text{and} \quad |G_{st}(z) - \Pi_{st}(z)| \prec \Psi(z), \quad (3.5.1)$$

uniformly for $z \in \mathcal{D}(\tilde{c}, \tau)$. To be more specific, the latter can be written as

$$\begin{aligned} \max_{i \in \mathcal{I}_M} |G_{ii} - \frac{b(z)}{d_i^2 - w(z)}| + \max_{i \in \mathcal{I}_M} |G_{\bar{i}\bar{i}} - \frac{\tilde{b}(z)}{d_i^2 - w(z)}| + \max_{i \in \mathcal{I}_M} |G_{i\bar{i}} - \frac{d_i}{d_i^2 - w(z)}| \\ + \max_{\mu \in \llbracket 2M+1, M+N \rrbracket} |G_{\mu\mu} + \frac{1}{b(z)}| + \max_{s \neq t, (s,t) \text{ is not index pair}} |G_{st}| \prec \Psi(z). \end{aligned} \quad (3.5.2)$$

To prove such results, it is typical to obtain a weak local law (Lemma 3.5.2) first, whose error bounds are not optimal. Then the fluctuation average lemma (see Lemma 3.5.10) will help to improve the error bounds to the optimal scale through an improved stability analysis on the self-consistent equation (see (3.5.60)). This is a routine procedure, see Benaych-Georges and Knowles (2016) as a summary for Wigner matrices. Our proof is similar to the local laws for deformed Wigner matrices in Lee and Schnelli (2013).

3.5.1 Weak local law

In this subsection, we establish the following weaker version of Lemma 3.5.1.

Lemma 3.5.2. *Under the assumptions of Lemma 3.5.1, we have*

$$|s_N(z) - s(z)| + \max_{s \neq t, (s,t) \text{ is not index pair}} |G_{st}(z)| \prec \left(\frac{1}{N\eta}\right)^{1/4} \quad (3.5.3)$$

uniformly for $z \in \mathcal{D}(\tilde{c}, \tau)$.

3.5.1.1 Basic tools

We collect two lemmas that characterize some basic properties of $s(z)$.

Lemma 3.5.3. *There exists sufficiently small constant $\tilde{c} > 0$ independent of N , such that for $z \in \mathcal{D}(\tilde{c}, \tau)$,*

$$\begin{aligned} |s(z)| &\sim 1, \quad |1 + cs(z)| \sim 1, \\ \text{and} \quad \left| -z + \frac{d_i^2}{1 + cs(z)} - z\tilde{s}(z) \right| &\sim 1, \quad \text{for any } i = 1, \dots, M. \end{aligned} \quad (3.5.4)$$

Proof. Lemma 2.1 of Dozier and Silverstein (2007a) implies that $s(z) = O(1)$ for $z \in \mathcal{D}(\tilde{c}, \tau)$. And the lower bound of $|s(z)|$ can be implied by the last one in (3.5.4) assuming its validity. In the Appendix A in Loubaton et al. (2011), it has been proved that

$$\Re(1 + cs) \geq 1/2 \quad (3.5.5)$$

for $z \in \mathbb{C}^+ \cup \mathbb{R}^+$. Therefore, $|1 + cs| \geq 1/2$. The upper bound of $|1 + cs(z)|$ follows easily from $|s(z)| = O(1)$. So we mainly focus on the last estimate in this lemma.

Recall w defined in (3.2.9). Notice that

$$\frac{d_i^2}{1 + cs(\lambda_r)} - \lambda_r(1 + cs(\lambda_r)) + 1 - c = \frac{d_i^2 - w(\lambda_r)}{1 + cs(\lambda_r)} < -C\hat{c}, \quad (3.5.6)$$

where in the second step, we use Assumption 3 and (3.5.5). First we consider the case when z is in a small disk with center λ_r and radius ϵ , denoted by $C(\lambda_r, \epsilon)$. By analytic property of s , there exists small constant $\tilde{c}$, such that for any $z \in C(\lambda_r, \tilde{c}) \cap \mathbb{C}^+$,

$$|s(z) - s(\lambda_r)| \leq C'\hat{c}, \quad (3.5.7)$$

where the choice of C' and the associated $\tilde{c}$ should guarantee the second inequality of the following,

$$\begin{aligned} & \left| \frac{d_i^2}{1+cs(z)} - z(1+cs(z)) + 1 - c \right| - \left| \frac{d_i^2}{1+cs(\lambda_r)} - \lambda_r(1+cs(\lambda_r)) + 1 - c \right| \\ & \leq (4d_i^2c + \lambda_rc)|s(z) - s(\lambda_r)| + |z - \lambda_r| \sup_{z \in C(\lambda_r, \tilde{c}) \cap \mathbb{C}^+} |1+cs(z)| \leq \frac{C\hat{c}}{2}. \end{aligned} \quad (3.5.8)$$

It follows that

$$\left| \frac{d_i^2}{1+cs(z)} - z(1+cs(z)) + 1 - c \right| > \frac{C\hat{c}}{2} \quad \text{for } z \in C(\lambda_r, \tilde{c}) \cap \mathbb{C}^+. \quad (3.5.9)$$

Next, if $E > \lambda_r$, $s(x)$ is negative and increasing on $(\lambda_r, 1/\tau)$, with the fact that (3.5.5), then

$$\begin{aligned} \frac{d_i^2}{1+cs(E)} - E(1+cs(E)) + 1 - c & < \frac{d_i^2}{1+cs(\lambda_r)} - \lambda_r(1+cs(\lambda_r)) + 1 - c \\ & < -C\hat{c}. \end{aligned}$$

By the uniform continuity of $s(z)$ on $\mathcal{D}(\tilde{c}, \tau)$, we can find a sufficiently small constant η_0 , such that

$$\sup_{E \in [\lambda_r + \tilde{c}/2, 1/\tau]} |s(E + i\eta) - s(E)| \leq C''\eta_0 \quad \text{for } \eta \leq \eta_0.$$

Following similar arguments from (3.5.7) to (3.5.9), we have

$$\begin{aligned} & \left| \frac{d_i^2}{1+cs(z)} - z(1+cs(z)) + 1 - c \right| > \frac{C\hat{c}}{2} \\ & \text{for } z \in \{z | E \in [\lambda_r + \tilde{c}/2, 1/\tau], \eta \in [0, \eta_0]\}. \end{aligned} \quad (3.5.10)$$

Last, since $\Im s > 0$ and $\Im(zs) > 0$,

$$\Im\left(\frac{d_i^2}{1+cs(z)} - z(1+cs(z)) + 1 - c\right) \leq -\eta_0 \quad \text{for } z \in \{z | \eta > \eta_0\}. \quad (3.5.11)$$

Combining (3.5.9)(3.5.10)(3.5.11) together, we conclude that $\left| \frac{d_i^2}{1+cs(z)} - z(1+cs(z)) + 1 - c \right|$ are bounded below by some positive constant in the domain $\mathcal{D}(\tilde{c}, \tau)$. The upper bound exists due to $|1+cs(z)| \sim 1$. $\square$

Lemma 3.5.4. *Recall ρ_0 defined in (3.2.7). There exists a small constant $\delta > 0$, independent of N , such that*

$$\rho_0(x) \sim \sqrt{\lambda_r - x}, \text{ for any } x \in [\lambda_r - \delta, \lambda_r]. \quad (3.5.12)$$

Furthermore,

$$\Im s(z) \sim \begin{cases} \frac{\eta}{\sqrt{\kappa + \eta}}, & \text{if } E \geq \lambda_r + \eta, \\ \sqrt{\kappa + \eta}, & \text{if } E \in [\lambda_r - \delta, \lambda_r + \eta]. \end{cases} \quad (3.5.13)$$

Proof. Recalling that $z = \phi(w(z))$, $\lambda_r = \phi(\xi_r)$ and $\phi^{(1)}(\xi_r) = 0$ (see (3.2.11), (3.2.12)), we have for some small constant $\delta > 0$,

$$z - \lambda_r = \frac{1}{2}\phi^{(2)}(\xi_r)(w - \xi_r)^2 + O(|w - \xi_r|^3), \text{ for } |z - \lambda_r| \leq \delta,$$

where we use a simple fact $\phi^{(3)}(\xi_r) < C$ where C is a large constant. By calculations, we find there exists a positive constant c_1 such that

$$\phi^{(2)}(\xi_r) = \frac{1}{N} \sum_i \frac{\frac{-4d_i^2}{b} - 2(1-c)}{(d_i^2 - \xi_r)^3} + 2\xi_r \left(\frac{1}{N} \sum_i \frac{1}{(d_i^2 - \xi_r)^2} \right)^2 > c_1 > 0,$$

where we use Assumptions 2 and 3. Therefore, for $z = E \in \mathbb{R}$, $|E - \lambda_r| \leq \delta$, we have

$$w(E) - w(\lambda_r) \sim \sqrt{E - \lambda_r}.$$

It follows that $\Im w(E) \sim \sqrt{\lambda_r - E}$. Due to the relation $w(z) = zb^2(z) - (1-c)b(z)$ (see (3.2.9) and (3.2.13)), we can solve

$$b(E) = \frac{(1-c) + \sqrt{(1-c)^2 + 4Ew}}{2E}.$$

This, together with Lemma 3.5.3, implies $\Im b(E) \sim \Im w(E)$ through some calculations. Therefore, we find

$$\Im s(E) \sim \Im b(E) \sim \Im w(E) \sim \sqrt{\lambda_r - E} \quad \text{for } \lambda_r - \delta < E < \lambda_r.$$

This implies (3.5.12). (3.5.13) can be derived from (3.5.12) without difficulty, see Theorem 3.2 in Bao et al. (2015) for example. $\square$

3.5.1.2 A prior estimates on domain $\Omega(z)$

We start from finding estimates for diagonal entries of G . Using (3.3.28), write

$$\begin{aligned}
 \frac{1}{G_{ii}} &= -z - (\mathbf{Y}G^{(i)}\mathbf{Y}^*)_{ii} \\
 &= -z - d_i^2 G_{\bar{i}\bar{i}}^{(i)} - \sum_{\mu,\nu} x_{i\mu} G_{\mu\nu}^{(i)} x_{i\nu} - 2d_i \sum_{\nu} G_{\bar{i}\nu}^{(i)} x_{i\nu} \\
 &= -z - d_i^2 G_{\bar{i}\bar{i}}^{(i)} - z\tilde{s}_N^{(i)} - (1 - E_i) \sum_{\mu,\nu} x_{i\mu} G_{\mu\nu}^{(i)} x_{i\nu} - 2d_i \sum_{\nu} G_{\bar{i}\nu}^{(i)} x_{i\nu} \\
 &= -z - d_i^2 G_{\bar{i}\bar{i}}^{(i)} - z\tilde{s} + z(\tilde{s} - \tilde{s}_N) + \mathcal{T}_i,
 \end{aligned} \tag{3.5.14}$$

where

$$\mathcal{T}_i := z(\tilde{s}_N - \tilde{s}_N^{(i)}) - (1 - E_i) \sum_{\mu,\nu} x_{i\mu} G_{\mu\nu}^{(i)} x_{i\nu} - 2d_i \sum_{\nu} G_{\bar{i}\nu}^{(i)} x_{i\nu}, \tag{3.5.15}$$

and where $\tilde{s}_N^{(T)} := \frac{1}{N} \sum G_{\mu\mu}^{(T)}$. Furthermore,

$$\begin{aligned}
 \frac{1}{G_{\bar{i}\bar{i}}^{(i)}} &= -1 - (\mathbf{Y}^{(i)*} G^{(i\mu_i)} \mathbf{Y}^{(i)})_{\bar{i}\bar{i}} \\
 &= -1 - \sum_{j,k}^{(i)} x_{j\bar{i}} G_{jk}^{(i\bar{i})} x_{k\bar{i}} \\
 &= -1 - c \frac{1}{M} \text{Tr} G^{(i\bar{i})} - (1 - E_{\bar{i}}) \sum_{j,k}^{(i)} x_{j\bar{i}} G_{jk}^{(i\bar{i})} x_{k\bar{i}} \\
 &= -1 - cs + c(s - s_N) + \mathcal{S}_i,
 \end{aligned} \tag{3.5.16}$$

where

$$\mathcal{S}_i := c(s_N - s_N^{(i\bar{i})}) - (1 - E_{\bar{i}}) \sum_{j,k}^{(i)} x_{j\bar{i}} G_{jk}^{(i\bar{i})} x_{k\bar{i}}. \tag{3.5.17}$$

Define control quantities:

$$\Lambda_o := \max_{s \neq t, (s,t) \text{ is not index pair}} |G_{st}|, \quad \Lambda := |s_N - s|, \quad \Psi_\Lambda := \sqrt{\frac{\Lambda + \Im s}{N\eta}} + \frac{1}{N\eta} \tag{3.5.18}$$

and an event $\Omega(z)$ by

$$\Omega(z) := \{\Lambda_o \leq N^{-\tau/10}\} \cap \{\Lambda \leq N^{-\tau/10}\}. \tag{3.5.19}$$

Lemma 3.5.5. *For any $t \in \mathcal{I}$, the following statement holds with high probability on $\Omega(z)$,*

$$C^{-1} \leq |G_{tt}| \leq C \quad (3.5.20)$$

Proof. We only prove the case when $t \in \mathcal{I}_M$. For the case $t \in \mathcal{I}_N$, the proof is similar. First, we bound $\mathcal{S}_i$ defined above. By Lemma 3.3.7 we have

$$\begin{aligned} |(1 - \mathbb{E}_i) \sum_{j,k}^{(i)} x_{j\bar{i}} G_{jk}^{(i\bar{i})} x_{k\bar{i}}| &\leq \sum_j^{(i)} (|x_{j\bar{i}}|^2 - \frac{1}{N}) |G_{jj}^{(i\bar{i})}| + |\sum_{j \neq k}^{(i)} x_{j\bar{i}}^* G_{jk}^{(i\bar{i})} x_{k\bar{i}}| \\ &\prec (\frac{1}{N^2} \sum_{j,k}^{(i)} |G_{jk}^{(i\bar{i})}|^2)^{1/2} \leq (\frac{1}{N^2 \eta} \sum_j^{(i)} \Im G_{jj}^{(i\bar{i})})^{1/2} \\ &\leq \left(\frac{\sum_j^{(i)} \Im(G_{jj}^{(i\bar{i})} - G_{jj}) + M \Im s_N}{N^2 \eta} \right)^{1/2} \\ &\leq \left(\frac{\Lambda_o^2}{N \eta} + \frac{c \Lambda + c \Im s}{N \eta} \right)^{1/2} \end{aligned} \quad (3.5.21)$$

where in the last step we use (3.3.32). And the above term can also be bounded by Ψ_Λ if we use interlacing relationship of eigenvalues of Hermitian matrices in the last step above. Therefore, on Ω we have

$$|\mathcal{S}_i| = o(1), \quad (3.5.22)$$

and

$$|\frac{1}{G_{ii}^{(i)}}| = |-1 - cs| + o(1). \quad (3.5.23)$$

Moreover we have

$$\Im G_{ii}^{(i)} \prec \Im s + \Lambda + \Psi_\Lambda, \quad (3.5.24)$$

which can be deduced from (3.5.16), (3.5.21) and (3.5.23).

Then, we consider $1/G_{ii}$. Since

$$|(1 - \mathbb{E}_i) \sum_{\mu, \nu} x_{i\mu} G_{\mu\nu}^{(i)} x_{i\nu}| \prec (\frac{1}{N^2} \sum_{\mu, \nu} |G_{\mu\nu}^{(i)}|^2)^{1/2} \prec \frac{1}{N \eta} + \sqrt{\frac{\Im \tilde{s} + \Lambda}{N \eta}},$$

and

$$|\sum_{\nu} G_{i\nu}^{(i)} x_{i\nu}| \prec (\frac{1}{N} \sum_{\nu} |G_{i\nu}^{(i)}|^2)^{1/2} \leq (\frac{\Im G_{ii}^{(i)}}{N\eta})^{1/2} \prec \Psi_{\Lambda}, \quad (3.5.25)$$

where we use (3.5.24) in the last step, we then have

$$|\frac{1}{G_{ii}}| = |-z - d_i^2 G_{ii}^{(i)} - z\tilde{s}| + o(1) = |-z + \frac{d_i^2}{1+cs} - z\tilde{s}| + o(1)$$

holds with high probability on $\Omega(z)$. Using Lemma 3.5.3 we conclude that G_{ii} is bounded from both above and below. $\square$

Lemma 3.5.6. *For all $z \in \mathcal{D}(\tilde{c}, \tau)$, we have for $i \in \mathcal{I}_M$,*

$$(1(\Omega) + \mathbf{1}(\eta \geq 1))(\Lambda_o + |\mathcal{S}_i| + |\mathcal{T}_i|) \prec \Psi_{\Lambda}$$

and

$$(1(\Omega) + \mathbf{1}(\eta \geq 1)) \left(|G_{ii} - \frac{b(z)}{d_i^2 - w(z)}| + |G_{i\bar{i}} - \frac{d_i}{d_i^2 - w(z)}| + |G_{i\bar{i}} - \frac{\tilde{b}(z)}{d_i^2 - w(z)}| \right) \prec \Psi_{\Lambda} + \Lambda. \quad (3.5.26)$$

For $\mu \in \llbracket 2M + 1, M + N \rrbracket$,

$$(1(\Omega) + \mathbf{1}(\eta \geq 1)) |G_{\mu\mu} + \frac{1}{b(z)}| \prec \Psi_{\Lambda} + \Lambda. \quad (3.5.27)$$

Remark 17. By the definition of $\prec$, the result can be directly extended to maximum of variables over countable indices, such as $\max_i |\mathcal{S}_i| \prec \Psi_{\Lambda}$, and $\max_i |G_{ii} - b(z)/(d_i^2 - w(z))| \prec \Psi_{\Lambda} + \Lambda$.

Proof. We consider the estimation of quantities in this lemma under the event Ω first and we omit the indicator function for simplicity. We start with G_{ij} for $i \neq j \in \mathcal{I}_M$. Applying (3.3.29), we have

$$\begin{aligned} G_{ij} &= G_{ii} G_{jj}^{(i)} (Y G^{(ij)} Y^*)_{ij} \\ &= G_{ii} G_{jj}^{(i)} (d_i G_{i\bar{j}}^{(ij)} d_j + d_i \sum_{\nu} G_{i\nu}^{(ij)} x_{j\nu} + d_j \sum_{\nu} x_{i\nu} G_{\nu\bar{j}}^{(ij)} + \sum_{\mu, \nu} x_{i\mu} G_{\mu\nu}^{(ij)} x_{j\nu}). \end{aligned} \quad (3.5.28)$$

For $G_{\bar{i}\bar{j}}^{(ij)}$, according to (ii) and (v) of Lemma 3.3.8 it can be written as

$$\begin{aligned} G_{\bar{i}\bar{j}}^{(ij)} &= G_{\bar{i}\bar{i}}^{(ij)} G_{\bar{j}\bar{j}}^{(ij\bar{i})} (Y^{*(ij)} G^{(ij\bar{i})} Y^{(ij)})_{\bar{i}\bar{j}} \\ &= G_{\bar{i}\bar{i}}^{(ij)} G_{\bar{j}\bar{j}}^{(ij\bar{i})} \sum_{k,l}^{(ij)} x_{k\bar{i}}^{(ij)} G_{kl}^{(ij\bar{i})} x_{l\bar{j}}^{(ij)}. \end{aligned} \quad (3.5.29)$$

By Lemma 3.3.7, $|\sum_{k,l}^{(ij)} x_{k\bar{i}}^{(ij)} G_{kl}^{(ij\bar{i})} x_{l\bar{j}}^{(ij)}| \prec \Psi_\Lambda$. We also have that $G_{\bar{i}\bar{i}}^{(ij)}$ and $G_{\bar{j}\bar{j}}^{(ij\bar{i})}$ are of constant order from (3.5.23) and (3.3.32). Then the first term inside the bracket of (3.5.28) is stochastically dominated by Ψ_Λ . For the second term,

$$|\sum_{\nu} G_{\bar{i}\bar{\nu}}^{(ij)} x_{\nu\bar{j}}^*| \prec \sqrt{\frac{\Im G_{\bar{i}\bar{i}}^{(ij)}}{N\eta}} \prec \Psi_\Lambda + \sqrt{\frac{\Lambda_o^2}{N\eta}},$$

where in the last step we use (3.5.24) and (3.3.32). The last term in (3.5.28) is also bounded by Ψ_Λ . Therefore, $|G_{ij}| \prec \Psi_\Lambda + (N\eta)^{-1/2} \Lambda_o \prec \Psi_\Lambda + N^{-\tau/4} \Lambda_o$. The proof of $|G_{\mu\nu}| \prec \Psi_\Lambda + N^{-\tau/4} \Lambda_o$ for $\mu \neq \nu \in \mathcal{I}_N$ is similar.

To bound $G_{\mu i}$ for $\mu \in \mathcal{I}_N$ and $i \in \mathcal{I}_M$, by (3.3.31) we have

$$G_{\mu i} = -G_{ii}(d_i G_{\mu\bar{i}}^{(i)} + \sum_{\nu} G_{\mu\nu}^{(i)} x_{i\nu}). \quad (3.5.30)$$

Using (3.3.32) and (3.3.31) we find

$$G_{\mu\bar{i}}^{(i)} = G_{\mu\bar{i}} - \frac{G_{\mu i} G_{\bar{i}\bar{i}}}{G_{ii}}, \quad (3.5.31)$$

$$G_{\bar{i}\bar{i}} = -G_{ii}(d_i G_{\bar{i}\bar{i}}^{(i)} + \sum_{\nu} x_{i\nu} G_{\nu\bar{i}}^{(i)}). \quad (3.5.32)$$

Substituting (3.5.32) into (3.5.31), and then together with (3.5.30), we get

$$(1 + d_i^2 G_{ii} G_{\bar{i}\bar{i}}^{(i)} + d_i G_{ii} \sum_{\nu} x_{i\nu} G_{\nu\bar{i}}^{(i)}) G_{\mu i} = -d_i G_{ii} G_{\mu\bar{i}} - G_{ii} \sum_{\nu} G_{\mu\nu}^{(i)} x_{\nu i}^*. \quad (3.5.33)$$

By equation (3.5.14),

$$1 + d_i^2 G_{ii} G_{\bar{i}\bar{i}}^{(i)} = -z(1 + \tilde{s}) G_{ii} + z(\tilde{s} - \tilde{s}_N) G_{ii} + G_{ii} \mathcal{T}_i. \quad (3.5.34)$$

Therefore, if $\mu \neq \bar{i}$, the absolute value of coefficient of $G_{\mu i}$ in (3.5.33) is bounded

from below by a constant based on (3.5.4), (3.5.19), (3.5.20) and similar argument to (3.5.25). The right hand side in (3.5.33) is stochastically controlled by Ψ_Λ . Then we get that $|G_{\mu i}| \prec \Psi_\Lambda$. Similarly, it can be shown that $|G_{i\mu}| \prec \Psi_\Lambda$, for $\mu \neq \bar{i}$. We conclude that

$$\Lambda_o \prec \Psi_\Lambda.$$

$|\mathcal{S}_i| \prec \Psi_\Lambda$ has been shown in Lemma 3.5.5, and for $|\mathcal{T}_i|$ the proof is similar.

What remains is the estimate (3.5.26) and (3.5.27). Recall w and $\tilde{b}$ defined in (3.2.9) and (3.3.9). Using the expansion (3.5.37) below it follows easily that $|G_{ii} - \frac{b}{d_i^2 - w}| \prec \Psi_\Lambda + \Lambda$. By a similar strategy, we can control $G_{\bar{i}\bar{i}}$ and $G_{\mu\mu}$. To estimate $G_{\bar{i}\bar{i}}$, we use (3.5.32). Since $|G_{ii} - \frac{b}{d_i^2 - w}| \prec \Psi_\Lambda + \Lambda$ and $|G_{\bar{i}\bar{i}}^{(i)} - \frac{1}{-\tilde{b}}| \prec \Psi_\Lambda$ where the latter is inferred from (3.5.16), together with (3.5.25), we can get

$$\left| G_{\bar{i}\bar{i}} - \frac{d_i}{d_i^2 - w} \right| \prec \Psi_\Lambda + \Lambda.$$

The proof under the event $1(\eta \geq 1)$ can be proceeded similar, but in some steps we use trivial bound $\|G^{(T)}\| \leq 1$. We omit these steps and conclude the proof. $\square$

Denote $[v] = \frac{1}{M} \sum_i G_{ii} - s$. We now put $[v]$ into a form that we may control it, see Lemma 3.5.7 for a rough control that is sufficient for the proof of Lemma 3.5.2. Recall $b(z) = 1 + cs(z)$ and for simplicity we omit the dependence on z . Let

$$t_i := \left(-z + \frac{d_i^2}{b} - z\tilde{s}\right)^{-1}.$$

Define

$$R_2(z) := \frac{1}{M} \sum_{i=1}^M c \left(\frac{d_i^2}{b^2} + z \right) t_i^2, \quad R_3(z) := \frac{1}{M} \sum_{i=1}^M \left(-c^2 \frac{d_i^2}{b^3} t_i^2 + \left(\frac{d_i^2 c}{b^2} + cz \right)^2 t_i^3 \right). \quad (3.5.35)$$

Lemma 3.5.7. *Recall $\mathcal{T}_i$ and $\mathcal{S}_i$ defined in (3.5.15) and (3.5.17). We have on $\Omega(z)$ or $1(\eta \geq 1)$, with high probability,*

$$|(1 - R_2(z))[v] - R_3(z)[v]^2| \leq C(\max_i \{|\mathcal{T}_i| + |\mathcal{S}_i| + |\Lambda|^3 + |\mathcal{S}_i||\Lambda|\}). \quad (3.5.36)$$

Proof. Using (3.5.14), (3.5.16) and trivial equality

$$\frac{1}{x-y} = \frac{1}{x} + \frac{y}{x^2} + \frac{y^2}{x^3} + \frac{y^3}{(x-y)x^3},$$

we have that

$$\begin{aligned} \frac{1}{G_{ii}} &= -z + \frac{d_i^2}{b - \mathcal{S}_i + c(s - s_N)} - z\tilde{s}_N + \mathcal{T}_i \\ &= -z + \frac{d_i^2}{b} - z\tilde{s} + \frac{d_i^2(\mathcal{S}_i + c(s - s_N))}{b^2} + \frac{d_i^2(\mathcal{S}_i + c(s - s_N))^2}{b^3} \\ &\quad + \frac{d_i^2(\mathcal{S}_i + c(s - s_N))^3}{b^3[(b - \mathcal{S}_i - c(s - s_N))]} + z(\tilde{s} - \tilde{s}_N) + \mathcal{T}_i. \end{aligned}$$

Denote

$$D_i = \frac{d_i^2(\mathcal{S}_i + c(s - s_N))}{b^2} + \frac{d_i^2(\mathcal{S}_i + c(s - s_N))^2}{b^3} + \frac{d_i^2(\mathcal{S}_i + c(s - s_N))^3}{b^3[(b - \mathcal{S}_i - c(s - s_N))]}$$

Then it follows that

$$G_{ii} = t_i - [z(\tilde{s} - \tilde{s}_N) + D_i + \mathcal{T}_i] t_i^2 + [z(\tilde{s} - \tilde{s}_N) + D_i + \mathcal{T}_i]^2 t_i^3 + G_{ii}^{-1} [z(\tilde{s} - \tilde{s}_N) + D_i + \mathcal{T}_i]^3 t_i^3 \quad (3.5.37)$$

Since $|-z + \frac{d_i^2}{b} - z\tilde{s}|$ is bounded from below known from Lemma 3.5.3, we can derive a second order equation of $[v]$ from above equation by averaging over $G_{ii}, i = 1 \dots, M$. Recalling $||v|| = \Lambda$, and using lemma 3.5.6, we conclude that (3.5.36) holds with high probability. $\square$

Lemma 3.5.8. *Let $\alpha(z) \equiv \alpha := |1 - R_2|$, we have the following estimates for R_2 and R_3 in (3.5.35):*

1. $\alpha(z) \sim \sqrt{\kappa + \eta}$ for $z \in \mathcal{D}(\tilde{c}, \tau)$.
2. $|R_3(z)|$ is uniformly bounded above for $z \in \mathcal{D}(\tilde{c}, \tau)$. And there exist small constants c' and ϵ_0 , such that $|R_3(z)| \geq c'$ whenever $z \in \mathcal{D}(\tilde{c}, \tau)$ satisfies $|z - \lambda_r| < \epsilon_0$.

Proof. We first introduce some notations for ease of presentation:

$$g(b) := \frac{1}{M} \sum_{i=1}^M \frac{cd_i^2 |t_i|^2}{|b|^2}, \quad G(b) := \frac{1}{M} \sum_{i=1}^M c |t_i|^2 \quad (3.5.38)$$

Both $g(b)$ and $G(b)$ are bounded from above and below, which can be inferred from Lemma 3.5.3.

For statement 1 about $\alpha(z)$, we first consider upper bound for $\alpha(z)$. We first show that for $z \in \mathcal{D}(\tilde{c}, \tau)$,

$$|1 - g(b) - zG(b)| = O(\sqrt{\kappa + \eta}). \quad (3.5.39)$$

Using (3.2.6) where we replace ρ by $\hat{\rho}$, we obtain

$$c\Im s = g(b)\Im b + G(b)\Im(zb),$$

then it follows that

$$\begin{aligned} |1 - g(b) - zG(b)| &= |1 - \frac{c\Im s - G(b)\Im(zb)}{\Im b} - zG(b)| = |\frac{G(b)\Im(zb)}{\Im b} - zG(b)| \\ &= |\frac{(E\Im b + \eta\Re b)G(b)}{\Im b} - \frac{(E + i\eta)\Im b G(b)}{\Im b}| \\ &= |\frac{\eta G(b)\bar{b}}{\Im b}|. \end{aligned} \quad (3.5.40)$$

Then Lemma 3.5.4 implies (3.5.39). Next, we consider

$$\begin{aligned} g(b) + zG(b) - R_2 &= \frac{1}{M} \sum_{i=1}^M (\frac{cd_i^2 |t_i|^2}{|b|^2} + cz|t_i|^2) - \frac{1}{M} \sum_{i=1}^M (\frac{cd_i^2 t_i^2}{b^2} + cz t_i^2) \\ &= \frac{1}{M} \sum_{i=1}^M cd_i^2 |t_i|^2 (\frac{1}{|b|^2} - \frac{1}{b^2}) + \frac{1}{M} \sum_{i=1}^M c(\frac{d_i^2}{b^2} + z)(|t_i|^2 - t_i^2) \\ &= A_1 + A_2 \end{aligned}$$

Using a simple fact

$$\frac{1}{|z|^2} - \frac{1}{z^2} = \frac{2(\eta + Ei)\eta}{|z|^4},$$

it follows that

$$|A_1| = O(\Im b), \quad |A_2| \leq C\Im(\frac{d_i^2}{b} - zb) = O(\Im b).$$

Then

$$|g(b) + zG(b) - R_2| = O(\Im b) = O(\sqrt{\kappa + \eta}). \quad (3.5.41)$$

We then have $|\alpha(z)| \leq C(\sqrt{\kappa + \eta})$ from (3.5.39) and (3.5.41).

For the lower bound of $\alpha(z)$, we first consider the case when $\lambda_r - \tilde{c} \leq E \leq \lambda_r$,

and $0 < \eta < \eta_0$, where η_0 is determined by the way in Lemma 3.5.3. We first collect some basic control on real and imaginary part of some terms used later. Based on the proof of Lemma 3.5.3, and recall that $\omega = zb^2 - (1 - c)b$. we have

$$\max\{\Re \frac{d_i^2 - \omega}{b}, \Re(d_i^2 - \omega)\} < -C\hat{c}. \quad (3.5.42)$$

We also have

$$\begin{aligned} \Im(d_i^2 - w) &= \Im\left(\frac{d_i^2 - \omega}{b}b\right) = 2\left(\Re \frac{d_i^2 - \omega}{b}\Im b + \Im \frac{d_i^2 - \omega}{b}\Re b\right) \\ &\leq -C\hat{c}\Im b, \end{aligned} \quad (3.5.43)$$

and

$$\Im\left(\frac{b}{d_i^2 - \omega}\right)^2 = -C\Im\left(\frac{d_i^2 - \omega}{b}\right)^2 = -2C\Re \frac{d_i^2 - \omega}{b}\Im \frac{d_i^2 - \omega}{b} < 0. \quad (3.5.44)$$

Now, for $\alpha(z)$, it can be bounded from below by

$$|\alpha(z)| = |1 - R_2(z)| \geq \left| -\frac{1}{M} \sum_{i=1}^M \Im \frac{d_i^2 + zb^2}{(d_i^2 - \omega)^2} \right|. \quad (3.5.45)$$

For the first imaginary part term, we have

$$-\frac{1}{M} \sum_{i=1}^M \Im \frac{d_i^2}{(d_i - \omega)^2} = \frac{1}{M} \sum_{i=1}^M \frac{2d_i^2 \Re(d_i^2 - \omega) \Im(d_i^2 - \omega)}{|d_i^2 - \omega|^4} \geq C\hat{c}^2 \Im b \geq C\sqrt{\kappa + \eta},$$

where in the second step, we use (3.5.42) and (3.5.43), and in the last step, we use Lemma 3.5.4. For the second part in (3.5.45), we have

$$-\frac{1}{M} \sum_{i=1}^M \Im \frac{zb^2}{(d_i - \omega)^2} = -\frac{1}{M} \sum_{i=1}^M E \Im \left(\frac{b}{d_i^2 - \omega} \right)^2 - \eta \Re \left(\frac{b}{d_i^2 - \omega} \right)^2 > -C\eta, \quad (3.5.46)$$

where in the last step, we use (3.5.44) and $\Re \left(\frac{b}{d_i^2 - \omega} \right)^2 = O(1)$. Hence, we can choose a small constant η'_0 smaller than η_0 such that $|\alpha(z)| \geq C\sqrt{\kappa + \eta}$ for all z satisfying $\lambda_r - \tilde{c} \leq E \leq \lambda_r$, and $0 < \eta < \eta'_0$.

For the other cases, we are going to obtain the lower bound from

$$|1 - R_2(z)| \geq 1 - g(b) - |z|G(b) = \frac{[1 + (1 - c)G(b)]\eta G(b)}{\Im b[1 - g(b) + |z|G(b)]}, \quad (3.5.47)$$

where the second step is obtained from the argument between (2.3) and (2.5) in Dozier and Silverstein (2007a). We have $0 < 1 - g(b) + |z|G(b) \leq C$ inferred from (2.5) in Dozier and Silverstein (2007a). Recall that $c \leq 1$, $G(b)$ is bounded from below, then, if $\lambda_r - \tilde{c} \leq E \leq \lambda_r$, and $\eta > \eta'_0$, we have

$$|1 - R_2(z)| \geq C \frac{\eta}{\sqrt{\kappa + \eta}} \geq C\sqrt{\kappa + \eta}$$

and if $\lambda_r \leq E \leq \tau^{-1}$, regardless of the order of η , we always have

$$|1 - R_2(z)| \geq C\sqrt{\kappa + \eta}.$$

For statement 2 about estimation of $R_3(z)$ in this lemma, the upper bound can be easily derived from Lemma 3.5.3. For the lower bound, we have

$$R_3(\lambda_r) = \frac{c^2}{M} \sum_{i=1}^M \left(\frac{3\lambda_r d_i^2}{b(\lambda_r)^2} - \frac{d_i^2(1-c)}{b(\lambda_r)^3} + \lambda_r^2 \right) (t_i(\lambda_r))^3. \quad (3.5.48)$$

By (3.5.6) and recall that $b(\lambda_r) > 0$ and $c \leq 1$, we have

$$\lambda_r^2 > \left(\frac{d_i^2}{b(\lambda_r)^2} + \frac{1-c}{b(\lambda_r)} \right)^2 \geq \frac{d_i^4}{b(\lambda_r)^4} + \frac{(1-c)^2}{b(\lambda_r)^2} \geq 2 \frac{d_i^2(1-c)}{b(\lambda_r)^3}.$$

Combining above with (3.5.48), and using Lemma 3.5.3, we have $|R_3(\lambda_r)|$ is bounded from below. Since $R_3(z)$ is an analytic function of z in the neighborhood of λ_r , this proves the statement 2 in the lemma. $\square$

We also need the following estimate for Λ when $\eta \sim 1$.

Lemma 3.5.9. *We can find η_b sufficiently large, such that for any z in $\mathcal{D}$ satisfying $\eta \geq \eta_b$, and any $\epsilon > 0$,*

$$\Lambda \leq \frac{N^\epsilon}{\sqrt{N}}$$

holds with high probability.

Proof. Using Lemma 3.5.6, Lemma 3.5.7 and Lemma 3.5.8, we have $\sqrt{\eta}\Lambda \leq C\Lambda^2 + CN^\epsilon\Psi_\Lambda$ holds with high probability, where C is a large constant independent of choice of z . Using trivial estimates

$$|G_{st}^{(\mathbb{T})}| \leq \frac{1}{\eta}, \quad |s_N^{(\mathbb{T})}| \leq \frac{1}{\eta}, \quad |s| \leq \frac{1}{\eta},$$

it follows that $\Lambda \leq 2/\eta$. If we can find sufficiently large η such that $2C/\eta \leq \sqrt{\eta}/2$, then $\sqrt{\eta}\Lambda \leq 2C\Lambda/\eta + CN^\epsilon\Psi_\Lambda \leq \sqrt{\eta}\Lambda/2 + CN^\epsilon\Psi_\Lambda$, which implies the lemma. $\square$

3.5.1.3 Proof of Lemma 3.5.2

Let $\beta(z) := (N\eta)^{-1/4}$. Fix $\epsilon \in (0, \tau/20)$. For any fixed E , since $\sqrt{\kappa + \eta}$ is increasing and $\beta(E + i\eta)$ is decreasing in η , the equation

$$\sqrt{\kappa + \eta} = N^{2\epsilon}\beta(E + i\eta) \quad (3.5.49)$$

has a unique solution $\tilde{\eta}$. Note that $\tilde{\eta} \ll 1$.

We have that

$$\Psi_\Lambda^2 \leq 2\left(\frac{\Lambda + \Im s}{N\eta} + \frac{1}{(N\eta)^2}\right) \leq 2(\beta^4\Lambda + \beta^4\alpha + \beta^8)$$

which implies

$$\Psi_\Lambda \leq 2\sqrt{\beta^4\Lambda + \beta^4\alpha + \beta^8} = 2\beta\sqrt{\beta^2\Lambda + \beta^2\alpha + \beta^6} \leq \beta^2\Lambda + \beta^2\alpha + 2\beta^2.$$

Before proceeding the proof, we need a dichotomy based on whether $\eta \geq \tilde{\eta}$ or $\eta < \tilde{\eta}$ first. To make the following statement clear, we specify some constants to control the $\alpha(z)$ and $R(z)$ in Lemma (3.5.8). Specifically, we suppose $C_1\sqrt{\kappa + \eta} \leq \alpha(z) \leq C_2\sqrt{\kappa + \eta}$, and $|R_3(z)| \leq C_3$ for any $z \in \mathcal{D}(\tilde{c}, \tau)$. Then depending on the relative size of α and β , we consider the bound for Λ separately.

Case 1. $\eta \geq \tilde{\eta}$. On $\Omega(z)$, with probability $1 - N^{-D}$, where D is any positive constant, by Lemma 3.5.7, we have that for sufficiently large N ,

$$\alpha\Lambda \leq C_3\Lambda^2 + N^\epsilon(\beta^2\Lambda + \beta^2\alpha + 2\beta^2).$$

Using $N^\epsilon\beta \ll \alpha$, it follows that

$$\alpha\Lambda \leq C_3\Lambda^2 + C_4\alpha\beta.$$

Thus, we either have

$$\Lambda \geq \frac{\alpha}{2C_3} \quad \text{or} \quad \Lambda \leq 2C_4\beta. \quad (3.5.50)$$

Case 2. $\eta < \tilde{\eta}$. In this case, note that $\sqrt{\kappa + \eta} < N^{2\epsilon}\beta(E + i\eta) = o(1)$, then $\kappa = o(1)$, and it follows that $R_3(z) \geq c'$. Using Lemma 3.5.7,

$$c'\Lambda^2 \leq \alpha\Lambda + CN^\epsilon(\beta^2\Lambda + \beta^2\alpha + 2\beta^2) \leq N^{2\epsilon}\beta\Lambda + CN^\epsilon(\beta^2\Lambda + \beta^2\alpha + 2\beta^2) \leq \frac{c'\Lambda^2}{2} + 4N^{4\epsilon}\beta^2,$$

which implies

$$\Lambda \leq \sqrt{\frac{8}{c'}} N^{2\epsilon}\beta. \quad (3.5.51)$$

After obtaining the dichotomy, we come to the main proof. Define discrete set $\hat{D} = \mathcal{D}(\tilde{c}, \tau) \cap (N^{-3}\mathbb{Z}^2)$. We will first show that for $z \in \hat{D}$, $\Lambda \prec (N\eta)^{-1/4}$. For fixed E , let $z_0 = E + \eta_b i$, and define $z_k := E + \eta_k i$ with $\eta_k = \eta_b - kN^{-3}$. Define the event $\Omega_k := \Omega(z_k) \cap \{\Lambda(z_k) \leq C^{(k)}N^{2\epsilon}\beta(z_k)\}$, where

$$C^{(k)} := \begin{cases} C_1/4C_3, & \text{if } \eta_k \geq \tilde{\eta}, \\ \sqrt{8/c'}, & \text{if } \eta_k < \tilde{\eta}. \end{cases}$$

Next we will show that for any $D > 0$ and any k ,

$$P(\Omega_k^c) \leq 3kN^{-D}. \quad (3.5.52)$$

We use induction method. When $k = 1$, this has been proved by Lemma 3.5.6 and Lemma 3.5.9. Assume that (3.5.52) holds for some $k \geq 1$, then

$$P(\Omega_{k+1}^c) \leq P(\Omega_k \cap \Omega(z_{k+1}) \cap \Omega_{k+1}^c) + P(\Omega_k \cap (\Omega(z_{k+1}))^c) + P(\Omega_k^c) := B + A + P(\Omega_k^c),$$

where

$$\begin{aligned} A &:= P(\{\Omega_k \cap \{\Lambda(z_{k+1}) > N^{-\frac{\tau}{10}}\}\} \cup \{\Omega_k \cap \{\Lambda_o(z_{k+1}) > N^{-\frac{\tau}{10}}\}\}), \\ B &:= P(\Omega_k \cap \Omega(z_{k+1}) \cap \{\Lambda(z_{k+1}) > C^{(k+1)}N^{2\epsilon}\beta(z_{k+1})\}). \end{aligned}$$

We estimate A first. By Lipschitz continuity of resolvent map: $z \mapsto G(z)$, $z \in \mathbb{C}^+$, we have

$$|G_{st}(z_{k+1}) - G_{st}(z_k)| \leq |z_{k+1} - z_k| \sup_{z \in \mathcal{D}} |G'_{st}(z)| \leq N^{-1}.$$

Thus

$$\Lambda(z_{k+1}) \leq \Lambda(z_k) + N^{-1} \leq CN^{2\epsilon}\beta_k < N^{\frac{\tau}{10}}N^{-\frac{\tau}{4}} < N^{-\frac{\tau}{10}},$$

and similarly, $\Lambda_o(z_{k+1}) < N^{-\frac{\tau}{10}}$. Thus $A \leq N^{-D}$. To estimate B , recalling that $\tilde{\eta}$ is the solution to the equation in (3.5.49), suppose that $\eta_k > \eta_{k+1} > \tilde{\eta}$, then

$$\Lambda(z_{k+1}) \leq \Lambda(z_k) + N^{-1} \leq \frac{C_1}{4C_3} N^{2\epsilon} \beta_k + N^{-1} < \frac{\alpha}{2C_3}.$$

By the dichotomy argument in (3.5.50), we have that with probability $1 - N^{-D}$, $\Lambda(z_{k+1}) \leq 2C_4\beta$. If $\eta_k < \tilde{\eta}$, (3.5.51) implies $\Lambda(z_{k+1}) \leq \sqrt{8/c'} N^{2\epsilon} \beta_{k+1}$. Thus $B \leq N^{-D}$. By the inductive assumption, (3.5.52) concludes.

To finish the proof, we need to extend the result for $z \in \hat{D}$ to $z \in \mathcal{D}(\tilde{c}, \tau)$. Similar to the argument for the proof of Corollary 3.24 in Lee and Schnelli (2013), we can also show that there exists a constant C ,

$$P\left(\bigcup_{z \in \mathcal{D}(\tilde{c}, \tau)} \Omega(z)^c\right) + P\left(\bigcup_{z \in \mathcal{D}(\tilde{c}, \tau)} \{\Lambda(z) > CN^\epsilon \beta(z)\}\right) \leq N^{-D'} \quad (3.5.53)$$

for arbitrary large D' . This can be proceeded by the Lipschitz continuity of the resolvent map: $z \mapsto G(z)$, $z \in \mathbb{C}^+$ and the error parameter β . Therefore, we conclude the proof for Λ part. We can also conclude the proof for the Λ_o part by using Lemma 3.5.6, (3.5.53) and the lattice argument. $\square$

3.5.2 Fluctuation lemma and entrywise local law

In this subsection we first prove a fluctuation lemma (see Lemma 3.5.10 below). Similar results for other random matrix models appear in Erdős et al. (2013a,b); Lee and Schnelli (2013), e.t.c., see also Section 6 of Benaych-Georges and Knowles (2016) whose argument we follow. However, there exist slight differences between our cases with previous results mainly because the particularity of y_{ii} , thus we need to analyze carefully. Then combining the fluctuation lemma with the weak local law in Lemma 3.5.2, we finish the proof of Lemma 3.5.1.

3.5.2.1 Fluctuation Lemma

Recall H defined in (3.3.2). For $s \in \mathcal{I}$, we define operations P_s and Q_s by

$$P_s[\cdot] := E[\cdot | H^{(s)}], \quad Q_s := 1 - P_s.$$

Lemma 3.5.10. *Suppose that for $z \in \mathcal{D}(\tilde{c}, \tau)$, we have $\Lambda_o \prec \Upsilon$, where $N^{-1/2} \leq \Upsilon \leq N^{-d}$ for some $d > 0$. Then for $z \in \mathcal{D}(\tilde{c}, \tau)$, we have*

$$\begin{aligned} \frac{1}{M} \sum_{i=1}^M Q_i\left(\frac{1}{G_{ii}}\right) &= O_{\prec}(\Upsilon^2), \\ \frac{1}{M} \sum_{i=1}^M g_{1i} \mathcal{T}_i &= O_{\prec}(\Upsilon^2), \\ \frac{1}{M} \sum_{i=1}^M g_{2i} \mathcal{S}_i &= O_{\prec}(\Upsilon^2), \end{aligned} \tag{3.5.54}$$

where (g_{1i}) and (g_{2i}) are uniformly bounded constants.

Proof. The first estimate in (3.5.54) can be proved by the analogous method used in the proof of Proposition 6.1 in Benaych-Georges and Knowles (2016). One can also refer to the proofs of Lemma 4.9 in Bloemendal et al. (2014) or Lemma 7.1 in Pillai et al. (2014). Recall $\mathcal{T}_i$ introduced in (3.5.15). For the second estimate in (3.5.54), we find

$$\mathcal{T}_i = Q_i\left(\frac{1}{G_{ii}}\right) + z(\tilde{s}_N - \tilde{s}_N^{(i)}) = Q_i\left(\frac{1}{G_{ii}}\right) - cz(s_N - s_N^{(i)}).$$

Then we have

$$\frac{1}{M} \sum_{i=1}^M \mathcal{T}_i = \frac{1}{M} \sum_{i=1}^M Q_i\left(\frac{1}{G_{ii}}\right) + O(\Lambda_o^2) = \frac{1}{M} \sum_{i=1}^M Q_i\left(\frac{1}{G_{ii}}\right) + O_{\prec}(\Upsilon^2),$$

where in the first equality we use (3.3.32), and in the second equality we use Lemma 3.5.5 and $M^{-1/2} \leq \Upsilon$. Then if $g_{1i} = 1, i = 1, \dots, M$, the second estimate in (3.5.54) follows. For general uniformly bounded (g_{1i}) , the conclusion also holds since we used high moment method to estimate $\sum_{i=1}^M Q_i(\frac{1}{G_{ii}})$, and we can deal with $\sum_{i=1}^M g_{1i} Q_i(\frac{1}{G_{ii}})$ similarly.

For the third estimate, similar to the above, we just need to consider $g_{2i} = 1, i = 1, \dots, M$. Recall $\mathcal{S}_i$ introduced in (3.5.17), we have

$$\frac{1}{M} \sum_{i=1}^M \mathcal{S}_i = -\frac{1}{M} \sum_{i=1}^M (1 - E_{\bar{i}}) \sum_{j,k}^{(i)} x_{j\bar{i}}^* G_{jk}^{(i\mu_i)} x_{k\bar{i}} + O(\Lambda_o^2). \tag{3.5.55}$$

In the rest of this proof, we let $\mathbf{y}_\mu$ be the μ -th column of $\mathbf{Y}$, and $\mathbf{x}_\mu$ is defined similarly. Denote $\tilde{V}_i := \mathbf{x}_{\bar{i}}^* G^{(i\mu_i)} \mathbf{x}_{\bar{i}}$ and $V_i := Q_{\bar{i}} \tilde{V}_i$. Next we are going to bound

$E|\frac{1}{M}\sum_i V_i|^p$. In the following, let p be an even integer and denote $V_{i_s} := V_{i_s}$ for $s \leq p/2$ and $V_{i_s} := \bar{V}_{i_s}$ for $s > p/2$. For a set $A \subset \{\mu_1, \dots, \mu_M\}$ where $\mu_i = \bar{i}$, we use notations $P_A := \prod_{\mu \in A} P_\mu$ and $Q_A := \prod_{\mu \in A} Q_\mu$. Let $\mathbf{i} = (i_1, \dots, i_p)$ and $[\mathbf{i}] = \{i_1, \dots, i_p\}$. We have

$$\begin{aligned} E|\frac{1}{M}\sum_i V_i|^p &= \frac{1}{M^p} \sum_{i_1, \dots, i_p} E \prod_{s=1}^p V_{i_s} = \frac{1}{M^p} \sum_{i_1, \dots, i_p} E \prod_{s=1}^p \left(\prod_{r=1}^s (P_{i_r} + Q_{i_r}) V_{i_s} \right) \\ &= \frac{1}{M^p} \sum_{\mathbf{i}} \sum_{A_1, \dots, A_p \subset [\mathbf{i}]} E \prod_{s=1}^p (P_{A_s^c} Q_{A_s} V_{i_s}). \end{aligned} \quad (3.5.56)$$

Following arguments from (7.21) to (7.23) in [Benaych-Georges and Knowles \(2016\)](#), we have

$$\sum_{s=1}^p |A_s| \geq 2|[\mathbf{i}]|. \quad (3.5.57)$$

The next step is to show that

$$|Q_A V_i| \prec \Upsilon^{|A|}. \quad (3.5.58)$$

If $A = |1|$ (corresponding to the case $A = \{\bar{i}\}$), it follows from Lemma 3.3.7 directly. For the case $|A| \geq 2$ we assume that $i = 1$ and $A = \{\mu_1, \mu_2, \dots, \mu_t\}$ with $t \geq 2$. We have

$$\begin{aligned} &Q_{\mu_2}(\mathbf{x}_{\mu_1}^* G^{(1\mu_1)} \mathbf{x}_{\mu_1}) \\ &= Q_{\mu_2} \mathbf{x}_{\mu_1}^* \left(G^{(1\mu_1\mu_2)} - \frac{G^{(1\mu_1\mu_2)} \mathbf{y}_{\mu_2} \mathbf{y}_{\mu_2}^* G^{(1\mu_1\mu_2)}}{1 + \mathbf{y}_{\mu_2}^* G^{(1\mu_1\mu_2)} \mathbf{y}_{\mu_2}} \right) \mathbf{x}_{\mu_1} \\ &= Q_{\mu_2} \mathbf{x}_{\mu_1}^* G^{(1\mu_1\mu_2)} \mathbf{y}_{\mu_2} \mathbf{y}_{\mu_2}^* G^{(1\mu_1\mu_2)} \mathbf{x}_{\mu_1} G_{\mu_2\mu_2}^{(1\mu_1)} \\ &= Q_{\mu_2} \left(\frac{G_{\mu_1\mu_2}^{(1)}}{G_{\mu_1\mu_1}^{(1)}} \mathbf{y}_{\mu_2}^* G^{(1\mu_1\mu_2)} \mathbf{x}_{\mu_1} \right), \end{aligned}$$

where in the second equality, $Q_{\mu_2}(\mathbf{x}_{\mu_1}^* G^{(1\mu_1\mu_2)} \mathbf{x}_{\mu_1})$ vanishes because $\mathbf{x}_{\mu_1}^* G^{(1\mu_1\mu_2)} \mathbf{x}_{\mu_1}$ is measurable with respect to $H^{(\mu_2)}$, and in the last equality, we use $G_{\mu_1\mu_2}^{(1)} = G_{\mu_1\mu_1}^{(1)} G_{\mu_2\mu_2}^{(1\mu_1)} \mathbf{y}_{\mu_2}^* G^{(1\mu_1\mu_2)} \mathbf{x}_{\mu_1}$. For the term $\mathbf{y}_{\mu_2}^* G^{(1\mu_1\mu_2)} \mathbf{x}_{\mu_1}$, by using Lemma 3.3.7, we have

$$\mathbf{y}_{\mu_2}^* G^{(1\mu_1\mu_2)} \mathbf{x}_{\mu_1} = \sum_j^{(1)} \mathbf{x}_{\mu_2 j}^* G_{jk}^{(1\mu_1\mu_2)} \mathbf{x}_{k\mu_1} + d_2 \sum_k^{(1)} G_{2k}^{(1\mu_1\mu_2)} \mathbf{x}_{k\mu_1} \prec \Upsilon.$$

Therefore $Q_{\mu_2}(\mathbf{x}_{\mu_1}^* G^{(1\mu_1)} \mathbf{x}_{\mu_1}) \prec \Upsilon^2$. Furthermore, we have

$$\begin{aligned}
 & Q_{\mu_3} Q_{\mu_2}(\mathbf{x}_{\mu_1}^* G^{(1\mu_1)} \mathbf{x}_{\mu_1}) \\
 &= Q_{\mu_3} Q_{\mu_2} \left(\frac{G_{\mu_1\mu_2}^{(1)}}{G_{\mu_1\mu_1}^{(1)}} \mathbf{y}_{\mu_2}^* G^{(1\mu_1\mu_2)} \mathbf{x}_{\mu_1} \right) \\
 &= Q_{\mu_2} Q_{\mu_3} \left\{ \left(G_{\mu_1\mu_2}^{(1\mu_3)} + \frac{G_{\mu_1\mu_3}^{(1)} G_{\mu_3\mu_2}^{(1)}}{G_{\mu_3\mu_3}^{(1)}} \right) \left(\frac{1}{G_{\mu_1\mu_1}^{(1\mu_3)}} - \frac{G_{\mu_1\mu_3}^{(1)} G_{\mu_3\mu_1}^{(1)}}{G_{\mu_1\mu_1}^{(1)} G_{\mu_1\mu_1}^{(1\mu_3)} G_{\mu_3\mu_3}^{(1)}} \right) \right. \\
 &\quad \times \mathbf{y}_{\mu_2}^* (G^{(1\mu_1\mu_2\mu_3)} - G^{(1\mu_1\mu_2\mu_3)} \mathbf{y}_{\mu_3} \mathbf{y}_{\mu_3}^* G^{(1\mu_1\mu_2\mu_3)} G_{\mu_3\mu_3}^{(1)}) \mathbf{x}_{\mu_1} \Big\} \\
 &= Q_{\mu_2} Q_{\mu_3} \left\{ \frac{G_{\mu_1\mu_2}^{(1\mu_3)}}{G_{\mu_1\mu_1}^{(1\mu_3)}} \mathbf{y}_{\mu_2}^* G^{(1\mu_1\mu_2\mu_3)} \mathbf{x}_{\mu_1} - \frac{G_{\mu_1\mu_2}^{(1\mu_3)}}{G_{\mu_1\mu_1}^{(1\mu_3)}} \frac{G_{\mu_2\mu_3}^{(1\mu_1)}}{G_{\mu_2\mu_2}^{(1\mu_1)}} \mathbf{y}_{\mu_3}^* G^{(1\mu_1\mu_2\mu_3)} \mathbf{x}_{\mu_1} \right. \\
 &\quad + \left(\frac{G_{\mu_1\mu_3}^{(1)} G_{\mu_3\mu_2}^{(1)}}{G_{\mu_3\mu_3}^{(1)} G_{\mu_1\mu_1}^{(1\mu_3)}} - \frac{G_{\mu_1\mu_2}^{(1\mu_3)} G_{\mu_1\mu_3}^{(1)} G_{\mu_3\mu_1}^{(1)}}{G_{\mu_1\mu_1}^{(1)} G_{\mu_1\mu_1}^{(1\mu_3)} G_{\mu_3\mu_3}^{(1)}} - \frac{(G_{\mu_1\mu_3}^{(1)})^2 G_{\mu_3\mu_2}^{(1)} G_{\mu_3\mu_1}^{(1)}}{G_{\mu_1\mu_1}^{(1)} G_{\mu_1\mu_1}^{(1\mu_3)} (G_{\mu_3\mu_3}^{(1)})^2} \right) \mathbf{y}_{\mu_2}^* G^{(1\mu_1\mu_2\mu_3)} \mathbf{x}_{\mu_1} \\
 &\quad \left. - \left(\frac{G_{\mu_1\mu_3}^{(1)} G_{\mu_3\mu_2}^{(1)}}{G_{\mu_3\mu_3}^{(1)} G_{\mu_1\mu_1}^{(1\mu_3)}} - \frac{G_{\mu_1\mu_2}^{(1\mu_3)} G_{\mu_1\mu_3}^{(1)} G_{\mu_3\mu_1}^{(1)}}{G_{\mu_1\mu_1}^{(1)} G_{\mu_1\mu_1}^{(1\mu_3)} G_{\mu_3\mu_3}^{(1)}} - \frac{(G_{\mu_1\mu_3}^{(1)})^2 G_{\mu_3\mu_2}^{(1)} G_{\mu_3\mu_1}^{(1)}}{G_{\mu_1\mu_1}^{(1)} G_{\mu_1\mu_1}^{(1\mu_3)} (G_{\mu_3\mu_3}^{(1)})^2} \right) \frac{G_{\mu_2\mu_3}^{(1\mu_1)}}{G_{\mu_2\mu_2}^{(1\mu_1)}} \mathbf{y}_{\mu_3}^* G^{(1\mu_1\mu_2\mu_3)} \mathbf{x}_{\mu_1} \right\}, \tag{3.5.59}
 \end{aligned}$$

where in the last step, we use $\frac{G_{\mu_2\mu_3}^{(1\mu_1)}}{G_{\mu_2\mu_2}^{(1\mu_1)}} = G_{\mu_3\mu_3}^{(1\mu_1\mu_2)} \mathbf{y}_{\mu_2}^* G^{(1\mu_1\mu_2\mu_3)} \mathbf{y}_{\mu_3}$. For the term

$Q_{\mu_2} Q_{\mu_3} \frac{G_{\mu_1\mu_2}^{(1\mu_3)}}{G_{\mu_1\mu_1}^{(1\mu_3)}} \mathbf{y}_{\mu_2}^* G^{(1\mu_1\mu_2\mu_3)} \mathbf{x}_{\mu_1}$ in the last step of (3.5.59), it vanishes because $\frac{G_{\mu_1\mu_2}^{(1\mu_3)}}{G_{\mu_1\mu_1}^{(1\mu_3)}} \mathbf{y}_{\mu_2}^* G^{(1\mu_1\mu_2\mu_3)} \mathbf{x}_{\mu_1}$

is measurable with respect to $H^{(\mu_3)}$. Noting that the other terms are all fractions of G entries multiplying with quadratic form of $\mathbf{y}_{\mu_j}^* G^{(T)} \mathbf{x}_{\mu_1} (j = 2, \dots, M)$, and the fractions have property that the entries in the numerators are off-diagonal and the entries in the denominator are diagonal. Furthermore, the numbers of off-diagonal entries in each term are at least two, and $\mathbf{y}_{\mu_j}^* G^{(T)} \mathbf{x}_{\mu_1} (j = 2, \dots, M)$ appears once. Therefore, we have $Q_{\mu_3} Q_{\mu_2}(\mathbf{x}_{\mu_1}^* G^{(1\mu_1)} \mathbf{x}_{\mu_1}) \prec \Upsilon^3$.

We may continue in this manner. At the step when Q_{μ_t} is appended, we expand the Green entries by (3.3.32), change the order of Q operators (attach the Q_{μ_t} to expended formulas), and keep the non-vanishing terms. After each step, the number of off-diagonal G entries in the numerator in each term increases by one. Consequently, we get

$$Q_{\mu_t} \cdots Q_{\mu_2}(\mathbf{x}_{\mu_1}^* G^{(1\mu_1)} \mathbf{x}_{\mu_1}) \prec \Upsilon^t.$$

Therefore, we conclude (3.5.58).

Using (3.5.56), (3.5.57) and (3.5.58), we find

$$\begin{aligned} E \left| \frac{1}{M} \sum_i V_i \right|^p &\prec C_p \frac{1}{M^p} \sum_{\mathbf{i}} \Upsilon^{2|\mathbf{i}|} = C_p \sum_{u=1}^p \Upsilon^{2u} \frac{1}{M^p} \sum_{\mathbf{i}} \mathbf{1}(|\mathbf{i}| = u) \\ &\leq C_p \sum_{u=1}^p \Upsilon^{2u} M^{u-p} \leq C_p (\Upsilon + M^{-1/2})^{2p} \leq C_p \Upsilon^{2p}, \end{aligned}$$

where in the fourth step we use the inequality $a^m b^n \leq (a+b)^{m+n}$ for positive a, b .

Then it follows that

$$\frac{1}{M} \sum_i V_i = O_{\prec}(\Upsilon^2).$$

This together with (3.5.55) concludes the third estimate in (3.5.54). $\square$

Then we obtain an equation involving $[v]$, which is an improvement on Lemma 3.5.7.

Lemma 3.5.11. *(Strong self-consistent equation) The following equation holds with high probability, uniformly for $z \in \mathcal{D}(\tilde{c}, \tau)$:*

$$\begin{aligned} (1 - R_2)[v] - R_3[v]^2 &= \frac{1}{M} \sum_{i=1}^M \frac{d_i^2 t_i^2 \mathcal{S}_i}{b^2} - \frac{1}{M} \sum_{i=1}^M t_i^2 \mathcal{T}_i + \frac{1}{M} \sum_{i=1}^M \left(\frac{2ct_i^2 d_i^2 \mathcal{S}_i}{b^3} - \frac{2ct_i^3 d_i^4 \mathcal{S}_i}{b^4} - \right. \\ &\quad \left. \frac{2ct_i^3 z d_i^2 \mathcal{S}_i}{b^2} - \frac{2ct_i^2 d_i^2 \mathcal{T}_i}{b^2} \right) [v] + O(\Lambda^3) + O(\max_i (|\mathcal{S}_i|^2) + O(\max_i |\mathcal{T}_i|^2)). \end{aligned} \quad (3.5.60)$$

Proof. This follows from Lemma 3.5.2 and (3.5.37). $\square$

3.5.2.2 Proof of Lemma 3.5.1

The main step is to show that for any $\sigma \in [1/4, 1]$, $\Lambda \prec (N\eta)^{-\sigma}$ will imply $\Lambda \prec (N\eta)^{-\frac{1}{2}-\frac{\sigma}{2}}$. Denote

$$g_{1i} = t_i^2, \quad g_{2i} = \frac{2ct_i^2 d_i^2}{b^2}, \quad g_{3i} = \frac{d_i^2 t_i^2}{b^2}, \quad g_{4i} = \frac{2ct_i^2 d_i^2}{b^3} - \frac{2ct_i^3 d_i^4}{b^4} - \frac{2ct_i^3 z d_i^2}{b^2}.$$

Fix $\epsilon \in (0, \tau/10)$ and $\sigma = 1/4$. Define an event Ξ as the intersection of following events:

$$\begin{aligned} E_1 &= \{\text{the strong self-consistent equation in Lemma 3.5.11 holds}\}, \\ E_2 &= \{\Lambda + \Lambda_o \leq N^{\epsilon/2}(N\eta)^{-\sigma}\}, \\ E_3 &= \left\{ \Lambda_o + \max_i |\mathcal{S}_i| + \max_i |\mathcal{T}_i| \leq N^{\epsilon/2} \sqrt{\frac{\Im m + \Lambda}{N\eta}} \right\}, \\ E_4 &= \left\{ \frac{1}{M} \left(\left| \sum_{i=1}^M g_{1i} \mathcal{Z}_i \right| + \left| \sum_{i=1}^M g_{2i} \mathcal{Z}_i \right| + \left| \sum_{i=1}^M g_{3i} \mathcal{Y}_i \right| + \left| \sum_{i=1}^M g_{4i} \mathcal{Y}_i \right| \right) \leq N^\epsilon \frac{\Im m + \Lambda}{N\eta} \right\}. \end{aligned} \quad (3.5.61)$$

We know that Ξ holds with high probability. For the following, we fix an realization of $H \in \Xi$. Denote $\alpha_0 = N^\epsilon (N\eta)^{-\frac{1}{2}-\frac{\sigma}{2}}$.

First consider the case $\alpha \leq \alpha_0$. We find that

$$\begin{aligned} |[v]|^2 &\leq \left| [v]^2 - \frac{(1-R_2)[v]}{R_3} \right| + \left| \frac{(1-R_2)[v]}{R_3} \right| \\ &\leq N^\epsilon \frac{\Im m + \Lambda}{N\eta} + \frac{|[v]|^2}{4N^{\epsilon/2}} + N^{\epsilon/2} \left(N^\epsilon \frac{\Im m + \Lambda}{N\eta} \right)^2 + \frac{\alpha |[v]|}{c'} \\ &\leq N^\epsilon \frac{C\alpha + N^{\epsilon/2}(N\eta)^{-\sigma}}{N\eta} + o(1)|[v]|^2 + \frac{\alpha |[v]|}{c'}, \end{aligned}$$

where in the second step we use (3.5.60) and the second statement in Lemma 3.5.8, and in the last step we use $\Im m \sim \sqrt{\kappa + \eta} \sim \alpha$ and $N^{2\epsilon} \frac{\Im m + \Lambda}{N\eta} \ll 1$. Hence

$$\left(\Lambda - \frac{\alpha}{2c'} \right)^2 \leq N^\epsilon \frac{C\alpha + N^{\epsilon/2}(N\eta)^{-\sigma}}{N\eta} + \frac{\alpha^2}{4c'}.$$

Taking square root, recalling that $\alpha \leq \alpha_0$ and using the fact that $\sigma \leq 1$, we get

$$\Lambda \leq C\alpha_0.$$

Next, we consider the case $\alpha > \alpha_0$. Suppose $|R_3(z)| \leq C_3$ for $z \in D(\tilde{c}, \tau)$. Assume first that $\Lambda \leq \alpha/2C_3$. By (3.5.60), we get

$$\alpha\Lambda \leq C_3\Lambda^2 + N^\epsilon \frac{\Im s + \Lambda}{N\eta} + o(1)\Lambda^2 \leq \frac{\alpha\Lambda}{2} + N^\epsilon \frac{C\alpha + N^{\epsilon/2}(N\eta)^{-\sigma}}{N\eta}.$$

This implies that

$$\Lambda \leq CN^\epsilon \frac{1}{N\eta} + \frac{\alpha_0^2}{N^{\epsilon/2}\alpha}.$$

Since $\alpha > \alpha_0$, we have $\Lambda \ll \alpha_0 < \alpha$. Thus if $\alpha > \alpha_0$, we have either $\Lambda > \alpha/2C_3$ or $\Lambda \ll \alpha$. By the continuity of $\Lambda(z)$ in η , and the fact that $\Lambda \ll 1 \sim \alpha$ when $\eta \sim 1$, we must have $\Lambda \ll \alpha$. Thus $\Lambda \leq \alpha/2C_3$, and it follows that $\Lambda \ll \alpha_0$.

We conclude that for any $\sigma \in [1/4, 1]$, $\Lambda \prec (N\eta)^{-\sigma}$ implies $\Lambda \prec (N\eta)^{-\frac{1}{2}-\frac{\sigma}{2}}$. Therefore, we can start from $\sigma = 1/4$ which is guaranteed by Lemma 3.5.2, and iterate the procedure finite number of times to get $\Lambda \prec (N\eta)^{-1}$. (3.5.2) follows from the first one in (3.5.1) and Lemma 3.5.6. This concludes the proof of Lemma 3.5.1.

Recalling (3.5.60) and the event set defined in (3.5.61), we have that

$$\Lambda \leq \frac{1}{\alpha} \left(\Lambda^2 + N^\epsilon \frac{\Im s + \Lambda}{N\eta} \right).$$

Thus by the Lemma 3.5.1 and Lemma 3.5.8, we have for $z \in \mathcal{D}(\tilde{c}, \tau) \cap \{z = E + i\eta : E \geq \lambda_r\}$,

$$\Lambda \prec \frac{1}{\sqrt{\kappa + \eta}} \left(\frac{\Im s}{N\eta} + \frac{1}{(N\eta)^2} \right) \sim \frac{1}{N(\kappa + \eta)} + \frac{1}{(N\eta)^2 \sqrt{\kappa + \eta}} \quad (3.5.62)$$

which concludes (3.3.16) for the Gaussian case. $\square$

3.6 Anisotropic local law for general $\mathbf{X}$ with support $q = N^{-\delta}$, $0 < \delta < 1/2$.

In this section we finish the proof of anisotropic local law (3.3.14). We start from the Gaussian case in Subsection 3.6.1, next extend to the case that $\mathbf{X}$ has support $q = N^{-1/2}$ with the third moment vanishing in Subsection 3.6.2, then in Subsection 3.6.3 we remove the third moment condition. The averaged local laws (3.3.15) and (3.3.16) are also given. In Subsection 3.6.4, we conclude (3.3.14) by extending the anisotropic local to the case when $\mathbf{X}$ has support $q = N^{-\delta}$, $0 < \delta < 1/2$. Subsection 3.6.5 contains a Green function comparison result which is used in the proof of Theorem 3.3.6.

3.6.1 Anisotropic local law for Gaussian $\mathbf{X}$

In this subsection, we extend the entrywise local law to an anisotropic version for Gaussian $\mathbf{X}$. The following definition simplifies notation.

Definition 3.6.1. For any $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^{\mathcal{I}}$, we define the generalized Green function entries by

$$G_{\mathbf{w}_1 \mathbf{w}_2} = \langle \mathbf{w}_1, G \mathbf{w}_2 \rangle, \quad G_{s \mathbf{w}_2} = \langle \mathbf{e}_s, G \mathbf{w}_2 \rangle, \quad G_{\mathbf{w}_1 s} = \langle \mathbf{w}_1, G \mathbf{e}_s \rangle, \quad (3.6.1)$$

where $\mathbf{e}_s$ denotes the standard unit vector with s -coordinate being 1. Moreover, if $\mathbf{w}_1 \in \mathbb{R}^{\mathcal{I}_M}$ and $\mathbf{w}_2 \in \mathbb{R}^{\mathcal{I}_N}$, we sometimes identify them with their natural embeddings $(\mathbf{w}_1^*, \mathbf{0})^*$ and $(\mathbf{0}, \mathbf{w}_2^*)^*$ in $\mathbb{R}^{\mathcal{I}}$.

We next give some basic identities for G and introduce a notation to relabel G into groups.

Let the SVD of $\mathbf{Y}$ be

$$\mathbf{Y} = \sum_{k=1}^M \sqrt{\mu_k} \boldsymbol{\xi}_k \boldsymbol{\zeta}_k^*$$

Then

$$G_{ij} = (\mathbf{Y} \mathbf{Y}^* - z \mathbf{I})_{ij}^{-1} = \sum_{k=1}^M \frac{\boldsymbol{\xi}_k(i) \boldsymbol{\xi}_k^*(j)}{\mu_k - z},$$

$$G_{i\mu} = [(\mathbf{Y} \mathbf{Y}^* - z \mathbf{I})^{-1} \mathbf{Y}]_{i\mu} = \sum_{k=1}^M \frac{\sqrt{\mu_k} \boldsymbol{\xi}_k(i) \boldsymbol{\zeta}_k^*(\mu)}{\mu_k - z}$$

$$G_{\mu\nu} = (\mathbf{Y}^* \mathbf{Y} / z - \mathbf{I})_{\mu\nu}^{-1} = \sum_{k=1}^N \frac{z \boldsymbol{\zeta}_k(\mu) \boldsymbol{\zeta}_k^*(\nu)}{\mu_k - z} = -(\mathbf{I}_N)_{\mu\nu} + \sum_{k=1}^N \frac{\mu_k \boldsymbol{\zeta}_k(\mu) \boldsymbol{\zeta}_k^*(\nu)}{\mu_k - z}$$

Define

$$\boldsymbol{\eta}_k = \begin{pmatrix} \mathbf{1}(k \leq M) \boldsymbol{\xi}_k \\ \mathbf{1}(k \leq N) \sqrt{\mu_k} \boldsymbol{\zeta}_k \end{pmatrix},$$

then we have

$$G = -\mathbf{I} + \sum_{k=1}^M \frac{\boldsymbol{\eta}_k \boldsymbol{\eta}_k^*}{\mu_k - z} \quad \text{where} \quad \mathbf{I} = \begin{pmatrix} 0 & \\ & \mathbf{I}_N \end{pmatrix}. \quad (3.6.2)$$

For $\mathbf{v} \in \mathbb{R}^{\mathcal{I}_M}$ and $\mathbf{w} \in \mathbb{R}^{\mathcal{I}_N}$ (identified with their natural embeddings $(\mathbf{v}^*, \mathbf{0})^*$ and $(\mathbf{0}, \mathbf{w}^*)^*$ in $\mathbb{R}^{\mathcal{I}}$), by calculations we obtain

$$\sum_i |G_{\mathbf{v}i}|^2 = \frac{\Im G_{\mathbf{v}\mathbf{v}}}{\eta}, \quad (3.6.3)$$

$$\sum_{\mu} |G_{\mathbf{v}\mu}|^2 = (G \mathbf{Y} \mathbf{Y}^* G^*)_{\mathbf{v}\mathbf{v}} = C \|\mathbf{Y} \mathbf{Y}^*\| \frac{\Im G_{\mathbf{v}\mathbf{v}}}{\eta}, \quad (3.6.4)$$

$$\sum_i |G_{\mathbf{w}i}|^2 = (\mathbf{Y}^* G G^* \mathbf{Y})_{\mathbf{w}\mathbf{w}} = (z^{-1} G \mathbf{Y}^* \mathbf{Y} G^* (z^*)^{-1})_{\mathbf{w}\mathbf{w}} \leq C \|\mathbf{Y}^* \mathbf{Y}\| \frac{\Im G_{\mathbf{w}\mathbf{w}}}{\eta}, \quad (3.6.5)$$

$$\sum_{\mu} |G_{\mathbf{w}\mu}|^2 = \sum_{k=1}^N \frac{|z|^2 |\mathbf{w}^* \boldsymbol{\zeta}_k|^2}{|\mu_k - z|^2} = \frac{|z|^2 \Im(G_{\mathbf{w}\mathbf{w}}/z)}{\eta} = \frac{\Im(G_{\mathbf{w}\mathbf{w}} \bar{z})}{\eta} = \frac{E \Im G_{\mathbf{w}\mathbf{w}}}{\eta} - \Re G_{\mathbf{w}\mathbf{w}}. \quad (3.6.6)$$

For an $\mathcal{I} \times \mathcal{I}$ matrix A and $i, j \in \mathcal{I}_M$, we denote

$$A_{[ij]} = \begin{pmatrix} A_{ij} & A_{i\bar{j}} \\ A_{\bar{i}j} & A_{\bar{i}\bar{j}} \end{pmatrix}. \quad (3.6.7)$$

Moreover, for $s \in \mathcal{I}$, we denote

$$A_{s[i]} = (A_{si}, A_{s\bar{i}}), \quad \text{and} \quad A_{[i]s} = \begin{pmatrix} A_{is} \\ A_{\bar{i}s} \end{pmatrix} \quad (3.6.8)$$

For a vector $\mathbf{w} \in \mathbb{R}^{\mathcal{I}}$, let $w_{[i]} = (w_i, w_{\bar{i}})^*$.

Proposition 3.6.2. *Anisotropic local law (3.3.14) holds for $\mathbf{X}$ with standard i.i.d Gaussian entries.*

Proof. Since $\mathbf{X}$ is assumed to be Gaussian, it is equivalent to showing that Theorem 3.3.3 holds with $\mathbf{R}$ replaced by $\mathbf{D}$. By the polarization identity, it suffices to consider $\mathbf{w}_1 = \mathbf{w}_2 = \mathbf{w}$ with bounded spectral norm. Using notation in (3.6.7) and (3.6.8), we have

$$\begin{aligned} |\mathbf{w}^*(G - \Pi)\mathbf{w}| &\leq \left| \sum_{i \in \mathcal{I}_M} w_{[i]}^* (G_{[ii]} - \Pi_{[ii]}) w_{[i]} \right| + \left| \sum_{\mu > 2M} w_{\mu}^* (G_{\mu\mu} - \Pi_{\mu\mu}) w_{\mu} \right| \\ &+ \left| \sum_{i \neq j \in \mathcal{I}_M} w_{[i]}^* (G_{[ij]} - \Pi_{[ij]}) w_{[j]} \right| + 2 \left| \sum_{i \in \mathcal{I}_M, \mu > 2M} w_{[i]}^* (G_{[i]\mu} - \Pi_{[i]\mu}) w_{\mu} \right| + \left| \sum_{\mu \neq \nu > 2M} w_{\mu}^* (G_{\mu\nu} - \Pi_{\mu\nu}) w_{\nu} \right|. \end{aligned}$$

The first and second terms are stochastically controlled by Ψ from an direct application of entrywise local law and boundness of the spectral norm of $\mathbf{w}$. For the remaining terms, for example the third term, one can obtain

$$\mathbb{E} \left| \sum_{i \neq j \in \mathcal{I}_M} w_{[i]}^* (G_{[ij]} - \Pi_{[ij]}) w_{[j]} \right|^p \prec \Psi^p. \quad (3.6.9)$$

Then an application of the Markov inequality yields the conclusion. The moment estimation in (3.6.9) requires to expand the Green function entries (blocks) using Lemma 3.3.8 and analysis expectations carefully using some combinatorial results. This procedure was first performed in Alex et al. (2014) for generalized Wigner matrices and was generalized to non-Hermitian matrices in Xi et al. (2017). Our model has no essential difference with the latter case so we omit the details here. $\square$

3.6.2 Anisotropic local law for general $\mathbf{X}$ with support $q = N^{-1/2}$ and $\mathbb{E}x_{ij}^3 = 0$

We finish the proof of Theorem (3.3.14) for general $\mathbf{X}$ with support $q = N^{-1/2}$ in this subsection. Suppose that $\mathbf{X}$ satisfies the additional condition

$$\mathbb{E}x_{i\mu}^3 = 0. \quad (3.6.10)$$

The proof follows analogous bootstrap arguments developed in Knowles and Yin (2017) and we give an outline here. Some notations and tools for the proof of edge universality results are also prepared in this section.

Let $\hat{S}$ be an N^{-10} -net of $\mathcal{D}(\tilde{c}, \tau)$. By the Lipschitz continuity of the function $z \mapsto \langle \mathbf{w}, (G(z) - \Pi(z))\mathbf{w} \rangle$ with Lipschitz constants of order $O(N^2)$, it suffices to show that (3.3.14) holds for any $z \in \hat{S}$. Take a small constant

$$\delta \in (0, \frac{\tau}{2C_0}) \quad (3.6.11)$$

where C_0 is a sufficiently large constant. ($C_0 = 24$ will work). For any given $\eta > N^{-1}$ we define $\eta_0 \leq \eta_1 \leq \dots \leq \eta_L$ where

$$L \equiv L(\eta) := \max\{l \in \mathbb{N} : \eta N^{\delta(l-1)} < 1\},$$

through

$$\eta_l := \eta N^{\delta l} \quad (l = 0, \dots, L-1), \quad \eta_L = 1.$$

Define

$$\hat{S}_m := \{z \in \hat{S} : \Im z \geq N^{-\delta m}\}.$$

We are now prepared to start the bootstrapping which takes the anisotropic local law in a relatively small domain $\hat{S}_m$ as input and aims to control the generalized Green function entries at slightly larger domain $\hat{S}_{m+1}$. We introduce the following

two properties:

($\mathbf{A}_m$) For all $z \in \hat{S}_m$ and all $\mathbf{w} \in \mathbb{R}^{\mathcal{I}}$,

$$\Im G_{\mathbf{w}\mathbf{w}}(z) \prec \Im s(z) + N^{C_0\delta} \Psi(z). \quad (3.6.12)$$

($\mathbf{C}_m$) For all $z \in \hat{S}_m$ and all $\mathbf{w} \in \mathbb{R}^{\mathcal{I}}$,

$$|G_{\mathbf{w}\mathbf{w}}(z) - \Pi_{\mathbf{w}\mathbf{w}}(z)| \prec N^{C_0\delta} \Psi(z). \quad (3.6.13)$$

We find property ($\mathbf{A}_0$) holds obviously. Property ($\mathbf{C}_m$) implies ($\mathbf{A}_m$) since we have

$$\Im \Pi_{\mathbf{w}\mathbf{w}} \leq C \Im s(z), \quad (3.6.14)$$

which can be verified using Lemma 3.5.4 and Lemma 3.5.3. We can conclude that (3.6.13) holds for all $z \in \hat{S}$ if the following Lemma holds.

Lemma 3.6.3. *For any $1 \leq m \leq \delta^{-1}$, property ($\mathbf{A}_{m-1}$) implies property ($\mathbf{C}_m$),*

To prove the above lemma, we shall estimate

$$L_{\mathbf{w}}(\mathbf{X}, z) := |G_{\mathbf{w}\mathbf{w}}(z) - \Pi_{\mathbf{w}\mathbf{w}}(z)|$$

where $\mathbf{w}$ is any deterministic vector in $\mathbb{R}^{\mathcal{I}}$. By the Markov inequality, it suffices to show the following lemma.

Lemma 3.6.4. *Assume that property ($\mathbf{A}_{m-1}$) holds, for a fixed constant $p \in 2\mathbb{N}$, we have*

$$\mathbb{E} L_{\mathbf{w}}^p(\mathbf{X}, z) \leq (N^{C_0\delta} \Psi)^p. \quad (3.6.15)$$

In the remaining part of this subsection, we will prove Lemma 3.6.4 using interpolation method. From Proposition 3.6.2 and the third statement in Remark 14, we see (3.6.15) holds if $\mathbf{X}$ are standard Gaussian. Now we define the interpolation matrix $\mathbf{X}^\theta$ between $\mathbf{X}^1 = (x_{i\mu}^1) = \mathbf{X}$ and $\mathbf{X}^0 = (x_{i\mu}^0) = \mathbf{X}^{Gauss}$.

Definition 3.6.5. For $i \in \mathcal{I}_M$ and $\mu \in \mathcal{I}_N$, denote the law of $X_{i\mu}$ by $\rho_{i\mu}^1$ and the law of $x_{i\mu}^0$ by $\rho_{i\mu}^0$. For $\theta \in [0, 1]$, we define the law of $x_{i\mu}^\theta$ by

$$\rho_{i\mu}^\theta = \theta \rho_{i\mu}^1 + (1 - \theta) \rho_{i\mu}^0.$$

Let the interpolation matrix be $\mathbf{X}^\theta = (x_{i\mu}^\theta)$ where $(x_{i\mu}^\theta)$ are mutually independent for all i, μ . Furthermore, denote

$$\mathbf{X}_{(i\mu)}^{\theta, \lambda} = \mathbf{X}^\theta + (\lambda - x_{i\mu}^\theta) \mathbf{e}_i \mathbf{e}_\mu^*$$

which differs from $\mathbf{X}^\theta$ only at the position (i, μ) , and Green functions

$$G^\theta(z) = G(X^\theta, z), \quad G_{(i\mu)}^{\theta, \lambda}(z) = G(\mathbf{X}_{(i\mu)}^{\theta, \lambda}, z)$$

We need the following interpolation formula. One can refer to Lemma 7.9 in Knowles and Yin (2017).

Lemma 3.6.6. *For $K : \mathbb{R}^{\mathcal{I}_M \times \mathcal{I}_N} \rightarrow \mathbb{C}$ we have*

$$\mathbb{E}[K(\mathbf{X}^1) - K(\mathbf{X}^0)] = \int_0^1 d\theta \sum_{i, \mu} \left[\mathbb{E}K\left(\mathbf{X}_{(i\mu)}^{\theta, x_{i\mu}^1}\right) - \mathbb{E}K\left(\mathbf{X}_{(i\mu)}^{\theta, x_{i\mu}^0}\right) \right]. \quad (3.6.16)$$

We shall apply Lemma 3.6.6 for $K(\mathbf{X}) = L_{\mathbf{w}}^p(\mathbf{X})$. According to the Grönwall's inequality and the fundamental theorem of calculus, to conclude (3.6.15) we just need to verify that for any $\theta \in [0, 1]$, we have the following:

$$\sum_{i, \mu} \left[\mathbb{E}L_{\mathbf{w}}^p(\mathbf{X}_{(i\mu)}^{\theta, x_{i\mu}^1}, z) - \mathbb{E}L_{\mathbf{w}}^p(\mathbf{X}_{(i\mu)}^{\theta, x_{i\mu}^0}, z) \right] = O\left((N^{C_0\delta}\Psi)^p + \mathbb{E}L_{\mathbf{w}}^p(\mathbf{X}^\theta, z)\right). \quad (3.6.17)$$

This is an easy consequence of the following lemma.

Lemma 3.6.7. *Fix $p \in 2\mathbb{N}$ and small constant δ satisfying (3.6.11). Suppose that property $(\mathbf{A}_{m-1})$ holds, then for $z \in \hat{S}_m$ there exists some function $A(\cdot, z)$ such that for $u \in \{0, 1\}$ and any $\theta \in [0, 1]$ we have*

$$\sum_{i, \mu} \left[\mathbb{E}L_{\mathbf{w}}^p(\mathbf{X}_{(i\mu)}^{\theta, x_{i\mu}^u}, z) - \mathbb{E}A(\mathbf{X}_{(i\mu)}^{\theta, 0}, z) \right] = O\left((N^{C_0\delta}\Psi)^p + \mathbb{E}L_{\mathbf{w}}^p(\mathbf{X}^\theta, z)\right). \quad (3.6.18)$$

To prove this, we need some rough estimates about $G - \Pi$ for $z \in \hat{S}_m$ provided that $(\mathbf{A}_{m-1})$ holds, given by Lemma 3.6.9 below. To prove that we need the following lemma first.

Lemma 3.6.8. *For any $z = E + i\eta \in \mathbf{S}$ and $\mathbf{v}, \mathbf{w} \in \mathbb{R}^{\mathcal{I}}$, we have*

$$|G_{\mathbf{vw}} - \Pi_{\mathbf{vw}}| \prec N^{2\delta} \sum_{l=1}^{L(\eta)} (\Im G_{\mathbf{vv}}(E + i\eta_l) + \Im G_{\mathbf{ww}}(E + i\eta_l)) + C \quad (3.6.19)$$

Proof. The proof is the same as that for Lemma 7.12 in Knowles and Yin (2017) $\square$

Lemma 3.6.9. *Suppose that the property $(\mathbf{A}_{m-1})$ holds. Then for any deterministic vector $\mathbf{v}, \mathbf{w}$, we have for all $z \in \hat{S}_m$*

$$|G_{\mathbf{vw}}(z) - \Pi_{\mathbf{vw}}(z)| \prec N^{2\delta}. \quad (3.6.20)$$

Moreover,

$$\Im G_{\mathbf{ww}} \prec N^{2\delta}(\Im s(z) + N^{C_0\delta}\Psi(z)) \quad (3.6.21)$$

Proof. The proof is the same as that for Lemma 7.13 in Knowles and Yin (2017) $\square$

We have resolvent expansion

$$G_{(i\mu)}^{\theta, \lambda'} = G_{(i\mu)}^{\theta, \lambda} + \sum_{j=1}^l G_{(i\mu)}^{\theta, \lambda} (\Delta_{(i\mu)}^{\lambda-\lambda'} G_{(i\mu)}^{\theta, \lambda})^j + G_{(i\mu)}^{\theta, \lambda'} (\Delta_{(i\mu)}^{\lambda-\lambda'} G_{(i\mu)}^{\theta, \lambda})^{l+1} \quad (3.6.22)$$

where

$$\Delta_{i\mu}^t = t \begin{pmatrix} 0 & \mathbf{u}_i \mathbf{v}_\mu^* \\ \mathbf{v}_\mu \mathbf{u}_i^* & 0 \end{pmatrix}. \quad (3.6.23)$$

For ease of notation, we denote

$$f_{i\mu}(\lambda) = L_{\mathbf{w}}^p(\mathbf{X}_{(i\mu)}^{\theta, \lambda}).$$

We have the following lemma estimating derivatives of f which can be verified from (3.6.22) and Lemma 3.6.9.

Lemma 3.6.10. *Suppose that λ is a random variable satisfying $\lambda \prec N^{-1/2}$, then for any $l \in \mathbb{N}$, we have*

$$|f^{(l)}(\lambda)| \prec N^{2\delta(p+l)}. \quad (3.6.24)$$

Moreover, from the Taylor expansion we can get

$$f_{(i\mu)}(\lambda) = \sum_{l=0}^{4p} \frac{\lambda^l}{l!} f^{(l)}(0) + O(\Psi^p). \quad (3.6.25)$$

From Lemma 3.6.10, we have for either $u = 0$ or 1 ,

$$\begin{aligned} \mathbb{E}L_{\mathbf{w}}^p(\mathbf{X}_{(i\mu)}^{\theta, X_{i\mu}^u}) - \mathbb{E}L_{\mathbf{w}}^p(\mathbf{X}_{(i\mu)}^{\theta, 0}) &= \mathbb{E}[f_{i\mu}(x_{i\mu}^\theta) - f_{i\mu}(0)] \\ &= \mathbb{E}f_{(i\mu)}^{(1)}(0) + \frac{1}{2N}\mathbb{E}f_{i\mu}^{(2)}(0) + \sum_{l=4}^{4p} \frac{1}{l!}\mathbb{E}f_{i\mu}^{(l)}(0)\mathbb{E}(x_{i\mu}^u)^l + O(\Psi^p). \end{aligned} \quad (3.6.26)$$

where we use (3.6.10). Therefore, to conclude (3.6.18), it suffices to show that for $l = 4, \dots, 4p$,

$$N^{-l/2} \sum_{i,\mu} |\mathbb{E}f_{(i\mu)}^{(l)}(0)| = O\left((N^{C_0\delta}\Psi)^p + \mathbb{E}L_{\mathbf{w}}^p(\mathbf{X}^\theta)\right). \quad (3.6.27)$$

Instead of proving (3.6.27), we turn to prove that for $l = 4, \dots, 4p$,

$$N^{-l/2} \sum_{i,\mu} |\mathbb{E}f_{(i\mu)}^{(l)}(x_{i\mu}^\theta)| = O\left((N^{C_0\delta}\Psi)^p + \mathbb{E}L_{\mathbf{w}}^p(\mathbf{X}^\theta)\right). \quad (3.6.28)$$

By the same argument as the proof of Lemma 7.16 in Knowles and Yin (2017), (3.6.28) will imply (3.6.27). We omit the details here.

Next we introduce some notations to denote products of generalized Green function entries that are generated from calculating derivatives in (3.6.27). Notice that (3.6.27) only involves $\mathbf{X}^\theta$, so we denote $\mathbf{X}^\theta$ by $\mathbf{X}$ for simplicity. For $l \geq 1$ we denote by $\mathcal{W}_l$ the set of paired indices $\{s_1t_1, s_2t_2, \dots, s_lt_l\}$ where $s_kt_k \in \{i\mu, \mu i\}$. For an element $w \in \cup_l \mathcal{W}_l$, we denote its length (number of pairs of indices) by $l(w)$. For w such that $l(w) = 0$, define

$$A_{\mathbf{w}, i, \mu}(w) := G_{\mathbf{w}\mathbf{w}} - \Pi_{\mathbf{w}\mathbf{w}},$$

and if $l(w) \equiv l \geq 1$, we define

$$A_{\mathbf{w}, i, \mu}(w) := G_{\mathbf{w}[s_1]}G_{[t_1][s_2]} \cdots G_{[t_{l-1}][s_l]}G_{[t_l]\mathbf{w}}, \quad (3.6.29)$$

where the pairs of indices $s_kt_k \in \{i\mu, \mu i\}$ for $k = 1, \dots, l$, and

$$[i] := (\mathbf{u}_i^*, \mathbf{0})^*, \quad [\mu] := (\mathbf{0}, \mathbf{v}_\mu^*)^*.$$

With these notations we can get

$$\left(\frac{\partial}{\partial X_{i\mu}}\right)^l (G_{\mathbf{w}\mathbf{w}} - \Pi_{\mathbf{w}\mathbf{w}}) = (-1)^l \sum_{w \in \mathcal{W}_l} A_{\mathbf{w},i,\mu}(w).$$

It follows that

$$\left(\frac{\partial}{\partial X_{i\mu}}\right)^l L_{\mathbf{w}}^p(\mathbf{X}) = (-1)^l \sum_{\substack{l_1, \dots, l_{p/2}, \tilde{l}_1, \dots, \tilde{l}_{p/2} \geq 0 \\ l_1 + \dots + l_{p/2} + \tilde{l}_1 + \dots + \tilde{l}_{p/2} = l}} \frac{l!}{\prod_r l_r! \tilde{l}_r!} \prod_r \left(\sum_{\substack{w_r \in \mathcal{W}_{l_r} \\ \tilde{w}_r \in \mathcal{W}_{\tilde{l}_r}}} A_{\mathbf{w},i,\mu}(w_r) \overline{A_{\mathbf{w},i,\mu}(\tilde{w}_r)} \right). \quad (3.6.30)$$

Therefore we obtain an expression of the derivatives in (3.6.28) using (3.6.30). For simplicity, we omit the conjugates in the following. Due to the fact that p is finite, we can get rid of the summation indices in (3.6.30) and it suffices to estimate for $l = 4, \dots, 4p$ that

$$N^{-l/2} \sum_{i,\mu} \mathbb{E} \left| A_{\mathbf{w},i,\mu}(w_0)^{p-q} \prod_{r=1}^q A_{\mathbf{w},i,\mu}(w_r) \right| = O \left((N^{C\delta} \Psi)^p + \mathbb{E} L_{\mathbf{w}}^p(X) \right), \quad (3.6.31)$$

where $l(w_0) = 0$, $l(w_r) \geq 1$ for $1 \leq r \leq q$ and $\sum_{r=1}^q l(w_r) = l$.

For any $w \in \cup_l \mathcal{W}_l$, from (3.6.20) we have the rough bound

$$|A_{\mathbf{w},i,\mu}(w)| \prec N^{2\delta(l(w)+1)}. \quad (3.6.32)$$

We introduce the quantity

$$R_s = |G_{\mathbf{w}[s]}| + |G_{[s]\mathbf{w}}|.$$

If w satisfies $l(w) \geq 1$ we have the following improved bound

$$|A_{\mathbf{w},i,\mu}(w)| \prec (R_i^2 + R_\mu^2) N^{2\delta(l(w)-1)}. \quad (3.6.33)$$

And for w such that $l(w) = 1$, we have the sharper bound

$$|A_{\mathbf{w},i,\mu}(w)| \prec R_i R_\mu. \quad (3.6.34)$$

Write $\mathbf{w}^\top = (\mathbf{w}_1^\top, \mathbf{w}_2^\top)^\top$ where $\mathbf{w}_1 \in \mathbb{R}^{\mathcal{I}_M}$ and $\mathbf{w}_2 \in \mathbb{R}^{\mathcal{I}_N}$. Using (3.6.6), (3.6.20) and (3.6.21), we get

$$\frac{1}{N} \left(\sum_i R_i^2 + \sum_\mu R_\mu^2 \right) \leq C \frac{\Im G_{\mathbf{w}_1 \mathbf{w}_1} + \Im G_{\mathbf{w}_2 \mathbf{w}_2} + \eta |G_{\mathbf{w}_2 \mathbf{w}_2}|}{N\eta} \prec N^{2\delta} \frac{\Im S + N^{C_0\delta} \Psi}{N\eta} = N^{(C_0+2)\delta} \Psi^2. \quad (3.6.35)$$

The pigeonholing principle tells us that if $l \leq 2q - 2$, there must be at least two w_r satisfying $l(w_r) = 1$. Therefore we have

$$\begin{aligned} & N^{-l/2} \sum_{i,\mu} \mathbb{E} \left| A_{\mathbf{w},i,\mu}(w_0)^{p-q} \prod_{r=1}^q A_{\mathbf{w},i,\mu}(w_r) \right| \\ & \leq N^{-l/2+2\delta(l+q)} \mathbb{E} \sum_{i,\mu} L_{\mathbf{w}}(\mathbf{X})^{p-q} [\mathbf{1}(l \leq 2q-2) R_i^2 R_\mu^2 + \mathbf{1}(l \geq 2q-1) (R_i^2 + R_\mu^2)] \\ & \prec N^{-l/2+2+2\delta(l+q)} \mathbb{E} L_{\mathbf{w}}(\mathbf{X})^{p-q} [\mathbf{1}(l \leq 2q-2) N^{(2C_0+4)\delta} \Psi^4 + \mathbf{1}(l \geq 2q-1) N^{(C_0+2)\delta} \Psi^2] \\ & \leq \mathbb{E} L_{\mathbf{w}}(\mathbf{X})^{p-q} [\mathbf{1}(l \leq 2q-2) N^{(2C_0+4+2l+2q)\delta} \Psi^l + \mathbf{1}(l \geq 2q-1) N^{(C_0+2+2l+2q)\delta} \Psi^{l-2}] \\ & \leq \mathbb{E} L_{\mathbf{w}}(\mathbf{X})^{p-q} [\mathbf{1}(l \leq 2q-2) (N^{C_0\delta} \Psi)^l + \mathbf{1}(l \geq 2q-1) (N^{C_0\delta} \Psi)^{l-2}] \end{aligned}$$

where in the second step we use (3.6.35), in the third step we use $\Psi \geq N^{-1/2}$ and in the last step, we take a large constant C_0 , for example, $C_0 = 24$ works. Since $l \geq q$ and $l \geq 4$, if $l \geq 2q - 1$, then $l - 2 \geq q$. By (3.6.11), we get $N^{C_0\delta} \Psi < 1$. Therefore we conclude that

$$N^{-l/2} \sum_{i,\mu} \mathbb{E} \left| A_{\mathbf{w},i,\mu}(w_0)^{p-q} \prod_{r=1}^q A_{\mathbf{w},i,\mu}(w_r) \right| \leq \mathbb{E} L_{\mathbf{w}}(\mathbf{X})^{p-q} (N^{C_0\delta} \Psi)^q.$$

Now (3.6.31) follows by the Young's inequality. We conclude the anisotropic law under the additional condition (3.6.10).

3.6.3 Anisotropic and averaged local law for general $\mathbf{X}$ with support $N^{-1/2}$

In this subsection, we remove the condition (3.6.10) but we restrict z to the domain $\mathcal{D}(\tilde{c}, \tau, c_0)$ (see (3.3.13)). Thus we can use

$$\Psi(z) \leq N^{-1/4-c_0} \quad (3.6.36)$$

to offset the additional $N^{1/2}$ in (3.6.31) for $l = 3$, which is non negligible due to the expansion (3.6.26). The proof requires a further iteration step at fixed z . Under the $(\mathbf{A}_{m-1})$, we have

$$\tilde{G} := G - \Pi = O_{\prec}(N^{2\delta}\Phi), \quad (3.6.37)$$

where Φ is a control parameter that is going to be reduced after each iteration.

3.6.3.1 Anisotropic local law in the domain $\mathcal{D}(\tilde{c}, \tau, c_0)$

We can take $\Phi = 1$ in (3.6.37) according to (3.6.20). We will prove the following self-improving bound

$$N^{-3/2} \sum_{i,\mu} \mathbb{E} \left| A_{\mathbf{w},i,\mu}(w_0)^{p-q} \prod_{r=1}^q A_{\mathbf{w},i,\mu}(w_r) \right| = O \left((N^{C_0\delta}\Psi)^p + (N^{-\frac{c_0}{2}}\Phi)^p + \mathbb{E}L_{\mathbf{w}}^p(X) \right). \quad (3.6.38)$$

Assume the validity of (3.6.38), and combining with (3.6.31), we get an improved priori bound (3.6.37) with Φ deduced to $N^{-\frac{c_0}{2}}\Phi$ by an application of the Gronwall's inequality. After doing this $O(1/c_0)$ times, (3.6.31) for $l = 3$ follows. In the following, we prove (3.6.38). By the Young's inequality, it suffices to show that

$$\left| N^{-3/2} \sum_{i,\mu} A_{\mathbf{w},i,\mu}(w_0)^{p-q} \prod_{r=1}^q A_{\mathbf{w},i,\mu}(w_r) \right| \prec L_{\mathbf{w}}^{p-q}(\mathbf{X})(N^{C_0\delta}\Psi + N^{-c_0}\Phi)^q, \quad (3.6.39)$$

where $q \leq 3$. We consider case by case.

Case 1: $q = 1$. In this case $A_{\mathbf{w},i,\mu}$ has the form $G_{\mathbf{w}[s_1]}G_{[t_1][s_2]}G_{[t_2][s_3]}G_{[t_3]\mathbf{w}}$ where $\{s_j t_j\} \in \{i\mu, \mu i\}$, see (3.6.29). It can be decomposed by the following

$$\begin{aligned} G_{\mathbf{w}[s_1]}G_{[t_1][s_2]}G_{[t_2][s_3]}G_{[t_3]\mathbf{w}} &= G_{\mathbf{w}[s_1]}\Pi_{[t_1][s_2]}\Pi_{[t_2][s_3]}G_{[t_3]\mathbf{w}} \\ &\quad + G_{\mathbf{w}[s_1]}\tilde{G}_{[t_1][s_2]}\Pi_{[t_2][s_3]}G_{[t_3]\mathbf{w}} + G_{\mathbf{w}[s_1]}\Pi_{[t_1][s_2]}\tilde{G}_{[t_2][s_3]}G_{[t_3]\mathbf{w}}. \end{aligned} \quad (3.6.40)$$

Consider the second term on the right side above, taking summation over i, μ , by carefully choose some small constant c_0 , we have

$$N^{-\frac{3}{2}} \sum_{i,\mu} |G_{\mathbf{w}[s_1]}\tilde{G}_{[t_1][s_2]}\Pi_{[t_2][s_3]}G_{[t_3]\mathbf{w}}| \leq N^{(C_0+2)\delta}N^{1/2}\Psi^2\Phi \prec N^{-c_0}\Phi, \quad (3.6.41)$$

where in the first step we use the Hölder's inequality, (3.6.35) and (3.6.37), and in the second step we use the fact that $\Psi^2 \leq N^{-1/2-2c_0}$. The last two terms in (3.6.40) can be dealt with similarly and we omit details here.

For the first term in (3.6.40), based on whether s_1 equals t_3 , we analysis two cases separately. When $s_1 = t_3$, we consider $s_1 = t_3 = i$ for an illustration. In this scenario $t_1 = s_3 = \mu$, and $\{s_2 t_2\} \in \{i\mu, \mu i\}$, therefore we consider the term of form $G_{\mathbf{w}[i]}\Pi_{[\mu][i]}\Pi_{[\mu][\mu]}G_{[i]\mathbf{w}}$. It is easy to conclude that

$$|\sum_{\mu} \Pi_{[\mu i]}\Pi_{[\mu \mu]}| = O(\sqrt{N}).$$

It follows from (3.6.35) that

$$\left| \sum_{i, \mu} G_{\mathbf{w}[i]}\Pi_{[\mu][i]}\Pi_{[\mu][\mu]}G_{[i]\mathbf{w}} \right| \leq \sum_i \left| \sum_{\mu} \Pi_{[\mu][i]}\Pi_{[\mu][\mu]} \right| |G_{\mathbf{w}[i]}|^2 \prec N^{3/2+(C_0+2)\delta} \Psi^2 \quad (3.6.42)$$

When $s_1 \neq t_3$, we consider $s_1 = i, t_3 = \mu$ as an illustration. Using (3.6.37) we find

$$\sum_i G_{\mathbf{w}[i]}\Pi_{[t_1][s_2]}\Pi_{[t_2][s_3]} \prec N^{1/2+2\delta} \Phi$$

Combining with $\sum_{\mu} |G_{[\mu]\mathbf{w}}| \leq N^{1+(C_0/2+1)\delta} \Psi$, we get

$$\left| \sum_{i, \mu} G_{\mathbf{w}[i]}\Pi_{[t_1][s_2]}\Pi_{[t_2][s_3]}G_{[\mu]\mathbf{w}} \right| \prec N^{3/2+(C_0/2+3)\delta} \Psi. \quad (3.6.43)$$

Up to now the case $q = 1$ has been proved.

Case 2: $q = 2$. In this case $\prod_{r=1}^2 A_{\mathbf{w}, i, \mu}(w_r)$ has the form $G_{\mathbf{w}[i]}G_{[\mu]\mathbf{w}}G_{\mathbf{w}[i]}G_{[\mu][\mu]}G_{[i]\mathbf{w}}$ or $G_{\mathbf{w}[i]}G_{[\mu]\mathbf{w}}G_{\mathbf{w}[i]}G_{[\mu][i]}G_{[\mu]\mathbf{w}}$, or by exchanging i and μ in one of these two terms. We consider the first term as an illustration and the second is similar. Using (3.6.35) and (3.6.37) we have estimations

$$|\sum_{\mu} G_{[\mu]\mathbf{w}}G_{[\mu][\mu]}| \leq \sum_{\mu} |G_{[\mu]\mathbf{w}}\tilde{G}_{[\mu][\mu]}| + |\sum_{\mu} G_{[\mu]\mathbf{w}}\Pi_{[\mu][\mu]}| \prec N^{1+(C_0+2)\delta/2+2\delta} \Psi \Phi + N^{1/2+2\delta} \Phi,$$

and

$$\sum_i |G_{\mathbf{w}[i]}|^3 \prec N^{1+(C_0+4)\delta} \Psi^2 \Phi. \quad (3.6.44)$$

From above and (3.6.36) we conclude that

$$N^{-3/2} \left| \sum_{i,\mu} G_{\mathbf{w}[i]} G_{[\mu]\mathbf{w}} G_{\mathbf{w}[i]} G_{[\mu][\mu]} G_{[i]\mathbf{w}} \right| \prec (N^{C_0\delta} \Psi + N^{-c_0} \Phi)^2.$$

So we conclude the case $q = 2$.

Case 3: $q = 3$. Now $\prod_{r=1}^3 A_{\mathbf{w},i,\mu}$ has the form $(G_{\mathbf{w}[i]} G_{[\mu]\mathbf{w}})^3$ or by exchanging i and μ in some of three factors. We have estimate

$$N^{-3/2} \sum_{i,\mu} |G_{\mathbf{w}[i]}|^3 |G_{[\mu]\mathbf{w}}|^3 \prec N^{1/2+2(C_0+4)\delta} \Psi^4 \Phi^2 \leq (N^{C_0\delta} \Psi + N^{-c_0} \Phi)^3, \quad (3.6.45)$$

where we use (3.6.44) in the first step, and use (3.6.36) in the second step. Now we conclude (3.6.39).

3.6.3.2 Averaged local law in the domain $\mathcal{D}(\tilde{c}, \tau, c_0)$

We turn to conclude the averaged local law for $\mathbf{X}$ with support $N^{-1/2}$ in the domain $\mathcal{D}(\tilde{c}, \tau, c_0)$. Let

$$F(\mathbf{X}) = \left| \frac{1}{M} \sum_k G_{kk}(z) - s(z) \right|.$$

The method is similar to that used in Section 3.6.2, but there is no bootstrap since anisotropic local laws provide priori bounds. It suffices to show the following lemma.

Lemma 3.6.11. *Suppose that $\mathbf{X}$ satisfies (3.2.3) and has support $N^{-1/2}$. For any $z \in \mathcal{D}(\tilde{c}, \tau)$, we have*

$$N^{-l/2} \sum_{i,\mu} \left| \mathbb{E} \left(\frac{\partial}{\partial X_{i\mu}} \right)^l F^p(\mathbf{X}) \right| = O \left((N^\delta \Psi^2)^p + \left(\frac{N^{-c_0}}{N\eta} \right)^p + \mathbb{E} F^p(\mathbf{X}) \right), \quad (3.6.46)$$

for $l \geq 4$. If $l = 3$, the above holds for z in $\mathcal{D}(\tilde{c}, \tau, c_0)$.

Proof. Similar to (3.6.30), we can obtain the derivatives of $F^p(\mathbf{X})$ and it suffices to show that

$$N^{-l/2} \sum_{i,\mu} \left| \mathbb{E} \prod_{r=1}^p \frac{1}{M} \sum_k A_{\mathbf{e}_k, i, \mu}(w_r) \right| = O \left((N^\delta \Psi^2)^p + \left(\frac{N^{-c_0}}{N\eta} \right)^p + \mathbb{E} F^p(\mathbf{X}) \right) \quad (3.6.47)$$

We first consider $l \geq 4$, from. we know that for w with $l(w) \geq 1$,

$$\left| \frac{1}{N} \sum_k A_{\mathbf{e}_k, i, \mu}(w) \right| \prec \frac{1}{N} \sum_k (|G_{k[i]}|^2 + |G_{k[\mu]}|^2) \prec \Psi^2, \quad (3.6.48)$$

where the second step is similar to (3.6.35) but we use the anisotropic local law instead of the bootstrapping bound in (3.6.21). Then (3.6.47) follows from an application of the Young's inequality. In the remaining, we work on the case $l = 3$ and restrict z in $\mathcal{D}(\tilde{c}, \tau, c_0)$. Our goal is to show that for $q = 1, 2, 3$,

$$\left| N^{-3/2-q} \sum_{i, \mu} \prod_{r=1}^q \sum_k A_{\mathbf{e}_k, i, \mu}(w_r) \right| \prec O \left((N^\delta \Psi^2 + \frac{N^{-c_0}}{N\eta})^q \right). \quad (3.6.49)$$

We consider case by case.

Case 1: $q = 1$. The term inside the summation of k is of the form $G_{k[s_1]} G_{[t_1][s_2]} G_{[t_2][s_3]} G_{[t_3]k}$.

We consider the following three cases for illustration:

$$G_{k[i]} G_{[\mu][\mu]} G_{[i][i]} G_{[\mu]k}, \quad G_{k[i]} G_{[\mu][\mu]} G_{[i][\mu]} G_{[i]k}, \quad \text{or} \quad G_{k[i]} G_{[\mu][i]} G_{[\mu][i]} G_{[\mu]k}. \quad (3.6.50)$$

For the first one, Let $\tilde{\mathbf{u}} = N^{-1/2} \sum_i \Pi_{[i][i]} \mathbf{u}_i$ which has bounded spectral norm. We have

$$\left| \sum_i G_{k[i]} \Pi_{[i][i]} \right| = |N^{1/2} G_{k\tilde{\mathbf{u}}}| \prec N^{1/2} |\Pi_{k\tilde{\mathbf{u}}}| + N^{1/2} \Psi = N^{1/2} |\tilde{\mathbf{u}}(k)| + N^{1/2} \Psi.$$

Therefore we have

$$\begin{aligned} \left| \sum_i G_{k[i]} G_{[i][i]} \right| &\leq \left| \sum_i G_{k[i]} \tilde{G}_{[i][i]} \right| + \left| \sum_i G_{k[i]} \Pi_{[i][i]} \right| \\ &\prec \sum_i (|\Pi_{k[i]}| + \Psi) \Psi + N^{1/2} |\tilde{\mathbf{u}}(k)| + N^{1/2} \Psi \prec N^{1/2} |\tilde{\mathbf{u}}(k)| + N \Psi^2. \end{aligned} \quad (3.6.51)$$

Similarly if we define $\tilde{\mathbf{v}} = N^{-1/2} \sum_\mu \Pi_{[\mu][\mu]} \mathbf{v}_\mu$, then we have $\sum_\mu G_{[\mu][\mu]} G_{[\mu]k} = O_{\prec}(N^{1/2} |\tilde{\mathbf{v}}(\bar{k})| + N \Psi^2)$. Combining with $\Psi^2 \leq N^{-1/2-2c_0}$, we get

$$N^{-5/2} \left| \sum_{i, k, \mu} G_{k[i]} G_{[\mu][\mu]} G_{[i][i]} G_{[\mu]k} \right| \prec \Psi^2.$$

For the second term in (3.6.50), it can be estimated by combining

$$|\sum_{\mu} G_{[\mu][\mu]} G_{[\mu]i}| \prec N\Psi^2 + N^{1/2}|\tilde{\mathbf{v}}(\bar{i})|, \quad \text{and} \quad \sum_i |G_{k[i]} G_{[i]k}| \prec N\Psi^2, \quad (3.6.52)$$

which implies

$$N^{-5/2} \left| \sum_{i,k,\mu} G_{k[i]} G_{[\mu][\mu]} G_{[i][\mu]} G_{[i]k} \right| \prec \Psi^2.$$

The third term in (3.6.50) can be estimated using

$$|\sum_k G_{k[i]} G_{[\mu]k}| \leq 2 \sum_k |G_{k[i]}|^2 + |G_{[\mu]k}|^2 \prec N\Psi^2, \quad \text{and} \quad \sum_{i,\mu} |G_{[i][\mu]}|^2 \prec N^2\Psi^2.$$

Therefore we conclude the case $q = 1$.

Case 2: $q=2$. In this case we need to estimate

$$\sum_{i,\mu} \sum_{k_1,k_2} G_{k_1[i]} G_{[\mu]k_1} G_{k_2[i]} G_{[\mu][\mu]} G_{[i]k_2} \quad \text{and} \quad \sum_{i,\mu} \sum_{k_1,k_2} G_{k_1[i]} G_{[\mu]k_1} G_{k_2[i]} G_{[\mu][i]} G_{[\mu]k_2}. \quad (3.6.53)$$

We consider the second term first. By the resolvent formula it is easy to conclude that

$$\sum_{k_1} G_{k_1[i]} G_{[\mu]k_1} = \mathbf{v}_{\mu}^* G \begin{pmatrix} \mathbf{I}_M & \\ & 0 \end{pmatrix} G \mathbf{u}_i = (G \bar{\mathbf{I}} G)_{[\mu][i]}. \quad (3.6.54)$$

Moreover,

$$\sum_{i,\mu} |G \bar{\mathbf{I}} G_{[i][\mu]}|^2 = \text{Tr}[G_1^2 \mathbf{Y} \mathbf{Y}^* (G_1^*)^2] = \text{Tr}[G_1 (G_1^*)^2] + z \text{Tr}[G_1^2 (G_1^*)^2],$$

where we have

$$|\text{Tr}[G_1 (G_1^*)^2]| \leq \sum_j \frac{1}{[(\mu_j - E)^2 + \eta^2]^{3/2}} \leq N\eta^{-2} \Im s_N \prec N^2 \eta^{-1} \left(\frac{\Im s + \Psi}{N\eta} \right) \prec N^2 \eta^{-1} \Psi^2,$$

and similarly we have

$$|\text{Tr}[G_1^2 (G_1^*)^2]| \leq N\eta^{-3} \Im s_N \prec N^2 \eta^{-2} \Psi^2.$$

We get

$$\sum_{i,\mu} |G\bar{\mathbf{I}}G_{[i][\mu]}|^2 \leq N^2 \eta^{-2} \Psi^2. \quad (3.6.55)$$

Then it follows that the second term in (3.6.53) is estimated by

$$N^{-7/2} \left| \sum_{i,\mu} [(G\bar{\mathbf{I}}G)_{[\mu][i]}]^2 G_{[\mu][i]} \right| \prec N^{-3/2} \eta^{-2} \Psi^2 \leq \left(\frac{N^{-c_0}}{N\eta} \right)^2.$$

Next for the first term in (3.6.53), we write it as

$$\sum_{i,\mu} (G\bar{\mathbf{I}}G)_{[\mu][i]} (G\bar{\mathbf{I}}G)_{[i][i]} (\tilde{G}_{[\mu][\mu]} + \Pi_{[\mu][\mu]}). \quad (3.6.56)$$

From (3.3.14) and (3.6.3) we have for any $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{C}^{\mathcal{I}}$,

$$|(G\bar{\mathbf{I}}G)_{\mathbf{w}_1 \mathbf{w}_2}| \prec N \Psi^2, \quad (3.6.57)$$

and $|\sum_{\mu} \Pi_{[\mu][\mu]} (G\bar{\mathbf{I}}G)_{[\mu][i]}| \prec N^{3/2} \Psi^2$. Together with (3.6.55) we get

$$\begin{aligned} & N^{-7/2} \left| \sum_{i,\mu} (G\bar{\mathbf{I}}G)_{[\mu][i]} (G\bar{\mathbf{I}}G)_{[i][i]} (\tilde{G}_{[\mu][\mu]} + \Pi_{[\mu][\mu]}) \right| \\ & \prec N^{-3/2} \left(\sum_{i,\mu} |(G\bar{\mathbf{I}}G)_{[\mu][i]}|^2 \right)^{1/2} \Psi^3 + \Psi^4 \prec N^{-1/2} \eta^{-1} \Psi^4 + \Psi^4 \leq \left(\Psi^2 + \frac{N^{-c_0}}{N\eta} \right)^2. \end{aligned}$$

Now we conclude the proof in the case $q = 2$.

Case 3: $q = 3$. In this case the term we need to consider is of the form $(\sum_k G_{k[i]} G_{[\mu]k})^3$ (or by exchanging i and μ in some of three factors). We can estimate it by combining (3.6.55) and (3.6.57) to get

$$N^{-9/2} \sum_{i,\mu} |(G\bar{\mathbf{I}}G)_{[\mu][i]}|^3 \prec N^{-3/2} \eta^{-2} \Psi^4 \leq \left(\Psi^2 + \frac{N^{-c_0}}{N\eta} \right)^3.$$

Now we conclude this lemma and finish the proof of averaged local law for $\mathbf{X}$ with support $N^{-1/2}$. $\square$

At the end of this subsection, we remark that we also obtain (3.3.16) when $\mathbf{X}$ has support $N^{-1/2}$ and $z \in \mathcal{D}(\tilde{c}, \tau, c_0) \cap \{z = E + i\eta : E \geq \lambda_r, N\eta\sqrt{\kappa + \eta} > N^\tau\}$

which is a direct consequence from (3.6.46) together with the fact that

$$\Psi^2 \prec \frac{1}{N(\kappa + \eta)} + \frac{1}{(N\eta)^2 \sqrt{\kappa + \eta}}.$$

3.6.4 Anisotropic local law for $\mathbf{X}$ with support $q = N^{-\delta}$ ($0 < \delta < 1/2$).

In this subsection, we prove the anisotropic local law for $\mathbf{X}$ satisfying Assumption 1 with support $q = N^{-\delta}$ ($0 < \delta < 1/2$). We can use the method analogous to that in Subsection 3.6.2. We only mention the difference without repeating most of the arguments again. Recall the two properties $(\mathbf{A}_m)$ and $(\mathbf{C}_m)$ in (3.6.12) and (3.6.13). For current scenario, we reformulate them as

$$\begin{aligned} \Im G_{\mathbf{w}\mathbf{w}}(z) &\prec \Im \Pi_{\mathbf{w}\mathbf{w}}(z) + N^{C_0\delta}(q + \Psi(z)), \\ |G_{\mathbf{w}\mathbf{w}}(z) - \Pi_{\mathbf{w}\mathbf{w}}(z)| &\prec N^{C_0\delta}(q + \Psi(z)) \end{aligned} \quad (3.6.58)$$

We use $\mathbf{X}^1$ to denote the matrix with large support $N^{-\delta}$ for any $0 < \delta < 1/2$. By Lemma 3.7.1 below, there exists another matrix $\mathbf{X}^0$ with support $N^{-1/2}$ and the fourth moments of entries of $\mathbf{X}^0$ match with those of $\mathbf{X}^1$. The anisotropic local law holds for such $\mathbf{X}^0$ according to previous proofs. Similar to (3.6.17), it remains to conclude that

$$\sum_{i,\mu} \left[\mathbb{E} L_{\mathbf{w}}^p(\mathbf{X}_{(i\mu)}^{\theta, x_{i\mu}^1}, z) - \mathbb{E} L_{\mathbf{w}}^p(\mathbf{X}_{(i\mu)}^{\theta, x_{i\mu}^0}, z) \right] = O \left([N^{C_0\delta}(q + \Psi)]^p + \sup_{\mathbf{v} \in \mathbb{S}} \mathbb{E} L_{\mathbf{v}}^p(\mathbf{X}^{\theta}, z) \right).$$

Different from (3.6.26), we have

$$\begin{aligned} \mathbb{E} L_{\mathbf{w}}^p(X_{(i\mu)}^{\theta, x_{i\mu}^u}) - \mathbb{E} L_{\mathbf{w}}^p(X_{(i\mu)}^{\theta, 0}) &= \mathbb{E}[f_{i\mu}(x_{i\mu}^{\theta}) - f_{i\mu}(0)] \\ &= \mathbb{E} f_{(i\mu)}^{(1)}(0) + \frac{1}{2N} \mathbb{E} f_{i\mu}^{(2)}(0) + \frac{f^{(3)}(0)}{6} \mathbb{E}(x_{i\mu}^u)^3 \\ &\quad + \sum_{l=4}^{4p} \frac{1}{l!} \mathbb{E} f_{i\mu}^{(l)}(0) \mathbb{E}(x_{i\mu}^u)^l + O(q^{p+4}). \end{aligned} \quad (3.6.59)$$

Since for any $\theta \in [0, 1]$, we have

$$\mathbb{E}(x_{i\mu}^1)^k = \mathbb{E}(x_{i\mu}^{\theta})^k,$$

the problem is reduced to prove

$$N^{-2}q^{l-4} \sum_{i,\mu} |\mathbb{E} f_{i\mu}^{(l)}(0)| = O([N^{C_0\delta}(q + \Psi)]^p + \mathbb{E} L_{\mathbf{w}}^p(\mathbf{X}^\theta, z)). \quad (3.6.60)$$

This can be shown by similar method used in the proof of (3.6.28), but here we need to use the bound

$$\frac{1}{N} \left(\sum_i R_i^2 + \sum_\mu R_\mu^2 \right) \prec N^{2\delta} \frac{\Im s + q + N^{C_0\delta}\Psi}{N\eta} = N^{(C_0+2)\delta} \left(\Psi^2 + \frac{q}{N\eta} \right), \quad (3.6.61)$$

where in the first step we use the first estimate in (3.6.58). We refer the readers to Section 6.2 of Yang (2019) for more details. We also remark here that the author performed interpolation between the large support matrices with Gaussian matrices directly, different from our method here, so the third moment condition and the control of $\Psi(z)$ should be imposed simultaneously. One may refer to the restricted spectral domain defined in (3.21) therein where the third moment of $\mathbf{X}$ is involved, and (6.49) therein where such condition is required.

3.6.5 Green function comparison for two matrices with support $N^{-1/2}$

In this subsection, we establish the Green function comparison result for two matrix ensembles $\mathbf{X}, \tilde{\mathbf{X}}$ that both satisfy Assumption 1 and the bounded support conditoin with $q = N^{-1/2}$.

Proposition 3.6.12. *Let $\mathbf{X}$ and $\tilde{\mathbf{X}}$ satisfy Assumption 1 and have bounded support conditoin with $q = N^{-1/2}$. Let $\epsilon > 0$ and set $\eta = N^{-2/3-\epsilon}$. Suppose $K : \mathbb{R} \rightarrow \mathbb{R}$ is a real function satisfying*

$$\sup_x |K^{(k)}(x)| / (|x| + 1)^C \leq C, \quad k = 0, 1, 2, 3, 4.$$

Then for any E, E_1 and E_2 satisfying

$$|E|, |E_1|, |E_2| \leq N^{-2/3+\epsilon}, \quad (3.6.62)$$

there exists a constant $\phi > 0$ such that for any sufficiently large N and for any sufficiently small $\epsilon > 0$, we have

$$|\mathbb{E}K(N\eta\Im s_N(E + \lambda_r + i\eta)) - \mathbb{E}K(N\eta\Im \tilde{s}_N(E + \lambda_r + i\eta))| \leq N^{-\phi} \quad (3.6.63)$$

and

$$\left| EK\left(N \int_{E_1}^{E_2} \Im s_N(x + \lambda_r + i\eta) dx\right) - EK\left(N \int_{E_1}^{E_2} \Im \tilde{s}_N(x + \lambda_r + i\eta) dx\right) \right| \leq N^{-\phi}. \quad (3.6.64)$$

Proof. The proof is similar to that of Theorem 7.2 in Han et al. (2018) where the continuous interpolation (Lemma 3.6.6) was used, so we omit details here. One can follow their arguments but should be careful when deal with the remaining terms involving third order derivatives, which is (11.4) in Han et al. (2018). In our case, this can be handled through analogous arguments used in the proof of Lemma 3.6.11. $\square$

3.7 Proof of Theorem 3.2.1 and Theorem 3.3.3

In this section we finish the proof of Theorems 3.3.3 and 3.2.1. In the first subsection, we conclude the averaged local law for general $\mathbf{X}$ with large support $N^{-\delta}$ for small positive constant δ . Together with the anisotropic local law established in the previous section, we finish the proof of Theorem 3.3.3. The Green function comparison result between the small support case and large support case is also provided, thus the proof of Theorem 3.3.6 is completed. In the second subsection, we conclude the main result, the Theorem 3.2.1.

3.7.1 Averaged local law (3.3.15) and (3.3.16)

We adopt the Linderberg replacement strategy used in Lee and Yin (2014) for Wigner matrices, or see some extensions for different matrix models, e.g. Ding and Yang (2018), Yang (2019). We prepare some notations and present some key lemmas here. For more details one may refer to Lee and Yin (2014), or Ding and Yang (2018). We record a crucial moment matching lemma first.

Lemma 3.7.1. (Lemma 3.12 in Yang (2019)) *For any $\mathbf{X}$ satisfies Assumption 1 with support $N^{-\delta}$ where $\delta \in (0, 1/2)$, there exists another matrix $\tilde{\mathbf{X}}$ having support*

$q = N^{-1/2}$, and the first four moments of entries of $\tilde{\mathbf{X}}$ match with $\mathbf{X}$, i.e.,

$$\mathbb{E}x_{i\mu}^k = \mathbb{E}\tilde{x}_{i\mu}^k, \quad k = 1, 2, 3, 4.$$

Define a bijection map to ordering the index of $\mathbf{X}$ entries:

$$\varphi : \{(i, \mu) : 1 \leq i \leq M, M+1 \leq \mu \leq M+N\} \rightarrow \{1, 2, \dots, \zeta_{\max} = MN\}.$$

Denote by $\mathbf{X}^\zeta$ the interpolation matrix whose entries $x_{i\mu}$ follow the same distribution as x_{11} if $\varphi(i, \mu) \leq \zeta$ and follows the same distribution as $\tilde{x}_{11}$ otherwise. Thus $\mathbf{X}^0 = \tilde{\mathbf{X}}$ and $\mathbf{X}^{\zeta_{\max}} = \mathbf{X}$. Correspondingly, denote by H^ζ and G^ζ the linearization matrix and the Green function associated with $\mathbf{X}^\zeta$, i.e.

$$H^\zeta := \begin{pmatrix} -z\mathbf{I}_M & \mathbf{D} + \mathbf{U}^*\mathbf{X}^\zeta\mathbf{V} \\ (\mathbf{D} + \mathbf{U}^*\mathbf{X}^\zeta\mathbf{V})^* & -\mathbf{I}_N \end{pmatrix}, \quad G^\zeta := (H^\zeta)^{-1}.$$

Recall (3.6.23). If $\varphi(i, \mu) = \zeta$, then we can write

$$H^{\zeta-1} = H^{(\zeta)} + \Delta_2^\zeta, \quad H^\zeta = H^{(\zeta)} + \Delta_1^\zeta, \quad (3.7.1)$$

where $\Delta_1^\zeta = \Delta_{i\mu}^{x_{i\mu}}$, $\Delta_2^\zeta = \Delta_{i\mu}^{\tilde{x}_{i\mu}}$, and $H^{(\zeta)}$ is independent of $x_{i\mu}$ and $\tilde{x}_{i\mu}$. Let

$$W^\zeta := (H^{(\zeta)})^{-1}, \quad S^\zeta := G^\zeta, \quad T^\zeta = G^{\zeta-1}. \quad (3.7.2)$$

For notational simplicity, we omit the superscript ζ sometimes if it is fixed. We have the expansion

$$S = \sum_{l=0}^{\xi} (-1)^l (W\Delta_1)^l W + (-W\Delta_1)^{\xi+1} S \quad (3.7.3)$$

The following definition is introduced to simplify the notations when we study the expectations of polynomials of Green entries.

Definition 3.7.2. (Matrix operators $*_\zeta$ and $\mathcal{P}_{\zeta, \mathbf{a}}$) For ζ with $\varphi(i, \mu) = \zeta$, we define $A *_\zeta B$ as

$$A *_\zeta B := A\mathbf{u}_i\mathbf{v}_\mu^*B + A\mathbf{v}_\mu\mathbf{u}_i^*B.$$

The l -th power of A under the $*$ -product is

$$A^{*\zeta^l} := A *_\zeta A *_\zeta A *_\zeta \cdots *_\zeta A$$

For $a \in \mathbb{N}$ and $\mathbf{a} = (a_1, a_2, \dots, a_s) \in \mathbb{N}^s$, we define

$$\mathcal{P}_{\zeta, a} G_{\mathbf{w}\mathbf{v}} := G_{\mathbf{w}\mathbf{v}}^{*\zeta(a+1)} = \langle \mathbf{w}, G^{*\zeta(a+1)} \mathbf{v} \rangle, \quad \mathcal{P}_{\zeta, \mathbf{a}} \left(\prod_{t=1}^s G_{\mathbf{w}_t \mathbf{v}_t} \right) := \prod_{t=1}^s \mathcal{P}_{\zeta, a_t} G_{\mathbf{w}_t \mathbf{v}_t} = \prod_{t=1}^s G_{\mathbf{w}_t \mathbf{v}_t}^{*\zeta(a_t+1)}.$$

If $\mathcal{G}_1$ and $\mathcal{G}_2$ are products of G entries as above, then we define

$$\mathcal{P}_{\zeta, \mathbf{a}}(\mathcal{G}_1 + \mathcal{G}_2) := \mathcal{P}_{\zeta, \mathbf{a}} \mathcal{G}_1 + \mathcal{P}_{\zeta, \mathbf{a}} \mathcal{G}_2.$$

Similarly, we define

$$\mathcal{P}_{\zeta, \mathbf{a}} \left(\prod_{t=1}^s (G - \Pi)_{\mathbf{w}_t \mathbf{v}_t} \right) := \prod_{t=1}^s \mathcal{P}_{\zeta, a_t} (G - \Pi)_{\mathbf{w}_t \mathbf{v}_t},$$

where

$$\mathcal{P}_{\zeta, a} (G - \Pi)_{\mathbf{w}\mathbf{v}} := \begin{cases} (G - \Pi)_{\mathbf{w}\mathbf{v}} & \text{if } a = 0, \\ G_{\mathbf{w}\mathbf{v}}^{*\zeta(a+1)} & \text{otherwise.} \end{cases}$$

With these notations, we have the following lemma from (3.3.14).

Lemma 3.7.3. *For any $\mathbf{a} = (a_1, \dots, a_s) \in \mathbb{N}^s$, ζ, ζ' and $\mathbf{w}_t, \mathbf{v}_t, 1 \leq t \leq s$, we have*

$$\left| \mathcal{P}_{\zeta, \mathbf{a}} \left(\prod_{t=1}^s A_{\mathbf{w}_t \mathbf{v}_t} \right) \right|, \left| \mathcal{P}_{\zeta', \mathbf{a}} \left(\prod_{t=1}^s (A - \Pi)_{\mathbf{w}_t \mathbf{v}_t} \right) \right| \prec C^{s+|\mathbf{a}|+1},$$

where A can be S, T or W , C is a fixed constant and $|\mathbf{a}| = \sum_k a_k$.

From (3.7.3) we have

$$S = \sum_{l=0}^{\xi} (W^{*(l+1)}) (-x_{i\mu})^l + O_{\prec}(N^{-\xi\delta}). \quad (3.7.4)$$

Another useful expansion for any finite s, ξ is

$$\prod_{t=1}^s S_{j_t k_t} = \prod_{t=1}^s W_{j_t k_t} + \sum_{1 \leq l \leq \xi} \left(\sum_{\mathbf{a} \in I_{\xi, l}^s} \mathcal{P}_{\zeta, \mathbf{a}} \prod_{t=1}^s W_{j_t k_t} \right) (-x_{ab})^l + O(N^{-\xi\delta}) \quad (3.7.5)$$

where

$$I_{\xi, l}^s = \{ \mathbf{a} = (a_1, \dots, a_s) \in \mathbb{N}^s : 0 \leq a_i \leq \xi, \sum a_i = l \}. \quad (3.7.6)$$

We will need this in the proof of Lemma 3.7.8 later.

We copy Lemma 6.5 in [Lee and Yin \(2014\)](#) which is powerful to handle the expectation of $S_{j_1 k_1} S_{j_2 k_2} \cdots S_{j_s k_s}$. One may also refer to Lemma 7.5 of [Yang \(2019\)](#) where the generalized Green function entries are involved.

Lemma 3.7.4. *For any finite $s, m \in \mathbb{N}$, we have with $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_s) \in \mathbb{N}^s$, $|\alpha| = \sum |\alpha_i|$,*

$$\mathbb{E} \prod_{t=1}^s S_{j_t, k_t} = \sum_{k=0}^4 A_k \mathbb{E}(-x_{i\mu})^k + \sum_{\substack{5 \leq |\alpha| \leq 2m/\delta \\ \alpha_1, \dots, \alpha_s \geq 0}} \mathcal{A}_\alpha \mathbb{E} \mathcal{P}_{\zeta, \alpha} \prod_{t=1}^s S_{j_t, k_t} + O(N^{-m}), \quad (3.7.7)$$

where $A_k (0 \leq k \leq 4)$ depend only on W , $\mathcal{A}$'s are independent of (j_t, k_t) , $1 \leq t \leq s$ and

$$|\mathcal{A}_\alpha| \leq N^{-|\alpha|\delta/10-2}. \quad (3.7.8)$$

Similarly, we have

$$\mathbb{E} \prod_{t=1}^s (S - \Pi)_{j_t, k_t} = \sum_{k=0}^4 A_k \mathbb{E}(-x_{i\mu})^k + \sum_{\substack{5 \leq |\alpha| \leq 2m/\delta \\ \alpha_1, \alpha_2, \dots, \alpha_s \geq 0}} \mathcal{A}_\alpha \mathbb{E} \mathcal{P}_{\zeta, \alpha} \prod_{t=1}^s (S - \Pi)_{j_t, k_t} + O(N^{-m}).$$

Moreover, as in (3.7.7), we have

$$\mathbb{E} \prod_{t=1}^s S_{j_t, k_t} = \mathbb{E} \prod_{t=1}^s W_{j_t, k_t} + \sum_{\substack{5 \leq |\alpha| \leq 2m/\delta \\ \alpha_1, \alpha_2, \dots, \alpha_s \geq 0}} \tilde{\mathcal{A}}_\alpha \mathbb{E} \mathcal{P}_{\zeta, \alpha} \prod_{t=1}^s S_{j_t k_t} + O(N^{-m}),$$

where $\tilde{\mathcal{A}}$ are independent of (j_t, k_t) , $1 \leq t \leq s$ and $|\tilde{\mathcal{A}}_\alpha| \leq N^{-|\alpha|\delta/10}$.

From this lemma, one can obtain some useful identities (One may refer to (6.35)-(6.38) in [Lee and Yin \(2014\)](#)):

$$\begin{aligned} \left| \mathbb{E} \prod_{t=1}^s G_{j_t k_t}^{\zeta_{\max}} \right| &\leq \left| \mathbb{E} \prod_{t=1}^s G_{j_t k_t}^0 \right| \\ &\quad + \max_{\mathbf{a}, n} (N^{-2})^n (N^{-\delta/20})^{\sum_i |\mathbf{a}_i|} \sum_{\zeta_1, \dots, \zeta_n} \left| \mathbb{E} \mathcal{P}_{\zeta_n, \mathbf{a}_n} \cdots \mathcal{P}_{\zeta_1, \mathbf{a}_1} \prod_{t=1}^s G_{j_t k_t}^0 \right| + O(N^{-m}), \end{aligned} \quad (3.7.9)$$

where

$$n \leq \frac{6m}{\delta} \quad \text{and} \quad \mathbf{a} = \{(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) \mid \mathbf{a}_1 \in \mathbb{N}^s, \mathbf{a}_2 \in \mathbb{N}^{s+|\mathbf{a}_1|}, \dots, \mathbf{a}_n \in \mathbb{N}^{s+\sum_{i=1}^{n-1} |\mathbf{a}_i|}, 5 \leq |\mathbf{a}_i| \leq \frac{2m}{\delta}\} \quad (3.7.10)$$

And for a coefficient function $h(I, J)$ with

$$\sum_{J, K} h(J, K) = 1, \quad h(J, K) \geq 0, \quad J = (j_1, \dots, j_s) \in \mathcal{I}_1^s, \quad K = (k_1, \dots, k_s) \in \mathcal{I}_1^s,$$

we have

$$\begin{aligned} \left| \mathbb{E} \sum_{J, K} h(J, K) \prod_{t=1}^s G_{j_t k_t}^{\zeta_{\max}} \right| &\leq \left| \mathbb{E} \sum_{J, K} h(J, K) \prod_{t=1}^s G_{j_t k_t}^0 \right| \\ &+ \max_{\mathbf{a}, n, \zeta} (N^{-\delta/20})^{\sum_i |\mathbf{a}_i|} \left| \mathbb{E} \sum_{J, K} h(J, K) \mathcal{P}_{\zeta_n, \mathbf{a}_n} \cdots \mathcal{P}_{\zeta_1, \mathbf{a}_1} \prod_{t=1}^s G_{j_t k_t}^0 \right| + O(N^{-m}). \end{aligned} \quad (3.7.11)$$

We also have (3.7.9) and (3.7.11) hold if we replace G by $(G - \Pi)$.

Lemma 3.7.5. *Under the assumptions of Theorem 3.3.3, we have for any $z \in \mathcal{D}(\tilde{c}, \tau, c_0)$,*

$$\mathbb{E} |\langle \mathbf{w}_1, (G(z) - \Pi(z)) \mathbf{w}_2 \rangle|^2 \prec \Psi^2. \quad (3.7.12)$$

Proof. The proof is the same as that of Lemma 3.7 in Yang (2019). $\square$

The above lemma gives more accurate second moment estimate than (3.3.14) for $\mathbf{X}$ with large support $q = N^{-\delta}$, which will be used in the proof of the main result (3.2.20) (see (3.7.45) below).

Lemma 3.7.6. *Let $\mathbf{X}$ and $\tilde{\mathbf{X}}$ satisfy the assumptions of Lemma 3.7.1. For any $z \in \mathcal{D}(\tilde{c}, \tau, c_0)$, if there exist deterministic quantities J_1 and J_2 , such that*

$$\tilde{G}(z) - \Pi(z) = O_{\prec}(J_1), \quad |\tilde{s}_N(z) - s(z)| \prec J_2, \quad (3.7.13)$$

then for any $p \in 2\mathbb{N}$, we have

$$\mathbb{E} |s_N(z) - s(z)|^p \prec \mathbb{E} |\tilde{s}_N(z) - s(z)|^p + (J_1^2 + J_2 + \Psi^2)^p. \quad (3.7.14)$$

Proof of Lemma 3.7.6. For simplicity, we show that

$$|\mathbb{E}(s_N - s)^p| \prec |\mathbb{E}(\tilde{s}_N - s)^p| + (J_1^2 + J_2 + \Psi^2)^p. \quad (3.7.15)$$

If we take a coefficient function $h(I, J)$ with

$$h(J, K) = M^{-p} \prod_{t=1}^p \delta_{j_t, k_t}, \quad J = (j_1, \dots, j_p) \in \mathcal{I}_1^p, \quad K = (k_1, \dots, k_p) \in \mathcal{I}_1^p,$$

then we see that

$$\mathbb{E} \sum h(J, K) \prod_{t=1}^p (G - \Pi)_{j_t k_t} = \mathbb{E}(s_N - s)^p, \quad (3.7.16)$$

and this is also true if we replace G, s_N with $\tilde{G}, \tilde{s}_N$ respectively. Hence by using (3.7.11) we just need to consider the terms

$$\sum_{j_t} \mathcal{P}_{\zeta_n, \mathbf{a}_n} \cdots \mathcal{P}_{\zeta_1, \mathbf{a}_1} \prod_{t=1}^p (\tilde{G} - \Pi)_{j_t j_t} \quad (3.7.17)$$

for any fixed $\zeta_1, \dots, \zeta_n$ and $\mathbf{a}_1, \dots, \mathbf{a}_n$ satisfying (3.7.10). We know that (3.7.17) is the sum of at most $C^{\sum_i |\mathbf{a}_i|} = O(1)$ products of generalized $\tilde{G}$ entries or $(\tilde{G} - \Pi)$, which is of order $O_{\prec}(1)$ according to the anisotropic local law (3.3.14). Furthermore, there are two j_t 's in the indices of $\tilde{G}$ in each product term. These two j_t 's can appear as (a) $\tilde{G}_{j_t j_t} - \Pi_{j_t j_t}$, or (b) $\tilde{G}_{j_t \mathbf{u}_i} \tilde{G}_{\mathbf{v}_\mu j_t}$ (or exchange $\mathbf{u}_i, \mathbf{v}_\mu$). In the case (a) after averaging over j_t the term becomes $\tilde{s}_N - s$ which is bounded by J_2 . In the case (b), after averaging over j_t the term is bounded by

$$\frac{\Im \tilde{G}_{\mathbf{u}_i \mathbf{u}_i} + \Im \tilde{G}_{\mathbf{v}_\mu \mathbf{v}_\mu}}{N\eta} = O_{\prec}(\Psi^2 + J_1^2).$$

Hence we conclude that

$$\frac{1}{M^p} \sum_{j_1, \dots, j_p} \mathcal{P}_{\zeta_n, \mathbf{a}_n} \cdots \mathcal{P}_{\zeta_1, \mathbf{a}_1} \prod_{t=1}^p (\tilde{G} - \Pi)_{j_t j_t} \prec (\Psi^2 + J_1^2 + J_2)^p. \quad (3.7.18)$$

This together with (3.7.11) for $(G - \Pi)$ concludes the proof. $\square$

The following Corollary concludes the proof of (3.3.15) and (3.3.16).

Corollary 3.7.7. *Let $\mathbf{X}$ satisfy Assumption 1 with support $N^{-\delta}$ for any $0 < \delta < 1/2$. We have for $z \in \mathcal{D}(\tilde{c}, \tau, c_0)$,*

$$|s_N(z) - s(z)| \prec \frac{1}{N\eta}. \quad (3.7.19)$$

Moreover, for $z \in \mathcal{D}(\tilde{c}, \tau, c_0) \cap \{z = E + i\eta : E \geq \lambda_r, N\eta\sqrt{\kappa + \eta} > N^\tau\}$, we have the following stronger estimate

$$|s_N(z) - s(z)| \prec \frac{N^{-c_0/2}}{N\eta} + \frac{1}{N(\kappa + \eta)} + \frac{1}{(N\eta)^2\sqrt{\kappa + \eta}}. \quad (3.7.20)$$

Proof. By taking $J_1 = \Psi(z)$ and $J_2 = \frac{1}{N\eta}$ in Lemma 3.7.6, together with an application of Markov inequality, one can obtain (3.7.19). Similarly, (3.7.20) is obtained by taking $J_1 = \Psi(z)$ and $J_2 = \frac{N^{-c_0/2}}{N\eta} + \frac{1}{N(\kappa + \eta)} + \frac{1}{(N\eta)^2\sqrt{\kappa + \eta}}$. $\square$

The following proposition is used in the proof of Theorem 3.3.6.

Proposition 3.7.8. *Let $\mathbf{X}$ and $\tilde{\mathbf{X}}$ satisfy the assumptions of Lemma 3.7.1. Let $\epsilon > 0$ and set $\eta = N^{-2/3-\epsilon}$. Suppose $K : \mathbb{R} \rightarrow \mathbb{R}$ is a real function satisfying*

$$\sup_x |K^{(k)}(x)|/(|x| + 1)^C \leq C, \quad k = 0, 1, 2, 3, 4.$$

Then for any E, E_1 and E_2 satisfying

$$|E|, |E_1|, |E_2| \leq N^{-2/3+\epsilon}, \quad (3.7.21)$$

there exists a constant $\phi > 0$ such that for any sufficiently large N and for any sufficiently small $\epsilon > 0$, we have

$$|\mathbb{E}K(N\eta\Im s_N(E + \lambda_r + i\eta)) - \mathbb{E}K(N\eta\Im \tilde{s}_N(E + \lambda_r + i\eta))| \leq N^{-\phi} \quad (3.7.22)$$

and

$$\left| EK\left(N \int_{E_1}^{E_2} \Im s_N(x + \lambda_r + i\eta) dx\right) - EK\left(N \int_{E_1}^{E_2} \Im \tilde{s}_N(x + \lambda_r + i\eta) dx\right) \right| \leq N^{-\phi}. \quad (3.7.23)$$

Proof. The proof is basically the same as that for Lemma 5.5 of Lee and Yin (2014) with subtle difference, so we just give an outline here. We only consider (3.7.22) and (3.7.23) is similar. Recall $W^\zeta, S^\zeta, T^\zeta$ defined in (3.7.2). Define

$$x^S := \eta \Im \text{Tr} S = \eta^2 \sum_{jk} S_{jk} \overline{S_{jk}}, \quad (3.7.24)$$

and x^T, x^W are defined by replacing S with T, W respectively. Using (3.3.15) we have estimate

$$\max_{\zeta} \{|x^S| + |x^W|\} \prec N^{C\epsilon}. \quad (3.7.25)$$

We write

$$\mathbb{E}K[N\eta\mathfrak{S}_N(E + \lambda_r + i\eta)] - \mathbb{E}K[N\eta\mathfrak{S}_N(E + \lambda_r + i\eta)] = \sum_{\zeta=1}^{\zeta_{\max}} [\mathbb{E}K(x^S) - \mathbb{E}K(x^T)]. \quad (3.7.26)$$

From the Taylor expansion, we get

$$K(x^S) - K(x^W) = \sum_{r=1}^2 \frac{1}{r!} K^{(r)}(x^W)(x^S - x^W)^r + \frac{1}{3!} K^{(3)}(\zeta_W)(x^S - x^W)^3,$$

where ζ_W lies between x^S and x^W .

Letting $s = 2$ and $\xi = 3/\delta$ in (3.7.5), one obtains

$$x^S = x^W + \sum_{1 \leq l \leq 3/\delta} \left(\sum_{\mathbf{a} \in I_{3/\delta, l}^2} \eta^2 \sum_{j,k} \mathcal{P}_{\zeta, \mathbf{a}}(W_{jk} \overline{W_{jk}}) \right) (-x_{i\mu})^l + O(N^{-3}). \quad (3.7.27)$$

We claim that

$$\mathbb{E} \left| \sum_{j,k} \mathcal{P}_{\zeta, \mathbf{a}}(W_{jk} \overline{W_{jk}}) \right|^p \leq N^{(1+C\epsilon)p}. \quad (3.7.28)$$

Assume this we have from Markov inequality that $|\mathcal{P}_{\zeta, \mathbf{a}} x^W| \prec N^{-1/3}$. Then from (3.7.27) we get

$$\mathbb{E}|x^S - x^W|^3 \leq N^{-5/2+C\epsilon}. \quad (3.7.29)$$

One can prove

$$|\mathbb{E}[K^{(r)}(x^W)(x^S - x^W)^r] - \mathbb{E}[K^{(r)}(x^W)(x^T - x^W)^r]| \leq N^{-2-\delta+C\epsilon} \quad (3.7.30)$$

the same as (6.103) in Lee and Yin (2014). Then we conclude (3.7.22) by combining (3.7.25), (3.7.26), (3.7.29) and (3.7.30)

What remains is to show the claim in (3.7.28). Applying (3.7.11) with $s = 2p$ and

$$h(J, K) = N^{-2p}, \quad J = (j_1, \dots, j_p), \quad K = (k_1, \dots, k_p),$$

the problem can be reduced to estimate the term

$$\sum_{J, K} \mathcal{P}_{\zeta_n, \mathbf{a}_n} \cdots \mathcal{P}_{\zeta_1, \mathbf{a}_1} \left(\mathcal{P}_{\zeta, \mathbf{a}} (\tilde{G}_{j_t k_t} \overline{\tilde{G}_{j_t k_t}}) \right), \quad (3.7.31)$$

with $\mathbf{a} \in \mathbb{N}^2$ and $\mathbf{a}_1, \dots, \mathbf{a}_n$ satisfy (3.7.10) ($s = 2p$). Since $\mathbf{a} \neq 0$, each product term contains one of the following three forms:

$$\tilde{G}_{j_t[i]} \tilde{G}_{[\mu]k_t} \overline{\tilde{G}_{j_t k_t}}, \quad \tilde{G}_{j_t k_t} \overline{\tilde{G}_{j_t[i]} \tilde{G}_{[\mu]k_t}}, \quad \tilde{G}_{j_t[i]} \tilde{G}_{[\mu]k_t} \overline{\tilde{G}_{j_t[i]} \tilde{G}_{[\mu]k_t}} \quad (3.7.32)$$

From (3.3.14) we have estimates

$$|\tilde{G}_{j_t k_t}| \prec \delta_{j_t k_t} + N^{-1/3+\epsilon}, \quad |\tilde{G}_{j_t[i]}| \prec |\mathbf{u}_i(j_t) \Pi_{j_t j_t}| + N^{-1/3+\epsilon}, \quad |\tilde{G}_{[\mu]k_t}| \prec |\mathbf{v}_\mu(\bar{k}_t) \Pi_{j_t j_t}| + N^{-1/3+\epsilon}.$$

Applying $N^{-2} \sum_{j_t, k_t}$, all in (3.7.32) are bounded by $N^{-1+3\epsilon}$. This assures (3.7.28) $\square$

3.7.2 Proof of Theorem 3.2.1

In this subsection, we denote by $Q(\mathbf{X})$ the matrix $(\mathbf{D} + \mathbf{U}\mathbf{X}\mathbf{V}^*)(\mathbf{D} + \mathbf{U}\mathbf{X}\mathbf{V}^*)^*$ and $\mathbf{X}$ will have some variations.

Proof of sufficient condition. For a fixed small constant $\delta > 0$, we define

$$\alpha := \alpha_N = \mathbb{P}(|x_{11}| > N^{-\delta}), \quad \beta := \beta_N = \mathbb{E}[\mathbf{1}(|x_{11}| > N^{-\delta})x_{11}]. \quad (3.7.33)$$

By (3.2.4) and integration by parts we have

$$\alpha \leq CN^{-2+4\delta}, \quad |\beta| \leq CN^{-2+3\delta}. \quad (3.7.34)$$

We denote the distribution of x_{11} by $\rho(x)$. Define random variables $x_{i\mu}^S, x_{i\mu}^L$ with densities respectively defined by

$$\rho_S(x + \beta) = \mathbf{1}(|x| \leq N^{-\delta}) \frac{\rho(x)}{1 - \alpha}, \quad \rho_L(x + \beta) = \mathbf{1}(|x| > N^{-\delta}) \frac{\rho(x)}{\alpha}, \quad (3.7.35)$$

and let $x_{i\mu}^C$ be bernoulli random variables that take value 1 with probability α and 0 otherwise. Moreover, let $x_{i\mu}^S, x_{i\mu}^L, x_{i\mu}^C$ be independent. Denote by $\mathbf{X}^S, \mathbf{X}^L$ and $\mathbf{X}^C$ the random matrices with entries of $\mathbf{X}$ replaced by $x_{i\mu}^S, x_{i\mu}^L, x_{i\mu}^C$ respectively. From the formulas of densities we have

$$x_{i\mu} \sim x_{i\mu}^S(1 - x_{i\mu}^C) + x_{i\mu}^L x_{i\mu}^C + \beta. \quad (3.7.36)$$

Using the Wely's inequality, the β will not influence the edge behaviour of μ_1 .

Let $\tilde{\mathbf{X}}$ be the matrix with entries $\tilde{x}_{i\mu} = x_{i\mu}^L - x_{i\mu}^S$. Using (3.2.4) we have for any constant $\epsilon_0 > 0$,

$$\mathbb{P}(|\tilde{x}_{i\mu} x_{i\mu}^C| \geq \epsilon_0) \leq \mathbb{P}(|x_{i\mu}| \geq \frac{\epsilon_0}{2}) = o(N^{-2}). \quad (3.7.37)$$

Now we define the truncated matrix $\mathbf{T}$ by

$$t_{i\mu} = \mathbf{1}(|\tilde{x}_{i\mu} x_{i\mu}^C| \leq \epsilon_0) \tilde{x}_{i\mu}.$$

Since $\mathbb{P}(\max_{i,\mu} |t_{i\mu} x_{i\mu}^C - \tilde{x}_{i\mu} x_{i\mu}^C|) = 0$ from (3.7.37), we see that it suffices to study the extreme eigenvalue of $Q(\mathbf{X}^S + \mathbf{T} \circ \mathbf{X}^C)$, where $\circ$ represents hadamard product.

By (3.2.4) and integration by parts we have

$$\mathbb{E}x_{i\mu}^S = 0, \quad \mathbb{E}|x_{i\mu}^S|^2 = \frac{1}{N} - O(N^{-2+2\delta}), \quad \mathbb{E}|x_{i\mu}^S|^3 = O(N^{-3/2}), \quad \mathbb{E}|x_{i\mu}^S|^4 = O(N^{-2} \log N).$$

Then by Theorem 3.3.6, the largest eigenvalue of $Q(\mathbf{X}^S)$ satisfies the Tracy-Widom law. Roughly speaking, it remains to show that adding the matrix $\mathbf{T} \circ \mathbf{X}^C$ will not influence the eigenvalue so much.

We further introduce a truncation of $\mathbf{X}^C$ denoted by $\tilde{\mathbf{X}}^C$ whose number of nonzero entries is bounded by

$$\#\{ \{i, \mu\} : \tilde{x}_{i\mu}^C = 1 \} \leq N^{4\epsilon}, \quad (3.7.38)$$

and there does not exists two nonzero entries on the same row or column. Such truncation exists because from the Chernoff bound we have

$$\mathbb{P}(\sum_{i,\mu} |x_{i\mu}^C| > N^{4\epsilon}) < e^{-N^\epsilon},$$

and it follows that

$$\mathbb{P} \left(\exists i = j, \mu \neq \nu \text{ or } i \neq j, \mu = \nu, \text{ s.t. } x_{i\mu}^C = x_{j\mu}^C = 1 \mid \sum_{i,\mu} |x_{i\mu}^C| = n = N^{4\epsilon} \right) = O(n^2 N^{-1}).$$

Therefore we only need to study the largest eigenvalue of $Q(\mathbf{X}^S + \mathbf{E})$, where $E_{i\mu} = t_{i\mu} \tilde{x}_{i\mu}^C$. In the following, to conclude the Theorem, it suffices to show that

$$\mathbb{P}(|\mu_1^S - \mu_1^E| \leq N^{-3/4}) = 1 - o(1), \quad (3.7.39)$$

where μ_1^S and μ_1^E are the largest eigenvalues of $Q(\mathbf{X}^S)$, and $Q(\mathbf{X}^S + \mathbf{E})$ respectively.

Without loss of generality, we may assume that the nonzero elements of $\mathbf{E}$ are $E_{11}, E_{22}, \dots, E_{nn}$. We define H^E and G^E by replacing the $\mathbf{X}$ inside with $\mathbf{X}^S + \mathbf{E}$, see (3.3.2) and (3.3.3). And by similar way we can define H^S and G^S . We have decomposition

$$\begin{pmatrix} 0 & \mathbf{E} \\ \mathbf{E}^* & 0 \end{pmatrix} = OD_E O^*, \quad (3.7.40)$$

where

$$D_E = \text{diag}\{e_{11}, e_{22}, \dots, e_{nn}, -e_{11}, -e_{22}, \dots, -e_{nn}\},$$

and O is an $(M + N) \times 2n$ matrix such that

$$O_{ab} = \begin{cases} \delta_{a,i}/\sqrt{2} + \delta_{a,M+i}/\sqrt{2} & \text{if } b = i, i \leq n, \\ \delta_{a,i}/\sqrt{2} - \delta_{a,M+i}/\sqrt{2} & \text{if } b = i + n, i \leq n. \end{cases}$$

Let

$$W = \begin{pmatrix} \mathbf{U}^* & 0 \\ 0 & \mathbf{V}^* \end{pmatrix} O, \quad F^\theta(z) = W^* G^S(z) W + (\theta D_E)^{-1}. \quad (3.7.41)$$

Notice that $\det(H) = (-1)^N \det[Q(\mathbf{X}) - z\mathbf{I}_M]$, we have a simple result (see Knowles and Yin (2013)): If μ is not an eigenvalue of $Q(\mathbf{X}^S)$, then μ is an eigenvalue of $Q(\mathbf{X}^S + \theta\mathbf{E})$ if and only if $\det[F^\theta(\mu)] = 0$. Let

$$\mu_0^\pm = \mu_1^S \pm N^{-3/4},$$

and we claim that

$$\mathbb{P}(\det[F^\theta(\mu_0^\pm)] \neq 0 \text{ for } 0 < \gamma \leq 1) = 1 - o(1). \quad (3.7.42)$$

Assuming this claim, we see that $\mu_0^\pm$ are not eigenvalues of $Q(\mathbf{X}^S + \theta \mathbf{E})$ for $0 < \theta \leq 1$ with probability $1 - o(1)$. By the continuity of eigenvalues with respect to θ , we conclude (3.7.39).

In the remaining, we prove the claim in (3.7.42). We take $\mu_0 = \mu_0^+$ as illustration and for μ_0^- it is the same. For simplicity we denote the eigenvalues of $Q(\mathbf{X}^S)$ by $\mu_1, \dots, \mu_M$. Let $z_1 = \lambda_r + i\eta_0$, $z_0 = \mu_0 + i\eta_0$ with $\eta_0 = N^{-3/4}$. We write $F^\theta(\mu_0)$ as

$$W^*(G^S(\mu_0) - G^S(z_1))W + W^*(G^S(z_1) - \Pi(z_1))W + W^*\Pi(z_1)W + (\theta D_E) \quad (3.7.43)$$

For sufficiently small $\epsilon > 0$, we define events

$$\begin{aligned} \mathcal{B}_1 &= \left\{ \min_k |\mu_1 - \mu_2| \geq 2N^{-3/4}, |\mu_1 - \lambda_r| \leq N^{-2/3+\epsilon} \right\}, \\ \mathcal{B}_2 &= \left\{ \max_{1 \leq j, k \leq n} |[W^*(G^S(z_1) - \Pi(z_1))W]_{jk}| \leq N^{-1/5} \right\}. \end{aligned} \quad (3.7.44)$$

From Remark 11 we see that $\mathbb{P}(\mu_1 - \mu_2 \geq 2N^{-3/4}) = 1 - o(1)$. Together with the rigidity property (3.3.20), we have $\mathbb{P}(\mathcal{B}_1) = 1 - o(1)$. From Lemma 3.7.5, we get for any $1 \leq j, k \leq n$,

$$\mathbb{E}[|W^*(G^S(z_1) - \Pi(z_1))W]_{jk}|^2] \prec N^{-1/2}. \quad (3.7.45)$$

Then by Markov inequality and taking union bound, $\mathbb{P}(\mathcal{B}_2) = 1 - o(1)$. In the following we estimate (3.7.43) under the event $\mathcal{B}_1 \cap \mathcal{B}_2$.

We deal with the first term in (3.7.43) first. We have for any two columns of W denoted by $\mathbf{w}_1, \mathbf{w}_2$ (it is allowed that $\mathbf{w}_1 = \mathbf{w}_2$),

$$\begin{aligned} |\langle \mathbf{w}_1, (G^S(\mu_0) - G^S(z_0))\mathbf{w}_2 \rangle| &\leq \sum_k |\langle \mathbf{w}_1, \boldsymbol{\eta}_k \rangle \langle \mathbf{w}_2, \boldsymbol{\eta}_k \rangle| \left| \frac{1}{\mu_k - \mu_0} - \frac{1}{\mu_k - z_0} \right| \\ &\leq C \sum_k (|\langle \mathbf{w}_1, \boldsymbol{\eta}_k \rangle|^2 + |\langle \mathbf{w}_2, \boldsymbol{\eta}_k \rangle|^2) \frac{\eta_0}{(\mu_k - \mu_0)^2 + \eta_0^2} = C \Im G_{\mathbf{w}_1 \mathbf{w}_1}^S(z_0) + \Im G_{\mathbf{w}_2 \mathbf{w}_2}^S(z_0). \end{aligned} \quad (3.7.46)$$

To bound above, we consider the difference of $G_{\mathbf{w}_1\mathbf{w}_2}^S(z_0)$ and $G_{\mathbf{w}_1\mathbf{w}_2}^S(z_1)$ by

$$\begin{aligned}
& |\langle \mathbf{w}_1, (G^S(z_0) - G^S(z_1))\mathbf{w}_2 \rangle| \\
& \leq C \sum_k (|\langle \mathbf{w}_1, \boldsymbol{\eta}_k \rangle|^2 + |\langle \mathbf{w}_2, \boldsymbol{\eta}_k \rangle|^2) \frac{|\mu_0 - \lambda_r|}{\sqrt{(\mu_k - \mu_0)^2 + \eta_0^2} \sqrt{(\mu_k - \lambda_r)^2 + \eta_0^2}} \\
& \leq C \sum_k (|\langle \mathbf{w}_1, \boldsymbol{\eta}_k \rangle|^2 + |\langle \mathbf{w}_2, \boldsymbol{\eta}_k \rangle|^2) \frac{N^{1/6+2\epsilon}\eta_0}{|\mu_k - \lambda_r|^2 + \eta_0^2} \quad (3.7.47) \\
& \leq CN^{1/6+2\epsilon} (\Im G_{\mathbf{w}_1\mathbf{w}_1}^S(z_1) + \Im G_{\mathbf{w}_2\mathbf{w}_2}^S(z_1)) \\
& \leq CN^{1/6+2\epsilon} (\Im s(z_1) + N^{-1/5}) \leq CN^{-\frac{1}{30}+2\epsilon},
\end{aligned}$$

where in the second step we use

$$\begin{aligned}
\sqrt{|\mu_k - \lambda_r|^2 + \eta_0^2} & \sim |\mu_k - \lambda_r| + \eta_0 \leq |\mu_k - \mu_0| + |\mu_0 - \lambda_r| + \eta_0 \\
& \leq N^{\frac{1}{12}+\epsilon} (|\mu_k - \mu_0| + \eta_0) \sim N^{\frac{1}{12}+\epsilon} \sqrt{|\mu_k - \mu_0|^2 + \eta_0^2}. \quad (3.7.48)
\end{aligned}$$

From (3.7.47) and $\mathcal{B}_2$ we conclude that the right side of (3.7.46) is bounded by $N^{-1/30+2\epsilon}$. Together with (3.7.47), we conclude with probability $1 - o(1)$,

$$\|W^*(G^S(\mu_0) - G^S(z_1))W\|_2 = nN^{-1/30+2\epsilon} = N^{-1/30+6\epsilon}. \quad (3.7.49)$$

The second term in (3.7.43) has spectral norm bounded by $N^{-1/5+4\epsilon}$. The third is $O(1)$ since

$$\|\Pi\| = O(1). \quad (3.7.50)$$

Then by choosing sufficiently small positive constant ϵ_0 (see (3.7.37)), we have for all $0 < \theta \leq 1$, with probability $1 - o(1)$,

$$\|W^*(G^S(\mu_0) - G^S(z_1))W + W^*(G^S(z_1) - \Pi(z_1))W + W^*\Pi(z_1)W\| < \frac{1}{\theta\epsilon_0} \quad (3.7.51)$$

This implies (3.7.42). We complete the proof of (3.2.20).

Proof of necessary condition. To avoid confusion, we use $x_{i\mu}^{(N)}$ to denote entries of $\mathbf{X}_{cN \times N}$. If (3.2.4) does not hold, we can find a sequence of $b_{N_1} \rightarrow \infty$, and some constant $c_1 > 0$, s.t.

$$\mathbb{P}(|\sqrt{N}x_{i\mu}^{(N)}| \geq b_{N_1}) \geq c_1 b_{N_1}^{-4}. \quad (3.7.52)$$

Choose fixed constant $d > \max\{2\|\mathbf{R}\|, 2\sqrt{T}\}$, and let

$$\Gamma_N = \{\text{There exists } i, \mu, \text{ s.t. } |x_{i\mu}^{(N)}| > d\}. \quad (3.7.53)$$

Choose $N_2 = \lceil (\frac{b_{N_1}}{d})^2 \rceil$. We have

$$\begin{aligned} 1 - \mathbb{P}(\Gamma_{N_2}) &= \left(1 - \mathbb{P}(|x_{i\mu}^{(N_2)}| \geq d)\right)^{N_2^2} \leq \left(1 - P(\sqrt{N_2}|x_{i\mu}^{(N_2)}| > b_{N_1})\right)^{N_2^2} \\ &\leq (1 - c_1 b_{N_1}^{-4})^{N_2^2} \leq (1 - \frac{c_1}{N_2^2 d^4})^{N_2^2} < c_2, \end{aligned} \quad (3.7.54)$$

for some constant $0 < c_2 < 1$ independent of N_2 , and under the event Γ_{N_2} , the largest eigenvalue of $cN_2 \times N_2$ matrix $(\mathbf{R} + \mathbf{X})(\mathbf{R} + \mathbf{X})^*$ satisfy

$$\mu_1 \geq \max_{i,\mu} |x_{i\mu} + r_{i\mu}|^2 > \frac{d^2}{4} > T. \quad (3.7.55)$$

From (3.7.54) and (3.7.55), we conclude (3.2.21).

3.8 Appendix

In this appendix we establish the so called “optical theorems” as in Appendix B of Lee and Schnelli (2016), or see Lee and Schnelli (2015). Optical theorems ensure that the expectations of certain linear combinations of terms in (3.4.32) and (3.4.33) have orders smaller than their naive size obtained from local laws directly, which play crucial roles in the proof of (3.4.4). In the meantime, we will derive many useful formulas about the expectations of terms in (3.4.32) and (3.4.33) in the process of obtaining the “optical theorems”, and they will be used in the proof of Lemma 3.4.3.

3.8.1 Preliminaries

Recall $\hat{G}$ defined in (3.4.19) which is denoted by G in this section, and other notations introduced in Subsection 3.4.3 that we will use tacitly till the end. Moreover we need some more notations and useful expressions that will be used frequently later. First we consider the expansion of term G_{aa} for any $a \in \mathcal{I}_M$. Using schur

complement formula, we have

$$\begin{aligned} \frac{1}{G_{aa}} &= -z - \gamma d_a^2 G_{\bar{a}\bar{a}}^{(a)} - \gamma \sum_{\mu, \nu} x_{a\mu} G_{\mu\nu}^{(a)} x_{a\nu} - 2\gamma d_a \sum_{\mu} G_{\bar{a}\mu}^{(a)} x_{a\mu}, \\ \frac{1}{G_{\bar{a}\bar{a}}^{(a)}} &= -1 - \gamma \sum_{j, k}^{(a)} x_{j\bar{a}} G_{jk}^{(a\bar{a})} x_{ka}. \end{aligned} \quad (3.8.1)$$

Recall notations b in (3.4.7) and $\tilde{b}$ in (3.4.9). Denote

$$\begin{aligned} \beta_a &:= \frac{1}{-\gamma d_a^2 G_{\bar{a}\bar{a}}^{(a)} - \tilde{b}}, \quad \mathcal{Y}_a := \gamma \left(cm(E_+) - \sum_{j, k}^{(a)} x_{j\bar{a}} G_{jk}^{(a\bar{a})} x_{k\bar{a}} \right), \\ \mathcal{Z}_a &:= E_+ - z + \gamma \left(E_+ \tilde{m}(E_+) - \sum_{\mu, \nu} x_{a\mu} G_{\mu\nu}^{(a)} x_{a\nu} - 2d_a \sum_{\mu} G_{\bar{a}\mu}^{(a)} x_{a\mu} \right). \end{aligned} \quad (3.8.2)$$

Lemma 3.8.1. *Recall (3.4.22). For z in $\mathcal{E}_\epsilon(t)$ defined in (3.4.20),*

$$m(E_+) - m(z) = O(\Psi_0), \quad (3.8.3)$$

and

$$\mathcal{Y}_a, \mathcal{Z}_a \prec \Psi_0. \quad (3.8.4)$$

Proof. We show (3.8.3) first. Recalling that $z = \hat{\phi}(w)$, $E_+ = \hat{\phi}(\xi)$ and $\hat{\phi}'(\xi) = 0$, we have for some sufficiently small constant ϵ_1 ,

$$z - E_+ = \frac{1}{2} \hat{\phi}''(\xi)(w - \xi)^2 + O(|w - \xi|^3), \text{ for } |z - E_+| \leq \epsilon_1,$$

where we use a simple fact $\hat{\phi}'''(\xi) < C$, where C is a large constant. By calculations, we find there exists a positive constant c_1 such that

$$\hat{\phi}''(\xi) = \frac{1}{N} \sum_i \frac{\frac{-4\gamma^2 d_i^2}{b} - 2(1-c)\gamma^2}{(\gamma d_i^2 - \xi)^3} + 2\xi\gamma^2 \left(\frac{1}{N} \sum_i \frac{1}{(\gamma d_i^2 - \xi)^2} \right)^2 > c_1 > 0,$$

where we use Assumptions 2 and 3. Therefore, we get

$$|w(z) - w(E_+)| = \sqrt{|z - E_+|}(1 + o(1)) = O(\Psi_0).$$

Then (3.8.3) follows from this.

Then we deal with $\mathcal{Y}_a$. Write

$$\mathcal{Y}_a = (cm - \frac{1}{N} \sum_j G_{jj}) + \left(\frac{1}{N} \sum_j G_{jj} - \frac{1}{N} \sum_j^{(a)} G_{jj}^{(a\bar{a})} \right) + \left(\frac{1}{N} \sum_j^{(a)} G_{jj}^{(a\bar{a})} - \sum_{j,k}^{(a)} x_{j\bar{a}} G_{jk}^{(a\bar{a})} x_{k\bar{a}} \right).$$

From (3.4.24) and (3.3.32), we see that terms in the first two brackets are of order $O(\Psi_0)$. Using (3.3.7) and Ward identity, we find for the last term that

$$\left| \frac{1}{N} \sum_j^{(a)} G_{jj}^{(a\bar{a})} - \sum_{j,k}^{(a)} x_{j\bar{a}} G_{jk}^{(a\bar{a})} x_{k\bar{a}} \right| \prec \left(\frac{1}{N^2} \sum_{j,k}^{(a)} |G_{jk}^{(a\bar{a})}|^2 \right)^{1/2} \prec \Psi_0,$$

Therefore we conclude $\mathcal{Y}_a \prec \Psi_0$. $\mathcal{Z}_a$ can be estimated similar to $\mathcal{Y}_a$ with the help of (3.5.25). $\square$

In the remaining of the appendix, we always consider for $z \in \mathcal{E}_\epsilon(t)$, and Ψ_0 is defined by (3.4.22). From (3.8.1), (3.8.2) and (3.8.4), it is easy to obtain the following from Taylor's expansion,

$$\begin{aligned} G_{aa} &= \beta_a - \beta_a^2 \mathcal{Z}_a + \beta_a^3 \mathcal{Z}_a^2 + O(\Psi_0^3), \\ G_{a\bar{a}}^{(a)} &= -\frac{1}{b} - \frac{1}{b^2} \mathcal{Y}_a - \frac{1}{b^3} \mathcal{Y}_a^2 + O(\Psi_0^3). \end{aligned} \quad (3.8.5)$$

Recalling that $\xi = b\tilde{b}$ from (3.4.10), using the Taylor expansion again, we get

$$\begin{aligned} \beta_a &= \frac{b}{\gamma d_a^2 - \xi} - \frac{\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} \mathcal{Y}_a + \frac{\gamma d_a^2 \tilde{b}}{(\gamma d_a^2 - \xi)^3} \mathcal{Y}_a^2 + O(\Psi_0^3), \\ \beta_a G_{a\bar{a}}^{(a)2} &= \frac{1}{(\gamma d_a^2 - \xi)b} + \frac{\gamma d_a^2 - 2\xi}{(\gamma d_a^2 - \xi)^2 b^2} \mathcal{Y}_a + O(\Psi_0^2), \\ \beta_a^2 G_{a\bar{a}}^{(a)2} &= \frac{1}{(\gamma d_a^2 - \xi)^2} - \frac{2\tilde{b}}{(\gamma d_a^2 - \xi)^3} \mathcal{Y}_a + \frac{3\tilde{b}^2}{(\gamma d_a^2 - \xi)^4} \mathcal{Y}_a^2 + O(\Psi_0^3). \end{aligned} \quad (3.8.6)$$

Next, we can expand $G_{\bar{a}\bar{a}}$ similarly. Let

$$\begin{aligned} \alpha_a &:= \frac{1}{-\gamma d_a^2 G_{a\bar{a}}^{(\bar{a})} - b}, \quad \mathcal{U}_a := E_+ - z + \gamma \left(E_+ \tilde{m}(E_+) - \sum_{\mu, \nu}^{(\bar{a})} x_{a\mu} G_{\mu\nu}^{(a\bar{a})} G_{a\nu} \right), \\ \mathcal{V}_a &:= \gamma \left(cm(E_+) - \sum_{j,k} x_{j\bar{a}} G_{jk}^{(\bar{a})} x_{k\bar{a}} - 2d_a \sum_j G_{aj}^{(\bar{a})} x_{j\bar{a}} \right). \end{aligned} \quad (3.8.7)$$

Then it is easy to obtain that

$$\begin{aligned} G_{\bar{a}\bar{a}} &= \alpha_a - \alpha_a^2 \mathcal{V}_a + \alpha_a^3 \mathcal{V}_a^2 + O(\Psi_0^3), \\ G_{aa}^{(\bar{a})} &= -\frac{1}{\tilde{b}} - \frac{1}{\tilde{b}^2} \mathcal{U}_a - \frac{1}{\tilde{b}^3} \mathcal{U}_a^2 + O(\Psi_0^3). \end{aligned} \quad (3.8.8)$$

And we have

$$\begin{aligned} \alpha_a &= \frac{\tilde{b}}{\gamma d_a^2 - \xi} - \frac{\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} \mathcal{U}_a + \frac{\gamma d_a^2 b}{(\gamma d_a^2 - \xi)^3} \mathcal{U}_a^2 + O(\Psi_0^3), \\ \alpha_a G_{aa}^{(\bar{a})^2} &= \frac{1}{(\gamma d_a^2 - \xi) \tilde{b}} + \frac{\gamma d_a^2 - 2\xi}{(\gamma d_a^2 - \xi)^2 \tilde{b}^2} \mathcal{U}_a + O(\Psi_0^2), \\ \alpha_a^2 G_{aa}^{(\bar{a})^2} &= \frac{1}{(\gamma d_a^2 - \xi)^2} - \frac{2b}{(\gamma d_a^2 - \xi)^3} \mathcal{U}_a + \frac{3b^2}{(\gamma d_a^2 - \xi)^4} \mathcal{U}_a + O(\Psi_0^3). \end{aligned} \quad (3.8.9)$$

For $\alpha > 2M$, we have

$$G_{\alpha\alpha} = -\frac{1}{b} - \frac{\mathcal{V}_\alpha}{b^2} - \frac{\mathcal{V}_\alpha^2}{b^3} + O(\Psi_0^3), \quad (3.8.10)$$

where

$$\mathcal{V}_\alpha = \gamma(cm - \sum_{j,k} x_{j\alpha} G_{jk}^{(\alpha)} G_{k\alpha}).$$

Similar to Lemma 3.8.1, we have

$$\mathcal{U}_a, \mathcal{V}_a \prec \Psi_0, \quad a \in \mathcal{I}_M; \quad \mathcal{V}_\alpha \prec \Psi_0, \quad \alpha \in \llbracket 2M+1, M+N \rrbracket. \quad (3.8.11)$$

We repeat equations (3.4.11) and (3.4.12) here, since they play crucial role in deriving the optical theorems,

$$\frac{1}{N} \sum_{i=1}^M \frac{\gamma b(E_+ b^2 + \xi)}{(\gamma d_i^2 - \xi)^2} = 1, \quad (3.8.12)$$

and

$$\frac{1}{\gamma^2} \frac{1}{N} \sum_{i=1}^M \frac{b^2}{(\gamma d_i^2 - \xi)^2} = -\frac{1}{\gamma b^3} - \frac{1}{N} \sum_{i=1}^M \frac{(2E_+ b - \gamma(1-c))^2}{(\gamma d_i^2 - \xi)^3} - \frac{1}{N} \sum_{i=1}^M \frac{E_+}{(\gamma d_i^2 - \xi)^2} \quad (3.8.13)$$

For $k \in \mathbb{N}$, let

$$\begin{aligned}\varphi_k &:= \frac{1}{N} \sum_a \frac{1}{(\gamma d_a^2 - \xi)^k}, \quad \psi_k := \frac{1}{N} \sum_a \frac{\gamma d_a^2}{(\gamma d_a^2 - \xi)^k}, \quad \varpi_2 := \sum_a \frac{\tilde{b}^2}{(\gamma d_a^2 - \xi)^2} + \frac{1-c}{b^2}. \\ \Phi_1 &:= \frac{1}{N} \sum \left(\frac{\gamma^3 d_a^2 \tilde{b}}{(\gamma d_a^2 - \xi)^3} + \frac{2\gamma^3 d_a^2 b E_+}{(\gamma d_a^2 - \xi)^3} + \frac{\gamma^2 b^3 E_+^2}{(\gamma d_a^2 - \xi)^3} \right), \\ \Phi_2 &:= \frac{1}{N} \sum \left(\frac{\gamma^3 d_a^2 b E_+^2}{(\gamma d_a^2 - \xi)^3} + \frac{2\gamma^3 d_a^2 \tilde{b} E_+}{(\gamma d_a^2 - \xi)^3} + \frac{\gamma^2 \tilde{b}^3}{(\gamma d_a^2 - \xi)^3} \right) - \gamma^2 (1-c) \frac{1}{b^3}.\end{aligned}\tag{3.8.14}$$

The above quantities $\varphi_2, \psi_2, \varpi_2, \Phi_1, \Phi_2$ will appear naturally in the derivation of optical theorems, and the following lemma will help simplify the calculations, see e.g. (3.8.78).

Lemma 3.8.2. *Denote $h = E_+ b + \tilde{b}$. We have*

$$\varphi_2 = \frac{1}{\gamma b^2 h}, \quad \psi_2 = \frac{1}{\gamma} - E_+ b^2 \varphi_2, \quad \varphi_3 = - \left(\frac{1}{\gamma^3 h^3} + \frac{1}{\gamma b^3 h^2} + \frac{E_+}{\gamma b^2 h^3} \right), \quad \varpi_2 = E_+^2 b^2 \varphi_2\tag{3.8.15}$$

$$\Phi_1 = -b^3 \varphi_2, \quad \text{and} \quad \Phi_2 = -\tilde{b} b^2 \varphi_2.\tag{3.8.16}$$

Proof. From (3.8.12) and (3.4.10), we see that $\varphi_2 = 1/\gamma b^2 h$. It follows that $\psi_2 = \frac{1}{\gamma} - E_+ b^2 \varphi_2$ by using a simple fact $\varphi_1 = (b-1)/\gamma b$. The third relation in (3.8.15) is obtained from (3.8.13). The last one in (3.8.15) can be verified without difficulty. Next we verify Φ_1 ,

$$\begin{aligned}\Phi_1 &= \frac{1}{N} \sum_a \left(\frac{\gamma^3 d_a^2 \tilde{b}}{(\gamma d_a^2 - \xi)^3} + \frac{2\gamma^3 b d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} + \frac{\gamma^2 b^3 E_+^2}{(\gamma d_a^2 - \xi)^3} \right) \\ &= \frac{1}{N} \sum_a \left(\frac{\gamma^2 \tilde{b}}{(\gamma d_a^2 - \xi)^2} + \frac{2\gamma^2 b E_+}{(\gamma d_a^2 - \xi)^2} + \frac{\gamma^2 b \tilde{b}^2 + 2\gamma^2 b^2 \tilde{b} E_+ + \gamma^2 b^3 E_+^2}{(\gamma d_a^2 - \xi)^3} \right) \\ &= \frac{\gamma^2}{N} \sum_a \left(\frac{\tilde{b} + E_+ b}{(\gamma d_a^2 - \xi)^2} + \frac{E_+ b}{(\gamma d_a^2 - \xi)^2} + \frac{b(\tilde{b} + E_+ b)^2}{(\gamma d_a^2 - \xi)^3} \right) \\ &= \frac{\gamma}{b^2} + \frac{\gamma^2}{N} \sum_a \left(\frac{E_+ b}{(\gamma d_a^2 - \xi)^2} + \frac{b(\tilde{b} + E_+ b)^2}{(\gamma d_a^2 - \xi)^3} \right) \\ &= -b^3 \varphi_2,\end{aligned}$$

where in the last step we use (3.8.13). Similarly, we can get $\Phi_2 = -\tilde{b} b^2 \varphi_2$. $\square$

There are some useful relations between the quantities defined in (3.4.32) and (3.4.33). We list those that will be used frequently later, for example, see (3.8.31) below.

Lemma 3.8.3.

$$\mathbb{E}X_{44}''' = E_+ \mathbb{E}X_{44}'' + O(\Psi_0^5) = E_+^2 \mathbb{E}X_{44}' + O(\Psi_0^5) = E_+^3 \mathbb{E}X_{44} + O(\Psi_0^5), \quad (3.8.17)$$

$$EX_{43}'' = E_+ \mathbb{E}X_{43}' + O(\Psi_0^5) = E_+^2 \mathbb{E}X_{43} + O(\Psi_0^5), \quad (3.8.18)$$

$$\mathbb{E}X_{42}' = E_+ \mathbb{E}X_{42} + O(\Psi_0^5), \quad (E_+ - z) \mathbb{E}X_{22}' = E_+ (E_+ - z) \mathbb{E}X_{22} + O(\Psi_0^5). \quad (3.8.19)$$

Remark 18. Actually, the relations depend on how many Greek lower indices in the quantity. When the number of Greek letters in the lower indices decreases by one, meanwhile the number of Latin indices increases by one, the expectation increases by a factor of E_+ with an negligible error of order $O(\Psi_0^5)$.

The following Lemma is also useful in the calculations to derive optical theorems, for example, see (3.8.30) below.

Lemma 3.8.4. *We have*

$$\begin{aligned} \mathbb{E} \frac{1}{N^2} \sum_{\alpha, \beta}^{(\bar{a})} G_{i\alpha}^{(a\bar{a})} G_{\alpha\beta}^{(a\bar{a})} G_{\beta i}^{(a\bar{a})} &= \mathbb{E}X_{33}'' - \frac{\gamma E_+^2 (E_+ b + \tilde{b})}{\gamma d_a^2 - \xi} \mathbb{E}X_{44} + O(\Psi_0^5), \\ \mathbb{E} \frac{1}{N^2} \sum_{j, \beta}^{(a, \bar{a})} G_{ij}^{(a\bar{a})} G_{j\beta}^{(a\bar{a})} G_{\beta i}^{(a\bar{a})} &= \mathbb{E}X_{33}' - \frac{\gamma E_+ (E_+ b + \tilde{b})}{\gamma d_a^2 - \xi} \mathbb{E}X_{44} + O(\Psi_0^5), \\ \mathbb{E} \frac{1}{N^2} \sum_{j, k}^{(a)} G_{ij}^{(a\bar{a})} G_{jk}^{(a\bar{a})} G_{ki}^{(a\bar{a})} &= \mathbb{E}X_{33} - \frac{\gamma (E_+ b + \tilde{b})}{\gamma d_a^2 - \xi} \mathbb{E}X_{44} + O(\Psi_0^5), \\ \mathbb{E} \frac{1}{N} (cm - cm_N^{(a\bar{a})}) \sum_{\alpha}^{(\bar{a})} G_{i\alpha}^{(a\bar{a})} G_{\alpha i}^{(a\bar{a})} &= \mathbb{E}X_{32}' + \frac{\gamma E_+ (E_+ b + \tilde{b})}{\gamma d_a^2 - \xi} (\mathbb{E}\tilde{X}_{44} - 2\mathbb{E}X_{43}) + O(\Psi_0^5), \\ \mathbb{E} \frac{1}{N} (cm - cm_N^{(a\bar{a})}) \sum_j^{(a)} G_{ij}^{(a\bar{a})} G_{ji}^{(a\bar{a})} &= \mathbb{E}X_{32} + \frac{\gamma (E_+ b + \tilde{b})}{\gamma d_a^2 - \xi} (\mathbb{E}\tilde{X}_{44} - 2\mathbb{E}X_{43}) + O(\Psi_0^5). \end{aligned} \quad (3.8.20)$$

The proof of Lemma 3.8.3 and 3.8.4 are postponed to Subsection 3.8.6.

3.8.2 Optical theorem from X_{22}

We start from expanding $\mathbb{E}G_{ia}G_{ai}$ with respect to index a . When $a = i$,

$$\mathbb{E}G_{ii}^2 = \frac{b^2}{(\gamma d_i^2 - \xi)^2} - \frac{2\gamma^2 b d_i^2}{(\gamma d_i^2 - \xi)^3} \mathbb{E}(cm - cm_N) - \frac{2\gamma b^3 E_+}{(\gamma d_i^2 - \xi)^3} \mathbb{E}(cm - cm_N) + O(\Psi_0^2). \quad (3.8.21)$$

For $a \neq i$, using (3.3.29), we write

$$\begin{aligned} G_{ia}G_{ai} &= \sum_{\alpha, \beta} G_{aa}^2 G_{i\alpha}^{(a)} y_{a\alpha} y_{a\beta} G_{i\beta}^{(a)} \\ &= \gamma G_{aa}^2 G_{i\bar{a}}^{(a)} d_a^2 G_{\bar{a}i}^{(a)} + 2\gamma \sum_{\alpha} G_{aa}^2 G_{i\alpha}^{(a)} x_{a\alpha} d_a G_{i\bar{a}}^{(a)} + \gamma \sum_{\alpha, \beta} G_{aa}^2 G_{i\alpha}^{(a)} x_{a\alpha} x_{a\beta} G_{i\beta}^{(a)}. \end{aligned}$$

Denote

$$B_1 := G_{aa}^2 G_{i\bar{a}}^{(a)} d_a^2 G_{\bar{a}i}^{(a)}, \quad B_2 := \sum_{\alpha} G_{aa}^2 G_{i\alpha}^{(a)} x_{a\alpha} d_a G_{i\bar{a}}^{(a)}, \quad B_3 := \sum_{\alpha, \beta} G_{aa}^2 G_{i\alpha}^{(a)} x_{a\alpha} x_{a\beta} G_{i\beta}^{(a)}. \quad (3.8.22)$$

3.8.2.1 Expansion of B_1

Consider B_1 first. Using (3.3.31), we write

$$B_1 = \gamma \sum_{p, q}^{(a)} G_{aa}^2 G_{\bar{a}\bar{a}}^{(a)2} d_a^2 G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})}. \quad (3.8.23)$$

Using (3.8.5) and (3.8.6), we decompose B_1 into summation of six terms defined below:

$$\begin{aligned} B_{11} &:= \sum_{p, q}^{(a)} \frac{\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})}, & B_{12} &:= \sum_{p, q}^{(a)} \frac{-2\gamma \tilde{b} d_a^2 \mathcal{Y}_a}{(\gamma d_a^2 - \xi)^3} G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})}, \\ B_{13} &:= \sum_{p, q}^{(a)} \frac{-2\gamma b d_a^2 \mathcal{Z}_a}{(\gamma d_a^2 - \xi)^3} G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})}, & B_{14} &:= \sum_{p, q}^{(a)} \frac{3\gamma \tilde{b}^2 d_a^2 \mathcal{Y}_a^2}{(\gamma d_a^2 - \xi)^4} G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})}, \\ B_{15} &:= \sum_{p, q}^{(a)} \frac{(2\gamma^2 d_a^4 + 4\gamma \xi d_a^2) \mathcal{Y}_a \mathcal{Z}_a}{(\gamma d_a^2 - \xi)^4} G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})}, & B_{16} &:= \sum_{p, q}^{(a)} \frac{3\gamma b^2 d_a^2 \mathcal{Z}_a^2}{(\gamma d_a^2 - \xi)^4} G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})}. \end{aligned} \quad (3.8.24)$$

In the following, we analyze these terms one by one. We start from B_{11} that requires more calculations to expand to desired order. Recall the partial expectation notations $\mathbb{E}_a$ and $\mathbb{E}_\mu$ defined in Section 3.3.

Expansion of B_{11} . By taking partial expectation $\mathbb{E}_{\bar{a}}$ of B_{11} , we obtain

$$\mathbb{E}_{\bar{a}} B_{11} = \frac{1}{N} \sum_p^{(a)} \frac{\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} G_{ip}^{(a\bar{a})} G_{pi}^{(a\bar{a})}. \quad (3.8.25)$$

Using (3.3.32), write

$$G_{ip}^{(a\bar{a})} G_{pi}^{(a\bar{a})} = \left(G_{ip} - \frac{G_{i\bar{a}} G_{\bar{a}p}}{G_{\bar{a}\bar{a}}} - \frac{G_{ia}^{(\bar{a})} G_{ap}^{(\bar{a})}}{G_{aa}^{(\bar{a})}} \right) G_{pi} - G_{ip}^{(a\bar{a})} \frac{G_{i\bar{a}} G_{\bar{a}p}}{G_{\bar{a}\bar{a}}} - G_{ip}^{(a\bar{a})} \frac{G_{ia}^{(\bar{a})} G_{ap}^{(\bar{a})}}{G_{aa}^{(\bar{a})}}. \quad (3.8.26)$$

Substituting (3.8.26) into (3.8.25), we decompose $E_{\bar{a}} B_{11}$ into summation of following five terms:

$$\begin{aligned} B_{111} &:= \frac{1}{N} \sum_p \frac{\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} G_{ip} G_{pi}, & B_{112} &:= \frac{1}{N} \sum_p \frac{-\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} \frac{G_{i\bar{a}} G_{\bar{a}p} G_{pi}}{G_{\bar{a}\bar{a}}}, \\ B_{113} &:= \frac{1}{N} \sum_p \frac{-\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} \frac{G_{ia}^{(\bar{a})} G_{ap}^{(\bar{a})} G_{pi}}{G_{aa}^{(\bar{a})}}, & B_{114} &:= \frac{1}{N} \sum_p \frac{-\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} \frac{G_{i\bar{a}} G_{\bar{a}p} G_{pi}^{(a\bar{a})}}{G_{\bar{a}\bar{a}}}, \\ B_{115} &:= \frac{1}{N} \sum_p \frac{-\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} \frac{G_{ia}^{(\bar{a})} G_{ap}^{(\bar{a})} G_{pi}^{(a\bar{a})}}{G_{aa}^{(\bar{a})}}. \end{aligned} \quad (3.8.27)$$

We have changed summation index sets by removing the restrictions $p \neq a$. This is feasible because the difference term is of negligible order $O(\Psi_0^5)$. Notice that in B_{114} and B_{115} , the summation index p cannot be a since a appears as an upper index. However, we use the full summation silently and similarly in the following for notational simplicity.

Notice that in B_{111} , the index a has been decoupled from resolvent entries. Thus there is no need to further expand B_{111} . Next we turn to B_{114} , which is simpler than B_{112} . Using (3.3.31), we find

$$\begin{aligned} B_{114} &= \frac{1}{N} \sum_{p,r,s} \frac{-\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} G_{\bar{a}\bar{a}} G_{ir}^{(\bar{a})} y_{r\bar{a}} y_{s\bar{a}} G_{sp}^{(\bar{a})} G_{pi}^{(a\bar{a})} \\ &= \frac{1}{N} \sum_p \frac{-\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} G_{\bar{a}\bar{a}} G_{ia}^{(\bar{a})} d_a^2 G_{ap}^{(\bar{a})} G_{pi}^{(a\bar{a})} + \frac{1}{N} \sum_{p,r} \frac{-\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} G_{\bar{a}\bar{a}} G_{ir}^{(\bar{a})} x_{r\bar{a}} d_a G_{pa}^{(\bar{a})} G_{pi}^{(a\bar{a})} \\ &\quad + \frac{1}{N} \sum_{p,s} \frac{-\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} G_{\bar{a}\bar{a}} G_{ia}^{(\bar{a})} d_a x_{s\bar{a}} G_{ps}^{(\bar{a})} G_{pi}^{(a\bar{a})} + \frac{1}{N} \sum_{p,r,s} \frac{-\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} G_{\bar{a}\bar{a}} G_{ir}^{(\bar{a})} x_{r\bar{a}} x_{s\bar{a}} G_{ps}^{(\bar{a})} G_{pi}^{(a\bar{a})}. \end{aligned} \quad (3.8.28)$$

By (3.3.29), (3.8.8), (3.8.9) and (3.8.11) we find that the first term in the right side of (3.8.28) is

$$\begin{aligned}
 & \frac{1}{N} \sum_p \frac{-\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} G_{\bar{a}\bar{a}} G_{i\bar{a}}^{(\bar{a})} d_a^2 G_{a\bar{p}}^{(\bar{a})} G_{p\bar{i}}^{(a\bar{a})} \\
 &= \frac{1}{N} \sum_p \frac{-\gamma^3 d_a^4}{(\gamma d_a^2 - \xi)^2} G_{\bar{a}\bar{a}} G_{a\bar{a}}^{(\bar{a})2} \sum_{\mu, \nu}^{(\bar{a})} G_{i\mu}^{(a\bar{a})} x_{a\mu} x_{a\nu} G_{p\nu}^{(a\bar{a})} G_{p\bar{i}}^{(a\bar{a})} \\
 &= \frac{1}{N} \sum_{p, \mu, \nu}^{(\bar{a})} \frac{-\gamma^3 d_a^4}{(\gamma d_a^2 - \xi)^2} \left[\frac{1}{(\gamma d_a^2 - \xi) \tilde{b}} + \frac{(\gamma d_a^2 - 2\xi) \mathcal{U}_a}{(\gamma d_a^2 - \xi)^2 \tilde{b}^2} - \frac{\mathcal{V}_a}{(\gamma d_a^2 - \xi)^2} \right] G_{i\mu}^{(a\bar{a})} x_{a\mu} x_{a\nu} G_{p\nu}^{(a\bar{a})} G_{p\bar{i}}^{(a\bar{a})} + O(\Psi_0^5).
 \end{aligned} \tag{3.8.29}$$

Then we handle the three terms in the last step of (3.8.29) separately. For the first term, by taking expectation, we get

$$\begin{aligned}
 & \mathbb{E} \frac{1}{N} \sum_{p, \mu, \nu} \frac{-\gamma^3 d_a^4}{(\gamma d_a^2 - \xi)^3 \tilde{b}} G_{i\mu}^{(a\bar{a})} x_{a\mu} x_{a\nu} G_{p\nu}^{(a\bar{a})} G_{p\bar{i}}^{(a\bar{a})} \\
 &= \mathbb{E} \frac{1}{N^2} \sum_{p, \mu} \frac{-\gamma^3 d_a^4}{(\gamma d_a^2 - \xi)^3 \tilde{b}} G_{i\mu}^{(a\bar{a})} G_{\mu p}^{(a\bar{a})} G_{p\bar{i}}^{(a\bar{a})} \\
 &= \frac{-\gamma^3 d_a^4}{(\gamma d_a^2 - \xi)^3 \tilde{b}} \mathbb{E} X'_{33} + \frac{3\gamma^4 d_a^4 E_+ (E_+ b + \tilde{b})}{(\gamma d_a^2 - \xi)^4 \tilde{b}} \mathbb{E} X_{44} + O(\Psi_0^5).
 \end{aligned} \tag{3.8.30}$$

where in the last step, we use the second equation in Lemma 3.8.4. We find the second term in (3.8.29) equals

$$\begin{aligned}
 & \frac{1}{N} \sum_p \frac{-\gamma^3 d_a^4 (\gamma d_a^2 - 2\xi)}{(\gamma d_a^2 - \xi)^4 \tilde{b}^2} \left[E_+ - z + \gamma \left(E_+ \tilde{m} - \sum x_{a\alpha} G_{\alpha\beta}^{(a\bar{a})} G_{a\beta} \right) \right] \sum_{\mu, \nu}^{(\bar{a})} G_{i\mu}^{(a\bar{a})} G_{i\mu}^{(a\bar{a})} x_{a\mu} x_{a\nu} G_{p\nu}^{(a\bar{a})} \\
 &= \frac{1}{N} \sum_p \frac{-\gamma^3 d_a^4 (\gamma d_a^2 - 2\xi)}{(\gamma d_a^2 - \xi)^4 \tilde{b}^2} \gamma \left(E_+ \tilde{m} - \sum x_{a\alpha} G_{\alpha\beta}^{(a\bar{a})} G_{a\beta} \right) \sum_{\mu, \nu}^{(\bar{a})} G_{i\mu}^{(a\bar{a})} G_{i\mu}^{(a\bar{a})} x_{a\mu} x_{a\nu} G_{p\nu}^{(a\bar{a})} + O(\Psi_0^5).
 \end{aligned} \tag{3.8.31}$$

Using (3.4.36), (3.8.17) and (3.8.18), its expectation equals

$$\left(\frac{-\gamma^4 d_a^4 E_+^2}{(\gamma d_a^2 - \xi)^3 \tilde{b}^2} + \frac{\gamma^4 d_a^4 b E_+^2}{(\gamma d_a^2 - \xi)^4 \tilde{b}} \right) (\mathbb{E} X_{43} - 2\mathbb{E} X_{44}) + O(\Psi_0^5). \tag{3.8.32}$$

Here we address that we will use Lemma 3.8.3 frequently later to express the terms by using $\mathbb{E} X_{42}, \mathbb{E} X_{43}$ and $\mathbb{E} X_{44}$ instead of their variations defined in (3.4.33). The

last term in (3.8.29) equals

$$\frac{\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} \mathbb{E} X_{43} + O(\Psi_0^5) \quad (3.8.33)$$

Combining (3.8.30), (3.8.32) and (3.8.33), the expectation of the first term in (3.8.28) is

$$\begin{aligned} & \mathbb{E} \frac{1}{N} \sum_p \frac{-\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} G_{\bar{a}\bar{a}} G_{ia}^{(\bar{a})} d_a^2 G_{ap}^{(\bar{a})} G_{pi}^{(a\bar{a})} \\ &= \frac{-\gamma^3 d_a^4}{(\gamma d_a^2 - \xi)^3 \tilde{b}} \mathbb{E} X'_{33} + \left(\frac{-\gamma^4 d_a^4 E_+^2}{(\gamma d_a^2 - \xi)^3 \tilde{b}^2} + \frac{\gamma^4 d_a^4 b E_+^2}{(\gamma d_a^2 - \xi)^4 \tilde{b}} + \frac{\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E} X_{43} \\ & \quad + \left(\frac{\gamma^4 d_a^4 b E_+^2}{(\gamma d_a^2 - \xi)^4 \tilde{b}} + \frac{3\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} + \frac{2\gamma^4 d_a^4 E_+^2}{(\gamma d_a^2 - \xi)^3 \tilde{b}^2} \right) \mathbb{E} X_{44} + O(\Psi_0^5). \end{aligned} \quad (3.8.34)$$

For the second term in (3.8.28), expanding the $G_{\bar{a}\bar{a}}$ by using (3.8.8) and (3.8.9), we get

$$\begin{aligned} & \frac{1}{N} \sum_{p,r} \frac{-\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} G_{\bar{a}\bar{a}} G_{ir}^{(\bar{a})} x_{r\bar{a}} d_a G_{pa}^{(\bar{a})} G_{pi}^{(a\bar{a})} \\ &= \frac{1}{N} \sum_{p,r} \frac{-\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} \left[\frac{\tilde{b}}{\gamma d_a^2 - \xi} - \frac{\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} \mathcal{U}_a - \frac{\tilde{b}^2}{(\gamma d_a^2 - \xi)^2} \mathcal{V}_a \right] \\ & \quad \times G_{ir}^{(\bar{a})} x_{r\bar{a}} d_a G_{pa}^{(\bar{a})} G_{pi}^{(a\bar{a})} + O(\Psi_0^5). \end{aligned} \quad (3.8.35)$$

The expectations of the first and second term in (3.8.35) vanish by taking partial expectation $\mathbb{E}_{\bar{a}}$ first. For the third term above, after taking expectation, we obtain

$$\begin{aligned} & \mathbb{E} \frac{1}{N} \sum_{p,r} \frac{\gamma^3 d_a^2 \tilde{b}^2}{(\gamma d_a^2 - \xi)^4} \left(cm - \sum x_{j\bar{a}} G_{jk}^{(a\bar{a})} x_{k\bar{a}} - 2d_a \sum G_{aj}^{(\bar{a})} x_{j\bar{a}} \right) G_{ir}^{(\bar{a})} x_{r\bar{a}} d_a G_{pa}^{(\bar{a})} G_{pi}^{(a\bar{a})} \\ &= \frac{1}{N^2} \sum_{p,r} \frac{-2\gamma^3 d_a^4 \tilde{b}^2}{(\gamma d_a^2 - \xi)^4} \mathbb{E} G_{ir}^{(\bar{a})} G_{ra}^{(\bar{a})} G_{ap}^{(\bar{a})} G_{pi}^{(a\bar{a})}. \end{aligned} \quad (3.8.36)$$

By expanding Green function entries that have lower index a , we obtain that

$$\begin{aligned}
 & \frac{1}{N^2} \sum_{p,r} \frac{-2\gamma^3 d_a^2 \tilde{b}^2}{(\gamma d_a^4 - \xi)^4} \mathbb{E} G_{ir}^{(\bar{a})} G_{ra}^{(\bar{a})} G_{ap}^{(\bar{a})} G_{pi}^{(a\bar{a})} \\
 &= \frac{1}{N^2} \sum_{p,r,\mu,\nu} \frac{-2\gamma^4 d_a^4 \tilde{b}^2}{(\gamma d_a^2 - \xi)^4} \mathbb{E} G_{ir}^{(\bar{a})} G_{aa}^{(\bar{a})2} G_{r\mu}^{(a\bar{a})} x_{a\mu} x_{a\nu} G_{p\nu}^{(a\bar{a})} G_{pi}^{(a\bar{a})} \\
 &= \frac{1}{N^3} \sum_{p,r,\mu} \frac{-2\gamma^4 d_a^4}{(\gamma d_a^2 - \xi)^4} \mathbb{E} G_{ir}^{(a\bar{a})} G_{r\mu}^{(a\bar{a})} G_{\mu p}^{(a\bar{a})} G_{pi}^{(a\bar{a})} \\
 &= \frac{-2\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} \mathbb{E} X_{44} + O(\Psi_0^5).
 \end{aligned} \tag{3.8.37}$$

Therefore, we conclude that the expectation of the second term in (3.8.28) is

$$\frac{-2\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} \mathbb{E} X_{44} + O(\Psi_0^5). \tag{3.8.38}$$

For the third term in (3.8.28), the estimation is essentially the same as the argument from (3.8.35) to (3.8.37), and we obtains that

$$\mathbb{E} \frac{1}{N} \sum_{p,s} \frac{-\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} G_{\bar{a}\bar{a}} G_{ia}^{(\bar{a})} d_a x_{s\bar{a}} G_{ps}^{(\bar{a})} G_{pi}^{(a\bar{a})} = \frac{-2\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} \mathbb{E} X_{44} + O(\Psi_0^5). \tag{3.8.39}$$

For the last term in (3.8.28), we have

$$\begin{aligned}
 & \frac{1}{N} \sum_{p,r,s} \frac{-\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} G_{\bar{a}\bar{a}} G_{ir}^{(\bar{a})} x_{r\bar{a}} x_{s\bar{a}} G_{ps}^{(\bar{a})} G_{pi}^{(a\bar{a})} \\
 &= \frac{1}{N} \sum_{p,r,s} \frac{-\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} \left[\frac{\tilde{b}}{\gamma d_a^2 - \xi} - \frac{\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} \mathcal{U}_a - \frac{\tilde{b}^2}{(\gamma d_a^2 - \xi)^2} \mathcal{V}_a \right] \\
 & \quad \times G_{ir}^{(\bar{a})} x_{r\bar{a}} x_{s\bar{a}} G_{ps}^{(\bar{a})} G_{pi}^{(a\bar{a})} + O(\Psi_0^5).
 \end{aligned} \tag{3.8.40}$$

The expectation of the first term of the right side of (3.8.40) is

$$\begin{aligned}
 & \frac{1}{N^2} \sum_{p,r} \frac{-\gamma^2 d_a^2 \tilde{b}}{(\gamma d_a^2 - \xi)^3} \mathbb{E} G_{ir}^{(\bar{a})} G_{rp}^{(\bar{a})} G_{pi}^{(a\bar{a})} \\
 &= \frac{1}{N^2} \sum_{p,r} \frac{-\gamma^2 d_a^2 \tilde{b}}{(\gamma d_a^2 - \xi)^3} \mathbb{E} G_{ir}^{(\bar{a})} G_{rp}^{(\bar{a})} G_{pi}^{(\bar{a})} + \frac{1}{N} \sum_{p,r} \frac{\gamma^2 d_a^2 \tilde{b}}{(\gamma d_a^2 - \xi)^3} \mathbb{E} G_{ir}^{(\bar{a})} G_{rp}^{(\bar{a})} \frac{G_{pa}^{(\bar{a})} G_{ai}^{(\bar{a})}}{G_{aa}^{(\bar{a})}} \\
 &= \frac{-\gamma^2 d_a^2 \tilde{b}}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X_{33} + 3 \left(\frac{\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} + \frac{\gamma^3 d_a^2 \tilde{b}^2}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E} X_{44} - \frac{\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X_{44} + O(\Psi_0^5).
 \end{aligned} \tag{3.8.41}$$

where in the second equality, we use the fact that

$$\mathbb{E} \frac{1}{N^2} \sum_{j,k} G_{ij}^{(\bar{a})} G_{jk}^{(\bar{a})} G_{ki}^{(\bar{a})} = \mathbb{E} X_{33} - 3 \left(\frac{\gamma^2 d_a^2 E_+}{(\gamma d_a^2 - \xi) \tilde{b}} + \frac{\gamma \tilde{b}}{\gamma d_a^2 - \xi} \right) \mathbb{E} X_{44} + O(\Psi_0^5). \quad (3.8.42)$$

To derive (3.8.42), we write

$$\begin{aligned} \mathbb{E} \frac{1}{N^2} \sum_{j,k} G_{ij}^{(\bar{a})} G_{jk}^{(\bar{a})} G_{ki}^{(\bar{a})} &= \mathbb{E} \frac{1}{N^2} \sum_{j,k} \left(G_{ij} - \frac{G_{i\bar{a}} G_{\bar{a}j}}{G_{\bar{a}\bar{a}}} \right) G_{jk} G_{ki} \\ &\quad - \mathbb{E} \frac{1}{N^2} \sum_{j,k} G_{ij}^{(\bar{a})} \frac{G_{j\bar{a}} G_{\bar{a}k}}{G_{\bar{a}\bar{a}}} G_{ki} - \mathbb{E} \frac{1}{N^2} \sum_{j,k} G_{ij}^{(\bar{a})} G_{jk}^{(\bar{a})} \frac{G_{ka} G_{\bar{a}i}}{G_{\bar{a}\bar{a}}}. \end{aligned} \quad (3.8.43)$$

Similar to (3.8.106) below, we calculate

$$\mathbb{E} \frac{1}{N^2} \sum_{j,k} \frac{G_{i\bar{a}} G_{\bar{a}j}}{G_{\bar{a}\bar{a}}} G_{jk} G_{ki} = \frac{\gamma^2 d_a^2 E_+}{(\gamma d_a^2 - \xi) \tilde{b}} \mathbb{E} X_{44} + \frac{\gamma \tilde{b}}{\gamma d_a^2 - \xi} \mathbb{E} X_{44} + O(\Psi_0^5). \quad (3.8.44)$$

The same estimation also hold for the second and third term in (3.8.43). Therefore, (3.8.42) holds true.

The expectation of the second term of the right side of (3.8.40) is

$$\mathbb{E} \frac{1}{N} \sum_{p,r,s} \frac{\gamma^3 d_a^4}{(\gamma d_a^2 - \xi)^4} \mathcal{U}_a G_{ir}^{(\bar{a})} x_{r\bar{a}} x_{s\bar{a}} G_{ps}^{(\bar{a})} G_{pi}^{(a\bar{a})} = \frac{\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} \mathbb{E} X_{43} + O(\Psi_0^5). \quad (3.8.45)$$

And the expectation of the third term of the right side of (3.8.40) is

$$\begin{aligned} &\mathbb{E} \frac{1}{N} \sum_{p,r,s} \frac{\gamma^2 d_a^2 \tilde{b}^2}{(\gamma d_a^2 - \xi)^4} \mathcal{V}_a G_{ir}^{(\bar{a})} x_{r\bar{a}} x_{s\bar{a}} G_{ps}^{(\bar{a})} G_{pi}^{(a\bar{a})} \\ &= -\frac{2\gamma^3 d_a^2 \tilde{b}^2}{(\gamma d_a^2 - \xi)^4} \mathbb{E} X_{44} + \frac{\gamma^3 d_a^2 \tilde{b}^2}{(\gamma d_a^2 - \xi)^4} \mathbb{E} X_{43} + O(\Psi_0^5). \end{aligned} \quad (3.8.46)$$

Then, combining (3.8.41), (3.8.45) and (3.8.46) together, we have that the expectation of the last term of (3.8.28) is

$$\begin{aligned} & \mathbb{E} \frac{1}{N} \sum_{p,r,s} \frac{-\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} G_{\bar{a}\bar{a}} G_{ir}^{(\bar{a})} x_{r\bar{a}} x_{s\bar{a}} G_{ps}^{(\bar{a})} G_{pi}^{(a\bar{a})} \\ &= \frac{-\gamma^2 d_a^2 \tilde{b}}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X_{33} + \left(\frac{3\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} + \frac{\gamma^3 d_a^2 \tilde{b}^2}{(\gamma d_a^2 - \xi)^4} - \frac{\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E} X_{44} \\ & \quad + \left(\frac{\gamma^3 d_a^2 \tilde{b}^2}{(\gamma d_a^2 - \xi)^4} + \frac{\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E} X_{43} + O(\Psi_0^5). \end{aligned} \quad (3.8.47)$$

Therefore, combining (3.8.34), (3.8.38), (3.8.39) and (3.8.47), we have

$$\begin{aligned} \mathbb{E} B_{114} &= \frac{-\gamma^2 d_a^2 \tilde{b}}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X_{33} + \frac{-\gamma^3 d_a^4}{(\gamma d_a^2 - \xi)^3 \tilde{b}} \mathbb{E} X'_{33} \\ & \quad + \left(\frac{-\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^3 \tilde{b}^2} + \frac{\gamma^4 d_a^4 b E_+}{(\gamma d_a^2 - \xi)^4 \tilde{b}} + \frac{2\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} + \frac{\gamma^3 d_a^2 \tilde{b}^2}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E} X_{43} \\ & \quad + \left(\frac{\gamma^4 d_a^4 b E_+}{(\gamma d_a^2 - \xi)^4 \tilde{b}} + \frac{2\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} + \frac{2\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^3 \tilde{b}^2} + \frac{\gamma^3 d_a^2 \tilde{b}^2}{(\gamma d_a^2 - \xi)^4} - \frac{\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E} X_{44} + O(\Psi_0^5). \end{aligned} \quad (3.8.48)$$

Then we consider B_{112} in (3.8.27). Using (3.3.32), we find

$$\mathbb{E} B_{112} = \mathbb{E} B_{114} + \frac{1}{N} \sum_p \frac{-\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} \frac{G_{i\bar{a}} G_{\bar{a}p} G_{p\bar{a}} G_{\bar{a}i}}{G_{\bar{a}\bar{a}}^2} + \frac{1}{N} \sum_p \frac{-\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} \frac{G_{i\bar{a}} G_{\bar{a}p} G_{p\bar{a}}^{(\bar{a})} G_{\bar{a}i}^{(\bar{a})}}{G_{\bar{a}\bar{a}} G_{\bar{a}\bar{a}}^{(\bar{a})}} \quad (3.8.49)$$

To calculate the expectation of the second term on the right side of (3.8.49), we consider

$$\begin{aligned} & \mathbb{E} \frac{1}{N} \sum_p \frac{G_{i\bar{a}} G_{\bar{a}p} G_{p\bar{a}} G_{\bar{a}i}}{G_{\bar{a}\bar{a}}^2} \\ &= \mathbb{E} \frac{1}{N} \sum_p G_{\bar{a}\bar{a}}^2 \gamma^2 \left(G_{i\bar{a}}^{(\bar{a})} d_a^2 G_{p\bar{a}}^{(\bar{a})} + \sum_r G_{ir}^{(\bar{a})} x_{r\bar{a}} d_a G_{p\bar{a}}^{(\bar{a})} + \sum_s G_{is}^{(\bar{a})} d_a x_{s\bar{a}} G_{p\bar{a}}^{(\bar{a})} + \sum_{r,s} G_{ir}^{(\bar{a})} x_{r\bar{a}} x_{s\bar{a}} G_{ps}^{(\bar{a})} \right)^2 \end{aligned} \quad (3.8.50)$$

By expanding the square, we find that

$$\begin{aligned}
\mathbb{E} \frac{\gamma^2}{N} \sum_p G_{\bar{a}\bar{a}}^2 G_{ia}^{(\bar{a})} d_a^2 G_{pa}^{(\bar{a})} G_{ia}^{(\bar{a})} d_a^2 G_{pa}^{(\bar{a})} &= \frac{\gamma^4 d_a^4 E_+^2}{(\gamma d_a^2 - \xi)^2 \tilde{b}^2} (2\mathbb{E}X_{44} + \mathbb{E}\tilde{X}_{44}), \\
\mathbb{E} \frac{\gamma^2}{N} \sum_p G_{\bar{a}\bar{a}}^2 \sum_{r,s} G_{ir}^{(\bar{a})} x_{r\bar{a}} d_a G_{pa}^{(\bar{a})} G_{is}^{(\bar{a})} x_{s\bar{a}} d_a G_{pa}^{(\bar{a})} &= \frac{\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^2} \mathbb{E}\tilde{X}_{44}, \\
\mathbb{E} \frac{\gamma^2}{N} \sum_p G_{\bar{a}\bar{a}}^2 \sum_{r,s} G_{ia}^{(\bar{a})} d_a x_{r\bar{a}} G_{pr}^{(\bar{a})} G_{ia}^{(\bar{a})} d_a x_{s\bar{a}} G_{ps}^{(\bar{a})} &= \frac{\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^2} \mathbb{E}\tilde{X}_{44}, \\
\mathbb{E} \frac{\gamma^2}{N} \sum_p G_{\bar{a}\bar{a}}^2 \sum_{r,s,j,k} G_{ir}^{(\bar{a})} x_{r\bar{a}} x_{s\bar{a}} G_{ps}^{(\bar{a})} G_{ij}^{(\bar{a})} x_{j\bar{a}} x_{k\bar{a}} G_{pk}^{(\bar{a})} &= \frac{\gamma^2 \tilde{b}^2}{(\gamma d_a^2 - \xi)^2} (2\mathbb{E}X_{44} + \mathbb{E}\tilde{X}_{44}), \\
\mathbb{E} \frac{2\gamma^2}{N} \sum_p G_{\bar{a}\bar{a}}^2 G_{ia}^{(\bar{a})} d_a^2 G_{pa}^{(\bar{a})} \sum_{r,s} G_{ir}^{(\bar{a})} x_{r\bar{a}} x_{s\bar{a}} G_{ps}^{(\bar{a})} &= \frac{2\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^2} \mathbb{E}X_{44} \\
\mathbb{E} \frac{2\gamma^2}{N} \sum_p G_{\bar{a}\bar{a}}^2 \sum_{r,s} G_{ir}^{(\bar{a})} x_{r\bar{a}} d_a G_{pa}^{(\bar{a})} G_{ia}^{(\bar{a})} d_a x_{s\bar{a}} G_{ps}^{(\bar{a})} &= \frac{2\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^2} \mathbb{E}X_{44}.
\end{aligned} \tag{3.8.51}$$

And for the other terms that contain one or three entries of X have negligible expectation. It follows that

$$\begin{aligned}
\mathbb{E} \frac{1}{N} \sum_p \frac{-\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} \frac{G_{i\bar{a}} G_{\bar{a}p} G_{p\bar{a}} G_{\bar{a}i}}{G_{\bar{a}\bar{a}}^2} \\
= \left(\frac{-\gamma^5 d_a^4 E_+^2}{(\gamma d_a^2 - \xi)^4 \tilde{b}^2} + \frac{-\gamma^3 d_a^2 \tilde{b}^2}{(\gamma d_a^2 - \xi)^4} + \frac{-2\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} \right) (2\mathbb{E}X_{44} + \mathbb{E}\tilde{X}_{44}) + O(\Psi_0^5).
\end{aligned} \tag{3.8.52}$$

The expectation of the last term in (3.8.49) equals

$$\begin{aligned}
\mathbb{E} \frac{1}{N} \sum_p \frac{-\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} \frac{G_{i\bar{a}} G_{\bar{a}p} G_{pa}^{(\bar{a})} G_{ai}^{(\bar{a})}}{G_{\bar{a}\bar{a}} G_{aa}^{(\bar{a})}} \\
= \mathbb{E} \frac{1}{N} \sum_p \frac{-\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} G_{\bar{a}\bar{a}} G_{aa}^{(\bar{a})} G_{ir}^{(\bar{a})} y_{r\bar{a}} y_{s\bar{a}} G_{ps}^{(\bar{a})} G_{p\mu}^{(a\bar{a})} x_{a\mu} x_{a\nu} G_{\nu i}^{(a\bar{a})} \\
= \left(\frac{2\gamma^4 d_a^4 E_+^2}{(\gamma d_a^2 - \xi)^3 \tilde{b}^2} + \frac{\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E}X_{44} + \frac{\gamma^4 d_a^4 E_+^2}{(\gamma d_a^2 - \xi)^3 \tilde{b}^2} \mathbb{E}\tilde{X}_{44} + O(\Psi_0^5).
\end{aligned} \tag{3.8.53}$$

Therefore we have

$$\begin{aligned}
\mathbb{E}B_{112} = \mathbb{E}B_{114} + \left(\frac{-2\gamma^4 d_a^4 b E_+^2}{(\gamma d_a^2 - \xi)^4 \tilde{b}} + \frac{-2\gamma^3 d_a^2 \tilde{b}^2}{(\gamma d_a^2 - \xi)^4} + \frac{-4\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} + \frac{\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E}X_{44} \\
+ \left(\frac{-\gamma^4 d_a^4 b E_+^2}{(\gamma d_a^2 - \xi)^4 \tilde{b}} + \frac{-\gamma^3 d_a^2 \tilde{b}^2}{(\gamma d_a^2 - \xi)^4} + \frac{-2\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E}\tilde{X}_{44} + O(\Psi_0^5).
\end{aligned} \tag{3.8.54}$$

We now turn to B_{115} in (3.8.27), which is similar to the first term in (3.8.28). Write

$$B_{115} = \frac{1}{N} \sum_p \frac{-\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} G_{aa}^{(\bar{a})} \sum_{\mu, \nu}^{(\bar{a})} G_{i\mu}^{(a\bar{a})} x_{a\mu} x_{a\nu} G_{p\nu}^{(a\bar{a})} G_{pi}^{(a\bar{a})}. \quad (3.8.55)$$

Using the second expansion in (3.8.8), in a similar manner to the expansion of the first term in (3.8.28), we have

$$\begin{aligned} \mathbb{E}B_{115} &= \frac{1}{N} \sum_p \frac{-\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} \left(-\frac{1}{\tilde{b}} - \frac{1}{\tilde{b}^2} \mathcal{U}_a \right) \sum_{\mu, \nu}^{(\bar{a})} G_{i\mu}^{(a\bar{a})} x_{a\mu} x_{a\nu} G_{p\nu}^{(a\bar{a})} G_{pi}^{(a\bar{a})} + O(\Psi_0^5) \\ &= \frac{\gamma d_a^2}{(\gamma d_a^2 - \xi)^2 \tilde{b}} \mathbb{E}X'_{33} - \frac{3\gamma^3 d_a^2 E_+ (E_+ b + \tilde{b})}{(\gamma d_a^2 - \xi)^3 \tilde{b}} \mathbb{E}X_{44} - \frac{2\gamma^3 d_a^2 E_+^2}{(\gamma d_a^2 - \xi)^2 \tilde{b}^2} \mathbb{E}X_{44} \\ &\quad + \frac{\gamma^3 d_a^2 E_+^2}{(\gamma d_a^2 - \xi)^2 \tilde{b}^2} \mathbb{E}X_{43} + O(\Psi_0^5). \end{aligned} \quad (3.8.56)$$

Next, we consider B_{113} . Notice that

$$B_{113} = B_{115} + \frac{1}{N} \sum \frac{-\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} \frac{G_{ia}^{(\bar{a})} G_{ap}^{(\bar{a})} G_{p\bar{a}}^{(\bar{a})} G_{\bar{a}i}^{(\bar{a})}}{G_{aa}^{(\bar{a})} G_{\bar{a}\bar{a}}^{(\bar{a})}} + \frac{1}{N} \sum \frac{-\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} \frac{G_{ia}^{(\bar{a})} G_{ap}^{(\bar{a})} G_{pa}^{(\bar{a})} G_{ai}^{(\bar{a})}}{G_{aa}^{(\bar{a})^2}}. \quad (3.8.57)$$

The expansion of the second term of (3.8.57) was already done in (3.8.53). For the third term, we have

$$\begin{aligned} &\mathbb{E} \frac{1}{N} \sum \frac{-\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} \frac{G_{ia}^{(\bar{a})} G_{ap}^{(\bar{a})} G_{pa}^{(\bar{a})} G_{ai}^{(\bar{a})}}{G_{aa}^{(\bar{a})^2}} \\ &= \mathbb{E} \frac{1}{N} \sum \frac{-\gamma^3 d_a^2}{(\gamma d_a^2 - \xi)^2} G_{aa}^{(\bar{a})^2} G_{i\alpha}^{(a\bar{a})} x_{a\alpha} x_{a\beta} G_{p\beta}^{(a\bar{a})} G_{p\mu}^{(a\bar{a})} x_{a\mu} x_{x\nu} G_{i\nu}^{(a\bar{a})} \\ &= \frac{-\gamma^3 d_a^2 E_+^2}{(\gamma d_a^2 - \xi)^2 \tilde{b}^2} (2\mathbb{E}X_{44} + \mathbb{E}\tilde{X}_{44}) + O(\Psi_0^5). \end{aligned} \quad (3.8.58)$$

Therefore, we get

$$\mathbb{E}B_{113} = \mathbb{E}B_{115} + \left(\frac{2\gamma^3 d_a^2 b E_+^2}{(\gamma d_a^2 - \xi)^3 \tilde{b}} + \frac{\gamma^3 d_a^2 E_+^2}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E}X_{44} + \frac{\gamma^3 d_a^2 b E_+^2}{(\gamma d_a^2 - \xi)^3 \tilde{b}} \mathbb{E}\tilde{X}_{44} + O(\Psi_0^5). \quad (3.8.59)$$

Combining (3.8.48), (3.8.54), (3.8.56), (3.8.59), we conclude

$$\begin{aligned} \mathbb{E}B_{11} &= \frac{\gamma d_a^2}{(\gamma d_a^2 - \xi)^2} \mathbb{E}X_{22} + \frac{-2\gamma^2 d_a^2 \tilde{b}}{(\gamma d_a^2 - \xi)^3} \mathbb{E}X_{33} + \frac{-2\gamma^2 d_a^2 b}{(\gamma d_a^2 - \xi)^3} \mathbb{E}X'_{33} \\ &+ \left(\frac{4\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} + \frac{2\gamma^3 d_a^2 b^2 E_+^2}{(\gamma d_a^2 - \xi)^4} + \frac{2\gamma^3 d_a^2 \tilde{b}^2}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E}X_{43} + \frac{-6\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \mathbb{E}X_{44} \\ &+ \left(\frac{-2\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} + \frac{-\gamma^3 d_a^2 b^2 E_+^2}{(\gamma d_a^2 - \xi)^4} + \frac{-\gamma^3 d_a^2 \tilde{b}^2}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E}\tilde{X}_{44} + O(\Psi_0^5). \end{aligned} \quad (3.8.60)$$

Expansions of $B_{12}, B_{13}, \dots, B_{16}$. For B_{12} in (3.8.24), we have

$$\begin{aligned} \mathbb{E}B_{12} &= \mathbb{E} \sum_{p,q} \frac{-2\gamma^2 \tilde{b} d_a^2}{(\gamma d_a^2 - \xi)^3} (cm - \sum x_{j\bar{a}} G_{jk}^{(a\bar{a})} x_{k\bar{a}}) G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})} \\ &= \mathbb{E} \frac{1}{N} \sum_{p,q} \frac{-2\gamma^2 \tilde{b} d_a^2}{(\gamma d_a^2 - \xi)^3} \left(cm - \frac{1}{N} \sum G_{jj}^{(a\bar{a})} \right) G_{ip}^{(a\bar{a})} G_{pi}^{(a\bar{a})} \\ &\quad + \mathbb{E} \frac{1}{N^2} \sum_{p,q} \frac{4\gamma^2 \tilde{b} d_a^2}{(\gamma d_a^2 - \xi)^3} G_{ip}^{(a\bar{a})} G_{pq}^{(a\bar{a})} G_{qi}^{(a\bar{a})} \\ &= \frac{-2\gamma^2 \tilde{b} d_a^2}{(\gamma d_a^2 - \xi)^3} \mathbb{E}X_{32} + \frac{-2\gamma^3 \tilde{b} d_a^2 (E_+ b + \tilde{b})}{(\gamma d_a^2 - \xi)^4} (\mathbb{E}\tilde{X}_{44} - 2\mathbb{E}X_{43}) \\ &\quad + \frac{4\gamma^2 \tilde{b} d_a^2}{(\gamma d_a^2 - \xi)^3} \mathbb{E}X_{33} - \frac{12\gamma^3 \tilde{b} d_a^2 (E_+ b + \tilde{b})}{(\gamma d_a^2 - \xi)^4} \mathbb{E}X_{44} + O(\Psi_0^5), \end{aligned} \quad (3.8.61)$$

where in the last step, we use Lemma 3.8.4.

For B_{13} in (3.8.24), write

$$\begin{aligned} B_{13} &= \frac{-2\gamma b d_a^2}{(\gamma d_a^2 - \xi)^3} (E_+ - z) \sum_{p,q} G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})} \\ &+ \frac{-2\gamma b d_a^2}{(\gamma d_a^2 - \xi)^3} \left(E_+ \tilde{m} - \sum_{\mu,\nu} x_{a\mu} G_{\mu\nu}^{(a)} x_{a\nu} - 2d_a \sum_{\mu} G_{\bar{a}\mu}^{(a)} x_{a\mu} \right) \sum_{p,q} G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})}. \end{aligned} \quad (3.8.62)$$

The expectation of the first term is

$$\frac{-2\gamma b d_a^2}{(\gamma d_a^2 - \xi)^3} (E_+ - z) \mathbb{E}X_{22} + O(\Psi_0^5). \quad (3.8.63)$$

For the second term, we have

$$\begin{aligned}
 & \mathbb{E} \frac{-2\gamma b d_a^2}{(\gamma d_a^2 - \xi)^3} \left(E_+ \tilde{m} - \sum x_{a\mu} G_{\mu\nu}^{(a)} x_{a\nu} - 2d_a \sum G_{\bar{a}\mu}^{(a)} x_{a\mu} \right) \sum_{p,q} G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})} \\
 &= \mathbb{E} \frac{-2\gamma^2 b d_a^2}{(\gamma d_a^2 - \xi)^3} \left(E_+ \tilde{m} - \frac{1}{N} \sum G_{\mu\mu}^{(a)} \right) \sum_{p,q} G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})} \\
 &= \mathbb{E} \frac{-2\gamma^2 b d_a^2}{(\gamma d_a^2 - \xi)^3} \left(E_+ \tilde{m} - \frac{1}{N} \sum G_{\mu\mu}^{(a\bar{a})} \right) \sum_{p,q} G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})} \\
 &\quad + \mathbb{E} \frac{2\gamma^2 b d_a^2}{(\gamma d_a^2 - \xi)^3} \frac{G_{\mu\bar{a}}^{(a)} G_{\bar{a}\mu}^{(a)}}{G_{\bar{a}\bar{a}}^{(a)}} \sum_{p,q} G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})} \\
 &= \frac{-2\gamma^2 b d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X_{32} + \frac{-2\gamma^3 b d_a^2 E_+ (E_+ b + \tilde{b})}{(\gamma d_a^2 - \xi)^4} (\mathbb{E} \tilde{X}_{44} - 2\mathbb{E} X_{43}) \\
 &\quad + \frac{-2\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} (2\mathbb{E} X_{44} + \mathbb{E} \tilde{X}_{44}) + O(\Psi_0^5).
 \end{aligned} \tag{3.8.64}$$

Combining (3.8.63) and (3.8.64), we have

$$\begin{aligned}
 \mathbb{E} B_{13} &= \frac{-2\gamma^2 b d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X_{32} + \frac{-2\gamma b d_a^2}{(\gamma d_a^2 - \xi)^3} (E_+ - z) \mathbb{E} X_{22} + \frac{4\gamma^3 b d_a^2 E_+ (E_+ b + \tilde{b})}{(\gamma d_a^2 - \xi)^4} \mathbb{E} X_{43} \\
 &\quad + \frac{-4\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X_{44} + \left(\frac{-2\gamma^3 d_a^2 b^2 E_+^2}{(\gamma d_a^2 - \xi)^3} + \frac{-2\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E} \tilde{X}_{44} + O(\Psi_0^5)
 \end{aligned} \tag{3.8.65}$$

For B_{14} in (3.8.24), we have

$$\begin{aligned}
 \mathbb{E} B_{14} &= \mathbb{E} \sum_{p,q}^{(a)} \frac{3\gamma^3 \tilde{b}^2 d_a^2}{(\gamma d_a^2 - \xi)^4} \left(cm(E_+) - \sum_{j,k}^{(a)} x_{j\bar{a}} G_{jk}^{(a\bar{a})} x_{k\bar{a}} \right)^2 G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})} \\
 &= \frac{3\gamma^3 \tilde{b}^2 d_a^2}{(\gamma d_a^2 - \xi)^4} \left(\mathbb{E} X_{42} - 4\mathbb{E} X_{43} + 8\mathbb{E} X_{44} + 2\mathbb{E} \tilde{X}_{44} \right) + O(\Psi_0^5).
 \end{aligned} \tag{3.8.66}$$

For B_{15} in (3.8.24), the expectation equals

$$\begin{aligned}
 \mathbb{E} B_{15} &= \mathbb{E} \sum_{p,q}^{(a)} \frac{(2\gamma^2 d_a^4 + 4\gamma \xi d_a^2) \mathcal{Y}_a \mathcal{Z}_a}{(\gamma d_a^2 - \xi)^4} G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})} \\
 &= \frac{(2\gamma^4 d_a^4 E_+ + 4\gamma^3 \xi d_a^2 E_+)}{(\gamma d_a^2 - \xi)^4} \mathbb{E} X_{42} - \frac{(4\gamma^4 d_a^4 E_+ + 8\gamma^3 \xi d_a^2 E_+)}{(\gamma d_a^2 - \xi)^4} \mathbb{E} X_{43} + O(\Psi_0^5) \\
 &= \left(\frac{6\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} - \frac{4\gamma^3 \xi d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E} X_{42} - \left(\frac{12\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} - \frac{8\gamma^3 \xi d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E} X_{43} + O(\Psi_0^5).
 \end{aligned} \tag{3.8.67}$$

For B_{16} in (3.8.24), the expectation is

$$\begin{aligned}
\mathbb{E}B_{16} &= \mathbb{E} \sum_{p,q}^{(a)} \frac{3\gamma b^2 d_a^2 \mathcal{Z}_a^2}{(\gamma d_a^2 - \xi)^4} G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})} \\
&= \mathbb{E} \sum_{p,q}^{(a)} \frac{3\gamma^3 b^2 d_a^2}{(\gamma d_a^2 - \xi)^4} \left(E_+ \tilde{m}(E_+) - \sum_{\mu,\nu} x_{a\mu} G_{\mu\nu}^{(a)} x_{a\nu} - 2d_a \sum G_{\bar{a}\mu}^{(a)} x_{a\mu} \right)^2 G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})} + O(\Psi_0^5) \\
&= \mathbb{E} \sum_{p,q}^{(a)} \frac{3\gamma^3 b^2 d_a^2}{(\gamma d_a^2 - \xi)^4} \left(E_+ \tilde{m}(E_+) - \sum_{\mu,\nu} x_{a\mu} G_{\mu\nu}^{(a)} x_{a\nu} \right)^2 G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})} \\
&\quad + \mathbb{E} \sum_{p,q}^{(a)} \frac{3\gamma^3 b^2 d_a^2 \mathcal{Z}_a^2}{(\gamma d_a^2 - \xi)^4} \left(2d_a \sum G_{\bar{a}\mu}^{(a)} x_{a\mu} \right)^2 G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{iq}^{(a\bar{a})} + O(\Psi_0^5) \\
&= \frac{3\gamma^3 b^2 d_a^2 E_+^2}{(\gamma d_a^2 - \xi)^4} \mathbb{E}X_{42} + \frac{24\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} \mathbb{E}X_{44} + \left(\frac{6\gamma^3 b^2 d_a^2 E_+^2}{(\gamma d_a^2 - \xi)^4} + \frac{12\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E}\tilde{X}_{44} + O(\Psi_0^5).
\end{aligned}$$

Up to now we finish the expansion of B_1 .

3.8.2.2 Expansion of B_2

We turn to estimate B_2 in (3.8.22). Using (3.8.5), write

$$\begin{aligned}
&\sum_{\alpha} G_{aa}^2 G_{i\alpha}^{(a)} x_{a\alpha} d_a G_{i\bar{a}}^{(a)} \\
&= \sum_{\alpha} (\beta_a^2 - 2\beta_a^3 \mathcal{Z}_a + 3\beta_a^4 \mathcal{Z}_a^2) G_{i\alpha}^{(a)} x_{a\alpha} d_a G_{i\bar{a}}^{(a)} + O(\Psi_0^5).
\end{aligned} \tag{3.8.68}$$

After taking expectation, the first term vanishes. Using (3.8.6), the expectation of the second term in (3.8.68) equals

$$\begin{aligned}
 & \mathbb{E} \sum_{\mu} 4\beta_a^3 \gamma d_a^2 G_{i\mu}^{(a)} G_{\mu\bar{a}}^{(a)} G_{\bar{a}i}^{(a)} - \mathbb{E} \sum_{\alpha} 2\beta_a^3 \gamma (E_+ \tilde{m} - \sum x_{a\mu} G_{\mu\nu}^{(a)} x_{a\nu}) G_{i\alpha}^{(a)} x_{a\alpha} d_a G_{i\bar{a}}^{(a)} \\
 &= \mathbb{E} \sum_{\mu} \frac{4b^3 \gamma d_a^2}{(\gamma d_a^2 - \xi)^3} G_{i\mu}^{(a)} G_{\mu\bar{a}}^{(a)} G_{\bar{a}i}^{(a)} + \mathbb{E} \sum_{\mu} \frac{-12\gamma^3 d_a^4 b^2}{(\gamma d_a^2 - \xi)^4} (cm - \sum x_{j\bar{a}} G_{jk}^{(a\bar{a})} x_{k\bar{a}}) G_{i\mu}^{(a)} G_{\mu\bar{a}}^{(a)} G_{\bar{a}i}^{(a)} + O(\Psi_0^5) \\
 &= \mathbb{E} \sum_{\mu} \frac{4\gamma^2 b^3 d_a^2}{(\gamma d_a^2 - \xi)^3} G_{i\mu}^{(a)} G_{\bar{a}\bar{a}}^{(a)^2} \sum G_{\mu r}^{(a\bar{a})} x_{r\bar{a}} x_{s\bar{a}} G_{si}^{(a\bar{a})} \\
 &\quad + \mathbb{E} \sum_{\mu} \frac{-12\gamma^4 d_a^4 b^2}{(\gamma d_a^2 - \xi)^4} (cm - \sum x_{j\bar{a}} G_{jk}^{(a\bar{a})} x_{k\bar{a}}) G_{i\mu}^{(a)} G_{\bar{a}\bar{a}}^{(a)^2} \sum G_{\mu r}^{(a\bar{a})} x_{r\bar{a}} x_{s\bar{a}} G_{si}^{(a\bar{a})} + O(\Psi_0^5) \\
 &= \mathbb{E} \sum_{\mu} \frac{4\gamma^2 b^3 d_a^2}{(\gamma d_a^2 - \xi)^3} \left[G_{i\mu}^{(a\bar{a})} \left(\frac{1}{b^2} + \frac{2\mathcal{Y}_a}{b^3} \right) + \frac{G_{i\bar{a}}^{(a)} G_{\bar{a}\mu}^{(a)}}{G_{\bar{a}\bar{a}}^{(a)}} \frac{1}{b^2} \right] \sum G_{\mu r}^{(a\bar{a})} x_{r\bar{a}} x_{s\bar{a}} G_{si}^{(a\bar{a})} + \\
 &\quad + \mathbb{E} \sum_{\mu} \frac{-12\gamma^4 d_a^4}{(\gamma d_a^2 - \xi)^4} (cm - \sum x_{j\bar{a}} G_{jk}^{(a\bar{a})} x_{k\bar{a}}) G_{i\mu}^{(a\bar{a})} \sum G_{\mu r}^{(a\bar{a})} x_{r\bar{a}} x_{s\bar{a}} G_{si}^{(a\bar{a})} + O(\Psi_0^5) \\
 &= \frac{4\gamma^2 b^3 d_a^2}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X'_{33} + \frac{-12\gamma^3 d_a^2 b E_+ (E_+ b + \tilde{b})}{(\gamma d_a^2 - \xi)^4} \mathbb{E} X_{44} + \frac{8\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X_{43} + \frac{-16\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X_{44} \\
 &\quad + \frac{-4\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} (2\mathbb{E} X_{44} + \mathbb{E} \tilde{X}_{44}) + \frac{-12\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} (\mathbb{E} X_{43} - 2\mathbb{E} X_{44}) + O(\Psi_0^5).
 \end{aligned}$$

For the third term in (3.8.68), after expanding square of $\mathcal{Z}_a$ and taking expectation, we find that all the terms with odd number of X entries have negligible expectations, and therefore

$$\begin{aligned}
 & \mathbb{E} \sum_{\alpha} 3\beta_a^4 \mathcal{Z}_a^2 G_{i\alpha}^{(a)} x_{a\alpha} d_a G_{i\bar{a}}^{(a)} \\
 &= \mathbb{E} \sum_{\alpha} \frac{-12\gamma^2 d_a^2 b^4}{(\gamma d_a^2 - \xi)^4} (E_+ \tilde{m} - \frac{1}{N} \sum G_{\mu\mu}^{(a)}) G_{i\alpha}^{(a)} G_{\alpha\bar{a}}^{(a)} G_{\bar{a}i}^{(a)} + \mathbb{E} \sum_{\alpha} \frac{24\gamma^2 d_a^2 b^4}{(\gamma d_a^2 - \xi)^4} G_{i\alpha}^{(a)} G_{\alpha\beta}^{(a)} G_{\beta\bar{a}}^{(a)} G_{\bar{a}i}^{(a)} \\
 &= \frac{-12\gamma^3 d_a^2 b^2 E_+^2}{(\gamma d_a^2 - \xi)^4} \mathbb{E} X_{43} + \frac{24\gamma^3 d_a^2 b^2 E_+^2}{(\gamma d_a^2 - \xi)^4} \mathbb{E} X_{44} + O(\Psi_0^5).
 \end{aligned}$$

Then, we have

$$\begin{aligned}
 \mathbb{E} B_2 &= \frac{4\gamma^2 b^3 d_a^2}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X'_{33} + \left(\frac{-12\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} + \frac{-12\gamma^3 d_a^2 b^2 E_+^2}{(\gamma d_a^2 - \xi)^4} + \frac{8\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E} X_{43} \\
 &\quad + \left(\frac{12\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} + \frac{12\gamma^3 d_a^2 b^2 E_+^2}{(\gamma d_a^2 - \xi)^4} + \frac{-12\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E} X_{44} + \frac{-4\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \mathbb{E} \tilde{X}_{44} + O(\Psi_0^5).
 \end{aligned} \tag{3.8.69}$$

3.8.2.3 Expansion of B_3

Next we estimate the expectation of B_3 in (3.8.22). Using (3.8.5) and (3.8.6), we write B_3 into summation of the following six terms:

$$\begin{aligned}
 B_{31} &:= \frac{b^2}{(\gamma d_a^2 - \xi)^2} \sum_{\alpha, \beta} G_{i\alpha}^{(a)} x_{a\alpha} x_{a\beta} G_{i\beta}^{(a)}, & B_{32} &:= \frac{-2\gamma d_a^2 b \mathcal{Y}_a}{(\gamma d_a^2 - \xi)^3} \sum_{\alpha, \beta} G_{i\alpha}^{(a)} x_{a\alpha} x_{a\beta} G_{i\beta}^{(a)}, \\
 B_{33} &:= \frac{-2b^3 \mathcal{Z}_a}{(\gamma d_a^2 - \xi)^3} \sum_{\alpha, \beta} G_{i\alpha}^{(a)} x_{a\alpha} x_{a\beta} G_{i\beta}^{(a)}, & B_{34} &:= \frac{\gamma^2 d_a^4 + 2\gamma \xi d_a^2}{(\gamma d_a^2 - \xi)^4} \mathcal{Y}_a^2 \sum_{\alpha, \beta} G_{i\alpha}^{(a)} x_{a\alpha} x_{a\beta} G_{i\beta}^{(a)} \\
 B_{35} &:= \frac{6\gamma d_a^2 b^2}{(\gamma d_a^2 - \xi)^4} \mathcal{Y}_a \mathcal{Z}_a \sum_{\alpha, \beta} G_{i\alpha}^{(a)} x_{a\alpha} x_{a\beta} G_{i\beta}^{(a)}, & B_{36} &:= \frac{3b^4}{(\gamma d_a^2 - \xi)^4} \mathcal{Z}_a^2 \sum_{\alpha, \beta} G_{i\alpha}^{(a)} x_{a\alpha} x_{a\beta} G_{i\beta}^{(a)}
 \end{aligned} \tag{3.8.70}$$

We handle the above six terms one by one. For B_{31} , taking partial expectation $\mathbb{E}_a$ and using (3.3.32), we get

$$\begin{aligned}
 \mathbb{E}_a B_{31} &= \frac{1}{N} \sum_{\alpha} \frac{b^2}{(\gamma d_a^2 - \xi)^2} G_{i\alpha}^{(a)} G_{\alpha i}^{(a)} \\
 &= \frac{1}{N} \sum_{\alpha} \frac{b^2}{(\gamma d_a^2 - \xi)^2} \left(G_{i\alpha} G_{\alpha i} - \frac{2G_{ia} G_{a\alpha}}{G_{aa}} G_{\alpha i}^{(a)} - \frac{G_{ia} G_{a\alpha} G_{\alpha a} G_{ai}}{G_{aa}^2} \right)
 \end{aligned} \tag{3.8.71}$$

Expanding the second term in the lower index a , we can write it into summation of four terms:

$$\begin{aligned}
 B_{311} &:= \frac{-2}{N} \sum_{\alpha} \frac{\gamma b^2}{(\gamma d_a^2 - \xi)^2} G_{aa} G_{i\bar{a}}^{(a)} d_a^2 G_{\alpha\bar{a}}^{(a)} G_{\alpha i}^{(a)}, \\
 B_{312} &:= \frac{-2}{N} \sum_{\alpha, \nu} \frac{\gamma b^2}{(\gamma d_a^2 - \xi)^2} G_{aa} G_{i\bar{a}}^{(a)} d_a x_{a\nu} G_{\alpha\nu}^{(a)} G_{\alpha i}^{(a)}, \\
 B_{313} &:= \frac{-2}{N} \sum_{\alpha, \mu} \frac{\gamma b^2}{(\gamma d_a^2 - \xi)^2} G_{aa} G_{i\mu}^{(a)} x_{a\mu} d_a G_{\alpha\bar{a}}^{(a)} G_{\alpha i}^{(a)}, \\
 B_{314} &:= \frac{-2}{N} \sum_{\alpha, \mu, \nu} \frac{\gamma b^2}{(\gamma d_a^2 - \xi)^2} G_{aa} G_{i\mu}^{(a)} x_{a\mu} x_{a\nu} G_{\alpha\nu}^{(a)} G_{\alpha i}^{(a)}.
 \end{aligned}$$

We have

$$\begin{aligned}
 \mathbb{E}B_{311} &= \mathbb{E} \sum_{\alpha,p,q} \frac{-2\gamma^2 b^2}{(\gamma d_a^2 - \xi)^2} G_{aa} G_{\bar{a}\bar{a}}^{(a)2} d_a^2 G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{\alpha q}^{(a\bar{a})} \left(G_{\alpha i}^{(a\bar{a})} + \frac{G_{\alpha\bar{a}}^{(a)} G_{\bar{a}i}^{(a)}}{G_{\bar{a}\bar{a}}^{(a)}} \right) \\
 &= \mathbb{E} \sum_{\alpha,p,q} \frac{-2\gamma^2 b^2 d_a^2}{(\gamma d_a^2 - \xi)^2} \left(\frac{1}{(\gamma d_a^2 - \xi)b} + \frac{(\gamma d_a^2 - 2\xi)\mathcal{Y}_a}{(\gamma d_a^2 - \xi)^2 b^2} + \frac{-\mathcal{Z}_a}{(\gamma d_a^2 - \xi)^2} \right) G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{\alpha q}^{(a\bar{a})} G_{\alpha i}^{(a\bar{a})} \\
 &\quad + \mathbb{E} \sum_{\alpha,p,q} \frac{2\gamma^3 b^2 d_a^2}{(\gamma d_a^2 - \xi)^3} G_{ip}^{(a\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{\alpha q}^{(a\bar{a})} G_{\alpha\bar{a}}^{(a)} G_{\bar{a}i}^{(a)} \\
 &= \frac{-2\gamma^2 b d_a^2}{(\gamma d_a^2 - \xi)^3} \mathbb{E}X'_{33} + \left(\frac{6\gamma^3 d_a^2 b^2 E_+^2}{(\gamma d_a^2 - \xi)^4} + \frac{6\gamma^3 \xi d_a^2 E_+}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E}X_{44} + \frac{2\gamma^3 (\gamma d_a^4 - 2\xi d_a^2) E_+}{(\gamma d_a^2 - \xi)^4} (2\mathbb{E}X_{44} - \mathbb{E}\tilde{X}_{44}) \\
 &\quad + \frac{2\gamma^3 b^2 d_a^2 E_+^2}{(\gamma d_a^2 - \xi)^4} \mathbb{E}X_{43} + \frac{2\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} (2\mathbb{E}X_{44} + \mathbb{E}\tilde{X}_{44}) + O(\Psi_0^5).
 \end{aligned}$$

$$\mathbb{E}B_{312} = \mathbb{E}B_{313} = \frac{-4\gamma^3 b^2 d_a^2 E_+^2}{(\gamma d_a^2 - \xi)^4} \mathbb{E}X_{44} + O(\Psi_0^5).$$

$$\begin{aligned}
 \mathbb{E}B_{314} &= \frac{-2}{N} \sum \frac{\gamma b^2}{(\gamma d_a^2 - \xi)^2} \left(\frac{b}{\gamma d_a^2 - \xi} - \frac{\gamma d_a^2 \mathcal{Y}_a}{(\gamma d_a^2 - \xi)^2} - \frac{b^2 \mathcal{Z}_a}{(\gamma d_a^2 - \xi)^2} \right) G_{i\mu}^{(a)} x_{a\mu} x_{a\nu} G_{\alpha\nu}^{(a)} G_{\alpha i}^{(a)} + O(\Psi_0^5) \\
 &= \frac{-2\gamma b^3}{(\gamma d_a^2 - \xi)^3} \mathbb{E}X''_{33} + \left(\frac{6\gamma^3 b^2 d_a^2 E_+^2}{(\gamma d_a^2 - \xi)^4} + \frac{6\gamma^2 b^4 E_+^3}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E}X_{44} \\
 &\quad + \frac{2\gamma^3 b^2 d_a^2 E_+^2}{(\gamma d_a^2 - \xi)^4} \mathbb{E}X_{43} + \frac{2\gamma^2 b^4 E_+^3}{(\gamma d_a^2 - \xi)^4} (\mathbb{E}X_{43} - 2\mathbb{E}X_{44}) + O(\Psi_0^5).
 \end{aligned}$$

For the last term in (3.8.71) it can be handled by similar steps to (3.8.50), (3.8.51) and (3.8.52). We obtain

$$\begin{aligned}
 &\mathbb{E} \frac{1}{N} \sum \frac{b^2}{(\gamma d_a^2 - \xi)^2} \frac{-G_{ia} G_{a\alpha} G_{\alpha a} G_{ai}}{G_{aa}^2} \\
 &= - \left(\frac{\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} + \frac{2\gamma^3 d_a^2 b^2 E_+^2}{(\gamma d_a^2 - \xi)^4} + \frac{\gamma^2 b^4 E_+^3}{(\gamma d_a^2 - \xi)^4} \right) (2\mathbb{E}X_{44} + \mathbb{E}\tilde{X}_{44}) + O(\Psi_0^5).
 \end{aligned} \tag{3.8.72}$$

Then we conclude $\mathbb{E}B_{31}$ with negligible order $O(\Psi_0^5)$ from the argument between (3.8.71) and (3.8.72).

We turn to find other terms defined in (3.8.70). By analogous decoupling procedure as before, we can obtain that

$$\begin{aligned}\mathbb{E}B_{32} &= \frac{-2\gamma^2 b d_a^2}{(\gamma d_a^2 - \xi)^3} \mathbb{E}X'_{32} + \left(\frac{4\gamma^3 d_a^2 b^2 E_+^2}{(\gamma d_a^2 - \xi)^4} + \frac{4\gamma^3 d_a^2 \xi E_+}{(\gamma d_a^2 - \xi)^4} + \frac{4\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E}X_{43} \\ &\quad + \frac{-8\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \mathbb{E}X_{44} + \left(\frac{-2\gamma^3 d_a^2 b^2 E_+^2}{(\gamma d_a^2 - \xi)^4} + \frac{-2\gamma^3 d_a^2 \xi E_+}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E}\tilde{X}_{44} + O(\Psi_0^5).\end{aligned}$$

$$\begin{aligned}\mathbb{E}B_{33} &= \frac{-2\gamma b^3 E_+}{(\gamma d_a^2 - \xi)^3} \mathbb{E}X'_{32} + \frac{4\gamma b^3}{(\gamma d_a^2 - \xi)^3} \mathbb{E}X''_{33} + \frac{-2\gamma b^4}{(\gamma d_a^2 - \xi)^3} (E_+ - z) \mathbb{E}X'_{22} \\ &\quad + \left(\frac{2\gamma^3 d_a^2 b^2 E_+^2}{(\gamma d_a^2 - \xi)^4} + \frac{2\gamma^2 b^4 E_+^3}{(\gamma d_a^2 - \xi)^4} \right) (2\mathbb{E}X_{43} - 6\mathbb{E}X_{44} - \mathbb{E}\tilde{X}_{44}) + O(\Psi_0^5).\end{aligned}$$

$$\mathbb{E}B_{34} = \frac{\gamma^4 d_a^4 E_+ + 2\gamma^3 \xi d_a^2 E_+}{(\gamma d_a^2 - \xi)^4} (\mathbb{E}X_{42} + 2\mathbb{E}\tilde{X}_{44}) + O(\Psi_0^5).$$

$$\mathbb{E}B_{35} = \frac{6\gamma^3 d_a^2 b^2 E_+^2}{(\gamma d_a^2 - \xi)^4} (\mathbb{E}X_{42} - 2\mathbb{E}X_{43}) + O(\Psi_0^5).$$

$$\begin{aligned}\mathbb{E}B_{36} &= \frac{3\gamma^2 b^4 E_+^3}{(\gamma d_a^2 - \xi)^4} \mathbb{E}X_{42} + \frac{-12\gamma^2 b^4 E_+^3}{(\gamma d_a^2 - \xi)^4} \mathbb{E}X_{43} + \left(\frac{24\gamma^2 b^4 E_+^3}{(\gamma d_a^2 - \xi)^4} + \frac{24\gamma^2 b^2 d_a^2 E_+^2}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E}X_{44} \\ &\quad + \left(\frac{6\gamma^2 b^4 E_+^3}{(\gamma d_a^2 - \xi)^4} + \frac{12\gamma^3 b^2 d_a^2 E_+^2}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E}\tilde{X}_{44} + O(\Psi_0^5).\end{aligned}$$

We conclude that

$$\begin{aligned}\mathbb{E}B_3 &= \frac{b^2}{(\gamma d_a^2 - \xi)^2} \mathbb{E}X'_{22} + \left(\frac{-2\gamma^2 b d_a^2}{(\gamma d_a^2 - \xi)^3} + \frac{-2\gamma b^3 E_+}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E}X'_{32} + \frac{-2\gamma^2 b d_a^2}{(\gamma d_a^2 - \xi)^3} \mathbb{E}X'_{33} + \\ &\quad + \left(\frac{-2\gamma^2 b d_a^2}{(\gamma d_a^2 - \xi)^3} + \frac{4\gamma b^3}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E}X''_{33} + \left(\frac{2\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} + \frac{-6\gamma^2 b^4 E_+^3}{(\gamma d_a^2 - \xi)^4} + \frac{4\gamma^3 d_a^2 \xi E_+}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E}X_{43} \\ &\quad + \left(\frac{\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} + \frac{6\gamma^3 b^2 d_a^2 E_+^2}{(\gamma d_a^2 - \xi)^4} + \frac{3\gamma^2 b^4 E_+^2}{(\gamma d_a^2 - \xi)^4} + \frac{2\gamma^3 \xi d_a^2 E_+}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E}X_{42} \\ &\quad + \left(\frac{3\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} + \frac{6\gamma^3 b^2 d_a^2 E_+^2}{(\gamma d_a^2 - \xi)^4} + \frac{3\gamma^2 b^4 E_+^3}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E}\tilde{X}_{44} \\ &\quad + \left(\frac{-2\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^4} + \frac{12\gamma^3 b^2 d_a^2 E_+^2}{(\gamma d_a^2 - \xi)^4} + \frac{12\gamma^2 b^4 E_+^3}{(\gamma d_a^2 - \xi)^4} + \frac{2\gamma^3 \xi d_a^2 E_+}{(\gamma d_a^2 - \xi)^4} \right) \mathbb{E}X_{44} \\ &\quad + \frac{-2\gamma b^4 E_+}{(\gamma d_a^2 - \xi)^3} (E_+ - z) \mathbb{E}X_{22} + O(\Psi_0^5).\end{aligned}$$

3.8.2.4 Optical theorem from X_{22}

By tedious calculations, we summarize that

$$\begin{aligned}
 \frac{1}{N} \mathbb{E} \sum_a^{(i)} G_{ia} G_{ai} &= \gamma \psi_2 \mathbb{E} X_{22} + \gamma b^2 \varphi_2 \mathbb{E} X'_{22} - \frac{1}{N} \sum_a \left(\frac{2\gamma^3 \tilde{b} d_a^2}{(\gamma d_a^2 - \xi)^3} + \frac{2\gamma^3 b d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E} X_{32} \\
 &- \frac{1}{N} \sum_a \left(\frac{2\gamma^3 b d_a^2}{(\gamma d_a^2 - \xi)^3} + \frac{2\gamma^2 b^3 E_+}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E} X'_{32} + \frac{1}{N} \sum_a \frac{2\gamma^3 \tilde{b} d_a^2}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X_{33} + \frac{1}{N} \sum_a \frac{4\gamma^3 b d_a^2}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X'_{33} \\
 &+ \frac{1}{N} \sum_a \frac{2\gamma^2 b^3}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X''_{33} - \frac{1}{N} \sum_a \left(\frac{2\gamma^2 b^2 d_a^2}{(\gamma d_a^2 - \xi)^3} + \frac{2\gamma b^4 E_+}{(\gamma d_a^2 - \xi)^3} \right) (E_+ - z) \mathbb{E} X_{22} \\
 &+ \frac{1}{N} \sum_a \left(\frac{\gamma^3 E_+ (\gamma d_a^2 + E_+ b^2)^2}{(\gamma d_a^2 - \xi)^4} + \frac{\gamma^4 d_a^2 (E_+ b + \tilde{b})^2}{(\gamma d_a^2 - \xi)^4} \right) (3\mathbb{E} X_{42} - \mathbb{E} 6X_{43} + \mathbb{E} 12X_{44} + 3\mathbb{E} \tilde{X}_{44}) + O(\Psi_0^5)
 \end{aligned} \tag{3.8.73}$$

By similar procedure, we obtain that

$$\begin{aligned}
 \frac{1}{N} \mathbb{E} \sum_a^{(i)} G_{ia} G_{ai} &= \gamma \psi_2 \mathbb{E} X'_{22} + \gamma \tilde{b}^2 \varphi_2 \mathbb{E} X_{22} - \frac{1}{N} \sum_a \left(\frac{2\gamma^3 \tilde{b} d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} + \frac{2\gamma^2 \tilde{b}^3}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E} X_{32} \\
 &- \frac{1}{N} \sum_a \left(\frac{2\gamma^3 b d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} + \frac{2\gamma^3 \tilde{b} d_a^2}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E} X'_{32} + \frac{1}{N} \sum_a \frac{2\gamma^3 d_a^2 b}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X''_{33} + \frac{1}{N} \sum_a \frac{4\gamma^3 d_a^2 \tilde{b}}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X'_{33} \\
 &+ \frac{1}{N} \sum_a \frac{2\gamma^2 \tilde{b}^3}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X_{33} - \frac{1}{N} \sum_a \left(\frac{2\gamma^2 \xi d_a^2}{(\gamma d_a^2 - \xi)^3} + \frac{2\gamma^2 b^2 d_a^2 E_+}{(\gamma d_a^2 - \xi)^3} \right) (E_+ - z) \mathbb{E} X_{22} \\
 &+ \frac{1}{N} \sum_a \left(\frac{\gamma^3 (\gamma E_+ d_a^2 + \tilde{b}^2)^2}{(\gamma d_a^2 - \xi)^4} + \frac{\gamma^4 E_+ d_a^2 (E_+ b + \tilde{b})^2}{(\gamma d_a^2 - \xi)^4} \right) (3\mathbb{E} X_{42} - 6\mathbb{E} X_{43} + 12\mathbb{E} X_{44} + 3\mathbb{E} \tilde{X}_{44}) + O(\Psi_0^5)
 \end{aligned} \tag{3.8.74}$$

It is easy to obtain that

$$\mathbb{E} G_{ii}^2 = \frac{\gamma d_i^2}{(\gamma d_i^2 - \xi)^2} - \frac{2\gamma^2 b d_i^2 E_+}{(\gamma d_i^2 - \xi)^3} \mathbb{E}(cm - cm_N) - \frac{2\gamma^2 \tilde{b} d_i^2}{(\gamma d_i^2 - \xi)^3} \mathbb{E}(cm - cm_N) + O(\Psi_0^2), \tag{3.8.75}$$

and

$$\begin{aligned}
 \frac{1}{N} \mathbb{E} \sum_{\alpha > 2M} G_{i\alpha} G_{\alpha i} &= \frac{\gamma(1-c)}{b^2} \mathbb{E} X_{22} + \frac{2\gamma^2(1-c)}{b^3} \mathbb{E} X_{32} - \frac{2\gamma^2(1-c)}{b^3} \mathbb{E} X_{33} \\
 &+ \frac{\gamma^3(1-c)}{b^4} (3\mathbb{E} X_{42} - 6\mathbb{E} X_{43} + 12\mathbb{E} X_{44} + 3\mathbb{E} \tilde{X}_{44}) + O(\Psi_0^5)
 \end{aligned} \tag{3.8.76}$$

We remark that in (3.8.73) and (3.8.74), terms contain $\mathbb{E} X'_{32}$ can be further represented as linear combination of $\mathbb{E} X_{32}$ and $\mathbb{E}(cm - cm_N)$, see (3.8.96) below. Terms

contain $\mathbb{E}X'_{33}$ and $\mathbb{E}X''_{33}$ can be replaced with $E_+\mathbb{E}X_{33}$ and $E_+^2\mathbb{E}X_{33}$ respectively according to (3.8.89) and (3.8.90).

Let

$$\theta_4 := \frac{1}{N} \sum_a \left(\frac{\gamma^3 E_+^2 (\gamma d_a^2 + E_+ b^2)^2}{(\gamma d_a^2 - \xi)^4} + \frac{2\gamma^4 d_a^2 E_+ (E_+ b + \tilde{b})^2}{(\gamma d_a^2 - \xi)^4} + \frac{\gamma^3 (\gamma E_+ d_a^2 + \tilde{b}^2)^2}{(\gamma d_a^2 - \xi)^4} + \frac{\gamma^3 (1-c)}{b^4} \right)$$

Multiplying E_+ on both sides of (3.8.73) and (3.8.21), then summing with (3.8.74), (3.8.75) and (3.8.76) together, we find that the coefficient of $\mathbb{E}X_{22}$ and $\mathbb{E}X'_{22}$ cancel by (3.8.15), and obtain

$$\begin{aligned} & \frac{E_+ b^2}{(\gamma d_i^2 - \xi)^2} + \frac{\gamma d_i^2}{(\gamma d_i^2 - \xi)^2} - \left(\frac{2\gamma^2 \tilde{b} d_i^2}{(\gamma d_i^2 - \xi)^3} + \frac{4\gamma^2 b d_i^2 E_+}{(\gamma d_i^2 - \xi)^3} + \frac{2\gamma b^3 E_+^2}{(\gamma d_i^2 - \xi)^3} \right) \mathbb{E}(cm - cm_N) \\ & - 2\Phi_2 X_{32} - 2\Phi_1 \mathbb{E}X'_{32} + (2\Phi_1 E_+ + 2\Phi_2) \mathbb{E}X_{33} - \frac{2b\Phi_1}{\gamma} (E_+ - z) \mathbb{E}X_{22} \\ & + \theta_4 (3\mathbb{E}X_{42} - 6\mathbb{E}X_{43} + 12\mathbb{E}X_{44} + 3\mathbb{E}\tilde{X}_{44}) = O(\Psi_0^5) \end{aligned}$$

Summing over index i , and using (3.8.96), we have that

$$\begin{aligned} & E_+ b^2 \varphi_2 + \psi_2 + (2\Phi_2 + 2E_+ \Phi_1)(-X_{32} + X_{33}) - \frac{2b\Phi_1}{\gamma} \mathbb{E}(cm - cm_N) \\ & - \frac{2b\Phi_1}{\gamma} (E_+ - z) X_{22} + \theta_4 (3X_{42} - 6X_{43} + 12X_{44} + 3\tilde{X}_{44}) = O(\Psi_0^5). \end{aligned} \quad (3.8.77)$$

Recall (3.4.34). Using Lemma 3.8.2, we get the optical theorem from X_{22} ,

$$\frac{1}{\gamma} + \frac{2}{\gamma} \mathbb{E}\mathcal{X}_3 + \frac{2b^2}{\gamma^2 h} \mathbb{E}\mathfrak{X}_4 + \theta_4 \mathbb{E}\mathcal{X}_4 = O(\Psi_0^5), \quad (3.8.78)$$

where

$$\theta_4 = \gamma^3 h^4 \varphi_4 - \frac{4E_+}{h} - \frac{4\gamma^2 E_+}{b^2} - \frac{2\gamma^2 E_+^2}{b^2 h} + \frac{\gamma^3 (1-c)}{b^4}. \quad (3.8.79)$$

The above expression of θ_4 is obtained through simple but lengthy calculations with the help of Lemma 3.8.2.

Combining (3.8.74), (3.8.75) and (3.8.76), we have

$$\begin{aligned} \frac{1}{N} \sum_i \mathbb{E} X'_{22} = & E_+ \frac{1}{N} \sum_i \mathbb{E} X_{22} + \frac{1}{N} \sum_i \frac{2\tilde{b}}{\gamma E_+} \mathbb{E}[X_{32} - X_{33}] + \frac{\psi_2}{1 - \gamma\psi_2} \\ & - \frac{1}{N^2} \sum_{i,a} \frac{2\gamma^2 d_a^2 b (E_+ b + \tilde{b})^2}{(\gamma d_a^2 - \xi)^3 E_+} [\mathbb{E}(cm - cm_N) + (E_+ - z) \mathbb{E} X_{22}] \\ & + \frac{1}{1 - \gamma\psi_2} \left\{ \frac{1}{N} \sum_a \left(\frac{\gamma^3 (\gamma E_+ d_a^2 + \tilde{b}^2)^2}{(\gamma d_a^2 - \xi)^4} + \frac{\gamma^4 E_+ d_a^2 (E_+ b + \tilde{b})^2}{(\gamma d_a^2 - \xi)^4} \right) + \frac{\gamma^3 (1 - c)}{b^4} \right\} \\ & \times \frac{1}{N} \sum_i \mathbb{E}(3X_{42} - 6X_{43} + 12X_{44} + 3\tilde{X}_{44}) + O(\Psi_0^5). \end{aligned}$$

Using Lemma 3.8.2, we have

$$\begin{aligned} \frac{1}{N} \sum_i \mathbb{E} X'_{22} = & E_+ \frac{1}{N} \sum_i \mathbb{E} X_{22} + \frac{2\tilde{b}}{\gamma E_+} \mathbb{E} \mathcal{X}_3 + \left(\frac{h}{\gamma E_+} - \frac{1}{\gamma} \right) - \frac{2b(\gamma^2 E_+^2 - b\tilde{b})}{\gamma^2 E_+ h} \mathbb{E} \mathcal{X}_4 \\ & + \left(\frac{\gamma^3 h^4 \tilde{b} \varphi_6}{E_+} - \frac{3\tilde{b}}{h} - \frac{3\gamma^2 \tilde{b}}{b^3} - \frac{3\gamma^2 E_+ \tilde{b}}{b^2 h} - \frac{E_+ b^2 + \gamma^2 E_+^2}{bh} + \frac{h\gamma^3 (1 - c)}{E_+ b^4} \right) \mathbb{E} \mathcal{X}_4 + O(\Psi_0^5) \end{aligned} \quad (3.8.80)$$

3.8.3 Optical theorem from X_{33}

Recall

$$X_{33} = \frac{1}{N^2} \sum_{a,b} G_{ia} G_{ab} G_{bi}.$$

Since the contributions to the sum when $a = i$ or $a = b$ are of negligible order $O(\Psi_0^2)$, we can assume that $i, b \neq a$. Expanding X_{33} with respect to index a , we have

$$\begin{aligned} \frac{1}{N^2} \sum_{a,b} G_{ia} G_{ab} G_{bi} &= \frac{1}{N^2} \sum_{a,b,\alpha,\beta} G_{aa}^2 G_{i\alpha}^{(a)} Y_{a\alpha} Y_{a\beta} G_{b\beta}^{(a)} G_{bi} \\ &= \frac{\gamma}{N^2} \sum_{a,b} G_{aa}^2 G_{i\bar{a}}^{(a)} d_a^2 G_{b\bar{a}}^{(a)} G_{bi} + \frac{\gamma}{N^2} \sum_{a,b,\alpha} G_{aa}^2 G_{i\alpha}^{(a)} x_{a\alpha} d_a G_{b\bar{a}}^{(a)} G_{bi} \\ &\quad + \frac{\gamma}{N^2} \sum_{a,b,\beta} G_{aa}^2 G_{i\bar{a}}^{(a)} d_a x_{a\beta} G_{b\beta}^{(a)} G_{bi} + \frac{\gamma}{N^2} \sum_{a,b,\alpha,\beta} G_{aa}^2 G_{i\alpha}^{(a)} x_{a\alpha} x_{a\beta} G_{b\beta}^{(a)} G_{bi} \end{aligned} \quad (3.8.81)$$

By decomposing G_{bi} into sum of $G_{bi}^{(\bar{a})}$ and $G_{b\bar{a}} G_{\bar{a}i} / G_{\bar{a}\bar{a}}$ for the first three terms, and sum of $G_{bi}^{(a)}$ and $G_{ba} G_{ai} / G_{aa}$ for the last term, we can write X_{33} into sum of eight

terms:

$$\begin{aligned}
D_1 &:= \frac{\gamma}{N^2} \sum_{a,b} G_{aa}^2 G_{i\bar{a}}^{(a)} d_a^2 G_{b\bar{a}}^{(a)} G_{bi}^{(\bar{a})}, & D_2 &:= \frac{\gamma}{N^2} \sum_{a,b} G_{aa}^2 G_{i\bar{a}}^{(a)} d_a^2 G_{b\bar{a}}^{(a)} \frac{G_{b\bar{a}} G_{\bar{a}i}}{G_{\bar{a}\bar{a}}}, \\
D_3 &:= \frac{\gamma}{N^2} \sum_{a,b,\alpha} G_{aa}^2 G_{i\alpha}^{(a)} x_{a\alpha} d_a G_{b\bar{a}}^{(a)} G_{bi}^{(\bar{a})}, & D_4 &:= \frac{\gamma}{N^2} \sum_{a,b,\alpha} G_{aa}^2 G_{i\alpha}^{(a)} x_{a\alpha} d_a G_{b\bar{a}}^{(a)} \frac{G_{b\bar{a}} G_{\bar{a}i}}{G_{\bar{a}\bar{a}}}, \\
D_5 &:= \frac{\gamma}{N^2} \sum_{a,b,\beta} G_{aa}^2 G_{i\bar{a}}^{(a)} d_a x_{a\beta} G_{b\beta}^{(a)} G_{bi}^{(\bar{a})}, & D_6 &:= \frac{\gamma}{N^2} \sum_{a,b,\beta} G_{aa}^2 G_{i\bar{a}}^{(a)} d_a x_{a\beta} G_{b\beta}^{(a)} \frac{G_{b\bar{a}} G_{\bar{a}i}}{G_{\bar{a}\bar{a}}}, \\
D_7 &:= \frac{\gamma}{N^2} \sum_{a,b,\alpha,\beta} G_{aa}^2 G_{i\alpha}^{(a)} x_{a\alpha} x_{a\beta} G_{b\beta}^{(a)} G_{bi}^{(\bar{a})}, & D_8 &:= \frac{\gamma}{N^2} \sum_{a,b,\alpha,\beta} G_{aa}^2 G_{i\alpha}^{(a)} x_{a\alpha} x_{a\beta} G_{b\beta}^{(a)} \frac{G_{ba} G_{ai}}{G_{aa}}.
\end{aligned} \tag{3.8.82}$$

We analyze above terms one by one. D_1 can be written as

$$\begin{aligned}
D_1 &= \frac{\gamma^2}{N^2} \sum_{a,b,j,k} G_{aa}^2 G_{\bar{a}\bar{a}}^{(a)^2} d_a^2 G_{ij}^{(a\bar{a})} x_{j\bar{a}} x_{k\bar{a}} G_{bk}^{(a\bar{a})} G_{bi}^{(\bar{a})} \\
&= \frac{\gamma^2}{N^2} \sum \left(\frac{1}{(\gamma d_a^2 - \xi)^2} + \frac{-2\tilde{b}\mathcal{Y}_a}{(\gamma d_a^2 - \xi)^3} + \frac{-2b\mathcal{Z}_a}{(\gamma d_a^2 - \xi)^3} \right) d_a^2 G_{ij}^{(a\bar{a})} x_{j\bar{a}} x_{k\bar{a}} G_{bk}^{(a\bar{a})} G_{bi}^{(\bar{a})} + O(\Psi_0^5).
\end{aligned} \tag{3.8.83}$$

The expectation of the first term on the right side of (3.8.83) equals

$$\begin{aligned}
&\mathbb{E} \frac{1}{N^3} \sum_{a,b,j} \frac{\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} G_{ij}^{(a\bar{a})} G_{jb}^{(a\bar{a})} G_{bi}^{(\bar{a})} \\
&= \mathbb{E} \frac{1}{N^3} \sum_{a,b,j} \frac{\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} G_{ij}^{(a\bar{a})} G_{jb}^{(a\bar{a})} G_{bi}^{(\bar{a})} + \mathbb{E} \frac{1}{N^3} \sum_{a,b,j} \frac{\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} G_{ij}^{(a\bar{a})} G_{jb}^{(a\bar{a})} \frac{G_{ba}^{(\bar{a})} G_{ai}^{(\bar{a})}}{G_{aa}^{(\bar{a})}} \\
&= \frac{1}{N} \sum_a \frac{\gamma^2 d_a^2}{(\gamma d_a^2 - \xi)^2} \mathbb{E} X_{33} + \frac{1}{N} \sum_a \frac{-3\gamma^3 d_a^2 (E_+ b + \tilde{b})}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X_{44} + \frac{1}{N} \sum_a \frac{-\gamma^3 d_a^2 E_+}{(\gamma d_a^2 - \xi)^2 \tilde{b}} \mathbb{E} X_{44} + O(\Psi_0^5),
\end{aligned}$$

where we use the third equation in (3.8.20) to handle the first term in the second line above. The expectation of the other two terms in (3.8.83) equals

$$\frac{1}{N} \sum_a \frac{-2\gamma^3 d_a^2 \tilde{b}}{(\gamma d_a^2 - \xi)^3} (\mathbb{E} X_{43} - 2\mathbb{E} X_{44}) + \frac{1}{N} \sum_a \frac{-2\gamma^3 d_a^2 b E_+}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X_{43} + O(\Psi_0^5). \tag{3.8.84}$$

Then we get $\mathbb{E} D_1$ from above easily. For the other terms, we find

$$\mathbb{E} D_2 = \frac{1}{N} \sum_a \frac{\gamma^4 d_a^4 E_+}{(\gamma d_a^2 - \xi)^3 \tilde{b}} \mathbb{E} X_{44} + \frac{1}{N} \sum_a \frac{2\gamma^3 d_a^2 \tilde{b}}{(\gamma d_a^2 - \xi)^3} \mathbb{E} X_{44} + \frac{1}{N} \sum_a \frac{\gamma^3 d_a^2 \tilde{b}}{(\gamma d_a^2 - \xi)^3} \mathbb{E} \tilde{X}_{44} + O(\Psi_0^5).$$

$$\begin{aligned}
 \mathbb{E}D_3 &= \mathbb{E}D_5 = \frac{1}{N} \sum_a \frac{4\gamma^3 d_a^2 b E_+}{(\gamma d_a^2 - \xi)^3} \mathbb{E}X_{44} + O(\Psi_0^5). \\
 \mathbb{E}D_4 &= \mathbb{E}D_6 = \frac{1}{N} \sum_a \frac{\gamma^3 d_a^2 b E_+}{(\gamma d_a^2 - \xi)^3} (\mathbb{E}X_{44} + \mathbb{E}\tilde{X}_{44}) + O(\Psi_0^5). \\
 \mathbb{E}D_7 &= \frac{1}{N} \sum_a \frac{\gamma b^2}{(\gamma d_a^2 - \xi)^2} \mathbb{E}X'_{33} + \frac{1}{N} \sum_a \left(\frac{-3\gamma^3 d_a^2 b E_+}{(\gamma d_a^2 - \xi)^3} + \frac{-3\gamma^2 b^3 E_+^2}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E}X_{44} \\
 &\quad + \frac{1}{N} \sum_a \left(\frac{-2\gamma^3 d_a^2 b E_+}{(\gamma d_a^2 - \xi)^3} + \frac{-2\gamma^2 b^3 E_+^2}{(\gamma d_a^2 - \xi)^3} \right) \mathbb{E}X_{43} + \frac{1}{N} \sum_a \frac{4\gamma^2 b^3 E_+^2}{(\gamma d_a^2 - \xi)^3} \mathbb{E}X_{44} + O(\Psi_0^5). \\
 \mathbb{E}D_8 &= \frac{1}{N} \sum_a \frac{\gamma^3 d_a^2 b E_+}{(\gamma d_a^2 - \xi)^3} \mathbb{E}X_{44} + \frac{1}{N} \sum_a \frac{\gamma^2 b^3 E_+^2}{(\gamma d_a^2 - \xi)^3} (2\mathbb{E}X_{44} + \mathbb{E}\tilde{X}_{44}) + O(\Psi_0^5).
 \end{aligned}$$

We conclude from above that

$$\mathbb{E}X_{33} = \gamma\psi_2 \mathbb{E}X_{33} + \gamma b^2 \varphi_2 \mathbb{E}X'_{33} + \Phi_1(-2\mathbb{E}X_{43} + 3\mathbb{E}X_{44} + \mathbb{E}\tilde{X}_{44}) + O(\Psi_0^5).$$

Using Lemma 3.8.2, it follows that

$$\mathbb{E}X_{33} = \frac{1}{E_+} \mathbb{E}X'_{33} + \frac{-b}{\gamma E_+} (-2\mathbb{E}X_{43} + 3\mathbb{E}X_{44} + \mathbb{E}\tilde{X}_{44}) + O(\Psi_0^5). \quad (3.8.85)$$

By similar procedure of expanding X_{33} , we can derive

$$\frac{1}{N^2} \sum_{\alpha > 2M, k} G_{i\alpha} G_{\alpha k} G_{ki} = \frac{\gamma(1-c)}{b^2} \mathbb{E}X_{33} + \frac{-\gamma^2(1-c)}{b^3} (-2\mathbb{E}X_{43} + 3\mathbb{E}X_{44} + \mathbb{E}\tilde{X}_{44}) + O(\Psi_0^5). \quad (3.8.86)$$

And

$$\mathbb{E}X'_{33} = \gamma\psi_2 \mathbb{E}X'_{33} + \gamma\varpi_2 \mathbb{E}X_{33} + \Phi_2(-2\mathbb{E}X_{43} + 3\mathbb{E}X_{44} + \mathbb{E}\tilde{X}_{44}) + O(\Psi_0^5).$$

It follows that

$$\mathbb{E}X'_{33} = E_+ \mathbb{E}X_{33} + \frac{-\tilde{b}}{\gamma E_+} (-2\mathbb{E}X_{43} + 3\mathbb{E}X_{44} + \mathbb{E}\tilde{X}_{44}) + O(\Psi_0^5). \quad (3.8.87)$$

From (3.8.85) and (3.8.87), we get the optical theorem obtained from X_{33} ,

$$-2\mathbb{E}X_{43} + 3\mathbb{E}X_{44} + \mathbb{E}\tilde{X}_{44} = O(\Psi_0^5). \quad (3.8.88)$$

Substituting (3.8.88) back into (3.8.87), we get

$$\mathbb{E}X'_{33} = E_+ \mathbb{E}X_{33} + O(\Psi_0^5). \quad (3.8.89)$$

Similarly, by expanding $\mathbb{E}X'_{33}$ with respect to the Latin lower index and an $\mathbb{E}X''_{33}$ with respect to a Greek lower index, together with the optical theorem (3.8.88), we can get

$$\mathbb{E}X''_{33} = E_+ \mathbb{E}X'_{33} + O(\Psi_0^5). \quad (3.8.90)$$

3.8.4 Optical theorem from X_{32}

Recall

$$X_{32} = (cm(E_+) - cm_N) \frac{1}{N} \sum_r G_{ir} G_{ri}.$$

Expanding with lower index r , through tedious calculations, we obtain

$$\mathbb{E}X_{32} - \frac{1}{N} \mathbb{E}(cm - cm_N) G_{ii}^2 = \gamma \psi_2 \mathbb{E}X_{32} + \gamma b^2 \varphi_2 \mathbb{E}X'_{32} + \Phi_1 \mathbb{E}(-2X_{42} + 2X_{43} - 2X_{44}) + O(\Psi_0^5). \quad (3.8.91)$$

Expanding of the left side of the following with lower index α , we get

$$\frac{1}{N} \sum_{\alpha > 2M} \mathbb{E}(cm - cm_N) G_{i\alpha} G_{\alpha i} = \frac{\gamma(1-c)}{b^2} \mathbb{E}X_{32} - \frac{\gamma^2(1-c)}{b^3} \mathbb{E}(-2X_{42} + 2X_{43} - 2X_{44}) + O(\Psi_0^5). \quad (3.8.92)$$

Similar to (3.8.91), we can obtain

$$\mathbb{E}X'_{32} - \frac{1}{N} \mathbb{E}(cm - cm_N) G_{ii}^2 = \gamma \psi_2 \mathbb{E}X'_{32} + \gamma \varpi_2 \mathbb{E}X_{32} + \Phi_2 \mathbb{E}(-2X_{42} + 2X_{43} - 2X_{44}) + O(\Psi_0^5). \quad (3.8.93)$$

From (3.8.93), we have

$$\mathbb{E}X'_{32} = E_+ \mathbb{E}X_{32} + \frac{\gamma d_i^2 (E_+ b + \tilde{b})}{E_+ (\gamma d_i^2 - \xi)^2} \mathbb{E} \frac{1}{N} (cm - cm_N) - \frac{\tilde{b}}{\gamma E_+} \mathbb{E}(-2X_{42} + 2X_{43} - 2X_{44}) \quad (3.8.94)$$

Using Lemma 3.8.2, we find from (3.8.91) and (3.8.94) that

$$\begin{aligned} \frac{E_+ b^2 + \gamma d_i^2}{(\gamma d_i - \xi)^2} \mathbb{E} \frac{1}{N} (cm - cm_N) &= -(E_+ \Phi_1 + \Phi_2) \mathbb{E}(-2X_{42} + 2X_{43} - 2X_{44}) + O(\Psi_0^5) \\ &= \frac{1}{\gamma} \mathbb{E}(-2X_{42} + 2X_{43} - 2X_{44}) + O(\Psi_0^5) \end{aligned} \quad (3.8.95)$$

Substituting (3.8.95) back into (3.8.94), we have

$$\mathbb{E}X'_{32} = E_+ \mathbb{E}X_{32} + \frac{b}{\gamma d_i^2 - \xi} \mathbb{E} \frac{1}{N} (cm - cm_N) + O(\Psi_0^5) \quad (3.8.96)$$

Moreover, if summing over i and taking the average for (3.8.95), we get the optical theorem derived from X_{32} ,

$$E \frac{1}{N} (cm - cm_N) = \frac{1}{N} \sum_i \mathbb{E}(-2X_{42} + 2X_{43} - 2X_{44}) + O(\Psi_0^5). \quad (3.8.97)$$

3.8.5 Optical theorem from $m_N X_{22}$

In this subsection we show that $(E_+ - z)\mathbb{E}X_{22}$ can be decomposed as linear combinations of $\mathbb{E}X_{42}$, $\mathbb{E}X_{44}$ and $\mathbb{E}\tilde{X}_{44}$, which is an optical theorem derived from $m_N X_{22}$. Recall that $m_N = \frac{1}{M} \sum_a G_{aa}$. Using (3.8.5) and (3.8.6), we get

$$\begin{aligned} \frac{1}{N^2} \sum_{a,j} G_{aa} G_{ij} G_{ji} &= \frac{1}{N^2} \sum_{a,j} \left(\frac{b}{\gamma d_a^2 - \xi} + \frac{-\gamma d_a^2 \mathcal{Y}_a}{(\gamma d_a^2 - \xi)^2} + \frac{-b^2 \mathcal{Z}_a}{(\gamma d_a^2 - \xi)^2} \right. \\ &\quad \left. + \frac{\gamma d_a^2 \tilde{b} \mathcal{Y}_a^2}{(\gamma d_a^2 - \xi)^3} + \frac{2\gamma d_a^2 b \mathcal{Y}_a \mathcal{Z}_a}{(\gamma d_a^2 - \xi)^3} + \frac{b^3 \mathcal{Z}_a^2}{(\gamma d_a^2 - \xi)^3} \right) \times G_{ij} G_{ji} + O(\Psi_0^5). \end{aligned} \quad (3.8.98)$$

After taking expectation, we find the first term on the right side of (3.8.98) is already fully expanded with index a . Recall notations in (3.8.14). By calculations, we find that the second to the sixth terms in (3.8.98) are respectively

$$\mathbb{E} \frac{1}{N^2} \sum_{a,j} \frac{-\gamma d_a^2 \mathcal{Y}_a}{(\gamma d_a^2 - \xi)^2} G_{ij} G_{ji} = -\gamma \psi_2 \mathbb{E}X_{32} + 4\gamma^2 \tilde{b} \psi_3 \mathbb{E}X_{44} - \gamma^2 (E_+ b + \tilde{b}) \psi_3 \mathbb{E}\tilde{X}_{44} + O(\Psi_0^5),$$

$$\begin{aligned} \mathbb{E} \frac{1}{N^2} \sum_{a,j} \frac{-b^2 \mathcal{Z}_a}{(\gamma d_a^2 - \xi)^2} G_{ij} G_{ji} &= -b^3 \varphi_2 (E_+ - z) \mathbb{E}X_{22} - \gamma E_+ b^2 \varphi_2 \mathbb{E}X_{32} + 8\gamma^2 E_+ b \psi_3 \mathbb{E}X_{44} \\ &\quad - (\gamma^2 E_+ b \psi_3 + \gamma^2 E_+^2 b^3 \varphi_3) \mathbb{E}\tilde{X}_{44} + O(\Psi_0^5), \end{aligned}$$

$$\frac{1}{N^2} \sum_{a,j} \frac{\gamma d_a^2 \tilde{b} \mathcal{Y}_a^2}{(\gamma d_a^2 - \xi)^3} G_{ij} G_{ji} = \gamma^2 \tilde{b} \psi_3 (\mathbb{E}X_{42} + 2\mathbb{E}\tilde{X}_{44}) + O(\Psi_0^5),$$

$$\frac{1}{N^2} \sum_{a,j} \frac{2\gamma d_a^2 b \mathcal{Y}_a \mathcal{Z}_a}{(\gamma d_a^2 - \xi)^3} G_{ij} G_{ji} = 2\gamma^2 E_+ b \psi_3 \mathbb{E} X_{42} + O(\Psi_0^5),$$

$$\frac{1}{N^2} \sum_{a,j} \frac{b^3 \mathcal{Z}_a^2}{(\gamma d_a^2 - \xi)^3} G_{ij} G_{ji} = \gamma^2 E_+^2 b^3 \varphi_3 (\mathbb{E} X_{42} + 2\mathbb{E} \tilde{X}_{44}) + 4\gamma^2 E_+ b \psi_3 \mathbb{E} \tilde{X}_{44} + O(\Psi_0^5).$$

Combining above with (3.8.98), we find

$$\begin{aligned} \mathbb{E} c m_N X_{22} &= b \varphi_1 \mathbb{E} X_{22} - b^3 \varphi_2 (E_+ - z) \mathbb{E} X_{22} - (\gamma \psi_2 + \gamma E_+ b^2 \varphi_2) \mathbb{E} X_{32} \\ &\quad + \Phi_1 (\mathbb{E} X_{42} + 4\mathbb{E} X_{44} + \mathbb{E} \tilde{X}_{44}) + O(\Psi_0^5). \end{aligned} \quad (3.8.99)$$

Since $c m_N X_{22} = c m X_{22} - X_{32}$, using Lemma 3.8.2 and $\varphi_1 = c m / b$, we obtain

$$(E_+ - z) \mathbb{E} X_{22} = -\mathbb{E} X_{42} - 4\mathbb{E} X_{44} - \mathbb{E} \tilde{X}_{44} + O(\Psi_0^5). \quad (3.8.100)$$

3.8.6 Proof of Lemma 3.8.3 and Lemma 3.8.4

Proof of Lemma 3.8.3. We only show that $\mathbb{E} X'_{44} = E_+ \mathbb{E} X_{44} + O(\Psi_0^5)$ and the others are similar. By expanding with lower index α using (3.3.31), we find

$$\begin{aligned} \sum_{\alpha} G_{i\alpha} G_{\alpha s} &= \sum_{\alpha} \gamma G_{\alpha\alpha}^2 G_{ij}^{(\alpha)} Y_{j\alpha} Y_{k\alpha} G_{ks}^{(\alpha)} = \sum_a \gamma G_{\bar{a}\bar{a}}^2 d_a^2 G_{ia}^{(\bar{a})} G_{as}^{(\bar{a})} + \sum_a \gamma G_{\bar{a}\bar{a}}^2 G_{ia}^{(\bar{a})} d_a x_{k\bar{a}} G_{ks}^{(\bar{a})} \\ &\quad + \sum_a \gamma G_{\bar{a}\bar{a}}^2 G_{ij}^{(\bar{a})} x_{j\bar{a}} d_a G_{as}^{(\bar{a})} + \sum_{\alpha} \gamma G_{\alpha\alpha}^2 G_{ij}^{(\alpha)} x_{j\alpha} x_{k\alpha} G_{ks}^{(\alpha)}. \end{aligned} \quad (3.8.101)$$

Since we have

$$\begin{aligned} \mathbb{E} \frac{1}{N^3} \sum_{a,s,t} \gamma G_{\bar{a}\bar{a}}^2 G_{ia}^{(\bar{a})} d_a x_{k\bar{a}} G_{ks}^{(\bar{a})} G_{st} G_{ti} &= \mathbb{E} \frac{1}{N^3} \sum_{a,s,t} \gamma G_{\bar{a}\bar{a}}^2 G_{ia}^{(\bar{a})} d_a x_{k\bar{a}} G_{ks}^{(\bar{a})} G_{st}^{(\bar{a})} G_{ti}^{(\bar{a})} + O(\Psi_0^5) \\ &= \mathbb{E} \frac{1}{N^3} \sum_{a,s,t} \frac{\gamma \tilde{b}^2}{(\gamma d_a^2 - \xi)^2} G_{ia}^{(\bar{a})} d_a x_{k\bar{a}} G_{ks}^{(\bar{a})} G_{st}^{(\bar{a})} G_{ti}^{(\bar{a})} + O(\Psi_0^5) = O(\Psi_0^5), \end{aligned}$$

where in the first step we use (3.3.32) and in the second step we use (3.8.8), (3.8.9) and (3.8.11), it follows by substituting (3.8.101) into X'_{44} that

$$\begin{aligned} \mathbb{E} \frac{1}{N^3} \sum_{\alpha,s,t} G_{i\alpha} G_{\alpha s} G_{st} G_{ti} &= \mathbb{E} \sum_{a,s,t} \frac{\gamma d_a^2}{N^3} G_{\bar{a}\bar{a}}^2 G_{ia}^{(\bar{a})} G_{as}^{(\bar{a})} G_{st} G_{ti} \\ &\quad + \mathbb{E} \sum_{\alpha,s,t} \frac{\gamma}{N^3} G_{\alpha\alpha}^2 G_{ij}^{(\alpha)} x_{j\alpha} x_{k\alpha} G_{ks}^{(\alpha)} G_{st} G_{ti} + O(\Psi_0^5) \end{aligned} \quad (3.8.102)$$

Recall ψ_2 and ϖ_2 defined in (3.8.14). By using (3.3.29), (3.3.32), (3.8.8) and (3.8.9) the first term on the right side above equals

$$\mathbb{E} \sum_{a,s,t,\alpha} \frac{\gamma^2 d_a^2}{N^4 (\gamma d_a^2 - \xi)^2} G_{i\alpha}^{(a\bar{a})} G_{\alpha s}^{(a\bar{a})} G_{st}^{(a)} G_{ti}^{(a)} = \gamma \psi_2 \mathbb{E} X'_{44} + O(\Psi_0^5). \quad (3.8.103)$$

Using (3.8.8) and (3.8.9) and (3.8.10), the second term equals

$$\mathbb{E} \sum \frac{\gamma \tilde{b}^2}{N^4 (\gamma d_a^2 - \xi)^2} G_{ij}^{(\bar{a})} G_{js}^{(\bar{a})} G_{st}^{(\bar{a})} G_{ti}^{(\bar{a})} + \mathbb{E} \sum_{\alpha > 2M, j, s, t} \frac{\gamma}{N^4 b^2} G_{ij}^{(\alpha)} G_{js}^{(\alpha)} G_{st}^{(\alpha)} G_{ti}^{(\alpha)} = \gamma \varpi_2 \mathbb{E} X_{44} + O(\Psi_0^5). \quad (3.8.104)$$

Combining (3.8.102), (3.8.103) and (3.8.104) together with the help of (3.8.15), we conclude that $\mathbb{E} X'_{44} = E_+ \mathbb{E} X_{44} + O(\Psi_0^5)$.

Proof of Lemma 3.8.4. We only show the first one and the others can be proved similarly. Using (3.3.32), we have

$$\begin{aligned} \frac{1}{N^2} \sum_{\alpha, \beta}^{(\bar{a})} G_{i\alpha}^{(a\bar{a})} G_{\alpha\beta}^{(a\bar{a})} G_{\beta i}^{(a\bar{a})} &= \frac{1}{N^2} \sum_{\alpha, \beta}^{(\bar{a})} \left(G_{i\alpha} - \frac{G_{i\bar{a}} G_{\bar{a}\alpha}}{G_{\bar{a}\bar{a}}} - \frac{G_{i\bar{a}} G_{\bar{a}\alpha}^{(\bar{a})}}{G_{\bar{a}\bar{a}}^{(\bar{a})}} \right) G_{\alpha\beta} G_{\beta i} \\ &- \frac{1}{N^2} \sum_{a, \alpha, \beta}^{(\bar{a})} G_{i\alpha}^{(a\bar{a})} \left(\frac{G_{\alpha\bar{a}} G_{\bar{a}\beta}}{G_{\bar{a}\bar{a}}} + \frac{G_{\alpha\bar{a}}^{(\bar{a})} G_{\bar{a}\beta}^{(\bar{a})}}{G_{\bar{a}\bar{a}}^{(\bar{a})}} \right) G_{\beta i} - \frac{1}{N^2} \sum_{a, \alpha, \beta}^{(\bar{a})} G_{i\alpha}^{(a\bar{a})} G_{\alpha\beta}^{(a\bar{a})} \left(\frac{G_{\beta\bar{a}} G_{\bar{a}i}}{G_{\bar{a}\bar{a}}} + \frac{G_{\beta\bar{a}}^{(\bar{a})} G_{\bar{a}i}^{(\bar{a})}}{G_{\bar{a}\bar{a}}^{(\bar{a})}} \right). \end{aligned} \quad (3.8.105)$$

We find that

$$\begin{aligned} &\mathbb{E} \frac{1}{N^2} \sum_{\alpha, \beta}^{(\bar{a})} \frac{G_{i\bar{a}} G_{\bar{a}\alpha}}{G_{\bar{a}\bar{a}}} G_{\alpha\beta} G_{\beta i} \\ &= \mathbb{E} \frac{1}{N^2} \sum G_{\bar{a}\bar{a}} G_{ip}^{(\bar{a})} Y_{p\bar{a}} Y_{q\bar{a}} G_{q\alpha}^{(\bar{a})} G_{\alpha\beta}^{(\bar{a})} G_{\beta i}^{(\bar{a})} + O(\Psi_0^5) \\ &= \mathbb{E} \frac{1}{N^2} \sum \gamma^2 G_{\bar{a}\bar{a}} G_{\bar{a}\bar{a}}^{(\bar{a})2} d_a^2 G_{i\mu}^{(\bar{a})} x_{a\mu} x_{a\nu} G_{\nu\alpha}^{(\bar{a})} G_{\alpha\beta}^{(\bar{a})} G_{\beta i}^{(\bar{a})} \\ &\quad + \mathbb{E} \frac{1}{N^2} \sum \gamma G_{\bar{a}\bar{a}} G_{ip}^{(\bar{a})} x_{p\bar{a}} x_{q\bar{a}} G_{q\alpha}^{(\bar{a})} G_{\alpha\beta}^{(\bar{a})} G_{\beta i}^{(\bar{a})} + O(\Psi_0^5) \\ &= \mathbb{E} \frac{1}{N^3} \sum \frac{\gamma^2 d_a^2}{(\gamma d_a^2 - \xi) \tilde{b}} G_{i\mu}^{(\bar{a})} G_{\mu\alpha}^{(\bar{a})} G_{\alpha\beta}^{(\bar{a})} G_{\beta i}^{(\bar{a})} + \mathbb{E} \frac{1}{N^3} \sum \frac{\gamma \tilde{b}}{\gamma d_a^2 - \xi} G_{ip}^{(\bar{a})} G_{p\alpha}^{(\bar{a})} G_{\alpha\beta}^{(\bar{a})} G_{\beta i}^{(\bar{a})} + O(\Psi_0^5) \\ &= \frac{\gamma^2 d_a^2 E_+^3}{(\gamma d_a^2 - \xi) \tilde{b}} \mathbb{E} X_{44} + \frac{\gamma \tilde{b} E_+^2}{\gamma d_a^2 - \xi} \mathbb{E} X_{44} + O(\Psi_0^5), \end{aligned} \quad (3.8.106)$$

where in the second equality we use that

$$\begin{aligned} & \mathbb{E} \frac{1}{N^2} \sum \gamma G_{\bar{a}\bar{a}} G_{ia}^{(\bar{a})} d_a x_{q\bar{a}} G_{q\alpha}^{(\bar{a})} G_{\alpha\beta}^{(\bar{a})} G_{\beta i}^{(\bar{a})} \\ &= \mathbb{E} \frac{1}{N^2} \sum \frac{\gamma \tilde{b}}{\gamma d_a^2 - \xi} G_{ia}^{(\bar{a})} d_a x_{q\bar{a}} G_{q\alpha}^{(\bar{a})} G_{\alpha\beta}^{(\bar{a})} G_{\beta i}^{(\bar{a})} + O(\Psi_0^5) = O(\Psi_0^5). \end{aligned} \quad (3.8.107)$$

And we also have

$$\begin{aligned} \mathbb{E} \frac{1}{N^2} \sum_{\alpha, \beta} \frac{G_{ia}^{(\bar{a})} G_{a\alpha}^{(\bar{a})}}{G_{aa}^{(\bar{a})}} G_{\alpha\beta} G_{\beta i} &= \mathbb{E} \frac{1}{N^2} \sum \gamma G_{aa}^{(\bar{a})} G_{i\mu}^{(\bar{a})} x_{a\mu} x_{a\nu} G_{\nu\alpha}^{(\bar{a})} G_{\alpha\beta} G_{\beta i} \\ &= -\mathbb{E} \frac{1}{N^3} \sum \frac{\gamma}{\tilde{b}} G_{i\mu}^{(\bar{a})} G_{\mu\alpha}^{(\bar{a})} G_{\alpha\beta}^{(\bar{a})} G_{\beta i}^{(\bar{a})} + O(\Psi_0^5) \quad (3.8.108) \\ &= -\frac{\gamma E_+^3}{\tilde{b}} \mathbb{E} X_{44} + O(\Psi_0^5). \end{aligned}$$

Combining (3.8.106) and (3.8.108), we find

$$\mathbb{E} \frac{1}{N^2} \sum_{\alpha, \beta} \left(\frac{G_{ia} G_{\bar{a}\alpha}}{G_{\bar{a}\bar{a}}} + \frac{G_{ia}^{(\bar{a})} G_{a\alpha}^{(\bar{a})}}{G_{aa}^{(\bar{a})}} \right) G_{\alpha\beta} G_{\beta i} = \frac{\gamma E_+^2 (E_+ b + \tilde{b})}{\gamma d_a^2 - \xi} \mathbb{E} X_{44} + O(\Psi_0^5).$$

By similar steps, we obtain

$$\mathbb{E} \frac{1}{N^2} \sum_{\alpha, \beta} G_{i\alpha}^{(a\bar{a})} \left(\frac{G_{\alpha\bar{a}} G_{\bar{a}\beta}}{G_{\bar{a}\bar{a}}} + \frac{G_{\alpha a}^{(\bar{a})} G_{a\beta}^{(\bar{a})}}{G_{aa}^{(\bar{a})}} \right) G_{\beta i} = \frac{\gamma E_+^2 (E_+ b + \tilde{b})}{\gamma d_a^2 - \xi} \mathbb{E} X_{44} + O(\Psi_0^5),$$

and

$$\mathbb{E} \frac{1}{N^2} \sum_{\alpha, \beta} G_{i\alpha}^{(a\bar{a})} G_{\alpha\beta}^{(a\bar{a})} \left(\frac{G_{\beta\bar{a}} G_{\bar{a}i}}{G_{\bar{a}\bar{a}}} + \frac{G_{\beta a}^{(\bar{a})} G_{ai}^{(\bar{a})}}{G_{aa}^{(\bar{a})}} \right) = \frac{\gamma E_+^2 (E_+ b + \tilde{b})}{\gamma d_a^2 - \xi} \mathbb{E} X_{44} + O(\Psi_0^5).$$

From (3.8.105) and above three equations, we conclude the first one in the lemma. $\square$

Chapter 4

Discussion and Future Research

This research work mainly develops the asymptotic distributions of the largest eigenvalues of two types of matrices.

For the sample covariance matrices, we consider a general spiked model and obtain the asymptotic normality of the spiked eigenvalues. It is also shown that the spiked eigenvalues are asymptotically independent of linear spectral statistics. Based on our theoretical results, we propose a statistic combining the information of spiked eigenvalues and linear spectral statistics to test equality of two spiked population covariance matrices, and the statistic performs well under our designed settings.

For the signal-plus-noise matrices, we consider a model that the signal matrix contains no spikes but is allowed to have full rank, and the noise matrix contains i.i.d standardized entries. Under our regularity condition (3.2.18) on the signal matrix, we prove the Tracy-Widom laws of the extreme eigenvalues under the necessary and sufficient condition (3.2.4).

The signal-plus-noise model considered in this thesis can be extended to models with more complicated noise structures. For example, it is very realistic that the entries of noise matrix $\mathbf{X}$ have different variances. The limiting spectral distribution of such model was derived in [Hachem et al. \(2007\)](#), and spectral properties of some similar models have been studied in literature. However, there is no result about the asymptotic behaviours of the largest eigenvalues of such model. Another research direction is to consider a spiked signal-plus-noise model where the signal matrix is also allowed to be full rank. If the largest eigenvalue of $\mathbf{R}\mathbf{R}^*$ is greater than some critical values ((3.2.18) is not satisfied), the largest eigenvalue will jump outside bulks of spectrum, demonstrated by simulations. It is of interest to know

the localizations and fluctuations of spiked sample eigenvalues motivated by some statistical problems such as estimating numbers of strong signals or matrix denoising. There has been a huge literature on these problems focusing on estimation of low rank signals and we hope to extend to large rank cases in the future.

Bibliography

- Alex, B., Erdős, L., Knowles, A., Yau, H.-T., and Yin, J. (2014). Isotropic local laws for sample covariance and generalized wigner matrices. *Electronic Journal of Probability*, 19. [119](#)
- Anderson, T. W. (1962). An introduction to multivariate statistical analysis. Technical report, Wiley New York. [1](#)
- Apple, D. (2017). Learning with privacy at scale. *Apple Machine Learning Journal*, 1(8). [71](#)
- Bai, Z. and Ding, X. (2012). Estimation of spiked eigenvalues in spiked models. *Random Matrices: Theory and Applications*, 1(02):1150011. [18](#)
- Bai, Z., Miao, B., Pan, G., et al. (2007). On asymptotics of eigenvectors of large sample covariance matrix. *The Annals of Probability*, 35(4):1532–1572. [36](#), [48](#), [50](#), [52](#)
- Bai, Z. and Silverstein, J. W. (1999). Exact separation of eigenvalues of large dimensional sample covariance matrices. *Annals of probability*, pages 1536–1555. [9](#), [60](#)
- Bai, Z. and Silverstein, J. W. (2010). *Spectral analysis of large dimensional random matrices*, volume 20. Springer. [1](#), [13](#), [48](#)
- Bai, Z. and Yao, J. (2012). On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis*, 106:167–177. [2](#), [7](#), [9](#), [12](#), [13](#), [39](#), [58](#)
- Bai, Z., Yao, J.-f., et al. (2008). Central limit theorems for eigenvalues in a spiked population model. In *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, volume 44, pages 447–474. Institut Henri Poincaré. [2](#), [7](#), [12](#), [13](#), [45](#)
- Bai, Z. D. and Silverstein, J. W. (2004). Clt for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.*, 32(1A):553–605. [3](#), [15](#), [27](#), [36](#), [39](#), [45](#), [46](#), [47](#), [48](#), [50](#), [57](#)
- Bai, Z.-D., Silverstein, J. W., et al. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *The Annals of Probability*, 26(1):316–345. [9](#), [39](#), [49](#), [58](#)

- Baik, J., Arous, G. B., and P, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697. [1](#)
- Baik, J., Lee, J. O., and Wu, H. (2018). Ferromagnetic to paramagnetic transition in spherical spin glass. *Journal of Statistical Physics*, 173(5):1484–1522. [3](#)
- Baik, J. and Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *Journal of multivariate analysis*, 97(6):1382–1408. [2](#)
- Banna, M., Najim, J., and Yao, J. (2020). A clt for linear spectral statistics of large random information-plus-noise matrices. *Stochastic Processes and their Applications*, 130(4):2250–2281. [4](#)
- Bao, Z., Ding, X., and Wang, K. (2018). Singular vector and singular subspace distribution for the matrix denoising model. *arXiv preprint arXiv:1809.10476*. [4](#)
- Bao, Z., Hu, J., Pan, G., Zhou, W., et al. (2019). Canonical correlation coefficients of high-dimensional gaussian vectors: finite rank case. *The Annals of Statistics*, 47(1):612–640. [5](#)
- Bao, Z., Pan, G., Zhou, W., et al. (2015). Universality for the largest eigenvalue of sample covariance matrices with general population. *The Annals of Statistics*, 43(1):382–421. [98](#)
- Benaych-Georges, F. and Knowles, A. (2016). Lectures on the local semicircle law for wigner matrices. *arXiv preprint arXiv:1601.04055*. [95](#), [110](#), [111](#), [112](#)
- Benaych-Georges, F. and Nadakuditi, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521. [4](#)
- Billingsley, P. (1995). Measure and probability. [26](#)
- Bloemendal, A., Erdos, L., Knowles, A., Yau, H.-T., and Yin, J. (2014). Isotropic local laws for sample covariance and generalized wigner matrices. *Electron. J. Probab*, 19(33):1–53. [111](#)
- Cai, T., Han, X., and Pan, G. (2017). Limiting laws for divergent spiked eigenvalues and largest non-spiked eigenvalue of sample covariance matrices. *arXiv preprint arXiv:1711.00217*. [3](#), [39](#)

- Cai, T., Liu, W., and Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association*, 108(501):265–277. [17](#), [22](#)
- Cape, J., Tang, M., and Priebe, C. E. (2019). Signal-plus-noise matrix models: eigenvector deviations and fluctuations. *Biometrika*, 106(1):243–250. [4](#)
- Chang, C.-I. and Du, Q. (1999). Interference and noise-adjusted principal components analysis. *IEEE transactions on geoscience and remote sensing*, 37(5):2387–2396. [4](#), [69](#)
- Chaudhuri, K., Sarwate, A. D., and Sinha, K. (2013). A near-optimal algorithm for differentially-private principal components. *Journal of Machine Learning Research*, 14. [71](#)
- Ding, X. (2017). Asymptotics of empirical eigen-structure for high dimensional sample covariance matrices of general form. *arXiv preprint arXiv:1708.06296*. [15](#)
- Ding, X. (2020). High dimensional deformed rectangular matrices with applications in matrix denoising. *Bernoulli*, 26(1):387–417. [4](#)
- Ding, X. and Yang, F. (2018). A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices. *Annals of Applied Probability*, 28(3):1679–1738. [5](#), [63](#), [134](#)
- Ding, X. and Yang, F. (2020a). Edge statistics of large dimensional deformed rectangular matrices. *arXiv preprint arXiv:2009.00389*. [5](#), [63](#)
- Ding, X. and Yang, F. (2020b). Tracy-widom distribution for the edge eigenvalues of gram type random matrices. *arXiv preprint arXiv:2008.04166*. [62](#), [63](#)
- Dozier, R. B. and Silverstein, J. W. (2007a). Analysis of the limiting spectral distribution of large dimensional information-plus-noise type matrices. *Journal of Multivariate Analysis*, 98(6):1099–1122. [4](#), [65](#), [66](#), [96](#), [107](#)
- Dozier, R. B. and Silverstein, J. W. (2007b). On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices. *Journal of Multivariate Analysis*, 98(4):678–694. [4](#), [65](#)

- Erdős, L., Knowles, A., and Yau, H.-T. (2013a). Averaging fluctuations in resolvents of random band matrices. In *Annales Henri Poincaré*, volume 14, pages 1837–1926. Springer. [75](#), [110](#)
- Erdős, L., Knowles, A., Yau, H.-T., and Yin, J. (2013b). The local semicircle law for a general class of random matrices. *Electronic Journal of Probability*, 18. [110](#)
- Erdős, L., Knowles, A., Yau, H.-T., and Yin, J. (2013c). Spectral statistics of erdős-rényi graphs i: Local semicircle law. *The Annals of Probability*, 41(3B):2279–2375. [80](#)
- Erdős, L., Yau, H.-T., and Yin, J. (2012). Rigidity of eigenvalues of generalized wigner matrices. *Advances in Mathematics*, 229(3):1435–1515. [5](#), [63](#), [79](#), [80](#), [85](#)
- Erlingsson, Ú., Pihur, V., and Korolova, A. (2014). Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 1054–1067. [71](#)
- Fan, Z. and Johnstone, I. M. (2017). Tracy-widom at each edge of real covariance and manova estimators. *arXiv preprint arXiv:1707.02352*. [5](#)
- Hachem, W., Loubaton, P., and Najim, J. (2007). Deterministic equivalents for certain functionals of large random matrices. *The Annals of Applied Probability*, 17(3):875–930. [4](#), [179](#)
- Hachem, W., Loubaton, P., Najim, J., and Vallet, P. (2013). On bilinear forms based on the resolvent of large random matrices. In *Annales de l’IHP Probabilités et statistiques*, volume 49, pages 36–63. [4](#)
- Han, X., Pan, G., and Yang, Q. (2018). A unified matrix model including both cca and f matrices in multivariate analysis: the largest eigenvalue and its applications. *Bernoulli*, 24(4B):3447–3468. [62](#), [134](#)
- Han, X., Pan, G., and Zhang, B. (2016). The tracy-widom law for the largest eigenvalue of f type matrices. *The Annals of Statistics*, 44(4):1564–1592. [5](#), [62](#)
- Jiang, D. and Bai, Z. (2018). Generalized four moment theorem and an application to clt for spiked eigenvalues of large-dimensional covariance matrices. *arXiv preprint arXiv:1808.05362*. [2](#), [7](#), [12](#), [13](#), [58](#)

- Johansson, K. (2000). Shape fluctuations and random matrices. *Communications in mathematical physics*, 209(2):437–476. [4](#)
- Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *The Annals of statistics*, 29(2):295–327. [1](#), [2](#), [4](#)
- Knowles, A. and Yin, J. (2013). The isotropic semicircle law and deformation of wigner matrices. *Communications on Pure and Applied Mathematics*, 66(11):1663–1749. [4](#), [144](#)
- Knowles, A. and Yin, J. (2017). Anisotropic local laws for random matrices. *Probability Theory and Related Fields*, 169(1-2):257–352. [62](#), [81](#), [119](#), [121](#), [122](#), [123](#)
- Landon, B. and Yau, H.-T. (2017). Edge statistics of dyson brownian motion. *arXiv preprint arXiv:1712.03881*. [5](#), [63](#)
- Lee, J. O. and Schnelli, K. (2013). Local deformed semicircle law and complete delocalization for wigner matrices with random potential. *Journal of Mathematical Physics*, 54(10):103504. [95](#), [110](#)
- Lee, J. O. and Schnelli, K. (2015). Edge universality for deformed wigner matrices. *Reviews in Mathematical Physics*, 27(08):1550018. [5](#), [61](#), [63](#), [81](#), [85](#), [147](#)
- Lee, J. O. and Schnelli, K. (2016). Tracy–widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *The Annals of Applied Probability*, 26(6):3786–3839. [61](#), [68](#), [81](#), [85](#), [87](#), [147](#)
- Lee, J. O. and Yin, J. (2014). A necessary and sufficient condition for edge universality of wigner matrices. *Duke Mathematical Journal*, 163(1):117–173. [5](#), [63](#), [134](#), [137](#), [140](#), [141](#)
- Li, J. and Chen, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *The Annals of Statistics*, 40(2):908–940. [17](#), [22](#)
- Li, Z., Han, F., Yao, J., et al. (2020). Asymptotic joint distribution of extreme eigenvalues and trace of large sample covariance matrix in a generalized spiked population model. *Annals of Statistics*, 48(6):3138–3160. [3](#), [12](#), [15](#)
- Loubaton, P., Vallet, P., et al. (2011). Almost sure localization of the eigenvalues in a gaussian information plus noise model. application to the spiked models. *Electronic Journal of Probability*, 16:1934–1959. [4](#), [65](#), [66](#), [96](#)

- Mani, V. V. and Bose, R. (2010). Direction of arrival estimation and beamforming of multiple coherent uwb signals. In *2010 IEEE International Conference on Communications*, pages 1–5. IEEE. [4](#)
- Mestre, X. (2008). Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Transactions on Information Theory*, 54(11):5113–5129. [15](#), [48](#), [49](#), [51](#)
- Montanari, A., Reichman, D., and Zeitouni, O. (2015). On the limitation of spectral methods: From the gaussian hidden clique problem to rank-one perturbations of gaussian tensors. *Advances in Neural Information Processing Systems*, 28:217–225. [72](#)
- Nadakuditi, R. R. (2014). Optshrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage. *IEEE Transactions on Information Theory*, 60(5):3002–3018. [3](#)
- Najim, J. and Yao, J. (2016). Gaussian fluctuations for linear spectral statistics of large random covariance matrices. *The Annals of Applied Probability*, 26(3):1837–1887. [3](#)
- Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77(5):1447–1479. [5](#), [69](#)
- Pan, G. (2014). Comparison between two types of large sample covariance matrices. In *Annales de l’IHP Probabilités et statistiques*, volume 50, pages 655–677. [3](#)
- Pan, G. and Zhou, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *The Annals of Applied Probability*, 18(3):1232–1270. [3](#), [14](#), [15](#), [17](#), [18](#), [21](#), [48](#)
- Pastor-Satorras, R. and Castellano, C. (2016). Distinct types of eigenvector localization in networks. *Scientific reports*, 6(1):1–9. [16](#)
- Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642. [2](#), [15](#)
- Paul, D. and Aue, A. (2014). Random matrix theory in statistics: A review. *Journal of Statistical Planning and Inference*, 150:1–29. [1](#)

- Pillai, N. S., Yin, J., et al. (2014). Universality of covariance matrices. *The Annals of Applied Probability*, 24(3):935–1001. [79](#), [111](#)
- Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N., Guhr, T., and Stanley, H. E. (2002). Random matrix approach to cross correlations in financial data. *Physical Review E*, 65(6):066126. [16](#)
- Svensson, A. and Jakobsson, A. (2011). Adaptive detection of a partly known signal corrupted by strong interference. *IEEE Signal Processing Letters*, 18(12):729–732. [69](#)
- Tao, T. and Vu, V. (2010). Random matrices: Universality of local eigenvalue statistics up to the edge. *Communications in Mathematical Physics*, 298(2):549–572. [5](#)
- Tracy, C. A. and Widom, H. (1994). Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics*, 159(1):151–174. [4](#)
- Tracy, C. A. and Widom, H. (1996). On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics*, 177(3):727–754. [4](#), [68](#)
- Vallet, P., Loubaton, P., and Mestre, X. (2012). Improved subspace estimation for multivariate observations of high dimension: the deterministic signals case. *IEEE Transactions on Information Theory*, 58(2):1043–1068. [3](#), [4](#)
- Vinogradova, J., Björnson, E., and Larsson, E. G. (2016). On the separability of signal and interference-plus-noise subspaces in blind pilot decontamination. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3421–3425. IEEE. [4](#)
- Xi, H., Yang, F., and Yin, J. (2017). Local circular law for the product of a deterministic matrix with a random matrix. *Electronic Journal of Probability*, 22. [119](#)
- Yang, D., Ma, Z., and Buja, A. (2016). Rate optimal denoising of simultaneously sparse and low rank matrices. *The Journal of Machine Learning Research*, 17(1):3163–3189. [3](#), [4](#)
- Yang, F. (2019). Edge universality of separable covariance matrices. *Electronic Journal of Probability*, 24. [62](#), [63](#), [133](#), [134](#), [137](#), [138](#)

- Yang, M., Lyu, L., Zhao, J., Zhu, T., and Lam, K.-Y. (2020). Local differential privacy and its applications: A comprehensive survey. *arXiv preprint arXiv:2008.03686*. [71](#)
- Yang, Q. and Pan, G. (2017). Weighted statistic in detecting faint and sparse alternatives for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 112(517):188–200. [17](#)
- Yao, J., Zheng, S., and Bai, Z. (2015). *Sample covariance matrices and high-dimensional data analysis*. Cambridge University Press Cambridge. [18](#)
- Zhang, Z., Zheng, S., Pan, G., and Zhong, P. (2020). Asymptotic independence of spiked eigenvalues and linear spectral statistics for large sample covariance matrices. *arXiv preprint arXiv:2009.11010*. [6](#)
- Zheng, S., Bai, Z., Yao, J., et al. (2015). Substitution principle for clt of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *The Annals of Statistics*, 43(2):546–591. [3](#)