Random Matrices in High-dimensional Data Analysis

Xiuyuan Cheng

A Dissertation
Presented to the Faculty of Princeton University
in Candidacy for the Degree of Doctor of Philosophy

Recommended for Acceptance
by the Program in Applied and Computational Mathematics
Adviser: Professor Amit Singer

November 2013
Abstract

This thesis studies the spectrum of kernel matrices built from high-dimensional data vectors, a mathematical problem that naturally arises in many applications.

In the first part, we consider the spectrum of large kernel matrices built from independent random high-dimensional vectors (the null model). Specifically, we consider \( n \)-by-\( n \) matrices whose \((i,j)\)-th entry is \( f(X_i^T X_j) \), where \( X_1, \ldots, X_n \) are i.i.d. random vectors in \( \mathbb{R}^p \), and \( f \) belongs to a large class of real-valued functions. As \( p, n \to \infty \) and \( p/n \to \gamma \), we obtain a family of limiting spectral densities which includes the Marčenko-Pastur density and semi-circle density as special cases. The convergence of the spectral density is firstly proved for i.i.d. normal Gaussian vectors, and then extended to i.i.d vectors that can be “compared” with the normal Gaussian vectors. The study of the null model is fundamental towards understanding noise-corrupted kernel matrices, which are built from vectors admitting a decomposition of “signal + noise” (the “spiking” model). We provide conjectures for the spiking model based on our results for the null model.

The second part addresses the application in cryo-EM, where certain kernel matrices built from microscopic image data are used to study the structure of biological molecules. We consider the situation where the molecule admits non-trivial group symmetries, and study (i) the symmetry detection problem and (ii) the structural reconstruction problem. For the former, we derive a theoretical solution based on estimating the rank of certain auto-correlation kernels. For the later, we propose two approaches extending the existing methods developed for non-symmetric molecules. For both problems the proposed methods are tested on simulated data sets. The cryo-EM problem together with other applications motivates the study of the random matrix model in the first part of the thesis.
Acknowledgements

It is my pleasure to thank the following people who have made my graduate study a valuable experience:

First and foremost, I would like to thank my advisor Professor Amit Singer. It is hard to imagine my Ph.D. work had I not received his academic guidance and personal support over the years. It is Amit who taught me the fundamentals of data sciences when I started working with him, who suggested considering kernel matrices as a random matrix problem and believed that it can be treated with mathematical rigor, who read and corrected every single error in the manuscript of our paper, and who listened and thought for me when I was making career decisions. I deeply appreciate the patience that he put into my education, and what I have learned from him is far more than mathematics. I feel fortunate to have been his student.

I would like to thank my committee members, Professor Yakov Sinai and Professor Ramon van Handel, for their help along the way. I thank Professor Sinai for the guidance on the random matrix problem, and for providing me the opportunity to talk in the Ergodic Theory and Statistical Mechanics Seminar. Thanks to Ramon for many useful discussions since my starting working on the random matrix problem, and for reading my thesis and providing detailed comments.

I am also grateful to many other people who have helped to make this thesis possible. I would like to thank Professor Sasha Sodin, from whom I learned about a lot of recent progress in random matrix theory. Thanks to Sasha for reading through my thesis and providing revision suggestions. I am thankful to Professor Yoel Shkolnisky, with whom I collaborate on the cryo-EM project. I had a wonderful time visiting Yale University for two weeks in the summer of 2012 when I worked together with Yoel. His talent and enthusiasm for research have a deep influence on me. This is the opportunity to thank Lanhui Wang and Zhizhen (Jane) Zhao, whose Ph.D. works are closely related to mine. I thank Lanhui and Jane for sharing with me the code of
their work on the cryo-EM project, and for the numerous discussions that we had in our shared office.

I would like to express my gratitude to: Professor Weinan E for the guidance on my graduate study; Professor Charlie Fefferman for discussing the random matrix problem and the proof of the theorem; Professor Gerard Ben Arous and Professor Jianqing Fan for useful communications.

I am very thankful to the graduate students and staff at PACM for making it a great place for graduate study. Special thanks go to Program Secretary Valerie Marino, Technical Coordinator Christina M. Lipsky and IT Support Josko Plazonic and Michael Barone.

At last, I would like to thank my parents. This achievement would not have been possible without their never-ending support from overseas.
Contents

Abstract ................................................................. iii
Acknowledgements .................................................... i
List of Tables .......................................................... vi
List of Figures .......................................................... vii

I Random Kernel Matrices .............................................. 1

1 Introduction ............................................................ 2
  1.1 Overview ........................................................... 2
  1.2 Notation ............................................................ 5

2 Kernel Matrices ......................................................... 7
  2.1 The Zero-diagonal Model ........................................... 7
  2.2 The Kernel Function ............................................... 11
    2.2.1 Orthogonal Polynomials ...................................... 11
    2.2.2 Conditions on $k(x; p)$ ....................................... 15
    2.2.3 $k(x; p) = k(x)$ ............................................... 17
    2.2.4 Examples of $k(x; p)$ .......................................... 21
  2.3 Numerical Experiments ............................................ 24
    2.3.1 Sign kernel ..................................................... 24
    2.3.2 Hard thresholding kernel ..................................... 25
7 Symmetry Detection

7.1 The Symmetry Detection Problem ............................... 87
7.2 The Auto-correlation Function ................................. 88
7.3 Approximating $C_l(r_1, r_2)$ From Projections .................. 90
7.4 Algorithm and Numerical Experiments .......................... 92

8 Structure Reconstruction

8.1 The Reconstruction Problem .................................... 99
8.2 Common-line and Top-view .................................... 100
  8.2.1 Common-line Between Two Projections .................... 100
  8.2.2 Top-view .................................................... 101
  8.2.3 Common-line Between a Projection and the Top-view ...... 103
8.3 The Synchronization Algorithm .................................. 104
  8.3.1 The Algorithm for $C_2$ Symmetry .......................... 104
  8.3.2 Numerical Experiments .................................... 107
8.4 The Top-view-based Algorithm .................................. 110
  8.4.1 The Algorithm for $C_2$ Symmetry .......................... 110
  8.4.2 Numerical Experiments .................................... 114

9 Conclusion ......................................................... 117

9.1 Conclusion .......................................................... 117
9.2 Future Directions .................................................. 117

A Stieltjes Transform ................................................. 119

A.1 Inversion Formula ............................................... 119
A.2 Boundedness by $1/\Im(z)$ ...................................... 120

B Hermite Polynomials .............................................. 121

Bibliography ............................................................ 122
List of Tables

7.1 The rank of $C_l(r_1, r_2)$ (i.e. the dimension of the range of the integral operator with kernel $C_l(r_1, r_2)$) for different groups, according to [21]. The first two members of cyclic groups and dihedral groups are shown. 89

8.1 MSE of the $N = 100$ estimated rotations. The MSE when using a noiseless top-view projection for Step 2-6 is listed as a comparison. 115
List of Figures

2.1 Random kernel matrix with the Sign kernel, and $X_i \sim \mathcal{N}(0, p^{-1}I_p)$. (Left) $p = 4 \times 10^2$, $n = 4 \times 10^3$, $\gamma = p/n = 0.1$. (Right) $p = 8 \times 10^3$, $n = 4 \times 10^3$, $\gamma = p/n = 2$. The blue-boundary bars are the empirical eigenvalue histograms, and the red broken-line curves are the theoretical prediction of the eigenvalue densities by Thm. 3.4. 24

2.2 Random kernel matrix where $k(x) = T_M(x) = x1_{\{|x|>M\}}$, with $M = 1.5$ (left) and $M = 2.5$ (right). $X_i \sim \mathcal{N}(0, n^{-1}I_p)$, and $p = 8 \times 10^3$, $n = 1 \times 10^3$. (p and n switched from the rest of the text.) 25

2.3 (Left) The function $\nu(M)$ as in Eq. (2.18). (Right) The limiting spectral density of random kernel matrices where $k(x) = T_M(x)$ for different values of $M$. When $M = 0$ the density is the M.P. density, which is not shown. 26

2.4 Random kernel matrix where $k(x) = k_o(x) = |x|^{-1/4} - E|\zeta|^{-1/4}$ (left) and $k_o(x) = \text{Sign}(x)|x|^{-1/4}$ (right). $X_i \sim \mathcal{N}(0, p^{-1}I_p)$, and $p = 2 \times 10^3$, $n = 4 \times 10^3$, $\gamma = p/n = 0.5$. 27

3.1 Function $y(u; a, \nu, \gamma)$ as in Eq. (3.7). 34
4.1 Random kernel matrix with $X_i$ uniformly sampled from the $2^p$ vertices of the hypercube $\{-p^{-1/2}, p^{-1/2}\}^p$. (Left) $p = 2 \times 10^2$, $n = 4 \times 10^3$, $k(x) = \text{Sign}(x)$. (Right) $p = 2 \times 10^3$, $n = 4 \times 10^3$, $k(x) = |x| - \sqrt{2/\pi}$.

The blue-boundary bars are the empirical eigenvalue histograms, and the red broken-line curves are the theoretical prediction of the eigenvalue densities by Thm. 3.4.

4.2 Rank-1 spiking model, where $Y_i = \pm u$. $p = 200$, $n = 400$, and $R$ varies. (Left) $k(x) = x$. (Right) $k(x) = \text{Sign}(x)$.

4.3 $R_c$ as a function of $a$ and $\gamma$, $0 < a \leq \nu = 1$.

4.4 Similar plots to Fig. 4.2. $d = 2$, $Y_i \sim \mathcal{U}(S^1)$. Consistency type 1 and 2 are defined in Eq. (4.10, 4.11). (Left) $k(x) = x$. (Right) $k(x) = \text{Sign}(x)$.


7.2 Bar plot of the eigenvalues of $C_l^{(proj)}$ which are computed from $N = 1000$ clean and noisy projections with SNR = 1, 1/4, 1/16 (from upper to lower) respectively, $l = 1, 2, 3$. No symmetry.

7.3 (Upper) Simulated projections SNR = 1/16. (Lower) Bar plot of the eigenvalues of $C_l^{(vol)}$. $C_2$ symmetry.

7.4 Bar plot of the eigenvalues of $C_l^{(proj)}$ which are computed from $N = 1000$ clean and noisy projections with SNR = 1, 1/4, 1/16 (from upper to lower) respectively, $l = 1, 2, 3$. $C_2$ symmetry.

8.1 Four clean (upper) and noisy (lower) projections with SNR = 1/2.

8.2 The largest 20 eigenvalues of the synchronization matrix $\mathcal{R}_{3\text{rd row}}$ (left) and those of the angular synchronization matrix to recover the first two rows of $R_i$’s (right), computed from $N = 200$ clean projections.
8.3 The error in the angles (left) and the comparison of the true and reconstructed volume (right) computed from \( N = 200 \) clean projections. 109

8.4 The same plot as in Fig. 8.2, computed from \( N = 200 \) noisy projections with \( \text{SNR} = 1/2 \). ......................... 109

8.5 The error in the angles (left) and the comparison of the true and reconstructed volume (right) computed from \( N = 200 \) noisy projections with \( \text{SNR} = 1/2 \), where the reconstructed volume is Gaussian-filtered with width 0.8. ......................... 110

8.6 Four clean (upper) and noisy (lower) projections with \( \text{SNR} = 1/32 \). 114

8.7 (Upper) The clean “top-view projection and “top-view” projections at \( \text{SNR} = 1/4, 1/16 \) and \( 1/64 \). (Lower) The estimated “top-view projections at the same noise level as the upper row. ............... 115

8.8 (Left) Histogram of rotation recovery errors in degrees. (Right) Volume reconstructed from estimated orientations and de-noised projections, seen from the side. \( \text{SNR} = 1/32 \). ......................... 116

8.9 The true volume (left) and the reconstructed one (right) seen from the top. \( \text{SNR} = 1/32 \). ......................... 116
Part I

Random Kernel Matrices
Chapter 1

Introduction

1.1 Overview

Given a set of \( n \) data points \( X_1, \ldots, X_n \) in \( \mathbb{R}^p \), a “kernel matrix” \( A \) is an \( n \)-by-\( n \) matrix whose \((i,j)\)-th entry is a function of \( X_i \) and \( X_j \), i.e.

\[
A_{ij} = f(X_i, X_j), \quad 1 \leq i, j \leq n.
\]

Kernel matrices are widely used in machine learning and high-dimensional data analysis, e.g. in kernel-PCA (Principal Component Analysis), kernel-SVM (Support Vector Machine), non-linear dimensionality reduction [24, 5, 9] and so on. The function \( f(\cdot, \cdot) \) is called the “kernel function”. In most applications, \( f(X_i, X_j) \) is a numerical function of the inner-product \( X_i^T X_j \), or of the Euclidean distance \( |X_i - X_j| \). For the former, we call the kernel matrices “inner-product kernel matrices”, and for the latter “Euclidean kernel matrices”. The eigenvalue distribution, namely the “spectrum” of the kernel matrices, is of key interest. In practice, the data vectors \( X_i \)'s usually lie in a high-dimensional space, for example, when \( X_i \)'s are images having \( 10^2 \sim 10^4 \) many pixels. At the same time, the data may contain a significant amount of noise. As a result, the spectrum of kernel matrices is different from when the data vectors are
free from noise. The problem “how high-dimensional noise influences the spectrum of kernel matrices” is an important one, and has not been clearly understood yet.

Towards understanding noise-corrupted kernel matrices, a fundamental mathematical problem is to study the spectrum of kernel matrices built from independent random high-dimensional vectors, which is the topic of this part of the thesis. The case where the data vectors contain both an information part and a noise part is briefly discussed as the “spiking model” in the last chapter.

The study of the spectrum of large random matrices dates back to Wigner’s seminal work on the semi-circle law. Another important result, which influences the application of PCA to high-dimensional vectors, is the Marčenko-Pastur (M.P.) law [23] for the spectrum of Wishart matrices. To be specific, a Wishart matrix has the form of \( S = \frac{1}{n}XX^T \), where \( X \) is a \( p \)-by-\( n \) (complex or real) matrix with i.i.d Gaussian entries. In the “large \( p \), large \( n \)” limit, i.e. \( p, n \to \infty \) and \( p/n \to y \) (\( 0 < y < \infty \)), the spectral density of \( S \) converges to a deterministic limit known as the Marčenko-Pastur distribution. Notice that the Wishart matrix \( S \) shares the non-zero eigenvalues with the Gram matrices \( G = X^TX \) up to a normalization constant, and the Gram matrix \( G \) can be seen as a kernel matrix when the kernel function is a linear function of the inner-product, i.e. \( f(X_i, X_j) = X_i^TX_j \). Thus the M.P. law implies that, when \( p, n \to \infty \) and \( p/n \) converges to a positive constant, the kernel matrix with linear kernel function (under proper normalization) has Marčenko-Pastur density as the limiting spectral density.

The spectrum of large kernel matrices built from real random vectors has been studied in [15], where the kernel function \( f \) does not change as the dimension \( p \) grows, and \( f \) satisfies certain regularity conditions. For this class of kernel matrices, [15] shows the convergence of the spectral density to the M.P. density, which is also the limiting density of the Gram matrix that “approximates” the kernel matrix. To be specific, [15] shows that the kernel matrix with kernel function \( f \) has the same
limiting spectrum as the one whose kernel function is \( f_1 \), where \( f_1 \) is a linear function that approximates \( f \). However, the result in [15] does not apply to the situation when \( f \) depends on the dimension \( p \), nor when \( f \) is non-differentiable. For example, when \( f \) is a step function, i.e. \( f(X_i, X_j) = \text{Sign}(X_i^T X_j) \), it is not even obvious what is the “linear approximation” of this kernel function. Actually, for the \( \text{Sign}(\cdot) \) kernel function, a normalization of \( n^{-1/2} \) of the kernel matrix is needed for the spectral density to converge, and the limiting density is not the M.P. density.

In this part of the thesis we study the case where the kernel function is non-linear and non-smooth, and can depend on the dimension \( p \). We consider the “large \( p \), large \( n \)” limit as in the M.P. law and in [15]. We consider inner-product kernel matrices, addressed as “kernel matrices” in the rest of the text, with a large class of kernel functions. For the model under study, we obtain a new family of limiting spectral densities which includes the Marčenko-Pastur density and semi-circle density as special cases, and the limiting density interestingly involves only a few parameters that are completely determined by the kernel function and the limiting ratio of \( p/n \).

See Thm. 3.4 (for i.i.d. normal Gaussian vectors) and Thm. 4.3 (for i.i.d vectors that can be “compared” with the normal Gaussian vectors) for details.

This part of the thesis mainly addresses the weak convergence of the spectral density, and the study of the law of extreme eigenvalues etc. is not covered. Also, the parallel study of random Euclidean kernel matrices is open.

The rest of this part is outlined as follows:

In Chapter 2, we introduce the mathematical model of kernel matrices, the assumptions on the kernel function, and present the numerical experiments of the eigenvalue distribution of “large \( p \), large \( n \)” kernel matrices.

In Chapter 3, we prove the convergence of the limiting spectrum under the assumption that the data vector \( X_i \)’s are normal Gaussian (Thm. 3.4). As a lemma,
we derive an $O(n^{1/4})$ upper bound for the mean spectral norm based on moment analysis.

In Chapter 4, we extend the study to the cases when $X_i$’s are not Gaussian (Thm. 4.3), while the limiting spectrum in Thm. 3.4 is conjectured to apply to a large class of $X_i$’s beyond what has been proved. We also provide conjectures and heuristics about the “spiking model”.

The content of Chapter 2 and 3 and the first section of Chapter 4 is based on [8].

1.2 Notation

We provide the list of notations throughout the text:

- $I_p$: identity matrix of size $p \times p$.
- $\mathcal{N}(0, \Sigma)$: multi-variate normal distribution with covariance matrix $\Sigma$. In the model of normal Gaussian vectors, $\Sigma = p^{-1}I_p$, which is the covariance matrix of a $p$-dimensional standard normal variable $X$ normalized so that $\mathbb{E}|X|^2 = 1$.
- $\zeta$: standard normal random variable, namely $\zeta \sim \mathcal{N}(0, 1)$.
- $X_i$: i.i.d. data vectors in $\mathbb{R}^p$, $i = 1, \cdots, n$.
- $|X|$: the $L^2$ norm of a vector $X$.
- $\xi_{ij}$: the inner-product of vectors $X_i$ and $X_j$.
- $\xi_p$: the random variable defined as $\xi_p = \sqrt{p}X^TY$, where $X$ and $Y$ are independent copies of $X_i$.
- $q_p(x)$ and $q(x)$: the probability density of $\xi_p$ and $\zeta$ respectively. $q(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. 

5
• \( \mathcal{H}_p \): the \( L^2 \) space of functions on \( \mathbb{R} \) equipped with the probability density of \( \xi_p \),
  i.e. \( \mathcal{H}_p = L^2(\mathbb{R}, q_p(x)dx) \).

• \( P_{l,p}(x), l = 0, 1, \cdots \): orthonormal polynomials in \( \mathcal{H}_p \) when \( X_i \sim \mathcal{N}(0, p^{-1}I_p) \).
  \( P_{l,p}(x) \) is of order \( l \), and \( P_{0,p}(x) = 1 \).

• \( X \equiv Y \): the random variable \( X \) has the same distribution as \( Y \).

• \( A_{:,j} \): the \( j \)-th column of the matrix \( A \).

• \( m(z) \): the Stieltjes transform. \( m_A(z) \) is the Stieltjes transform of the matrix \( A \).

• \( s(M) \): the spectral norm of an \( n \)-by-\( n \) (real-symmetric) matrix \( M \), namely
  \( s(M) = \max_{1 \leq i \leq n} |\lambda_i(M)| \), where \( \{\lambda_i(M), i = 1, \cdots, n\} \) are the eigenvalues of \( M \).

• \( \mathcal{O}(\cdot) \): we write \( x = \mathcal{O}(1)p^\alpha \) if \( |x| \leq C p^\alpha \) for some positive constant \( C \) and large enough \( p \), as \( p \to \infty \). \( \mathcal{O}_a(1) \) means that the constant \( C \) depends on the quantity \( a \), and the latter independent from \( p \).

• \( o(\cdot) \): we write \( x = o(1)p^\alpha \) if \( |x|/p^\alpha \to 0 \) as \( p \to \infty \).
Chapter 2

Kernel Matrices

2.1 The Zero-diagonal Model

An inner-product kernel matrix $A$ is constructed as follows:

Let $X_1, \ldots, X_n$ be $n$ i.i.d normalized Gaussian random vectors in $\mathbb{R}^p$, i.e. $X_i \sim \mathcal{N}(0, p^{-1}I_p)$ and $I_p$ is the $p \times p$ identity matrix. The other models of $X_i$’s are discussed in Chapter 4.

The $n$-by-$n$ matrix $A$ is defined as

$$A_{ij} = \begin{cases} f(X_i^T X_j; p), & i \neq j, \\ 0, & i = j, \end{cases} \quad (2.1)$$

where $f(\xi; p)$ is a real-valued function possibly depending on $p$. $A$ is a real symmetric matrix, and has $n$ (real) eigenvalues $\{\lambda_i(A), i = 1, \cdots, n\}$. The empirical spectral density (ESD) of $A$ is defined as

$$\text{ESD}_A = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(A)}(\lambda), \quad (2.2)$$
which is a random probability measure on \( \mathbb{R} \). We will prove the weak almost sure (a.s.) limit of ESD in the limit of \( p, n \to \infty \) and \( p/n \to \gamma \) \((0 < \gamma < \infty)\).

We study the Stieltjes transform of ESD\(_A\), which is defined as

\[
m_\lambda(z) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i(A) - z} = \frac{1}{n} \text{Tr}(A - zI)^{-1}, \quad \Im(z) > 0.
\] (2.3)

It is a well-known theorem that point-wise convergence of the Stieltjes transform implies weak convergence of the probability density. To be precise, to show the convergence of ESD\(_A\) in a.s. sense to \( \rho_\infty \), it suffices to show that for every fixed \( z \) with \( \Im(z) > 0 \), \( m_\lambda(z) \) converges to the Stieltjes transform of \( \rho_\infty \) in a.s. sense (see e.g. Thm. 2.4.4 in [2]). The properties of Stieltjes transforms which will be used in the text are reviewed in Appendix A.

The following proposition reduces the a.s. convergence of \( m_\lambda(z) \) to the convergence of \( \mathbb{E}m_\lambda(z) \):

**Proposition 2.1.** For the \( n \)-by-\( n \) inner-product kernel matrix \( A \) defined in Eq. (2.1), where \( X_i \)'s are independent random vectors, and a fixed complex number \( z \) with \( \Im(z) > 0 \), we have that as \( n \to \infty \),

\[
m_\lambda(z) - \mathbb{E}m_\lambda(z) \to 0
\]

almost surely, and also

\[
\mathbb{E}|m_\lambda(z) - \mathbb{E}m_\lambda(z)| \leq O(1)n^{-1/2}.
\] (2.4)

The above proposition relies on that \( \Im(z) > 0 \) and that the \( X_i \)'s are independent, while there is no restriction on the specific form of the kernel function, nor on the distribution of \( X_i \). The proof uses Burkholder’s inequality and the interlacing law of
eigenvalues of the minor of a symmetric matrix, combined with the observation that among all the entries of $A$ only the $k$-th column/row depend on $X_k$.

**Proof of Prop. 2.1.** We need the Burkholder’s Inequality (Lemma 2.12. of [3]), which states that for $\{\gamma_k, 1 \leq k \leq n\}$ being a (complex-valued) martingale difference sequence, for $\beta > 1$,

$$
\mathbb{E}\left|\sum_{k=1}^{n} \gamma_k\right|^{\beta} \leq K_\beta \mathbb{E}\left(\sum_{k=1}^{n} |\gamma_k|^2\right)^{\beta/2}, \quad (2.5)
$$

where $K_\beta$ is a positive constant depending on $\beta$.

Using the i.i.d. random vectors $\{X_i, 1 \leq i \leq n\}$, we will define the martingale to be

$$
M_k = \mathbb{E}(\text{Tr}(A - zI)^{-1}|\sigma\{X_{k+1}, \cdots, X_n\}) := \mathbb{E}_k\text{Tr}(A - zI)^{-1}, \quad 0 \leq k \leq n,
$$

where $\sigma\{X_{k+1}, \cdots, X_n\} := \mathcal{F}_{n-k}$ denotes the $\sigma$-algebra generated by $\{X_i, k+1 \leq i \leq n\}$ and $\mathbb{E}(\cdot|\mathcal{G})$ the conditional expectation with respect to the sub-$\sigma$-algebra $\mathcal{G}$. We have $M_n = \mathbb{E}\text{Tr}(A - zI)^{-1}$ and $M_0 = \text{Tr}(A - zI)^{-1}$, and $M_n, \cdots, M_0$ form a martingale with respect to the filtration $\{\mathcal{F}_t, t = 0, \cdots, n\}$. The martingale difference

$$
\gamma_k = M_{k-1} - M_k
\quad = \mathbb{E}_{k-1}\text{Tr}(A - zI)^{-1} - \mathbb{E}_k\text{Tr}(A - zI)^{-1}
\quad = \mathbb{E}_k\left(\text{Tr}(A - zI)^{-1} - \text{Tr}(A^{(k)} - zI)^{-1}\right)
\quad - \mathbb{E}_{k-1}\left(\text{Tr}(A - zI)^{-1} - \text{Tr}(A^{(k)} - zI)^{-1}\right), \quad (2.6)
$$

where $A^{(k)}$ is an $(n-1)$-by-$(n-1)$ matrix that is obtained from the matrix $A$ by eliminating its $k$-th column and $k$-th row. Notice that $A^{(k)}$ is independent of $X_k$, $\mathbb{E}_{k-1}\text{Tr}(A^{(k)} - zI)^{-1} = \mathbb{E}_k\text{Tr}(A^{(k)} - zI)^{-1}$, which verifies the last line of Eq. (2.6).
At the same time, we have

\[ |\text{Tr}(A - zI)^{-1} - \text{Tr}(A^{(k)} - zI)^{-1}| \leq \frac{4}{v}, \tag{2.7} \]

where \( v = \Im(z) > 0 \), using an argument similar to that in Sec. 2.4. of [33] (see Eq. (2.96)). The way to show Eq. (2.7) is by making use of (1) that the ordered \( n - 1 \) eigenvalues of a minor of a symmetric (or Hermitian) matrix \( A \) “interlace” the ordered \( n \) eigenvalues of \( A \), which follows from the Courant-Fischer theorem (see, for example, Exercise 1.3.14 of [33]), and (2) that for fixed \( z \) both real and imaginary parts of \((t - z)^{-1}\) as functions of \( t \) have bounded total variation.

As a result,

\[
|\gamma_k| \leq |\mathbb{E}_k(\text{Tr}(A - zI)^{-1} - \text{Tr}(A^{(k)} - zI)^{-1})| + |\mathbb{E}_{k-1}(\text{Tr}(A - zI)^{-1} - \text{Tr}(A^{(k)} - zI)^{-1})| \\
\leq 2\frac{4}{v} := C,
\]

and then with Eq. (2.5), choosing \( \beta = 4 \),

\[
\mathbb{E}|m_A - \mathbb{E}m_A|^4 = \frac{1}{n^4} \mathbb{E}\left|\sum_{k=1}^{n} \gamma_k\right|^4 \\
\leq \frac{1}{n^4} K_4 \left(\sum_{k=1}^{n} |\gamma_k|^2\right)^2 \\
\leq \frac{1}{n^4} K_4 (nC^2)^2 = \mathcal{O}(1)n^{-2}.
\]

This implies the almost sure convergence of \( m_A - \mathbb{E}m_A \) to 0 by Borel-Cantelli lemma. Also, Eq. (2.4) follows by Jensen’s inequality.

\[
\mathbb{E}|m_A - \mathbb{E}m_A| \leq \left(\mathbb{E}|m_A - \mathbb{E}m_A|^4\right)^{1/4} \leq \mathcal{O}(1)n^{-1/2}. \tag{2.8}
\]
2.2 The Kernel Function

We define
\[ k(x; p) := \sqrt{p} f \left( \frac{x}{\sqrt{p}} ; p \right) \] (2.9)
and the notation \( k(x; p) \) is used throughout the text. We require \( k(x; p) \) to satisfy certain conditions. We take a general formulation and keep the dependency of \( k(x; p) \) on \( p \), while in many cases of interest (see the examples in Sec. 2.2.4), \( k(x; p) = k(x) \) which does not change with \( p \).

The conditions on \( k(x; p) \) will be introduced in Sec. 2.2.2, which involves an orthogonal polynomial expansion of \( k(x; p) \). We firstly introduce the orthogonal polynomials in Sec. 2.2.1. In Sec. 2.2.3 we show that the conditions on \( k(x; p) \) are satisfied by a large class of kernel functions of the form of \( k(x; p) = k(x) \). In Sec. 2.2.4 we provide a list of \( k(x; p) \) which satisfy the conditions, and discuss the motivation to study them.

2.2.1 Orthogonal Polynomials

The renormalized inner-product \( \xi_p \)

Let \( X \) and \( Y \) be two independent random vectors distributed as \( \mathcal{N}(0, p^{-1}I_p) \), and define
\[ \xi_p = \sqrt{p} X^T Y. \] (2.10)

Denote the probability density of \( \xi_p \) by \( q_p(x) \), and the \( L^2 \) spaces \( \mathcal{H}_p = L^2(\mathbb{R}, q_p(x)dx) \). The random variable \( \xi_p \) \((p > 2)\) is sub-exponential, namely
Lemma 2.2. Let $\xi_p$ be as in Eq. (2.10), and equivalently $\xi_p = p^{-1/2} \sum_{i=1}^{p} x_i y_i$ where $x_i$ and $y_i$ i.i.d. $\sim \mathcal{N}(0, 1)$. For $p > 2$,

$$\Pr[|\xi_p| > R] \leq (2e)e^{-R}.$$ 

Proof. Since for $|t| < \sqrt{p}$, $\mathbb{E}e^{t \frac{x_1 y_1}{\sqrt{p}}} = (1 - t^2/p)^{-1/2}$, by choosing $t = 1$ we have

$$\Pr[\xi_p > R] \leq e^{-M(\mathbb{E}e^{x_1 y_1/\sqrt{p}})^p}$$

$$= e^{-M(1 - \frac{1}{p})^{-p/2}}$$

$$\leq e^{-M} e,$$

where the last line is due to that $x = 1/p$ satisfies $\log(1-x)/x > -2$ when $0 < x < 1/2$. The argument for bounding $\Pr[\xi_p < -R]$ is similar.

At the same time, as $p \to \infty$, the random variable $\xi_p$ converges in distribution to $\mathcal{N}(0, 1)$. We have the moment matching lemma:

Lemma 2.3. The moments of $\xi_p$ approximate those of $\mathcal{N}(0, 1)$:

$$\mathbb{E}\xi_p^k = \begin{cases} (k - 1)!! + \mathcal{O}_k(1)p^{-1}, & k \text{ even;} \\ 0, & k \text{ odd.} \end{cases} \quad (2.11)$$

Proof. One way to verify Eq. (2.11) is by directly computing the moments of $\xi_p$ using the definition. Another way is the following:

Let $X_1, X_2 \sim \mathcal{N}(0, p^{-1}I_p)$, and let

$$X_2 = \eta_2 \frac{X_1}{|X_1|} + \tilde{X}_2, \quad \tilde{X}_2^T X_1 = 0,$$

then

$$\xi_p \equiv \sqrt{p}\eta_2 |X_1|,$$
where \( \eta_2 \sim \mathcal{N}(0, p^{-1}) \) and \( \eta_2 \) and \(|X_1|\) are independent. Thus

\[
\mathbb{E} \xi^k_p = \mathbb{E} \xi^k \mathbb{E} |X_1|^k,
\]

where \( \zeta \sim \mathcal{N}(0, 1) \), and \( \mathbb{E} |X_1|^k = 1 + \mathcal{O}_k(1)p^{-1} \) by Lemma 3.1. This verifies the formula in Eq. (2.11) when \( k \) is even. The vanishing of odd moments is by that \( \xi_p \equiv -\xi_p \) since \( X_2 \equiv -X_2 \).

\[\square\]

**The orthogonal polynomials \( P_{l,p}(x) \)**

Let \( \{P_{l,p}(x), l = 0, 1, \cdots\} \) be the sequence of orthonormal polynomials in \( \mathcal{H}_p \), that is

\[
\int_{\mathbb{R}} P_{l_1,p}(x)P_{l_2,p}(x)q_p(x)dx = \delta_{l_1,l_2},
\]

where \( \delta_{l,k} \) equals 1 when \( l = k \) and 0 otherwise. We define \( P_{l,p} \) \((l \geq 0)\) using the Gram-Schmidt procedure on the monomials \( \{1, x, x^2, \ldots\} \), so that \( P_{0,p} = 1, P_{1,p} = x \) (notice that \( \mathbb{E} \xi^2_p = 1 \)), and \( P_{l,p} \) is a polynomial of degree \( l \). We define \( \mathcal{H}_N = L^2(\mathbb{R}, q(x)dx) \) where \( q(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \). The normalized \( l \)-degree Hermite polynomial \( h_l(x), l = 0, 1, \cdots \), form an orthonormal basis of \( \mathcal{H}_N \), see Sec. B for a brief review. Lemma 2.3 implies that the coefficients of the polynomial \( P_{l,p}(x) \), for finite degree \( l \), converge to those of \( h_l(x) \), which is the following lemma:

**Lemma 2.4.** Let \( \{p_{l,p}, l = 0, 1, \cdots\} \) be the orthonormal polynomials of \( L^2(\mathbb{R}, d\mu_p) \), where \( \mu_p \) is a sequence of probability measures. Define \( m_k := \int_{\mathbb{R}} x^k d\mu_p(x) \). Suppose that for fixed positive integer \( l \) and any \( \epsilon > 0 \), as \( p \to \infty \),

\[
|m_k - \mathbb{E} \xi^k| < \epsilon, \quad 1 \leq k \leq l,
\]

then

\[
p_{l,p}(x) = h_l(x) + \sum_{j=0}^{l} (r_{l,p})_j x^j, \quad \max_{0 \leq j \leq l} |(r_{l,p})_j| < C_l \epsilon, \quad (2.12)
\]
where $C_l$ is a positive constant that depends on $l$, and $h_l(x)$ is the normalized Hermite polynomial defined in Eq. (B.1). Particularly, for $d\mu_p$ being the probability measure of $\xi_p$, by Lemma 2.3,

$$P_{l,p}(x) = h_l(x) + \sum_{j=0}^{l} (r_{l,p})_j x^j, \quad \max_{0 \leq j \leq l} |(r_{l,p})_j| < \mathcal{O}_l(1)p^{-1}.$$

**Proof.** Eq. (2.12) is verified by the formula that (see eg. [32])

$$p_{l,p}(x) = c_l \det \begin{pmatrix}
    m_0 & m_1 & \cdots & m_l \\
    \vdots & \vdots & & \vdots \\
    m_{l-1} & m_l & \cdots & m_{2l-1} \\
    1 & x & \cdots & x^l
\end{pmatrix}, \quad c_l^2 = (\det M_{l-1} \det M_l)^{-1},$$

where

$$M_l = \begin{pmatrix}
    m_0 & \cdots & m_l \\
    \vdots & \vdots & \\
    m_l & \cdots & m_{2l}
\end{pmatrix},$$

and the sign of $c_l$ is chosen so that the coefficient of $x^l$ is positive.

As a consequence, we have the following properties of $P_{l,p}(x)$ which will be used in proving the main theorem:

**Proposition 2.5.** For finite $l$ and $p \to \infty$, if $|x| \leq M$, then

$$|P_{l,p}(x)| \leq \mathcal{O}_l(1)M^l. \quad (2.13)$$

**Proof.** By Lemma 2.4, the coefficients of $P_{l,p}(x)$ for each $l$ converge to those of $h_l(x)$. Combining with the expression of $h_l(x)$ as in Eq. (B.1, B.2) the bound follows.
Proposition 2.6. For \( l \geq 2 \) and \( |x| \leq M \),

\[
P_l'(x) = \sqrt{l}P_{l-1,p}(x) + \mathcal{O}_l(1)M^{l-1}p^{-1},
\]

\[
P_l''(x) = \sqrt{l(l-1)}P_{l-2,p}(x) + \mathcal{O}_l(1)M^{l-2}p^{-1}.
\]

(2.14)

Proof. By Lemma 2.4 and Eq. (B.3).

\[\square\]

### 2.2.2 Conditions on \( k(x; p) \)

We formally expand \( k(x; p) \) as

\[
k(x; p) = \sum_{l=0}^{\infty} a_{l,p}P_{l,p}(x),
\]

(2.15)

\[
a_{l,p} = \int_{\mathbb{R}} k(x;p)P_{l,p}(x)q_p(x)dx,
\]

and require the following conditions on \( k(x; p) \):

1. **(C.Variance)** For all \( p, k(x; p) \in \mathcal{H}_p \), and as \( p \to \infty \), \( \text{Var}(k(\xi_p; p)) = \nu_p \to \nu \) which is a finite non-negative number.

2. **(C.p-Uniform)** The expansion in Eq. (2.15) converges in \( \mathcal{H}_p \) uniformly in \( p \).

   Equivalently, let

   \[
k_L(x; p) = \sum_{l=0}^{L} a_{l,p}P_{l,p}(x),
\]

   then for any \( \epsilon > 0 \), there exist \( L \) and \( p_0 \) such that \( \sum_{l=L+1}^{\infty} a_{l,p}^2 < \epsilon \) for \( p > p_0 \).

3. **(C.a1)** As \( p \to \infty \), \( a_{1,p} \to a \) which is a constant.

By condition **(C.Variance)**, the integrals in Eq. (2.15) are well-defined and the series in the expansion converges. The requirement \( \nu_p \to \nu \) can be fulfilled as long as \( k(x; p) \in \mathcal{H}_p \) and is properly scaled.

Notice that

\[
\nu_p = \text{Var}(k(\xi_p; p)) = \sum_{l=1}^{\infty} a_{l,p}^2.
\]
which implies that

**Lemma 2.7.**

\[ \sum_{l=1}^{\infty} a_{l,p}^2 = O(1), \quad p \to \infty. \] (2.16)

Particularly, in condition \((C.a_1)\), \(a^2 \leq \nu\).

In the rest of this section we will discuss the class of \(k(x;p)\) which satisfies the above conditions. The case where \(k(x;p) = k(x)\) is addressed in the next subsection. The inner-product kernel function studied in [15], namely \(f(\xi;p) = f(\xi)\) and is differentiable at \(\xi = 0\), is included after a truncation. Specifically, we have

**Lemma 2.8.** If \(f(\xi;p) = f(\xi)\) is \(C^1\) at \(\xi = 0\), let \(\hat{f}(\xi;p) = f(\xi)1_{|\xi|\leq \delta}\), where \(\delta = \delta(p) = \frac{M}{\sqrt{p}}, M = \sqrt{20 \ln p}\), then \(k(x;p)\) associated with \(\hat{f}\) satisfies the conditions \((C.\text{Variance})\), \((C.p-\text{Uniform})\) and \((C.a_1)\), and \(a = f'(0)\) and \(\nu = a^2\).

**Remark 2.1.** If we denote by \(\hat{A}\) the kernel matrix with kernel function \(\hat{f}\), then \(\hat{A}\) and \(A\) have the same limiting spectral density: using a similar argument as in Lemma 3.7, we have

\[ \Pr[\exists i \neq j, |X_i^T X_j| > \delta] \leq O(1)p^{-7}. \]

Thus, for fixed \(z = u + iv\)

\[ \mathbb{E}|m_A(z) - m_{\hat{A}}(z)| \leq \frac{2}{v} \Pr[\exists i \neq j, |X_i^T X_j| > \delta] \to 0, \]

where \(m_A(z)\) and \(m_{\hat{A}}(z)\) are the Stieltjes transforms of \(A\) and \(\hat{A}\) respectively.

**Proof of Lemma 2.8.** Since \(f(\xi)\) is \(C^1\) at \(\xi = 0\), for any \(\epsilon > 0\) there exists a neighborhood \([-R, R]\) on which

\[ f(\xi) = f(0) + f'(0)\xi + r(\xi), \quad |r(\xi)| \leq \epsilon|\xi|. \]
Since $\delta \to 0$ when $p \to \infty$, we assume that $p$ is large enough so that $\delta < R$.

Let $k(x; p) = \sqrt{p} \hat{f}(x/\sqrt{p})$, and assume that $f(0) = 0$ since it only contributes to $E_k(\zeta_p, p) = a_{0,p}$, we have

$$k(x; p) = \left( f'(0)x + \sqrt{p}r\left(\frac{x}{\sqrt{p}}\right) \right) 1_{\{|x| \leq M\}} : = k_1 + k_2,$$

where $|k_2(x; p)| \leq \epsilon|x|$, so $E\xi_2(k_p; p)^2 \leq \epsilon^2$. Thus, the $L^2$ norm of $k_2$ is arbitrarily small in $\mathcal{H}_p$, and $\nu_p = Var(k_\zeta(p))$ and $a_{1,p} = \mathbb{E}\xi_p(k_\zeta(p) - \mathbb{E}k_\zeta(p))$ are decided by $k_1$. For $k_1(x; p) = f'(0)x 1_{\{|x| \leq M\}}$, $E_1(k_\zeta(p)) = 0$, and since $M \to \infty$ as $p \to \infty$, $E_1(k_\zeta(p))^2 \to (f'(0))^2$ and $\mathbb{E}_p k_1(\zeta_p; p) \to f'(0)$. Thus $\nu_p \to (f'(0))^2 = \nu$, and $a_{1,p} \to f'(0) = a$.

### 2.2.3 $k(x; p) = k(x)$

When $k(x; p) = k(x)$ does not change with $p$, and $k(x) \in \mathcal{H}_N$, under certain non-restrictive assumptions on $k(x)$, mainly Eq. (2.17), the conditions (C.Variance), (C.p-Uniform) and (C.a1) are satisfied. Furthermore, the $P_{i,p}$-expansion coefficients $a_{i,p}$ converges to the Hermite-expansion coefficients of $k(x)$. This is the content of the following proposition:

**Proposition 2.9.** Suppose that

1. $k(x) \in \mathcal{H}_N$. Let $\zeta \sim \mathcal{N}(0,1)$, define

$$\nu_N := \mathbb{E}k(\zeta)^2 < \infty,$$

Also, $\mathbb{E}k(\zeta) = 0$. 

17
2. \(k(x)\) satisfies that

\[
\int_{\mathbb{R}} k(x)^2 |q_p(x) - q(x)| dx \to 0, \quad p \to \infty. \tag{2.17}
\]

Eq. (2.17) implies that \(\mathbb{E} k(\xi_p)^2 \to \nu_N\). Without loss of generality, \(k(x)\) is in \(\mathcal{H}_p\) for all \(p\). Define \(b_{l,p}\) and \(a_{l,p}\) as in Lemma 2.10, and notice that since \(k(x)\) does not depend on \(p\), \(b_{l,p} = b_l\) independent of \(p\). Then conditions (C.Variance), (C.\(p\)-Uniform) and (C.\(\alpha_1\)) are satisfied by \(k(x; p) = k(x) - a_{0,p}\). Also,

\[
\nu_p \to \nu_N, \quad a_{1,p} \to a_N = b_1.
\]

**Proof.** By definition \(\mathbb{E} k(\xi_p; p) = 0\). In this case,

\[
\nu_p = \mathbb{E} k(\xi_p; p)^2 = \mathbb{E} k(\xi_p)^2 - a_{0,p}^2.
\]

Since Lemma 2.10 applies to \(k(x)\), we know that

\[
a_{0,p} \to b_0 = \mathbb{E} k(\zeta) = 0.
\]

Together with the fact that \(\mathbb{E} k(\xi_p)^2 \to \mathbb{E} k(\zeta)^2 = \nu_N\), we know that \(\nu_p \to \nu_N\) as \(p \to \infty\). Thus (C.Variance) is satisfied.

Also \(a_{1,p} \to b_1\) which is a constant, thus (C.\(\alpha_1\)) holds.

For (C.\(p\)-Uniform) to be satisfied, it suffices to show that \(\sum_{l=L+1}^{\infty} a_{l,p}^2\) can be made \(p\)-uniformly small. Notice that

\[
\sum_{l=0}^{\infty} a_{l,p}^2 = \mathbb{E} k(\xi_p)^2 \to \nu_N,
\]

\[
\sum_{l=0}^{\infty} b_l^2 = \nu_N,
\]
and meanwhile for each $l$, $a_{l,p} \to b_l$ by Lemma 2.10, thus for any finite $L$

$$
\sum_{l=L+1}^{\infty} a_{l,p}^2 = \mathbb{E} h(\xi_p)^2 - \sum_{l=0}^{L} a_{l,p}^2 \\
\to \nu_N - \sum_{l=0}^{L} b_l^2 = \sum_{l=L+1}^{\infty} b_l^2,
$$

which can be made small by choosing $L$ large independently of $p$. \hfill \Box

**Lemma 2.10.** Suppose that $k(x; p)$ is in $H_N$ and $H_p$ for all $p$, and satisfies Eq. (2.17). Let

$$
b_{l,p} = \int_R k(x; p) h_l(x) q(x) \, dx,
$$

$$
a_{l,p} = \int_R k(x; p) P_{l,p}(x) q_p(x) \, dx,
$$

for $l = 0, 1, \ldots$. Then for each $l$, $|b_{l,p} - a_{l,p}| \to 0$ as $p \to \infty$.

**Proof.**

$$
|b_{l,p} - a_{l,p}|
= |\int_R k h_l(q - q_p) \, dx + \int_R k (h_l - P_{l,p}) q_p \, dx|
\leq \int_R |k h_l||q - q_p| \, dx + \int_R |k||h_l - P_{l,p}| q_p \, dx
\leq (1) + (2).
$$

For (1), by Cauchy-Schwarz

$$
(1)^2 \leq \left( \int k^2 |q - q_p| \, dx \right) \left( \int h_l^2 |q - q_p| \, dx \right),
$$

where

$$
\int h_l^2 |q - q_p| \, dx \leq \int h_l^2 q \, dx + \int h_l^2 q_p \, dx = 1 + (1 + O_l(1)p^{-1}),
$$
which is bounded as \( p \to \infty \), and \( \int k^2 |q - q_p|dx \to 0 \), thus (1) \( \to 0 \). For (2),

\[
(2)^2 \leq (\int k^2 q_p dx)(\int (h_l - P_{l,p})^2 q_p dx),
\]

where \( \int k^2 q_p dx \to \int k^2 q dx \) which is bounded, and by Lemma 2.4

\[
\left( \int (h_l - P_{l,p})^2 q_p dx \right)^{1/2} \leq O(1)p^{-1},
\]

so (2) \( \to 0 \).

Eq. (2.17) holds as long as the singularity in the integral, say at \( x = \infty \) or \( k(x) = \infty \), can be controlled \( p \)-uniformly. This is the case, for example, when \( k(x) \) is bounded, or when \( k(x) \) is bounded on \( |x| \leq R \) for any \( R > 0 \) and \( k(x)^2 \) is \( p \)-uniformly integrable at \( x \to \infty \). We have the following lemma:

**Lemma 2.11.** Suppose \( k(x) \) is (Case 1) bounded, or (Case 2) in \( \mathcal{H}_N \) and \( \mathcal{H}_p \) for all \( p \), is bounded on \( |x| \leq R \) for any \( R > 0 \), and satisfies

\[
\int_{|x| > R} k(x)^2 q_p(x) dx \to 0, \quad R \to \infty
\]

uniformly in \( p \), then Eq. (2.17) holds. When \( k(x) \) is any polynomial, it belongs to (Case 2).

**Remark 2.2.** It is also possible for \( k(x) \) to be singular at \( x = 0 \). See the examples in Sec. 2.2.4.

**Proof of Lemma 2.11.** First, we reduce (Case 2) to (Case 1). Notice that

\[
\int_R k(x)^2 |q_p(x) - q(x)| dx \\
\leq \int_{|x| \leq R} k(x)^2 |q_p(x) - q(x)| dx + \int_{|x| > R} k(x)^2 q_p(x) dx \\
+ \int_{|x| > R} k(x)^2 q(x) dx.
\]
The last two terms can be made arbitrarily small independently of $p$ by choosing $R$ large, and for fixed $R$, the first term goes to 0 given that (Case 1) is proved.

To show the claim for (Case 1), it suffices to show that $\int |q_p - q| dx \to 0$. Since $\xi_p$ converge in distribution to $\mathcal{N}(0,1)$, we know that for any finite $R$, $\int_{|x|<R} |q_p(x) - q(x)| dx \to 0$. Thus, it suffices to show that

$$\int_{|x|>R} q_p(x) dx \to 0, \quad R \to \infty$$

uniformly in $p$. This follows from the large deviation bound given in Lemma 2.2.

(Case 2) includes all the polynomials. The $p$-uniform integrability is verified by a Cauchy-Schwarz inequality, combined with 1) the fact that all (even) moments of $\xi_p$ are finite and converge to those of the standard Gaussian (Eq. (2.11)), thus are $p$-uniformly bounded, and 2) the bound given in Lemma 2.2.

2.2.4 Examples of $k(x; p)$

To summarize, we provide a list of examples of $k(x; p)$ that fulfill the conditions (C.Variance), (C.$p$-Uniform) and (C.$a_1$) and the constants $a$ and $\nu$ associated:

1. (Differentiable $f$ independent of $p$) $f(\xi; p) = f(\xi)$ is $C^1$ at $\xi = 0$, and $k(x; p) = \sqrt{p} f(p^{-1/2}x)$. $a = f'(0)$ and $\nu = a^2$. (Lemma 2.8).

2. (Sign kernel) $k(x; p) = \text{Sign}(x)$. $\text{Sign}(x) = 1$ for $x > 0$, $\text{Sign}(x) = -1$ for $x < 0$ and $\text{Sign}(x) = 0$ for $x = 0$. $k(x)$ is bounded, and Prop. 2.9 and Lemma 2.11 apply. $a = \mathbb{E}|\zeta| = \sqrt{2/\pi}$, and $\nu = 1$.

The following serves as a motivation for the sign kernel matrix. Consider a network of $n$ “subjects” represented by $X_1, \ldots, X_n$ lying in $\mathbb{R}^p$. Subjects $i$ and $j$ have a friendship relationship if they are positively correlated, i.e., if $X_i^T X_j > 0$, and a non-friendship relationship if $X_i^T X_j < 0$. The off-diagonal entries of the
n-by-n kernel matrix $A$ are all $\pm 1$ representing the friendship/non-friendship relationships. This model has the merit that if $i$ and $j$ are friends, and $j$ and $k$ are also friends, then chances are greater that $i$ and $k$ are also friends. When the $X_i$’s are i.i.d uniformly distributed on the unit sphere in $\mathbb{R}^p$ and $p$ is fixed, according to [22], as $n$ grows to infinity the top $p$ eigenvectors of the kernel matrix $A$ converge, up to a multiplying constant and a global rotation, to the coordinates of the $n$ data points. In this case, the eigenvalues of the sign kernel matrix converges to those of the integral operator on the manifold, and the eigenvectors recover the positioning of the subjects in the whole community from their pairwise relationships.

On the other hand, Thm. 3.4 deals with the case where the data are distributed as $p$-dimensional standard Gaussian, and the dimension $p$ increases with $n$ where $p/n \to \gamma$. For this case, the spectrum of the random kernel matrices observes a limiting law, where the dependence of the kernel function involves only the parameters $\nu$ and $a$.

3. (Hard thresholding kernel) $k(x) = T_M(x) = x 1_{|x| > M}$, where $M$ is a positive constant. Prop. 2.9 and Lemma 2.11 apply. Meanwhile,

$$\nu = \nu(M) = \mathbb{E}T_M(\zeta)^2$$

$$= 2 \int_{M}^{\infty} x^2 e^{-x^2/2} \frac{1}{\sqrt{2\pi}} dx$$

$$= \sqrt{\frac{2}{\pi}} M e^{-M^2/2} + 2(1 - \Phi(M)),$$

(2.18)

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$. For this case $a = \nu$.

The study of the hard thresholding kernel is motivated by the application of thresholding in estimating sample covariance matrices [7, 13]. Let $X_1, \cdots, X_p$ be i.i.d. distributed as $\mathcal{N}(0, n^{-1}I_n)$, the off-diagonal entries of the random kernel
matrix $A$ equal $^1$

$$A_{ij} = T_{\frac{M}{\sqrt{n}}} (X_i^T X_j), \quad i \neq j, 1 \leq i, j \leq p,$$

and $A$ equals a sample covariance matrix of $n$ standard Gaussian vectors hard-thresholded at $t = \frac{M}{\sqrt{n}}$, except for the diagonal entries $^2$.

Eq. (2.18) implies that $\nu(M) < \sqrt{(M + 1)e^{-M^2/2}}$, which means that as $M$ grows, typically beyond 2, $a = \nu \rightarrow 0$ exponentially fast. A plot of $\nu(M)$ is shown on the left of Fig. 2.3.

4. (Singular kernel) $k(x) = |x|^{-r}$ ($r < 1/2$). For example, the even function

$$k_e(x) = |x|^{-r} - \mathbb{E}|\zeta|^{-r}$$

and the odd function

$$k_o(x) = \text{Sign}(x)|x|^{-r},$$

where $r < 1/2$ so as to guarantee the integrability of $k(x)^2$ at $x = 0$. For both cases, $|k(x)|$ is bounded on $\{|x| > R\}$ for any $R > 0$, and diverge at $x = 0$. Meanwhile, $k(x)^2 = |x|^{-2r}$ is integrable at $x = 0$, and with the fact that $q_p(x) \leq q_p(0) \rightarrow q(0) = 1/\sqrt{2\pi}$, Eq. (2.17) still holds, thus Prop. 2.9 applies.

The constants $\nu$ and $a$ for both $k_e$ and $k_o$ have explicit expression involving the Gamma function.

$^1$Due to the convention that the sample covariance matrices is $p$-by-$p$, we switch the notion of $n$ and $p$ in this example. We consider standard normal $X_i$’s, so the distribution of the vectors is again normal after switching the rows with the columns.

$^2$For this model, the analysis of the diagonal entries can be separated from that of the off-diagonal ones, by the fact that the diagonal entries of the sample covariance matrix asymptotically concentrate at 1. The details of extending the analysis to including the diagonal entries are omitted here.
Figure 2.1: Random kernel matrix with the Sign kernel, and $X_i \sim \mathcal{N}(0, p^{-1}I_p)$. (Left) $p = 4 \times 10^2$, $n = 4 \times 10^3$, $\gamma = p/n = 0.1$. (Right) $p = 8 \times 10^3$, $n = 4 \times 10^3$, $\gamma = p/n = 2$. The blue-boundary bars are the empirical eigenvalue histograms, and the red broken-line curves are the theoretical prediction of the eigenvalue densities by Thm. 3.4.

2.3 Numerical Experiments

In the subsequent plots, we numerically compare the eigenvalue histogram and the theoretical limiting spectral density. The “theoretical curve” is calculated using the expression of the limiting density defined in Eq. (3.4) based on Thm. 3.4. The eigenvalues that produce the empirical histogram are computed by MATLAB’s eig function and correspond to a single realization of the random kernel matrix.

2.3.1 Sign kernel

Let

$$k(x) = \text{Sign}(x).$$

Recall that $a = \sqrt{2/\pi}$ and $\nu = 1$. Fig. 2.1 is for $X_i \sim \mathcal{N}(0, p^{-1}I_p)$. Notice that for the sign kernel, the two models $X_i \sim \mathcal{N}(0, p^{-1}I_p)$ and $X_i \sim \mathcal{U}(S^{p-1})$ result in the same probability law of the random kernel matrix. This is due to the fact that $\text{Sign}(X_i^T X_j) = \text{Sign}((X_i/|X_i|)^T(X_j/|X_j|))$ and that if $X_i \sim \mathcal{N}(0, p^{-1}I_p)$ then $X_i/|X_i| \sim \mathcal{U}(S^{p-1})$. As such, the results for $X_i \sim \mathcal{U}(S^{p-1})$ are omitted.
2.3.2 Hard thresholding kernel

Let $k(x)$ be the hard thresholding function

$$k(x) = T_M(x) = x 1_{|x| > M}, \quad M > 0$$

Recall that $a = \nu$ which is a function of $M$, and the value is computed in Eq. (2.18).

The comparison of the empirical eigenvalues and the limiting density are shown in Fig. 2.2, where $n = 1 \times 10^3$, $p = 8 \times 10^3$, and $M = 1.5$ and 2.5 respectively. The limiting densities for different values of $M$ are shown on the right of Fig. 2.3.

According to M.P. Law, without thresholding, the right edge of the limiting density is at $(1 + \sqrt{p/n})^2 - 1 = 13.6569$. The right edge decreases to about 1 when thresholded at $M = 3$.

2.3.3 Singular kernel

For the even function $k_e(x) = |x|^{-r} - \mathbb{E}|\zeta|^{-r}$,

$$\nu = \text{Var}(|\zeta|^{-r}), \quad a = 0,$$
and $\nu$ has the expression as

$$E|\zeta|^{-r} = \sqrt{\frac{2}{\pi}}2^{-(r+1)/2}\Gamma\left(\frac{1-r}{2}\right), \quad \zeta \sim \mathcal{N}(0, 1),$$  \hspace{1cm} (2.19)$$

where $\Gamma(\cdot)$ is the Gamma function.

For the odd function $k_o(x) = \text{Sign}(x)|x|^{-r}$, $\nu = |\zeta|^{-2r}$

and the expression is known by Eq. (2.19). Meanwhile,

$$a = E|\zeta|^{1-r} = \sqrt{\frac{2}{\pi}}2^{-r/2}\Gamma(1 - \frac{r}{2}).$$

The numerical results for $r = 1/4$ with $X_i \sim \mathcal{N}(0, p^{-1}I_p)$ are shown in Fig. 2.4. The empirical histograms for $X_i \sim \mathcal{U}(S^{p-1})$ look almost identical and are therefore omitted. In the left panel of Fig. 2.4, the empirical spectral density is close to a semi-circle, as our theory predicts.
Figure 2.4: Random kernel matrix where \( k(x) = k_{e}(x) = |x|^{-1/4} - \mathbb{E}|\zeta|^{-1/4} \) (left) and \( k_o(x) = \text{Sign}(x)|x|^{-1/4} \) (right). \( X_i \sim \mathcal{N}(0, p^{-1}I_p) \), and \( p = 2 \times 10^3, n = 4 \times 10^3, \gamma = p/n = 0.5. \)
Chapter 3

Limiting Spectrum

3.1 Analysis of the Fourth Moment

In this section we bound the mean spectral norm of the kernel matrix to be of order \( n^{1/4} \), see Prop. 3.3. The method to prove Prop. 3.3 is by analyzing the fourth moment of the random matrix. We firstly introduce a useful lemma concerning the asymptotic concentration of the length of the Gaussian vector \( X \sim \mathcal{N}(0, p^{-1}I_p) \) at 1 as \( p \to \infty \):

**Lemma 3.1.** Let \( X \sim \mathcal{N}(0, p^{-1}I_p) \), then for any positive integer \( k \),

\[
\mathbb{E}|X|^k = 1 + O_k(1) p^{-1}, \quad p \to \infty. \tag{3.1}
\]

The following estimate is important in analyzing the 4th moment:

**Lemma 3.2.** Let \( X_1, X_2, X_3, X_4 \) be i.i.d distributed as \( \mathcal{N}(0, p^{-1}I_p) \), and \( P_{l,p}(x) \) is as defined in Sec. 2.2.1, \( l \geq 2 \). Then

\[
\mathbb{E} P_{l,p}(\sqrt{p} \xi_{12}) P_{l,p}(\sqrt{p} \xi_{23}) P_{l,p}(\sqrt{p} \xi_{34}) P_{l,p}(\sqrt{p} \xi_{41}) = O_l(1) p^{-2},
\]

where \( \xi_{ij} = X_i^T X_j \).

The proofs of Lemma 3.1 and Lemma 3.2 are postponed until Sec. 3.4.
Proposition 3.3. Let $A$ be the random kernel matrix defined in Eq. (2.1) with the kernel function $f(\xi; p) = p^{-1/2}P_{l,p}(\sqrt{p} \xi)$, $l \geq 1$, where $P_{l,p}$ is defined as in Sec. 2.2.1. $X_i$’s are i.i.d. distributed as $\mathcal{N}(0, p^{-1}I_p)$. Then, as $p, n \to \infty$, $p/n \to \gamma$,

$$
\mathbb{E}s(A) \leq O_{l,\gamma}(1)n^{1/4}.
$$

The spectral norm of random matrices is an important topic in random matrix theory. The asymptotic concentration of the largest eigenvalue at its mean value is quantified by the Tracy-Widom Law for Gaussian ensembles [35] and a large class of Wigner-type matrices (see e.g. [30], [34] and references therein). The Tracy-Widom Law was also established for Wishart matrices [18], and was shown to be universal for sample covariance matrices with non-Gaussian entries, see e.g. [31, 12].

For kernel matrices studied in Prop. 3.3, the spectral norm is conjectured to be $O(1)$. The bound in Prop. 3.3, though not tight, is sufficient for the proof of the main theorem of this chapter.

Proof of Prop. 3.3. Let $\{\lambda_i, 1 \leq i \leq n\}$ be the eigenvalues of $A$. Since

$$
s(A)^4 \leq \sum_{i=1}^{n} \lambda_i^4 = \text{Tr}(A^4) = \sum_{i,j,k,l} A_{ij}A_{jk}A_{kl}A_{li},
$$

we have

$$
\mathbb{E}s(A) \leq (\mathbb{E}s(A)^4)^{1/4} \leq (\sum_{i,j,k,l} \mathbb{E}A_{ij}A_{jk}A_{kl}A_{li})^{1/4}.
$$

In Eq. (3.2) for $\mathbb{E}A_{ij}A_{jk}A_{kl}A_{li}$ to be non-zero, in $\{i, j, k, l\}$ the neighboring indices must differ since $A_{ii} = 0$. Meanwhile, by Lemma 2.3 and Lemma 2.4, for any fixed $l$,

$$
\mathbb{E}P_{l,p}(\sqrt{p}\xi_{12})^4 = O_l(1).
$$

We have the following cases:
1. $i = k, j = l$:

$$E A_{ij} A_{jk} A_{kl} A_{li} = E A_{12}^4$$

$$= p^{-2} E P_{l,p} (\sqrt{p} \xi_{12})^4$$

$$= O_l(1) p^{-2},$$

2. $i = k, j \neq l$ or $i \neq k, j = l$:

$$E A_{ij} A_{jk} A_{kl} A_{li} = E A_{12}^2 A_{13}^2$$

$$= p^{-2} E P_{l,p} (\sqrt{p} \xi_{12})^2 P_{l,p} (\sqrt{p} \xi_{13})^2$$

$$\leq p^{-2} E P_{l,p} (\sqrt{p} \xi_{12})^4 \quad \text{(by Cauchy-Schwarz)}$$

$$= O_l(1) p^{-2}.$$

3. $i \neq k, l \neq j$: when $l = 1$, $E A_{ij} A_{jk} A_{kl} A_{li} = p^{-3}$. When $l \geq 2$, we have the following estimate (Lemma 3.2)

$$E A_{ij} A_{jk} A_{kl} A_{li} = O_l(1) p^{-4}.$$

As a result, when $l = 1$,

$$\sum_{i,j,k,l} E A_{ij} A_{jk} A_{kl} A_{li}$$

$$\leq n^2 O(1) p^{-2} + 2n^3 O(1) p^{-2} + n^4 p^{-3}$$

$$= O_\gamma(1)n + O_\gamma(1),$$
and when \( l \geq 2 \),

\[
\sum_{i,j,k,l} E A_{ij} A_{jk} A_{kl} A_{li} \leq n^2 \mathcal{O}(1)p^{-2} + 2n^3 \mathcal{O}(1)p^{-2} + n^4 \mathcal{O}(1)p^{-4} = \mathcal{O}_{l,\gamma}(1)n + \mathcal{O}_{l,\gamma}(1).
\]

Combining the above estimates with Eq. (3.2) leads to the bound wanted. \( \square \)

### 3.2 Convergence of ESD

#### 3.2.1 Statement of the main theorem

**Theorem 3.4.** Suppose that \( X_1, \cdots, X_n \) are i.i.d Gaussian vectors distributed as \( \mathcal{N}(0, p^{-1}I_p) \), \( k(x; p) \) satisfies conditions (C.Variance), (C.p-Uniform) and (C.a1) defined in Sec. 2.2.2, and \( A \) is the kernel matrix defined as in Eq. (2.1). As \( p, n \to \infty \) with \( p/n \to \gamma \), \( ESD_A \) (defined in Eq. (2.2)) converges weakly to a continuous probability measure on \( \mathbb{R} \) in the almost sure sense. The Stieltjes transform of the limiting spectral density is the solution of the following algebraic equation

\[
-\frac{1}{m(z)} = z + a \left( 1 - \frac{1}{1 + \frac{a}{\gamma}m(z)} \right) + \frac{\nu - a^2}{\gamma} m(z). \tag{3.3}
\]

Eq. (3.3) is at most cubic, and involves three parameters: \( \nu \) (defined in (C.Variance)), \( a \) (defined in (C.a1)) and \( \gamma \). Eq. (3.3) has a unique solution \( m(z) \) with positive imaginary part (Lemma 3.5), and the explicit formula of

\[
y(u) := \lim_{v \to 0^+} \Im(m(u + iv)) \tag{3.4}
\]

is given in Sec. 3.2.2.
Remark 3.1. Without loss of generality we assume that $a_{0,p} = 0$. Otherwise it results in adding to the kernel matrix a perturbation $\frac{1}{\sqrt{p}}a_{0,p}(1_n1_n^T - I_n)$, where $1_n$ is the all-ones vector of length $n$ and $I_n$ is the identity matrix. The limiting spectral density of a sequence of Hermitian matrices with growing size ($n \to \infty$) is stable under a finite-rank perturbation (with rank that does not depend on $n$), see Thm. A.43 in [3]. When $a_{0,p}$ is not zero, it results in the shift of the spectrum by $a_{0,p}/\sqrt{p}$ and the largest (smallest) eigenvalue of $A$ being of the magnitude of $\sqrt{p}|a_{0,p}|$ if $a_{0,p} > 0$ ($a_{0,p} < 0$).

Remark 3.2. Corresponding to the $l$-th term in Eq. (2.15), we define the random kernel matrix $A_l$ to be

$$
(A_l)_{ij} = \begin{cases} 
    f_l(X_i^TX_j;p), & i \neq j, \\
    0, & i = j,
\end{cases}
$$

(3.5)

where $f_l(\xi;p) = \frac{a_{1,p}}{\sqrt{p}}P_l(\sqrt{p}\xi)$.

The limiting spectral density of $A_1$ is the M.P. distribution. For this case, $f(\xi;p) = a\xi$, or equivalently $k(x;p) = ax$, for some constant $a$. Then, the expansion in Eq. (2.15) has one term, $a_{1,p} = a$, $\nu_p = a^2$, and Eq. (3.3) is reduced to the equation of the Stieltjes transform of the M.P. density$^1$.

The limiting spectral density of $A_l$ ($l \geq 2$) is a semi-circle. Moreover, the limiting density of any partial sum (finite or infinite) of $A_2, A_3, \cdots$ is a semi-circle, whose squared radius equals the sum of the squared radii of the semi-circle of each $A_l, l \geq 2$.

Remark 3.3. For random kernel matrices where $f(\xi;p) = f(\xi)$ and is differentiable at $\xi = 0$, the limiting spectral density is the M.P. distribution. Specifically, by Lemma 2.8 the result in the theorem holds for the kernel matrix associated with the truncated kernel function, and then extends to the kernel matrix associated with $f(\xi)$ by Remark

$^1$To be more precise it is the shifted and re-scaled M.P. density, see Eq. (2.5, 2.6) of [8]
2.1, where \( a^2 = \nu = (f'(0))^2 \) in Eq. (3.3). In other words, the linear term in Eq. (2.15) determines the limiting spectral density, in consistence with the result in [15].

### 3.2.2 Solution of Eq. (3.3)

We rewrite Eq. (3.3) as

\[
a\frac{(\nu - a^2)}{\gamma} m^3 + (\nu + az)m^2 + (a + \gamma z)m + \gamma = 0, \quad \Im(z) > 0, \Im(m) > 0,
\]

where \( a^2 \leq \nu \). When \( a = 0 \) \((a^2 = \nu)\) the equation corresponds to the semi-circle distribution (M.P. distribution), and the existence and uniqueness of the solution with positive imaginary part are known. We consider the case where \( 0 < a^2 < \nu \), thus the cubic term in Eq. (3.6) does not vanish.

**Lemma 3.5.** For every \( z \) with \( \Im(z) > 0 \), there exists a unique \( m \) with \( \Im(m) > 0 \) for which Eq. (3.6) holds.

**Proof.** It can be verified that whenever \( a, \nu, \gamma \) are real and \( \Im(z) > 0 \), the solution \( m \) must not be real. Define the domain \( \mathcal{D} := \{(a, \nu, \gamma, z), \gamma > 0, 0 < a^2 < \nu, \Im(z) > 0\} \) which has two connected components \( \mathcal{D}_+ = \mathcal{D} \cap \{a > 0\} \) and \( \mathcal{D}_- = \mathcal{D} \cap \{a < 0\} \).

The three solutions of the cubic equation depend continuously on the coefficients, thus if we let \((a, \nu, \gamma, z)\) vary continuously in \( \mathcal{D}_+ \), the imaginary parts of the three solutions never change sign, and similarly for \( \mathcal{D}_- \). As a result, it suffices to show that for one choice of \((a, \nu, \gamma, z)\) \( \in \mathcal{D}_+ \) and one choice in \( \mathcal{D}_- \), there is a unique solution with positive imaginary part. This can be done, for example, by choosing \( a = \pm 1/2, \nu = 1, \gamma = 1 \) and \( z = i \). \( \square \)
The explicit expression for $y(u)$ defined in Eq. (3.4) is given by

$$y(u; a, \nu, \gamma) = \begin{cases} 
0, & D \leq 0, \\
\frac{\sqrt{3}}{2}((\sqrt{D} + R)\frac{1}{3} + (\sqrt{D} - R)\frac{1}{3}), & D > 0,
\end{cases} \quad (3.7)$$

where

$$D = Q^3 + R^2,$$

$$R = (9\alpha_2\alpha_1 - 27\alpha_0 - 2\alpha_2^3)/54,$$

$$Q = (3\alpha_1 - \alpha_2^2)/9,$$

and

$$m^3 + \alpha_2m^2 + \alpha_1m + \alpha_0 = 0$$
is derived from Eq. (3.6) by multiplying \((\frac{a\nu}{\gamma})^{-1}\) on both sides. Explicitly,

\[
\alpha_2 = \frac{(\nu + au)\gamma}{a(\nu - a^2)},
\]

\[
\alpha_1 = \frac{(a + \gamma u)\gamma}{a(\nu - a^2)},
\]

\[
\alpha_0 = \frac{\gamma^2}{a(\nu - a^2)}.
\]

So all of \(\alpha_2, \alpha_1, \alpha_0\), and thus \(R, Q\) and \(D\) are real numbers. \(D\) is the “discriminant” of cubic equation, where \(D\) turning from negative to positive signals the emergence of a pair of complex solutions. The function \(y(u; a, \nu, \gamma)\) is plotted in Fig. 3.1 where \(\nu = 1, a = \sqrt{2/\pi}\) and \(\gamma = 0.1, 0.2, 0.3\). Notice the invariance of Eq. (3.6) under the transformation

\[\nu c^2 \to \nu, \; ac \to a, \; zc \to z, \; m/c \to m\]

where \(c\) is any positive constant, which corresponds to multiplying the kernel function by \(c\).

### 3.3 Proof of Thm. 3.4

By Prop. 2.1, it suffices to show the mean convergence of the Stieltjes transform, namely that \(E m_A(z)\) converges to the unique solution of Eq. (3.3) for any fixed \(z\).

Define

\[
\text{RHS}(m; a, \nu) = \left(-z - a \left(1 - \frac{1}{1 + \frac{\nu - a^2}{\gamma} m}\right) - \frac{\nu - a^2}{\gamma} m\right)^{-1},
\]

and Eq. (3.3) can be rewritten as

\[m = \text{RHS}(m; a, \nu).\]
Thus it suffices to show that

$$|\mathbb{E}m_A - \text{RHS}(\mathbb{E}m_A; a, \nu)| \to 0. \quad (3.9)$$

We firstly introduce a lemma concerning the stability of the Stieltjes transform under the mean squared perturbation of the matrix entries. A similar result has been presented in existing literature (see e.g. Lemma 2.1 of [15], the first statement), and we include the lemma and its proof here so as to make the text self-contained.

**Lemma 3.6.** Suppose that $A$ and $B$ are two $n$-by-$n$ symmetric matrices, and

$$\mathbb{E}|A_{ij} - B_{ij}|^2 \leq \epsilon n^{-1}, \quad 1 \leq i, j \leq n,$$

for some $\epsilon > 0$. Let $m_A(z)$ and $m_B(z)$ be the Stieltjes Transforms of $A$ and $B$ respectively. Then for a fixed $z$,

$$\mathbb{E}|m_A(z) - m_B(z)| \leq O(1) \sqrt{\epsilon}.$$

**Proof.** We have that

$$m_A(z) - m_B(z)$$

$$= \frac{1}{n} \left( \text{Tr}((A - zI)^{-1}) - \text{Tr}((B - zI)^{-1}) \right)$$

$$= \frac{1}{n} \text{Tr}((A - zI)^{-1}(B - A)(B - zI)^{-1}),$$
thus

\[ E|m_A(z) - m_B(z)|^2 \]
\[ = E \frac{1}{n^2} (\text{Tr}((A - zI)^{-1}(B - A)(B - zI)^{-1}))^2 \]
\[ \leq E \frac{1}{n^2} \text{Tr}(((B - zI)^{-1}(A - zI)^{-1})^2) \text{Tr}((B - A)^2) \]
\[ \leq E \frac{1}{n^2} n \sum_{i,j=1}^{n} |A_{ij} - B_{ij}|^2 \]
\[ \leq \frac{1}{v^4 n} n^2 \epsilon n^{-1} \]
\[ = O(1) \epsilon. \]

We will prove Thm. 3.4 in the following steps:

### 3.3.1 Step 1. Reduction to the case of finite expansion up to order \( L \)

Denote the truncated kernel function up to finite order \( L \) by

\[ f_L(\xi; p) = p^{-1/2} k_L(\sqrt{p} \xi; p) \]

where

\[ k_L(x; p) = \sum_{l=0}^{L} a_{l,p} P_{l,p}(x). \]

By Remark 3.1, we assume that \( a_{0,p} = 0 \) in the rest of the proof.

Let \( A_{\leq L} \) be the kernel matrix with the kernel function \( f_L(\xi; p) \), and \( m_A(z) \) and \( m_L(z) \) be the Stieltjes transforms of matrix \( A \) and \( A_{\leq L} \) respectively. By condition (C.p-Uniform), for arbitrary \( \epsilon > 0 \), there exists some \( L = L(\epsilon) \), so that

\[ E(k(\xi_p; p) - k_{L(\epsilon)}(\xi_p; p))^2 \leq \epsilon^2, \quad \forall p, \]
and then

\[ \mathbb{E}|A_{ij} - (A_{\leq L})_{ij}|^2 = \mathbb{E}(f(X^TY;p) - f_{L(\epsilon)}(X^TY;p))^2 \]
\[ \leq \epsilon^2 p^{-1}, \quad \forall 1 \leq i, j \leq n. \]

Since \( n, p \to \infty \) with a constant ratio, by Lemma 3.6,

\[ |\mathbb{E}m_A(z) - \mathbb{E}m_{L(\epsilon)}(z)| \leq \mathbb{E}|m_A(z) - m_{L(\epsilon)}(z)| \leq \mathcal{O}(1)\epsilon. \]

If in addition we can show that, for any fixed \( L \) and some sequence of \( a_L(p) \) and \( \nu_L(p) \),

\[ |\mathbb{E}m_L - \text{RHS}(\mathbb{E}m_L; a_L(p), \nu_L(p))| \to 0, \]
\[ a_L(p) \to a, \quad \nu_L(p) \to \nu, \quad \text{(3.10)} \]

then Eq. (3.9) holds asymptotically.

### 3.3.2 Step 2. Calculation of \( \mathbb{E}m_L(z) \) for finite \( L \)

With slight abuse of notation, we denote \( A_{\leq L} \), the random matrix with kernel function \( f_L(\xi;p) \), by \( A \). We omit the dependence on \( \epsilon \) and denote the Stieltjes transform by \( m_L(z) \). In what follows we sometimes drop the dependence on \( p \) and write \( f_L(\xi;p) \) as \( f_L(\xi) \), and similar for other functions.
the formula of \(((A - zI)^{-1})_{nn}\)

Observe that

\[
\mathbb{E}m_A(z) = \mathbb{E} \frac{1}{n} \text{Tr}(A - zI)^{-1} = \mathbb{E} \frac{1}{n} \sum_{i=1}^{n} ((A - zI)^{-1})_{ii} = \mathbb{E} ((A - zI)^{-1})_{nn}, \tag{3.11}
\]

where the last equality follows from that the rows/columns of \(A\) are exchangeable and so are those of \((A - zI)^{-1}\).

We have the following formula by Schur’s compliment

\[
((A - zI)^{-1})_{nn} = \frac{1}{(A_{nn} - z) - A^T_{,n}(A^{(n)} - zI_{n-1})^{-1}A_{,n}}, \tag{3.12}
\]

where \(I_{n-1}\) is the \((n - 1) \times (n - 1)\) identity matrix, \(A^{(n)}\) is the top left \((n - 1) \times (n - 1)\) minor of \(A\), and \(A_{,n}\) is the last column of \(A\) without the \(n\)-th entry. In other words, the matrix \(A\) is written in blocks as

\[
A = \begin{bmatrix}
A^{(n)} & A_{,n} \\
A_{,n}^T & A_{nn}
\end{bmatrix}.
\]

Notice that since \(\Im(z) > 0\), both \(A - zI\) and \(A^{(n)} - zI_{n-1}\) are invertible. Formula (3.12) can be verified by elementary linear algebra manipulation, see e.g. Thm. A.4 of [3].

Recall that \(A_{nn} = 0\), then

\[
\mathbb{E}m_L(z) = \mathbb{E} ((A - zI)^{-1})_{nn} = \mathbb{E}(-z - A^T_{,n}(A^{(n)} - zI_{n-1})^{-1}A_{,n})^{-1}. \tag{3.13}
\]
We need to study the convergence of the term

\[(\#) := A_{n}^T(A^{(n)} - zI_{n-1})^{-1}A_{n}.\]  \hfill (3.14)

**Conditioning on** \(X_n\)

In \((\#),\) \(A_{n}\) and \(A^{(n)}\) are dependent since they both involve the vectors \(X_1, \cdots, X_{n-1}\). The idea is to “separate” the part of \(A^{(n)}\) which is independent from \(A_{n}\) and control the residual.

To do so, we condition on the choice of \(X_n\), and write

\[X_i = \eta_i (X_n)_0 + \tilde{X}_i, \quad 1 \leq i \leq n - 1,\]  \hfill (3.15)

where \((X_n)_0 = \frac{X_n}{|X_n|}\) is the unit vector in the same direction of \(X_n\), and \(\tilde{X}_i\) lie in the \((p - 1)\)-dimensional subspace orthogonal to \(X_n\). Due to the orthogonal invariance of the standard multivariate Gaussian distribution, we know that \(\eta_i \sim \mathcal{N}(0, p^{-1})\), \(\tilde{X}_i \sim \mathcal{N}(0, p^{-1}I_{p-1})\), and they are independent. Now we have

\[X_i^T X_n = \eta_i |X_n|, \quad 1 \leq i \leq n - 1,\]  \hfill (3.16)

and

\[X_i^T X_j = \eta_i \eta_j + \tilde{X}_i^T \tilde{X}_j, \quad 1 \leq i, j \leq n - 1, i \neq j.\]  \hfill (3.17)

Define

\[\eta := (\eta_1, \cdots, \eta_{n-1})^T, \quad D_\eta := \text{diag}\{\eta_1^2, \cdots, \eta_{n-1}^2\},\]

and we have
1. $A_{n}$:

\[ A_{n} = f(1) + f(2), \]

\[ f(1) := a_{1,p} |X_n| \eta, \]

\[ f(2) := (f_{>1}(\xi_1), \cdots, f_{>1}(\xi_{n-1}))^T, \]

where $\xi_{in} := X_i^T X_n = |X_n| \eta_i$ for $1 \leq i \leq n - 1$, and

\[ f_{>1}(\xi) = \frac{1}{\sqrt{p}} \sum_{t=2}^{L} a_{t,p} P_{t,p}(\sqrt{p} \xi). \tag{3.18} \]

2. $A^{(n)}$: the off-diagonal entries of $A^{(n)}$ are

\[ A^{(n)}_{ij} = f_L(X_i^T X_j) = f_L(\eta_i \eta_j + \tilde{\xi}_{ij}), \quad 1 \leq i, j \leq n - 1, i \neq j, \]

where $\tilde{\xi}_{ij} = \tilde{X}_i^T \tilde{X}_j$. Further more,

\[ f_L(\eta_i \eta_j + \tilde{\xi}_{ij}) = a_{1,p} \eta_i \eta_j + a_{1,p} \tilde{\xi}_{ij} \]

\[ + f_{>1}(\tilde{\xi}_{ij}) + f'_{>1}(\tilde{\xi}_{ij}) \eta_i \eta_j + t_{ij}, \]

where

\[ t_{ij} = \frac{1}{2} f''_{>1}(\theta_{ij})(\eta_i \eta_j)^2. \]

**truncation of $|\eta_i|$ and $|\tilde{\xi}_{ij}|$ to $\sim \sqrt{\ln p/p}$**

We define the large probability set $\Omega_\delta$ as

\[ \Omega_\delta = \{|\eta_i| < \delta, |\tilde{\xi}_{ij}| < \delta, ||X_n|^2 - 1| < \sqrt{2}\delta, 1 \leq i, j \leq n - 1, i \neq j\}, \tag{3.19} \]

where $\delta = \frac{M}{\sqrt{p}}$, $M = \sqrt{20 \ln p}$. 

41
Lemma 3.7.

\[
\Pr(\Omega^c_\delta) \leq O(1)p^{-7}.
\]

The proof is postponed until Sec. 3.4.

decomposition of \(A^{(n)}\)

On \(\Omega_\delta\), \(|\sqrt{p}\tilde{\xi}_{ij}| < M\), then by Prop. 2.6

\[
f'_{>1}(\tilde{\xi}_{ij}) = \sum_{l=2}^{L} a_{l,p}(\sqrt{l}P_{l-1,p}(\sqrt{p}\tilde{\xi}_{ij}) + O_l(1)M^{l-1}p^{-1}),  
\]

and

\[
f''_{>1}(\tilde{\xi}_{ij}) = \sqrt{p} \sum_{l=2}^{L} a_{l,p}(\sqrt{l(l-1)}P_{l-2,p}(\sqrt{p}\tilde{\xi}_{ij}) + O_l(1)M^{l-2}p^{-1}).  
\]

We define

\[
\tilde{A}^{(n)}_{ij} = a_{1,p}\tilde{\xi}_{ij} + f_{>1}(\tilde{\xi}_{ij}),
\]

\[
\tilde{F}_{ij} = \frac{1}{\sqrt{p}} \sum_{l=2}^{L} a_{l,p}\sqrt{l}P_{l-1,p}(\sqrt{p}\tilde{\xi}_{ij}),
\]

and set the diagonal entries to be zeros for both \(\tilde{A}^{(n)}\) and \(\tilde{F}\). Then \(A^{(n)}\) admits the following decomposition

\[
A^{(n)} = \tilde{A}^{(n)} + a_{1,p}(\eta\eta^T - D_\eta) + \sqrt{p}W\tilde{F}W + T,
\]

where \(T\) is Hermitian, \(T_{ii} = 0\) and for \(i \neq j\)

\[
T_{ij} = t_{ij} + \eta_i\eta_j(f'_{>1}(\tilde{\xi}_{ij}) - \sqrt{p}\tilde{F}_{ij}),
\]

and \(W = \text{diag}\{\eta_1, \cdots, \eta_{n-1}\}\).
We define
\[
\hat{A}^{(n)} := a_{1,p}y^T + \tilde{A}^{(n)},
\]
(3.24)
We are going to “replace” \( A^{(n)} \) by \( \hat{A}^{(n)} \) in computing the limit of \((\#)\), and bound the difference. Define
\[
(\hat{\#}) := A^T_n(\hat{A}^{(n)} - zI_{n-1})^{-1}A_{..n},
\]
(3.25)
we have

**Lemma 3.8.** Let \( r_1 = (\hat{\#}) - (\#), (\hat{\#}) \) defined as in Eq. (3.25), and \((\#)\) defined as in Eq. (3.14). \( \Omega_\delta \) defined as in Eq. (3.19), as \( p, n \to \infty \) with \( p/n \to \gamma \),
\[
\mathbb{E}|r_1| \cdot 1_{\Omega_\delta} = \mathcal{O}_L(1)M^{2L+2}p^{-1/4}.
\]
The proof is postponed until Sec. 3.4.

**computation of \((\hat{\#})\)**

Recall that \( A_{..n} = f(1) + f(2) \) as defined in Eq. (3.18),
\[
(\hat{\#}) = A^T_n(\hat{A}^{(n)} - zI_{n-1})^{-1}A_{..n}
= f_{(1)}(\hat{A}^{(n)} - zI_{n-1})^{-1}f(1) + f_{(2)}(\hat{A}^{(n)} - zI_{n-1})^{-1}f(2) + r_2
:= (\hat{\#})_1 + (\hat{\#})_2 + r_2.
\]
(3.26)
where
\[
|r_2| = 2f_{(1)}^T(\hat{A}^{(n)} - zI_{n-1})^{-1}f_{(2)}.
\]
(3.27)

We analyze Eq. (3.26) term by term in the following three lemmas, and the proofs are postponed until Sec. 3.4:

**Lemma 3.9.**
\[
\mathbb{E}|r_2| \cdot 1_{\Omega_\delta} \leq \mathcal{O}_L(1)M^2p^{-1/2}.
\]
Lemma 3.10. Define
\[ \tilde{m}(z) := \frac{1}{n-1} \text{Tr}(\tilde{A}^{(n)} - zI_{n-1})^{-1}. \] (3.28)

We have
\[ (\hat{#})_1 = a_{1,p} \left( 1 - (1 + \frac{a_{1,p}}{\gamma} \tilde{m}(z))^{-1} \right) + r_{(1)}, \] (3.29)
where
\[ \mathbb{E}|r_{(1)}| \cdot 1_{\Omega_{1}^c \cap \Omega_{(1)}} \leq O(1)M^{4p-1/2}, \] (3.30)
and \( \Omega_{(1)} \) is a large probability set satisfying \( \Pr(\Omega_{(1)}^c) \leq O(1)p^{-1/4} \).

Lemma 3.11.
\[ (\hat{#})_2 = \frac{\nu_{>1,p}}{\gamma} \tilde{m}(z) + r_{(2)}, \] (3.31)
where \( \nu_{>1,p} = \nu_p - a_{1,p}^2 \), and
\[ \mathbb{E}|r_{(2)}| \cdot 1_{\Omega_{(1)}} \leq O(1)M^{L+1}p^{-1/2}. \]

convergence of \( \mathbb{E}m_L(z) \)

Suppose that \( m(z) \) is the Stieltjes transform of a probability measure for fixed \( z = u + iv \). Notice that

1. \( |m(z)| \leq \frac{1}{v} \) by Eq. (A.2).

2. For any constants \( a \) and \( \nu \) where \( a^2 \leq \nu \),

\[ |\text{RHS}(m, a, \nu)| \leq \frac{1}{v}. \]
The proof is as follows: define the denominator of RHS\((m, a, \nu)\) to be \(-z - \text{POUND}(m, a, \nu)\), where

\[
\text{POUND}(m, a, \nu) = a \left( 1 - \frac{1}{a + \frac{a}{\gamma m}} \right) + \frac{\nu - a^2}{\gamma} m. \tag{3.32}
\]

It suffices to show that \(\Im(\text{POUND}(m, a, \nu)) \geq 0\), which can be verified by the expression and that \(\nu - a^2 \geq 0\), \(\gamma > 0\) and \(\Im(m) > 0\).

Eq. (3.13) combined with lemmas 3.8, 3.9, 3.10, 3.11 gives that

\[
\mathbb{E} m_L(z) = \mathbb{E} \frac{1}{-z - \text{POUND}(\mathbb{E} \tilde{m}(z), a_{1, p}, \nu_p) + r},
\]

where \(r := -r_{(1)} - r_{(2)} + r_2 - r_1\), and

\[
\mathbb{E}|r| \mathbb{1}_{\Omega_{(1)} \cap \Omega} \leq O(1)M^{2L+2}p^{-1/4}.
\]

By Lemma 3.7 and Lemma 3.10,

\[
\text{Pr}\{(\Omega_{(1)} \cap \Omega_{(1)})^c\} \leq \text{Pr}(\Omega_{(1)}) + \text{Pr}(\Omega_{(1)}^c) \leq O(1)p^{-7} + O(1)p^{-1/4} = O(1)p^{-1/4}.
\]

Thus we have

\[
|\mathbb{E} m_L(z) - \text{RHS}(\mathbb{E} \tilde{m}(z), a_{1, p}, \nu_p)|

= \left| \mathbb{E} \frac{1}{-z - \text{POUND}(\mathbb{E} \tilde{m}(z), a_{1, p}, \nu_p) + r} - \text{RHS}(\mathbb{E} \tilde{m}(z), a_{1, p}, \nu_p) \right|

\leq \frac{2}{v} \text{Pr}\{(\Omega_{(1)} \cap \Omega_{(1)})^c\}

+ \mathbb{E} \left| \frac{1}{-z - \text{POUND}(\mathbb{E} \tilde{m}(z), a_{1, p}, \nu_p) + r} - \text{RHS}(\mathbb{E} \tilde{m}(z), a_{1, p}, \nu_p) \right| \cdot \mathbb{1}_{\Omega_{(1)} \cap \Omega}

\leq \frac{2}{v} \text{Pr}\{(\Omega_{(1)} \cap \Omega_{(1)})^c\} + \frac{2}{v} \mathbb{E}|r| \mathbb{1}_{\Omega_{(1)} \cap \Omega}

\leq O(1)p^{-1/4} + O(1)M^{2L+2}p^{-1/4} \to 0. \tag{3.33}
\]
We also have the following lemma, the proof of which is postponed until Sec. 3.4.

**Lemma 3.12.** For fixed $z = u + iv$, as $p, n \to \infty$, $p/n \to \gamma$,

\[
\mathbb{E}|m_L(z) - \tilde{m}(z)| \to 0.
\]

Eq. (3.33) and Lemma 3.12 imply that (dropping the dependence on $z$)

\[
|\mathbb{E}\tilde{m} - \text{RHS}(\mathbb{E}\tilde{m}; a_1, \nu_p)| \to 0,
\]

and thus

\[
|\mathbb{E}m_L - \text{RHS}(\mathbb{E}m_L; a_1, \nu_p)| \to 0.
\]

At last, by condition (C.Variance) and (C.a1), $a_{1,p} \to a$ and $\nu_p \to \nu$. Thus Eq. (3.10) is verified if we set $a_L(p) = a_{1,p}$ and $\nu_L(p) = \nu_p$. This finishes the proof of Thm. 3.4.

### 3.4 Proof of the Lemmas

*Proof of Lemma 3.1.* When $k = 2$, $\mathbb{E}|X|^2 = 1$.

When $k = 1$,

\[
\mathbb{E}|X| \leq \sqrt{\mathbb{E}|X|^2} = 1.
\]

At the same time, notice that $\sqrt{p}|X| \sim \chi(p)$, thus by the expression of the expectation of the Chi distribution [36], we have that

\[
\mathbb{E}|X| = \sqrt{\frac{2}{p}} \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)}.
\]
When $p$ is even, Eq. (3.1) is verified by the formula

$$
\Gamma(m + \frac{1}{2}) = \frac{(2m - 1)!! \sqrt{\pi}}{2^m}
$$

and the Stirling Formula

$$
n! = \sqrt{2\pi n} (n/e)^n (1 + \mathcal{O}(n^{-1})).
$$

When $p$ is odd,

$$
1 \geq \mathbb{E}|X| = \frac{\mathbb{E} \chi(p)}{\sqrt{p}} \geq \frac{\mathbb{E} \chi(p - 1)}{\sqrt{p}},
$$

and thus is reduced to the case when $p$ is even.

When $k = 2m - 1, m \geq 2$,

$$
1 = (\mathbb{E}|X|^2)^{(2m-1)/2} \leq \mathbb{E}|X|^{2m-1} \leq \sqrt{\mathbb{E}|X|^2 \mathbb{E}|X|^{4m-4}} = \sqrt{\mathbb{E}|X|^{4m-4}}
$$

and thus is reduced to the case where $k \geq 4$ and is even, which is analyzed below.

When $k = 2m, m \geq 2$, let

$$
y := |X|^2 = \frac{1}{p} \sum_{j=1}^{p} x_j^2, \quad x_1, \cdots, x_p \text{ i.i.d. } \sim \mathcal{N}(0, 1),
$$

and the characteristic function of $y$ is

$$
\phi(t) := \mathbb{E} e^{ty} = (\mathbb{E} e^{tx_j^2/p})^p = \left(1 - \frac{2t}{p}\right)^{-p/2}, \quad 0 < \frac{t}{p} < \frac{1}{2}.
$$
This gives that
\[
\mathbb{E}|X|^{2m} = \mathbb{E}y^m = \left. \frac{d^m}{dt^m} \phi(t) \right|_{t=0}
= \prod_{j=1}^{m-1} \left( 1 + \frac{2j}{p} \right)
= 1 + \frac{m(m-1)}{p} + O_m(1)p^{-2}.
\]

Lemma 3.13. Suppose that \(Z\) is a random variable, and for positive integer \(l\) and \(\epsilon > 0,\)
\[
|\mathbb{E}Z^k - 1| < \epsilon, \quad 0 \leq k \leq l.
\]
Let \(q(x) = \sum_{k=0}^l b_k x^k\) is a polynomial of degree \(l,\) then
\[
|\mathbb{E}q(Z)| \leq |q(1)| + \epsilon \sum_{j=0}^l |b_j|.
\]

The proof is elementary and is omitted.

Proof of Lemma 3.2. There exists an orthogonal transform \(P_1\) that depends on \(X_1\) so that
\[
P_1X_1 = (|X_1|, 0, \cdots, 0)^T,
\]
and let
\[
P_1X_i = (\eta_i, \tilde{X}_i)^T, \quad i = 2, 3, 4.
\]
Since \(X_1, \cdots, X_4\) i.i.d. \(\sim \mathcal{N}(0, p^{-1}I_p),\) \(|X_1| \sim \chi(p)/\sqrt{p}, \) \(\eta_i \sim \mathcal{N}(0, p^{-1}),\) \(\tilde{X}_i \sim \mathcal{N}(0, p^{-1}I_{p-1})\) for \(i = 2, 3, 4\) are independent. Also, there exists an orthogonal transform \(P_3\) which applies to the 2-to-\(p\) coordinates so that
\[
P_3P_1X_3 = (\eta_3, |\tilde{X}_3|, 0, \cdots, 0)^T,
\]
and let
\[ P_3 P_1 X_i = (\eta_i, \tilde{\eta}_i, \cdots)^T, \quad i = 2, 4. \]

By the independence of \(X_1, \cdots, X_4\) and that \(P_3\) only depends on \(X_1, X_3, |\tilde{X}_3| \sim \chi(p - 1)/\sqrt{p}, \tilde{\eta}_n \sim \mathcal{N}(0, p^{-1})\) for \(i = 2, 4\), and \(|X_1|, \eta_i, i = 2, 3, 4, |\tilde{X}_3|\) and \(\tilde{\eta}_i, i = 2, 4\) are jointly independent.

Thus,
\[
\begin{align*}
\xi_{12} &= |X_1|\eta_2, \quad \xi_{14} = |X_1|\eta_4, \quad \xi_{23} = \eta_2\eta_3 + |\tilde{X}_3|\tilde{\eta}_2, \quad \xi_{34} = \eta_3\eta_4 + |\tilde{X}_3|\tilde{\eta}_4,
\end{align*}
\]

and define
\[
\begin{align*}
\zeta_2 := \sqrt{p}\eta_2, \quad \zeta_4 := \sqrt{p}\eta_4, \quad \tilde{\zeta}_2 := \sqrt{p}\tilde{\eta}_2, \quad \tilde{\zeta}_4 := \sqrt{p}\tilde{\eta}_4,
\end{align*}
\]

which are i.i.d. distributed as \(\mathcal{N}(0, 1)\), we have that
\[
\sqrt{p}\xi_{12} = |X_1|\zeta_2, \quad \sqrt{p}\xi_{14} = |X_1|\zeta_4, \quad \sqrt{p}\xi_{23} = \eta_3\zeta_2 + |\tilde{X}_3|\tilde{\zeta}_2, \quad \sqrt{p}\xi_{34} = \eta_3\zeta_4 + |\tilde{X}_3|\tilde{\zeta}_4.
\]

Since \(P_{l,p}(x)\) is a polynomial of degree \(l\),
\[
P_{l,p}(x_1 + x_2) = \sum_{k=0}^{l} \frac{P_{l,p}^{(k)}(x_2)}{k!} x_1^k,
\]

thus
\[
\begin{align*}
\mathbb{E}P_{l,p}(\sqrt{p}\xi_{12})P_{l,p}(\sqrt{p}\xi_{23})P_{l,p}(\sqrt{p}\xi_{34})P_{l,p}(\sqrt{p}\xi_{41})
&= \sum_{j,k=0}^{l} \mathbb{E}\eta_3^{k+j} \mathbb{E}\left( \zeta_2^k P_{l,p}(|X_1|\zeta_2) \zeta_4^j P_{l,p}(|X_1|\zeta_4) \right) \\
&\quad \frac{\mathbb{E}\left( \frac{P_{l,p}^{(k)}(|\tilde{X}_3|\tilde{\zeta}_2)}{k!} \frac{P_{l,p}^{(j)}(|\tilde{X}_3|\tilde{\zeta}_4)}{j!} \right)}{k!j!}.
\end{align*}
\] (3.34)
Meanwhile, let
\[ P_{t,p}(x) = \sum_{j=0}^{l} (c_{t,p})_j x^j, \]
and recall that by Lemma 2.4 and Eq. (B.3),
\[ P_{t,p}(x) = h_t(x) + \sum_{j=0}^{l} (r_{t,p})_j x^j, \]
\[ P'_{t,p}(x) = \sqrt{h_{t-1}(x)} + \sum_{j=0}^{l-1} (r_{t,p})_{j+1} (j+1)x^j, \]
\[ \max_{0 \leq j \leq l} |(r_{t,p})_j| = O(l)p^{-1}, \]
and as a result
\[ \max_{0 \leq j \leq l} |(c_{t,p})_j| = O(1). \]

For the claim in the lemma to hold, it suffices to show that each term in Eq. (3.34) is \( O(l)p^{-2} \). Notice that when \( k + j \) is odd, \( \mathbb{E} \eta_3^{k+j} \) vanishes. Then we have the following cases:

1. \( k = j = 0 \): In \( \mathbb{E} (P_{t,p}(|X_1|\zeta_2)P_{t,p}(|X_1|\zeta_4)) \), since \( \zeta_2, \zeta_4 \) and \( |X_1| \) are independent, we can take expectation with respect to \( \zeta_2, \zeta_4 \) first, which gives that
\[ \mathbb{E} (P_{t,p}(|X_1|\zeta_2)P_{t,p}(|X_1|\zeta_4)) = \mathbb{E} q_0(|X_1|^2), \]
where
\[ q_0(r) = \sum_{j=0}^{l} (c_{t,p})_j \mathbb{E} \zeta^j \cdot r^j, \quad \zeta \sim \mathcal{N}(0,1), \]
and \( q_0(1) = \sum_{j=0}^{l} (c_{t,p})_j \mathbb{E} \zeta^j = \mathbb{E} h_t(\zeta) + \sum_{j=0}^{l} (r_{t,p})_j \mathbb{E} \zeta^j. \) Because \( l \geq 2, \mathbb{E} h_t(\zeta) = 0, \) and then
\[ |q_0(1)| \leq \sum_{j=0}^{l} |(r_{t,p})_j| \mathbb{E} \zeta^j \leq O(l)p^{-1}. \]
Applying Lemma 3.13 to $q = q_{0l}$ and $Z = |X_1|$, together with Lemma 3.1, we have that

$$\mathbb{E}_{q_{0l}}(|X_1|^2) \leq (O_l(1)p^{-1})^2 + \max_{0 \leq k \leq 2l} |E|X_1|^k - 1| \cdot O_l(1) = O_l(1)p^{-1}.$$

For same reason, we have that $|E \left( P_{l,p}(|\tilde{X}_3|\tilde{\zeta}_2) P_{l,p}(|\tilde{X}_3|\tilde{\zeta}_4) \right) | = O_l(1)p^{-1}$. This implies the $O_l(1)p^{-2}$ bound for the term of $j = k = 0$.

2. $k = 0, j = 2$ or $k = 2, j = 0$: $\mathbb{E}_{\eta_3}^2 = p^{-1}$. Taking $k = 0, j = 2$ as an example, for $E (P_{l,p}(|X_1|\zeta_2)P_{l,p}(|X_1|\zeta_4))$ and $E \left( P_{l,p}(|\tilde{X}_3|\tilde{\zeta}_2) P_{l,p}^{(2)}(|\tilde{X}_3|\tilde{\zeta}_4) \right)$, a similar argument as above gives a bound of $O_l(1)p^{-1}$ for each of them. Then the term of $k = 0, j = 2$ is bounded by $O_l(1)p^{-3}$.

3. $k = 1, j = 1$: $\mathbb{E}_{\eta_3}^2 = p^{-1}$. For $E (\zeta_2 P_{l,p}(|X_1|\zeta_2)\zeta_4 P_{l,p}(|X_1|\zeta_4))$, as $l \geq 2$, $E \zeta_h l(\zeta) = 0$, and a similar argument as above gives a bound of $O_l(1)p^{-1}$. For $E \left( P_{l,p}^{(2)}(|\tilde{X}_3|\tilde{\zeta}_2) P_{l,p}^{(2)}(|\tilde{X}_3|\tilde{\zeta}_4) \right)$, as $E h_{l-1}(\zeta) = 0$, we have a bound of $O_l(1)p^{-1}$. Then the term of $k = 1, j = 1$ is bounded by $O_l(1)p^{-3}$.

4. $2 \leq k, j \leq l$ (and $k + j$ is even): $\mathbb{E}_{\eta_3}^{k+j} = (k + j - 1)!! p^{-(k+j)/2}$ which is at least $O_l(1)p^{-2}$. For similar reason as above, both $E (\zeta_2^k P_{l,p}(|X_1|\zeta_2)\zeta_4 P_{l,p}(|X_1|\zeta_4))$ and $E \left( P_{l,p}^{(k)}(|\tilde{X}_3|\tilde{\zeta}_2) P_{l,p}^{(j)}(|\tilde{X}_3|\tilde{\zeta}_4) \right)$ can be shown to be $O_l(1)$. Then the term is bounded by $O_l(1)p^{-2}$.

Proof of Lemma 3.7. For $\eta_i$ we have the concentration inequality

$$\Pr[|\eta_i| > \delta] = 2 \int_M^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \leq \frac{1}{\sqrt{2}} e^{-\frac{M^2}{2}} = \frac{1}{\sqrt{2}} p^{-10}, \quad 1 \leq i \leq n - 1.$$  \hfill (3.35)
For each \( \tilde{\xi}_{ij} \), recall that \( \tilde{\xi}_{ij} = \tilde{X}_{i}^{T} \tilde{X}_{j} \), where \( \tilde{X}_{i}, \tilde{X}_{j} \) are independently distributed as \( \mathcal{N}(0, p^{-1}I_{p-1}) \). Conditioning on the draw of \( \tilde{X}_{i} \), we have

\[
\tilde{\xi}_{ij} = |\tilde{X}_{i}| \tilde{\eta}_{ij},
\]

where \( \tilde{\eta}_{ij} \) has marginal distribution \( \mathcal{N}(0, p^{-1}) \) and is independent of \( |\tilde{X}_{i}| \). We firstly observe that when \( p \) is large, \( |X_{i}|^{2} \) concentrates at 1, and specifically

\[
\Pr \left[ ||X_{i}|^{2} - 1| > \sqrt{\frac{40 \ln p}{p}} \right] < p^{-9}, \tag{3.36}
\]

which can be verified by standard large deviation inequality techniques. Then we have

\[
\Pr[||\tilde{X}_{i}|| \tilde{\eta}_{ij} > \delta] \leq \Pr \left[ |\tilde{X}_{i}| > 1 + \sqrt{\frac{40 \ln p}{p}} \right]
+ \Pr \left[ |\tilde{X}_{i}| |\tilde{\eta}_{ij}| > \delta, |\tilde{X}_{i}| < 1 + \sqrt{\frac{40 \ln p}{p}} \right]
\leq p^{-9} + \Pr[|\tilde{\eta}_{ij}| > \delta - \frac{\delta}{1.01}]
\leq p^{-9} + \frac{1}{\sqrt{2}} p^{-9},
\]

thus

\[
\Pr[|\tilde{\xi}_{ij}| > \delta] < \mathcal{O}(1)p^{-9}.
\]

Then, a union bound gives

\[
\Pr(\Omega_{\xi}^{c}) \leq (n - 1) \Pr[|\tilde{\eta}_{i}| > \delta]
+ \frac{(n - 1)(n - 2)}{2} \Pr[|\tilde{\xi}_{ij}| > \delta] + \Pr[||X_{n}|^{2} - 1| > \sqrt{2}\delta]
\leq \mathcal{O}(1)p^{-9} + \mathcal{O}(1)p^{-7} + p^{-9} = \mathcal{O}(1)p^{-7}.
\]

\( \square \)
Proof of Lemma 3.8.

\[ r_1 = A^T_{-n}((\hat{A}^{(n)} - zI_{n-1})^{-1} - (A^{(n)} - zI_{n-1})^{-1})A_{-n} \]
\[ = A^T_{-n}(\hat{A}^{(n)} - zI_{n-1})^{-1}(\sqrt{pW\tilde{F}W} + T - a_1pD_\eta) \]
\[ \cdot (A^{(n)} - zI_{n-1})^{-1}A_{-n}, \]

thus,

\[ |r_1| \leq ||A_{-n}||^2s((\hat{A}^{(n)} - zI_{n-1})^{-1}) \]
\[ \cdot (s(\sqrt{pW\tilde{F}W}) + s(T) + s(a_1pD_\eta))s((A^{(n)} - zI_{n-1})^{-1}) \]
\[ \leq \frac{1}{v^2}||A_{-n}||^2(s(\sqrt{pW\tilde{F}W}) + s(T) + s(a_1pD_\eta)). \tag{3.37} \]

On \( \Omega_\delta \), we have the following bounds

1. \( s(T) \): since \( \theta_{ij} \) is between \( \tilde{\xi}_{ij} \) and \( \tilde{\xi}_{ij} + \eta_i\eta_j \), and both \( \tilde{\xi}_{ij} \) and \( \eta \) are bounded in magnitude by \( \delta = p^{-1/2}M \),

\[ |\theta_{ij}| \leq \delta + \delta^2 \leq 1.01\delta = p^{-1/2}1.01M. \]

By Eq. (3.21), Prop. 2.5, and Lemma 2.7,

\[ |f''_{>1}(\theta_{ij})| \leq \sqrt{pO_L(1)}M^{L-2}, \]

and then

\[ |t_{ij}| \leq \delta^4 |f''_{>1}(\theta_{ij})| \leq O_L(1)M^{L+2}p^{-3/2}. \]

Meanwhile, Eq. (3.20) and Lemma 2.7 imply that

\[ |f'_{>1}(\tilde{\xi}_{ij}) - \sqrt{p\tilde{F}_{ij}}||\eta_i||\eta_j| \leq O_L(1)M^{L-1}p^{-1} \cdot M^2p^{-1} = O_L(1)M^{L+1}p^{-2}. \]
Thus

\[ |T_{ij}| \leq O_L(1)M^{L+2}p^{-3/2} + O_L(1)M^{L+1}p^{-2} = O_L(1)M^{L+2}p^{-3/2}. \quad (3.38) \]

Thus

\[ s(T) \leq \sqrt{\sum_{ij} |T_{ij}|^2} \leq n \max_{ij} |T_{ij}| \leq O_L(1)M^{L+2}p^{-1/2}. \]

2. \( s(a_{1,p}D_\eta) \): \( |a_{1,p}| \) is \( O(1) \) by Lemma 2.7, thus

\[ |a_{1,p}|\eta_i^2 \leq |a_{1,p}|\delta^2 \leq O(1)M^2p^{-1}, \quad i = 1, \ldots, n-1, \]

which means that

\[ s(a_{1,p}D_\eta) \leq O(1)M^2p^{-1}. \]

3. \( s(\sqrt{pW\tilde{F}W}) \): \( \tilde{F} \) can be written as \( \sum_{l=1}^{L-1} a_{l,p}\sqrt{l}\tilde{F}_l \), where Prop. 3.3 applies to each \( \tilde{F}_l \). Lemma 2.7 means that the coefficients \( |a_{l,p}| \) for \( 1 \leq l \leq L-1 \) are uniformly bounded by \( O(1) \), thus we have

\[ \mathbb{E}s(\tilde{F}) \leq \sum_{i=2}^{L} \sqrt{i}O_l(1)p^{1/4} = O_L(1)p^{1/4}. \quad (3.39) \]

and as a result,

\[ \mathbb{E}s(\sqrt{pW\tilde{F}W}) \cdot 1_{\Omega_3} \leq M^2p^{-1/2}\mathbb{E}s(\tilde{F}) \leq O_L(1)M^2p^{-1/4}. \quad (3.40) \]

4. \( ||A_{n,n}||^2 \):

\[ ||A_{n,n}||^2 = \sum_{i=1}^{n-1} f_L(\xi_m)^2 \leq O_L(1)M^L. \]
The first three bounds imply that
\[
\mathbb{E} s(\sqrt{p}W \tilde{F}W + T - a_{1,p}D_\eta) \cdot 1_{\Omega_5} \\
\leq O_L(1)M^2p^{-1/4} + O_L(1)M^{L+2}p^{-1/2} + O(1)M^2p^{-1} \\
= O_L(1)M^2p^{-1/4}.
\]
(3.41)

Plugging into Eq. (3.37) gives that
\[
\mathbb{E} |r_1| \cdot 1_{\Omega_5} \leq O(1)\mathbb{E} s(\sqrt{p}W \tilde{F}W + T - a_{1,p}D_\eta)||A_{\gamma,n}||^2 \cdot 1_{\Omega_5} \\
= O_L(1)M^{2L+2}p^{-1/4}.
\]
(3.42)

Proof of Lemma 3.9.

\[
r_2 = 2a_{1,p}|X_n|\eta^T(\hat{A}^{(n)} - zI_{n-1})^{-1}f_{(2)} \\
= 2a_{1,p}f_{(2)}^T(\hat{A}^{(n)} - zI_{n-1})^{-1}(|X_n|\eta) \\
= 2a_{1,p}\{f_{(2)}^T(\hat{A}^{(n)} - zI_{n-1})^{-1}(|X_n|\eta) \\
- f_{(2)}^T(\hat{A}^{(n)} - zI_{n-1})^{-1}a_{1,p}\eta\eta^T(\hat{A}^{(n)} - zI_{n-1})^{-1}(|X_n|\eta)\} \\
:= 2a_{1,p}(r_{2,1} - r_{2,2}),
\]
(3.43)

where \(\hat{A}^{(n)}\) is defined as in Eq. (3.22).

Firstly,
\[
r_{2,1} = f_{(2)}^T(\hat{A}^{(n)} - zI_{n-1})^{-1}(|X_n|\eta)
\]
satisfies \(\mathbb{E}|r_{2,1}| \leq O_L(1)p^{-1/2}\) by a moment bound: recall the definition of \(\xi_{in}\) as in Eq. (3.16), and that \(f_{>1}(\xi)\) is a linear combination of rescaled and renormalized Hermite-like polynomials of degree \(\geq 2\). Also, \(\mathbb{E}|X_n|^{2m} = 1 + O_m(1)p^{-1}\) (Lemma 3.1), and \(|X_n|\) is independent from \(\eta_i\)’s and \(\hat{X}_i\)’s. Denote \(\tilde{B} = (\hat{A}^{(n)} - zI_{n-1})^{-1}\). By taking
expectation over $|X_n|$ first and then over $\eta_i$’s, we have

$$\mathbb{E}|r_{2,1}|^2 = \mathbb{E} \left| \sum_{i_1, i_2=1}^{n-1} f_{>1}(\xi_{i_1 n})\xi_{i_2 n} \tilde{B}_{i_1 i_2} \right|^2$$

$$= \mathbb{E} \sum_{i_1, i_2, i'_1, i'_2} f_{>1}(\xi_{i_1 n})\xi_{i_2 n} f_{>1}(\xi_{i'_1 n})\xi_{i'_2 n} \tilde{B}_{i_1 i_2} \tilde{B}_{i'_1 i'_2}$$

$$= \{i_1 = i_2 = i'_1 = i'_2\} + \{i_1, i_2 = i'_1 = i'_2, \text{or } i'_2 \text{ as } i_1\} + \{i_2 = i'_2, i_1, i'_1\} + \{i_1 = i_2, i'_1 = i'_2, \text{or } i'_1 \text{ as } i_1\} + \{i_1 = i'_1, i_2 = i'_2\}$$

$$= \mathcal{O}_L(1)p^{-1} + \nu_{>1, p}^{-2} \mathbb{E} \text{Tr}(\tilde{B} \tilde{B})$$

$$\leq \mathcal{O}_L(1)p^{-1} + \mathcal{O}(1) \cdot p^{-2} \frac{n}{\nu^2} = \mathcal{O}_L(1)p^{-1}.$$ 

By $\{i_1, i_2 = i'_1 = i'_2\}$ we denote the term in summation where the last three indices take the same value while $i_1$ is distinct from them, and similar for others.

Secondly,

$$r_{2,2} = (f_{(2)}(\hat{A}(n) - zI_{n-1})^{-1}(|X_n|\eta)) (a_1(p)\eta^T(\hat{A}(n) - zI_{n-1})^{-1}\eta)$$

$$= r_{2,1}(a_1(p)\eta^T(\hat{A}(n) - zI_{n-1})^{-1}\eta),$$

where

$$|a_1(p)\eta^T(\hat{A}(n) - zI_{n-1})^{-1}\eta| \cdot 1_{\Omega_3} \leq \mathcal{O}(1)s((\hat{A}(n) - zI_{n-1})^{-1})||\eta||^2 \cdot 1_{\Omega_3}$$

$$\leq \mathcal{O}(1)M^2 = \mathcal{O}(1)M^2,$$

thus

$$\mathbb{E}|r_{2,2}| \cdot 1_{\Omega_3} \leq \mathcal{O}(1)M^2 \mathbb{E}|r_{2,1}| \leq \mathcal{O}(1)M^2 \cdot \mathcal{O}_L(1)p^{-1/2} = \mathcal{O}_L(1)M^2p^{-1/2}. $$
Then\[\mathbb{E}|r_2| \cdot 1_{\Omega_3} \leq \mathcal{O}(1)(\mathbb{E}|r_{2,1}| + \mathbb{E}|r_{2,2}| \cdot 1_{\Omega_3}) \leq \mathcal{O}_L(1)M^2p^{-1/2}.\]

**Proof of Lemma 3.10.** Recall that \( \hat{A}^{(n)} = a_{1,p}\eta\eta^T + \tilde{A}^{(n)} \) (Eq. (3.24)). We have

\[
\left(\hat{\#}\right)_1 = |X_n|^2a_{1,p}^2\eta^T(a_{1,p}\eta\eta^T + \hat{A}^{(n)} - zI_{n-1})^{-1}\eta
= |X_n|^2a_{1,p} \cdot \frac{a_{1,p}\eta^T(\hat{A}^{(n)} - zI_{n-1})^{-1}\eta}{1 + a_{1,p}\eta^T(\hat{A}^{(n)} - zI_{n-1})^{-1}\eta}
= |X_n|^2a_{1,p} \left(1 - \frac{1}{1 + a_{1,p}\eta^T(\hat{A}^{(n)} - zI_{n-1})^{-1}\eta}\right),
\]

where to get the 2nd line we use the *Sherman-Morrison* formula

\[
q^T(pq^T + M - zI)^{-1} = \frac{q^T(M - zI)^{-1}}{1 + q^T(M - zI)^{-1}p}, \quad \forall p, q.
\]

Further more,

\[
\eta^T(\hat{A}^{(n)} - zI_{n-1})^{-1}\eta = \gamma^{-1}\mathbb{E}\tilde{m}(z) + \gamma^{-1}\tilde{r} + r_{(1),2},
\]

where \( \tilde{m}(z) \) is its Stieltjes transform of \( \hat{A}^{(n)} \), and

1. \( \tilde{r} = \tilde{m}(z) - \mathbb{E}\tilde{m}(z) \), and by Prop. 2.1

\[
\mathbb{E}|\tilde{r}| \leq \mathcal{O}(1)n^{-1/2}.
\]

2. \( r_{(1),2} = \eta^T(\hat{A}^{(n)} - zI_{n-1})^{-1}\eta - \frac{1}{p}\text{Tr}(\hat{A}^{(n)} - zI_{n-1})^{-1}, \)

and by moment method one can show that (Lemma 3.14)

\[
\mathbb{E}|r_{(1),2}| \leq \mathcal{O}(1)p^{-1/2}.
\]
To obtain a bound of \( r(1) \) defined in Eq. (3.29), we bound the quantities \(|(1 + a_{1,p} \eta^T (A^{(n)} - z I_{n-1})^{-1} \eta)| \) and \(|(1 + \frac{a_{1,p}}{\gamma} \mathbb{E} \tilde{m}(z))^{-1}| \) as follows:

1. 

\[
|(1 + a_{1,p} \eta^T (A^{(n)} - z I_{n-1})^{-1} \eta)| \cdot 1_{\Omega_\delta} \leq O(1) M^2 := M'.
\] (3.49)

Recall that as in Eq. (3.44),

\[
a_{1,p} \eta^T (A^{(n)} - z I_{n-1})^{-1} \eta = 1 - \frac{1}{1 + a_{1,p} \eta^T (A^{(n)} - z I_{n-1})^{-1} \eta},
\]

which gives that

\[
\left| \frac{1}{1 + a_{1,p} \eta^T (A^{(n)} - z I_{n-1})^{-1} \eta} \right| = \left| 1 - a_{1,p} \eta^T (A^{(n)} - z I_{n-1})^{-1} \eta \right|
\]

\[
\leq 1 + |a_{1,p}| s(\hat{A}^{(n)} - z I_{n-1}) ||\eta||^2.
\]

Recall that on \( \Omega_\delta \), \(|\eta| < Mp^{-1/2}\) so that \(||\eta||^2 \leq O(1) M^2\), and that \(|a_{1,p}| \leq O(1)\) (Lemma 2.7) and \(s(\hat{A}^{(n)} - z I_{n-1})^{-1}) \leq \frac{1}{z}\), then the bound Eq. (3.49) follows.

2. Define

\[
\Omega_{(1)} := \{|\tilde{r}| \leq p^{-1/4}, |r_{(1),2}| \leq p^{-1/4}\}.
\]

By Eq. (3.46, 3.48) and Markov inequality, we have

\[
\mathbb{P}r(\Omega_{(1)}^c) \leq p^{1/4} \mathbb{E}|\tilde{r}| + p^{1/4} \mathbb{E}|r_{(1),2}| \leq O(1)p^{-1/4}.
\]

By Eq. (3.45),

\[
\left| \frac{1}{1 + \frac{a_{1,p}}{\gamma} \mathbb{E} \tilde{m}(z)} \right| = \left| \frac{1}{1 + a_{1,p} \eta^T (A^{(n)} - z I_{n-1})^{-1} \eta - \frac{a_{1,p}}{\gamma} \tilde{r} - a_{1,p} r_{(1),2}} \right|
\]

\[
\leq \left| \frac{1}{1 + a_{1,p} \eta^T (A^{(n)} - z I_{n-1})^{-1} \eta} - \left| \frac{a_{1,p}}{\gamma} \tilde{r} - |a_{1,p} r_{(1),2}| \right| \right|.
\] (3.50)

58
and by Eq. (3.49), on $\Omega(1) \cap \Omega_\delta$,

$$|1 + a_{1,p}^T (\hat{A}^{(n)} - zI_{n-1})^{-1} \eta| \geq \frac{1}{M'}$$

$$\gg \left| \frac{a_{1,p}}{\gamma} \right| + |a_{1,p}r(1,2)| = O(1)p^{-1/4}.$$

Plugging into Eq. (3.50) gives that

$$|(1 + \frac{a_{1,p}}{\gamma} \tilde{E} \tilde{m}(z))^{-1}| \cdot 1_{\Omega(1) \cap \Omega_\delta} \leq 2M' = O(1)M^2. \quad (3.51)$$

Recall that

$$r(1) = (\hat{#})_1 - a_{1,p} \left(1 - \frac{1}{1 + \frac{a_{1,p}}{\gamma} \tilde{E} \tilde{m}(z)}\right)$$

$$= |X_n|^2 a_{1,p} \left(1 - \frac{1}{1 + a_{1,p}^T (\hat{A}^{(n)} - zI_{n-1})^{-1} \eta} \right)$$

$$- a_{1,p} \left(1 - \frac{1}{1 + \frac{a_{1,p}}{\gamma} \tilde{E} \tilde{m}(z)}\right)$$

$$= (|X_n|^2 - 1)a_{1,p} \left(1 - \frac{1}{1 + a_{1,p}^T (\hat{A}^{(n)} - zI_{n-1})^{-1} \eta} \right)$$

$$+ a_{1,p} \left(\frac{1}{1 + \frac{a_{1,p}}{\gamma} \tilde{E} \tilde{m}(z)} - \frac{1}{1 + a_{1,p}^T (\hat{A}^{(n)} - zI_{n-1})^{-1} \eta} \right) .$$

By Eq. (3.49),

$$\mathbb{E} \left| (|X_n|^2 - 1)a_{1,p} \left(1 - \frac{1}{1 + a_{1,p}^T (\hat{A}^{(n)} - zI_{n-1})^{-1} \eta} \right) \right| \cdot 1_{\Omega(1) \cap \Omega_\delta}$$

$$\leq |a_{1,p}| (1 + O(1)M^2) \mathbb{E} \|X_n\|^2 - 1|$$

$$\leq O(1)M^2 \sqrt{\mathbb{E}(|X_n|^2 - 1)^2}$$

$$= O(1)M^2 p^{-1/2}.$$
where the last line follows from that $\mathbb{E}(|X_n|^2 - 1)^2 = \mathcal{O}(1)p^{-1}$ by Lemma 3.1. Meanwhile, by Eq. (3.45, 3.46, 3.48) and Eq. (3.49, 3.51), we have that

\[
\mathbb{E} \left| a_{1,p} \left( 1 + \frac{a_{1,p}}{\gamma} \mathbb{E}\tilde{m}(z) \right) - 1 + a_{1,p}\eta^T (\tilde{A}^{(n)} - zI_{n-1})^{-1} \eta \right| \cdot 1_{\Omega_{(1)} \cap \Omega_{\delta}}
\]

\[
\leq \mathbb{E} |a_{1,p}| (|\frac{a_{1,p}}{\gamma} \tilde{r}| + |a_{1,p}r_{(1),2}|) (1 + \frac{a_{1,p}}{\gamma} \mathbb{E}\tilde{m}(z))^{-1} | (1 + a_{1,p}\eta^T (\tilde{A}^{(n)} - zI_{n-1})^{-1} \eta)^{-1} \cdot 1_{\Omega_{(1)} \cap \Omega_{\delta}}
\]

\[
\leq (\mathcal{O}(1)M^2)^2 \mathbb{E} (|\frac{a_{1,p}}{\gamma} \tilde{r}| + |a_{1,p}r_{(1),2}|)
\]

\[
= \mathcal{O}(1)M^4 p^{-1/2}.
\]

Thus we have proved that

\[
\mathbb{E}|r_{(1)}| \cdot 1_{\Omega_{(1)} \cap \Omega_{\delta}} \leq \mathcal{O}(1)M^4 p^{-1/2}.
\]

Proof of Lemma 3.11.

\[
(\hat{#})_2 = f_{(2)}^T (\tilde{A}^{(n)} - zI_{n-1})^{-1} f_{(2)}
\]

\[
- f_{(2)}^T (\tilde{A}^{(n)} - zI_{n-1})^{-1} a_{1,p}\eta^T (\tilde{A}^{(n)} - zI_{n-1})^{-1} f_{(2)}
\]

\[
= \frac{\nu_{>1,p}}{\gamma} \mathbb{E}\tilde{m}(z) + \frac{\nu_{>1,p}}{\gamma} \tilde{r} + r_{(2),2} - r_{(2),3}
\]

(3.52)

where

\[
\nu_{>1,p} = \mathbb{E}(f_{(2)})_i^2 = \mathbb{E}f_{>1} (\xi_i)^2 = \nu_p - a_{1,p}^2
\]

and

\[
r_{(2),2} = f_{(2)}^T (\tilde{A}^{(n)} - zI_{n-1})^{-1} f_{(2)} - \frac{\nu_{>1,p}}{p} \text{Tr}(\tilde{A}^{(n)} - zI_{n-1})^{-1},
\]

\[
r_{(2),3} = f_{(2)}^T (\tilde{A}^{(n)} - zI_{n-1})^{-1} a_{1,p}\eta^T (\tilde{A}^{(n)} - zI_{n-1})^{-1} f_{(2)}
\]

\[
= a_{1,p}(\eta^T (\tilde{A}^{(n)} - zI_{n-1})^{-1} f_{(2)})r_{2,1}.
\]
For \( r_{(2),2} \), by a moment method argument similar to the first part in the proof of Lemma 3.9, we have

\[
\mathbb{E}|r_{(2),2}| \leq \mathcal{O}_L(1)p^{-1/2}.
\] (3.53)

To bound \( r_{(2),3} \), we restrict ourselves to \( \Omega_\delta \), where

\[
|f_2(\xi_m)| \leq \mathcal{O}_L(1)M^Lp^{-1/2}, \quad |\eta_i| \leq Mp^{-1/2}, \quad 1 \leq i \leq n-1,
\]

thus

\[
|a_{1,p}\eta^T(\hat{A}^{(n)} - zI_{n-1})^{-1}f_2(\xi_m)| \cdot 1_{\Omega_\delta} \leq \mathcal{O}(1)s((\hat{A}^{(n)} - zI_{n-1})^{-1})||\eta|| \cdot ||f_2||
\]

\[
\leq \frac{\mathcal{O}(1)}{v} \sqrt{\mathcal{O}(1)M^2\sqrt{\mathcal{O}_L(1)M^{2L}}} = \mathcal{O}_L(1)M^{L+1},
\]

and then

\[
\mathbb{E}|r_{(2),3}| \cdot 1_{\Omega_\delta} = \mathbb{E}|r_{2,1}| |a_{1,p}(\eta^T(\hat{A}^{(n)} - zI_{n-1})^{-1}f_2(\xi_m)| \cdot 1_{\Omega_\delta}
\]

\[
\leq \mathcal{O}_L(1)M^{L+1}\mathbb{E}|r_{2,1}|
\]

\[
\leq \mathcal{O}_L(1)M^{L+1}p^{-1/2}.
\] (3.54)

Recall that

\[
r_{(2)} = \gamma \bar{r} + r_{(2),2} - r_{(2),3},
\]

then the claim in Lemma 3.11 follows by Eq. (3.46, 3.53, 3.54).

\[\square\]

**Lemma 3.14.** \( r_{(1),2} \) defined as in Eq. (3.47),

\[
\mathbb{E}|r_{(1),2}| \leq \mathcal{O}(1)p^{-1/2}.
\]

**Remark 3.4.** The technique is similar to the moment bound method in [3, Chapter 3.3], where the main observation is that \( \hat{A}^{(n)} \) is independent of the vector \( \eta \).
Proof. Write \( r_{(1,2)} \) as \( r_2 \) for simplicity. Define \( (\tilde{A}^{(n)} - z I_{n-1})^{-1} \) as \( \tilde{B} \) which is Hermitian, we have

\[
E |r_2|^2 = \mathbb{E} \left[ \sum_{i=1}^{n-1} \left( \tilde{\eta}_i^2 - \frac{1}{p} \right) \tilde{B}_{ii} + \sum_{i_1 \neq i_2} \eta_{i_1} \eta_{i_2} \tilde{B}_{i_1i_2} \right]^2
\]

\[
= \mathbb{E} \sum_{i, i'} \left( \tilde{\eta}_i^2 - \frac{1}{p} \right) \left( \tilde{\eta}_{i'}^2 - \frac{1}{p} \right) \tilde{B}_{ii} \tilde{B}_{i'i'}
\]

\[
+ \mathbb{E} \sum_i \sum_{i_1 \neq i_2} \eta_{i_1} \eta_{i_2} \tilde{B}_{i_1i_2} \tilde{B}_{i'i_2}
\]

\[
= \mathbb{E} \sum_i \left( \tilde{\eta}_i^2 - \frac{1}{p} \right) \eta_{i_1} \eta_{i_2} \tilde{B}_{i_1i_2} \tilde{B}_{i'i_2}
\]

Due to that \( \eta \) is independent from \( \tilde{B} \), and the independence of \( \eta_{i_1} \) and \( \eta_{i_2} \) for \( i_1 \neq i_2 \), we take expectation over \( \eta_i \)'s conditioning on \( \tilde{B} \) and it gives

\[
E |r_2|^2 \leq \mathbb{E} \left( \sum_i \mathbb{E} \left( \eta_i^2 - \frac{1}{p} \right)^2 |\tilde{B}_{ii}|^2 + \sum_{i_1 \neq i_2} \frac{1}{p^2} |\tilde{B}_{i_1i_2}|^2 \right)
\]

\[
= \mathbb{E} \frac{2}{p^2} \left( \sum_i |\tilde{B}_{ii}|^2 + \sum_{i_1 \neq i_2} |\tilde{B}_{i_1i_2}|^2 \right)
\]

\[
= \mathbb{E} \frac{2}{p^2} \text{Tr}(\tilde{B}^T \tilde{B}).
\]

Observe that

\[
\text{Tr}(\tilde{B}^T \tilde{B}) = \sum_{i=1}^{n-1} \frac{1}{|\tilde{\lambda}_i - z|^2} \leq \sum_{i=1}^{n-1} \frac{1}{v^2} = \frac{n - 1}{v^2},
\]

where \( v = \Im(z) > 0 \) and \( \tilde{\lambda}_i \) are the eigenvalues of \( \tilde{A}^{(n)} \). Then

\[
\frac{2}{p^2} \text{Tr}(\tilde{B}^T \tilde{B}) \leq \frac{2}{v^2} \frac{n - 1}{v^2} \leq \frac{2}{v^2\gamma} \frac{1}{p},
\]

which means that

\[
E |r_2|^2 \leq \frac{O(1)}{p},
\]

so we have \( E |r_2| \leq \sqrt{E |r_2|^2} \leq O(1)p^{-1/2} \). \( \square \)
Proof of Lemma 3.12. Firstly, due to Eq. (2.7)

\[ |m_A(z) - m_{A(n)}(z)| \leq \frac{4}{v} \cdot n^{-1} \rightarrow 0, \]

where \( m_{A(n)}(z) = \frac{1}{n-1} \text{Tr}(A^{(n)} - zI_{n-1})^{-1} \). It suffices to show that

\[ \mathbb{E}|m_{A(n)}(z) - \tilde{m}(z)| \rightarrow 0. \]

Recall the definition of \( \tilde{m}(z) \) in Eq. (3.28), and denote \( \tilde{m}(z) \) as \( m_{\tilde{A}(n)}(z) \). By Eq. (3.23), and denote \( a_{1,p} by a \) for simplicity,

\[
\text{Tr}(A^{(n)} - zI_{n-1})^{-1} - \text{Tr}(\tilde{A}^{(n)} - zI_{n-1})^{-1} \\
= \text{Tr}(-(A^{(n)} - zI_{n-1})^{-1}(a\eta^T - aD) + \sqrt{p}WFW + T)(\tilde{A}^{(n)} - zI_{n-1})^{-1} \\
= \text{Tr}(-(A^{(n)} - zI_{n-1})^{-1}(\sqrt{p}WFW + T - aD)(\tilde{A}^{(n)} - zI_{n-1})^{-1}) \\
- a\eta^T(A^{(n)} - zI_{n-1})^{-1}(\tilde{A}^{(n)} - zI_{n-1})^{-1}\eta \\
:= (I) + (II).
\]

By \( |\text{Tr}(M)| \leq ns(M) \) for \( n \)-by-\( n \) symmetric matrix \( M \), and that \( s((M-z)^{-1}) \leq \frac{1}{v} \), we have that

\[
|(I)| = |\text{Tr}((A^{(n)} - zI_{n-1})^{-1}(\sqrt{p}WFW + T - aD\eta)(\tilde{A}^{(n)} - zI_{n-1})^{-1})| \\
\leq (n - 1)s(((A^{(n)} - zI_{n-1})^{-1})s(\sqrt{p}WFW + T - aD\eta)s(\tilde{A}^{(n)} - zI_{n-1})^{-1}) \\
\leq (n - 1)\frac{1}{v^2}s(\sqrt{p}WFW + T - aD\eta).
\]

By Eq. (3.41), we have that

\[ \mathbb{E}|(I)| \cdot 1_{\Omega_1} \leq O_L(1)M^2p^{3/4}. \]
Also,

\[
|\langle II \rangle| = |a \eta^T (A^{(n)} - zI_{n-1})^{-1} (\tilde{A}^{(n)} - zI_{n-1})^{-1} \eta| \\
\leq |a| s((A^{(n)} - zI_{n-1})^{-1}) s((\tilde{A}^{(n)} - zI_{n-1})^{-1}) ||\eta||^2 \\ 
\leq \mathcal{O}(1) ||\eta||^2,
\]

and on \( \Omega_\delta \) which is defined in Eq. (3.19), \( ||\eta||^2 \leq \delta^2 n = \mathcal{O}(1)M^2 \). Thus

\[
\mathbb{E}||\langle II \rangle|| \cdot 1_{\Omega_\delta} \leq \mathcal{O}(1)M^2.
\]

As a result, using the fact that \( |m_{A^{(n)}}(z)|, |m_{\tilde{A}^{(n)}}(z)| \leq \frac{1}{v} \) (Eq. (A.2)),

\[
\mathbb{E}|m_{A^{(n)}}(z) - m_{\tilde{A}^{(n)}}(z)| \leq \frac{2}{v} \Pr(\Omega_\delta) + \mathbb{E}|m_{A^{(n)}} - m_{\tilde{A}^{(n)}}| \cdot 1_{\Omega_\delta} \\ 
\leq \mathcal{O}(1)p^{-9} + \frac{1}{n-1}(\mathcal{O}_L(1)M^2p^{3/4} + \mathcal{O}(1)M^2) \\
= \mathcal{O}_L(1)M^2p^{-1/4}
\]
Chapter 4

Other Models of $X_i$’s

4.1 Uniform Distribution on the $p$-dimensional Unit Sphere

In this section, we extend the result in Thm. 3.4 to the situation where the random vectors $X_i$’s are i.i.d. uniformly distributed on a $p$-dimensional sphere.

For this model, the marginal distribution of the inner-product $\xi_{ij} = X_i^T X_j$ has probability density $Q_p'(u) = A_p (1 - u^2)^{(p-3)/2}$, where $A_p$ is a normalization constant. Let $\xi_p' \equiv \sqrt{p} \xi_{ij}$, whose probability density is $q_p'(x) = \frac{1}{\sqrt{p}} Q_p' \left( \frac{x}{\sqrt{p}} \right)$, and let $\mathcal{H}_p' = L^2(\mathbb{R}, q_p'(x)dx)$.

Parallel to Lemma 2.3, we have

**Lemma 4.1.**

$$\mathbb{E}(\xi_p')^k = \begin{cases} 
(k - 1)!! + \mathcal{O}_k(1)p^{-1}, & k \text{ even;} \\
0, & k \text{ odd.}
\end{cases}$$

*Proof.* The odd moments vanish since the distribution of $\xi_p'$ is symmetric with respect to 0. For even moments, let $k = 2m$. Let

$$\xi_{N,p} = \sqrt{p}X^TY$$

65
where $X$ and $Y$ are i.i.d $\mathcal{N}(0, p^{-1}I_p)$, and we have that $\xi_{N,p}$ and $\xi'_p \mid X \mid Y$ observe the same probability distribution. Notice that $\xi'_p \mid X \mid$ and $\mid Y$ are independent, so

$$
E\xi_{N,p}^{2m} = E|X|^{2m}E|Y|^{2m}E(\xi'_p)^{2m} = (E|X|^{2m})^2E(\xi'_p)^{2m}.
$$

The claim follows by Eq. (2.11) and that $E|X|^{2m} = 1 + O_m(1)p^{-1}$ (Lemma 3.1).

As a result, by Lemma 2.4, the orthonormal polynomials of $\mathcal{H}'_p$ are asymptotically consistent with the Hermite polynomials. If we expand $k(x; p)$ into the orthonormal polynomials of $\mathcal{H}'_p$, and require the conditions (C.Variance), (C.$p$-Uniform) and (C.$a_1$) accordingly, the result in Thm. 3.4 still holds. This is the following theorem:

**Theorem 4.2.** Suppose that $X_1, \cdots, X_n$ are i.i.d vectors distributed uniformly on $S^{p-1}$. Let $\{p_{l,p}(x), l = 0, 1, \cdots\}$ be the set of orthonormal polynomials in $\mathcal{H}'_p$, starting from $p_{0,p} = 1$, $p_{1,p} = x$. Draw the expansion

$$
k(x; p) = \sum_{l=0}^{\infty} a_{l,p} p_{l,p}(x)
$$

and require that the conditions (C.Variance), (C.$p$-Uniform) and (C.$a_1$) defined in Sec. 2.2.2 hold with respect to $\mathcal{H}'_p$ and $p_{l,p}(x)$. Then as $p, n \to \infty$ with $p/n \to \gamma$, the kernel matrix $A$ built from $X_i$’s as in Eq. (2.1) has the same (weak and almost sure) limiting spectral density as defined in Thm. 3.4.

One way of proving Thm. 4.2 is sketched as follows:

Condition on the draw of $X_n$, and without loss of generality let $X_n = (1, 0, \cdots, 0)^T$. Then

$$
X_i = (u_i, \sqrt{1 - u_i^2} Y_i^T)^T, \quad 1 \leq i \leq n - 1,
$$

66
where $u_i$’s are i.i.d distributed, and $\tilde{X}_i$’s are i.i.d. uniformly distributed on the unit sphere in $\mathbb{R}^{p-1}$ independently from $u_i$’s. As a result, let $\xi_{ij} = X_i^T X_j$ and $\tilde{\xi}_{ij} = \tilde{X}_i^T \tilde{X}_j$, then

$$
\xi_{ij} = u_i u_j + \sqrt{1 - u_i^2} \sqrt{1 - u_j^2} \tilde{\xi}_{ij}, \quad 1 \leq i, j \leq n - 1, i \neq j,
$$

which is different from before. However, on the large probability set

$$
\Omega_\delta = \{|u_i| \leq \delta, |\tilde{\xi}_{ij}| \leq \delta, 1 \leq i, j \leq n - 1, i \neq j, \delta = p^{-1/2} M, M = \sqrt{20 \ln p}\},
$$

it can be shown that

$$
\xi_{ij} = u_i u_j + \tilde{\xi}_{ij} + r_{ij}, \quad |r_{ij}| \leq \delta^3.
$$

Thus, the Taylor expansion can be carried out in the same way, where the contribution of the extra $r_{ij}$ term is put into $T_{ij}$ and the bound Eq. (3.38) remains true.

We still need the mean spectral norm bound to show that Eq. (3.39) holds, and to use the bound given by the 4th moment (Prop. 3.3), it suffices to establish the bound in Lemma 3.2. Notice that Gegenbauer polynomials [1] are orthogonal in the space $L^2([-1, 1], Q_p(u) du)$. Gegenbauer polynomials are related to the $p$-spherical harmonics $\{\phi_j, j \in J\}$, which form an orthonormal basis of $L^2(S^{p-1}, dP)$. $J = \bigcup_{l=0}^{\infty} J_l$, and $\{\phi_j(X), j \in J_l\}$ are $p$-spherical harmonics of degree $l$, which are homogeneous harmonic polynomials restricted to the surface of the unit sphere. The Gegenbauer polynomial of degree $l$ as a function of $X^T Y$, $X, Y \in S^{p-1}$, up to a multiplicative constant, equals

$$
Z_{l,X}(Y) = \sum_{j \in J_l} \phi_j(X) \phi_j(Y)
$$

which is named “the $l$-degree zonal harmonic function with axis $X$”. We thus define $G_{t,p}(\xi)$ to be

$$
G_{t,p}(X^T Y) = \sum_{j \in J_l} \phi_j(X) \phi_j(Y). \quad (4.1)
$$
Notice that $G_{1,p}(X^TY) = pX^TY$, and by convention $G_{0,p} = 1$. $G_{l,p}(\xi)$ is a polynomial of degree $l$ for all $l$, and

$$\int_{S^{p-1}} \int_{S^{p-1}} G_{l,p}(X^TY)G_{k,p}(X^TY)dP(X)dP(Y) = \delta_{l,k}|J_l|.$$ 

$|J_l|$ is the number of $p$-spherical harmonics of degree $l$, $|J_1| = p$, and for $l \geq 2$

$$|J_l| = \binom{p + l - 1}{l} - \binom{p + l - 3}{l - 2} = \left(\frac{1}{l!} + \frac{O_l(1)}{p}\right)p^l.$$ 

Thus, the orthonormal polynomials $p_{l,p}(x)$ of the space $\mathcal{H}_p^l$ can be written as

$$p_{l,p}(x) = \frac{1}{\sqrt{|J_l|}}G_{l,p}\left(\frac{x}{\sqrt{p}}\right).$$

By Eq. (4.1), we have

$$\int_{S^{p-1}} G_{l,p}(X_1^TX_2)G_{l,p}(X_2^TX_3)dP(X_2) = G_{l,p}(X_1^TX_3), \quad X_1, X_2, X_3 \in S^{p-1},$$

which gives that (define $\xi_{ij} = X_i^TX_j$)

$$\mathbb{E}[p_{l,p}(\sqrt{p}\xi_{12})p_{l,p}(\sqrt{p}\xi_{23})|X_1, X_3] = \frac{1}{\sqrt{|J_l|}}p_{l,p}(\sqrt{p}\xi_{13}).$$

As a result, $\mathbb{E}p_{l,p}(\sqrt{p}\xi_{12})p_{l,p}(\sqrt{p}\xi_{23})p_{l,p}(\sqrt{p}\xi_{34})p_{l,p}(\sqrt{p}\xi_{41})$ is bounded by

$$\frac{1}{|J_l|} = O(1)p^{-l},$$

which is stronger than the estimate in Lemma 3.2.

To continue to show the result in Thm. 3.4, the mechanism in Sec. 3.3 applies to what follows in almost the same way.
Another way is by comparing $X_i$ to the standard Gaussian vectors, which is the topic of the next section.

### 4.2 Comparison With the Gaussian Vectors

We introduce the following comparison theorem:

**Theorem 4.3.** Suppose that $X_1, \ldots, X_n$ are i.i.d. random vectors in $\mathbb{R}^p$, and $Z_1, \ldots, Z_n$ are i.i.d Gaussian vectors $Z_i \sim \mathcal{N}(0, p^{-1}I_p)$. Let

$$
\xi_p \equiv \sqrt{p}X_1^T X_2, \quad \xi_{N,p} \equiv \sqrt{p}Z_1^T Z_2,
$$

and $\mathcal{H}_p$ be the $L^2$ space on $\mathbb{R}$ associated with the probability density of $\xi_p$. Let $\{p_l(x), l = 0, 1, \cdots\}$ be the set of orthonormal polynomials in $\mathcal{H}_p$. If for fixed positive integer $l$ any $\epsilon > 0$, as $p \to \infty$,

$$
\mathbb{E}(\xi_p^k - \xi_{N,p}^k)^2 < \epsilon^2, \quad 1 \leq k \leq l,
$$

and $k(x;p)$ satisfies (C.Variance), (C.$p$-Uniform) and (C.$a_1$) defined in Sec. 2.2.2 with respect to $\mathcal{H}_p$ and $p_l(x)$, then at the limit $p,n \to \infty$ with $p/n \to \gamma$, the kernel matrix $A$ built from $X_i$'s as in Eq. (2.1) has the same (weak and almost sure) limiting spectral density as defined in Thm. 3.4.

The following examples satisfy the conditions in the theorem:

1. (unit vector) $X_i \sim \mathcal{U}(S^{p-1})$. Let $X_i = Z_i/|Z_i|$ for $i = 1, \cdots, n$. Because $\xi_{N,p} = \sqrt{p}|Z_1||Z_2|X_1^T X_2$, and $|Z_1|, |Z_2|$ are independent from $X_1, X_2$, we have
that $\xi_{N,p} = |Z_1| |Z_2| \xi_p$ and $|Z_1|, |Z_2|$ are independent from $\xi_p$. Thus

$$E(\xi_p^k - \xi_{N,p}^k)^2 = E(|Z_1|^k |Z_2|^k - 1)^2 E \xi_{N,p}^{2k}$$

$$= E(|Z_1|^k |Z_2|^k - 1)^2 \cdot (E \xi^{2k} + O_k(1)p^{-1})$$

$$= O_k(1) E(|Z_1|^2k |Z_2|^{2k} - 2|Z_1|^k |Z_2|^k + 1)$$

which is of the order $O_k(p^{-1})$ by Lemma 3.1, and this verifies Eq. (4.2).

2. ("slightly" correlated Gaussian vector) $X_i \sim N(0, p^{-1} \Sigma_p)$, where $\Sigma_p$ satisfies that 1) $\xi_p$ is $p$-uniformly sub-gaussian, i.e.

$$Pr[|\xi_p| > t] \leq C_1 e^{-C_2 t^2}, \quad t > 0$$

for some positive constants $C_1$ and $C_2$, and 2)

$$\frac{1}{p}||\Sigma_p - I||^2_{Fro} = O(p^{-\alpha}) \quad (4.3)$$

for some positive constant $\alpha$.

For this case, let $Z_i = \Sigma_p^{-1/2} X_i$, and we firstly have that

$$E(\xi_p - \xi_{N,p})^2 = E(\sqrt{p} Z_1^T (\Sigma_p - I_p) Z_2)^2 = \frac{1}{p}||\Sigma_p - I||^2_{Fro} = O(p^{-\alpha}). \quad (4.4)$$

Since both $\xi_p$ and $\xi_{N,p}$ are sub-gaussian, for any $\epsilon > 0$, we can choose $M = C_k \sqrt{\ln p}$ where $C_k$ is a positive number depending on $k$, so that

$$\Omega_{k,p} := \{|\xi_p| < M, |\xi_{N,p}| < M\}, \quad Pr[\Omega_{k,p}] = O(1)p^{-4k}\epsilon^4.$$
Then as $p \to \infty$,

$$
\mathbb{E}(\xi^k_p - \xi^k_{N,p})^2 \mathbf{1}_{\Omega_{k,p}} \leq (kM^{k-1})^2 \mathbb{E}(\xi_p - \xi_{N,p})^2
$$

$$
= (kM^{k-1})^2 \mathcal{O}(p^{-a})
$$

$$
= k^2 o(1) \leq \mathcal{O}_k(1) \epsilon^2,
$$

(4.5)

and meanwhile, $|\xi_p| \leq \sqrt{p}|Z_1||Z_2|||\Sigma_p||_{\text{Fro}}, ||\Sigma_p||_{\text{Fro}} \leq \mathcal{O}(p^{1/2})$,

$$
\mathbb{E}(\xi^k_p - \xi^k_{N,p})^2 \mathbf{1}_{\Omega_{k,p}} \leq (2p)^{2k} \mathbb{E}|Z_1|^{2k}|Z_2|^{2k} \mathbf{1}_{\Omega_{k,p}^c}
$$

$$
\leq (2p)^{2k} \sqrt{\mathbb{E}|Z_1|^{4k}|Z_2|^{4k} \mathbb{P}[\Omega_{k,p}^c]}^k
$$

$$
\leq \mathcal{O}_k(1) \epsilon^2.
$$

(4.6)

Eq. (4.5, 4.6) together verify Eq. (4.2).

One example of the “slightly” correlated Gaussian vector is the “finite-rank spiking” model, i.e. where $\sigma^2_j$ for $1 \leq j \leq d$ are positive constants bigger than one and $d > 0$ is some fixed number, and $\sigma^2_j = 1$ for $d + 1 \leq j \leq p$. For this model the bound in Eq. (4.3) is $\mathcal{O}(p^{-1})$. This model will be further discussed in Sec. 4.3.

At the same time, numerical experiments suggest that the result in Thm. 3.4 extends to a large class of distribution of $X_i$, which goes beyond Thm. 4.3. For example, the “Bernoulli” model, where $X_i$’s are uniformly sampled from the $2^p$ vertices of the hypercube $\{-p^{-1/2}, p^{-1/2}\}^p$, and see Fig. 4.1 for supportive numerical results.

The conjectured validity of Thm. 3.4 for the “Bernoulli” model has been positively answered by Do and Vu (see Thm. 3 in a recent publication [11]). In [11], Thm. 3.4 was extended to $X_i$’s whose $p$ entries are independent copies of a random variable $x/\sqrt{p}$ and $x$ has finite $k$-th moments for all $k$. The method used by Do and Vu is
Figure 4.1: Random kernel matrix with $X_i$ uniformly sampled from the $2^p$ vertices of the hypercube $\{-p^{-1/2}, p^{-1/2}\}^p$. (Left) $p = 2 \times 10^2$, $n = 4 \times 10^3$, $k(x) = \text{Sign}(x)$. (Right) $p = 2 \times 10^3$, $n = 4 \times 10^3$, $k(x) = |x| - \sqrt{2/\pi}$. The blue-boundary bars are the empirical eigenvalue histograms, and the red broken-line curves are the theoretical prediction of the eigenvalue densities by Thm. 3.4.

a Lindeberg-swapping approach to control the deviation of the Stieltjes transform before and after replacing the $np$ entries of $X_1, \cdots, X_n$ with Gaussian ones.

Notice that Thm. 4.3 allows dependency among the entries of $X_i$’s, which is not covered in Do and Vu’s result. A complete understanding of the “universality” of $X_i$’s is still open.

Proof of Thm. 4.3. As in proving Thm. 3.4, it suffices to consider the truncated kernel function

$$k_L(x; p) = \sum_{l=1}^{L} a_{l,p} p_{l,p}(x),$$

where $L$ is a finite integer, $\sum_{l=1}^{L} a_{l,p}^2 \sim O(1)$. Eq. (4.2) means that for any $\epsilon > 0$, as $p \to \infty$,

$$|\mathbb{E}_{X_{N,p}}^{k} - \mathbb{E}_{X,p}^{k}| < \epsilon, \quad 1 \leq k \leq L,$$

and by Lemma 2.3, for large enough $p$,

$$|\mathbb{E}_{X_{N,p}}^{k} - \mathbb{E}_{X_{N}}^{k}| < 2\epsilon, \quad 1 \leq k \leq L.$$
Then by Lemma 2.4,

\[ p_{l,p}(x) = h_l(x) + \sum_{j=0}^{l} (r_{l,p})_j x^j, \quad \max_{0 \leq j \leq l} |(r_{l,p})_j| < O_l(1)\epsilon. \]  

(4.7)

Thus,

\[ k_L(\xi_p; p) - k_L(\xi_{N,p}; p) = \sum_{l=1}^{L} a_{l,p} \sum_{j=0}^{l} (c_{l,p})_j (\xi^j - \xi^j_{N,p}), \]

where \(|a_{l,p}| \sim O(1)\) and \(|(c_{l,p})_j| \sim O_l(1)\) (by Eq. (4.7)), and by Cauchy-Schwarz and Eq. (4.2),

\[ \mathbb{E}(k_L(\xi_p; p) - k_L(\xi_{N,p}; p))^2 = O_L(1)\epsilon^2. \]

This means that we can replace \(X_i^T X_j\) by \(Z_i^T Z_j\) in building the kernel matrix using \(k_L\) without changing the limiting spectrum by Lemma 3.6.

Meanwhile, let \(\mathcal{H}_{N,p}\) be the \(L^2\) space on \(\mathbb{R}\) associated with the probability density of \(\xi_{N,p}\), and \(\{P_{l,p}(x), l = 0, 1, \cdots\}\) be the set of orthonormal polynomials in \(\mathcal{H}_{N,p}\). Applying Lemma 2.4 to both \(p_{l,p}(x)\) and \(P_{l,p}(x)\) gives that for large enough \(p\),

\[ p_{l,p}(x) - P_{l,p}(x) = \sum_{j=0}^{l} (r'_{l,p})_j x^j, \quad \max_{0 \leq j \leq l} |(r'_{l,p})_j| < O_l(1)\epsilon. \]

Define

\[ k_{N,L}(x; p) := \sum_{l=1}^{L} a_{l,p} P_{l,p}(x), \]

and by Cauchy-Schwarz we have that

\[ \mathbb{E}(k_L(\xi_{N,p}; p) - k_{N,L}(\xi_{N,p}; p))^2 = O_L(1)\epsilon^2. \]

By Lemma 3.6 again, we can replace \(k_L(x; p)\) by \(k_{N,L}(x; p)\) in building the kernel matrix from \(Z_i\)’s. We end up with the kernel matrix \(A\), \(A_{ij} = \frac{1}{\sqrt{p}} k_{N,L}(\sqrt{p}Z_i^T Z_j; p), i \neq j\), to which Thm. 3.4 applies.
4.3 Spiking Model

In this section we discuss the situation when $X_i$ admits a decomposition of “signal + noise”, where the “noise” is in the $p$-dimensional space and the “signal” lies on a low-dimensional manifold which does not change with the dimension $p$.

The problem is related to that of the finite-rank perturbation of large random matrices. In the latter, the objects of interest are the eigenvalue and eigenvectors of the matrix $A + P$, where $A$ is the “substrate matrix” having certain limiting spectral density, and $P$ is the “perturbation matrix” which is assumed to be of low-rank. When the substrate matrix is a large sample covariance matrix, the limiting law of the spiked model when $p, n \to \infty$ with $p/n \to \gamma$ was firstly obtained in [4]. It has been proved in [4] that there is a “phase transition” of the limiting position of the largest eigenvalue (when $P$ is of rank-1), from “at the right-end of the M.P. density” to “outside the bulk of the M.P. density”, when the magnitude of the perturbation increases over certain thresholding level. Associated with the phase transition of the spiked eigenvalue is that of the correlation between the “spiked” eigenspace and the eigenspace of the $P$. It is named B-B-P phase transition after the three authors of [4]. Extension of the B-B-P phase transition to spiking models where the substrate matrix $A$ and the perturbation matrix $P$ are independent and orthogonally invariant has been studied in [6]. The implication of B-B-P phase transition to high-dimensional PCA has been discussed in [19].

For non-linear kernel matrix, the “signal + noise” model has been studied in [14], where the kernel function satisfies certain regular conditions. We consider kernel functions as in Sec. 2.2, which can be non-smooth and can depend on the dimension $p$. We provide conjectures concerning the limiting “spiking” law as well as numerical results.
4.3.1 Rank-1 ($Z_2$) Spiking

Let $u \in \mathbb{R}^p$ be a unit vector, $y_i, i = 1, \cdots, n$ are arbitrarily $\pm 1$, and

$$X_i = y_i u + \sigma N_i, \quad N_i \sim \mathcal{N}(0, I_p).$$

In $X_i$, the “signal” part is $y_i u$, which is a unit vector up to a sign, and the “noise” part is $\sigma N_i$. Due to the isotropy of the Gaussian distribution, one may assume that $u = (1, 0, \cdots, 0)^T$. The “signal-to-noise-ratio” is defined as

$$R = \frac{1}{\sigma^2}.$$ We define the kernel matrix $A$ as

$$A_{ij} = \begin{cases} f(X_i^T X_j), & i \neq j, \\ 0, & i = j, \end{cases}$$

where, assuming $\sigma > 0$,

$$f(\xi) = \frac{1}{\sqrt{p}} k \left( \frac{1}{\sigma^2 \sqrt{p}} \xi; p \right),$$

and $k(x; p)$ is as in Sec. 2.2, e.g. $k(x) = \text{Sign}(x)$. Recall that

$$a_{1,p} \to a, \quad \nu_p \to \nu \geq a^2.$$  

Conjecture 4.4. Assume that $\sigma^2$ is fixed to be a constant. $k(x; p)$ satisfies the conditions in 2.2. Let $v_1$ be the top eigenvector of $A$, and $y_0 = \frac{1}{\sqrt{n}} (y_1, \cdots, y_n)^T$. As $p, n \to \infty$, and $p/n \to \gamma,$
1. When $a > 0$, then the model “spikes” when $R$ is bigger than a thresholding value.

\[
|v_1^Ty_0|^2 \rightarrow \begin{cases} 
\frac{R}{R+1} \left[ 1 - \frac{\gamma}{R^2} - \frac{(\nu-a^2)\gamma}{a^2(R+1)^2} \right], & R > R_c \\
0, & R < R_c 
\end{cases} \quad (4.8)
\]

where $R_c$ depends on $\gamma$ and $\frac{a^2}{\nu}$, and the function is explicitly defined by

\[
\frac{1}{\gamma} = \frac{1}{R_c^2} + \frac{\nu-a^2}{a^2} \frac{1}{(R_c+1)^2}. \quad (4.9)
\]

It is well-defined since the right-hand-side as a function of $R_c$ is monotonically decreasing on $(0, \infty)$.

2. When $a = 0$, then the model never spikes, i.e. $|v_1^Ty_0|^2 \to 0$.

In Fig. 4.2 we compare the value of $|v_1^Ty_0|^2$ to the conjectured limit, together with the correlation between $y_0$ and the normalized vector $\hat{v}_i = \text{Sign}((v_1)_i)/\sqrt{n}$.

A few remarks:

1. When $\nu = a^2$ (the linear kernel) we have

\[
R_c = \sqrt{\gamma},
\]

which is the “spiking threshold” for the sample covariance matrix [4].

2. When $\nu - a^2$ increases, $R_c$ increases, so the spiking will be postponed to higher SNR. The values of $R_c$ with varying $a$ and $\gamma$ are shown in Fig. 4.3.

3. The spiking threshold

\[
R_c = \text{Const}(\gamma, \frac{a^2}{\nu}),
\]

and we have the following bound

\[
\sqrt{\gamma} \leq R_c \leq \sqrt{\gamma} \sqrt{\frac{\nu}{a^2}}, \quad \gamma = \frac{p}{n},
\]
Figure 4.2: Rank-1 spiking model, where $Y_i = \pm u$. $p = 200$, $n = 400$, and $R$ varies. (Left) $k(x) = x$. (Right) $k(x) = \text{Sign}(x)$.

Figure 4.3: $R_c$ as a function of $a$ and $\gamma$, $0 < a \leq \nu = 1$.

which means that

$$\sqrt{\frac{a^2}{\nu} \cdot \frac{n}{p}} \leq (\sigma^2)_c \leq \sqrt{\frac{n}{p}}.$$

Some heuristics which lead to the conjecture:
Notice that if there is no signal, then $A$ has the same model as the random kernel matrices in our main theorem. In the spiking model,

$$f(X_1^T X_2 : p) = \frac{1}{\sqrt{p}} k \left( \frac{N_1^T N_2}{\sqrt{p}} + \frac{y_1 y_2}{\sigma^2 \sqrt{p}} + \text{"cross term"} \right),$$

and the quantity

$$\frac{y_1 y_2}{\sigma^2 \sqrt{p}} \sim O(p^{-1/2}),$$

while

$$\frac{N_1^T N_2}{\sqrt{p}} \rightarrow N(0,1) \sim O(1).$$

Thus $f(X_1^T X_2; p)$ can be thought as

$$\frac{1}{\sqrt{p}} k \left( \frac{N_1^T N_2}{\sqrt{p}} ; p \right) + \text{"perturbation"}$$

and it turns out that the “linear part” in the “perturbation”, namely the rank-1 matrix

$$\left\{ a_{1,i}, \frac{y_i y_j}{\sigma^2 \sqrt{p}} \right\}_{i,j=1}^n$$

is the “first” to spike as the SNR increases. The computation of the consistency $|v_1 y_0|^2$ and the threshold value $R_c$ follows a procedure similar to that in [6].

### 4.3.2 $S^{d-1}$ Spiking

The model of rank-1 spiking extends to the rank-$d$ case. Let $Y_i$ be i.i.d. uniformly drawn from $S^{d-1}$ and corrupted by high-dimensional noise, i.e.

$$X_i = Y_i + \sigma N_i.$$

Construct the kernel matrix $A$ as before. Assume that the kernel has a non-zero linear part, i.e. $a^2 > 0$. Otherwise the model never spikes.
Figure 4.4: Similar plots to Fig. 4.2. \( d = 2 \), \( Y_i \sim \mathcal{U}(S^1) \). Consistency type 1 and 2 are defined in Eq. (4.10, 4.11). (Left) \( k(x) = x \). (Right) \( k(x) = \text{Sign}(x) \).

We define the first-type consistency of top-\( d \) eigenspace as

\[
\text{consist}_1 = \frac{\text{Tr}(V_d V_d^T Y_0 Y_0^T)}{d},
\]

where the columns of \( V_d \) are the top-\( d \) eigenvectors of \( A \), \( Y_0 \) is made from \( Y \) by normalizing each column to be of unit length, and \( Y \) is an \( n \)-by-\( d \) matrix, each row of which consisting of the coordinates of the point \( Y_i \) on \( S^{d-1} \). The second-type consistency is defined as

\[
\text{consist}_2 = \frac{\text{Tr}(\bar{V}_d \bar{V}_d^T Y Y^T)}{|V_d V_d^T|_{Fro} |Y Y^T|_{Fro}}.
\]

where \( \bar{V}_d \) is made from \( V_d \) by normalizing each row to be of unit length.

We define the SNR to be

\[
R = (1/d)/\sigma^2.
\]

The phase transition point \( R_c \) is the same as defined in Eq. (4.9) and the limit of \( \text{consist}_1 \) is the same as in Eq. (4.8). In Fig. 4.4 we compare the value of two types of consistency to the conjectured limit, where \( d = 2 \).
4.3.3 Spiking of the $l$-degree Polynomial Kernel

Conjecture 4.5. The model is same as in Sec. 4.3.1. Suppose that $k(x;p) = h_l(x)$, and

$$\sigma^2 = s_l p^{-1/2+1/(2l)}.$$  

Then the “spiking” condition is that

$$s_l < (s_l)_c$$

where $(s_l)_c$ is some constant depending on $\gamma$.

Some heuristic which leads to the conjecture:

We take rank-1 ($Z_2$) model as an example.

$$f(X_1^T X_2;p) = \frac{1}{\sqrt{p}} h_l(y_1y_2) + \frac{N_1^T N_2}{\sqrt{p}}$$

$$\sim \frac{1}{\gamma s_l} \frac{1}{n} (y_1y_2)^l + \frac{1}{\sqrt{p}} h_l(\frac{N_1^T N_2}{\sqrt{p}}),$$

ignoring the “cross terms”. Then the model is equivalent to $A_Y = \frac{1}{n} \{(Y_i^T Y_j)^l\}_{i,j=1}^n$ which is rank-1 matrix, “spiking” the semicircle spectrum associated with $A_N = \gamma s_l^l A_{N,h_l}$, where

$$A_{N,h_l} = \left\{ \frac{1}{\sqrt{p}} h_l(\frac{N_i^T N_j}{\sqrt{p}}) \right\}_{i,j=1}^n, \quad (A_{N,h_l})_{ii} = 0.$$  

4.3.4 Consistency of Non-linear Kernel on a Bounded Manifold

Let $Y_i \in \mathcal{M} \subset \mathbb{R}^p$, $\mathcal{M}$ is a bounded manifold and without loss of generality assume that

$$\max\{|x|, x \in M\} < 1.$$  

80
We are given the noise-corrupted data
\[ X_i = Y_i + \sigma N_i, \quad N_i \sim \mathcal{N}(0, I_p), \]
and we are interested in the “information kernel matrix”
\[ A_Y = \{ f(Y_i^T Y_j; p) \}_{i,j=1}^n, \quad (A_Y)_{ii} = 0. \]
Define
\[ A = \{ f(X_i^T X_j; p) \}_{i,j=1}^n, \quad A_{ii} = 0, \]
and we need the condition when \( A \) is a consistent estimator of \( A_Y \).

Based on Conjecture 4.5, we set \( l \) to be infinity and have the following conjecture

**Conjecture 4.6.** Suppose that \( \sigma^2 = sp^{-1/2} \), and \( s \sim O(1) \). Let
\[ k(x; p) := \sqrt{p} f(x \sqrt{p} \sigma^2; p) \]
and satisfies the conditions in Sec. 2.2. There exists some constant \( s_c = s_c(\gamma, a, \nu) \) so that if \( s < s_c \), as \( p, n \to \infty \), \( p/n \to \gamma \), almost surely
\[ s(A_Y - A) \to 0, \]
where \( s(M) \) stands for the spectral norm of matrix \( M \).
Chapter 5

Conclusion

5.1 Conclusion

This part of the thesis studies the limiting spectrum of random inner-product kernel matrices in the limit $p, n \to \infty$ with $p/n \to \gamma$, where the kernel function admits the model in Chapter 2. The convergence of the limiting spectrum is proved firstly for Gaussian vectors (Chapter 3, Thm. 3.4), and then extended to some other models of $X_i$’s (Chapter 4, Thm. 4.3). Our approach is based on analyzing the Stieltjes transform, employing an orthogonal polynomial expansion of the kernel function. The results have implication for the “signal+noise” model which is closely related to many high-dimensional data analysis applications.

5.2 Future Directions

There are many potential directions to continue the study of this thesis, e.g. the “universality” problem of other models of $X_i$’s which is discussed in Sec. 4.2. We list a few more here:

Firstly, while this part of the thesis mainly focuses on the limiting spectral density, another question of practical importance concerns the extreme (largest/smallest)
eigenvalue. For the null model, the asymptotic statistics of the extreme eigenvalue is unknown, and the conjecture that the spectral norm is asymptotically $O(1)$ needs to be proved. For the “spiking model”, our conjectures in Sec. 4.3 call for a rigorous study.

Secondly, the limiting spectrum of large Euclidean kernel matrices has not been understood. Due to the relation

$$|X_i - X_j|^2 = |X_i|^2 + |X_j|^2 - 2X_i^T X_j,$$

for the model where $|X_i| \equiv 1$, the problem is reduced to the inner-product model. However, for the model where $X_i \sim \mathcal{N}(0, p^{-1}I_p)$, the law of the limiting spectrum is unknown.

Finally, we hope to extend the study to kernels that are of more general forms. In the second part of this thesis we study the “symmetry detection problem” and the “structural reconstruction problem” in electron microscopy. The symmetry detection problem involves estimating the spectrum of an auto-correlation kernel under spherical Bessel basis, and in the reconstruction problem a synchronization operator over the group of $SO(3)$ is built based on “common-line” detection between the pairs of projections. Thus in both problems the model of the kernel function goes beyond the form of inner-product or Euclidean kernels. As another example, a complex-valued kernel has been used in [29] for a dataset of tomographic images. The modulus of the kernel function corresponds to the similarity of the images when they are optimally aligned, while the phase of the kernel is the optimal in-plane alignment angle. Similar kernels have also been used for dimensionality reduction [28] and sensor network localization [10]. In many senses these applications motivate the study of the random matrix model in this part of the thesis.
Part II

Cryo-EM With Symmetry
Chapter 6

Introduction

Cryo-electron microscopy (cryo-EM) is a technique to image biological macromolecules using an electron microscope, where two-dimensional projections of a three-dimensional molecule are taken at unknown random orientations. To determine the molecular structure from cryo-EM images is of scientific importance in many fields, such as chemistry and biology. However, due to experimental reasons cryo-EM images are of low contrast and have very high noise, which makes the structure reconstruction problem challenging.

Mathematically, the molecular structure is described by a scalar function $V(\vec{r})$, $\vec{r} = (x, y, z)^T \in [-1, 1]^3$, which evaluates the electric potential in space. $V$ stands for “volume”. When the molecule admits a symmetry group $G$, which is a subgroup of $SO(3)$, the function $V(\vec{r})$ satisfies

$$V(\vec{r}) = V(g\vec{r}), \quad \forall g \in G. \quad (6.1)$$

Given $N$ cryo-EM images $P_1, \cdots, P_N$ ($P$ stands for “projection”), each of them is formed by projecting the volume $V(\vec{r})$ in some unknown direction $R_i \in SO(3)$, i.e.
the pixel intensities in $P_i(x,y)$ are given by

$$P_i(x,y) = \int_{-\infty}^{\infty} V(R_i\vec{r})dz, \quad \vec{r} = (x, y, z)^T.$$  \hspace{1cm} (6.2) 

In practice, each $P_i$ is further corrupted by noise. The structure reconstruction problem is to reconstruct the volume $V(\vec{r})$ from the projections $P_1, \cdots, P_N$ where the projecting directions $R_1, \cdots, R_N$ are unknown.

Under the assumption that the molecule has no symmetry, an \textit{ab-initio} reconstruction algorithm has been introduced in [27] and there is a series of studies developing theories and algorithms for various cryo-EM-related tasks [16, 17, 29, 25, 37]. When the molecule admits non-trivial symmetries, the problem is different from what has been considered in [27, 25] and needs a new solution. Another problem is to detect the symmetry group of the molecule from the set of cryo-EM images.

This part of the thesis focuses on the symmetry detection problem and the reconstruction problem for symmetric molecules. In Chapter 7 we propose to detect the symmetry group by computing certain auto-correlation functions of the projections. In Chapter 8, we generalize the ideas in [27] and [25] and propose two reconstruction algorithms. We focus on the theoretical solutions and test the methods on simulated data sets.
Chapter 7

Symmetry Detection

7.1 The Symmetry Detection Problem

The problem is to identify from finitely many noisy projections the symmetry group of the molecule as defined in Eq. (6.1), which is a subgroup of $SO(3)$. According to a classical theorem of Klein [21], the symmetry group $G$, when not being the trivial group $\{I_3\}$, is isomorphic to one of the following:

- $C_k$, $k \geq 2$: Cyclic group consisting of $k$ elements;
- $D_k$, $k \geq 2$: Dihedral group consisting of $2k$ elements;
- $T_{12}$: Tetrahedral group consisting of 12 elements, which is the symmetry group of the standard tetrahedron;
- $O_{24}$: Octahedral group consisting of 24 elements, which is the symmetry group of the standard cube;
- $I_{60}$: Icosahedral group consisting of 60 elements, which is the symmetry group of the standard icosahedron.

Our approach is to compute the “$l$-th order auto-correlation function” of the projections, $C_l(r_1,r_2)$, whose “rank” encodes the type of the symmetry group that
the volume has. The usage of the auto-correlation function was suggested in [20] for structure reconstruction, and we extend the idea to the symmetry setting. In estimating $C_l(r_1, r_2)$ we compute the “steerable” covariance matrix of the (Fourier transformed) projections in the Fourier-Bessel basis, which is similar to the treatment in [37]. We introduce the theory in Sec. 7.2 and Sec. 7.3, and we explain the algorithm and the numerical results using simulated projections in Sec. 7.4.

### 7.2 The Auto-correlation Function

We introduce the auto-correlation function based on the theory in [20]:

Let $\hat{V}(r, \theta, \phi)$ be the three-dimensional Fourier transform of the density function $V(x, y, z)$. $\hat{V}$ has the spherical harmonic expansion

$$\hat{V}(r, \theta, \phi) = \sum_{l,m} A_{lm}(r) Y_{l,m}(\theta, \phi).$$

According to the Fourier Slice Theorem, the Fourier-transformed projection at orientation $\omega \in SO(3)$ is

$$\hat{P}^\omega(r, \phi) = \sum_{l,m} A_{lm}(r) \sum_{m'} D_{m,m'}^l(\omega) Y_{l,m'}(\theta, \phi) \bigg|_{\theta = \frac{\pi}{2}},$$

where $D_{m,m'}^l(\omega)$ is the Wigner-D matrix of order $l$ for rotation $\omega \in SO(3)$. The auto-correlation function is defined as

$$C(r_1, r_2, \phi_1, \phi_2) := \int_{SO(3)} \left\{ \hat{P}^\omega(r_1, \phi_1) \cdot \hat{P}^\omega(r_2, \phi_2) \right\} d\omega$$

$$= \sum_{l=0}^{\infty} \sum_{m=-l}^{l} A_{lm}(r_1) A_{lm}(r_2) \frac{1}{4\pi} P_l(\cos(\phi_1 - \phi_2)), \quad (7.1)$$
Table 7.1: The rank of $C_l(r_1, r_2)$ (i.e. the dimension of the range of the integral operator with kernel $C_l(r_1, r_2)$) for different groups, according to [21]. The first two members of cyclic groups and dihedral groups are shown.

<table>
<thead>
<tr>
<th>Group</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>No symmetry</td>
<td>0 1 3 5 7 9 11 13 15 17 19 21</td>
</tr>
<tr>
<td>$C_2$</td>
<td>1 1 2 3 3 5 5 7 7 9 9 11</td>
</tr>
<tr>
<td>$C_4$</td>
<td>1 1 1 1 3 3 3 3 5 5 5 5</td>
</tr>
<tr>
<td>$D_2$</td>
<td>1 0 2 1 3 2 4 3 5 4 6</td>
</tr>
<tr>
<td>$D_4$</td>
<td>1 0 1 0 2 1 2 1 3 2 3</td>
</tr>
<tr>
<td>$T_{12}$</td>
<td>1 0 0 1 1 0 2 1 1 2 2</td>
</tr>
<tr>
<td>$O_{24}$</td>
<td>1 0 0 0 1 0 1 0 1 1 1</td>
</tr>
<tr>
<td>$I_{60}$</td>
<td>1 0 0 0 0 0 1 0 0 0 1</td>
</tr>
</tbody>
</table>

where $d\omega$ is the Haar measure and $P_l(x)$ is the Legendre polynomial of degree $l$. Notice that $C(r_1, r_2, \phi_1, \phi_2)$ only depends on $\psi := \phi_1 - \phi_2$. Using the orthogonality of Legendre polynomials, we define

$$C_l(r_1, r_2) := 2\pi(2l + 1) \int_0^\pi C(r_1, r_2, \psi)P_l(\cos \psi)\sin \psi d\psi = \sum_{m=-l}^l A_{lm}(r_1)\overline{A_{lm}(r_2)}. \quad (7.2)$$

We call $C_l(r_1, r_2)$ the “auto-correlation function of $l$-th order”, and it is the kernel of an integral operator in the functional space $L^2([0, 1], r^2 dr)$. The range of the integral operator is of dimension at most $2l + 1$, and we denote it by the “rank” of $C_l(r_1, r_2)$. When the volume has symmetry, some of $A_{lm}(r)$ may vanish so that the rank of $C_l(r_1, r_2)$ degenerates. According to the classical representation theory [21], the rank of $C_l(r_1, r_2)$ encodes the type of symmetry that the volume has. Table 7.1 lists the rank of $C_l(r_1, r_2)$ for different groups.

Finally, we expand $C_l(r_1, r_2)$ using the spherical Bessel basis $\{j_{ls}(r), s = 1, 2, \cdots \}$, which forms a set of ortho-normal basis of the functional space $L^2([0, 1], r^2 dr)$. Specifically, let

$$A_{lms} := \int_0^1 A_{lm}(r)j_{ls}(r)r^2 dr, \quad (7.3)$$
where
\[
    j_{ls}(r) = \frac{\sqrt{2}}{|j_{l+1}(R_{l,s})|} j_l(R_{l,s} r), \quad 0 < r < 1, \ s = 1, 2, \cdots
\]
where \( R_{l,s} \) is the \( s \)-th positive root of the equation \( j_l(x) = 0 \). The auto-correlation function of \( l \)-th order in the basis of \( \{ j_{ls}(r), s = 1, 2, \cdots \} \) is given by

\[
(C_l)_{s_1,s_2} = \sum_{m=-l}^{l} A_{lm,s_1} \overline{A_{lm,s_2}}. \tag{7.4}
\]

### 7.3 Approximating \( C_l(r_1, r_2) \) From Projections

Given \( N \) projections \( P_1, \cdots, P_N \) we can compute an empirical \( l \)-th order auto-correlation function which approximates \( C_l(r_1, r_2) \).

Let the Fourier transformed projections be \( \hat{P}_i, i = 1, \cdots, N \). Define the (rotation and mirror-reflection invariant) empirical auto-correlation function as

\[
C_l^{(proj)}(r_1, r_2, \phi_1, \phi_2) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} \int_0^{2\pi} \frac{dt}{2\pi} \left\{ \hat{P}_i(r_1, \phi_1 + t) \cdot \overline{\hat{P}_i(r_2, \phi_2 + t)} + \hat{P}_i^r(r_1, \phi_1 + t) \cdot \overline{\hat{P}_i^r(r_2, \phi_2 + t)} \right\}, \tag{7.5}
\]

where \( \hat{P}_i^r \) is the mirror-reflected projection, i.e. \( \hat{P}_i^r(r, \phi) = \hat{P}(r, 2\pi - \phi) \). We expand \( \hat{P}_i \) using the Fourier-Bessel basis

\[
\hat{P}_i(r, \phi) = \sum_{k,q} \frac{1}{\sqrt{2\pi}} (a_i)_{kq} J_{kq}(r) e^{ik\phi}, \tag{7.6}
\]

where
\[
J_{kq}(r) = \frac{\sqrt{2}}{|J_{k+1}(R_{k,q})|} J_k(R_{k,q} r), \quad 0 < r < 1, \ q = 1, 2, \cdots
\]
and $R_{k,q}$ is the $q$-th positive root of the equation $J_k(x) = 0$. Plugging Eq. (7.6) into Eq. (7.5) gives that

$$C^{(proj)}(r_1, r_2, \phi_1, \phi_2) = \sum_k \sum_{q_1, q_2} \left( \frac{1}{N} \sum_i (a_i)_k (a_i)_{kq_1} \overline{(a_i)_{kq_2}} \right) \cdot J_{kq_1}(r_1) J_{kq_2}(r_2) \cdot \frac{1}{2\pi} \cos(k(\phi_1 - \phi_2)).$$

(7.7)

Following [37], we define

$$(C^{FB})_{k_1q_1, k_2q_2} = \frac{1}{N} \sum_i (a_i)_k (a_i)_{kq_1} \overline{(a_i)_{kq_2}} \delta_{k_1, k_2},$$

(7.8)

and call $C^{FB}$ the “steerable” sample covariance matrix in the Fourier-Bessel basis, which has a block diagonal structure [37]. The expansion coefficients $(a_i)_{kq}$ can be computed from $\hat{P}_l$ by solving a least square problem, similar to the approach in [37]. The difference is that [37] works with the projections in real space, and we work with the Fourier-transformed projections.

By Eq. (7.8) we rewrite Eq. (7.7) as

$$C^{(proj)}(r_1, r_2, \phi_1, \phi_2) = \sum_k \sum_{q_1, q_2} (C^{FB})_{kq_1, kq_2} J_{kq_1}(r_1) J_{kq_2}(r_2) \cdot \frac{1}{2\pi} \cos(k(\phi_1 - \phi_2)).$$

(7.9)

Next, parallel to Eq. (7.2), we define the empirical $l$-th order auto-correlation function as

$$C_l^{(proj)}(r_1, r_2) := 2\pi (2l + 1) \int_0^\pi C^{(proj)}(r_1, r_2, \psi) P_l(\cos \psi) \sin \psi d\psi$$

$$= \sum_{k, q_1, q_2} (C^{FB})_{kq_1, kq_2} \cdot J_{kq_1}(r_1) J_{kq_2}(r_2) \alpha_{kl},$$

(7.10)

where

$$\alpha_{kl} = \int_0^\pi \cos(k\psi) \cdot (2l + 1) P_l(\cos \psi) \sin \psi d\psi.$$
The values of \( \alpha_{kl} \) can be calculated analytically by definition. Notice that when \( l > 0 \), \( \alpha_{0l} = 0 \). Thus

\[
C_l^{(proj)}(r_1, r_2) = \sum_{k=1}^{\infty} \sum_{q_1, q_2} \alpha_{kl} ((C^{FB})_{q_1, q_2} + (C^{FB})_{-q_1, -q_2}) J_{kq_1}(r_1) J_{kq_2}(r_2),
\]

\[
= \sum_{k=1}^{\infty} \alpha_{kl} 2\Re(C^{FB})_{kq_1, kq_2} J_{kq_1}(r_1) J_{kq_2}(r_2). \tag{7.12}
\]

The second equality makes use of the fact that \( a_{-kq} = a_{kq} \), which follows since \( \hat{P}(r, \phi) = \hat{P}(r, \phi + \pi) \) (\( \hat{P} \) is the Fourier transform of a real-valued image) and \( J_{-k}(x) = (-1)^k J_k(x) \).

Same as in Eq. (7.4), we expand \( C_l^{(proj)}(r_1, r_2) \) using the spherical Bessel series \( \{j_{ls}(r), s = 1, 2, \cdots \} \) and obtain

\[
(C_l^{(proj)})_{s_1, s_2} = \sum_{k=1}^{\infty} \alpha_{kl} \sum_{k, q_1, q_2} 2\Re(C^{FB})_{kq_1, kq_2} \beta_{kq_1, ls_1} \beta_{kq_2, ls_2}, \tag{7.13}
\]

where

\[
\beta_{kq, ls} := \int_0^1 J_{kq}(r) j_{ls}(r)r^2dr.
\]

### 7.4 Algorithm and Numerical Experiments

In practice, we truncate the Fourier-Bessel expansion up to \( |k| \leq K \) and \( q \leq q(k) \) and the spherical Bessel expansion up to \( l \leq L \) and \( s \leq s(l) \) where \( q(k) \) and \( s(l) \) are integers to be specified. Specifically, suppose that the projection \( P_i(x, y), (x, y) \in [-1, 1]^2 \) is sampled on a Cartesian grid of size \( 2N_x \times 2N_x \), we choose \( K = N_x \) and \( q(k) \) to be the maximal integer \( q \) so that \( R_{k,q} \leq \pi N_x \) by Nyquist criterion, same as in [37]. The choice of \( s(l) \) is by that \( R_{l,s} \leq \pi N_x \). The truncation of \( L \) depends on the assumption on the decay of the magnitude of the radial functions \( A_{lm}(r) \), and we set \( L \) to be some integer large enough.
After applying the truncation, both $C_l$ as in Eq. (7.4) and $C_l^{(proj)}$ as in Eq. (7.13) are $s(l)$-by-$s(l)$ matrices, and we can write Eq. (7.13) in matrix form as

$$C_l^{(proj)} = \sum_{k=1}^{K} \alpha_{kl} B_{k,l} \cdot 2\Re C^{FB}(k) \cdot B_{k,l}^T,$$  

(7.14)

where $C^{FB}(k)$ is the $k$-th diagonal block of $C^{FB}$ which is of size $q(k)$-by-$q(k)$, and $B_{k,l}$ is a matrix of size $l(s)$-by-$q(k)$ defined as

$$(B_{k,l})_{sq} = \beta_{kq,ls}, \quad 1 \leq s \leq s(l), 1 \leq q \leq q(k).$$  

(7.15)

We compute the numbers $\beta_{kq,ls}$ explicitly by numerical integration.

The symmetry detection algorithm is summarized as follows:

Given $N$ projections $P_i, i = 1, \cdots, N$, which are sampled on a $2N_x$-by-$2N_x$ Cartesian grid,

Step 1. Choose some positive integer $L$, and set $K = N_x$. For $1 \leq k \leq K$, $1 \leq l \leq L$, compute the matrices $B_{k,l}$ as in Eq. (7.15) and the values $\alpha_{kl}$ as in Eq. (7.11).

Step 2. Compute the Fourier transformed projections $\hat{P}_i, i = 1, \cdots, N$ by two-dimensional FFT. $\hat{P}_i$ is sampled on the same Cartesian grid as $P_i$.

Step 3. Compute the steerable Fourier-Bessel sample covariance matrix $C^{FB}$ defined in Eq. (7.8), which has $K$ diagonal blocks $C^{FB}(k), 1 \leq k \leq K$, and $C^{FB}(k)$ is of size $q(k)$-by-$q(k)$. This is similar to the approach in [37] but we compute the covariance matrix for $\hat{P}_i$ instead of $P_i$.

Step 4. For $1 \leq l \leq L$, compute the empirical $l$-th order auto-correlation function $C_l^{(proj)}$ as in Eq. (7.14), which is a of size $s(l)$-by-$s(l)$.

Step 5. Compute the eigenvalues of the matrices $C_l^{(proj)}$, and estimate the rank for $l = 1, 2, \cdots$. Compare the rank to Table 7.1 to decide on the symmetry group.
The above algorithm is implemented using MATLAB and tested on simulated projections. In generating the projections, we use two Gaussian mixture volumes, which admits no symmetry and $C_2$ symmetry respectively and are constructed to be compactly-supported in the unit ball. We sample each volume on a three-dimensional grid of size $65 \times 65 \times 65$, and compute the Fourier-transformed volume $\hat{V}(\vec{r})$, from which the expansion coefficients $A_{lms}$ as in Eq. (7.3) can be computed by solving a least square problem. We use the values of $A_{lms}$ to compute the auto-correlation function of order $l$ according to Eq. (7.4), and call the resulting $s(l)$-by-$s(l)$ matrix $C_l^{(vol)}$. We use $C_l^{(vol)}$, $l = 1, 2, \ldots$ as the underlying truth and compare it with the matrices $C_l^{(proj)}$ which are computed from projections.

We generate $N = 1000$ projections of size $65 \times 65$, and test both clean and noisy projections. For the latter we use additive Gaussian white noise and set SNR = $1, 1/4, 1/16$ respectively, where “SNR” stands for the the signal-to-noise-ratio and is defined as in [25].

For the non-symmetric volume, the simulated noisy projections and the eigenvalues of $C_l^{(vol)}$, $l = 1, 2, 3$ are shown in Fig. 7.1. When $l = 1$, $C_1^{(vol)}$ has three non-zero eigenvalues, which indicates that the volume has no symmetry. The eigenvalues of $C_l^{(proj)}$ are shown in Fig. 7.2, where the large eigenvalues match consistently with those of $C_l^{(vol)}$ and the non-zero eigenvalues of small magnitude are due to the noise.

For the $C_2$-symmetric volume, we have the same plots in Fig. 7.3 and Fig. 7.4. Notice that $C_1^{(vol)}$ has one non-zero eigenvalue as the theory predicts, and $C_2^{(vol)}$ is expected to have three non-zero eigenvalues, but the third one is so small that it can hardly be seen from the plots. According to Table 7.1, when $C_1$ is rank-1 and the rank of $C_2$ is more than 1 it indicates the symmetry group to be $C_2$. In Fig. 7.4 the large eigenvalues of $C_l^{(proj)}$ are consistent with those of $C_l^{(vol)}$ except for that when SNR = $1/16$ (the bottom row) the second largest eigenvalue can hardly be
Figure 7.1: (Upper) Simulated projections SNR = 1/16. (Lower) Bar plot of the eigenvalues of $C_l^{(vol)}$. No symmetry.

distinguished from the continues spectrum produced by the noise. This means that more projections are needed to better estimate the rank of $C_l(r_1, r_2)$. 
Figure 7.2: Bar plot of the eigenvalues of $C_l^{(proj)}$ which are computed from $N = 1000$ clean and noisy projections with SNR = 1, 1/4, 1/16 (from upper to lower) respectively, $l = 1, 2, 3$. No symmetry.
Figure 7.3: (Upper) Simulated projections SNR = 1/16. (Lower) Bar plot of the eigenvalues of $C_l^{(vol)}$. $C_2$ symmetry.
Figure 7.4: Bar plot of the eigenvalues of $C_l^{(proj)}$ which are computed from $N = 1000$ clean and noisy projections with SNR = 1, 1/4, 1/16 (from upper to lower) respectively, $l = 1, 2, 3$. $C_2$ symmetry.
Chapter 8

Structure Reconstruction

8.1 The Reconstruction Problem

Suppose that the molecule admits some known symmetry group $G$. By Eq. (6.1, 6.2), for each orientation $R_i \in SO(3)$, the orientation $gR_i$ for any $g \in G$ will produce the same projection. The reconstruction problem is to estimate the “orbit” of $R_i$

$$\{gR_i, g \in G\}$$

for each $i$, and after that the volume can be reconstructed by aligning the projections $P_i$ according to $R_i$ and using the three-dimensional pseudo-polar Fourier transform, which is the same method as in [25].

In this chapter we focus on the cyclic group and assume that the rotation axis is the $z$-axis. For example, when $G = C_2$,

$$C_2 = \{I_3, g\}, \quad I_3 = \text{diag}\{1,1,1\}, \quad g = \text{diag}\{-1,-1,1\}.$$ 

We firstly introduce the concepts of “common-line” and “top-view”, and then the pipeline of two reconstruction algorithms for $G = C_2$ with numerical tests.
8.2 Common-line and Top-view

8.2.1 Common-line Between Two Projections

Assuming that the molecule has no symmetry, the concept of “common-line” has been introduced in [27], and common-line detection is the foundation of the reconstruction algorithms in [27, 25]. We generalize the concept to the situation of symmetry molecule. In Sec. 8.3 we will extend the synchronization approach in [25] based on common-line detection.

When \( G = C_2 \), between (Fourier transformed) projections \( \hat{P}_j \) and \( \hat{P}_k \) there are two pairs of “mutual common-lines”, denoted as \( \{c_{jk}, c_{kj}\} \) and \( \{c_{\text{g}jk}, c_{\text{g}kj}\} \). \( c_{jk} = (x_{jk}, y_{jk}, 0)^T \) and \( c_{\text{g}jk} \) are radial lines on the plain of \( \hat{P}_j \), and \( c_{kj} \) and \( c_{\text{g}kj} \) lines on that of \( \hat{P}_k \). They are the “common” lines shared by projections \( \hat{P}_j \) and \( \hat{P}_k \) in the sense that

\[
\hat{P}_j(\xi c_{jk}) = \hat{V}(\xi R_j c_{jk}) = \hat{V}(\xi R_k c_{kj}) = \hat{P}_k(\xi c_{kj}), \quad \xi \in \mathbb{R},
\]

and the same with \( \{c_{\text{g}jk}, c_{\text{g}kj}\} \). They satisfy the equations

\[
R_j c_{jk} = R_k c_{kj}, \quad R_j c_{\text{g}jk} = g R_k c_{\text{g}kj}.
\]

Furthermore, let

\[
U := R_j^T R_k, \quad U^g := R_j^T g R_k
\]

which are 3-by-3 matrices, then \( c_{jk} \) and \( c_{kj} \) have the expressions that

\[
c_{jk} = \frac{1}{\sqrt{1 - U(3,3)^2}} \begin{bmatrix} -U(2,3) \\ U(1,3) \end{bmatrix}, \quad c_{kj} = \frac{1}{\sqrt{1 - U(3,3)^2}} \begin{bmatrix} U(3,2) \\ -U(3,1) \end{bmatrix}
\]

and similar for \( c_{\text{g}jk} \) and for \( c_{\text{g}kj} \) in terms of the entries of \( U^g \).
Generally, when the molecule has symmetry group \( G \), there are \(|G|\) pairs of mutual common-lines between two projections, where \(|G|\) is order of the group.

Let

\[
J = \text{diag}\{1, 1, -1\}.
\]

Notice that in the above formula changing \( \{R_j, R_k\} \) into \( \{JR_jJ, JR_kJ\} \) results in \( \{c_{jk}, c_{kj}\} \) changing into \( \{-c_{jk}, -c_{kj}\} \) (\( \{-c_{jk}^g, -c_{kj}^g\} \)), which again satisfy Eq. (8.1). Strictly speaking, the common-line pair is well-defined up to a sign, namely \( \{\pm c_{jk}, \pm c_{kj}\} \), and we use \( \{c_{jk}, c_{kj}\} \) as a short-hand notation. The equivalence of mutual common-lines with/without \( J \)-conjugacy means that using common-lines one can at most determine a pair of rotations \( \{R_j, R_k\} \) up to a simultaneous \( J \)-conjugacy, i.e. either \( \{R_j, R_k\} \) or \( \{JR_jJ, JR_kJ\} \). As a result, the common-line-based method recovers all the rotations \( R_1, \cdots, R_N \) up to a global \( J \)-conjugacy, apart from an individual \( g \)-ambiguity (for each rotation \( R_i \), it is aimed at recovering either \( R_i \) or \( gR_i \)) and a global \( x, y \)-in-plain rotation \( h \) (which commutes with \( g \)). The global \( J \)-conjugacy is known as the “handedness” of the system.

### 8.2.2 Top-view

When \( G \) is a cyclic group, we introduce the concept of “top-view”, based on which we will propose a reconstruction algorithm in Sec. 8.4.

“Top-views” are the projections associated with \( R \in SO(3) \) which is an \( x, y \)-in-plain rotation, that is, \( R(3, 3) = 1 \). It is a family of projections consisting of the “standard top-view” (associated with \( I_3 \)) and its rotated versions. The mirrored images of the “top-views” are the “bottom-views”, which are rotated versions of the “standard bottom-view” (associated with \( \text{diag}\{-1, 1, -1\} \)). When \( G = C_k \), both the “top-view” and the “bottom-view” projections are \( C_k \) symmetric. Taking \( G = C_2 \) as
an example, the top-view projection $P$ satisfies the equation

$$P(x, y) = P(-x, -y) := P^g(x, y), \quad (8.3)$$

because for each line $c$ on a “top/bottom-view” $P$,

$$P(\xi c) = V(\xi Rc) = V(\xi g Rc) = V(-\xi Rc) = P(-\xi c), \quad \xi \in \mathbb{R}. \quad (8.4)$$

The top-view projection can be estimated from “nearly-top/bottom-views” selected from the projections, and the method is the following: we define “$\theta_{\text{th}}$-top-views” (“$\theta_{\text{th}}$-bottom-views”) as projections associated with $R$ whose the third column $R(:,3)$ has inner-product with $(0,0,1)^T$ ($(0,0,-1)^T$) larger than $1 - \cos(\theta_{\text{th}})$, where $\theta_{\text{th}}$ is a small angle, say 10 degrees. Due to continuity, we expect these projections to be close to the top/bottom-views and thus are nearly $C_2$ symmetric, that is

$$||P - P^g||^2 = \int |P(x, y) - P(-x, -y)|^2 dxdy$$

to be small, and this $\ell^2$ distance distinguishes the “$\theta_{\text{th}}$-top-views” from the others. Suppose that the rotations are uniformly distributed over $SO(3)$ with respect to the Haar measure, then the third columns $R_k(:,3)$ are uniformly distributed on the unit sphere $S^2$. Then from a large number of projections one expects the “$\theta_{\text{th}}$-top/bottom-views” occupying a fraction of $A(\theta_{\text{th}})/(2\pi)$ of all the projections, where $A(\theta_{\text{th}})$ is the surface area of a spherical cap of unit sphere $S^2$ with height $1 - \cos(\theta_{\text{th}})$. The method labels $N_{\text{top}} := A(\theta_{\text{th}})/(2\pi) \cdot N'_0$ many “$\theta_{\text{th}}$-top/bottom-views” as those which have the smallest value of $||P - P^g||$.

After selecting the “$\theta_{\text{th}}$-top/bottom-views” as approximate top/bottom-views, the top-view projection can be estimated (up to an in-plain rotation) by aligning the in-plain rotation angles of the different top-views and averaging them, in which process
the mirror symmetry needs to be properly handled. To be more specific, the projections are firstly marked “mirrored-or-not” by a \( \mathbb{Z}_2 \) synchronization, and then the rotation angles are aligned using a weighted angular synchronization. This selecting and aligning method is shown to have good performance when the projections are corrupted by noise.

### 8.2.3 Common-line Between a Projection and the Top-view

Taking \( G = C_2 \) as an example. Suppose that \( \hat{P}_{\text{top}} \) is the (Fourier transformed) top-view, and \( \hat{P}_k \) is a non-top-view projection. With out loss of generality, assuming that \( \hat{P}_{\text{top}} \) is associated with \( R_0 = I_3 \). With \( R_j = R_0, U = R_k \) and \( U^g = gR_k \), Eq. (8.2) gives that

\[
\begin{align*}
c_{0k} &= \frac{1}{\sqrt{1 - R_k(3,3)^2}} \begin{bmatrix} -R_k(2,3) \\ R_k(1,3) \end{bmatrix}, \\
c_{k0} &= \frac{1}{\sqrt{1 - R_k(3,3)^2}} \begin{bmatrix} R_k(3,2) \\ -R_k(3,1) \end{bmatrix},
\end{align*}
\]

and

\[
c_{k0} = c_{k0}^g, \quad c_{0k} = -c_{0k}^g.
\]

This means that there is one pair of common-lines between \( \hat{P}_k \) and \( \hat{P}_{\text{top}} \) instead of two. Since \( \hat{P}_{\text{top}}(\xi c_{0k}) \) as a function of \( \xi \) is even (Eq. (8.3)), \( c_{0k} \) and \( -c_{0k} \) on \( \hat{P}_{\text{top}} \) are the same line, and so are \( c_{k0} \) and \( -c_{k0} \) on \( \hat{P}_k \).

Actually, the two lines \( \{c_{k0}, -c_{k0}\} \) satisfy

\[
R_k c_{k0} = -gR_k c_{k0},
\]
and we call them the pair of “self common-lines” of projection \( \hat{P}_k \). When the molecule has a general symmetry group \( G \), there are \(|G| - 1\) pairs of self common-lines on a projection.

When \( c_{0k} \) and \( c_{k0} \) are known, by Eq. (8.4), determining the rotation \( R_k \) is reduced to determining \( R_k(3,3) \). Notice that in this case both \( \{c_{0k}, c_{k0}\} \) and \( \{c_{0k}, -c_{k0}\} \) can be the common-line pair, corresponding to the ambiguity of recovering \( R_k \) or \( gR_k \). At the same time, similar to the case of mutual common-lines, \( \{c_{0k}, c_{k0}\} \) and \( \{-c_{0k}, -c_{k0}\} \) are both the “with-\( \hat{P}_{top} \)-common lines” for \( \hat{P}_k \), which corresponds to a choice of either \( R_k \) or \( JR_kJ \). The method to estimate \( R_k(3,3) \) will be explained in Sec. 8.4.1.

8.3 The Synchronization Algorithm

8.3.1 The Algorithm for \( C_2 \) Symmetry

For \( G = C_2 \), the reconstruction algorithm is as follows: given \( N \) projections \( P_1, \ldots, P_N \),

Step 1. Compute the 2 pairs of mutual common-lines between \( \hat{P}_i \) and \( \hat{P}_j \), \( \{c_{ij}, c_{ji}\} \) and \( \{c^g_{ij}, c^g_{ji}\} \), for \( 1 \leq i, j \leq N \), which satisfy

\[
R_i c_{ij} = R_j c_{ji}, \quad R_i c^g_{ij} = g R_j c^g_{ji}.
\]

Step 2. From the mutual common-line data estimate the relative rotations of \( R^T_i R_j \) and \( R^T_i g R_j \), up to \( J \)-conjugacy, which is explained as below:

Let \( T_R \) be the plane in \( \mathbb{R}^3 \) normal to the viewing direction of \( R \). \( R^T_i R_j \) (\( R^T_i g R_j \)) is determined when \( \{c_{ij} \in T_{R_i}, c_{ji} \in T_{R_j}\} \) (\( \{c^g_{ij} \in T_{gR_i}, c^g_{ji} \in T_{gR_j}\} \)) and the angle \( \phi_{ij} \) (\( \phi^g_{ij} \)) between the planes \( T_{R_i} \) and \( T_{R_j} \) (\( T_{gR_i} \) and \( T_{gR_j} \)) are known. The angle \( \phi_{ij} \) (\( \phi^g_{ij} \)) is estimated using a voting procedure involving \( \{c_{ik}, c_{ki}, c^g_{ik}, c^g_{ki}\} \).
\{c_{jk}, c_{kj}, c_{gjk}, c_{gkj}\}, i.e. the common-lines of \(\hat{P}_i\) and \(\hat{P}_j\) with a third projection \(\hat{P}_k\), for all \(k \neq i, j\). Due to the fact that \(R_i\) and \(R_j\) (\(gR_j\)) produce the same pairs of common-lines \(\{\pm c_{ij}, \pm c_{ji}\}\) as \(JR_iJ\) and \(JR_jJ\) (\(JgR_jJ\)), where \(J = \text{diag}\{1, 1, -1\}\), the estimated rotation \(R_{ij} (R_{ij}^g)\) can be \(JR_{ij}I\) (\(JR_{ij}^gJ\)). The \(R_{ij}\) and \(R_{ij}^g\) can be enforced to be of a simultaneous \(J\)-conjugacy by requiring that \(R_{ij} + R_{ij}^g\) is close to rank-1.

In this step, the two relative rotations \(R_{ij} := R_i^T R_j\) and \(R_{ij}^g := R_i^T gR_j\) are estimated, up to a simultaneous \(J\)-conjugacy, without knowing which is which.

Step 3. Eliminating the \(J\)-conjugacy through all pairs \((i, j)\), i.e. make \({R_{ij}, R_{ij}^g}\) for all \((i, j)\) be up to a simultaneous \(J\)-conjugacy. The consistency within a triple \({(i, j), (j, k), (k, i)}\} can be enforced by checking the relation

\[
(v_i v_j^T)(v_j v_k^T) = v_i v_k^T,
\]

\[
(v_j v_k^T)(v_k v_i^T) = v_j v_i^T,
\]

\[
(v_k v_i^T)(v_i v_j^T) = v_k v_j^T,
\]

where \(v_i^T\) is the 3rd row of \(R_i\), i.e. the viewing direction, and

\[
v_i v_j^T = \frac{1}{2}(R_{ij} + R_{ij}^g).
\]

After enforcing a simultaneous \(J\)-conjugacy on each triple, a synchronization step is used to propagate the consistency though all the pairs. The synchronization is on a graph consisting of pairs \((i, j)\) as nodes, where edges are within triangles of triple \({(i, j), (j, k), (k, i)}\}, i.e. each triple is a clique.

Step 4. From the relative rotations \(R_{ij} = R_i^TR_j\) and \(R_{ij}^g = R_i^T gR_j\) estimate the rotations \(R_i\)'s up to a simultaneous group action \(gh\) in front, i.e. the estimated \(R_i = ghR_i,\)

\(i = 1, \cdots, N\), where \(g \in G = C_2,\) and \(h \in \mathcal{H} = SO(2)\) is an in-plane rotation.
in the $x$-$y$-plane. The global $h$ is due to the fact that the group $H$ commutes with $G$. This step consists of the following two procedures:

Firstly, estimate the 3rd row of $R_i$’s, by building a $3N$-by-$3N$ matrix $R_{3\text{rd row}}$ where the $(i,j)$-th subblock is

$$(R_{3\text{rd row}})_{ij} = v_i v_j^T = \frac{1}{2}(R_{ij} + R_{ij}^T).$$

When the estimation of $R_{ij}$ and $R_{ij}^T$ is exact, the matrix $R_{3\text{rd row}}$ is rank-1, and that eigenvector reveals the 3rd rows of $R_i$’s. The error in evaluating commonlines, both due to the discretization into $N_\theta$ radial lines (which gives small error) and corruption of the image by noise (which gives outliers), will render the matrix not exactly rank-1. When the top eigenvalue is far from the rest of the spectrum, the eigen-space associated with the top eigenvalue is taken as a good approximation of that for the exact case. Specifically, the 3rd row of $R_i$ is approximated by the $i$-th 3-by-1 block of the 1st eigenvector of $R_{3\text{rd row}}$, normalized to be of unit-length.

Secondly, estimate the 1st two rows of $R_i$’s. For $R_i \in SO(3)$, using the 3rd-row of $R_i$’s which are estimated, the only left degree of freedom is an in-plane rotation where the rotation axis is the 3rd row. To be specific, after specifying $R^{(3)}(v_i)$ as some rotation in $SO(3)$ which has $v_i$ as its 3rd row, the rotation $R_i$ can be written as

$$R_i = R^{(12)}(\alpha_i)R^{(3)}(v_i),$$

where

$$R^{(12)}(\alpha_i) = \begin{bmatrix}
\cos(\alpha) & -\sin(\alpha) \\
\sin(\alpha) & \cos(\alpha)
\end{bmatrix}.$$
The goal is to determine $\alpha_i$ up to the ambiguity of $\alpha_i + \pi$, which is due to the $C_2$ symmetry. Notice fact that

$$R^T_i R_j = R_3(v_i)^T R^{(12)}_i(-\alpha_i + \alpha_j) R_3(v_j),$$
$$R^T_i g R_j = R_3(v_i)^T R^{(12)}_i(-\alpha_i + \alpha_j + \pi) R_3(v_j),$$

then both $(R_3(v_i) R_{ij} R_3(v_j)^T)^2$ and $(R_3(v_i) R_{ij}^2 R_3(v_j)^T)^2$ provide an estimation of

$$R^{(12)}_i(2(-\alpha_i + \alpha_j)).$$

An angular synchronization procedure [26] is used to obtain $R^{(12)}(2\alpha_i)$’s which determines $\alpha_i$’s up to $\pi$.

Step 5. From the estimated rotations $\hat{R}_i$ and the associated denoised projections, reconstruct the volume $V$.

### 8.3.2 Numerical Experiments

The algorithm is implemented in Matlab and was tested on simulated projections. We create an “artificial” $C_2$-symmetric volume $V(\vec{r})$, which is produced from an experiment volume data $V_0(\vec{r})$ by

$$V(\vec{r}) = \frac{1}{|G|} \sum_{g \in G} V(g\vec{r}) = \frac{1}{2} \left( V(x, y, z) + V(-x, -y, z) \right).$$

Projections are generated in $N = 200$ random orientations drawn independently uniformly from $SO(3)$. Each projection has $129 \times 129$ pixels. Additive Gaussian white noise is used and the signal-to-noise-ratio (SNR) is defined as in [25]. The clean and simulated noisy projections are as in Fig. 8.1.

From $N = 200$ clean projections, the eigenvalues of the two synchronization procedures in Step 4. are shown in Fig. 8.2.
Figure 8.1: Four clean (upper) and noisy (lower) projections with SNR = 1/2.

Figure 8.2: The largest 20 eigenvalues of the synchronization matrix $R_{\text{3rd row}}$ (left) and those of the angular synchronization matrix to recover the first two rows of $R_i$’s (right), computed from $N = 200$ clean projections.

The mean squared error (MSE) is defined as

$$\text{MSE}(R_i, 1 \leq i \leq N) = \min_{O \in SO(3)} \min_{g_i \in G = C_2} \frac{1}{N} \|R_i^{(\text{true})} - O g_i R_i\|_{\text{Fro}}^2,$$  

(8.5)

where $\|\cdot\|_{\text{Fro}}$ stands for the Frobenius norm of matrices. In this test, MSE = $1.8747e-06$. The mean error in angles is $3.528098e-02$ (degrees) with standard deviation $2.348523e-02$ (degrees). The error in angles and the comparison of true/reconstructed volumes can be see in Fig. 8.3.
Figure 8.3: The error in the angles (left) and the comparison of the true and reconstructed volume (right) computed from $N = 200$ clean projections.

Figure 8.4: The same plot as in Fig. 8.2, computed from $N = 200$ noisy projections with SNR = 1/2.

From $N = 200$ noisy projections, SNR = 1/2, the same plots are produced in Fig. 8.4, 8.5. In this test, MSE = 8.4822e − 04, and the mean error in angles is 7.954328e-01 (degrees) with standard deviation 4.549464e-01 (degrees).
Figure 8.5: The error in the angles (left) and the comparison of the true and reconstructed volume (right) computed from $N = 200$ noisy projections with SNR = 1/2, where the reconstructed volume is Gaussian-filtered with width 0.8.

8.4 The Top-view-based Algorithm

8.4.1 The Algorithm for $C_2$ Symmetry

We start from $N'_0$ noisy projections,

Step 1. Use the method described in Sec. 8.2.2 to select $N_{\text{top}}$ number of “$\theta_{\text{th}}$-top-views” from the de-noised projections, and from them to estimate the “top-view” projection $P_{\text{top}}$.

After selecting the “$\theta_{\text{th}}$-top-views”, we aim at recovering the rotations associated with the rest non-top-view projections. We firstly process $N_0 \leq N'_0 - N_{\text{top}}$ ones, and then select $N$-many “good” ones from them. The algorithm continues as:

Step 2. Choose an integer $N_0 \leq N'_0 - N_{\text{top}}$ and $N \leq N_0$. Choose $N_0$ non-top-view projections.
For each projection $P_k$, $1 \leq k \leq N_0$, estimate the common-lines $\{c_{0k}, c_{k0}\}$ between $\hat{P}_k$ and $\hat{P}_{\text{top}}$ by selecting the pair of lines so that

$$||\hat{P}_{\text{top}}(\xi c_{0k}) - \hat{P}_k(\xi c_{k0})||$$

is minimized, where $||\cdot||$ stands for the $\ell^2$ norm of functions. Let the minimizer be $\{\hat{c}_{0k}, \hat{c}_{k0}\}$. In practice the projection (in Fourier space) is discretized into $N_\theta$ radial lines, each having $N_r$ grid points. The minimization is carried out by a brute force search of the $(N_\theta/2)^2$ pairs, as only half of the plate needs to be considered due to the symmetry of the lines (see more in Sec. 8.2.3).

Notice that if $\{c_{0k}, c_{k0}\}$ is associated with $R_k$, then $\{-c_{0k}, -c_{k0}\}$ is associated with $J R_k J$. We define a vector

$$v_J(k), \quad k = 1, \ldots, N_0$$

as $v_J(k) = 1$ if $\{\hat{c}_{0k}, \hat{c}_{k0}\}$ (if corrected estimated) is consistent with the true rotation $R_k^*$, and $v_J(k) = -1$ if with $J R_k^* J$. The values of $v_J$ will be estimated in Step 4.

**Step 3.** For a pair of (non-top-view) projections $P_k$ and $P_j$, estimate $R_k(3, 3)$, $R_j(3, 3)$ and the number $v_J(k) v_J(j)$ in the following way:

For $\hat{P}_k$, the with-$\hat{P}_{\text{top}}$-commonlines $\{c_{0k}, c_{k0}\}$ are estimated in Step 2, and there is an unknown sign $v_J(k)$ which indicates whether $\{c_{0k}, c_{k0}\}$ or $\{-c_{0k}, -c_{k0}\}$ is consistent with $R_k^*$. With $\{c_{0k}, c_{k0}\}$, whenever $R_k(3, 3)$ is specified the rotation $R_k$ is determined (see more in Sec. 8.2.3). It is the same with $R_j$ for $j \neq k$.

At the same time, the rotations $R_j$ and $R_k$ dictate the “mutual common-lines” $\{c_{jk}, c_{kj}\}$ and $\{c_{jk}^g, c_{kj}^g\}$, which are supposed to be the shared lines by the projections $\hat{P}_j$ and $\hat{P}_k$. In other words, given $\{c_{0k}, c_{k0}\}$ and $\{c_{0j}, c_{j0}\}$, $\{c_{jk}, c_{kj}\}$
and \{c_{jk}^2, c_{kj}^2\} are functions of \(R_j(3, 3)\) and \(R_k(3, 3)\). Notice that \(\{\pm c_{0k}, \pm c_{0j}\}\) and \(\{\pm c_{0j}, \pm c_{0j}\}\) (namely \(v_j(k)v_j(j) = 1\)) give one type of such function, while \(\{\pm c_{0k}, \pm c_{0k}\}\) and \(\{\mp c_{0j}, \mp c_{0j}\}\) (namely \(v_j(k)v_j(j) = -1\)) give another type. Thus, we define the “cost function” of \((R_j(3, 3), R_k(3, 3)) := (x_1, x_2)\) as

\[
F_\pm(x_1, x_2) = ||\hat{P}_j(\xi c_{jk}) - \hat{P}_k(\xi c_{kj})||^2 + ||\hat{P}_j(\xi c_{jk}^g) - \hat{P}_k(\xi c_{kj}^g)||^2, \quad (x_1, x_2) \in [0, 1]^2,
\]

where the subscript \(\pm\) stands for the case of \(v_j(k)v_j(j) = \pm 1\) respectively. The mappings \(F_\pm(x_1, x_2)\) are non-linear, and to find their minima we use a brute force search on a mesh of \((x_1, x_2)\) as

\[(x_1, x_2) = (\cos t_1, \cos t_2), \quad -\frac{\pi}{2} + \theta_{th} \leq t_1, t_2 \leq \frac{\pi}{2} - \theta_{th},\]

where \(N_{33}\) even-spaced grid points are used for \(t_1\) and \(t_2\) respectively. Let the minimizer of \(F_\pm\) be \((x^*_1, x^*_2)\). If \(F_+(x^*_1, x^*_2) < F_-(x^*_1, x^*_2)\), we use \((x^*_1, x^*_2)\) as the estimated value of \(R_j(3, 3)\) \((R_k(3, 3))\) from \(\hat{P}_k\) \((\hat{P}_j)\), and let \((\hat{S}_j)(j, k) = (\hat{S}_j)(k, j) = 1\). If \(F_+(x^*_1, x^*_2) > F_-(x^*_1, x^*_2)\) we use the values of \((x^*_1, x^*_2)\) similarly, and let \((\hat{S}_j)(j, k) = (\hat{S}_j)(k, j) = -1\).

At the end of this step, we obtain a table of estimated values of \(R_k(3, 3)\) from \(N_0 - 1\) projections \(P_j, j \neq k\), and a \(N_0\)-by-\(N_0\) symmetric matrix \(\hat{S}_j\) where the off-diagonal entries take values as \pm 1 (the diagonal entries are set to be zeros).

Step 4. Compute the eigenvector \(v_1\) of \(\hat{S}_j\) with the largest eigenvalue, and let \(\hat{v}_j\) be the sign vector so that

\[\hat{v}_j(k) = \text{Sign}(v_1(k)), \quad k = 1, \ldots, N_0.\]
\( \hat{v}_J \) is an estimation of \( v_J \) up to a global sign, since if \((\hat{S}_J)(j, k) = v_J(j)v_J(k)\) for all \((j, k)\), then \( \hat{v}_J = v_J \) or \(-v_J\).

**Step 5.** Averaging the \( N_0 - 1 \) estimations of \( R_k \) computed in Step 3:

For each \( k \), Step 3 generates \( N_0 - 1 \) estimations of \( R_k(3, 3) \) by the other projections, from which a “voting-and-averaging” procedure can be used to estimate \( R_k(3, 3) \) along with a “confidence”. To be specific, if the “with-\( \hat{P}_{\text{top-commonline}} \) {\( c_{0k}, c_{k0} \)} are well estimated, the histogram of the \( N_0 - 1 \) “voted” values of \( R_k(3, 3) \) will be peaked at the true value. Otherwise, from a wrong estimation of \{\( c_{0k}, c_{k0} \)\} the cost function minimization approach in Step 3 will generate some random estimates of \( R_k(3, 3) \), and the histogram will be flat.

“Good voters” are the \( P_j \)'s from which the estimated values of \( R_k(3, 3) \) fall into a small bin of the histogram peak. The final estimate of \( \hat{R}_k(3, 3) \) is averaged from the “good votes”, and the “confidence” is in terms of the fraction of “good voters” among the total \( N_0 - 1 \) ones.

The final estimation of \( \hat{R}_k \) is determined by \( \hat{R}_k(3, 3) \) and \{\( \hat{v}_J(k)\hat{c}_{0k}, \hat{v}_J(k)\hat{c}_{k0} \)}

where \( \hat{v}_J \) is computed in Step 4. We choose the \( N \) estimations out of the \( N_0 \) ones as those that have the best voting-based “confidence”.

**Step 6.** From the \( N \) estimated rotations \( \hat{R}_k \) and the associated denoised projections, reconstruct the volume \( V \).

Step 2, 3 can be naturally parallelized. Further acceleration of computing can be obtained by doing a first-pass voting of \( R_k(3, 3) \) from \( R_j \) where \( j \) goes over some subset of \{1, \cdots, N_0\}, and then complete the voting table only for those \( R_k \)'s that show good confidence from the first-pass voting.
8.4.2 Numerical Experiments

The simulated projections are prepared in the same way as in Sec. 8.3.2. Gaussian noise is added to the projections with SNR = $1/4, 1/8, 1/16, 1/32, 1/64, 1/128$. The clean and simulated noisy projections are shown in Fig. 8.6.

The parameters of the algorithm are set as

- Number of projections: $N_0' = 4000, N_0 = 2000, N = 100$.
- $\theta_{th}$-top-view: $\theta_{th} = 10$ degrees.
- Discretization of Fourier projections: $N_r = 60, N_\theta = 360$.
- Discretization of $R_k(3, 3)$: $N_{33} = 72$.

The “top-view” projections estimated by Step 1 in Sec. 8.4.1 are shown in Fig. 8.7 for SNR = $1/4, 1/16, 1/64$ respectively. The estimated “top-view” projection is a rotated and possibly also mirrored version of the clean one, containing a certain level of noise.

The accuracy of rotation recovery is evaluated by MSE defined as in Eq. (8.5). Tab. 8.1 shows the MSE with varying values of SNR. We also list the MSE when using a noiseless top-view projection for Step 2-6 as a comparison, which indicates that using the estimated top-view projection does not degrade the performance. In
Figure 8.7: (Upper) The clean “top-view projection and “top-view” projections at SNR = 1/4, 1/16 and 1/64. (Lower) The estimated “top-view projections at the same noise level as the upper row.

<table>
<thead>
<tr>
<th>SNR</th>
<th>MSE (using clean “top-view”)</th>
<th>MSE (using estimated “top-view”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/4</td>
<td>0.0017</td>
<td>0.0019</td>
</tr>
<tr>
<td>1/8</td>
<td>0.0027</td>
<td>0.0030</td>
</tr>
<tr>
<td>1/16</td>
<td>0.0077</td>
<td>0.0073</td>
</tr>
<tr>
<td>1/32</td>
<td>0.0153</td>
<td>0.0181</td>
</tr>
<tr>
<td>1/64</td>
<td>0.0604</td>
<td>0.0599</td>
</tr>
<tr>
<td>1/128</td>
<td>0.3258</td>
<td>0.3256</td>
</tr>
<tr>
<td>1/256</td>
<td>0.9530</td>
<td>1.5735</td>
</tr>
</tbody>
</table>

Table 8.1: MSE of the $N = 100$ estimated rotations. The MSE when using a noiseless top-view projection for Step 2-6 is listed as a comparison.

this test $N = 100$ recovered rotations are selected from the $N_0 = 2000$ ones based on the “confidence” computed by the voting procedure.

For SNR = 1/32, Fig. 8.8 (left) shows the error histogram of the $N = 100$ recovered orientations, where most errors are less than 10 degrees. Fig. 8.8 (right) shows the side view of the reconstructed 3D volume. In Fig. 8.9, the reconstructed volume is compared with the true volume which is used to generate the simulated projections.
Figure 8.8: (Left) Histogram of rotation recovery errors in degrees. (Right) Volume reconstructed from estimated orientations and de-noised projections, seen from the side. SNR = 1/32.

Figure 8.9: The true volume (left) and the reconstructed one (right) seen from the top. SNR = 1/32.
Chapter 9

Conclusion

9.1 Conclusion

We derive a theoretical solution to the symmetry detection problem based on estimating the rank of the $l$-th order auto-correlation function $C_l(r_1, r_2)$. Given finite many (Fourier transformed) projections sampled on a Cartesian grid, we expand them in the truncated Fourier-Bessel basis, from which we can compute $C_l(r_1, r_2)$ in the basis of truncated spherical Bessel series. In the preliminary tests on simulated noisy projections the algorithm discriminates the $C_2$ symmetric volume from the non-symmetric one.

We propose two approaches to attack the structural reconstruction problem. One of them extends the synchronization method for non-symmetric molecules, and the other firstly estimates the top-view which is a feather of cyclic groups. Both methods are tested on the simulated noisy projections generated from a $C_2$ symmetric volume.

9.2 Future Directions

The study will continue in the following directions:
For the symmetry detection problem, we plan to improve the estimation of the \( l \)-th order self-correlation function to be more robust to noise. Meanwhile, the current algorithm outputs the spectrum of the matrices \((C_l)^{\text{proj}}\), which needs to be translated into the rank of \( C_l \) and then the type of symmetry. We plan to develop an automatic inference scheme for the rank estimation and symmetry group identification.

For the structural reconstruction problem, we plan to extend the current algorithm to higher-order symmetry groups. The top-view approach is potentially valid for any cyclic group, and the synchronization approach for any point symmetry group. In the synchronization approach it is desirable to estimate the relative rotations \( \{R^T_i g R_j, g \in G\} \) based on likelihood rather than the common-line detection, since the latter is challenging when the group order \( |G| \) is high. For example, as is explained in Sec. 8.2 there are \( |G| \) pairs of mutual common-lines between two projections and \( |G| - 1 \) pairs of self common-lines on a projection, and the successful detection of all of them has slim chance at low SNR.

For both problems, we plan to go beyond the numerical tests on simulated data and apply the algorithms to experimental cryo-EM data.
Appendix A

Stieltjes Transform

In this section we review the two basic properties of the Stieltjes transform to make the text self-contained. Stieltjes transforms are widely used in the study of random matrix theory and are reviewed in many places in the existing literature, see e.g. Appendix B.2 of [3].

For a probability measure $d\mu(x) = p(x)dx$ (assumed to be absolutely continuous) on $\mathbb{R}$, the Stieltjes transform is defined as

$$m(z) = \int_{\mathbb{R}} \frac{1}{x - z}d\mu(x), \quad \Im(z) > 0,$$

and $\Im(m(z)) > 0$.

A.1 Inversion Formula

The probability density function can be recovered from its Stieltjes transform via the “inversion formula”

$$\lim_{b \to 0^+} \frac{1}{\pi} \Im(m(x + ib)) = p(x), \quad x \in \mathbb{R} \quad (A.1)$$
where the convergence is in the weak sense.

### A.2 Boundedness by $1/\Im(z)$

For $z = u + iv$, $v > 0$,

$$|m(z)| \leq \int_{\mathbb{R}} \left| \frac{1}{x - z} \right| d\mu(x) \leq \frac{1}{v}$$

for any probability measure $\mu$. Specifically,

$$|m_A(z)| \leq \frac{1}{n} \sum_{i=1}^{n} \frac{1}{|\lambda_i(A) - z|} \leq \frac{1}{n} \sum_{i=1}^{n} \frac{1}{v} = \frac{1}{v}.$$  \hfill (A.2)
Appendix B

Hermite Polynomials

Define the normalized Hermite polynomials as

\[ h_l(x) = \frac{1}{\sqrt{l!}} H_l(x), \quad l = 0, 1, \cdots \]  \hspace{1cm} (B.1)

where \( H_l(x) \) is the \( l \)-degree Hermite polynomial, satisfying

\[ \int_{\mathbb{R}} H_{l_1}(x) H_{l_2}(x) q(x) dx = \delta_{l_1, l_2} \cdot l_1!. \]

Thus, \( \{h_l(x), l = 0, 1, \cdots \} \) form an orthonormal basis of \( \mathcal{H}_N \). The explicit formula of \( H_l \) is \[1\]

\[ H_l(x) = l! \sum_{k=0}^{[l/2]} \left( -\frac{1}{2} \right)^k \frac{1}{k!(l - 2k)!} x^{l-2k}. \]  \hspace{1cm} (B.2)

Also, the derivative of \( H_l(x) \) satisfies the recurrence relation \( H_l'(x) = lH_{l-1}(x) \) for \( l \geq 1 \), and as a result,

\[ h_l'(x) = \sqrt{l} h_{l-1}(x). \]  \hspace{1cm} (B.3)
Bibliography


