Abstract

Random Matrices and Statistical Inference with Correlated Structures

Haoyu Wang

2025

Random matrices play a central role in modern probability theory and high-dimensional statistics, both as fundamental mathematical objects and as powerful tools for inference. This dissertation presents several works on random matrix theory and its connections to statistical inference with correlated structures in high dimensions.

On the random matrix side, we investigate the spectral properties of sample covariance matrices. Our approach is built upon the resolvent method, and we utilize local laws for various spectral analyses. Specifically, our contributions are three-fold: (1) we attain the first explicit rate of convergence for the largest eigenvalue of sample covariance matrices; (2) we derive the first quantitative estimate for the universality of the smallest singular value for random matrices at the hard edge; (3) we obtain the first sensitivity analysis for the principal component with respect to resampling of the data as the perturbation effect. These results deepen our understanding of the classical universality phenomenon.

On the statistical side, we study inference of latent structures and correlations from noisy data. In particular, we focus on random graph matching of geometric models and empirical Bayes estimation in high-dimensional linear models. For these statistical inference tasks, we establish both positive and negative results: we derive optimal information-theoretic limits and, in parallel, develop algorithms that achieve these limits whenever feasible. These results solve open problems in the fields and also provide new directions to explore in the future.

Random Matrices and Statistical Inference with Correlated Structures

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Haoyu Wang

Dissertation Director: Yihong Wu

December 2025

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my advisor, Prof. Yihong Wu. No words can fully capture how much I have learned from Yihong. Yihong is always an infinite source of inspiration and fantastic ideas, and his tireless and energetic work ethic always motivates me to push forward. His meticulous attention to detail and his pursuit of clarity in writing have profoundly shaped my own approach to research. I am especially grateful for his generosity with his time—always willing to engage in discussions and patient even when I asked naive questions or made mistakes. Beyond being an extraordinary research mentor, Yihong's humor and warm care have been a constant source of support throughout these years. Talking to Yihong and working with him are always a source of satisfaction for me and will be one of the most pleasant memories in my life.

I am deeply thankful to Prof. Zhou Fan and Prof. Yandi Shen for agreeing to serve on my dissertation reading committee. I learned a great deal from Zhou's classes on high-dimensional probability and statistics, which significantly shaped my research foundation. I also greatly appreciate Yandi's invaluable guidance during our collaboration. It is a true honor to have both of them as members of my committee.

I would like to thank my professors at the Courant Institute, where I made up my mind to pursue a PhD. I am particularly grateful to Prof. Paul Bourgade. Paul is my first mentor in research and showed me the beautiful world of random matrix theory. More importantly, Paul taught me how to be a researcher and helped me build confidence in my own potential. As a role model of great researcher, his attention to detail and lucid writing style had a huge influence on me. Even now, I am still benefiting from what Paul taught me. I also thank Prof. Scott Armstrong and Prof. Vlad Vicol for their mentorship. Scott generously shared his wisdom in probability and analysis with me, and he informed me of the significance of clear communication and presentation skills. Vlad mentored me on PDEs and offered generous help when I was confused and in need. Without them, it would be impossible for me to get the chance to do this PhD.

This thesis would not have been possible without the support of my dear friends. I am especially indebted to Yuqiu Fu and Jiaheng Wei for our long-standing friendship of over a decade. We have known each other since our teenage years in the Special Class for the Gifted Young and have walked together through many stages of life. Without their consistent encouragement, I would never have finished my Ph.D. studies. I thank them for bringing me numerous joyful memories and for their considerate listening and help when I have concerns. I sincerely hope we may continue to learn and grow together in the future, and I hope we may share the happiness and support each other in our careers. I also want to thank Yutong Nie for being a good friend at Yale. Her encouragement and thoughtful advice have cheered me up when I was upset. The meals and wines we shared will always be my precious memories.

Finally, I want to thank my parents, Hong Pan and Yong Wang, for their unwavering support behind me, for backing up all my decisions during my growth, for their selfless love, and for being the best possible parents in the world. I also thank my two adorable fluffy puppies, Tuan Tuan and Duo Duo, for curing my mind when I am upset and for helping me always have faith in the beauty of this world.

A toast to the tenacity and endeavor.

A toast to the ambitious dream in youth.

Dedicated to my parents Hong Pan and Yong Wang

The best parents in the world

Contents

Chapter 1

Overview

Over the past few decades, random matrix theory has emerged as an important subject in modern probability and it has been applied in a broad range of fields such as physics [Wig51, Wig55, Dys62a, Dys62b, Dys62c, GMGW98], economics [BP09, OM00, YP16], genetics [LZYZ06, Zho17, ABBR20], etc. In particular, the interplay between random matrix theory and high-dimensional statistics is becoming more and more important due to the development of machine learning and data science [DGNT11, PW17, PB17, MM21, CLMW11a, CCF$^+$21, CL22, CD11]. As the datasets grow in both size and complexity, classical statistical theory that either restricts to fixed-dimensional structures or focuses on normality assumption becomes inadequate. Random matrix theory provide powerful tools for high-dimensional statistical inference [Joh07, BS10a, CL22].

A central topic in random matrix theory is the spectral analysis of large random matrices [AGZ10, BS10a]. In particular, with a deep connection to statistics, the eigenstructure of large sample covariance matrices play a fundamental role in signal detection [Joh01, BBAP05, Pau07, BGN11, KN09], principal component analysis [JL09, FLW16, ZCW22, AFW22], and linear regression [DIC16, BHMM19, HMRT22].

Before diving into the focus of this dissertation, we first have a brief review of two important aspects of random matrix theory: (1) the universality phenomenon (2) classical application of random matrix theory in statistics.

1.1 Universality in random matrix theory

One of the central discovery in random matrix theory is the universality phenomenon: certain spectral behaviors of large-scale random matrices are insensitive to the microscopic distribution of the matrix entries. In particular, the famous Wigner-Dyson-Mehta conjecture asserts that the spectral behavior should only depend on the symmetry class of the matrix ensemble (this conjecture has been proved [TV12, EYY11, ESYY12, BYY14, BEYY16, BY17]). In this thesis, since our primary concern is random matrices related to statistics, we focus more on the sample covariance matrices of type $X^\top X$. As a warm-up, we briefly present three important universality results (the detailed version will be discussed in later chapters):

Empirical spectral distribution For a symmetric matrix $H \in \mathbb{R}^{N\times N}$, its empirical spectral distribution is defined as $\frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}$, where $\{\lambda_i\}_{i=1}^N$ are the eigenvalues of H. It is well known that for Wigner matrices, the empirical spectral distribution converge to the semicircle distribution and for sample covariance matrices the limit is the Marchenko-Pastur distribution [MP67]. Specifically, consider a large data matrix $X \in \mathbb{R}^{m\times n}$ with entries satisfying mean zero and variance 1, in which the dimensions m and n are proportional and both grow to infinity $\lim_{n\to\infty}\frac{m}{n} = \xi$ (for simplicity we assume $0 \le \xi \le 1$). With the eigenvalues of $H = \frac{1}{n}XX^\top$ denoted as $\lambda_1 \le \cdots \le \lambda_m$, its empirical spectral distribution converges to a deterministic limiting law

$$\frac{1}{m}\sum_{i=1}^m \delta_{\lambda_i} \Rightarrow \mu_{\mathsf{MP}}$$

where

$$d\mu_{\mathsf{MP}}(x) = \frac{1}{2\pi\xi}\frac{\sqrt{(\lambda_+ - x)(x-\lambda_-)}}{x}, \quad \lambda_\pm = (1\pm\sqrt{\xi})^2.$$

The Marchenko-Pastur limit is of fundamental importance for the spectral analysis of sample covariance matrices. For modern applications, a refined version of the Marchenko-Pastur law is proposed. This is called the local Marchenko-Pastur law [PY14, PY12, BEK$^+$14, BKYY16, KY17], which is stated in terms of the Stieltjes transform of the spectral distribution and quantifies the convergence on local microscopic scales. This quantitative measurement of spectral convergence is one of the most powerful tools in modern random matrix theory [EY17a, BGK17], and has been widely applied in various fields [FMWX20, FMWX22, CL22]. For a more detailed introduction to this local law, we defer it to the technical chapters (Chapter 2 and Chapter 3).

Tracy-Widom fluctuation The empirical spectral distribution aims to describe the global eigenvalue distribution. In some cases, the behavior of the top eigenvalue may be even more important. Another classical result in random matrix universality is the limiting law of the largest eigenvalue of random matrix ensembles. For the sample covariance matrices, it is well known that the fluctuation of the largest eigenvalue around the spectral edge (after appropriate rescaling) converges to a deterministic limit named Tracy-Widom distribution [Joh01, Ma12, PY14, BEK$^+$14, BKYY16, KY17, DY18]. This limit describes the concentration of the largest eigenvalue, and plays a vital role in many applications [BBAP05, MRZ15].

Eigenvector delocalization The eigenvectors also exhibit universal behavior for large random matrices. One of the key properties of eigenvectors is the delocalization phenomenon [PY14, BEK$^+$14, BKYY16, KY17, BY17], which means that the size of each component of the eigenvector is approximately the same. Roughly speaking, this means that the distribution of the eigenvector is isotropic on a sphere. As one of the most important property for large-scale random matrices, delocalization inspires the development of many spectral method in statistics [AKS98, MRZ15, FMWX20, FMWX22].

1.2 Random matrices in statistical inference

Random matrices naturally appear in various statistical inference tasks. As an introduction for how random matrix theory can be used in high-dimensional statistics, we present a simple example of signal detection in the signal-plus-noise model. Let $W \in \mathbb{R}^{N\times N}$ be a matrix whose elements are independent (up to symmetry) $\mathcal{N}(0, \frac{1}{N})$ variables. Suppose $x \in \mathbb{R}^N$ is a uniformly random vector on the unit sphere. Consider the following hypothesis testing problem:

$$H_0 : X = W \ \text{ vs } \ H_1 : X = \sqrt{\lambda} x x^\top + W.$$

This problem can be easily solved based on a classical result in random matrix theory called the BBP phase transition [BBAP05]. For the null model in H_0, we have the well-known semicircle law and we know that the largest eigenvalue λ_N of X concentrates around 2. On the other hand, when $\lambda > 1$, it was proved that there will be an outlier in the spectrum and the largest eigenvalue λ_N will be strictly greater than 2 with high probability. Therefore, using random matrix theory can easily solve the testing problem in this case.

Beyond this simple setting, random matrix theory provides powerful tools for various statistical tasks such as signal detection [Joh01, BBAP05, Pau07, BGN11, KN09], covariance estimation [LW03a, LW03b, LW04, Kar08, Nad14, LW17], and regression [DIC16, BHMM19, HMRT22, RR07, MP21]. The interplay between random matrix theory and high-dimensional statistics is a broad and profound topic. We do not aim to introduce every aspect of it, and we refer to [Joh07] for a more comprehensive review. In this dissertation, we will investigate how random matrices appear in two statistical problems that are less studied in previous literature: random graph matching and empirical Bayes estimation for linear models.

1.3 Organization

This thesis can be divided into two parts: the first part focus on the mathematical theory of random matrices; the second part focus on how random matrices appear in statistical models and discuss various aspect of the inference problems such as information-theoretic threshold and inference algorithms.

In Chapter 2, we study the quantitative edge universality for sample covariance matrices. We prove the first explicit rate of convergence to the Tracy-Widom distribution for the fluctuation of the largest eigenvalue of sample covariance matrices that are not integrable, and also derived the first quantitative universality for the smallest singular value. Our primary focus is matrices of type X^*X and the proof follows the Erdős-Schlein-Yau dynamical method. We use the dynamical appproach to analyze the Dyson Brownian motion and obtain a quantitative error estimate for the local relaxation flow at the edge. Together with a quantitative version of the Green function comparison theorem, this gives the rate of convergence.

In Chapter 3, we focus on eigenvectors of sample covariance matrices, which are closely related to the principal component analysis in statistics. The stability and sensitivity of statistical methods or algorithms with respect to their data is an essential problem in machine learning and statistics. One fundamental way to measure the stability of an algorithm is to study its performance under resampling of the data. In this chapter, we study the resampling sensitivity for the principal component analysis (PCA). When the population covariance matrix of the data does not have a strong spike (i.e. in the subcritical regime), we show that PCA is resampling sensitive by establishing a sharp threshold for the resampling strength, above which resampling even a negligible portion of the input may completely change the principal component; below this threshold, a moderate resampling has almost no effect on the output. In contrast, if the population covariance matrix possesses a strong spike, PCA will be stabilized by the planted signal. All of our results hold with universality, regardless of the underlying data distribution. Such sensitivity phenomenon has been discovered for Wigner matrices [BLZ20, BL22], and we generalize it to the case of PCA.

In Chapter 4, we focus on matching random graphs with latent geometric structures. This paper studies the problem of matching two complete graphs with edge weights correlated through latent geometries, extending a recent line of research on random graph matching with independent edge weights to geometric models. Specifically, given a random permutation π^* on $[n]$ and n iid pairs of correlated Gaussian vectors $\{X_{\pi^*(i)}, Y_i\}$ in $\mathbb{R}^d$ with noise parameter σ, the edge weights are given by $A_{ij} = \kappa(X_i, X_j)$ and $B_{ij} = \kappa(Y_i, Y_j)$ for some link function κ. The goal is to recover the hidden vertex correspondence π^* based on the observation of A and B. We focus on the dot-product model with $\kappa(x, y) = \langle x, y \rangle$ and Euclidean distance model with $\kappa(x, y) = \|x - y\|^2$, in the low-dimensional regime of $d = o(\log n)$ wherein the underlying geometric structures are most evident. We derive an approximate maximum likelihood estimator, which provably achieves, with high probability, perfect recovery of π^* when $\sigma = o(n^{-2/d})$ and almost perfect recovery with a vanishing fraction of errors when $\sigma = o(n^{-1/d})$. Furthermore, these conditions are shown to be information-theoretically optimal even when the latent coordinates $\{X_i\}$ and $\{Y_i\}$ are observed, complementing the recent results of [DCK19] and [KNW22] in geometric models of the planted bipartite matching problem.

In Chapter 5, we study the empirical Bayes estimation in high-dimensional linear models. Empirical Bayes is a powerful framework for large-scale inference, and its core idea is to estimate the prior distribution in the traditional Bayesian method via data-driven approaches and then apply the downstream Bayesian inference based on the estimated prior. Empirical Bayes is well-developed for sequence models, but it is challenging to adapt it to linear models. For the Bayesian linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where β_i are i.i.d. from some distribution g, we aim to learn the unknown prior distribution. The two most classical methods in statistics, maximum likelihood estimation and method of moments, both have technical difficulties when analyzing their statistical theory in linear models. Previous attempts mainly follow the nonparametric maximum likelihood estimation (NPMLE) approach [MSS23, FGSW23]. Despite making the first steps toward understanding the statistical guarantees, their works require strong assumptions and do now have a good understanding of the sample complexity. In our work, we rely on the idea of method of moments and design an estimation algorithm with polynomial complexity. We show the consistency of our estimator under mild conditions. In particular, our estimator only requires sublinear sample complexity and we also prove that this complexity is information-theoretically optimal.

Chapter 2

Quantitative Edge Universality of Sample Covariance Matrices

2.1 Introduction

The study of extremal eigenvalues is one of the most important questions in random matrix theory, not only because the eigenvalues themselves reveal essential properties of the matrix but also they play significant roles in developing spectral algorithms in statistics or maching learning. For high-dimensional large sample covariance matrices, it is well known that its empirical spectral distribution converges to the Marchenko-Pastur law [MP67] and its largest eigenvalue converges has a limiting Tracy-Widom distribution [Joh01].

The edge universality phenomenon has been discovered for more than a decade [Sos02, Ṕ09, PY14, BEK+14, BKYY16, KY17, DY18, TV10, CL19], but quantifying the rate of convergence to limiting laws remained challenging. Quantitative estimates for edge universality were only obtained for the Wishart ensemble [Joh00, EK06, Ma12]. Also, as noted in [CMS13], for large-scale sample covariance matrices with a asymptotically square data matrix (called the hard edge case in random matrix literature), its smallest eigenvalue will exhibit very different behavior from the Tracy-Widom limit, due to the singuarlity of the Markchenko-Pastur distribution at the origin. This makes quantifying the universality of the smallest eigenvalue/singular value even more difficult.

Following the framework proposed by Bourgade [Bou22], we obtain quantitative universality for the extremal eigenvalues of sample covariance matrices in both the soft edge case and the hard edge case [Wan24, Wan22]. We remark that the results we stated in this dissertation were chronologically the first, but improvements have been obtained in other works [SX21a, SX21b, SX22].

2.1.1 Model and Main Results

Let $X = (x_{ij})$ be an $M \times N$ data matrix with independent real valued entries with mean 0 and variance M^{-1},

$$x_{ij} = M^{-1/2} q_{ij}, \quad \mathbb{E}[q_{ij}] = 0, \quad \mathbb{E}[q_{ij}^2] = 1. \tag{2.1}$$

Furthermore, we assume the entries q_{ij} have a sub-exponential decay, that is, there exists a constant $\theta > 0$ such that for $u > 1$,

$$\mathbb{P}(|q_{ij}| > u) \leq \theta^{-1} \exp(-u^{\theta}). \tag{2.2}$$

This sub-exponential decay assumption is mainly for convenience, other conditions such as the finiteness of a sufficiently high moment would be enough. (For a necessary and sufficient condition for the edge universality we refer to [DY18].)

The sample covariance matrix corresponding to data matrix X is defined by $H := X^* X$. Throughout this paper, to avoid trivial eigenvalues, we will be working in the regime

$$\xi = \xi(N) := N/M, \quad \lim_{N \to \infty} \xi \in (0, 1) \text{ or } \xi \equiv 1.$$

We will mainly work with the rectangular case $0 < \xi < 1$, but will also show how to adapt the arguments to the square case $M \equiv N$. (The reason why we do not discuss the general case $\lim \xi = 1$ is merely technical due to the lack of local laws at the hard edge. In particular, the rigidity estimate at the hard edge is only known for a fixed $\xi \equiv 1$ but not for $\xi = \xi(N) \to 1$.)

We order the eigenvalues of H as $\lambda_1 \le \cdots \le \lambda_N$, and use λ_+ to denote the typical location of the largest eigenvalue (see (2.9) for the definition). For the main result of this paper, we consider the Kolmogorov distance

$$\mathsf{d}_{\mathsf{K}}(X, Y) := \sup_x |\mathbb{P}(X \le x) - \mathbb{P}(Y \le x)| \,.$$

Theorem 2.1. *Let H_N be sample covariance matrices satisfying (2.1) and (2.2). Let TW be the Tracy-Widom distribution. For any $\varepsilon > 0$, for large enough N we have*

$$\mathsf{d}_{\mathsf{K}}\left(\gamma N^{2/3}(\lambda_N - \lambda_+), \mathsf{TW}\right) \le N^{-\frac{2}{9}+\varepsilon},$$

where γ is an en explicit normalizing constant given by

$$\gamma = \xi^{7/6}(1 + \sqrt{\xi})^{-4/3}. \tag{2.3}$$

The null case X^*X is our primary concern in this paper, but quantitative estimates are also valid for general diagonal population matrices $X^*\Sigma X$ thanks to the comparison theorem for the Green function flow by Lee and Schnelli (see Section 2.6 for more details). Combining our quantitative edge universality for the null case (Theorem 2.1) with the Green function comparison by Lee-Schnelli (Proposition 2.4), we derive the rate of convergence to Tracy-Widom distribution for separable sample covariance matrices with general diagonal population.

Corollary 2.1. *Let $Q := X^*\Sigma X$ be an $N \times N$ separable sample covariance matrix, where X is an $M \times N$ real random matrix satisfying (2.1) and (2.2), and Σ is a real diagonal $M \times M$ matrix satisfying (2.80). Let μ_N be the largest eigenvalue of Q. For any $\varepsilon > 0$, for* *large enough N we have*

$$\mathsf{d}_{\mathsf{K}}\left(\gamma_0 N^{2/3}(\mu_N - E_+), \mathsf{TW}\right) \leq N^{-\frac{1}{138}+\varepsilon}, \tag{2.4}$$

where E_+ defined in (2.79) denotes the rightmost endpoint of the spectrum and γ_0 is a normalization constant defined in (2.81).

For the hard edge case (i.e. $\xi \equiv 1$), for technical reasons, we state the results in terms of singular values. For an $N \times N$ matrix X, let $\sigma_1(X) \leq \cdots \leq \sigma_N(X)$ denote the singular values in non-decreasing order and use $\kappa(X) := \sigma_N(X)/\sigma_1(X)$ to denote the condition number. Throughout this paper, we let G be an $N \times N$ Gaussian matrix with i.i.d. entries $\mathcal{N}(0, N^{-1})$.

The main results in the hard edge case are the following.

Theorem 2.2. *Let X be an $N \times N$ matrix satisfying (2.1) and (2.2). For any $\delta \in (0,1)$ and $\varepsilon > 0$, we have*

$$\begin{aligned}\mathbb{P}\left(N\sigma_1(G) > r + N^{-\delta}\right) - N^{\varepsilon}\left(N^{-1+\delta} \vee N^{-\frac{1}{2}}\right) &\leq \mathbb{P}\left(N\sigma_1(X) > r\right) \\ &\leq \mathbb{P}\left(N\sigma_1(G) > r - N^{-\delta}\right) + N^{\varepsilon}\left(N^{-1+\delta} \vee N^{-\frac{1}{2}}\right),\end{aligned} \tag{2.5}$$

where $a \vee b = \max\{a, b\}$ denotes the maximum between a and b.

For the smallest singular value, the complex-valued case is particularly interesting in the sense that the complex Gaussian model is explicitly integrable, i.e., the distribution of its smallest singular value is given by an exact formula. Specifically, let $G_{\mathbb{C}}$ be an $N \times N$ matrix whose entries are i.i.d complex Gaussians whose real and imaginary parts are i.i.d. copies of $\frac{1}{\sqrt{2}}\mathcal{N}(0, N^{-1})$. For the complex Gaussian ensemble, Edelman proved in [Ede88] that the distribution of the (renormalized) smallest singular value of a complex Gaussian ensemble is independent of N and can be computed explicitly

$$\mathbb{P}(N\sigma_1(G_{\mathbb{C}}) \leq r) = \int_0^r e^{-x}\mathrm{d}x = 1 - e^{-r}.$$

Thanks to this exact formula for the integrable model, the edge universality for the smallest singular value can be quantified in terms of the Kolmogorov-Smirnov distance to the explicit law.

More precisely, let $X_{\mathbb{C}}$ be an $N \times N$ random matrix satisfying

$$(X_{\mathbb{C}})_{ij} = N^{-1/2}x_{ij}, \quad \mathbb{E}x_{ij} = 0, \quad \mathbb{E}[(\operatorname{Re} x_{ij})^2] = \mathbb{E}[(\operatorname{Im} x_{ij})^2] = \frac{1}{2}, \quad \mathbb{E}[(\operatorname{Re} x_{ij})(\operatorname{Im} x_{ij})] = 0.$$

and the sub-exponential decay assumption (2.2). Then we have the following rate of convergence to the explicit law.

Corollary 2.2. *Let $X_{\mathbb{C}}$ be a complex $N \times N$ random matrix defined as above, then for any $\varepsilon > 0$ we have*

$$\mathbb{P}\left(N\sigma_1(X_{\mathbb{C}}) \leq r\right) = 1 - e^{-r} + O(N^{-\frac{1}{2}+\varepsilon}). \tag{2.6}$$

2.1.2 Outline of Proofs

The central idea of this paper is based on the Erdős-Schlein-Yau dynamical approach in random matrix theory. In their seminal work [ESYY12], the so-called *three-step strategy* is developed to prove universality phenomena for random matrices. Roughly speaking, this framework is the following three ideas.

(i) **A priori estimates for spectral statistics**. This is based on the analyzing the resolvent of the matrix, and such analysis is called local law in random matrix theory. Local law states that the spectral density converges to the limiting law on microscopic scale. This local law implies the eigenvalue rigidity phenomenon, which states that the eigenvalues are close to their typical locations. Such a priori control of the eigenvalue locations will play a significant role in further analysis.

(ii) **Local relaxation of eigenvalues**. This step is designed to prove universality for matrices with a tiny Gaussian component. We perturb the matrix by some independent Gaussian noise, and then under this perturbation, the dynamics of the eigenvalues is governed by the Dyson Brownian motion (DBM). Moreover, the spectral distribution of the Gaussian ensemble is the equilibrium measure of DBM. The ergodicity of DBM results in a fast convergence to the local equilibrium, and hence implies the universality for matrix with small Gaussian perturbation.

(iii) **Density arguments**. For any probability distribution of the matrix elements, there exists a distribution with a small Gaussian component (in the sense of Step (ii)) such that the two associated random matrices have asymptotically identical spectral statistics. Typically, such an asymptotic identity is guaranteed by some moment matching conditions and the comparison of resolvents.

For a systematic discussion of this method, we refer to the monograph [EY17a]. Following this strategy, our main techniques can also be divided into the following three steps.

- The first step is the local semicircle law for the symmetrization of the random matrix X. This local law guarantees the optimal rigidity estimates for the singular values. This step is be based on classical works in random matrix theory such as [BEK$^+$14, AEK14b, AEK17, PY14].

- The second step is to interpolate the general matrix X with the Gaussian matrix G, and estimate the dynamics of the singular value. More specifically, we consider the interpolation $X_t = e^{-t/2}X + \sqrt{1-e^{-t}}G$, which solves the matrix Ornstein-Uhlenbeck stochastic differential equation

$$\mathrm{d}X_t = \frac{1}{\sqrt{N}}\mathrm{d}B_t - \frac{1}{2}X_t\mathrm{d}t.$$

Note that this interpolation X_t is equivalent to the matrix perturbation in our smoothed analysis. We consider a weighted Stieltjes transform (defined in (2.21)). A key innovation of our work is that, combined with a symmetrization trick, the evolution of the weighted Stieltjes along the dynamics of H_t satisfies a stochastic PDE that can be well approximated by a deterministic advection equation. This deterministic PDE yields a rough estimate for $|\sigma_k(X_t)-\sigma_k(G)|$. Finally, using a delicate bootstrap argument, we show that the estimates for $|\sigma_k(X_t) - \sigma_k(G)|$ are self-improving. Iterating the bootstrap argument to optimal scale, we derive the optimal smoothed analysis for the smallest singular value.

- The last step is a quantitative resolvent comparison. In particular, the difference between the resolvent of two different random matrices are explicitly controlled in terms of the difference of their fourth moments. This comparison is proved via the Lindeberg exchange method. Together with the optimal smoothed analysis, this comparison theorem establishes the quantitative universality.

2.1.3 Notations

We introduce two important probabilistic notions that are commonly used throughout this Chapter.

Definition 2.1 (Overwhelming probability)**.** *Let $\{\mathcal{E}_N\}$ be a sequence of events. We say that $\mathcal{E}_N$ holds with overwhelming probability if for any (large) $D > 0$, there exists $N_0(D)$ such that for all $N \geq N_0(D)$ we have*

$$\mathbb{P}(\mathcal{E}_N) \geq 1 - N^{-D}.$$

Definition 2.2 (Stochastic domination)**.** *Let $\mathcal{X} = \{\mathcal{X}_N\}$ and $\mathcal{Y} = \{\mathcal{Y}_N\}$ be two families of nonnegative random variables. We say that $\mathcal{X}$ is stochastically dominated by $\mathcal{Y}$ if for all (small) $\varepsilon > 0$ and (large) $D > 0$, there exists $N_0(\varepsilon, D) > 0$ such that for $N \geq N_0(\varepsilon, D)$*

we have

$$\mathbb{P}\left(\mathcal{X}_N > N^{\varepsilon}\mathcal{Y}_N\right) \leq N^{-D}.$$

The stochastic domination is always uniform in all parameters. If $\mathcal{X}$ is stochastically dominated by $\mathcal{Y}$, we use the notation $\mathcal{X} \prec \mathcal{Y}$.

We also denote

$$\varphi = e^{C_0(\log\log N)^2}$$

a subpolynomial error parameter, for some fixed $C_0 > 0$. This constant C_0 is chosen large enough so that the eigenvalues (and singular values) rigidity and the local Marchenko-Pastur law (see next section for more details) holds. This parameter is essentially a factor of N^{ε} for an arbitrarily fixed $\varepsilon > 0$, and we present some results with φ for a pursuit of better accuracy.

2.2 Preliminaries

2.2.1 Dyson Brownian motion of Gram matrices

Let B be an $M \times N$ real matrix Brownian motion: B_{ij} are independent standard Brownian motions. We define the $M \times N$ matrix X_t by

$$X_t = X_0 + \frac{1}{\sqrt{N}} B_t.$$

The eigenvalue dynamics for the real Wishart process $H_t := X_t^* X_t$ was first derived in [Bru89]. Under our normalization convention, the equation is in the following form given in [BY17, Appendix C]

$$\mathrm{d}\lambda_k = 2\sqrt{\lambda_k}\frac{\mathrm{d}B_{kk}}{\sqrt{N}} + \left(\frac{M}{N} + \frac{1}{N}\sum_{\ell \neq k} \frac{\lambda_k + \lambda_\ell}{\lambda_k - \lambda_\ell}\right)\mathrm{d}t. \tag{2.7}$$

For technical convenience, when applying the coupling method from [BEYY16] to analyze the dynamics (2.7), focusing on the singular value will be a more suggested way so that we can eliminate the coefficient $\sqrt{\lambda_k}$ in the martingale part.

For $1 \leq k \leq N$, let $s_k := \sqrt{\lambda_k}$ denote the singular values of X. The Dyson Brownian motion for singular values dynamics of such sample covariance matrices is the following Ornstein-Uhlenbeck process [ESYY12, equation (5.8)].

$$\mathrm{d}s_k = \frac{\mathrm{d}B_k}{\sqrt{N}} + \left[-\frac{1}{2\xi} s_k + \frac{1}{2}\left(\frac{1}{\xi} - 1\right)\frac{1}{s_k} + \frac{1}{2N}\sum_{\ell \neq k}\left(\frac{1}{s_k - s_\ell} + \frac{1}{s_k + s_\ell}\right)\right]\mathrm{d}t.$$

An important idea in this paper is the following symmetrization trick (see [CL19, equation (3.9)]):

$$s_{-i}(t) = -s_i(t), \quad B_{-i}(t) = -B_i(t), \quad \forall t \geq 0,\ 1 \leq i \leq N.$$

From now we label the indices from -1 to $-N$ and 1 to N, so that the zero index is omitted. Unless otherwise stated, this will be the convention and we will not emphasize it explicitly. After symmetrization, the dynamics turns to the following form

$$\mathrm{d}s_k = \frac{\mathrm{d}B_k}{\sqrt{N}} + \left[-\frac{1}{2\xi} s_k + \frac{1}{2}\left(\frac{1}{\xi} - 1\right)\frac{1}{s_k} + \frac{1}{2N}\sum_{\ell \neq \pm k}\frac{1}{s_k - s_\ell}\right]\mathrm{d}t,$$
$$-N \leq k \leq N, k \neq 0. \quad (2.8)$$

2.2.2 Local law and spectral rigidity

The local law and rigidity estimates are classical results for the eigenvalues of sample covariance matrices. In this section, for later use, we rephrase these results into the corresponding version in terms of singular values.

It is well known that the empirical measure of the eigenvalues converges to the Marchenko-Pastur distribution

$$\rho_{\mathsf{MP}}(x) = \frac{1}{2\pi\xi}\sqrt{\frac{[(x-\lambda_-)(\lambda_+ - x)]_+}{x^2}},$$

where

$$\lambda_{\pm} = (1 \pm \sqrt{\xi})^2. \tag{2.9}$$

Define the typical locations of the singular values:

$$\gamma_k := \inf \left\{ E > 0 : \int_{-\infty}^{E^2} \rho_{\mathsf{MP}}(x)\mathrm{d}x \geq \frac{k}{N} \right\}, \quad 1 \leq k \leq N.$$

Following the symmetrization trick, we also define $\gamma_{-k} = -\gamma_k$. By a change of variable, it is easy to check that

$$\int_{-\infty}^{\gamma_k} \rho(x)\mathrm{d}x = \frac{N+k}{2N}, \quad \int_{-\infty}^{\gamma_{-k}} \rho(x)\mathrm{d}x = \frac{N-k}{2N}, \tag{2.10}$$

where $\rho(x)$ is the counterpart of Marchenko-Pastur law for singular values, defined by

$$\rho(x) = \frac{1}{2\pi\xi}\sqrt{\frac{[(x^2-\lambda_-)(\lambda_+ - x^2)]_+}{x^2}}, \quad \sqrt{\lambda_-} \leq |x| \leq \sqrt{\lambda_+}. \tag{2.11}$$

Denote $s_1 \leq \cdots \leq s_N$ the singular values of the data matrix X, and extend the singular values following the symmetrization trick by $s_{-k} = -s_k$. For $z = E + \mathrm{i}\eta \in \mathbb{C}$ with $\eta > 0$, let $m_N(z)$ and $S_N(z)$ denote the Stieltjes transform of the empirical measure of the (symmetrized) singular values and eigenvalues, respectively:

$$m_N(z) := \frac{1}{2N} \sum_{-N \leq k \leq N} \frac{1}{s_k - z}, \quad S(z) := \frac{1}{N}\sum_{k=1}^{N} \frac{1}{\lambda_k - z}.$$

As mentioned previously, in the summation from $-N$ to N the 0 index is always excluded. Note that due to the symmetrization, this is equivalent to

$$m_N(z) = \frac{1}{2N}\sum_{k=1}^{N}\left(\frac{1}{s_k - z} + \frac{1}{-s_k - z}\right) = \frac{1}{N}\sum_{k=1}^{N}\frac{z}{s_k^2 - z^2} = zS(z^2). \tag{2.12}$$

On the other hand, use $m_{\mathsf{MP}}(z)$ to denote the Stieltjes transform of the Marchenko-Pastur law

$$m_{\mathsf{MP}}(z) := \int_{\mathbb{R}} \frac{\rho_{\mathsf{MP}}(x)}{x-z}\mathrm{d}x = \frac{1-\xi-z+\sqrt{(z-\lambda_-)(z-\lambda_+)}}{2\xi z},$$

where $\sqrt{}$ denotes the square root on the complex plane whose branch cut is the negative real line. With this choice we always have $\operatorname{Im} m_{\mathsf{MP}}(z) > 0$ when $\operatorname{Im} z > 0$. For the singular values, recall the limit distribution $\rho(x)$ for the empirical measure and use $m(z)$ to denote its corresponding Stieltjes transform

$$m(z) := \int_{\mathbb{R}} \frac{\rho(x)}{x-z}\mathrm{d}x = \int_{\mathbb{R}} \frac{1}{x-z}\frac{1}{2\pi\xi}\sqrt{\frac{[(x^2-\lambda_-)(\lambda_+-x^2)]_+}{x^2}}\mathrm{d}x.$$

We have the following relation between $m(z)$ and $m_{\mathsf{MP}}(z)$

$$\begin{aligned} m(z) &= \int_{\sqrt{\lambda_-}}^{\sqrt{\lambda_+}} \left(\frac{1}{x-z} - \frac{1}{x+z}\right)\frac{1}{2\pi\xi}\sqrt{\frac{[(x^2-\lambda_-)(\lambda_+-x^2)]_+}{x^2}}\mathrm{d}x \\ &= \int_{\mathbb{R}} \frac{z}{x-z^2}\rho_{\mathsf{MP}}(x)\mathrm{d}x = z m_{\mathsf{MP}}(z^2). \end{aligned} \tag{2.13}$$

It is well known that we have the strong local Marchenko-Pastur law [BEK+14, PY14] for $S(z)$. Denote by $\kappa(z)$ the distance of z to the spectral edges $\lambda_\pm$, i.e.,

$$\kappa(z) := \min\left\{\left|z-\sqrt{\lambda_-}\right|, \left|z-\sqrt{\lambda_+}\right|\right\},$$

For any fixed $\omega \in (0,1)$, define the domain

$$\mathbf{S} = \mathbf{S}(\omega, N) := \left\{z = E + \mathrm{i}\eta \in \mathbb{C} : \kappa(z) \le \omega^{-1}, N^{-1+\omega} \le \eta \le \omega^{-1}.|z| \ge \omega\right\}.$$

For any $D > 0$, there exists $N_0(D) > 0$ such that for every $N \ge N_0$ the following holds uniformly in $z \in \mathbf{S}$,

$$\mathbb{P}\left(|S(z) - m_{\mathsf{MP}}(z)| \le \frac{\varphi}{N\eta}\right) > 1 - N^{-D}. \tag{2.14}$$

By the relations (2.12) and (2.13), we know that

$$|m_N(z) - m(z)| = \left|z\left(S(z^2) - m_{\mathsf{MP}}(z^2)\right)\right| \le |z|\left|S(z^2) - m_{\mathsf{MP}}(z^2)\right|.$$

Combining with the strong local Marchenko-Pastur law (2.14), this gives us

$$\mathbb{P}\left(|m_N(z) - m(z)| \lesssim \frac{\varphi}{N\eta}\right) > 1 - N^{-D}. \tag{2.15}$$

We remark that in the square case $\xi \equiv 1$, the Stieltjes transform $m(z)$ reduces to the Stieltjes transform of the semicircle law $m_{\mathsf{sc}}(z) = \frac{-z+\sqrt{z^2-4}}{2}$.

For the rigidity estimates, a key observation is that the critical case $\xi = 1$ is significantly different from other cases. This is because the Marchenko-Pastur law ρ_{MP} has a singularity at the point $x = 0$ in this situation. When $\xi < 1$, the rigidity of singular values can be easily obtained from the analogous estimates for eigenvalues (see [PY14]). Let $\widehat{k} := \min(k, N+1-k)$, for any $D > 0$ there exists $N_0(D)$ such that the following holds for any $N \ge N_0$,

$$\mathbb{P}\left(|s_k - \gamma_k| \le \varphi^{\frac{1}{2}} N^{-\frac{2}{3}} (\widehat{k})^{-\frac{1}{3}} \text{ for all } k \in [1, N]\right) > 1 - N^{-D}. \tag{2.16}$$

For the critical case $\xi = 1$, now the Marchenko-Pastur distribution is supported on $[0, 4]$ and is given by $\rho_{\mathsf{MP}}(x) = \frac{1}{2\pi}\sqrt{(4-x)/x}$. A key observation is that the scales of eigenvalue spacings are different at the two edges. Due to this phenomenon, we use the following two different results, depending on the location in the spectrum.

On the one hand, the Marchenko-Pastur distribution still behaves like a square root near the soft edge $x = 4$, which implies that the result is the same as the rectangular case. The rigidity estimate near the soft edge can be easily adapted from the result for eigenvalues in

[BEK+14, Theorem 2.10], i.e. for some (small) $\omega > 0$ and any $\varepsilon > 0$ we have

$$\mathbb{P}\left(|s_k - \gamma_k| \le N^{\varepsilon}(N-k+1)^{-\frac{1}{3}}N^{-\frac{2}{3}} \text{ for all } k \in [(1-\omega)N, N]\right) > 1 - N^{-D}. \quad (2.17)$$

On the other hand, as explained in [CMS13], at the hard edge $x = 0$ the typical distance between eigenvalues and the edge is of order N^{-2}, which is much smaller than the typical distance between neighbouring eigenvalues in the bulk (or at the soft edge). Note that in this situation, the measure for the symmetrized singular values coincides with the standard semicircle law, that is $\rho(x) = \frac{1}{2\pi}\sqrt{(4-x^2)_+}$. By the relation (2.10), this means that the typical k-th singular value γ_k of an $N \times N$ data matrix shares the same position with the typical $(N+k)$-th eigenvalue of a $2N \times 2N$ generalized Wigner matrix. The link between these two models can be illustrated by the symmetrization trick: Define the $2N \times 2N$ matrix

$$\widetilde{H} = \begin{pmatrix} 0 & X^* \\ X & 0 \end{pmatrix},$$

then we know that the eigenvalues of $\widetilde{H}$ are precisely the symmetrized singular values of X^*. Note that we have $\widetilde{H} = \widetilde{H}^*$, $\mathbb{E}\widetilde{H}_{ij} = 0$ and $\sum_{i=1}^{2N} \mathbb{E}\widetilde{H}_{ij}^2 = 1$ for every $j \in [1, 2N]$. This shows that $\widetilde{H}$ is indeed a Wigner-type matrix except the lack of nondegeneracy condition caused by the zero blocks. By considering the matrix of this type, the rigidity at the hard edge can be proved directly from [AEK17, Theorem 2.7]

$$\mathbb{P}\left(|s_k - \gamma_k| \le N^{-1+\varepsilon} \text{ for all } k \in [1, (1-\omega)N]\right) > 1 - N^{-D}. \quad (2.18)$$

2.2.3 Auxliary results

To make this Chapter self-contained, we collect some well-known results that are used in the Chapter.

The first result is about controlling the size of a martingale. This is from [SW09, Appendix B.6, Equation (18)].

Lemma 2.1. *For any continuous martingale M and any $\lambda, \mu > 0$, we have*

$$\mathbb{P}\left(\sup_{0\le u\le t} |M_u| \ge \lambda, \langle M\rangle_t \le \mu\right) \le 2e^{-\frac{\lambda^2}{2\mu}}.$$

The second result is the Helffer-Sjőstrand formula, which is a classical result in functional calculus. This formula is used in Proposition 2.3 to compute the trace of functions via the Stieltjes transform. We are using the version in [EY17a, Section 11.2].

Lemma 2.2 (Helffer-Sjőstrand formula)**.** *Let $f \in C^1(\mathbb{R})$ with compact support and let $\chi(y)$ be a smooth cutoff function with support in $[-1, 1]$, with $\chi(y) = 1$ for $|y| \le \frac{1}{2}$ and with bounded derivatives. Then*

$$f(\lambda) = \frac{1}{\pi}\int_{\mathbb{R}^2} \frac{\mathrm{i}yf''(x)\chi(y) + \mathrm{i}(f(x) + \mathrm{i}f'(x))\chi'(y)}{\lambda - x - \mathrm{i}y}\mathrm{d}x\mathrm{d}y.$$

We also have the following resolvent expansion identity. This is a well-known result in linear algebra, and it is used in Proposition 2.3 to compare the resolvents of two matrices.

Lemma 2.3 (Resolvent expansion)**.** *For any two matrices A and B, we have*

$$(A + B)^{-1} = A^{-1} - (A + B)^{-1}BA^{-1}$$

provided that all the matrix inverses exist.

Finally, we have some estimates for the Stieltjes transform of the semicircle law. For $z = E + \mathrm{i}\eta$ with $\eta > 0$, recall that $m_{\mathsf{sc}}(z)$ denotes the Stieltjes transform of the semicircle distribution. The following estimates are well known in random matrix theory (see e.g. [EY17a, Lemma 6.2]).

Lemma 2.4. *We have for all $z = E + \mathrm{i}\eta$ with η that*

$$|m_{\mathsf{sc}}(z)| = |m_{\mathsf{sc}}(z) + z|^{-1} \leq 1.$$

Furthermore, there is a constant $c > 0$ such that for $E \in [-10, 10]$ and $\eta \in (0, 10]$ we have

$$c \leq |m_{\mathsf{sc}}(z)| \leq 1 - c\eta, \;\; |1 - m_{\mathsf{sc}}^2(z)| \sim \sqrt{\kappa(E) + \eta},$$

as well as

$$\operatorname{Im} m_{\mathsf{sc}}(z) \sim \begin{cases} \sqrt{\kappa(E) + \eta} & \text{if } |E| \leq 2, \\ \frac{\eta}{\sqrt{\kappa(E)+\eta}} & \text{if } |E| \geq 2, \end{cases}$$

where $\kappa(E) := ||E| - 2|$ is the distance of E to the spectral edge.

2.3 Singular Value Dynamics

Following the three-step strategy, a key component of our proof is quantifying the universality of Gaussian divisible ensembles, which leads us to analyze the Dyson Brownian motion.

2.3.1 Stochastic advection equation

The analysis of singular value dynamics relies on comparing the singular values of a general matrix and the Gaussian ensemble. This comparison follows the coupling method developed in [BEYY16, LSY19].

Consider the interpotation between a general matrix X and a Gaussian matrix G. Let $\{\sigma_k(X)\}_{k=-N}^{N}$ and $\{\sigma_k(G)\}_{k=-N}^{N}$ be the (symmetrized) singular values of X and G, respectively. For $\nu \in [0, 1]$, define

$$s_k^{(\nu)}(0) = (1 - \nu)\sigma_k(X) + \nu\sigma_k(G).$$

With this initial condition, we denote the unique solution of (2.8) by $\{s_k^{(\nu)}(t)\}$. Also, let $\{\sigma_k(X,t)\}$ and $\{\sigma_k(G,t)\}$ denote the solutions of (2.8) with initial conditions $\{\sigma_k(X)\}$ and $\{\sigma_k(G)\}$, respectively.

The rigidity estimates (2.16), (2.17), (2.18) hold uniformly (with respect to both ν and t) for the interpolation dynamics. For any fixed $\varepsilon > 0$, denote $\widehat{k} := min(k, N+1-k)$ and consider the set of good trajectories

$$\begin{aligned} &\mathscr{A}_\varepsilon \\ &= \begin{cases} \left\{\left|s_k^{(\nu)}(t) - \gamma_k\right| < N^{-\frac{2}{3}+\varepsilon}(\widehat{k})^{-\frac{1}{3}} \text{ for } 0 \le t \le 1, |k| \le N, 0 \le \nu \le 1\right\}, \text{ if } \xi \not\equiv 1 \\ \left\{\left|s_k^{(\nu)}(t) - \gamma_k\right| < N^{-\frac{2}{3}+\varepsilon}(N+1-|k|)^{-\frac{1}{3}} \text{ for } 0 \le t \le 1, |k| \le N, 0 \le \nu \le 1\right\}, \text{ if } \xi \equiv 1 \end{cases} \end{aligned} \tag{2.19}$$

and let $\mathscr{A}$ denote the set similarly as in $\mathscr{A}_\varepsilon$ but with N^ε replaced by φ^C. Rigidity estimates for fixed t and $\nu = 0$ or 1 were proved in [AEK14a, AEK17, BEK+14, BYY14, CMS13].The extension to uniform estimates in parameters can be done by a discretization argument: (1) discretize in t and ν; (2) use weyl's inequality to control increments over small time intervals; (3) use a maximum principle for the derivative with respect to the ν parameter (see Lemma 2.6) to control increments in small ν-intervals. As a consequence, we have

Lemma 2.5. *For any $\varepsilon > 0$, the event $\mathscr{A}_\varepsilon$ happens with overwhelming probability, i.e. for any $D > 0$, there exists $N_0(\varepsilon, D)$ such that for any $N > N_0$ we have*

$$\mathbb{P}(\mathscr{A}_\varepsilon) > 1 - N^{-D}.$$

We consider

$$\varphi_k^{(\nu)}(t) := e^{\frac{t}{2\xi}} \frac{\mathrm{d}}{\mathrm{d}\nu} s_k^{(\nu)}(t).$$

For the simplicity of notations, we omit the parameter ν if the context is clear. Then φ_k

satisfies the following non-local parabolic type equation.

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi_k = \frac{1}{2}\left(1-\frac{1}{\xi}\right)\frac{\varphi_k}{s_k^2} + \frac{1}{2N}\sum_{\ell\neq\pm k}\frac{\varphi_\ell-\varphi_k}{(s_\ell-s_k)^2} \tag{2.20}$$

Let $\psi_k = \psi_k^{(\nu)}$ solve the same equation as φ_k in (2.20) but with initial condition $\psi_k(0) = |\varphi_k(0)| = |\sigma_k(X)-\sigma_k(G)|$.

Lemma 2.6. *For all $t \geq 0$ and $-N \leq k \leq N$, we have*

$$\psi_k(t) = \psi_{-k}(t), \quad \psi_k(t) \geq 0, \quad |\psi_k(t)| \leq \max_{-N\leq\ell\leq N}|\psi_\ell(0)|, \quad |\varphi_k(t)| \leq \psi_k(t).$$

Proof. The first claim can be checked directly. Note that for the coefficients in the summation part of equation (2.20) we have $\frac{1}{(s_\ell-s_k)^2} > 0$. Therefore, for $f(t) := \min_k \psi_k(t)$, we have

$$f'(t) \geq \frac{1}{2}\left(1-\frac{1}{\xi}\right)\frac{1}{s_1^2}f(t).$$

Combined with the fact $\psi_k(0) \geq 0$, this gives us the first claim. For the second claim, since ψ_k's are nonnegative, we know that $\frac{\mathrm{d}}{\mathrm{d}t}\max_k \psi_k \leq 0$ and this yields the desired result. The third claim follows from linearity. We know that both $\psi+\varphi$ and $\psi-\varphi$ satisfy the equation (2.20), and we also have $(\psi+\varphi)(0) \geq 0$ and $(\psi-\varphi)(0) \geq 0$. Similarly to the first claim, this gives us $\psi_k(t)+\varphi_k(t) \geq 0$ and $\psi_k(t)-\varphi_k(t) \geq 0$, which completes the proof. $\square$

We consider the following weighted Stieltjes transform

$$\mathfrak{S}_t(z) := e^{-\frac{t}{2}}\sum_{-N\leq k\leq N}\frac{\varphi_k(t)}{s_k(t)-z}, \quad \widetilde{\mathfrak{S}}_t(z) := e^{-\frac{t}{2}}\sum_{-N\leq k\leq N}\frac{\psi_k(t)}{s_k(t)-z}. \tag{2.21}$$

Let $S_t(z)$ denote the Stieltjes transforms of the empirical measure for the singular values

$$S_t(z) = \frac{1}{2N}\sum_{-N\leq k\leq N}\frac{1}{s_k-z}.$$

Then the observables $\mathfrak{S}_t(z)$ and $\widetilde{\mathfrak{S}}_t(z)$ satisfy the following stochastic advection equation

Lemma 2.7. *For any* $\operatorname{Im} z \neq 0$, *we have*

$$\begin{aligned}
\mathrm{d}\widetilde{\mathfrak{S}}_t = {} & \left(S_t(z) + \frac{z}{2\xi}\right)(\partial_z \widetilde{\mathfrak{S}}_t)\mathrm{d}t + \frac{1}{4N}(\partial_{zz}\widetilde{\mathfrak{S}}_t)\mathrm{d}t + \left[\frac{e^{-\frac{t}{2\xi}}}{2N}\sum_{-N\le k\le N}\frac{\psi_k}{(x_k-z)^2(x_k+z)}\right]\mathrm{d}t \\
& + \left[\left(1-\frac{1}{\xi}\right)e^{-\frac{t}{2\xi}}\left(\sum_{-N\le k\le N}\frac{3z\psi_k}{2x_k^2(x_k-z)(x_k+z)}\right.\right. \\
& \left.\left. + \sum_{-N\le k\le N}\frac{z^3\psi_k}{x_k^2(x_k-z)^2(x_k+z)^2}\right)\right]\mathrm{d}t - \frac{e^{-\frac{t}{2\xi}}}{\sqrt{N}}\sum_{-N\le k\le N}\frac{\psi_k}{(x_k-z)^2}\mathrm{d}B_k.
\end{aligned} \tag{2.22}$$

Recall that the Stieltjes transform of the empirical measure for singular values satisfies the local law (2.15), so that the leading term of the stochastic differential equation (2.22) satisfied by $\widetilde{\mathfrak{S}}_t$ is close to

$$\begin{aligned}
\frac{z}{2\xi} + z m_{\mathsf{MP}}(z^2) &= \frac{z}{2\xi} + \frac{1-\xi-z^2+\sqrt{(z^2-\lambda_-)(z^2-\lambda_+)}}{2\xi z} \\
&= \frac{(1-\xi)+\sqrt{(z^2-\lambda_-)(z^2-\lambda_+)}}{2\xi z}.
\end{aligned}$$

Thus, the dynamics can be approximated by the following advection equation

$$\partial_t r = \frac{(1-\xi)+\sqrt{(z^2-\lambda_-)(z^2-\lambda_+)}}{2\xi z}\partial_z r. \tag{2.23}$$

In order to estimate the evolution of the observable, we analyze its dynamics (2.22) by studying the characteristics of the approximate advection PDE (2.23), similarly to [HL19, Bou22]. To do this, we first need some bounds on the shape of the characteristics $(z_t)_{t\ge 0}$, and some estimates for the initial value. We mainly focus on the case $\xi \neq 1$, and the adaptation to the square case $\xi \equiv 1$ is straightforward.

Recall that

$$\kappa(z) = \min\left\{\left|z-\sqrt{\lambda_-}\right|, \left|z-\sqrt{\lambda_+}\right|\right\},$$

and denote

$$a(z) := \operatorname{dist}\left(z, \left[\sqrt{\lambda_-}, \sqrt{\lambda_+}\right]\right), \quad b(z) := \operatorname{dist}\left(z, \left[\sqrt{\lambda_-}, \sqrt{\lambda_+}\right]^c\right)$$

We consider the curve

$$\mathscr{S} := \left\{ z = E + \mathrm{i}y : \sqrt{\lambda_-} + \varphi^2 N^{-\frac{2}{3}} < E < \sqrt{\lambda_+} - \varphi^2 N^{-\frac{2}{3}},\ y = \tfrac{\varphi^2}{N\kappa(E)^{1/2}} \right\},$$

and the domain $\mathscr{R} := \cup_{0<t<1}\{z_t : z \in \mathscr{S}\}$.

Lemma 2.8. *Uniformly in $0 < t < 1$ and $z = z_0$ satisfying $\eta := \operatorname{Im} z > 0$ and $|z - \sqrt{\lambda_+}| < \sqrt{\xi}/10$, we have*

$$\operatorname{Re}(z_t - z_0) \sim t\frac{a(z)}{\kappa(z)^{1/2}} + t^2, \quad \operatorname{Im}(z_t - z_0) \sim t\frac{b(z)}{\kappa(z)^{1/2}}.$$

In particular, if in addition we have $z \in \mathscr{S}$, then

$$(z_t - z_0) \sim \left(t\frac{\varphi^2}{N\kappa(E)} + t^2\right) + \mathrm{i}\kappa(E)^{\frac{1}{2}}t,$$

where φ is the subpolynomial error parameter introduced in Section 2.1.3. Moreover, for any $\kappa > 0$, uniformly in $0 < t < 1$ and $z = E + \mathrm{i}\eta \in [\sqrt{\lambda_-} + \kappa, \sqrt{\lambda_+} - \kappa] \times [0, \kappa^{-1}]$, we have $\operatorname{Im}(z_t - z_0) \sim t$.

Furthermore, we have the following lemma regarding the growth of the characteristics, which will be useful for the error estimates in the local relaxation.

Lemma 2.9. *For any $z = E + \mathrm{i}\eta \in \mathscr{S}$, we have*

$$\frac{\varphi^4}{N^2}\int_0^t ds \int \frac{d\rho(x)}{|z_{t-s} - x|^4 \max(\kappa(x), s^2)} \lesssim \frac{\kappa(E)}{\max(\kappa(E), t^2)}.$$

Conditioned on the rigidity phenomenon, we have the following estimate for the initial conditions.

Lemma 2.10. *In the set $\mathscr{A}$, for any $z = E + \mathrm{i}\eta \in \mathscr{R}$, we have $\operatorname{Im}\widetilde{\mathfrak{S}}_0(z) \lesssim \varphi^{1/2}$ if $\eta > \max(E - \sqrt{\lambda_+}, -E + \sqrt{\lambda_-})$, and $\operatorname{Im}\widetilde{\mathfrak{S}}_0(z) \lesssim \varphi^{1/2}\frac{\eta}{\kappa(z)}$ otherwise. The same bound also holds for $|\operatorname{Im}\mathfrak{S}_0|$.*

2.3.2 Local relaxation at soft edge

To prove the edge universality, we first have the following estimate for the size of the observable $\widetilde{\mathfrak{S}}_t$.

Proposition 2.1. *For any (large) $D > 0$ there exists $N_0(D)$ such that for any $N \geq N_0$ we have*

$$\mathbb{P}\left(\operatorname{Im}\widetilde{\mathfrak{S}}_t(z) \lesssim \varphi\frac{\kappa(E)^{1/2}}{\max(\kappa(E)^{1/2}, t)} \text{ for all } 0 < t < 1 \text{ and } z = E + \mathrm{i}\eta \in \mathscr{S}\right) > 1 - N^{-D}.$$

Proof. For any $1 \leq \ell, m \leq N^{10}$, we define $t_\ell = \ell N^{-10}$ and $z^{(m)} = E_m + \mathrm{i}\eta_m = E_m + \mathrm{i}N^{-1+4\varepsilon}\kappa(E_m)^{-1/2}$, where $\int_{-\infty}^{E_m} \mathrm{d}\rho_{\mathsf{MP}} = mN^{-10}$. Consider the stopping times

$$\begin{aligned}
\tau_0 &= \inf\left\{0 \leq t \leq 1 : \exists -N \leq k \leq N \ s.t.\ |s_k(t) - \gamma_k| > N^{-\frac{2}{3}+\varepsilon}(N + 1 - |k|)^{-\frac{1}{3}}\right\},\\
\tau_{\ell,m} &= \inf\left\{0 \leq s \leq t_\ell : \operatorname{Im}\widetilde{\mathfrak{S}}_s\left(z^{(m)}_{t_\ell - s}\right) > \frac{N^{2\varepsilon}}{2}\frac{\kappa(E_m)^{1/2}}{(\kappa(E_m)^{1/2} \vee t_\ell)}\right\},\\
\tau &= \min\left\{\tau_0, \tau_{\ell,m} : 0 \leq \ell, m \leq N^{10}, \kappa(E_m) > N^{-\frac{2}{3}+4\varepsilon}\right\},
\end{aligned}$$

with the convention $\inf\emptyset = \infty$. We claim that it suffices to show that $\tau = \infty$ with overwhelming probability.

To prove this claim, for any $z \in \mathscr{S}_\varepsilon$ and $0 \leq t \leq 1$, we pick $t_\ell \leq t \leq t_{\ell+1}$ and $|z - z^{(m)}| < N^{-5}$. Note that the maximum principle Lemma 2.6 implies $|\psi_k(t)| \lesssim 1$ for all k and t. Then we have $|\widetilde{\mathfrak{S}}_t(z) - \widetilde{\mathfrak{S}}_t(z^{(m)})| \lesssim N^{-2}$. Also, note that for $z = E + i\eta$ we have $|S_t(z)| \leq \eta^{-1}$, and

$$\left|\partial_z \widetilde{\mathfrak{S}}_t(z)\right| \lesssim N \max_k |\psi_k(0)| \eta^{-2} \lesssim N\eta^{-2}, \quad \left|\partial_{zz} \widetilde{\mathfrak{S}}_t(z)\right| \lesssim N\eta^{-3}.$$

Consider the events

$$\mathcal{E}_{\ell,m,k} := \left\{ \sup_{t_\ell \leq u \leq t_{\ell+1}} \left| \int_{t_\ell}^{u} \frac{\psi_k(v)}{(s_k(v) - z^{(m)})^2} \mathrm{d}B_k(v) \right| < N^{-4} \right\}.$$

On the event $\bigcap_k \mathcal{E}_{\ell,m,k}$, the above estimates imply that $|\widetilde{\mathfrak{S}}_t(z^{(m)}) - \widetilde{\mathfrak{S}}_{t_\ell}(z^{(m)})| < N^{-2}$. It further shows that

$$\left|\widetilde{\mathfrak{S}}_t(z) - \widetilde{\mathfrak{S}}_{t_\ell}(z^{(m)})\right| < N^{-2}.$$

Since this holds for all z and t, we have shown that

$$\{\tau = \infty\} \bigcap \left(\bigcap_{1 \leq \ell, m \leq N^{10}, -N \leq k \leq N} \mathcal{E}_{\ell,m,k} \right) \subset \bigcap_{z \in \mathscr{S}_\varepsilon, 0 \leq t \leq 1} \left\{ \operatorname{Im} \widetilde{\mathfrak{S}}_t(z) \leq N^{2\varepsilon} \frac{\kappa(E)^{1/2}}{(\kappa(E)^{1/2} \vee t)} \right\}. \tag{2.24}$$

Moreover, note that

$$\begin{aligned} &\left\langle \int_{t_\ell}^{t_{\ell+1}} \frac{\psi_k(v)}{(s_k(v) - z^{(m)})^2} \mathrm{d}B_k(v) \right\rangle_{t_{\ell+1}} \\ &\qquad \leq \int_{t_\ell}^{t_{\ell+1}} \frac{|\psi_k(v)|^2}{(s_k(v) - z^{(m)})^4} \mathrm{d}v \leq N^{-10} \left(N^{-1+4\varepsilon}\right)^{-4} \left(\max_k |\psi_k(0)|\right)^2 \leq N^{-6+16\varepsilon}. \end{aligned}$$

Using Lemma 2.1, we conclude that the event $\mathcal{E}_{\ell,m,k}$ happens with overwhelming probability. By a union bound, we further have that $\bigcap_{l,m,k} \mathcal{E}_{\ell,m,k}$ happens with overwhelming probability. Together with the set inclusion (2.24), we conclude that the claim is true, i.e. it suffices to prove $\tau = \infty$ with overwhelming probability.

To prove $\tau = \infty$ with overwhelming probability, consider some fixed $t = t_\ell$ and $z = z^{(m)} = E + \mathrm{i}\eta$, and define the function $f_u(z) := \widetilde{\mathfrak{S}}_u(z_{t-u})$. By Lemma 2.10, the initial condition is well controlled $\operatorname{Im} \widetilde{\mathfrak{S}}_0(z) \lesssim N^{2\varepsilon} \frac{\kappa(E_m)^{1/2}}{(\kappa(E_m)^{1/2} \vee t)}$. To bound the increments, note that the dynamics (2.22) yields

$$\mathrm{d}f_{u\wedge\tau}(z)=\epsilon_u(z_{t-u})\mathrm{d}(u\wedge\tau)-\frac{e^{-\frac{u}{2}}}{\sqrt{N}}\sum_{-N\leq k\leq N}\frac{\psi_k(u)}{(z_{t-u}-s_k(u))^2}\mathrm{d}B_k(u\wedge\tau),\tag{2.25}$$

where

$$\epsilon_u(z):=(S_u(z)-m(z))\partial_z\widetilde{\mathfrak{S}}_u+\frac{1}{4N}(\partial_{zz}\widetilde{\mathfrak{S}}_u)+\frac{e^{-\frac{u}{2\xi}}}{2N}\sum_{-N\leq k\leq N}\frac{\psi_k(u)}{(s_k-z)^2(s_k+z)}$$

By the local Marchenko-Pastur law (2.14), we have

$$\begin{aligned}
&\sup_{0\leq s\leq t}\left|\int_0^s\left(S_u(z_{t-u})-m(z_{t-u})\right)\partial_z\widetilde{\mathfrak{S}}_u(z_{t-u})\mathrm{d}(u\wedge\tau)\right|\\
&\leq\int_0^t\frac{N^{\varepsilon}}{N\mathrm{Im}\,(z_{t-u})}\sum_{-N\leq k\leq N}\frac{\psi_k(u)}{|s_k(u)-z_{t-u}|^2}\mathrm{d}(u\wedge\tau)\\
&\leq\int_0^t\frac{N^{\varepsilon}\mathrm{Im}\,\widetilde{\mathfrak{S}}_u(z_{t-u})}{N(\mathrm{Im}\,z_{t-u})^2}\mathrm{d}(u\wedge\tau)\\
&\leq\int_0^t\frac{N^{2\varepsilon}\mathrm{d}u}{N\left(\eta+(t-u)\frac{b(z)}{\kappa(z)^{1/2}}\right)^2}\frac{\kappa(E)^{\frac{1}{2}}}{(\kappa(E)^{\frac{1}{2}}\vee t)}\\
&\lesssim\frac{\kappa(E)^{\frac{1}{2}}}{(\kappa(E)^{\frac{1}{2}}\vee t)}.
\end{aligned}\tag{2.26}$$

Also, we have

$$\sup_{0\leq s\leq t}\left|\int_0^s\frac{1}{4N}(\partial_{zz}\widetilde{\mathfrak{S}}_u(z_{t-u}))\mathrm{d}(u\wedge\tau)\right|\lesssim\int_0^t\frac{\mathrm{Im}\,\widetilde{\mathfrak{S}}_u(z_{t-u})}{N\left(\mathrm{Im}\,(z_{t-u})\right)^2}\mathrm{d}(u\wedge\tau)\lesssim N^{-\varepsilon}\frac{\kappa(E)^{\frac{1}{2}}}{(\kappa(E)^{\frac{1}{2}}\vee t)},$$

and

$$\sup_{0\leq s\leq t}\left|\int_0^s \frac{1}{2N}\sum_{-N\leq k\leq N}\frac{\psi_k(u)}{(s_k-z_{t-u})^2(s_k+z_{t-u})}\mathrm{d}(u\wedge\tau)\right|$$
$$\lesssim \frac{1}{N}\int_0^t \frac{1}{\mathrm{Im}\,(z_{t-u})}\sum_{-N\leq k\leq N}\frac{\psi_k(u)}{|s_k(u)-z_{t-u}|^2}\mathrm{d}(u\wedge\tau)$$
$$\lesssim \int_0^t \frac{\mathrm{Im}\,\widetilde{\mathfrak{S}}_u(z_{t-u})}{N\left(\mathrm{Im}\,(z_{t-u})\right)^2}\mathrm{d}(u\wedge\tau)\lesssim N^{-\varepsilon}\frac{\kappa(E)^{\frac12}}{(\kappa(E)^{\frac12}\vee t)}.$$

For the martingale part

$$M_s:=\int_0^s \frac{e^{-\frac{u}{2\xi}}}{\sqrt{N}}\sum_{-N\leq k\leq N}\frac{\psi_k(u)}{(z_{t-u}-s_k(u))^2}\mathrm{d}B_k(u\wedge\tau),$$

using the rigidity of singular values (Lemma 2.5), with overwhelming probability we have

$$\sup_{0\leq s\leq t}|M_s|^2\lesssim N^{\frac{\varepsilon}{2}}\int_0^t \frac{1}{N}\sum_{-N\leq k\leq N}\frac{|\psi_k(u)|^2}{|z_{t-u}-\gamma_k|^4}\mathrm{d}(u\wedge\tau).$$

To estimate this integral, we chop the interval $[-N,N]$ into $2N^{1-4\varepsilon}$ subintervals $I_j=[k_j,k_{j+1}]$ where $k_j=-N+\lfloor jN^{4\varepsilon}\rfloor$. We can bound the summation in the integral in the following way

$$\frac{1}{N}\sum_{-N\leq k\leq N}\frac{|\psi_k(u)|^2}{|z_{t-u}-\gamma_k|^4}\leq \frac{1}{N}\sum_{0\leq j\leq 2N^{1-4\varepsilon}}\left(\max_{k\in I_j}\psi_k(u)\right)\left(\max_{k\in I_j}\frac{1}{|z_{t-u}-\gamma_k|^4}\right)\left(\sum_{k\in I_j}\psi_k(u)\right).$$

Using similar discretization arguments as above, we can derive

$$\max_{k\in I_j}\psi_k(u)\leq\sum_{k\in I_j}\psi_k(u)\leq\frac{N^{6\varepsilon}}{N(\kappa(\gamma_{k_j})^{1/2}\vee u)},$$

and

$$\max_{k\in I_j}\frac{1}{|z_{t-u}-\gamma_k|^4}\leq N^{-4\varepsilon}\sum_{k\in I_j}\frac{1}{|z_{t-u}-\gamma_k|^4}.$$

Therefore, we obtain

$$\sup_{0\leq s\leq t}|M_s|^2\leq N^{-2+9\varepsilon}\int_0^t\int_{-2}^2\frac{1}{|z_{t-u}-x|^4(\kappa(x)\vee u^2)}d\rho_{\mathsf{sc}}(x)\mathrm{d}u\lesssim N^{\varepsilon}\frac{\kappa(E)}{(\kappa(E)\vee t^2)}.$$

Combining this estimate for the martingale term with previous estimates, a union bound shows that with overwhelming probability we have

$$\sup_{\ell,m,0\leq s\leq t_\ell,\kappa(E_m)>\varphi^2N^{-2/3}}\operatorname{Im}\widetilde{\mathfrak{S}}_{s\wedge\tau}(z^{(m)}_{t_\ell-s\wedge\tau})\lesssim N^{2\varepsilon}\frac{\kappa(E)^{1/2}}{(\kappa(E)^{1/2}\vee t)}.$$

Now we have proved $\tau=\infty$ with overwhelming probability and hence the desired result is true. □

This estimate yields a rough control for the decay of $\varphi_k(t)$, which will be an important input for more refined estimates.

Lemma 2.11. *For all $-N\leq k\leq N$ and $0\leq t\leq 1$, we have*

$$|\varphi_k(t)|\prec\frac{1}{N}\frac{1}{\left((\frac{N+1-|k|}{N})^{1/3}\vee t\right)}\tag{2.27}$$

Proof. By Lemma 2.6, it suffices to control $\psi_k(t)$. By the nonnegativity of $\psi_k(t)$, we have

$$\operatorname{Im}\widetilde{\mathfrak{S}}_t(z)=\sum_{-N\leq k\leq N}\frac{\psi_k(t)\operatorname{Im}z}{|s_k(t)-z|^2}\geq\psi_k(t)\frac{\operatorname{Im}z}{|s_k(t)-z|^2},$$

which implies

$$\psi_k(t)\leq\operatorname{Im}\widetilde{\mathfrak{S}}_t(z)\frac{|s_k(t)-z|^2}{\operatorname{Im}z}.$$

Let $\varepsilon>0$. For $(N+1-|k|)>N^{10\varepsilon}$, pick the point $z=\gamma_k+\mathrm{i}N^{-1+4\varepsilon}\kappa(\gamma_k)^{-1/2}\in\mathscr{S}_\varepsilon$. In this case we have $\kappa(\gamma_k)^{1/2}\sim(\frac{N+1-|k|}{N})^{1/3}$. Therefore, in the set $\mathscr{A}_\varepsilon$, by Proposition 2.1, uniformly for all $-N+N^{10\varepsilon}\leq k\leq N-N^{10\varepsilon}$ and $0\leq t\leq 1$, with overwhelming probability we have

$$|\psi_k(t)| < \frac{N^{8\varepsilon}}{N} \frac{1}{\left((\frac{N+1-|k|}{N})^{1/3} \vee t \right)}.$$

For $(N+1-|k|) \leq N^{10\varepsilon}$, without loss of generality we consider $N+1-k \leq N^{10\varepsilon}$. In this case, let $k_0 = N - N^{10\varepsilon} + 1$ and consider $z = \gamma_{k_0} + \mathrm{i}N^{-1+4\varepsilon}\kappa(\gamma_{k_0})^{-1/2}$. The same argument results in a similar bound with a larger $N^{20\varepsilon}$ factor. By the arbitrariness of ε, this completes the proof. □

A direct consequence of Lemma 2.11 is the following estimate of the local relaxation flow at the soft edge.

Theorem 2.3. *For any $D > 0$ and $\varepsilon > 0$ there exists $N_0 > 0$ such that for any $N > N_0$ we have*

$$\mathbb{P}\left(|\sigma_k(X,t) - \sigma_k(G,t)| \lesssim \frac{N^\varepsilon}{Nt} \textit{ for all } k \in [N] \textit{ in } t \in [0,1] \right) > 1 - N^{-D}.$$

2.3.3 Local relaxation at hard edge

In this subsection we prove a quantitative estimate for the local relaxation flow (2.8) at the hard edge. The main estimate in this section is the following.

Theorem 2.4. *For $\varepsilon_0 > 0$ arbitrarily small and any $N^{-1+\varepsilon_0} < t < 1$, we have*

$$|\sigma_1(X,t) - \sigma_1(G,t)| \prec \frac{1}{N^2 t}. \tag{2.28}$$

To estimate $\varphi_k(t)$ near the hard edge, we introduce the following quantity to approximate it. Let $\gamma_k^t = (\gamma_k)_t$ with the convention $\gamma^t = (\gamma + \mathrm{i}0^+)_t$, and define

$$\widehat{\varphi}_k(t) := \frac{1}{2N\mathrm{Im}\, m_{\mathsf{sc}}(\gamma_k^t)} \sum_{-N \leq j \leq N} \mathrm{Im}\left(\frac{1}{\gamma_j - \gamma_k^t} \right) (\sigma_j(X) - \sigma_j(G)). \tag{2.29}$$

Our goal is to prove the following estimates

Proposition 2.2. *Let $0 < c < 1$ be a fixed small constant. For $\varepsilon_0 > 0$ arbitrarily small with any $N^{-1+\varepsilon_0} < t < 1$ and $k \in [(c-1)N, (1-c)N]$, we have*

$$|\sigma_k(X,t) - \sigma_k(G,t) - \widehat{\varphi}_k(t)| \prec \frac{1}{N^2 t}.$$

To obtain the optimal control for the local relaxation flow, we need to carefully estimate $\widehat{\varphi}_k$ near the hard edge. A first step towards such estimates is given in the following lemma.

Lemma 2.12. *Let $\varepsilon > 0$ and $0 < c < 1$. For any $(k,\ell) \in [(c-1)N, (1-c)N]^2$, $|E| < 2-c$, and $s,t,\eta \in [N^{-1+4\varepsilon}, 1]$, in the set $\mathscr{A}_\varepsilon$, for $z = E + \mathrm{i}\eta$ we have*

$$|\widehat{\varphi}_k(t) - \widehat{\varphi}_\ell(s)| \lesssim N^\varepsilon \left(\frac{|k-\ell|}{N^2(s\wedge t)} + \frac{|s-t|}{N(s\wedge t)} \right), \tag{2.30}$$

$$\left| \frac{\operatorname{Im} \mathfrak{S}_0(z_t)}{2N \operatorname{Im} S_0(z_t)} - \widehat{\varphi}_\ell(s) \right| \lesssim N^\varepsilon \left(\frac{|E-\gamma_\ell|}{N(s\wedge(\eta+t))} + \frac{|\eta+t-s|}{N(s\wedge(\eta+t))} \right). \tag{2.31}$$

Proof. By the properties of the Stieltjes transform $m_{\mathsf{sc}}(z)$ (see e.g. [EY17a, Section 6]) and direct computation, we have

$$\operatorname{Im} m_{\mathsf{sc}}(\gamma_k^t) \gtrsim 1, \quad \operatorname{Im} m_{\mathsf{sc}}(\gamma_\ell^s) \gtrsim 1,$$

and

$$\left|\operatorname{Im} m_{\mathsf{sc}}(\gamma_k^t) - \operatorname{Im} m_{\mathsf{sc}}(\gamma_\ell^s)\right| + \left|\gamma_k^t - \gamma_\ell^s\right| \lesssim \frac{|k-\ell|}{N} + |s-t|. \tag{2.32}$$

In the set $\mathscr{A}_\varepsilon$, the rigidity estimates imply that

$$|\sigma_j(X) - \sigma_j(G)| \le N^{-\frac{2}{3}+\varepsilon}(N+1-|j|)^{-\frac{1}{3}}.$$

Then we have

$$
\begin{aligned}
\left|\frac{1}{2N}\mathrm{Im}\left(\sum_{-N\le j\le N}\frac{\sigma_j(X)-\sigma_j(G)}{\gamma_j-\gamma_k^t}\right)\right| &\lesssim \frac{1}{2N}(\mathrm{Im}\,\gamma_k^t)\sum_{-j\le j\le N}\frac{N^{-\frac{2}{3}+\varepsilon}(N+1-|j|)^{-\frac{1}{3}}}{(\gamma_j-\mathrm{Re}\,\gamma_k^t)^2+(\mathrm{Im}\,\gamma_k^t)^2}\\
&\lesssim N^{-1+\varepsilon}(\mathrm{Im}\,\gamma_k^t)\int_{-2}^{2}\frac{\kappa(x)^{-\frac{1}{2}}}{(x-\mathrm{Re}\,\gamma_k^t)^2+(\mathrm{Im}\,\gamma_k^t)^2}\mathrm{d}\rho_{\mathsf{sc}}(x)\\
&\lesssim N^{-1+\varepsilon}(\mathrm{Im}\,\gamma_k^t)\int_{-2}^{2}\frac{1}{(x-\mathrm{Re}\,\gamma_k^t)^2+(\mathrm{Im}\,\gamma_k^t)^2}\mathrm{d}x\\
&\lesssim N^{-1+\varepsilon}.
\end{aligned}
\tag{2.33}
$$

By triangle inequality, we have

$$
\begin{aligned}
|\widehat{\varphi}_k(t)-\widehat{\varphi}_\ell(s)| \le{}& \frac{1}{2N}\left|\left(\frac{1}{\mathrm{Im}\,m_{\mathsf{sc}}(\gamma_k^t)}-\frac{1}{\mathrm{Im}\,m_{\mathsf{sc}}(\gamma_\ell^s)}\right)\mathrm{Im}\left(\sum_{-N\le j\le N}\frac{\sigma_j(X)-\sigma_j(G)}{\gamma_j-\gamma_k^t}\right)\right|\\
&+\frac{1}{2N\mathrm{Im}\,m_{\mathsf{sc}}(\gamma_\ell^s)}\sum_{-N\le j\le N}\left|\mathrm{Im}\left(\frac{1}{\gamma_j-\gamma_k^t}\right)-\mathrm{Im}\left(\frac{1}{\gamma_j-\gamma_\ell^s}\right)\right||\sigma_j(X)-\sigma_j(G)|\\
=:{}& I_1+I_2.
\end{aligned}
$$

Using (2.32) and (2.33), we obtain

$$
I_1 \lesssim N^{\varepsilon}\left(\frac{|k-\ell|}{N^2}+\frac{|s-t|}{N}\right).
$$

For the second term I_2, in the set $\mathscr{A}_\varepsilon$, the rigidity and (2.32) imply that

$$
\begin{aligned}
I_2 &\lesssim N^{-1+\varepsilon}\sum_{-N\le j\le N}N^{-\frac{2}{3}}(N+1-|j|)^{-\frac{1}{3}}\left|\frac{\gamma_k^t-\gamma_\ell^s}{(\gamma_j-\gamma_k^t)(\gamma_j-\gamma_\ell^s)}\right|\\
&\lesssim N^{-1+\varepsilon}\left(\frac{|k-\ell|}{N}+|s-t|\right)\sum_{-N\le j\le N}N^{-\frac{2}{3}}(N+1-|j|)^{-\frac{1}{3}}\left(\frac{1}{|\gamma_j-\gamma_k^t|^2}+\frac{1}{|\gamma_j-\gamma_\ell^s|^2}\right).
\end{aligned}
$$

Recall from Lemma 2.8 that $\mathrm{Im}\,\gamma_k^t \sim t$ and $\mathrm{Im}\,\gamma_\ell^s \sim s$. Using a similar argument as in

(2.33) we obtain

$$I_2 \lesssim N^{-1+\varepsilon}\left(\frac{|k-\ell|}{N}+|s-t|\right)\left(\frac{1}{t}+\frac{1}{s}\right) \lesssim N^{\varepsilon}\left(\frac{|k-\ell|}{N^2(s\wedge t)}+\frac{|s-t|}{N(s\wedge t)}\right).$$

Hence we have proved (2.30). For the other part (2.31), it can be proved via the same arguments. $\square$

As a consequence, we have a good control for the size of $\widehat{\varphi}_k(t)$ away from the soft edge. This is based on the symmetric structure of $\{\widehat{\varphi}_k\}$.

Lemma 2.13. *Let $\varepsilon>0$ and $0<c<1$. For any $t\in[N^{-1+4\varepsilon},1]$ and $k\in[(c-1)N,(1-c)N]$, with overwhelming probability we have*

$$|\widehat{\varphi}_k(t)| \lesssim N^{\varepsilon}\frac{k}{N^2 t}. \tag{2.34}$$

Proof. A key observation is the following

$$\operatorname{Re}\gamma_{-k}^t=-\operatorname{Re}\gamma_k^t,\ \operatorname{Im}\gamma_{-k}^t=\operatorname{Im}\gamma_k^t,\ \operatorname{Re}m_{\mathsf{sc}}(\gamma_{-k}^t)=-\operatorname{Re}m_{\mathsf{sc}}(\gamma_k^t),\ \operatorname{Im}m_{\mathsf{sc}}(\gamma_{-k}^t)=\operatorname{Im}m_{\mathsf{sc}}(\gamma_k^t).$$

Therefore, we have

$$\begin{aligned}
\widehat{\varphi}_{-k}(t) &= \frac{1}{2N\operatorname{Im}m_{\mathsf{sc}}(\gamma_{-k}^t)}\sum_{-N\le j\le N}\operatorname{Im}\left(\frac{1}{\gamma_j-\gamma_{-k}^t}\right)(\sigma_j(X)-\sigma_j(G))\\
&= \frac{1}{2N\operatorname{Im}m_{\mathsf{sc}}(\gamma_k^t)}\sum_{-N\le j\le N}\frac{\operatorname{Im}(\gamma_{-k}^t)}{(\gamma_j-\operatorname{Re}(\gamma_{-k}^t))^2+(\operatorname{Im}\gamma_{-k}^t)^2}(\sigma_j(X)-\sigma_j(G))\\
&= \frac{1}{2N\operatorname{Im}m_{\mathsf{sc}}(\gamma_k^t)}\sum_{-N\le j\le N}\frac{\operatorname{Im}(\gamma_k^t)}{(\gamma_j+\operatorname{Re}(\gamma_k^t))^2+(\operatorname{Im}\gamma_k^t)^2}(\sigma_j(X)-\sigma_j(G))
\end{aligned}$$

Using the symmtrization of the singular values, we further have

$$\begin{aligned}
\widehat{\varphi}_{-k}(t) &= -\frac{1}{2N\mathrm{Im}\, m_{\mathsf{sc}}(\gamma_k^t)} \sum_{-N\le j\le N} \frac{\mathrm{Im}\,(\gamma_k^t)}{(\gamma_{-j}-\mathrm{Re}\,(\gamma_k^t))^2+(\mathrm{Im}\,\gamma_k^t)^2}(\sigma_{-j}(X)-\sigma_{-j}(G)) \\
&= -\frac{1}{2N\mathrm{Im}\, m_{\mathsf{sc}}(\gamma_k^t)} \sum_{-N\le j\le N} \mathrm{Im}\left(\frac{1}{\gamma_{-j}-\gamma_k^t}\right)(\sigma_{-j}(X)-\sigma_{-j}(G)) \\
&= -\frac{1}{2N\mathrm{Im}\, m_{\mathsf{sc}}(\gamma_k^t)} \sum_{-N\le j\le N} \mathrm{Im}\left(\frac{1}{\gamma_{j}-\gamma_k^t}\right)(\sigma_{j}(X)-\sigma_{j}(G)) \\
&= -\widehat{\varphi}_k(t)
\end{aligned}$$

Consequently, by (2.30) we have

$$|\widehat{\varphi}_k(t)-\widehat{\varphi}_{-k}(t)| = 2|\widehat{\varphi}_k(t)| \lesssim N^{\varepsilon}\frac{2k}{N^2 t}.$$

This shows the desired result. □

Finally, it's straightfoward to derive Theorem 2.4 from Proposition 2.2 and Lemma 2.13. Thus, our primary goal is to prove Proposition 2.2.

We will prove the main technical estimate Proposition 2.2 via a bootstrap argument.

Definition 2.3 (Hypothesis $\mathcal{H}_\alpha$)**.** *Consider the following hypothesis: For any fixed small $0<c<1$, the following holds for $\varepsilon_0>0$ arbitrarily small. For any $N^{-1+\varepsilon_0}<t<1$, $k\in[(c-1)N,(1-c)N]$ and $\nu\in[0,1]$, we have*

$$\left|\varphi_k^{(\nu)}(t)-\widehat{\varphi}_k(t)\right| \prec \frac{(Nt)^\alpha}{N^2 t}. \tag{2.35}$$

Proposition 2.2 is derived via a bootstrap of the hypothesis $\mathcal{H}_\alpha$. Specifically, we have the following two lemmas.

Lemma 2.14. *The hypothesis $\mathcal{H}_1$ is true.*

Proof. Recall from Lemma 2.11 that (2.27) implies that $\varphi_k^{(\nu)}(t)\prec N^{-1}$. On the other hand, from the definition (2.29) of $\widehat{\varphi}_k$, using the rigidity and (2.33) we obtain $\widehat{\varphi}_k(t)\prec N^{-1}$ thanks to the arbitrariness of ε_0. Therefore, the triangle inequality yields $|\varphi_k^{(\nu)}(t) - \widehat{\varphi}_k(t)| \prec N^{-1}$, which completes the proof. $\square$

Lemma 2.15. *If $\mathcal{H}_\alpha$ is true, then $\mathcal{H}_{3\alpha/4}$ is true, i.e.*

$$\left|\varphi_k^{(\nu)}(t) - \widehat{\varphi}_k(t)\right| \prec \frac{(Nt)^{\frac{3\alpha}{4}}}{N^2 t}.$$

The self-improving property of the hypothesis $\mathcal{H}_\alpha$ stated in Lemma 2.15 is the main technical part of the proof for Proposition 2.2. We defer its proof to Section 2.3.4.

Finally, the optimal control (2.28) for the local relaxation flow at the hard edge follows from these two lemma together with Lemma 2.13.

Proof of Proposition 2.2. Note that

$$\sigma_k(X,t) - \sigma_k(G,t) - \widehat{\varphi}_k(t) = \int_0^1 \varphi_k^{(\nu)}(t)\mathrm{d}\nu - \widehat{\varphi}_k(t) = \int_0^1 \left(\varphi_k^{(\nu)}(t) - \widehat{\varphi}_k(t)\right)\mathrm{d}\nu \quad (2.36)$$

Consider an arbitrarily fixed $\delta > 0$, based on Lemma 2.14 and Lemma 2.15, after a finite time of iterations, with overwhelming probability we have

$$|\varphi_k(t) - \widehat{\varphi}_k(t)| \leq \frac{N^\delta}{N^2 t}$$

This shows that for any fixed $\widetilde{D}$ and p, and for large enough N, we have

$$\mathbb{E}\left(|\varphi_k(t) - \widehat{\varphi}_k(t)|^{2p}\right) \leq \left(\frac{N^\delta}{N^2 t}\right)^{2p} + N^{-\widetilde{D}}.$$

By (2.36) we obtain

$$\mathbb{E}\left(|\sigma_k(X,t) - \sigma_k(G,t) - \widehat{\varphi}_k(t)|^{2p}\right) \leq \int_0^1 \mathbb{E}\left(|\varphi_k(t) - \widehat{\varphi}_k(t)|^{2p}\right)\mathrm{d}\nu \leq \left(\frac{N^\delta}{N^2 t}\right)^{2p} + N^{-\widetilde{D}}.$$

We choose $p = D/\delta$ and $\widetilde{D} = D + 100p$, and then the Markov inequality yields

$$\mathbb{P}\left(|\sigma_k(X,t) - \sigma_k(G,t) - \widehat{\varphi}_k(t)| \leq \frac{N^{2\delta}}{N^2 t} \right) > 1 - N^{-D},$$

which completes the proof thanks to the arbitrariness of δ and D. $\square$

2.3.4 Proof of the Bootstrap Argument

The proof of Lemma 2.15 is a delicate task. The key part of the proof is a careful analysis of the dynamics. The main idea is to approximate the dynamics with a short-range version, which will be easier to control. To do this, we show the finite speed of propagation estimate for the short-range kernel of the parabolic-type equation (2.20) satisfied by $\{\varphi_k\}$. Then we prove a short-range approximation of the original dynamics and introduce a regularized equation. Finally, we show that, with a well-behaved initial condition, the regularized equation gives us the desired good approximation.

To begin with, the core input of the bootstrap argument is the following technical lemma, which states that the estimate of the local average will improve along with the induction hypothesis $\mathcal{H}_\alpha$.

Lemma 2.16. *Assume $\mathcal{H}_\alpha$. Let $b > 0$ be any fixed small constant. For any $0 < t < 1$, any $\varepsilon_0 > 0$ arbitrarily small and $z = E + \mathrm{i}\eta$ satisfying $N^{-1+\varepsilon_0} < \eta < 1$, $|E| < 2 - b$, we have*

$$\left| \operatorname{Im} \mathfrak{S}_t(z) - e^{-\frac{t}{2}} \frac{\operatorname{Im} S_t(z)}{\operatorname{Im} S_0(z_t)} \operatorname{Im} \mathfrak{S}_0(z_t) \right| \prec \left(\frac{(Nt)^\alpha}{N^2 t\eta} + \frac{1}{Nt} \right) \tag{2.37}$$

Proof. Fix t and consider the function

$$g_u(z) := \mathfrak{S}_u(z_{t-u}) - e^{-\frac{u}{2}} \frac{\operatorname{Im} \mathfrak{S}_0(z_t)}{\operatorname{Im} S_0(z_t)} S_u(z_{t-u}), \quad 0 \leq u \leq t.$$

An observation is that $e^{-u/2} S_u(z)$ satisfy the same stochastic advection equation (2.22) with φ_k replaced by $\frac{1}{2N}$. Therefore, we have

$$
\begin{aligned}
\mathrm{d}g_u &= (S_u(z_{t-u}) - m_{\mathsf{sc}}(z_{t-u}))\left(\partial_z \mathfrak{S}_u(z_{t-u}) - e^{-\frac{u}{2}}\frac{\mathrm{Im}\,\mathfrak{S}_0(z_t)}{\mathrm{Im}\,S_0(z_t)}\partial_z S_u(z_{t-u})\right)\mathrm{d}u \\
&+ \frac{1}{4N}\left(\partial_{zz}\mathfrak{S}_u(z_{t-u}) - e^{-\frac{u}{2}}\frac{\mathrm{Im}\,\mathfrak{S}_0(z_t)}{\mathrm{Im}\,S_0(z_t)}\partial_{zz} S_u(z_{t-u})\right)\mathrm{d}u \\
&+ \frac{e^{-\frac{u}{2}}}{2N}\sum_{-N\le k\le N}\frac{\theta_k(u)}{(s_k(u)-z_{t-u})^2(s_k(u)+z_{t-u})}\mathrm{d}u \\
&- \frac{e^{-\frac{u}{2}}}{\sqrt{N}}\sum_{-N\le k\le N}\frac{\theta_k(u)}{(s_k(u)-z_{t-u})^2}\mathrm{d}B_k,
\end{aligned}
\tag{2.38}
$$

where

$$
\theta_k(u) = \varphi_k(u) - \frac{\mathrm{Im}\,\mathfrak{S}_0(z_t)}{2N\mathrm{Im}\,S_0(z_t)}.
$$

Similarly as in the proof of Proposition 2.1, for $\varepsilon > 0$ and $0 \le \ell, m, p \le N^{10}$, define $t_\ell = \ell N^{-10}$ and $z^{(m,p)} = E_m + \mathrm{i}\eta_p$ where $\int_{-\infty}^{E_m}\mathrm{d}\rho = mN^{-10}$ and $\eta_p = N^{-1+4\varepsilon} + pN^{-10}$. We also pick $c > 0$ such that $\lfloor (1-c)N \rfloor = \arg\min_k |\gamma_k - (2 - \frac{b}{10})|$. Assuming $\mathcal{H}_\alpha$, let $\varepsilon_0 > 0$ be the arbitrarily small scale in the hypothesis. Let $C > 0$ be some suitably large constant. Recall the stopping times

$$
\tau_0 = \inf\left\{0 \le u \le 1 : \exists - N \le k \le N \ s.t.\ |s_k(u) - \gamma_k| > N^{-\frac{2}{3}+\varepsilon}(N+1-|k|)^{-\frac{1}{3}}\right\},
$$

$$
\tau_1 = \inf\left\{N^{-1+\varepsilon_0} \le u \le 1 : \exists - N \le k \le N \ s.t.\ |\varphi_k(u)| > \frac{N^{C\varepsilon}}{N}\frac{1}{\left((\frac{N+1-|k|}{N})^{1/3} \vee u\right)}\right\},
$$

and consider the new stopping times

$$
\begin{aligned}
\tau_2 &= \inf\left\{N^{-1+\varepsilon_0} \le u \le 1 : \exists k \in [(c-1)N, (1-c)N] \ s.t.\ |\varphi_k(u) - \widehat{\varphi}_k(u)| > N^{C\varepsilon}\frac{(Nt)^\alpha}{N^2 t}\right\}, \\
\tau_{\ell,m,p} &= \inf\left\{0 \le u \le t_\ell : \left|\mathrm{Im}\, g_u^{(t_\ell)}(z^{(m,p)})\right| > N^{C\varepsilon}\left(\frac{(Nt)^\alpha}{N^2 t\eta} + \frac{1}{Nt}\right)\right\} \\
\tau &= \min\{\tau_0, \tau_1, \tau_2, \tau_{\ell,m,p} : 0 \le \ell, m, p \le N^{10}, |E_m| < 2 - b\}.
\end{aligned}
$$

Recall the convention $\inf\emptyset = \infty$. As shown in the proof of Proposition 2.1, it suffices to show that $\tau = \infty$ with overwhelming probability.

A key ingredient for the analysis of the dynamics of g_u is the following estimates on $\theta_k(u)$. To do this, we fix some $t = t_\ell$ and $z = z^{(m,p)}$ with $|E_m| < 2 - b$, and let $N^{-1+\varepsilon_0} \le u \le t \wedge \tau$ and $k \in [(c-1)N, (1-c)N]$.

On the one hand, we have a direct a priori estimate. Since $u \le \tau_1$, we have

$$|\varphi_k(u)| \lesssim N^{-\frac{2}{3}+C\varepsilon}(N+1-|k|)^{-\frac{1}{3}}.$$

Moreover, note that for $z = E + i\eta$ with E in the bulk and $t < 1$, uniformly we have $\operatorname{Im} S_0(z_t) \gtrsim 1$. By Lemma 2.10, this shows

$$\frac{\operatorname{Im} \mathfrak{S}_0(z_t)}{2N \operatorname{Im} S_0(z_t)} \lesssim N^{-1+\varepsilon} \lesssim N^{-\frac{2}{3}+C\varepsilon}(N+1-|k|)^{-\frac{1}{3}}.$$

As a consequence, we have

$$|\theta_k(u)| \lesssim N^{-\frac{2}{3}+C\varepsilon}(N+1-|k|)^{-\frac{1}{3}}. \tag{2.39}$$

On the other hand, the estimate can also be obtained via approximation

$$|\theta_k(u)| \le |\varphi_k(u) - \widehat{\varphi}_k(u)| + |\widehat{\varphi}_k(u) - \widehat{\varphi}_j(t)| + \left|\widehat{\varphi}_j(t) - \frac{\operatorname{Im} \mathfrak{S}_0(z_t)}{2N \operatorname{Im} S_0(z_t)}\right|.$$

For the first term, since $u \le \tau_2$ we have

$$|\varphi_k(u) - \widehat{\varphi}_k(u)| \le N^{C\varepsilon}\frac{(Nu)^a}{N^2 u}.$$

Choosing $|\gamma_j - E| \le N^{-1+2\varepsilon}$, the remaining two terms are controlled by Lemma 2.13

$$\begin{aligned}|\widehat{\varphi}_k(u) - \widehat{\varphi}_j(t)| + \left|\widehat{\varphi}_j(t) - \frac{\operatorname{Im}\mathfrak{S}_0(z_t)}{2N\operatorname{Im}S_0(z_t)}\right| &\lesssim N^{\varepsilon}\left(\frac{|k-j|}{N^2u} + \frac{|t-u|}{Nu} + \frac{|E-\gamma_j|}{Nt} + \frac{\eta}{Nt}\right)\\ &\lesssim N^{C\varepsilon}\left(\frac{|\gamma_k - E|}{Nu} + \frac{t-u}{Nu} + \frac{\eta}{Nt}\right).\end{aligned}$$

Together, we decompose the error terms into two parts and obtain the following

$$|\theta_k(u)| \le N^{C\varepsilon}\left(\frac{|\gamma_k - E|}{Nu} + \frac{t-u}{Nu} + \frac{\eta}{Nt} + \frac{(Nu)^{\alpha}}{N^2u}\right) =: N^{C\varepsilon}\left(\frac{|\gamma_k - E|}{Nu} + \Lambda(a, N, t, \eta, u)\right) \tag{2.40}$$

With the above control on $\theta_k(u)$, the dynamics (2.38) can be used to bound $\operatorname{Im}(g_t - g_0)$ similarly as in Proposition 2.1. For the first term, we have

$$\begin{aligned}&\int_0^{t\wedge\tau} |S_u(z_{t-u}) - m_{\mathsf{sc}}(z_{t-u})| \left|\partial_z\mathfrak{S}_u(z_{t-u}) - e^{-\frac{u}{2}}\frac{\operatorname{Im}\mathfrak{S}_0(z_t)}{\operatorname{Im}S_0(z_t)}\partial_z S_u(z_{t-u})\right| \mathrm{d}u\\ &\le \int_0^{t\wedge\tau} \frac{N^{C\varepsilon}}{N\operatorname{Im}(z_{t-u})} \sum_{-N\le k\le N} \frac{|\theta_k(u)|}{|s_k(u) - z_{t-u}|^2}\mathrm{d}u\\ &\le \int_0^{t\wedge\tau} \frac{N^{C\varepsilon}}{N\operatorname{Im}(z_{t-u})}\left(\sum_{|k|\ge(1-c)N} \frac{|\theta_k(u)|}{|\gamma_k - z_{t-u}|^2} + \sum_{|k|<(1-c)N}\frac{|\theta_k(u)|}{|\gamma_k - z_{t-u}|^2}\right)\mathrm{d}u\\ &=: I_1 + I_2.\end{aligned} \tag{2.41}$$

For the soft edge part $|k| \ge (1-c)N$, using (2.39) we obtain

$$\begin{aligned}I_1 &\le N^{C\varepsilon}\int_0^{t\wedge\tau} \frac{1}{N\operatorname{Im}(z_{t-u})} \sum_{|k|\ge(1-\alpha)N} N^{-\frac{2}{3}+C\varepsilon}(N+1-|k|)^{-\frac{1}{3}}\mathrm{d}u\\ &\le N^{C\varepsilon}\int_0^{t\wedge\tau}\frac{1}{N\operatorname{Im}(z_{t-u})}\mathrm{d}u\\ &\le N^{C\varepsilon}\frac{\log(1+Nt)}{N}.\end{aligned} \tag{2.42}$$

For I_2, note that

$$\sum_{|k|<(1-c)N} \frac{|\theta_k(u)|}{|\gamma_k - z_{t-u}|^2} \leq N^{C\varepsilon} \left(\frac{1}{Nu} \sum_{|k|<(1-c)N} \frac{|\gamma_k - E|}{|\gamma_k - z_{t-u}|^2} + \Lambda \sum_{|k|<(1-c)N} \frac{1}{|\gamma_k - z_{t-u}|^2} \right)$$
$$\leq N^{C\varepsilon} \left(\frac{1}{u} + N \frac{\Lambda}{\eta + t - u} \right)$$

This yields

$$\begin{aligned}
&\int_{N^{-1+\varepsilon_0}}^{t\wedge\tau} \frac{N^{C\varepsilon}}{N\mathrm{Im}\,(z_{t-u})} \sum_{|k|<(1-\alpha)N} \frac{|\theta_k(u)|}{|\gamma_k - z_{t-u}|^2} \mathrm{d}u \\
&\leq N^{C\varepsilon} \int_{N^{-1+\varepsilon_0}}^{t\wedge\tau} \frac{1}{N(\eta + t - u)} \left(\frac{1}{u} + N \frac{\Lambda}{\eta + t - u} \right) \mathrm{d}u \\
&\leq N^{C\varepsilon} \int_{N^{-1+\varepsilon_0}}^{t\wedge\tau} \frac{1}{N(\eta + t - u)} \left(\frac{1}{u} + N \frac{1}{\eta + t - u} \left(\frac{t-u}{Nu} + \frac{\eta}{Nt} + \frac{(Nu)^\alpha}{N^2 u} \right) \right) \mathrm{d}u \\
&\leq N^{C\varepsilon} \left(\frac{(Nt)^\alpha}{N^2 t\eta} + \frac{1}{Nt} \right)
\end{aligned} \tag{2.43}$$

Moreover, without loss of generality, we may assume that $\varepsilon_0 \sim \varepsilon$. Then, using (2.39) we obtain

$$\int_0^{N^{-1+\varepsilon_0}} \frac{N^{C\varepsilon}}{N\mathrm{Im}\,(z_{t-u})} \sum_{|k|<(1-\alpha)N} \frac{|\theta_k(u)|}{|\gamma_k - z_{t-u}|^2} \mathrm{d}u \leq \frac{N^{C\varepsilon}}{N^2(\eta + t)^2} \tag{2.44}$$

Together with previous estimates, this shows

$$\int_0^{t\wedge\tau} |S_u(z_{t-u}) - m_{\mathsf{sc}}(z_{t-u})| \left| \partial_z \mathfrak{S}_u(z_{t-u}) - e^{-\frac{u}{2}} \frac{\mathrm{Im}\,\mathfrak{S}_0(z_t)}{\mathrm{Im}\,S_0(z_t)} \partial_z S_u(z_{t-u}) \right| \mathrm{d}u \leq N^{C\varepsilon} \left(\frac{(Nt)^\alpha}{N^2 t\eta} + \frac{1}{Nt} \right). \tag{2.45}$$

Similarly, we have

$$\int_0^{t\wedge\tau} \frac{1}{4N} \left| \partial_{zz} \mathfrak{S}_u(z_{t-u}) - e^{-\frac{u}{2}} \frac{\mathrm{Im}\,\mathfrak{S}_0(z_t)}{\mathrm{Im}\,S_0(z_t)} \partial_{zz} S_u(z_{t-u}) \right| \mathrm{d}u \leq N^{C\varepsilon} \left(\frac{(Nt)^\alpha}{N^2 t\eta} + \frac{1}{Nt} \right)$$

and

$$\int_0^{t\wedge\tau}\left|\frac{e^{-\frac{u}{2}}}{2N}\sum_{-N\le k\le N}\frac{\theta_k(u)}{(s_k(u)-z_{t-u})^2(s_k(u)+z_{t-u})}\right|\mathrm{d}u\le N^{C\varepsilon}\left(\frac{(Nt)^\alpha}{N^2t\eta}+\frac{1}{Nt}\right)$$

It suffices to bound the martingale term

$$M_s:=\int_0^s\frac{e^{-\frac{u}{2}}}{\sqrt{N}}\sum_{-N\le k\le N}\frac{\theta_k(u)}{(s_k(u)-z_{t-u})^2}\mathrm{d}B_k.$$

Again we decompose the integral into two parts

$$\begin{aligned}\langle M\rangle_{t\wedge\tau}&\lesssim\frac{1}{N}\int_0^{t\wedge\tau}\sum_{|k|\ge(1-c)N}\frac{|\theta_k(u)|^2}{|\gamma_k-z_{t-u}|^4}\mathrm{d}u+\frac{1}{N}\int_0^{t\wedge\tau}\sum_{|k|<(1-c)N}\frac{|\theta_k(u)|^2}{|\gamma_k-z_{t-u}|^4}\mathrm{d}u\\&=:J_1+J_2.\end{aligned}$$

The contribution from the soft edge is easy to control

$$J_1\le\frac{N^{C\varepsilon}}{N}\int_0^{t\wedge\tau}\sum_{|k|\ge(1-c)N}\left(N^{-\frac{2}{3}}(N+1-|k|)^{-\frac{1}{3}}\right)^2\mathrm{d}u\le N^{C\varepsilon}\frac{t}{N^2}$$

For the other term, we use both (2.39) and (2.40)

$$\begin{aligned}J_2\lesssim&\frac{N^{C\varepsilon}}{N}\int_0^{t\wedge\eta}\frac{1}{N^2}\sum_{|k|<(1-c)N}\frac{1}{|\gamma_k-z_{t-u}|^4}\mathrm{d}u\\&+\frac{N^{C\varepsilon}}{N}\int_{t\wedge\eta}^{t\wedge\tau}\sum_{|k|<(1-c)N}\frac{1}{|\gamma_k-z_{t-u}|^4}\left[\left(\frac{|\gamma_k-E|^2}{N^2u^2}+\Lambda^2\right)\wedge\frac{1}{N^2}\right]\mathrm{d}u.\end{aligned}$$

Note that

$$\frac{N^{C\varepsilon}}{N}\int_0^{t\wedge\eta}\frac{1}{N^2}\sum_{|k|<(1-c)N}\frac{1}{|\gamma_k-z_{t-u}|^4}\mathrm{d}u\le\frac{N^{C\varepsilon}}{N}\int_0^{t\wedge\eta}\frac{1}{N(\eta+t-u)^3}\mathrm{d}u\le\frac{N^{C\varepsilon}}{N^2t^2}.$$

For the other term, without loss of generality we may assume $\eta<t$. For the $\frac{|\gamma_k-E|^2}{N^2u^2}$ term,

we have

$$
\begin{aligned}
&\frac{N^{C\varepsilon}}{N}\int_\eta^t \sum_{|k|<(1-c)N} \frac{1}{|\gamma_k - z_{t-u}|^4}\left(\frac{|\gamma_k-E|^2}{N^2u^2}\wedge\frac{1}{N^2}\right)\mathrm{d}u \\
&\leq \frac{N^{C\varepsilon}}{N^2t^2} + \frac{N^{C\varepsilon}}{N}\int_\eta^t \sum_{k:|\gamma_k-E|\leq u} \frac{1}{|\gamma_k - z_{t-u}|^4}\frac{|\gamma_k-E|^2}{N^2u^2}\mathrm{d}u \\
&\leq \frac{N^{C\varepsilon}}{N^2t^2} + N^{C\varepsilon}\int_\eta^t \frac{1}{N^2u^2}\left(\int_{-u}^u \frac{x^2}{x^4+(\eta+t-u)^4}\mathrm{d}x\right)\mathrm{d}u \\
&\leq \frac{N^{C\varepsilon}}{N^2t^2}.
\end{aligned}
$$

For the contribution of $\Lambda(a,N,t,\eta,u)$, we have

$$
\begin{aligned}
&\frac{N^{C\varepsilon}}{N}\int_\eta^t \sum_{|k|<(1-c)N} \frac{1}{|\gamma_k - z_{t-u}|^4}\left(\Lambda^2\wedge\frac{1}{N^2}\right)\mathrm{d}u \\
&\leq \frac{N^{C\varepsilon}}{N}\int_\eta^t \sum_{|k|<(1-c)N} \frac{1}{|\gamma_k - z_{t-u}|^4}\left[\left(\frac{(t-u)^2}{N^2u^2}+\frac{(Nu)^{2\alpha}}{N^4u^2}+\frac{\eta^2}{N^2t^2}\right)\wedge\frac{1}{N^2}\right]\mathrm{d}u \\
&\leq N^{C\varepsilon}\int_\eta^t \frac{1}{(\eta+t-u)^3}\left[\frac{(Nu)^{2\alpha}}{N^4u^2}+\frac{\eta^2}{N^2t^2}+\left(\frac{(t-u)^2}{N^2u^2}\wedge\frac{1}{N^2}\right)\right]\mathrm{d}u.
\end{aligned}
$$

The first two terms in the bracket give us

$$
\int_\eta^t \frac{1}{(\eta+t-u)^3}\left(\frac{(Nu)^{2\alpha}}{N^4u^2}+\frac{\eta^2}{N^2t^2}\right)\mathrm{d}u \leq \frac{(Nt)^{2\alpha}}{N^4t^2\eta^2}+\frac{1}{N^2t^2}.
$$

For the remaining term, we have

$$
\begin{aligned}
&\int_\eta^t \frac{1}{(\eta+t-u)^3}\left(\frac{(t-u)^2}{N^2u^2}\wedge\frac{1}{N^2}\right)\mathrm{d}u \\
&\qquad\qquad \leq \int_\eta^{\frac{t}{2}} \frac{1}{(\eta+t-u)^3}\frac{1}{N^2}\mathrm{d}u + \int_{\frac{t}{2}}^t \frac{1}{(\eta+t-u)^3}\frac{(t-u)^2}{N^2u^2}\mathrm{d}u \leq \frac{\log N}{N^2t^2}
\end{aligned}
$$

Combining these results shows

$$
\langle M\rangle_{t\wedge\tau} \leq N^{C\varepsilon}\left(\frac{(Nt)^{2\alpha}}{N^4t^2\eta^2}+\frac{1}{N^2t^2}\right).
$$

Using Lemma 2.1 and a union bound, for any fixed large $D > 0$ and sufficiently large $N > N_0(\varepsilon, D)$, we have

$$\mathbb{P}\left(\sup_{\ell,m,p,0\leq s\leq t\wedge\tau,|E_m|<2-b} |M_s| \leq N^{C\varepsilon}\left(\frac{(Nt)^\alpha}{N^2 t\eta} + \frac{1}{Nt}\right)\right) > 1 - N^{-D}.$$

Together with previous estimates, with overwhelming probability we have

$$\sup_{\ell,m,p,0\leq s\leq t\wedge\tau,|E_m|<2-b} |g_s(z)| \leq N^{C\varepsilon}\left(\frac{(Nt)^\alpha}{N^2 t\eta} + \frac{1}{Nt}\right).$$

This implies $\min_{\ell,m,p}\{\tau_{\ell,m,p}\} = \infty$ with overwhelming probability. Moreover, we have shown in Lemma 2.5, Lemma 2.11 that $\tau_0 = \infty$ and $\tau_1 = \infty$ with overwhelming probability. Assuming the hypothesis $\mathcal{H}_\alpha$, we also have $\tau_2 = \infty$ with overwhelming probability. These imply that $\tau = \infty$ with overwhelming probability. Hence we complete the proof. $\square$

Now we move on to the short-range approximation of the dynamics. Recall that $\{\varphi_k\}$ satisfies the parabolic equation (2.20), and we rewrite it as

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi_k = (\mathcal{P}\varphi)_k$$

where the time-dependent operator $\mathcal{P}$ is defined in the following way: For $f : \mathbb{R} \to \mathbb{R}^{2N}$,

$$(\mathcal{P}f)_k := \sum_{j\neq\pm k} c_{jk}(t)(f_j(t) - f_k(t)), \quad c_{jk}(t) := \frac{1}{2N(s_j(t) - s_k(t))^2}.$$

Consider some parameter $l = l(N, \alpha)$ which will be determined later, we decompose the operator $\mathcal{P}$ into two parts $\mathcal{P} = \mathcal{P}_{\text{short}} + \mathcal{P}_{\text{long}}$. The operators $\mathcal{P}_{\text{short}}$ and $\mathcal{P}_{\text{long}}$ represent the short-range interactions and long-range interactions respectively and are defined as follows

$$(\mathcal{P}_{\text{short}} f)_k = \sum_{|j-k|\le l} c_{jk}(t)(f_j(t) - f_k(t)),$$
$$(\mathcal{P}_{\text{long}} f)_k = \sum_{|j-k|> l} c_{jk}(t)(f_j(t) - f_k(t)).$$

Note that the operators $\mathcal{P}_{\text{short}}, \mathcal{P}_{\text{long}}$ are also time dependent. Let $\mathscr{T}_{\text{short}}(s,t)$ denote the semigroup associated with the operator $\mathcal{P}_{\text{short}}$ in the sense

$$\partial_t \mathscr{T}_{\text{short}}(s,t) = \mathcal{P}_{\text{short}}(t)\mathscr{T}_{\text{short}}(s,t), \quad \mathscr{T}_{\text{short}}(s,s) = \mathsf{Id}.$$

Also, let $\mathscr{T}$ denote the semigroup associated with $\mathcal{P}$.

To prove the short-range approximation, we need the following finite speed of propagation estimate for the semigroup. Such estimates were proved in [CL19] with minor changes.

Lemma 2.17. *For any fixed small $c > 0$ and large $D > 0$, there exists $N_0(c, D)$ such that the following holds with probability at least $1 - N^{-D}$. For any $\varepsilon > 0$, $N > N_0$, $0 < u < v < 1$, $l \ge N|u - v|$, $|k| \le (1 - c)N$ and $-N \le j \le N$ such that $|k - j| > N^{\varepsilon} l$, we have*

$$\left(\mathscr{T}_{\text{short}}(u,v)\delta_k\right)(j) < N^{-D}. \tag{2.46}$$

With Lemma 2.17, we have the following short-range approximation estimate. In particular, this short-range approximation can be improved based on the hypothesis $\mathcal{H}_\alpha$.

Lemma 2.18. *Assume $\mathcal{H}_\alpha$. Let $c > 0$ be any fixed small constant. There exists a constant $C > 0$ such that for any $\varepsilon > 0$, $N^{-1+C\varepsilon} < t < 1$, $\frac{t}{2} \le u < v \le t$, $l \ge N^{\varepsilon}$ and $|k| \le (1 - c)N$, we have*

$$\left|\left(\left(\mathscr{T}(u,v) - \mathscr{T}_{\text{short}}(u,v)\right)\varphi(u)\right)_k\right| \prec (v - u)\left(\frac{N}{l}\frac{(Nt)^{\alpha}}{N^2 t} + \frac{1}{Nt}\right). \tag{2.47}$$

Proof. The Duhamel's principle implies

$$((\mathscr{T}(u,v)-\mathscr{T}_{\text{short}}(u,v))\,\varphi(u))_k=\int_u^v\left(\mathscr{T}_{\text{short}}(s,v)[(\mathcal{P}_{\text{long}}\,\varphi)\,(s)]\right)_k\mathrm{d}s.$$

On the event that Lemma 2.17 holds, for $|k|\leq(1-3c)N$, the finite speed of propagation yields

$$\left(\mathscr{T}_{\text{short}}(s,v)[(\mathcal{P}_{\text{long}}\,\varphi)\,(s)]\right)_k=\left(\mathscr{T}_{\text{short}}(s,v)[\left(\mathcal{P}_{\text{long}}\,\varphi\mathbb{1}_{[(2c-1)N,(1-2c)N]}\right)(s)]\right)_k+N^{-D},$$

where $(\varphi\mathbb{1}_{[(2c-1)N,(1-2c)N]})_j=\varphi_j\mathbb{1}_{[(2c-1)N,(1-2c)N]}(j)$. Moreover, using the property that $\mathscr{T}_{\text{short}}$ is an L^∞ contraction, we have

$$|((\mathscr{T}(u,v)-\mathscr{T}_{\text{short}}(u,v))\,\varphi(u))_k|\leq|u-v|\sup_{|j|\leq(1-2c)N,u<s<v}\left|(\mathcal{P}_{\text{long}}\,\varphi)_j\,(s)\right|+N^{-D}. \tag{2.48}$$

For $|i|\leq(1-c)N$, assuming $\mathcal{H}_\alpha$ and on the event that Lemma 2.12 holds, there exists a constant $C>0$ so that

$$|\varphi_i(s)-\varphi_j(s)|\leq|\varphi_i(s)-\widehat{\varphi}_i(s)|+|\varphi_j(s)-\widehat{\varphi}_j(s)|+|\widehat{\varphi}_i(s)-\widehat{\varphi}_j(s)|\lesssim N^{C\varepsilon}\left(\frac{(Nt)^\alpha}{N^2t}+\frac{|i-j|}{N^2t}\right).$$

For $|i|>(1-c)N$, Lemma 2.11 yields

$$|\varphi_i(s)-\varphi_j(s)|\leq|\varphi_i(s)|+|\varphi_j(s)|\lesssim N^{-\frac{2}{3}+C\varepsilon}(N+1-|i|)^{-\frac{1}{3}}.$$

Therefore, using rigidity of singular values, we have

$$\begin{aligned}
\left(\mathcal{P}_{\text{long}}\varphi\right)_j(s) &= \sum_{|i-j|>l} \frac{\varphi_i(s)-\varphi_j(s)}{2N(s_i(s)-s_j(s))^2} \\
&\lesssim N^{1+C\varepsilon} \sum_{|i-j|>l} \frac{1}{(i-j)^2}\left(\frac{(Nt)^\alpha}{N^2t}+\frac{|i-j|}{N^2t}\right) + N^{-1+C\varepsilon} \sum_{|i|>(1-c)N} N^{-\frac{2}{3}}(N+1-|i|)^{-\frac{1}{3}} \\
&\lesssim N^{C\varepsilon}\left(\frac{N}{l}\frac{(Nt)^\alpha}{N^2t}+\frac{1}{Nt}\right).
\end{aligned}$$

Combined with (2.48), this implies the desired claim with $|k| \leq (1-3c)N$. Note that $c>0$ is arbitrary, and thus it concludes the proof. □

Further, we will show that we have nice control for a regularization of the short-range dynamics with a well-behaved initial data. To do this, we follow the techniques developed in [BY17]. Consider some fixed times $u<t$, the short-range parameter l, and define an averaging space window scale r. Throughout the remaining parts of this section, for any fixed arbitrarily small $\varepsilon>0$, we make the following assumption on these parameters

$$N^{30\varepsilon}(t-u) < N^{20\varepsilon}\frac{l}{N} < N^{10\varepsilon} r < t. \tag{2.49}$$

For a fixed index k, as in [BY17], we define the flattening operator with parameter $a>0$ by

$$\left(\mathcal{F}_a f\right)_j(v) := \begin{cases} f_j(v) & \text{if } |j-k| \leq a \\ \widehat{\varphi}_k(t) & \text{if } |j-k| > a \end{cases} \quad \text{for } u \leq v \leq t,$$

and the averaging operator

$$(\mathcal{A}f)_j := \frac{1}{|[Nr,2Nr]|} \sum_{a\in[Nr,2Nr]} \left(\mathcal{F}_a f\right)_j$$

As shown in [BY17, Equation (7.4)], the averaging operator can also be represented as a combination of Lipschitz function, i.e. there exists a Lipschitz function h with $|h_i - h_j| \leq$

$\frac{|i-j|}{Nr}$ such that

$$(\mathcal{A}f)_j = h_j f_j + (1-h_j)\widehat{\varphi}_k(t). \tag{2.50}$$

Finally, for $u < v < t$, consider the regularized dynamics

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}v}\Gamma_j(v) = (\mathcal{P}_{\text{short}}(v)\Gamma)_j\,, \\ \Gamma(u) = \mathcal{A}\varphi(u) \end{cases}$$

The following lemma shows that the averaging the regularized dynamics gives good approximation for $\widehat{\varphi}_k$.

Lemma 2.19. *Assume $\mathcal{H}_\alpha$. Let $c > 0$ be any fixed small constant. There exists a constant $C > 0$ such that for any $\varepsilon > 0$, $N^{-1+C\varepsilon} < \eta, t < 1$, $\frac{t}{2} \leq u < v \leq t$, $l > N^{\varepsilon}$, $j, k \in [(2c-1)N, (1-2c)N]$ such that $|\gamma_j - \gamma_k| < 10r$, and $z = \gamma_j + \mathrm{i}\eta$, we have*

$$\left|\frac{1}{2N}\mathrm{Im}\sum_{|i-j|<l}\frac{\Gamma_i(v)}{s_i(v)-z} - \left(\frac{1}{2N}\mathrm{Im}\sum_{|i-j|<l}\frac{1}{s_i(v)-z}\right)\widehat{\varphi}_k(v)\right| \\ \prec \left(\frac{r}{Nt} + \frac{\eta}{Nt} + \frac{(Nt)^{\alpha}}{N^2 t}\left(\frac{l}{Nr} + \frac{N\eta}{l} + \frac{N(t-u)}{l} + \frac{1}{N\eta}\right)\right). \tag{2.51}$$

Proof. We decompose the upper line of (2.51) into three parts

$$\frac{1}{2N}\mathrm{Im}\sum_{|i-j|<l}\frac{\Gamma_i(v)}{s_i(v)-z} - \left(\frac{1}{2N}\mathrm{Im}\sum_{|i-j|<l}\frac{1}{s_i(v)-z}\right)\widehat{\varphi}_k(v) =: I_1 + I_2 + I_3,$$

where

$$\begin{aligned} I_1 &= \frac{1}{2N}\mathrm{Im}\sum_{|i-j|<l}\frac{(\mathscr{T}_{\text{short}}(u,v)\mathcal{A}\varphi(u) - \mathcal{A}\mathscr{T}_{\text{short}}(u,v)\varphi(u))_i}{s_i(v)-z} \\ I_2 &= \frac{1}{2N}\mathrm{Im}\sum_{|i-j|<l}\frac{(\mathcal{A}\mathscr{T}_{\text{short}}(u,v)\varphi(u) - \mathcal{A}\mathscr{T}(u,v)\varphi(u))_i}{s_i(v)-z} \\ I_3 &= \frac{1}{2N}\mathrm{Im}\sum_{|i-j|<l}\frac{(\mathcal{A}\mathscr{T}(u,v)\varphi(u))_i - \widehat{\varphi}_k(v)}{s_i(v)-z}. \end{aligned}$$

For the first term I_1, Note that

$$
\begin{aligned}
&(\mathscr{T}_{\text{short}}(u,v)\mathcal{A}\varphi(u) - \mathcal{A}\mathscr{T}_{\text{short}}(u,v)\varphi(u))_i = \\
&\qquad \frac{1}{|[Nr,2Nr]|}\sum_{a\in[Nr,2Nr]} (\mathscr{T}_{\text{short}}(u,v)\mathcal{F}_a\varphi(u) - \mathcal{F}_a\mathscr{T}_{\text{short}}(u,v)\varphi(u))_i \,.
\end{aligned}
$$

When $|i-k| < a - N^{\varepsilon}l$, we have

$$
(\mathcal{F}_a\mathscr{T}_{\text{short}}(u,v)\varphi(u))_i = (\mathscr{T}_{\text{short}}(u,v)\varphi(u))_i \,,
$$

and the finite speed of propagation (2.46) yields

$$
(\mathscr{T}_{\text{short}}(u,v)\mathcal{F}_a\varphi(u))_i = (\mathscr{T}_{\text{short}}(u,v)\varphi(u))_i + N^{-D}.
$$

This gives

$$
(\mathscr{T}_{\text{short}}(u,v)\mathcal{F}_a\varphi(u) - \mathcal{F}_a\mathscr{T}_{\text{short}}(u,v)\varphi(u))_i < N^{-D}
$$

Similarly, this bound also holds in the case $|i-k| > a + N^{\varepsilon}l$. Now suppose $a - N^{\varepsilon}l \leq |i-k| \leq a + N^{\varepsilon}l$. Applying (2.30) and Hypothesis $\mathcal{H}_\alpha$, we obtain

$$
\begin{aligned}
&|(\mathscr{T}_{\text{short}}(u,v)\mathcal{F}_a\varphi(u) - \mathcal{F}_a\mathscr{T}_{\text{short}}(u,v)\varphi(u))_i| \\
\leq& \max_{m:||m-k|-a|\leq 2N^{\varepsilon}l} |\varphi_m(v) - \widehat{\varphi}_k(t)| + N^{-D} \\
\leq& \max_{m:||m-k|-a|\leq 2N^{\varepsilon}l} |\varphi_m(v) - \widehat{\varphi}_m(v)| + \max_{m:||m-k|-a|\leq 2N^{\varepsilon}l} |\widehat{\varphi}_m(v) - \widehat{\varphi}_k(t)| + N^{-D} \\
\leq& N^{C\varepsilon}\left(\frac{(Nt)^{\alpha}}{N^2 t} + \frac{r+(t-u)}{Nt}\right).
\end{aligned}
$$

Combined with the estimate above, this implies

$$
I_1 \leq N^{C\varepsilon}\frac{l}{Nr}\left(\frac{(Nt)^{\alpha}}{N^2 t} + \frac{r+(t-u)}{Nt}\right). \tag{2.52}
$$

For the term I_2, note that $|i-j|<l$ implies $|i|\leq (1-c)N$. Therefore, using the Lipschitz representation of the averaging operator (2.50), the short-range approximation (2.47) gives us

$$
\begin{aligned}
&|(\mathcal{A}\mathscr{T}_{\text{short}}(u,v)\varphi(u)-\mathcal{A}\mathscr{T}(u,v)\varphi(u))_i| \\
&\qquad\qquad \leq |\left(\mathscr{T}_{\text{short}}(u,v)\varphi(u)-\mathscr{T}(u,v)\varphi(u)\right)_i| \leq N^{C\varepsilon}(t-u)\left(\frac{N}{l}\frac{(Nt)^\alpha}{N^2t}+\frac{1}{Nt}\right).
\end{aligned}
$$

This shows

$$
I_2 \leq N^{C\varepsilon}(t-u)\left(\frac{N}{l}\frac{(Nt)^\alpha}{N^2t}+\frac{1}{Nt}\right). \tag{2.53}
$$

Finally, for the term I_3, by the Lipschitz representation of the averaging opeator (2.50), it can be rewritten in the following way,

$$
\begin{aligned}
I_3 &= \frac{1}{2N}\mathrm{Im}\sum_{-N\leq i\leq N}\frac{h_j(\varphi_i(v)-\widehat{\varphi}_k(v))}{s_i-z}-\frac{1}{2N}\mathrm{Im}\sum_{|i-j|\geq l}\frac{h_j(\varphi_i(v)-\widehat{\varphi}_k(v))}{s_i-z} \\
&\quad+\frac{1}{2N}\mathrm{Im}\sum_{|i-j|<l}\frac{(h_i-h_j)(\varphi_i(v)-\widehat{\varphi}_k(v))}{s_i-z}+\frac{1}{2N}\mathrm{Im}\sum_{|i-j|<l}\frac{(1-h_i)(\widehat{\varphi}_k(t)-\widehat{\varphi}_k(v))}{s_i-z} \\
&=: J_1+J_2+J_3+J_4.
\end{aligned}
$$

Using (2.31) and (2.37), we control J_1 in the following way

$$
\begin{aligned}
J_1 &\leq \frac{e^{\frac{v}{2}}}{2N}\left(\mathrm{Im}\,\mathfrak{S}_v(z)-e^{-\frac{v}{2}}\frac{\mathrm{Im}\,S_v(z)}{\mathrm{Im}\,S_0(z_v)}\mathrm{Im}\,\mathfrak{S}_0(z_v)\right)+\mathrm{Im}\,S_v(z)\left(\frac{\mathrm{Im}\,\mathfrak{S}_0(z_v)}{N\mathrm{Im}\,S_0(z_v)}-\widehat{\varphi}_k(v)\right) \\
&\leq N^{C\varepsilon}\left(\frac{(Nt)^\alpha}{N^3t\eta}+\frac{1}{N^2t}+\frac{\eta+r}{Nt}\right).
\end{aligned}
$$

Applying (2.30) to estimate J_2, we obtain

$$
\begin{aligned}
J_2 &\le \frac{1}{2N}\sum_{|i-j|\ge l}\frac{\eta}{(\gamma_i-\gamma_j)^2+\eta^2}\left(|\varphi_i(v)-\widehat{\varphi}_i(v)|+|\widehat{\varphi}_i(v)-\widehat{\varphi}_k(v)|\right)\\
&\le \frac{N^{C\varepsilon}}{2N}\sum_{|i-j|\ge l}\frac{\eta}{(\gamma_i-\gamma_j)^2+\eta^2}\left(\frac{(Nt)^\alpha}{N^2t}+\frac{|i-j|}{N^2t}+\frac{Nr}{N^2t}\right)\\
&\le N^{C\varepsilon}\left(\frac{(Nt)^\alpha}{N^2t}\frac{N\eta}{l}+\frac{\eta}{Nt}+\frac{r}{Nt}\right).
\end{aligned}
$$

Similarly, by the Lipschitz property of $\{h_i\}$, we estimate J_3 as follows

$$
\begin{aligned}
J_3 \le \frac{1}{2N}\sum_{|i-j|<l}\frac{\eta}{(\gamma_i-\gamma_j)^2+\eta^2}\frac{|i-j|}{Nr}\left(\frac{(Nt)^\alpha}{N^2t}+\frac{|i-j|}{N^2t}+\frac{Nr}{N^2t}\right)\\
\le N^{C\varepsilon}\frac{l}{Nr}\left(\frac{(Nt)^\alpha}{N^2t}+\frac{r}{Nt}\right)\le N^{C\varepsilon}\left(\frac{(Nt)^\alpha}{N^2t}\frac{l}{Nr}+\frac{r}{Nt}\right).
\end{aligned}
$$

Using the same arguments, by (2.30), we have

$$
J_4 \le \frac{N^{C\varepsilon}}{2N}\sum_{|i-j|<l}\frac{\eta}{(\gamma_i-\gamma_j)^2+\eta^2}\frac{t-v}{Nt}\le N^{C\varepsilon}\frac{r}{Nt}.
$$

Together with the previous estimates, this leads to

$$
I_3 \le N^{C\varepsilon}\left(\frac{r}{Nt}+\frac{\eta}{Nt}+\frac{(Nt)^\alpha}{N^2t}\left(\frac{l}{Nr}+\frac{N\eta}{l}+\frac{1}{N\eta}\right)\right)
$$

Combined with (2.52) and (2.53), we obtain the desired result. □

Finally, we have all the tools to prove Lemma 2.15.

Proof of Lemma 2.15. We fix some small $c>0$ and consider an arbitrarily small $\varepsilon>0$. Throughout the whole proof, we do all estimates on the overwhelming probability event where Lemma 2.17, Lemma 2.18 and Lemma 2.19 hold. For a fixed index $k\in[(2c-$

$1)N, (1-2c)N]$, we have

$$|\varphi_k(t) - \Gamma_k(t)| \leq |((\mathscr{T}(u,t) - \mathscr{T}_{\text{short}}(u,t))\varphi(u))_k| + |(\mathscr{T}_{\text{short}}(u,t)(\varphi(u) - \Gamma(u)))_k| \, .$$

By the definition of the averaging opeartor, we know that $\Gamma(u) = \mathcal{A}\varphi(u) = \varphi(u)$ on the set $\{j : |j-k| \leq Nr\}$. Therefore, combined with the finite speed of propagation estimate (2.46) for the second term and the short-range approximation (2.47) for the first term, we obtain

$$|\varphi_k(t) - \Gamma_k(t)| \leq N^{C\varepsilon}(t-u)\left(\frac{(Nt)^{\alpha}}{N^2 t}\frac{N}{l} + \frac{1}{Nt}\right) + N^{-2022}. \tag{2.54}$$

It suffices to estimate $|\Gamma_k(t) - \widehat{\varphi}_k(t)|$. Consider the function

$$M(v) := \max_{-N \leq i \leq N} \left(\Gamma_i(v) - \widehat{\varphi}_k(t)\right).$$

Similarly as in Lemma 2.6, we can show a parabolic maximum principle for M and consequently M decreases in time. Moreover, note that $\Gamma_i(u) = \widehat{\varphi}_k(t)$ if $|i-k| \geq 2Nr$.

Let $j = j(v)$ to denote the index that attains the maximum. If there exists a time $u \leq v \leq t$ such that $|j-k| > 3Nr$, then in this case the finite speed propagation (2.46) gives us

$$M(t) \leq M(v) = \Gamma_j(v) - \widehat{\varphi}_k(t) \leq N^{-2022}. \tag{2.55}$$

On the other hand, now we assume that $|j(v) - k| < 3Nr$ for all $u \leq v \leq t$. In this case, we have

$$\frac{\mathrm{d}}{\mathrm{d}v}\left(\Gamma_j(v) - \widehat{\varphi}_k(t)\right) = \sum_{|i-j|<l} \frac{\Gamma_i(v) - \Gamma_j(v)}{2N(s_i(v) - s_j(v))^2} \leq \frac{1}{2N}\sum_{|i-j|<l} \frac{\Gamma_i(v) - \Gamma_j(v)}{(s_i(v) - s_j(v))^2 + \eta^2}.$$

This gives us

$$\frac{\mathrm{d}}{\mathrm{d}v}\left(\Gamma_j(v) - \widehat{\varphi}_k(t)\right) \leq \frac{1}{2N\eta}\mathrm{Im}\sum_{|i-j|<l}\frac{\Gamma_i(v)}{s_i(v) - z} - \left(\frac{1}{2N\eta}\mathrm{Im}\sum_{|i-j|<l}\frac{1}{s_i(v) - z}\right)\Gamma_j(v)$$

and therefore

$$\frac{\mathrm{d}}{\mathrm{d}v}\left(\Gamma_j(v)-\widehat{\varphi}_k(t)\right)\le\frac{1}{\eta}\left(\frac{1}{2N}\mathrm{Im}\sum_{|i-j|<l}\frac{\Gamma_i(v)}{s_i(v)-z}-\left(\frac{1}{2N}\mathrm{Im}\sum_{|i-j|<l}\frac{1}{s_i(v)-z}\right)\widehat{\varphi}_k(v)\right)$$
$$+\frac{1}{\eta}\left(\frac{1}{2N}\mathrm{Im}\sum_{|i-j|<l}\frac{1}{s_i(v)-z}\right)\left(\widehat{\varphi}_k(v)-\widehat{\varphi}_k(t)\right)$$
$$+\frac{1}{\eta}\left(\frac{1}{2N}\mathrm{Im}\sum_{|i-j|<l}\frac{1}{s_i(v)-z}\right)\left(\widehat{\varphi}_k(t)-\Gamma_j(v)\right).$$

Applying Lemma 2.19 and Lemma 2.12 yields

$$\frac{\mathrm{d}}{\mathrm{d}v}M(v^+)\lesssim-\frac{1}{\eta}M(v)+\frac{N^{C\varepsilon}}{\eta}\left(\frac{r}{Nt}+\frac{\eta}{Nt}+\frac{(Nt)^\alpha}{N^2t}\left(\frac{l}{Nr}+\frac{N\eta}{l}+\frac{N(t-u)}{l}+\frac{1}{N\eta}\right)\right),$$

where the left-hand side represents the right derivative of M at time v. Let $\eta=\frac{(t-u)}{N^\varepsilon}$, then above inequality leads to

$$M(t)\le N^{C\varepsilon}\left(\frac{r}{Nt}+\frac{(Nt)^\alpha}{N^2t}\left(\frac{l}{Nr}+\frac{N(t-u)}{l}+\frac{1}{N(t-u)}\right)\right)$$

Choosing

$$r=\frac{(Nt)^{\frac{3\alpha}{4}}}{N},\quad l=(Nt)^{\frac{\alpha}{2}},\quad (t-u)=\frac{(Nt)^{\frac{\alpha}{4}}}{N},\tag{2.56}$$

then we have

$$M(t)<N^{C\varepsilon}\frac{(Nt)^{\frac{3\alpha}{4}}}{N^2t}$$

Similarly, this bound also holds for $-\max_{-N\le i\le N}\left(\Gamma_i(s)-\widehat{\varphi}_k(t)\right)$. Combined with (2.54) and (2.55), this completes the proof. □

2.4 Green's Function Comparison

Following the general three-step strategy in the dynamical approach, the derivation of the rate of convergence relies on both the relaxation and the Green function comparison theorem from [EYY12]. In the context of sample covariance matrices, this Lindeberg exchange strategy based on the fourth moments matching condition was first used by Tao and Vu in [TV12]. To obtain an explicit convergence rate, we need a quantitative version of the comparison theorem.

2.4.1 Soft edge

For the statement, we consider a fixed $|E-\lambda_+| < \varphi N^{-2/3}$, a scale $\rho = \rho(N) \in [N^{-1}, N^{-2/3}]$, and a function $f = f(N) : \mathbb{R} \to \mathbb{R}$ satisfying

$$\|f^{(k)}\|_{L^\infty([E,E+\rho])} \leq C_k \rho^{-k}, \quad \|f^{(k)}\|_{L^\infty([E^+,E^++1])} = O(1), \quad 0 \leq k \leq 2.$$

where $E^+ = E + \varphi N^{-2/3}$. We assume that f is non-decreasing on $(-\infty, E^+]$, $f(x) \equiv 0$ for $x < E$ and $f(x) \equiv 1$ for $E + \rho < x \leq E^+$; and also assume f is non-increasing on $[E^+, \infty)$, $f \equiv 0$ for $x > E^+ + 1$. Furthermore, let F be a fixed smooth non-increasing function such that $F(x) \equiv 1$ for $x \leq 0$ and $F(x) \equiv 0$ for $x \geq 1$.

Theorem 2.5 (Quantitative Green function comparison)**.** *There exists $C > 0$ such that the following holds. Let $X^{\mathrm{v}}, X^{\mathrm{w}}$ be data matrices satisfying assumptions* (2.1) *and* (2.2)*, and $H^{\mathrm{v}}, H^{\mathrm{w}}$ be the corresponding sample covariance matrices. Assume that the first three moments of the entries are the same, i.e. for all $1 \leq i \leq M$, $1 \leq j \leq N$ and $1 \leq k \leq 3$ we have*

$$\mathbb{E}^{\mathrm{v}}(x_{ij}^k) = \mathbb{E}^{\mathrm{w}}(x_{ij}^k).$$

Assume also that for some parameter $t = t(N)$ we have

$$\left|\mathbb{E}^{\mathrm{v}}(\sqrt{M}x_{ij})^4 - \mathbb{E}^{\mathrm{w}}(\sqrt{M}x_{ij})^4\right| \le t.$$

With the above notations for the test functions f and F, we have

$$|(\mathbb{E}^{\mathrm{v}} - \mathbb{E}^{\mathrm{w}})F(\operatorname{Tr} f(H))| \le \varphi^C \left(\frac{1}{N^{18}\rho^{20}} + \frac{t}{N\rho} + \frac{1}{(N\rho)^2} + \frac{1}{N^2}\right).$$

Proof. We follow the notations in [PY14] and the reasoning from [EY17b, Theorem 17.4]. Fix a bijective ordering map on the index set of the independent matrix elements, $\phi : \{(i,j) : 1 \le i \le M, 1 \le j \le N\} \to \{1, \cdots, MN\}$ and define the family of random matrices X_γ, $0 \le \gamma \le MN$

$$[X_\gamma]_{ij} = \begin{cases} [X^{\mathrm{v}}]_{ij} & \text{if } \phi(i,j) > \gamma, \\ [X^{\mathrm{w}}]_{ij} & \text{if } \phi(i,j) \le \gamma. \end{cases}$$

Note that in particular we have $X_0 = X^{\mathrm{v}}$ and $X_{MN} = X^{\mathrm{w}}$. Denote sample covariance matrices H_γ as

$$H_\gamma := X_\gamma^* X_\gamma.$$

Let χ be a fixed, smooth, symmetric cutoff function such that $\chi(x) = 1$ if $|x| < 1$ and $\chi(x) = 0$ if $|x| > 2$. By the Helffer-Sjöstrand formula, if λ_i's are the (real) eigenvalues of a matrix H, we have

$$\sum f(\lambda_i) = \int_{\mathbb{C}} g(z) \operatorname{Tr} \frac{1}{H - z} \mathrm{d}m(z),$$

where $\mathrm{d}m$ is the Lebesgue measure on $\mathbb{C}$, and the function g is defined as

$$g(z) := \frac{1}{\pi}\left(\mathrm{i}yf''(y)\chi(y) + \mathrm{i}(f(x) + \mathrm{i}yf'(x))\chi'(y)\right), \quad z = x + \mathrm{i}y.$$

Define

$$\Xi^H := \int_{|y|>N^{-1}} g(z)\,\mathrm{Tr}(H-z)^{-1}\mathrm{d}m(z),$$

and we have the bound (see [Bou22, Section 5.2])

$$\left|\sum f(\lambda_i) - \Xi^H\right| \leq O\left(\frac{\varphi^C}{(N\rho)^2}\right).$$

This shows that it suffices to show

$$\left|\mathbb{E}F(\Xi^{H_\gamma}) - \mathbb{E}F(\Xi^{H_{\gamma-1}})\right| \leq \frac{\varphi^C}{N^2}\left(\frac{1}{N^{18}\rho^{20}} + \frac{t}{N\rho} + \frac{1}{(N\rho)^2} + \frac{1}{N^2}\right). \tag{2.57}$$

For an arbitrarily fixed γ corresponding to (i,j), we can write

$$X_{\gamma-1} = Q + V, \quad V := X_{ij}^{\mathrm{v}} E^{(ij)}, \quad X_\gamma = Q + W, \quad W := X_{ij}^{\mathrm{w}} E^{(ij)}.$$

where Q coincides with $X_{\gamma-1}$ and X_γ except on the (i,j) position (where it is 0). We define the Green functions

$$R := (Q^*Q - z)^{-1}, \quad S := (H_{\gamma-1} - z)^{-1}.$$

By Taylor expansion, for some fixed order m, we have

$$\begin{aligned}\mathbb{E}F(\Xi^{H_\gamma}) - \mathbb{E}F(\Xi^{H_{\gamma-1}}) &= \sum_{l=1}^{m-1} \mathbb{E}\frac{F^{(l)}(\Xi^Q)}{l!}\left((\Xi^{H_\gamma} - \Xi^Q)^l - (\Xi^{H_{\gamma-1}} - \Xi^Q)^l\right) \\ &\quad + O\left(\|F^{(m)}\|_\infty\right)\left(\mathbb{E}\left((\Xi^{H_\gamma} - \Xi^Q)^m + (\Xi^{H_{\gamma-1}} - \Xi^Q)^m\right)\right). \end{aligned} \tag{2.58}$$

First we estimate the m-th order error term. For a matrix M, we denote $\|M\|_\infty =$

$\max_{i,j} |M_{ij}|$. By the first order resolvent expansion we have

$$
\begin{aligned}
&|\Xi^{H_\gamma} - \Xi^Q| \\
&\leq \int_{|y|>N^{-1}, |x|<\lambda_+ + 2} |g(z)|\, |\widetilde{r}R(z)(V^*Q + Q^*V + V^*V)S(z)|\, \mathrm{d}m(z) \\
&\leq \varphi^C N \int_{|y|>N^{-1}, |x|<\lambda_+ + 2} |g(z)| \|S(z)\|_\infty \|R(z)\|_\infty \mathrm{d}m(z)
\end{aligned}
$$

with overwhelming probability, where we use the fact that there are only $O(N)$ nonzero entries with size $O(N^{-1})$ in the matrix $V^*Q+Q^*V+V^*V$. By the strong local Marchenko-Pastur law ([PY14, Theorem 3.1]), for any $D > 0$ we have

$$
\mathbb{P}\left(\max_j |S_{jj}(z) - m_{\mathsf{MP}}(z)| + \max_{j \neq k} |S_{jk}(z)| \leq \varphi^C \left(\frac{1}{Ny} + \sqrt{\frac{\operatorname{Im} m_{\mathsf{MP}}(y)}{Ny}} \right) \right) > 1 - N^{-D}.
$$

Same bound for $\|R(z)\|_\infty$ also holds (see [PY14, Lemma 5.4]). This shows that

$$
\mathbb{E}(\Xi^{H_\gamma} - \Xi^Q)^m = O\left(\varphi^C/(N^m \rho^m)\right), \quad \mathbb{E}(\Xi^{H_{\gamma-1}} - \Xi^Q)^m = O\left(\varphi^C/(N^m \rho^m)\right).
$$

Therefore the m-th order term in (2.58) can be bounded by $\varphi^C N^{-2}(N^{-m+2}\rho^{-m})$.

Next we consider the first order term in the Taylor expansion. By the resolvent expansion, we have

$$
S = R - RA^{\mathrm{v}}R + (RA^{\mathrm{v}})^2 R - (RA^{\mathrm{v}})^3 R + \cdots - (RA^{\mathrm{v}})^{11} R + (RA^{\mathrm{v}})^{12} S,
$$

where

$$
A^{\mathrm{v}} = V^*Q + Q^*V + V^*V.
$$

Denote

$$
\widehat{R}_{\mathrm{v}}^{(n)} := (-1)^n \widetilde{r}(RA^{\mathrm{v}})^n R, \quad \Omega_{\mathrm{v}} := \widetilde{r}(RA^{\mathrm{v}})^{12} S.
$$

Then we have

$$\mathbb{E}F'(\Xi^Q)\left(\Xi^{H_{\gamma-1}}-\Xi^{H_\gamma}\right)$$
$$=\mathbb{E}F'(\Xi^Q)\int g(z)\left(\sum_{n=1}^{11}\left(\widehat{R}_{\mathrm{v}}^{(n)}-\widehat{R}_{\mathrm{w}}^{(n)}\right)+(\Omega_{\mathrm{v}}-\Omega_{\mathrm{w}})\right)\mathrm{d}m(z).$$

Since the first three moments of the two matrices are identical, we know that the case $n=1$ gives null contribution.

For $n=2$, note that the entries of the matrix A satisfy the following relation

$$A_{ab}=\begin{cases} x_{ij}x_{ib} & \text{if } a=j, b\neq j,\\ x_{ij}x_{ia} & \text{if } a\neq j, b=j,\\ 0 & \text{otherwise.}\end{cases}$$

This shows that

$$\mathbb{E}(\widehat{R}_{\mathrm{v}}^{(2)}-\widehat{R}_{\mathrm{w}}^{(2)})\leq N\left(\frac{t}{N^2}\right)\left(\max_{i\neq j}|R_{ij}|\right)^2\left(\max_i|R_{ii}|\right),$$

where we used that in the expansion

$$\begin{aligned}\widetilde{r}(RA^{\mathrm{v}})^2R&=\sum_k\sum_{a_1,b_1,a_2,b_2}R_{ka_1}A^{\mathrm{v}}_{a_1b_1}R_{b_1a_2}A^{\mathrm{v}}_{a_2b_2}R_{b_2k}\\&=\sum_k\sum_{(a_1,b_1,a_2,b_2)\neq(j,j,j,j)}R_{ka_1}A^{\mathrm{v}}_{a_1b_1}R_{b_1a_2}A^{\mathrm{v}}_{a_2b_2}R_{b_2k}\\&\quad+\sum_k R_{kj}A^{\mathrm{v}}_{jj}R_{jj}A^{\mathrm{v}}_{jj}R_{jk}\end{aligned}$$

due to the moment matching condition, the terms that make nontrivial contribution are only in the second summation, which is

$$\sum_k R_{kj}R_{jj}R_{jk}(X^{\mathrm{v}}_{ij})^4.$$

Here we also use the fact that the contribution for the terms with k equal i or j is combinatorially negligible. By the local law, we conclude

$$\mathbb{E}F'(\Xi^Q)\int g(z)\left(\widehat{R}_{\mathrm{v}}^{(2)}-\widehat{R}_{\mathrm{w}}^{(2)}\right)\mathrm{d}m(z)=O\left(\frac{\varphi^C t}{N}\right)\int\frac{|g(z)|}{(Ny)^2}\mathrm{d}m(z)=O\left(\frac{\varphi^C}{N^2}\frac{t}{N\rho}\right).$$

For the terms $n=3,\ldots,11$, as explained in [PY14, Lemma 5.4], their contributions are of smaller order. Similarly, for the term $(\Omega_{\mathrm{v}}-\Omega_{\mathrm{w}})$, as shown in [PY14, Lemma 5.4] we have $\Omega_{\mathrm{v}}=O(N^{-4})$. Therefore we have

$$\mathbb{E}F'(\Xi^Q)\int g(z)\left((\Omega_{\mathrm{v}}-\Omega_{\mathrm{w}})\right)\mathrm{d}m(z)=O\left(\frac{\varphi^C}{N^2}\frac{1}{N^2\rho}\right)=O\left(\frac{\varphi^C}{N^2}\right)\left(\frac{1}{(N\rho)^2}+\frac{1}{N^2}\right).$$

Moreover, as explained in [EY17b, Theorem 17.4], the contributions of higher order terms in Taylor expansion are of smaller order. By taking $m=20$, we can ensure that the error caused by Taylor expansion will be dominated by other terms (we will see the detailed reason in (2.73) in the next section). Combining all estimates above gives us (2.57). Finally, a telescopic summation yields the desired result. □

2.4.2 Weak local law and hard edge

For a fixed constant $a\in(1,2)$, let $\rho=\rho(N)\in[N^{-a},N^{-1}]$ be a cutoff scaling. Let $r>0$ and consider two symmetric functions $f_1(x), f_2(x)$ that are non-increasing in $|x|$, given by

$$f_1(x):=\begin{cases}0 & \text{if } |x|>rN^{-1}\\ 1 & \text{if } |x|<rN^{-1}-\rho\end{cases},\quad f_2(x):=\begin{cases}0 & \text{if } |x|>rN^{-1}+\rho\\ 1 & \text{if } |x|<rN^{-1}\end{cases}.$$

Also, consider a fixed non-increasing smooth function F such that $F(x)=1$ for $x\le 0$ and $F(x)=0$ for $x\ge 1$.

A key observation is that the functions f_1, f_2 and F can bound the distribution of the smallest singular value $\sigma_1(X)$. For any function $f:\mathbb{R}\to\mathbb{R}$, we denote $\operatorname{Tr} f(X):=$

$\sum_{i=-N}^{N} f(\sigma_i(X))$.

Lemma 2.20. *We have*

$$\mathbb{E}\left[F\left(\operatorname{Tr} f_2(X)\right)\right] \le \mathbb{P}\left(\sigma_1(X) > rN^{-1}\right) \le \mathbb{E}\left[F\left(\operatorname{Tr} f_1(X)\right)\right], \tag{2.59}$$

Proof. For the right-hand side, assume $\sigma_1(X) > rN^{-1}$. By definition of the function f_1, we have $\sum_{i=-N}^{N} f_1(\sigma_i(X)) = 0$, which implies $F(\operatorname{Tr} f_1(X)) = 1$. Also note that $F \ge 0$. Therefore, we conclude $\mathbb{1}\left\{\sigma_1(X) > rN^{-1}\right\} \le F(\operatorname{Tr} f_1(X))$ and this yields

$$\mathbb{P}\left(\sigma_1(X) > rN^{-1}\right) = \mathbb{E}\left[\mathbb{1}\left\{\sigma_1(X) > rN^{-1}\right\}\right] \le \mathbb{E}\left[F(\operatorname{Tr} f_1(X))\right].$$

The left-hand side can be proved similarly. □

When estimating the distribution $\mathbb{P}\left(\sigma_1(X) > rN^{-1}\right)$, thanks to the rigidity of singular values, we can assume $r < N^{\varepsilon}$ without loss of generality, where $\varepsilon > 0$ is a constant that can be arbitrarily small. Based on Lemma 2.20, to compare the distribution of the smallest singular values of different random matrices, it suffices to compare the functions $\operatorname{Tr} f_1$ and $\operatorname{Tr} f_2$. In the remaining part of this section, we provide a systematic treatment of such a comparison.

Pick a point $E \in \mathbb{R}$ with $0 < E < N^{-1+\varepsilon}$. Let $f(x)$ be a smooth symmetric function that is non-increasing in $|x|$ satisfying

$$f(x) = \begin{cases} 0 & \text{if } |x| > E \\ 1 & \text{if } |x| < E - \rho \end{cases}, \quad \text{and } \|f^{(k)}\|_\infty \lesssim \rho^{-k} \text{ for } k = 1, 2. \tag{2.60}$$

For the test functions f and F defined as above, we have the following quantitative comparison of the resolvents.

Proposition 2.3. *Let X and Y be two independent random matrices satisfying* (2.1) *and* (2.2)*. Assume the first three moments of the entries are identical, i.e.* $\mathbb{E}[X_{ij}^k] = \mathbb{E}[Y_{ij}^k]$ *for* *all* $1 \le i, j \le N$ *and* $1 \le k \le 3$*. Suppose also that for some parameter* $t = t(N)$ *we have*

$$\left| \mathbb{E}[(\sqrt{N} X_{ij})^4] - \mathbb{E}[(\sqrt{N} Y_{ij})^4] \right| \le t, \text{ for all } 1 \le i, j \le N. \tag{2.61}$$

With the test functions f *and* F *defined as above, there exists a constant* $C > 0$ *such that the following is true for any* $\varepsilon > 0$

$$|\mathbb{E}[F(\operatorname{Tr} f(X))] - \mathbb{E}[F(\operatorname{Tr} f(Y))]| \le N^{C\varepsilon} \left(\frac{1}{\rho N^2} + \frac{(\rho N^a)^5}{\sqrt{N}} + t\rho N^a \right). \tag{2.62}$$

Proof. The key idea is based on the Lindeberg exchange method (for a detailed introduction we refer to the monograph [EY17a, VH14]). We first fix an ordering map of the indices $\phi : \{(i,j) : 1 \le i, j \le N\} \to [N^2]$. For $0 \le k \le N^2$, let H_k be the random matrix defined as

$$(A_k)_{ij} = \begin{cases} X_{ij} & \text{if } \phi(i,j) \le k \\ Y_{ij} & \text{otherwise} \end{cases},$$

so that $A_0 = Y$ and $A_{N^2} = X$. By telescoping summation, it suffices to show the following is true uniformly in $1 \le k \le N^2$,

$$|\mathbb{E}[F(\operatorname{Tr} f(A_k))] - \mathbb{E}[F(\operatorname{Tr} f(A_{k-1}))]| \le \frac{N^{C\varepsilon}}{N^2} \left(\frac{1}{\rho N^2} + \frac{(\rho N^a)^5}{\sqrt{N}} + t\rho N^a \right) \tag{2.63}$$

To prove (2.63), we use the Helffer-Sjöstrand formula. Let χ be a smooth symmetric cutoff function such that $\chi(y) = 1$ if $|y| < N^{-a}$ and $\chi(y) = 0$ if $|y| > 2N^{-a}$, with $\|\chi'\|_\infty \le N^a$. For any matrix $A \in \{A_k\}_{k=0}^{N^2}$, let $\widetilde{H}$ denote its Girko symmetrization

$$\widetilde{A} = \begin{pmatrix} 0 & A^\top \\ A & 0 \end{pmatrix}$$

Recall that the symmetrized singular values $\{\sigma_i(A)\}_{i=-N}^{N}$ are the eigenvalues of $\widetilde{A}$. With the cutoff function χ, applying Lemma 2.2 to $\widetilde{A}$ yields

$$\operatorname{Tr} f(A) = \int_{\mathbb{C}} g(z) \operatorname{Tr}(\widetilde{A} - z)^{-1} \, \mathrm{d}^2 z, \tag{2.64}$$

where $\mathrm{d}^2 z$ is the Lebesgue measure on $\mathbb{C}$ and

$$g(z) := \frac{1}{\pi} \left[\mathrm{i} y f''(x) \chi(y) + \mathrm{i} \left(f(x) + \mathrm{i} y f'(x) \right) \chi'(y) \right], \quad z = x + y\mathrm{i}.$$

The analysis of the comparison can be proceeded in the following steps:

Step 1: Approximation of $\operatorname{Tr} f(A)$. We first truncate the integral in (2.64) and define

$$\mathcal{T}(A) := \int_{|y| > N^{-2}} g(z) \operatorname{Tr}(\widetilde{A} - z)^{-1} \mathrm{d}^2 z.$$

The approximation error can be bounded by

$$\begin{aligned} |\operatorname{Tr} f(A) - \mathcal{T}(A)| &\lesssim \iint_{|y| < N^{-2}, E < |x| < E + \rho} |f''(x)| \sum_{-N \le k \le N} \frac{y^2}{|\sigma_k - (x + y\mathrm{i})|^2} \mathrm{d}x \mathrm{d}y \\ &\lesssim \int_{E < |x| < E + \rho} \frac{1}{\rho^2 N^2} \left(\frac{1}{N^2} \sum_{-N \le k \le N} \frac{1}{|\sigma_k - (x + \frac{\mathrm{i}}{N})|^2} \right) \mathrm{d}x. \end{aligned}$$

For singular values near the origin, i.e. $|k| \le N^{C\varepsilon}$, we have

$$\int_{E < |x| < E + \rho} \frac{1}{|\sigma_k - (x + \frac{\mathrm{i}}{N})|^2} \mathrm{d}x \le \int_{E < |x| < E + \rho} N^2 \mathrm{d}x \lesssim \rho N^2.$$

On the other hand, for $|k| > N^{C\varepsilon}$, by the rigidity of singular values, we have the following overwhelming probability bound

$$\int_{E < |x| < E + \rho} \frac{1}{|\sigma_k - (x + \frac{\mathrm{i}}{N})|^2} \mathrm{d}x \le \frac{\rho}{|E - \gamma_k|^2}.$$

Combining the above two bounds together, we obtain

$$|\operatorname{Tr} f(A) - \mathcal{T}(A)| \le \frac{N^{C\varepsilon}}{\rho N^2} + \frac{1}{\rho N^4} \sum_{|k|>N^{C\varepsilon}} \frac{1}{|E-\gamma_k|^2} \lesssim \frac{N^{C\varepsilon}}{\rho N^2} + \frac{1}{\rho N^3} \left(\frac{1}{N} \sum_{k\ge 1} \frac{1}{(k/N)^2} \right) \lesssim \frac{N^{C\varepsilon}}{\rho N^2}$$

with overwhelming probability.

Step 2: Expansions and moment matching. With the approximation by $\mathcal{T}(A)$, it suffices to show that

$$|\mathbb{E}[F(\mathcal{T}(A_k))] - \mathbb{E}[F(\mathcal{T}(A_{k-1}))]| \lesssim \frac{N^{C\varepsilon}}{N^2} \left(\frac{(\rho N^a)^5}{\sqrt{N}} + t\rho N^a \right) \tag{2.65}$$

uniformly for all $1 \le k \le N^2$. Now consider a fixed $1 \le \omega \le N^2$ corresponding to the index (i,j), i.e. $\phi(i,j) = \omega$. We rewrite the matrices A_ω and $A_{\omega-1}$ in the following way

$$A_\omega = W + \frac{1}{\sqrt{N}} U, \quad A_{\omega-1} = W + \frac{1}{\sqrt{N}} V,$$

where the matrix W coincides with $A_{\omega-1}$ and A_ω except on the (i,j) entry with $W_{ij} = 0$. Then note that the matrices U, V satisfy $U_{ij} = \sqrt{N} X_{ij}$ and $V_{ij} = \sqrt{N} Y_{ij}$ and all other entries are zero. Recall the notation $\widetilde{A}$ for the Girko symmetrization. Consider the resolvents of the matrices $\widetilde{W}$ and $\widetilde{A}_\omega$

$$R := (\widetilde{W} - z)^{-1}, \quad S := (\widetilde{A}_\omega - z)^{-1}.$$

The Taylor expansion yields

$$\begin{aligned}
&\mathbb{E}[F(\mathcal{T}(A_\omega))] - \mathbb{E}[F(\mathcal{T}(A_{\omega-1}))] \\
&\quad = \sum_{k=1}^{4} \mathbb{E} \left[\frac{F^{(k)}(\mathcal{T}(W))}{k!} \left((\mathcal{T}(A_\omega) - \mathcal{T}(W))^k - (\mathcal{T}(A_{\omega-1}) - \mathcal{T}(W))^k \right) \right] \\
&\qquad + O(\|F^{(5)}\|_\infty)\, \mathbb{E} \left[(\mathcal{T}(A_\omega) - \mathcal{T}(W))^5 + (\mathcal{T}(A_{\omega-1}) - \mathcal{T}(W))^5 \right].
\end{aligned} \tag{2.66}$$

We first control the term corresponding to the fifth derivative. By Lemma 2.3, the first order resolvent expansion gives us

$$\frac{1}{N}\operatorname{Tr} S - \frac{1}{N}\operatorname{Tr} R = \frac{1}{\sqrt{N}}\operatorname{Tr} S\widetilde{U}R.$$

Consequently,

$$|\mathcal{T}(A_\omega) - \mathcal{T}(W)| \leq \frac{1}{\sqrt{N}}\int_{|y|>N^{-2}} |g(z)| \left|\operatorname{Tr} S(z)\widetilde{U}R(z)\right| \mathrm{d}^2 z.$$

We can restrict the integral on the domain $\{z = x + y\mathrm{i} : N^{-2} < |y| < 2N^{-a}, E < |x| < E + \rho\}$ as the contribution outside this region is negligible. Moreover, a key observation is that the matrix $\widetilde{U}$ only has two non-zero entries. Thus,

$$|\mathcal{T}(A_\omega) - \mathcal{T}(W)| \lesssim N^{\frac{1}{2}+C\varepsilon}\int_{N^{-2}<|y|<2N^{-a}, E<|x|<E+\rho} |g(z)| \Big(\max_{k\neq\ell}|S_{k\ell}(z)|\Big)\Big(\max_{k\neq\ell}|R_{k\ell}(z)|\Big)\mathrm{d}^2 z$$
$$+ N^{-\frac{1}{2}+C\varepsilon}\int_{N^{-2}<|y|<2N^{-a}, E<|x|<E+\rho} |g(z)| \left(\max_{k}|S_{kk}(z)|\right)\left(\max_{k}|R_{kk}(z)|\right)\mathrm{d}^2 z.$$

Note that in this integral domain, the scale of $|y|$ is smaller than the natural size of the local law. Therefore, we will use a suboptimal version of the local semicircle law for a larger spectral domain, which was discussed in [EKYY13, LS18]. For $z = x + \mathrm{i}y$ in this integral domain, with overwhelming probability we have

$$\max_{k,\ell}|S_{k\ell}(z) - \delta_{k\ell}\, m_{sc}(z)| \leq N^{C\varepsilon}\left(\frac{1}{\sqrt{N}} + \Psi(z)\right), \quad \Psi(z) = \frac{1}{Ny} + \sqrt{\frac{\operatorname{Im} m_{sc}(z)}{Ny}}.$$

The same result also holds for $R_{k\ell}(z)$. By Lemma 2.4, we have $\Psi(z) \lesssim \frac{1}{\sqrt{Ny}}$ for z in the integral domain. Note that the contribution of the diagonal resolvent entries is negligible. Therefore, with overwhelming probability we have

$$|\mathcal{T}(A_\omega) - \mathcal{T}(W)| \lesssim N^{C\varepsilon}N^{1/2}\int_{N^{-2}<|y|<2N^{-a}, E<|x|<E+\rho}\frac{|g(z)|}{Ny}\mathrm{d}^2 z \lesssim N^{C\varepsilon}N^{-1/2}\rho N^{a}.$$

Similarly, this bound also holds for $|\mathcal{T}(A_{\omega-1}) - \mathcal{T}(W)|$, and we obtain

$$\mathbb{E}[(\mathcal{T}(A_\omega)-\mathcal{T}(W))^5] \lesssim \frac{N^{C\varepsilon}}{N^2}\left(N^{-\frac{1}{2}}(\rho N^a)^5\right), \ \mathbb{E}[(\mathcal{T}(A_{\omega-1})-\mathcal{T}(W))^5] \lesssim \frac{N^{C\varepsilon}}{N^2}\left(N^{-\frac{1}{2}}(\rho N^a)^5\right).$$

Hence the fifth order term in (2.66) is bounded by

$$O(\|F^{(5)}\|_\infty)\,\mathbb{E}\left[(\mathcal{T}(A_\omega) - \mathcal{T}(W))^5 + (\mathcal{T}(A_{\omega-1}) - \mathcal{T}(W))^5\right] \lesssim \frac{N^{C\varepsilon}}{N^2}\frac{(\rho N^a)^5}{\sqrt{N}}. \tag{2.67}$$

Now we consider the first term $k = 1$ in the Taylor expansion (2.66). Denote

$$\widehat{R} := \frac{1}{N}\operatorname{Tr} R, \ \ \widehat{R}_X^{(m)} = \frac{(-1)^m}{N}\operatorname{Tr}(R\widetilde{U})^m R, \ \ \Omega_X := -\frac{1}{N}\operatorname{Tr}(R\widetilde{U})^5 S,$$

and also define

$$\widehat{R}_Y^{(m)} := \frac{(-1)^m}{N}\operatorname{Tr}(R\widetilde{V})^m R, \ \ \Omega_Y := -\frac{1}{N}\operatorname{Tr}(R\widetilde{V})^5(\widetilde{A}_{\omega-1} - z)^{-1}.$$

Using the resolvent expansion (Lemma 2.3) up to the fifth order, we obtain

$$\frac{1}{N}\operatorname{Tr} S = \widehat{R} + \sum_{m=1}^{4} N^{-\frac{m}{2}}\widehat{R}_X^{(m)} + N^{-\frac{5}{2}}\Omega_X.$$

A Similar expansion also holds for $(\widetilde{A}_{\omega-1} - z)^{-1}$. Then we have

$$\begin{aligned}
&\mathbb{E}\left[F'(\mathcal{T}(W))\,(\mathcal{T}(A_\omega) - \mathcal{T}(A_{\omega-1}))\right] \\
&= \mathbb{E}\left[F'(\mathcal{T}(W))\int_{|y|>N^{-2}} g(z)\left(\sum_{m=1}^{4} N^{-\frac{m}{2}+1}(\widehat{R}_X^{(m)} - \widehat{R}_Y^{(m)}) + N^{-\frac{3}{2}}(\Omega_X - \Omega_Y)\right) \mathrm{d}^2 z\right]
\end{aligned} \tag{2.68}$$

A key observation is that for $1 \le m \le 3$, the terms $\widehat{R}_X^{(m)}$ and $\widehat{R}_Y^{(m)}$ only depend on the first three moments of X_{ij} and Y_{ij}. Recall that the first three moments of X_{ij} and Y_{ij}

are identical. Therefore, the terms corresponding to $1 \le m \le 3$ in (2.68) makes no contribution.

Step 3: Higher order error. For the $m = 4$ term in (2.68), note that

$$\operatorname{Tr}(R\widetilde{U})^4 R = \sum_{1 \le \ell \le 2N} \sum_{\{\alpha_k, \beta_k\} = \{i+N, j\}} R_{\ell\alpha_1} \widetilde{U}_{\alpha_1\beta_1} R_{\beta_1\alpha_2} \widetilde{U}_{\alpha_2\beta_2} R_{\beta_2\alpha_3} \widetilde{U}_{\alpha_3\beta_3} R_{\beta_3\alpha_4} \widetilde{U}_{\alpha_4\beta_4} R_{\beta_4\ell}. \tag{2.69}$$

A similar formula is also true for $\operatorname{Tr}(R\widetilde{V})^4 R$. Note that typically we have $\ell \neq \alpha_1$ and $\ell \neq \beta_4$, but we may have $\beta_1 = \alpha_2$, $\beta_2 = \alpha_3$, $\beta_3 = \alpha_4$. Moreover, the terms with either $\ell = 1 + N$ or $\ell = j$ are combinatorially negligible in the summation and therefore we can ignore these terms in the following computations. Recall that the difference between the fourth moments of X_{ij} and Y_{ij} is bounded by t. Thus, we have

$$\mathbb{E}\left[N(\widehat{R}_X^{(4)} - \widehat{R}_Y^{(4)})\right] = \mathbb{E}\left[\operatorname{Tr}(R\widetilde{U})^4 R - \operatorname{Tr}(R\widetilde{V})^4 R\right] \lesssim Nt \left(\max_{k \neq \ell} |R_{k\ell}|\right)^2 \left(\max_k |R_{kk}|\right)^3.$$

As mentioned above, for the integral in (2.68) we can restrict the integral domain to $N^{-2} < |y| < 2N^{-a}$ and $E < |x| < E + \rho$. In this region, the entries of the resolvent are bound by $\max_{k \neq \ell} |R_{k\ell}(z)| \lesssim \frac{N^{C\varepsilon}}{\sqrt{Ny}}$ and $\max_k |R_{kk}(z)| \lesssim N^{C\varepsilon}$. As a consequence,

$$\begin{aligned}
&\left|\mathbb{E}\left[F'(\mathcal{T}(W)) \int_{|y|>N^{-2}} g(z) N^{-\frac{4}{2}+1} (\widehat{R}_X^{(4)} - \widehat{R}_Y^{(4)}) \mathrm{d}^2 z\right]\right| \\
&\qquad \lesssim N^{C\varepsilon} \frac{t}{N} \int_{N^{-2}<|y|<2N^{-a}, E<|x|<E+\rho} \frac{|g(z)|}{Ny} \mathrm{d}^2 z \lesssim \frac{N^{C\varepsilon}}{N^2} (t\rho N^a).
\end{aligned} \tag{2.70}$$

For the term $\Omega_X - \Omega_Y$, since these terms involve the higher moments of X_{ij} and Y_{ij}, we simply bound it by the size of Ω_X and Ω_Y. By a similar expansion as in (2.69) and the local law, we have $|\Omega_X|, |\Omega_Y| \lesssim \frac{N^{C\varepsilon}}{Ny}$. Therefore,

$$\begin{aligned}&\left|\mathbb{E}\left[F'(\mathcal{T}(W))\int_{|y|>N^{-2}} g(z)N^{-\frac{3}{2}}\Omega_X \mathrm{d}^2 z\right]\right| \\ &\lesssim N^{C\varepsilon}N^{-\frac{3}{2}}\int_{N^{-2}<|y|<2N^{-a}, E<|x|<E+\rho} \frac{|g(z)|}{Ny} d^2 z \lesssim \frac{N^{C\varepsilon}}{N^2}\frac{\rho N^a}{\sqrt{N}} \leq \frac{N^{C\varepsilon}}{N^2}\frac{(\rho N^a)^5}{\sqrt{N}}. \end{aligned} \tag{2.71}$$

The same bound also holds for Ω_Y.

Finally, as explained in classical literature of random matrix theory (see e.g. [EY17a, Theorem 17.4]), the contributions of higher order terms in the Taylor expansion (2.66) are of smaller order. Consequently, combining (2.67), (2.70) and (2.71) yields the claim (2.65), which implies the desired result (2.62). □

2.5 Quantitative Universality

2.5.1 Extremal eigenvalue at soft edge

Recall the normalizing constant γ defined in (2.3). Let $s \in \mathbb{R}$. If $|s| > \varphi$, due to the rigidity we know that for any $D > 0$ and large enough N, we have $\mathbb{P}\left(\gamma N^{2/3}(\lambda_N - \lambda_+) \leq s\right) = \mathbb{P}(\mathsf{TW} \leq s) + O(N^{-D})$. So in the following discussion we assume $|s| \leq \varphi$.

Denoting a non-decreasing function f_1 such that $f_1(x) = 1$ for $x > \lambda_+ + s\gamma^{-1}N^{-2/3}$ and $f_1(x) = 0$ for $x < \lambda_+ + s\gamma^{-1}N^{-2/3} - \rho$. We also define $f_2(x) := f_1(x - \rho)$. Then we have

$$\mathbb{E}_H F\left(\sum_{i=1}^N f_1(\lambda_i)\right) \leq \mathbb{P}_H\left(\lambda_N < \lambda_+ + s\gamma^{-1}N^{-2/3}\right) \leq \mathbb{E}_H F\left(\sum_{i=1}^N f_2(\lambda_i)\right). \tag{2.72}$$

Moreover, as discussed in [EYY11, PY14], we can find an $M \times N$ matrix $\widetilde{X}_0$ such that the Gaussian divisible ensemble $\widetilde{X}_t := e^{-t/2}\widetilde{X}_0 + (1 - e^{-t})^{1/2}X_G$, where X_G is a matrix whose entries are independent Gaussian random variables with mean 0 and variance 1, satisfies the following: for $1 \le k \le 3$,

$$\mathbb{E}(\sqrt{M}X_{ij})^k = \mathbb{E}[\widetilde{X}_t]_{ij}^k, \quad |\mathbb{E}(\sqrt{M}X_{ij})^4 - \mathbb{E}[\widetilde{X}_t]_{ij}^4| \lesssim t.$$

By the quantitative Green function comparison theorem 2.5, we obtain the following bound

$$\begin{aligned}\mathbb{E}_{\widetilde{X}_t} F\left(\sum_{i=1}^{N} f_1(\lambda_i)\right) - \varphi^C\left(\frac{1}{N^{18}\rho^{20}} + \frac{t}{N\rho} + \frac{1}{(N\rho)^2} + \frac{1}{N^2}\right) \\ \le \mathbb{P}_H\left(\lambda_N < \lambda_+ + s\gamma^{-1}N^{-2/3}\right) \le \\ \mathbb{E}_{\widetilde{X}_t} F\left(\sum_{i=1}^{N} f_2(\lambda_i)\right) + \varphi^C\left(\frac{1}{N^{18}\rho^{20}} + \frac{t}{N\rho} + \frac{1}{(N\rho)^2} + \frac{1}{N^2}\right).\end{aligned}$$

Using (2.72) for $\widetilde{X}_t$, the estimate becomes

$$\begin{aligned}\mathbb{P}_{\widetilde{X}_t}\left(\lambda_N < \lambda_+ + s\gamma^{-1}N^{-2/3} - \rho\right) - \varphi^C\left(\frac{1}{N^{18}\rho^{20}} + \frac{t}{N\rho} + \frac{1}{(N\rho)^2} + \frac{1}{N^2}\right) \\ \le \mathbb{P}_H\left(\lambda_N < \lambda_+ + s\gamma^{-1}N^{-2/3}\right) \le \\ \mathbb{P}_{\widetilde{X}_t}\left(\lambda_N < \lambda_+ + s\gamma^{-1}N^{-2/3} + \rho\right) + \varphi^C\left(\frac{1}{N^{18}\rho^{20}} + \frac{t}{N\rho} + \frac{1}{(N\rho)^2} + \frac{1}{N^2}\right).\end{aligned}$$

After combined with the edge relaxation Theorem 2.3, the estimate now gives us

$$\begin{aligned}\mathbb{P}_{\mathsf{Wishart}}\left(\gamma N^{2/3}(\lambda_N - \lambda_+) < s - \gamma N^{2/3}\rho - \frac{\gamma N^{\varepsilon}}{N^{1/3}t}\right) \\ - \varphi^C\left(\frac{1}{N^{18}\rho^{20}} + \frac{t}{N\rho} + \frac{1}{(N\rho)^2} + \frac{1}{N^2}\right) \\ \le \mathbb{P}_H\left(\gamma N^{2/3}(\lambda_N - \lambda_+) < s\right) \le \\ \mathbb{P}_{\mathsf{Wishart}}\left(\gamma N^{2/3}(\lambda_N - \lambda_+) < s + \gamma N^{2/3}\rho + \frac{\gamma N^{\varepsilon}}{N^{1/3}t}\right) \\ + \varphi^C\left(\frac{1}{N^{18}\rho^{20}} + \frac{t}{N\rho} + \frac{1}{(N\rho)^2} + \frac{1}{N^2}\right).\end{aligned}$$

Moreover, as shown in [EK06, Ma12], we know

$$\mathbb{P}_{\mathsf{Wishart}}\left(\gamma N^{2/3}(\lambda_N-\lambda_+)<s\right)=\mathbb{P}(\mathsf{TW}<s)+O(N^{-2/3}).$$

By using this Wishart result and the boundedness of the density for TW, we obtain

$$\begin{aligned}&\mathbb{P}_H\left(\gamma N^{2/3}(\lambda_N-\lambda_+)<s\right)-\mathbb{P}\left(\mathsf{TW}<s\right)\\&\qquad=O\left(N^{\varepsilon}\right)\left(N^{2/3}\rho+\frac{1}{N^{1/3}t}+\frac{1}{N^{18}\rho^{20}}+\frac{t}{N\rho}+\frac{1}{(N\rho)^2}+\frac{1}{N^{2/3}}\right).\end{aligned}\tag{2.73}$$

The optimal bound $N^{-2/9+\varepsilon}$ is obtained for $t=N^{-1/9}$ and $\rho=N^{-8/9}$. This completes the whole proof for Theorem 2.1.

2.5.2 Smallest singular value at hard edge

Using the quantitative comparison theorem (Proposition 2.3) and the smoothed analysis (Theorem 2.4), we now prove the quantitative universality.

For a general random matrix X satisfying Assumptions (2.1) and (2.2), there exists another matrix X_0' that also satisfies the same assumptions such that the matrix $X_t':=e^{-t/2}X_0'+(1-e^{-t})^{1/2}G$ has the same first three moments as X and the difference between the fourth moments (in the sense of (2.61)) is $O(t)$. This is guaranteed by [EYY11, Lemma 3.4].

Lemma 2.20 and Proposition 2.3 yields

$$\begin{aligned}\mathbb{E}\left[F(\operatorname{Tr} f_2(X_t'))\right]-N^{C\varepsilon}\left(\frac{1}{\rho N^2}+\frac{(\rho N^a)^5}{\sqrt{N}}+t\rho N^a\right)&\le\mathbb{P}\left(\sigma_1(X)>rN^{-1}\right)\\&\le\mathbb{E}\left[F(\operatorname{Tr} f_1(X_t'))\right]+N^{C\varepsilon}\left(\frac{1}{\rho N^2}+\frac{(\rho N^a)^5}{\sqrt{N}}+t\rho N^a\right).\end{aligned}\tag{2.74}$$

Using Lemma 2.20 for X_t' with f_1 and f_2 shifted by $\pm\rho$, we have

$$\mathbb{P}\left(\sigma_1(X_t') > \frac{r}{N} + \rho\right) - N^{C\varepsilon}\left(\frac{1}{\rho N^2} + \frac{(\rho N^a)^5}{\sqrt{N}} + t\rho N^a\right) \leq \mathbb{P}\left(\sigma_1(X) > \frac{r}{N}\right)$$
$$\leq \mathbb{P}\left(\sigma_1(X_t') > \frac{r}{N} - \rho\right) + N^{C\varepsilon}\left(\frac{1}{\rho N^2} + \frac{(\rho N^a)^5}{\sqrt{N}} + t\rho N^a\right). \quad (2.75)$$

Using smoothed analysis Theorem 2.4, we have

$$\mathbb{P}\left(\sigma_1(G) > \frac{r}{N} + \rho + \frac{1}{N^2 t}\right) - N^{C\varepsilon}\left(\frac{1}{\rho N^2} + \frac{(\rho N^a)^5}{\sqrt{N}} + t\rho N^a\right) \leq \mathbb{P}\left(\sigma_1(X) > \frac{r}{N}\right)$$
$$\leq \mathbb{P}\left(\sigma_1(G) > \frac{r}{N} - \rho - \frac{1}{N^2 t}\right) + N^{C\varepsilon}\left(\frac{1}{\rho N^2} + \frac{(\rho N^a)^5}{\sqrt{N}} + t\rho N^a\right). \quad (2.76)$$

Taking $\rho = N^{-a}$, $t = N^{a-2}$ and setting $a = 1 + \delta$, we obtain the optimal bounds

$$\mathbb{P}\left(N\sigma_1(G) > r + N^{-\delta}\right) - N^{C\varepsilon}\left(N^{-1+\delta} \vee N^{-\frac{1}{2}}\right) \leq \mathbb{P}\left(N\sigma_1(X) > r\right)$$
$$\leq \mathbb{P}\left(N\sigma_1(G) > r - N^{-\delta}\right) + N^{C\varepsilon}\left(N^{-1+\delta} \vee N^{-\frac{1}{2}}\right). \quad (2.77)$$

Hence, thanks to the arbitrariness of ε, we have proved Theorem 2.2.

Finally, for the complex case, using the exact formula for the distribution of $\sigma_1(G_{\mathbb{C}})$, we obtain a rate of convergence to the limiting law. Recall that

$$\mathbb{P}(N\sigma_1(G_{\mathbb{C}}) \leq r) = \int_0^r e^{-x}\mathrm{d}x = 1 - e^{-r}.$$

Proof of Corollary 2.2. For the complex case, the previous arguments are still valid. Therefore, we still have (2.77). Since $N\sigma_1(G_{\mathbb{C}})$ has a bounded density, we have

$$\mathbb{P}\left(N\sigma_1(G_{\mathbb{C}}) \leq r\right) - N^{\varepsilon}\left(N^{-\delta} + (N^{-1+\delta} \vee N^{-1/2})\right) \leq \mathbb{P}\left(N\sigma_1(X_{\mathbb{C}}) \leq r\right)$$
$$\mathbb{P}\left(N\sigma_1(G_{\mathbb{C}}) \leq r\right) + N^{\varepsilon}\left(N^{-\delta} + (N^{-1+\delta} \vee N^{-1/2})\right) \quad (2.78)$$

Choosing $\delta = 1/2$, we obtain the optimal estimate

$$\mathbb{P}\left(N\sigma_1(X_{\mathbb{C}}) \leq r\right) = 1 - e^{-r} + N^{-\frac{1}{2}+\varepsilon},$$

which proves the desired result. $\square$

2.6 Largest Eigenvalue with General Population

In this section, we proceed to generalize our previous results for sample covariance matrices of type X^*X (which corresponds to the identity population) and aim to derive the rate of convergence to the Tracy-Widom distribution for the (rescaled) largest eigenvalue of separable sample covariance matrices with general population. Throughout this section, we will follow the notations and the setup in the the work by Lee and Schnelli [LS16].

Let $X = (x_{ij})$ be defined as in (2.1) and (2.2). For some deterministic $M \times M$ matrix T, the sample covariance matrices associated with data matrix X and population matrix $\Sigma := T^*T$ is defined as $\mathcal{Q} := (TX)(TX)^*$. Note that the $M \times M$ matrix $\mathcal{Q}$ and the matrix

$$Q := X^*\Sigma X$$

share the same non-trivial eigenvalues. Besides avoiding trivial eigenvalues, the matrix Q is sometimes more technically amenable than $\mathcal{Q}$. For more detailed explanations, we refer to [DY18]. Since we are studying the largest eigenvalue of the sample covariance matrix, it is more convenient to work with the matrix Q. We denote the eigenvalues of Q in increasing order by $\mu_1 \leq \cdots \leq \mu_N$.

As mentioned previously, for the null case (i.e. the population matrix is identity), it is well known that the empirical eigenvalue distribution of a sample covariance matrix converges weakly in probability to the Marchenko-Pastur law. Under the general setting, however, this result need to be modified and the limiting measure (called the deformed

Marchenko-Pastur law) will depend on the spectrum of the population matrix. Let $s_1 \leq \cdots \leq s_M$ be the eigenvalues of the population matrix Σ, we denote by $\widehat{\rho} = \widehat{\rho}(M)$ the empirical eigenvalue distribution of Σ, which is defined as

$$\widehat{\rho} := \frac{1}{M} \sum_{j=1}^{M} \delta_{s_j}.$$

The deformed Marchenko-Pastur law $\widehat{\rho}_{\mathsf{fc}}$ is defined in the following way. The Stieltjes transform $\widehat{m}_{\mathsf{fc}}$ of the probability measure is given by the unique solution of the equation

$$\widehat{m}_{\mathsf{fc}}(z) = \frac{1}{-z + \xi^{-1} \int \frac{1}{t\widehat{m}_{\mathsf{fc}}(z)+1} \mathrm{d}\widehat{\rho}(t)}, \quad \operatorname{Im} \widehat{m}_{\mathsf{fc}}(z) \geq 0, \quad z \in \mathbb{C}^+.$$

It has been discussed in [KY17] that $\widehat{m}_{\mathsf{fc}}$ is associated to a continuous probability density $\widehat{\rho}_{\mathsf{fc}}$ with compact support in $[0, \infty)$. Moreover, the density $\widehat{\rho}_{\mathsf{fc}}$ can be obtained from $\widehat{m}_{\mathsf{fc}}$ via the Stieltjes inversion formula

$$\widehat{\rho}_{\mathsf{fc}}(E) = \lim_{\eta \downarrow 0} \frac{1}{\pi} \operatorname{Im} \widehat{m}_{\mathsf{fc}}(E + \mathrm{i}\eta).$$

The typical location of the largest eigenvalue, which is the rightmost endpoint of the support of the density $\widehat{\rho}_{\mathsf{fc}}$ is determined in the following way. Recall that $\xi := N/M$, we define ξ_+ as the largest solution of the equation

$$\int \left(\frac{t\xi_+}{1 - t\xi_+} \right)^2 \mathrm{d}\widehat{\rho}_{\mathsf{fc}}(t) = \xi.$$

We remark that ξ_+ is unique and $\xi_+ \in [0, s_M^{-1}]$. We then introduce the typical location for the largest eigenvalue E_+ by

$$E_+ := \frac{1}{\xi_+} \left(1 + \xi^{-1} \int \frac{t\xi_+}{1 - t\xi_+} \mathrm{d}\widehat{\rho}_{\mathsf{fc}}(t) \right). \tag{2.79}$$

Now we state our assumptions on the population matrix Σ that are needed to prove the explicit rate of convergence. For general random matrices X we require Σ to be diagonal, and we will show later this diagonal condition can be removed if X is Gaussian. We further need the following assumption for the spectrum of the population matrix Σ. Throughout this section, we assume the following:

$$\liminf_{M} s_1 > 0, \quad \limsup_{M} s_M < \infty, \quad \text{and} \quad \limsup_{M} s_M \xi_+ < 1. \tag{2.80}$$

The assumption (2.80) is the same as [LS16, Assumption 2.2]. It is first used in [BPZ15, KY17] to prove the local deformed Marchenko-Pastur law. In particular, the last inequality ensures that the density $\widehat{\rho}_{\mathsf{fc}}$ exhibits a square-root behavior near the right edge of its support, which is crucial to derive the local law.

It is natural to note that with a general population matrix, the distribution of the largest eigenvalue should not behave exactly like the null case. Besides the typical location of the largest eigenvalue is changed, the normalization constant of the fluctuation is also different. Therefore, we introduce the following normalization constant γ_0 given by

$$\frac{1}{\gamma_0^3} = \frac{1}{\xi} \int \left(\frac{t}{1 - t\xi_+} \right) \mathrm{d}\widehat{\rho}(t) + \frac{1}{\xi_+^3}. \tag{2.81}$$

Moreover, we remark that the Tracy-Widom limit for the general case is rescaled and it is not the same as the previous one we used for the null case. However they differ only by a simple scaling so that we do not emphasize this difference and still use the notation TW to denote this distribution. Under this framework, our main result given in Corollary 2.1.

Unlike the proof for the null case in the previous sections, we will not strictly follow the three-step strategy in the dynamical approach. Instead, we use the comparison theorem for the Green function flow, which is a method based on continuous interpolation, linearization and renormalization developed in [LS16].

2.6.1 Local deformed Marchenko-Pastur law

For completeness, we will briefly introduce the local deformed Marchenko-Pastur law. Though we will not give a detailed proof, we emphasize that the local law is an indispensable part to prove the Green function comparison Proposition 2.4, which further leads to the edge universality.

For small nonnegative $c, \epsilon \geq 0$ and sufficiently large $E_+ < C < \infty$, we consider the domain

$$\mathcal{D}(c,\epsilon) := \left\{ z = E + \mathrm{i}\eta \in \mathbb{C}^+ : E_+ - c \leq E \leq C, N^{-1+\epsilon} \leq \eta \leq 1 \right\}.$$

We also denote $\kappa = \kappa(E) := |E - E_+|$. Then we have the following estimates for the density and the Stieltjes transform of the deformed Marchenko-Pastur law.

Lemma 2.21 (Theorem 3.1 in [BPZ15])**.** *Under the assumption* (2.80)*, there exists a constant $c > 0$ such that*

$$\widehat{\rho}_{\mathsf{fc}}(E) \sim \sqrt{E_+ - E} \quad E \in [E_+ - 2c, E_+].$$

Moreover, the Stieltjes transform $\widehat{m}_{\mathsf{fc}}$ satisfies the following: for $z \in \mathcal{D}(c, 0)$, we have

$$|\widehat{m}_{\mathsf{fc}}(z)| \sim 1, \quad \operatorname{Im} \widehat{m}_{\mathsf{fc}}(z) \sim \begin{cases} \dfrac{\eta}{\sqrt{\kappa+\eta}} & \text{if } E \geq E_+ + \eta, \\ \sqrt{\kappa+\eta}, & \text{if } E \in [E_+ - c, E_+ + \eta). \end{cases}.$$

The Green function and the Stieltjes transform are defined in the usual way:

$$G_Q(z) := (Q - z)^{-1}, \quad m_Q(z) := \frac{1}{N} \operatorname{Tr} G_Q(z).$$

Then we have the following local law for the separable sample covariance matrix Q.

Lemma 2.22 (Theorem 3.2 and Theorem 3.3 in [BPZ15]). *Under the assumption* (2.80)*, for any sufficiently small* $\epsilon > 0$*, and for any (large)* $D > 0$*, there exists* $N_0(D) > 0$ *such that for any* $N \geq N_0(D)$ *we have the following estimate uniformly in* $z \in \mathcal{D}(c, \epsilon)$*:*

$$\mathbb{P}\left(|m_Q(z) - \widehat{m}_{\mathsf{fc}}(z)| \leq \frac{N^\epsilon}{N\eta}\right) > 1 - N^{-D},$$

and

$$\mathbb{P}\left(\max_{i,j} |(G_Q)_{ij}(z) - \delta_{ij}\widehat{m}_{\mathsf{fc}}(z)| \leq N^\epsilon \left(\sqrt{\frac{\operatorname{Im}\widehat{m}_{\mathsf{fc}}(z)}{N\eta}} + \frac{1}{N\eta}\right)\right) > 1 - N^{-D}.$$

It is clear to see that the estimates for the deformed Marchenko-Pastur law (Lemma 2.21) and the local law (Lemma 2.22) are greatly similar as the corresponding results for the null case (see e.g. [PY14, Theorem 3.1]). In the framework of the Erdős-Schlein-Yau approach, the proof of edge universality essentially only depends on the local law. Since the integrable Gaussian model has Tracy-Widom fluctuation at the edge, heuristically, this deformed local Marchenko-Pasture law implies the Tracy-Widom limit in the edge universality for the non-null case.

2.6.2 Interpolation flow and Green's functions

In classical theory of random matrix universality, the tool needed to prove the edge universality is the Green function comparison theorem. The usual approach is to compare two ensembles with some moments matching conditions, and then use the construction of Gaussian divisible ensembles together with estimates of the local relaxation flow to remove the moments matching requirement. In this section, however, we do not follow this traditional step. Instead, we compare the Green function of a general ensemble with its corresponding null sample covariance matrix. This argument was first introduced in [LS15] to handle the deformed Wigner matrices, and used in [LS16] to identify the Tracy-Widom limit for general separable sample covariance matrices. The basic idea is to introduce a time evolution that deforms the population matrices continuously to the identity and offset the change of the Green function by a renormalization of the matrix.

Recall the normalization constant γ_0 defined in (2.81). We consider the following two rescaled matrices

$$\widetilde{\Sigma} := \gamma_0 \Sigma, \quad \widetilde{Q} := X^* \widetilde{\Sigma} X.$$

We also denote the eigenvalues of $\widetilde{Q}$ by $\widetilde{\mu}_1 \leq \cdots \leq \widetilde{\mu}_N$, and let $L_+ := \gamma_0 E_+$. We remark that in the literature about sample covariance matrices with general population (e.g. [BPZ15, EK07, LS16]), the scaling of the Tracy-Widom distribution is chosen in the way such that it is the limit for the distribution of the (rescaled) largest eigenvalue of the matrix

$$W := \sqrt{\xi}(1+\sqrt{\xi})^{-4/3} X^* X.$$

Specifically, we order the eigenvalues of the matrix W by $\lambda_1 \leq \cdots \leq \lambda_N$, and let M_+ denote the rightmost endpoint of the rescaled Marchenko-Pastur law for W. It has been shown in [BPZ15, equation (1.9)] that

$$\gamma_0 = \sqrt{\xi}(1+\sqrt{\xi})^{-4/3} + o(1).$$

This can be regarded as a good motivation for considering the normalization constant γ_0.

For the diagonal population matrix $\Sigma = \operatorname{diag}(s_j)$, we introduce the following time evolution $t \mapsto (s_j(t))$ that deforms Σ to the identity matrix I and the Green function flow by

$$\frac{1}{s_j(t)} = e^{-t} \frac{1}{s_j(0)} + (1 - e^{-t}), \quad \widetilde{Q}(t) = \gamma_0 X^* \Sigma(t) X, \tag{2.82}$$

and

$$m_{\widetilde{Q}(t)}(z) := \frac{1}{N} \operatorname{Tr}(\widetilde{Q}(t) - z)^{-1}.$$

Based on the local law Lemma 2.21 and Lemma 2.22, and a delicate analysis for the time derivative of the Green function for $\widetilde{Q}(t)$, the Green function comparison theorem (see Proposition 2.4) is proved in [LS16]. We note that though the original estimate in [LS16] is not explicit, a careful examination of the proof will reveal that the result is actually quantitative.

Proposition 2.4 (Theorem 4.1 in [LS16])**.** *Let $\varepsilon > 0$ and set $\eta = N^{-2/3-\varepsilon}$. Let $E_1, E_2 \in \mathbb{R}$ satisfy $E_1 < E_2$ and $|E_1|, |E_2| \leq N^{-2/3+\varepsilon}$. Let $F : \mathbb{R} \to \mathbb{R}$ be a smooth function satisfying*

$$\max_x |F^{(l)}(x)|(|x|+1)^{-C} \leq C, \quad l = 1, 2, 3, 4.$$

Then for any (small) $\delta > 0$ and for sufficiently large N we have

$$\Bigg|\mathbb{E}F\left(N\int_{E_1}^{E_2} \operatorname{Im} m_{\widetilde{Q}}(x + L_+ + \mathrm{i}\eta)\mathrm{d}x\right) \\ -\mathbb{E}F\left(N\int_{E_1}^{E_2} \operatorname{Im} m_W(x + M_+ + \mathrm{i}\eta)\mathrm{d}x\right)\Bigg| \leq N^{-\frac{1}{3}+2\varepsilon+\delta}.$$

2.6.3 Deriving the rate of convergence

In this section we can finally prove Corollary 2.1. The proof is based on our previous rate of convergence for the null case (Theorem 2.1) and the estimate on the comparison theorem for the Green function flow (Proposition 2.4).

Proof of Corollary 2.1. We first remark that we are not supposed to use the rate of convergence for the null case (Theorem 2.1) and the triangle inequality in a naive way to derive the convergence rate for the general case. This is because Theorem 2.1 is obtain by choosing the optimal parameters in the estimate (2.73), and the scale parameter ρ in (2.73) is also related to the scale in the Green function comparison (Proposition 2.4). The optimal parameter in Theorem 2.1 may be suboptimal in Proposition 2.4, and therefore the naive triangle inequality does not produce the optimal estimate. To illustrate this link more clearly, we first briefly review how Green function comparison is used to obtain the edge universality.

We introduce a smooth cutoff function $K : \mathbb{R} \to \mathbb{R}$ satisfying

$$K(x) = \begin{cases} 1 & \text{if } x \le 1/9, \\ 0 & \text{if } x \ge 2/9. \end{cases}$$

and we also define the Poisson kernel θ_η, for $\eta > 0$

$$\theta_\eta(x) := \frac{\eta}{\pi(x^2+\eta^2)}.$$

Let $E_* := L_+ + \varphi^C N^{-2/3}$, and denote $\chi_E := 1_{[E,E_*]}$. For $\varepsilon > 0$, let $l := \frac{1}{2}N^{-2/3-\varepsilon}$ and $\eta := N^{-2/3-9\varepsilon}$. Then for any (large) $D > 0$, it is proved in [LS16, PY14] that for large enough N we have

$$\mathbb{E}K\left(\mathrm{Tr}(\chi_{E-l} * \theta_\eta(\widetilde{Q}))\right) \le \mathbb{P}\left(\widetilde{\mu}_N \le E\right) \le \mathbb{E}K\left(\mathrm{Tr}(\chi_{E+l} * \theta_\eta(\widetilde{Q}))\right) + N^{-D}. \qquad (2.83)$$

Here the parameter l plays the same role as the ρ in (2.73), and therefore we have $N^{-\varepsilon} = N^{2/3}\rho$ and $\eta = N^{-2/3}N^6\rho^9$. By the Green function comparison Proposition 2.4 and the [LS16, Theorem 2.4], we have

$$\begin{aligned} \mathbb{P}\left(N^{2/3}(\lambda_N - M_+) \le s\right) - N^{-\frac{1}{3}+45\varepsilon+\delta} &\le \mathbb{P}\left(N^{2/3}(\widetilde{\mu}_N - L_+) \le s\right) \\ &\le \mathbb{P}\left(N^{2/3}(\lambda_N - M_+) \le s\right) + N^{-\frac{1}{3}+45\varepsilon+\delta}. \end{aligned}$$

This gives us

$$\mathsf{d}_{\mathsf{K}}\left(\gamma_0 N^{2/3}(\mu_N - E_+), N^{2/3}(\lambda_N - M_+)\right) \le N^\delta N^{-1/3} N^{-30} \rho^{-45}. \qquad (2.84)$$

Combined with (2.73), by triangle inequality we finally obtain

$$
\begin{aligned}
&\mathsf{d_K}\left(\gamma_0 N^{2/3}(\mu_N - E_+), \mathsf{TW}\right) \\
&\leq \mathsf{d_K}\left(\gamma_0 N^{2/3}(\mu_N - E_+), N^{2/3}(\lambda_N - M_+)\right) + \mathsf{d_K}\left(N^{2/3}(\lambda_N - M_+), \mathsf{TW}\right) \\
&\leq N^{\delta}\left(N^{-\frac{1}{3}}N^{-30}\rho^{-45} + N^{\frac{2}{3}}\rho + \frac{N^{-\frac{1}{3}}}{t} + \frac{1}{N^{18}\rho^{20}} + \frac{t}{N\rho} + \frac{1}{(N\rho)^2} + N^{-\frac{2}{3}}\right).
\end{aligned}
$$

The optimal result is obtained now by choosing $\rho = N^{-31/46}$ and $t = N^{-15/46}$, which gives us

$$
\mathsf{d_K}\left(\gamma_0 N^{2/3}(\mu_N - E_+), \mathsf{TW}\right) \leq N^{-\frac{1}{138}+\delta}.
$$

This completes the proof. □

Based on the Corollary 2.1 for diagonal population matrices Σ and general random matrices X, we can easily obtain the following result for the case in which we can have a general population if the random matrix X is restricted to be Gaussian.

Corollary 2.3. *Let $Q := X^*\Sigma X$ be an $N \times N$ separable sample covariance matrix, where X is an $M \times N$ real random matrix with independent Gaussian entries satisfying* (2.1)*, and Σ is a real positive-definite deterministic $M \times M$ matrix satisfying* (2.80)*. For any $\varepsilon > 0$ and large enough N, we have*

$$
\mathsf{d_K}\left(\gamma_0 N^{2/3}(\mu_N - E_+), \mathsf{TW}\right) \leq N^{-\frac{1}{138}+\varepsilon}
$$

Proof. Under these assumptions, we know that the population matrix Σ is diagonizable, i.e. there exists an $M \times M$ real diagonal matrix D and an $N \times N$ orthogonal matrix U such that $\Sigma = U^* D U$. Since X is a matrix whose entries are independent Gaussian random variable, we know that UX is also a real random matrix with Gaussian entries satisfying the assumption (2.1). Therefore, by applying our Corollary 2.1 to the matrix $X^*\Sigma X = (UX)^* D (UX)$, we will get the desired result. □

Chapter 3

Resampling Sensitivity of High-Dimensional PCA

3.1 Introduction

The study of stability and sensitivity of statistical methods and algorithms with respect to the input data is an important task in machine learning and statistics [BE02, EEPK05, MNPR06, HRS16, DHS21]. The notion of stability for algorithms is also closely related to differential privacy [DR14] and generalization error [KN02]. To measure algorithmic stability, one fundamental question is to study the performance of the algorithm under resampling of its input data [BCRT21, KB21]. Originating from the analysis of Boolean functions [BKS99, GS14], resampling sensitivity (also called noise sensitivity) is an important concept in theoretical computer science, which refers to the phenomenon that resampling a small portion of the random input data may lead to decorrelation of the output.

In this work, we study the resampling sensitivity of principal component analysis (PCA). As one of the most commonly used statistical methods, PCA is widely applied for dimension reduction, feature extraction, etc [Joh07, DT11]. It is also used in other fields such as economics [VK06], finance [ASX17], genetics [Rin08], and so on. The impact of noise on PCA is a significant problem in statistics and machine learning, and has been a subject of extensive research. The performance of PCA under the additive or multiplicative independent perturbation of the data matrix has been well studied (see e.g. [BBAP05, BS06, Pau07, BGN11, CLMW11b, FWZ18]). However, the influence of resampling on the outcome, as a distinct form of data corruption, remains poorly understood. In this paper, we aim to address this issue for the first time. Here, we emphasize that the resampling of the input data may not have any structure, and the specific resampling procedure is given in the next subsection. Our primary findings reveal that PCA is sensitive to resampling when the population covariance matrix of the data lacks a strong signal. In such cases, even resampling a negligible portion of the data can cause a significant alteration in the resulting principal component, rendering it orthogonal to the original direction. Conversely, when the population covariance of the data possesses a strong spike, the planted signal acts to stabilize PCA.

3.1.1 Model and Main Results

Let $\mathbf{z}_1, \dots, \mathbf{z}_n \in \mathbb{R}^p$ be independent random vectors with covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{p \times p}$, i.e. $\mathbb{E}[\mathbf{z}_i \mathbf{z}_i^\top] = \mathbf{\Sigma}$. The sample covariance matrix of the data $\mathbf{z}_1, \dots, \mathbf{z}_n$ is defined as $\mathbf{H} := \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top$, and the principal component of the data refers to the unit eigenvector corresponding to the top eigenvalue of the sample covariance matrix. Equivalently, we can rewrite the sample covariance matrix as $\mathbf{H} = (\mathbf{\Sigma}^{1/2}\mathbf{X})(\mathbf{\Sigma}^{1/2}\mathbf{X})^\top$, where the square root matrix $\Sigma^{1/2}$ is well defined via the spectral decomposition and $\mathbf{X} \in \mathbb{R}^{p \times n}$ is a random matrix whose columns have an isotropic covariance matrix. The assumptions on the data matrix are stated as follows.

Assumption 3.1. *Let $\mathbf{X} = (\mathbf{X}_{ij})$ be an $p \times n$ data matrix with independent real valued entries with mean 0 and variance n^{-1},*

$$\mathbf{X}_{ij} = n^{-1/2} x_{ij}, \quad \mathbb{E}[x_{ij}] = 0, \quad \mathbb{E}[x_{ij}^2] = 1. \tag{3.1}$$

Furthermore, we assume the entries x_{ij} have a sub-exponential decay, that is, there exists a constant $\theta > 0$ such that for $u > 1$,

$$\mathbb{P}(|x_{ij}| > u) \leq \theta^{-1} \exp(-u^{\theta}). \tag{3.2}$$

Note that we do not require the i.i.d. condition for the data. The sub-exponential decay assumption is mainly for convenience, and other conditions such as the finiteness of a sufficiently high moment would be enough.

Motivated by high-dimensional statistics, we will work in the proportional growth regime $n \asymp p$.

Assumption 3.2. *Throughout this paper, to avoid trivial eigenvalues, we will be working in the regime*

$$\lim_{n\to\infty} p/n = \xi \in (0,1) \;\; \textit{or} \;\; p/n \equiv 1.$$

In the case $\lim p/n = 1$, our assumption $p/n \equiv 1$ is due to some technicalities in random matrix theory. Specifically, our proof relies on the delocalization of eigenvectors in the whole spectrum. As one of the major open problems in random matrix theory, delocalization of eigenvectors near the lower spectral edge is not known in the general case with just $\lim p/n = 1$ [AEK14b, BEK+14]. The strictly square assumption $p \equiv n$ can be slightly relaxed to $|n-p| = p^{o(1)}$ (see e.g. [Wan22]), but we do not pursue such an extension for simplicity.

For the population covariance matrix Σ, we are interested in the spiked covariance model, which was initiated by Johnstone [Joh01].

$$\mathbf{\Sigma} = \mathbf{I}_p + \sum_{i=1}^{r} \widetilde{\sigma}_i \mathbf{y}_i \mathbf{y}_i^{\top},$$

where r is a fixed integer, the constants $\{\widetilde{\sigma}_i\}_{i=1}^r$ represent strengths of the signals, and $\{\mathbf{y}_i\}_{i=1}^r$ is an orthonormal basis of eigenvectors. It is well known that the BBP phase transition [BBAP05, BGN11, BKYY16] affirms that if $\widetilde{\sigma}_i > \sqrt{\xi}$, then the i-th spike will give rise to an outlier of the spectrum of the sample covariance matrix $\mathbf{H}$. When all $\widetilde{\sigma}_i \leq \sqrt{\xi}$, we call it a weakly spiked model and in particular when $\mathbf{\Sigma} = \mathbf{I}_p$ we call it the null model. On the other hand, if $\widetilde{\sigma}_i > \sqrt{\xi}$ for some i, we call it a strongly spiked model.

In our work, due to technical reasons, we assume that the population covariance matrix is diagonal.

Assumption 3.3. *The population covariance matrix is diagonal, i.e.* $\mathbf{\Sigma} = \mathsf{diag}(d_1, \cdots, d_p)$ *with constants* $d_1 \geq \cdots \geq d_p \geq 1$. *Moreover, the population covariance matrix differs from the identity by a finite rank, i.e.* $|\{i : d_i \neq 1\}| = r$ *for some fixed integer* r.

The reasons for the diagonal assumption on the population covariance are two-fold: (1) our proof hinges on several technical results in random matrix theory such as delocalization of all eigenvectors and Tracy-Widom concentration of the top eigenvalue, etc. Beyond the null model, these results are only known in the case where the population covariance matrix is diagonal [BKYY16, DY18], etc. (2) The diagonal population covariance implies that the entries in each data vector $\mathbf{z}_i$ are independent. In this way, when implementing the resampling procedure, the diagonal assumption makes resampling of the original data $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n]$ equivalent to resampling of the entries in $\mathbf{X}$. This facilitates the model easier to state and the analysis more tractable.

We order the eigenvalues of the sample covariance matrix $\mathbf{H} := (\mathbf{\Sigma}^{1/2}\mathbf{X})(\mathbf{\Sigma}^{1/2}\mathbf{X})^\top$ as $\lambda_1 \geq \cdots \geq \lambda_p$, and use $\mathbf{v}_i \in \mathbb{R}^p$ to denote the unit eigenvector corresponding to the eigenvalue λ_i. If the context is clear, we just use $\lambda := \lambda_1$ and $\mathbf{v} := \mathbf{v}_1$ to denote the largest eigenvalue and the top eigenvector. We also consider the eigenvalues and eigenvectors of the Gram matrix $\widehat{\mathbf{H}} := (\mathbf{\Sigma}^{1/2}\mathbf{X})^\top(\mathbf{\Sigma}^{1/2}\mathbf{X})$. Note that $\widehat{\mathbf{H}}$ and $\mathbf{H}$ have the same non-trivial eigenvalues, and the spectrum of $\widehat{\mathbf{H}}$ is given by $\{\lambda_i\}_{i=1}^n$ with $\lambda_{p+1} = \cdots = \lambda_n = 0$. We denote the unit eigenvectors of $\widehat{\mathbf{H}}$ associated with the eigenvalue λ_i by $\mathbf{u}_i \in \mathbb{R}^n$. Writing $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_p] \in \mathbb{R}^{n\times p}$ and $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_p] \in \mathbb{R}^{p\times p}$, then these eigenvectors may be connected by the singular value decomposition of the data matrix $\mathbf{\Sigma}^{1/2}\mathbf{X} = \mathbf{VSU}^\top$, where

$\mathbf{S} := \mathsf{diag}(\sigma_1, \cdots, \sigma_p)$ with $\sigma_i = \sqrt{\lambda_i}$ corresponds to the singular values. For convenience, we also denote $\sigma := \sigma_1$. And therefore, up to the sign of the eigenvectors, we have

$$(\mathbf{\Sigma}^{1/2}\mathbf{X})^\top \mathbf{v}_i = \sqrt{\lambda_i}\mathbf{u}_i, \ \ (\mathbf{\Sigma}^{1/2}\mathbf{X})\mathbf{u}_i = \sqrt{\lambda_i}\mathbf{v}_i.$$

We now describe the resampling procedure. We first emphasize that resampling the data $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n]$ is equivalent to resampling the matrix $\mathbf{X}$ as the diagonal assumption of the population $\mathbf{\Sigma}$ ensures that the entries in a data vector $\mathbf{z}_i$ are independent. For a positive number $k \le np$, define the random matrix $\mathbf{X}^{[k]}$ in the following way. Let $S_k = \{(i_1, \alpha_1), \cdots, (i_k, \alpha_k)\}$ be a set of k pairs chosen uniformly at random without replacement from the set of all ordered pairs (i, α) of indices with $1 \le i \le p$ and $1 \le \alpha \le n$. We assume that the set S_k is independent of the entries of $\mathbf{X}$. The entries of $\mathbf{X}^{[k]}$ are given by

$$\mathbf{X}^{[k]}_{i,\alpha} = \begin{cases} \mathbf{X}'_{i,\alpha} & \text{if } (i,\alpha) \in S_k, \\ \mathbf{X}_{i,\alpha} & \text{otherwise,} \end{cases}$$

where $(\mathbf{X}'_{i,\alpha})_{1\le i\le p, 1\le\alpha\le n}$ are independent random variables, independent of $\mathbf{X}$, and $\mathbf{X}'_{i,\alpha}$ has the same distribution as $\mathbf{X}_{i,\alpha}$. In other words, $\mathbf{X}^{[k]}$ is obtained from $\mathbf{X}$ by resampling k random entries of the matrix, and therefore $\mathbf{X}^{[k]}$ clearly has the same distribution as $\mathbf{X}$. Let $\mathbf{H}^{[k]} := (\mathbf{\Sigma}^{1/2}\mathbf{X}^{[k]})(\mathbf{\Sigma}^{1/2}\mathbf{X}^{[k]})^\top$ be the sample covariance matrix corresponding to the resampled matrix $\mathbf{X}^{[k]}$. Denote the eigenvalues and the corresponding normalized eigenvectors of $\mathbf{H}^{[k]}$ by $\lambda_1^{[k]} \ge \cdots \ge \lambda_p^{[k]}$ and $\mathbf{v}_1^{[k]}, \cdots, \mathbf{v}_p^{[k]}$. When the context is clear, the principal component is just denoted by $\lambda^{[k]}$ and $\mathbf{v}^{[k]}$. Similarly, denote the eigenvector of the Gram matrix $\widehat{\mathbf{H}}^{[k]} := (\mathbf{\Sigma}^{1/2}\mathbf{X}^{[k]})^\top(\mathbf{\Sigma}^{1/2}\mathbf{X}^{[k]})$ associated with the eigenvalue $\lambda_i^{[k]}$ by $\mathbf{u}_i^{[k]}$.

The sensitivity and stability of PCA crucially depend on how strong the planted signal in the spiked covariance is. To measure the strength of the spikes in the population covariance matrix, we define the set of outlier indices

$$\mathcal{O} := \{i : d_i > 1 + \sqrt{\xi}\}.$$

If $\mathcal{O} = \emptyset$, then the model is weakly spiked. On the other hand, if $\mathcal{O} \neq \emptyset$, the eigenvalues with indices in $\mathcal{O}$ will be an outlier.

For the weakly spiked model, with the resampling parameter in two different regimes, we have the following results.

Theorem 3.1 (Weakly spiked model: sensitivity)**.** *Suppose the data profile $\mathbf{X}, \mathbf{\Sigma}$ satisfy Assumptions 3.1, 3.2 and 3.3 with $\mathcal{O} = \emptyset$, and let $\mathbf{X}^{[k]}$ be the resampled matrix defined as above. For any $\epsilon_0 > 0$, if $k \geq n^{5/3+\epsilon_0}$, then the associated principal components are asymptotically orthogonal, i.e.*

$$\lim_{n\to\infty} \mathbb{E}\left|\langle \mathbf{v}, \mathbf{v}^{[k]}\rangle\right| = 0, \quad \textit{and} \quad \lim_{n\to\infty} \mathbb{E}\left|\langle \mathbf{u}, \mathbf{u}^{[k]}\rangle\right| = 0. \tag{3.3}$$

Moreover, in the null model, the threshold for k can be improved to $k \gg n^{5/3}$.

Theorem 3.2 (Weakly spiked model: stability)**.** *Suppose the data profile $\mathbf{X}, \mathbf{\Sigma}$ satisfy Assumptions 3.1, 3.2 and 3.3 with $\mathcal{O} = \emptyset$, and let $\mathbf{X}^{[k]}$ be the resampled matrix defined as above. For any $\epsilon_0 > 0$,*

$$\max_{1\leq k\leq n^{5/3-\epsilon_0}} \min_{s\in\{-1,1\}} \sqrt{n}\|\mathbf{v} - s\mathbf{v}^{[k]}\|_\infty \xrightarrow{\text{prob}} 0, \tag{3.4}$$

where $\xrightarrow{\text{prob}}$ means convergence in probability. The same result also holds for $\mathbf{u}$ and $\mathbf{u}^{[k]}$.

These two theorems together state that the critical threshold for the resampling strength is of order $k \asymp n^{5/3}$. Note that $n^{5/3}$ compared with the total number of inputs $np \asymp n^2$ is negligible. We show that, above the threshold $n^{5/3}$, resampling even a negligible portion of the data will result in a dramatic change of the resulting principal component, in the sense that the new principal component is asymptotically orthogonal to the old one; while below the threshold, a relatively mild resampling has almost no effect on the corresponding new principal component. If considering the eigenvector overlaps $|\langle \mathbf{v}, \mathbf{v}^{[k]}\rangle|$ and $|\langle \mathbf{u}, \mathbf{u}^{[k]}\rangle|$, these quantities exhibit sharp phase transitions from 1 to 0 near the critical threshold $k \asymp n^{5/3}$.

We want to make two concluding remarks for the weakly spiked model:

- The phase transition stated in the above theorems is not restricted to the top eigenvectors $\mathbf{v}, \mathbf{v}^{[k]}, \mathbf{u}, \mathbf{u}^{[k]}$. With essentially the same arguments, we can prove that for any fixed $m > 0$, the m-th leading eigenvectors $\mathbf{v}_m, \mathbf{v}_m^{[k]}$ and $\mathbf{u}_m, \mathbf{u}_m^{[k]}$ exhibit the same phase transition at the critical threshold of the same order $n^{5/3}$.

- Although the resampling procedure is done uniformly at random for all entries, the proof proceeds as well if we choose to resample the columns. If we resample K columns uniformly at random, then all results are still valid with the threshold $K \asymp n^{2/3}$.

Finally, we turn our attention to the strongly spiked model. In contrast to the weak signal case, for the strongly spiked model, the spike forces the principal components to be correlated with the planted signals. As a direct consequence of the BBP phase transition, in this case, PCA performs better stability than in the weakly spiked model in the following sense:

Suppose the data profile $\mathbf{X}, \mathbf{\Sigma}$ satisfy Assumptions 3.1, 3.2 and 3.3, and let $\mathbf{X}^{[k]}$ be the resampled matrix defined as above. If $\mathcal{O} \neq \emptyset$, for any $i \in \mathcal{O}$ and $k \geq 0$, almost surely we have

$$\left|\langle \mathbf{v}_i, \mathbf{v}_i^{[k]}\rangle\right| \geq 1 - 4\left(1 - \sqrt{\frac{1 - \frac{\xi}{(d_i-1)^2}}{1 + \frac{\xi}{d_i-1}}}\right) + o(1). \tag{3.5}$$

3.1.2 Related Literature

The resampling sensitivity of the leading eigenvector for Wigner matrices and Erdős-Rényi graphs has been studied in [BLZ20, BL22]. The problem discussed in our paper shares a similar prototype, and in particular the threshold $k \asymp n^{5/3}$ for the stability-sensitivity transition in the weakly spiked model coincides with the the threshold for Wigner matrices. Here we want to highlight the differences between our work and theirs.

(i) Our model, the sample covariance matrix, has a Gram structure. This nonlinearity in the matrix model requires more delicate techniques to analyze and the proofs in previous work on symmetric linear model cannot be directly applied here. To overcome the nonlinearity, an important linearization technique is introduced to reduce the interdependency of entries in the sample covariance matrix.

(ii) Due to the Gram matrix structure, the entries in the sample covariance matrix are correlated. In contrast to the case of symmetric matrices, resampling one entry in the data matrix will result in changes of $\Theta(n)$ entries in the sample covariance matrix. Therefore, it is highly non-trivial that our threshold coincides with the threshold for linear models.

(iii) The most important difference: our work is capable of dealing with heteroskedastic data, while previous works are restricted to matrices with identical variances. This makes the applicability of our result much wider. In particular, we establish a clear understanding for the effect of spikes. Such a study of matrix models with planted signals in our work is beyond the scope of previous works.

Compared with previously mentioned work such as [BBAP05, Pau07, BGN11, FWZ18] that mainly focused on PCA with additive or multiplicative independent noise, our setting is very different. In our model, if writing the resampling effect as an additive or multiplicative perturbation, then this noise is not independent of the signal and does not possess any special structure. In contrast, in previous work, sometimes low-rank assumptions on the structure of the matrix or the noise, or some kind of incoherence conditions were imposed. In our work, we have almost no assumption on the data other than a sub-exponential decay condition. Moreover, we highlight that our results have universality. In particular, we do not need to know the specific distribution of the data and we do not require the data is i.i.d sampled.

A similar framework of PCA with corrupted data is the robust PCA [CLMW11b, XCS10]. Regarding connection with robust PCA, our setting does share some similarities with RPCA, as both settings consider corruptions of the original data. In RPCA, corruption is usually related to outlier distribution and we focus on recovering the data. On the other hand, the resampling sensitivity setting in our work studies corruption of data by an independent copy of the same distribution. The key point of our result is that even for two data matrices with the same marginal distribution, a negligible portion of different entries may result in having drastically different principal components.

3.2 Preliminaries

In this section, we collect some important notions used throughout this chapter as well as some auxiliary results.

For the analysis of the sample covariance matrix, it is useful to apply the linearization trick (see e.g. [Tro12, DY18]). Specifically, we will analyze the symmetrization of $\mathbf{\Sigma}^{1/2}\mathbf{X}$. To ease the representation, we drop the dependency on $\mathbf{\Sigma}$ in the notation since the population matrix is fixed throughout, and the symmetrization is denoted as

$$\widetilde{\mathbf{X}} := \begin{pmatrix} 0 & (\mathbf{\Sigma}^{1/2}\mathbf{X})^\top \\ (\mathbf{\Sigma}^{1/2}\mathbf{X}) & 0 \end{pmatrix} \tag{3.6}$$

The spectrum of the symmetrization $\widetilde{\mathbf{X}}$ are given by the singular values $\{\sqrt{\lambda_m}\}_{m=1}^p$ of

$\mathbf{\Sigma}^{1/2}\mathbf{X}$, the symmetrized singular values $\{-\sqrt{\lambda_m}\}_{m=1}^{p}$, and trivial eigenvalue 0 with multiplicity $n-p$. Let $\mathbf{w}_i := (\mathbf{u}_i^\top, \mathbf{v}_i^\top)^\top \in \mathbb{R}^{n+p}$ be the concatenation of the eigenvectors $\mathbf{u}_i$ and $\mathbf{v}_i$. Then $\mathbf{w}_i$ is the eigenvector of $\widetilde{\mathbf{X}}$ associated with the eigenvalue $\sqrt{\lambda_i}$. Indeed, we have

$$\begin{pmatrix} 0 & (\mathbf{\Sigma}^{1/2}\mathbf{X})^\top \\ (\mathbf{\Sigma}^{1/2}\mathbf{X}) & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{pmatrix} = \begin{pmatrix} (\mathbf{\Sigma}^{1/2}\mathbf{X})^\top \mathbf{v}_i \\ (\mathbf{\Sigma}^{1/2}\mathbf{X})\mathbf{u}_i \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda_i}\mathbf{u}_i \\ \sqrt{\lambda_i}\mathbf{v}_i \end{pmatrix}.$$

An important probabilistic concept that will be used repeatedly is the notion of overwhelming probability.

Definition 3.1 (Overwhelming probability)**.** *Let $\{\mathcal{E}_N\}$ be a sequence of events. We say that $\mathcal{E}_N$ holds with overwhelming probability if for any (large) $D > 0$, there exists $N_0(D)$ such that for all $N \geq N_0(D)$ we have*

$$\mathbb{P}(\mathcal{E}_N) \geq 1 - N^{-D}.$$

3.2.1 Variance formula and resampling

An essential technique for our proof is the formula for the variance of a function of independent random variables. This formula represents the variance via resampling of the random variables. This idea is first due to Chatterjee [Cha05], and in this paper we will use a slight extension of it as in [BLZ20].

Let $X_1, \cdots, X_N$ be independent random variables taking values in some set $\mathcal{X}$, and let $f : \mathcal{X}^N \to \mathbb{R}$ be some measurable function. Let $X = (X_1, \cdots, X_N)$ and X' be an independent copy of X. We denote

$$X^{(i)} = (X_1, \cdots, X_{i-1}, X_i', X_{i+1}, \cdots, X_N), \text{ and } X^{[i]} = (X_1', \cdots, X_i', X_{i+1}, \cdots, X_N).$$

And in general, for $A \subset [N]$, we define X^A to be the random vector obtained from X by replacing the components indexed by A by corresponding components of X'. By a result of Chatterjee [Cha05], we have the following variance formula

$$\mathsf{Var}\left(f(X)\right) = \frac{1}{2}\sum_{i=1}^{N}\mathbb{E}\left[\left(f(X) - f(X^{(i)})\right)\left(f(X^{[i-1]}) - f(X^{[i]})\right)\right].$$

We remark that this variance formula does not depend on the order of the random variables. Therefore, we can consider an arbitrary permutation of $[N]$. Specifically, let $\pi = (\pi(1), \cdots, \pi(N))$ be a random permutation sampled uniformly from the symmetric group $\mathcal{S}_N$ and denote $\pi([i]) := \{\pi(1), \cdots, \pi(i)\}$. Then we have

$$\mathsf{Var}\left(f(X)\right) = \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}\left[\left(f(X) - f(X^{(\pi(i))})\right)\left(f(X^{\pi([i-1])}) - f(X^{\pi([i])})\right)\right].$$

Note that, in the formula above, the expectation is taken with respect to both X, X' and the random permutation π.

Let j have uniform distribution over $[N]$. Let $X^{(j)\circ\pi([i-1])}$ denote the vector obtained from $X^{\pi([i-1])}$ by replacing its j-th component by another independent copy of the random variable X_j in the following way: If j belongs to $\pi([i-1])$, then we replace X'_j by X''_j; if j is not in $\pi([i-1])$, then we replace X_j by X'''_j, where X'' and X''' are independent copies of X such that (X, X', X'', X''') are independent. With this notation, we have the following estimates.

Lemma 3.1 (Lemma 3 in [BLZ20])**.** *Assume that j is chosen uniformly at random from the set $[N]$ and independently of other random variables involved, we have for any $k \in [N]$,*

$$B_k \leq \frac{2\mathsf{Var}\left(f(X)\right)}{k}\left(\frac{N+1}{N}\right)$$

where for any $i \in [N]$,

$$B_i := \mathbb{E}\left[\left(f(X) - f(X^{(j)})\right)\left(f(X^{\pi([i-1])}) - f(X^{(j)\circ\pi([i-1])})\right)\right]$$

and the expectation is taken with respect to components of vectors, random permutations π and the random variable j.

3.2.2 Tools from random matrix theory

In this section we collect some classical results in random matrix theory, which will be indispensable for proving the main theorems. These include concentration of the top eigenvalue, eigenvalue rigidity estimates, and eigenvector delocalization. We focus on the *weakly spiked model*, and the BBP phase transition for the strongly spiked model will be deferred to Section 3.5.

To begin with, we first state some basic settings and notations. It is well known that the empirical distribution of the spectrum of the null model (i.e. $\mathbf{\Sigma} = \mathbf{I}$) converges to the Marchenko-Pastur distribution

$$\rho_{\mathsf{MP}}(x) = \frac{1}{2\pi\xi}\sqrt{\frac{[(x-\lambda_-)(\lambda_+-x)]_+}{x^2}}, \tag{3.7}$$

where the endpoints of the spectrum are given by

$$\lambda_\pm = (1 \pm \sqrt{\xi})^2. \tag{3.8}$$

Beyond the null model, where the population covariance matrix is not identity, the convergence of the empirical spectral measure deviates from the ordinary Marchenko-Pastur law in general and instead converges to a distinct limiting distribution known as the deformed Marchenko-Pastur law. The deformed Marchenko-Pastur distribution, denoted as ρ_{fc}, is characterized by its Stieltjes transform. For a probability measure ρ on the real line, the Stieltjes transofrm is defined as

$$m_\rho(z) := \int_{\mathbb{R}} \frac{1}{x-z}\mathrm{d}\rho(x), \quad z \in \mathbb{C}^+,$$

where $\mathbb{C}^+ := \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$ is the upper half plane of $\mathbb{C}$. The Stieltjes transform is an important object in probability theory with two useful applications: (1) the convergence of the probability measure is equivalent to the convergence of the Stieltjes transform; (2) if the probability measure ρ is absolutely continuous with respect to the Lebesgue measure, it can be recovered from its Stieltjes transform by the inversion formula

$$\rho(x) = \lim_{\eta \downarrow 0} \frac{1}{\pi} \operatorname{Im} m_\rho(x + \mathrm{i}\eta).$$

For the deformed Marchenko-Pastur law, we denote its Stieltjes transform by m_{fc}. The notation m_{fc} stands for free convolution, as the spectral measure of the sample covariance matrix is given by the multiplicative free convolution of the Marchenko-Pastur law and the spectral measure of the population covariance matrix. Suppose the empirical spectral measure $\widehat{\mu}$ of the population covariance matrix $\mathbf{\Sigma}$ converges to a limiting law μ. Then m_{fc} is determined by the following self-consistent equation

$$\frac{1}{m_{\mathsf{fc}}(z)} = -z + \xi \int \frac{t}{1 + m_{\mathsf{fc}}(z)t} \mathrm{d}\mu(t). \tag{3.9}$$

Recall that $\mathbf{\Sigma} = \mathsf{diag}(d_1, \cdots, d_p)$, and indices with $d_i > 1 + \sqrt{\xi}$ are called outliers $i \in \mathcal{O}$. For the weakly spiked model, we have $\mathcal{O} = \emptyset$ and in this case the support of the deformed Marchenko-Pastur law has only one connected component (see e.g. [LS16, DY18]). The right endpoint of the spectrum λ_{R}, which is closely related to the concentration of the top eigenvalue, is determined by

$$\lambda_{\mathrm{R}} = \frac{1}{a} \left(1 + \xi \int \frac{ta}{1 - ta} \mathrm{d}\mu(t) \right), \tag{3.10}$$

where $a \geq 0$ is the unique solution of the equation

$$\int \left(\frac{ta}{1 - ta} \right)^2 \mathrm{d}\mu(t) = \xi^{-1}.$$

The left endpoint of the spectrum, denoted as λ_{L}, can be determined via a similar way.

An important result in random matrix theory is that the eigenvalues are concentrated. To state the result, we define the typical locations of the eigenvalues:

$$\gamma_m := \inf\left\{E > 0 : \int_{-\infty}^{E} \rho_{\mathsf{fc}}(x)\mathrm{d}x \geq \frac{m}{p}\right\}, \quad 1 \leq m \leq p.$$

The eigenvalue rigidity estimates [PY14, BEK+14, LS16, DY18] are stated as follows.

Let $\widehat{m} := \min(m, p+1-k)$, for any small $\varepsilon > 0$ and large $D > 0$ there exists $n_0(\varepsilon, D)$ such that the following holds for any $n \geq n_0$,

$$\mathbb{P}\left(|\lambda_m - \gamma_m| \leq n^{-\frac{2}{3}+\varepsilon}(\widehat{m})^{-\frac{1}{3}} \text{ for all } 1 \leq m \leq p\right) > 1 - n^{-D}. \tag{3.11}$$

We remark that the square case $\xi \equiv 1$ is actually significantly different, due to the singularity of the Marchenko-Pastur law at $x = 0$. Near the left spectral edge, the typical eigenvalue spacing would be of order $O(n^{-2})$, which leads to a stronger concentration. In this case, the tight rigidity was proved in [AEK14b], and see [Wan22] for more explanations. However, the estimate (3.11) is good enough for our purpose.

Another important result is the Tracy-Widom limit for the top eigenvalue (see e.g. [PY14, DY18, Wan24, SX21a]). As a consequence, we have the following concentration of the top eigenvalue and also a spectral gap estimate.

Lemma 3.2. *Consider the weakly spiked model $\mathcal{O} = \emptyset$. For any $\varepsilon > 0$, with overwhelming probability, we have*

$$|\lambda - \lambda_{\mathrm{R}}| \leq n^{-2/3+\varepsilon}, \quad \textit{and} \ \ \mathsf{Var}(\lambda) \leq n^{-4/3+\varepsilon}.$$

In particular, for the null model, we have $\mathsf{Var}(\lambda) \lesssim n^{-4/3}$. Moreover, for any $\delta > 0$, there *exists a constant $c_0 > 0$ such that*

$$\mathbb{P}\left(\lambda_1 - \lambda_2 \geq c_0 n^{-2/3}\right) \geq 1 - \delta.$$

Proof. The first result follows from the well-known Tracy-Widom limit for the top eigenvalue. Specifically,

$$\lim_{n\to\infty} \mathbb{P}\left(\gamma n^{2/3}(\lambda - \lambda_{\mathrm{R}}) \leq s\right) = F_1(s),$$

where γ is a constant depending only on the ratio ξ and the limiting spectral measure μ of the population covariance matrix, and F_1 is the type-1 Tracy-Widom distribution (in particular, [Wan24, SX21a] provided quantitative rate of convergence). The variance estimate is then a natural consequence. For the improved variance bound of the null model, the Gaussian case (i.e. the white Wishart ensemble) was proved in [LR10], and the general case follows from universality, i.e. for any fixed m

$$\lim_{n\to\infty} \mathbb{P}\left(\left(n^{2/3}(\lambda_\ell - \lambda_{\mathrm{R}}) \leq s_\ell\right)_{1\leq \ell\leq m}\right) = \lim_{n\to\infty} \mathbb{P}^G\left(\left(n^{2/3}(\lambda_\ell - \lambda_{\mathrm{R}}) \leq s_\ell\right)_{1\leq \ell\leq m}\right),$$

where $\mathbb{P}^G$ denotes the probability measure associated with the Gaussian matrix. The spectral gap estimate also follows from universality that the spectral statistics of the sample covariance matrix is the same as the Gaussian Orthogonal Ensemble (GOE), i.e. for any fixed m

$$\lim_{n\to\infty} \mathbb{P}\left(\left(\gamma n^{2/3}(\lambda_\ell - \lambda_{\mathrm{R}}) \leq s_\ell\right)_{1\leq \ell\leq m}\right) = \lim_{n\to\infty} \mathbb{P}\left(\left(n^{2/3}(\lambda_\ell^{GOE} - 2) \leq s_\ell\right)_{1\leq \ell\leq m}\right)$$

For GOE, the desired spectral gap estimate was proved in e.g. [AGZ10]. $\square$

Moreover, a gap estimate for the eigenvalues is needed, which was proved in [TV12, Wan12]

Lemma 3.3. *Consider the weakly spiked model $\mathcal{O} = \emptyset$. For any $c > 0$, there exists $\kappa > 0$*

such that for every $1 \le i \le p$, with probability at least $1 - n^{-\kappa}$, we have

$$|\lambda_i - \lambda_{i+1}| \ge n^{-1-c}.$$

The property of eigenvectors is also a key ingredient for our proof. In particular, we extensively rely on the following delocalization property, which implies that the eigenvectors are distributed roughly uniformly on the unit sphere (see e.g. [PY14, BEK+14, DY18]). This is one of the most significant difference between the weakly spiked model and the strongly spiked model. With a strong spike in the population covariance matrix, the top eigenvector will be forced to lie on a cone around the signal and hence is localized in some sense.

Lemma 3.4. *Consider the weakly spiked model $\mathcal{O} = \emptyset$. For any $\varepsilon > 0$, with overwhelming probability, we have*

$$\max_{1 \le i \le p} \|\mathbf{v}_i\|_\infty + \max_{1 \le j \le p} \|\mathbf{u}_j\|_\infty \le n^{-\frac{1}{2}+\varepsilon}.$$

3.3 Sensitivity Regime of Weakly Spiked Model

We begin with a heuristic argument for deriving the threshold for the sensitivity regime. We consider the derivative of the top eigenvalue as a function of the matrix entries. For a symmetric matrix $\mathbf{A}$ with an eigenpair $(\lambda, \mathbf{v})$, the derivative of λ with respect to the matrix entries is given by $\mathrm{d}\lambda = \mathbf{v}^\top(\mathrm{d}\mathbf{A})\mathbf{v}$. Motivated by this, we have the approximation

$$\lambda^{[1]} - \lambda \approx \mathbf{v}^\top \left[(\mathbf{\Sigma}^{1/2}\mathbf{X}^{[1]})(\mathbf{\Sigma}^{1/2}\mathbf{X}^{[1]})^\top - (\mathbf{\Sigma}^{1/2}\mathbf{X})(\mathbf{\Sigma}^{1/2}\mathbf{X})^\top\right] \mathbf{v}.$$

Note that the matrix in the parenthesis has only $\Theta(p)$ non-zero entries, and each entry is roughly of size $O(n^{-1+\varepsilon/2})$ for an arbitrarily small $\varepsilon > 0$ thanks to the sub-exponential decay assumption (3.2). Also, the eigenvector $\mathbf{v}$ is delocalized in the sense that $|\mathbf{v}(m)| \approx$

$p^{-1/2+\varepsilon/4}$ for all $m = 1, \ldots, p$. The CLT scaling yields that approximately we have

$$\lambda^{[1]} - \lambda \approx O\left(\sqrt{p} n^{-1+\varepsilon/2} p^{-1+\varepsilon/4} p^{-1+\varepsilon/4}\right) = O(n^{-3/2+\varepsilon}).$$

By this heuristic argument and central limit theorem, we have

$$\lambda^{[k]} - \lambda \approx O(\sqrt{k} n^{-3/2+\varepsilon}).$$

Note that from random matrix theory, we know that $\lambda_1 - \lambda_2$ is approximately of order $n^{-2/3}$. Therefore, if we have $\sqrt{k} n^{-3/2+\varepsilon} \ll n^{-2/3}$ (this corresponds to $k \ll n^{5/3-\varepsilon}$), then the difference the two top eigenvalues λ and $\lambda^{[k]}$ is much smaller than the first two eigenvalues λ_1 and λ_2 of the matrix $(\mathbf{\Sigma}^{1/2}\mathbf{X})(\mathbf{\Sigma}^{1/2}\mathbf{X})^\top$. This implies that the perturbation effect on $\mathbf{X}^{[k]}$ is small, and therefore in this case it is plausible to believe that $\mathbf{v}^{[k]}$ is just a small perturbation of $\mathbf{v}$. Thus, for the threshold of the sensitivity regime, we must expect $k \gg n^{5/3}$.

To make the above heuristics rigorous, a key observation is that the inner products $\langle \mathbf{v}, \mathbf{v}^{[k]} \rangle$ and $\langle \mathbf{u}, \mathbf{u}^{[k]} \rangle$ can be related to the variance of the leading eigenvalue. Specifically, we will prove

$$\mathbb{E}\left[|\langle \mathbf{v}, \mathbf{v}^{[k]} \rangle|^2\right] \leq C \frac{n^3 \mathsf{Var}(\sigma)}{k} + o(1),$$

where $C > 0$ is some universal constant and $\sigma = \sqrt{\lambda}$ is the leading singular value. A similar result is also true for $\mathbf{u}$ and $\mathbf{u}^{[k]}$. The variance $\mathsf{Var}(\sigma)$ in Lemma 3.2 leads to the desired threshold.

3.3.1 Sensitivity analysis for neighboring data matrices

As a first step, we will first show that resampling of a single entry has little perturbation effect on the top eigenvectors in the weakly spiked model. This will be helpful to control the single entry resampling term in the variance formula (see Lemma 3.1).

For any fixed $1 \leq i \leq p$ and $1 \leq \alpha \leq n$, let $\mathbf{X}_{(i,\alpha)}$ be the matrix obtained from $\mathbf{X}$ by replacing the (i, α) entry $\mathbf{X}_{i\alpha}$ with $\mathbf{X}'_{i\alpha}$. Define the corresponding covariance matrix $\mathbf{H}_{(i,\alpha)} := (\mathbf{\Sigma}^{1/2}\mathbf{X}_{(i,\alpha)})(\mathbf{\Sigma}^{1/2}\mathbf{X}_{(i,\alpha)})^\top$, and use $\mathbf{v}^{(i,\alpha)}$ to denote its unit top eigenvector. Similarly, we denote by $\mathbf{u}^{(i,\alpha)}$ the unit top eigenvector of $\widehat{\mathbf{H}}_{(i,\alpha)} := (\mathbf{\Sigma}^{1/2}\mathbf{X}_{(i,\alpha)})^\top(\mathbf{\Sigma}^{1/2}\mathbf{X}_{(i,\alpha)})$.

Lemma 3.5. *Let $c > 0$ small and $0 < \delta < \frac{1}{2} - c$. For all $1 \leq i \leq n$ and $1 \leq \alpha \leq p$, on the event $\{\lambda_1 - \lambda_2 \geq n^{-1-c}\}$, with overwhelming probability*

$$\max_{i,\alpha} \min_{s \in \{\pm 1\}} \|\mathbf{v} - s\mathbf{v}^{(i,\alpha)}\|_\infty \leq n^{-\frac{1}{2}-\delta} \tag{3.12}$$

and similarly

$$\max_{i,\alpha} \min_{s \in \{\pm 1\}} \|\mathbf{u} - s\mathbf{u}^{(i,\alpha)}\|_\infty \leq n^{-\frac{1}{2}-\delta}$$

Proof. Let $\lambda_1^{(i,\alpha)} \geq \cdots \geq \lambda_p^{(1,\alpha)}$ denote the eigenvalues of the matrix $\mathbf{H}_{(i,\alpha)}$, and let $\mathbf{v}_j^{(i,\alpha)}$ denote the unit eigenvector associated with the eigenvalue $\lambda_j^{(i,\alpha)}$. Similarly, we define the unit eigenvectors $\{\mathbf{u}_\beta^{(i,\alpha)}\}$ for the matrix $\widehat{\mathbf{H}}_{(i,\alpha)}$. Note that

$$\lambda_1 \geq \langle \mathbf{v}_1^{(i,\alpha)}, \mathbf{H}\mathbf{v}_1^{(i,\alpha)} \rangle = \lambda_1^{(i,\alpha)} + \langle \mathbf{v}_1^{(i,\alpha)}, (\mathbf{H} - \mathbf{H}_{(i,\alpha)})\mathbf{v}_1^{(i,\alpha)} \rangle.$$

Recall $\mathbf{\Sigma} = \mathsf{diag}(d_1, \ldots, d_p)$. Since $\mathbf{X}$ and $\mathbf{X}_{(i,\alpha)}$ differ only at the (i, α) entry, we have

$$(\mathbf{H} - \mathbf{H}_{(i,\alpha)})_{j\ell} = \begin{cases} d_i(\mathbf{X}_{i\alpha} - \mathbf{X}'_{i\alpha})\mathbf{X}_{\ell\alpha} & \text{if } j = i, \ell \neq i, \\ d_i(\mathbf{X}_{i\alpha} - \mathbf{X}'_{i\alpha})\mathbf{X}_{j\alpha} & \text{if } j \neq i, \ell = i, \\ d_i(\mathbf{X}_{i\alpha}^2 - (\mathbf{X}'_{i\alpha})^2) & \text{if } j = i, \ell = i, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, setting $\Delta_{i\alpha} := (\mathbf{X}_{i\alpha} - \mathbf{X}'_{i\alpha})$, we have

$$
\begin{aligned}
&\langle \mathbf{v}_1^{(i,\alpha)}, (\mathbf{H} - \mathbf{H}_{(i,\alpha)})\mathbf{v}_1^{(i,\alpha)} \rangle \\
&\quad = 2\mathbf{v}_1^{(i,\alpha)}(i) d_i \Delta_{i\alpha} \left(\sum_{j=1}^{p} (\mathbf{X}_{(i,\alpha)})_{j\alpha} \mathbf{v}_1^{(i,\alpha)}(j) - \mathbf{X}'_{i\alpha} \mathbf{v}_1^{(i,\alpha)}(i) \right) + d_i \left(\mathbf{v}_1^{(i,\alpha)}(i) \right)^2 (\mathbf{X}_{i\alpha}^2 - (\mathbf{X}'_{i\alpha})^2) \\
&\quad = 2\mathbf{v}_1^{(i,\alpha)}(i) d_i \Delta_{i\alpha} \left(\mathbf{X}_{(i,\alpha)}^\top \mathbf{v}_1^{(i,\alpha)} \right)(\alpha) + d_i \left(\mathbf{v}_1^{(i,\alpha)}(i) \right)^2 \left(\mathbf{X}_{i\alpha}^2 - (\mathbf{X}'_{i\alpha})^2 - 2(\mathbf{X}_{i\alpha} - \mathbf{X}'_{i\alpha})\mathbf{X}'_{i\alpha} \right) \\
&\quad = 2d_i \sqrt{\lambda_1^{(i,\alpha)}} \mathbf{v}_1^{(i,\alpha)}(i) \mathbf{u}_1^{(i,\alpha)}(\alpha) \Delta_{i\alpha} + d_i \left(\mathbf{v}_1^{(i,\alpha)}(i)^2 \right) \Delta_{i\alpha}^2.
\end{aligned}
\tag{3.13}
$$

This gives us

$$
\lambda_1 \geq \lambda_1^{(i,\alpha)} - 2d_i \sqrt{\lambda_1^{(i,\alpha)}} \left(|\mathbf{X}_{i\alpha}| + |\mathbf{X}'_{i\alpha}| \right) \|\mathbf{v}_1^{(i,\alpha)}\|_\infty \|\mathbf{u}_1^{(i,\alpha)}\|_\infty - d_i \left(|\mathbf{X}_{i\alpha}| + |\mathbf{X}'_{i\alpha}| \right)^2 \|\mathbf{v}_1^{(i,\alpha)}\|_\infty^2. \tag{3.14}
$$

Similarly,

$$
\lambda_1^{(i,\alpha)} \geq \lambda_1 - 2d_i \sqrt{\lambda_1} \left(|\mathbf{X}_{i\alpha}| + |\mathbf{X}'_{i\alpha}| \right) \|\mathbf{v}_1\|_\infty \|\mathbf{u}_1\|_\infty - d_i \left(|\mathbf{X}_{i\alpha}| + |\mathbf{X}'_{i\alpha}| \right)^2 \|\mathbf{v}_1\|_\infty^2. \tag{3.15}
$$

By Assumption 1, the sub-exponential decay implies $|\mathbf{X}_{i\alpha}|, |\mathbf{X}'_{i\alpha}| \leq n^{-1/2+\varepsilon}$ with overwhelming probability for any $\varepsilon > 0$. Also, by Assumption 3, we know that $d_i = \Theta(1)$. Moreover, by the delocalization of eigenvectors, with overwhelming probability, we have

$$
\max \left(\|\mathbf{v}_1\|_\infty, \|\mathbf{u}_1\|_\infty, \|\mathbf{v}_1^{(i,\alpha)}\|_\infty, \|\mathbf{u}_1^{(i,\alpha)}\|_\infty \right) \leq n^{-\frac{1}{2}+\varepsilon}.
$$

Moreover, by the rigidity estimates (3.11), with overwhelming probability we have

$$
|\lambda_1 - \lambda_{\mathrm{R}}| \leq n^{-\frac{2}{3}+\varepsilon}, \quad |\lambda_1^{(i,\alpha)} - \lambda_{\mathrm{R}}| \leq n^{-\frac{2}{3}+\varepsilon}
$$

Therefore, combining with a union bound, the above two inequalities (3.14) and (3.15) together yield

$$
\max_{1 \leq i \leq n, 1 \leq \alpha \leq p} |\lambda_1 - \lambda_1^{(i,\alpha)}| \leq n^{-3/2+3\varepsilon} \tag{3.16}
$$

with overwhelming probability.

We write $\mathbf{v}_1^{(i,\alpha)}$ in the orthonormal basis of eigenvectors $\{\mathbf{v}_j\}$:

$$\mathbf{v}_1^{(i,\alpha)} = \sum_{j=1}^{p} a_j \mathbf{v}_j.$$

Using this representation and the spectral theorem,

$$\sum_{j=1}^{p} \lambda_j a_j \mathbf{v}_j = \mathbf{H}\mathbf{v}_1^{(i,\alpha)} = \left(\mathbf{H} - \mathbf{H}_{(i,\alpha)}\right) \mathbf{v}_1^{(i,\alpha)} + \left(\lambda_1^{(i,\alpha)} - \lambda_1\right) \mathbf{v}_1^{(i,\alpha)} + \lambda_1 \mathbf{v}_1^{(i,\alpha)}.$$

As a consequence,

$$\lambda_1 \mathbf{v}_1^{(i,\alpha)} = \sum_{j=1}^{p} \lambda_j a_j \mathbf{v}_j + \left(\mathbf{H}_{(i,\alpha)} - \mathbf{H}\right) \mathbf{v}_1^{(i,\alpha)} + \left(\lambda_1 - \lambda_1^{(i,\alpha)}\right) \mathbf{v}_1^{(i,\alpha)}.$$

For $j \neq 1$, taking inner product with $\mathbf{v}_j$ yields

$$\lambda_1 a_j = \langle \mathbf{v}_j, \lambda_1 \mathbf{v}_1^{(i,\alpha)} \rangle = \lambda_j a_j + \langle \mathbf{v}_j, (\mathbf{H}_{(i,\alpha)} - \mathbf{H})\mathbf{v}_1^{(i,\alpha)} \rangle + (\lambda_1 - \lambda_1^{(i,\alpha)}) a_j,$$

which implies

$$\left((\lambda_1 - \lambda_j) + (\lambda_1^{(i,\alpha)} - \lambda_1)\right) a_j = \langle \mathbf{v}_j, (\mathbf{H}_{(i,\alpha)} - \mathbf{H})\mathbf{v}_1^{(i,\alpha)} \rangle. \tag{3.17}$$

By a similar computation as in (3.13), we have

$$\begin{aligned}\left|\langle \mathbf{v}_j, (\mathbf{H}_{(i,\alpha)} - \mathbf{H})\mathbf{v}_1^{(i,\alpha)} \rangle\right| &= \left| d_i \Delta_{i\alpha} \left(\sqrt{\lambda_1^{(i,\alpha)}} \mathbf{v}_j(i) \mathbf{u}_1^{(i,\alpha)}(\alpha) + \sqrt{\lambda_\beta} \mathbf{v}_1^{(i,\alpha)}(i) \mathbf{u}_j(\alpha) \right) \right| \\ &\lesssim (|\mathbf{X}_{i\alpha}| + |\mathbf{X}'_{i\alpha}|) \left(\|\mathbf{v}_j\|_\infty \|\mathbf{u}_1^{(i,\alpha)}\|_\infty + \|\mathbf{v}_1^{(i,\alpha)}\|_\infty \|\mathbf{u}_j\|_\infty \right) \\ &\leq n^{-\frac{3}{2}+3\varepsilon}\end{aligned} \tag{3.18}$$

with overwhelming probability, where the second step follows from rigidity of eigenvalues and the last step follows from the sub-exponential decay assumption and delocalization of eigenvectors.

Consider the event $\mathcal{E} := \{\lambda_1 - \lambda_2 \geq n^{-1-c}\}$. Fix some $\omega > 0$ small. By rigidity of eigenvalues (3.11), on the event $\mathcal{E}$, with overwhelming probability

$$\lambda_1 - \lambda_j \gtrsim \begin{cases} n^{-1-c} & \text{if } 2 \leq j \leq n^{\omega}, \\ j^{2/3} n^{-2/3} & \text{if } n^{\omega} < j \leq p. \end{cases} \tag{3.19}$$

On the event $\mathcal{E}$, using (3.16), (3.18) and (3.19), with overwhelming probability we have

$$|a_j| \lesssim \begin{cases} n^{-\frac{1}{2}+c+3\varepsilon} & \text{if } 2 \leq j \leq n^{\omega}, \\ j^{-\frac{2}{3}} n^{-\frac{5}{6}+3\varepsilon} & \text{if } n^{\omega} < j \leq p. \end{cases} \tag{3.20}$$

Choose $s = a_1/|a_1|$. Note that $1 - |a_1| \leq \sum_{j=2}^{p} |a_j|$. Thanks to the delocalization of eigenvectors, with overwhelming probability, we have

$$\|s\mathbf{v}_1 - \mathbf{v}_1^{(i,\alpha)}\|_\infty = \left\|(s-a_1)\mathbf{v}_1 + \sum_{j=2}^{p} a_j \mathbf{v}_j\right\|_\infty \leq (1-|a_1|)\|\mathbf{v}_1\|_\infty + \sum_{j=2}^{p} |a_j|\|\mathbf{v}_j\|_\infty \leq n^{-\frac{1}{2}+\varepsilon} \sum_{j=2}^{p} |a_j|.$$

Thus, on the event $\mathcal{E}$, it follows from (3.19) that

$$\begin{aligned} \|s\mathbf{v}_1 - \mathbf{v}_1^{(i,\alpha)}\|_\infty &\lesssim n^{-\frac{1}{2}+\varepsilon}\left(n^{-\frac{1}{2}+3\varepsilon+c+\omega} + n^{-\frac{5}{6}+3\varepsilon} \sum_{j=n^{\omega}}^{p} j^{-\frac{2}{3}}\right) \\ &\lesssim n^{-1+4\varepsilon+c+\omega} + n^{-1+4\varepsilon}. \end{aligned}$$

Choosing ε and ω small enough so that $4\varepsilon + c + \omega < \frac{1}{2}$, we conclude that (3.12) is true.

A similar bound for $\mathbf{u}$ can be shown by the same arguments for $\widehat{\mathbf{H}} = (\mathbf{\Sigma}^{1/2}\mathbf{X})^\top(\mathbf{\Sigma}^{1/2}\mathbf{X})$. Hence, we have shown the desired results. $\square$

3.3.2 Proof of sensitivity

Let $\mathbf{X}'' \in \mathbb{R}^{p\times n}$ be a copy of $\mathbf{X}$ that is independent of $\mathbf{X}$ and $\mathbf{X}'$. For an arbitrary index (i,α) with $1 \le i \le p$ and $1 \le \alpha \le n$, we introduce another random matrix $\mathbf{Y}_{(i,\alpha)}$ obtained from $\mathbf{X}$ by replacing the (i,α) entry $\mathbf{X}_{i\alpha}$ by $\mathbf{X}''_{i\alpha}$. Similarly, we denote $\mathbf{Y}^{[k]}_{(i,\alpha)}$ the matrix obtained via the same procedure from $\mathbf{X}^{[k]}$. For the matrix $\mathbf{X}^{[k]}$, we do the similar resampling procedure in the following way: if $(i,\alpha) \in S_k$, then replace $\mathbf{X}^{[k]}_{i\alpha}$ with $\mathbf{X}''_{i\alpha}$; if $(i,\alpha) \notin S_k$, then replace $\mathbf{X}^{[k]}_{i\alpha}$ with $\mathbf{X}'''_{i\alpha}$, where $\mathbf{X}'''$ is another independent copy of $\mathbf{X}$, $\mathbf{X}'$ and $\mathbf{X}''$.

In the following analysis, we choose an index (s,θ) uniformly at random from the set of all pairs $\{(i,\alpha) : 1 \le i \le p, 1 \le \alpha \le n\}$. Let μ be the top singular value of $\mathbf{Y}_{(s,\theta)}$ and use $\mathbf{f} \in \mathbb{R}^p$ and $\mathbf{g} \in \mathbb{R}^n$ to denote the normalized top left and right singular vectors of $\mathbf{Y}_{(s,\theta)}$. Similarly, we define $\mu^{[k]}$, $\mathbf{f}^{[k]}$ and $\mathbf{g}^{[k]}$ for $\mathbf{Y}^{[k]}_{(s,\theta)}$. We also denote by $\mathbf{h}$ and $\mathbf{h}^{[k]}$ the concatenation of $\mathbf{f},\mathbf{g}$ and $\mathbf{f}^{[k]},\mathbf{g}^{[k]}$, respectively. Finally, let $\widetilde{\mathbf{X}}^{[k]}$, $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Y}}^{[k]}$ be the symmetrization (3.6) of $\mathbf{X}^{[k]}$, $\mathbf{Y}$ and $\mathbf{Y}^{[k]}$, respectively. When the context is clear, we will omit the index (s,θ) for the convenience of notations.

Step 1. By Lemma 3.1, we have

$$\frac{2\mathsf{Var}(\sigma)}{k} \cdot \frac{np+1}{np} \ge \mathbb{E}\left[(\sigma-\mu)\left(\sigma^{[k]}-\mu^{[k]}\right)\right]. \tag{3.21}$$

Using the variational characterization of the top singular value

$$\langle \mathbf{f}, \mathbf{\Sigma}^{1/2}\mathbf{X}\mathbf{g}\rangle \le \sigma = \langle \mathbf{v}, \mathbf{\Sigma}^{1/2}\mathbf{X}\mathbf{u}\rangle, \quad \langle \mathbf{v}, \mathbf{\Sigma}^{1/2}\mathbf{Y}\mathbf{u}\rangle \le \mu = \langle \mathbf{f}, \mathbf{\Sigma}^{1/2}\mathbf{Y}\mathbf{g}\rangle.$$

This implies

$$\langle \mathbf{f}, \mathbf{\Sigma}^{1/2}(\mathbf{X}-\mathbf{Y})\mathbf{g}\rangle \le \sigma-\mu \le \langle \mathbf{v}, \mathbf{\Sigma}^{1/2}(\mathbf{X}-\mathbf{Y})\mathbf{u}\rangle. \tag{3.22}$$

Applying the same arguments to $\mathbf{X}^{[k]}$ and $\mathbf{Y}^{[k]}$, we have

$$\left\langle \mathbf{f}^{[k]}, \boldsymbol{\Sigma}^{1/2}\left(\mathbf{X}^{[k]} - \mathbf{Y}^{[k]}\right)\mathbf{g}^{[k]}\right\rangle \le \sigma^{[k]} - \mu^{[k]} \le \left\langle \mathbf{v}^{[k]}, \boldsymbol{\Sigma}^{1/2}\left(\mathbf{X}^{[k]} - \mathbf{Y}^{[k]}\right)\mathbf{u}^{[k]}\right\rangle.$$

Since the matrices $\mathbf{X}$ and $\mathbf{Y}$ differ only at the (s, θ) entry, for any vectors $\mathbf{a} \in \mathbb{R}^p$ and $\mathbf{b} \in \mathbb{R}^n$ we have

$$\langle \mathbf{a}, \boldsymbol{\Sigma}^{1/2}(\mathbf{X}-\mathbf{Y})\mathbf{b}\rangle = \sqrt{d_s}\Delta_{s\theta}\,\mathbf{a}(s)\mathbf{b}(\theta),\ \ \left\langle \mathbf{a}, \boldsymbol{\Sigma}^{1/2}\left(\mathbf{X}^{[k]} - \mathbf{Y}^{[k]}\right)\mathbf{b}\right\rangle = \sqrt{d_s}\Delta_{s\theta}^{[k]}\,\mathbf{a}(s)\mathbf{b}(\theta),$$

where

$$\Delta_{s\theta} := \mathbf{X}_{s\theta} - \mathbf{X}''_{s\theta}, \ \text{ and } \ \Delta_{s\theta}^{[k]} := \begin{cases} \mathbf{X}'_{s\theta} - \mathbf{X}''_{s\theta} & \text{if } (s,\theta) \in S_k, \\ \mathbf{X}_{s\theta} - \mathbf{X}'''_{s\theta} & \text{if } (s,\theta) \notin S_k. \end{cases}$$

Therefore,

$$\sqrt{d_s}\Delta_{s\theta}\,\mathbf{f}(s)\mathbf{g}(\theta) \le \sigma - \mu \le \sqrt{d_s}\Delta_{s\theta}\mathbf{v}(s)\mathbf{u}(\theta),$$

and

$$\sqrt{d_s}\Delta_{s\theta}^{[k]}\,\mathbf{f}^{[k]}(s)\mathbf{g}^{[k]}(\theta) \le \sigma^{[k]} - \mu^{[k]} \le \sqrt{d_s}\Delta_{s\theta}^{[k]}\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta).$$

Consider

$$T_1 := d_s\left(\Delta_{s\theta}\mathbf{v}(s)\mathbf{u}(\theta)\right)\left(\Delta_{s\theta}^{[k]}\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta)\right),\ \ T_2 := d_s\left(\Delta_{s\theta}\,\mathbf{f}(s)\mathbf{g}(\theta)\right)\left(\Delta_{s\theta}^{[k]}\,\mathbf{f}^{[k]}(s)\mathbf{g}^{[k]}(\theta)\right),$$

$$T_3 := d_s\left(\Delta_{s\theta}\mathbf{v}(s)\mathbf{u}(\theta)\right)\left(\Delta_{s\theta}^{[k]}\,\mathbf{f}^{[k]}(s)\mathbf{g}^{[k]}(\theta)\right),\ \ T_4 := d_s\left(\Delta_{s\theta}\,\mathbf{f}(s)\mathbf{g}(\theta)\right)\left(\Delta_{s\theta}^{[k]}\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta)\right).$$

Then we have

$$\min(T_1, T_2, T_3, T_4) \le (\sigma - \mu)\left(\sigma^{[k]} - \mu^{[k]}\right) \le \max(T_1, T_2, T_3, T_4). \tag{3.23}$$

To estimate (3.23), we introduce the following events

$$\mathcal{E}_1 := \left\{ \max\left(\|\mathbf{v}\|_\infty, \|\mathbf{u}\|_\infty, \|\mathbf{f}\|_\infty, \|\mathbf{g}\|_\infty, \|\mathbf{v}^{[k]}\|_\infty, \|\mathbf{u}^{[k]}\|_\infty, \|\mathbf{f}^{[k]}\|_\infty, \|\mathbf{g}^{[k]}\|_\infty \right) \leq n^{-\frac{1}{2}+\varepsilon} \right\}, \tag{3.24}$$

and

$$\mathcal{E}_2 := \left\{ \max\left(\|\mathbf{v}-\mathbf{g}\|_\infty, \|\mathbf{u}-\mathbf{f}\|_\infty, \|\mathbf{v}^{[k]}-\mathbf{g}^{[k]}\|_\infty, \|\mathbf{u}^{[k]}-\mathbf{f}^{[k]}\|_\infty \right) \leq n^{-\frac{1}{2}-\delta} \right\}. \tag{3.25}$$

Define the event $\mathcal{E} := \mathcal{E}_1 \cap \mathcal{E}_2$. On the event $\mathcal{E}$, for all

$$J \in \left\{ \mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta), \mathbf{v}(s)\mathbf{u}(\theta)\mathbf{f}^{[k]}(s)\mathbf{g}^{[k]}(\theta), \mathbf{f}(s)\mathbf{g}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta), \mathbf{f}(s)\mathbf{g}(\theta)\mathbf{f}^{[k]}(s)\mathbf{g}^{[k]}(\theta) \right\}$$

we have

$$\left| J - \mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta) \right| = O\left(n^{-2-\delta+3\varepsilon}\right). \tag{3.26}$$

Let $T := \min(T_1, T_2, T_3, T_4)$. On the event $\mathcal{E}$, using (3.26) we have

$$T \geq d_s \left(\Delta_{s\theta} \Delta_{s\theta}^{[k]} \right) \mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta) - O\left(d_s \left| \Delta_{s\theta} \Delta_{s\theta}^{[k]} \right| n^{-2-\delta+3\varepsilon} \right). \tag{3.27}$$

Step 2. Next we claim that the contribution of T when $\mathcal{E}$ does not hold is negligible. Specifically, we have

$$\mathbb{E}\left[T \mathbb{1}_{\mathcal{E}^c} \right] = o(n^{-3}). \tag{3.28}$$

Recall that $d_s = \Theta(1)$. Without loss of generality, it suffices to show that

$$\mathbb{E}\left[\Delta_{s\theta}\, \Delta_{s\theta}^{[k]} \mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta) \mathbb{1}_{\mathcal{E}^c} \right] = o(n^{-3}). \tag{3.29}$$

To see this, using $\mathbb{1}_{\mathcal{E}^c} = \mathbb{1}_{\mathcal{E}_1 \setminus \mathcal{E}} + \mathbb{1}_{\mathcal{E}_1^c}$, we decompose the expectation into two parts

$$\mathbb{E}\left[\Delta_{s\theta}\, \Delta_{s\theta}^{[k]} \mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta) \mathbb{1}_{\mathcal{E}^c} \right] = I_1 + I_2,$$

where

$$I_1 := \mathbb{E}\left[\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta)\mathbb{1}_{\mathcal{E}_1\backslash\mathcal{E}}\right],\ \ I_2 := \mathbb{E}\left[\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta)\mathbb{1}_{\mathcal{E}_1^c}\right].$$

For the first term I_1, by delocalization and the relation $\mathcal{E}_1\backslash\mathcal{E} = \mathcal{E}_1\backslash\mathcal{E}_2$, we have

$$|I_1| \leq n^{-2+4\varepsilon}\mathbb{E}\left[\left|\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\right|\mathbb{1}_{\mathcal{E}_1\backslash\mathcal{E}_2}\right] \leq n^{-2+4\varepsilon}\mathbb{E}\left[\left|\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\right|\mathbb{1}_{\mathcal{E}_2^c}\right]. \tag{3.30}$$

Note that the random variable $\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}$ and the event $\mathcal{E}_2$ are dependent. To decouple these variables, we introduce a new event. Consider the event $\mathcal{E}_3 := \mathcal{A}\cup\mathcal{B}$, where

$$\mathcal{A} := \left\{\min\left(\sigma_1-\sigma_2, \sigma_1^{[k]}-\sigma_2^{[k]}\right) \geq n^{-1-c}\right\},\ \ \mathcal{B} := \left\{\min\left(\mu_1-\mu_2, \mu_1^{[k]}-\mu_2^{[k]}\right) \geq n^{-1-c}\right\}$$

Then,

$$\begin{aligned}
\mathbb{E}\left[\left|\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\right|\mathbb{1}_{\mathcal{E}_3^c}\right] &\lesssim \mathbb{E}\left[\left(\Delta_{s\theta}^2 + (\Delta_{s\theta}^{[k]})^2\right)\mathbb{1}_{\mathcal{E}_3^c}\right] \\
&\lesssim \mathbb{E}\left[\left(\mathbf{X}_{s\theta}^2 + (\mathbf{X}'_{s\theta})^2 + (\mathbf{X}''_{s\theta})^2 + (\mathbf{X}'''_{s\theta})^2\right)\mathbb{1}_{\mathcal{E}_3^c}\right] \\
&\lesssim \mathbb{E}\left[\left(\mathbf{X}_{s\theta}^2 + (\mathbf{X}'_{s\theta})^2\right)\mathbb{1}_{\mathcal{B}^c}\right] + \mathbb{E}\left[\left((\mathbf{X}''_{s\theta})^2 + (\mathbf{X}'''_{s\theta})^2\right)\mathbb{1}_{\mathcal{A}^c}\right].
\end{aligned}$$

Observe that the random variables $\mathbf{X}_{s\theta}$ and $\mathbf{X}'_{s\theta}$ are independent of the event $\mathcal{B}$, and the random variable $\mathbf{X}''_{s\theta}$ is independent of $\mathcal{A}$. Therefore, by Lemma 3.3,

$$\mathbb{E}\left[\left(\mathbf{X}_{s\theta}^2 + (\mathbf{X}'_{s\theta})^2\right)\mathbb{1}_{\mathcal{B}^c}\right] = O(n^{-1-\kappa}),\quad \mathbb{E}\left[\left((\mathbf{X}''_{s\theta})^2 + (\mathbf{X}'''_{s\theta})^2\right)\mathbb{1}_{\mathcal{A}^c}\right] = O(n^{-1-\kappa}).$$

By Lemma 3.5, we have $\mathbb{P}(\mathcal{E}_3 \backslash \mathcal{E}_2) = O(N^{-D})$ for any fixed large $D > 0$. Using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\mathbb{E}\left[\left|\Delta_{s\theta}\, \Delta_{s\theta}^{[k]}\right| \mathbb{1}_{\mathcal{E}_2^c}\right] &= \mathbb{E}\left[\left|\Delta_{s\theta}\, \Delta_{s\theta}^{[k]}\right| \mathbb{1}_{\mathcal{E}_3^c}\right] + \mathbb{E}\left[\left|\Delta_{s\theta}\, \Delta_{s\theta}^{[k]}\right| \mathbb{1}_{\mathcal{E}_3 \backslash \mathcal{E}_2}\right] \\ &= O(n^{-1-\kappa}) + \sqrt{\mathbb{E}\left[\left|\Delta_{s\theta}\, \Delta_{s\theta}^{[k]}\right|^2\right]} \sqrt{\mathbb{P}(\mathcal{E}_3 \backslash \mathcal{E}_2)} \\ &= O(n^{-1-\kappa}) + O(N^{-D}) \\ &= O(n^{-1-\kappa}).\end{aligned}$$

Choosing $4\varepsilon < \kappa$, then (3.30) yields

$$|I_1| \le O(n^{-2+4\varepsilon-1-\kappa}) = o(n^{-3}).$$

For the term I_2, note that $\mathbf{u}$, $\mathbf{v}$, $\mathbf{u}^{[k]}$ and $\mathbf{v}^{[k]}$ are unit vectors. We have that

$$\max(\|\mathbf{u}\|_\infty, \|\mathbf{v}\|_\infty, \|\mathbf{u}^{[k]}\|_\infty, \|\mathbf{v}^{[k]}\|_\infty) \le 1.$$

Recall that $\mathcal{E}_1$ holds with overwhelming probability. By the Cauchy-Schwarz inequality, for any large $D > 0$, we have

$$|I_2| \le \mathbb{E}\left[\left|\Delta_{s\theta}\, \Delta_{s\theta}^{[k]}\right| \mathbb{1}_{\mathcal{E}_1^c}\right] \le \sqrt{\mathbb{E}\left[\left|\Delta_{s\theta}\, \Delta_{s\theta}^{[k]}\right|^2\right]} \sqrt{\mathbb{P}(\mathcal{E}_1^c)} = O(N^{-D}).$$

Hence we have shown the desired claim (3.29).

Step 3. Combining (3.23), (3.27) and (3.28), we obtain

$$\mathbb{E}\left[(\sigma - \mu)\left(\sigma^{[k]} - \mu^{[k]}\right)\right] \ge \mathbb{E}\left[d_s \Delta_{s\theta} \Delta_{s\theta}^{[k]}\, \mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta)\right] + o(n^{-3}).$$

Since $\frac{np+1}{np} \leq 2$ and $d_s = \Theta(1)$, by (3.21) we have

$$\mathbb{E}\left[\Delta_{s\theta}\Delta_{s\theta}^{[k]}\, \mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta)\right] \lesssim \frac{\mathsf{Var}(\sigma)}{k} + o(n^{-3}). \tag{3.31}$$

Since the random index (s, θ) is uniformly sampled, we have

$$\mathbb{E}\left[\Delta_{s\theta}\Delta_{s\theta}^{[k]}\, \mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta)\right] = \frac{1}{np}\mathbb{E}\left[\sum_{1\leq i\leq n, 1\leq \alpha\leq p} \Delta_{i\alpha}\Delta_{i\alpha}^{[k]}\, \mathbf{v}(i)\mathbf{u}(\alpha)\mathbf{v}^{[k]}(i)\mathbf{u}^{[k]}(\alpha)\right]. \tag{3.32}$$

Note that

$$\Delta_{i\alpha}\Delta_{i\alpha}^{[k]} = \begin{cases} (\mathbf{X}_{i\alpha} - \mathbf{X}'_{i\alpha})(\mathbf{X}'_{i\alpha} - \mathbf{X}''_{i\alpha}) & \text{if } (i,\alpha) \in S_k, \\ (\mathbf{X}_{i\alpha} - \mathbf{X}'_{i\alpha})(\mathbf{X}_{i\alpha} - \mathbf{X}'''_{i\alpha}) & \text{if } (i,\alpha) \notin S_k. \end{cases}$$

In either case, we have $\mathbb{E}[\Delta_{i\alpha}\Delta_{i\alpha}^{[k]}] = p^{-1}$. Therefore,

$$\sum_{1\leq i\leq n, 1\leq \alpha\leq p} \mathbb{E}\left[\Delta_{i\alpha}\Delta_{i\alpha}^{[k]} \mid S_k\right] \mathbf{v}(i)\mathbf{u}(\alpha)\mathbf{v}^{[k]}(i)\mathbf{u}^{[k]}(\alpha) = \frac{1}{p}\langle \mathbf{v}, \mathbf{v}^{[k]}\rangle\langle \mathbf{u}, \mathbf{u}^{[k]}\rangle.$$

Consequently, this implies

$$\mathbb{E}\left[\sum_{1\leq i\leq n, 1\leq \alpha\leq p} \mathbb{E}\left[\Delta_{i\alpha}\Delta_{i\alpha}^{[k]} \mid S_k\right] \mathbf{v}(i)\mathbf{u}(\alpha)\mathbf{v}^{[k]}(i)\mathbf{u}^{[k]}(\alpha)\right] = \frac{1}{p}\mathbb{E}\left[\langle \mathbf{v}, \mathbf{v}^{[k]}\rangle\langle \mathbf{u}, \mathbf{u}^{[k]}\rangle\right]. \tag{3.33}$$

Moreover, we claim that

$$\mathbb{E}\left[\sum_{1\leq i\leq n, 1\leq \alpha\leq p} \left(\Delta_{i\alpha}\Delta_{i\alpha}^{[k]} - \mathbb{E}\left[\Delta_{i\alpha}\Delta_{i\alpha}^{[k]} \mid S_k\right]\right) \mathbf{v}(i)\mathbf{u}(\alpha)\mathbf{v}^{[k]}(i)\mathbf{u}^{[k]}(\alpha)\right] = o(n^{-1}). \tag{3.34}$$

For the ease of notations, we set $\Xi_{i\alpha} := \Delta_{i\alpha}\Delta_{i\alpha}^{[k]} - \mathbb{E}[\Delta_{i\alpha}\Delta_{i\alpha}^{[k]} | S_k]$. It suffices to show that for all pairs (i, α) we have

$$\mathbb{E}\left[\Xi_{i\alpha}\mathbf{v}(i)\mathbf{u}(\alpha)\mathbf{v}^{[k]}(i)\mathbf{u}^{[k]}(\alpha)\right] = o(n^{-3}). \tag{3.35}$$

To see this, we introduce another copy of $\mathbf{X}$, denoted by $\mathbf{X}''''$, which is independent of all previous random variables $(\mathbf{X}, \mathbf{X}', \mathbf{X}'', \mathbf{X}''')$. For an arbitrarily fixed index (i, α), we define matrices $\widehat{\mathbf{X}}_{(i,\alpha)}$ and $\widehat{\mathbf{X}}^{[k]}_{(i,\alpha)}$ by resampling the (i, α) entry of $\mathbf{X}$ and $\mathbf{X}^{[k]}$ with $\mathbf{X}''''_{i\alpha}$. Let $\widehat{\mathbf{u}}$, $\widehat{\mathbf{v}}$ be the left and right top singular vector of $\widehat{\mathbf{X}}$, and similarly $\widehat{\mathbf{u}}^{[k]}$, $\widehat{\mathbf{v}}^{[k]}$ for $\widehat{\mathbf{X}}^{[k]}$. Define

$$\psi_{i\alpha} := \mathbf{v}(i)\mathbf{u}(\alpha)\mathbf{v}^{[k]}(i)\mathbf{u}^{[k]}(\alpha), \;\; \widehat{\psi}_{i\alpha} := \widehat{\mathbf{v}}(i)\widehat{\mathbf{u}}(\alpha)\widehat{\mathbf{v}}^{[k]}(i)\widehat{\mathbf{u}}^{[k]}(\alpha).$$

A crucial observation is that $\Xi_{i\alpha}$ and $\widehat{\psi}_{i\alpha}$ are independent. This is because, conditioned on S_k, the matrices $\widehat{\mathbf{X}}$ and $\widehat{\mathbf{X}}^{[k]}$ are independent of $(\mathbf{X}_{i\alpha}, \mathbf{X}'_{i\alpha}, \mathbf{X}''_{i\alpha}, \mathbf{X}'''_{i\alpha})$. Such a conditional independence is also true for the singular vectors, and hence also holds for $\widehat{\psi}_{i\alpha}$. On the other hand, by definition, the variable $\Xi_{i\alpha}$ only depends on $(\mathbf{X}_{i\alpha}, \mathbf{X}'_{i\alpha}, \mathbf{X}''_{i\alpha}, \mathbf{X}'''_{i\alpha})$. Therefore,

$$\mathbb{E}\left[\Xi_{i\alpha}\widehat{\psi}_{i\alpha}\right] = \mathbb{E}\left[\mathbb{E}[\Xi_{i\alpha}|S_k]\,\mathbb{E}[\widehat{\psi}_{i\alpha}|S_k]\right] = 0$$

Thus, we reduce (3.35) to showing

$$\mathbb{E}\left[\Xi_{i\alpha}\left(\psi_{i\alpha} - \widehat{\psi}_{i\alpha}\right)\right] = o(n^{-3}). \tag{3.36}$$

The proof of (3.36) is similar as previous arguments. Consider the events

$$\widehat{\mathcal{E}}_1 := \left\{\max\left(\|\mathbf{v}\|_\infty, \|\mathbf{u}\|_\infty, \|\widehat{\mathbf{u}}\|_\infty, \|\widehat{\mathbf{v}}\|_\infty, \|\mathbf{v}^{[k]}\|_\infty, \|\mathbf{u}^{[k]}\|_\infty, \|\widehat{\mathbf{u}}^{[k]}\|_\infty, \|\widehat{\mathbf{v}}^{[k]}\|_\infty\right) \le n^{-\frac{1}{2}+\varepsilon}\right\},$$

$$\widehat{\mathcal{E}}_2 := \left\{\max\left(\|\mathbf{v} - \widehat{\mathbf{v}}\|_\infty, \|\mathbf{u} - \widehat{\mathbf{u}}\|_\infty, \|\mathbf{v}^{[k]} - \widehat{\mathbf{v}}^{[k]}\|_\infty, \|\mathbf{u}^{[k]} - \widehat{\mathbf{u}}^{[k]}\|_\infty\right) \le n^{-\frac{1}{2}-\delta}\right\}.$$

On the event $\widehat{\mathcal{E}} := \widehat{\mathcal{E}}_1 \cap \widehat{\mathcal{E}}_2$, we have $|\psi_{i\alpha} - \widehat{\psi}_{i\alpha}| = O(n^{-2-\delta+3\varepsilon})$. Note that $\mathbb{E}[|\Xi_{i\alpha}|] = O(n^{-1})$ since $\mathbb{E}[|\Delta_{i\alpha}\Delta^{[k]}_{i\alpha}|] = O(n^{-1})$. As a consequence,

$$\mathbb{E}\left[\left|\Xi_{i\alpha}(\psi_{i\alpha} - \widehat{\psi}_{i\alpha})\right|\mathbb{1}_{\widehat{\mathcal{E}}}\right] = O(n^{-3-\delta+3\varepsilon}) = o(n^{-3}). \tag{3.37}$$

Using the same argument as in (3.29), we have

$$\mathbb{E}\left[\left|\Xi_{i\alpha}(\psi_{i\alpha}-\widehat{\psi}_{i\alpha})\right|\mathbb{1}_{\widehat{\mathcal{E}}^c}\right] \lesssim N^{-2+4\varepsilon}\mathbb{E}\left[|\Xi_{i\alpha}|\mathbb{1}_{\widehat{\mathcal{E}}^c}\right] = O(N^{-2+4\varepsilon}N^{-1-\kappa}) = o(n^{-3}), \quad (3.38)$$

where κ is the constant in the gap property (Lemma 3.3) Thus, by (3.37) and (3.38), we have shown the desired claim (3.36).

Based on (3.31) and (3.32), combining (3.33) and (3.34) yields

$$\frac{1}{np^2}\mathbb{E}\left[\langle \mathbf{v},\mathbf{v}^{[k]}\rangle\langle \mathbf{u},\mathbf{u}^{[k]}\rangle\right] \lesssim \frac{\mathsf{Var}(\sigma)}{k} + o\left(\frac{1}{n^3}\right) + o\left(\frac{1}{n^2 p}\right)$$

By Lemma 3.2 we have $\mathsf{Var}(\sigma) = O(n^{-4/3+\varepsilon_0/2})$ and the assumption $k \geq n^{5/3+\varepsilon_0}$, we have

$$\mathbb{E}\left[\langle \mathbf{v},\mathbf{v}^{[k]}\rangle\langle \mathbf{u},\mathbf{u}^{[k]}\rangle\right] \leq \frac{np^2}{k}O(n^{-4/3+\varepsilon_0/2}) + o(1) = o(1).$$

This implies

$$\mathbb{E}\left[|\langle \mathbf{v},\mathbf{v}^{[k]}\rangle\langle \mathbf{u},\mathbf{u}^{[k]}\rangle|\right] \to 0. \quad (3.39)$$

In the null model, we have $\mathsf{Var}(\sigma) = O(n^{-4/3})$, and therefore the threshold can be improved to $k \gg n^{5/3}$.

Step 4. Consider the symmetrization matrix $\widetilde{\mathbf{X}}$ defined in (3.6). The variational representation of the top eigenvalue yields

$$\sigma = \frac{\langle \mathbf{w},\widetilde{\mathbf{X}}\mathbf{w}\rangle}{\|\mathbf{w}\|_2^2}, \quad \sigma^{[k]} = \frac{\langle \mathbf{w}^{[k]},\widetilde{\mathbf{X}}^{[k]}\mathbf{w}^{[k]}\rangle}{\|\mathbf{w}^{[k]}\|_2^2} \text{ with } \|\mathbf{w}\|_2^2 = \|\mathbf{w}^{[k]}\|_2^2 = 2.$$

Using the same arguments as in Step 1-3, we can conclude that

$$\mathbb{E}\left[|\langle \mathbf{w},\mathbf{w}^{[k]}\rangle|^2\right] = \mathbb{E}\left[|\langle \mathbf{v},\mathbf{v}^{[k]}\rangle + \langle \mathbf{u},\mathbf{u}^{[k]}\rangle|^2\right] \to 0.$$

Combined with (3.39), this gives us

$$\mathbb{E}\left[|\langle \mathbf{v}, \mathbf{v}^{[k]}\rangle|^2 + |\langle \mathbf{u}, \mathbf{u}^{[k]}\rangle|^2\right] \to 0,$$

which proves the desired results.

3.4 Stability Regime of Weakly Spiked Model

To establish the stability of PCA when the resampling strength is mild, we will utilize tools from random matrix theory and specifically the proof relies on the analysis of the resolvent matrix. Furthermore, to reduce the nonlinearity caused by Gram structure of the sample covariance matrix, when considering the resolvent we use a linearization trick. For any $z \in \mathbb{C}$ with $\operatorname{Im} z > 0$, the resolvent is defined as

$$\mathbf{R}(z) := \begin{pmatrix} -\mathbf{I}_n & (\boldsymbol{\Sigma}^{1/2}\mathbf{X})^\top \\ (\boldsymbol{\Sigma}^{1/2}\mathbf{X}) & -z\mathbf{I}_p \end{pmatrix}^{-1}.$$

Similarly, we denote the resolvent of $\mathbf{X}^{[k]}$ by $\mathbf{R}^{[k]}(z)$. The key idea for the stability regime is that eigenvectors can be approximated by resolvents and the resolvents are stable under moderate resampling:

Resolvent Approximation The entries of the resolvent can be used to approximate the product of entries in the eigenvector. For some small $\delta > 0$, let $z_0 = \lambda + \mathrm{i}\eta$ with $\eta = n^{-2/3-\delta}$. In the regime $k \leq n^{5/3-\epsilon_0}$ for some $\epsilon_0 > 0$, there exists some small $c > 0$ such that for all $i, j = 1, \ldots, p$, we have

$$|\eta \operatorname{Im} \mathbf{R}_{n+i,n+j}(z_0) - \mathbf{v}(i)\mathbf{v}(j)| \leq n^{-1-c},$$

and

$$\left|\eta \mathrm{Im}\, \mathbf{R}^{[k]}_{n+i,n+j}(z_0) - \mathbf{v}^{[k]}(i)\mathbf{v}^{[k]}(j)\right| \leq n^{-1-c}.$$

A similar result also holds for $\mathbf{u}$ and $\mathbf{u}^{[k]}$.

Stability of Resolvent Since the eigenvector can be approximated by the resolvent, it suffices to show the stability of the resolvent. Consider the regime $k \leq n^{5/3-\epsilon_0}$ for some $\epsilon_0 > 0$. For some small $\delta > 0$ and all $z = E + \mathrm{i}\eta$ that is close to the upper spectral edge and $\eta = n^{-2/3-\delta}$, there exists a small constant $c > 0$ such that the following is true for all $i, j = 1, \ldots, p$ and $\alpha, \beta = 1, \ldots, n$,

$$\left|\mathbf{R}^{[k]}_{\alpha\beta}(z) - \mathbf{R}_{\alpha\beta}(z)\right| \leq \frac{1}{n^{1+c}\eta},$$

and

$$\left|\mathbf{R}^{[k]}_{n+i,n+j}(z) - \mathbf{R}_{n+i,n+j}(z)\right| \leq \frac{1}{n^{1+c}\eta}.$$

This is the main technical part of the whole argument, and its proof relies on the Lindeberg exchange method and a martingale concentration argument.

In this section, we will focus on the behavior of $\mathbf{v}$ and $\mathbf{v}^{[k]}$ in detail. Similar results also hold for $\mathbf{u}$ and $\mathbf{u}^{[k]}$ via the same arguments.

3.4.1 Linearization and local law

In the study of sample covariance matrices, a convenient trick is to consider the symmetrization $\widetilde{\mathbf{X}}$ of the data matrix $\mathbf{\Sigma}^{1/2}\mathbf{X}$ (defined as in (3.6)) when exploring its spectral properties. For $z \in \mathbb{C}$ with $\mathrm{Im}\, z > 0$, We introduce the resolvent of this symmetrization

$$\mathbf{R}(z) := \begin{pmatrix} -\mathbf{I}_n & (\mathbf{\Sigma}^{1/2}\mathbf{X})^\top \\ (\mathbf{\Sigma}^{1/2}\mathbf{X}) & -z\mathbf{I}_p \end{pmatrix}^{-1}. \tag{3.40}$$

Note that $\mathbf{R}(z)$ is not the conventional definition of the resolvent matrix, but we still call it resolvent for convenience. For the ease of notations, we will relabel the indices in $\mathbf{R}$ in the following way:

Definition 3.2 (Index sets)**.** *We define the index sets*

$$\mathcal{I}_1 := \{1, \ldots, n\}, \;\; \mathcal{I}_2 := \{1, \ldots, p\}, \;\; \mathcal{I} := \mathcal{I}_1 \cup \{n + i : i \in \mathcal{I}_2\}.$$

For a general matrix $\mathbf{M} \in \mathbb{R}^{|\mathcal{I}| \times |\mathcal{I}|}$*, we label the indices of the matrix elements in the following way: for* $a, b \in \mathcal{I}$*, if* $1 \le a, b \le n$*, then typically we use Greek letters* α, β *to represent them; if* $n + 1 \le a, b \le n + p$*, we use the corresponding Latin letters* $i = a - n$ *and* $j = b - n$ *to represent them.*

The resolvent $\mathbf{R}$ is closely related to the eigenvalue and eigenvectors of the sample covariance matrix. As discussed in [DY18, Equation (3.9),(3.10)], we have

$$\mathbf{R}_{\alpha\beta}(z) = \sum_{\ell=1}^{n} \frac{z\mathbf{u}_\ell(\alpha)\mathbf{u}_\ell(\beta)}{\lambda_\ell - z}, \;\; \mathbf{R}_{ij}(z) = \sum_{\ell=1}^{p} \frac{\mathbf{v}_\ell(i)\mathbf{v}_\ell(j)}{\lambda_\ell - z}, \tag{3.41}$$

and

$$\mathbf{R}_{i\alpha}(z) = \sum_{\ell=1}^{p} \frac{\sqrt{\lambda_\ell}\mathbf{u}_\ell(\alpha)\mathbf{v}_\ell(i)}{\lambda_\ell - z}, \;\; \mathbf{R}_{\alpha i}(z) = \sum_{\ell=1}^{p} \frac{\sqrt{\lambda_\ell}\mathbf{v}_\ell(i)\mathbf{u}_\ell(\alpha)}{\lambda_\ell - z}.$$

An important result is the local deformed Marchenko-Pastur law for the resolvent matrix $\mathbf{R}$. This was first proved in [BKYY16], and we refer to [DY18, Lemma 3.11] for a version that is consistent with our setting. Specifically, the resolvent matrix $\mathbf{R}$ has a deterministic limit, defined by

$$\mathbf{G}(z) := \begin{pmatrix} -(1 + m_{\mathsf{fc}}(z)\mathbf{\Sigma})^{-1} & 0 \\ 0 & m_{\mathsf{fc}}(z)\mathbf{I}_p \end{pmatrix}, \tag{3.42}$$

where $m_{\mathsf{fc}}(z)$ is the Stieltjes transform of the deformed Marchenko-Pasture law given by (3.9)

To state the local law, we will focus on the spectral domain

$$\mathcal{S} := \{E + \mathrm{i}\eta : \lambda_{\mathrm{R}} - 1 \le E \le \lambda_{\mathrm{R}} + 1, 0 < \eta < 1\}\,. \tag{3.43}$$

Lemma 3.6 (Local deformed Marchenko-Pastur law)**.** *For any $\varepsilon > 0$, the following estimate holds With overwhelming probability uniformly for $z \in \mathcal{S}$,*

$$\max_{a,b\in\mathcal{I}} |\mathbf{R}_{ab}(z) - \mathbf{G}_{ab}(z)| \le n^{\varepsilon}\left(\sqrt{\frac{\operatorname{Im} m_{\mathsf{fc}}(z)}{n\eta}} + \frac{1}{n\eta}\right). \tag{3.44}$$

To give a precise characterization of the resolvent, we rely on the following estimates for the Stieltjes transform $m_{\mathsf{MP}}(z)$ of the Marchenko-Pasture law. We refer to e.g. [BKYY16, Lemma 3.6] and [DY18, Lemma 3.6] for more details.

Lemma 3.7 (Estimate for $m_{\mathsf{fc}}(z)$)**.** *For $z = E + \mathrm{i}\eta$, let $\kappa(z) := \min(|E - \lambda_{\mathrm{L}}|, |E - \lambda_{\mathrm{R}}|)$ denote the distance to the spectral edge. If $z \in \mathcal{S}$, we have*

$$|m_{\mathsf{fc}}(z)| \asymp 1, \quad \textit{and} \quad \operatorname{Im} m_{\mathsf{fc}}(z) \asymp \begin{cases} \sqrt{\kappa(z) + \eta} & \textit{if } E \in [\lambda_{\mathrm{L}}, \lambda_{\mathrm{R}}], \\ \frac{\eta}{\sqrt{\kappa(z)+\eta}} & \textit{if } E \notin [\lambda_{\mathrm{L}}, \lambda_{\mathrm{R}}]. \end{cases} \tag{3.45}$$

In the following analysis, we will work with $z = E + \mathrm{i}\eta$ satisfying $|E - \lambda_{\mathrm{R}}| \le n^{-2/3+\delta}$ and $\eta = n^{-2/3-\delta}$, where $0 < \delta < \frac{1}{3}$ is some parameter. Uniformly in this regime, the local law yields that the following is true with overwhelming probability for all $\varepsilon > 0$ and some universal constant $C_0 > 0$,

$$\sup_{z} \max_{a\neq b\in\mathcal{I}} |\mathbf{R}_{ab}(z)| \le n^{-\frac{1}{3}+\delta+\varepsilon}, \text{ and } \sup_{z} \max_{a\in\mathcal{I}} |\mathbf{R}_{aa}(z)| \le C_0. \tag{3.46}$$

These estimates will be used repeatedly in the following subsections.

3.4.2 Stability of resolvent

In this subsection, we will prove the main technical result for the proof of resampling stability in the weakly spiked model. Specifically, we will show that under moderate resampling, the resolvent matrices are stable. Since resolvent is closely related to various spectral statistics, this stability lemma for resolvent will be a key ingredient for our proof.

Lemma 3.8. *Consider the weakly spiked model $\mathcal{O} = \emptyset$. Assume $k \le n^{5/3-\epsilon_0}$ for some $\epsilon_0 > 0$. There exists $\delta_0 > 0$ such that for all $0 < \delta < \delta_0$, uniformly for $z = E + \mathrm{i}\eta$ with $|E - \lambda_{\mathrm{R}}| \le n^{-2/3+\delta}$ and $\eta = n^{-2/3-\delta}$, there exists $c > 0$ such that the following is true with overwhelming probability*

$$\max_{i,j} \left| \mathbf{R}^{[k]}_{ij}(z) - \mathbf{R}_{ij}(z) \right| \le \frac{1}{n^{1+c}\eta}, \quad \max_{\alpha,\beta} \left| \mathbf{R}^{[k]}_{\alpha\beta}(z) - \mathbf{R}_{\alpha\beta}(z) \right| \le \frac{1}{n^{1+c}\eta}. \tag{3.47}$$

Proof. Recall that $S_k := \{(i_1, \alpha_1), \dots, (i_k, \alpha_k)\}$ is the random subset of matrix indices whose elements are resampled in the matrix $\mathbf{X}$. For $1 \le t \le k$, let $\mathbf{X}^{[t]}$ be the matrix obtained from $\mathbf{X}$ by resampling the $\{(i_s, \alpha_s)\}_{1 \le s \le t}$ entries and let $\mathcal{F}_t$ be the σ-algebra generated by the random variables $\mathbf{X}$, S_k and $\{\mathbf{X}'_{i_s\alpha_s}\}_{1\le s\le t}$. For $a, b \in \mathcal{I}$, we define

$$T_{ab} := \{t : \{i_t, \alpha_t\} \cap \{a, b\} \neq \emptyset\}.$$

Let $\varepsilon > 0$ be an arbitrarily fixed parameter, and let C_0 be the constant in (3.46). Consider the event $\mathcal{E}_t \in \mathcal{F}_t$ where for all $z = E + \mathrm{i}\eta$ with $|z - \lambda_{\mathrm{R}}| \le n^{-2/3+\delta}$ and $\eta = n^{-2/3-\delta}$ we have

$$\max_{a \neq b} \left| \mathbf{R}^{[t]}_{ab}(z) \right| \le n^{-\frac{1}{3}+\delta+\varepsilon} =: \Psi, \quad \text{and} \quad \max_{a} \left| \mathbf{R}^{[t]}_{aa}(z) \right| \le C_0.$$

Define $\mathbf{X}^{[t]}_0$ as the matrix obtained from $\mathbf{X}^{[t]}$ by replacing the (i_t, α_t) entry with 0, and also define its symmetrization $\widetilde{\mathbf{X}}^{[t]}_0 \in \mathbb{R}^{|\mathcal{I}|\times|\mathcal{I}|}$ as in (3.6). Note that $\widetilde{\mathbf{X}}^{[t+1]}_0$ is $\mathcal{F}_t$-measurable.

We write

$$\widetilde{\mathbf{X}}^{[t]} = \widetilde{\mathbf{X}}_0^{[t+1]} + \widetilde{\mathbf{P}}^{[t+1]}, \quad \widetilde{\mathbf{X}}^{[t+1]} = \widetilde{\mathbf{X}}_0^{[t+1]} + \widetilde{\mathbf{Q}}^{[t+1]},$$

where $\widetilde{\mathbf{P}}^{[t]}, \widetilde{\mathbf{Q}}^{[t]}$ are $|\mathcal{I}| \times |\mathcal{I}|$ symmetric matrices whose elements are all 0 except at the (i_t, α_t) and (α_t, i_t) entries, satisfying

$$(\widetilde{\mathbf{P}}^{[t]})_{ab} = \begin{cases} \sqrt{d_{i_t}}\mathbf{X}_{i_t\alpha_t} & \text{if } \{a,b\} = \{i_t, \alpha_t\}, \\ 0 & \text{otherwise} \end{cases} \quad (\widetilde{\mathbf{Q}}^{[t]})_{ab} = \begin{cases} \sqrt{d_{i_t}}\mathbf{X}'_{i_t\alpha_t} & \text{if } \{a,b\} = \{i_t, \alpha_t\}, \\ 0 & \text{otherwise} \end{cases}.$$

Define the resolvents for the matrices $\widetilde{\mathbf{X}}^{[t]}$ and $\widetilde{\mathbf{X}}_0^{[t]}$ as in (3.40):

$$\mathbf{R}^{[t]} := \begin{pmatrix} -\mathbf{I}_n & (\mathbf{\Sigma}^{1/2}\mathbf{X}^{[t]})^\top \\ (\mathbf{\Sigma}^{1/2}\mathbf{X}^{[t]}) & -z\mathbf{I}_p \end{pmatrix}^{-1}, \ \mathbf{R}_0^{[t]} := \begin{pmatrix} -\mathbf{I}_n & (\mathbf{\Sigma}^{1/2}\mathbf{X}_0^{[t]})^\top \\ (\mathbf{\Sigma}^{1/2}\mathbf{X}_0^{[t]}) & -z\mathbf{I}_p \end{pmatrix}^{-1}.$$

Using first-order resolvent expansion, we obtain

$$\mathbf{R}_0^{[t+1]} = \mathbf{R}^{[t]} + \mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}^{[t]} + \left(\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\right)^2 \mathbf{R}_0^{[t+1]}. \tag{3.48}$$

The triangle inequality yields

$$\left|\left(\mathbf{R}_0^{[t+1]} - \mathbf{R}^{[t]}\right)_{ij}\right| \leq \left|\left(\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}^{[t]}\right)_{ij}\right| + \left|\left(\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}_0^{[t+1]}\right)_{ij}\right|.$$

Note that $\widetilde{\mathbf{P}}^{[t+1]}$ has only two non-zero entries,

$$\left(\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}^{[t]}\right)_{ij} = \sum_{\ell_1,\ell_2} \mathbf{R}_{i\ell_1}^{[t]}\widetilde{\mathbf{P}}_{\ell_1\ell_2}^{[t+1]}\mathbf{R}_{\ell_2 j}^{[t]} = \sqrt{d_{i_{t+1}}}X_{i_{t+1}\alpha_{t+1}}\left(\mathbf{R}_{ii_{t+1}}^{[t]}\mathbf{R}_{\alpha_{t+1}j}^{[t]} + \mathbf{R}_{i\alpha_{t+1}}^{[t]}\mathbf{R}_{i_{t+1}j}^{[t]}\right)$$

Recall that and $|X_{i_{t+1}\alpha_{t+1}}| \leq n^{-1/2+\varepsilon}$ with overwhelming probability thanks to the sub-exponential decay (see Assumption 1). Then on the event $\mathcal{E}_t$, we have

$$\left|\left(\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}^{[t]}\right)_{ij}\right| \leq 2\sqrt{d_{i_{t+1}}}C_0\Psi n^{-\frac{1}{2}+\varepsilon}.$$

Similarly,

$$\begin{aligned}&\left(\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}_0^{[t+1]}\right)_{ij}\\ &\qquad\qquad = \sum_{\{m_1,m_2\},\{m_3,m_4\}=\{i_{t+1},\alpha_{t+1}\}} \mathbf{R}^{[t]}_{im_1}\widetilde{\mathbf{P}}^{[t+1]}_{m_1m_2}\mathbf{R}^{[t]}_{m_2m_3}\widetilde{\mathbf{P}}^{[t+1]}_{m_3m_4}(\mathbf{R}_0^{[t+1]})_{m_4j}.\end{aligned}$$

We use the trivial bound $|\mathbf{R}_0^{[t+1]}| \leq \eta^{-1}$ for the last term. Then, on the event $\mathcal{E}_t$, we have

$$\left|\left(\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}_0^{[t+1]}\right)_{ij}\right| \leq 2d_{i_{t+1}}n^{-1+2\varepsilon}\eta^{-1}(\Psi^2 + C_0^2) \ll \Psi.$$

Therefore, we have shown that, on the event $\mathcal{E}_t$,

$$\max_{i\neq j}\left|(\mathbf{R}_0^{[t+1]})_{ij}\right| \leq 2\Psi, \;\; \max_{i}\left|(\mathbf{R}_0^{[t+1]})_{ii}\right| \leq 2C_0. \tag{3.49}$$

Similarly, using the first-order resolvent expansion for $\mathbf{R}^{[t+1]}$ around $\mathbf{R}^{[t]}$, we have

$$\mathbf{R}^{[t+1]} = \mathbf{R}^{[t]} + \mathbf{R}^{[t]}(\widetilde{\mathbf{P}}^{[t+1]} - \widetilde{\mathbf{Q}}^{[t+1]})\mathbf{R}^{[t]} + \left(\mathbf{R}^{[t]}(\widetilde{\mathbf{P}}^{[t+1]} - \widetilde{\mathbf{Q}}^{[t+1]})\right)^2\mathbf{R}^{[t+1]}.$$

By the same arguments as above, on the event $\mathcal{E}_t$, we can derive

$$\max_{i\neq j}\left|\mathbf{R}^{[t+1]}_{ij}\right| \leq 2\Psi, \;\; \max_{i}\left|\mathbf{R}^{[t+1]}_{ii}\right| \leq 2C_0.$$

Next, we use the resolvent identity (or zeroth-order expansion) for $\mathbf{R}^{[t+1]}$ and $\mathbf{R}_0^{[t+1]}$:

$$\mathbf{R}^{[t+1]} = \mathbf{R}_0^{[t+1]} - \mathbf{R}_0^{[t+1]}\widetilde{\mathbf{Q}}^{[t+1]}\mathbf{R}^{[t+1]}.$$

This leads to

$$\left|\left(\mathbf{R}^{[t+1]}-\mathbf{R}_0^{[t+1]}\right)_{ij}\right| = \left|\sum_{\{\ell_1,\ell_2\}=\{i_{t+1}\alpha_{t+1}\}}(\mathbf{R}_0^{[t+1]})_{i\ell_1}\widetilde{\mathbf{Q}}^{[t+1]}_{\ell_1\ell_2}\mathbf{R}^{[t+1]}_{\ell_2 j}\right|$$

Thus, on the event $\mathcal{E}_t$, we conclude

$$\left|\left(\mathbf{R}^{[t+1]}-\mathbf{R}_0^{[t+1]}\right)_{ij}\right| \le 4\sqrt{d_{i_{t+1}}}n^{-\frac{1}{2}+\varepsilon}\left(\Psi^2+C_0\Psi\mathbb{1}_{((t+1)\in T_{ij})}\right) =: \mathfrak{f}_{ij}^{[t+1]} \tag{3.50}$$

Meanwhile, the second-order resolvent expansion of $\mathbf{R}^{[t+1]}$ around $\mathbf{R}_0^{[t+1]}$ yields

$$\mathbf{R}^{[t+1]} = \mathbf{R}_0^{[t+1]} - \mathbf{R}_0^{[t+1]}\widetilde{\mathbf{Q}}^{[t+1]}\mathbf{R}_0^{[t+1]} + \left(\mathbf{R}_0^{[t+1]}\widetilde{\mathbf{Q}}^{[t+1]}\right)^2\mathbf{R}_0^{[t+1]} - \left(\mathbf{R}_0^{[t+1]}\widetilde{\mathbf{Q}}^{[t+1]}\right)^3\mathbf{R}^{[t+1]}.$$

A key observation is that $\mathbf{R}_0^{[t+1]}$ is $\mathcal{F}_t$-measurable, and $\mathbb{E}[\widetilde{\mathbf{Q}}^{[t+1]}|\mathcal{F}_t]=0$. For simplicity of notations, we set

$$\mathfrak{q}_{ij}^{[t]} := \left((\mathbf{R}_0^{[t]}\widetilde{\mathbf{E}}^{(i_t,\alpha_t)})^2\mathbf{R}_0^{[t]}\right)_{ij}$$

where $\widetilde{\mathbf{E}}^{(i_t,\alpha_t)} \in \mathbb{R}^{|\mathcal{I}|\times|\mathcal{I}|}$ is the symmetrization of the matrix $\mathbf{E}^{(i_t,\alpha_t)} \in \mathbb{R}^{p\times n}$ whose elements are all 0 except $\widetilde{\mathbf{E}}^{(i_t,\alpha_t)}_{i_t\alpha_t} = \widetilde{\mathbf{E}}^{(i_t,\alpha_t)}_{\alpha_t i_t} = 1$. Then we have

$$\left|\mathbb{E}\left[\mathbf{R}_{ij}^{[t+1]}|\mathcal{F}_t\right] - (\mathbf{R}_0^{[t+1]})_{ij} - p^{-1}\mathfrak{q}_{ij}^{[t+1]}\right| \le 32 d_{i_{t+1}}^{\frac{3}{2}} n^{-\frac{3}{2}+3\varepsilon}\left(\Psi^2C_0^2 + C_0^4\mathbb{1}_{((t+1)\in T_{ij})}\right) =: \mathfrak{g}_{ij}^{[t+1]}. \tag{3.51}$$

Similarly, using resolvent expansion of $\mathbf{R}^{[t]}$ around $\mathbf{R}_0^{[t+1]}$, we obtain

$$\mathbf{R}^{[t]} = \mathbf{R}_0^{[t+1]} - \mathbf{R}_0^{[t+1]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}_0^{[t+1]} + (\mathbf{R}_0^{[t+1]}\widetilde{\mathbf{P}}^{[t+1]})^2\mathbf{R}_0^{[t+1]} - (\mathbf{R}_0^{[t+1]}\widetilde{\mathbf{P}}^{[t+1]})^3\mathbf{R}^{[t]}.$$

By the same arguments as above, on the event $\mathcal{E}_t$, we deduce that

$$\left|\mathbf{R}_{ij}^{[t]} - (\mathbf{R}_0^{[t+1]})_{ij} + \mathbf{X}_{i_{t+1}\alpha_{t+1}}\mathfrak{p}_{ij}^{[t+1]} - \mathbf{X}^2_{i_{t+1}\alpha_{t+1}}\mathfrak{q}_{ij}^{[t+1]}\right| \le \mathfrak{g}_{ij}^{[t+1]} \tag{3.52}$$

where

$$\mathfrak{p}_{ij}^{[t]} := \left(\mathbf{R}_0^{[t]}\widetilde{\mathbf{E}}^{(i_t,\alpha_t)}\mathbf{R}_0^{[t]}\right)_{ij}. \tag{3.53}$$

Combining (3.51) and (3.52) yields

$$\left|\mathbb{E}\left[\mathbf{R}_{ij}^{[t+1]}|\mathcal{F}_t\right] - \mathbf{R}_{ij}^{[t]} - \mathbf{X}_{i_{t+1}\alpha_{t+1}}\mathfrak{p}_{ij}^{[t+1]} + (\mathbf{X}_{i_{t+1}\alpha_{t+1}}^2 - p^{-1})\mathfrak{q}_{ij}^{[t+1]}\right| \leq 2\mathfrak{g}_{ij}^{[t+1]}. \tag{3.54}$$

By a telescopic summation, we obtain

$$\mathbf{R}_{ij}^{[k]} - \mathbf{R}_{ij} = \sum_{t=0}^{k-1}\left(\mathbf{R}_{ij}^{[t+1]} - \mathbf{R}_{ij}^{[t]}\right)$$
$$= \sum_{t=0}^{k-1}\left(\mathbf{R}_{ij}^{[t+1]} - \mathbb{E}\left[\mathbf{R}_{ij}^{[t+1]}|\mathcal{F}_t\right]\right) + \sum_{t=0}^{k-1}\mathbf{X}_{i_{t+1}\alpha_{t+1}}\mathfrak{p}_{ij}^{[t+1]} + \sum_{t=0}^{k-1}(\mathbf{X}_{i_{t+1}\alpha_{t+1}}^2 - p^{-1})\mathfrak{q}_{ij}^{[t+1]} + \mathfrak{r}_{ij} \tag{3.55}$$

where the remainder $\mathfrak{r}_{ij}$ is bounded by (3.54)

$$|\mathfrak{r}_{ij}| \leq 2\sum_{t=0}^{k-1}\mathfrak{g}_{ij}^{[t+1]}.$$

Recall the expression of $\mathfrak{g}_{ij}^{[t]}$, to estimate the remainder, we need to control the size of the set T_{ij}. Note that $\mathbb{E}\left[|T_{ij}|\right] = 2k/p$. By a Bernstein-type inequality (see e.g. [Cha07, Proposition 1.1]), for any $x > 0$, we have

$$\mathbb{P}\left(|T_{ij}| \geq \mathbb{E}\left[|T_{ij}|\right] + x\right) \leq \exp\left(-\frac{x^2}{4\mathbb{E}\left[|T_{ij}|\right] + 2x}\right)$$

Recall that $k \leq n^{5/3-\epsilon_0}$. The inequality together with a union bound implies that

$$\max_{i,j}|T_{ij}| \leq \frac{3\max(k, p(\log n)^2)}{p} =: \mathsf{T}$$

with overwhelming probability. We denote this event by $\mathcal{T}$. On the event $\mathcal{T}$, we have

$$|\mathfrak{r}_{ij}| \lesssim 2kn^{-\frac{3}{2}+3\varepsilon}\Psi^2 C_0^2 + 2n^{-\frac{3}{2}+3\varepsilon}C_0^4\mathsf{T} \lesssim n^{3\varepsilon}\sqrt{\mathsf{T}}\Psi^2. \tag{3.56}$$

For the first term in (3.55), we set

$$\mathfrak{w}_{ij}^{[t+1]} := \left(\mathbf{R}_{ij}^{[t+1]} - \mathbb{E}\left[\mathbf{R}_{ij}^{[t+1]}|\mathcal{F}_t\right]\right)\mathbb{1}_{\mathcal{E}_t}.$$

Note that $\mathcal{E}_t \in \mathcal{F}_t$. This implies that $\mathbb{E}[\mathfrak{w}_{ij}^{[t+1]}|\mathcal{F}_t] = 0$. Moreover, by (3.50), on the event $\mathcal{E}_t$ we have $|\mathfrak{w}_{ij}^{[t+1]}| \le 2\mathfrak{f}_{ij}^{[t+1]}$. Further, on the event $\mathcal{T}$,

$$\left(\sum_{t=0}^{k-1}(\mathfrak{f}_{ij}^{[t+1]})^2\right)^{1/2} \lesssim n^{-\frac{1}{2}+\varepsilon}\Psi^2\sqrt{k} + n^{-\frac{1}{2}+\varepsilon}C_0\Psi\sqrt{\mathsf{T}} \le 2n^{\varepsilon}\Psi^2\sqrt{\mathsf{T}}.$$

Using the Azuma-Hoeffding inequality, for any $x \ge 0$, we have

$$\mathbb{P}\left(\left|\sum_{t=0}^{k-1}\mathfrak{w}_{ij}^{[t+1]}\right| \ge 2n^{\varepsilon}\Psi^2\sqrt{\mathsf{T}}x\right) \le 2\exp\left(-\frac{x^2}{2}\right).$$

Moreover,

$$\mathbb{P}\left(\left|\sum_{t=0}^{k-1}\left(\mathbf{R}_{ij}^{[t+1]} - \mathbb{E}\left[\mathbf{R}_{ij}^{[t+1]}|\mathcal{F}_t\right]\right)\right| \ge 2n^{\varepsilon}\Psi^2\sqrt{\mathsf{T}}x\right) \le \mathbb{P}\left(\left|\sum_{t=0}^{k-1}\mathfrak{w}_{ij}^{[t+1]}\right| \ge 2n^{\varepsilon}\Psi^2\sqrt{\mathsf{T}}x\right) + \sum_{t=0}^{k-1}\mathbb{P}(\mathcal{E}_t^c).$$

Recall that $\mathcal{E}_t$ holds with overwhelming probability, and consequently $\sum_{t=0}^{k-1}\mathbb{P}(\mathcal{E}_t^c) \le n^{-D}$ for any $D > 0$. Choosing $x = n^{\varepsilon}$ implies that with overwhelming probability

$$\left|\sum_{t=0}^{k-1}\left(\mathbf{R}_{ij}^{[t+1]} - \mathbb{E}\left[\mathbf{R}_{ij}^{[t+1]}|\mathcal{F}_t\right]\right)\right| \le 2n^{2\varepsilon}\Psi^2\sqrt{\mathsf{T}}. \tag{3.57}$$

For the next two terms in (3.55), we will deal with them by introducing a backward filtration. Let $\mathcal{F}_t'$ be the σ-algebra generated by the random variables $\mathbf{X}'$, S_k and $\{\mathbf{X}_{i\alpha}\}$ with $i \notin \{i_1, \ldots, i_t\}$ and $\alpha \notin \{\alpha_1, \ldots, \alpha_t\}$. Similarly as above, we consider the event $\mathcal{E}_t'$

that for all $z = E + \mathrm{i}\eta$ with $|z - \lambda_{\mathrm{R}}| \le n^{-2/3+\delta}$ and $\eta = n^{-2/3-\delta}$ we have

$$\max_{a \ne b} \left| \mathbf{R}^{[t]}_{ab}(z) \right| \le \Psi, \quad \text{and} \quad \max_{a} \left| \mathbf{R}^{[t]}_{aa}(z) \right| \le C_0.$$

Using resolvent expansion, the same arguments for (3.49) yield that, on the event $\mathcal{E}'_t$, we have

$$\max_{i \ne j} \left| (\mathbf{R}^{[t]}_0)_{ij} \right| \le 2\Psi, \quad \max_{i} \left| (\mathbf{R}^{[t]}_0)_{ii} \right| \le 2C_0.$$

A key observation is that $\mathfrak{p}^{[t]}_{ij}$ defined in (3.53) is $\mathcal{F}'_t$-measurable. Also, we have $\mathbb{E}[\mathbf{X}_{i_t\alpha_t} | \mathcal{F}'_t] = 0$. Consider

$$\widetilde{\mathfrak{p}}^{[t]}_{ij} := \mathbf{X}_{i_t\alpha_t} \mathfrak{p}^{[t]}_{ij} \mathbb{1}_{\mathcal{E}'_t}.$$

Then we have $\mathbb{E}[\widetilde{\mathfrak{p}}^{[t]}_{ij} | \mathcal{F}'_t] = 0$ since we also have $\mathcal{E}'_t \in \mathcal{F}'_t$. Note that

$$\mathbb{P}\left(\left| \sum_{t=0}^{k-1} \mathbf{X}_{i_{t+1}\alpha_{t+1}} \mathfrak{p}^{[t+1]}_{ij} \right| \ge x \right) \le \mathbb{P}\left(\left| \sum_{t=0}^{k-1} \widetilde{\mathfrak{p}}^{[t+1]}_{ij} \right| \ge x \right) + \sum_{t=0}^{k-1} \mathbb{P}((\mathcal{E}'_{t+1})^c),$$

The second term is negligible since $\mathcal{E}'_t$ holds with overwhelming probability. To estimate the first term, we use Azuma-Hoeffding inequality as before. Based on similar arguments as in (3.50), we deduce

$$\left| \widetilde{\mathfrak{p}}^{[t]}_{ij} \right| \le 4\sqrt{d_{i_t}} n^{-\frac{1}{2}+\varepsilon} \left(\Psi^2 + C_0 \Psi \mathbb{1}_{(t \in T_{\alpha\beta})} \right).$$

By considering the event $\mathcal{T}$ and using Azuma-Hoeffding inequality as in (3.57), we can conclude that with overwhelming probability,

$$\left| \sum_{t=0}^{k-1} \widetilde{\mathfrak{p}}^{[t+1]}_{ij} \right| \le n^{2\varepsilon} \Psi^2 \sqrt{\mathsf{T}}$$

As a consequence, with overwhelming probability

$$\left| \sum_{t=0}^{k-1} \mathbf{X}_{i_{t+1}\alpha_{t+1}} \mathfrak{p}_{ij}^{[t+1]} \right| \lesssim n^{2\varepsilon} \Psi^2 \sqrt{\mathsf{T}}. \tag{3.58}$$

For the third term in (3.55), by the same arguments, we have

$$\left| \sum_{t=0}^{k-1} (\mathbf{X}_{i_{t+1}\alpha_{t+1}}^2 - p^{-1}) \mathfrak{q}_{ij}^{[t+1]} \right| \lesssim n^{2\varepsilon} \Psi^2 \sqrt{\mathsf{T}}. \tag{3.59}$$

Finally, combining (3.55), (3.56), (3.57), (3.58) and (3.59), we have shown that

$$\left| \mathbf{R}_{ij}^{[k]}(z) - \mathbf{R}_{ij}(z) \right| \lesssim n^{3\varepsilon} \Psi^2 \sqrt{\mathsf{T}}.$$

Recall that $\eta = n^{-2/3-\delta}$, $\Psi = O(n^{-\frac{1}{3}+\delta+\varepsilon})$, and $\mathsf{T} = O(n^{\frac{2}{3}-\epsilon_0})$. Then we obtain

$$n\eta \left| \mathbf{R}_{ij}^{[k]}(z) - \mathbf{R}_{ij}(z) \right| \le n^{-\frac{\epsilon_0}{2}+\delta+5\varepsilon}. \tag{3.60}$$

Choosing $\delta + 5\varepsilon < \frac{\epsilon_0}{2}$ yields the desired bound (3.47) for a fixed z.

So far, we have proved the desired result for a fixed z. To extend this result to a uniform estimate, we simply invoke a standard net argument. To do this, we divide the interval $[-n^{-2/3+\delta}, n^{-2/3+\delta}]$ into n^2 sub-intervals, and consider $z = E + \mathrm{i}\eta$ with $\kappa(z)$ taking values in each sub-interval. Note that

$$|\mathbf{R}_{ij}(z_1) - \mathbf{R}_{ij}(z_2)| \le \frac{|z_1 - z_2|}{\min(\mathrm{Im}\,(z_1), \mathrm{Im}\,(z_2))^2}.$$

For z_1, z_2 associated with the same sub-interval, we have

$$n\eta |\mathbf{R}_{ij}(z_1) - \mathbf{R}_{ij}(z_2)| \le n\eta \frac{n^{-2/3+\delta} n^{-2}}{\eta^2} \le n^{-1+2\delta},$$

which is of lower order compared with the error bound in (3.60). This shows that, up to a small multiplicative factor, the desired error bound (3.47) holds uniformly in each sub-interval with overwhelming probability. Finally, thanks to the overwhelming probability, a union bound over the n^2 sub-intervals yields the desired uniform estimate (3.47) for all $z = E + \mathrm{i}\eta$ with $|E - \lambda_{\mathrm{R}}| \leq n^{-2/3+\delta}$ and $\eta = n^{-2/3-\delta}$.

Using the same arguments, we can prove a similar bound for the $\mathbf{R}^{[k]}_{\alpha\beta}$ and $\mathbf{R}_{\alpha\beta}$ blocks. Hence, we have shown the desired results. □

3.4.3 Stability of top eigenvalue

As a consequence of the stability of the resolvent, we also have the stability of the top eigenvalue. This stability of the eigenvalue will play a crucial rule for the resolvent approximation of eigenvector statistics in the next subsection.

Lemma 3.9. *Consider the weakly spiked model $\mathcal{O} = \emptyset$. Assume $k \leq n^{5/3-\epsilon_0}$ for some $\epsilon_0 > 0$. Let $0 < \delta < \delta_0$ with δ_0 as in Lemma 3.8. For any $\varepsilon > 0$, with overwhelming probability, we have*

$$\left|\lambda - \lambda^{[k]}\right| \leq n^{-\frac{2}{3}-\delta+\varepsilon}.$$

Proof. Without loss of generality, we assume that $\lambda > \lambda^{[k]}$. Set $\eta = n^{-2/3-\delta}$. By the spectral representation of the resolvent (3.41), we have

$$\operatorname{Im} \mathbf{R}_{ii}(z) = \eta \sum_{\ell=1}^{p} \frac{|\mathbf{v}_\ell(i)|^2}{(\lambda_\ell - E)^2 + \eta^2} \geq \frac{\eta |\mathbf{v}(i)|^2}{(\lambda - E)^2 + \eta^2} \geq \frac{\eta |\mathbf{v}(i)|^2}{2\left(\max(|\lambda - E|, \eta)\right)^2}.$$

By the pigeonhole principle, we know that there exists $1 \leq i \leq p$ such that $|\mathbf{v}(i)| \geq p^{-1/2}$. Choosing this i and $z = \lambda + \mathrm{i}\eta$, we obtain

$$p\eta^{-1} \operatorname{Im} \mathbf{R}_{ii}(\lambda + \mathrm{i}\eta) \geq \frac{1}{2\eta^2}. \tag{3.61}$$

On the other hand, using the spectral representation of resolvent again for $\mathbf{R}^{[k]}$, we have

$$p\eta^{-1}\mathrm{Im}\,\mathbf{R}^{[k]}_{jj}(z) = \sum_{m=1}^{p} \frac{p|\mathbf{v}^{[k]}_m(j)|^2}{\left(\lambda^{[k]}_m - \lambda\right)^2 + \eta^2}.$$

Pick $\omega > 0$, we decompose the summation into two parts

$$J_1 = \sum_{m=1}^{n^\omega} \frac{p|\mathbf{v}^{[k]}_m(j)|^2}{\left(\lambda^{[k]}_m - \lambda\right)^2 + \eta^2}, \quad J_2 = \sum_{m=n^\omega+1}^{p} \frac{p|\mathbf{v}^{[k]}_m(j)|^2}{\left(\lambda^{[k]}_m - \lambda\right)^2 + \eta^2}.$$

Using delocalization of eigenvectors, for any $\varepsilon > 0$, with overwhelming probability, we have

$$J_1 \lesssim \frac{n^{\omega+\varepsilon}}{(\min_{1\leq m\leq p}|\lambda^{[k]}_m - \lambda|)^2}. \tag{3.62}$$

By the Tracy-Widom limit of the top eigenvalue (Lemma 3.2), for any $\varepsilon > 0$, with overwhelming probability, we have $|\lambda - \lambda_{\mathrm{R}}| \leq n^{-2/3+\varepsilon}$. Also, as discussed in (3.19), the rigidity of eigenvalues yields that for all $m \geq n^\omega$, with overwhelming probability,

$$\lambda - \lambda^{[k]}_m \gtrsim m^{2/3}p^{-2/3}.$$

Then using delocalization again, with overwhelming probability, we have

$$J_2 \leq \sum_{m=n^\omega+1}^{p} \frac{n^\varepsilon}{(\lambda^{[k]}_m - \lambda)^2} \lesssim n^\varepsilon (n^\omega)^{-1/3} n^{4/3}. \tag{3.63}$$

Again, since $|\lambda^{[k]} - \lambda| \leq 2n^{-2/3+\varepsilon}$, by choosing $\omega = 2\varepsilon$ we have $J_2 \leq J_1$. Therefore, by (3.62) and (3.63), we have shown that with overwhelming probability

$$p\eta^{-1}\mathrm{Im}\,\mathbf{R}^{[k]}_{jj}(\lambda + \mathrm{i}\eta) \lesssim n^{3\varepsilon}\left(\min_{1\leq m\leq p}|\lambda^{[k]}_m - \lambda|\right)^{-2}.$$

Note that the minimum is attained by $\lambda^{[k]}$. This shows that

$$n\eta^{-1}\mathrm{Im}\,\mathbf{R}_{ii}^{[k]}(\lambda+\mathrm{i}\eta)\lesssim n^{3\varepsilon}|\lambda^{[k]}-\lambda|^{-2}.$$

Using Lemma 3.8 and (3.61), we have

$$\begin{aligned}&n\eta^{-1}\mathrm{Im}\,\mathbf{R}_{ii}^{[k]}(\lambda+\mathrm{i}\eta)\\ &\geq n\eta^{-1}\left(\mathrm{Im}\,\mathbf{R}_{ii}(\lambda+\mathrm{i}\eta)-\left|\mathrm{Im}\,\mathbf{R}_{ii}^{[k]}(\lambda+\mathrm{i}\eta)-\mathrm{Im}\,\mathbf{R}_{ii}(\lambda+\mathrm{i}\eta)\right|\right)\geq\frac{1}{2\eta^2}-\frac{1}{n^c\eta^2}\gtrsim\frac{1}{\eta^2}.\end{aligned}$$

Therefore, we have shown that, with overwhelming probability,

$$\frac{1}{\eta^2}\lesssim n^{3\varepsilon}\frac{1}{|\lambda-\lambda^{[k]}|^2}.$$

Recall $\eta=n^{-2/3-\delta}$, and we conclude that

$$\left|\lambda-\lambda^{[k]}\right|\leq n^{-2/3-\delta+3\varepsilon},$$

which proves the desired result thanks to the arbitrariness of $\varepsilon>0$. □

3.4.4 Proof of Stability

The final ingredient to prove the resampling stability is the following approximation lemma, which asserts that the product of entries in the eigenvector can be well approximated by the resolvent entries.

Lemma 3.10. *Consider the weakly spiked model $\mathcal{O}=\emptyset$. Assume that $k\ll n^{5/3-\epsilon_0}$ for some $\epsilon_0>0$. Let $0<\delta<\delta_0$ be as in Lemma 3.8. Then, for $z_0=\lambda+\mathrm{i}\eta$ with $\eta=n^{-2/3-\delta}$, there exists $c'>0$ such that with probability $1-o(1)$ we have*

$$\max_{i,j}|\eta\mathrm{Im}\,\mathbf{R}_{ij}(z_0)-\mathbf{v}(i)\mathbf{v}(j)|\leq n^{-1-c'},\ \textit{ and }\ \max_{i,j}\left|\eta\mathrm{Im}\,\mathbf{R}_{ij}^{[k]}(z_0)-\mathbf{v}^{[k]}(i)\mathbf{v}^{[k]}(j)\right|\leq n^{-1-c'}.$$

Similarly, we also have

$$\max_{\alpha,\beta}\left|\eta \mathrm{Im}\,\frac{\mathbf{R}_{\alpha\beta}(z_0)}{z_0}-\mathbf{u}(\alpha)\mathbf{u}(\beta)\right|\leq n^{-1-c'},\ \text{and}\ \max_{\alpha,\beta}\left|\eta \mathrm{Im}\,\frac{\mathbf{R}^{[k]}_{\alpha\beta}(z_0)}{z_0}-\mathbf{u}^{[k]}(\alpha)\mathbf{u}^{[k]}(\beta)\right|\leq n^{-1-c'}.$$

Proof. For any $\varepsilon>0$, we consider a general $z=E+\mathrm{i}\eta$ with $|E-\lambda_{\mathrm{R}}|\leq n^{-2/3+\varepsilon}$. From the spectral representation of the resolvent (3.41), we have

$$\mathrm{Im}\,\mathbf{R}_{ij}(z)=\eta\sum_{\ell=1}^{p}\frac{\mathbf{v}_\ell(i)\mathbf{v}_\ell(j)}{(\lambda_\ell-E)^2+\eta^2}.$$

Pick some $\omega>0$, we decompose the summation on the right-hand side into three parts

$$\sum_{\ell=1}^{p}\frac{\mathbf{v}_\ell(i)\mathbf{v}_\ell(j)}{(\lambda_\ell-E)^2+\eta^2}=\frac{\mathbf{v}(i)\mathbf{v}(j)}{(\lambda-E)^2+\eta^2}+J_1+J_2,$$

where

$$J_1=\sum_{\ell=2}^{n^\omega}\frac{\mathbf{v}_\ell(i)\mathbf{v}_\ell(j)}{(\lambda_\ell-E)^2+\eta^2},\quad J_2=\sum_{\ell=n^\omega+1}^{p}\frac{\mathbf{v}_\ell(i)\mathbf{v}_\ell(j)}{(\lambda_\ell-E)^2+\eta^2}.$$

Using the same arguments as in (3.63), for any $\varepsilon>0$, with overwhelming probability we have

$$|J_2|\lesssim n^{\varepsilon}(n^{\omega})^{-1/3}n^{1/3}.$$

For the term J_1, we consider the following event

$$\mathcal{E}:=\left\{\lambda_1-\lambda_2\geq c_0n^{-2/3}\right\}\cap\left\{\max_{1\leq\ell\leq p}\|\mathbf{v}_\ell\|_\infty\leq n^{-1/2+\varepsilon}\right\}\cap\left\{|J_2|\lesssim n^{\varepsilon}(n^{\omega})^{-1/3}n^{4/3}\right\}.$$

For any $\varepsilon>0$, we can find an appropriate $c_0>0$ such that $\mathbb{P}(\mathcal{E})>1-\varepsilon/2$. Then, for $z=E+\mathrm{i}\eta$ with $|\lambda-E|\leq\frac{c_0}{2}n^{-2/3}$, on the event $\mathcal{E}$, we have

$$|J_1|\lesssim n^{\varepsilon}n^{\omega}n^{1/3}.$$

Let $\delta'>0$ with $\delta'+\delta<\delta_0$. On the event $\mathcal{E}$, for all $z=E+\mathrm{i}\eta$ with $|\lambda-E|\leq\eta n^{-\delta'}$ and

$\eta = n^{-2/3-\delta}$, we have

$$\left|\mathbf{v}(i)\mathbf{v}(j) - \frac{\eta^2\mathbf{v}(i)\mathbf{v}(j)}{(\lambda-E)^2+\eta^2}\right| \le n^{-1+2\varepsilon}\left|1 - \frac{\eta^2}{(\lambda-E)^2+\eta^2}\right| \le n^{-1+2\varepsilon-2\delta'}.$$

This yields

$$\begin{aligned}|\eta\mathrm{Im}\,\mathbf{R}_{ij}(z) - \mathbf{v}(i)\mathbf{v}(j)| &\le \left|\mathbf{v}(i)\mathbf{v}(j) - \frac{\eta^2\mathbf{v}(i)\mathbf{v}(j)}{(\lambda-E)^2+\eta^2}\right| + \eta^2(|J_1|+|J_2|)\\ &\le n^{-1+2\varepsilon-2\delta'} + n^{-1+\varepsilon+\omega-2\delta} + n^{-1+\varepsilon-\frac{\omega}{3}-2\delta}.\end{aligned}$$

Choosing $\omega = \varepsilon < \min(\delta,\delta')/2$, we obtain

$$\max_{i,j}|\eta\mathrm{Im}\,\mathbf{R}_{ij}(z) - \mathbf{v}(i)\mathbf{v}(j)| \le n^{-1-\min(\delta,\delta')}. \tag{3.64}$$

Similarly, we can apply the same arguments to $\mathbf{R}^{[k]}$. Consider the event

$$\mathcal{E}' := \left\{\max_{i,j}\left|\eta\mathrm{Im}\,\mathbf{R}^{[k]}_{ij}(z) - \mathbf{v}^{[k]}(i)\mathbf{v}^{[k]}(j)\right| \le n^{-1-\min(\delta,\delta')} \text{ for all } |\lambda^{[k]}-E| \le \eta n^{-\delta'}, \eta = n^{-2/3-\delta}\right\}.$$

By previous arguments, we know $\mathbb{P}(\mathcal{E}') > 1-\varepsilon/2$. This gives us $\mathbb{P}(\mathcal{E}\cap\mathcal{E}') > 1-\varepsilon$. Finally, note that $\delta+\delta' < \delta_0$, by Lemma 3.9, with overwhelming probability we have $|\lambda - \lambda^{[k]}| \le n^{-2/3-\delta-\delta'} = \eta n^{-\delta'}$. This implies that we can choose $z = \lambda + \mathrm{i}\eta$ in both (3.64) and $\mathcal{E}'$. Thus, we have shown the desired result for $\mathbf{v}$ and $\mathbf{v}^{[k]}$ by choosing $0 < c' < \min(\delta,\delta')$.

On the other hand, from (3.41) we also have

$$\mathrm{Im}\,\frac{\mathbf{R}_{\alpha\beta}(z)}{z} = \eta\sum_{\ell=1}^{n}\frac{\mathbf{u}_\ell(\alpha)\mathbf{u}_\ell(\beta)}{(\lambda_\ell-E)^2+\eta^2}.$$

Using the same methods as above yields the desired result for $\mathbf{u}$ and $\mathbf{u}^{[k]}$. □

Now we prove the main result Theorem 3.2 on the stability of PCA under moderate resampling for the weakly spiked model.

Proof of Theorem 3.2. Let $z_0 = \lambda + \mathrm{i}\eta$ as in Lemma 3.10. By Lemma 3.8 and 3.10, we know that, with probability $1 - o(1)$, for all $\alpha, \beta \in \mathcal{I}_2$, we have

$$\begin{aligned}
&\left|\mathbf{v}(i)\mathbf{v}(j) - \mathbf{v}^{[k]}(i)\mathbf{v}^{[k]}(j)\right| \\
&\leq |\mathbf{v}(i)\mathbf{v}(j) - \eta \mathrm{Im}\, \mathbf{R}_{ij}(z_0)| + \left|\eta \mathrm{Im}\, \mathbf{R}_{ij}(z_0) - \eta \mathrm{Im}\, \mathbf{R}_{ij}^{[k]}(z_0)\right| + \left|\eta \mathrm{Im}\, \mathbf{R}_{ij}^{[k]}(z_0) - \mathbf{v}^{[k]}(i)\mathbf{v}^{[k]}(j)\right| \\
&\leq n^{-1-c} + n^{-1-c'} + n^{-1-c}.
\end{aligned}$$

Denote $c'' := \min(c, c')$, and we have

$$\max_{i,j} \left|\mathbf{v}(i)\mathbf{v}(j) - \mathbf{v}^{[k]}(i)\mathbf{v}^{[k]}(j)\right| \lesssim n^{-1-c''}.$$

For any fixed $\varepsilon > 0$, we consider the event

$$\mathcal{E} := \left\{\max_{i,j} \left|\mathbf{v}(i)\mathbf{v}(j) - \mathbf{v}^{[k]}(i)\mathbf{v}^{[k]}(j)\right| \lesssim n^{-1-c''}\right\} \cap \left\{\|\mathbf{v}^{[k]}\|_\infty \leq n^{-1/2+\varepsilon}\right\}.$$

Since delocalization of eigenvectors holds with overwhelming probability, we know that $\mathbb{P}(\mathcal{E}) = 1 - o(1)$.

By the pigeonhole principle, there exists $1 \leq i \leq p$ such that $|\mathbf{v}(i)| > p^{-1/2}$. We choose the $\pm$ phases of $\mathbf{v}$ and $\mathbf{v}^{[k]}$ in the way that $\mathbf{v}(i)$ and $\mathbf{v}^{[k]}(i)$ are non-negative. On the event $\mathcal{E}$, we obtain

$$\left|\mathbf{v}(i) - \mathbf{v}^{[k]}(i)\right| = \frac{\left|(\mathbf{v}(i))^2 - (\mathbf{v}^{[k]}(i))^2\right|}{\mathbf{v}(i) + \mathbf{v}^{[k]}(i)} \lesssim n^{-1/2-c''}.$$

Moreover, for any entry $\mathbf{v}(j)$ and $\mathbf{v}^{[k]}(j)$, if $\mathcal{E}$ holds, the triangle inequality gives us

$$\begin{aligned}
\left|\mathbf{v}(j) - \mathbf{v}^{[k]}(j)\right| &= \frac{\left|\mathbf{v}(i)\mathbf{v}(j) - \mathbf{v}(i)\mathbf{v}^{[k]}(j)\right|}{\mathbf{v}(i)} \\
&\leq \frac{\left|\mathbf{v}(i)\mathbf{v}(j) - \mathbf{v}^{[k]}(i)\mathbf{v}^{[k]}(j)\right|}{\mathbf{v}(i)} + \frac{|\mathbf{v}^{[k]}(j)|}{\mathbf{v}(i)}|\mathbf{v}(i) - \mathbf{v}^{[k]}(i)| \\
&\lesssim n^{-1/2-c''} + n^{-1/2-c''+\varepsilon}.
\end{aligned}$$

Choosing ε sufficiently small, this implies the desired result.

For $\mathbf{u}$ and $\mathbf{u}^{[k]}$, note that

$$\begin{aligned}
&\left|\mathbf{u}(\alpha)\mathbf{u}(\beta)-\mathbf{u}^{[k]}(\alpha)\mathbf{u}^{[k]}(\beta)\right| \\
\leq &\left|\mathbf{u}(\alpha)\mathbf{u}(\beta)-\eta\mathrm{Im}\,\frac{\mathbf{R}_{\alpha\beta}(z_0)}{z_0}\right|+\left|\eta\mathrm{Im}\,\frac{\mathbf{R}_{\alpha\beta}(z_0)}{z_0}-\eta\mathrm{Im}\,\frac{\mathbf{R}^{[k]}_{\alpha\beta}(z_0)}{z_0}\right|+\left|\eta\mathrm{Im}\,\frac{\mathbf{R}^{[k]}_{\alpha\beta}(z_0)}{z_0}-\mathbf{u}^{[k]}(\alpha)\mathbf{u}^{[k]}(\beta)\right|
\end{aligned}$$

By Lemma 3.8, we have

$$\left|\mathrm{Im}\,\frac{\mathbf{R}_{\alpha\beta}(z_0)}{z_0}-\mathrm{Im}\,\frac{\mathbf{R}^{[k]}_{\alpha\beta}(z_0)}{z_0}\right|\leq\left|\frac{\mathbf{R}_{\alpha\beta}(z_0)-\mathbf{R}^{[k]}_{\alpha\beta}(z_0)}{z_0}\right|\lesssim\left|\mathbf{R}_{\alpha\beta}(z_0)-\mathbf{R}^{[k]}_{\alpha\beta}(z_0)\right|\leq\frac{1}{n^{1+c}\eta}.$$

As a consequence, we have

$$\left|\mathbf{u}(\alpha)\mathbf{u}(\beta)-\mathbf{u}^{[k]}(\alpha)\mathbf{u}^{[k]}(\beta)\right|\lesssim n^{-1-c''}.$$

The desired result for $\mathbf{u}$ and $\mathbf{u}^{[k]}$ then follows from the same arguments above for $\mathbf{v}$ and $\mathbf{v}^{[k]}$. □

3.5 Stability of Strongly Spiked Model

As discussed in Section 3.2, one of the key differences between the weakly spiked model and the strongly spiked model is the distribution of eigenvectors. We see from the previous sections that the proof for the weakly spiked model crucially depends on the delocalization property. In contrast, this is not valid in the strongly spiked case and consequently results in a distinct phenomenon.

In the strongly spiked model, the celebrated BBP phase transition [BBAP05] shows that the leading sample eigenvectors in the outlier of the spectrum have non-trivial correlation with the corresponding population eigenvectors. Recall that the population covariance matrix is in the form $\mathbf{\Sigma}=\sum_{i=1}^{p}d_i\mathbf{e}_i\mathbf{e}_i^\top$, and indices i with $d_i>1+\sqrt{\xi}$ correspond to the outlier (denoted as $\mathcal{O}$). For $i \in \mathcal{O}$, it was first derived in [Lu02] and later generalized in [JL09, BGN11] that

$$|\langle \mathbf{v}_i, \mathbf{e}_i \rangle|^2 = \frac{1 - \frac{\xi}{(d_i-1)^2}}{1 + \frac{\xi}{d_i-1}} + o(1), \quad a.s. \tag{3.65}$$

Since $\mathbf{X}$ and $\mathbf{X}^{[k]}$ have the same marginal distribution, the same also holds for $\mathbf{v}_i^{[k]}$. Note that the eigenvector overlap $|\langle \mathbf{v}, \mathbf{v}^{[k]} \rangle|$ is independent of the sign the of principal components. Therefore, without loss of generality, we may assume that

$$\langle \mathbf{v}_i, \mathbf{e}_i \rangle = \langle \mathbf{v}_i^{[k]}, \mathbf{e}_i \rangle = \sqrt{\frac{1 - \frac{\xi}{(d_i-1)^2}}{1 + \frac{\xi}{d_i-1}}} + o(1)$$

Since both principal components $\mathbf{v}$ and $\mathbf{v}^{[k]}$ lie on the unit sphere, we obtain

$$\|\mathbf{v}_i - \mathbf{e}_i\|^2 = 2 - 2\langle \mathbf{v}_i, \mathbf{e}_i \rangle = 2 - 2\sqrt{\frac{1 - \frac{\xi}{(d_i-1)^2}}{1 + \frac{\xi}{d_i-1}}} + o(1)$$

and same also holds for $\mathbf{v}_i^{[k]}$. By triangle inequality,

$$\|\mathbf{v} - \mathbf{v}^{[k]}\| \le \|\mathbf{v}_i - \mathbf{e}_i\| + \|\mathbf{v}_i^{[k]} - \mathbf{e}_i\| \le 2\sqrt{2 - 2\sqrt{\frac{1 - \frac{\xi}{(d_i-1)^2}}{1 + \frac{\xi}{d_i-1}}}} + o(1)$$

Hence,

$$|\langle \mathbf{v}_i, \mathbf{v}_i^{[k]} \rangle| = \frac{2 - \|\mathbf{v}_i - \mathbf{v}_i^{[k]}\|^2}{2} \ge 1 - 4\left(1 - \sqrt{\frac{1 - \frac{\xi}{(d_i-1)^2}}{1 + \frac{\xi}{d_i-1}}}\right) + o(1)$$

which completes the proof.

Chapter 4

Random Graph Matching for Geometric Models

4.1 Introduction

Graph matching (or network alignment) refers to finding the best vertex correspondence between two graphs that maximizes the total number of common edges. While this problem, as an instance of quadratic assignment problem, is computationally intractable in the worst case, significant headways, both information-theoretic and algorithmic, have been achieved in the average-case analysis under meaningful statistical models [CK16, CK17, DMWX21, BCL+19, FMWX19a, FMWX19b, HM20, WXY21, GM20, GML22, MRT21b, MRT21a]. One of the most popular models is the *correlated Erdős-Rényi graph* model [PG11], where both observed graphs are Erdős-Rényi graphs with edges correlated through a latent vertex matching; more generally, in the *correlated Wigner* model, the observations are two weighted graph with correlated edge weights (e.g. Gaussians [DMWX21, DCK19, FMWX19a, Gan21a]). Despite their simplicity, these models inspired a number of new algorithms that achieve strong performance both theoretically and practically [DMWX21, FMWX19a, FMWX19b, GM20, GML22, MRT21b, MRT21a]. Nevertheless, one of the major limitations of models with independent edges is that they fail to capture graphs with spatial structure [AG14], such as those arising in computer vision datasets (e.g. mesh graphs obtained by triangulating 3D shapes [LRB+16]). In contrast to Erdős-Rényi-style model with iid edges, *geometric graph models*, such as random dot-product graphs and random geometric graphs, take into account the latent geometry by embedding each node in a Euclidean space and determines edge connection between two nodes by the proximity of their geographical location. While the coordinates are typically assumed to be independent (e.g. Gaussians or uniform over spheres or hypercubes), the edges or edge weights are now dependent. The main objective for the present paper is to study graph matching in correlated geometric graph models, where the network correlation is due to that of the latent coordinates.

4.1.1 Model

Given two point clouds $\{X_1, \ldots, X_n\}$ and $\{Y_1, \ldots, Y_n\}$ in $\mathbb{R}^d$, we construct two weighted graphs on the vertex set $[n]$ with weighted adjacency matrices A and B as follows. For each i, j, let $A_{ij} \overset{\text{ind}}{\sim} W(\cdot|X_i, X_j)$ and $B_{ij} \overset{\text{ind}}{\sim} W(\cdot|Y_i, Y_j)$, for some probability transition kernel W. The coordinates are correlated through a latent matching as follows: Consider a Gaussian model

$$Y_i = X_{\pi^*(i)} + \sigma Z_i, \quad i = 1, \ldots, n,$$

where X_i, Z_i's are iid $\mathcal{N}(0, I_d)$ vectors and π^* is uniform on S_n, the set of all permutations on $[n]$. In matrix form, we have

$$Y = \Pi^* X + \sigma Z, \tag{4.1}$$

where $X, Y, Z \in \mathbb{R}^{n \times d}$ are matrices whose rows are X_i's, Y_i's and Z_i's respectively, $\Pi^* \in \mathcal{S}_n$ denotes the permutation matrix corresponding to π^*, and $\mathcal{S}_n$ is the collection of all permutation matrices. Given the observation A and B, the goal is to recover the latent correspondence π^*.

Of particular interest are the following special cases:

- *Dot-product model*: The observations are complete graphs with pairwise inner products as edge weights, namely, $A_{ij} = \langle X_i, X_j \rangle$ and $B_{ij} = \langle Y_i, Y_j \rangle$. As such, the weighted adjacency matrices are $A = XX^\top$ and $B = YY^\top$, both Wishart matrices. It is clear that from A and B one can reconstruct X and Y respectively, each up to a global orthogonal transformation on the rows. In this light, the model is also equivalent to the so-called *Procrustes Matching* problem [MDK+16, DL17, GJB19], where Y in (4.1) undergoes a further random orthogonal transformation

- *Distance model*: The edge weights are pairwise squared distances $A_{ij} = \|X_i - X_j\|^2$ and $B_{ij} = \|X_i - X_j\|^2$. This setting corresponds to the classical problem of multidimensional scaling (MDS), where the goal is to reconstruct the coordinates (up to global shift and orthogonal transformation) from the distance data (cf. [BG05]).

- *Random Dot Product Graph (RDPG)*: In this model, the observed data are two graphs with adjacency matrices A and B, where $A_{ij} \overset{\text{ind}}{\sim} \text{Bern}\,(\kappa(\langle X_i, X_j \rangle))$ and $B_{ij} \overset{\text{ind}}{\sim} \text{Bern}\,(\kappa(\langle X_i, X_j \rangle))$ conditioned on X and Y, and $\kappa : \mathbb{R} \to [0, 1]$ is some link function, e.g. $\kappa(t) = e^{-t^2/2}$. In this way, we observe two instances of RDPG that are correlated through the underlying points and the latent matching. See [AFT+17] for a recent survey on RDPG.

- *Random Geometric Graph (RGG)*: Similar to RDPG, $A_{ij} \overset{\text{ind}}{\sim} \text{Bern}(\kappa(\|X_i - X_j\|))$ conditioned on $X_1, \ldots, X_n$ for some link function $\kappa : \mathbb{R}_+ \to [0, 1]$ applied to the pairwise distances. The second RGG instance B is constructed in the same way using $Y_1, \ldots, Y_n$. A simple example is $\kappa(t) = \mathbf{1}_{\{t \le r\}}$ for some threshold $r > 0$, where each pair of points within distance r is connected [Gil61]; see the monograph [Pen03] for a comprehensive discussion on RGG.

Let us mention that the model where the two point clouds are directly observed has been recently studied by [DCK19, DCK20] in the context of feature matching and inde-

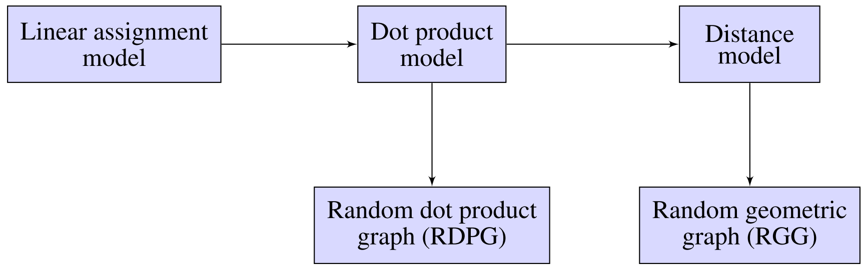

Figure 4.1: Geometric matching models. Here arrows denote statistical ordering.

pendently by [KNW22] as a geometric model for the planted matching problem, extending the previous work in [CKK$^+$10, MMX21, DWXY21] with iid weights to a geometric (low-rank) setting. In this model, X and Y in (4.2) are observed and the maximum likelihood estimator (MLE) of π^* amounts to solving

$$\max_{\Pi \in \mathcal{S}_n} \langle Y, \Pi X \rangle. \tag{4.2}$$

which is a linear assignment (max-weight matching) problem on the weighted complete bipartite graph with weight matrix $YX^\top$. In the sequel we shall refer to this setting as the linear assignment model, which we also study in this paper for the sake of proving impossibility results for the more difficult graph matching problem, as the coordinates are latent and only *pairwise information* are available.

Figure 4.1.1 elucidates the logical connections between the aforementioned models. Among these, linear assignment model is the most informative, followed by the dot product model and the distance model, whose further stochastically degraded versions are RDPG and RGG, respectively. As a first step towards understanding graph matching on geometric models, in this paper we study the case of weighted complete graphs in the dot product and distance models.

4.1.2 Main Results

By analyzing the MLE (4.2) in the stronger linear assignment model (4.1), [KNW22] identified a critical scaling of dimension d at $\log n$:

- In the low-dimensional regime of $d \ll \log n$, accurate reconstruction requires the noise level σ to be vanishingly small. More precisely, with high probability, the MLE (4.2) recovers the latent π^* perfectly (resp. with a vanishing fraction of errors) provided that $\sigma = o(n^{-2/d})$ (resp. $\sigma = o(n^{-1/d})$).

- In the high-dimensional regime of $d \gg \log n$, it is possible for σ^2 to be as large as $\frac{d}{(4+o(1))\log n}$. Since the dependency between the edges weakens as the latent dimension increases[1], this is consistent with the known results in the correlated Erdős-Rényi and Wigner model. For example, to match two GOE matrices with correlation coefficient ρ, the sharp reconstruction threshold is at $\rho^2 = \frac{(4+o(1))\log n}{n}$ [Gan21b, WXY21]

In this paper we mostly focus on the low-dimensional setting as this is the regime where geometric graph ensembles are structurally distinct from Erdős-Rényi graphs. Our main findings are two-fold:

- The same reconstruction thresholds remain achievable even when the coordinates are latent and only inner-product or distance data are accessible.

- Furthermore, these thresholds cannot be improved even when the coordinates are observed.

To make these results precise, we start with the dot-product model with $A = XX^\top$ and $B = YY^\top$, and $Y = \Pi^* X + \sigma Z$ according to (4.1). In this case the MLE turns out to be

1. For the Wishart matrix, it is known [JL15, BG18] that the total variation between the joint law of the off-diagonals and their iid Gaussian counterpart converges to zero provided that $d = \omega(n^3)$. Analogous results have also been obtained in [BDER16] showing that high-dimensional RGG is approximately Erdős-Rényi.

much more complicated than (4.2) for the linear assignment model. The MLE takes the form

$$\widehat{\Pi}_{\text{ML}} = \arg\max_{\Pi \in \mathcal{S}_n} \int_{O(d)} \mathrm{d}Q \exp\left(\frac{\langle B^{1/2}, \Pi A^{1/2} Q\rangle}{\sigma^2}\right), \tag{4.3}$$

where the integral is with respect to the Haar measure on the orthogonal group $O(d)$, $A^{1/2} \triangleq U\Lambda^{1/2} \in \mathbb{R}^{n\times d}$ based on the SVD $A = U\Lambda U^\top$, and similarly for $B^{1/2}$. It is unclear whether the above Haar integral has a closed-form solution[2], let alone how to optimize it over all permutations. Next, we turn to its approximation.

As we will show later, in the low-dimensional case of $d = o(\log n)$, meaningful reconstruction of the latent matching is information-theoretically impossible unless σ vanishes with n at a certain speed. In the regime of small σ, Laplace's method suggests that the predominant contribution to the integral in (4.3) comes from the maximum $\langle B^{1/2}, \Pi A^{1/2} Q\rangle$ over $Q \in O(d)$. Using the dual form of the nuclear norm $\|X\|_* = \max_{Q\in O(d)}\langle X, Q\rangle$, where $\|X\|_*$ denotes the sum of all singular values of X, we arrive at the following approximate MLE:

$$\widehat{\Pi}_{\text{AML}} = \arg\max_{\Pi \in \mathcal{S}_n} \|(A^{1/2})^\top \Pi^\top B^{1/2}\|_*. \tag{4.4}$$

We stress that the above approximation to the MLE (4.3) is justified for the low-dimensional regime where σ is small. In the high-dimensional (high-noise) case, the approximate MLE actually takes on the form of a quadratic assignment problem (QAP), which is the MLE for the well-studied iid model [CK16]; in the special case of the dot-product model, it amounts to replacing the nuclear norm in (4.4) by the Frobenius norm.

To measure the accuracy of a given estimator $\widehat{\pi}$, we define

$$\mathsf{overlap}(\widehat{\pi}, \pi) \triangleq \frac{1}{n} |\{i \in [n] : \widehat{\pi}(i) = \pi(i)\}|$$

2. The integral in (4.3) can be reduced to computing $\int \mathrm{d}Q \exp(\langle \Lambda, Q\rangle)$ for a diagonal Λ, which, in principle, can be evaluated by Taylor expansion and applying formulas for the joint moments of Q in [Mat13, Theorem 2.2].

as the fraction of nodes whose matching is correctly recovered. The following result identifies the threshold at which the approximate MLE achieves perfect or almost perfect recovery.

Theorem 4.1 (Recovery guarantee of AML in the dot-product model)**.** *Assume the dot-product model with $d = o(\log n)$. Let $\widehat{\pi}_{\rm AML}$ be the approximate MLE defined in (4.4).*

(i) If $\sigma \ll n^{-2/d}$, the estimator $\widehat{\pi}_{\rm AML}$ achieves perfect recovery with high probability:

$$\mathbb{P}\left\{\mathsf{overlap}(\widehat{\pi}_{\rm AML}, \pi^*) = 1\right\} = 1 - o(1). \tag{4.5}$$

(ii) If $\sigma \ll n^{-1/d}$, the estimator $\widehat{\pi}_{\rm AML}$ achieves almost perfect recovery with high probability:

$$\mathbb{P}\left\{\mathsf{overlap}(\widehat{\pi}_{\rm AML}, \pi^*) \geq 1 - o(1)\right\} = 1 - o(1). \tag{4.6}$$

A few remarks are in order:

- In fact we will show the following nonasymptotic estimate that implies (4.6): For all sufficiently small ε, if $\sigma^{-d} > 16n2^{2/\varepsilon}$, then $\mathsf{overlap}(\widehat{\pi}_{\rm AML}, \pi^*) \geq 1 - \varepsilon$ with probability tending to one.
- The estimator (4.4) has previously appeared in the literature of Procrustes matching [GJB19], albeit not as an approximation to the MLE in a generative model.
- Unlike linear assignment, it is unclear how to solve the optimization in (4.4) over permutations efficiently. Nevertheless, for constant d we show that it is possible to find an approximate solution in time that is polynomial in n that achieves the same statistical guarantee as in Theorem 4.1. Indeed, note that (4.3) is equivalent to the double maximization

$$\widehat{\Pi}_{\rm AML} = \arg\max_{\Pi \in \mathcal{S}_n} \max_{Q \in O(d)} \langle B^{1/2}, \Pi A^{1/2} Q \rangle. \tag{4.7}$$

Approximating the inner maximum over a suitable discretization of $O(d)$, each maximization over Π for fixed Q is a linear assignment problem, which can be solved in $O(n^3)$ time. It can be argued that (4.7) can be further approximated by the classical spectral algorithm of Umeyama [Ume88] which is much faster in practice and achieves good empirical performance. For d that grows with n, it is an open question to find a polynomial-time algorithm that attains the (optimal, as we show next) threshold in Theorem 4.1.

Next, we proceed to the more difficult distance model, where $A_{ij} = \|X_i - X_j\|^2$ and $B_{ij} = \|Y_i - Y_j\|^2$. Deriving the exact MLE in this model appears to be challenging; instead, we apply the estimator (4.4) to an appropriately centered version of the data matrices. Let $\mathbf{1} \in \mathbb{R}^n$ denotes the all-one vector and define $\mathbf{F} = \frac{1}{n}\mathbf{1}\mathbf{1}^\top$. Then $A = -2XX^\top + a\mathbf{1}^\top + \mathbf{1}a^\top$ and $B = -2YY^\top + b\mathbf{1}^\top + \mathbf{1}b^\top$, where $a = (\|X_i\|^2)$ and $b = (\|Y_i\|^2)$. Strictly speaking, the vectors a and b are correlated with the ground truth π^*, since b can be viewed as a noisy version of $\Pi^* a$; however, we expect them to inform very little about π^* because such scalar-valued observations are highly sensitive to noise (analogous to degree matching in correlated Erdős-Rényi graphs [DMWX21, Section 1.3]). As such, we ignore a and b by projecting A and B to the orthogonal complement of the vector $\mathbf{1}$. Specifically, we compute, as commonly done in the MDS literature (see e.g. [SRZF03, OMK10]),

$$\widetilde{A} = -\frac{1}{2}(I - \mathbf{F})A(I - \mathbf{F}), \quad \widetilde{B} = -\frac{1}{2}(I - \mathbf{F})B(I - \mathbf{F}). \tag{4.8}$$

It is easy to verify that $\widetilde{A} = \widetilde{X}\widetilde{X}^\top$ and $\widetilde{B} = \widetilde{Y}\widetilde{Y}^\top$, where $\widetilde{X} = (I - \mathbf{F})X$ and $\widetilde{Y} = (I - \mathbf{F})Y$ consist of centered coordinates $\widetilde{X}_i = X_i - \bar{X}$ and $\widetilde{Y}_i = Y_i - \bar{Y}$ respectively, with $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$. Overall, we have reduced the distance model to a dot product model where the latent coordinates are now centered.

One can show that the MLE of Π^* given the reduced data $(\widetilde{A}, \widetilde{B})$ is of the same Haar-integral form (4.3). Using again the small-σ approximation, we arrive at the following estimator by applying (4.4) to the centered data $\widetilde{A}$ and $\widetilde{B}$:

$$\widetilde{\Pi}_{\mathrm{AML}} = \arg\max_{\Pi \in \mathcal{S}_n} \|(\widetilde{A}^{1/2})^\top \Pi^\top \widetilde{B}^{1/2}\|_*. \tag{4.9}$$

Theorem 4.2 (Recovery guarantee in the distance model)**.** *Assuming the distance model, Theorem 4.1 holds under the same condition on d and σ, with the estimator $\widetilde{\Pi}_{\mathrm{AML}}$ in (4.9) replacing $\widehat{\Pi}_{\mathrm{AML}}$ in (4.4).*

Finally, we state an impossibility result for the linear assignment model, proving that the perfect and almost perfect recovery threshold of $\sigma = o(n^{-2/d})$ and $\sigma = o(n^{-1/d})$ obtained by analyzing the MLE in [KNW22] are in fact information-theoretically necessary. Complementing Theorem 4.1 and Theorem 4.2, this result also establishes the optimality of the estimator (4.4) and (4.9) for their respective model.

Theorem 4.3 (Impossibility result in the linear assignment model)**.** *Consider the linear assignment model with $d = o(\log n)$.*

(i) If there exists an estimator that achieves perfect recovery with high probability, then $\sigma \le n^{-2/d}$.

(ii) If there exists an estimator that achieves almost perfect recovery with high probability, then $\sigma \le n^{-(1-o(1))/d}$.

Furthermore, in the special case of $d = \Theta(1)$, necessary conditions in (i) and (ii) can be improved to $\sigma \le o(n^{-2/d})$ and $\sigma \le o(n^{-1/d})$, respectively.

Theorem 4.3(i) slightly improves the necessary condition for perfect recovery in [KNW22] from $\sigma = O(n^{-2/d})$ to $\sigma = o(n^{-2/d})$. For almost perfect recovery, the negative result in [KNW22] is limited to MLE, while Theorem 4.3 holds for all algorithms. Moreover, the necessary condition in Theorem 4.3(ii) was conjectured in [KNW22, Conjecture 1.4, item 1], which we now resolve in the positive. Finally, while our focus is in the low-dimensional case of $d = o(\log n)$, we also provide necessary conditions that hold for general d.

In view of Fig. 4.1.1, since the negative results in Theorem 4.3 are proved for the strongest model and the positive results in Theorem 4.2 are for the weakest model, we conclude that for all three models, namely, linear assignment, dot-product, and distance model, the thresholds for exact and almost perfect reconstruction is given by $n^{-2/d}$ and $n^{-1/d}$, respectively.

4.2 Outline of Proofs

4.2.1 Derivation of the Maximum Likilihood Estimator

To compute the "likelihood" of the observation (A, B) given the ground truth Π^*, it is useful to keep in mind of the graphical model

$$\begin{array}{ccccc} \Pi^* & \longrightarrow & Y & \longrightarrow & B \\ & & \uparrow & & \\ & & X & \longrightarrow & A \end{array}$$

where X, Y, Π^* is related via (4.1), $A = XX^\top$, and $B = YY^\top$.

Note that A are B are rank-deficient. To compute the density of (A, B) conditioned on Π^* meaningfully, one needs to choose an appropriate reference measure μ and evaluate the relative density $\frac{\mathrm{d}P_{A,B|\Pi^*}}{\mathrm{d}\mu}$. Let us choose μ to be the product of the marginal distributions of A and B, which does not depend on Π^*. For any rank-d positive semidefinite matrices A_0 and B_0, define $A_0^{1/2} \triangleq U_0\Lambda^{1/2}$ and $B_0^{1/2} \triangleq V_0 D^{1/2}$ based on the SVD $A_0 = U_0\Lambda_0^{1/2} Q_0^\top$ and $B_0 = V_0 D_0 O_0^\top$, where $Q_0, O_0 \in O(d)$ and $U_0, V_0 \in V_{n,d} \triangleq \{U \in \mathbb{R}^{n\times d} : U^\top U = I_d\}$ (the Stiefel manifold). We aim to show

$$\frac{\mathrm{d}P_{A,B|\Pi^*}(A_0, B_0|\Pi)}{\mathrm{d}\mu(A_0, B_0)} = h(A_0, B_0) \int_{O(d)} \mathrm{d}Q \exp\left(\frac{\langle B_0^{1/2}, \Pi A_0^{1/2} Q\rangle}{\sigma^2}\right) \tag{4.10}$$

for some fixed function h, where the integral is with respect to the Haar measure on $O(d)$.

This justifies the MLE in (4.3) for the dot-product model.

To show (4.10), denote by $N_\delta(U_0) = \{U \in V_{n,d} : \|U - U_0\|_{\mathrm{F}} \le \delta\}$ and $N_\delta(\Lambda_0) = \{\Lambda \text{ diagonal} : \|\Lambda - \Lambda_0\|_{\ell_\infty} \le \delta\}$ neighborhoods of U_0 and Λ_0 respectively. (Their specific definitions are not crucial.) Consider a δ-neighborhood of A_0 of the following form:

$$N_\delta(A_0) \triangleq \{U\Lambda U^\top : U \in N_\delta(U_0), \Lambda \in N_\delta(\Lambda_0)\}$$

and similarly define $N_\delta(B_0)$. Write the SVD for X as $X = URQ^\top$, where $U \in V_{n,d}, Q \in O(d)$ and the diagonal matrix R are mutually independent; in particular, Q is uniformly distributed over $O(d)$. Then for constant $C = C(n, d, \sigma)$,

$$\begin{aligned}
&\mathbb{P}[A \in N_\delta(A_0), B \in N_\delta(B_0) | \Pi^* = \Pi] \\
&= \mathbb{E}[\mathbf{1}_{\{XX^\top \in N_\delta(A_0)\}} \mathbf{1}_{\{YY^\top \in N_\delta(B_0)\}} | \Pi^* = \Pi] \\
&= \mathbb{E}[\mathbf{1}_{\{U \in N_\delta(U_0)\}} \mathbf{1}_{\{R \in N_\delta(D_0^{1/2})\}} \mathbf{1}_{\{YY^\top \in N_\delta(B_0)\}} | \Pi^* = \Pi] \\
&= C \cdot \mathbb{E}\left[\mathbf{1}_{\{U \in N_\delta(U_0)\}} \mathbf{1}_{\{R \in N_\delta(D_0^{1/2})\}} \int_{\mathbb{R}^{n\times d}} \mathrm{d}y \mathbf{1}_{\{yy^\top \in N_\delta(B_0)\}} \exp\left(-\frac{\|y - \Pi URQ^\top\|_{\mathrm{F}}^2}{2\sigma^2}\right)\right] \\
&= C \cdot \mathbb{E}\left[\mathbf{1}_{\{U \in N_\delta(U_0)\}} \mathbf{1}_{\{R \in N_\delta(D_0^{1/2})\}} \int_{\mathbb{R}^{n\times d}} \mathrm{d}y \mathbf{1}_{\{yy^\top \in N_\delta(B_0)\}} \exp\left(-\frac{\|y\|_{\mathrm{F}}^2 + \|R\|_{\mathrm{F}}^2}{2\sigma^2}\right) F(y, \Pi UR)\right],
\end{aligned}$$

where $F : \mathbb{R}^{n\times d} \times \mathbb{R}^{n\times d} \to \mathbb{R}_+$ is defined by

$$F(y, x) \triangleq \mathbb{E}_Q\left[\exp\left(\frac{\langle y, xQ^\top\rangle}{\sigma^2}\right)\right] = \int_{O(d)} \mathrm{d}Q \exp\left(\frac{\langle y, xQ^\top\rangle}{\sigma^2}\right).$$

Note that this function is continuous, strictly positive, and right-invariant, in the sense that $F(YO, XO') = F(Y, X)$ for any $O, O' \in O(d)$. Thus, as $\delta \to 0$, we have for some constant $C' = C'(n, d, \sigma)$,

$$\begin{aligned}
&\mathbb{P}[A \in N_\delta(A_0), B \in N_\delta(B_0) | \Pi^* = \Pi] \\
&= (1+o(1)) \underbrace{C' \exp\left(\frac{\mathrm{Tr}(A_0)}{2\sigma^2} - \frac{\mathrm{Tr}(B_0)}{2\sigma^2(\sigma^2+1)}\right)}_{\triangleq h(A_0, B_0)} F(B_0^{1/2}, \Pi A_0^{1/2}) \\
&\quad \underbrace{\cdot\, \mathbb{E}\left[\mathbf{1}_{\{U \in N_\delta(U_0)\}} \mathbf{1}_{\left\{R \in N_\delta(D_0^{1/2})\right\}}\right] \cdot (2\pi(1+\sigma^2))^{-nd/2} \int_{\mathbb{R}^{n\times d}} \mathrm{d}y \mathbf{1}_{\left\{yy^\top \in N_\delta(B_0)\right\}} \exp\left(-\frac{\|y\|_{\mathrm{F}}^2}{2(1+\sigma^2)}\right)}_{\mu[A \in N_\delta(A_0), B \in N_\delta(B_0)]},
\end{aligned}$$

proving (4.10).

4.2.2 Positive Results

Here we briefly describe the proof strategy in the dot product model. Suppose we want to bound the probability that the approximate MLE $\widehat{\Pi}_{\mathrm{AML}}$ in (4.4) makes more than t number of errors. Denote by $\mathrm{d}(\pi_1, \pi_2) \triangleq \sum_{i=1}^n \mathbf{1}_{\{\pi_1(i) \neq \pi_2(i)\}}$ the Hamming distance between two permutations $\pi_1, \pi_2 \in S_n$. Without loss of generality, we will assume that $\pi^* = \mathrm{Id}$. By the orthogonal invariance of $\|\cdot\|_*$, we can assume, *for the sake of analysis*, that $A^{1/2} = X$ and $B^{1/2} = Y$. Applying (4.7),

$$\begin{aligned}
\mathbb{P}\left\{\mathrm{d}(\widehat{\Pi}_{\mathrm{AML}}, \mathrm{Id}) > t\right\} \leq & \mathbb{P}\left\{\max_{\pi: \mathrm{d}(\pi, \mathrm{Id}) > t} \|X^\top \Pi^\top Y\|_* \geq \|X^\top Y\|_*\right\} \\
\leq & \mathbb{P}\left\{\max_{\pi: \mathrm{d}(\pi, \mathrm{Id}) > t} \max_{Q \in O(d)} \langle X^\top \Pi^\top Y, Q\rangle \geq \langle X^\top Y, I_d\rangle\right\}. \qquad (4.11)
\end{aligned}$$

For each fixed Π and Q, averaging over the noise yields, for some absolute constant c_0,

$$\mathbb{P}\left\{\langle X^\top \Pi^\top Y, Q\rangle \geq \langle X^\top Y, I_d\rangle\right\} \leq \mathbb{E} \exp\left\{-\frac{c_0}{\sigma^2} \|X - \Pi X Q\|_{\mathrm{F}}^2\right\}. \qquad (4.12)$$

In the remaining argument, there are three places where the structure of the orthogonal group $O(d)$ plays a crucial role:

1. The quantity in (4.12) turns out to depend on Π through its cycle type and on Q through its eigenvalues. Crucially, the eigenvalues of an orthogonal matrix Q lie on the unit circle, denoted by $(e^{\mathrm{i}\theta_1}, \ldots, e^{\mathrm{i}\theta_d})$, with $|\theta_\ell| \le \pi$. We then show that the error probability in (4.12) can be further bounded by, for some absolute constant C_0,

$$(C_0\sigma)^{d(n-\mathfrak{c})} \left(\prod_{\ell=1}^{d} \frac{C_0\sigma}{\sigma + |\theta_\ell|} \right)^{n_1}, \tag{4.13}$$

where n_1 is the number of fixed points in π and $\mathfrak{c}$ is the total number of cycles.

2. In order to bound (4.11), we take a union bound over π and another union bound over an appropriate discretization of $O(d)$. This turns out to be much subtler than the usual δ-net-based argument, as one needs to implement a localized covering and take into account the local geometry of the orthogonal group. Specifically, note that the error probability in (4.12) becomes larger when π is near Id and when Q is near I_d (i.e. the phases $|\theta_\ell|$'s are small); fortunately, the entropy (namely, the number of such π and such Q within a certain resolution) also becomes smaller, balancing out the deterioration in the probability bound. This is the second place where the structure of $O(d)$ is used crucially, as the local metric entropy of $O(d)$ in the vicinity of I_d is much lower than that elsewhere.

3. Controlling the approximation error of the nuclear norm is another key step. Note that for any matrix norm of the dual form $\|A\| = \sup_{\|Q\|' \le 1} \langle A, Q \rangle$, where $\|\cdot\|'$ is the dual norm of $\|\cdot\|$, the standard δ-net argument (cf. [Ver18a, Lemma 4.4.1]) yields a multiplicative approximation $\max_{Q \in N} \langle A, Q \rangle \ge (1-\delta)\|A\|$, where N is any δ-net of the dual norm ball. In general, this result cannot be improved (e.g. for Frobenius norm); nevertheless, for the special case of nuclear norm, this approximation ratio can be improved from $1-\delta$ to $1-\delta^2$, as the following result of independent interest shows. This improvement turns out to be crucial for obtaining the sharp threshold.

Lemma 4.1. *Let $N \subset O(d)$ be a δ-net in operator norm of the orthogonal group*

$O(d)$. *For any* $A \in \mathbb{R}^{d\times d}$,

$$\max_{Q\in N} \langle A, Q\rangle \geq \left(1 - \frac{\delta^2}{2}\right) \|A\|_*. \tag{4.14}$$

The proof of Theorem 4.1 is completed by combining (4.13) with a union bound over a specific discretization of $O(d)$, whose cardinality satisfies the desired eigenvalue-based local entropy estimate, followed by a union bound over π which can be controlled using moment generating function of the number of cycles in a random derangement.

4.2.3 Negative Results

Here we sketch the main ideas for proving the information-theoretic lower bounds in Theorem 4.3 for the linear assignment model. We first derive a necessary condition for almost perfect recovery that holds for any d via a simple mutual information argument [HWX17]: On one hand, the mutual information $I(\pi^*; X, Y)$ can be upper bounded by the Gaussian channel capacity as $\frac{nd}{2}\log(1+\sigma^{-2})$. On the other hand, to achieve almost perfect recovery, $I(\pi^*; X, Y)$ needs be asymptotically equal to the full entropy $H(\pi^*)$ which is $(1-o(1))\log n$. These two assertions together immediately imply that $\frac{nd}{2}\log(1+\sigma^{-2}) \geq ((1-o(1))\log n$, which further simplifies to $\sigma = n^{-(1-o(1))/d}$ when $d = o(\log n)$. However, for constant d, this necessary condition turns out to be loose and the main bulk of our proof is to improve it to the optimal condition $\sigma = o(n^{-1/d})$. To this end, we follow the program recently developed in [DWXY21] in the context of the planted matching model by analyzing the posterior measure of the latent π^* given the data (X, Y).

To start, a simple yet crucial observation in [DWXY21] is that to prove the impossibility of almost perfect recovery, it suffices to show a random permutation sampled from the posterior distribution is at Hamming distance $\Omega(n)$ away from the ground truth with constant probability. As such, it suffices to show there is more posterior mass over the bad permutations (those far away from the ground truth) than that over the good permutations (those near the ground truth) in the posterior distribution. To proceed, we first bound from above the total posterior mass of good permutations by a truncated first moment calculation applying the large deviation analysis developed in the proof of the positive results. To bound from below the posterior mass of bad permutations, we aim to construct exponentially many bad permutations π whose log likelihood $L(\pi)$ is no smaller than $L(\pi^*)$. A key observation is that $L(\pi) - L(\pi^*)$ can be decomposed according to the orbit decomposition of $(\pi^*)^{-1} \circ \pi$:

$$L(\pi) - L(\pi^*) = \frac{1}{\sigma^2} \langle \Pi X - \Pi^* X, Y \rangle = \frac{1}{\sigma^2} \sum_{O \in \mathcal{O}} \Delta(O), \tag{4.15}$$

where $\mathcal{O}$ denotes the set of orbits in $(\pi^*)^{-1} \circ \pi$ and for any orbit $O = (i_1, i_2, \ldots, i_t)$,

$$\Delta(O) \triangleq \sum_{k=1}^{t} \left\langle X_{\pi^*(i_{k+1})} - X_{\pi^*(i_k)}, Y_{i_k} \right\rangle . \tag{4.16}$$

Thus, the goal is to find a collection of vertex-disjoint orbits O whose total lengths add up to $\Omega(n)$ and each of which is *augmenting* in the sense that $\Delta(O) \geq 0$. Here, a key difference to [DWXY21] is that in the planted matching model with independent edge weights studied there, short augmenting orbits are insufficient to meet the $\Omega(n)$ total length requirement; instead, [DWXY21] resorts to a sophisticated two-stage process that first finds many augmenting paths then connects then into long cycles. Fortunately, for the linear assignment model in low dimensions of $d = \Theta(1)$, as also observed in [KNW22] in their analysis of the MLE, it suffices to look for augmenting 2-orbits and take their disjoint unions. More precisely, we show that there are $\Omega(n)$ many vertex-disjoint augmenting 2-orbits. This has already been done in [KNW22] using a second-moment method enhanced by an additional concentration inequality. It turns out that the correlation among the augmenting 2-orbits is mild enough so that a much simpler argument via a basic second-moment calculation followed by an application of Turán's theorem suffices to extract a large vertex-disjoint subcollection. Finally, these vertex-disjoint augmenting 2-orbits give rise to exponentially many permutations that differ from the ground truth by $\Omega(n)$.

Finally, we briefly remark on perfect recovery, for which it suffices to focus on the MLE (4.2) which minimizes the error probability for uniform π^*. In view of the likelihood decomposition given in (4.15), it further suffices to prove the existence of *an* augmenting 2-orbit. This can be easily done using the second-moment method. A similar strategy was adopted in [DCK19], but our first-moment and second-moment estimates are tighter and hence yield nearly optimal conditions.

4.3 Positive Results: Approximate Maximum Likelihood

4.3.1 Discretization of orthogonal group

We first prove Lemma 4.1 on the approximation of nuclear norm on a discretization of $O(d)$.

Proof of Lemma 4.1. Consider the singular value decomposition $A = UDV^\top$, where $U, V \in O(d)$ and D is diagonal. Then the nuclear norm $\|A\|_* = \max_{Q\in O(d)} \langle A, Q\rangle = \mathrm{Tr}(D)$ is attained at $Q_* = UV^\top$. Pick an element $Q \in N$ with $Q = Q_* + \Delta$, where $\|\Delta\| \le \delta$. By orthogonality of Q and Q_*, we have

$$\Delta Q_*^\top + Q_*\Delta^\top + \Delta\Delta^\top = 0. \tag{4.17}$$

Note that

$$AQ_*^\top = Q_*A^\top = UDU^\top =: B. \tag{4.18}$$

Also, we have

$$\langle A, \Delta\rangle = \left\langle AQ_*^\top, \Delta Q_*^\top\right\rangle, \quad \langle A, \Delta\rangle = \left\langle A^\top, \Delta^\top\right\rangle = \left\langle Q_*A^\top, Q_*\Delta^\top\right\rangle.$$

Adding the above equations and applying (4.17)-(4.18) yield

$$\langle A, \Delta \rangle = \frac{1}{2} \left\langle B, \Delta Q_*^\top + Q_* \Delta^\top \right\rangle = -\frac{1}{2} \left\langle B, \Delta \Delta^\top \right\rangle.$$

This implies

$$|\langle A, \Delta \rangle| \leq \frac{1}{2} \|B\|_* \|\Delta\|^2 = \frac{1}{2} \|A\|_* \|\Delta\|^2,$$

which completes the proof. □

Next we give a specific construction of a δ-net for $O(d)$ that is suitable for the purpose of proving Theorem 4.1. Since orthogonal matrices are normal, by the spectral decomposition theorem, each orthogonal matrix $Q \in O(d)$ can be written as $Q = U^* \Lambda U$, where $\Lambda = \mathsf{diag}(e^{\mathrm{i}\theta_1}, \ldots, e^{\mathrm{i}\theta_d})$ with $\theta_j \in [-\pi, \pi]$ for all $j = 1, \ldots, d$ and $U \in U(d)$ is an unitary matrix. To construct a net for $O(d)$, we first discretize the eigenvalues uniformly and then discretize the eigenvectors according to the optimal local entropy of orthogonal matrices with prescribed eigenvalues.

For any fixed $\delta > 0$, let $\Theta \triangleq \{\theta_k = \frac{k\delta}{4} : k = \lfloor -\frac{4\pi}{\delta} \rfloor, \lfloor -\frac{4\pi}{\delta} \rfloor + 1, \ldots, \lceil \frac{4\pi}{\delta} \rceil \}$. Then the set

$$\mathbf{\Lambda} \triangleq \left\{ (\lambda_1, \ldots, \lambda_d) \in \mathbb{C}^d : \lambda_j = e^{\mathrm{i}\theta_j}, \theta_j \in \Theta, j = 1, \ldots, d \right\}$$

is a $\frac{\delta}{4}$-net in ℓ_∞ norm for the set of all possible spectrum $\{(\lambda_1, \ldots, \lambda_d) \in \mathbb{C}^d : |\lambda_j| = 1\}$. For each $(\lambda_1, \ldots, \lambda_d) \in \mathbb{C}^d$, let $O(\lambda_1, \ldots, \lambda_d)$ denote the set of orthogonal matrices with a prescribed spectrum $\{\lambda_j\}_{j=1}^d$, i.e.

$$O(\lambda_1, \ldots, \lambda_d) \triangleq \{ O \in O(d) : \lambda_i(O) = \lambda_i, i = 1, \ldots, d \},$$

where $\lambda_i(O)$'s are the eigenvalues of O sorted in the counterclockwise way from $-\pi$ to π. Similarly, define $U(\lambda_1, \ldots, \lambda_d)$ to be the set of unitary matrices with a given spectrum

$$U(\lambda_1, \ldots, \lambda_d) \triangleq \{ U^* \mathsf{diag}(\lambda_1, \ldots, \lambda_d) U : U \in U(d) \}.$$

Then $O(\lambda_1, \ldots, \lambda_d) \subset U(\lambda_1, \ldots, \lambda_d) \subset U(d)$. Let $N'(\lambda_1, \ldots, \lambda_d)$ be the optimal $\frac{\delta}{4}$-net in operator norm for $U(\lambda_1, \ldots, \lambda_d)$, and let $N(\lambda_1, \ldots, \lambda_d)$ be the projection (with respect to $\|\cdot\|_{\mathrm{op}}$) of $N'(\lambda_1, \ldots, \lambda_d)$ to $O(d)$. Define

$$N \triangleq \bigcup_{(\lambda_1,\ldots,\lambda_d)\in\mathbf{\Lambda}} N(\lambda_1, \ldots, \lambda_d). \tag{4.19}$$

We claim that N is a δ-net in operator norm for the orthogonal group.

Lemma 4.2. *The set $N \subset O(d)$ defined in (4.19) is a δ-net in operator norm for $O(d)$.*

Proof. Given $Q \in O(d)$, let its eigenvalue decomposition be $Q = U^*\Lambda U$. where $\Lambda = \mathsf{diag}(\lambda_1, \ldots, \lambda_d)$. Then there exists $\widetilde{\Lambda} = \mathsf{diag}(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_d)$ where $(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_d) \in \mathbf{\Lambda}$, such that $\|\Lambda - \widetilde{\Lambda}\| \le \frac{\delta}{4}$. By definition, there exists $\widetilde{U} \in U(d)$ such that $\widetilde{U}^*\widetilde{\Lambda}\widetilde{U} \in N'(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_d)$ and $\|\widetilde{U}^*\widetilde{\Lambda}\widetilde{U} - U^*\widetilde{\Lambda}U\| \le \frac{\delta}{4}$. Let $\widetilde{Q} \in N$ denote the projection of $\widetilde{U}^*\widetilde{\Lambda}\widetilde{U}$. Then

$$\begin{aligned}
\|Q - \widetilde{Q}\| &\le \|Q - \widetilde{U}^*\widetilde{\Lambda}\widetilde{U}\| + \|\widetilde{U}^*\widetilde{\Lambda}\widetilde{U} - \widetilde{Q}\| \\
&\le 2\|\widetilde{U}^*\widetilde{\Lambda}\widetilde{U} - Q\| \\
&= 2\|\widetilde{U}^*\widetilde{\Lambda}\widetilde{U} - U^*\Lambda U\| \\
&\le 2(\|\widetilde{U}^*\widetilde{\Lambda}\widetilde{U} - U^*\widetilde{\Lambda}U\| + \|U^*(\widetilde{\Lambda} - \Lambda)U\|) \le \delta,
\end{aligned}$$

where the second inequality follows from projection. □

The size of this δ-net is estimated in the following lemma.

Lemma 4.3 (Local entropy of $O(d)$)**.** *For each $(\lambda_1, \ldots, \lambda_d)$ where $\lambda_\ell = e^{\mathrm{i}\theta_\ell}$, we have*

$$|N(\lambda_1, \ldots, \lambda_d)| \le \left(1 + \frac{2\max|\theta_\ell|}{\delta}\right)^{2d^2} \tag{4.20}$$

Proof. Note that

$$U(\lambda_1, \ldots, \lambda_d) = I + \{U^*\mathsf{diag}\,(\lambda_1 - 1, \ldots, \lambda_d - 1)\,U : U \in U(d)\} =: I + \widetilde{U}(\lambda_1, \ldots, \lambda_d).$$

For any matrix $Q \in \widetilde{U}(\lambda_1, \ldots, \lambda_d)$, we have

$$\|Q\|_{\text{op}}^2 = \max \left|e^{\mathrm{i}\theta_\ell} - 1\right|^2 = \max |2 - 2\cos\theta_\ell| \le \max |\theta_\ell|^2.$$

where $\|\cdot\|_{\text{op}}$ is the the operator norm with respect to $\mathbb{C}^d \to \mathbb{C}^d$. This implies

$$U(\lambda_1, \ldots, \lambda_d) \subset \mathrm{B}(I, \max |\theta_\ell|),$$

where $\mathrm{B}(I, r)$ is the operator norm ball centered at I_d with radius r. As a normed vector space over $\mathbb{R}$, the space of $d \times d$ complex matrices has dimension $2d^2$ since $\mathbb{C}^{d\times d} \simeq \mathbb{R}^{2d^2}$. Then the desired result follows from a standard volume bound (c.f. e.g. [Pis99, Lemma 4.10]) for the metric entropy

$$|N(\lambda_1, \ldots, \lambda_d)| \le |N'(\lambda_1, \ldots, \lambda_d)| \le \left(1 + \frac{2\max|\theta_\ell|}{\delta}\right)^{2d^2}.$$

□

4.3.2 Moment generating function and cycle decomposition

Based on the reduction (4.49), it suffices to estimate

$$\sum_{\Pi \ne I_n} \sum_{(\lambda_1, \ldots, \lambda_d) \in \mathbf{\Lambda}} \sum_{Q \in N(\lambda_1, \ldots, \lambda_d)} p(\Pi, Q),$$

where

$$p(\Pi, Q) \triangleq \mathbb{E} \exp\left\{-\frac{1}{32\sigma^2} \|X - \Pi X Q\|_{\mathrm{F}}^2\right\}. \tag{4.21}$$

This moment generating function (MGF) is estimated in the following lemma.

Lemma 4.4. *For any fixed $\Pi \in \mathcal{S}_n$, let $\mathcal{O}$ denote the set of orbits of the permutation and n_k be the number of orbits with length k. Let $Q \in O(d)$ and denote by $e^{\mathrm{i}\theta_1}, \ldots, e^{\mathrm{i}\theta_d}$ the* *eigenvalues of* Q*, where* $\theta_1, \ldots, \theta_d \in [-\pi, \pi]$*. Then*

$$p(\Pi, Q) = \prod_{O \in \mathcal{O}, |O| \geq 1} a_{|O|}(Q) = \prod_{k=1}^{n} a_k(Q)^{n_k}, \tag{4.22}$$

where

$$a_k(Q) \triangleq (4\sigma)^{kd} \prod_{\ell=1}^{d} \left[(\sqrt{1+4\sigma^2} + 2\sigma)^{2k} + (\sqrt{1+4\sigma^2} - 2\sigma)^{2k} - 2\cos(k\theta_\ell) \right]^{-1/2}, \tag{4.23}$$

satisfying, for all $1 \leq k \leq n$*,*

$$a_k(Q) \leq a_k(I) \leq (4\sigma)^{(k-1)d}. \tag{4.24}$$

Furthermore,

$$a_1(Q) \leq (C\sigma)^d \prod_{\ell=1}^{d} \frac{1}{\sigma + |\theta_\ell|}, \tag{4.25}$$

where $C > 0$ *is a universal constant independent of* d, n, σ*.*

Proof. For simplicity, denote $t = \frac{1}{32\sigma^2}$. Let $x = \mathsf{vec}(X) \in \mathbb{R}^{nd}$ be the vectorization of X, and note that $x \sim \mathcal{N}(0, I_{nd})$. Through the vectorization, we have

$$\|X - \Pi X Q\|_{\mathrm{F}}^2 = \left\|(I_{nd} - Q^\top \otimes \Pi)x\right\|^2.$$

Let $H \triangleq I_{nd} - Q^\top \otimes \Pi$, then

$$p(\Pi, Q) = \mathbb{E}\exp\left(-t x^\top H^\top H x\right) = \left[\det\left(I + 2tH^\top H\right)\right]^{-\frac{1}{2}}. \tag{4.26}$$

Note that the eigenvalues of H are

$$\lambda_{ij}(H) = 1 - \lambda_i(Q^\top)\lambda_j(\Pi), \quad i = 1, \ldots, d, \ j = 1, \ldots, n.$$

This leads to

$$p(\Pi, Q) = \prod_{i=1}^{d}\prod_{j=1}^{n}\left(1 + 2t\left|1 - \lambda_i(Q^\top)\lambda_j(\Pi)\right|^2\right)^{-\frac{1}{2}}. \tag{4.27}$$

Through a cycle decomposition, the spectrum of Π is the same as a block diagonal matrix $\widetilde{\Pi}$ of the following form

$$\widetilde{\Pi} = \mathsf{diag}\left(P_1^{(1)}, \ldots, P_{n_1}^{(1)}, \ldots, P_1^{(k)}, \ldots, P_{n_k}^{(k)}, \ldots, P_1^{(n)}, \ldots, P_{n_n}^{(n)}\right),$$

where n_k is the number of k-cycles in π, and $P_1^{(k)} = \cdots = P_{n_k}^{(k)} = P^{(k)}$ is a $k \times k$ circulant matrix given by

$$P^{(k)} = \begin{bmatrix} 0 & 1 & \cdots & & 0 \\ 0 & 0 & 1 & & \\ \vdots & 0 & 0 & \ddots & \vdots \\ & & \ddots & \ddots & 1 \\ 1 & & \cdots & 0 & 0 \end{bmatrix}.$$

It is well known that the eigenvalues of $P^{(k)}$ are the k-th roots of unity $\{e^{\mathrm{i}\frac{2\pi}{k}j}\}_{j=0}^{k-1}$. Therefore, the spectrum of Π is the following multiset

$$\mathsf{Spec}(\Pi) = \{e^{\mathrm{i}\frac{2\pi}{k}j_k} \text{ with multiplicity } n_k : 1 \le k \le n, j_k = 0, \ldots, k-1\}. \tag{4.28}$$

Recall that $e^{\mathrm{i}\theta_1}, \ldots, e^{\mathrm{i}\theta_d}$ are the eigenvalues of Q. Note that the eigenvalues of $Q^\top$ are the complex conjugate of the eigenvalues of Q. Combined with (4.27) and (4.28), we have

$$\begin{aligned} p(\Pi, Q) &= \left[\prod_{\ell=1}^{d}\prod_{k=1}^{n}\prod_{j=0}^{k-1}\left(1 + 2t\left|1 - e^{-\mathrm{i}\theta_\ell}e^{\mathrm{i}\frac{2\pi}{k}j}\right|^2\right)^{n_k}\right]^{-1/2} \\ &= \prod_{k=1}^{n}\left[\prod_{\ell=1}^{d}\prod_{j=0}^{k-1}\left(1 + 4t - 4t\cos(-\theta_\ell + \tfrac{2\pi}{k}j)\right)^{-1/2}\right]^{n_k} \triangleq \prod_{k=1}^{n} a_k(Q)^{n_k}. \end{aligned} \tag{4.29}$$

Define

$$f(\theta) \triangleq \prod_{j=0}^{k-1} \left(1 + 4t - 4t\cos(\theta + \tfrac{2\pi}{k}j)\right),$$

To simplify $f(\theta)$, let $p = \frac{\sqrt{1+8t}+1}{2}$ and $q = \frac{\sqrt{1+8t}-1}{2}$ so that $p^2 + q^2 = 1 + 4t$ and $pq = 2t$. Thus,

$$f(\theta) = \prod_{j=0}^{k-1} \left(p^2 + q^2 - 2pq\cos\left(\tfrac{2\pi}{k}j + \theta\right)\right).$$

Note that

$$p^k - q^k e^{\mathrm{i}k\theta} = \prod_{j=0}^{k-1} \left(p - qe^{\mathrm{i}\frac{2\pi}{k}j + \mathrm{i}\theta}\right), \quad p^k - q^k e^{-\mathrm{i}k\theta} = \prod_{j=0}^{k-1} \left(p - qe^{\mathrm{i}\frac{2\pi}{k}j - \mathrm{i}\theta}\right).$$

Multiplying the above two equations gives us

$$p^{2k} + q^{2k} - 2p^k q^k \cos k\theta = \prod_{j=0}^{k-1} \left(p^2 + q^2 - 2pq\cos\left(\tfrac{2\pi}{k}j + \theta\right)\right) = f(\theta).$$

which implies

$$\begin{aligned} f(\theta) &= \left(\frac{\sqrt{1+8t}+1}{2}\right)^{2k} + \left(\frac{\sqrt{1+8t}-1}{2}\right)^{2k} - 2(2t)^k \cos(k\theta) \\ &= \left(\frac{1}{4\sigma}\right)^{2k} \left[\left(\sqrt{1+4\sigma^2}+2\sigma\right)^{2k} + \left(\sqrt{1+4\sigma^2}-2\sigma\right)^{2k} - 2\cos k\theta\right]. \end{aligned}$$

Note that $a_k(Q) = \prod_{\ell=1}^d f(-\theta_\ell)^{-1/2}$, and therefore we have shown (4.23). In particular,

$$a_1(Q) = (4\sigma)^d \prod_{\ell=1}^d (2 - 2\cos\theta_\ell + 16\sigma^2)^{-\frac{1}{2}}. \tag{4.30}$$

Since $\sin^2\theta \geq \frac{\theta^2}{4}$ for $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, we have

$$\begin{aligned} \sqrt{2 - 2\cos\theta_\ell + 16\sigma^2} = \sqrt{4\sin^2(\theta_\ell/2) + 16\sigma^2} &\geq \sqrt{2\sin^2(\theta_\ell/2)} + \sqrt{8\sigma^2} \\ &\geq \sqrt{2(\theta_\ell/4)^2} + \sqrt{8\sigma^2} = \sqrt{2}|\theta_\ell|/4 + 2\sqrt{2}\sigma \end{aligned}$$

Consequently, this gives us (4.25). In general, note that

$$(\sqrt{1+4\sigma^2}+2\sigma)^{2k}+(\sqrt{1+4\sigma^2}-2\sigma)^{2k}-2\geq(4k\sigma)^2,$$

which completes the proof for (4.24). To see this, define $g(x)=x^k-x^{-k}$ which is increasing in x. Then

$$(\sqrt{1+4\sigma^2}+2\sigma)^{2k}+(\sqrt{1+4\sigma^2}-2\sigma)^{2k}-2=g\left(\sqrt{1+4\sigma^2}+2\sigma\right)^2\geq g\left(1+2\sigma\right)^2\geq(4k\sigma)^2,$$

where the last inequality holds because $(1+a)^k-(1-a)^k\geq 2ak$ for $a\geq 0$. Finally, (4.22) follows from (4.29). □

Based on the above representation via cycle decomposition, we have the following estimate for the moment generating function.

Lemma 4.5. *Suppose $d=o(\log n)$. For some $\sigma_0>0$, let $\delta=\sigma_0/\sqrt{n}$ and $N\subset O(d)$ be the δ-net defined in (4.19).*

(i) If $\sigma_0=o(n^{-2/d})$, then

$$\sum_{\Pi\neq I_n}\sum_{Q\in N}\mathbb{E}\exp\left\{-\frac{1}{32\sigma_0^2}\|X-\Pi XQ\|_{\mathrm{F}}^2\right\}=o(1). \tag{4.31}$$

(ii) For any $\varepsilon=\varepsilon(n)>0$, if $\sigma_0^{-d}>16n2^{2/\varepsilon}$, then the following is true

$$\sum_{\mathrm{d}(\pi,\mathrm{Id})\geq\varepsilon n}\sum_{Q\in N}\mathbb{E}\exp\left\{-\frac{1}{32\sigma_0^2}\|X-\Pi XQ\|_{\mathrm{F}}^2\right\}=o(1). \tag{4.32}$$

Proof. (i) For any fixed $\Pi \in \mathcal{S}_n$, combining (4.24) and (4.25) yields

$$\begin{aligned}
\prod_{k\geq 1} a_k(Q)^{n_k} &\leq (C\sigma_0)^{n_1 d+\sum_{k\geq 2} n_k(k-1)d} \left(\prod_{\ell=1}^{d} \frac{1}{\theta_\ell+\sigma_0}\right)^{n_1} \\
&= (C\sigma_0)^{d(n-\sum_{k\geq 2} n_k)} \left(\prod_{\ell=1}^{d} \frac{1}{\theta_\ell+\sigma_0}\right)^{n_1} \\
&\leq (C\sigma_0)^{\frac{n+n_1}{2}d} \left(\prod_{\ell=1}^{d} \frac{1}{\theta_\ell+\sigma_0}\right)^{n_1}. \qquad (4.33)
\end{aligned}$$

Note that by Lemma 4.3, we have

$$\left|N\left(e^{\mathrm{i}\frac{m_1\delta}{4}},\ldots,e^{\mathrm{i}\frac{m_d\delta}{4}}\right)\right| \leq \left(1+\frac{\max|m_\ell|}{2}\right)^{2d^2} \leq \left(1+\frac{\sum_{\ell=1}^{d}|m_\ell|}{2}\right)^{2d^2} \leq \prod_{\ell=1}^{d}\left(1+\frac{|m_\ell|}{2}\right)^{2d^2}. \qquad (4.34)$$

Using Lemma 4.4 and (4.33), this leads to

$$\begin{aligned}
&\sum_{\Pi\neq I_n}\sum_{Q\in N} p(\Pi,Q) \\
&\leq \sum_{n_1=0}^{n-2} \sum_{m_1,\ldots,m_d=\lfloor-\frac{4\pi}{\delta}\rfloor}^{\lceil\frac{4\pi}{\delta}\rceil} \left|N\left(e^{\mathrm{i}\frac{m_1\delta}{4}},\ldots,e^{\mathrm{i}\frac{m_d\delta}{4}}\right)\right|(n-n_1)!\binom{n}{n_1}(C\sigma_0)^{\frac{n+n_1}{2}d}\left(\prod_{\ell=1}^{d}\frac{1}{\frac{\delta|m_\ell|}{4}+\sigma_0}\right)^{n_1} \\
&\leq \sum_{n_1=0}^{n-2}(C\sigma_0)^{\frac{n+n_1}{2}d}(n-n_1)!\binom{n}{n_1}\left[\sum_{m=\lfloor-\frac{4\pi}{\delta}\rfloor}^{\lceil\frac{4\pi}{\delta}\rceil}\frac{1}{(\frac{\delta|m|}{4}+\sigma_0)^{n_1}}(1+\tfrac{|m|}{2})^{2d^2}\right]^d \\
&\leq \sum_{n_1=0}^{n-2}(C\sigma_0)^{\frac{n-n_1}{2}d}(n-n_1)!\binom{n}{n_1}\left[\sum_{m=\lfloor-\frac{4\pi}{\delta}\rfloor}^{\lceil\frac{4\pi}{\delta}\rceil}\frac{1}{(1+\frac{\delta}{4\sigma_0}|m|)^{n_1}}(1+\tfrac{|m|}{2})^{2d^2}\right]^d \\
&\leq \sum_{n_1=0}^{n-2}\left((C\sigma_0)^d n^2\right)^{\frac{n-n_1}{2}}\left[\sum_{m=\lfloor-\frac{4\pi}{\delta}\rfloor}^{\lceil\frac{4\pi}{\delta}\rceil}\frac{1}{(1+\frac{\delta}{4\sigma_0}|m|)^{n_1}}(1+\tfrac{|m|}{2})^{2d^2}\right]^d,
\end{aligned}$$

where the second line follows from Lemma 4.3 and the fourth line follows from the fact that the number of permutations with n_1 fixed points is at most $(n-n_1)!\binom{n}{n_1} \leq n^{n-n_1}$.

Recall that $\delta = \sigma_0/\sqrt{n}$ and $\sigma_0 = o(n^{-2/d})$. For any fixed $1 \le n_1 \le n-2$,

$$T_{n_1} \triangleq \sum_{m=\lfloor -\frac{4\pi}{\delta} \rfloor}^{\lceil \frac{4\pi}{\delta} \rceil} \frac{1}{(1+\frac{\delta}{4\sigma_0}|m|)^{n_1}} (1+\tfrac{|m|}{2})^{2d^2} = \sum_{m=\lfloor -\frac{4\pi}{\delta} \rfloor}^{\lceil \frac{4\pi}{\delta} \rceil} \frac{1}{(1+\frac{|m|}{4\sqrt{n}})^{n_1}} (1+\tfrac{|m|}{2})^{2d^2}.$$

If $n_1 \le \sqrt{n}$, we have

$$T_{n_1} \le \sum_{m=\lfloor -\frac{4\pi}{\delta} \rfloor}^{\lceil \frac{4\pi}{\delta} \rceil} \left(1+\frac{|m|}{2}\right)^{2d^2} \le \frac{8\pi}{\delta}\left(1+\frac{2\pi}{\delta}\right)^{2d^2} \le 2\left(\frac{4\pi\sqrt{n}}{\sigma_0}\right)^{2d^2+1}. \tag{4.35}$$

Therefore, let $\sigma_0^{-d} = L$ and $L = n^2 K$ where $K \gg 1$, then

$$\begin{aligned}
\sum_{n_1=0}^{\sqrt{n}} \left((C\sigma_0)^d n^2\right)^{\frac{n-n_1}{2}} T_{n_1}^d &\le \sqrt{n}\left[2\left(\frac{4\pi\sqrt{n}}{\sigma_0}\right)^{2d^2+1}\right]^d \left((C\sigma_0)^d n^2\right)^{\frac{n-\sqrt{n}}{2}} \\
&\le C^{d^3} n^{2d^3} L^{2d^2+1} K^{-\frac{n-\sqrt{n}}{2}} \\
&\le C^{d^3} n^{2d^3} L^{3d^2} K^{-\frac{n}{3}} \\
&\le C^{d^3} n^{2d^3} \exp\left(3d^2 \log(n^2 K)\right) \exp\left(-\frac{n}{3}\log K\right) \\
&\le C^{d^3} \exp\left((6d^2+2d^3)\log n - \left(\frac{n}{3}-3d^2\right)\log K\right) \\
&= o(1),
\end{aligned} \tag{4.36}$$

where the last line follows from $K \gg 1$ and $d = o(\log n)$.

On the other hand, for $\sqrt{n} \le n_1 \le n-2$,we decompose it into two parts $T_{n_1} = J_1 + J_2$, where

$$\begin{aligned}
J_1 &\triangleq \sum_{|m| \le 8\sqrt{n}} \frac{1}{(1+\frac{|m|}{4\sqrt{n}})^{n_1}} (1+\tfrac{|m|}{2})^{2d^2}, \\
J_2 &\triangleq \sum_{8\sqrt{n} < |m| \le \frac{4\pi}{\delta}} \frac{1}{(1+\frac{|m|}{4\sqrt{n}})^{n_1}} (1+\tfrac{|m|}{2})^{2d^2}.
\end{aligned}$$

We first show that the contribution of J_2 is negligible. To see this, note that

$$
\begin{aligned}
J_2 &\le C(4\sqrt{n})^{n_1} \sum_{m=1+8\sqrt{n}}^{4\pi/\delta} m^{-n_1+2d^2} \\
&\le C(4\sqrt{n})^{n_1} \int_{8\sqrt{n}}^{4\pi/\delta} x^{-(n_1-2d^2)} \mathrm{d}x \\
&\le C(4\sqrt{n})^{n_1} \frac{1}{n_1-2d^2-1} (8\sqrt{n})^{-n_1+2d^2+1} \\
&\le C2^{-n_1+6d^2+3} \frac{1}{n_1-2d^2-1} n^{d^2+\frac{1}{2}} \\
&\le C2^{-n_1/2} n^{d^2}.
\end{aligned}
$$

Recall that $n_1 \ge \sqrt{n}$ and $d = o(\log n)$. Therefore we have $J_2 = o(1)$. Moreover, a simple observation is that $T_{n_1} \ge 1$. This concludes that J_2 is negligible and it suffices to bound J_1. Note that for $0 \le x \le 2$ we have $1 + x \ge e^{x/2}$. Therefore, this implies

$$
J_1 \le C \sum_{m=0}^{8\sqrt{n}} \exp\left(-\left(\frac{n_1}{8\sqrt{n}} - 2d^2\right) m\right).
$$

For $n_1 \ge 32\sqrt{n}(\log n)^2$, we have $\frac{n_1}{8\sqrt{n}} - 2d^2 > \frac{n_1}{16\sqrt{n}}$ since $d = o(\log n)$. Consequently, in this regime we have

$$
J_1 \le C \sum_{m=0}^{8\sqrt{n}} \exp\left(-\frac{n_1}{16\sqrt{n}} m\right) \le \frac{C}{1 - e^{-\frac{n_1}{16\sqrt{n}}}} \le \frac{C}{1 - e^{-4(\log n)^2}}.
$$

Thus, for $n_1 \ge 32\sqrt{n}(\log n)^2$, we have

$$
T_{n_1}^d \le (2J_1)^d \le C^d \left(1 - e^{-4(\log n)^2}\right)^{-d} \le C^d \exp\left(de^{-4(\log n)^2}\right) \le C^d. \tag{4.37}
$$

For $\sqrt{n} \le n_1 < 32\sqrt{n}(\log n)^2$, we use a trivial bound

$$
J_1 \le C \sum_{m=0}^{8\sqrt{n}} \left(1 + \frac{m}{2}\right)^{2d^2} \le C(8\sqrt{n})^{2d^2+1}.
$$

In this case,

$$T_{n_1}^d \le C^d(8\sqrt{n})^{2d^2+1}. \tag{4.38}$$

Thus, (4.37) and (4.38) together imply

$$\begin{aligned}
&\sum_{n_1=\sqrt{n}}^{n-2} \left((C\sigma_0)^d n^2\right)^{\frac{n-n_1}{2}} T_{n_1}^d \\
&\le \sum_{n_1=\sqrt{n}}^{32\sqrt{n}(\log n)^2} \left((C\sigma_0)^d n^2\right)^{\frac{n-n_1}{2}} T_{n_1}^d + \sum_{n_1=32\sqrt{n}(\log n)^2}^{n-2} \left((C\sigma_0)^d n^2\right)^{\frac{n-n_1}{2}} T_{n_1}^d \\
&\le 32\sqrt{n}(\log n)^2 C^{d^2}(8\sqrt{n})^{2d^3+d}2^{-n} + C^d\sigma_0^d n^2 \\
&= o(1)
\end{aligned} \tag{4.39}$$

Combining (4.36) and (4.39) together, we obtain

$$\sum_{\Pi \neq I_n} \sum_{Q \in N} p(\Pi, Q) = o(1),$$

which completes the proof.

(ii) Due to the stronger noise level, we need to be more careful in (4.33):

$$\begin{aligned}
\prod_{j\ge 1} a_k(Q)^{n_j} &\le (C\sigma_0)^{n_1 d + \sum_{j\ge 2} n_j (j-1)d} \left(\prod_{\ell=1}^d \frac{1}{|\theta_\ell| + \sigma_0}\right)^{n_1} \\
&= (C\sigma_0)^{dn - d\sum_{j=1}^n n_j} \prod_{\ell=1}^d \frac{1}{(1+\frac{|\theta_\ell|}{\sigma_0})^{n_1}}.
\end{aligned} \tag{4.40}$$

For simplicity, denote by $k \triangleq \mathrm{d}(\pi, \mathrm{Id}) = n - n_1$ the number of non-fixed points of π. Let $\widetilde{\pi}$ be the restriction of the permutation $\pi \in S_n$ on its non-fixed points, which by definition is a derangement. Denote the number of cycles of a permutation π by $\mathfrak{c}(\pi)$. An observation is that $\mathfrak{c}(\pi) = \sum_{j=1}^{n} n_j = n_1 + \mathfrak{c}(\widetilde{\pi})$. Then Lemma 4.4 and (4.40) yield

$$\begin{aligned}&\sum_{\mathrm{d}(\pi,\mathrm{Id})\geq\varepsilon n}\sum_{Q\in N} p(\Pi,Q)\\ &\leq \sum_{k=\varepsilon n}^{n}\binom{n}{k}\sum_{\widetilde{\pi}\text{ derangement}}\sum_{m_1,\ldots,m_d=\lfloor-\frac{4\pi}{\delta}\rfloor}^{\lceil\frac{4\pi}{\delta}\rceil}\left|N\left(e^{\mathrm{i}\frac{m_1\delta}{4}},\ldots,e^{\mathrm{i}\frac{m_d\delta}{4}}\right)\right|(C\sigma_0)^{d(k-\mathfrak{c}(\widetilde{\pi}))}\prod_{\ell=1}^{d}\frac{1}{(1+\frac{\delta|m_\ell|}{4\sigma_0})^{n-k}}.\end{aligned}$$

Denote $L = \sigma_0^{-d}$. Using (4.34) and rearranging the above inequality give us

$$\begin{aligned}&\sum_{\mathrm{d}(\pi,\mathrm{Id})\geq\varepsilon n}\sum_{Q\in N} p(\Pi,Q)\\ &\quad\leq \sum_{k=\varepsilon n}^{n}\binom{n}{k}L^{-k}\sum_{\widetilde{\pi}\text{ derangement}}L^{\mathfrak{c}(\widetilde{\pi})}\left[\sum_{m=\lfloor-\frac{4\pi}{\delta}\rfloor}^{\lceil\frac{4\pi}{\delta}\rceil}\frac{1}{(1+\frac{\delta}{4\sigma_0}|m|)^{n-k}}(1+\tfrac{|m|}{2})^{2d^2}\right]^{d}. \qquad (4.41)\end{aligned}$$

Note that

$$\sum_{\widetilde{\pi}\text{ derangement}} L^{\mathfrak{c}(\widetilde{\pi})} = k!\,\mathbb{E}_\tau\left[L^{\mathfrak{c}(\tau)}\mathbb{1}_{\{\tau\text{ is a derangement}\}}\right],$$

where the expectation $\mathbb{E}_\tau$ is taken for a uniformly random permutation $\tau \in S_k$. To bound the above truncated generating function, recall that the generating function of $\mathfrak{c}(\tau)$ is given by (see, e.g., [FS09, Eq. (39)])

$$\mathbb{E}_\tau[L^{\mathfrak{c}(\tau)}] = \binom{L+k-1}{k} = \frac{L(L+1)\cdots(L+k-1)}{k!}. \qquad (4.42)$$

Pick some $\alpha \in (0,1)$ to be determined later and obtain the following

$$\begin{aligned}\mathbb{E}_\tau\left[L^{\mathfrak{c}(\tau)}\mathbb{1}_{\{\tau\text{ is a derangement}\}}\right] &\leq \mathbb{E}_\tau\left[L^{\mathfrak{c}(\tau)}\mathbb{1}_{\{\mathfrak{c}(\tau)\leq k/2\}}\right]\\ &\leq \mathbb{E}_\tau\left[L^{\alpha\mathfrak{c}(\tau)+(1-\alpha)\frac{k}{2}}\right] = L^{(1-\alpha)\frac{k}{2}}\mathbb{E}_\tau\left[L^{\alpha\mathfrak{c}(\tau)}\right] = L^{(1-\alpha)\frac{k}{2}}\binom{L^\alpha+k-1}{k}.\end{aligned}$$

Choosing $\alpha = \frac{\log k}{\log L}$, we have

$$\mathbb{E}_\tau \left[L^{\mathfrak{c}(\tau)} \mathbb{1}_{\{\tau \text{ is a derangement}\}} \right] \le \binom{2k-1}{k} \left(\frac{L}{k} \right)^{k/2} \le \left(\frac{16L}{k} \right)^{k/2}. \tag{4.43}$$

Recall that

$$T_{n-k} = \sum_{m=\lfloor -\frac{4\pi}{\delta} \rfloor}^{\lceil \frac{4\pi}{\delta} \rceil} \frac{1}{(1+\frac{\delta}{4\sigma_0}|m|)^{n-k}} (1+\tfrac{|m|}{2})^{2d^2}.$$

For $k \le n - \sqrt{n}$, each term T_{n-k} is bounded by (4.37) and (4.38). On the other hand, if $k \ge n - \sqrt{n}$, we control T_{n-k} via (4.35). Here in the case of almost perfect recovery, combined with (4.43), the assumption on σ_0 yields a superexponentially decaying term in the summation (4.41). Specifically, combined this with (4.41) and (4.43), we obtain

$$\sum_{\mathrm{d}(\pi, \mathrm{Id}) \ge \varepsilon n} \sum_{Q \in N} p(\Pi, Q) \le J_1 + J_2,$$

where

$$J_1 \triangleq C^d \sum_{k=\varepsilon n}^{n-32\sqrt{n}(\log n)^2} \binom{n}{k} L^{-k} k! \left(\frac{16L}{k} \right)^{k/2},$$

$$J_2 \triangleq C^{d^3} n^{2d^3} L^{2d^2+1} \sum_{k=n-32\sqrt{n}(\log n)^2+1}^{n} \binom{n}{k} L^{-k} k! \left(\frac{16L}{k} \right)^{k/2}.$$

Let $L = nK$ where $\frac{\varepsilon}{2} \log \frac{K}{16} > \log 2$. Recall that $d = o(\log n)$. Then applying Stirling's approximation gives us

$$J_1 \le C^d n 2^n \left(\frac{16n}{L} \right)^{\varepsilon n/2} \le C^d n \exp\left(n \log 2 - \frac{\varepsilon n}{2} \log \frac{K}{16} \right) = o(1), \tag{4.44}$$

and

$$J_2 \leq C^{d^3} n^{2d^3+1} L^{2d^2+1} 2^n \left(\frac{16n}{L}\right)^{n/3}$$
$$\leq C^{d^3} n^{2d^3+1} \exp\left[(2d^2+1)\log n + (2d^2+1)\log K + n\log 2 - \frac{n}{3}\log\frac{K}{16}\right] = o(1). \tag{4.45}$$

Combining (4.44) and (4.45) implies

$$\sum_{\mathrm{d}(\pi,\mathrm{Id})\geq \varepsilon n} \sum_{Q\in N} p(\Pi, Q) = o(1),$$

which completes the proof. □

The estimate of the moment generating functions results in the following lemma, which plays a crucial rule in the probability reduction estimate (4.49).

Lemma 4.6. *For some $\sigma_0 > 0$, let $\delta = \sigma_0/\sqrt{n}$ and N be the δ-net defined in (4.19).*

(i) If $\sigma_0 = o(n^{-2/d})$, for any constant $c > 0$, the following inequality is true with high probability

$$\min_{\Pi\neq I_n} \min_{Q\in N} \|X - \Pi X Q\|_{\mathrm{F}} \geq c\sqrt{d}\sigma_0. \tag{4.46}$$

(ii) For any $\varepsilon = \varepsilon(n) > 0$, if $\sigma_0^{-d} > 16n2^{2/\varepsilon}$, the following is true for any fixed constant $c > 0$ with high probability

$$\min_{\mathrm{d}(\pi,\mathrm{Id})\geq \varepsilon n} \min_{Q\in N} \|X - \Pi X Q\|_{\mathrm{F}} \geq c\sqrt{d}\sigma_0. \tag{4.47}$$

Proof. (i) For fixed $\Pi \neq I_n$ and $Q \in N$, by the Chernoff bound, for every $t \geq 0$ we have

$$\mathbb{P}\left\{\|X - \Pi X Q\|_{\mathrm{F}} < c\sqrt{d}\sigma_0\right\}$$
$$= \mathbb{P}\left\{e^{-t\|X-\Pi XQ\|_{\mathrm{F}}^2} > e^{-tc^2 d\sigma_0^2}\right\} \leq e^{tc^2 d\sigma_0^2}\mathbb{E}\exp\left(-t\|X - \Pi X Q\|_{\mathrm{F}}^2\right).$$

Taking $t = \frac{1}{32\sigma_0^2}$, by the union bound we have

$$\begin{aligned}\mathbb{P}\left\{\min_{\Pi\neq I_n}\min_{Q\in N}\|X-\Pi XQ\|_{\mathrm{F}}\geq c\sqrt{d}\sigma_0\right\} &= 1-\mathbb{P}\left\{\exists\Pi\neq I_n, \exists Q\in N \ s.t.\ \|X-\Pi XQ\|_{\mathrm{F}}\leq c\sigma_0\right\}\\ &\geq 1-e^{\frac{c^2d}{32}}\sum_{\Pi\neq I_d}\sum_{Q\in N}\mathbb{E}\exp\left\{-\frac{1}{32\sigma_0^2}\|X-\Pi XQ\|_{\mathrm{F}}^2\right\}\\ &\geq 1-o(1),\end{aligned}$$

where the last step follows from Lemma 4.5.

(ii) The arguments are similar with Part (i). Using Chernoff bound and Lemma 4.5, we have

$$\begin{aligned}&\mathbb{P}\left\{\min_{\mathrm{d}(\pi,\mathrm{Id})\geq\varepsilon n}\min_{Q\in N}\|X-\Pi XQ\|_{\mathrm{F}}\geq c\sqrt{d}\sigma_0\right\}\\ &\geq 1-e^{\frac{c^2d}{32}}\sum_{\mathrm{d}(\pi,\mathrm{Id})\geq\varepsilon n}\sum_{Q\in N}\mathbb{E}\exp\left\{-\frac{1}{32\sigma_0^2}\|X-\Pi XQ\|_{\mathrm{F}}^2\right\}\\ &\geq 1-o(1),\end{aligned}$$

which completes the proof. □

4.3.3 Consistency of Approximate MLE

Now we prove Theorem 4.1.

Perfect Recovery For $\sigma \ll n^{-2/d}$, let $\delta = \sigma/\sqrt{n}$ and let N be the δ-net in operator norm for $O(d)$ defined in (4.19). Applying Lemma 4.1, we have

$$\begin{aligned}\mathbb{P}\left\{\|X^\top\Pi^\top Y\|_*\geq\|X^\top Y\|_*\right\} &\leq \mathbb{P}\left\{\max_{Q\in O(d)}\langle X^\top\Pi^\top Y, Q\rangle\geq\langle X^\top Y, I_d\rangle\right\}\\ &\leq \mathbb{P}\left\{\max_{Q\in N}\langle X^\top\Pi^\top Y, Q\rangle\geq(1-\delta^2)\langle X^\top Y, I_d\rangle\right\}.\end{aligned}$$

For fixed Π and Q, we have

$$\mathbb{P}\left\{\langle X^\top \Pi^\top Y, Q\rangle \geq (1-\delta^2)\langle X^\top Y, I_d\rangle\right\}$$
$$= \mathbb{P}\left\{\sigma\langle Z, (1-\delta^2)X - \Pi X Q\rangle \geq (1-\delta^2)\|X\|_{\mathrm{F}}^2 - \langle X, \Pi X Q\rangle\right\}.$$

Note that we have the following observations

$$\|X\|_{\mathrm{F}}^2 - \langle X, \Pi X Q\rangle = \frac{1}{2}\|X - \Pi X Q\|_{\mathrm{F}}^2,$$

and

$$\begin{aligned}\left\|(1-\delta^2)X - \Pi X Q\right\|_{\mathrm{F}}^2 &= (1-\delta^2)^2\|X\|_{\mathrm{F}}^2 + \|X\|_{\mathrm{F}}^2 - 2(1-\delta^2)\langle X, \Pi X Q\rangle \\ &= (1-\delta^2)\|X - \Pi X Q\|_{\mathrm{F}}^2 - \delta^4\|X\|_{\mathrm{F}}^2.\end{aligned}$$

Therefore,

$$\begin{aligned}&\mathbb{P}\left\{\langle X^\top \Pi^\top Y, Q\rangle \geq (1-\delta^2)\langle X^\top Y, I_d\rangle\right\} \\ =&\mathbb{P}\left\{\sigma\mathcal{N}\left(0, (1-\delta^2)\|X - \Pi X Q\|_{\mathrm{F}}^2 - \delta^4\|X\|_{\mathrm{F}}^2\right) \geq \frac{1}{2}\|X - \Pi X Q\|_{\mathrm{F}}^2 - \delta^2\|X\|_{\mathrm{F}}^2\right\} \\ \leq&\mathbb{P}\left\{\sigma\mathcal{N}\left(0, \|X - \Pi X Q\|_{\mathrm{F}}^2\right) \geq \frac{1}{2}\|X - \Pi X Q\|_{\mathrm{F}}^2 - \delta^2\|X\|_{\mathrm{F}}^2\right\}.\end{aligned} \tag{4.48}$$

Consider the following events

$$\mathcal{E}_1 \triangleq \left\{cdn \leq \|X\|_{\mathrm{F}}^2 \leq Cdn\right\}, \quad \mathcal{E}_2 \triangleq \left\{\min_{\Pi\neq I}\min_{Q\in N}\|X - \Pi X Q\|_{\mathrm{F}} \geq C\sqrt{d}\sigma\right\}.$$

It is well known that $\mathbb{P}\{\mathcal{E}_1\} = 1 - o(1)$, and by Lemma 4.6 we also have $\mathbb{P}\{\mathcal{E}_2\} = 1 - o(1)$.

On the events $\mathcal{E}_1$ and $\mathcal{E}_2$, the previous estimate (4.48) for $\Pi \neq I$ reduces to

$$\begin{aligned}
&\mathbb{P}\left\{\langle X^\top \Pi^\top Y, Q\rangle \geq (1-\delta^2)\langle X^\top Y, I_d\rangle, \mathcal{E}_1, \mathcal{E}_2\right\} \\
&\leq \mathbb{P}\left\{\sigma\mathcal{N}\left(0, \|X - \Pi X Q\|_{\mathrm{F}}^2\right) \geq \frac{1}{4}\|X - \Pi X Q\|_{\mathrm{F}}^2\right\} \leq \mathbb{E}\exp\left\{-\frac{1}{32\sigma^2}\|X - \Pi X Q\|_{\mathrm{F}}^2\right\}.
\end{aligned} \tag{4.49}$$

By Lemma 4.5, the reduction (4.49) and a union bound, we have

$$\begin{aligned}
&\mathbb{P}\left\{\max_{\Pi \neq I}\|X^\top \Pi^\top Y\|_* \geq \|X^\top Y\|_*\right\} \\
&\leq \mathbb{P}\left\{\max_{\Pi \neq I}\|X^\top \Pi^\top Y\|_* \geq \|X^\top Y\|_*, \mathcal{E}_1, \mathcal{E}_2\right\} + \mathbb{P}\left\{\mathcal{E}_1^c\right\} + \mathbb{P}\left\{\mathcal{E}_2^c\right\} \\
&\leq \mathbb{P}\left\{\max_{\Pi \neq I_n}\max_{Q \in N}\langle X^\top \Pi^\top Y, Q\rangle \geq (1-\delta^2)\langle X^\top Y, I_d\rangle, \mathcal{E}_1, \mathcal{E}_2\right\} + o(1) \\
&\leq \sum_{\Pi \neq I_n}\sum_{Q \in N}\mathbb{P}\left\{\langle X^\top \Pi^\top Y, Q\rangle \geq (1-\delta^2)\langle X^\top Y, I_d\rangle, \mathcal{E}_1, \mathcal{E}_2\right\} + o(1) \\
&\leq \sum_{\Pi \neq I_n}\sum_{Q \in N}\mathbb{E}\exp\left\{-\frac{1}{32\sigma^2}\|X - \Pi X Q\|_{\mathrm{F}}^2\right\} + o(1) \\
&= o(1).
\end{aligned}$$

This implies that the ground truth $\Pi^* = I_n$ is the approximate MLE with probability $1 - o(1)$, i.e.,

$$\mathbb{P}\left\{\mathrm{argmax}_{\Pi \in S_n}\|X^\top \Pi^\top Y\|_* = I_n\right\} = 1 - o(1),$$

which shows the success of perfect recovery with high probability.

Almost Perfect Recovery The arguments are essentially the same. For a sufficiently small $\varepsilon = \varepsilon(n) > 0$, take $\sigma^{-d} > 16n2^{2/\varepsilon}$ and consider the event

$$\mathcal{E}_2' \triangleq \left\{\min_{\mathrm{d}(\pi, \mathrm{Id}) \geq \varepsilon n}\min_{Q \in N}\|X - \Pi X Q\|_{\mathrm{F}} \geq C\sqrt{d}\sigma\right\}.$$

Then Lemma 4.6 implies $\mathbb{P}\{\mathcal{E}_2'\} = 1 - o(1)$. On the event $\mathcal{E}_1$ and $\mathcal{E}_2'$, the reduction estimate for Π with $\mathrm{d}(\pi, \mathrm{Id}) \geq \varepsilon n$ still holds

$$\mathbb{P}\left\{\langle X^\top \Pi^\top Y, Q\rangle \geq (1-\delta^2)\langle X^\top Y, I_d\rangle, \mathcal{E}_1, \mathcal{E}_2'\right\} \leq \mathbb{E}\exp\left\{-\frac{1}{32\sigma^2}\|X - \Pi X Q\|_{\mathrm{F}}^2\right\}.$$

Combining this with Lemma 4.5, we have

$$\begin{aligned}\mathbb{P}&\left\{\max_{\mathrm{d}(\pi,\mathrm{Id})\geq\varepsilon n}\|X^\top\Pi^\top Y\|_* \geq \|X^\top Y\|_*\right\}\\ &\qquad\qquad \leq \sum_{\mathrm{d}(\pi,\mathrm{Id})\geq\varepsilon n}\sum_{Q\in N}\mathbb{E}\exp\left\{-\frac{1}{32\sigma^2}\|X-\Pi XQ\|_{\mathrm{F}}^2\right\} + o(1) = o(1).\end{aligned}$$

Thus,

$$\mathbb{P}\left\{\mathsf{overlap}(\widehat{\pi}_{\mathrm{AML}}, \pi^*) \geq 1 - \varepsilon\right\} = 1 - o(1).$$

Taking $\sigma \ll n^{-1/d}$ so that $\epsilon = o(1)$, this implies the desired (4.6).

4.4 Negative Results: Analysis of Posterior Sampling

In this section, we derive necessary conditions for both almost perfect recovery and perfect recovery for the linear assignment model (4.1). These conditions also hold for the weaker dot-product and distance models.

As a warm-up, we first derive a necessary condition for almost perfect recovery that holds for any d via a simple mutual information argument. Then we focus on the special case where d is a constant and give a much sharper analysis, improving the necessary condition from $\sigma \leq n^{-(1-o(1))/d}$ to $\sigma = o(n^{-1/d})$. Note that achieving a vanishing recovery error in expectation is equivalent to that with high probability (see e.g. [HWX17, Appendix A]). Thus without loss of generality, we focus on the expected number of errors $\mathbb{E}\mathrm{d}(\pi^*, \widehat{\pi})$ in this subsection.

Proposition 4.1. *For any $\epsilon \in (0,1)$, if there exists an estimator $\widehat{\pi} \equiv \widehat{\pi}(X, Y)$ such that*

$\mathbb{E}\mathrm{d}(\pi^*, \widehat{\pi}) \le \epsilon n$*, then*

$$\frac{d}{2}\log\left(1+\frac{1}{\sigma^2}\right) - (1-\epsilon)\log n + 1 + \frac{\log(n+1)}{n} \ge 0. \tag{4.50}$$

Proof. Since $\pi^* \to (X, Y) \to \widehat{\pi}$ form a Markov chain, by the data processing inequality of mutual information, we have

$$I(\pi^*; X, Y) \ge I(\pi^*; \widehat{\pi}) = H(\pi^*) - H(\pi^*|\widehat{\pi}). \tag{4.51}$$

On the one hand, note that $H(\pi^*) = \log(n!) \ge n\log n - n$. Moreover, for any fixed realization of $\widehat{\pi}$, the number of π^* such that $\mathrm{d}(\pi^*, \widehat{\pi}) = \ell$ is $\binom{n}{\ell}!\ell \le n^\ell$, where $!\ell$ denotes the number of derangements of ℓ elements, given by

$$!\ell = \ell! \sum_{i=0}^{\ell} \frac{(-1)^i}{i!} = \left[\frac{\ell!}{e}\right],$$

and $[\cdot]$ denotes rounding to the nearest integer. Therefore,

$$H(\pi^*|\widehat{\pi}, \mathrm{d}(\pi^*, \widehat{\pi})) \le \mathbb{E}\mathrm{d}(\pi^*, \widehat{\pi}) \log n \le \epsilon n \log n.$$

Furthermore, $\mathrm{d}(\pi^*, \widehat{\pi})$ takes values in $\{0, 1, \ldots, n\}$. Thus from the chain rule,

$$H(\pi^*|\widehat{\pi}) = H(\mathrm{d}(\pi^*, \widehat{\pi})|\widehat{\pi}) + H(\pi^*|\widehat{\pi}, \mathrm{d}(\pi^*, \widehat{\pi})) \le \log(n+1) + \epsilon n \log n. \tag{4.52}$$

On the other hand, the information provided by the observation (X, Y) about π^* satisfies

$$\begin{aligned} I(\pi^*; X, Y) &= I(\Pi^* X; \Pi^* X + \sigma Z | X) \\ &\overset{\text{(a)}}{\le} \frac{nd}{2}\log\left(1 + \frac{\mathbb{E}[\|X\|^2]}{nd\sigma^2}\right) \\ &= \frac{nd}{2}\log\left(1+\frac{1}{\sigma^2}\right), \end{aligned} \tag{4.53}$$

where (a) follows from the Gaussian channel capacity formula and the fact that the mutual information in the Gaussian channel under a second moment constraint is maximized by the Gaussian input distribution. Combining (4.51)–(4.53), we get that

$$\frac{nd}{2}\log\left(1+\frac{1}{\sigma^2}\right) \geq (1-\epsilon)\, n\log n - n - \log(n+1),$$

arriving at the desired necessary condition (4.50). □

The necessary condition (4.50) further specializes to:

- $d = o(\log n)$:

$$\sigma = O\left(n^{-(1-\epsilon)/d}\right). \tag{4.54}$$

  This yields Theorem 4.3(ii) and resolves [KNW22, Conjecture 1.4, item 1] in the positive;

- $d = \Theta(\log n)$:

$$\sigma \leq \frac{1-\epsilon+o(1)}{\sqrt{n^{2/d}-1}};$$

- $d = \omega(\log n)$:

$$\sigma \leq \sqrt{\frac{d}{2(1-\epsilon-o(1))\log n}}.$$

  In this case, this necessary condition matches the sufficient condition of almost perfect recovery in [DCK20, Theorem 1] and [KNW22, Section A.2] up to $1+o(1)$ factor, thereby determining the sharp information-theoretic limit for the linear assignment model in high dimensions.

While the negative result in Proposition 4.1 holds for any d, the necessary condition (4.50) turns out to be loose for bounded d.

4.4.1 Analysis of Posterior Distribution

To prove the optimal information-theoretic limits, we follow the program in [DWXY21] of analyzing the posterior distribution. The likelihood function of (X, Y) given $\Pi^* = \Pi$ is proportional to $\exp(-\frac{1}{2\sigma^2}\|Y - \Pi X\|_{\mathrm{F}}^2)$. Therefore, conditional on (X, Y), the posterior distribution of Π^* is a Gibbs measure, given by

$$\mu_{X,Y}(\Pi) = \frac{1}{Z(X,Y)} \exp\left(L(\Pi)\right), \qquad \text{where } L(\Pi) = \frac{1}{\sigma^2}\left\langle \Pi X, Y \right\rangle,$$

and $Z(X, Y)$ is the normalization factor.

As observed in [DWXY21, Section 3.1], in order to prove the impossibility of almost perfect recovery, it suffices to consider the estimator $\widetilde{\Pi}$ which is sampled from the posterior distribution $\mu_{X,Y}(\Pi)$. To see this, given any estimator $\widehat{\Pi} \equiv \widehat{\Pi}(X, Y)$, $(\widehat{\Pi}, \Pi^*)$ and $(\widehat{\Pi}, \widetilde{\Pi})$ are equal in law, and hence

$$\mathbb{E}[\mathrm{d}(\widetilde{\Pi}, \Pi^*)] \le \mathbb{E}[\mathrm{d}(\widetilde{\Pi}, \widehat{\Pi})] + \mathbb{E}[\mathrm{d}(\Pi^*, \widehat{\Pi})] = 2\mathbb{E}[\mathrm{d}(\Pi^*, \widehat{\Pi})],$$

which shows that $\widetilde{\Pi}$ is optimal within a factor of two. Thus it suffices to bound $\mathbb{E}[\mathrm{d}(\widetilde{\Pi}, \Pi^*)]$ from below.

To this end, fix some δ to be specified later and define the sets of good and bad solutions respectively as

$$\begin{aligned} \mathbf{\Pi}_{\mathsf{good}} &= \{\Pi \in \mathcal{S}_n : \mathrm{d}(\Pi, \Pi^*) < \delta n\}, \\ \mathbf{\Pi}_{\mathsf{bad}} &= \{\Pi \in \mathcal{S}_n : \mathrm{d}(\Pi, \Pi^*) \ge \delta n\}. \end{aligned}$$

By the definition of $\widetilde{\Pi}$, we have

$$\mathbb{E}[\mathrm{d}(\widetilde{\Pi}, \Pi^*)] \ge \delta n \cdot \mathbb{E}[\mu_{X,Y}(\mathbf{\Pi}_{\mathsf{bad}})].$$

Next we show two key lemmas, which bound the posterior mass of $\mathbf{\Pi}_{\mathsf{good}}$ and $\mathbf{\Pi}_{\mathsf{good}}$ from above and below, respectively.

Lemma 4.7. *Assume $\sigma = \sigma_0 n^{-1/d}$ for any constant $\sigma_0 \in (0, 1/2)$. For any constant δ such that $\delta \leq 16(2\sigma_0)^d$, with probability at least $1 - 4\delta n e^{-\delta n/\log n}$,*

$$\frac{\mu_{X,Y}(\mathbf{\Pi}_{\mathsf{good}})}{\mu_{X,Y}(\Pi^*)} \leq 2\left(\frac{16e^2(2\sigma_0)^d}{\delta}\right)^{\delta n}. \tag{4.55}$$

Lemma 4.8. *Assume $\sigma = \sigma_0 n^{-1/d}$ for some constant σ_0. There exist constants $\delta_0(\sigma_0, d)$ and $c(\sigma_0, d)$ that only depend on σ_0, d such that for all $\delta \leq \delta_0$ and sufficiently large n, with probability at least $1/2 - c/n$,*

$$\frac{\mu_{X,Y}(\mathbf{\Pi}_{\mathsf{bad}})}{\mu_{X,Y}(\Pi^*)} \geq e^{\delta_0 n/2}. \tag{4.56}$$

We prove Lemma 4.7 by a truncated first moment calculation. To do this, we need the following key auxiliary result.

Lemma 4.9. *Assume that $n(2\sigma)^d \leq 1$. Then for any $\ell \in [0, n]$,*

$$\sum_{\Pi : \mathsf{d}(\Pi, \Pi^*) = \ell} \mathbb{E}\exp\left(-\frac{1}{8\sigma^2}\|\Pi X - \Pi^* X\|_{\mathrm{F}}^2\right) \leq \left(\frac{16n^2(2\sigma)^d}{\ell}\right)^{\ell/2}.$$

Proof. It follows from (4.24) in Lemma 4.4 that

$$\mathbb{E}\exp\left(-\frac{1}{8\sigma^2}\|\Pi X - \Pi^* X\|_{\mathrm{F}}^2\right) \leq \prod_{k=1}^{n}\left[(2\sigma)^{k-1}\right]^{dn_k} \leq (2\sigma)^{d(\ell - \mathsf{c}(\widetilde{\pi}))},$$

where $\ell = n - n_1$ is the number of non-fixed points, $\widetilde{\pi}$ is the restriction of the permutation

π on its non-fixed points, and $\mathfrak{c}(\widetilde{\pi})$ denotes the number of cycles of $\widetilde{\pi}$. It follows that

$$\begin{aligned}\sum_{\Pi:\mathrm{d}(\Pi,\Pi^*)=\ell} \mathbb{E}\exp\left(-\frac{1}{8\sigma^2}\|\Pi X - \Pi^* X\|_{\mathrm{F}}^2\right) &\leq \binom{n}{\ell}\frac{\ell!}{L^\ell}\mathbb{E}_\tau\left[L^{\mathfrak{c}(\tau)}\mathbb{1}_{\{\tau \text{ is a derangement}\}}\right] \\ &\leq \left(\frac{n}{L}\right)^\ell \mathbb{E}_\tau\left[L^{\mathfrak{c}(\tau)}\mathbb{1}_{\{\tau \text{ is a derangement}\}}\right] \\ &\leq \left(\frac{16n^2}{\ell L}\right)^{\ell/2},\end{aligned}$$

where $L = (2\sigma)^{-d}$, the expectation $\mathbb{E}_\tau$ is taken for a uniformly random permutation $\tau \in S_\ell$, and the last inequality follows from (4.43). □

Proof of Lemma 4.7. Note that

$$\frac{\mu_{X,Y}(\mathbf{\Pi}_{\mathsf{good}})}{\mu_{X,Y}(\Pi^*)} = \sum_{\Pi \in \mathbf{\Pi}_{\mathsf{good}}} e^{L(\Pi)-L(\Pi^*)} = R_1 + R_2,$$

where

$$\begin{aligned} R_1 &\triangleq \sum_{\Pi:\mathrm{d}(\Pi,\Pi^*)<\beta n/\log n} e^{L(\Pi)-L(\Pi^*)} \\ R_2 &\triangleq \sum_{\Pi:\frac{\beta n}{\log n}\leq \mathrm{d}(\Pi,\Pi^*)<\delta n} e^{L(\Pi)-L(\Pi^*)}\end{aligned}$$

for some β to be specified. Next we bound R_1 and R_2 separately.

First, the number of permutations Π such that $\Pi^{-1}\circ\Pi^*$ has ℓ non-fixed points is

$$|\{\Pi \in \mathcal{S}_n : \mathrm{d}(\Pi,\Pi^*) = \ell\}| = !\ell \cdot \binom{n}{\ell}, \tag{4.57}$$

where $!\ell = \left[\frac{\ell!}{e}\right]$. Thus

$$\frac{1}{2e}n(n-1)\cdots(n-\ell+1) \leq |\{\Pi \in \mathcal{S}_n : \mathrm{d}(\Pi,\Pi^*) = \ell\}| \leq \frac{2}{e}n(n-1)\cdots(n-\ell+1). \tag{4.58}$$

Furthermore, for any Π,

$$\begin{aligned}
\mathbb{E}e^{L(\Pi)-L(\Pi^*)} &= \mathbb{E}\mathrm{exp}\left(\frac{1}{\sigma^2}\langle \Pi X - \Pi^* X, Y\rangle\right) \\
&= \mathbb{E}\mathrm{exp}\left(\frac{1}{\sigma^2}\langle \Pi X - \Pi^* X, \Pi^* X\rangle + \frac{1}{2\sigma^2}\|\Pi X - \Pi^* X\|_{\mathrm{F}}^2\right) \\
&= 1, \qquad (4.59)
\end{aligned}$$

where the first equality holds due to $Y = \Pi^* X + \sigma^2 Z$ and $\mathbb{E}\mathrm{exp}(\langle A, Z\rangle) = \exp(\|A\|_{\mathrm{F}}^2/2)$ and the second equality follows from $\langle \Pi X - \Pi^* X, \Pi^* X\rangle = -\frac{1}{2}\|\Pi X - \Pi^* X\|_{\mathrm{F}}^2$.

To bound R_1, using (4.58) and (4.59) we have

$$\mathbb{E}R_1 = \sum_{\mathrm{d}(\Pi,\Pi^*)<\frac{\beta n}{\log n}} \mathbb{E}e^{L(\Pi)-L(\Pi^*)} \le \sum_{\ell<\frac{\beta n}{\log n}} \frac{2}{e}n^\ell \le \frac{2\beta n}{e\log n}\exp(\beta n).$$

By Markov's inequality,

$$\mathbb{P}\left\{R_1 \ge e^{2\beta n}\right\} \le \frac{2n}{e}\exp(-\beta n). \qquad (4.60)$$

To bound R_2, the calculation above shows that directly applying the Markov inequality is too crude since $\mathbb{E}[R_2] = e^{\Theta(n\log n)}$. Note that although $L(\Pi) - L(\Pi^*)$ is negatively biased, when $L(\Pi) - L(\Pi^*)$ is atypically large it results in an excessive contribution to the exponential moments. Thus we truncate on the following event:

$$\mathcal{E} \triangleq \bigcap_{\Pi:\frac{\beta n}{\log n}\le \mathrm{d}(\Pi,\Pi^*)<\delta n} \left\{L(\Pi) - L(\Pi^*) \le \tau\left(\mathrm{d}(\Pi,\Pi^*)\right)\right\}$$

for some threshold $\tau(\ell)$ to be chosen.

Then for any $c' > 0$,

$$
\begin{aligned}
&\mathbb{P}\left\{R_2 \geq e^{c'n}\right\} \\
&\leq \mathbb{P}\left\{\mathcal{E}^c\right\} + \mathbb{P}\left\{\{R_2 \geq e^{c'n}\} \cap \mathcal{E}\right\} \\
&\leq \mathbb{P}\left\{\mathcal{E}^c\right\} + \mathbb{P}\left\{\sum_{\frac{\beta n}{\log n} \leq \mathrm{d}(\Pi,\Pi^*) < \delta n} e^{L(\Pi)-L(\Pi^*)} \mathbf{1}_{\{L(\Pi)-L(\Pi^*) \leq \tau(\mathrm{d}(\Pi,\Pi^*))\}} \geq e^{c'n}\right\} \\
&\leq \mathbb{P}\left\{\mathcal{E}^c\right\} + e^{-c'n} \sum_{\frac{\beta n}{\log n} \leq \mathrm{d}(\Pi,\Pi^*) < \delta n} \mathbb{E} e^{L(\Pi)-L(\Pi^*)} \mathbf{1}_{\{L(\Pi)-L(\Pi^*) \leq \tau(\mathrm{d}(\Pi,\Pi^*))\}}. \qquad (4.61)
\end{aligned}
$$

To bound the first term, note that for any $t > 0$,

$$
\begin{aligned}
&\mathbb{P}\left\{L(\Pi) - L(\Pi^*) \geq \tau\right\} \\
&\qquad \leq e^{-t\tau} \mathbb{E}\exp\left(\frac{t}{\sigma^2}\left\langle \Pi X - \Pi^* X, Y\right\rangle\right) = e^{-t\tau} \mathbb{E}\exp\left(\frac{t^2 - t}{2\sigma^2}\|\Pi X - \Pi^* X\|_{\mathrm{F}}^2\right).
\end{aligned}
$$

By choosing $t = 1/2$, we get that

$$
\mathbb{P}\left\{L(\Pi) - L(\Pi^*) \geq \tau\right\} \leq e^{-\tau/2} \mathbb{E}\exp\left(-\frac{1}{8\sigma^2}\|\Pi X - \Pi^* X\|_{\mathrm{F}}^2\right).
$$

Recall from Lemma 4.9, we have that

$$
\sum_{\Pi : \mathrm{d}(\Pi,\Pi^*) = \ell} \mathbb{E}\exp\left(-\frac{1}{8\sigma^2}\|\Pi X - \Pi^* X\|_{\mathrm{F}}^2\right) \leq \left(\frac{16 n^2 (2\sigma)^d}{\ell}\right)^{\ell/2} = \left(\frac{16 n (2\sigma_0)^d}{\ell}\right)^{\ell/2}.
$$

Therefore, it follows from a union bound that

$$
\begin{aligned}
\mathbb{P}\left\{\mathcal{E}^c\right\} &= \sum_{\frac{\beta n}{\log n} \le \mathrm{d}(\Pi,\Pi^*) < \delta n} \mathbb{P}\left\{L(\Pi) - L(\Pi^*) \ge \tau\left(\mathrm{d}(\Pi,\Pi^*)\right)\right\} \\
&\le \sum_{\frac{\beta n}{\log n} \le \ell < \delta n} e^{-\tau(\ell)/2} \left(\frac{16n(2\sigma_0)^d}{\ell}\right)^{\ell/2} \\
&= \sum_{\frac{\beta n}{\log n} \le \ell < \delta n} e^{-\ell} \le \delta n e^{-\frac{\beta n}{\log n}},
\end{aligned} \tag{4.62}
$$

where the last equality holds by choosing $\tau(\ell) = \ell \log(16e^2 n(2\sigma_0)^d/\ell)$.

For the second term in (4.62), we bound the truncated MGF as follows:

$$
\begin{aligned}
&\sum_{\Pi:\mathrm{d}(\Pi,\Pi^*)=\ell} \mathbb{E} e^{L(\Pi)-L(\Pi^*)} \mathbf{1}_{\{L(\Pi)-L(\Pi^*) \le \tau(\mathrm{d}(\Pi,\Pi^*))\}} \\
&\le \sum_{\Pi:\mathrm{d}(\Pi,\Pi^*)=\ell} \mathbb{E}\exp\left(\frac{1}{2}\left(L(\Pi) - L(\Pi^*) + \tau(\ell)\right)\right) \\
&\le \sum_{\Pi:\mathrm{d}(\Pi,\Pi^*)=\ell} \mathbb{E}\exp\left(-\frac{1}{8\sigma^2}\|\Pi X - \Pi^* X\|_{\mathrm{F}}^2\right) e^{\tau(\ell)/2} \\
&\le \left(\frac{16n(2\sigma_0)^d}{\ell}\right)^{\ell/2} e^{\tau(\ell)/2} \\
&\le \left(\frac{16en(2\sigma_0)^d}{\ell}\right)^{\ell}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\sum_{\frac{\beta n}{\log n} \le \mathrm{d}(\Pi,\Pi^*) < \delta n} \mathbb{E} e^{L(\Pi)-L(\Pi^*)} \mathbf{1}_{\{L(\Pi)-L(\Pi^*) \le r\mathrm{d}(\Pi,\Pi^*)\}} &\le \sum_{\frac{\beta n}{\log n} \le \ell < \delta n} \left(\frac{16en(2\sigma_0)^d}{\ell}\right)^{\ell} \\
&\le \delta n \left(\frac{16e(2\sigma_0)^d}{\delta}\right)^{\delta n},
\end{aligned}
$$

where the last inequality holds for all $\delta \le 16(2\sigma_0)^d$. Choosing $c' = \delta \log(16e^2(2\sigma_0)^d/\delta)$, we get that

$$e^{-c'n} \sum_{\frac{\beta n}{\log n} \leq \mathrm{d}(\Pi,\Pi^*) < \delta n} \mathbb{E} e^{L(\Pi)-L(\Pi^*)} \mathbf{1}_{\{L(\Pi)-L(\Pi^*) \leq r\mathrm{d}(\Pi,\Pi^*)\}} \leq \delta n e^{-\delta n} \tag{4.63}$$

Substituting (4.62) and (4.63) into (4.61), we get

$$\mathbb{P}\left\{ R_2 \geq \left(\frac{16e^2(2\sigma_0)^d}{\delta} \right)^{\delta n} \right\} \leq 2\delta n e^{-\beta n/\log n}.$$

Combining this with (4.60) and upon choosing $\beta = \delta$, we have

$$\mathbb{P}\left\{ R_1 + R_2 \geq 2\left(\frac{16e^2(2\sigma_0)^d}{\delta} \right)^{\delta n} \right\} \leq 4\delta n e^{-\delta n/\log n},$$

concluding the proof. □

To prove Lemma 4.8, we aim to construct exponentially many bad permutations π whose log likelihood $L(\pi)$ is no smaller than $L(\pi^*)$. It turns out that $L(\pi) - L(\pi^*)$ can be decomposed according to the orbit decomposition of $(\pi^*)^{-1} \circ \pi$ as per (4.15). Thus, following [DWXY21], we look for vertex-disjoint orbits O whose total lengths add up to $\Omega(n)$ and each of them is *augmenting* in the sense that $\Delta(O) \geq 0$.

In the planted matching model with independent weights [DWXY21], a great challenge lies in the fact that short augmenting orbits (even after taking their disjoint unions) are insufficient to meet the $\Omega(n)$ total length requirement. As a result, one has to search for long augmenting orbits of length $\Omega(n)$. However, due to the excessive correlations among long augmenting orbits, the second-moment calculation fundamentally fails. To overcome this challenge, [DWXY21] invents a two-stage finding scheme which first finds many but short augmenting paths and then patches them together to form a long augmenting orbit using the so-called sprinkling idea. Fortunately, in our low-dimensional case of $d = \Theta(1)$, as also observed in [KNW22], it suffices to look for augmenting 2-orbits and take their disjoint unions. More precisely, the following lemma shows that there are $\Omega(n)$ vertex-disjoint augmenting 2-orbits, from which we can easily extract exponentially many different unions of total length $\Omega(n)$. In contrast, to prove the failure of the MLE for almost perfect recovery in [KNW22], a single union of $\Omega(n)$ vertex-disjoint augmenting 2-orbits is sufficient.

Lemma 4.10. *If $\sigma = \sigma_0 n^{-1/d}$, then there exist constants $c(\sigma_0, d)$, $\delta_0(\sigma_0, d)$, and $n_0(\sigma_0, d)$ that only depend on σ_0 and d such that for all $n \geq n_0$, with probability at least $1/2 - c/n$, there are at least $\delta_0 n$ many vertex-disjoint augmenting 2-orbits.*

This lemma is proved in [KNW22, Section 4] using the so-called concentration-enhanced second-moment method. For completeness, here we provide a much simpler proof via the vanilla second-moment method combined with Turán's theorem.

Proof. Let I_{ij} denote the indicator that (i,j) is an augmenting 2-orbit and $I = \sum_{i<j} I_{ij}$. To extract a collection of vertex-disjoint augmenting 2-orbits, we construct a graph $G = (V, E)$, where the vertices correspond to (i,j) for which $I_{ij} = 1$, and (i,j) and (k,ℓ) are connected if (i,j) and (k,ℓ) share a common vertex. By construction, any collection of vertex-disjoint 2-orbits corresponds to an independent set in G. By Turán's theorem (see e.g. [AS08, Theorem 1, p. 95]), there exists an independent set S in G of size at least $|V|^2/(2|E| + |V|)$. It remains to bound $|V|$ from below and $|E|$ from above.

Note that $|V| = I = \sum_{i<j} I_{ij}$. For all n sufficiently large, $\sigma^2 \leq d/40$ and it follows from [KNW22, Prop. 4.3] that

$$p \triangleq \mathbb{P}\left\{I_{ij} = 1\right\} \geq \frac{1}{1000\sqrt{d}} \left(1 + \frac{1}{\sigma^2}\right)^{-d/2}.$$

Therefore,

$$\mathbb{E} I = \sum_{i<j} \mathbb{P}\left\{I_{ij} = 1\right\} \geq \binom{n}{2} \frac{1}{1000\sqrt{d}} \left(1 + \frac{1}{\sigma^2}\right)^{-d/2}. \tag{4.64}$$

Under the assumption that $\sigma = \sigma_0 n^{-1/d}$, it follows that $\mathbb{E} I \geq c_0(d, \sigma_0) n$ for some constant

$c_0(d, \sigma_0)$ that only depends on d and σ_0. Moreover,

$$\begin{aligned}\mathsf{Var}(I) &= \sum_{i<j, k<\ell} \mathsf{Cov}\left(I_{ij}, I_{k\ell}\right) \\ &= \sum_{i<j} \mathsf{Var}(I_{ij}) + \sum_{i<j} \sum_{k:k\neq i,j} \left(\mathsf{Cov}\left(I_{ij}, I_{ik}\right) + \mathsf{Cov}\left(I_{ij}, I_{jk}\right)\right) \\ &\leq \sum_{i<j} \mathbb{E} I_{ij}^2 + \sum_{i<j} \sum_{k:k\neq i,j} \left(\mathbb{E} I_{ij} I_{ik} + \mathbb{E} I_{ij} I_{jk}\right),\end{aligned}$$

where the second equality holds because I_{ij} and $I_{k\ell}$ are independent when $\{i,j\} \cap \{k,\ell\} = \emptyset$. Recall that $\mathbb{E} I_{ij}^2 = \mathbb{E} I_{ij} = p$. Moreover, it follows from [KNW22, Prop. 4.5] that

$$\mathbb{E} I_{ij} I_{ik} \leq \left(1 + \frac{3}{4\sigma^2}\right)^{-d}.$$

Combining the last three displayed equation yields that

$$\mathsf{Var}(I) \leq \mathbb{E} I + n^3 \left(1 + \frac{3}{4\sigma^2}\right)^{-d}. \tag{4.65}$$

Under the assumption that $\sigma = \sigma_0 n^{-1/d}$, it follows that $\mathsf{Var}(I) \leq \mathbb{E} I + c_1(d, \sigma_0) n$ for some $c_1(d, \sigma_0)$ that only depends on d and σ_0. By Chebyshev's inequality,

$$\mathbb{P}\left\{I \leq \frac{1}{2} \mathbb{E} I\right\} \leq \frac{4 \mathsf{Var}(I)}{(\mathbb{E} I)^2} \leq \frac{4(c_0 + c_1)}{c_0^2 n}.$$

Moreover,

$$|E| = \sum_{i<j} \sum_{k:k\neq i,j} \left(I_{ij} I_{ik} + I_{ij} I_{jk}\right)$$

and hence

$$\mathbb{E}|E| = \sum_{i<j} \sum_{k:k\neq i,j} \left(\mathbb{E} I_{ij} I_{ik} + \mathbb{E} I_{ij} I_{jk}\right) \leq n^3 \left(1 + \frac{3}{4\sigma^2}\right)^{-d} \leq c_1(d, \sigma_0) n.$$

By Markov's inequality, $|E| \leq 2\mathbb{E}|E|$ with probability at least $1/2$. Therefore, with probability at least $1/2 - 4c_1/(c_0^2 n)$,

$$|S| \geq \frac{|V|^2}{2|E| + |V|} \geq \frac{(\mathbb{E}I)^2/4}{4\mathbb{E}|E| + \mathbb{E}I/2} \geq \frac{c_0^2 n^2/4}{4c_1 n + c_0 n/2} \geq \delta_0 n,$$

for some constant $\delta_0(d, \sigma_0)$ that only depends on d and σ_0. □

Proof of Lemma 4.8. By Lemma 4.10, from $\delta_0 n$ such vertex-disjoint augmenting 2-orbits, we choose $\delta_0 n/2$ many of them and form a union of augmenting 2-orbits with the total length $\delta_0 n/2 \times 2 = \delta_0 n$. There are $\binom{\delta_0 n}{\delta_0 n/2}$ many different unions, and each of such union corresponds to a permutation Π with $\mathrm{d}(\Pi, \Pi^*) = \delta_0 n$ and $L(\Pi) \geq L(\Pi^*)$ in view of (4.15). Therefore, for any $\delta \leq \delta_0$,

$$\frac{\mu_{X,Y}(\mathbf{\Pi}_{\mathsf{bad}})}{\mu_{X,Y}(\Pi^*)} \geq \binom{\delta_0 n}{\delta_0 n/2} \geq 2^{\delta_0 n/2}.$$

□

4.4.2 Impossibility of almost perfect recovery

The optimal information-theoretic threshold for almost perfect recovery is stated in the following theorem.

Theorem 4.4. *Assume $\sigma = \sigma_0 n^{-1/d}$ for any constant $\sigma_0 \in (0, 1/2)$. There exists a constant $\delta_0(\sigma_0, d)$ that only depends on σ_0, d such that for any estimator $\widehat{\Pi}$ and all sufficiently large n,*

$$\mathbb{E}\mathrm{d}\left(\Pi^*, \widehat{\Pi}\right) \geq \delta_0 n.$$

Given the above two lemmas, Theorem 4.4 readily follows. Indeed, combining Lemma 4.7 and Lemma 4.8 and choosing δ such that $\delta \log(16e^2(2\sigma_0)^d/\delta) = \delta_0/4$ we get $\mu_{X,Y}(\mathbf{\Pi}_{\mathsf{bad}}) \geq \frac{e^{\delta_0 n/4}}{2+e^{\delta_0 n/4}}$ with probability at least $1/2 - c/n - 4\delta n e^{-\delta n/\log n}$, which shows that $\mathbb{E}[\mathrm{d}(\widetilde{\Pi}, \Pi^*)] \gtrsim \delta n$ as desired.

Theorem 4.4 readily implies that for constant d, $\sigma = o(n^{-1/d})$ is necessary for achieving the almost perfect recovery, i.e., $\mathbb{E}\mathsf{d}(\Pi^*, \widehat{\Pi}) = o(n)$.

4.4.3 Impossibility of perfect recovery

In this section, we prove an impossibility condition of perfect recovery.

Theorem 4.5. *Suppose that $\sigma^2 \le d/40$ and*

$$\frac{d}{4}\log\left(1+\frac{1}{\sigma^2}\right) - \log n + \log d \le C, \tag{4.66}$$

for a constant $C > 0$. Then there exists a constant c that only depends on C such that for any estimator $\widehat{\pi}$, $\mathbb{P}\left\{\widehat{\pi} \neq \pi^*\right\} \ge c$.

Theorem 4.5 immediately implies that if there exists an estimator that achieves perfect recovery with high probability, then

$$\frac{d}{4}\log\left(1+\frac{1}{\sigma^2}\right) - \log n + \log d \to +\infty. \tag{4.67}$$

In comparison, it is shown in [DCK19, Theorem 1] that perfect recovery is possible if $\frac{d}{4}\log\left(1+\frac{1}{\sigma^2}\right) - \log n \to +\infty$. Thus our necessary condition agrees with their sufficient condition up to an additive $\log d$ factor. Our necessary condition (4.67) further specializes to

- $d \ll \log n$:

$$\sigma \le \begin{cases} o(n^{-2/d}) & \text{if } d = O(1) \\ n^{-2/d} & \text{if } d \gg 1 \end{cases}.$$

  This yields Theorem 4.3(i) and slightly improves over the necessary condition of MLE in [KNW22, Theorem 1.1], that is, $\sigma = O(n^{-2/d})$.

- $d = \Theta(\log n)$:

$$\sigma \le \frac{1}{\sqrt{n^{4/d} - 1}};$$

- $d \gg \log n$:

$$\sigma \le \sqrt{\frac{d}{4\log(n/d) + \omega(1)}}.$$

Note that the previous work [DCK19] shows that $\frac{d}{4}\log\left(1 + \frac{1}{\sigma^2}\right) \ge (1 - \Omega(1))\log n$ is necessary for perfect recovery, under the additional assumption that $1 \ll d = O(\log n)$. The analysis therein is based on showing the existence of an augmenting 2-orbit via the second-moment method. We follow a similar strategy, but our first and second moment estimates are sharper and thus yield a tighter condition.

Proof. Recall that I_{ij} denote the indicator that (i, j) is an augmenting 2-orbit and $I = \sum_{i<j} I_{ij}$. For the purpose of lower bound, consider the Bayesian setting where π^* is drawn uniformly at random. Then the MLE $\widehat{\pi}_{\rm ML}$ given in (4.2) minimizes the probability of error. Hence, it suffices to bound from below $\mathbb{P}\left\{\widehat{\pi}_{\rm ML} \neq \pi^*\right\}$. Note that on the event $\{I > 0\}$, there exists at least one permutation $\pi \neq \pi^*$ whose likelihood is at least as large as that of π^* and hence the error probability of MLE is at least $1/2$. Therefore,

$$\mathbb{P}\left\{\widehat{\pi}_{\rm ML} \neq \pi^*\right\} \ge \frac{1}{2}\mathbb{P}\left\{I > 0\right\}.$$

It remains to bound $\mathbb{P}\left\{I > 0\right\}$ from below. To this end, we first bound $\mathsf{Var}(I)/\left(\mathbb{E}I\right)^2$. In view of (4.65),

$$\frac{\mathsf{Var}(I)}{\left(\mathbb{E}I\right)^2} \le \frac{1}{\mathbb{E}I} + \frac{1}{\left(\mathbb{E}I\right)^2} n^3 \left(1 + \frac{3}{4\sigma^2}\right)^{-d}.$$

By assumption $\sigma^2 \le d/40$ and (4.66), it follows from (4.64) that

$$\mathbb{E}I \gtrsim \frac{n^2}{\sqrt{d}}\left(1 + \frac{1}{\sigma^2}\right)^{-d/2} \ge \exp\left(\frac{3}{2}\log d - 2C\right) \ge \exp\left(-2C\right).$$

Moreover,

$$\frac{1}{(\mathbb{E}I)^2} n^3 \left(1+\frac{3}{4\sigma^2}\right)^{-d} \lesssim \frac{d}{n}\left(\frac{1+1/\sigma^2}{1+3/(4\sigma^2)}\right)^{d} \overset{(a)}{\le} \frac{d}{n}\left(1+\frac{1}{\sigma^2}\right)^{d/4} \overset{(b)}{\le} e^C,$$

where (a) holds because $1+3x/4 \ge (1+x)^{3/4}$ for all $x \ge 0$ and (b) holds due to assumption (4.66),

Combining the last three displayed equation yields that $\mathsf{Var}(I)/(\mathbb{E}I)^2 \le c_0$ for some constant c_0 that only depends on C. By the Paley-Zygmund inequality,

$$\mathbb{P}\{I>0\} \ge \mathbb{P}\left\{I \ge \frac{1}{2}\mathbb{E}I\right\} \ge \frac{(\mathbb{E}I)^2}{4\left(\mathsf{Var}(I)+(\mathbb{E}I)^2\right)} \ge \frac{1}{4c_0+1}.$$

□

4.5 Extensions

The proof strategies for the dot-product model can be further extended to other cases. For example, after simple modifications, we can also prove the same threshold for the distance model and models with non-isotropic distributions for the point cloud.

4.5.1 Distance Model

In this section, we prove Theorem 4.2. Let $\widetilde{X} \triangleq (I-\mathbf{F})X$, $\widetilde{Y} \triangleq (I-\mathbf{F})Y$ and $\widetilde{Z} \triangleq (I-\mathbf{F})Z$. Recall that the approximate MLE for the distance model is given by (4.9). As in the proof of Theorem 4.1, thanks to the orthogonal invariance of the nuclear norm $\|\cdot\|_*$, we may assume $\widetilde{A}^{1/2} = \widetilde{X}$ and $\widetilde{B}^{1/2} = \widetilde{Y}$ without loss of generality, so that

$$\widetilde{\Pi}_{\mathrm{AML}} = \arg\max_{\Pi \in \mathcal{S}_n} \|\widetilde{X}^\top \Pi^\top \widetilde{Y}\|_*.$$

Following the arguments for the dot-product model, a key step is to extend the estimate for $p(\Pi, Q)$ in (4.21) to the following MGF:

$$\widetilde{p}(\Pi, Q) \triangleq \mathbb{E}\exp\left\{-\frac{1}{32\sigma^2}\left\|\widetilde{X} - \Pi\widetilde{X}Q\right\|_{\mathrm{F}}^2\right\}, \tag{4.68}$$

where $\Pi \in \mathcal{S}_n$ and $Q \in O(d)$. The following lemma gives a comparison between the MGF for the distance model and that for the dot-product model defined in (4.21), the latter of which was previous estimated in Lemma 4.4.

Lemma 4.11. *Fix a permutation matrix $\Pi \in \mathcal{S}_n$. For $Q \in O(d)$, denote by $e^{\mathrm{i}\theta_1}, \ldots, e^{\mathrm{i}\theta_d}$ the eigenvalues of Q, where $\theta_1, \ldots, \theta_d \in [-\pi, \pi]$. Then*

$$\widetilde{p}(\Pi, Q) \le p(\Pi, Q)\prod_{\ell=1}^{d}\left(1 + \frac{\theta_l^2}{16\sigma^2}\right)^{1/2}. \tag{4.69}$$

Proof. Let $t = \frac{1}{32\sigma^2}$. Denote by $\widetilde{x} = \mathsf{vec}(\widetilde{X}) \in \mathbb{R}^{nd}$ the vectorization of $\widetilde{X}$ and recall that $x = \mathsf{vec}(X) \in \mathbb{R}^{nd}$ satisfies $x \sim \mathcal{N}(0, I_{nd})$. Then

$$\left\|\widetilde{X} - \Pi\widetilde{X}Q\right\|_{\mathrm{F}}^2 = \left\|(I_{nd} - Q^\top \otimes \Pi)\widetilde{x}\right\|^2 = \left\|(I_{nd} - Q^\top \otimes \Pi)(I_d \otimes (I_n - \mathbf{F}))x\right\|^2.$$

Denote $\widetilde{H} \triangleq (I_{nd} - Q^\top \otimes \Pi)(I_d \otimes (I_n - \mathbf{F}))$, then

$$\widetilde{p}(\Pi, Q) = \mathbb{E}\exp\left(-tx^\top\widetilde{H}^\top\widetilde{H}x\right) = \left[\det\left(I + 2t\widetilde{H}^\top\widetilde{H}\right)\right]^{-\frac{1}{2}}.$$

It suffices to compute the eigenvalues of $\widetilde{H}$. Recall that the spectrum of Π is given by (4.28). We claim that the spectrum of $\widetilde{H}$ is the following multiset

$$\mathsf{Spec}(\widetilde{H}) = \big(\mathsf{Spec}(H)\backslash\{1 - e^{-\mathrm{i}\theta_\ell} : \ell = 1, \ldots, d\}\big) \cup \{0 \text{ with multiplicity } d\}, \tag{4.70}$$

where $\mathsf{Spec}(H)$ is the spectrum of H defined in Lemma 4.4, given by

$$\mathsf{Spec}(H) = \left\{1 - e^{-\mathrm{i}\theta_\ell}\lambda_j : \lambda_j \in \mathsf{Spec}(\Pi),\ j = 1, \ldots, n,\ \ell = 1, \ldots, d\right\}.$$

Now we prove (4.70). As shown in (4.28), Π has eigenvalue 1 with multiplicity $\mathfrak{c}(\Pi)$, where $\mathfrak{c}(\Pi)$ denote the number of cycles. We denote these by $\lambda_1 = \cdots = \lambda_{\mathfrak{c}(\Pi)} = 1$. Using the cycle decomposition and the block diagonal structure as in Lemma 4.4, we know that the eigenvectors corresponding to $\lambda_1, \ldots, \lambda_{\mathfrak{c}(\Pi)}$ are of the following form

$$v_i = (0, \ldots, 0, 1, \ldots, 1, 0, \ldots, 0)^\top, \quad i = 1, \ldots, \mathfrak{c}(\Pi)$$

where the number of 1's equals the length of the corresponding cycle. In particular, due to the block diagonal structure, the 1 blocks in v_i's do not overlap. Therefore, we know that the vector $\widetilde{v}_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{\mathfrak{c}(\Pi)} v_i = \frac{1}{\sqrt{n}} \mathbf{1} = \frac{1}{\sqrt{n}}(1, \ldots, 1)^\top \in \mathbb{R}^n$ is in the eigenspace of 1. Using the Gram-Schmidt process, we can construct vectors $\widetilde{v}_2, \ldots, \widetilde{v}_{\mathfrak{c}(\Pi)}$ such that $\{\widetilde{v}_i\}_{i=1}^n$ is a orthonormal basis of the eigenspace, i.e.

$$\langle \widetilde{v}_i, \widetilde{v}_j \rangle = \delta_{ij}, \quad \mathsf{span}(\widetilde{v}_1, \ldots, \widetilde{v}_{\mathfrak{c}(\Pi)}) = \mathsf{span}(v_1, \ldots, v_{\mathfrak{c}(\Pi)}).$$

Pick an arbitrary eigenvalue μ of $Q^\top$ with eigenvector $w \in \mathbb{R}^d$, and also pick an arbitrary eigenvalue λ of Π with eigenvector $v \in \mathbb{R}^n$. Based on the arguments above, if $\lambda \neq \lambda_1$, then $v \perp \widetilde{v}_1$, and therefore

$$\widetilde{H}(w \otimes v) = w \otimes (I - \mathbf{F})v - (Q^\top w) \otimes \Pi(I - \mathbf{F})v = w \otimes v - \mu w \otimes \lambda v = (1 - \mu\lambda)(w \otimes v). \tag{4.71}$$

For the eigenpair $(\lambda_1, \widetilde{v}_1)$, we have

$$\widetilde{H}(w \otimes \widetilde{v}_1) = w \otimes (I - \mathbf{F})\widetilde{v}_1 - (Q^\top w) \otimes \Pi(I - \mathbf{F})\widetilde{v}_1 = w \otimes 0 - \mu w \otimes 0 = 0. \tag{4.72}$$

Combining (4.71) and (4.72), we conclude that for $\ell = 1, \ldots, d$ and $j = 2, \ldots, n$, the eigenvalue $1 - e^{-\mathrm{i}\theta_\ell}\lambda_j$ of H remains to be an eigenvalue of $\widetilde{H}$, while the eigenvalues $1 - e^{-\mathrm{i}\theta_\ell}\lambda_1 = 1 - e^{-\mathrm{i}\theta_\ell}$ of H are replaced by 0 in the spectrum of $\widetilde{H}$. Hence we have shown

(4.70) is true.

Using (4.70) and (4.27), we obtain

$$
\begin{aligned}
\widetilde{p}(\Pi, Q) &= \prod_{j=2}^{n} \prod_{\ell=1}^{d} \left(1 + 2t|1 - e^{-\mathrm{i}\theta_\ell}\lambda_j|^2\right)^{-1/2} \\
&= p(\Pi, Q) \prod_{\ell=1}^{d} \left(1 + 2t|1 - e^{-\mathrm{i}\theta_\ell}|^2\right)^{1/2} \\
&= p(\Pi, Q) \prod_{\ell=1}^{d} (1 + 2t(2 - 2\cos\theta_\ell))^{1/2} \\
&\le p(\Pi, Q) \prod_{\ell=1}^{d} \left(1 + \frac{\theta_\ell^2}{16\sigma^2}\right)^{1/2},
\end{aligned}
$$

which completes the proof. □

Applying Lemma 4.11, the following lemma is the counterpart of Lemma 4.5.

Lemma 4.12. *Suppose $d = o(\log n)$. For some $\sigma_0 > 0$, let $\delta = \sigma_0/\sqrt{n}$ and $N \subset O(d)$ be the δ-net defined in (4.19).*

(i) If $\sigma_0 = o(n^{-2/d})$, then

$$
\sum_{\Pi \neq I_n} \sum_{Q \in N} \mathbb{E} \exp\left\{-\frac{1}{32\sigma_0^2} \left\|\widetilde{X} - \Pi \widetilde{X} Q\right\|_{\mathrm{F}}^2\right\} = o(1). \tag{4.73}
$$

(ii) For any $\varepsilon = \varepsilon(n) > 0$, if $\sigma_0^{-d} > 16n2^{2/\varepsilon}$, then the following is true

$$
\sum_{\mathrm{d}(\pi, \mathrm{Id}) \ge \varepsilon n} \sum_{Q \in N} \mathbb{E} \exp\left\{-\frac{1}{32\sigma_0^2} \left\|\widetilde{X} - \Pi \widetilde{X} Q\right\|_{\mathrm{F}}^2\right\} = o(1). \tag{4.74}
$$

Proof. (i) Similarly as in Lemma 4.5 Part (i), using (4.69) we have

$$
\begin{aligned}
&\sum_{\Pi \neq I_n} \sum_{Q \in N} \widetilde{p}(\Pi, Q) \\
&\leq \sum_{n_1=0}^{n-2} \sum_{m_1,\dots,m_d=\lfloor -\frac{4\pi}{\delta} \rfloor}^{\lceil \frac{4\pi}{\delta} \rceil} \left\{ \left| N\left(e^{\mathrm{i}\frac{m_1\delta}{4}}, \dots, e^{\mathrm{i}\frac{m_d\delta}{4}}\right) \right| (n-n_1)! \binom{n}{n_1} (C\sigma_0)^{\frac{n+n_1}{2}d} \right. \\
&\qquad \left. \times \left[\prod_{\ell=1}^{d} \frac{1}{(\frac{\delta|m_\ell|}{4}+\sigma_0)^{n_1}} \left(1+\frac{\delta^2 m_\ell^2}{256\sigma_0^2}\right)^{\frac{1}{2}} \right] \right\} \\
&\leq \sum_{n_1=0}^{n-2} \left((C\sigma_0)^d n^2\right)^{\frac{n-n_1}{2}} \left[\sum_{m=\lfloor -\frac{4\pi}{\delta} \rfloor}^{\lceil \frac{4\pi}{\delta} \rceil} \frac{1}{(1+\frac{\delta}{4\sigma_0}|m|)^{n_1}} \left(1+\frac{\delta^2 m^2}{256\sigma_0^2}\right)^{\frac{1}{2}} \left(1+\frac{|m|}{2}\right)^{2d^2} \right]^d \\
&= \sum_{n_1=0}^{n-2} \left((C\sigma_0)^d n^2\right)^{\frac{n-n_1}{2}} \left[\sum_{m=\lfloor -\frac{4\pi}{\delta} \rfloor}^{\lceil \frac{4\pi}{\delta} \rceil} \frac{1}{(1+\frac{|m|}{4\sqrt{n}})^{n_1}} \left(1+\frac{m^2}{256n}\right)^{\frac{1}{2}} \left(1+\frac{|m|}{2}\right)^{2d^2} \right]^d \\
&\leq \sum_{n_1=0}^{n-2} \left((C\sigma_0)^d n^2\right)^{\frac{n-n_1}{2}} \left[\sum_{m=\lfloor -\frac{4\pi}{\delta} \rfloor}^{\lceil \frac{4\pi}{\delta} \rceil} \frac{1}{(1+\frac{|m|}{4\sqrt{n}})^{n_1}} \left(1+\frac{|m|}{2}\right)^{2d^2+1} \right]^d .
\end{aligned}
$$

Let

$$
\widetilde{T}_{n_1} \triangleq \sum_{m=\lfloor -\frac{4\pi}{\delta} \rfloor}^{\lceil \frac{4\pi}{\delta} \rceil} \frac{1}{(1+\frac{|m|}{4\sqrt{n}})^{n_1}} \left(1+\frac{|m|}{2}\right)^{2d^2+1} .
$$

Using the same arguments as in (4.35), (4.37) and (4.38), $\widetilde{T}_{n_1}$ can be bounded by

$$
\widetilde{T}_{n_1}^d \leq \begin{cases} C^{d^3} n^{d^3+d} L^{2d^2+2} & \text{if } n_1 \leq \sqrt{n}, \\ C^d (8\sqrt{n})^{2d^2+2} & \text{if } \sqrt{n} < n_1 < 32\sqrt{n}(\log n)^2, \\ C^d & \text{if } n_1 \geq 32\sqrt{n}(\log n)^2, \end{cases} \tag{4.75}
$$

where $L = \sigma_0^{-d}$. Consequently, following the similar estimates in (4.36) and (4.39),

$$
\sum_{\Pi \neq I_n} \sum_{Q \in N} \widetilde{p}(\Pi, Q) = o(1),
$$

which completes the proof.

(ii) Combined with (4.69), using the same arguments as in Lemma 4.5 Part (ii) yields

$$\sum_{\mathrm{d}(\pi,\mathrm{Id})\geq\varepsilon n}\sum_{Q\in N}\widetilde{p}(\Pi,Q)\leq\sum_{k=\varepsilon n}^{n}\binom{n}{k}L^{-k}k!\left(\frac{16L}{k}\right)^{k/2}\widetilde{T}^d_{n-k}=\widetilde{J}_1+\widetilde{J}_2$$

where

$$\widetilde{J}_1\triangleq\sum_{k=\varepsilon n}^{n-32\sqrt{n}(\log n)^2}\binom{n}{k}L^{-k}k!\left(\frac{16L}{k}\right)^{k/2}\widetilde{T}^d_{n-k},$$
$$\widetilde{J}_2\triangleq\sum_{k=n-32\sqrt{n}(\log n)^2+1}^{n}\binom{n}{k}L^{-k}k!\left(\frac{16L}{k}\right)^{k/2}\widetilde{T}^d_{n-k}.$$

By (4.75), these two term can be bounded in the same way as in (4.44) and (4.45). Thus,

$$\sum_{\mathrm{d}(\pi,\mathrm{Id})\geq\varepsilon n}\sum_{Q\in N}\widetilde{p}(\Pi,Q)=o(1),$$

which completes the proof. □

Lemma 4.12 implies the following high probability estimates. The proof is the same as in Lemma 4.6 via Chernoff bound and therefore we omit it here.

Lemma 4.13. *Suppose $d=o(\log n)$. For some $\sigma_0>0$, let $\delta=\sigma_0/\sqrt{n}$ and $N\subset O(d)$ be the δ-net defined in (4.19).*

(i) If $\sigma_0=o(n^{-2/d})$, for any constant $c>0$, the following inequality is true with high probability

$$\min_{\Pi\neq I_n}\min_{Q\in N}\left\|\widetilde{X}-\Pi\widetilde{X}Q\right\|_{\mathrm{F}}\geq c\sqrt{d}\sigma_0. \tag{4.76}$$

(ii) For any $\varepsilon=\varepsilon(n)>0$, if $\sigma_0^{-d}>16n2^{2/\varepsilon}$, the following is true for any fixed constant $c>0$ with high probability

$$\min_{\mathrm{d}(\pi,\mathrm{Id})\geq\varepsilon n}\min_{Q\in N}\left\|\widetilde{X}-\Pi\widetilde{X}Q\right\|_{\mathrm{F}}\geq c\sqrt{d}\sigma_0. \tag{4.77}$$

Now we are ready to prove Theorem 4.2. Similarly as in the dot-product model (see the remark following Theorem 4.1), for almost perfect recovery, we actually prove a stronger nonasymptotic bound: For sufficiently small ε, if $\sigma^{-d} > 16n2^{2/\varepsilon}$, then $\mathsf{overlap}(\widetilde{\pi}_{\mathrm{AML}}, \pi^*) \geq 1 - \varepsilon$ with high probability, which clearly implies Theorem 4.2 by taking $\sigma \ll n^{-1/d}$.

Proof of Theorem 4.2. (i) Let N be the δ-net for $O(d)$ defined in (4.19). Following the same argument as in Theorem 4.1

$$\mathbb{P}\left\{\|\widetilde{X}^\top\Pi^\top\widetilde{Y}\|_* \geq \|\widetilde{X}^\top\widetilde{Y}\|_*\right\} \leq \mathbb{P}\left\{\max_{Q\in N}\langle\widetilde{X}^\top\Pi^\top\widetilde{Y}, Q\rangle \geq (1-\delta^2)\langle\widetilde{X}^\top\widetilde{Y}, I_d\rangle\right\}.$$

For fixed Π and Q, we have

$$\begin{aligned}&\mathbb{P}\left\{\langle\widetilde{X}^\top\Pi^\top\widetilde{Y}, Q\rangle \geq (1-\delta^2)\langle\widetilde{X}^\top\widetilde{Y}, I_d\rangle\right\}\\ &\qquad = \mathbb{P}\left\{\sigma\langle\widetilde{Z}, (1-\delta^2)\widetilde{X} - \Pi\widetilde{X}Q\rangle \geq (1-\delta^2)\|\widetilde{X}\|_{\mathrm{F}}^2 - \langle\widetilde{X}, \Pi\widetilde{X}Q\rangle\right\}.\end{aligned}$$

Since the entries of $\widetilde{Z}$ are not independent, we need to be more careful:

$$\begin{aligned}\langle\widetilde{Z}, (1-\delta^2)\widetilde{X} - \Pi\widetilde{X}Q\rangle &= \langle(I-\mathbf{F})Z, (1-\delta^2)\widetilde{X} - \Pi\widetilde{X}Q\rangle\\ &= \langle Z, (I-\mathbf{F})((1-\delta^2)\widetilde{X} - \Pi\widetilde{X}Q)\rangle\\ &= \langle Z, (1-\delta^2)\widetilde{X} - \Pi\widetilde{X}Q\rangle,\end{aligned}$$

because $(I-\mathbf{F})\widetilde{X} = \widetilde{X}$ and $I - \mathbf{F}$ commutes with any permutation matrix Π. Therefore, similarly as in (4.48),

$$\begin{aligned}&\mathbb{P}\left\{\langle\widetilde{X}^\top\Pi^\top\widetilde{Y}, Q\rangle \geq (1-\delta^2)\langle\widetilde{X}^\top\widetilde{Y}, I_d\rangle\right\}\\ =&\mathbb{P}\left\{\sigma\mathcal{N}\left(0, (1-\delta^2)\left\|\widetilde{X} - \Pi\widetilde{X}Q\right\|_{\mathrm{F}}^2 - \delta^4\left\|\widetilde{X}\right\|_{\mathrm{F}}^2\right) \geq \frac{1}{2}\left\|\widetilde{X} - \Pi\widetilde{X}Q\right\|_{\mathrm{F}}^2 - \delta^2\left\|\widetilde{X}\right\|_{\mathrm{F}}^2\right\}\\ \leq&\mathbb{P}\left\{\sigma\mathcal{N}\left(0, \left\|\widetilde{X} - \Pi\widetilde{X}Q\right\|_{\mathrm{F}}^2\right) \geq \frac{1}{2}\left\|\widetilde{X} - \Pi\widetilde{X}Q\right\|_{\mathrm{F}}^2 - \delta^2\left\|\widetilde{X}\right\|_{\mathrm{F}}^2\right\}.\end{aligned} \tag{4.78}$$

Consider the events

$$\mathcal{E}_1 \triangleq \left\{ cdn \le \left\| \widetilde{X} \right\|_{\mathrm{F}}^2 \le Cdn \right\}, \quad \mathcal{E}_2 \triangleq \left\{ \min_{\Pi \neq I} \min_{Q \in N} \left\| \widetilde{X} - \Pi \widetilde{X} Q \right\|_{\mathrm{F}} \ge C\sqrt{d}\sigma \right\}.$$

We claim that $\mathbb{P}\{\mathcal{E}_1\} = 1 - o(1)$. To see this, note that

$$\begin{aligned} \|\widetilde{X}\|_{\mathrm{F}}^2 &= \langle (I - \mathbf{F})X, (I - \mathbf{F})X \rangle = \langle X, (I - \mathbf{F})X \rangle \\ &= \mathrm{Tr}(X^\top (I - \mathbf{F})X) = \sum_{i=1}^{d} \sum_{\alpha,\beta=1}^{n} X_{\alpha i} X_{\beta i} (I - F)_{\alpha\beta}. \end{aligned}$$

For each $i = 1, \ldots, d$, we have $\sum_{\alpha,\beta=1}^{n} X_{\alpha i} X_{\beta i} (I - \mathbf{F})_{\alpha\beta} = \mathsf{Col}_i(X)^\top (I - \mathbf{F}) \mathsf{Col}_i(X)$, where $\mathsf{Col}_i(X) \sim \mathcal{N}(0, I_n)$ is the i-th column of X. By Hanson-Wright inequality (see e.g. [RV13, Theorem 1.1]), for each $t \ge 0$,

$$\begin{aligned} &\mathbb{P}\left\{ |\mathsf{Col}_i(X)^\top (I - \mathbf{F}) \mathsf{Col}_i(X) - \mathbb{E}\mathsf{Col}_i(X)^\top (I - \mathbf{F}) \mathsf{Col}_i(X)| > t \right\} \\ &\qquad \le 2 \exp\left[-c \min\left(\frac{t^2}{\|I - \mathbf{F}\|_{\mathrm{F}}^2}, \frac{t}{\|I - \mathbf{F}\|} \right) \right]. \end{aligned}$$

Taking $t = n^{3/4}$ and simplifying the above inequality yield

$$\mathbb{P}\left\{ |\mathsf{Col}_i(X)^\top (I - \mathbf{F}) \mathsf{Col}_i(X) - (n-1)| > n^{3/4} \right\} \le 2 \exp\left(-cn^{1/2} \right). \tag{4.79}$$

Note that (4.79) is true for every $i = 1, \ldots, d$, and the columns $\mathsf{Col}_i(X)$'s are independent. This immediately gives us $\mathbb{P}\{\mathcal{E}_1\} = 1 - o(1)$. Moreover, by Lemma 4.13 we also have $\mathbb{P}\{\mathcal{E}_2\} = 1 - o(1)$. On the events $\mathcal{E}_1$ and $\mathcal{E}_2$, the estimate (4.78) reduces to

$$\begin{aligned} &\mathbb{P}\left\{ \langle \widetilde{X}^\top \Pi^\top \widetilde{Y}, Q \rangle \ge (1 - \delta^2) \langle \widetilde{X}^\top \widetilde{Y}, I_d \rangle, \mathcal{E}_1, \mathcal{E}_2 \right\} \\ &\le \mathbb{P}\left\{ \sigma \mathcal{N}\left(0, \left\| \widetilde{X} - \Pi \widetilde{X} Q \right\|_{\mathrm{F}}^2 \right) \ge \frac{1}{4} \left\| \widetilde{X} - \Pi \widetilde{X} Q \right\|_{\mathrm{F}}^2 \right\} \le \mathbb{E} \exp\left\{ -\frac{1}{32\sigma^2} \left\| \widetilde{X} - \Pi \widetilde{X} Q \right\|_{\mathrm{F}}^2 \right\}. \end{aligned} \tag{4.80}$$

Combining this with Lemma 4.12 and applying a union bound, we have

$$
\begin{aligned}
&\mathbb{P}\left\{\max_{\Pi\neq I}\|\widetilde{X}^\top\Pi^\top\widetilde{Y}\|_*\geq\|\widetilde{X}^\top\widetilde{Y}\|_*\right\}\\
&\leq\mathbb{P}\left\{\max_{\Pi\neq I_n}\max_{Q\in N}\langle\widetilde{X}^\top\Pi^\top\widetilde{Y},Q\rangle\geq(1-\delta^2)\langle\widetilde{X}^\top\widetilde{Y},I_d\rangle,\mathcal{E}_1,\mathcal{E}_2\right\}+\mathbb{P}\left\{\mathcal{E}_1^c\right\}+\mathbb{P}\left\{\mathcal{E}_2^c\right\}\\
&\leq\sum_{\Pi\neq I_n}\sum_{Q\in N}\mathbb{P}\left\{\langle\widetilde{X}^\top\Pi^\top\widetilde{Y},Q\rangle\geq(1-\delta^2)\langle\widetilde{X}^\top\widetilde{Y},I_d\rangle,\mathcal{E}_1,\mathcal{E}_2\right\}+o(1)\\
&\leq\sum_{\Pi\neq I_n}\sum_{Q\in N}\mathbb{E}\exp\left\{-\frac{1}{32\sigma^2}\left\|\widetilde{X}-\Pi\widetilde{X}Q\right\|_{\mathrm{F}}^2\right\}+o(1)\\
&=o(1).
\end{aligned}
$$

This implies $\widetilde{\pi}_{\mathrm{AML}}=\mathrm{Id}$ with high probability, which completes the proof.

(ii) The idea is the same as Theorem 4.1 Part (ii). For a sufficiently small $\varepsilon=\varepsilon(n)>0$, take $\sigma^{-d}>16n2^{2/\varepsilon}$ and consider the event

$$
\mathcal{E}_2'\triangleq\left\{\min_{\mathrm{d}(\pi,\mathrm{Id})\geq\varepsilon n}\min_{Q\in N}\left\|\widetilde{X}-\Pi\widetilde{X}Q\right\|_{\mathrm{F}}\geq C\sqrt{d}\sigma\right\}.
$$

Then Lemma 4.13 implies $\mathbb{P}\left\{\mathcal{E}_2'\right\}=1-o(1)$. On the event $\mathcal{E}_1$ and $\mathcal{E}_2'$, the reduction estimate (4.78) for Π with $\mathrm{d}(\pi,\mathrm{Id})\geq\varepsilon n$ still holds

$$
\mathbb{P}\left\{\langle\widetilde{X}^\top\Pi^\top\widetilde{Y},Q\rangle\geq(1-\delta^2)\langle\widetilde{X}^\top\widetilde{Y},I_d\rangle,\mathcal{E}_1,\mathcal{E}_2'\right\}\leq\mathbb{E}\exp\left\{-\frac{1}{32\sigma^2}\left\|\widetilde{X}-\Pi\widetilde{X}Q\right\|_{\mathrm{F}}^2\right\}.
$$

Combined with Lemma 4.12, we have

$$
\begin{aligned}
&\mathbb{P}\left\{\max_{\mathrm{d}(\pi,\mathrm{Id})\geq\varepsilon n}\|\widetilde{X}^\top\Pi^\top\widetilde{Y}\|_*\geq\|\widetilde{X}^\top\widetilde{Y}\|_*\right\}\\
&\qquad\qquad\leq\sum_{\mathrm{d}(\pi,\mathrm{Id})\geq\varepsilon n}\sum_{Q\in N}\mathbb{E}\exp\left\{-\frac{1}{32\sigma^2}\left\|\widetilde{X}-\Pi\widetilde{X}Q\right\|_{\mathrm{F}}^2\right\}+o(1)=o(1).
\end{aligned}
$$

Thus,

$$
\mathbb{P}\left\{\mathsf{overlap}(\widetilde{\pi}_{\mathrm{AML}},\pi^*)\geq1-\varepsilon\right\}=1-o(1),
$$

which completes the proof. $\square$

4.5.2 Case of Non-isotropic Distribution

In this section we argue that Theorem 4.1 continues to hold under the same conditions in the non-isotropic case of $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma)$, provided that $\Sigma \succ cI$ for some small constant $c > 0$. In the general non-isotropic case, we denote by $p(\Sigma, \sigma, \Pi, Q)$ the moment generating function given by (4.21) to highlight the dependency on the covariance matrix Σ and the noise level σ. As in the proof of Lemma 4.5, recall $x = \mathsf{vec}(X)$ denotes the vectorization of X. Since $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma)$, the vector $x \in \mathbb{R}^{nd}$ has distribution $x \sim \mathcal{N}(0, I_n \otimes \Sigma)$. Note that $I_n \otimes \Sigma \succ cI_{nd}$. Modifying (4.26) accordingly, we have

$$
\begin{aligned}
p(\Sigma, \sigma, \Pi, Q) = \mathbb{E} \exp\left(-\frac{1}{32\sigma^2} x^\top H^\top H x\right) &= \left[\det\left(I + \frac{1}{16\sigma^2} H^\top H (I_n \otimes \Sigma)\right)\right]^{-\frac{1}{2}} \\
&\le \left[\det\left(I + \frac{c}{16\sigma^2} H^\top H\right)\right]^{-\frac{1}{2}} = p(I, \sigma', \Pi, Q),
\end{aligned}
$$

where $H = I_{nd} - Q^\top \otimes \Pi$ and $\sigma' = \sigma/\sqrt{c}$. This shows that the MGF $p(\Sigma, \sigma, \Pi, Q)$ satisfies the same estimates (4.24), (4.25) and Lemma 4.5 for the isotropic case with the original noise σ replaced by a constant multiple of it σ'. This constant multiplicative factor keeps σ' satisfying the same noise threshold in Theorem 4.1, which implies both prefect recovery and almost perfect recovery can still be achieved for the non-isotropic case under the same conditions, hence confirming our claim.

Chapter 5

Method of Moments and Sublinear Sample Complexity for High-Dimensional Empirical Bayes Linear Models

5.1 Introduction

Empirical Bayes, introduced by Robbins [Rob51, Rob56], is a powerful paradigm for large-scale inference. Developed originally for sequence models, empirical Bayes have been extensively studied for deconvolution/denoising problems. The core idea of empirical Bayes is to estimate the prior distribution in the traditional Bayesian method via data-driven approaches, and then apply the downstream Bayesian inference based on the estimated prior. Such methods have been widely explored in statistics [Cas85, Zha03, Efr12] and also received attention from other applied areas [VH96, ETST01, Bro08].

While the theoretical study of empirical Bayes are well-developed for sequence models, many modern applications involve more complex dependence structures between the data and the latent variables. One attempt for such a more complicated model is high-dimensional linear regression. Consider the Bayesian linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_p] \in \mathbb{R}^{n\times p}$ is a fixed or random design matrix, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p) \in \mathbb{R}^p$ where $\beta_1, \ldots, \beta_p$ are iid from some prior g, and $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 I_n)$ independent of $\boldsymbol{\beta}$. The goal is to estimate the prior g from the data $(\mathbf{X}, \mathbf{y})$, under the assumption that the sample size n and the dimension p are both large.

To our knowledge, the development of statistical theory of empirical Bayes for linear models in high dimensions is still limited. The linear model is analogous to a sequence model with correlated errors (and the sequence model can be seen as a special case of the linear model with design $\mathbf{X} = \mathbf{I}$), but extending the empirical Bayes theory and computational efficient methods to high-dimensional linear regression has substantial difficulties. Although both maximum likelihood estimation and method of moments, the two most commonly used methodologies, have achieved huge success in sequence models (in terms of both statistical optimality and computational efficiency), they are technically challenging to directly adapt to linear models. Due to difficulties such as computing the objective maximum likelihood or handling the interdependency of moments estimators, the literature has very few rigorous studies for the statistical theory of empirical inference for large-scale multiple linear regression, and previous attempts either imposed inflexible restriction to the priors or struggled with considerable computational challenges [NS86, GF00, YL05]

Existing works for empirical Bayes in linear models typically follow the idea to approximately compute the maximum likelihood estimator of the prior [MSS23, FGSW23], which originates from the nonparametric maximum likelihood estimator (NPMLE) that was introduced in [Rob50, KW56] to study sequence models. Unlike the sequence model, the marginal likelihood of the data does not decompose across the coordinates and is nonconvex with respect to the prior in general. In particular, the work of Mukherjee-Sen-

Sen [MSS23] took a mean-field approach and developed a varional inference method. In the context of empirical Bayes linear regression, the variational empirical Bayes method jointly estimate the prior and apprximate the posterior. In [MSS23], the authors showed a naive mean-field approximation to the full empirical Bayes objective can recover the NPMLE asymptotically. Another method, developed in [FGSW23], introduced a gradient flow framework. This gradient flow approach aims to directly optimizing the empirical Bayes likelihood. By coupling a Langevin diffusion for the posterior with a Fisher-Rao gradient flow for the prior, it can be shown this continuous-time framework converges in polynomial time to a near-NPMLE. Despite making the first steps towards understanding the consistency of empirical Bayes estimators, the previous approaches have two notable drawbacks: (1) they require strong assumptions, some of which are even hard to verify in practice; (2) they require a relatively high sample complexity. In particular, [MSS23] requires $n \geq p$ and was improved to $n \asymp p$ in [FGSW23]. The optimal sample complexity is still poorly understood.

As a counterpart of the MLE approach in statistical estimation, the method of moments (MoM) and its variants are also extensively used in sequence models. However, applying MoM to linear models is also technically challenging due to the following two distinct phenomenons:

- Unlike the sequence models in which each moment can be estimated separately, the interactions of the parameters $\beta_1, \ldots, \beta_p$ result in a nonlinear relation among the moments. Consequently, to estimate one moment m_k, an estimation of all previous moments $m_1, \ldots, m_{k-1}$ are necessary.

- In the sequence model $y_i = \theta_i + \varepsilon_i$ for $i = 1, \ldots, n$ where $\theta_i \overset{\text{i.i.d.}}{\sim} g$, estimating g is a simple deconvolution problem as the coordinates y_i are i.i.d drawn from $g * \mathcal{N}(0, \sigma^2)$. In contrast, for the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, the interaction of the coefficient makes the distribution of y_i not just a convolution between g and $\mathcal{N}(0, \sigma^2)$ but also g convolves with itself. As a consequence, to recover the prior g, we need to deconvolve

g from both the noise and itself. This new phenomenon of *self-deconvolution* is a key difference from the sequence models. We refer to Section 5.3 for detailed discussions.

In this work, we adapt the method of moments to linear models. We proposed Empirical Bayes Method of Moments (EBMoM) that estimates the central moments of the prior $\mu_k(g) = \mathbb{E}[(\beta - \mathbb{E}[\beta])^k]$ where $\beta \sim g$. Combined with efficient moment-to-distribution pipelines (e.g. the Denoised Method of Moments in [WY20a]), our algorithm recovers the prior with polynomial time complexity and provable consistency guarantee under mild conditions. Our estimator deals with the two challenges mentioned above efficiently, and more importantly, we made the first step to understand how much sample is required for consistency in high-dimensional empirical Bayes regression by proving an optimal sublinear sample complexity.

Our work makes a progress towards more a comprehensive understanding of empirical Bayes in high-dimensional linear models, as well as of the generalization for the method of moments framework to correlated structures. We summarize our contributions as follows:

1. We design an computationally efficient estimator of the prior g, with polynomial time complexity.
2. By establishing non-asymptotic error estimates for the moments, we show that a sublinear sample complexity $n \geq p^{1-o(1)}$ suffices for consistency.
3. We prove that the threshold is information theoretically optimal.

5.1.1 Related work

MoM for sequence models MoM is one of the most classic method for parameter estimation and can be traced back to Pearson [Pea94]. While MoM is conceptually simple, it suffers from several issues in practice such as solvability of the polynomial system derived from the moments, the intractable computational efficiency, or strong assumptions to ensure theoretical consistency. There were various modifications to remedy the drawbacks of traditional MoM, including the Generalized Method of Moment (GMM) by Hansen [Han82]. GMM was broadly applied in analyzing financial and economic data (see [Hal05] for a more comprehensive review), but it lacked a rigorous theoretical understanding for a long time.

In theoretical computer science, various moment-based methods obtained fruitful progress for estimating sequence models, especially Gaussian location-scale mixture models [BS10b, KMV10, MV10, HP15, LS17]. However, these methods typically still have high computational complexity (e.g. [MV10, KMV10, BS10b] rely on a grid search of parameters). In a seminal work [WY20a], an innovation of the classic MoM named *denoised method of moments* (DMM) was developed that overcomes these drawbacks. A key insight in DMM is to project moment estimations onto the moment space, and we refer to [WY20b] for a more comprehensive discussion.

MoM for Bayesian linear models MoM has been widely used in parametric Bayesian linear models, in which a common assumption is to assume the parameters are drawn from some Gaussian prior with unknown variance. MoM is a classic method for this variance estimation task, and was first developed in [Hen53]. A more comprehensive exploration of MoM for linear models was pioneered by Rao in his work of the minimum-variance estimators (MINQUE) [Rao71, Rao72]. Specifically, for the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with Gaussian prior $\beta_1, \ldots, \beta_p \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, s^2)$ and noise $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$, the MoM estimator takes the form

$$\widehat{s}^2 = \mathbf{y}^\top \mathrm{B}\mathbf{y}, \quad \text{where B satisfies } \mathbb{E}[\mathbf{y}^\top \mathrm{B}\mathbf{y}] = s^2. \tag{5.1}$$

One popular modern procedure for estimating the Bayesian linear model is the LD-score regression [BSLF+15]. Assuming a standardization of variants $\|\mathbf{x}_j\| = 1$, LD-score regression is based on the observation

$$\mathbb{E}[(\mathbf{x}_j^\top \mathbf{y})^2] = s^2 \sum_{k=1}^{p} \left[(\mathbf{x}_j^\top \mathbf{x}_k)^2 - \frac{1}{n} \right] + \frac{p}{n} s^2 + \sigma^2.$$

This method is recognized to be within the general framework of MoM (5.1) in [Zho17].

For application purposes, models beyond the linear model with isotropic Gaussian prior were proposed and more complex estimation tasks were raised. For example, [OSH+19] introduced a fourth moment/excess-kurtosis type statistic to measure quantitative polygenicity in genetics

$$\kappa_e = \frac{3\mathbb{E}[\alpha^2\beta^2] - 2\mathbb{E}[\beta^4]}{(E[\alpha\beta])^2},$$

where $\alpha_j = \mathbf{x}_j^\top \mathbf{X}\boldsymbol{\beta}$, (α, β) is a uniformly random entry (α_j, β_j) with $j = 1, \ldots, p$ and the expectation is over this uniformly random choice and the randomness of $\boldsymbol{\beta}$. Under a Bayesian setting $\beta_j \sim \mathcal{N}(0, s_j^2)$, an MoM estimator for κ_e was proposed in [OSH+19, O'C21] based on the observation

$$\kappa_e = \frac{3p}{\sum s_k^2} \sum_{j=1}^{p} \frac{s_j^2}{\sum s_k^2} \mathbb{E}[\alpha_j^2].$$

Another extension is the linear mixed model $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \cdots + \mathbf{X}_K\boldsymbol{\beta}_K + \boldsymbol{\varepsilon}$. The linear mixed model is now one of the predominant model for application in various fields such as genetics [YLGV11], and MoM is applied for various inference problems in this model [LBG+15].

5.1.2 Notations

We use the standard big-O notations: for two positive sequences $\{a_n\}$ and $\{b_n\}$, we denote $a_n = O(b_n)$ if $a_n \le Cb_n$ for some constant $C > 0$ and we write $a_n = O_k(b_n)$ if C depends on some other parameter k.

For $x \in \mathbb{R}^p$, $x^{\otimes k}$ is a rank-one tensor given by $x^{\otimes k}_{a_1,\ldots,a_k} = x_{a_1} \cdots x_{a_k}$. For a tensor

$T \in (\mathbb{R}^p)^{\otimes k}$, its trace is defined as $\mathrm{Tr}(T) := \sum_{a=1}^{p} T_{a,\ldots,a}$. For $x \in \mathbb{R}^p$, $x^{\otimes k}$ is a rank-one tensor given by $x^{\otimes k}_{a_1,\ldots,a_k} = x_{a_1} \cdots x_{a_k}$. For tensors S and T, the tensor inner product is defined as $\langle S, T\rangle = \sum_{a_1,\ldots,a_k=1}^{p} S_{a_1,\ldots,a_k} T_{a_1,\ldots,a_k}$.

For a probability distribution g, we use $m_k(g)$ to denote its k-th moment and use $\mu_k(g) = \mathbb{E}[(\beta - \mathbb{E}[\beta])^k]$ where $\beta \sim g$ to denote its k-th central moment. For $\ell \geq 1$, the moment vector up to order L is defined as $\boldsymbol{m}_\ell(g) = (m_1(g), \ldots, m_\ell(g))$. The moment tensor of a random vector $\boldsymbol{\beta}$ is defined as $M^{(\ell)}(\boldsymbol{\beta}) = \mathbb{E}[\boldsymbol{\beta}^{\otimes \ell}]$; that is, $M^{(\ell)}_{s_1,\ldots,s_p} = \mathbb{E}[\beta_{s_1} \ldots \beta_{s_p}]$.

5.1.3 Organization

The paper is organized as follows. In Section 5.2, we proposed our estimation method *Empirical Bayes Method of Moments* (EBMoM) and provide its statistical guarantees and the efficient computation algorithm. In Section 5.3, we discussed the sharp sample complexity for empirical Bayes in high-dimensional linear models. The proofs for statistical consistency of EBMoM are given in Section 5.4 and the proof for lower bound of sample complexity is in Section 5.5.

5.2 Method of moments for EB linear regression

5.2.1 Construction of moment estimators

In a nutshell, to estimate the first L moments of the prior, we aim to set up L estimating equations, which, upon taking expectations, result in a system of polynomial equations that are *lower triangular*. That is, for each i, the expectation of the ith summary statistic is a polynomial of $m_1, \ldots, m_{i-1}$ plus a *linear* term of m_i. This suggests we may use previously estimates of lower-order moments to successively peel off the nonlinear terms and obtain an estimate of the next moment.

For a fixed degree $L \in \mathbb{N}$, we construct estimators for the mean $m_1(g)$ and central moments $\mu_2(g), \ldots, \mu_L(g)$. These central moments are estimated recursively, with variance as the base case. We refer to this estimator the *Empirical Bayes Method of Moments* (EBMoM).

Step 1: Mean estimator. We estimate the first moment m_1 by

$$\widehat{m}_1 = \frac{\mathbf{1}^\top \mathbf{X}^\top \mathbf{y}}{\|\mathbf{X}\mathbf{1}\|^2}, \tag{5.2}$$

where $\mathbf{1}$ denotes the all one vector in $\mathbb{R}^p$. This estimator, which is linear in the response $\mathbf{y}$, is an unbiased estimator of m_1. There is in fact a continuum of unbiased linear estimators.

Step 2: Variance estimator. Denote the central moments of the prior by $\mu_\ell := \mathbb{E}[(\beta - m_1)^\ell]$, where $\beta \sim g$. The estimator for μ_2, the variance of g, is defined as

$$\widehat{\mu}_2 := \frac{\|\mathbf{X}^\top \widehat{\mathbf{y}}\|^2 - \|\mathbf{X}\|_{\mathrm{F}}^2}{\|\mathbf{X}^\top \mathbf{X}\|_{\mathrm{F}}^2}, \qquad \widehat{\mathbf{y}} := \mathbf{y} - \widehat{m}_1 \mathbf{1}. \tag{5.3}$$

Step 3: Estimating higher moments. The estimator for central moments $\{\mu_\ell\}$ of order $\ell \geq 3$ are recursively defined as follows, with the above $\widehat{\mu}_2$ as the base case.

First, we introduce the needed tensor notations. Denote the Gram matrix by $\mathbf{H} := \mathbf{X}^\top \mathbf{X}$. For each ℓ, define a symmetric order-ℓ tensor $T^{(\ell)} \in (\mathbb{R}^p)^{\otimes \ell}$ as

$$(T^{(\ell)})_{s_1,\ldots,s_\ell} := \sum_{j=1}^{p} \mathbf{H}_{j,s} \ldots \mathbf{H}_{j,s_\ell}, \qquad s_1, \ldots, s_\ell \in [p]. \tag{5.4}$$

Given any partition of ℓ into t positive integers $d_1, \ldots, d_t \geq 1$ such that $d_1 + \ldots + d_t = \ell$, define a upper-triangular order-t tensor $\widetilde{T}^{(d_1,\ldots,d_t)} \in (\mathbb{R}^p)^{\otimes t}$ by

$$\widetilde{T}^{(d_1,\ldots,d_t)}_{s_1,\ldots,s_t} := T^{(\ell)}_{\underbrace{s_1,\ldots,s_1}_{d_1\text{-many}},\ldots,\underbrace{s_t,\ldots,s_t}_{d_t\text{-many}}} = \binom{\ell}{d_1,\ldots,d_t} \sum_{j=1}^{p} \mathbf{H}_{j,s_1}^{d_1} \ldots \mathbf{H}_{j,s_t}^{d_t}, \qquad 1 \leq s_1 < \ldots < s_t \leq p, \tag{5.5}$$

and zero otherwise. Note that for the finest partition $d_1 = \ldots = d_\ell = 1$, $\widetilde{T}^{(1,\ldots,1)}$ coincides with $T^{(\ell)}$. Let

$$A_\ell := \operatorname{Tr} T^{(\ell)} = \sum_{i,j=1}^{p} \mathbf{H}_{i,j}^{\ell}. \tag{5.6}$$

Note that we always have $A_\ell \geq 0$ since $A_\ell = \mathbf{1}^\top \mathbf{H}^{(\ell)} \mathbf{1}$, where $\mathbf{H}^{(\ell)}$ denotes the ℓ-th Hadamard product of $\mathbf{H}$ and is PSD.

Next, we estimate the central moments. Given $\widehat{\mu}_2, \ldots, \widehat{\mu}_{\ell-1}$, we construct an estimator $\widehat{\mu}_\ell$ for μ_ℓ as follows. Let $K_\ell = M^\ell \ell^{\ell/2}$ where M is the sub-Gaussian constant of the prior g (see Assumption 5.2 below for precise definitions), and $\widehat{\mu}_\ell^\circ := (\widehat{\mu}_\ell \wedge 2K_\ell) \vee (-2K_\ell)$ be the truncated version of $\widehat{\mu}_\ell$. Define

$$\widehat{\mu}_\ell = \frac{1}{A_\ell}\Big[F_\ell(\widehat{\mathbf{y}}) - \sum_{t=2}^{\ell} \sum_{d_1+\ldots+d_t=\ell : d_j \geq 2} \widehat{\mu}_{d_1}^\circ \ldots \widehat{\mu}_{d_t}^\circ \sum_{s_1,\ldots,s_t=1}^{p} \widetilde{T}_{s_1,\ldots,s_t}^{(d_1,\ldots,d_t)}\Big], \tag{5.7}$$

where the summation is over ordered t-tuple $d_1, \ldots, d_t$ as in (5.5), and

$$F_\ell(\widehat{\mathbf{y}}) := \sum_{j=1}^{p} (\sigma \|\mathbf{x}_j\|)^\ell H_\ell\Big(\frac{\mathbf{x}_j^\top \widehat{\mathbf{y}}}{\sigma \|\mathbf{x}_j\|}\Big). \tag{5.8}$$

Here $H_k(x) \equiv (-1)^k e^{x^2/2} \frac{\mathrm{d}^k}{\mathrm{d}x^k} e^{-x^2/2}$ denotes the degree-k Hermite polynomial. Note that (5.7) recovers the the variance estimator $\widehat{\mu}_2$ in (5.3), since $\widehat{\mu}_1 \equiv 0$ and $H_2(x) = x^2 - 1$.

Let us discuss several salient ramifications of the proposed estimator.

From moments to distribution After obtaining the moment estimates, we can convert it to an estimate of the prior. There are multiple ways to do so. For example, we may apply the *denoised method of moment* [WY20a], a special case of the generalized method of moments [Han82], to produce an estimator that is a discrete distribution. Specifically,

1. For some odd $L = 2k - 1$, we first project the noisy moment estimates $\widehat{\boldsymbol{\mu}}_L = (\widehat{\mu}_1, \ldots, \widehat{\mu}_L)$ with $\widehat{\mu}_1 \equiv 0$ onto the space of moments by solving the following

semidefinite program [WY20a]:

$$\widetilde{\boldsymbol{\mu}}_L := \operatorname{argmin}\{\|\mathbf{m} - \widehat{\mathbf{m}}_L\| : \mathbf{m} \in \mathcal{M}_L([-A, A])\}, \tag{5.9}$$

where

$$\mathcal{M}_L([-A, A]) := \{\boldsymbol{m}_L(\pi) : \pi \text{ supported on } [-A, A]\}, \tag{5.10}$$

is the space of moment vector $\boldsymbol{m}_L(\pi) := (m_1(\pi), \ldots, m_L(\pi))$ over all distributions π supported on $[-A, A]$. Here the hyperparameters $L \in \mathbb{N}$ and $A > 0$ may be chosen depending on n, p and the tail assumption of the true prior g.

2. Next, we construct a discrete probability distribution $\widetilde{g}$ with matching moments, such that $m_\ell(\widetilde{g}) = \widetilde{\mu}_\ell$ for $\ell = 1, \ldots, L$. This $\widehat{g}$ is a k-atomic distribution that can be found using Gauss quadrature [WY20a].

3. Finally, we shift the atoms of $\widetilde{g}$ by the estimated mean $\widehat{m}_1$ to produce the final estimator $\widehat{g}$.

Alternatively, if continuous estimates are desired, one may fit a continuous parametric family of densities (e.g. mixture of Gaussians) by moment matching.

Motivation of the EBMoM estimator The intuition behind the EBMoM estimator is the following. Unlike the sequence model, where each moment can be estimated separately, in the linear model, the independent entries of $\boldsymbol{\beta}$ interact through the design matrix, so that estimating a given moment requires estimates of lower order moments.

Recall that the Hermite polynomial provides an unbiased estimator of monomials in the normal mean model:

$$\mathbb{E}_{Z \sim \mathcal{N}(\mu, 1)}[H_k(Z)] = \mu^k, \quad \mu \in \mathbb{R}. \tag{5.11}$$

For simplicity, let us assume the true prior has zero mean and consider a simplified version of the summary statistic (5.8) without centering $\mathbf{y}$:

$$F_\ell(\mathbf{y}) = \sum_{j=1}^{p} (\sigma\|\mathbf{x}_j\|)^\ell H_\ell\Big(\frac{\mathbf{x}_j^\top \mathbf{y}}{\sigma\|\mathbf{x}_j\|}\Big). \tag{5.12}$$

Note that conditioned on $\boldsymbol{\beta}$, we have

$$\frac{\mathbf{x}_j^\top \mathbf{y}}{\sigma\|\mathbf{x}_j\|} \mid \boldsymbol{\beta} \sim \mathcal{N}\left(\frac{\mathbf{x}_j^\top \mathbf{X}\boldsymbol{\beta}}{\sigma\|\mathbf{x}_j\|}, 1\right), \qquad j = 1, \ldots, p \tag{5.13}$$

Averaging over the noise $\boldsymbol{\varepsilon}$ and applying (5.11), we obtain

$$\mathbb{E}[F_\ell(\mathbf{y}) \mid \boldsymbol{\beta}] = \sum_{j=1}^{p} (\mathbf{x}_j^\top \mathbf{X}\boldsymbol{\beta})^\ell = \left\langle \boldsymbol{\beta}^{\otimes \ell}, \sum_{j=1}^{p} (\mathbf{X}^\top \mathbf{X} e_j)^{\otimes \ell} \right\rangle = \left\langle \boldsymbol{\beta}^{\otimes \ell}, T^{(\ell)} \right\rangle. \tag{5.14}$$

where $T^{(\ell)}$ is the tensor defined in (5.4). Further averaging over $\boldsymbol{\beta}$ yields

$$\mathbb{E}[F_\ell(\mathbf{y})] = \left\langle M^{(\ell)}, T^{(\ell)} \right\rangle, \tag{5.15}$$

where

$$M^{(\ell)} := \mathbb{E}[\boldsymbol{\beta}^{\otimes \ell}]$$

is the degree-ℓ moment tensor of $\boldsymbol{\beta} \sim g^{\otimes p}$; equivalently, (5.15) also equals to the trace of the moment tensor of $\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}$, which no longer has iid coordinates. Since $M^{(\ell)}_{s_1,\ldots,s_p} = \mathbb{E}[\beta_{s_1} \ldots \beta_{s_p}]$, depending on the multiplicity of the indices, this can be expressed as monomials of the moments of g thanks to β_i's being iid. For example, $M^{(2)}_{i,j} = m_2$ if $i = j$ and m_1^2 otherwise. Crucially, the order-ℓ moment of g only appears on the diagonal of $M^{(\ell)}$: $M^{(\ell)}_{i,\ldots,i} = m_\ell$ for $i = 1, \ldots, p$. Furthermore, the inner product (5.15) can be expanded as

$$\mathbb{E}[F_\ell(\mathbf{y})] = A_\ell m_\ell + \sum_{t=2}^{\ell} \sum_{d_1+\ldots+d_t=\ell : d_j \geq 2} m_{d_1} \ldots m_{d_t} \sum_{s_1,\ldots,s_t=1}^{p} \widetilde{T}^{(d_1,\ldots,d_t)}_{s_1,\ldots,s_t}, \tag{5.16}$$

where $A_\ell = \sum_{i,j=1}^{p} (\mathbf{X}^\top \mathbf{X})_{ij}^\ell$ is the trace of $T^{(\ell)}$ defined in (5.6), and $\widetilde{T}^{(d_1,\ldots,d_t)}$ is the tensor defined in (5.5). In this decomposition of $\mathbb{E}[F_\ell(\mathbf{y})]$, the leading term $m_\ell A_\ell$ is desired and the remaining terms only depend on the lower-order moments. This motivates the iterative construction of the estimator (5.7), wherein we substitute the lower moments by their estimates. In other words, if the estimated lower moments in (5.7) were the true values, the estimator would be unbiased.

For analysis, the major difficulty is that, even if conditioned on $\boldsymbol{\beta}$, the p columnwise projections in (5.13) are still dependent because the columns of $\mathbf{X}$ are not orthogonal. This correlation is particularly pronounced in the sublinear regime of $p \gg n$. We present the statistical guarantees in Section 5.2.2.

Connection to Rao's MINQUE For clarity, we work out the estimate of the first two moments. It turns out that if one is content with consistently estimating the first two moments as opposed to the whole prior, the sample complexity is much lower. For example, for isotropic random design, we only need n and p to both tend to infinity individually, without requiring their relative growth.

Example 5.1 (First moment)**.** *For $\widehat{m}_1$ defined in (5.2), which is unbiased, we get*

$$\mathsf{Var}(\widehat{m}_1) = \frac{\sigma^2\|\mathbf{X}\mathbf{1}\|^2 + s\|\mathbf{X}^\top\mathbf{X}\mathbf{1}\|^2}{\|\mathbf{X}\mathbf{1}\|^4}$$

where $s = m_2 - m_1^2$ is the variance of the prior. Suppose that $\mathbf{X}$ is isotropic Gaussian with iid $N(0, \frac{1}{p})$ entries. Then with high probability, $\|\mathbf{X}\mathbf{1}\| \asymp \sqrt{n}$, and furthermore, $\|\mathbf{X}\|_{\mathrm{op}} \lesssim 1 + \sqrt{\frac{n}{p}}$, so that $\|\mathbf{X}^\top\mathbf{X}\mathbf{1}\| \lesssim \sqrt{n}(1 + \sqrt{\frac{n}{p}})$. This leads to $\mathsf{Var}(\widehat{m}_1) = O_{\mathbb{P}}(\frac{1}{\min(n,p)})$.

Example 5.2 (Second moment)**.** *For simplicity, assuming that the true prior is centered so $m_1 = 0$. Then (5.7) specializes to:*

$$\widehat{m}_2 = \frac{\|\mathbf{X}^\top\mathbf{y}\|^2 - \|\mathbf{X}\|_{\mathrm{F}}^2}{\|\mathbf{X}^\top\mathbf{X}\|_{\mathrm{F}}^2} \tag{5.17}$$

which is again unbiased. Note that $\|X^\top\mathbf{y}\|^2 = \|X^\top\boldsymbol{\beta}\|^2 + \|X^\top\boldsymbol{\varepsilon}\|^2 + 2\left\langle \mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}, \mathbf{X}^\top\boldsymbol{\varepsilon}\right\rangle$.

Recalling that $\mathbf{H} = \mathbf{X}^\top\mathbf{X}$*, we have* $\mathsf{Var}(\|\mathbf{X}^\top\mathbf{y}\|^2) = \mathsf{Var}(\|\mathbf{H}\boldsymbol{\beta}\|^2) + \mathsf{Var}(\|X^\top\boldsymbol{\varepsilon}\|^2) + 4\mathsf{Var}(\langle\mathbf{H}\boldsymbol{\beta}, \mathbf{X}^\top\boldsymbol{\varepsilon}\rangle)$*, because odd moments of* $\boldsymbol{\varepsilon}$ *vanish. It can be verified that*

$$\mathsf{Var}(\langle\mathbf{H}\boldsymbol{\beta}, \mathbf{X}^\top\boldsymbol{\varepsilon}\rangle) = \sigma^2 m_2 \mathrm{Tr}(\mathbf{H}^3), \quad \mathsf{Var}(\|\mathbf{X}^\top\boldsymbol{\varepsilon}\|^2) = 2\sigma^4\mathrm{Tr}(\mathbf{H}^2).$$

Finally, with $\mathbf{S} \equiv \mathbf{H}^2$*,*

$$\begin{aligned}
\mathbb{E}[\|\mathbf{H}\boldsymbol{\beta}\|^4] = &\sum_{a,b,c,d=1}^{p} S_{ab}S_{cd}\mathbb{E}[\beta_a\beta_b\beta_c\beta_d] \\
=& m_4\sum_{a=1}^{p} S_{aa}^2 + m_2^2 \sum_{a\neq c=1}^{p} S_{aa}S_{cc} + 2m_2^2\sum_{a\neq b=1}^{p} S_{ab}^2 \\
=& (m_4 - 3m_2^2) \underbrace{\sum_{a=1}^{p} S_{aa}^2}_{\leq \|\mathbf{S}\|_{\mathrm{F}}^2 = \mathrm{Tr}(\mathbf{H}^4)} + m_2^2 \underbrace{\sum_{a,c=1}^{p} S_{aa}S_{cc}}_{\mathrm{Tr}(\mathbf{H}^2)^2} + 2m_2^2 \underbrace{\sum_{a,b=1}^{p} S_{ab}^2}_{\mathrm{Tr}(\mathbf{H}^4)}
\end{aligned}$$

Applying $\mathbb{E}[\|\mathbf{H}\boldsymbol{\beta}\|^2] = m_2 \|\mathbf{H}\|_{\mathrm{F}}^2$*, we get*

$$\mathsf{Var}(\|\mathbf{H}\boldsymbol{\beta}\|^2) \leq (m_4 + 5m_2^2)\mathrm{Tr}(\mathbf{H}^4).$$

Overall, we get

$$\begin{aligned}
\mathsf{Var}(\widehat{m}_2) =& \frac{(m_4 - 3m_2^2)\sum_{a=1}^{p} S_{aa}^2 + 2m_2^2\mathrm{Tr}(\mathbf{H}^4) + 4\sigma^2 m_2\mathrm{Tr}(\mathbf{H}^3) + 2\sigma^4\mathrm{Tr}(\mathbf{H}^2)}{\mathrm{Tr}(\mathbf{H}^2)^2} && (5.18) \\
\leq& \frac{(m_4 + 5m_2^2)\mathrm{Tr}(\mathbf{H}^4) + 4\sigma^2 m_2\mathrm{Tr}(\mathbf{H}^3) + 2\sigma^4\mathrm{Tr}(\mathbf{H}^2)}{\mathrm{Tr}(\mathbf{H}^2)^2} && (5.19)
\end{aligned}$$

Consider again an isotropic Gaussian design $\mathbf{X}$*, in which case with high probability* $\mathrm{Tr}(\mathbf{H}^2) \asymp n + \frac{n^2}{p}$*. Consider two cases: (a) If* $n \lesssim p$*, then all singular values of* $\mathbf{X}$ *and hence* $\mathbf{H}$ *are* $O(1)$*, so* $\mathrm{Tr}(H^k) \lesssim_k n$ *and hence* $\mathsf{Var}(\widehat{m}_2) \lesssim \frac{1}{n}$*. (b) If* $n \gg p$*, then all singular values of* $\mathbf{X}$ *are at most* $\sqrt{n/p}$*. Hence* $|\mathrm{Tr}(\mathbf{H}^k)| \lesssim (n/p)^k p$*. Thus,* $\mathsf{Var}(\widehat{m}_2) \lesssim \frac{1}{p}$*. So we again have* $\mathsf{Var}(\widehat{m}_2) = O_{\mathbb{P}}(\frac{1}{\min(n,p)})$*.*

5.2.2 Statistical guarantees

We first bound the estimation error of first two moments based on simple bias and variance calculations, as shown by the following result proved in Section 5.4.1. Recall that $A_1 = \|\mathbf{X}\mathbf{1}\|^2$ and $A_2 = \|\mathbf{X}^\top \mathbf{X}\|_F^2$ as defined by (5.6).

Proposition 5.1. *Suppose that for some fixed $c_0 \in (0,1)$ it holds that*

$$\frac{\|\mathbf{X}\|_{\mathrm{op}}}{\sigma} \geq c_0, \quad \|\mathbf{X}\mathbf{1}\|^2 \geq c_0(n \wedge p)\|\mathbf{X}\|_{\mathrm{op}}^2, \quad \|\mathbf{X}^\top \mathbf{X}\|_F^2 \geq c_0(n \wedge p)\|\mathbf{X}\|_{\mathrm{op}}^4. \tag{5.20}$$

Then for some $C > 0$ that depends on (c_0, M), it holds that

$$\mathbb{E}(\widehat{m}_1 - m_1)^2 + \mathbb{E}(\widehat{\mu}_2 - \mu_2)^2 \leq \frac{C}{n \wedge p}.$$

Consequently, under the condition (5.20), $(\widehat{m}_1, \widehat{\mu}_2)$ is consistent for (m_1, μ_2) as long as $n \wedge p \to \infty$ without imposing on their relative growth. We note that in (5.20), the condition on A_1 is identical to the general Assumption 5.1 below, but the one for A_2 is slightly stronger ($n \wedge p$ versus $n^2/p \wedge p$). It can be readily checked that this stronger assumption is still satisfied with high probability by the random assemble of Lemma 5.1 below, and is essential to achieve the minimal consistency condition $n \wedge p \to \infty$.

Next we state the MSE bound for estimating higher central moments μ_k for $k \geq 3$. We make the following assumptions on the design matrix $\mathbf{X}$ and the prior distribution.

Assumption 5.1 (Design matrix)**.** *There exists some fixed $c_0 > 0$ such that $\|\mathbf{X}\|_{\mathrm{op}}/\sigma \geq c_0$ and*

$$A_k = \sum_{i,j=1}^{p} \mathbf{H}_{ij}^k \geq c_0^k \|\mathbf{X}\|_{\mathrm{op}}^{2k} \cdot p\left(\frac{n}{p} \wedge 1\right)^k. \tag{5.21}$$

Assumption 5.2 (Prior distribution)**.** *We assume $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ has i.i.d. entries such* *that $\beta_1 - \mathbb{E}[\beta_1]$ is M-sub-Gaussian*[1] *for some $M \geq 1$.*

Let us pause to discuss Assumption 5.1. As shown in (5.16), A_k is the leading coefficient of the tensor product $\langle M^{(\ell)}, T^{(\ell)} \rangle$, and represents the signal strength of the problem. The following lemma shows that Assumption 5.1 is satisfied with high probability by a wide class of random design matrices, which includes isotropic Gaussian design ($X_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \frac{1}{p})$) as a special case. Its proof is given in Section 5.4.4

Lemma 5.1. *Suppose the row of $\mathbf{X}$ is iid drawn from $\mathcal{N}(0, \frac{1}{p}\Sigma)$, where $\Sigma \in \mathbb{R}^{p\times p}$ is a PSD matrix and well-conditioned in the sense $cI \prec \Sigma \prec c^{-1}I$ where $c > 0$ is some absolute constant. Then, the design matrix $\mathbf{X}$ satisfies Assumption 5.1.*

Equipped with Assumptions 5.1 and 5.2, we are ready to state the following MSE bound of $\widehat{\mu}_\ell$ defined in (5.7), whose proof is given in Section 5.4.2.

Theorem 5.1. *Suppose Assumptions 5.1 and 5.2 hold with some constants c_0, M. For any $k \geq 3$, it holds that*

$$\mathbb{E}[(\widehat{\mu}_k - \mu_k)^2] \leq \left(\frac{CMk}{c_0^2}\right)^{k^3} \frac{1}{n \wedge p}\left(\frac{p}{n} \vee 1\right)^{2k^2}, \tag{5.22}$$

where $C > 0$ is universal.

Finally, we convert the statistical guarantee on estimating the moments to that of estimating the prior. If $n = p^{1-o(1)}$, the moment-based estimator is consistent under the 1-Wasserstein loss.[2]

Corollary 5.1. *Let $n = p^{1-\epsilon}$. Let $\widehat{g}$ be the DMM estimator (5.9) applied to the L moments*

1. Recall that a random variable X is called sub-Gaussian with constant M if $(\mathbb{E}|X - \mathbb{E}[X]|^p)^{1/p} \leq M\sqrt{p}$ for all positive integers p.

2. The W_1 distance between one-dimensional distributions is the L_1-distance between their cumulative distribution functions.

$\widehat{m}_1, \ldots, \widehat{m}_L$ *estimated by the EBMoM, where*

$$L = c \min\left\{\frac{1}{\sqrt{\epsilon}}, \frac{(\log n)^{1/3}}{\log\log n}\right\}, \quad M = c\sqrt{L}$$

for some universal constant c*. Under the assumption of Theorem 5.1,*

$$\mathbb{E}[W_1(\widehat{g}, g)] \leq \frac{C}{\sqrt{L}}$$

where C *is a constant depending only on* σ *and the subgaussian constant* M *of* g*.*

5.2.3 Computation of the EB-MoM estimator

We discuss the computational details of the moment estimates. Naively implementing the form (5.7) (i.e., compute $\sum_{s_1,\ldots,s_t=1}^{p} \widetilde{T}_{s_1,\ldots,s_t}^{(d_1,\ldots,d_t)}$ when $t = \ell$) suggests a time complexity of $O(p^\ell)$. It turns out that it can be computed in time linear in p, similar to the calculation of U-statistics. Taking into account of the computation of F_ℓ's, the overall time complexity is $O_k(np)$.

To compute $\widehat{\mu}_k$, note that for any $k \geq 2$,

$$\begin{aligned}
&\sum_{d_1+\ldots+d_t=k:d_j\geq 1} \widehat{\mu}^\circ_{d_1} \ldots \widehat{\mu}^\circ_{d_t} \sum_{s_1,\ldots,s_t=1}^{p} \widetilde{T}_{s_1,\ldots,s_t}^{(d_1,\ldots,d_t)} \\
&= \sum_{f_1+\ldots+f_t=k:f_j\geq 1} \widehat{\mu}^\circ_{f_1} \ldots \widehat{\mu}^\circ_{f_t} \Bigg[\sum_{\{d_1,\ldots,d_t\}=\{f_1,\ldots,f_t\}} \sum_{s_1,\ldots,s_t=1}^{p} \widetilde{T}_{s_1,\ldots,s_t}^{(d_1,\ldots,d_t)} \Bigg],
\end{aligned}$$

where the summation $\{d_1, \ldots, d_t\} = \{f_1, \ldots, f_t\}$ is over all ordered $d_1, \ldots, d_t$ with configuration $\{f_1, \ldots, f_t\}$. Moreover, we have

$$\begin{aligned}
\sum_{\{d_1,\ldots,d_t\}=\{f_1,\ldots,f_t\}} \sum_{s_1,\ldots,s_t=1}^{p} \widetilde{T}_{s_1,\ldots,s_t}^{(d_1,\ldots,d_t)} &= \binom{k}{f_1, \ldots, f_t} \sum_{\{d_1,\ldots,d_t\}=\{f_1,\ldots,f_t\}} \sum_{1\leq s_1<\ldots<s_t\leq p} \sum_{j=1}^{p} H_{j,s_1}^{d_1} \ldots H_{j,s_t}^{d_t} \\
&= \frac{\binom{k}{d_1,\ldots,d_t}}{n(f_1, \ldots, f_t)} \sum_{j=1}^{p} \sum_{s_\ell\in[p]:s_1\neq\ldots\neq s_t} H_{j,s_1}^{f_1} \ldots H_{j,s_t}^{f_t},
\end{aligned}$$

where $n(f_1, \ldots, f_t) = \prod_{j=1}^{p}(\#\{i : f_i = j\})!$ with convention $0! \equiv 1$, and we use the fact that, for any $j \in [p]$,

$$\sum_{\{d_1,\ldots,d_t\}=\{f_1,\ldots,f_t\}} \sum_{1\le s_1<\ldots<s_t\le p} H_{j,s_1}^{d_1} \ldots H_{j,s_t}^{d_t} = \frac{1}{n(f_1,\ldots,f_t)} \sum_{s_\ell\in[p]:s_1\neq\ldots\neq s_t} H_{j,s_1}^{f_1} \ldots H_{j,s_t}^{f_t}.$$

Note that by using inclusion-exclusion, we may compute $\sum_{j=1}^{p} \sum_{s_\ell\in[p]:s_1\neq\ldots\neq s_t} H_{j,s_1}^{f_1} \ldots H_{j,s_t}^{f_t}$ with time-complexity $O_k(p)$ instead of $O(p^k)$. A better approach is given in Algorithm 1.

We provide efficient algorithms to compute our estimators and present simulation experiments for various designs. In particular, we focus on computing terms in the form

$$\sum_{s_\ell\in[p]:s_1\neq\ldots\neq s_t} x_{s_1}^{f_1} \ldots x_{s_t}^{f_t},$$

for which an naive method has complexity $O(p^t)$. An improved algorithm is to use the sieve method in combinatorics (see e.g. [Sta11, Chapter 2]) and the summation above can be rewritten as

$$\sum_{s_\ell\in[p]:s_1\neq\ldots\neq s_t} x_{s_1}^{f_1} \ldots x_{s_t}^{f_t} = \sum_{\pi\in\mathcal{P}_t} \mu(\pi) \prod_{B\in\pi} \Big(\sum_{i=1}^{p} x_i^{\sum_{j\in B} f_j} \Big),$$

where $\mathcal{P}_t$ consists of all partitions of the set $[t]$, and $\mu(\pi)$ is the Möbius function, defined as

$$\mu(\pi) := \prod_{B\in\pi} \mu(B), \;\; \mu(B) := (-1)^{|B|}(|B|-1)!.$$

Computing the right-hand side requires complexity $O(tB_t p)$, where B_t is the t-th Bell number representing the number of all set partitions of size t and B_t grows superexponentially.

A further optimized algorithm is based on generating function. Consider the polynomial

$$\prod_{i=1}^{p} \Big(1 + \sum_{j=1}^{t} y_j x_i^{f_j}\Big),$$

we observe that the coefficient for the term $y_1 y_2 \dots y_t$ is exactly $\sum_{s_\ell \in [p]: s_1 \neq \dots \neq s_t} x_{s_1}^{f_1} \dots x_{s_t}^{f_t}$. To compute this coefficient, a standard dynamic programming approach will be used, in which we have 2^t states representing if y_j has been chosen. The algorithm is stated as follows. Since there are only 2^t states in the program, this algorithm has time complexity $O(t 2^t p)$, which is faster than the sieve method.

Algorithm 1 Dynamic Programming for Extracting the Coefficient of $y_1 y_2 \cdots y_t$

Input: A list $x = \{x_1, \dots, x_p\}$ and an exponent vector $d = (d_1, \dots, d_t)$.

Output: The coefficient of $y_1 y_2 \cdots y_t$ in $F(y_1, \dots, y_t) = \prod_{i=1}^{p} \Big(1 + \sum_{j=1}^{t} y_j \, x_i^{d_j}\Big)$.

$N \leftarrow 2^t$ ▷ N is the total number of states (bitmask representation)

Initialize $\texttt{dp}[0, \dots, N-1]$ with $\texttt{dp}[0] \leftarrow 1$ and $\texttt{dp}[mask] \leftarrow 0$ for $mask \neq 0$

for $i = 1$ to p **do**

    $\texttt{new\_dp} \leftarrow \texttt{dp}$

    for each $mask \in \{0, \dots, N-1\}$ **do**

        for $j = 0$ to $t-1$ **do**

            if the jth bit of $mask$ is 0 **then**

                $new\_mask \leftarrow mask \mid (1 \ll j)$

                $\texttt{new\_dp}[new\_mask] \mathrel{+}= \texttt{dp}[mask] \cdot x_i^{d_{j+1}}$

            end if

        end for

    end for

    $\texttt{dp} \leftarrow \texttt{new\_dp}$

end for

return $\texttt{dp}[N-1]$

5.3 Lower bounds and sharp sample complexity

In this section, we provide statistical lower bound for learning the prior that leads to matching sample complexity bounds and certifies the optimality of the method-of-moments estimator.

Let us start by explaining the major distinction between the linear model and the sequence model from a deconvolution point of view. As an instructive example, we consider the following *block design* where $p = nm$ for some integer $m \geq 1$, and each of the n rows has exactly m non-zeros all equal to $\frac{1}{\sqrt{m}}$ and disjoint support:

$$\mathbf{X} = \frac{1}{\sqrt{m}} \begin{bmatrix} 1 \cdots 1\, 0 \cdots 0 \cdots 0 \\ 0 \cdots 0\, 1 \cdots 1 \cdots 0 \\ \ddots \\ 0 \cdots 0\, 0 \cdots 0\, 1 \cdots 1 \end{bmatrix} \tag{5.23}$$

Thus, $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta}$ has iid coordinates drawn from

$$g^{(m)} := \text{Law}\left(\frac{\beta_1 + \ldots + \beta_m}{\sqrt{m}}\right), \quad \beta_i \overset{\text{i.i.d.}}{\sim} g, \tag{5.24}$$

which the m-fold convolution of g with itself rescaled by $1/\sqrt{m}$, and $\mathbf{y} = \boldsymbol{\theta} + \epsilon$ has iid coordinates drawn from $g^{(m)} * \mathcal{N}(0, \sigma^2)$. As such, the problem reduces to a sequence model, where $g^{(m)}$ plays the role of the prior. As such, in order to learn the prior g, one needs to deconvolve g from both the noise distribution and itself. This new phenomenon of *self-deconvolution* is a major distinction between linear models and sequence models.

For dense design, the situation is more complicated as all p regression coefficients participated in the self-convolution and contributes to all responses; nevertheless, this simple example of block design is already enough to demonstrate the sample complexity lower bound $n = p^{1-o(1)}$. To see this, the main observation is that, due to the Central Limit Theorem (CLT), the L-fold self-convolution $g^{(L)}$ in (5.24) approaches standard normal as L grows. The typical convergence rate (Berry-Esseen) in CLT is $\frac{1}{\sqrt{m}}$, but this can be sped up arbitrarily by choosing g to have many matching moments with the normal limit. The same idea can be used to construct a pair of priors g, g' matching a constant L number of moments such that the moments of their self-convolutions $g^{(m)}, g'^{(m)}$ are even closer, leading to the indistinguishabilty of the Gaussian convolutions $g^{(m)} * \mathcal{N}$ and $g'^{(m)} * \mathcal{N}$. This is formalized by the next result proved in Section 5.5.1.

Proposition 5.2. *Consider the block design (5.23), with $m = \frac{p}{n} \in \mathbb{N}$. For every $L \geq 2$, there exists a pair of distinct priors g, g' on $[-1, 1]$, such that*

$$\mathsf{TV}(P_g(\mathbf{y}), P_{g'}(\mathbf{y})) \leq Cnm^{-(L-1)}.$$

where $P_g(\mathbf{y}), P_{g'}(\mathbf{y})$ denotes the law of $\mathbf{y}$ (conditioning on $\mathbf{X}$) under prior g, g' respectively, and C is a universal constant.

Proposition 5.2 provides a (Le Cam) two-point lower bound for estimating the prior. Suppose that $n \leq p^{1-\delta}$ for any constant δ. We can choose $g \neq g'$ with $L = 2/\delta$, such that $\mathsf{TV}(P_g(\mathbf{y}), P_{g'}(\mathbf{y})) = o(1)$; that is, one cannot distinguish whether g or g' is the true prior based on $\mathbf{y}$. This shows that for the block design, consistent estimation of the prior requires sample size at least $n = p^{1-o(1)}$.

The following theorem extends this lower bound to general designs. Compared with the block design example in Proposition 5.2 , the major technical difficulty is that $\mathbf{y}$ no longer has a product law. See Section 5.5.2 for a proof.

Theorem 5.2. *Suppose that $n \leq p^{1-\delta}$ for some fixed $\delta > 0$ and $\max_{j \in [p]} \|\mathbf{x}_j\| \leq C_0 (\frac{n}{p})^{\alpha} (\log p)^{\gamma}$ for some fixed $C_0, \alpha, \gamma > 0$. Then there exist a pair of distinct distributions g, g' supported* *on* $[0,1]$ *such that, for all large enough* n, p*,*

$$d_{\mathrm{TV}}(P_g(\mathbf{y}), P_{g'}(\mathbf{y})) \leq 0.1.$$

To apply this result, consider the random design in Lemma 5.1, where the rows $\mathbf{X}_1, \dots, \mathbf{X}_n$ are centered, independent, and can be written as $\mathbf{X}_i = \Sigma^{1/2}\mathbf{Z}_i/\sqrt{p}$, with $\{\mathbf{Z}_i\}_{i=1}^n$ satisfying a log-Sobolev inequality with constant C_{LSI}. A simple example is (correlated) Gaussian design, with rows drawn independently from $N(0, \frac{1}{p}\Sigma)$. Then with probability converging to 1, it holds that $\max_{j\in[p]}\|\mathbf{x}_j\| \leq C(\sqrt{n} + \sqrt{\log p})/\sqrt{p}$ where $C > 0$ only depends on $(\|\Sigma\|_{\mathrm{op}}, C_{LSI})$, so the condition of Theorem 5.2 is satisfied with $\alpha = \gamma = 1/2$. This shows that for such random design, the sample complexity for estimating the prior is also $n = p^{1-o(1)}$, which matches .

Sample complexity for non-trivial Bayes error Our main result shows that under mild assumptions on the design (such as random design), the prior can be estimtated consistently with sublinear sample size $n \ll p$ and the optimal sample complexity for this is $n = p^{1-o(1)}$. In comparison, even if the prior is known, linear sample complexity is still necessary in order to achieve a non-trivial estimation error. This follows from applying the mutual information method by observing that for any estimator $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\mathbf{X}, \mathbf{y})$, the data processing inequality for mutual information gives

$$I(\boldsymbol{\beta}; \widehat{\boldsymbol{\beta}}) \leq I(\boldsymbol{\beta}; \mathbf{X}, \mathbf{y}).$$

For any prior with non-zero variance σ^2, we know from rate-distortion theory that the left side must be $\Omega(p)$ whenever $\widehat{\boldsymbol{\beta}}$ has non-trivial variance reduction, namely, $\mathbb{E}[\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\|^2] \leq p(\sigma^2 - \Omega(1))$. On the right side, we have $I(\boldsymbol{\beta}; \mathbf{X}, \mathbf{y}) = I(\mathbf{X}\boldsymbol{\beta}; \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}|\mathbf{X}) \leq \frac{n}{2}\log(1 + \frac{\mathbb{E}[\|\mathbf{X}(\boldsymbol{\beta} - \mathbb{E}[\boldsymbol{\beta}])\|_2^2]}{n}) = \frac{n}{2}\log(1 + \frac{\sigma^2\|\mathbf{X}\|_{\mathrm{F}}^2}{n})$, by applying the Gaussian channel capacity formula. Thus, we must have $\|\mathbf{X}\|_{\mathrm{F}}^2 = \Omega(p)$, which requires $n = \Omega(p)$ when the rows of the design matrix $\mathbf{X}$ have bound norm (e.g. random design).

When n is indeed at least proportional to p, the method of moments in Theorem 5.1 yields a consistent estimator of π which allows methods such as Bayesian AMP to achieve near-Bayes estimation error.

5.4 Proof of upper bounds

5.4.1 Proof of Proposition 5.1

Simple calculations yields that $\widehat{m}_1$ is unbiased and

$$\mathsf{Var}(\widehat{m}_1) = \frac{\sigma^2\|\mathbf{X}\mathbf{1}\|^2 + s\|\mathbf{X}^\top\mathbf{X}\mathbf{1}\|^2}{\|\mathbf{X}\mathbf{1}\|^4}$$

where $s = m_2 - m_1^2$ is the variance of the prior. Using $\|\mathbf{X}^\top\mathbf{X}\mathbf{1}\|^2 \le \|\mathbf{X}\|_{\mathrm{op}}^2\|\mathbf{X}\mathbf{1}\|^2$ and $\|\mathbf{X}\|_{\mathrm{op}} \ge c\sigma$, we have

$$\mathsf{Var}(\widehat{m}_1) \le \|\mathbf{X}\|_{\mathrm{op}}^2\left(s + \frac{\sigma^2}{\|\mathbf{X}\|_{\mathrm{op}}}\right)\frac{1}{\|\mathbf{X}\mathbf{1}\|^2} \le C\frac{\|\mathbf{X}\|_{\mathrm{op}}^2}{\|\mathbf{X}\mathbf{1}\|^2},$$

and the conclusion follows from the condition on $\|\mathbf{X}\mathbf{1}\|^2$.

Next we bound $\mathbb{E}[(\widehat{\mu}_2 - \mu_2)^2]$. Using $\bar{\mu}_2 = F_2(\bar{\mathbf{y}}) = \frac{\|\mathbf{X}^\top\bar{\mathbf{y}}\|^2 - \|\mathbf{X}\|_{\mathrm{F}}^2}{\|\mathbf{X}^\top\mathbf{X}\|_{\mathrm{F}}^2}$ with $\bar{\mathbf{y}} = y - m_1\mathbf{1}$ as an intermediary, we have

$$\mathbb{E}[(\widehat{\mu}_2 - \mu_2)^2] \le 2\underbrace{\mathbb{E}[(\bar{\mu}_2 - \mu_2)^2]}_{(I)} + 2\underbrace{\mathbb{E}[(\widehat{\mu}_2 - \bar{\mu}_2)^2]}_{(II)}.$$

To bound (I), note that $\mathbb{E}[\bar{\mu}_2] = \mu_2$, and $\|\mathbf{X}^\top\bar{\mathbf{y}}\|^2 = \|\mathbf{X}^\top\bar{\boldsymbol{\beta}}\|^2 + \|\mathbf{X}^\top\boldsymbol{\varepsilon}\|^2 + 2\left\langle\mathbf{X}^\top\mathbf{X}\bar{\boldsymbol{\beta}}, \mathbf{X}^\top\boldsymbol{\varepsilon}\right\rangle$. Recalling that $\mathbf{H} = \mathbf{X}^\top\mathbf{X}$, we have $\mathsf{Var}(\|\mathbf{X}^\top\bar{\mathbf{y}}\|^2) = \mathsf{Var}(\|\mathbf{H}\bar{\boldsymbol{\beta}}\|^2) + \mathsf{Var}(\|\mathbf{X}^\top\boldsymbol{\varepsilon}\|^2) +$

$4\mathsf{Var}(\langle \mathbf{H}\bar{\boldsymbol{\beta}}, \mathbf{X}^\top \boldsymbol{\varepsilon}\rangle)$, because odd moments of ε vanish. It can be verified that

$$\mathsf{Var}\Big(\langle \mathbf{H}\bar{\boldsymbol{\beta}}, \mathbf{X}^\top \boldsymbol{\varepsilon}\rangle \Big) = \sigma^2 \mu_2 \mathrm{Tr}(\mathbf{H}^3), \quad \mathsf{Var}(\|\mathbf{X}^\top \boldsymbol{\varepsilon}\|^2) = 2\sigma^4 \mathrm{Tr}(\mathbf{H}^2).$$

Finally, with $\mathbf{S} \equiv \mathbf{H}^2$,

$$\begin{aligned}
\mathbb{E}[\|\mathbf{H}\bar{\boldsymbol{\beta}}\|^4] = &\sum_{a,b,c,d=1}^{p} S_{ab} S_{cd} \mathbb{E}[\bar{\beta}_a \bar{\beta}_b \bar{\beta}_c \bar{\beta}_d] \\
=&\mu_4 \sum_{a=1}^{p} S_{aa}^2 + \mu_2^2 \sum_{a\neq c=1}^{p} S_{aa} S_{cc} + 2\mu_2^2 \sum_{a\neq b=1}^{p} S_{ab}^2 \\
=&(\mu_4 - 3\mu_2^2) \underbrace{\sum_{a=1}^{p} S_{aa}^2}_{\leq \|\mathbf{S}\|_{\mathrm{F}}^2 = \mathrm{Tr}(\mathbf{H}^4)} + \mu_2^2 \underbrace{\sum_{a,c=1}^{p} S_{aa} S_{cc}}_{\mathrm{Tr}(\mathbf{H}^2)^2} + 2\mu_2^2 \underbrace{\sum_{a,b=1}^{p} S_{ab}^2}_{\mathrm{Tr}(\mathbf{H}^4)} .
\end{aligned}$$

Applying $\mathbb{E}[\|\mathbf{H}\bar{\boldsymbol{\beta}}\|^2] = \mu_2 \|\mathbf{H}\|_{\mathrm{F}}^2$, we get

$$\mathsf{Var}(\|\mathbf{H}\bar{\boldsymbol{\beta}}\|^2) = (\mu_4 - 3\mu_2^2) \sum_{a=1}^{p} S_{aa}^2 + 2\mu_2^2 \,\mathrm{Tr}(\mathbf{H}^4) \leq (\mu_4 - \mu_2^2)\, \mathrm{Tr}(\mathbf{H}^4).$$

Overall, we get

$$\begin{aligned}
\mathsf{Var}(\bar{\mu}_2) =& \frac{(\mu_4 - 3\mu_2^2) \sum_{a=1}^{p} S_{aa}^2 + 2\mu_2^2 \mathrm{Tr}(\mathbf{H}^4) + 4\sigma^2 \mu_2 \mathrm{Tr}(\mathbf{H}^3) + 2\sigma^4 \mathrm{Tr}(\mathbf{H}^2)}{\mathrm{Tr}(\mathbf{H}^2)^2} && (5.25)\\
\leq& C \frac{\mathrm{Tr}(\mathbf{H}^4) + \sigma^2 \mathrm{Tr}(\mathbf{H}^3) + \sigma^4 \mathrm{Tr}(\mathbf{H}^2)}{\mathrm{Tr}(\mathbf{H}^2)^2} \leq C \frac{\mathrm{Tr}(\mathbf{H}^4) + \sigma^4 \,\mathrm{Tr}(\mathbf{H}^2)}{\mathrm{Tr}(\mathbf{H}^2)^2}, && (5.26)
\end{aligned}$$

where in the last step we apply $\mathrm{Tr}(\mathbf{H}^3)^2 \leq \|\mathbf{H}\|_F^2 \|\mathbf{H}^2\|_F^2 = \mathrm{Tr}(\mathbf{H}^2)\,\mathrm{Tr}(\mathbf{H}^4)$. Now applying $\mathrm{Tr}(\mathbf{H}^4) \leq (n \wedge p) \|\mathbf{X}\|_{\mathrm{op}}^8$ and $\mathrm{Tr}(\mathbf{H}^2) \leq (n \wedge p) \|\mathbf{X}\|_{\mathrm{op}}^4$ along with the lower bound of $\|\mathbf{H}\|_F^2$ in (5.20), we have

$$(I) = \mathsf{Var}(\bar{\mu}_2) \leq C \frac{\|\mathbf{X}\|_{\mathrm{op}}^4 \Big(1 + \frac{\sigma^4}{\|\mathbf{X}\|_{\mathrm{op}}^4}\Big)(n \wedge p)}{\|\mathbf{H}\|_F^4} \leq \frac{C}{n \wedge p}.$$

Next to bound (II), using $\widehat{\mathbf{y}}-\bar{\mathbf{y}}=-(\widehat{m}_1-m_1)\mathbf{X}\mathbf{1}$, we have

$$\begin{aligned}\widehat{\mu}_2-\bar{\mu}_2&=\frac{\|\mathbf{X}^\top\widehat{\mathbf{y}}\|^2-\|\mathbf{X}^\top\bar{\mathbf{y}}\|^2}{\|\mathbf{X}^\top\mathbf{X}\|_F^2}=\frac{\langle\mathbf{X}^\top(\widehat{\mathbf{y}}-\bar{\mathbf{y}}),\mathbf{X}^\top(\widehat{\mathbf{y}}+\bar{\mathbf{y}})\rangle}{\|\mathbf{X}^\top\mathbf{X}\|_F^2}\\&=-(\widehat{m}_1-m_1)\Big[\frac{\langle\mathbf{X}^\top\mathbf{X}\mathbf{1},\mathbf{X}^\top(\mathbf{X}(2\boldsymbol{\beta}-(\widehat{m}_1+m_1)\mathbf{1})+\boldsymbol{\varepsilon})\rangle}{\|\mathbf{X}^\top\mathbf{X}\|_F^2}\Big],\end{aligned}$$

which implies that $(II)=\mathbb{E}[(\widehat{\mu}_2-\bar{\mu}_2)^2]\le 2((II_1)+(II_2))$, where

$$\begin{aligned}(II_I)&=\frac{1}{\|\mathbf{X}^\top\mathbf{X}\|_F^4}\mathbb{E}\Big[(\widehat{m}_1-m_1)^2\langle\mathbf{X}^\top\mathbf{X}\mathbf{1},\mathbf{X}^\top\boldsymbol{\varepsilon}\rangle^2\Big],\\(II_2)&=\frac{1}{\|\mathbf{X}^\top\mathbf{X}\|_F^4}\mathbb{E}\Big[(\widehat{m}_1-m_1)^2\langle\mathbf{X}^\top\mathbf{X}\mathbf{1},\mathbf{X}^\top\mathbf{X}(2\boldsymbol{\beta}-(\widehat{m}_1+m_1)\mathbf{1})\rangle^2\Big].\end{aligned}$$

To bound (II_1), we use $\widehat{m}_1-m_1=\frac{\mathbf{1}^\top\mathbf{X}^\top\bar{\mathbf{y}}}{\|\mathbf{X}\mathbf{1}\|^2}$ to get

$$\begin{aligned}(II_2)&=\frac{1}{\|\mathbf{X}^\top\mathbf{X}\|_F^4\|\mathbf{X}\mathbf{1}\|^4}\mathbb{E}\Big[(\mathbf{1}^\top\mathbf{X}\bar{\mathbf{y}})^2\langle\mathbf{X}^\top\mathbf{X}\mathbf{1},\mathbf{X}^\top\boldsymbol{\varepsilon}\rangle^2\Big]\\&\le\frac{1}{\|\mathbf{X}^\top\mathbf{X}\|_F^4\|\mathbf{X}\mathbf{1}\|^4}\mathbb{E}^{1/2}[(\mathbf{1}^\top\mathbf{X}\bar{\mathbf{y}})^4]\mathbb{E}^{1/2}\Big[\langle\mathbf{X}^\top\mathbf{X}\mathbf{1},\mathbf{X}^\top\boldsymbol{\varepsilon}\rangle^4\Big].\end{aligned}$$

Since $\mathbf{1}^\top\mathbf{X}^\top\bar{\mathbf{y}}=\mathbf{1}^\top\mathbf{X}^\top\mathbf{X}\bar{\boldsymbol{\beta}}+\mathbf{1}^\top\mathbf{X}^\top\boldsymbol{\varepsilon}$, where $\mathbf{1}^\top\mathbf{X}^\top\mathbf{X}\bar{\boldsymbol{\beta}}$ is sub-Gaussian with constant $2M\|\mathbf{X}^\top\mathbf{X}\mathbf{1}\|$ and $\mathbf{1}^\top\mathbf{X}^\top\boldsymbol{\varepsilon}\sim\mathcal{N}(0,\sigma^2\|\mathbf{X}\mathbf{1}\|^2)$, we have $\mathbb{E}[(\mathbf{1}^\top\mathbf{X}\bar{\mathbf{y}})^4]\le C(M^4\|\mathbf{X}^\top\mathbf{X}\mathbf{1}\|^4+\sigma^4\|\mathbf{X}\mathbf{1}\|^4)$. Similarly, $\langle\mathbf{X}^\top\mathbf{X}\mathbf{1},\mathbf{X}^\top\boldsymbol{\varepsilon}\rangle\sim\mathcal{N}(0,\sigma^2\|\mathbf{X}\mathbf{X}^\top\mathbf{X}\mathbf{1}\|^2)$, so $\mathbb{E}[\langle\mathbf{X}^\top\mathbf{X}\mathbf{1},\mathbf{X}^\top\boldsymbol{\varepsilon}\rangle^4]\le C\sigma^4\|\mathbf{X}\mathbf{X}^\top\mathbf{X}\mathbf{1}\|^4$. Combining the estimates yields

$$\begin{aligned}(II_1)&\le C\frac{M^2\sigma^2\|\mathbf{X}^\top\mathbf{X}\mathbf{1}\|^2\|\mathbf{X}\mathbf{X}^\top\mathbf{X}\mathbf{1}\|^2+\sigma^4\|\mathbf{X}\mathbf{1}\|^2\|\mathbf{X}\mathbf{X}^\top\mathbf{X}\mathbf{1}\|^2}{\|\mathbf{X}^\top\mathbf{X}\|_F^4\|\mathbf{X}\mathbf{1}\|^4}\\&\le C\frac{\|\mathbf{X}\|_{\mathrm{op}}^6+\|\mathbf{X}\|_{\mathrm{op}}^4}{\|\mathbf{X}^\top\mathbf{X}\|_F^4}\le\frac{C}{(n\wedge p)^2},\end{aligned}$$

where we apply the lower bound of $\|\mathbf{X}^\top\mathbf{X}\|_F^2$ in (5.20). A similar argument yields the same bound for (II_2), concluding that $(II)\le\frac{C}{(n\wedge p)^2}$. Combining the bounds of (I) and (II) concludes the proof.

5.4.2 Proof of Theorem 5.1

The proof strategy consists of the following three steps:

(i) We considered the centered model $\bar{\mathbf{y}} = \mathbf{X}(\boldsymbol{\beta} - m_1\mathbf{1}) + \boldsymbol{\varepsilon}$ pretending m_1 is known. Based on observations $(\mathbf{X}, \bar{\mathbf{y}})$, we apply the recipe in (5.7) to estimate the centered moments defined as $\mu_k := \mathbb{E}[(\beta_1 - m_1)^k]$. More precisely, let $\{\bar{\mu}_\ell\}_{\ell \geq 2}$ be recursively defined by

$$\bar{\mu}_\ell := \frac{1}{A_\ell}\Big[F_\ell(\bar{\mathbf{y}}) - \sum_{t=2}^{\ell} \sum_{d_1+\ldots+d_t=\ell: d_j \geq 2} \bar{\mu}^\circ_{d_1} \ldots \bar{\mu}^\circ_{d_t} \sum_{s_1,\ldots,s_t=1}^{p} \widetilde{T}^{(d_1,\ldots,d_t)}_{s_1,\ldots,s_t}\Big], \tag{5.27}$$

where $\bar{\mu}^\circ_d = (\bar{\mu}_d \vee -2K_d) \wedge 2K_d$ with $K_d = M^d d^{d/2}$ is the truncated version of $\bar{\mu}_d$. The MSE of $\bar{\mu}_\ell$ for μ_ℓ is detailed in Section 5.4.2.1. We emphasize that $\bar{\mu}_\ell$ is not an actual estimator (for μ_ℓ) and is only used as a proof device.

(ii) Since $\widehat{m}_1$ given in (5.2) is consistent for m_1, we expect $\widehat{\mathbf{y}} = \mathbf{X}(\boldsymbol{\beta} - \widehat{m}_1\mathbf{1}) + \boldsymbol{\varepsilon}$ to be a good proxy for $\bar{\mathbf{y}}$, and therefore $\widehat{\mu}_\ell$ in (5.7), namely,

$$\widehat{\mu}_\ell := \frac{1}{A_\ell}\Big[F_\ell(\widehat{\mathbf{y}}) - \sum_{t=2}^{\ell} \sum_{d_1+\ldots+d_t=\ell: d_j \geq 2} \widehat{\mu}^\circ_{d_1} \ldots \widehat{\mu}^\circ_{d_t} \sum_{s_1,\ldots,s_t=1}^{p} \widetilde{T}^{(d_1,\ldots,d_t)}_{s_1,\ldots,s_t}\Big], \tag{5.28}$$

to be close to $\bar{\mu}_\ell$. We quantify this approximation error in Section 5.4.2.2. When combined with the previous step, this yields the MSE of $\{\widehat{\mu}_\ell\}_{\ell \geq 2}$ for $\{\mu_\ell\}_{\ell \geq 2}$.

(iii) Since $m_\ell = \sum_{k=0}^{\ell} \binom{\ell}{k} \mu_k m_1^{\ell-k}$ by the binomial formula, we can bound the MSE of $\widehat{m}_\ell$ in (5.7) via the above MSE of $\widehat{m}_1$ and $\{\widehat{\mu}_\ell\}_{\ell \geq 2}$.

To ease notations, we use μ_ℓ to denote the centered moments $\mathbb{E}[(\beta_1 - m_1)^\ell]$. With $\widehat{m}_1$ given by (5.28), let

$$\widehat{\boldsymbol{\beta}} := \boldsymbol{\beta} - \widehat{m}_1\mathbf{1}_p, \qquad \overline{\boldsymbol{\beta}} := \boldsymbol{\beta} - m_1\mathbf{1}_p. \tag{5.29}$$

5.4.2.1 MSE for centered model

The following result states the MSE of (idealized estimator) $\bar{\mu}_k$ in (5.27) for the centered moments μ_k. Recall that M denotes the sub-Gaussian constant in Assumption 5.2.

Proposition 5.3. *Suppose that Assumptions 5.1 and 5.2 hold. For any $k \geq 2$, it holds that $\mathbb{E}[(\bar{\mu}_k - \mu_k)^2] \leq \Delta_k$, where*

$$\Delta_k = \Big(\frac{CMk}{c_0}\Big)^{(k+1)^3} \frac{1}{n \wedge p} \Big(\frac{p}{n} \vee 1\Big)^{k^2}. \tag{5.30}$$

We will apply induction to prove this result, by bounding separately the bias and variance. The following result states the bias bound.

Proposition 5.4. *Suppose that the bound (5.30) holds up to some $k-1$. Then the bias of $\bar{\mu}_k$ satisfies*

$$|\mathbb{E}[\bar{\mu}_k] - \mu_k|^2 \leq \Big(\frac{2Mk}{c_0}\Big)^{5k+2} \Big(\frac{p}{n} \vee 1\Big)^{2k-2} \Delta_{k-1}. \tag{5.31}$$

Proof. First note that if $\overline{\mathbf{y}} = \boldsymbol{\theta} + \boldsymbol{\varepsilon}$ with $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 I_n)$, then for any $\mathbf{a} \in \mathbb{S}^{n-1}$, $\mathbb{E}_{\boldsymbol{\varepsilon}} H_k(\mathbf{a}^\top \overline{\mathbf{y}}/\sigma) = (\mathbf{a}^\top \boldsymbol{\theta}/\sigma)^k$. In our case, we have $\boldsymbol{\theta} = \mathbf{X}\overline{\boldsymbol{\beta}}$. Moreover, with T and $\widetilde{T}$ defined in (5.4) and (5.5), it can be readily checked that for any vector $\boldsymbol{\beta} \in \mathbb{R}^p$,

$$\mathrm{Tr}(\mathbf{X}^\top \mathbf{X} \boldsymbol{\beta})^{\otimes k} = \big\langle T^{(k)}, \boldsymbol{\beta}^{\otimes k} \big\rangle = \sum_{t=1}^{k} \sum_{d_1,\ldots,d_t = k : d_j \geq 1} \sum_{s_1,\ldots,s_t=1}^{p} \widetilde{T}^{(d_1,\ldots,d_t)}_{s_1,\ldots,s_t} \beta^{d_1}_{s_1} \ldots \beta^{d_t}_{s_t}, \tag{5.32}$$

where the summation over $(d_1, \ldots, d_t)$ is ordered. Applying the above two facts, the first term in the definition (5.7) satisfies

$$
\begin{aligned}
\mathbb{E}[F_k(\bar{\mathbf{y}})] &= \mathbb{E}_{\boldsymbol{\beta}}\mathbb{E}_{\boldsymbol{\varepsilon}}\Big[\sum_{j=1}^{p}(\sigma\|\mathbf{x}_j\|^k)H_k\Big(\frac{\mathbf{x}_j^\top\bar{\mathbf{y}}}{\sigma\|\mathbf{x}_j\|}\Big)\Big] = \mathbb{E}_{\boldsymbol{\beta}}\Big[\sum_{j=1}^{p}(\mathbf{x}_j^\top\boldsymbol{\theta})^k\Big] = \mathbb{E}_{\boldsymbol{\beta}}[\mathrm{Tr}(\mathbf{X}^\top\mathbf{X}\bar{\boldsymbol{\beta}})^{\otimes k}] \\
&= \mathbb{E}_{\boldsymbol{\beta}}\Big[\sum_{t=1}^{k}\sum_{d_1,\dots,d_t=k:d_j\geq 1}\sum_{s_1,\dots,s_t=1}^{p}\widetilde{T}_{s_1,\dots,s_t}^{(d_1,\dots,d_t)}\bar{\beta}_{s_1}^{d_1}\dots\bar{\beta}_{s_t}^{d_t}\Big] \\
&= \sum_{t=1}^{k}\sum_{d_1+\dots+d_t=k:d_j\geq 2}\gamma^{(d_1,\dots,d_t)}\mu_{d_1}\cdots\mu_{d_t},
\end{aligned}
\tag{5.33}
$$

where the $t=1$ term corresponds to $A_k\mu_k$, and $\gamma^{(d_1,\dots,d_t)} = \sum_{s_1,\dots,s_t=1}^{p}\widetilde{T}_{s_1,\dots,s_t}^{(d_1,\dots,d_t)}$. Therefore, by the definition in (5.27), we have

$$
\big|\mathbb{E}[\bar{\mu}_k]-\mu_k\big| = \frac{1}{|A_k|}\Big|\sum_{t=2}^{k}\sum_{d_1+\dots+d_t=k,d_j\geq 2}\gamma^{(d_1,\dots,d_t)}\left(\mathbb{E}[\bar{\mu}_{d_1}^\circ\dots\bar{\mu}_{d_t}^\circ]-\mu_{d_1}\dots\mu_{d_t}\right)\Big|.
$$

Note that

$$
\mathbb{E}[\bar{\mu}_{d_1}^\circ\cdots\bar{\mu}_{d_t}^\circ]-\mu_{d_1}\cdots\mu_{d_t} = \mathbb{E}\Big[\sum_{j=1}^{t}\bar{\mu}_{d_1}^\circ\cdots\bar{\mu}_{d_{j-1}}^\circ(\bar{\mu}_{d_j}^\circ-\mu_{d_j})\mu_{d_{j+1}}\cdots\mu_{d_t}\Big],
$$

and since $|\bar{\mu}_d^\circ| = |(\bar{\mu}_d\vee -2K_d)\wedge 2K_d| \leq 2K_d$ (where $K_d = M^d d^{d/2}$) and $\mu_d \leq K_d$, it holds that $\mathbb{E}[(\bar{\mu}_d^\circ-\mu_d)^2] \leq \mathbb{E}[(\bar{\mu}_d-\mu_d)^2]$. Therefore, since $d_1,\dots,d_t \leq k-1$, the induction hypothesis yields that

$$
\begin{aligned}
\big|\mathbb{E}[\bar{\mu}_{d_1}^\circ\cdots\bar{\mu}_{d_t}^\circ]-\mu_{d_1}\cdots\mu_{d_t}\big| &\leq \mathbb{E}\Big[\Big|\sum_{j=1}^{t}\bar{\mu}_{d_1}^\circ\cdots\bar{\mu}_{d_{j-1}}^\circ(\bar{\mu}_{d_j}^\circ-\mu_{d_j})\mu_{d_{j+1}}\cdots\mu_{d_t}\Big|\Big] \\
&\leq \mathbb{E}\Big[\Big(\sum_{j=1}^{t}(\bar{\mu}_{d_j}^\circ-m_{d_j})^2\Big)^{1/2}\Big(\sum_{j=1}^{t}(\bar{\mu}_{d_1}^\circ\cdots\bar{\mu}_{d_{j-1}}^\circ m_{d_{j+1}}\cdots m_{d_t})^2\Big)^{1/2}\Big] \\
&\leq (t2^{2t}M^{2k}k^k)^{\frac{1}{2}}\Big(\mathbb{E}\sum_{j=1}^{t}(\bar{\mu}_{d_j}^\circ-\mu_{d_j})^2\Big)^{1/2} \\
&\leq \sqrt{t}2^tM^kk^{k/2}\Big(\mathbb{E}\sum_{j=1}^{t}(\bar{\mu}_{d_j}-\mu_{d_j})^2\Big)^{1/2} \leq t2^tM^kk^{k/2}\Delta_{k-1}^{\frac{1}{2}}.
\end{aligned}
$$

Applying this bound, Lemma 5.2 below, and the fact that $\sum_{t=2}^{k}\sum_{d_1+\ldots+d_t=k:d_j\geq 2} 1 \leq k^k$, we have

$$\begin{aligned}
|\mathbb{E}[\bar{\mu}_k]-m_k| &\leq \frac{1}{A_k}\sum_{t=2}^{k}\sum_{d_1+\cdots+d_t=k,d_j\geq 2}|\gamma^{(d_1,\ldots,d_t)}|\cdot|\mathbb{E}[\bar{m}^\circ_{d_1}\cdots\bar{m}^\circ_{d_t}]-m_{d_1}\cdots m_{d_t}| \\
&\leq \frac{1}{A_k}k^k(k^k(n\wedge p)\,\|\mathbf{X}\|_{\mathrm{op}}^{2k})(k2^kM^kk^{k/2}\Delta_{k-1}^{\frac{1}{2}}) \\
&= \frac{1}{A_k}(2M)^k k^{\frac{5}{2}k+1}(n\wedge p)\,\|\mathbf{X}\|_{\mathrm{op}}^{2k}\,\Delta_{k-1}^{\frac{1}{2}}. \qquad (5.34)
\end{aligned}$$

The desired bound now follows from the lower bound on A_k in Assumption 5.1. □

Lemma 5.2. *Recall that $\gamma^{(d_1,\ldots,d_t)}=\sum_{s_1,\ldots,s_t=1}^{p}\widetilde{T}^{(d_1,\ldots,d_t)}_{s_1,\ldots,s_t}$. For any positive integers t,k, and $d_1,\ldots,d_t$ satisfying $d_j\geq 2$ and $d_1+\cdots+d_t=k$, it holds that*

$$|\gamma^{(d_1,\ldots,d_t)}|\leq\binom{k}{d_1,\ldots,d_t}(n\wedge p)\|\mathbf{X}\|_{\mathrm{op}}^{2k}\leq k!(n\wedge p)\|\mathbf{X}\|_{\mathrm{op}}^{2k}. \qquad (5.35)$$

Proof. Let $\mathbf{H}^{(d)}:=\mathbf{H}^{\circ d}$ be the d-th Hadamard power of H (i.e., $(\mathbf{H}^{(d)})_{ij}=(\mathbf{H}_{ij})^d$), and recall from (5.5) that $\widetilde{T}^{(d_1,\ldots,d_t)}_{s_1,\ldots,s_t}=\binom{k}{d_1,\ldots,d_t}\sum_{i=1}^{p}\mathbf{H}^{d_1}_{i,s_1}\cdots\mathbf{H}^{d_t}_{i,s_t}$ for $1\leq s_1<\ldots<s_t\leq p$ and zero otherwise. We have

$$\begin{aligned}
|\gamma^{(d_1,\ldots,d_t)}| &\leq \binom{k}{d_1,\ldots,d_t}\left|\sum_{i=1}^{p}\sum_{1\leq s_1<\cdots<s_t\leq p}\mathbf{H}^{d_1}_{i,s_1}\cdots\mathbf{H}^{d_t}_{i,s_t}\right| \\
&\leq \binom{k}{d_1,\ldots,d_t}\sum_{i=1}^{p}\left|\|\mathbf{H}\|_{\mathrm{op}}^{d_1-2}\cdots\|\mathbf{H}\|_{\mathrm{op}}^{d_t-2}\sum_{1\leq s_1<\cdots<s_t\leq p}\mathbf{H}^2_{i,s_1}\cdots\mathbf{H}^2_{i,s_t}\right| \\
&\leq \binom{k}{d_1,\ldots,d_t}\|\mathbf{H}\|_{\mathrm{op}}^{k-2t}\sum_{i=1}^{p}\left|\sum_{1\leq s_1<\cdots<s_{t-1}}\mathbf{H}^2_{i,s_1}\cdots\mathbf{H}^2_{i,s_{t-1}}\sum_{s_t=1}^{p}\mathbf{H}^2_{i,s_t}\right| \\
&\leq \binom{k}{d_1,\ldots,d_t}\|\mathbf{H}\|_{\mathrm{op}}^{k-2t}\|\mathbf{H}\|_{\mathrm{op}}^{2}\sum_{i=1}^{p}\left|\sum_{1\leq s_1<\cdots<s_{t-1}}\mathbf{H}^2_{i,s_1}\cdots\mathbf{H}^2_{i,s_{t-1}}\right| \\
&\leq \binom{k}{d_1,\ldots,d_t}\|\mathbf{X}\|_{\mathrm{op}}^{2k-4t}\|\mathbf{H}\|_{\mathrm{op}}^{2(t-1)}\sum_{i=1}^{p}\|\mathbf{H}_{i\cdot}\|^2 \\
&= \binom{k}{d_1,\ldots,d_t}\|\mathbf{X}\|_{\mathrm{op}}^{2k-4}\|\mathbf{H}\|_{F}^{2}\leq\binom{k}{d_1,\ldots,d_t}\|\mathbf{X}\|_{\mathrm{op}}^{2k}(n\wedge p),
\end{aligned}$$

as desired. $\square$

The following result gives a variance bound for $\bar{\mu}_k$.

Proposition 5.5. *Suppose that* $\mathbf{X}$ *satisfies Assumption 5.1 and* $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ *has i.i.d. entries distributed as* g*, where* g *is centered and* M*-sub-Gaussian for some* $M > 0$*. Suppose that the bound (5.30) holds up to some* $k-1$ *with* Δ_{k-1}*, then we have*

$$\mathsf{Var}(\bar{\mu}_k) \leq \left(\frac{CMk}{c_0}\right)^{2k^2}\left(\frac{p}{n} \vee 1\right)^{2k-2}\Delta_{k-1} + \left(\frac{CMk}{c_0}\right)^{4k}\frac{1}{n \wedge p}\left(\frac{p}{n} \vee 1\right)^{2k-2}. \tag{5.36}$$

Proof. In the expression (5.27), we will bound the variances of $F_k(\bar{\mathbf{y}})$ and $\sum_{t=2}^{k}\sum_{d_1+\ldots+d_t=k:d_j\geq 2}\gamma^{(d_1,\ldots,d_t)}$ separately, where $\gamma^{(d_1,\ldots,d_t)} = \sum_{s_1,\ldots,s_t=1}^{p}\widetilde{T}_{s_1,\ldots,s_t}^{(d_1,\ldots,d_t)}$. For the latter, recall that $|\bar{\mu}_d^\circ| \leq 2K_d = 2M^d d^{d/2}$ and $d_1, \ldots, d_t \leq k-1$, so by Lemma 5.3 below and the induction hypothesis in (5.30),

$$\begin{aligned}\mathsf{Var}(\bar{\mu}_{d_1}^\circ \cdots \bar{\mu}_{d_t}^\circ) &\leq 2^{3k}M^{2k^2}k^{k^2}\max_j \mathsf{Var}(\bar{\mu}_{d_j}^\circ) \leq 2^{3k}M^{2k^2}k^{k^2}\max_j \mathbb{E}[(\bar{\mu}_{d_j}^\circ - \mu_{d_j})^2] \\ &\leq 2^{3k}M^{2k^2}k^{k^2}\max_j \mathbb{E}[(\bar{\mu}_{d_j} - \mu_{d_j})^2] \leq 2^{3k}M^{2k^2}k^{k^2}\Delta_{k-1}.\end{aligned}$$

Consequently, by the lower bound on A_k via Assumption 5.1 and upper bound of $|\gamma^{(d_1,\ldots,d_t)}|$ in Lemma 5.2,

$$\begin{aligned}\mathsf{Var}\left(\frac{1}{A_k}\sum_{t=2}^{k}\sum_{d_1+\cdots+d_t=k,d_j\geq 2}\gamma^{(d_1,\ldots,d_t)}\bar{\mu}_{d_1}^\circ\cdots\bar{\mu}_{d_t}^\circ\right) &\leq \frac{k^k}{A_k^2}\sum_{t=2}^{k}\sum_{d_1+\cdots+d_t=k,d_j\geq 2}|\gamma^{(d_1,\ldots,d_t)}|^2\mathsf{Var}\left(\bar{\mu}_{d_1}^\circ\cdots\bar{\mu}_{d_t}^\circ\right) \\ &\leq \left(\frac{8}{c_0^2}\right)^k k^{4k+k^2}M^{2k^2}\left(\frac{p}{n}\vee 1\right)^{2k-2}\Delta_{k-1} \\ &\leq \left(\frac{CMk}{c_0}\right)^{2k^2}\left(\frac{p}{n}\vee 1\right)^{2k-2}\Delta_{k-1}.\end{aligned}$$

For the variance of $F_k(\bar{\mathbf{y}})$ defined in (5.8), we have $\mathsf{Var}(F_k(\bar{\mathbf{y}})) = \mathsf{Var}(\mathbb{E}[F_k(\bar{\mathbf{y}})|\bar{\boldsymbol{\beta}}]) + \mathbb{E}[\mathsf{Var}(F_k(\bar{\mathbf{y}})|\bar{\boldsymbol{\beta}})]$. For the first term, recall from (5.33) that $\mathbb{E}[F_k(\bar{\mathbf{y}})|\bar{\boldsymbol{\beta}}] = \left\langle T^{(k)}, \bar{\boldsymbol{\beta}}^{\otimes k}\right\rangle$, so by Lemma 5.4 below,

$$\mathsf{Var}(\mathbb{E}[F_k(\bar{\mathbf{y}})|\bar{\boldsymbol{\beta}}]) = \mathsf{Var}(\langle T^{(k)}, \bar{\boldsymbol{\beta}}^k\rangle) \leq 4^k k^{4k} M^{2k}(n \wedge p)\|\mathbf{X}\|_{\mathrm{op}}^{4k}.$$

On the other hand, by Lemma 5.6, we have

$$\mathbb{E}[\mathsf{Var}(F_k(\bar{\mathbf{y}})|\bar{\boldsymbol{\beta}})] \leq \left(\frac{CMk}{c_0}\right)^{2k+2}(n \wedge p)\,\|\mathbf{X}\|_{\mathrm{op}}^{4k}.$$

Combining these terms together, we conclude

$$\begin{aligned}\mathsf{Var}(\bar{\mu}_k) &\leq \left(\frac{CMk}{c_0}\right)^{2k^2}\left(\frac{p}{n} \vee 1\right)^{2k-2}\Delta_{k-1} + \left(\frac{2M}{c_0}\right)^{2k} k^{4k}\frac{1}{n \wedge p}\left(\frac{p}{n} \vee 1\right)^{2k-2} \\ &\quad + \left(\frac{CMk}{c_0}\right)^{2k+2}\frac{1}{n \wedge p}\left(\frac{p}{n} \vee 1\right)^{2k-2} \\ &\leq \left(\frac{CMk}{c_0}\right)^{2k^2}\left(\frac{p}{n} \vee 1\right)^{2k-2}\Delta_{k-1} + \left(\frac{CMk}{c_0}\right)^{4k}\frac{1}{n \wedge p}\left(\frac{p}{n} \vee 1\right)^{2k-2},\end{aligned}$$

as desired.

□

Lemma 5.3. *Suppose $X_1, \ldots, X_n$ are bounded random variables for some positive integer n. Then*

$$\mathsf{Var}(X_1 \cdots X_n) \leq n2^{n-1}\max_{i\in[n]}\|X_i\|_\infty^{2(n-1)} \cdot \max_{i\in[n]}\mathsf{Var}(X_i), \tag{5.37}$$

where $\|X\|_\infty$ denotes the essential supremum of X.

Proof. We prove by induction. For the base case $n = 2$, we have

$$\begin{aligned}
\mathsf{Var}(X_1X_2) &\le 2\mathsf{Var}((X_1 - \mathbb{E}[X_1])X_2) + 2\mathsf{Var}(\mathbb{E}[X_1]X_2) \\
&\le 2\mathbb{E}[((X_1 - \mathbb{E}[X_1])X_2)^2] + 2\|X_1\|_\infty^2 \mathsf{Var}(X_2) \\
&\le 2\|X_2\|_\infty^2 \mathbb{E}[(X_1 - \mathbb{E}[X_1])^2] + 2\|X_1\|_\infty^2 \mathsf{Var}(X_2) \\
&= 2\|X_2\|_\infty^2 \mathsf{Var}(X_1) + 2\|X_1\|_\infty^2 \mathsf{Var}(X_2) \qquad (5.38) \\
&\le 4\max(\|X_1\|_\infty^2, \|X_2\|_\infty^2) \cdot \max(\mathsf{Var}(X_1), \mathsf{Var}(X_2)),
\end{aligned}$$

which agrees with the desired bound. Let $A := \max_{i\in[n]} \|X_i\|_\infty$. Suppose the result holds for $n-1$, applying (5.38) yields

$$\begin{aligned}
\mathsf{Var}(X_1 \cdots X_n) &\le 2\|X_1\|_\infty^2 \mathsf{Var}(X_2 \cdots X_n) + 2\|X_2 \cdots X_n\|_\infty^2 \mathsf{Var}(X_1) \\
&\le 2\|X_1\|_\infty^2 \Big[(n-1)2^{n-2}A^{2(n-2)} \max_{2\le j\le n} \mathsf{Var}(X_j)\Big] + 2A^{2(n-1)}\mathsf{Var}(X_1) \\
&\le n2^{n-1}A^{2(n-1)} \cdot \max_{j\in[n]} \mathsf{Var}(X_j),
\end{aligned}$$

which proves the desired bound. ◻

Lemma 5.4. *Suppose that $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ has i.i.d. entries distributed as g, where g is centered and M-sub-Gaussian for some $M > 0$. Then for any $k \in \mathbb{N}$, we have*

$$\mathsf{Var}\Big(\big\langle T^{(k)}, \boldsymbol{\beta}^{\otimes k}\big\rangle\Big) \le 4^k k^{4k} M^{2k}(n \wedge p)\|\mathbf{X}\|_{\mathrm{op}}^{4k}.$$

Proof. Using the diagonal-free decomposition (5.32) of the tensor $T^{(k)}$ and then expanding using $\beta_{s_j}^{d_j} = \beta_{s_j}^{d_j} - m_{d_j} + m_{d_j}$ (where m_{d_j} denotes the d_j-th moment of β_1), we have

$$\begin{aligned}
\big\langle T^{(k)}, \boldsymbol{\beta}^{\otimes k}\big\rangle &= \sum_{t=1}^{k} \sum_{d_1+\ldots+d_t=k:d_j\ge 1} \sum_{s_1,\ldots,s_t=1}^{p} \widetilde{T}^{(d_1,\ldots,d_t)}_{s_1,\ldots,s_t} \beta_{s_1}^{d_1} \ldots \beta_{s_t}^{d_t} \\
&= \sum_{t=1}^{k} \sum_{d_1+\cdots+d_t=k, d_j\ge 1} \sum_{\mathcal{I}\subset[t]} \sum_{s_1,\ldots,s_t=1}^{p} \widetilde{T}^{(d_1,\ldots,d_t)}_{s_1\ldots s_t} \prod_{i\in\mathcal{I}} (\beta_{s_i}^{d_i} - m_{d_i}) \cdot \prod_{i\notin\mathcal{I}} m_{d_i}. \qquad (5.39)
\end{aligned}$$

Therefore, for some $u \in [t]$, up to an ordering a typical term can be bounded by

$$
\begin{aligned}
&\mathsf{Var}\Big(\sum_{s_1,\ldots,s_t=1}^{p} \widetilde{T}^{(d_1,\ldots,d_t)}_{s_1\ldots s_t} (\beta_{s_1}^{d_1} - m_{d_1}) \cdots (\beta_{s_u}^{d_u} - m_{d_u}) m_{d_{u+1}} \cdots m_{d_t} \Big) \\
&= m_{d_{u+1}}^2 \cdots m_{d_t}^2 \mathsf{Var}\Big(\sum_{s_1,\ldots,s_u=1}^{p} \Big(\sum_{s_{u+1},\ldots,s_t=1}^{p} \widetilde{T}^{(d_1,\ldots,d_t)}_{s_1\ldots s_t} \Big) (\beta_{s_1}^{d_1} - m_{d_1}) \cdots (\beta_{s_u}^{d_u} - m_{d_u}) \Big) \\
&= \mathsf{Var}(\beta_1^{d_1}) \cdots \mathsf{Var}(\beta_1^{d_u})\, m_{d_{u+1}}^2 \cdots m_{d_t}^2 \sum_{s_1,\ldots,s_u=1}^{p} \Big(\sum_{s_{u+1},\ldots,s_t=1}^{p} \widetilde{T}^{(d_1,\ldots,d_t)}_{s_1\ldots s_t} \Big)^2, \qquad (5.40)
\end{aligned}
$$

and we note that, since $m_1 \equiv 0$, this term is 0 unless $d_{u+1}, \ldots, d_t \geq 2$. Next note that, with $\mathbf{H}^{(d)}$ denoting the d-th Hadamard product of $\mathbf{H}$ satisfying $\|\mathbf{H}^{(d)}\|_{\mathrm{op}} \leq \|\mathbf{H}\|_{\mathrm{op}}^d$,

$$
\begin{aligned}
&\binom{k}{d_1, \ldots, d_t}^{-2} \sum_{s_1,\ldots,s_u} \Big(\sum_{s_{u+1},\ldots,s_t} \widetilde{T}^{(d_1,\ldots,d_t)}_{s_1\ldots s_t} \Big)^2 \\
&= \sum_{1 \leq s_1 < \cdots < s_u} \Big(\sum_{i=1}^{p} \sum_{(s_u) < s_{u+1} < \cdots < s_t \leq p} \mathbf{H}_{i,s_1}^{d_1} \cdots \mathbf{H}_{i,s_u}^{d_u} \mathbf{H}_{i,s_{u+1}}^{d_{u+1}} \cdots \mathbf{H}_{i,s_t}^{d_t} \Big)^2 \\
&= \sum_{1 \leq s_1 < \ldots < s_u} \Big(\sum_{i=1}^{p} \mathbf{H}_{i,s_1}^{(d_1)} \ldots \mathbf{H}_{i,s_u}^{(d_u)} \underbrace{\sum_{(s_u) < s_{u+1} < \ldots < s_t \leq p} \mathbf{H}_{i,s_{u+1}}^{(d_{u+1})} \ldots \mathbf{H}_{i,s_t}^{(d_t)}}_{\lambda_{i;s_u}} \Big)^2 \\
&\leq \sum_{s_1,\ldots,s_u=1}^{p} \Big(\sum_{i=1}^{p} \lambda_{i;s_u} \mathbf{H}_{i,s_1}^{(d_1)} \ldots \mathbf{H}_{i,s_u}^{(d_u)} \Big)^2 \\
&\overset{(*)}{\leq} \sum_{i=1}^{p} \max_{j \in [p]} \lambda_{i;j}^2 \cdot \|\mathbf{H}^{(d_1)}\|_{\mathrm{op}}^2 \ldots \|\mathbf{H}^{(d_u)}\|_{\mathrm{op}}^2 \\
&\leq \|\mathbf{X}\|_{\mathrm{op}}^{4(d_1 + \ldots + d_u)} \cdot \sum_{i=1}^{p} \max_{j \in [p]} \lambda_{i;j}^2,
\end{aligned}
$$

where $(*)$ follows from Lemma 5.5. It remains to notice that: for $d_{u+1}, \ldots, d_t \geq 2$, we have the following bound uniform in j:

$$
\begin{aligned}
|\lambda_{i;j}| &= \Big| \sum_{j<s_{u+1}<\ldots<s_t} \mathbf{H}_{i,s_{u+1}}^{d_{u+1}} \ldots \mathbf{H}_{i,s_t}^{d_t} \Big| \\
&\leq \|\mathbf{X}\|_{\text{op}}^{2((d_{u+1}-2)+\ldots+(d_t-2))} \sum_{j<s_{u+1}<\ldots<s_t} \mathbf{H}_{i,s_{u+1}}^2 \ldots \mathbf{H}_{i,s_t}^2 \\
&\leq \|\mathbf{X}\|_{\text{op}}^{2((d_{u+1}-2)+\ldots+(d_t-2))} \|\mathbf{X}\|_{\text{op}}^4 \cdot \sum_{j<s_{u+1}<\ldots<s_{t-1}} \mathbf{H}_{i,s_{u+1}}^2 \ldots \mathbf{H}_{i,s_{t-1}}^2 \\
&\leq \ldots \leq \|\mathbf{X}\|_{\text{op}}^{2(d_{u+1}+\ldots+d_t-2)} \sum_{j<s_{u+1}} \mathbf{H}_{i,s_{u+1}}^2 \leq \|\mathbf{X}\|_{\text{op}}^{2(d_{u+1}+\ldots+d_t-2)} \|\mathbf{H}_{i,\cdot}\|^2, \qquad (5.41)
\end{aligned}
$$

which implies that

$$
\begin{aligned}
&\sum_{s_1,\ldots,s_u} \Big(\sum_{s_{u+1},\ldots,s_t} \widetilde{T}_{s_1\ldots s_t}^{(d_1,\ldots,d_t)} \Big)^2 \leq \binom{k}{d_1,\ldots,d_t}^2 \|\mathbf{X}\|_{\text{op}}^{4k-8} \sum_{i=1}^p \|\mathbf{H}_{i,\cdot}\|^4 \\
&\leq \binom{k}{d_1,\ldots,d_t}^2 \|\mathbf{X}\|_{\text{op}}^{4k-4} \sum_{i=1}^p \|\mathbf{H}_{i,\cdot}\|^2 = \binom{k}{d_1,\ldots,d_t}^2 \|\mathbf{X}\|_{\text{op}}^{4k-4} \|\mathbf{H}\|_F^2 \leq k^{2k}(n \wedge p) \|\mathbf{X}\|_{\text{op}}^{4k}.
\end{aligned}
$$

Recall that $K_d = M^d d^{d/2}$ and $m_d(g) \leq K_d$. Then using $\mathsf{Var}(\beta_1^d) \leq m_{2d}(g) \leq K_{2d}$, the above display further implies that the typical term in (5.40) can be bounded by

$$
\begin{aligned}
&\mathsf{Var}\Big(\sum_{s_1,\ldots,s_t=1}^p \widetilde{T}_{s_1\ldots s_t}^{(d_1,\ldots,d_t)} (\beta_{s_1}^{d_1} - m_{d_1}) \cdots (\beta_{s_u}^{d_u} - m_{d_u}) m_{d_{u+1}} \cdots m_{d_t} \Big) \\
&\leq \prod_{i=1}^u M^{2d_i} (2d_i)^{d_i} \cdot \prod_{i=u+1}^t M^{2d_i} d_i^{d_i} \cdot k^{2k} (n \wedge p) \|\mathbf{X}\|_{\text{op}}^{4k} \leq 2^k k^{3k} M^{2k} (n \wedge p) \|\mathbf{X}\|_{\text{op}}^{4k},
\end{aligned}
$$

where we apply the fact that $\max_{d_1+\ldots+d_t \leq k} d_1^{d_1} \ldots d_t^{d_t} = (k/t)^k \leq k^k$. Coming back to (5.39) and using the count that $\sum_{t=1}^k \sum_{d_1+\cdots+d_t=k, d_j \geq 1} \sum_{\mathcal{I} \subset [t]} 1 \leq k^k 2^k$, we have

$$
\mathsf{Var}\Big(\langle T^{(k)}, \boldsymbol{\beta}^k \rangle \Big) \leq k^k 2^k \cdot 2^k k^{3k} M^{2k} (n \wedge p) \|\mathbf{X}\|_{\text{op}}^{4k} = 4^k k^{4k} M^{2k} (n \wedge p) \|\mathbf{X}\|_{\text{op}}^{4k},
$$

as desired. $\square$

Lemma 5.5. *Let $A^{(1)}, \ldots, A^{(t)}$ be $m \times p$ matrices. For any set of real numbers $\{\lambda_{ij},$*

$i \in [p], j \in [m]\}$, it holds that

$$\sum_{s_1,\ldots,s_t=1}^{m} \Big(\sum_{i=1}^{p} \lambda_{i,s_t} A^{(1)}_{s_1,i} \cdots A^{(t)}_{s_t,i} \Big)^2 \leq \|A^{(1)}\|_{\mathrm{op}}^2 \cdots \|A^{(t)}\|_{\mathrm{op}}^2 \sum_{i=1}^{p} \max_{j\in[m]} \lambda_{ij}^2.$$

Proof. For any $s_2, \ldots, s_t \in [m]$, let $\Lambda^{(s_2,\ldots,s_t)} = (\{\lambda_{i,s_t} A^{(2)}_{s_2,i} \cdots A^{(t)}_{s_t,i}\}_{i=1}^{p}) \in \mathbb{R}^p$. Then,

$$\begin{aligned}
\sum_{s_1,\ldots,s_t=1}^{m} \Big(\sum_{i=1}^{p} \lambda_{i,s_t} A^{(1)}_{s_1,i} \cdots A^{(t)}_{s_t,i} \Big)^2 &= \sum_{s_2,\ldots,s_t=1}^{m} \Big\| \sum_{i=1}^{p} \lambda_{i,s_t} A^{(2)}_{s_2,i} \cdots A^{(t)}_{s_t,i} \cdot A^{(1)}_{\cdot i} \Big\|^2 \\
&= \sum_{s_2,\ldots,s_t=1}^{m} \|A^{(1)} \Lambda^{(s_2,\ldots,s_t)}\|^2 \\
&\leq \|A^{(1)}\|_{\mathrm{op}}^2 \sum_{i=1}^{p} \sum_{s_2,\ldots,s_t=1}^{m} \lambda_{i,s_t}^2 (A^{(2)}_{s_2,i} \cdots A^{(t)}_{s_t,i})^2 \\
&= \|A^{(1)}\|_{\mathrm{op}}^2 \sum_{i=1}^{p} \|A^{(2)}_{\cdot i}\|^2 \cdots \|A^{(t-1)}_{\cdot i}\|^2 \sum_{s_t=1}^{m} \lambda_{i,s_t}^2 (A^{(t)}_{s_t,i})^2 \\
&\leq \|A^{(1)}\|_{\mathrm{op}}^2 \cdots \|A^{(t)}\|_{\mathrm{op}}^2 \sum_{i=1}^{p} \max_{j\in[m]} \lambda_{ij}^2,
\end{aligned}$$

as desired. □

Lemma 5.6. *Recall $F_k(\cdot)$ in (5.8) and assume $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$. For any integer $k \geq 2$, it holds that*

$$\mathsf{Var}(F_k(\mathbf{y})|\boldsymbol{\beta}) \leq \sum_{\ell=1}^{k-1} \binom{k}{\ell}^2 \ell! \sigma^{2\ell} \|\mathbf{X}\|_{\mathrm{op}}^{2\ell} \sum_{j=1}^{p} (\mathbf{H}\boldsymbol{\beta})_j^{2(k-\ell)} + k! \sigma^{2k} (n \wedge p) \|\mathbf{X}\|_{\mathrm{op}}^{2k}.$$

Moreover, if $\beta_1, \ldots, \beta_p$ are i.i.d. centered and M-sub-Gaussian for some $M > 0$ and $\mathbf{X}$ satisfies Assumption 5.1, then

$$\mathbb{E}_{\boldsymbol{\beta}}[\mathsf{Var}(F_k|\boldsymbol{\beta})] \leq \Big(\frac{CMk}{c_0} \Big)^{2k+2} \|\mathbf{X}\|_{\mathrm{op}}^{4k},$$

where $C > 0$ is universal.

Proof. By Lemma 5.7, we have

$$\begin{aligned}
\mathsf{Var}(F_k(\mathbf{y})|\boldsymbol{\beta}) &= \sum_{j,j'=1}^{p} (\sigma\|\mathbf{x}_j\|)^k(\sigma\|\mathbf{x}_j'\|)^k \mathrm{Cov}\Big(H_k\Big(\frac{\mathbf{x}_j^\top\mathbf{y}}{\sigma\|\mathbf{x}_j\|}\Big), H_k\Big(\frac{\mathbf{x}_{j'}^\top\mathbf{y}}{\sigma\|\mathbf{x}_{j'}\|}\Big)\Big) \\
&= \sum_{j,j'=1}^{p}\sum_{\ell=1}^{k} \binom{k}{\ell}^2 \ell!(\sigma\|\mathbf{x}_j\|)^k(\sigma\|\mathbf{x}_j'\|)^k \Big(\frac{\mathbf{x}_j^\top\boldsymbol{\theta}}{\sigma\|\mathbf{x}_j\|}\Big)^{k-\ell}\Big(\frac{\mathbf{x}_{j'}^\top\boldsymbol{\theta}}{\sigma\|\mathbf{x}_{j'}\|}\Big)^{k-\ell}\Big(\frac{\mathbf{x}_j^\top\mathbf{x}_{j'}}{\|\mathbf{x}_j\|\|\mathbf{x}_{j'}\|}\Big)^{\ell} \\
&= \sum_{\ell=1}^{k}\binom{k}{\ell}^2 \ell!\sigma^{2\ell}\cdot\sum_{j,j'=1}^{p}(\mathbf{x}_j^\top\boldsymbol{\theta})^{k-\ell}(\mathbf{x}_{j'}^\top\boldsymbol{\theta})^{k-\ell}(\mathbf{x}_j^\top\mathbf{x}_{j'})^{\ell}.
\end{aligned}$$

Let us bound the inner summation for each $\ell \in [k]$. If $\ell = k \geq 2$, we have

$$\Big|\sum_{j,j'=1}^{p}(\mathbf{x}_j^\top\mathbf{x}_{j'})^k\Big| \leq \|\mathbf{X}\|_{\mathrm{op}}^{2(k-2)}\sum_{j,j'=1}^{p}(\mathbf{x}_j^\top\mathbf{x}_{j'})^2 = \|\mathbf{X}\|_{\mathrm{op}}^{2(k-2)}\|\mathbf{H}\|_F^2 \leq (n\wedge p)\|\mathbf{X}\|_{\mathrm{op}}^{2k}.$$

For $\ell \leq k-1$, denote $(\mathbf{H}\boldsymbol{\beta})^{\circ\alpha} \in \mathbb{R}^p$ denote the α-th Hadamard product of $\mathbf{H}\boldsymbol{\beta}$, then using $\mathbf{x}_j^\top\boldsymbol{\theta} = (\mathbf{H}\boldsymbol{\beta})_j$,

$$\begin{aligned}
&\Big|\sum_{j,j'=1}^{p}(\mathbf{x}_j^\top\boldsymbol{\theta})^{k-\ell}(\mathbf{x}_{j'}^\top\boldsymbol{\theta})^{k-\ell}(\mathbf{x}_j^\top\mathbf{x}_{j'})^{\ell}\Big| = \Big|\sum_{j,j'=1}^{p}[(\mathbf{H}\boldsymbol{\beta})^{\circ(k-\ell)}]_j[(\mathbf{H}\boldsymbol{\beta})^{\circ(k-\ell)}]_{j'}[\mathbf{H}^{\circ\ell}]_{j,j'}\Big| \\
&= (\mathbf{H}\boldsymbol{\beta})^{\circ(k-\ell)\top}\mathbf{H}^{\circ\ell}(\mathbf{H}\boldsymbol{\beta})^{\circ(k-\ell)} \leq \|\mathbf{H}^{\circ\ell}\|_{\mathrm{op}}\|(\mathbf{H}\boldsymbol{\beta})^{\circ(k-\ell)}\|^2 \leq \|\mathbf{X}\|_{\mathrm{op}}^{2\ell}\sum_{j=1}^{p}(\mathbf{H}\boldsymbol{\beta})_j^{2(k-\ell)},
\end{aligned}$$

leading to the desired bound for $\mathsf{Var}(F_k(\mathbf{y})|\boldsymbol{\beta})$. To bound its expectation over $\boldsymbol{\beta}$, note that for each $j \in [p]$, $(\mathbf{H}\boldsymbol{\beta})_j$ is sub-Gaussian with constant $CM\|\mathbf{H}_{j\cdot}\|$ for some universal $C > 0$ (in the following, $C > 0$ denotes a universal constant whose values changes from instance to instance). Hence the sub-Gaussian assumption yields

$$\sum_{j=1}^{p}\mathbb{E}(\mathbf{H}\boldsymbol{\beta})_j^{2(k-\ell)} \leq (CM)^{2(k-\ell)}\|\mathbf{X}\|_{\mathrm{op}}^{4(k-\ell-1)}\|\mathbf{H}\|_F^2(2k)^k \leq (CM)^{2k}(2k)^k(n\wedge p)\|\mathbf{X}\|_{\mathrm{op}}^{4(k-\ell)},$$

which further implies that

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\beta}}\mathsf{Var}(F_k|\boldsymbol{\beta}) &\leq (CM)^{2k}(2k)^k \sum_{\ell=1}^{k} \binom{k}{\ell}^2 \ell! \sigma^{2\ell} \|\mathbf{X}\|_{\text{op}}^{4k-2\ell} \\
&\leq (CM)^{2k} k^{2k} 8^k \sum_{\ell=1}^{k} \sigma^{2\ell} \|\mathbf{X}\|_{\text{op}}^{4k-2\ell} \\
&\leq (CM)^{2k} k^{2k} 8^k \|\mathbf{X}\|_{\text{op}}^{4k} \frac{\sigma^2}{\|\mathbf{X}\|_{\text{op}}^2} \sum_{\ell=0}^{k-1} \left(\frac{\sigma}{\|\mathbf{X}\|_{\text{op}}}\right)^{2k} \\
&\leq (CM)^{2k} k^{2k+1} 8^k \sigma^2 \|\mathbf{X}\|_{\text{op}}^{4k-2} \left(1 \vee \frac{\sigma}{\|\mathbf{X}\|_{\text{op}}}\right)^{2(k-1)} \\
&\leq \left(\frac{CMk}{c_0}\right)^{2k+2} \|\mathbf{X}\|_{\text{op}}^{4k},
\end{aligned}
$$

where we apply Assumption 5.1 in the last step. The proof is complete. $\square$

Lemma 5.7 (Hermite polynomials of correlation Gaussians)**.** *Let H_k denote the monic degree-k Hermite polynomial. Let (Z_1, Z_2) be joint normally distributed with $Z_1, Z_2 \sim \mathcal{N}(0,1)$ and $\mathbb{E}[Z_1 Z_2] = \rho \in [-1,1]$. Then for $k, k' \geq 1$,*

$$
Cov(H_k(Z_1), H_{k'}(Z_2)) = \mathbf{1}_{k=k'} \rho^k k!, \tag{5.42}
$$

and for any $\theta_1, \theta_2 \in \mathbb{R}$,

$$
Cov(H_k(\theta_1 + Z_1), H_k(\theta_2 + Z_2)) = \sum_{\ell=1}^{k} \binom{k}{\ell}^2 \ell! \cdot \theta_1^{k-\ell} \theta_2^{k-\ell} \rho^\ell. \tag{5.43}
$$

Proof. We first prove (5.42). Let $f_\rho(x,y)$ denote the joint density of (Z_1, Z_2) and $f_0(x,y) = \varphi(x)\varphi(y)$, where $\varphi(\cdot)$ is the standard normal density. It is well-known that the likelihood ratio $\frac{f_\rho}{f_0}$ (Mehler kernel) admits the following spectral decomposition:

$$
K(x,y) = \frac{f_\rho(x,y)}{f_0(x,y)} = \sum_{k \geq 0} \frac{\rho^k}{k!} H_k(x) H_k(y).
$$

Let W_1, W_2 be iid standard normal. Then

$$\begin{aligned}
\text{Cov}(H_k(Z_1), H_{k'}(Z_2)) &= \mathbb{E}[H_k(Z_1)H_{k'}(Z_2)] \\
&= \mathbb{E}[H_k(W_1)H_{k'}(W_2)K(W_1, W_2)] \\
&= \sum_{\ell \geq 0} \frac{\rho^\ell}{\ell!} \mathbb{E}[(H_k(W_1)H_\ell(W_1)]\mathbb{E}[H_{k'}(W_2)H_\ell(W_2)] \\
&= \begin{cases} \rho^k k! & k = k' \\ 0 & k \neq k' \end{cases},
\end{aligned}$$

where we used the orthogonality $\mathbb{E}[H_k(W_1)H_\ell(W_1)] = \mathbf{1}_{k=\ell} k!$.

Finally, (5.43) follows (5.42) and the identity that $H_k(x + y) = \sum_{\ell=0}^{k} \binom{k}{\ell} x^{k-\ell} H_\ell(y)$, which is a consequence of $H_k(x) = \mathbb{E}_{W \sim \mathcal{N}(0,1)}[(x + iW)^k]$. □

Equipped with the bias and variance bounds in Propositions 5.4 and 5.5, we are now ready to prove Proposition 5.3 by induction.

Proof of Proposition 5.3. **Base line** For $k = 2$, we have $\Delta_2 = (\frac{2CM}{c_0})^{27} \left(\frac{p}{n} \vee 1\right)^4 \frac{1}{n \wedge p}$, and $\bar{\mu}_2 = \frac{F_2(\bar{\mathbf{y}})}{A_2}$ as in (5.27) is unbiased and satisfies the variance bound as in (5.19):

$$\mathsf{Var}(\bar{\mu}_2) \leq \frac{(m_4 + 5m_2^2)\text{Tr}(\mathbf{H}^4) + 4\sigma^2 m_2 \text{Tr}(\mathbf{H}^3) + 2\sigma^4 \text{Tr}(\mathbf{H}^2)}{\text{Tr}(\mathbf{H}^2)^2}.$$

We now apply

$$\text{Tr}(\mathbf{H}^4) \leq (n \wedge p) \|\mathbf{X}\|_{\text{op}}^8, \quad \text{Tr}(\mathbf{H}^3) \leq (n \wedge p) \|\mathbf{X}\|_{\text{op}}^6, \quad \text{Tr}(\mathbf{H}^2) \leq (n \wedge p) \|\mathbf{X}\|_{\text{op}}^4,$$

along with Assumption 5.1, which implies $A_2 = \text{Tr}(\mathbf{H}^2) \geq c_0^2 \|\mathbf{X}\|_{\text{op}}^4 p \left(\frac{n}{p} \wedge 1\right)^2$. Applying the bound $m_k \leq M^k k^{k/2}$ yields the desired result $\mathbb{E}[(\bar{\mu}_2 - \mu_2)^2] = \mathsf{Var}(\bar{\mu}_2) \leq \Delta_2$.

Induction By Propositions 5.4 and 5.5, we have

$$
\begin{aligned}
\mathbb{E}[(\bar{\mu}_k - \mu_k)^2] &= |\mathbb{E}[\bar{\mu}_k] - \mu_k|^2 + \mathsf{Var}(\bar{\mu}_k) \\
&\leq \left(\frac{2Mk}{c_0}\right)^{5k+2}\left(\frac{p}{n}\vee 1\right)^{2k-2}\Delta_{k-1} + \left(\frac{CMk}{c_0}\right)^{2k^2}\left(\frac{p}{n}\vee 1\right)^{2k-2}\Delta_{k-1} \\
&\quad + \left(\frac{CMk}{c_0}\right)^{4k}\frac{1}{n\wedge p}\left(\frac{p}{n}\vee 1\right)^{2k-2} \\
&\leq \left(\frac{CMk}{c_0}\right)^{2k^2}\left(\frac{p}{n}\vee 1\right)^{2k-2}\Delta_{k-1} + \left(\frac{CMk}{c_0}\right)^{4k}\frac{1}{n\wedge p}\left(\frac{p}{n}\vee 1\right)^{2k-2} \\
&\leq \Delta_k,
\end{aligned}
$$

where the last step follows from the expression $\Delta_k = \left(\frac{CMk}{c_0}\right)^{(k+1)^3}\frac{1}{n\wedge p}\left(\frac{p}{n}\vee 1\right)^{k^2}$ in (5.30). The proof is complete. □

5.4.2.2 Pertubation analysis for debiased model

Now we focus on the difference between the moment estimators of the approximately centered model $\widehat{\mathbf{y}} := \mathbf{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}$ and the true centered model $\bar{\mathbf{y}} := \mathbf{X}\bar{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}$, where $\widehat{\boldsymbol{\beta}} := \boldsymbol{\beta} - \widehat{m}_1\mathbf{1}_p$ with $\widehat{m}_1$ given in (5.2), and $\bar{\boldsymbol{\beta}} = \boldsymbol{\beta} - m_1\mathbf{1}_p$. The following result states the mean squared difference between the true estimator $\widehat{\mu}_\ell$ in (5.28) and the idealized estimator $\bar{\mu}_\ell$ in (5.27).

Proposition 5.6. *Suppose that Assumptions 5.1 and 5.2 hold. For any $\ell \geq 3$, it holds that $\mathbb{E}[(\widehat{\mu}_\ell - \bar{\mu}_\ell)^2] \leq \delta_\ell$, where*

$$
\delta_\ell = \left(\frac{CM\ell}{c_0^2}\right)^{3\ell^2}\frac{1}{n\wedge p}\left(\frac{p}{n}\vee 1\right)^{2\ell^2} \tag{5.44}
$$

for some universal $C > 0$.

Proof. We will apply induction to prove this result. The base case $\ell = 2$ follows from Proposition 5.1. Assuming the claim holds for some $\ell - 1$ with $\ell \geq 3$, we will bound separately the two terms

$$\frac{1}{A_\ell}(F_\ell(\widehat{\mathbf{y}}) - F_\ell(\bar{\mathbf{y}})) \quad \text{and} \quad \frac{1}{A_\ell}\sum_{t=2}^{\ell}\sum_{d_1+\ldots+d_t=\ell: d_j\geq 2}\gamma^{(d_1,\ldots,d_t)}(\widehat{\mu}^\circ_{d_1}\ldots\widehat{\mu}^\circ_{d_t} - \bar{\mu}^\circ_{d_1}\ldots\bar{\mu}^\circ_{d_t}),$$

where we recall $\gamma^{(d_1,\ldots,d_t)} = \sum_{s_1,\ldots,s_t=1}^{p}\widetilde{T}^{(d_1,\ldots,d_t)}_{s_1,\ldots,s_t}$.

Bound of first term Recall that $F_\ell(\widehat{\mathbf{y}}) := \sum_{j=1}^{p}(\sigma\|\mathbf{x}_j\|)^\ell H_\ell\Big(\frac{\mathbf{x}_j^\top\widehat{\mathbf{y}}}{\sigma\|\mathbf{x}_j\|}\Big)$. Denoting $\mathbf{u}_j := \mathbf{x}_j/\|\mathbf{x}_j\|$ and $\delta := |\widehat{m}_1 - m_1|$, we have

$$\begin{aligned}\frac{1}{A_\ell}|F_\ell(\widehat{\mathbf{y}}) - F_\ell(\bar{\mathbf{y}})| &\leq \frac{1}{A_\ell}\sum_{j=1}^{p}\sigma^\ell\|\mathbf{x}_j\|^\ell\left|H_\ell(\mathbf{u}_j^\top\widehat{\mathbf{y}}/\sigma) - H_\ell(\mathbf{u}_j^\top\bar{\mathbf{y}}/\sigma)\right| \\ &\leq \frac{1}{A_\ell}\sum_{j=1}^{p}\sigma^\ell\|\mathbf{x}_j\|^\ell|H'_\ell(\xi_j)||\mathbf{u}_j^\top(\widehat{\mathbf{y}}-\bar{\mathbf{y}})/\sigma| \\ &\leq \frac{\delta}{A_\ell}\sum_{j=1}^{p}\sigma^{\ell-1}\|\mathbf{x}_j\|^\ell|H'_\ell(\xi_j)||\mathbf{u}_j^\top\mathbf{X}\mathbf{1}|,\end{aligned}$$

where ξ_j is some point between $\mathbf{u}_j^\top\widehat{\mathbf{y}}/\sigma$ and $\mathbf{u}_j^\top\bar{\mathbf{y}}/\sigma$. Note that $H'_\ell(x) = \sum_{k=0}^{\ell-1}a_kx^k$ is a polynomial of degree $\ell-1$. For each term x^k with $x\in[a,b]$, we have $|x|^k \leq |a|^k + |b|^k$. Then we have

$$\begin{aligned}&\frac{1}{A_\ell}|F_\ell(\widehat{\mathbf{y}}) - F_\ell(\bar{\mathbf{y}})| \\ &\leq \frac{\delta}{A_\ell}\sum_{j=1}^{p}\sigma^{\ell-1}\|\mathbf{x}_j\|^\ell\sum_{k=0}^{\ell-1}|a_k|\Big(\Big|\frac{\mathbf{u}_j^\top\widehat{\mathbf{y}}}{\sigma}\Big|^k + \Big|\frac{\mathbf{u}_j^\top\bar{\mathbf{y}}}{\sigma}\Big|^k\Big)|\mathbf{u}_j^\top\mathbf{X}\mathbf{1}| \\ &\leq \frac{\delta}{A_\ell}\sum_{k=0}^{\ell-1}|a_k|C^k\sigma^{\ell-k-1}\sum_{j=1}^{p}\|\mathbf{x}_j\|^{\ell-k-1}\Big(|\mathbf{x}_j^\top\mathbf{X}\widehat{\boldsymbol{\beta}}|^k + |\mathbf{x}_j^\top\mathbf{X}\bar{\boldsymbol{\beta}}|^k + |\mathbf{x}_j^\top\boldsymbol{\varepsilon}|^k\Big)|\mathbf{x}_j^\top\mathbf{X}\mathbf{1}| \\ &= \frac{\delta}{A_\ell}\sum_{k=0}^{\ell-1}|a_k|C^k\sigma^{\ell-k-1}\sum_{j=1}^{p}\|\mathbf{x}_j\|^{\ell-k-1}\Big(|\mathbf{x}_j^\top\mathbf{X}(\widehat{\boldsymbol{\beta}}-\bar{\boldsymbol{\beta}}+\bar{\boldsymbol{\beta}})|^k + |\mathbf{x}_j^\top\mathbf{X}\bar{\boldsymbol{\beta}}|^k + |\mathbf{x}_j^\top\boldsymbol{\varepsilon}|^k\Big)|\mathbf{x}_j^\top\mathbf{X}\mathbf{1}| \\ &\leq \frac{\delta}{A_\ell}\sum_{k=0}^{\ell-1}|a_k|C^k\sigma^{\ell-k-1}\sum_{j=1}^{p}\|\mathbf{x}_j\|^{\ell-k-1}\Big(|\mathbf{x}_j^\top\mathbf{X}\bar{\boldsymbol{\beta}}|^k + \delta^k|\mathbf{x}_j^\top\mathbf{X}\mathbf{1}|^k + |\mathbf{x}_j^\top\boldsymbol{\varepsilon}|^k\Big)|\mathbf{x}_j^\top\mathbf{X}\mathbf{1}|.\end{aligned}$$

Consequently, using $|a_k| \le (Ck)^k$, we have

$$
\begin{aligned}
&\mathbb{E}[(F_\ell(\widehat{\mathbf{y}}) - F_\ell(\bar{\mathbf{y}}))^2] \\
&\le \mathbb{E}\Big[\delta^2 \Big(\sum_{k=0}^{\ell-1} C^k (\sigma \|\mathbf{X}\|_{\mathrm{op}})^{\ell-k-1} \Big(\sum_{j=1}^{p} \big(|\mathbf{x}_j^\top \mathbf{X}\bar{\boldsymbol{\beta}}|^k + \delta^k |\mathbf{x}_j^\top \mathbf{X}\mathbf{1}|^k + |\mathbf{x}_j^\top \boldsymbol{\varepsilon}|^k \big) |\mathbf{x}_j^\top \mathbf{X}\mathbf{1}| \Big) \Big)^2 \Big] \\
&\le C^\ell \ell \mathbb{E}\Big[\delta^2 \sum_{k=0}^{\ell-1} (\sigma \|\mathbf{X}\|_{\mathrm{op}})^{2(\ell-k-1)} \Big(\sum_{j=1}^{p} \big(|\mathbf{x}_j^\top \mathbf{X}\bar{\boldsymbol{\beta}}|^k + \delta^k |\mathbf{x}_j^\top \mathbf{X}\mathbf{1}|^k + |\mathbf{x}_j^\top \boldsymbol{\varepsilon}|^k \big) |\mathbf{x}_j^\top \mathbf{X}\mathbf{1}| \Big)^2 \Big] \\
&\le C^\ell \ell \sum_{k=0}^{\ell-1} (\sigma \|\mathbf{X}\|_{\mathrm{op}})^{2(\ell-k-1)} \Big\{ \underbrace{\mathbb{E}\Big[\delta^2 \Big(\sum_{j=1}^{p} |\mathbf{x}_j^\top \mathbf{X}\bar{\boldsymbol{\beta}}|^k |\mathbf{x}_j^\top \mathbf{X}\mathbf{1}| \Big)^2 \Big]}_{(I_k)} + \underbrace{\mathbb{E}\Big[\delta^{2k+2} \Big(\sum_{j=1}^{p} |\mathbf{x}_j^\top \mathbf{X}\mathbf{1}|^{k+1}| \Big)^2 \Big]}_{(II_k)} \\
&\quad + \underbrace{\mathbb{E}\Big[\delta^2 \Big(\sum_{j=1}^{p} |\mathbf{x}_j^\top \boldsymbol{\varepsilon}|^k |\mathbf{x}_j^\top \mathbf{X}\mathbf{1}| \Big)^2 \Big]}_{(III_k)} \Big\}.
\end{aligned}
$$

To bound (I_k), we have

$$
\begin{aligned}
(I_k)^2 &\le \mathbb{E}[\delta^4] \mathbb{E}\Big[\Big(\sum_{j=1}^{p} |\mathbf{x}_j^\top \mathbf{X}\bar{\boldsymbol{\beta}}|^k |\mathbf{x}_j^\top \mathbf{X}\mathbf{1}| \Big)^4 \Big] \le \mathbb{E}[\delta^4] \mathbb{E}\Big[\Big(\sum_{j=1}^{p} |\mathbf{x}_j^\top \mathbf{X}\bar{\boldsymbol{\beta}}|^{2k} \Big)^2 \Big(\sum_{j=1}^{p} (\mathbf{x}_j^\top \mathbf{X}\mathbf{1})^2 \Big)^2 \Big] \\
&= \mathbb{E}[\delta^4] \|\mathbf{X}^\top \mathbf{X}\mathbf{1}\|^4 \mathbb{E}\Big[\Big(\sum_{j=1}^{p} |\mathbf{x}_j^\top \mathbf{X}\bar{\boldsymbol{\beta}}|^{2k} \Big)^2 \Big] \le p \mathbb{E}[\delta^4] \|\mathbf{X}^\top \mathbf{X}\mathbf{1}\|^4 \mathbb{E}\Big[\sum_{j=1}^{p} (\mathbf{x}_j^\top \mathbf{X}\bar{\boldsymbol{\beta}})^{4k} \Big] \\
&\le (Ck)^{2k} M^{4k} p \mathbb{E}[\delta^4] \|\mathbf{X}^\top \mathbf{X}\mathbf{1}\|^4 \sum_{j=1}^{p} \|\mathbf{X}^\top \mathbf{x}_j\|^{4k} \le (Ck)^{2k} M^{4k} p^2 \|\mathbf{X}\|_{\mathrm{op}}^{8k+4} \|\mathbf{X}\mathbf{1}\|^4 \mathbb{E}[\delta^4].
\end{aligned}
$$

Using the definition $\delta = \widehat{m}_1 - m_1 = \frac{\mathbf{1}^\top \mathbf{X}^\top \bar{\mathbf{y}}}{\|\mathbf{X}\mathbf{1}\|^2}$ and Assumption 5.2, we have

$$
\mathbb{E}[\delta^4] \le C \frac{\mathbb{E}(\mathbf{1}^\top \mathbf{X}\mathbf{X}\bar{\boldsymbol{\beta}})^4 + \mathbb{E}(\mathbf{1}^\top \mathbf{X}^\top \boldsymbol{\varepsilon})^4}{\|\mathbf{X}\mathbf{1}\|^8} \le C \frac{\|\mathbf{X}\|_{\mathrm{op}}^4}{\|\mathbf{X}\mathbf{1}\|^4} \Big(M + \frac{\sigma}{\|\mathbf{X}\|_{\mathrm{op}}} \Big)^4,
$$

which further implies

$$
(I_k)^2 \le (Ck)^{2k} M^{4k} p^2 \|\mathbf{X}\|_{\mathrm{op}}^{8k+8} \Big(M + \frac{\sigma}{\|\mathbf{X}\|_{\mathrm{op}}} \Big)^4,
$$

or equivalently, $(I_k) \le (Ck)^k M^{2k} p \|\mathbf{X}\|_{\mathrm{op}}^{4k+4} \Big(M + \frac{\sigma}{\|\mathbf{X}\|_{\mathrm{op}}} \Big)^2$. A similar consideration yields

$$(III_k) \le (Ck)^k p \sigma^{2k} \|\mathbf{X}\|_{\mathrm{op}}^{2k+4} \Big(M + \frac{\sigma}{\|\mathbf{X}\|_{\mathrm{op}}} \Big)^2.$$

Lastly, to bound (II_k), we have

$$\begin{aligned}
(II_k) &= \mathbb{E}[\delta^{2k+2}] \Big(\sum_{j=1}^p |\mathbf{x}_j^\top \mathbf{X}\mathbf{1}|^{k+1} \Big)^2 \le p \mathbb{E}[\delta^{2k+2}] \sum_{j=1}^p (\mathbf{x}_j^\top \mathbf{X}\mathbf{1})^{2(k+1)} \\
&\le p \mathbb{E}[\delta^{2k+2}] \|\mathbf{X}\|_{\mathrm{op}}^{2k} \|\mathbf{X}\mathbf{1}\|^{2k} \sum_{j=1}^p (\mathbf{x}_j^\top \mathbf{X}\mathbf{1})^2 \\
&= p \mathbb{E}[\delta^{2k+2}] \|\mathbf{X}\|_{\mathrm{op}}^{2k} \|\mathbf{X}\mathbf{1}\|^{2k} \|\mathbf{X}^\top \mathbf{X}\mathbf{1}\|^2 \le p \mathbb{E}[\delta^{2k+2}] \|\mathbf{X}\|_{\mathrm{op}}^{2k+2} \|\mathbf{X}\mathbf{1}\|^{2k+2},
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}[\delta^{2k+2}] &= \frac{\mathbb{E}(\mathbf{1}^\top \mathbf{X}^\top \bar{\mathbf{y}})^{2k+2}}{\|\mathbf{X}\mathbf{1}\|^{4k+4}} \le (Ck)^{k+1} \frac{M^{2k+2} \|\mathbf{X}^\top \mathbf{X}\mathbf{1}\|^{2k+2} + \sigma^{2k+2} \|\mathbf{X}\mathbf{1}\|^{2k+2}}{\|\mathbf{X}\mathbf{1}\|^{4k+4}} \\
&\le \frac{(Ck)^{k+1} \|\mathbf{X}\|_{\mathrm{op}}^{2k+2} \Big(M + \frac{\sigma}{\|\mathbf{X}\|_{\mathrm{op}}} \Big)^{2k+2}}{\|\mathbf{X}\mathbf{1}\|^{2k+2}}.
\end{aligned}$$

Thus we conclude

$$(II_k) \le (Ck)^{k+1} p \|\mathbf{X}\|_{\mathrm{op}}^{4k+4} \Big(M + \frac{\sigma}{\|\mathbf{X}\|_{\mathrm{op}}} \Big)^{2k+2}.$$

Combining the estimates of (I_k)-(III_k) yields

$$(I_k) + (II_k) + (III_k) \le (C\ell)^{\ell+1} p \|\mathbf{X}\|_{\mathrm{op}}^{4k+4} \Big(M + \frac{\sigma}{\|\mathbf{X}\|_{\mathrm{op}}} \Big)^{2k},$$

which implies that

$$
\begin{aligned}
\mathbb{E}[(F_\ell(\widehat{\mathbf{y}}) - F_\ell(\bar{\mathbf{y}}))^2] &\le C^\ell \ell \sum_{k=0}^{\ell-1} (\sigma\|\mathbf{X}\|_{\mathrm{op}})^{2(\ell-k-1)} (C\ell)^{\ell+1} p \|\mathbf{X}\|_{\mathrm{op}}^{4k+4} \Big(M + \frac{\sigma}{\|\mathbf{X}\|_{\mathrm{op}}}\Big)^{2k} \\
&\le p(C\ell)^{\ell+2} \|\mathbf{X}\|_{\mathrm{op}}^{4\ell} \sum_{k=0}^{\ell-1} \Big(\frac{\sigma}{\|\mathbf{X}\|_{\mathrm{op}}}\Big)^{2(\ell-k-1)} \Big(M + \frac{\sigma}{\|\mathbf{X}\|_{\mathrm{op}}}\Big)^{2k} \\
&\le p\Big(\frac{C\ell}{c_0}\Big)^{2\ell} \|\mathbf{X}\|_{\mathrm{op}}^{4\ell} \Big(M + \frac{\sigma}{\|\mathbf{X}\|_{\mathrm{op}}}\Big)^{2\ell} \\
&\le p\Big(\frac{C\ell M}{c_0^2}\Big)^{2\ell} \|\mathbf{X}\|_{\mathrm{op}}^{4\ell},
\end{aligned}
$$

where we apply Assumption 5.1. Hence using the lower bound of A_ℓ in Assumption 5.1, we have

$$
\mathbb{E}\Big[\Big(\frac{F_\ell(\widehat{\mathbf{y}}) - F_\ell(\bar{\mathbf{y}})}{A_\ell}\Big)^2\Big] \le \Big(\frac{C\ell M}{c_0^4}\Big)^{2\ell} \frac{1}{p} \Big(\frac{p}{n} \vee 1\Big)^{2\ell}.
$$

Bound of second term For the individual difference $\widehat{\mu}^\circ_{d_1} \dots \widehat{\mu}^\circ_{d_t} - \bar{\mu}^\circ_{d_1} \dots \bar{\mu}^\circ_{d_t}$ for $(d_1, \dots, d_t)$ such that $\sum_{i=1}^t d_i = \ell$ and $d_i \ge 2$, we have

$$
\widehat{\mu}^\circ_{d_1} \dots \widehat{\mu}^\circ_{d_t} - \bar{\mu}^\circ_{d_1} \dots \bar{\mu}^\circ_{d_t} = \sum_{i=1}^t \widehat{\mu}^\circ_{d_1} \cdots \widehat{\mu}^\circ_{d_{i-1}} (\widehat{\mu}^\circ_{d_i} - \bar{\mu}^\circ_{d_i}) \bar{\mu}^\circ_{d_{i+1}} \cdots \bar{\mu}^\circ_{d_t},
$$

where $|\widehat{\mu}^\circ_d| \vee |\bar{\mu}^\circ_d| \le 2K_d$ with $K_d := M^d d^{d/2}$. Using $\mathbb{E}[(\widehat{\mu}^\circ_d - \bar{\mu}^\circ_d)^2] \le \mathbb{E}[(\widehat{\mu}_d - \bar{\mu}_d)^2]$, we have

$$
\begin{aligned}
\mathbb{E}\Big[\Big(\widehat{\mu}^\circ_{d_1} \dots \widehat{\mu}^\circ_{d_t} - \bar{\mu}^\circ_{d_1} \dots \bar{\mu}^\circ_{d_t}\Big)^2\Big] &\le t \sum_{i=1}^t \mathbb{E}\Big[(\widehat{\mu}^\circ_{d_1} \cdots \widehat{\mu}^\circ_{d_{i-1}} (\widehat{\mu}^\circ_{d_i} - \bar{\mu}^\circ_{d_i}) \bar{\mu}^\circ_{d_{i+1}} \cdots \bar{\mu}^\circ_{d_t})^2\Big] \\
&\le t 2^{t-1} M^{2\ell} \ell^\ell \sum_{i=1}^t \mathbb{E}[(\widehat{\mu}_{d_i} - \bar{\mu}_{d_i})^2] \le t^2 2^{2t} M^{2\ell} \ell^\ell \delta_{\ell-1},
\end{aligned}
$$

applying the induction hypothesis since $d_i \le \ell - 1$. Recall the bound of $\gamma^{(d_1,\dots,d_t)} =$

$\sum_{s_1,\ldots,s_t=1}^{p} \widetilde{T}_{s_1,\ldots,s_t}^{(d_1,\ldots,d_t)}$ in Lemma 5.2. Using a similar argument as in (5.34), we obtain

$$
\begin{aligned}
&\mathbb{E}\Big[\Big(\frac{1}{A_\ell}\sum_{t=2}^{\ell}\sum_{d_1+\ldots+d_t=\ell:d_j\geq 2}\gamma^{(d_1,\ldots,d_t)}(\widehat{\mu}^\circ_{d_1}\ldots\widehat{\mu}^\circ_{d_t}-\bar{\mu}^\circ_{d_1}\ldots\bar{\mu}^\circ_{d_t})\Big)^2\Big]\\
&\quad\leq \frac{1}{A_\ell^2}\ell^{2\ell}(\ell^{2\ell}(n\wedge p)^2\,\|\mathbf{X}\|_{\mathrm{op}}^{4\ell})(\ell^2 2^{2\ell}M^{2\ell}\ell^\ell\delta_{\ell-1})\leq\Big(\frac{CM}{c_0}\Big)^{2\ell}\ell^{5\ell+2}\Big(\frac{p}{n}\vee 1\Big)^{2\ell-2}\delta_{\ell-1}.
\end{aligned}
$$

Hence, combining the two bounds and using the expression of $\delta_{\ell-1}$, we have

$$
\begin{aligned}
\mathbb{E}[(\widehat{\mu}_\ell-\bar{\mu}_\ell)^2]&\leq\Big(\frac{C\ell}{c_0^2}\Big)^{2\ell}\Big(M+\frac{\sigma}{\|\mathbf{X}\|_{\mathrm{op}}}\Big)^{2\ell}\frac{1}{p}\Big(\frac{p}{n}\vee 1\Big)^{2\ell}+\Big(\frac{CM}{c_0}\Big)^{2\ell}\ell^{5\ell+2}\Big(\frac{p}{n}\vee 1\Big)^{2\ell-2}\delta_{\ell-1}\\
&\leq\Big(\frac{C\ell}{c_0^2}\Big)^{2\ell}\Big(M+\frac{\sigma}{\|\mathbf{X}\|_{\mathrm{op}}}\Big)^{2\ell}\frac{1}{p}\Big(\frac{p}{n}\vee 1\Big)^{2\ell}\\
&\quad+\Big(\frac{CM}{c_0}\Big)^{2\ell}\ell^{5\ell+2}\Big(\frac{p}{n}\vee 1\Big)^{2\ell-2}\cdot\Big(\frac{CM\ell}{c_0^2}\Big)^{3(\ell-1)^2}\Big(M+\frac{\sigma}{\|\mathbf{X}\|_{\mathrm{op}}}\Big)^{2(\ell-1)}\frac{1}{n\wedge p}\Big(\frac{p}{n}\vee 1\Big)^{2(\ell-1)^2}\\
&\leq\Big(\frac{CM\ell}{c_0^2}\Big)^{3\ell^2}\Big(\frac{p}{n}\vee 1\Big)^{2\ell^2}\frac{1}{n\wedge p}=\delta_\ell,
\end{aligned}
$$

concluding the induction. $\square$

5.4.2.3 Completing the proof

By combining Propositions 5.3 and 5.6, we have

$$
\mathbb{E}[(\widehat{\mu}_\ell-\mu_\ell)^2]\leq 2\left(\mathbb{E}[(\widehat{\mu}_\ell-\bar{\mu}_\ell)^2]+\mathbb{E}[(\bar{\mu}_\ell-\mu_\ell)^2]\right)\leq\Big(\frac{CM\ell}{c_0^2}\Big)^{\ell^3}\frac{1}{n\wedge p}\Big(\frac{p}{n}\vee 1\Big)^{2\ell^2},
$$

as desired.

5.4.3 Proof of Corollary 5.1

Lemma 5.8. *For $a>0$, let g' be the conditional version of g on $[-a,a]$, i.e., $g'(A)=\frac{g(A\cap[-a,a])}{g([-a,a])}$. Then $W_1(g,g')\leq$. Furthermore, if g be a 1-subgaussian distribution, then*

Proof. Let g be a probability distribution on $\mathbb{R}$ with CDF F, and let

$$p := \mathbb{P}_g(|X| > a), \qquad m := 1 - p = \mathbb{P}_g(|X| \le a).$$

Define g' to be the truncated and renormalized version of g on $[-a, a]$, with CDF

$$F'(x) = \begin{cases} 0, & x < -a, \\ \dfrac{F(x) - F(-a)}{m}, & -a \le x \le a, \\ 1, & x > a. \end{cases}$$

In one dimension,

$$W_1(g, g') = \int_{-\infty}^{\infty} \bigl|F(x) - F'(x)\bigr| \, dx.$$

Outside $[-a, a]$:

$$\int_{-\infty}^{-a} |F(x) - F'(x)| \, dx + \int_{a}^{\infty} |F(x) - F'(x)| \, dx = \int_{-\infty}^{-a} F(x) \, dx + \int_{a}^{\infty} (1 - F(x)) \, dx = \mathbb{E}_g\bigl[(|X| - a)_+\bigr].$$

Inside $[-a, a]$: for $x \in [-a, a]$,

$$F(x) - F'(x) = \frac{F(-a) - p\,F(x)}{m} =: \Delta(x).$$

Since Δ is monotone and bounded between $F(-a)$ and $F(-a) - p$, we have

$$|\Delta(x)| \le p, \qquad x \in [-a, a].$$

Therefore,

$$\int_{-a}^{a} |F(x) - F'(x)| \, dx \le 2ap.$$

So $W_1(g,g') \le \mathbb{E}_g\big[(|X|-a)_+\big] + 2a\,\mathbb{P}_g(|X|>a)$. $\square$

5.4.4 Proof of Lemma 5.1

We first compute $\mathbb{E}[\sum_{i,j=1}^p \mathbf{H}_{ij}^k]$. Note that $\mathbf{H}_{ij}^k = \langle \mathbf{x}_i, \mathbf{x}_j\rangle^k = \big\langle \mathbf{x}_i^{\otimes k}, \mathbf{x}_j^{\otimes k}\big\rangle$, where $\mathbf{x}_i$ is the i-th column of $\mathbf{X}$, so

$$\begin{aligned}
\mathbb{E}\Big[\sum_{i,j=1}^p \mathbf{H}_{ij}^k\Big] &= \mathbb{E}\left\langle \sum_{i=1}^p \mathbf{x}_i^{\otimes k}, \sum_{j=1}^p \mathbf{x}_j^{\otimes k}\right\rangle = \mathbb{E}\Big\|\sum_{i=1}^p \mathbf{x}_i^{\otimes k}\Big\|_F^2 = \sum_{a_1,\dots,a_k=1}^n \mathbb{E}\Big(\sum_{i=1}^p \mathbf{X}_{a_1,i}\cdots\mathbf{X}_{a_k,i}\Big)^2 \\
&\ge \sum_{a_1,\dots,a_k\in[n]:a_1\ne\ldots\ne a_k} \mathbb{E}\Big(\sum_{i=1}^p \mathbf{X}_{a_1,i}\cdots\mathbf{X}_{a_k,i}\Big)^2 = n(n-1)\dots(n-k+1)\sum_{i,j=1}^p\Big(\frac{\Sigma_{ij}}{p}\Big)^k \\
&\ge (n-k+1)^k p^{-k}\mathbf{1}^\top\Sigma^{\circ k}\mathbf{1} \ge (n-k+1)^k p^{-k+1}\lambda_{\min}(\Sigma^{\circ k}) \ge (n-k+1)^k p^{-k+1}c^k,
\end{aligned}$$

where $\Sigma^{\circ k}$ denotes the kth Hadamard product of Σ, and we use the fact that $\lambda_{\min}(\Sigma^{\circ k}) \ge (\lambda_{\min}(\Sigma))^k$.

Next we bound $\mathsf{Var}(\sum_{i,j=1}^p \mathbf{H}_{ij}^k)$. Writing $\mathbf{X}_i = p^{-1/2}\Sigma^{1/2}\mathbf{Z}_i$ and $\mathbf{Z} := [\mathbf{Z}_1;\dots;\mathbf{Z}_n] \in \mathbb{R}^{n\times p}$, we have

$$\mathsf{Var}\Big(\sum_{i,j=1}^p (\mathbf{X}^\top\mathbf{X})_{ij}^k\Big) = \mathsf{Var}\Big(\sum_{i,j=1}^p p^{-k}(\Sigma^{1/2}\mathbf{Z}^\top\mathbf{Z}\Sigma^{1/2})_{ij}^k\Big) = p^{-2k}\mathsf{Var}\Big(\sum_{i,j=1}^p (\Sigma^{1/2}\mathbf{Z}^\top\mathbf{Z}\Sigma^{1/2})_{ij}^k\Big).$$

Let $f(\mathbf{Z}) = \sum_{i,j=1}^p (\Sigma^{1/2}\mathbf{Z}^\top\mathbf{Z}\Sigma^{1/2})_{ij}^k$. Then for any $\alpha\in[n]$ and $\beta\in[p]$, using $\partial_{\alpha\beta}(\Sigma^{1/2}\mathbf{Z}^\top\mathbf{Z}\Sigma^{1/2})_{ij} = (\mathbf{Z}\Sigma^{1/2})_{\alpha j}(\Sigma^{1/2})_{i\beta} + (\mathbf{Z}\Sigma^{1/2})_{\alpha i}(\Sigma^{1/2})_{j\beta}$, we have

$$\begin{aligned}
\frac{f(\mathbf{Z})}{\partial \mathbf{Z}_{\alpha\beta}} &= k\sum_{i,j=1}^p (\Sigma^{1/2}\mathbf{Z}^\top\mathbf{Z}\Sigma^{1/2})_{ij}^{k-1}\partial_{\alpha\beta}[(\Sigma^{1/2}\mathbf{Z}^\top\mathbf{Z}\Sigma^{1/2})_{ij}] \\
&= 2k\sum_{i,j=1}^p (\Sigma^{1/2}\mathbf{Z}^\top\mathbf{Z}\Sigma^{1/2})_{ij}^{k-1}(\mathbf{Z}\Sigma^{1/2})_{\alpha i}(\Sigma^{1/2})_{j\beta} \\
&= 2k\Big[\mathbf{Z}\Sigma^{1/2}(\Sigma^{1/2}\mathbf{Z}^\top\mathbf{Z}\Sigma^{1/2})^{\circ(k-1)}\Sigma^{1/2}\Big]_{\alpha\beta}.
\end{aligned}$$

By tensorization, the distribution of $\mathbf{Z}$ in $\mathbb{R}^{n\times p}$ satisfies the Poincaré inequality with constant C_{PI} (see [BGL14, Proposition 4.3.1]), hence

$$\begin{aligned}\mathsf{Var}(f(\mathbf{Z})) &\leq 4k^2\mathbb{E}\|\mathbf{Z}\Sigma^{1/2}(\Sigma^{1/2}\mathbf{Z}^\top\mathbf{Z}\Sigma^{1/2})^{\circ(k-1)}\Sigma^{1/2}\|_F^2 \\ &\leq 4k^2\mathbb{E}\|\mathbf{Z}\|_F^2\|\Sigma\|_{\mathrm{op}}^2\|\Sigma\|_{\mathrm{op}}^{2(k-1)}\|\mathbf{Z}\|_{\mathrm{op}}^{4(k-1)} \\ &\leq 4k^2\|\Sigma\|_{\mathrm{op}}^{2k}\cdot\mathbb{E}^{1/2}\|\mathbf{Z}\|_F^4\cdot\mathbb{E}^{1/2}\|\mathbf{Z}\|_{\mathrm{op}}^{8(k-1)} \\ &\leq C^k\|\Sigma\|_{\mathrm{op}}^{2k}(np)\cdot(\sqrt{n}+\sqrt{p})^{4(k-1)} \\ &\leq C^k\|\Sigma\|_{\mathrm{op}}^{2k}(n+p)^{2k},\end{aligned}$$

where the second last inequality follows from [Ver18b, Theorem 4.4.5]. This further implies that $\mathsf{Var}(\sum_{i,j=1}^p(\mathbf{X}^\top\mathbf{X})_{ij}^k) \leq C^k(1+n/p)^{2k}$.

Next we upper bound $\|\mathbf{X}\|_{\mathrm{op}}$ on the right side. Since $\|\mathbf{X}\|_{\mathrm{op}} = \|\mathbf{Z}\Sigma^{1/2}\| \leq \|\Sigma\|_{\mathrm{op}}^{1/2}\|\mathbf{Z}\|_{\mathrm{op}}$, it suffices to bound $\|\mathbf{Z}\|_{\mathrm{op}} = \sup_{u\in\mathbb{S}^{p-1}}\|\mathbf{Z}u\|$. Let $\mathcal{N}$ be a $1/2$-net of $\mathbb{S}^{p-1}$ with cardinality less than 5^p, then for any $u \in \mathbb{S}^{p-1}$, let $u' \in \mathcal{N}$ be such that $\|u-u'\| \leq 1/2$,

$$\|\mathbf{Z}u\| \leq \|\mathbf{Z}(u-u')\| + \|\mathbf{Z}u'\| \leq \frac{1}{2}\sup_{u\in\mathbb{S}^{p-1}}\|\mathbf{Z}u\| + \sup_{u\in\mathcal{N}}\|\mathbf{Z}u\|,$$

yielding that $\|\mathbf{Z}\|_{\mathrm{op}} \leq 2\sup_{u\in\mathbb{S}^{p-1}}\|\mathbf{Z}u\|$. Since for any fixed $u \in \mathcal{N}$, $Z \mapsto \|\mathbf{Z}u\|$ is 1-Lipschitz (with respect to the Frobenius distance on $\mathbb{R}^{n\times p}$), and by tensorization the law of $\mathbf{Z}$ in $\mathbb{R}^{n\times p}$ satisfies a log-Sobolev inequality with constant C_{LSI}, we have $\mathbb{P}(|\|\mathbf{Z}u\| - \mathbb{E}[\|\mathbf{Z}u\|]| \geq t) \leq 2\exp(-ct^2)$ for any $t \geq 0$, where $c > 0$ only depends on C_{LSI}. $\square$

5.5 Proof of lower bounds

5.5.1 Proof of Proposition 5.2

For simplicity, we take the noise variance to be $\sigma^2 = 1$. Recall the block design $\mathbf{X}$ in (5.23), where $m = p/n \in \mathbb{N}$. Let $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)\overset{\text{i.i.d.}}{\sim} g$. Then $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} \sim (g^{(m)})^{\otimes n}$, where $g^{(m)} = \text{Law}\left(\frac{\beta_1+\ldots+\beta_m}{\sqrt{m}}\right)$ is the rescaled m-fold self-convolution of g defined in (5.24).

Thus, $\mathbf{y} = \boldsymbol{\theta} + \epsilon$ has a product distribution $P_g(\mathbf{y}) = (g^{(m)} * \varphi)^{\otimes n}$, where φ is the normal density. So we can bound the total variation between product distribution by

$$\mathsf{TV}(P_g(\mathbf{y}), P_{g'}(\mathbf{y}))) \leq n\mathsf{TV}(g^{(m)} * \varphi, g'^{(m)} * \varphi).$$

The distance between the one-dimensional normal mixture can be bounded using standard moment matching argument:

$$\begin{aligned} 4\mathsf{TV}(g^{(m)} * \varphi, g'^{(m)} * \varphi)^2 &\overset{\text{(a)}}{\leq} \int \frac{\left(g^{(m)} * \varphi - g'^{(m)} * \varphi\right)(y)^2}{\varphi(y)} \mathrm{d}y \\ &\overset{\text{(b)}}{=} \sum_{k=0}^{\infty} \frac{1}{k!}\left(m_k(g^{(m)}) - m_k(\widetilde{g}^{(m)})\right)^2, \end{aligned}$$

where (a) follows from the variational representation $\mathsf{TV}(p,q) = \frac{1}{2}\inf_r \sqrt{\int \frac{(p-q)^2}{r}}$ with infimum is over all probability densities r (cf. [PW24, Exercise I.36]); (b) follows from expanding the mixture density under the Hermite basis whose coefficients are given by the moments of the mixing distribution (see [WY20a, Lemma 9]).

Next we bound the moment difference of the m-fold self-convolutions. Let $\beta_i \overset{\text{i.i.d.}}{\sim} g$ and $\beta_i' \overset{\text{i.i.d.}}{\sim} g'$. Then

$$m_k(g^{(m)}) = m^{-k/2}\mathbb{E}[(\beta_1 + \ldots + \beta_m)^k], \quad m_k(g'^{(m)}) = m^{-k/2}\mathbb{E}[(\beta_1' + \ldots + \beta_m')^k].$$

Let $g, \widetilde{g}$ be two symmetric distributions on $[-1,1]$ with first L moments. Then it is clear that $m_k(g^{\otimes m}) = m_k(\widetilde{g}^{\otimes m})$ for $k \leq L$. For $k > L$,

$$\mathbb{E}[(\beta_1' + \ldots + \beta_m')^k] - \mathbb{E}[(\beta_1' + \ldots + \beta_m')^k] = \sum_{i_1,\ldots,i_k=1}^{m} \mathbb{E}[\beta_{i_1}\ldots\beta_{i_k}] - \mathbb{E}[\beta_{i_1}\ldots\beta_{i_k}]. \tag{5.45}$$

For a given $i_1, \ldots, i_k \in [m]$, denote the multiplicity $f_j = \sum_{\ell=1}^{k} \mathbf{1}_{\{i_\ell = j\}}$ denote the number of times $j \in [m]$ occurs. It is clear that by construction, the contribution is zero unless (a) at least one of the $f_j \geq L+1$ (b) all f_j are even. Therefore, the number b of distinct indices in $[m]$ is at most $\frac{k-L-1}{2}+1$. Enumerating the distinct indices $j_1, \dots, j_b$ and their multiplicities $f_{j_1}, \dots, f_{j_b}$ satisfying $f_{j_\ell} \geq 2$ and $\sum_{\ell=1}^{b} f_{j_\ell} = k$, we conclude that the number of non-vanishing terms in (5.45) is at most

$$\sum_{b=1}^{\frac{k-L+1}{2}} m^b p(k)$$

where $p(k)$ is the partition number of k and satisfies $p(k) \leq \exp(C_0\sqrt{k})$ for some absolute constant C_0 (Hardy-Ramanujun). Combining the above and using that g, g' are supported on $[-1, 1]$, we get

$$|m_k(g^{(m)}) - m_k(g'^{(m)})| \leq m^{-(L-1)/2} k \exp(C_0\sqrt{k}).$$

Finally,

$$4\mathsf{TV}(g^{(m)} * \varphi, g'^{(m)} * \varphi)^2 \leq m^{-L+1} \sum_{k=L+1}^{\infty} \frac{1}{k!}\big(k \exp(C_0\sqrt{k})\big)^2 \leq C_1 m^{-L+1}$$

for some absolute constant.

5.5.2 Proof of Theorem 5.2

For simplicity, we take the error variance $\sigma^2 = 1$. Let g, g' be two distributions supported on $[0, 1]$ with matching first L moments. Let $\boldsymbol{\beta}' \in \mathbb{R}^p$ have iid entries distributed as g' and also be independent of $\boldsymbol{\beta}$. For any $0 \leq k \leq p$, let $\beta^{[k]} \in \mathbb{R}^p$ be a random vector such that $(\boldsymbol{\beta}^{[k]})_\ell = \boldsymbol{\beta}_\ell$ for $1 \leq \ell \leq k$, and $(\boldsymbol{\beta}^{[k]})_\ell = \boldsymbol{\beta}'_\ell$ for $k+1 \leq \ell \leq p$, so that $\boldsymbol{\beta}^{[0]} = \boldsymbol{\beta}'$ and

$\boldsymbol{\beta}^{[p]} = \boldsymbol{\beta}$. Then

$$d_{\mathrm{TV}}(P_g(\mathbf{y}), P_{g'}(\mathbf{y})) = d_{\mathrm{TV}}(\mathbf{X}\boldsymbol{\beta}^{[p]} + \boldsymbol{\varepsilon}, \mathbf{X}\boldsymbol{\beta}^{[0]} + \boldsymbol{\varepsilon}) \le \sum_{k=1}^{p} d_{\mathrm{TV}}(\mathbf{X}\boldsymbol{\beta}^{[k]} + \boldsymbol{\varepsilon}, \mathbf{X}\boldsymbol{\beta}^{[k-1]} + \boldsymbol{\varepsilon}). \tag{5.46}$$

Let $\mathbf{x}_j$ denote the jth column of $\mathbf{X}$. Then for any $1 \le k \le p$, by the data processing inequality and rotational invariance of the total variation distance,

$$\begin{aligned}
&d_{\mathrm{TV}}(\mathbf{X}\boldsymbol{\beta}^{[k]} + \boldsymbol{\varepsilon}, \mathbf{X}\boldsymbol{\beta}^{[k-1]} + \boldsymbol{\varepsilon}) && (5.47)\\
&= d_{\mathrm{TV}}\Big(\sum_{\ell=1}^{k-1} \mathbf{x}_\ell \boldsymbol{\beta}_\ell + \mathbf{x}_k \boldsymbol{\beta}_k + \sum_{\ell=k+1}^{p} \mathbf{x}_\ell \boldsymbol{\beta}'_\ell + \boldsymbol{\varepsilon}, \sum_{\ell=1}^{k-1} \mathbf{x}_\ell \boldsymbol{\beta}_\ell + \mathbf{x}_k \boldsymbol{\beta}'_k + \sum_{\ell=k+1}^{p} \mathbf{x}_\ell \boldsymbol{\beta}'_\ell + \boldsymbol{\varepsilon}\Big)\\
&\le d_{\mathrm{TV}}(\mathbf{x}_k \boldsymbol{\beta}_k + \boldsymbol{\varepsilon}, \mathbf{x}_k \boldsymbol{\beta}'_k + \boldsymbol{\varepsilon}) = d_{\mathrm{TV}}(\|\mathbf{x}_k\| \boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_1, \|\mathbf{x}_k\| \boldsymbol{\beta}'_k + \boldsymbol{\varepsilon}_1). && (5.48)
\end{aligned}$$

Next we apply a standard moment matching argument to bound the distance between these two univariate normal mixtures. Recall the assumption that $\max_{k\in[p]} \|\mathbf{x}_k\| \le C_0(\frac{n}{p})^\alpha (\log p)^\gamma \le 1/2$ which holds for $n \le p^{1-\delta}$ and all large enough p. Applying [WY20b, Theorem 3.3.3, Part 2], we get for some universal constant C,

$$\chi^2(\|\mathbf{x}_k\| \boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_1, \|\mathbf{x}_k\| \boldsymbol{\beta}'_k + \boldsymbol{\varepsilon}_1) \le (C\|\mathbf{x}_k\|)^{2(L+1)}.$$

Finally, combining this with (5.46)–(5.48) and applying $\mathsf{TV}^2 \le \frac{1}{4}\chi^2$ (see, e.g., [PW24, Proposition 7.15]), we get

$$d_{\mathrm{TV}}(P_g(\mathbf{y}), P_{g'}(\mathbf{y})) \le pC^{L+1} \cdot \max_{k\in[p]} \|\mathbf{x}_k\|^{L+1} \le p \left[CC_0 \left(\frac{n}{p}\right)^\alpha (\log p)^\gamma \right]^{L+1} \le 0.1,$$

by choosing L to be large enough (depending only on $C_0, \alpha, \delta, \gamma$). The proof is complete.

□

Bibliography

[ABBR20] Luis Aparicio, Mykola Bordyuh, Andrew J Blumberg, and Raul Rabadan. A random matrix theory approach to denoise single-cell data. *Patterns*, 1(3), 2020. 1

[AEK14a] Oskari Ajanki, László Erdős, and Torben Krüger. Local semicircle law with imprimitive variance matrix. *Electron. Commun. Probab.*, 19:no. 33, 9, 2014. 23

[AEK14b] Oskari Ajanki, Lászlo Erdős, and Torben Krüger. Local semicircle law with imprimitive variance matrix. *Electronic Communications in Probability*, 19:1–9, 2014. 13, 83, 94

[AEK17] Johannes Alt, László Erdős, and Torben Krüger. Local law for random Gram matrices. *Electron. J. Probab.*, 22:Paper No. 25, 41, 2017. 13, 20, 23

[AFT$^+$17] Avanti Athreya, Donniell E Fishkind, Minh Tang, Carey E Priebe, Youngser Park, Joshua T Vogelstein, Keith Levin, Vince Lyzinski, and Yichen Qin. Statistical inference on random dot product graphs: a survey. *The Journal of Machine Learning Research*, 18(1):8393–8484, 2017. 132

[AFW22] Emmanuel Abbe, Jianqing Fan, and Kaizheng Wang. An ℓ^p theory of pca and spectral clustering. *The Annals of Statistics*, 50(4):2359–2385, 2022. 1

[AG14] Ayser Armiti and Michael Gertz. Geometric graph matching and similarity: A probabilistic approach. In *Proceedings of the 26th International Conference on Scientific and Statistical Database Management*, pages 1–12, 2014. 131

[AGZ10] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*. Number 118. Cambridge university press, 2010. 1, 95

[AKS98] Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. *Random Structures & Algorithms*, 13(3-4):457–466, 1998. 3

[AS08] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley-Interscience Series in Discrete Mathematics and Optimization, 3 edition, 2008. 173

[ASX17] Yacine Ait-Sahalia and Dacheng Xiu. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, 201(2):384–399, 2017. 81

[BBAP05] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, pages 1643–1697, 2005. 1, 3, 4, 82, 84, 88, 128

[BCL+19] Boaz Barak, Chi-Ning Chou, Zhixian Lei, Tselil Schramm, and Yueqi Sheng. (nearly) efficient algorithms for the graph matching problem on correlated random graphs. In *Advances in Neural Information Processing Systems*, pages 9186–9194, 2019. 130

[BCRT21] Rina Foygel Barber, Emmanuel J Candes, Aaditya Ramdas, and Ryan J Tibshirani. Predictive inference with the jackknife+. *The Annals of Statistics*, 49(1):486–507, 2021. 81

[BDER16] Sébastien Bubeck, Jian Ding, Ronen Eldan, and Miklós Z Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3):503–532, 2016. 134

[BE02] Olivier Bousquet and André Elisseeff. Stability and generalization. *The Journal of Machine Learning Research*, 2:499–526, 2002. 81

[BEK+14] Alex Bloemendal, László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.*, 19:no. 33, 53, 2014. 3, 8, 13, 18, 20, 23, 83, 94, 96

[BEYY16] Paul Bourgade, Laszlo Erdős, Horng-Tzer Yau, and Jun Yin. Fixed energy universality for generalized Wigner matrices. *Comm. Pure Appl. Math.*, 69(10):1815–1881, 2016. 2, 16, 22

[BG05] Ingwer Borg and Patrick JF Groenen. *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media, 2005. 132

[BG18] Sébastien Bubeck and Shirshendu Ganguly. Entropic clt and phase transition in high-dimensional wishart matrices. *International Mathematics Research Notices*, 2018(2):588–606, 2018. 134

[BGK17] Florent Benaych-Georges and Antti Knowles. Local semicircle law for Wigner matrices. In *Advanced topics in random matrices*, volume 53 of *Panor. Synthèses*, pages 1–90. Soc. Math. France, Paris, 2017. 3

[BGL14] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014. 233

[BGN11] Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011. 1, 4, 82, 84, 88, 129

[BHMM19] Mikhail Belkin, Daniel Hsu, Siyuan Ma, and Soumik Mandal. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32):15849–15854, 2019. 1, 4

[BKS99] Itai Benjamini, Gil Kalai, and Oded Schramm. Noise sensitivity of Boolean functions and applications to percolation. *Inst. Hautes Études Sci. Publ. Math.*, (90):5–43 (2001), 1999. 81

[BKYY16] Alex Bloemendal, Antti Knowles, Horng-Tzer Yau, and Jun Yin. On the principal components of sample covariance matrices. *Probab. Theory Related Fields*, 164(1-2):459–552, 2016. 3, 8, 84, 112, 113

[BL22] Charles Bordenave and Jaehun Lee. Noise sensitivity for the top eigenvector of a sparse random matrix. *Electronic Journal of Probability*, 27:1–50, 2022. 6, 88

[BLZ20] Charles Bordenave, Gábor Lugosi, and Nikita Zhivotovskiy. Noise sensitivity of the top eigenvector of a Wigner matrix. *Probability Theory and Related Fields*, 177(3):1103–1135, 2020. 6, 88, 90, 91

[Bou22] Paul Bourgade. Extreme gaps between eigenvalues of Wigner matrices. *Journal of the European Mathematical Society*, 24(8):2823–2873, 2022. 9, 25, 57

[BP09] Jean-Philippe Bouchaud and Marc Potters. Financial applications of random matrix theory: a short review. *arXiv preprint arXiv:0910.1205*, 2009. 1

[BPZ15] Zhigang Bao, Guangming Pan, and Wang Zhou. Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.*, 43(1):382–421, 2015. 74, 75, 76, 77

[Bro08] Lawrence D Brown. In-season prediction of batting averages: A field test of empirical bayes and bayes methodologies. *The Annals of Applied Statistics*, pages 113–152, 2008. 188

[Bru89] Marie-France Bru. Diffusions of perturbed principal component analysis. *J. Multivariate Anal.*, 29(1):127–136, 1989. 15

[BS06] Jinho Baik and Jack W Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of multivariate analysis*, 97(6):1382–1408, 2006. 82

[BS10a] Zhidong Bai and Jack William Silverstein. *Spectral analysis of large dimensional random matrices*. Springer, 2010. 1

[BS10b] Mikhail Belkin and Kaushik Sinha. Polynomial learning of distribution families. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 103–112. IEEE, 2010. 192

[BSLF+15] Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Nick Patterson, Mark J Daly, Alkes L Price, and Benjamin M Neale. Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics*, 47(3):291–295, 2015. 192

[BY17] P. Bourgade and H.-T. Yau. The eigenvector moment flow and local quantum unique ergodicity. *Comm. Math. Phys.*, 350(1):231–278, 2017. 2, 3, 15, 48

[BYY14] Paul Bourgade, Horng-Tzer Yau, and Jun Yin. Local circular law for random matrices. *Probability Theory and Related Fields*, 159(3-4):545–595, 2014. 2, 23

[Cas85] George Casella. An introduction to empirical bayes data analysis. *The American Statistician*, 39(2):83–87, 1985. 188

[CCF+21] Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, et al. Spectral methods for data science: A statistical perspective. *Foundations and Trends® in Machine Learning*, 14(5):566–806, 2021. 1

[CD11] Romain Couillet and Merouane Debbah. *Random matrix methods for wireless communications*. Cambridge University Press, 2011. 1

[Cha05] Sourav Chatterjee. Concentration inequalities with exchangeable pairs. page 105, 2005. Thesis (Ph.D.)–Stanford University. 90, 91

[Cha07] Sourav Chatterjee. Stein's method for concentration inequalities. *Probab. Theory Related Fields*, 138(1-2):305–321, 2007. 118

[CK16] Daniel Cullina and Negar Kiyavash. Improved achievability and converse bounds for Erdös-Rényi graph matching. In *Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science*, pages 63–72. ACM, 2016. 130, 135

[CK17] Daniel Cullina and Negar Kiyavash. Exact alignment recovery for correlated Erdös-Rényi graphs. *arXiv preprint arXiv:1711.06783*, 2017. 130

[CKK+10] M. Chertkov, L. Kroc, F. Krzakala, M. Vergassola, and L. Zdeborová. Inference in particle tracking experiments by passing messages between images. *PNAS*, 107(17):7663–7668, 2010. 133

[CL19] Ziliang Che and Patrick Lopatto. Universality of the least singular value for sparse random matrices. *Electron. J. Probab.*, 24:Paper No. 9, 53, 2019. 8, 16, 46

[CL22] Romain Couillet and Zhenyu Liao. *Random matrix methods for machine learning*. Cambridge University Press, 2022. 1, 3

[CLMW11a] E. J Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):11, 2011. 1

[CLMW11b] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):1–37, 2011. 82, 89

[CMS13] Claudio Cacciapuoti, Anna Maltsev, and Benjamin Schlein. Local Marchenko-Pastur law at the hard edge of sample covariance matrices. *J. Math. Phys.*, 54(4):043302, 13, 2013. 8, 20, 23

[DCK19] Osman E Dai, Daniel Cullina, and Negar Kiyavash. Database alignment with Gaussian features. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3225–3233. PMLR, 2019. 6, 130, 132, 145, 176, 177

[DCK20] Osman Emre Dai, Daniel Cullina, and Negar Kiyavash. Achievability of nearly-exact alignment for correlated gaussian databases. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 1230–1235. IEEE, 2020. 132, 165

[DGNT11] Thong T Do, Lu Gan, Nam H Nguyen, and Trac D Tran. Fast and efficient compressive sensing using structurally random matrices. *IEEE Transactions on signal processing*, 60(1):139–154, 2011. 1

[DHS21] Zhun Deng, Hangfeng He, and Weijie Su. Toward better generalization bounds with locally elastic stability. In *International Conference on Machine Learning*, pages 2590–2600. PMLR, 2021. 81

[DIC16] LEE H DICKER. Ridge regression and asymptotic minimax estimation over spheres of growing dimension. *Bernoulli*, pages 1–37, 2016. 1, 4

[DL17] Nadav Dym and Yaron Lipman. Exact recovery with symmetries for procrustes matching. *SIAM Journal on Optimization*, 27(3):1513–1530, 2017. 132

[DMWX21] Jian Ding, Zongming Ma, Yihong Wu, and Jiaming Xu. Efficient random graph matching via degree profiles. *Probability Theory and Related Fields*, 179(1):29–115, 2021. 130, 137

[DR14] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407, 2014. 81

[DT11] Momar Dieng and Craig A Tracy. Application of random matrix theory to multivariate statistics. In *Random Matrices, Random Processes and Integrable Systems*, pages 443–507. Springer, 2011. 81

[DWXY21] Jian Ding, Yihong Wu, Jiaming Xu, and Dana Yang. The planted matching problem: Sharp threshold and infinite-order phase transition. *arXiv preprint arXiv:2103.09383*, 2021. 133, 143, 144, 166, 172

[DY18] Xiucai Ding and Fan Yang. A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices. *Ann. Appl. Probab.*, 28(3):1679–1738, 2018. 3, 8, 9, 72, 84, 89, 93, 94, 96, 112, 113

[Dys62a] Freeman J Dyson. Statistical theory of the energy levels of complex systems. i. *Journal of Mathematical Physics*, 3(1):140–156, 1962. 1

[Dys62b] Freeman J Dyson. Statistical theory of the energy levels of complex systems. ii. *Journal of Mathematical Physics*, 3(1):157–165, 1962. 1

[Dys62c] Freeman J Dyson. Statistical theory of the energy levels of complex systems. iii. *Journal of Mathematical Physics*, 3(1):166–175, 1962. 1

[Ede88] Alan Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4):543–560, 1988. 11

[EEPK05] Andre Elisseeff, Theodoros Evgeniou, Massimiliano Pontil, and Leslie Pack Kaelbing. Stability of randomized learning algorithms. *Journal of Machine Learning Research*, 6(1), 2005. 81

[Efr12] Bradley Efron. *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, volume 1. Cambridge University Press, 2012. 188

[EK06] Noureddine El Karoui. A rate of convergence result for the largest eigenvalue of complex white Wishart matrices. *Ann. Probab.*, 34(6):2077–2117, 2006. 8, 70

[EK07] Noureddine El Karoui. Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.*, 35(2):663–714, 2007. 77

[EKYY13] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Spectral statistics of erdős–rényi graphs i: local semicircle law. *The Annals of Probability*, 41(3B):2279–2375, 2013. 65

[ESYY12] László Erdős, Benjamin Schlein, Horng-Tzer Yau, and Jun Yin. The local relaxation flow approach to universality of the local statistics for random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.*, 48(1):1–46, 2012. 2, 12, 16

[ETST01] Bradley Efron, Robert Tibshirani, John D Storey, and Virginia Tusher. Empirical bayes analysis of a microarray experiment. *Journal of the American statistical association*, 96(456):1151–1160, 2001. 188

[EY17a] László Erdős and Horng-Tzer Yau. *A dynamical approach to random matrix theory*, volume 28 of *Courant Lecture Notes in Mathematics*. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2017. 3, 13, 21, 33, 62, 68

[EY17b] László Erdős and Horng-Tzer Yau. *A dynamical approach to random matrix theory*, volume 28 of *Courant Lecture Notes in Mathematics*. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2017. 56, 60

[EYY11] László Erdős, Horng-Tzer Yau, and Jun Yin. Universality for generalized Wigner matrices with Bernoulli distribution. *J. Comb.*, 2(1):15–81, 2011. 2, 68, 70

[EYY12] László Erdős, Horng-Tzer Yau, and Jun Yin. Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields*, 154(1-2):341–407, 2012. 55

[FGSW23] Zhou Fan, Leying Guan, Yandi Shen, and Yihong Wu. Gradient flows for empirical bayes in high-dimensional linear models. *arXiv preprint arXiv:2312.12708*, 2023. 7, 189, 190

[FLW16] Jianqing Fan, Yuan Liao, and Weichen Wang. Projected principal component analysis in factor models. *Annals of statistics*, 44(1):219, 2016. 1

[FMWX19a] Zhou Fan, Cheng Mao, Yihong Wu, and Jiaming Xu. Spectral graph matching and regularized quadratic relaxations I: The Gaussian model. *arxiv preprint arXiv:1907.08880*, 2019. 130

[FMWX19b] Zhou Fan, Cheng Mao, Yihong Wu, and Jiaming Xu. Spectral graph matching and regularized quadratic relaxations II: Erdős-Rényi graphs and universality. *arxiv preprint arXiv:1907.08883*, 2019. 130

[FMWX20] Zhou Fan, Cheng Mao, Yihong Wu, and Jiaming Xu. Spectral graph matching and regularized quadratic relaxations: Algorithm and theory. In *International conference on machine learning*, pages 2985–2995. PMLR, 2020. 3

[FMWX22] Zhou Fan, Cheng Mao, Yihong Wu, and Jiaming Xu. Spectral graph matching and regularized quadratic relaxations II. *Foundations of Computational Mathematics*, pages 1–51, 2022. 3

[FS09] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University Press, 2009. 157

[FWZ18] Jianqing Fan, Weichen Wang, and Yiqiao Zhong. An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *Journal of Machine Learning Research*, 18(207):1–42, 2018. 82, 88

[Gan21a] Luca Ganassali. Sharp threshold for alignment of graph databases with Gaussian weights. *Mathematical and Scientific Machine Learning (MSML21)*, 2021. arXiv preprint arXiv:2010.16295. 130

[Gan21b] Luca Ganassali. Sharp threshold for alignment of graph databases with gaussian weights. In *MSML21 (Mathematical and Scientific Machine Learning)*, 2021. 134

[GF00] EdwardI George and Dean P Foster. Calibration and empirical bayes variable selection. *Biometrika*, 87(4):731–747, 2000. 189

[Gil61] Edward N Gilbert. Random plane networks. *Journal of the society for industrial and applied mathematics*, 9(4):533–543, 1961. 132

[GJB19] Edouard Grave, Armand Joulin, and Quentin Berthet. Unsupervised alignment of embeddings with wasserstein procrustes. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1880–1890. PMLR, 2019. 132, 136

[GM20] Luca Ganassali and Laurent Massoulié. From tree matching to sparse graph alignment. *arXiv preprint arXiv:2002.01258*, 2020. 130

[GMGW98] Thomas Guhr, Axel Müller-Groeling, and Hans A Weidenmüller. Random-matrix theories in quantum physics: common concepts. *Physics Reports*, 299(4-6):189–425, 1998. 1

[GML22] Luca Ganassali, Laurent Massoulié, and Marc Lelarge. Correlation detection in trees for planted graph alignment. In *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022. 130

[GS14] Christophe Garban and Jeffrey E Steif. *Noise sensitivity of Boolean functions and percolation*, volume 5. Cambridge University Press, 2014. 81

[Hal05] Alastair R Hall. *Generalized Method of Moments*. Oxford University Press, 2005. 192

[Han82] Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the econometric society*, pages 1029–1054, 1982. 192, 196

[Hen53] Charles R Henderson. Estimation of variance and covariance components. *Biometrics*, 9(2):226–252, 1953. 192

[HL19] Jiaoyang Huang and Benjamin Landon. Rigidity and a mesoscopic central limit theorem for Dyson Brownian motion for general β and potentials. *Probab. Theory Related Fields*, 175(1-2):209–253, 2019. 25

[HM20] Georgina Hall and Laurent Massoulié. Partial recovery in the graph alignment problem. *arXiv preprint arXiv:2007.00533*, 2020. 130

[HMRT22] Trevor Hastie, Andrea Montanari, Saharon Rosset, and Ryan J Tibshirani. Surprises in high-dimensional ridgeless least squares interpolation. *Annals of statistics*, 50(2):949, 2022. 1, 4

[HP15] Moritz Hardt and Eric Price. Tight bounds for learning a mixture of two gaussians. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 753–760, 2015. 192

[HRS16] Moritz Hardt, Ben Recht, and Yoram Singer. Train faster, generalize better: Stability of stochastic gradient descent. In *International conference on machine learning*, pages 1225–1234. PMLR, 2016. 81

[HWX17] B. Hajek, Y. Wu, and J. Xu. Information limits for recovering a hidden community. *IEEE Trans. on Information Theory*, 63(8):4729 – 4745, 2017. 143, 163

[JL09] Iain M Johnstone and Arthur Yu Lu. Sparse principal components analysis. *arXiv preprint arXiv:0901.4392*, 2009. 1, 129

[JL15] Tiefeng Jiang and Danning Li. Approximation of rectangular beta-laguerre ensembles and large deviations. *Journal of Theoretical Probability*, 28(3):804–847, 2015. 134

[Joh00] Kurt Johansson. Shape fluctuations and random matrices. *Comm. Math. Phys.*, 209(2):437–476, 2000. 8

[Joh01] Iain M Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of statistics*, 29(2):295–327, 2001. 1, 3, 4, 8, 83

[Joh07] Iain M. Johnstone. High dimensional statistical inference and random matrices. In *International Congress of Mathematicians. Vol. I*, pages 307–333. Eur. Math. Soc., Zürich, 2007. 1, 4, 81

[Kar08] Noureddine El Karoui. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *The Annals of Statistics*, pages 2757–2790, 2008. 4

[KB21] Byol Kim and Rina Foygel Barber. Black box tests for algorithmic stability. *arXiv preprint arXiv:2111.15546*, 2021. 81

[KMV10] Adam Tauman Kalai, Ankur Moitra, and Gregory Valiant. Efficiently learning mixtures of two gaussians. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 553–562, 2010. 192

[KN02] Samuel Kutin and Partha Niyogi. Almost-everywhere algorithmic stability and generalization error. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, pages 275–282, 2002. 81

[KN09] Shira Kritchman and Boaz Nadler. Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory. *IEEE Transactions on Signal Processing*, 57(10):3930–3941, 2009. 1, 4

[KNW22] Dmitriy Kunisky and Jonathan Niles-Weed. Strong recovery of geometric planted matchings. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 834–876. SIAM, 2022. 6, 133, 134, 138, 144, 165, 172, 173, 174, 176

[KW56] Jack Kiefer and Jacob Wolfowitz. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *The Annals of Mathematical Statistics*, pages 887–906, 1956. 189

[KY17] Antti Knowles and Jun Yin. Anisotropic local laws for random matrices. *Probab. Theory Related Fields*, 169(1-2):257–352, 2017. 3, 8, 73, 74

[LBG$^+$15] Po-Ru Loh, Gaurav Bhatia, Alexander Gusev, Hilary K Finucane, Brendan K Bulik-Sullivan, Samuela J Pollack, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Teresa R de Candia, Sang Hong Lee, Naomi R Wray, et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nature genetics*, 47(12):1385–1392, 2015. 193

[LR10] Michel Ledoux and Brian Rider. Small deviations for beta ensembles. *Electronic Journal of Probability*, 15:1319–1343, 2010. 95

[LRB$^+$16] Z Lähner, Emanuele Rodolà, MM Bronstein, Daniel Cremers, Oliver Burghard, Luca Cosmo, Andreas Dieckmann, Reinhard Klein, and Y Sahillioglu. SHREC'16: Matching of deformable shapes with topological noise. *Proc. 3DOR*, 2(10.2312), 2016. 131

[LS15] Ji Oon Lee and Kevin Schnelli. Edge universality for deformed Wigner matrices. *Rev. Math. Phys.*, 27(8):1550018, 94, 2015. 76

[LS16] Ji Oon Lee and Kevin Schnelli. Tracy-Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *Ann. Appl. Probab.*, 26(6):3786–3839, 2016. 72, 74, 76, 77, 78, 79, 93, 94

[LS17] Jerry Li and Ludwig Schmidt. Robust and proper learning for mixtures of gaussians via systems of polynomial inequalities. In *Conference on Learning Theory*, pages 1302–1382. PMLR, 2017. 192

[LS18] Ji Oon Lee and Kevin Schnelli. Local law and tracy–widom limit for sparse random matrices. *Probability Theory and Related Fields*, 171(1):543–616, 2018. 65

[LSY19] Benjamin Landon, Philippe Sosoe, and Horng-Tzer Yau. Fixed energy universality of Dyson Brownian motion. *Adv. Math.*, 346:1137–1332, 2019. 22

[Lu02] Arthur Yu Lu. *Sparse principal component analysis for functional data*. Stanford University, 2002. 129

[LW03a] Olivier Ledoit and Michael Wolf. Honey, i shrunk the sample covariance matrix. 2003. 4

[LW03b] Olivier Ledoit and Michael Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5):603–621, 2003. 4

[LW04] Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411, 2004. 4

[LW17] Olivier Ledoit and Michael Wolf. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *The Review of Financial Studies*, 30(12):4349–4388, 2017. 4

[LZYZ06] Feng Luo, Jianxin Zhong, Yunfeng Yang, and Jizhong Zhou. Application of random matrix theory to microarray data for discovering functional gene modules. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 73(3):031924, 2006. 1

[Ma12] Zongming Ma. Accuracy of the Tracy-Widom limits for the extreme eigenvalues in white Wishart matrices. *Bernoulli*, 18(1):322–359, 2012. 3, 8, 70

[Mat13] Sho Matsumoto. Weingarten calculus for matrix ensembles associated with compact symmetric spaces. *Random Matrices: Theory and Applications*, 2(02):1350001, 2013. 135

[MDK+16] Haggai Maron, Nadav Dym, Itay Kezurer, Shahar Kovalsky, and Yaron Lipman. Point registration via efficient convex relaxation. *ACM Transactions on Graphics (TOG)*, 35(4):1–12, 2016. 132

[MM21] Charles H Martin and Michael W Mahoney. Implicit self-regularization in deep neural networks: Evidence from random matrix theory and implications for learning. *Journal of Machine Learning Research*, 22(165):1–73, 2021. 1

[MMX21] Mehrdad Moharrami, Cristopher Moore, and Jiaming Xu. The planted matching problem: Phase transitions and exact results. *The Annals of Applied Probability*, 31(6):2663–2720, 2021. 133

[MNPR06] Sayan Mukherjee, Partha Niyogi, Tomaso Poggio, and Ryan Rifkin. Learning theory: stability is sufficient for generalization and necessary and sufficient for consistency of empirical risk minimization. *Advances in Computational Mathematics*, 25(1):161–193, 2006. 81

[MP67] Vladimir A Marcenko and Leonid Andreevich Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, 1967. 2, 8

[MP21] Gabriel Mel and Jeffrey Pennington. Anisotropic random feature regression in high dimensions. In *International Conference on Learning Representations*, 2021. 4

[MRT21a] Cheng Mao, Mark Rudelson, and Konstantin Tikhomirov. Exact matching of random graphs with constant correlation. *arXiv preprint arXiv:2110.05000*, 2021. 130

[MRT21b] Cheng Mao, Mark Rudelson, and Konstantin Tikhomirov. Random graph matching with improved noise robustness. In *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 3296–3329, 2021. 130

[MRZ15] Andrea Montanari, Daniel Reichman, and Ofer Zeitouni. 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, 2015. 3

[MSS23] Sumit Mukherjee, Bodhisattva Sen, and Subhabrata Sen. A mean field approach to empirical bayes estimation in high-dimensional linear regression. *arXiv preprint arXiv:2309.16843*, 2023. 7, 189, 190

[MV10] Ankur Moitra and Gregory Valiant. Settling the polynomial learnability of mixtures of gaussians. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 93–102. IEEE, 2010. 192

[Nad14] Raj Rao Nadakuditi. 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, 2014. 4

[NS86] Fassil Nebebe and TWF Stroud. Bayes and empirical bayes shrinkage estimation of regression coefficients. *Canadian Journal of Statistics*, 14(4):267–280, 1986. 189

[O'C21] Luke J O'Connor. The distribution of common-variant effect sizes. *Nature genetics*, 53(8):1243–1249, 2021. 193

[OM00] Paul Ormerod and Craig Mounfield. Random matrix theory and the failure of macro-economic forecasts. *Physica A: Statistical Mechanics and its Applications*, 280(3-4):497–504, 2000. 1

[OMK10] Sewoong Oh, Andrea Montanari, and Amin Karbasi. Sensor network localization from local connectivity: Performance analysis for the MDS-MAP algorithm. In *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)*, pages 1–5. IEEE, 2010. 137

[OSH+19] Luke J O'Connor, Armin P Schoech, Farhad Hormozdiari, Steven Gazal, Nick Patterson, and Alkes L Price. Extreme polygenicity of complex traits is explained by negative selection. *The American journal of human genetics*, 105(3):456–476, 2019. 193

[Ṕ09] Sandrine Péché. Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probab. Theory Related Fields*, 143(3-4):481–516, 2009. 8

[Pau07] Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642, 2007. 1, 4, 82, 88

[PB17] Jeffrey Pennington and Yasaman Bahri. Geometry of neural network loss surfaces via random matrix theory. In *International conference on machine learning*, pages 2798–2806. PMLR, 2017. 1

[Pea94] Karl Pearson. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, 185:71–110, 1894. 191

[Pen03] Mathew Penrose. *Random geometric graphs*, volume 5. OUP Oxford, 2003. 132

[PG11] Pedram Pedarsani and Matthias Grossglauser. On the privacy of anonymized networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1235–1243. ACM, 2011. 130

[Pis99] G. Pisier. *The volume of convex bodies and Banach space geometry*. Cambridge University Press, 1999. 148

[PW17] Jeffrey Pennington and Pratik Worah. Nonlinear random matrix theory for deep learning. *Advances in neural information processing systems*, 30, 2017. 1

[PW24] Yury Polyanskiy and Yihong Wu. *Information Theory: From Coding to Learning*. Cambridge University Press, 2024. Available at: http://www.stat.yale.edu/~yw562/teaching/itbook-export.pdf. 234, 236

[PY12] Natesh S. Pillai and Jun Yin. Edge universality of correlation matrices. *Ann. Statist.*, 40(3):1737–1763, 2012. 3

[PY14] Natesh S. Pillai and Jun Yin. Universality of covariance matrices. *Ann. Appl. Probab.*, 24(3):935–1001, 2014. 3, 8, 13, 18, 19, 56, 58, 60, 68, 76, 79, 94, 96

[Rao71] C Radhakrishna Rao. Estimation of variance and covariance components–MINQUE theory. *Journal of multivariate analysis*, 1(3):257–275, 1971. 192

[Rao72] C Radhakrishna Rao. Estimation of variance and covariance components in linear models. *Journal of the American Statistical Association*, 67(337):112–115, 1972. 192

[Rin08] Markus Ringnér. What is principal component analysis? *Nature biotechnology*, 26(3):303–304, 2008. 81

[Rob50] Herbert Robbins. A generalization of the method of maximum likelihood-estimating a mixing distribution. In *Annals of mathematical statistics*, volume 21, pages 314–315. INST MATHEMATICAL STATISTICS IMS BUSINESS OFFICE-SUITE 7, 3401 INVESTMENT ..., 1950. 189

[Rob51] Herbert Robbins. Asymptotically subminimax solutions of compound statistical decision problems. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 131–149. University of California Press, 1951. 188

[Rob56] Herbert Robbins. An empirical bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, volume 3, pages 157–164. University of California Press, 1956. 188

[RR07] Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20, 2007. 4

[RV13] Mark Rudelson and Roman Vershynin. Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.*, 18:no. 82, 9, 2013. 185

[Sos02] Alexander Soshnikov. A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Statist. Phys.*, 108(5-6):1033–1056, 2002. 8

[SRZF03] Yi Shang, Wheeler Ruml, Ying Zhang, and Markus PJ Fromherz. Localization from mere connectivity. In *Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing*, pages 201–212, 2003. 137

[Sta11] Richard P Stanley. Enumerative combinatorics volume 1 second edition. *Cambridge studies in advanced mathematics*, 2011. 204

[SW09] Galen R. Shorack and Jon A. Wellner. *Empirical processes with applications to statistics*, volume 59 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009. Reprint of the 1986 original [MR0838963]. 20

[SX21a] Kevin Schnelli and Yuanyuan Xu. Convergence rate to the Tracy–Widom laws for the largest eigenvalue of sample covariance matrices. *arXiv preprint arXiv:2108.02728*, 2021. 9, 94, 95

[SX21b] Kevin Schnelli and Yuanyuan Xu. Convergence rate to the Tracy-Widom laws for the largest eigenvalue of sample covariance matrices. *arXiv preprint arXiv:2108.02728*, 2021. 9

[SX22] Kevin Schnelli and Yuanyuan Xu. Convergence rate to the Tracy-Widom laws for the largest eigenvalue of Wigner matrices. *Comm. Math. Phys.*, 393(2):839–907, 2022. 9

[Tro12] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012. 89

[TV10] Terence Tao and Van Vu. Random matrices: the distribution of the smallest singular values. *Geom. Funct. Anal.*, 20(1):260–297, 2010. 8

[TV12] Terence Tao and Van Vu. Random covariance matrices: universality of local statistics of eigenvalues. *Ann. Probab.*, 40(3):1285–1315, 2012. 2, 55, 95

[Ume88] Shinji Umeyama. An eigendecomposition approach to weighted graph matching problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(5):695–703, 1988. 137

[Ver18a] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018. 142

[Ver18b] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018. 233

[VH96] Jay M Ver Hoef. Parametric empirical bayes methods for ecological applications. *Ecological Applications*, 6(4):1047–1055, 1996. 188

[VH14] Ramon Van Handel. Probability in high dimension. Technical report, PRINCETON UNIV NJ, 2014. 62

[VK06] Seema Vyas and Lilani Kumaranayake. Constructing socio-economic status indices: how to use principal components analysis. *Health policy and planning*, 21(6):459–468, 2006. 81

[Wan12] Ke Wang. Random covariance matrices: Universality of local statistics of eigenvalues up to the edge. *Random Matrices: Theory and Applications*, 1(01):1150005, 2012. 95

[Wan22] Haoyu Wang. Optimal smoothed analysis and quantitative universality for the smallest singular value of random matrices. *arXiv preprint arXiv:2211.03975*, 2022. 9, 83, 94

[Wan24] Haoyu Wang. Quantitative universality for the largest eigenvalue of sample covariance matrices. *The Annals of Applied Probability*, 34(3):2539–2565, 2024. 9, 94, 95

[Wig51] Eugene P Wigner. On the statistical distribution of the widths and spacings of nuclear resonance levels. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 47, pages 790–798. Cambridge University Press, 1951. 1

[Wig55] Eugene P Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, 62(3):548–564, 1955. 1

[WXY21] Yihong Wu, Jiaming Xu, and Sophie H. Yu. Settling the sharp reconstruction thresholds of random graph matching. *arXiv preprint2102.00082*, 2021. 130, 134

[WY20a] Yihong Wu and Pengkun Yang. Optimal estimation of gaussian mixtures via denoised method of moments. *Annals of Statistics*, 48(4), 2020. 191, 192, 196, 197, 234

[WY20b] Yihong Wu and Pengkun Yang. Polynomial methods in statistical inference: theory and practice. *Monograph in* Foundations and Trends in Communications and Information Theory, 17(4):402–586, Oct 2020. 192, 236

[XCS10] Huan Xu, Constantine Caramanis, and Sujay Sanghavi. Robust pca via outlier pursuit. *Advances in neural information processing systems*, 23, 2010. 89

[YL05] Ming Yuan and Yi Lin. Efficient empirical bayes variable selection and estimation in linear models. *Journal of the American Statistical Association*, 100(472):1215–1225, 2005. 189

[YLGV11] Jian Yang, S Hong Lee, Michael E Goddard, and Peter M Visscher. Gcta: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, 88(1):76–82, 2011. 193

[YP16] Joongyeub Yeo and George Papanicolaou. Random matrix approach to estimation of high-dimensional factor models. *arXiv preprint arXiv:1611.05571*, 2016. 1

[ZCW22] Anru R Zhang, T Tony Cai, and Yihong Wu. Heteroskedastic pca: Algorithm, optimality, and applications. *The Annals of Statistics*, 50(1):53–80, 2022. 1

[Zha03] Cun-Hui Zhang. Compound decision theory and empirical bayes methods. *Annals of Statistics*, pages 379–390, 2003. 188

[Zho17] Xiang Zhou. A unified framework for variance component estimation with summary statistics in genome-wide association studies. *The Annals of Applied Statistics*, 11(4):2027, 2017. 1, 193