Singular vector and singular subspace distribution for the matrix denoising model

Zhigang Bao∗
Hong Kong University of Science and Technology
mazgbao@ust.hk

Xiucai Ding†
University of Toronto
xiucai.ding@mail.toronto.edu

Ke Wang‡
Hong Kong University of Science and Technology
kewang@ust.hk

In this paper, we study the matrix denoising model \( Y = S + X \), where \( S \) is a low rank deterministic signal matrix and \( X \) is a random noise matrix, and both are \( M \times n \). In the scenario that \( M \) and \( n \) are comparably large and the signals are supercritical, we study the fluctuation of the outlier singular vectors of \( Y \), under fully general assumption on the structure of \( S \) and the distribution of \( X \). More specifically, we derive the limiting distribution of angles between the principal singular vectors of \( Y \) and their deterministic counterparts, the singular vectors of \( S \). Further, we also derive the distribution of the distance between the subspace spanned by the principal singular vectors of \( Y \) and that spanned by the singular vectors of \( S \). It turns out that the limiting distributions depend on the structure of the singular vectors of \( S \) and the distribution of \( X \), and thus they are non-universal. Statistical applications of our results to singular vector and singular subspace inferences are also discussed.

1. Introduction

Consider an \( M \times n \) noisy matrix \( Y \) modeled as

\[
Y = S + X,
\]

where \( S \) is a low-rank deterministic matrix with fixed rank \( r \) and \( X \) is a real random noise matrix. We assume that \( S \) admits the singular value decomposition

\[
S = UDV^* = \sum_{i=1}^{r} d_i u_i v_i^*,
\]

where \( D = \text{diag}(d_1, \ldots, d_r) \) consists of the singular values of \( S \) and we assume \( d_1 > \ldots > d_r > 0 \); \( U = (u_1, \ldots, u_r) \in \mathbb{R}^{M \times r} \) and \( V = (v_1, \ldots, v_r) \in \mathbb{R}^{n \times r} \) are the matrices consisting of the \( \ell^2 \)-normalized left and right singular vectors. For the noise matrix \( X = (x_{ij}) \) in (1.1), we assume that the entries \( x_{ij} \)'s are i.i.d real random variables with

\[
\mathbb{E}x_{ij} = 0, \quad \mathbb{E}|x_{ij}|^2 = \frac{1}{n}.
\]

For simplicity, we also assume the existence of all moments, i.e., for every integer \( q \geq 3 \), there is some constant \( C_q > 0 \), such that

\[
\mathbb{E}|\sqrt{n}x_{ij}|^q \leq C_q < \infty.
\]

This condition can be weakened to the existence of some sufficiently high order moment. But we do not pursue this direction here. We remark that although we are primarily interested in the real case, our method also applies to the case when \( X \) is a complex noise matrix.

In practice, \( S \) is often called the signal matrix which contains the information of interest. In the high-dimensional setup, when \( M \) and \( n \) are comparably large, we are primarily

∗Supported by Hong Kong RGC ECS Grant 26301517.
†Supported in part by NSERC of Canada (Jeremy Quastel).
‡Supported by HKUST Initiation Grant IGN16SC05.
interested in the inference of $S$ or its left and right singular spaces, which are the subspaces spanned by $\mathbf{u}_i$'s or $\mathbf{v}_i$'s, respectively. Such a problem arises in many scientific applications such as sparse PCA [68, 74], matrix denoising [25, 27], multiple signal classification (MUSIC) [36, 72], Zq synchronization [64, 65] and multidimensional scaling [26, 30, 57]. We call the model in (1.1) the matrix denoising model, which is also often referred to as the signal-plus-noise model in the literature. We refer to subsection 1.2 for more introduction on the application aspects.

In this paper, we are interested in the following high-dimensional regime: for some small constant $\tau \in (0,1)$ we have

$$M \equiv M(n), \quad y \equiv y_n := \frac{M}{n} \to c \in [\tau, \tau^{-1}], \quad \text{as} \quad n \to \infty. \quad (1.6)$$

Our main results are on the limiting distributions of individual $|\langle \mathbf{u}_i, \mathbf{u}_j \rangle|^2$ (resp. $|\langle \mathbf{v}_i, \mathbf{v}_j \rangle|^2$) and $L$ (resp. $R$) when the signal strength, $d_i$'s, are supercritical (cf. Assumption 2.1). They are detailed in Theorems 2.3, 2.9, 2.10 and 2.11, after necessary notations are introduced. In the rest of this section, we review some related literature from both theoretical and applied perspectives.
1.1. On finite-rank deformation of random matrices. From the theoretical perspective, our model in (1.1) is in the category of the fixed-rank deformation of the random matrix models in the Random Matrix Theory, which also includes the deformed Wigner matrix and the spiked sample covariance matrix as typical examples. There are a vast of work devoted to this topic and the primary interest is to investigate the limiting behavior of the extreme eigenvalues and the associated eigenvectors of the deformed models. Since the seminal work of Baik, Ben Arous and Péché [4], it is now well-understood that the extreme eigenvalues undergo a so-called BBP transition along with the change of the strength of the deformation. Roughly speaking, there is a critical value such that the extreme eigenvalue of the deformed matrix will stick to the right end point of the limiting spectral distribution of the undeformed random matrix if the strength of the deformation is less than or equal to the critical value, and will otherwise jump out of the support of the limiting spectral distribution. In the latter case, we call the extreme eigenvalue as an outlier, and the associated eigenvector as an outlier eigenvector. Moreover, the fluctuation of the extreme eigenvalues in different regimes (subcritical, critical and supercritical) are also identified in [4] for the complex spiked covariance matrix. We also refer to [5, 10, 11, 3, 21, 25, 42, 55] and the reference therein for the first-order limit of the extreme eigenvalue of various fixed-rank deformation models. The fluctuation of the extreme eigenvalues of various models have been considered in [2, 3, 7, 8, 9, 23, 24, 28, 14, 15, 42, 55, 34, 56, 59, 44]. Especially, the fluctuations of the outliers are shown to be non-universal for the deformed Wigner matrices, first in [23] under certain special assumptions on the structure of the deformation and the distribution of the matrix entries, and then in [42] in full generality.

The study on the behavior of the extreme eigenvectors has been mainly focused on the level of the first order limit [10, 11, 20, 25, 33, 55]. In parallel to the results of the extreme eigenvalues, it is known that the eigenvectors are delocalized in the subcritical case and have a bias on the direction of the deformation in the supercritical case. It is recently observed in [13] that a deformation close to the critical regime will cause a bias even for the non-outlier eigenvectors. On the level of the fluctuation, the limiting behavior of the extreme eigenvectors has not been fully studied yet. By establishing a general universality result of the eigenvectors of the sample covariance matrix in the null case, the authors of [13] are able to show that the law of the eigenvectors of the spiked covariance matrices are asymptotically Gaussian in the subcritical regime. More specifically, the generalized components of the eigenvectors are $\chi^2$ distributed. In the supercritical regime, under some special assumptions on the structure of the deformation and the distribution of the random matrix entries, it is shown in [22] that the eigenvector distribution of a generalized deformed Wigner matrix model is non-universal in the supercritical regime. In the current work, we aim at establishing the non-universality for the outlier singular vectors for the matrix denoising model under fully general assumptions on the structure of the deformation $S$ and the distribution of the random matrix $X$. This can be regarded as an eigenvector counterpart of the result on the outlying eigenvalue distribution in [42].

1.2. On singular subspace inference. From the applied perspective, our model (1.1) appears prominently in the study of signal processing [40, 53], machine learning [69, 71] and statistics [17, 18, 27, 32]. For instance, in the study of image denoising, $S$ is treated as the true image [50] and in the problem of classification, $S$ contains the the underlying true mean vectors of samples [17]. In both situations, we need to understand the asymptotics of the singular vectors and subspace of $S$, given the observation $Y$. In addition, the statistics $R$ and $L$ defined in (1.5) can be used for the inference of the structure of the singular subspace
of $S$. In the high-dimensional regime (1.6), to the best of our knowledge, the distributions of $R$ and $L$ have not been studied yet in the literature.

In the situation when $M$ is fixed, the sample eigenvectors of $XX^*$ are normally distributed [1]. When $M$ diverges with $n$, many interesting results have been proposed under various assumptions. One line of the work is to derive the perturbation bounds for the perturbed singular vectors based on Davis-Kahan’s theorem. For instance, in [54], the authors improve the perturbation bounds of Davis-Kahan theorem to be nearly optimal. In [17], the authors study similar problems and their related statistical applications. Most recently, in the papers [31, 32, 73], the authors derive the $\ell_\infty$ perturbation bounds assuming that the population vectors were delocalized (i.e. incoherent). The other line of the work is to study the asymptotic normality of the spectral projection under various regularity conditions. In such cases, the singular vectors of $S$ can be estimated using those of $Y$ and some Gaussian approximation technique can be employed. Considering the Gaussian data samples $x_i \sim \mathcal{N}(0, \Sigma), i = 1, 2, \cdots, n$ and $X = (x_i)$, under the assumption that the order of $\frac{\|\Sigma\|}{\sqrt{n}}$ is much smaller than $n$, in [45, 46, 47], the authors prove that the eigenvectors of $XX^*$ are asymptotically normally distributed, whose variance depends the eigenvectors of $\Sigma$. Furthermore, in [70], assuming that $m$ such random matrices $X_i, i = 1, 2, \cdots, m$ are available, the author shows that the singular vectors of $S$ can be estimated via trace regression using matrix nuclear norm penalized least squares estimation (NNPLS). Under the assumption that $rK \log^3 m = o(m), K = \max\{M, n\}$, the author shows that the principal angles of the subspace estimated using NNPLS are asymptotically normal. In [39], for $n$ i.i.d sub-Gaussian samples $x_i$ with population covariance matrix $\Sigma$, the authors estimate the first loading factor $\beta_1$ of $\Sigma$ using a Lasso type de-biased estimator. Under the assumption that $M \gg n$ and $\beta_1$ is sparse, i.e $|\beta_1|_0 = o(\sqrt{n}/\log M)$, the authors prove that the Lasso type de-biased estimator is asymptotically normal.

1.3. Organization. The rest of the paper is organized as follows. In Section 2, we state our main results and summarize our proof strategy. In Section 3, we design Monte Carlo simulations to demonstrate the accuracy of our main results and illustrate their applications through a few hypothesis testing problems. In Section 4, we introduce some main technical results including the isotropic local law and also derive the Green function representation for our statistics. In Section 5, we prove Theorems 2.3 and 2.10, based on the recursive estimate in Proposition 5.2. Section 6 is then devoted to the proof of Proposition 5.2. In Section 7, we state the proof for a main technical lemma, Lemma 6.2, which is used in the proof of Proposition 5.2. In Section 8, we prove Theorems 2.9 and 2.11.

2. Main results and methodology

In this section, we state our main results, and briefly summarize our proof strategy.

2.1. Notations. For a positive integer $n$, we denote by $[n]$ the set $\{1, \cdots, n\}$. Let $\mathbb{C}^+$ be the complex upper-half plane. Further, we define the following linearization for our model

$$\mathcal{Y}(z) := UD(z)U^* + H(z), \quad z = E + i\eta \in \mathbb{C}^+, \quad (2.1)$$

where

$$U := \begin{pmatrix} U & V \end{pmatrix}, \quad D(z) := \sqrt{z} \begin{pmatrix} D & D \\ D & D \end{pmatrix}, \quad H(z) := \sqrt{z} \begin{pmatrix} X^* & X \end{pmatrix}. \quad (2.2)$$

In the sequel, we will often omit $z$ and simply write $\mathcal{Y} \equiv \mathcal{Y}(z), \mathcal{D} \equiv \mathcal{D}(z)$ and $H \equiv H(z)$ when there is no confusion.
We denote the empirical spectral distributions (ESD) of the matrices $XX^*$ and $X^*X$ by
\[
F_1(x) := \frac{1}{M} \sum_{i=1}^{M} 1_{\{\lambda_i(XX^*) \leq x\}}, \quad F_2(x) := \frac{1}{n} \sum_{i=1}^{n} 1_{\{\lambda_i(X^*X) \leq x\}}.
\]
$F_1(x)$ and $F_2(x)$ are known to satisfy the Marchenko-Pastur (MP) law [52]. More precisely, almost surely, $F_1(x)$ converges weakly to a non-random limit $F_{1y}(x)$ which has a density function given by
\[
\rho_1(x) := \begin{cases} \frac{1}{2\pi xy} \sqrt{(\lambda_+ - x)(x - \lambda_-)}, & \text{if } \lambda_- \leq x \leq \lambda_+, \\ 0, & \text{otherwise}, \end{cases}
\]
and has a point mass $1 - 1/y$ at the origin if $y > 1$, where $\lambda_+ = (1 + \sqrt{y})^2$ and $\lambda_- = (1 - \sqrt{y})^2$. Furthermore, the Stieltjes’s transform of $F_{1y}$ is given by
\[
m_1(z) := \int \frac{1}{x - z} dF_{1y}(x) = \frac{1 - y - z + i \sqrt{(\lambda_+ - z)(z - \lambda_-)}}{2yz}, \quad \text{for } z \in \mathbb{C}^+, \quad (2.3)
\]
where the square root denotes the complex square root with a branch cut on the negative real axis. Similarly, almost surely, $F_2(x)$ converges weakly to a non-random limit $F_{2y}(x)$ which has a density function given by
\[
\rho_2(x) := \begin{cases} \frac{1}{2\pi xy} \sqrt{(\lambda_+ - x)(x - \lambda_-)}, & \text{if } \lambda_- \leq x \leq \lambda_+, \\ 0, & \text{otherwise}, \end{cases}
\]
and a point mass $1 - y$ at the origin if $y < 1$. The corresponding Stieltjes’s transform is
\[
m_2(z) := \int \frac{1}{x - z} dF_{2y}(x) = \frac{y - 1 - z + i \sqrt{(\lambda_+ - z)(z - \lambda_-)}}{2z}. \quad (2.4)
\]

In this paper, the singular values of $S$ are assumed to satisfy the supercritical condition.

**Assumption 2.1** (Supercritical condition). *There exists a constant $\delta > 0$, such that
\[
d_1 > d_2 > \cdots > d_r \geq y^{1/4 + \delta}, \quad \min_{1 \leq j \neq k \leq r} |d_j - d_k| \geq \delta.
\]

**Remark 2.2.** The first inequality above ensures that the first $r$ singular values of $Y$ are outliers. The second inequality guarantees that the outliers of $Y$ are well separated from each other. Both conditions can be weakened. For instance, we do allow the existence of the subcritical and critical $d_i$’s if we only focus on the outlier singular vectors of $Y$. Also, the separation of $d_i$’s by an order 1 distance $\delta$ is not necessary. In [13], a much weaker separation of order $n^{-1/2+\epsilon}$ is enough for the discussion of the eigenvalues. But we do not pursue these directions in the current paper.

To state our results, we need more notations. First, we define
\[
p(d) := \frac{(d^2 + 1)(d^2 + y)}{d^2}. \quad (2.5)
\]
For each $i \in [r]$, we will write $p_i \equiv p(d_i)$ for short. In [25, Theorem 3.4], it has been shown that $p_i$ is the limit of $\mu_i$. Further, we set
\[
a_1(d) := \frac{d^4 - y}{d^2(d^2 + y)} \quad \text{and} \quad a_2(d) := \frac{d^4 - y}{d^2(d^2 + 1)}. \quad (2.6)
\]
It has been proved in [25] that $a_1(d_i)$ and $a_2(d_i)$ are the limits of $|\langle u_i, \tilde{u}_i \rangle|^2$ and $|\langle v_i, \tilde{v}_i \rangle|^2$ respectively (see Lemma 4.9 below). We also denote by $\kappa_l$ the $l$th cumulant of the random variables $\sqrt{n}x_{ij}$.
2.2. Main results. In this section, we state our main results.

For a vector \( \mathbf{w} = (w(1), \ldots, w(m))^T \) and \( l \in \mathbb{N} \), we introduce the notation
\[
\mathbf{s}_l(\mathbf{w}) := \sum_{i=1}^{m} w(i)^l.
\]

Set
\[
\theta(d) := \frac{d^4 + 2yd^2 + y}{d^4(d^2 + 1)^2}, \quad \psi(d) := \frac{d^6 - 3yd^2 - 2y}{d^4(d^2 + 1)^2}, \quad (2.7)
\]
and
\[
\mathcal{V}^E(d) := \frac{2}{d^2} \left( 2y(y+1)\theta(d)^2 - \frac{y(y-1)(5y+1)}{d(d^2+1)^2} \theta(d) \right.
\]
\[
+ \frac{(d^4+y)(d^2+y)^2}{d^4(d^2+1)^2} \psi(d) + \frac{2y^2(y-1)^2}{d^2(d^2+1)^2} \bigg). \quad (2.8)
\]
For the right singular vectors, we have the following theorem.

**Theorem 2.3** (Right singular vectors). Assume (1.2), (1.3), (1.6) and Assumption 2.1 hold. For \( i \in [r] \), define the random variable
\[
\Delta_i := -2\sqrt{n}\theta(d_i)\mathbf{v}_i^*X\mathbf{v}_i - \frac{2\psi(d_i)}{d_i^2} \frac{\kappa_3}{n} \mathbf{s}_1(\mathbf{u}_i)\mathbf{s}_1(\mathbf{v}_i), \quad (2.9)
\]
and let \( Z_i \) be a random variable, independent of \( \Delta_i \), with law \( Z_i \sim \mathcal{N}(0, \mathcal{V}_i) \), where
\[
\mathcal{V}_i := \mathcal{V}^E(d_i) - \frac{4}{d_i} \theta(d_i)\psi(d_i) \left( \frac{\kappa_3}{\sqrt{n}} \mathbf{s}_1(\mathbf{u}_i)\mathbf{s}_1(\mathbf{v}_i) \right) + \frac{4}{d_i} \theta(d_i)^2 \left( \frac{\kappa_3}{\sqrt{n}} \mathbf{s}_1(\mathbf{u}_i)\mathbf{s}_3(\mathbf{v}_i) \right)
\]
\[
+ \frac{\psi(d_i)^2}{d_i^2} \kappa_4 \mathbf{s}_4(\mathbf{u}_i) + \frac{\theta(d_i)^2}{d_i^2} \kappa_4 \mathbf{s}_4(\mathbf{v}_i).
\]

Then for any \( i \in [r] \) and for any bounded continuous function \( f \), we have that
\[
\lim_{n \to \infty} \left( E f \left( \sqrt{n} \left( \parallel \mathbf{v}_i \parallel^2 - a_2(d_i) \right) \right) - E f(\Delta_i + Z_i) \right) = 0.\]

**Remark 2.4.** In [42], the authors obtain the non-universality for the limiting distributions of the outliers (outlying eigenvalues) of the deformed Wigner matrices. The limiting distributions admit similar forms as the limiting distributing for the outlier singular vectors for our models. One might notice that the third or the fourth cumulants of the entries of the Wigner matrices are allowed to be different in [42]. An extension along this direction is also straightforward for our result.

We discuss a few special cases of interest. For simplicity, we assume that \( S \) has rank \( r = 1 \) and drop all the subindices.

**Remark 2.5.** If the entries of \( \sqrt{n}X \) are standard Gaussian random variables (i.e. \( \kappa_3 = \kappa_4 = 0 \)), then \( \Delta \sim \mathcal{N}(0, 4\theta(d)^2) \) and thus \( \Delta + Z \) is asymptotically distributed as
\[
\mathcal{N}(0, 4\theta(d)^2 + \mathcal{V}^E(d)).
\]

**Remark 2.6.** If both \( u \) and \( v \) are delocalized in the sense that \( \parallel u \parallel_\infty = o(1) \) and \( \parallel v \parallel_\infty = o(1) \), then \( s_l(u) = o(1) \) and \( s_l(v) = o(1) \) for \( l = 3, 4 \). We conclude from central limit theorem that
\[
\Delta \sim \mathcal{N} \left( -\frac{2\psi(d)}{d^2} \frac{\kappa_3}{n} \mathbf{s}_1(\mathbf{u})\mathbf{s}_1(\mathbf{v}), 4\theta(d)^2 \right), \quad (2.10)
\]
and therefore $\Delta + \mathcal{Z}$ has asymptotically the same distribution as
\[
\mathcal{N} \left( -\frac{2\psi(d)}{d^2} \left( \frac{\kappa_3}{n} s_1(u)s_1(v) \right), 4\theta(d)^2 + \mathcal{V}(d) \right).
\]
The only difference from the Gaussian case is a shift caused by the non-vanishing third cumulant.

**Remark 2.7.** If one of $u$ and $v$ is delocalized, say $\|u\|_\infty = o(1)$, then $\Delta$ still has the limiting distribution in (2.10). Therefore $\Delta + \mathcal{Z}$ has asymptotically the same distribution as a Gaussian random variable with mean
\[
-\frac{2\psi(d)}{d^2} \left( \frac{\kappa_3}{n} s_1(u)s_1(v) \right)
\]
and variance
\[
4\theta(d)^2 + \mathcal{V}(d) + \frac{4}{d^2} \theta(d)^2 \left( \frac{\kappa_3}{\sqrt{n}} s_1(u)s_3(v) \right) + \frac{y\theta(d)^2}{d^2} \kappa_4 s_4(v).
\]

**Remark 2.8.** If neither $u$ nor $v$ is delocalized, then $\Delta + \mathcal{Z}$ is no longer Gaussian in most cases. For example, if $u = e_1$ and $v = f_1$ where $e_1$ and $f_1$ are the canonical basis vectors in $\mathbb{R}^M$ and $\mathbb{R}^n$ respectively, then $\Delta + \mathcal{Z}$ is asymptotically distributed as
\[
-2\theta(d)\sqrt{n}X_{11} + \mathcal{N} \left( 0, \mathcal{V}(d) + \kappa_4 \frac{\psi(d)^2 + y\theta(d)^2}{d^2} \right),
\]
which depends on the distribution of the noise matrix and thus is non-universal.

For two vectors $w_1 = (w_1(1), \ldots, w_1(m))^T$ and $w_2 = (w_2(1), \ldots, w_2(m))^T$, we denote
\[
s_{k,l}(w_1, w_2) := \sum_{i=1}^m w_1(i)^k w_2(i)^l.
\]
Recall $R$ from (1.5). We have the following theorem.

**Theorem 2.9 (Right singular subspace).** Assume (1.2), (1.3), (1.6) and Assumption 2.1 hold. Let $\Delta = \sum_{i=1}^r \Delta_i$, where $\Delta_i$ is defined in (2.9). Let $\mathcal{Z}$ be a random variable independent of $\Delta$ with law $\mathcal{Z} \sim \mathcal{N}(0, \mathcal{V})$, where
\[
\mathcal{V} := \sum_{i=1}^r \mathcal{V}(d_i) + \kappa_4 \sum_{i,j=1}^r \left( \frac{\psi(d_i)\psi(d_j)}{d_id_j} s_{2,2}(u_i, u_j) + \frac{\theta(d_i)\theta(d_j)}{d_id_j} s_{2,2}(v_i, v_j) \right)
+ \frac{\kappa_3}{\sqrt{n}} \sum_{i,j=1}^r \frac{4}{d_i} \theta(d_j) \left( \theta(d_i)s_{2,1}(v_i, v_j)s_1(u_i) - \psi(d_i)s_{2,1}(u_i, u_j)s_1(v_j) \right).
\]

Then for any bounded continuous function $f$, we have that
\[
\lim_{n \to \infty} \left( \mathbb{E}f(\sqrt{n}(R - \sum_{i=1}^r a_2(d_i))) - \mathbb{E}f(\Delta + \mathcal{Z}) \right) = 0.
\]
Similarly, we set
\[
\zeta(d) := \frac{y(d^4 + 2d^2 + y)}{d^3(d^2 + y)^2}, \quad \phi(d) := \frac{d^6 - 3y(d^2 + 2y^2)}{d^3(d^2 + y)^2}.
\]
Then for any bounded continuous function $f$ hold. Let
\[ \Lambda := \sum_{i,j=1}^{r} \kappa_{i,j} s_{i,j}(u_i) s_{i,j}(v_i), \]
and let $Z_i$ be a random variable independent of $\Lambda_i$, with law $Z_i \sim \mathcal{N}(0, V_i)$, where
\[ V_i := V^E(d_i) - \frac{4}{d_i} \kappa_{3} \frac{s_{i}(u_i) s_{i}(v_i)}{\sqrt{2n}} + \frac{4}{d_i} \kappa_{4} s_{i}(u_i) s_{i}(v_i) \]
\[ + \frac{2\phi(d_i)}{d_i^2} \kappa_{4} s_{i}(v_i) + \frac{2\phi(d_i)}{d_i^2} \kappa_{4} s_{i}(u_i). \]
Then for any $i \in [r]$ and any bounded continuous function $f$, we have that
\[ \lim_{n \to \infty} \left( \mathbb{E} f \left( \sqrt{n} \left( |\langle u_i, \tilde{u}_i \rangle|^2 - a_1(d_i) \right) \right) - \mathbb{E} f(\Lambda_1 + Z_i) \right) = 0. \]

Next, we state the result on the asymptotic distribution of $L$ defined in (1.5).

**Theorem 2.11** (Left singular subspace). Assume (1.2), (1.3), (1.6) and Assumption 2.1 hold. Let $\Lambda = \sum_{i=1}^{r} \Lambda_i$, where $\Lambda_i$ is defined in (2.11). Let $Z$ be a random variable independent of $\Lambda$ with law $Z \sim \mathcal{N}(0, V)$, where
\[ V := \sum_{i=1}^{r} V^E(d_i) + \kappa_{4} \sum_{i,j=1}^{r} \left( \frac{\kappa_{3}}{\sqrt{2n}} s_{i,j}(u_i) s_{i,j}(v_j) + \frac{\kappa_{4}}{d_i} s_{i,j}(u_i) s_{i,j}(v_j) \right) \]
\[ + \frac{2\phi(d_i)}{d_i^2} \kappa_{4} s_{i,j}(v_j). \]
Then for any bounded continuous function $f$, we have that
\[ \lim_{n \to \infty} \left( \mathbb{E} f \left( \sqrt{n} \left( \sum_{i=1}^{r} a_1(d_i) \right) \right) - \mathbb{E} f(\Lambda + Z) \right) = 0. \]

### 2.3. Proof strategy
In this subsection, we briefly describe our proof strategy. We first review the method used in a related work [42], and then we highlight the novelty of our strategy.

As we previously mentioned, in [42], the authors derive the distribution of outliers (outlying eigenvalues) of the fixed-rank deformation of Wigner matrices. The main technical input is the isotropic local law for Wigner matrices, which provides a precise large deviation estimate for the quadratic form $\langle u, (W - z)^{-1} v \rangle$ for any deterministic vectors $u, v$. Here $W$ is a Wigner matrix. It turns out that an outlier of the deformed Wigner matrix can also be approximated by a quadratic form of the Green function, of the form $\langle u, (W - z)^{-1} u \rangle$. So one can turn to establish the law of the quadratic form of the Green function instead. In [42], the authors decompose the proof into three steps. First, the law is established for the GOE/GUE, the Gaussian Wigner matrix, for which orthogonal/unitary invariance of the matrix can be used to facilitate the proof. In the second step of going beyond Gaussian
matrix, in order to capture the independence of the Gaussian part and the non-Gaussian part of the limiting distribution of the outliers, the authors construct an intermediate matrix in which most of the matrix entries are replaced by the Gaussian ones while those with coordinates corresponding to the large components of \( \mathbf{u} \) are kept as generally distributed. The intermediate matrix allows one to use the nice properties of the Gaussian ensembles such as orthogonal/unitary invariance for the major part of the matrix, and meanwhile keeps the non-Gaussianity induced by the small amount of generally distributed entries. In the last step, the authors of [42] derive the law for the fully generally distributed Wigner matrix by further conducting a Green function comparison with the intermediate matrix.

For our problem, similarly, we will use the isotropic law of the sample covariance matrix in [12, 43] as a main technical input. It turns out that for the singular vectors, we can approximately represent \( \sqrt{n} \langle \mathbf{u}_i, \mathbf{u}_i \rangle \) (after appropriate centralization) in terms of a quantity of the form

\[
\mathcal{Q}_i = \sqrt{n} \left( \text{Tr}(G(p(d_i))) - \Pi_1(p(d_i))A_i^R + \text{Tr}(G'(p(d_i)) - \Pi_1(p(d_i)))B_i^R \right),
\]

where \( G \) is the Green function of the linearization of the sample covariance matrix and \( \Pi_1 \) is the deterministic approximation of \( G \); see (4.1) and (4.6) for the definitions. Here both \( A_i^R \) and \( B_i^R \) are deterministic fixed-rank matrices. Hence, differently from the outlying eigenvalues or singular values, the Green function representation of the singular vectors also contains the derivative of the Green function. More importantly, instead of the three step strategy in [42], here we derive the law of the above \( \mathcal{Q}_i \) directly for generally distributed matrix. Recall \( \Delta_i \) defined in (2.9), whose random part is proportional to \( \mathbf{u}_i^* \mathbf{X} \mathbf{v}_i \), which is simply a linear combination of the entries of \( \mathbf{X} \). Inspired by [42], we decompose \( \Delta_i \) into two parts, say \( \hat{\Delta}_i \) and \( \tilde{\Delta}_i \). The former contains the linear combination of \( x_{ki} \)’s for those indices \( k, \ell \) corresponding to the large components \( u_{ik} \) and \( v_{i\ell} \). The latter contains the linear combinations of the rest of \( x_{ki} \)’s. Note that \( \hat{\Delta}_i \) is asymptotically normal by CLT since the coefficients of \( x_{ki} \)’s are small. However, \( \tilde{\Delta}_i \) may not be normal. The key idea of our strategy is to show the following recursive estimate: For any fixed \( k \in \mathbb{N} \), we have

\[
\mathbb{E}(\mathcal{Q}_i - \hat{\Delta}_i)^k e^{it\hat{\Delta}_i} = (k - 1) \tilde{Y}_i \mathbb{E}(\mathcal{Q}_i - \hat{\Delta}_i)^{k-2} e^{it\hat{\Delta}_i} + o(1), \tag{2.12}
\]

for some positive number \( \tilde{Y}_i \). Choosing \( t = 0 \), we can derive the asymptotic normality of \( \mathcal{Q}_i - \hat{\Delta}_i \) for (2.12) by the recursive moment estimate. Choosing \( t \) to be arbitrary, we can further deduce from (2.12)

\[
\mathbb{E} e^{is(\mathcal{Q}_i - \hat{\Delta}_i)} e^{it\tilde{\Delta}_i} = \mathbb{E} e^{is(\mathcal{Q}_i - \hat{\Delta}_i)} e^{it\tilde{\Delta}_i} + o(1).
\]

Then asymptotic independence between \( \mathcal{Q}_i - \hat{\Delta}_i \) and \( \tilde{\Delta}_i \) follows. Hence, we prove both the asymptotic normality and asymptotic independence from (2.12), and thus we kill two birds with one stone. The method of using the recursive estimate to get the large deviation bounds for Green function or some functional of the Green functions has been previously used in the context of the Random Matrix Theory. For instance, we refer to [48]. However, as far as we know, it is the first time to use the recursive estimate to show the normality and the independence simultaneously for the functionals of the Green functions.

Finally, we remark that the approach in this paper can also be applied to derive the distribution of the outlier eigenvectors of the spiked sample covariance matrix and the deformed Wigner matrix. We will consider these extensions in the future work (c.f. [6]).

3. Simulations and statistical applications
3.1. Numerical simulations. In this subsection, we present some numerical simulations for our results stated in Section 2.2. For simplicity, we focus on the right singular vectors. For the simulations, we consider two specific distributions for our noise matrix. We assume that \( \sqrt{n}x_{ij} \)'s are i.i.d. \( \mathcal{N}(0,1) \) or i.i.d. with the distribution \( \frac{1}{2} \delta_\sqrt{2} + \frac{1}{2} \delta_{-\sqrt{2}} \). We call these two types of noise as Gaussian noise and Two-Point noise, respectively. It is easy to check that the 3rd and 4th cumulants of the distribution \( \frac{1}{2} \delta_\sqrt{2} + \frac{1}{2} \delta_{-\sqrt{2}} \) are \( \kappa_3 = \frac{1}{\sqrt{2}} \) and \( \kappa_4 = -\frac{3}{2} \).

In the sequel, let \( e_i \) (1 \( \leq \) i \( \leq \) M) and \( f_j \) (1 \( \leq \) j \( \leq \) n) be the canonical basis vectors in \( \mathbb{R}^M \) and \( \mathbb{R}^n \) respectively. Denote by \( 1_m \) the all-one vector in \( \mathbb{R}^m \).

Assume \( S \) has rank \( r = 1 \) and admits the singular value decomposition of the form \( S = du^T v \). Set the dimension ratio \( y = M/n = 0.5 \). We present the simulations corresponding to the special cases discussed in Remark 2.5 - 2.8. The normalization of \( \sqrt{n}((\hat{v}, v)^2 - a_2(d)) \) listed in the following cases are chosen according to the calculations in Remark 2.5 - 2.8, in such a way that the asymptotic distributions are standard normal.

Case 1. Gaussian noise. Recall the discussion in Remark 2.5. In this case, the structure of the singular vectors do not play a role. We choose \( u = e_1 \) and \( v = f_1 \). Denote by

\[
\mathcal{R}_g := \frac{\sqrt{n}}{\sigma} \left( (\hat{v}, v)^2 - a_2(d) \right),
\]

where

\[
\sigma^2 = (8d^2 + 24d^{10} + 26d^8 + 20d^6 + 15d^4 + 8d^2 + 2)/(2d^4(2d^2 - 1)(d^2 + 1)^4). \tag{3.1}
\]

The conclusion is that \( \mathcal{R}_g \) is asymptotically \( \mathcal{N}(0, 1) \).

Case 2. Two-point noise and both singular vectors of \( S \) are delocalized. In the presence of Two-Point noise, the structure of the singular vectors will influence the distributions. We consider both \( u \) and \( v \) are delocalized, which corresponds to the discussion in Remark 2.7. Let \( u = 1_M/\sqrt{M} \) and \( v = 1_n/\sqrt{n} \). Then

\[
\mathcal{R}_{dt} := \frac{1}{\sigma} \left( \sqrt{n}((\hat{v}, v)^2 - a_2(d)) + \frac{d^6 - 1.5d^2 - 1}{d^5(d^2 + 1)^2} \right)
\]

is asymptotically \( \mathcal{N}(0, 1) \), where \( \sigma \) is defined in (3.1).

Case 3. Two-point noise and one of the singular vectors of \( S \) is delocalized. We set \( u = 1_M/\sqrt{M} \) and \( v = f_1 \). From Remark 2.7, we know that the random variable

\[
\mathcal{R}_{pt} := \frac{\sqrt{n}}{\sigma_t} \left( (\hat{v}, v)^2 - a_2(d) \right)
\]

is asymptotically \( \mathcal{N}(0, 1) \), where

\[
\sigma_t^2 = \sigma^2 + 2(d^4 + d^2 + 0.5)^2/(d^5(d^2 + 1)^4) - 0.75(d^4 + d^2 + 0.5)^2/(d^6(d^2 + 1)^4).
\]

Case 4. Two-point noise and both singular vectors of \( S \) are sparse (localized). Let \( u = e_1 \) and \( v = f_1 \). From the discussion in Remark 2.8, by setting

\[
\mathcal{R}_{st} := \frac{1}{\sigma_s} \left( \sqrt{n} \left( (\hat{v}, v)^2 - a_2(d) \right) - \frac{2\sqrt{n}}{d^3} X_{11} \right),
\]

with

\[
\sigma_s^2 = (d^{16} + 4d^{14} + 6d^{12} + d^{10} - 6d^8 - 2d^6 + 6.25d^4 + 6.25d^2 + 1.6875)/(d^8(d^2 + 1)^4(2d^2 - 1)),
\]

we have that \( \mathcal{R}_{st} \) is asymptotically \( \mathcal{N}(0, 1) \).

In Table 1-4, we record the probabilities for different quantiles of the empirical cumulative distributions (ECDF) of \( \mathcal{R}_g, \mathcal{R}_{dt}, \mathcal{R}_{pt}, \mathcal{R}_{st} \) respectively. We choose \( n = 200 \) or 500. For each choice of \( n \), we take \( d = 2, 3, 5, 10 \). The first column corresponds to the theoretical quantile
probabilities for a standard normal distribution. Each simulation is obtained with 10,000 repetitions. From Table 1, we observe that \( R_g \) is fairly close to standard Gaussian even for a small sample size \( n = 200 \). (The same is also observed for \( R_{dt}, R_{pt}, R_{st} \).)

In Figure 1, we plot the ECDF of of \( R_g, R_{dt}, R_{pt}, R_{st} \) in subfigures (A), (B), (C), (D) respectively, for \( n = 500 \) and various values of \( d = 2, 3, 5, 10 \). The distributions of these quantities are fairly close to the standard normal distribution.

| \( n = 200 \) | \( n = 500 \) |
|------|------|------|------|------|------|------|------|------|------|
| \( d = 2 \) | \( d = 3 \) | \( d = 5 \) | \( d = 10 \) | SE | \( d = 2 \) | \( d = 3 \) | \( d = 5 \) | \( d = 10 \) | SE |
| 0.01 | 0.012 | 0.0134 | 0.0106 | 0.0128 | 0.003 | 0.128 | 0.0115 | 0.012 | 0.0115 | 0.002 |
| 0.05 | 0.0536 | 0.0499 | 0.0466 | 0.0495 | 0.002 | 0.0525 | 0.0474 | 0.0496 | 0.0498 | 0.0014 |
| 0.10 | 0.0969 | 0.095 | 0.0909 | 0.0909 | 0.0066 | 0.0968 | 0.0975 | 0.0976 | 0.0961 | 0.003 |
| 0.30 | 0.281 | 0.280 | 0.273 | 0.268 | 0.025 | 0.292 | 0.294 | 0.275 | 0.284 | 0.014 |
| 0.50 | 0.477 | 0.472 | 0.462 | 0.463 | 0.032 | 0.486 | 0.483 | 0.480 | 0.477 | 0.020 |
| 0.70 | 0.684 | 0.679 | 0.674 | 0.670 | 0.023 | 0.691 | 0.691 | 0.683 | 0.682 | 0.013 |
| 0.90 | 0.899 | 0.899 | 0.896 | 0.901 | 0.002 | 0.898 | 0.901 | 0.898 | 0.896 | 0.002 |
| 0.95 | 0.955 | 0.955 | 0.953 | 0.953 | 0.004 | 0.953 | 0.951 | 0.952 | 0.949 | 0.002 |
| 0.99 | 0.994 | 0.993 | 0.993 | 0.992 | 0.003 | 0.991 | 0.991 | 0.992 | 0.994 | 0.002 |

Table 1. Distribution of \( R_g \): Gaussian noise.

| \( n = 200 \) | \( n = 500 \) |
|------|------|------|------|------|------|------|------|------|------|
| \( d = 2 \) | \( d = 3 \) | \( d = 5 \) | \( d = 10 \) | SE | \( d = 2 \) | \( d = 3 \) | \( d = 5 \) | \( d = 10 \) | SE |
| 0.01 | 0.011 | 0.011 | 0.013 | 0.013 | 0.002 | 0.0106 | 0.012 | 0.012 | 0.0106 | 0.001 |
| 0.05 | 0.0555 | 0.0499 | 0.049 | 0.05 | 0.001 | 0.0473 | 0.053 | 0.0486 | 0.0496 | 0.002 |
| 0.10 | 0.0873 | 0.0923 | 0.0925 | 0.096 | 0.008 | 0.0905 | 0.099 | 0.0938 | 0.0945 | 0.006 |
| 0.30 | 0.26 | 0.273 | 0.268 | 0.273 | 0.03 | 0.2645 | 0.28 | 0.274 | 0.276 | 0.03 |
| 0.50 | 0.462 | 0.469 | 0.461 | 0.466 | 0.04 | 0.46 | 0.478 | 0.47 | 0.474 | 0.03 |
| 0.70 | 0.668 | 0.665 | 0.67 | 0.68 | 0.03 | 0.6755 | 0.682 | 0.679 | 0.675 | 0.02 |
| 0.90 | 0.892 | 0.887 | 0.887 | 0.897 | 0.009 | 0.899 | 0.898 | 0.892 | 0.895 | 0.004 |
| 0.95 | 0.95 | 0.949 | 0.947 | 0.954 | 0.002 | 0.954 | 0.952 | 0.947 | 0.949 | 0.003 |
| 0.99 | 0.9914 | 0.993 | 0.9914 | 0.99 | 0.001 | 0.992 | 0.992 | 0.992 | 0.992 | 0.002 |

Table 2. Distribution of \( R_{dt} \): Two-Point noise and delocalized singular vectors.

3.2. Statistical applications. In this section, we discuss the applications of our results in Section 2.2 to the singular subspace estimation and inference.

We start with the estimation part and focus on the right singular vectors. The estimation of singular subspace is important in the recovery of low-rank matrix based on noisy observations (see for instance [17, 19, 27] and reference therein). It is clear from Lemma 4.9 that the sample singular vector is concentrated on a cone with axis parallel to the true singular vector. The aperture of the cone is determined by the deterministic function \( a_2(d) \) defined in (2.6). Further, when \( d \) increases, the sample singular vector will be closer to the true singular vector in \( l^2 \) norm. It can be seen from the results in Section 2.2, the variance of the fluctuation also decays when \( d \) increases. This phenomenon is recorded in Figure 2.

Empirically, it can be seen from Figure 2 that for a sequence of \( y \in [\frac{1}{d}, 10] \), when \( d > 5 \), the variance part is already very small and hence the fluctuation can be ignored. Further,
when \( d > 7.5 \), we can use the sample singular vector to estimate the true singular vector since their inner product is rather close to 1. Finally, note that the noise type will affect the variance of the fluctuation. Especially when the noise has negative \( \kappa_3 \) and \( \kappa_4 \), we can ignore the fluctuation for a smaller value of \( d \).

Next, we employ the statistics (1.5) to infer the singular vectors and subspace of \( S \). Such statistics have been used extensively to explore the properties of singular subspace. To name a few, in [38], the authors studied the problem of testing whether the sample singular subspace is equal to some given subspace; in [19], the authors studied the eigenvector inference problems for the correlated stochastic block models; in [37], the authors analyzed the impact of dimensionality reduction for subspace clustering algorithms; and in [17], the authors studied the high dimensional clustering problem and the canonical correlation analysis. In the aforementioned literature, the statistical inferences are based on the perturbation results of (1.5). Since we have detailed results on the distributions of the statistics (1.5), we can do the singular subspace inference based on those results in Section 2.2. Especially, we will test whether \( V \) is equal to a given matrix [62, 63, 67]. We also point out that this problem is especially important in the study of gene expression datasets [49]. For instance, in the study of lung cancer data, using the microarray gene expression data, researchers are interested in looking for genes that are significantly expressed for certain types of cancer, or that can

---

### Table 3. Distribution of \( R_{pt} \): Two-Point noise and delocalized left singular vector and sparse right singular vector.

| Normal | \( d = 2 \) | \( d = 3 \) | \( d = 5 \) | \( d = 10 \) | SE  | \( d = 2 \) | \( d = 3 \) | \( d = 5 \) | \( d = 10 \) | SE  |
|--------|-------------|-------------|-------------|-------------|-----|-------------|-------------|-------------|-------------|-----|
| 0.01   | 0.016       | 0.0151      | 0.011       | 0.0123      | 0.004| 0.011       | 0.011       | 0.011       | 0.011       | 0.001|
| 0.05   | 0.053       | 0.0513      | 0.051       | 0.0464      | 0.002| 0.051       | 0.0505      | 0.0478      | 0.0536      | 0.002|
| 0.10   | 0.0976      | 0.0968      | 0.0955      | 0.0953      | 0.004| 0.094       | 0.0959      | 0.0934      | 0.1         | 0.004|
| 0.30   | 0.273       | 0.275       | 0.270       | 0.268       | 0.03 | 0.277       | 0.283       | 0.274       | 0.282       | 0.02 |
| 0.50   | 0.468       | 0.473       | 0.469       | 0.463       | 0.03 | 0.479       | 0.481       | 0.469       | 0.47         | 0.03 |
| 0.70   | 0.686       | 0.687       | 0.677       | 0.672       | 0.02 | 0.68         | 0.68         | 0.676       | 0.674       | 0.02 |
| 0.90   | 0.9035      | 0.9025      | 0.895       | 0.897       | 0.004| 0.908       | 0.897       | 0.892       | 0.891       | 0.007|
| 0.95   | 0.959       | 0.957       | 0.954       | 0.95        | 0.005| 0.955       | 0.952       | 0.95        | 0.949       | 0.002|
| 0.99   | 0.995       | 0.991       | 0.994       | 0.993       | 0.003| 0.993       | 0.992       | 0.993       | 0.991       | 0.002|

### Table 4. Distribution of \( R_{st} \): Two-Point noise and both singular vectors sparse.

| Normal | \( d = 2 \) | \( d = 3 \) | \( d = 5 \) | \( d = 10 \) | SE  | \( d = 2 \) | \( d = 3 \) | \( d = 5 \) | \( d = 10 \) | SE  |
|--------|-------------|-------------|-------------|-------------|-----|-------------|-------------|-------------|-------------|-----|
| 0.01   | 0.0115      | 0.009       | 0.008       | 0.0825      | 0.002| 0.0099      | 0.0099      | 0.0098      | 0.0088      | 0.001|
| 0.05   | 0.0454      | 0.0448      | 0.042       | 0.0443      | 0.006| 0.0469      | 0.0468      | 0.045       | 0.044        | 0.004|
| 0.10   | 0.0873      | 0.0886      | 0.081       | 0.0864      | 0.004| 0.0908      | 0.095       | 0.091       | 0.0896       | 0.005|
| 0.30   | 0.266       | 0.270       | 0.269       | 0.275       | 0.030| 0.280       | 0.278       | 0.282       | 0.270        | 0.02 |
| 0.50   | 0.460       | 0.463       | 0.460       | 0.453       | 0.042| 0.467       | 0.473       | 0.478       | 0.463        | 0.03 |
| 0.70   | 0.666       | 0.670       | 0.660       | 0.656       | 0.037| 0.673       | 0.680       | 0.673       | 0.663        | 0.03 |
| 0.90   | 0.885       | 0.883       | 0.884       | 0.879       | 0.017| 0.890       | 0.890       | 0.894       | 0.888        | 0.009|
| 0.95   | 0.944       | 0.940       | 0.940       | 0.939       | 0.009| 0.943       | 0.943       | 0.948       | 0.948        | 0.005|
| 0.99   | 0.989       | 0.987       | 0.988       | 0.989       | 0.002| 0.989       | 0.989       | 0.99        | 0.989        | 0.001|
help clustering different types of cancer. Hence, the rows of data matrix represent the subjects clustered together by the cancer type, and the columns correspond to the genes. As a consequence, the left singular vectors are employed to study the cancer clustering and right singular vectors are used to study the gene association. We focus on the statistical inference problems on the right singular subspace and consider the hypothesis testing problems listed in [62, Section 2]. Specifically, we compare $V$ with a specified $V_0$. Such a problem is significant in many applications. For instance, we want to test whether the gene association of the data is equal to some known gene association. The testing problem can be formulated as

$$H_0 : V = V_0, \quad H_a : V \neq V_0,$$  \hspace{1cm} (T1)$$

where $V_0$ is a given matrix consisting of orthonormal vectors.

$\bullet$ z-score test with a single observation $Y$. Assuming that the singular values are known, we shall carry out the z-score test. In the following simulations, we consider the setting that the signal matrix $S$ has rank $r = 2$ with the singular values $d_1 = 5$ and $d_2 = 3$. Assume the left singular vectors of $S$ are $u_1 = \frac{1}{\sqrt{M}}1_M$ and $u_2 = \frac{1}{\sqrt{M}}(1_{M/2} \oplus (-1_{M/2}))$. Set $V_0 = (f_1, f_2)$. 

Figure 1. Plots of the ECDF of $R_g, R_{dt}, R_{pt}, R_{st}$. 
Figure 2. Mean-Variance Discussion. In both of the figures, we plot the mean function $a_2(d)$ in the upper panel for $y = 0.1, 0.5, 5, 10$ respectively for a sequence of values of $d$ lie between 3 and 13. In the lower panel, we plot the standard deviation of the fluctuation correspondingly. Recall the definitions in (2.7) and (2.8). Specially, for the Gaussian noise, the standard deviation is $\sqrt{4\theta(d)^2 + V_E(d)}$ and $\sqrt{4\theta(d)^2 + V_E(d) - 3y\theta(d)^2/(2d^2)}$ for the Two-Point noise. We choose the true right singular vector to be $f_1$ and left singular vector to be $e_1$ and hence for the Two-Point noise, we need to add a part depending on $\kappa_4 = -3/2$. This makes the variance smaller.

Recall the definitions in (2.7) and (2.8). When the noise is Gaussian, we use the statistic

$$T_{1g} = \frac{\sqrt{n}}{\sigma} \left( \sum_{i,j=1}^{2} |\langle \hat{v}_i, v_j \rangle|^2 - a_2(d_1) - a_2(d_2) \right), \quad (3.2)$$

where

$$\sigma^2 = \sum_{i=1}^{2} (4\theta(d_i)^2 + V_E(d_i)).$$

When the noise is Two-point type, we use the statistic

$$T_{1t} := \frac{\sqrt{n}}{\sigma_t} \left( \sum_{i,j=1}^{2} |\langle \hat{v}_i, v_j \rangle|^2 - a_2(d_1) - a_2(d_2) \right), \quad (3.3)$$

where

$$\sigma_t^2 = \sum_{i=1}^{2} V_E(d_i)^2 - \frac{3y}{2} \sum_{i,j=1}^{2} \frac{\theta(d_i)\theta(d_j)}{d_i d_j} + \frac{4\sqrt{y}}{\sqrt{2}} \sum_{i,j=1}^{r} \frac{\theta(d_i)\theta(d_j)}{d_i d_j}.$$
Under the nominal level $\alpha$, we will reject $H_0$ when

$$|T_{g(t)}| > z_{1-\alpha/2},$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of a standard Gaussian random variable. In Table 5, we record the type I error rates which show the accuracy of our proposed z-score test for different values of $y$ based on 10,000 simulations.

| Gaussian noise | Two-point noise |
|----------------|-----------------|
| $\alpha = 0.05$ | $\alpha = 0.1$  |
| $n = 200$  | $n = 500$  | $n = 200$  | $n = 500$  |
| $y = 0.5$  | 0.047  | 0.048  | 0.098  | 0.0967 |
| $y = 1$  | 0.057  | 0.046  | 0.092  | 0.096 |
| $y = 2$  | 0.0494  | 0.052  | 0.0984  | 0.0955 |

| Two-point noise |
|-----------------|
| $\alpha = 0.05$ | $\alpha = 0.1$ |
| $n = 200$  | $n = 500$  | $n = 200$  | $n = 500$  |
| 0.0501  | 0.0496  | 0.105  | 0.0945 |
| 0.0488  | 0.0491  | 0.097  | 0.099 |
| 0.0474  | 0.049  | 0.091  | 0.094 |

Table 5. Type I error under $H_0$ for (T1) using z-score test.

- **Student-t test with multiple i.i.d observations $Y_1, \cdots, Y_m$.** In many applications, the responses are in matrix forms [60] and hence we can observe a sequence of i.i.d matrices. Assume that we have $m$ observations $Y_k = S + X_k$, where $X_k$ are i.i.d copies of $X$. Let $\mathbf{v}^{(k)}$ be the right singular vectors of $Y_k$. In this situation, we can still carry out the testing (T1) even without the knowledge of the singular values of $S$ and underlying noise of $X$. But we assume that the rank of $S$ is known.

Without loss of generality, suppose $m$ is even and denote $N := m/2$. We consider the same simulation setting as in the z-score test. That is, the signal matrix $S$ has rank $r = 2$ with the singular values $d_1 = 5$ and $d_2 = 3$. The left singular vectors of $S$ are $\mathbf{u}_1 = \frac{1}{\sqrt{M}} 1_M$ and $\mathbf{u}_2 = \frac{1}{\sqrt{M}} (1_M \oplus (-1_M))$. The given matrix is $V_0 = (f_1, f_2)$. Let $m = 10,000$. We first define a sequence of the statistics $\{T^{(k)}_2\}_{k=1}^N$ where

$$T^{(k)}_2 := \sqrt{n} \sum_{i,j=1}^2 |\langle \mathbf{v}^{(k)}_i, \mathbf{v}_j \rangle|^2$$

for $k = 1, 2, \cdots, m$.

We further denote

$$\hat{T}^{(l)}_2 := T^{N+l}_2 - T^{(l)}_2$$

for $l = 1, 2, \cdots, N$.

By independence of $X_k$’s and Theorem 2.9, we find that $\{\hat{T}^{(l)}_2\}_{l=1}^N$ is a sequence of mean-zero Gaussian random variables. We shall use the student t-test in this situation. Our test statistic is denoted by

$$T_2 := \frac{\overline{T}_2}{\overline{S}},$$

where

$$\overline{T}_2 = \frac{1}{N} \sum_{k=1}^N \hat{T}^{(k)}_2 \quad \text{and} \quad \overline{S}^2 = \frac{1}{N-1} \sum_{k=1}^N (\hat{T}^{(k)}_2 - \overline{T}_2)^2.$$

We will reject $H_0$ if

$$|T_2| > t_{1-\alpha/2}(N-1),$$

where $t_{1-\alpha/2}(N-1)$ is the $1-\alpha/2$ quantile of a student t random variable with degree $N - 1$. In Table 6, we record the type I error rates which show the accuracy of our proposed student t-test.
Finally, to study the power of our test against the alternatives, we consider the matrices $V_\delta = (f_1, \sqrt{1-\delta^2}f_2 + \delta f_3)$ for a parameter $\delta \in (0, 1)$. In Figure 3, we record the simulated power for different values of $\delta$ under the nominal level $\alpha = 0.05$. Figure 3(A) is for the $z$-score test and Figure 3(B) is for the student $t$-test; both for the Two-Point noise. In both the $z$-score test and the student $t$-test, we find that the power of our tests increases when $\delta$ increases. Furthermore, at the same level of $\delta$, the power is improved when $n$ increases.

![Figure 3](image)

**Figure 3.** Power vs $\delta$ under the nominal level $\alpha = 0.05$ for $y = 0.5, 1, 2$ respectively.

### 4. Technical tools and Green function representations

This section is devoted to providing some basic notions and technical tools, which will be needed often in our proofs for the theorems. The basic notions are given in Section 4.1. A main technical input for our proof is the isotropic local law for the sample covariance matrix obtained in [12, 43]. It will be stated in Section 4.2. In subsection 4.3, we represent (asymptotically) $|\langle \hat{u}_i, u_j \rangle|^2$’s and $|\langle \hat{v}_i, v_j \rangle|^2$’s, and also $R$ and $L$ (c.f. (1.5)), in terms of the Green function. The discussion is based on the second author’s previous work [25], where the limits for $|\langle \hat{u}_i, u_j \rangle|^2$ and $|\langle \hat{v}_i, v_j \rangle|^2$ are studied. We then collect a few auxiliary lemmas in Section 4.4.
4.1. Basic notions. Our estimation relies on the local MP law [58] and its isotropic version [12, 43], which provide sharp large deviation estimates for the Green functions

\[ G(z) = (H - z)^{-1}, \quad G_1(z) = (XX^* - z)^{-1}, \quad G_2(z) = (X^*X - z)^{-1}. \]

Here we recall the definition in (2.2). By Schur complement, it is easy to derive

\[ G(z) = \begin{pmatrix} G_1(z) & z^{-1/2}G_1(z)X \\ z^{-1/2}X^*G_1(z) & G_2(z) \end{pmatrix}. \] (4.1)

The Stieltjes transforms for the ESD of \( XX^* \) and \( X^*X \) are defined by

\[ m_{1n}(z) = \frac{1}{M} \text{Tr} G_1(z), \quad m_{2n}(z) = \frac{1}{n} \text{Tr} G_2(z) = \frac{1}{n} \sum_{\mu=M+1}^{M+n} G_{\mu\mu}(z). \] (4.2)

It is well-known that \( m_{1n}(z) \) and \( m_{2n}(z) \) have nonrandom approximates \( m_1(z) \) and \( m_2(z) \), which are the Stieltjes transforms for the MP laws defined in (2.3) and (2.4). Specifically, for any fixed \( z \in \mathbb{C}^+ \), the following hold almost surely,

\[ m_{1n}(z) - m_1(z) \to 0, \quad m_{2n}(z) - m_2(z) \to 0. \]

Furthermore, one can easily check that \( m_1(z) \) and \( m_2(z) \) satisfy the following self-consistent equations

\[ m_1(z) + \frac{1}{z - (1 - y) + zm_1(z)} = 0, \] (4.3)

\[ m_2(z) + \frac{1}{z + (1 - y) + zm_2(z)} = 0. \] (4.4)

We can also derive the following simple relation from the definitions

\[ m_1(z) = \frac{y^{-1} - 1/z}{z} + y^{-1}m_2(z). \] (4.5)

Next we summarize some basic identities in the following lemma without proof. They can be checked from (2.3) and (2.4) via elementary calculations.

Lemma 4.1. Denote \( p \equiv p(x) \) in (2.5). For any \( x > y^{1/4} \), we have

\[ m_1(p) = \frac{-1}{x^2 + y}, \quad m_2(p) = \frac{-1}{x^2 + 1}, \]

and

\[ m'_1(p) = \frac{x^4}{(x^2 + y)^3(x^4 - y)}, \quad m'_2(p) = \frac{x^4}{(x^2 + 1)^3(x^4 - y)}. \]

Furthermore, denote by \( T(t) = tm_1(t)m_2(t) \). We have

\[ T(p) = x^{-2}, \quad T'(p) = (y - x^4)^{-1}. \]

In the sequel, we also need the following notion on high probability events.

Definition 4.2 (High probability event). We say that an \( n \)-dependent event \( E \equiv E(n) \) holds with high probability if, for any large \( \varphi > 0 \),

\[ \mathbb{P}(E) \geq 1 - n^{-\varphi}, \]

for sufficiently large \( n \geq n_0(\varphi) \).

We also adopt the notion of stochastic domination introduced in [29]. It provides a convenient way of making precise statements of the form “\( X(n) \) is bounded by \( Y(n) \) up to small powers of \( n \) with high probability”.
Definition 4.3 (Stochastic domination). Let
\[ X = (X^{(n)}(u) : n \in \mathbb{N}, u \in U^{(n)}), \quad Y = (Y^{(n)}(u) : n \in \mathbb{N}, u \in U^{(n)}), \]
be two families of nonnegative random variables, where \( U^{(n)} \) is a possibly \( n \)-dependent parameter set. We say that \( X \) is stochastically dominated by \( Y \), uniformly in \( u \), if for all small \( \epsilon > 0 \) and large \( \varphi \), we have
\[ \sup_{u \in U^{(n)}} \mathbb{P}(X^{(n)}(u) > n^\epsilon Y^{(n)}(u)) \leq n^{-\varphi}, \]
for large enough \( n \geq n_0(\epsilon, \varphi) \). In addition, we use the notation \( X = O_\prec(Y) \) if \( |X| \) is stochastically dominated by \( Y \), uniformly in \( u \). Throughout this paper, the stochastic domination will always be uniform in all parameters (mostly are matrix indices and the spectral parameter \( z \)) that are not explicitly fixed.

4.2. Isotropic local laws. The key ingredient in our estimation is a special case of the anisotropic local law derived in [43], which is essentially the isotropic local law previously derived in [12]. Set
\[ \Pi_1(z) := m_1(z)I_M \oplus m_2(z)I_n. \]
We will need the isotropic local law outside the spectrum of the MP law. For \( \lambda_+ = (1 + y^{1/2})^2 \), define the spectral domain
\[ S_o \equiv S_o(\tau, \eta) := \{ z = E + i \eta \in \mathbb{C}^+ : \lambda_+ \leq E \leq \tau, 0 \leq \eta \leq \tau^{-1} \}, \]
where \( \tau > 0 \) is a fixed small constant. Recall the notations \( m_{1n} \) and \( m_{2n} \) defined in (4.2).

Lemma 4.4 (Theorem 3.7 of [43], Theorem 3.12 of [12] and Theorem 3.1 of [58]). Fix \( \tau > 0 \), for any unit deterministic vectors \( u, v \in \mathbb{R}^{M+n} \), we have
\[ \langle u, (G(z) - \Pi_1(z))v \rangle = O_\prec \left( \sqrt{\frac{\text{Im} m_2(z)}{n \eta}} \right), \]
and
\[ |m_{2n}(z) - m_2(z)| = O_\prec \left( \frac{1}{n} \right), \quad |m_{1n}(z) - m_1(z)| = O_\prec \left( \frac{1}{n^2} \right). \]
uniformly in \( z \in S_o \).

Remark 4.5. The bounds in (4.9) cannot be directly read from any of Theorem 3.7 of [43], Theorem 3.12 of [12] or Theorem 3.1 of [58]. In all these theorems, a weaker bound \( O_\prec \left( \frac{1}{n^2} \right) \) is stated, but not only for \( z \) outside of the support of the limiting spectral distribution. Here since our parameter \( z \) can be real, we use the stronger bound \( \frac{1}{n} \) instead of \( \frac{1}{n^2} \). For \( z \in S_o \), such a bound follows from the rigidity estimates of eigenvalues in [58] and the definition of the Stieltjes transform easily. We omit the details.

Following from Lemma 4.4, by further using Cauchy’s integral formula for derivatives, we have for any given \( l \in \mathbb{N} \),
\[ \langle u, (G^{(l)}(z) - \Pi_1^{(l)}(z))v \rangle = O_\prec \left( \sqrt{\frac{\text{Im} m_2(z)}{n \eta}} \right), \]
unformly in \( z \in S_o \).
Denote by \( \kappa = |E - \lambda_+| \). We summarize some basic estimates of \( m_{1,2}(z) \) without proof.
**Lemma 4.6.** The following estimates hold uniformly in \( z \in S_\kappa \)

\[
\begin{align*}
|m_{1,2}(z)| & \sim |m_{1,2}(z)| \sim 1, \\
\text{Im} m_1(z) & \sim \text{Im} m_2(z) \sim \frac{\eta}{\sqrt{k + \eta}}. 
\end{align*}
\] (4.11) (4.12)

Given any deterministic bounded Hermitian matrix \( A \) with fixed rank, it is easy to see from Lemma 4.4 and Lemma 4.6, the spectral decomposition and (4.10) that the following estimates hold uniformly in \( z \in S_\kappa \): For any fixed \( k, \ell \in \mathbb{N} \),

\[
\begin{align*}
\max_{\mu, \nu} |(G^{(l)}(z)A)_{\mu\nu} - (\Pi^{(l)}(z)A)_{\mu\nu}| &= O_{\prec}(n^{-\frac{1}{2}}), \\
\text{Tr} G^{(l)}(z)A - \text{Tr} \Pi^{(l)}(z)A &= O_{\prec}\left(\frac{1}{\sqrt{n}}\right), \\
\max_{\mu, \nu} |(G^{(l)}(z)AG^{(l)}(z))_{\mu\nu} - (\Pi^{(l)}(z)AG^{(l)}(z))_{\mu\nu}| &= O_{\prec}\left(\frac{1}{\sqrt{n}}\right). 
\end{align*}
\] (4.13)

In our proof, we will rely on the estimates of powers of \( G \), i.e \( G^l, l = 2, 3, 4 \). We have the following lemma.

**Lemma 4.7.** We have the following recursive relation

\[
G^2 = 2G^1 + \frac{G}{z}, \quad G^3 = (G^2)' + \frac{G^2}{z}, \quad G^4 = \frac{2}{3}(G^3)' + \frac{G^3}{z}. 
\] (4.14)

**Proof.** We focus our discussion on the first identity. Differentiating \( z \) on both sides of the equation

\[ G(H - z) = I, \]

we can get that

\[ G'(H - z) + \frac{1}{2z} G(H - 2z) = 0. \]

The proof follows by multiplying \( G \) on both sides of the above equation. For \( G^3 \) and \( G^4 \), we can compute them recursively by differentiating the following two equations respectively

\[ G^2(H - z) = G, \quad G^3(H - z) = G^2. \]

This completes the proof. \( \Box \)

Recall \( \Pi_1 \) defined in (4.6) and further define

\[
\begin{align*}
\Pi_2 & := 2\Pi_1' + \frac{1}{z} \Pi_1 = (2m_1' + \frac{m_1}{z})I_M \oplus (2m_2' + \frac{m_2}{z})I_n, \\
\Pi_3 & := \Pi_1' + \frac{1}{z} \Pi_2 = (2m_1' + \frac{3m_1'}{z})I_M \oplus (2m_2' + \frac{3m_2'}{z})I_n, \\
\Pi_4 & := \frac{2}{3}\Pi_3' + \frac{1}{z} \Pi_3 = (\frac{4}{3}m_1'' + \frac{4m_1}{z} + \frac{m_1'}{z^2})I_M \oplus (\frac{4}{3}m_2'' + \frac{4m_2}{z} + \frac{m_2'}{z^2})I_n.
\end{align*}
\] (4.15) (4.16) (4.17)

With Lemma 4.7, similarly to (4.8) and (4.10), we can get the following estimates for \( l = 1, 2, 3, 4 \),

\[
\langle u, (G^l - \Pi_l)v \rangle = O_{\prec}(n^{-\frac{1}{2}}), 
\] (4.18)

uniformly in \( z \in S_\kappa \).

For brevity, in the sequel, we will use the notation

\[
\Xi_l \equiv \Xi_l(z) := G^l(z) - \Pi_l(z), \quad l \in \mathbb{N}. 
\] (4.19)
4.3. **Green function representation.** In this section, we represent (asymptotically) the terms $|\langle u_i, u_j \rangle|^2$, $|\langle v_i, v_j \rangle|^2$, $R$ and $L$ (c.f (1.5)) in terms of the Green function. The derivation relies on the results obtained in [25]. Recall $p(d)$ in (2.5) and $a_1(d)$, $a_2(d)$ in (2.6). For $i \in [r]$, define

$$h_i(x) = \frac{x^2 p'(x)p(x)}{(x + d_i)^2}$$

(4.20)

and we use the shorthand notation

$$\hat{i} = i + r.$$

To state results for the right singular vectors, we introduce a $2r \times 2r$ matrix function $W_i(x)$ for $x > 0$, which has only four non-zero entries given by

$$\begin{align*}
(W_i(x))_{ii} &= m_2^2(x), \\
(W_i(x))_{\hat{i}i} &= \frac{1}{d_i^2 x}, \\
(W_i(x))_{i\hat{i}} &= (W_i(x))_{\hat{i}i} = -\frac{m_2(x)}{d_i \sqrt{x}}
\end{align*}$$

(4.21)

We further denote the matrix function

$$M_i(x) = \mathcal{U} W_i(x) \mathcal{U}^*.$$  

(4.22)

With the above notations, we further introduce two $(M + n) \times (M + n)$ matrices

$$A_i^R = -d_i^2 \left( h_i'(d_i) M_i(p_i) + h_i(d_i) p'(d_i) M_i'(p_i) \right),$$

$$B_i^R = -d_i^2 b(d_i) p'(d_i) M_i(p_i).$$

(4.23)

In light of the definition of $\mathcal{U}$ in (2.2), we have

$$A_i^R = \begin{pmatrix}
\omega_{i1} u_i u_i^T & \omega_{i2} u_i v_i^T \\
\omega_{i3} v_i u_i^T & \omega_{i4} v_i v_i^T
\end{pmatrix},$$

$$B_i^R = \begin{pmatrix}
\varpi_{i1} u_i u_i^T & \varpi_{i2} u_i v_i^T \\
\varpi_{i3} v_i u_i^T & \varpi_{i4} v_i v_i^T
\end{pmatrix}. $$

(4.24)

Here we use the notation

$$\begin{align*}
\omega_{i1} &:= -d_i^2 \left( h_i'(d_i)(W_i(p_i))_{ii} + h_i(d_i) p'(d_i)(W_i'(p_i))_{ii} \right), \\
\omega_{i4} &:= -d_i^2 \left( h_i'(d_i)(W_i(p_i))_{\hat{i}i} + h_i(d_i) p'(d_i)(W_i'(p_i))_{\hat{i}i} \right), \\
\omega_{i2} &:= \omega_{i3} := -d_i^2 \left( h_i'(d_i)(W_i(p_i))_{i\hat{i}} + h_i(d_i) p'(d_i)(W_i'(p_i))_{i\hat{i}} \right),
\end{align*}$$

and

$$\begin{align*}
\varpi_{i1} &:= -d_i^2 h_i(d_i) p'(d_i)(W_i(p_i))_{ii}, \\
\varpi_{i4} &:= -d_i^2 h_i(d_i) p'(d_i)(W_i(p_i))_{i\hat{i}}, \\
\varpi_{i2} &:= \varpi_{i3} := -d_i^2 h_i(d_i) p'(d_i)(W_i(p_i))_{i\hat{i}}.
\end{align*}$$

We then define their counterparts for the left singular vectors. Similarly, we define a $2r \times 2r$ matrix function $T_i(x)$ for $x > 0$, which has only four non-zero entries given by

$$\begin{align*}
(T_i(x))_{ii} &= \frac{1}{d_i^2 x}, \\
(T_i(x))_{i\hat{i}} &= m_1^2(x), \\
(T_i(x))_{\hat{i}i} &= (T_i(x))_{\hat{i}i} = -\frac{m_1(x)}{d_i \sqrt{x}}.
\end{align*}$$

(4.25)

Analogously, we also define

$$N_i(x) = \mathcal{U} T_i(x) \mathcal{U}^*$$

(4.26)
and
\[ A_i^L = -d_i^2 \left( h_i'(d_i) N_i(p_i) + h_i(d_i) p'(d_i) N'_i(p_i) \right), \]
\[ B_i^L = -d_i^2 h(d_i) p'(d_i) N_i(p_i). \]

Recall the notation introduced in (4.19). We have the following lemma.

**Lemma 4.8.** Under assumptions of (1.2), (1.3), (1.6) and Assumption 2.1, we have
\[ |\langle u_i, \hat{u}_i \rangle|^2 = a_2(d_i) + \text{Tr} (\Xi_i(p_i) A_i^R) + \text{Tr} (\Xi'_i(p_i) B_i^R) + O_\prec(\frac{1}{n}), \]
\[ |\langle u_i, \hat{v}_i \rangle|^2 = a_1(d_i) + \text{Tr} (\Xi_i(p_i) A_i^L) + \text{Tr} (\Xi'_i(p_i) B_i^L) + O_\prec(\frac{1}{n}). \]

Furthermore, we have
\[ R = \sum_{i=1}^r a_2(d_i) + \sum_{i=1}^r (\text{Tr}(\Xi_i(p_i) A_i^R) + \text{Tr} (\Xi'_i(p_i) B_i^R)) + O_\prec(\frac{1}{n}), \]
\[ L = \sum_{i=1}^r a_1(d_i) + \sum_{i=1}^r (\text{Tr}(\Xi_i(p_i) A_i^L) + \text{Tr} (\Xi'_i(p_i) B_i^L)) + O_\prec(\frac{1}{n}). \]  
(4.25)
(4.26)

**Proof of Lemma 4.8.** To prove Lemma 4.8, we first need the following result from [25].

**Lemma 4.9** (Theorem 3.3 and 3.4 of [25]). Under assumptions of (1.2), (1.3), (1.6) and Assumption 2.1, for \( i, j \in [r] \), we have
\[ |\mu_i - p(d_i)| = O_\prec(n^{-\frac{1}{2}}). \]

In addition, for the singular vectors, we have
\[ |\langle u_i, \hat{u}_i \rangle|^2 - a_1(d_i) = O_\prec(n^{-\frac{1}{2}}), \quad |\langle u_i, \hat{v}_i \rangle|^2 - a_2(d_i) = O_\prec(n^{-\frac{1}{2}}), \]
and for \( 1 \leq i \neq j \leq r \),
\[ |\langle u_i, \hat{u}_j \rangle|^2 = O_\prec(\frac{1}{n}), \quad |\langle u_i, \hat{v}_j \rangle|^2 = O_\prec(\frac{1}{n}). \]

With Lemma 4.9, we can rewrite (1.5) as
\[ L = \sum_{i=1}^r |\langle \hat{v}_i, u_i \rangle|^2 + O_\prec(\frac{1}{n}) \quad \text{and} \quad R = \sum_{i=1}^r |\langle \hat{v}_i, v_i \rangle|^2 + O_\prec(\frac{1}{n}). \]  
(4.27)

We next write the above quantities in terms of the Green functions. Recall from (2.1) \( \mathcal{Y} = \mathcal{Y}(z) \) and denote by \( \hat{G}(z) = (\mathcal{Y} - z)^{-1} \). By spectral decomposition, we write
\[ \hat{G}(z) = \sum_{i=1}^{M \times N} \frac{1}{\mu_i - z} \left( \begin{array}{c} \hat{u}_i \hat{u}_i^* \vspace{1mm} \\ \sqrt{\mu_i} \hat{v}_i \hat{v}_i^* \end{array} \right) (z^{-1/2} \sqrt{\mu_i} \hat{v}_i \hat{v}_i^*). \]  
(4.28)

For any \( i \in [r] \), denote \( \Gamma_i := \partial B_p(d_i) \), where \( B_p(d_i) \) is the open disc of radius \( \rho \) around \( d_i \). Here \( \rho \) is chosen to be a small but fixed positive number such that different discs corresponding to different \( d_i \) do not have overlaps. This is achievable due to Assumption 2.1. We start with the right singular vectors. By equation (6.7) of [25], with high probability, we have the following integral representation
\[ |\langle u_i, \hat{v}_i \rangle|^2 = \frac{1}{2d_i^2 \pi} \int_{\partial B_p(d_i)} \left( (\mathcal{D}^{-1} + \mathcal{U}^* G(z) \mathcal{U})^{-1} \right)_{ii} \frac{dz}{z}. \]
Recall (4.6) and denote by
\[ \Psi(z) = -\mathcal{U}^{*} \Xi_{1}(z) \mathcal{U}. \tag{4.29} \]
Using Lemma 4.4, we have
\[ \|\Psi(z)\|_{op} = O_{\prec}(n^{-\frac{1}{2}}), \quad z \in S_{\rho}. \tag{4.30} \]
We further employ the resolvent expansion (see (6.9) in [25] for more details) to write
\[ |\langle v_{i}, \hat{u}_{i} \rangle|^{2} = \frac{1}{d_{i}^{2}} (S_{0} + S_{1}) + O_{\prec}(\frac{1}{n}), \]
where
\[ S_{0} = \frac{1}{2\pi i} \oint_{\rho(\Gamma_{i})} \left( (\mathcal{D}^{-1} + \mathcal{U}^{*} \Sigma_{1}(z) \mathcal{U})^{-1} \right) \frac{dz}{z}, \]
\[ S_{1} = \frac{1}{2\pi i} \oint_{\rho(\Gamma_{i})} \left( (\mathcal{D}^{-1} + \mathcal{U}^{*} \Sigma_{1}(z) \mathcal{U})^{-1} \Psi(z)(\mathcal{D}^{-1} + \mathcal{U}^{*} \Sigma_{1}(z) \mathcal{U})^{-1} \right) \frac{dz}{z}. \tag{4.31} \]
By the residual theorem, we have \( S_{0} = d_{i}^{2} a_{1}(d_{i}). \) Recall (4.21) and denote
\[ f_{i}(z) := -\text{Tr}(\Xi_{1}(z) \mathcal{U} \Sigma_{R}(z) \mathcal{U}^{*}). \]
We can then write
\[ S_{1} = \frac{1}{2\pi i} \oint_{\rho(\Gamma_{i})} \left( z f_{i}(z) \right) \frac{dz}{z}. \]
As \( p(d) \) is a monotone function when \( d > y^{1/4} \) and by Lemma 2.5, we find that
\[ S_{1} = \frac{d_{i}^{4}}{2\pi i} \oint_{\Gamma_{i}} \left( p(\xi) f_{i}(\xi) \xi^{4} p^{\prime}(\xi) \right) d\xi. \]
Then, by residue theorem, we obtain
\[ S_{1} = d_{i}^{4} \left( f_{i}(p(\xi)) \frac{\xi^{4} p^{\prime}(\xi)}{(d_{i} + \xi)^{2}} \right) \bigg|_{\xi = d_{i}} = d_{i}^{2} \text{Tr}(\Xi_{1}(p_{i}) A_{R}^{R}) + d_{i}^{4} \text{Tr}(\Xi_{1}(p_{i}) B_{R}^{R}), \tag{4.32} \]
where we recall (4.20) and the definitions of \( A_{R}^{R} \) and \( B_{R}^{R} \) in (4.23). The conclusion for \( |\langle v_{i}, \hat{u}_{i} \rangle|^{2} \) follows immediately.

The above discussion holds for all \( i \in [r] \). Rearranging the terms of (4.32) and using Lemma 4.1, we can conclude our proof for \( R \) using (4.27). Similar discussion yields the conclusion of \( |\langle u_{i}, \hat{u}_{i} \rangle|^{2} \) for each \( i \in [r] \) and \( L \). This completes the proof of Lemma 4.8. \( \square \)

4.4. Auxiliary lemmas. A key tool for our computation is the following cumulant expansion formula, whose proof can be found in [51, Proposition 3.1] and [41, Section II], for instance.

Lemma 4.10. Let \( \ell \in \mathbb{N} \) be fixed and let \( f \in C^{\ell+1}(\mathbb{R}) \). Let \( \xi \) be a centered random variable with finite first \( \ell + 2 \) moments. Let \( \kappa_{k}(\xi) \) be the \( k \)-th cumulant of \( \xi \). Then we have the expansion
\[ \mathbb{E}(\epsilon_{\ell}(f(\xi))) = \sum_{k=1}^{\ell} \frac{\kappa_{k+1}(\xi)}{k!} \mathbb{E}(f^{(k)}(\xi)) + \mathbb{E}(\epsilon_{\ell}(f(\xi))), \tag{4.33} \]
where \( \epsilon_{\ell}(f(\xi)) \) satisfies
\[ |\mathbb{E}(\epsilon_{\ell}(f(\xi)))| \leq C_{\ell} \mathbb{E}(|\xi|^{\ell+2}) \sup_{|t| \leq \chi} |f^{(\ell+1)}(t)| + C_{\ell} \mathbb{E}(|\xi|^{\ell+2} 1(|\xi| > \chi)) \sup_{t \in \mathbb{R}} |f^{(\ell+1)}(t)| \]
for any \( \chi > 0 \).
Note that when $\xi$ is a standard Gaussian random variable (i.e. $\kappa_i = 0, i \geq 3$), (4.33) boils down to the celebrated Stein’s lemma [66]. Next we introduce the identities on the derivatives of the Green functions in (4.1). These can be verified by elementary calculus so we omit the proofs. We use the shorthand notation

$$j' = j + M.$$ 

(4.34)

For $i \in [M]$ and $j \in [n]$, denote by $E_{ij}$ the $(M + n) \times (M + n)$ matrix with entry 1 on the $(i, j')$ position and 0 elsewhere.

**Lemma 4.11.** Let $E_{ij} = E_{ij'} + E_{j'i}$ and $k \in \mathbb{N}$. We have

$$\frac{\partial^k G}{\partial x^k_{ij}} = (-1)^k k! z^k (GE_{ij})^k G,$$

$$\frac{\partial^k (G^2)}{\partial x^k_{ij}} = (-1)^k k! z^k \sum_{s=0}^k (GE_{ij})^s G (GE_{ij})^{k-s} G.$$ 

It is convenient to introduce the following notion of convergence in distribution.

**Definition 4.12** ([42, Definition 7.3]). Two sequences of random variables, $\{X_n\}$ and $\{Y_n\}$, are asymptotically equal in distribution, denoted as $X_n \simeq Y_n$, if they are tight and satisfy

$$\lim_{n \to \infty} (\mathbb{E} f(X_n) - \mathbb{E} f(Y_n)) = 0$$

for any bounded continuous function $f$.

Next, we collect some basic results on convergence and equivalence in distribution for sum of random variables. They can be found in [42, Lemma 7.7, 7.8 and 7.10].

**Lemma 4.13.** (1). Let $X_n \simeq Y_n$ and $R_n$ satisfy $\lim_{n \to \infty} \mathbb{P}(|R_n| \leq \epsilon_n) = 1$, where $\{\epsilon_n\}$ is a positive null sequence. Then

$$X_n \simeq Y_n + R_n.$$ 

(2). Let $\{X_n\}, \{X_n'\}, \{Y_n\}$ and $\{Y_n'\}$ be sequences of random variables. Suppose that $X_n \simeq X_n', Y_n \simeq Y_n'$, $X_n$ and $Y_n$ are independent, and $X_n'$ and $Y_n'$ are independent. Then

$$X_n + Y_n \simeq X_n' + Y_n'.$$ 

(3). Let $\{Z_n\}$ be a bounded deterministic sequence. Let $\{X_n\}$ be random variables such that $X_n$ converges weakly to $X$. Then for any bounded continuous function $f$, we have

$$\mathbb{E} f(Z_n X_n) - \mathbb{E} f(Z_n X) \to 0,$$

as $n \to \infty$.

The following notation from [42, Definition 7.11] will be convenient for us when we replace random variables with their i.i.d copies.

**Definition 4.14.** Let $\{\sigma_n\}$ be a sequence of bounded positive numbers. If $X_n$ and $Y_n$ are independent random variables with $Y_n \simeq \mathcal{N}(0, \sigma_n^2)$, and if $S_n \simeq X_n + Y_n$, we write

$$S_n \simeq X_n + \mathcal{N}(0, \sigma_n^2).$$
5. Proof of Theorem 2.3 and 2.10

We focus our discussion on the right singular vectors $\mathbf{v}$ and $\hat{\mathbf{v}}$; the proof for the left counterpart is analogous. For brevity, in this section, we omit the subindices of $d_i, \mathbf{u}_i, \hat{\mathbf{u}}_i$ and write $d, \mathbf{u}, \mathbf{v}, \hat{\mathbf{u}}, \hat{\mathbf{v}}$ instead. Similarly, we write the matrices $A^R$ and $B_i^R$ (c.f. (4.23)) as $A$ and $B$, respectively. We also write $m_{1,2}(z)$ as $m_{1,2}$ for brevity.

By Lemma 4.8, we can reduce the problem to study

$$Q \equiv Q(z) := \sqrt{n} \left( \text{Tr}(\Xi_1(z)A) + \text{Tr}(\Xi'_1(z)B) \right),$$

(5.1)

at $z = p(d)$ (c.f. (2.5)).

In the sequel, we will prove the limiting distribution of $Q(z)$ at $z = p(d)$. The key task is to prove Proposition 5.1 below. In this section, we will show that Theorem 2.3 follows from Proposition 5.1. Let index $i \in [M]$ and $j \in [n]$. Recall the notation in (4.34). For short, we also write

$$\sum_{i,j} = \sum_{i=1}^{M} \sum_{j=1}^{n}.$$ 

In order to state Proposition 5.1, we first introduce some notations. For a fixed small constant $\nu > 0$, denote by

$$B(\nu) := \left\{ (i,j) \in [M] \times [n] : |u(i)| > n^{-\nu}, |v(j)| > n^{-\nu} \right\},$$

the set of the indices of those components with large magnitude. Since $\mathbf{u}$ and $\mathbf{v}$ are unit vectors, we have $|B(\nu)| \leq C n^4 \nu$ for some constant $C > 0.$ Let $\mathcal{S}(\nu)$ be the complement of $B(\nu)$, i.e.,

$$\mathcal{S}(\nu) = ([M] \times [n]) \setminus B(\nu).$$

(5.2)

For brevity, we introduce the notation

$$P(\alpha_1, \ldots, \alpha_m),$$

(5.3)

to represent the set of all the permutations of $(\alpha_1, \ldots, \alpha_m)$, where $\alpha_i$‘s can be alike. Recall (4.6) and (4.15). We set the deterministic quantity

$$\Delta_d \equiv \Delta_d(z) := -\frac{\kappa_2 z^{3/2}}{n} \sum_{i,j} \left( (\Pi_1)_{ii} (\Pi_1)_{jj'} (2(\Pi_1A\Pi_1)_{ij} + (\Pi_1B\Pi_1')_{ij} + (\Pi_1' B\Pi_1)_{ij}) \right. \left. + \frac{1}{2} \sum_{(a_1, a_2, a_3) \in P(2, 1, 1)} (\Pi_{a_1})_{ii} (\Pi_{a_2})_{jj'} ((\Pi_1 B\Pi_{a_3})_{ij} + (\Pi_{a_3} B\Pi_1)_{ij'}) \right),$$

(5.4)

and the random variable

$$\Delta_r \equiv \Delta_r(z) := \sqrt{n} \sum_{(i,j) \in \mathcal{B}(\nu)} x_{ij} c_{ij},$$

(5.5)

where

$$c_{ij} \equiv c_{ij}(z) := \sum_{l_1, l_2 \in \{i,j\}} \left( (\Pi_1 A\Pi_1)_{l_1l_2} - \frac{1}{2z} (\Pi_1 B\Pi_1)_{l_1l_2} \right. \left. + \frac{1}{2} (\Pi_1 B\Pi_2)_{l_1l_2} + \frac{1}{2} (\Pi_2 B\Pi_1)_{l_1l_2} \right).$$

(5.6)
Define the $M \times n$ matrix function $S \equiv S(z) = (s_{ij})$ with

$$s_{ij} \equiv s_{ij}(z) := \sum_{l_1, \ldots, l_4 \in \{i, j\} \setminus \{i, j\}} \left( (\Pi_1 \Lambda \Pi_1)_{l_1 l_2} (\Pi_1)_{l_3 l_4} - \frac{1}{2z} (\Pi_1 B \Pi_1)_{l_1 l_2} (\Pi_1)_{l_3 l_4} \right) + \frac{1}{2} \sum_{(a_1, a_2, a_3) \in \mathcal{P}(2, 1, 1)} (\Pi_{a_1} B \Pi_{a_2})_{l_1 l_2} (\Pi_{a_3})_{l_3 l_4}. \quad (5.7)$$

We also define a function

$$V \equiv V(z) := \mathcal{V}^E(z) + 2 \frac{\kappa_1^2}{\sqrt{n}} \sum_{(i, j) \in \mathcal{S}(\nu)} c_{ij} s_{ij} + \frac{\kappa_2^2}{n} z \sum_{(i, j) \in \mathcal{S}(\nu)} s_{ij}^2 + \sum_{(i, j) \in \mathcal{S}(\nu)} c_{ij}^2, \quad (5.8)$$

where

$$\mathcal{V}^E \equiv \mathcal{V}^E(z) := -\sqrt{z} \sum_{\alpha = 1, 2} (m_{\alpha} a_{1\alpha} + \frac{m_{\alpha}}{2} \beta_{1\alpha} + m_{\alpha} b_{1\alpha}). \quad (5.9)$$

Here we refer to (6.9) for the definitions of $a_{1\alpha}$, $b_{1\alpha}$ and $\beta_{1\alpha}$ for $\alpha = 1, 2$.

With $\Delta_d$ and $\Delta_r$ defined in (5.4) and (5.5), we further introduce the notation

$$\Delta \equiv \Delta(z) := \Delta_r(z) + \Delta_d(z) \quad (5.10)$$

and define

$$Q \equiv Q(z) := Q(z) - \Delta(z). \quad (5.11)$$

**Proposition 5.1.** Under the assumptions of Theorem 2.3, we have that $Q(p_i)$ and $\Delta(p_i)$ are asymptotically independent. Furthermore,

$$Q(p_i) \simeq \mathcal{N}(0, V(p_i)). \quad (5.12)$$

We first show how Proposition 5.1 implies Theorem 2.3.

**Proof of Theorem 2.3.** By Lemma 4.8 and (5.1),

$$\sqrt{n} \left( \| (x_i, \bar{x}_i) \|^2 - a_2(d_i) \right) = Q(p_i) + O_{\prec}(n^{-\frac{1}{2}}).$$

Here $Q(p_i)$ is defined in (5.1) with $(A, B) = (A^R, B^R)$ (c.f.(4.23)). By Proposition 5.1, we have that at $z = p_i$,

$$Q = \Delta_d + \Delta_r + Q \simeq \Delta_d + \sqrt{n} \sum_{(i, j) \in \mathcal{B}(\nu)} x_{ij} c_{ij} + \mathcal{N}(0, V).$$

Next, by Central Limit Theorem and Lemma 4.13, one has

$$\sqrt{n} \sum_{i, j} x_{ij} c_{ij} \simeq \sqrt{n} \sum_{(i, j) \in \mathcal{B}(\nu)} x_{ij} c_{ij} + \mathcal{N}(0, \mathbf{z} \sum_{(i, j) \in \mathcal{S}(\nu)} (c_{ij})^2).$$

Furthermore, by the definition of $\mathcal{S}(\nu)$, we notice that

$$n^{-1/2} \sum_{(i, j) \in \mathcal{S}(\nu)} c_{ij} s_{ij} = n^{-1/2} \sum_{i, j} c_{ij} s_{ij} + O(n^{-\frac{1}{2}} + 4\nu).$$

Let $C(z) = (c_{ij}(z))$ with $c_{ij}(z)$ defined in (5.6) and recall $S(z)$ from (5.7). Using Lemma 4.13, we conclude that

$$Q(p_i) \simeq \Delta_d(p_i) + \sqrt{n} \mathbf{p}_i \operatorname{Tr}(X^* C(p_i)) + \mathcal{N}(0, V(p_i)).$$
where
\[ V(p_i) = V^E(p_i) + 2\frac{\kappa_4 p_i^{3/2}}{\sqrt{n}} \text{Tr}(C(p_i)^*S(p_i)) + \frac{\kappa_4 p_i^2}{n} \text{Tr}(S(p_i)^*S(p_i)). \]

Denote
\[ \Delta_i = \sqrt{np_i} \text{Tr}(X^*C(p_i)) + \Delta_d(p_i) \]
and
\[ Z_i \sim \mathcal{N}(0, V(p_i)), \]
which is independent of \( \Delta_i \). Next, plugging \( z = p_i \) into (5.4), (5.6), (5.7), using Lemma 4.1 and taking into account the definitions of \( A_i^R, B_i^R \) in (4.23), we find that
\[ \Delta_i = -\sqrt{n} - \frac{2(d_i^4 + 2y d_i^2 + y)}{d_i^4(d_i^2 + 1)^2} u_i^* X v_i - \frac{2(d_i^6 - 3y d_i^2 - 2y)}{d_i^4(d_i^2 + 1)^2} \left( \frac{\kappa_3}{n} \sum_{k,l} u_i(l) v_i(l) \right). \]

The variance \( V(p_i) \) is the sum of
\[
2 \frac{\kappa_4}{\sqrt{n}} p_i^{3/2} \text{Tr}(C(p_i)^*S(p_i)) \frac{\kappa_4 p_i^2}{n} \text{Tr}(S(p_i)^*S(p_i)) = - \frac{4(d_i^4 + 2y d_i^2 + y)(d_i^6 - 3y d_i^2 - 2y)}{d_i^4(d_i^2 + 1)^4} \left( \frac{\kappa_3}{n} \sum_{k,l} u_i(k) v_i(l) \right)
\]
\[ + \frac{4(d_i^4 + 2y d_i^2 + y)^2}{d_i^8(d_i^2 + 1)^4} \left( \frac{\kappa_3}{n} \sum_{k,l} u_i(k) v_i(l) \right)^2
\]
\[ + \frac{(d_i^6 - 3y d_i^2 - 2y)^2}{d_i^8(d_i^2 + 1)^4} \left( \frac{\kappa_4}{n} \sum_{k} u_i(k) \right)^2 + \frac{(d_i^4 + 2y d_i^2 + y)^2}{d_i^8(d_i^2 + 1)^4} \left( \frac{\kappa_4 y_i}{n} \sum_{l} v_i(l) \right)^2 \]
and
\[ V^E(p_i) = \frac{2}{d_i^2 - y} \left( \frac{2y(y + 1)(d_i^4 + 2y d_i^2 + y)}{d_i^4(d_i^2 + 1)^2} - \frac{y(y - 1)(5y + 1)}{d_i(d_i^2 + 1)^2} \right)
\]
\[ + \frac{(d_i^2 + y)(d_i^2 + y)^2}{d_i^2(d_i^2 + 1)^2} \left( \frac{d_i^6 - 3y d_i^2 - 2y}{d_i^3(d_i^2 + 1)^2} + \frac{2y^2(y - 1)^2}{d_i^2(d_i^2 + 1)^2} \right). \]
The last expression is obtained using the definitions of \( a_{1\alpha}, b_{1\alpha} \) and \( \hat{b}_{1\alpha} \), for \( \alpha = 1, 2 \) in (6.9) and performing tedious yet elementary calculations. Recall (2.7). The conclusion of Theorem 2.3 follows immediately by rewriting \( \Delta_i \) and \( V(p_i) \) in terms of \( \theta(d_i) \) and \( \psi(d_i) \). \( \square \)

The rest of this section is devoted to the proof of Proposition 5.1. Our proof relies on the cumulant expansion in Lemma 4.10, where we need to control the expectation. Throughout the proof, we will frequently use the estimates in (4.13). These estimates hold with high probability, which do not yield bounds for the expectations directly. In order to translate the high probability bounds into those for the expectations, one needs a crude deterministic bound for the Green function on the bad event with tiny probability. To this end, we will work with a slight modification of the real \( z = p(d) \) for Green function. Specifically, in the proof of the following Proposition 5.2, we will also use the parameter
\[ z = p(d) + \ln C, \tag{5.13} \]
for a large constant \( C \). On the bad event, we will use the naive bound of the Green function \( ||G|| \leq N^C \), which will be compensated by the tiny probability of the bad event. At the end, by the continuity of \( G(\bar{z}) \) at \( \bar{z} \) away from the support of the MP law, it is (asymptotically)
equivalent to work with \( (5.13) \), for the proof of Proposition 5.1. We first claim that it suffices to establish the following recursive estimate.

**Proposition 5.2.** Suppose the assumptions of Theorem 2.3 hold. Let \( z_0 = p(d) \) and \( z_0 \) be defined in \( (5.13) \). We have

\[
\mathbb{E}Q(z)e^{it\Delta(z_0)} = O_{\prec}(n^{-\frac{1}{2}+4\nu}),
\]

and for any fixed integer \( k \geq 2 \),

\[
\mathbb{E}Q^k(z)e^{it\Delta(z_0)} = (k-1)V\mathbb{E}Q^{k-2}(z)e^{it\Delta(z_0)} + O_{\prec}(n^{-\frac{1}{2}+4\nu}).
\]

The proof of Proposition 5.2 is our main technical task, which will be stated in Section 6. Now we first show the proof of Proposition 5.1 based on Proposition 5.2.

**Proof of Proposition 5.1.** Recall the following elementary bound, for any \( x \in \mathbb{R} \) and sufficiently large \( N \in \mathbb{N} \), we have

\[
\left| e^{ix} - \frac{N}{k!} \right| \leq \min \left\{ \frac{|x|^N}{(N+1)!}, \frac{2|x|^N}{N!} \right\}.
\]

First, we write

\[
Q(z) = Q_R(z) + iQ_I(z),
\]

where \( Q_R(z) \) and \( Q_I(z) \) stand for the real and imaginary parts of \( Q(z) \) respectively. According to the choice of \( z \) in \( (5.13) \), we have the deterministic bound \( |Q_I(z)| \leq N^C \) for some large positive constant \( C \). Moreover, by continuity of the Green function and the Stieltjes transform, one can easily check that \( |Q_I(z)| \leq N^{-C} \) for some large positive constant \( C' \) with high probability. Using the small bound \( N^{-C} \) on the high probability event and the large deterministic bound \( N^C \) on the tiny probability event, one can easily derive from \( (5.14) \) and \( (5.15) \) that

\[
\mathbb{E}Q_R(z)e^{it\Delta(z_0)} = O_{\prec}(n^{-\frac{1}{2}+4\nu}),
\]

\[
\mathbb{E}Q_R^k(z)e^{it\Delta(z_0)} = (k-1)V\mathbb{E}Q_R^{k-2}(z)e^{it\Delta(z_0)} + O_{\prec}(n^{-\frac{1}{2}+4\nu}).
\]

For any \( s, t \in \mathbb{R} \), by \( (5.16) \), we have

\[
\mathbb{E}e^{isQ_R(z)+it\Delta(z_0)} = \sum_{k=0}^{2N-1} \frac{(is)^k}{k!} \mathbb{E}Q_R^k(z)e^{it\Delta(z_0)} = O \left( \frac{s^{2N}}{(2N)!} \mathbb{E}Q_R^N(z) \right).
\]

For the error term on the right side of \( (5.19) \), using \( (5.18) \) recursively for \( t = 0 \), we first find

\[
\mathbb{E}Q_R^N(z) = (2N-1)!!V^N + O_{\prec}(n^{-\frac{1}{2}+4\nu}).
\]

Thus, for arbitrarily small \( \epsilon > 0 \), by taking \( N \) sufficiently large, we have

\[
\frac{(2N-1)!!V^N}{(2N)!} < \epsilon
\]

and it follows that

\[
\left| \mathbb{E}e^{isQ_R(z)+it\Delta(z_0)} - \sum_{k=0}^{2N-1} \frac{(is)^k}{k!} \mathbb{E}Q_R^k(z)e^{it\Delta(z_0)} \right| < \epsilon + O_{\prec}(n^{-\frac{1}{2}+4\nu}).
\]

Using \( (5.18) \), we get the following estimate

\[
\sum_{k=0}^{2N-1} \frac{(is)^k}{k!} \mathbb{E}Q_R^k(z)e^{it\Delta(z_0)} = \sum_{k=0}^{N-1} \frac{(is)^{2k}}{(2k)!!} \mathbb{E}Q_R^k(z)e^{it\Delta(z_0)} + O_{\prec}(n^{-\frac{1}{2}+4\nu}).
\]
Next, combing (5.21) with the fact
\[ \exp \left( \frac{x^2}{2} \right) = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!!}, \]
together with (5.20), we conclude that
\[ \left| E e^{iQ(z) + t\Delta(z)} - e^{-\frac{1}{2}V_s} e^{it\Delta(z)} \right| < 2\epsilon + O\prec(n^{-\frac{1}{2}} + 4\nu). \quad (5.22) \]
The asymptotic independence of \( Q_R(z) \) and \( \Delta(z_0) \) is a consequence of (5.22) and the fact \( \epsilon \) is arbitrarily small. (5.12) can be proved by setting \( s = 0 \).

Although Proposition 5.2 is proved under the choice (5.13), by continuity of \( G \) outside of the support of MP law, we know \( Q(z_0) = Q_R(z) + O(n^{-C'}) \) with high probability for some positive constant \( C' \). This concludes the proof of Proposition 5.1.

At the end, we claim that the proof of Theorem 2.10 is analogous.

**Proof of Theorem 2.10.** By considering \( Y^* \) instead of \( Y \), the proof of Theorem 2.3 applies to the right singular vectors of \( Y^* \), which are the left singular vectors of \( Y \). Hence, we conclude the proof of Theorem 2.10.

6. **Proof of Proposition 5.2**

This section is devoted to the proof of Proposition 5.2. In Proposition 5.2, we choose different parameters, \( z \) and \( z_0 \), for \( Q \) and \( \Delta \), separately. However, for brevity, we will omit both two parameters for simplicity in the sequel.

First of all, applying (4.13) to the definition in (5.1), we have
\[ Q = O_{\prec}(1). \quad (6.1) \]

Denote \((M + n) \times (M + n)\) diagonal matrices
\[ \Gamma_u := \begin{pmatrix} I_M \\ 0 \end{pmatrix} \quad \text{and} \quad \Gamma_l := \begin{pmatrix} 0 \\ I_n \end{pmatrix}. \quad (6.2) \]
We further define \( A_1 = A\Gamma_u \), \( A_2 = A\Gamma_l \) and define \( B_1, B_2 \) analogously. In addition, we set
\[
\begin{align*}
f_\alpha &:= -m_\alpha \text{Tr} H \Xi_1 A_\alpha + (1 + zm_\alpha) \text{Tr} G A_\alpha, \\
g_\alpha &:= -m_\alpha \text{Tr} H \Xi_2 B_\alpha + \frac{1 + zm_\alpha}{2} \text{Tr} G^2 B_\alpha + \frac{zm_\alpha - 1}{2z} \text{Tr} GB_\alpha \\
 &\quad - m'_\alpha \text{Tr} B_\alpha + m'_\alpha \text{Tr} H \Pi_1 B_\alpha, \quad \alpha = 1, 2. \quad (6.3)
\end{align*}
\]

The proof of Proposition 5.2 is based on the following two lemmas.

**Lemma 6.1.** Recall (5.4) and (5.5). For \( z \) defined in (5.13), we have
\[ Q = \sqrt{n} \left( f_1 + f_2 + g_1 + g_2 \right) + \sqrt{nm} \sum_{(i,j) \in S(\nu)} c_{ij} x_{ij} - \Delta_d. \quad (6.4) \]

To state the second crucial lemma, Lemma 6.2. We first introduce some notations. Recall that \( \Pi_a \) \((1 \leq a \leq 4)\) in (4.6) and (4.15) approximates \( G^a \). We introduce the following matrices to approximate the powers of \( G \) interacting with block diagonal matrices \( \Gamma_u \) and \( \Gamma_l \). For \( 1 \leq a_1, a_2 \leq 2 \), define
\[ \Pi_{a_1, a_2}^u := \Pi_{a_1} \Gamma_u \Pi_{a_2} \quad \text{and} \quad \Pi_{a_1, a_2}^l := \Pi_{a_1} \Gamma_l \Pi_{a_2}. \quad (6.5) \]
Note that they approximate $G^{a_1} \Gamma^u \Pi_{a_2}$ and $G^{a_2} \Gamma^u \Pi_{a_2}$ respectively. We further define

$$\Pi_2^i := m'_i I_M \oplus (m''_i + \frac{1}{z} m_2) I_n \quad \text{and} \quad \Pi_2^1 := (m'_1 + \frac{1}{z} m_1) I_M \oplus m'_2 I_n, \quad (6.6)$$

which approximate $G \Gamma^u G$ and $G \Gamma^1 G$.

We need to introduce more notations. The first set of notations will show up in the calculation of $\Delta Q$, which is the mean value of $Q$. We set

$$\tilde{\phi}_1^a := \frac{2z}{n} \sum_{i,j} (\Pi_1)_{ii}(\Pi_1)_{jj}'(\Pi_1 A_1)_{ij}', \quad \tilde{\phi}_1^b := \frac{2z}{n} \sum_{i,j} (\Pi_1)_{ii}(\Pi_1)_{jj}'(\Pi_1 A_2)_{ij}',
$$

$$\tilde{\phi}_2^a := \frac{2z}{n} \sum_{(a_1, a_2, a_3) \in P(2,1,1)} \sum_{i,j} (\Pi_{a_1})_{ii}(\Pi_{a_2})_{jj}'(\Pi_{a_3} B_1)_{ij}', \quad \tilde{\phi}_2^b := \frac{2z}{n} \sum_{(a_1, a_2, a_3) \in P(2,1,1)} \sum_{i,j} (\Pi_{a_1})_{ii}(\Pi_{a_2})_{jj}'(\Pi_{a_3} B_2)_{ij}'. \quad (6.7)$$

And $\tilde{\phi}_1^a$ (resp. $\tilde{\phi}_1^b$) is defined by replacing $A_1$ (resp. $A_2$) to $B_1$ (resp. $B_2$) in the expression of $\tilde{\phi}_1^a$ (resp. $\tilde{\phi}_1^b$). Using (6.2), we further set

$$\Pi_3^a := (m''_1 + \frac{1}{z} m'_1) I_M \oplus (m''_2 + \frac{2}{z} m'_2) I_n,$n
$$

$$\Pi_3^b := (m''_1 + \frac{2}{z} m'_1) I_M \oplus (m''_2 + \frac{1}{z} m'_2) I_n,$n
$$

$$\Pi_4^a := (\frac{2}{3} m''_1 + \frac{2}{z} m'_1) I_M \oplus (\frac{2}{3} m''_2 + \frac{2}{z} m'_2) I_n, \quad \Pi_4^b := (\frac{2}{3} m''_1 + \frac{2}{z} m'_1) I_M \oplus (\frac{2}{3} m''_2 + \frac{2}{z} m'_2) I_n. \quad (6.8)$$

The next set of notations will appear in the derivation of the variance of $Q$. We denote

$$a_{11} := -(k-1) \sqrt{\frac{z}{n}} \left(2 \text{Tr}(\Pi_2^1 - \Pi_1 ^1) A_1 \Pi_1 A - \frac{1}{z} \text{Tr}(\Pi_2^1 - \Pi_1 ^1) A_1 \Pi_1 B \right.
$$

$$\left. + \text{Tr}(\Pi_1^1 - \Pi_1 ^1) A_1 \Pi_2 B + \text{Tr}(\Pi_3^1 - \Pi_1 ^1) A_1 \Pi_1 B \right),
$$

$$b_{11} := -(k-1) \sqrt{\frac{z}{n}} \left(2 \text{Tr}(\Pi_2^1 - \Pi_1 ^1) B_1 \Pi_1 A - \frac{1}{z} \text{Tr}(\Pi_2^1 - \Pi_1 ^1) B_1 \Pi_1 B \right.
$$

$$\left. + \text{Tr}(\Pi_3^1 - \Pi_1 ^1) B_1 \Pi_2 B + \text{Tr}(\Pi_4^1 - \Pi_1 ^1) B_1 \Pi_1 B \right). \quad (6.9)$$

In addition, $a_{12}$ is defined via replacing $A_1$ with $A_2$ and $\Pi_1 ^1, \Pi_2 ^1, a_2$ with $\Pi_1 ^a, \Pi_2 ^a, a_2$ in the definition of $a_{11}$. We further define $b_{12}$ (resp. $b_{12}$) via replacing $A_1$ (resp. $A_2$) with $B_1$ (resp. $B_2$) in the definition of $a_{11}$ (resp. $a_{12}$). Similarly, $b_{12}$ is obtained by replacing $B_1$ with $B_2$ and $\Pi_1 ^a, \Pi_2 ^a, b_2$ with $\Pi_1 ^a, \Pi_2 ^a, a_2$ in the definition of $b_{11}$.

Next, recall $c_{ij}$ defined in (5.6) and set

$$a_{21} := -(k-1) \frac{z}{\sqrt{n}} \sum_{(i,j) \in \mathcal{S}(\nu)} (\Pi_1)_{ij}'(\Pi_1 A_1)_{ii} c_{ij},$$

$$b_{21} := -(k-1) \frac{z}{\sqrt{n}} \sum_{(i,j) \in \mathcal{S}(\nu)} (\Pi_1)_{ij}'(\Pi_2 B_1)_{ii} + (\Pi_2)_{ij}'(\Pi_1 B_1)_{ii} c_{ij}. \quad (6.10)$$
Further, $a_{22}$ (resp. $\tilde{b}_{22}$) is defined by replacing $(A_1)_{ii}$ (resp. $(B_1)_{ii}$) with $(A_2)_{j',j'}$ (resp. $(B_2)_{j',j'}$) in the definition of $a_{21}$ (resp. $b_{21}$). Then we recall $s_{ij}$ in (5.7) and set

$$a_{31} := -\frac{2(k - 1)\varepsilon^{3/2}}{n} \sum_{i,j} (\Pi_1)_{j',j'}(\Pi_1 A_1)_{ii} s_{ij},$$

$$\tilde{b}_{31} = -\frac{2(k - 1)\varepsilon^{3/2}}{n} \sum_{i,j} (\Pi_1)_{j',j'}(\Pi_2 B_1)_{ii} + (\Pi_2)_{j',j'}(\Pi_1 B_1)_{ii} s_{ij}. \quad (6.11)$$

Further, $a_{32}$ (resp. $\tilde{b}_{32}$) is defined via replacing $(A_1)_{ii}$ (resp. $(B_1)_{ii}$) with $(A_2)_{j',j'}$ (resp. $(B_2)_{j',j'}$) in the definition of the $a_{31}$ (resp. $b_{31}$). Also, $b_{31}$ (resp. $b_{32}$) is defined by replacing $A_1$ (resp. $A_2$) with $B_1$ (resp. $B_2$) in the definition of $a_{31}$ (resp. $a_{32}$).

For $\alpha = 1, 2$, we further write

$$a_{\alpha 0} := a_{10} + \kappa_1 a_{20} + \frac{\kappa_1}{2} a_{30},$$

$$b_{\alpha 0} := \frac{m_0}{2} b_0 + m_0' b_1 + \frac{\kappa_1}{2} m_0' b_2 + \frac{\kappa_1 m_0}{4} b_3 + \frac{\kappa_1 m_0'}{2} b_3. \quad (6.12)$$

For brevity, we also adopt the notation

$$q^{(j)} = Q^j(z)e^{it\Delta(z_0)}.$$ 

Recall the notations in (6.3). With the above notations, we now state the following lemma.

**Lemma 6.2.** Under the assumptions of Theorem 2.3, we have for $\alpha = 1, 2$,

\[
\sqrt{n}E f_{\alpha} q^{(k-1)} = -\sqrt{\varepsilon} m_0 E \left( \frac{\kappa_3}{2} \delta_{\alpha} q^{(k-2)} + a_{\alpha 0} q^{(k-2)} \right) + O_{\omega}(n^{-\frac{1}{4} + 4\nu}), \quad (6.13)
\]

\[
\sqrt{n}E g_{\alpha} q^{(k-1)} = -\sqrt{\varepsilon} m_0 E \left( \frac{\kappa_3}{4} \left( m_{\alpha} + 2 m_{\alpha}' \right) q^{(k-1)} + b_{\alpha 0} q^{(k-2)} \right) + O_{\omega}(n^{-\frac{1}{4} + 4\nu}), \quad (6.14)
\]

In addition, we also have

\[
\sqrt{n} \sum_{(i,j) \in \mathcal{S}(v)} c_{ij} E x_{ij} q^{(k-1)} = (k - 1) \left( \sum_{(i,j) \in \mathcal{S}(v)} c_{ij}^2 + \frac{\varepsilon^{3/2} \kappa_3}{\sqrt{n}} \right) \sum_{(i,j) \in \mathcal{S}(v)} s_{ij} c_{ij} q^{(k-2)} + O_{\omega}(n^{-\frac{1}{4} + 4\nu}). \quad (6.15)
\]

With Lemmas 6.1 and 6.2, we can now prove Proposition 5.2.

**Proof of Proposition 5.2.** By simply combining Lemma 6.1 and 6.2, we can write

\[
E q^{(k)} = c_1 E q^{(k-1)} + c_2 E q^{(k-2)} - \Delta_q E q^{(k-1)} + O_{\omega}(n^{-\frac{1}{4} + 4\nu}),
\]

where

\[
c_1 = -\sqrt{\varepsilon} \sum_{\alpha = 1} \left( \frac{1}{2} m_{\alpha} \delta_{\alpha} + \frac{\kappa_3 m_{\alpha}}{4} \delta_{\alpha} + \frac{\kappa_4 m_{\alpha}}{2} b_{10} + m_{\alpha} \tilde{b}_{10} + m_{\alpha}' b_{10} \right),
\]

\[
c_2 = -\sqrt{\varepsilon} \sum_{\alpha = 1} \left( m_{\alpha} a_{10} + \kappa_3 m_{\alpha} a_{20} + \frac{\kappa_4 m_{\alpha}}{2} a_{30} + \frac{\kappa_4 m_{\alpha}}{2} \beta_{10} + m_{\alpha}' b_{10} \right)
\]

\[+ \frac{\kappa_3 m_{\alpha}}{2} b_{20} + \kappa_3 m_{\alpha}' b_{20} + \frac{\kappa_4 m_{\alpha}}{4} b_{30} + \frac{\kappa_4 m_{\alpha}'}{2} b_{30}. \]

Also recall $\Delta_q$ from (5.4) and $V$ from (5.8). By substituting the definitions of the notations in (6.7), (6.9), (6.10), (6.11), and also their analogues, it is elementary to check

\[
c_1 = \Delta_d, \quad c_2 = V. \quad (6.16)
\]
This completes the proof of (5.15). Further we can regard (5.14) as a degenerate case of (5.15). The proof can be done in the same way. We thus conclude the proof of Proposition 5.2.

Therefore, what remains is to prove Lemmas 6.1 and 6.2. We prove Lemma 6.1 in the rest of this section, and state the proof of Lemma 6.2 in Section 7.

Proof of Lemma 6.1. Recall from (5.10) and (5.11) that
\[ Q = Q - \Delta_r - \Delta_d. \]  
(6.17)

For brevity, we also write
\[ F_1 = 1 + zm_1, \quad F_2 = 1 + zm_2. \]  
(6.18)

By (4.3) and (4.4), it is easy to check that
\[ F_1 = -zm_1 m_2, \quad F_2 = -zm_1 m_2. \]  
(6.19)

Note that by definition \( \text{Tr} GA = \text{Tr} GA_1 + \text{Tr} GA_2 \) and \( \text{Tr} H_1 A = m_1 \text{Tr} A_1 + m_2 \text{Tr} A_2 \). Thus using (6.18), we have
\[ \text{Tr} \Xi_1 A = \text{Tr} GA_1 + \text{Tr} GA_2 - m_1 \text{Tr} A_1 - m_2 \text{Tr} A_2 \]
\[ = -m_1 \text{Tr} H A_1 - m_2 \text{Tr} H A_2 + F_1 \text{Tr} GA_1 + F_2 \text{Tr} GA_2, \]  
(6.20)

where in the last step, we used the fact \( zG = HG - I \).

Using (4.14) and (4.15), one can write
\[ \text{Tr} \Xi'_1 B = \frac{1}{2} \text{Tr} G^2 B_1 + \frac{1}{2} \text{Tr} G^2 B_2 - \frac{1}{2z} \text{Tr} GB_1 - \frac{1}{2z} \text{Tr} GB_2 - m_1' \text{Tr} B_1 - m_2' \text{Tr} B_2. \]

By further using the identity \( zG^2 = HG^2 - G \), it is not difficult to check
\[ \text{Tr} \Xi'_1 B = -\frac{m_1}{2} \text{Tr} H G^2 B_1 + \frac{F_1}{2} \text{Tr} G^2 B_1 + \frac{1}{2} (m_1 - \frac{1}{z}) \text{Tr} GB_1 - m_1' \text{Tr} B_1 \]
\[ - \frac{m_2}{2} \text{Tr} H G^2 B_2 + \frac{F_2}{2} \text{Tr} G^2 B_2 + \frac{1}{2} (m_2 - \frac{1}{z}) \text{Tr} GB_2 - m_2' \text{Tr} B_2. \]  
(6.21)

Recall the definition (5.1). Putting (6.20) and (6.21) together, we get
\[ Q = \sqrt{n} \left( -m_1 \text{Tr} H A_1 + F_1 \text{Tr} A_1 - m_2 \text{Tr} H A_2 + F_2 \text{Tr} A_2 \right. \]
\[ - \frac{m_1}{2} \text{Tr} H G^2 B_1 + \frac{F_1}{2} \text{Tr} G^2 B_1 + \frac{1}{2} (m_1 - \frac{1}{z}) \text{Tr} GB_1 - m_1' \text{Tr} B_1 \]
\[ - \frac{m_2}{2} \text{Tr} H G^2 B_2 + \frac{F_2}{2} \text{Tr} G^2 B_2 + \frac{1}{2} (m_2 - \frac{1}{z}) \text{Tr} GB_2 - m_2' \text{Tr} B_2. \]  
(6.22)

Recall the definition of \( \Delta_r \) from (5.5). We write
\[ \Delta_r = \sqrt{n} \sum_{i,j} x_{ij} c_{ij} - \sqrt{n} \sum_{(i,j) \in S(v)} x_{ij} c_{ij}. \]

Further recall the definition of \( c_{ij} \) from (5.6). It is elementary to check that
\[ \sqrt{n} \sum_{i,j} x_{ij} c_{ij} = -\sqrt{n} \left( m_1 \text{Tr} H A_1 + m_2 \text{Tr} H A_2 + \frac{m_1}{2} \text{Tr} H B_1 \right. \]
\[ + \frac{m_2}{2} \text{Tr} H B_2 + m_1' \text{Tr} B_1 + m_2' \text{Tr} B_2 \right). \]  
(6.23)

Using (6.22) and (6.23), with the notations defined in (6.3), we can write
\[ Q - \sqrt{n} \sum_{i,j} x_{ij} c_{ij} = \sqrt{n} (f_1 + f_2 + g_1 + g_2). \]  
(6.24)
Combining (5.5), (6.17) and (6.24) we can conclude the proof. □

7. Proof of Lemma 6.2

To prove Lemma 6.2, we need the following lemma summarizing some estimates on the derivative of $Q$ w.r.t. $x_{ij}$'s, which will be frequently used in the subsequent discussion. We first write $\frac{\partial Q}{\partial x_{ij}}$ in terms of Green functions. Recall the definition of $Q$ in (5.11) that

$$Q = \sqrt{n} \left( \text{Tr}(\Xi_1 A) + \text{Tr}(\Xi_1 B) \right) - \sqrt{n\pi} \sum_{(i,j) \in B(\nu)} x_{ij} c_{ij} - \Delta_d,$$

where $\Xi_1 = G - \Pi_1$ and $\Delta_d$ is a deterministic quantity in (5.4). Using $G' = \frac{1}{2}(G^2 - z^{-1}G)$ in Lemma 4.7, we find that

$$\frac{\partial Q}{\partial x_{ij}} = \sqrt{n} \left( \frac{\partial G}{\partial x_{ij}} A + \frac{1}{2} \left( \frac{\partial G^2}{\partial x_{ij}} B - z^{-1} \frac{\partial G}{\partial x_{ij}} B \right) \right) - 1 \left( (i,j) \in B(\nu) \right) \sqrt{n\pi} c_{ij}.$$

By Lemma 4.11, it can be further seen that

$$\frac{\partial Q}{\partial x_{ij}} = -\sqrt{n\pi} \sum_{l_1, l_2 \in \{i,j\} \setminus \{i,j\}} \left( (GAG)_{l_1 l_2} - \frac{1}{2\pi} (GBG)_{l_1 l_2} + \frac{1}{2} (G^2 BG)_{l_1 l_2} + \frac{1}{2} (G^2 G)_{l_1 l_2} \right)$$

$$- 1 \left( (i,j) \in B(\nu) \right) \sqrt{n\pi} c_{ij}. \quad (7.1)$$

Lemma 7.1. Under the assumptions of Proposition 5.1, we have

$$\frac{\partial Q}{\partial x_{ij}} = \sqrt{n\pi} 1 \left( (i,j) \in S(\nu) \right) c_{ij} + O(1). \quad (7.2)$$

Consequently, we have the bounds

$$\frac{\partial Q}{\partial x_{ij}} = \begin{cases} O(1), & \forall (i,j) \in B(\nu) \\ O(n^{-\frac{1}{2}}), & \forall (i,j) \in S(\nu). \end{cases} \quad (7.3)$$

Proof of Lemma 7.1. First, recall the definitions in (4.15) and (4.14). By (4.13), we have that for $a_1, a_2 = 1, 2$,

$$(G^{a_1} A G^{a_2})_{l_1 l_2} = (\Pi_{a_1} A \Pi_{a_2})_{l_1 l_2} + O(n^{-\frac{1}{2}}).$$

Applying the above estimates to (7.1), we find that

$$\frac{\partial Q}{\partial x_{ij}} = -\sqrt{n\pi} \sum_{l_1, l_2 \in \{i,j\} \setminus \{i,j\}} \left( (\Pi_{a_1} A \Pi_{a_2})_{l_1 l_2} - \frac{1}{2\pi} (\Pi_{a_1} B \Pi_{a_2})_{l_1 l_2} + \frac{1}{2} (\Pi_{a_1} B \Pi_{a_2} \Pi_{a_2})_{l_1 l_2} + \frac{1}{2} (\Pi_{a_2} B \Pi_{a_2})_{l_1 l_2} \right)$$

$$- 1 \left( (i,j) \in B(\nu) \right) \sqrt{n\pi} c_{ij} + O(1). \quad (7.4)$$

Comparing (7.4) with the definition of $c_{ij}$ in (5.6), we prove (7.2) and the first case of (7.3).

Next, by the definitions of $A, B$ in (4.24) and the set $S(\nu)$ in (5.2), it follows immediately that there exists some constant $C > 0$, such that

$$|A_{ij'}| \leq C n^{-\nu}, \quad |B_{ij'}| \leq C n^{-\nu}, \quad \forall (i,j) \in S(\nu).$$

By the estimates in (4.11), we get the second case of (7.3). This concludes the proof of Lemma 7.1. □

The remaining of the section is devoted to the proof of Lemma 6.2.
Proof of Lemma 6.2. We will focus on the proof of (6.13). Since the proof of (6.14) is analogous, we shall only outline the main steps. Recall from the definition in (6.3) and (6.18) that
\[
\sqrt{n}E_1 q^{(k-1)} = E \left( -m_1 \sqrt{n} \sum_{i,j} x_{ij}(\Xi_1 A_1)_{j'i'} + \sqrt{n} F_1 \text{Tr} GA_1 \right) q^{(k-1)}. \quad (7.5)
\]

For brevity, we use the notations
\[
h_1 = (\Xi_1 A_1)_{j'i'}, \quad h_2 = Q^{k-1}, \quad h_3 = e^{it\Delta}. \quad (7.6)
\]

Note that \( h_1 \) actually depends on the index \((j', i')\). However, we drop this dependence from notation for brevity. By Lemma 4.10, one has
\[
\sqrt{n} \sum_{i,j} E x_{ij}(\Xi_1 A_1)_{j'i'} q^{(k-1)} = \sqrt{n} \sum_{i,j} E x_{ij}(h_1 h_2 h_3)
\]
\[
= \sum_{l=1}^{3} \sum_{i,j} E \left( \frac{\partial^l}{\partial x_{ij}^l} (h_1 h_2 h_3) \right) + \mathbb{E} R_4, \quad (7.7)
\]
where \( R_4 \) satisfies that, for any sufficiently small \( \epsilon > 0 \) and sufficiently large \( K > 0 \),
\[
|\mathbb{E} R_4| \leq \sum_{i,j} E \left( n^{-\frac{2}{3}} \sup_{|x_{ij}| \leq n^{1/3}} \left| \frac{\partial^l}{\partial x_{ij}^l} (h_1 h_2 h_3) \right| + n^{-K} \sup_{x_{ij} \in \mathbb{R}} \left| \frac{\partial^4}{\partial x_{ij}^4} (h_1 h_2 h_3) \right| \right). \quad (7.8)
\]

Here we used the assumption that \( E|\sqrt{n}x_{ij}|^p \leq C_p \) for all \( p \geq 3 \). Therefore, the main technical estimates are the first four derivatives of \( h_1 h_2 h_3 \). By product rule, for each \( l \in \mathbb{N} \), we have
\[
\frac{\partial^l}{\partial x_{ij}^l} (h_1 h_2 h_3) = \sum_{l_1 + l_2 + l_3 = l} \left( \frac{l}{l_1, l_2, l_3} \right) \frac{\partial^{l_1} h_1}{\partial x_{ij}^{l_1}} \frac{\partial^{l_2} h_2}{\partial x_{ij}^{l_2}} \frac{\partial^{l_3} h_3}{\partial x_{ij}^{l_3}}. \quad (7.9)
\]

First, it is elementary to verify
\[
\frac{\partial^3 h_3}{\partial x_{ij}^3} = 1(i, j) \in B(\nu) \left( it \sqrt{n} \hat{c}_{ij} \right)^l e^{it\Delta}, \quad (7.10)
\]
and
\[
\frac{\partial h_1}{\partial x_{ij}} = \left( \frac{\partial G}{\partial x_{ij}} A_1 \right)_{j'i'}.
\]

The derivatives of \( h_2 \) can be computed using Faà di Bruno's formulas. For the reader’s convenience, we list them here. The first derivative of \( h_2 \) is
\[
\frac{\partial h_2}{\partial x_{ij}} = (k - 1) \frac{\partial Q}{\partial x_{ij}} Q^{k-2}.
\]

The second derivative of \( h_2 \) is
\[
\frac{\partial^2 h_2}{\partial x_{ij}^2} = \frac{(k - 1)!}{(k - 3)!} Q^{k-3} \left( \frac{\partial Q}{\partial x_{ij}} \right)^2 + (k - 1)Q^{k-2} \frac{\partial^2 Q}{\partial x_{ij}^2}.
\]

The third derivative of \( h_2 \) is
\[
\frac{\partial^3 h_2}{\partial x_{ij}^3} = \frac{(k - 1)!}{(k - 4)!} Q^{k-4} \left( \frac{\partial Q}{\partial x_{ij}} \right)^3 + 3 \frac{(k - 1)!}{(k - 3)!} Q^{k-3} \frac{\partial Q}{\partial x_{ij}} \frac{\partial^2 Q}{\partial x_{ij}^2} + (k - 1)Q^{k-2} \frac{\partial^3 Q}{\partial x_{ij}^3}.
\]
The fourth derivative of $h_2$ is
\[
\frac{\partial^4 h_2}{\partial x_{ij}^4} = \frac{(k-1)!}{(k-5)!} Q^{k-5} \left( \frac{\partial Q}{\partial x_{ij}} \right)^4 + 6 \frac{(k-1)!}{(k-4)!} Q^{k-4} \left( \frac{\partial Q}{\partial x_{ij}} \right)^2 \frac{\partial^2 Q}{\partial x_{ij}^2} + \frac{(k-1)!}{(k-3)!} Q^{k-3} \left( 4 \frac{\partial Q}{\partial x_{ij}} \frac{\partial^3 Q}{\partial x_{ij}^3} + 3 \left( \frac{\partial^2 Q}{\partial x_{ij}^2} \right)^2 \right) + (k-1) Q^{k-2} \frac{\partial^4 Q}{\partial x_{ij}^4}.
\]

As we can see from the above identities, the key ingredients are the partial derivatives of $Q$ and $GA_1$. We further summarize some identities on the derivatives of $Q$ in Appendix A.

For brevity, we introduce the notation
\[
h(l_1, l_2, l_3) := n^{-\frac{l_1 l_2 l_3 - l_1 - l_2 - l_3}{2}} \sum_{i,j} \frac{\partial^i h_1}{\partial x_{ij}^i} \frac{\partial^j h_2}{\partial x_{ij}^j} \frac{\partial^k h_3}{\partial x_{ij}^k}. \tag{7.11}
\]

In the following two lemmas, we summarize the estimates of $h(l_1, l_2, l_3)$ for $l_1 + l_2 + l_3 \leq 4$. The proofs of the two lemmas will be given in Sections 7.1 and 7.2.

**Lemma 7.2.** For the first derivative of $h_1 h_2 h_3$, we have that
\[
h(1, 0, 0) = -\sqrt{\pi} m_2 \Tr(GA_1) q^{(k-1)} + O_\prec(n^{-\frac{k}{2}}), \tag{7.12}
\]
\[
h(0, 1, 0) = a_{11} q^{(k-2)} + O_\prec(n^{-\frac{k}{2}+4\nu}), \tag{7.13}
\]
\[
h(0, 0, 1) = O_\prec(n^{-\frac{k}{2}+4\nu}). \tag{7.14}
\]

**Lemma 7.3.** On higher order derivatives of $h_1 h_2 h_3$, we have the following estimates.

1. For the second derivative, we have
\[
h(2, 0, 0) = \mathcal{O}q^{(k-1)} + O_\prec(n^{-\frac{k}{2}}), \tag{7.15}
\]
\[
h(1, 1, 0) = a_{21} q^{(k-2)} + O_\prec(n^{-\frac{k}{2}+4\nu}), \tag{7.16}
\]
\[
h(1, 0, 1) = O_\prec(n^{-\frac{k}{2}+4\nu}), \quad h(0, 2, 0) = O_\prec(n^{-\frac{k}{2}}), \tag{7.17}
\]
\[
h(0, 1, 1) = O_\prec(n^{-\frac{k}{2}+4\nu}), \quad h(0, 0, 2) = O_\prec(n^{-1+4\nu}). \tag{7.18}
\]

2. For the third derivative, we have
\[
h(1, 2, 0) = a_{31} q^{(k-2)} + O_\prec(n^{-\frac{k}{2}}), \tag{7.19}
\]
\[
h(3, 0, 0) = O_\prec(n^{-\frac{k}{2}}), \quad h(0, 3, 0) = O_\prec(n^{-\frac{k}{2}}), \tag{7.20}
\]
\[
h(2, 1, 0) = O_\prec(n^{-1}), \quad h(2, 0, 1) = O_\prec(n^{-\frac{k}{2}+4\nu}), \tag{7.21}
\]
\[
h(1, 1, 1) = O_\prec(n^{-1+4\nu}), \quad h(1, 0, 2) = O_\prec(n^{-\frac{k}{2}+4\nu}), \tag{7.22}
\]
\[
h(0, 2, 1) = O_\prec(n^{-\frac{k}{2}+4\nu}). \tag{7.23}
\]

3. For the fourth derivative, all the terms in the RHS of (7.9) can be bounded by $O_\prec(n^{-\frac{k}{2}+4\nu})$.

By Lemma 7.2 and Lemma 7.3, the first term in (7.7) is estimated by
\[
\sum_{l=1}^{3} \kappa^{l+1}_{n^{l/2}} \sum_{i,j} \mathbb{E} \left( \frac{\partial^4}{\partial x_{ij}^4} h_1 h_2 h_3 \right) = \sum_{l=1}^{3} \sum_{i,j} \sum_{l_1 + l_2 + l_3 = l} \kappa^{l+1}_{n^{l/2}} h(l_1, l_2, l_3) = -\sqrt{\pi} m_2 \Tr(GA_1) q^{(k-1)} + \frac{\kappa_3}{2} \mathcal{O} q^{(k-1)} + \left( a_{11} + \kappa_3 a_{21} + \frac{\kappa_4}{2} a_{31} \right) q^{(k-2)}.
\]

For the second term in (7.7), we claim that
\[
|\mathbb{E} R_4| \leq n^{-1+4\nu}.
\]


To prove (7.18), it is enough to bound the two terms on the right hand side of (7.8). We apply Lemma 7.3 to the first term on the right hand side of (7.8) to get
\[ \sum_{i,j} \mathbb{E} n^{-\frac{5}{4}} \sup_{|x_{ij}| \leq n^{-\frac{1}{4} \rho}} \left| \frac{\partial^4}{\partial x_{ij}^4} (h_1 h_2 h_3) \right| \leq n^{-1+4\nu}. \]

A minor issue with the above step is that Lemma 7.3 is proved for the matrix $X$ with all entries random variables. In our application of Lemma 7.3, for each pair of fixed indices $(i,j)$, we actually consider a random matrix $X$ whose $(i,j)$th entry is a deterministic number with small magnitude and all the others random variables. However, this can be justified by a perturbation argument with the aid of resolvent expansion. Indeed, replacing one random entry $x_{ij}$ by any deterministic number bounded by $n^{-1/2+\varepsilon}$ and keeping the other $X$ entries random will not change the isotropic local law. Thus Lemma 7.3 holds for such random matrix $X$.

For the second term on the right hand side of (7.8), we use the trivial bounds for $G$ and its derivatives to obtain
\[ \sum_{i,j} \mathbb{E} n^{-K} \sup_{x_{ij} \in \mathbb{R}} \left| \frac{\partial^4}{\partial x_{ij}^4} (h_1 h_2 h_3) \right| \leq n^{-K+2+C} \]
for a positive constant $C$. By taking $K$ sufficiently large, we conclude (7.18).

Plugging (7.7) into (7.5), we finally get
\[ \sqrt{n} \mathbb{E} f_1 q^{(k-1)} = -m_1 \sqrt{\mathbb{E}} \left( \frac{\kappa_3}{2} q^{(k-1)} + \left( a_{11} + \kappa_3 a_{21} + \frac{\kappa_4}{2} a_{31} \right) q^{(k-2)} \right) + O_{\prec} (n^{-\frac{1}{2}+4\nu}). \]

Note that by (6.19), the term $\sqrt{n} m_1 m_2 \text{Tr} G A_{14} q^{(k-1)}$ is cancelled with $F_{14} G A_{14} q^{(k-1)}$ in (7.5). This verifies (6.13) in case of $\alpha = 1$ by recalling the definition in (6.12).

Next, we turn to (6.14) for $\alpha = 1$. Recall the definition of $g_1$ in (6.3). We have
\[ \sqrt{n} \mathbb{E} g_1 q^{(k-1)} = \sqrt{n} \mathbb{E} \left( -m_1 \sqrt{\mathbb{E}} \sum_{i,j} x_{ij} (\Xi_2 B_1)_{j'i} + \frac{m_1}{2} \text{Tr} G^2 B_1 \right) \]
\[ + \frac{z m_1 - 1}{2z} \text{Tr} G B_1 - m_1 \text{Tr} B_1 + m_1 \text{Tr} H \Pi_1 B_1 \right) q^{(k-1)}. \]

(7.19)
The main task is to estimate the cumulant expansion of the term
\[ \sqrt{n} \sum_{i,j} \mathbb{E} x_{ij} (\Xi_2 B_1)_{j'i} q^{(k-1)}, \]
which is analogous to (7.7). Recall $h_2$ and $h_3$ in (7.6) and denote
\[ \tilde{h}_1 = (\Xi_2 B_1)_{j'i}. \]

(7.20)
Note that $\tilde{h}_1$ depends on the indices $i, j$. However, we drop these dependence from the notation for brevity. Similarly to (7.11), we introduce the notation
\[ \tilde{h}(l_1, l_2, l_3) := n^{-\frac{l_1 + l_2 + l_3}{2}} \sum_{i,j} \frac{\partial^{l_1} h_1}{\partial x_{ij}^{l_1}} \frac{\partial^{l_2} h_2}{\partial x_{ij}^{l_2}} \frac{\partial^{l_3} h_3}{\partial x_{ij}^{l_3}}. \]

(7.21)
We collect the estimates of $\tilde{h}(l_1, l_2, l_3)$ for $l_1 + l_2 + l_3 \leq 4$ in the following two lemmas, whose proofs are postponed to Section 7.3.

**Lemma 7.4.** For the first derivative of $\tilde{h}_1 h_2 h_3$, we have
\[ \tilde{h}(1,0,0) = -\sqrt{n^2} \left( (2m_2^2 + \frac{m_3}{z}) \text{Tr} (GB_1) + m_2 \text{Tr} (G^2 B_1) \right) q^{(k-1)} + O_{\prec} (n^{-\frac{1}{2}}), \]
\[\tilde{h}(0, 1, 0) = \tilde{b}_{11} q^{(k-2)} + O_{\prec}(n^{-\frac{k}{2} + 4\nu}),\]
\[\tilde{h}(0, 0, 1) = O_{\prec}(n^{-\frac{k}{2} + 4\nu}).\]

**Lemma 7.5.** For higher order derivatives of \(\tilde{h}_1 h_2 h_3\), we have the following estimates.

1. For the second derivative, we have
   \[\tilde{h}(2, 0, 0) = \tilde{b}_1 q^{(k-1)} + O_{\prec}(n^{-\frac{k}{2}}),\]
   \[\tilde{h}(1, 1, 0) = \tilde{b}_{21} q^{(k-2)} + O_{\prec}(n^{-\frac{k}{2}}),\]
   \[\tilde{h}(0, 2, 0) = O_{\prec}(n^{-\frac{k}{2}}).\]

   All the other terms with \(l_3 \geq 1\) can be bounded by \(O_{\prec}(n^{-\frac{k}{2} + 4\nu})\).

2. For the third derivative, we have
   \[\tilde{h}(1, 2, 0) = \tilde{b}_{31} q^{(k-2)} + O_{\prec}(n^{-\frac{k}{2}}),\]
   \[\tilde{h}(3, 0, 0) = O_{\prec}(n^{-\frac{k}{2}}), \quad \tilde{h}(0, 3, 0) = O_{\prec}(n^{-\frac{k}{2}}),\]
   \[\tilde{h}(2, 1, 0) = O_{\prec}(n^{-1}).\]

   All the other terms with \(l_3 \geq 1\) can be bounded by \(O_{\prec}(n^{-\frac{k}{2} + 4\nu})\).

3. For the fourth derivative, all the terms can be bounded by \(O_{\prec}(n^{-\frac{k}{2} + 4\nu})\).

With these preparations, using arguments similar to those of (7.5), we find that

\[
\sqrt{nE}\partial_1 q^{(k-1)} = \frac{m_1}{2} \sqrt{\mathbb{E}} \left( \frac{\kappa_2 \bar{F}_1}{2} q^{(k-1)} + \left( \tilde{b}_{11} + \kappa_2 \tilde{b}_{21} + \frac{\kappa_4}{2} \bar{b}_{11} \right) q^{(k-2)} \right)
+ \sqrt{nE} \left( \frac{m_1}{m_1} \text{Tr} GB_1 - m_1' \text{Tr} B_1 + m_1' \text{Tr} H\Pi_i B_1 \right) q^{(k-1)} + O_{\prec}(n^{-\frac{k}{2} + 4\nu}).
\]

In the above, we use (6.19) and an identity

\[
\frac{m_1}{2} \left( z m_1 (2m_2 + \frac{m_2}{z}) + m_1 - \frac{1}{z} \right) = m_1',
\]

which can be checked from (4.3) and (4.4). Next, observe that

\[
\sqrt{nE} \left( \frac{m_1}{m_1} \text{Tr} GB_1 + m_1' \text{Tr} H\Pi_i B_1 - m_1' \text{Tr} B_1 \right) q^{(k-1)}
= \sqrt{nE} \left( - z m_1' \text{Tr} GB_1 + m_1' F_1 \text{Tr} GB_1 + m_1' \text{Tr} H\Pi_i B_1 - m_1' \text{Tr} B_1 \right) q^{(k-1)},
= \frac{m_1'}{m_1} \sqrt{nE} \left( - m_1 \text{Tr} H\bar{E}_i B_1 + F_1 \text{Tr} GB_1 \right) q^{(k-1)}.
\]

In the first step above, we simply use the definition of \(F_1\) in (6.18). In the second step, we use the fact \(zG = HG - I\). Note that the remaining derivation can be done via replacing \(A_1\) with \(B_1\) (mutatis mutandis) in the counterpart for \(f_1\). Therefore, we finally get

\[
\sqrt{nE} \partial_1 q^{(k-1)} = - \sqrt{\mathbb{E}} \left( \frac{m_1 \kappa_0}{4} \left( \tilde{b}_1 + 2 \frac{m_1'}{m_1} \tilde{b}_{11} \right) q^{(k-1)} + \left( \frac{m_1}{2} \tilde{b}_{11} + m_1' \tilde{b}_{11} \right) q^{(k-2)}
+ \left( \frac{m_1 \kappa_2}{2} \tilde{b}_{21} + \kappa_2 m_1' \tilde{b}_{21} + \frac{m_1 \kappa_4}{4} \bar{b}_{31} + \frac{\kappa_4 m_1'}{2} \bar{b}_{31} \right) q^{(k-2)} \right) + O_{\prec}(n^{-\frac{k}{2} + 4\nu}).
\]

This verifies (6.14) in case of \(\alpha = 1\) by recalling the definition in (6.12).
The proofs of (6.13) and (6.14) in case of $\alpha = 2$ are analogous to those of (7.5) and (7.19). We outline the main steps. First observe that
\[
\sqrt{n} \mathbb{E} f_2 q^{(k-1)} = \mathbb{E} \left( -m_2 \sqrt{n} \bar{z} \sum_{i,j} x_{ij}(E_1 A_2)_{ij} + F_2 \text{Tr} G A_2 \right) q^{(k-1)},
\]
\[
\sqrt{n} \mathbb{E} g_2 q^{(k-1)} = \mathbb{E} \left( -\frac{m_3}{2} \sqrt{n} \bar{z} \sum_{i,j} x_{ij}(E_2 B_2)_{ij} + \frac{F_2}{2} \text{Tr} G^2 B_2 \right.
\]
\[
+ \frac{zn_2 - 1}{2z} \text{Tr} GB_2 - m_1' \text{Tr} B_2 + m_1' \text{Tr} H\Pi B_2 \big) q^{(k-1)}. \]

Recall $h_2$ and $h_3$ in (7.6) and denote
\[
h_1 = (E_1 A_2)_{ij}, \quad \tilde{h}_1 = (E_2 B_2)_{ij}.
\]
Analogously to (7.11) and (7.21), we introduce the notations
\[
h(l_1, l_2, l_3) := n^{-\frac{l_1+l_2+l_3}{2}} \sum_{i,j} \frac{\partial^i h_{1}}{\partial x_{ij}^i} \frac{\partial^j h_{2}}{\partial x_{ij}^j} \frac{\partial^k h_{3}}{\partial x_{ij}^k},
\]
and $\tilde{h}(l_1, l_2, l_3)$ which is defined via replacing $h_1$ by $\tilde{h}_1$ in the above definition.

Then we have the estimates for the first order derivatives involving $h_1$ and $\tilde{h}_1$.

**Lemma 7.6.** For $h$, we have
\[
h(1,0,0) = -\sqrt{n} \bar{z} m_1 \langle \text{Tr} G A_2 \rangle q^{(k-1)} + O_{\prec} (n^{-\frac{1}{2}}),
\]
\[
h(0,1,0) = a_{12} q^{(k-2)} + O_{\prec} (n^{-\frac{1}{2} + 4\nu}),
\]
\[
h(0,0,1) = O_{\prec} (n^{-\frac{1}{2} + 4\nu}).
\]

Similarly, for $\tilde{h}$, we have
\[
\tilde{h}(1,0,0) = -\sqrt{n} \bar{z} y \left( (2m_1' + \frac{m_1}{z}) \text{Tr} GB_2 + m_1 \text{Tr} G^2 B_2 \right) q^{(k-1)} + O_{\prec} (n^{-\frac{1}{2}}),
\]
\[
\tilde{h}(0,1,0) = \tilde{b}_{12} q^{(k-2)} + O_{\prec} (n^{-\frac{1}{2} + 4\nu}),
\]
\[
\tilde{h}(0,0,1) = O_{\prec} (n^{-\frac{1}{2} + 4\nu}).
\]

For the higher order derivatives, we have the following lemma.

**Lemma 7.7.** We have the following estimates in case $l_1 + l_2 + l_3 \geq 2$.

1. For $h(l_1, l_2, l_3)$, we have
\[
h(2,0,0) = \partial^2 q^{(k-1)} + O_{\prec} (n^{-\frac{1}{2}}),
\]
\[
h(1,1,0) = a_{22} q^{(k-2)} + O_{\prec} (n^{-\frac{1}{2} + 4\nu}),
\]
\[
h(1,2,0) = a_{32} q^{(k-2)} + O_{\prec} (n^{-\frac{1}{2}}).
\]

All the other terms with $l_1 + l_2 + l_3 \geq 2$ can be bounded by $O_{\prec} (n^{-\frac{1}{2} + 4\nu})$.

2. For $\tilde{h}(l_1, l_2, l_3)$ we have
\[
\tilde{h}(2,0,0) = \tilde{\partial}^2 q^{(k-1)} + O_{\prec} (n^{-\frac{1}{2}}),
\]
\[
\tilde{h}(1,1,0) = \tilde{b}_{22} q^{(k-2)} + O_{\prec} (n^{-\frac{1}{2}}),
\]
\[
\tilde{h}(1,2,0) = \tilde{b}_{32} q^{(k-2)} + O_{\prec} (n^{-\frac{1}{2}}).
\]

All the other terms with $l_1 + l_2 + l_3 \geq 2$ can be bounded by $O_{\prec} (n^{-\frac{1}{2} + 4\nu})$. 
The proofs of the above lemmas will be given in Section 7.3. The remaining estimates for $\sqrt{n}\mathbb{E}_q q^{(k-1)}$ and $\sqrt{n}\mathbb{E}_q q^{(k-1)}$ follow the same arguments as those of (7.5) and (7.19), and are therefore omitted. As a side note, we mention an identity (comparable to (7.22))

$$\frac{m_2}{2} \left( z y m_2 (2m_1 + \frac{m_1}{z}) + m_2 - \frac{1}{z} \right) = m_2'$$

used in the derivation of the $g_2$ term.

Lastly, we prove (6.15). Recall $h_2 = Q^{k-1}$ and $h_3 = e^{i\Delta}$. By Lemma 4.10, we have

$$\sqrt{n} z \sum_{(i,j)\in S(\nu)} c_{ij} E x_{ij} q^{(k-1)} = \sqrt{z} \sum_{(i,j)\in S(\nu)} c_{ij} E \left( \frac{1}{\sqrt{n}} \frac{\partial(h_2 h_3)}{\partial x_{ij}} + \frac{\kappa_3}{2n} \frac{\partial^2(h_2 h_3)}{\partial x_{ij}^3} + \mathcal{R} \right),$$

where $\mathcal{R}$ satisfies that, for any sufficiently small $\epsilon > 0$ and sufficiently large $K > 0$,

$$|\mathcal{R}| \leq \sum_{i,j} \mathbb{E} \left( n^{-\frac{\epsilon}{2}} \sup_{|x_{ij}| \leq n^{-\frac{\epsilon}{2}}} \left| c_{ij} \frac{\partial^2(h_2 h_3)}{\partial x_{ij}^3} \right| + n^{-K} \sup_{|x_{ij}| \leq \mathbb{R}} \left| c_{ij} \frac{\partial^2(h_2 h_3)}{\partial x_{ij}^3} \right| \right).$$

We first show that

$$|\mathcal{R}| = O_{\prec}(n^{-\frac{1}{2}+4\nu}). \quad (7.25)$$

Similar to the discussion of (7.18), the proof boils down to estimate the third order derivative of $h_2 h_3$. Using the same proof as (7.14) in Lemma 7.3 (given in Section 7.1), we observe that in the derivatives of $h_2 h_3$, any term containing the derivatives of $h_3$ can bounded by $O_{\prec}(n^{-\frac{1}{2}+4\nu}).$ Thus, by product rule,

$$\frac{\partial^3(h_2 h_3)}{\partial x_{ij}^3} = \frac{\partial^3 h_2}{\partial x_{ij}^3} h_3 + O_{\prec}(n^{-\frac{1}{2}+4\nu}) = O_{\prec} \left( u(i) v(j) + n^{-\frac{1}{2}+4\nu} \right).$$

The last step is obtained analogously to (7.17). We omit the details. To conclude (7.25), we also use $c_{ij} = O_{\prec}(u(i) v(j))$ by recalling its definition (5.6) and the fact that $u, v$ are both unit vectors.

Next, using arguments similar to (7.16) and (7.17), we get

$$\frac{1}{\sqrt{n}} \frac{\partial(h_2 h_3)}{\partial x_{ij}} = \frac{1}{\sqrt{n}} \frac{\partial h_2}{\partial x_{ij}} h_3 + O_{\prec}(n^{-\frac{1}{2}+4\nu}) = (k-1)\sqrt{z} c_{ij} q^{(k-2)} + O_{\prec}(n^{-\frac{1}{2}+4\nu}), \quad (7.26)$$

and

$$\frac{1}{n} \frac{\partial^2(h_2 h_3)}{\partial x_{ij}^2} = \frac{1}{n} \frac{\partial^2 h_2}{\partial x_{ij}^2} h_3 + O_{\prec}(n^{-\frac{1}{2}+4\nu}) = 2 \frac{(k-1)z}{\sqrt{n}} s_{ij} q^{(k-2)} + O_{\prec}(n^{-\frac{1}{2}+4\nu}). \quad (7.27)$$

Plugging (7.25)-(7.27) into (7.24), we obtain (6.15). The proof of Lemma 6.2 is now complete.

7.1. Proof of Lemma 7.2. We start with a simple identity which will be frequently referred to later. For any deterministic matrix $W \in \mathbb{R}^{(M+n)\times (M+n)}$, it is elementary to check that

$$\left( \frac{\partial G}{\partial x_{ij}} W \right)_{ab} = -\sqrt{z} \left( G_{aj'}(GW)_{ib} + G_{ai}(GW)_{jb} \right). \quad (7.28)$$

We emphasize that both (4.11) and a basic fact (as a consequence of (6.1))

$$q^{(l)} = Q^l e^{i\Delta} = O_{\prec}(1) \quad \text{for } l \geq 1$$

will be applied to bound the error terms throughout the proofs of Lemma 7.2-Lemma 7.7.
For convenience, we denote the blocks of $A$ and $B$ (c.f. (4.24)) by $A_k$’s and $B_k$’s, i.e.,

$$A = \begin{pmatrix} \omega_1 uu^* & \omega_2 uv^* \\ \omega_3 vv^* & \omega_4 vv^* \end{pmatrix} := \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix}, \quad B = \begin{pmatrix} \omega_1 uu^* & \omega_2 uv^* \\ \omega_3 vv^* & \omega_4 vv^* \end{pmatrix} := \begin{pmatrix} B_1 & B_2 \\ B_3 & B_4 \end{pmatrix}. \quad (7.29)$$

With the above preparation, we now prove Lemma 7.2.

Proof of Lemma 7.2. First, by recalling the notations in (7.6) and (7.11), and using (7.28), we have

$$h(1,0,0) = \frac{1}{\sqrt{n}} \sum_{i,j} \left( \frac{\partial G}{\partial x_{ij}} A_1 \right)_{ij} q^{(k-1)}$$

$$= -\sqrt{n} \sum_{i,j} \left( G_{ij}(GA_1)_{ii} + G_{ij}(GA_1)_{ij} \right) q^{(k-1)}.$$  

Moreover, by (4.2) and (4.9), we further get

$$h(1,0,0) = -\sqrt{n} m_{2n} (\text{Tr} G A_1) q^{(k-1)} - \sqrt{\frac{\pi}{n}} (\text{Tr} G A_1) q^{(k-1)}$$

$$= -\sqrt{n} m_{2n} (\text{Tr} G A_1) q^{(k-1)} + O(n^{-\frac{3}{4}}), \quad (7.30)$$

where the last step follows from (4.13).

Next, using the fact $|B(\nu)| \leq C n^{4\nu}$ together with the definition of $c_{ij}$ in (5.6) and (4.13), we obtain

$$h(0,0,1) = \sqrt{\frac{\pi}{n}} \sum_{i,j} i t c_{ij} (\Xi_1)_{ij} q^{(k-1)} = O(n^{-\frac{1}{4} + 4\nu}). \quad (7.31)$$

The main task is the estimate of

$$h(0,1,0) = \frac{k - 1}{\sqrt{n}} \sum_{i,j} (\Xi_1 A_1)_{ij} \frac{\partial Q}{\partial x_{ij}} q^{(k-2)}.$$

In light of the expression of $\partial Q/\partial x_{ij}$ in (7.1), by symmetry, we get

$$h(0,1,0) = -(k - 1) \sqrt{n} \sum_{i,j} (\Xi_1 A_1)_{ij} \left[ 2(G AG)_{ij} - \frac{1}{z}(GBG)_{ij} + (GBG^2)_{ij} + (G^2 BG)_{ij} \right]$$

$$- (k - 1) \sqrt{n} \sum_{(i,j) \in B(\nu)} (\Xi_1 A_1)_{ij} c_{ij}. \quad (7.32)$$

The last term on the right hand side of (7.32) is bounded by $O(n^{-\frac{1}{4} + 4\nu})$, by exactly the same estimate of (7.31). Now we turn towards the first term on the right hand side of (7.32). We first claim that

$$\sum_{i,j} (\Xi_1 A_1)_{ij} (G AG)_{ij} = \text{Tr}(\Pi^2_2 - \Pi^1_{1,1}) A_1 A + O(n^{-\frac{1}{2}}). \quad (7.33)$$

To derive the above statement, a key observation is that the summation on the left hand side of (7.33) can be written in terms of a trace, with the aid of the block diagonal matrices $\Gamma^a$ and $\Gamma^1$ in (6.2). Indeed, we find

$$\sum_{i,j} (\Xi_1 A_1)_{ij} (G AG)_{ij} = \text{Tr}(\Gamma^2 \Xi_1 A_1 \Gamma^a G A G) = \text{Tr}(G^1 G - G^1 \Pi_1) A_1 G A.$$

Thus $\Pi^1_{1,1}$ and $\Pi^1_2$ (c.f. (6.5) and (6.6)) appear naturally in (7.33).
To prove (7.33), using the expressions of $G$ in (4.1) and $A$ in (7.29), we have that

$$\Xi_1 A_{ji} = \left( \frac{1}{\sqrt{z}} X^* G_1 A + (G_2 - m_2) A_3 \right)_{ji},$$  \hspace{1cm} (7.34)

$$G A G_{ji} = \left( \frac{1}{\sqrt{z}} X^* G_1 A_1 G_1 + G_2 A_3 G_1 + \frac{1}{z} X^* G_1 A_2 X^* G_1 + \frac{1}{\sqrt{z}} G_2 A_4 X^* G_1 \right)_{ji}.$$  \hspace{1cm} (7.35)

Expanding the left hand side of (7.33) with the above expressions, we shall show that there are two main terms and all others are negligible. Let us concentrate on the following term. All other terms in the expansion of the left hand side of (7.33) are negligible.

The first contributing term is

$$\sum_{i,j} ((G_2 - m_2) A_3)_{ji} (G_2 A_3 G_1)_{ji} = \omega_2^2 \text{Tr}(G_2 - m_2) v u^* G_1 u v^* G_2)$$

$$= \omega_2^2 (u^* G_1 u) (v^* G_2 (G_2 - m_2) v) = \sum_{i,j} ((\Pi_2^1 - \Pi_1^1) A_1 A)_{ji} A_{ji} + O_x(n^{-\frac{1}{2}}),$$

where in the last step we use $G_2^2 = G_2$ and the definition of $A$ in (7.29), followed by (4.13) and (4.18).

The second contributing term is

$$\frac{1}{z} \sum_{i,j} (X^* G_1 A_1)_{ji} (X^* G_1 A_1 G_1)_{ji} = \omega_2^2 (u^* G_1 u) \left( \frac{1}{z} u^* G_1 X X^* G_1 u \right).$$

Let $\bar{v} = (0, v)^*$ and $\bar{u} = (u, 0)^*$ denote the augmented vectors in $\mathbb{R}^{M+n}$. Note that by (4.18), we first have

$$u^* G^2 \bar{u} = u^* G^2 u + \frac{1}{z} u^* G_1 X X^* G_1 u = 2m^1 + \frac{m_1}{z} + O_x(n^{-\frac{1}{2}}).$$

Further observe that

$$u^* G^2 \bar{u} = u^* G^2 u = m^1 + O_x(n^{-\frac{1}{2}}),$$

where the last equation follows from (4.10). Putting them together, we conclude that

$$\frac{1}{z} u^* G_1 X X^* G_1 u = m^1 + \frac{m_1}{z} + O_x(n^{-\frac{1}{2}}).$$

As a consequence,

$$\frac{1}{z} \sum_{i,j} (X^* G_1 A_1)_{ji} (X^* G_1 A_1 G_1)_{ji} = \sum_{i,j} ((\Pi_2^1 - \Pi_1^1) A_1)_{ji} \bar{A}_{ji} + O_x(n^{-\frac{1}{2}}).$$

Note that

$$\sum_{i,j} ((\Pi_2^1 - \Pi_1^1) A_1)_{ji} \bar{A}_{ji} + \sum_{i,j} ((\Pi_2^1 - \Pi_1^1) A_1)_{ji} \bar{A}_{ji} = \text{Tr}(\Pi_2^1 - \Pi_1^1) A_1 A_1 A.$$

What remains is to show all other terms in the expansion of the left hand side of (7.33) with (7.34) and (7.35) are negligible. Let us concentrate on the following term. All other remaining terms are estimated similarly; we omit the details.

$$\frac{1}{\sqrt{z}} \sum_{i,j} (X^* G_1 A_1)_{ji} (G_2 A_3 G_1)_{ji} = \frac{\omega_3^2 (\omega_2^2) \text{Tr}(X^* G_1 u u^* G_1 u v^* G_2)}{\sqrt{z}}$$

$$= \frac{\omega_3 (\omega_2^2)}{\sqrt{z}} \text{Tr}(v^* G_2 X u u^* G_1 u) = \frac{\omega_3 (\omega_2^2)}{\sqrt{z}} (u^* G_1 u) (v^* G_2^2 X u).$$
In the second step above, we use the fact $X^*G_1 = G_2X^*$ which can be checked easily via the singular value decomposition. Therefore, using $G_2^2 = G_2'$ and $G' = (G^2 - z^{-1}G)/2$, together with (4.8) and (4.10), we get that

$$v^*G_2^2X^* u = (\bar{v}^* \sqrt{z} G \bar{u})' = \frac{1}{2\sqrt{z}} \bar{v}^* G \bar{u} + \sqrt{z} v^* G' \bar{u} = O_{\prec}(n^{-\frac{1}{2}}).$$

Hence, we conclude that

$$\frac{1}{\sqrt{z}} \sum_{i,j}(X^*G_1A_1)_{ji}(G_2A_2G_1)_{ji} = O_{\prec}(n^{-\frac{1}{2}}).$$

The proof of (7.33) is complete.

Next, analogously, we shall show that

$$\sum_{i,j}(\Xi_1A_1)_{ji}(G^2B)_{ji} = \text{Tr}(\Pi_3^2 - \Pi_2^2)A_1B + O_{\prec}(n^{-\frac{1}{2}}). \quad (7.36)$$

A simple calculation using (4.1) and (4.24) yields

$$(G^2BG)_{ji} = (\frac{1}{\sqrt{z}} X^*G_1B_1G_1 + \frac{1}{\sqrt{z}} G_2X^*B_1G_1 + \frac{1}{z} X^*G_1B_2G_1G_2 + \frac{1}{z} X^*G_2B_2X^*G_1)$$

$$+ \frac{1}{z} G_2X^*B_1X^*G_1 + \frac{1}{z^2} X^*G_2B_2X^*G_1 + \frac{1}{z^2} G_2B_4X^*G_1)_{ji}. \quad (7.37)$$

In a similar way to the discussion of (7.33), we expand $(\Xi_1A)_{ji}(G^2BG)_{ji}$ using (7.34) and (7.37). There are only four non-negligible terms in the expansion.

Recall $A_1$ and $B_1$ in (4.24). The first non-negligible term is

$$\frac{1}{z} \sum_{i,j}(X^*G_1A_1)_{ji}(X^*G_2^2B_1G_1)_{ji} = \frac{\omega_{12}}{z}(u^*G_1u)(u^*G_2^2X^*G_1u).$$

To estimate $u^*G_2^2XX^*G_1u$ in the above, we observe that (via elementary calculations and the fact $G_2X^* = X^*G_1$)

$$u^*G_2^2u = u^*G_1u + \frac{3}{z}(u^*G_2^2X^*G_1u).$$

Moreover, by $G_1^2 = \frac{1}{2}G''$, (4.10) and (4.18), we find

$$u^*G_2^2u = \frac{1}{2} u^*G''u = \frac{1}{2} m'' + O_{\prec}(n^{-\frac{1}{2}}),$$

$$u^*G_1u = 2m'' + \frac{3}{z} m'' + O_{\prec}(n^{-\frac{1}{2}}).$$

Hence,

$$u^*G_2^2X^*G_1u = \frac{z}{2} m'' + m'' + O_{\prec}(n^{-\frac{1}{2}}). \quad (7.38)$$

We conclude that

$$\frac{1}{z} \sum_{i,j}(X^*G_1A_1)_{ji}(X^*G_2^2B_1G_1)_{ji} = \frac{1}{2} \sum_{i,j}((\Pi_3^2 - \Pi_2^2)A_1B_{ij} + O_{\prec}(n^{-\frac{1}{2}})).$$

Using the fact $XG_2 = G_1X$ and the same arguments as above, we can show the second non-negligible term is

$$\frac{1}{z} \sum_{i,j}(X^*G_1A_1)_{ji}(G_2X^*G_1B_1G_1)_{ji} = \frac{1}{2} \sum_{i,j}((\Pi_3^1 - \Pi_2^1)A_1B_{ij} + O_{\prec}(n^{-\frac{1}{2}})),\quad (7.39)$$

and so on.
The third non-negligible term is
\[
\frac{1}{z} \sum_{i,j} ((G_2 - m_2)A_3)_{ji}(X^*G_1^2XB_3G_1)_{ji} = \frac{1}{z}(u^*G_1u)(v^*X^*G_1^2X(G_2 - m_2)v)
\]
\[
= \omega_3 \varpi_3 m_1 \left( \frac{m''_2}{2} + \frac{m'_2}{z} - \frac{m_2^2 + zm_2 m'_2}{z} \right) + O_\prec(n^{-\frac{1}{2}}),
\]
where we used the facts \(X^*G_1^2XG_2 = X^*G_1^3X\) and \(G_3^2 = \frac{1}{2}G''_2\), as well as
\[
\bar{v}^*G^3v = \frac{3}{z}v^*X^*G_1^3Xv + v^*G_2^3v = 2m''_2 + \frac{3m'_2}{z} + O_\prec(n^{-\frac{1}{2}}).
\]
The last non-negligible term can be estimated similarly as
\[
\sum_{i,j} ((G_2 - m_2)A_3)_{ji}(G_2^2B_3G_1)_{ji} = \omega_3 \varpi_3 m_1 \left( \frac{m''_2}{2} - m_2 m'_2 \right) + O_\prec(n^{-\frac{1}{2}}).
\]
Consequently, we have
\[
\frac{1}{z} \sum_{i,j} ((G_2 - m_2)A_3)_{ji}(X^*G_1^2XB_3G_1)_{ji} + \sum_{i,j} ((G_2 - m_2)A_3)_{ji}(G_2^2B_3G_1)_{ji}
\]
\[
= \sum_{i,j} ((\Pi_3^1 - \Pi_3^{1*})A_1A_1)_{ji}B_{ji} + O_\prec(n^{-1/2}).
\]
Note that the sum of the four contributing terms is exactly
\[
\text{Tr}((\Pi_3^1 - \Pi_3^{1*})A_1A_1B) + O_\prec(n^{-\frac{1}{2}}).
\]
To wrap up the proof of (7.36), it suffices to show all the other terms in the expansion of \(\sum_{i,j} (\Xi_1 A_1)_{ji}(G^2BG)_{ji}\) can be bounded by \(O_\prec(n^{-\frac{1}{2}})\). To see that, for instance, we focus on
\[
z^{-3/2} \sum_{i,j} (X^*G_1A_1)_{ji}(X^*G_1^2XB_3G_1)_{ji} = \omega_1 \varpi_1 (z^{-3/2}v^*X^*G_1^2X^*G_1u)(u^*G_1u).
\]
Note that
\[
z^{-3/2}v^*X^*G_1^2X^*G_1u = \bar{u}^*G^3\bar{v} - u^*\left( \frac{1}{\sqrt{z}}G_1^3X + \frac{1}{\sqrt{z}}G_1^2XG_2 + \frac{1}{\sqrt{z}}G_1XG_2^2 \right)v,
\]
\[
= \bar{u}G^3\bar{v} - \frac{3}{z^2}u^*G_1^3Xv = \bar{u}G^3\bar{v} - \frac{3}{2\sqrt{z}}u^*(\sqrt{z}G^2)''v = O_\prec(n^{-\frac{1}{2}}),
\]
where in the third step we use \(G_3^3 = \frac{1}{2}G''_2\) and in the last step we use (4.18). Consequently,
\[
z^{-3/2} \sum_{i,j} (X^*G_1A_1)_{ji}(X^*G_1^2XB_3G_1)_{ji} = O_\prec(n^{-\frac{1}{2}}).
\]
All the rest terms can be bounded by \(O_\prec(n^{-\frac{1}{2}})\) analogously; we omit the details. The proof of (7.36) is now complete.

The remaining two terms in (7.32) can be estimated the same way as (7.33) and (7.36); the details are omitted. We get
\[
\sum_{i,j} (\Xi_1 A_1)_{ji}(GBG^2)_{ji} = \text{Tr}(\Pi_2^1 - \Pi_2^{1*})A_1A_2B + O_\prec(n^{-\frac{1}{2}}),
\]
\[
\sum_{i,j} (\Xi_1 A_1)_{ji}(GBG)_ji = \text{Tr}(\Pi_2^1 - \Pi_2^{1*})A_1B + O_\prec(n^{-\frac{1}{2}}).
\]
Plugging (7.33), (7.36) and (7.39) into (7.32), recalling the definition of $a_{11}$ in (6.9), we conclude that

$$h(0, 1, 0) = a_{11} q^{(k-2)} + O_w(n^{-\frac{3}{2}+4\nu}).$$  

(7.40)

This completes the proof.  

\[ \Box \]

7.2. Proof of Lemma 7.3. We use this subsection to prove Lemma 7.3.

Proof of Lemma 7.3. We first study the second derivatives. By (7.11) and (A.1), we have

$$h(2, 0, 0) = \frac{1}{n} \sum_{i,j} \left( \frac{\partial^2 G}{\partial x_{ij}^2} A_1 \right) j_i q^{(k-1)} = \frac{2\pi}{n} \sum_{i,j} \left( (G_{j,j'}G_{j,j'} + G_{j',j}G_{j,j'}) (GA_1)_{ii} + (G_{j,j'}G_{j,i} + G_{j',j}G_{j,i}) (GA_1)_{j,i} \right) q^{(k-1)}.$$  

First of all, by (4.13) and (4.18), we find that

$$-\frac{1}{n} \sum_{i,j} G_{j,j'}G_{ii}(GA_1)_{j,i} = \frac{1}{n} \sum_{i,j} (\Pi_1)_{ii}(\Pi_1)_{j,j'}(\Pi_1A_1)_{j,i} + O_w(n^{-\frac{3}{2}}).$$

It is simple to check that $(GA)_{ii} = (G_1A_1 + z^{-1/2}G_1X_A)_i$. By (4.11) and (4.18), we get

$$\frac{1}{n} \sum_{i,j} G_{j,j'}G_{ii}(GA_1)_{ii} = O_w \left( n^{-\frac{7}{2}} \sum_{i,j} (G_1uu^* + G_1Xvv^*)_i \right) = O_w(n^{-\frac{7}{2}}).$$

Similarly, we also have

$$-\frac{1}{n} \sum_{i,j} G_{j,j'}G_{ji}(GA_1)_{ii} = O_w(n^{-\frac{3}{2}}),$$

$$-\frac{1}{n} \sum_{i,j} G_{j,j'}G_{ji}(GA_1)_{j,i} = O_w(n^{-1}).$$

Putting the above estimates together, and recalling $\delta^n_i$ in (6.7), we conclude that

$$h(2, 0, 0) = \delta^n_i q^{(k-1)} + O_w(n^{-\frac{7}{2}}).$$

Next, the estimation of

$$h(1, 1, 0) = \frac{k-1}{n} \sum_{i,j} \left( \frac{\partial G}{\partial x_{ij}} A_1 \right) j_i \frac{\partial Q}{\partial x_{ij}} q^{(k-2)}$$

follows closely the same steps as the derivation of (7.13). By (7.28),

$$h(1, 1, 0) = -\frac{(k-1)\sqrt{\pi}}{n} \sum_{i,j} \left( G_{j,j'}(GA_1)_{ii} + G_{j,i}(GA_1)_{j,i} \right) \frac{\partial Q}{\partial x_{ij}} q^{(k-2)}.$$  

(7.42)

We shall prove that

$$-\frac{1}{n} \sum_{i,j} G_{j,j'}(GA_1)_{ii} \frac{\partial Q}{\partial x_{ij}} = \frac{\sqrt{\pi}}{n} \sum_{i,j} \left( \Pi_1 A \right)_{ii}(\Pi_1)_{j,j'} c_{ij} + O_w(n^{-\frac{1}{2}+4\nu}),$$  

(7.43)

which will be used several times later. We postpone the proof of (7.43) till the end of this subsection.

Again by (4.18), recalling the definitions of $c_{ij}$ in (5.6) and $A$ in (4.24), we have that

$$-\frac{1}{n} \sum_{i,j} G_{j,i}(GA_1)_{j,i} \frac{\partial Q}{\partial x_{ij}} = O_w(n^{-3/2}) \sum_{i,j} A_{ij} c_{ij} = O_w(n^{-1}).$$  

(7.44)
Inserting (7.43) and (7.44) back into (7.42), by recalling \( a_{21} \) in (6.10), we conclude that
\[
h(1, 1, 0) = a_{21}q^{(k-2)} + O_{\prec}(n^{-\frac{1}{2}+4\nu}).
\]

Using a discussion similar to (7.14), we also have
\[
h(1, 0, 1) = \sqrt{\frac{2}{n}} \sum_{(i,j) \in B(\nu)} \frac{\partial G}{\partial x_{ij}} A_1 \, j_i \, q^{(k-1)} = O_{\prec}(n^{-\frac{1}{2}+4\nu}).
\]

Actually, all the terms containing the derivatives of \( h_3 \) can be estimated in the same way. Thus both \( h(0, 1, 1) \) and \( h(0, 0, 2) \) are also bounded by \( O_{\prec}(n^{-\frac{1}{2}+4\nu}) \). We omit the details.

It remains to estimate
\[
h(0, 2, 0) = \frac{1}{n} \sum_{i,j} (\Xi A_1) j_i \left( (k-1) \frac{\partial^2 Q}{\partial x_{ij}} q^{(k-2)} + (k-1)(k-2) \left( \frac{\partial Q}{\partial x_{ij}} \right)^2 q^{(k-3)} \right). \quad (7.45)
\]

The calculation of (7.45) is similar to that of (7.32) and due to an extra factor \( n^{-1/2} \) in front, we shall show that \( h(0, 2, 0) \) can be bounded by \( O_{\prec}(n^{-1/2}) \). We only list the main differences here. We expand the product on the right hand side of (7.45) using the expressions of \((\Xi A_1) j_i\) in (7.34), \( \partial Q/\partial x_{ij} \) in (7.1) and \( \partial^2 Q/\partial x_{ij}^2 \) in (A.5).

Most derivations of the items in (7.32) can be directly applied to those in (7.45) except three items, which are discussed below. Denote \( e_i \) with \( i \in [M] \) as the standard basis in \( \mathbb{R}^M \) and \( f_j \) with \( j \in [N] \) as those in \( \mathbb{R}^N \).

First, by (4.13) and (4.18),
\[
\sum_{i,j} (X^* G_i A_1 j_i (X^* G_i A_1 G_1) j_i (X^* G_i A_1 A_1 G_1) j_i) = \sum_{i,j} (e_j^* X^* G_i A_1 e_i)(e_j^* X^* G_i A_1 G_1 e_i)^2
\]
\[
= O_{\prec} \left( n^{-\frac{1}{2}} \sum_{i,j} u^2(i) \right) = O_{\prec}(n^{-\frac{1}{2}}). \quad (7.46)
\]

Second, using (4.13) and the fact that \( u, v \) are unit vectors, we get
\[
\frac{1}{\sqrt{n}} \sum_{i,j} (X^* G_i A_1 j_i (GAG) j_i A_1 j_i A_1 j_i) = \frac{m_1 m_2}{\sqrt{n}} \sum_{i,j} (X^* G_i A_1 j_i A_1 j_i) + O_{\prec}(n^{-\frac{1}{2}}),
\]
\[
= \frac{m_1 m_2}{\sqrt{n}} \sum_{i,j} f_j^* X^* G_i u u(i) v^2(j) + O_{\prec}(n^{-\frac{1}{2}})
\]
\[
= \frac{m_1 m_2}{n} \sum_{i,j} u(i) v^2(j) + O_{\prec}(n^{-\frac{1}{2}})
\]
\[
= O_{\prec}(n^{-\frac{1}{2}}).
\]
Third, we invoke (4.13) to get that \( \sum_j f_j X^* G_1 u = \sqrt{n} f X^* G_1 u = O_\prec(1) \), where \( f = \frac{1}{\sqrt{n}} 1 \). Then it follows that

\[
\frac{1}{\sqrt{n}} \sum_{i,j} (X^* G_1 A_1)_{ji} (GAG)_{ii} G_{j'} = \frac{m_1^2 m_2}{\sqrt{n}} \sum_{i,j} (X^* G_1 A_1)_{ji} A_{ii} + O_\prec(n^{-\frac{1}{2}}),
\]

\[
= \frac{m_1^2 m_2 \omega^2}{\sqrt{n}} \sum_{i,j} f_j X^* G_1 uu^3(i) + O_\prec(n^{-\frac{1}{2}})
\]

\[
= \frac{m_1^2 m_2 \omega^2}{\sqrt{n}} \left( \sum_i u^3(i) \right) \left( \sum_j f_j X^* G_1 u \right) + O_\prec(n^{-\frac{1}{2}})
\]

\[
= O_\prec(n^{-\frac{1}{2}}).
\]

Finally, we can conclude that

\[
h(0,2,0) = O_\prec(n^{-\frac{1}{2}}).
\]

This finishes the discussion of the second order derivatives. We continue with the third derivatives. We start with

\[
h(1,2,0) = n^{-\frac{3}{2}} \sum_{i,j} \left( \frac{\partial G}{\partial x_{ij}} A_1 \right)_{ji} \left( (k - 1) \frac{\partial^2 Q}{\partial x^2_{ij}} q^{(k-2)} + (k - 1)(k - 2) \left( \frac{\partial Q}{\partial x_{ij}} \right)^2 q^{(k-3)} \right).
\]

(7.48)

Recalling (7.28) and (A.5), by (4.13) and (4.18), the first term on the right hand side of (7.48) is estimated by

\[
n^{-\frac{3}{2}} \sum_{i,j} \left( \frac{\partial G}{\partial x_{ij}} A_1 \right)_{ji} \left( \frac{\partial^2 Q}{\partial x^2_{ij}} \right)
\]

\[
= -\frac{2z^2}{n} \sum_{i,j} G_{j'}(GA_1)_{ii} \left( (GAG)_{ii} G_{j'} - \frac{1}{2z} (GBG)_{ii} G_{j'} + \frac{1}{2} (GBG)_{ii} G_{j'}^2 \right)
\]

\[
+ \frac{1}{2} (G^2 BG)_{ii} G_{j'} + \frac{1}{2} (G^2 BG)_{ii} G_{j'}^2 ) + O_\prec(n^{-\frac{1}{2}})
\]

\[
= -\frac{2z^2}{n} \sum_{i,j} (\Pi_1 A)_{ii} (\Pi_1)_{j'j} s_{ij} + O_\prec(n^{-1/2}).
\]

(7.49)

In the last equation above, we recall the definition of \( s_{ij} \) in (5.7). Furthermore, recalling (7.2) and (7.28), by (4.18), it is easy to see that the second term on the right hand side of (7.48) is

\[
n^{-3/2} \sum_{i,j} \left( \frac{\partial G}{\partial x_{ij}} A_1 \right)_{ji} \left( \frac{\partial Q}{\partial x_{ij}} \right)^2 q^{(k-3)} = O_\prec \left( n^{-3/2} \sum_{i,j} (A_1)_{ii} c^2_{ij} \right) = O_\prec(n^{-1}).
\]

For the last equation above, we refer to the definition of \( c_{ij} \) in (5.6). Using \( a_{31} \) defined in (6.11), we hence conclude that

\[
h(1,2,0) = a_{31} q^{(k-2)} + O_\prec(n^{-1/2}).
\]
Next we study
\[ h(0, 3, 0) = n^{-\frac{3}{2}} \sum_{i,j} (\Xi_1 A_1)_{ij} \left( (k - 1)! \frac{\partial^2 Q}{\partial x_{ij}^2} q^{(k-2)} + \frac{3}{2} (k - 1)! \frac{\partial^2 Q}{\partial x_{ij}^2} \frac{\partial Q}{\partial x_{ij}} q^{(k-3)} + \frac{(k - 1)!}{(k - 4)!} \left( \frac{\partial Q}{\partial x_{ij}} \right)^3 q^{(k-4)} \right). \]

We briefly argue that \( h(0, 3, 0) \) is bounded by \( O(n^{-\frac{3}{2}}) \), using a discussion similar to those of \( h(0, 1, 0) \) in (7.32) and \( h(0, 2, 0) \) in (7.45).

Recalling (7.1) and (A.5), it is easy to see that
\[ n^{-\frac{3}{2}} \sum_{i,j} (\Xi_1 A_1)_{ij} \frac{\partial^2 Q}{\partial x_{ij}^2} \frac{\partial Q}{\partial x_{ij}} = O(n^{-1} \sum_{i,j} u(i) v^3(j)) = O(n^{-\frac{1}{2}}). \]

Similarly, by (7.1) and (A.6), it can also be shown that
\[ n^{-\frac{3}{2}} \sum_{i,j} (\Xi_1 A_1)_{ij} \frac{\partial^3 Q}{\partial x_{ij}^3} = O(n^{-\frac{1}{2}} \sum_{i,j} u^3(i) v^3(j)) = O(n^{-\frac{1}{2}}), \]
\[ n^{-\frac{3}{2}} \sum_{i,j} (\Xi_1 A_1)_{ij} \frac{\partial^4 Q}{\partial x_{ij}^4} = O(n^{-\frac{1}{2}}). \]

This completes the discussion of \( h(0, 3, 0) \). The same arguments can be applied to show that
\[ h(2, 1, 0) = n^{-\frac{3}{2}} \sum_{i,j} \left( \frac{\partial^2 G}{\partial x_{ij}^2} A_1 \right)_{ij} \frac{\partial Q}{\partial x_{ij}} q^{(k-2)} = O(n^{-1} \sum_{i,j} u^2(i) v^2(j)) = O(n^{-1}). \]
and
\[ h(3, 0, 0) = n^{-\frac{3}{2}} \sum_{i,j} \left( \frac{\partial^3 G}{\partial x_{ij}^3} A_1 \right)_{ij} q^{(k-1)} = O(n^{-3/2} \sum_{i,j} (A_1)_{ij}) = O(n^{-\frac{1}{2}}) \]
by using the expressions (A.1) and (A.2) respectively.

For all the rest of the items containing the derivatives of \( h_3 \), they can be easily estimated using a discussion similar to (7.14).

Finally, using (A.1)-(A.7), (7.1), (7.28) and (4.13), all the fourth order derivatives can bounded by \( O(n^{-\frac{1}{2}}) \). The discussion is similar to that of (7.45); we omit further details here. This concludes our proof.

**Proof of (7.43).** We split the left hand side of (7.43) as the sum of the following three items
\[ \frac{1}{n} \sum_{i,j} (G_{j'j'} - m_2)(GA_1)_{ij} \frac{\partial Q}{\partial x_{ij}}, \quad \frac{1}{n} \sum_{i,j} m_2(\Xi_1 A_1)_{ij} \frac{\partial Q}{\partial x_{ij}}, \quad \frac{1}{n} \sum_{i,j} m_1 m_2 (A_1)_{ij} \left( \frac{\partial Q}{\partial x_{ij}} - \sqrt{n} z_{ij} \right). \]

First of all, by (4.18) and (7.1), we have
\[ \frac{1}{n} \sum_{i,j} (G_{j'j'} - m_2)(GA_1)_{ij} \frac{\partial Q}{\partial x_{ij}} = O(n^{-1} \sum_{(i,j) \in S(\nu)} u^3(i) v(j)) = O(n^{-\frac{1}{2}}). \]

Similarly, we also have
\[ \frac{1}{n} \sum_{i,j} m_2(\Xi_1 A_1)_{ij} \frac{\partial Q}{\partial x_{ij}} = \frac{m_2 \omega_1}{n} \sum_{i,j} c_i \Xi_1 u u(i) \frac{\partial Q}{\partial x_{ij}} = O(n^{-1} \sum_{i,j} u(i)^2 v(j)) = O(n^{-\frac{1}{2}}). \]
Furthermore, using a discussion similar to that of (7.47), we get
\[
\frac{1}{n} \sum_{i,j} m_1 m_2 (A_{11})_{ij} \left( \frac{\partial Q}{\partial x_{ij}} - \sqrt{n} \varepsilon_{ij} \right) = O_\prec (n^{-\frac{1}{2}}) \sum_i u^3(i) = O_\prec (n^{-\frac{1}{2}}),
\]
where we apply the fact that
\[(GAG)_{j^1} - m_1 m_2 A_{j^1} = O_\prec \left( \frac{u(i)}{\sqrt{n}} \right).\]

Summing up the above three estimates, we can conclude the proof of (7.43). \qed

7.3. Proofs of Lemma 7.4-7.7. In this subsection, we will prove Lemma 7.4-7.7. The proofs are analogous to those of Lemma 7.2 and Lemma 7.3; we only outline the main steps.

We record a basic identity for later estimates. For any deterministic matrix \( W \in \mathbb{R}^{(M+n) \times (M+n)} \), it is elementary to check that
\[(\partial G^2/\partial x_{ij})_{ab} = -\sqrt{z}(G_{aj'}^2(GW)_{ib} + G_{aj}^2(GM)_{j'b} + G_{aj'}(G^2W)_{ib} + G_{ai}(G^2W)_{j'b}).\] (7.50)

**Proof of Lemma 7.4.** Recalling \( \tilde{h}_1 \) in (7.20), by a discussion similar to (7.30), we get
\[
\tilde{h}(0, 0) = \frac{1}{\sqrt{n}} \sum_{i,j} \tilde{h}_1 \partial_{x_{ij}} h_2 h_3 = \frac{1}{\sqrt{n}} \sum_{i,j} (\partial G^2/\partial x_{ij} \partial Q / \partial x_{ij})_{j^1} q^{(k-1)} - \sqrt{z} \left( (2m'_2 + \frac{m_2}{z}) \text{Tr}(GB_1) + m_2 \text{Tr}(G^2B_1) \right) q^{(k-1)} + O_\prec (n^{-\frac{1}{2}}).
\]
In the last step above, we use (7.50), (4.13) and (4.18). Next, we turn towards the term
\[
\tilde{h}(0, 1) = \frac{1}{\sqrt{n}} \sum_{i,j} \tilde{h}_1 \partial_{x_{ij}} h_2 h_3 = \frac{1}{n} \sum_{i,j} (\Xi_2B_1)_{j^1} \partial Q / \partial x_{ij} q^{(k-2)},
\]
which will be estimated following exactly the same steps as those of (7.13). Observe that
\[(\Xi_2B)_{j^1} = \left( \left( \frac{1}{\sqrt{z}} X^*G_2^2 + \frac{1}{\sqrt{z}} G_2X^*G_1 \right) B_1 + \left( \frac{1}{z} X^*G_2^2X + G_2^2 - 2m'_2 - \frac{m_2}{z} \right) B_1 \right)_{j^1}.
\]

By (7.1), after expanding the product on the right hand side of (7.51), we find the following estimates.
\[
\frac{1}{\sqrt{n}} \sum_{i,j} (\Xi_2B_1)_{j^1} (GAG)_{j^1} = \text{Tr}(\Pi_3^1 - \Pi_{1,2}^1) B_1 \Pi_1 A + O_\prec (n^{-\frac{1}{2}}),
\]
\[
\frac{1}{\sqrt{n}} \sum_{i,j} (\Xi_2B_1)_{j^1} (GBG)_{j^1} = \text{Tr}(\Pi_3^1 - \Pi_{1,2}^1) B_1 \Pi_1 B + O_\prec (n^{-\frac{1}{2}}),
\]
\[
\frac{1}{\sqrt{n}} \sum_{i,j} (\Xi_2B_1)_{j^1} (G^2BG)_{j^1} = \text{Tr}(\Pi_3^1 - \Pi_{1,2}^1) B_1 \Pi_2 B + O_\prec (n^{-\frac{1}{2}}),
\]
\[
\frac{1}{\sqrt{n}} \sum_{i,j} (\Xi_2B_1)_{j^1} (G^2G^2)_{j^1} = \text{Tr}(\Pi_3^1 - \Pi_{1,2}^1) B_1 \Pi_2 B + O_\prec (n^{-\frac{1}{2}}).
\]

Putting these estimates together and invoking \( \hat{b}_{111} \) in (6.9), we have
\[
\tilde{h}(0, 1, 0) = \hat{b}_{111} q^{(k-2)} + O_\prec (n^{-\frac{1}{2}+4v}).
\]

Lastly, \( \tilde{h}(0, 0, 1) \) can be estimated using a discussion similar to (7.14). We can therefore conclude our proof. \qed
The proof of Lemma 7.5 follows along the exact lines of Lemma 7.3 with minor changes. We only sketch it below.

**Proof of Lemma 7.5.** First of all, by (4.18) and (A.4), using a discussion similar to (7.15), we have that

\[
\tilde{h}(2, 0, 0) = \frac{1}{n} \sum_{i,j} \frac{\partial^2 \tilde{h}_1}{\partial x_{ij}^2} h_2 h_3 = \frac{1}{n} \sum_{i,j} \left( \frac{\partial^2 G^2}{\partial x_{ij}^2} B_1 \right) \frac{q^{(k-1)}}{i'j'} = \tilde{q}_1 q^{(k-1)} + O_\prec(n^{-\frac{1}{2}}),
\]

Second, following the same steps in (7.42), together with (7.50) and (7.1), we find that

\[
\tilde{h}(1, 1, 0) = \frac{1}{n} \sum_{i,j} \frac{\partial \tilde{h}_1}{\partial x_{ij}} \frac{\partial h_2}{\partial x_{ij}} h_3 = \frac{k - 1}{n} \sum_{i,j} \left( \frac{\partial G}{\partial x_{ij}} B_1 \right) \frac{\partial Q}{\partial x_{ij}} q^{(k-2)} = \tilde{b}_{21} q^{(k-2)} + O_\prec(n^{-\frac{1}{2}}).
\]

Third, the same arguments of (7.17) using (7.50) and (A.5) yield

\[
\tilde{h}(1, 2, 0) = n^{-\frac{1}{2}} \sum_{i,j} \frac{\partial \tilde{h}_1}{\partial x_{ij}} \frac{\partial^2 h_2}{\partial x_{ij}^2} h_3 = \tilde{b}_{31} q^{k-2} + O_\prec(n^{-1/2}).
\]

For the rest of the items, we can apply discussions similar to those of the corresponding items from Lemma 7.3. We omit the details here.

\[
\square
\]

Lemma 7.6 is an analogue of Lemma 7.2 and 7.4 for the matrices \(A_2\) and \(B_2\); the proof is analogous.

**Proof of Lemma 7.6.** Recall (7.28). Using a discussion similar to that of (7.30), by (4.13) and (4.18), we find that

\[
b(1, 0, 0) = \frac{1}{\sqrt{n}} \sum_{i,j} \frac{\partial \tilde{h}_1}{\partial x_{ij}} h_2 h_3 = \frac{1}{\sqrt{n}} \sum_{i,j} \left( \frac{\partial G}{\partial x_{ij}} A_2 \right) \frac{\partial q^{(k-1)}}{i'j'} = -\sqrt{n^2} \frac{1}{n} \sum_{i,j} \left( G_{ii}(GA_2)_{j'j'} + G_{jj}(GA_2)_{ii'} \right) q^{(k-1)}
\]

\[
= -\sqrt{n} y m_1 (\text{Tr} G A_2) q^{(k-1)} - \frac{\sqrt{n}}{\sqrt{7}} (\text{Tr} G A^2_2 G) q^{(k-1)}
\]

\[
= -\sqrt{n} y m_1 (\text{Tr} G A_2) q^{(k-1)} + O_\prec(n^{-\frac{1}{2}}),
\]

where we recall that \(y = \frac{M}{n}\). Similarly, using (7.50), we also have

\[
b(1, 0, 0) = \frac{1}{\sqrt{n}} \sum_{i,j} \frac{\partial \tilde{h}_1}{\partial x_{ij}} h_2 h_3 = \frac{1}{\sqrt{n}} \sum_{i,j} \left( \frac{\partial G^2}{\partial x_{ij}} B_2 \right) \frac{q^{(k-1)}}{i'j'}
\]

\[
= -\sqrt{n^2} \frac{1}{n} \sum_{i,j} \left( G_{ii}(GB_2)_{j'j'} + G_{jj}(GB_2)_{ii'} \right) q^{(k-1)} + O_\prec(n^{-\frac{1}{2}})
\]

\[
= -\sqrt{n} y \left( (2m_1 + m_1^2) (\text{Tr} GB_2) + m_1 \text{Tr} G^2 B_2 \right) q^{(k-1)} + O_\prec(n^{-\frac{1}{2}}),
\]

Next, we estimate

\[
b(0, 1, 0) = \frac{k - 1}{\sqrt{n}} \sum_{i,j} (\Xi_1 A_2)_{i'j'} \frac{\partial Q}{\partial x_{ij}} q^{(k-2)}.
\]
Recall the expression of $\partial Q / \partial x_{ij}$ in (7.1). As seen in the discussion below (7.32), the key observation is that

$$\sum_{i,j} (\Xi_1 A_2)_{ij'} (GAG)_{ij'} = \text{Tr}(G^4 G - G^4 \Pi_1) A_2 G A$$

Thus we shall prove

$$\sum_{i,j} (\Xi_1 A_2)_{ij'} (GAG)_{ij'} = \text{Tr}(\Pi_2^2 - \Pi_{1,1}^2) A_2 A_1 A + O_{\sim}(n^{-\frac{3}{2}}).$$

The proof follows from

$$(\Xi_1 A_2)_{ij'} = ((G_1 - m_1) A_2 + z^{-1/2} G_1 X A_4)_{ij}$$

and exactly the same arguments as (7.33). Likewise, we also get

$$\frac{1}{\sqrt{n}} \sum_{i,j} (\Xi_1 A_2)_{ij'} (G G B G)_{ij'} = \text{Tr}(\Pi_2^3 - \Pi_{1,1}^3) A_2 A_1 B + O_{\sim}(n^{-\frac{3}{2}}),$$

$$\frac{1}{\sqrt{n}} \sum_{i,j} (\Xi_1 A_2)_{ij'} (G^2 G B G)_{ij'} = \text{Tr}(\Pi_3^2 - \Pi_{2,1}^2) A_2 A_1 B + O_{\sim}(n^{-\frac{3}{2}}),$$

$$\frac{1}{\sqrt{n}} \sum_{i,j} (\Xi_1 A_2)_{ij'} (G B G^2)_{ij'} = \text{Tr}(\Pi_2^2 - \Pi_{1,1}^2) A_2 A_2 B + O_{\sim}(n^{-\frac{3}{2}}).$$

Putting the above estimates together and recalling $\alpha_{12}$ below (6.9), we finish the computation for (7.23). The other term $\tilde{h}(0, 1, 0)$ can be estimated analogously by noting

$$(\Xi_2 B_2)_{ij'} = ((G_1^2 + \frac{1}{z} G_1 X X^* G_1 - 2m_1^2 - \frac{m_1}{z}) B_2 + (\frac{1}{\sqrt{z}} G_1^2 X + \frac{1}{\sqrt{z}} G_1 X G_1) B_2)_{ij}.$$  

Finally, $\tilde{h}(0, 0, 1)$ and $\tilde{h}(0, 0, 1)$ can be estimated using a discussion similar to that of (7.14). The details are omitted.  

The remaining part of this section is the proof of Lemma 7.7, which is an analogue of Lemma 7.3 and 7.5 for the matrices $A_2$ and $B_2$.

**Proof of Lemma 7.7.** We shall outline our computation on the dominating terms. The discussions of the negligible terms are similar to those in Lemma 7.3 and 7.5, and are therefore omitted.

Recall (A.1). Using the same proof of (7.15), we first get

$$\tilde{h}(2, 0, 0) = \frac{1}{n} \sum_{i,j} \left( \frac{\partial^2 h_{ij}}{\partial x_{ij}^2} \right) h_{ij} h_{ij} h_{ij} h_{ij} = \frac{1}{n} \sum_{i,j} \left( \frac{\partial^2 G}{\partial x_{ij}^2} A_2 \right)_{ij'} q^{(k-1)}$$

$$= \frac{2z}{n} \sum_{i,j} G_{ij} G_{ij'} (G A_2)_{ij'} q^{(k-1)} + O_{\sim}(n^{-\frac{3}{2}})$$

$$= \frac{2z}{n} \sum_{i,j} (\Pi_1)_{ij} (\Pi_1)_{ij'} (\Pi_1 A_2)_{ij'} q^{(k-1)} + O_{\sim}(n^{-\frac{3}{2}})$$

$$= \alpha_{12}^2 q^{(k-1)} + O_{\sim}(n^{-\frac{3}{2}}).$$

Likewise, applying (A.4), we find that

$$\tilde{h}(2, 0, 0) = \frac{1}{n} \sum_{i,j} \left( \frac{\partial^2 h_{ij}}{\partial x_{ij}^2} \right) h_{ij} h_{ij} h_{ij} h_{ij} = \frac{1}{n} \sum_{i,j} \left( \frac{\partial^2 G^2}{\partial x_{ij}^2} B_2 \right)_{ij'} q^{(k-1)} = \bar{\delta}_{2} q^{(k-1)} + O_{\sim}(n^{-\frac{3}{2}}).$$
Next, recall (7.28) and (7.1). By a discussion similar to that of (7.42), we conclude that
\[ h(1, 1, 0) = \frac{1}{n} \sum_{i,j} \frac{\partial h_1}{\partial x_{ij}} \frac{\partial h_2}{\partial x_{ij}} h_3 \]
\[ = -\frac{(k-1)\epsilon}{\sqrt{n}} \sum_{i,j} (\Pi_1)_{ij} (\Pi_1 A_2)_{ij} c_{ij} q^{(k-2)} + O(n^{-\frac{1}{2}+4\nu}) \]
\[ = a_{22} q^{(k-2)} + O(n^{-\frac{1}{2}+4\nu}). \]

Similarly, recalling (7.50), we have
\[ \tilde{h}(1, 1, 0) = \frac{1}{n} \sum_{i,j} \frac{\partial \tilde{h}_1}{\partial x_{ij}} \frac{\partial h_2}{\partial x_{ij}} h_3 = a_{22} q^{(k-2)} + O(n^{-\frac{1}{2}}). \]

Finally, recall (A.3) and (A.5). The same arguments as (7.49) yield
\[ h(1, 2, 0) = n^{-\frac{3}{2}} \sum_{i,j} \frac{\partial h_1}{\partial x_{ij}} \frac{\partial^2 h_2}{\partial x_{ij}^2} h_3 = a_{32} q^{(k-2)} + O(n^{-\frac{1}{2}}), \]
\[ \tilde{h}(1, 2, 0) = n^{-\frac{3}{2}} \sum_{i,j} \frac{\partial \tilde{h}_1}{\partial x_{ij}} \frac{\partial^2 h_2}{\partial x_{ij}^2} h_3 = \tilde{a}_{32} q^{(k-2)} + O(n^{-\frac{1}{2}}). \]

This concludes our proof.

8. **Proof of Theorems 2.9 and 2.11**

In this section, we prove Theorem 2.9. The proof follows along the same lines of the proof of Theorem 2.3, and is summarized as follows. First, by Lemma 4.8, we reduce the problem to study the quantity \( Q \) defined below. After necessary notations are introduced, as done in the beginning of Section 5, it suffices to prove Proposition 8.1, which is an analogue of Proposition 5.1. The proof of Proposition 8.1 essentially relies on a recursive estimate presented in Proposition 8.2. Thus the main goal of this section is to prove Proposition 8.2. We shall apply the same arguments as the proof of its counterpart, Proposition 5.2, and emphasize the differences.

Let \( z = (z_1, \cdots, z_r) \) denote a vector with all the entries \( z_\beta \in S_0 \). Following the discussion in the beginning of Section 5, with a slight abuse of notation, we introduce a few definitions.

Let
\[ Q \equiv Q(z) := \sqrt{n} \sum_{\beta=1}^r \left( \text{Tr}(\Xi_1(z_\beta)) A_\beta^R + \text{Tr}(\Xi_1'(z_\beta)) B_\beta^N \right). \]

Denote the index set as
\[ B(\nu) := \bigcup_{\beta=1}^r B_\beta(\nu), \]
where \( B_\beta(\nu) \) is defined as
\[ B_\beta(\nu) := \left\{ (i, j) \in [M] \times [n] : |u_\beta(i)| > n^{-\nu}, |v_\beta(j)| > n^{-\nu} \right\}. \]

Since \( r \) is fixed and all the vectors \( u_\beta \) and \( v_\beta \) for \( \beta \in [r] \) are unit vectors, it is easy to conclude that \( |B(\nu)| \leq C n^{4\nu} \) for some constant \( C > 0 \).

For \( \beta \in [r] \), invoke \( \Delta_\beta(z_\beta) \) by plugging \( z_\beta \) in (5.4). We also introduce the random variable
\[ \Delta_r(z_\beta) := \sqrt{n} \sum_{(i, j) \in B(\nu)} x_{ij}(c_\beta)_{ij}, \]
where \((c_\beta)_{ij} \equiv (c_\beta(z_\beta))_{ij}\) is defined by inserting \(z_\beta\) into \(c_{ij}\) in (5.6). Similarly, we denote \((s_\beta)_{ij} \equiv (s_\beta(z_\beta))_{ij}\) by plugging \(z_\beta\) into \(s_{ij}\) in (5.7). Let \(C_\beta\) and \(S_\beta\) be \(M \times n\) matrices with entries \((c_\beta)_{ij}\) and \((s_\beta)_{ij}\) respectively. Denote
\[
\Delta_r \equiv \Delta_r(z) := \sum_{\beta=1}^r \Delta_r(z_\beta), \quad \Delta_d \equiv \Delta_d(z) := \sum_{\beta=1}^r \Delta_d(z_\beta),
\]
and
\[
\Delta = \Delta_d + \Delta_r.
\]
Furthermore, we denote
\[
Q = Q - \Delta.
\]
Note \(V_E(z_\beta)\) is defined in (5.9) by plugging \(z_\beta\).

Set
\[
V_E(z) = \sum_{\beta=1}^r V_E(z_\beta).
\]

Then we define the function
\[
V \equiv V(z)
= V_E(z) + 2\frac{\kappa_3}{\sqrt{n}} \text{Tr} \left( \left( \sum_{\beta=1}^r z_\beta S_\beta \right) \left( \sum_{\beta=1}^r \sqrt{z_\beta} C_\beta \right)^* \right) + \frac{\kappa_4}{n} \text{Tr} \left( \left( \sum_{\beta=1}^r z_\beta S_\beta \right) \left( \sum_{\beta=1}^r z_\beta S_\beta \right)^* \right)
+ \sum_{(i,j) \in S(\nu)} \left( \sum_{\beta=1}^r \sqrt{z_\beta(c_\beta)_{ij}} \right)^2.
\]

Recall \(p_\beta = p(d_\beta)\) in (2.5). Let
\[
z_0 := (p_1, \ldots, p_r).
\]

**Proposition 8.1.** Under the assumptions of Theorem 2.9, we have that \(Q(z_0)\) and \(\Delta(z_0)\) are asymptotically independent. Furthermore,
\[
Q(z_0) \sim N(0, V(z_0)).
\]

Theorem 2.9 follows from Proposition 8.1. The arguments are the same as the proof of Theorem 2.3 in Section 5. Again, the final presentation of the results in Theorem 2.9 are obtained by plugging the values \(p_\beta\) for \(1 \leq \beta \leq r\) using the continuity of Green functions and performing tedious calculations. We omit the details.

Similar to the discussion of Proposition 5.1, to prove Proposition 8.1, it suffices to establish the following recursive moment estimates. It is an analogue of Proposition 5.2.

**Proposition 8.2.** Suppose the assumptions of Theorem 2.9 hold. Let \(z_\beta = p_\beta + in^{-C}\) and \(z_{0\beta} = p_\beta\) for all \(\beta \in [r]\). We have
\[
E Q(z_\beta)e^{it\Delta(z_{0\beta})} = O_{\prec}(n^{-1/2+\nu}),
\]
and for any fixed integer \(k \geq 2\),
\[
E Q^k(z_\beta)e^{it\Delta(z_{0\beta})} = (k-1)V E Q^{k-2}(z_\beta)e^{it\Delta(z_{0\beta})} + O_{\prec}(n^{-1/2+\nu}).
\]

Proposition 8.2 can be proved in a way similar to Proposition 5.2. Recall from Section 6 that the proof of Proposition 5.2 is based on Lemma 6.1 and Lemma 6.2. We present the analogues of these two lemmas and their proofs in the following two steps. We shall only outline the key estimates and focus on discussing the differences.
**Step 1.** In the first step, we will rewrite $Q$ in (8.1). Recall (6.2) and for each $\beta \in [r]$, denote

$$A_{\beta,1} := A^R_{\beta 1}, \quad A_{\beta,1} := A^R_{\beta 1},$$

$$B_{\beta,1} := B^R_{\beta 1}, \quad B_{\beta,1} := B^R_{\beta 1}.$$  

Furthermore, for $\alpha = 1, 2$, we define

$$f_{\beta, \alpha} := -m_\alpha(\beta)\text{Tr}[H(\beta)\Xi_1(\beta)A_{\beta, \alpha}] + F_{\beta, \alpha}\text{Tr}[G(\beta)A_{\beta, \alpha}],$$

and

$$g_{\beta, \alpha} := -\frac{1}{2}m_\alpha(\beta)\text{Tr}[H(\beta)\Xi_2(\beta)B_{\beta, \alpha}] + \frac{F_{\beta, \alpha}}{2}\text{Tr}[G^2(\beta)B_{\beta, \alpha}]
+ \frac{1}{2}(m_\alpha(\beta) - \frac{1}{\beta})\text{Tr}[G(\beta)B_{\beta, \alpha}] - m_\alpha(\beta)\text{Tr}[B_{\beta,\alpha}(\beta)]
+ m'_\alpha(\beta)\text{Tr}[H(\beta)\Pi_1(\beta)B_{\beta, \alpha}(\beta)],$$

where $F_{\beta, \alpha}$ is defined in (6.18) with $z = \beta$. Finally, for $\beta \in [r]$, we denote

$$Q_{\beta} := \sqrt{n} \sum_{\alpha=1, 2} (f_{\beta, \alpha} + g_{\beta, \alpha}) + \sqrt{m_{\beta}} \sum_{(i, j) \in S(\nu)} x_{ij}(\beta)_{ij} - \Delta_d(\beta). \quad (8.2)$$

**Lemma 8.3.** Under the assumptions of Proposition 8.1, we have

$$Q = \sum_{\beta=1}^r Q_{\beta}. \quad (8.3)$$

Indeed, Lemma 8.3 is the analogue of Lemma 6.1, and its proof is also a straightforward extension of the rank one case. We omit the details here.

As a consequence, to prove Proposition 8.2, it suffices to study the following

$$\mathbb{E}Q^k e^{it\Delta} = \sqrt{n} \sum_{\beta=1}^r \sum_{\alpha=1}^2 \mathbb{E}(f_{\beta, \alpha} + g_{\beta, \alpha})Q^{k-1} e^{it\Delta}
+ \sqrt{n} \sum_{(i, j) \in S(\nu)} \left(\sum_{\beta=1}^r \sqrt{m_{\beta}}(\beta)_{ij}\right)\mathbb{E}x_{ij}Q^{k-1} e^{it\Delta} - \Delta_d \mathbb{E}Q^k e^{it\Delta}. \quad (8.4)$$

**Step 2.** In the second step, we will use the cumulant expansion to estimate the items on the right hand side of (8.4) and prove the analogue of Lemma 6.2.

Observe that for the rank $r$ case, we have

$$\frac{\partial Q}{\partial x_{ij}} = \sum_{\beta=1}^r \frac{\partial Q_{\beta}}{\partial x_{ij}}. \quad (8.5)$$

The estimates of the cumulant expansion for the terms in (8.4) follow along the exact lines of Lemma 7.2-7.7, together with linearity of expectation. The main difference is that we will have cross terms from $A^R_{\beta 1} A^R_{\beta 2}$, $B^R_{\beta 1} B^R_{\beta 2}$ and $A^R_{\beta 1} B^R_{\beta 2}$ for $\beta_1, \beta_2 \in [r]$. However, by the orthogonality of the singular vectors, it is easy to check (via the definitions of $A_{\beta}^R$ and $B_{\beta}^R$ in (4.24)) that

$$A^R_{\beta 1} A^R_{\beta 2} = B^R_{\beta 1} B^R_{\beta 2} = A^R_{\beta 1} B^R_{\beta 2} = 0$$

if $\beta_1 \neq \beta_2$. Consequently, these cross terms essentially make no contribution. We specify one example here. In the proof of the analogue of (7.13), we shall encounter an term of the
following form

$$
\frac{1}{\sqrt{n}} \sum_{i,j} \left( \Xi_1(z_{\beta})A_{\beta,1} \right)_{j',i} \frac{\partial Q}{\partial x_{ij}} \partial_{Q}^{k-2} e^{i\Delta} = \frac{1}{\sqrt{n}} \sum_{i,j} \left( \Xi_1(z_{\beta})A_{\beta,1} \right)_{j',i} \frac{\partial Q}{\partial x_{ij}} \partial_{Q}^{k-2} e^{i\Delta}.
$$

Applying (7.1) for each $\frac{\partial Q}{\partial x_{ij}}$, by (4.13) and orthogonality of the singular vectors, we find the only contributing part is

$$
\frac{1}{\sqrt{n}} \sum_{i,j} \left( \Xi_1(z_{\beta})A_{\beta,1} \right)_{j',i} \frac{\partial Q}{\partial x_{ij}} \partial_{Q}^{k-2} e^{i\Delta}
$$

and what remains is exactly the same as the proof of (7.13). This explains why most quantities appearing in Theorem 2.9 and its proof are similar to, and most of time are simply the sum of those in the proof of Theorem 2.3. In the following discussion, we shall concentrate on these cross terms from different singular values and vectors, and show they are actually negligible due to the orthogonality of singular vectors.

We first introduce some notations. Recall (6.7). For $\beta \in [r]$, we denote $\tilde{a}_{\beta,\alpha}, \tilde{b}_{\beta,\alpha}, \tilde{b}_{\beta,\alpha}$ by replacing $z$ with $z_{\beta}$ and $A_{\alpha}, B_{\alpha}$ with $A_{\beta,\alpha}, B_{\beta,\alpha}$ ($\alpha = 1, 2$) correspondingly. We also define $a_{\beta,1,\alpha}, b_{\beta,1,\alpha}, \tilde{b}_{\beta,1,\alpha}$ for $\alpha = 1, 2$ in the same fashion using (6.9). Next, we denote

$$a_{\beta,21} := -\frac{(k-1)z_{\beta}}{\sqrt{n}} \sum_{(i,j) \in S(\nu)} (\Pi_1(z_{\beta}))_{j',i} (\Pi_1(z_{\beta})A_{\beta,1})_{i,i} \left( \sum_{\gamma=1}^{r} \sqrt{\gamma} C_{\gamma} \right)_{ij},$$

$$\tilde{b}_{\beta,21} := -\frac{(k-1)z_{\beta}}{\sqrt{n}} \sum_{(i,j) \in S(\nu)} \left[ (\Pi_1(z_{\beta}))_{j',i} (\Pi_2(z_{\beta})B_{\beta,1})_{ii} + (\Pi_2(z_{\beta}))_{j',i} (\Pi_1(z_{\beta})B_{\beta,1})_{i,i} \right] \times \left( \sum_{\gamma=1}^{r} \sqrt{\gamma} C_{\gamma} \right)_{ij},$$

and define $a_{\beta,22}, \tilde{b}_{\beta,22}$ analogously. Further, we denote

$$a_{\beta,31} := -\frac{2(k-1)z_{\beta}^3}{n} \sum_{i,j} (\Pi_1(z_{\beta}))_{j',i} (\Pi_1(z_{\beta})A_{\beta,1})_{i,i} \left( \sum_{\gamma=1}^{r} \gamma S_{\gamma} \right)_{ij},$$

$$\tilde{b}_{\beta,31} = -\frac{2(k-1)z_{\beta}^{3/2}}{n} \sum_{i,j} \left[ (\Pi_1(z_{\beta}))_{j',i} (\Pi_2(z_{\beta})B_{\beta,1})_{ii} + (\Pi_2(z_{\beta}))_{j',i} (\Pi_1(z_{\beta})B_{\beta,1})_{i,i} \right] \times \left( \sum_{\gamma=1}^{r} \gamma S_{\gamma} \right)_{ij},$$

and define $a_{\beta,32}, \tilde{b}_{\beta,32}$ analogously. Finally, we denote

$$a_{\beta,0a} := a_{\beta,1a} + \kappa_2 a_{\beta,2a} + \frac{\kappa_3}{2} a_{\beta,3a},$$

$$b_{\beta,0a} := \frac{m_{z_{\beta}}(z_{\beta})}{2} b_{\beta,1a} + m_{z_{\beta}}(z_{\beta}) b_{\beta,1a} + \frac{\kappa_4 m_{z_{\beta}}(z_{\beta})}{2} b_{\beta,2a}$$

$$+ \kappa_4 m_{z_{\beta}}(z_{\beta}) b_{\beta,2a} + \frac{\kappa_4 m_{z_{\beta}}(z_{\beta})}{4} b_{\beta,3a},$$

We adopt the notation

$$q^{(l)} = Q(l)e^{i\Delta}.$$  

With these preparations, we present the following analogue of Lemma 6.2.
Lemma 8.4. Under the assumptions of Proposition 8.2, for each $\beta \in [r]$ and $\alpha = 1, 2$, we have
\[
\sqrt{n} E_{\beta,\alpha} q^{(k-1)} = -\sqrt{\pi/2} m_\alpha (z_\beta) E \left( \frac{K_3}{2} \overline{\delta}_{\beta,\alpha} q^{(k-1)} + a_{\beta,0\alpha} q^{(k-2)} \right) + O_\prec (n^{-\frac{1}{2}+4\nu}),
\]
(8.6) \\
\[
\sqrt{n} E_{\beta,\alpha} q^{(k-1)} = -\sqrt{\pi/2} E \left( \frac{K_3}{4} (m_\alpha (z_\beta) \overline{\delta}_{\beta,\alpha} + 2m_\alpha (z_\beta) \overline{d}_{\beta,\alpha} q^{(k-1)} + d_{\beta,0\alpha} q^{(k-2)} \right) \\
+ O_\prec (n^{-\frac{1}{2}+4\nu}).
\]
In addition, we have
\[
\sqrt{n} \sum_{(i,j) \in S(\nu)} (\sum_{\beta=1}^r \sqrt{\pi/2} (c_{(\beta)}_{ij}) x_{ij} q^{(k-1)} = (k-1) \left[ \sum_{(i,j) \in S(\nu)} \left( \sum_{\beta=1}^r \sqrt{\pi/2} (c_{(\beta)}_{ij}) \right)^2 \right] \\
+ \frac{K_3}{\sqrt{n}} \sum_{(i,j) \in S(\nu)} \left( \sum_{\beta=1}^r (z_{(\beta)}_{ij}) \left( \sum_{\beta=1}^r \sqrt{\pi/2} (c_{(\beta)}_{ij}) \right) \right) \xi q^{(k-2)} + O_\prec (n^{-\frac{1}{2}+4\nu}).
\]
Similar to the proof of Proposition 8.2, Proposition 8.2 immediately follows from Lemma 8.3 and 8.4. We omit the details here.

Next, we turn to the proof of Lemma 8.4. We will only focus our discussion on the term $\sqrt{n} E_{\beta,1} q^{(k-1)}$ and the other terms can be estimated likewise. Using a discussion similar to (7.5), for each fixed $\beta \in [r]$, we have
\[
\sqrt{n} E_{\beta,1} q^{(k-1)} = E \left( -m_1 \sqrt{\pi/2} \sum_{i,j} x_{ij} (\Xi_1 (z_\beta) A_{\beta,1})_{i,j} + \sqrt{n} F_1 \text{Tr}(G(z_\beta) A_{\beta,1}) \right) q^{(k-1)}.
\]
As seen in the proof of (6.13), we need the following estimates which are analogues of those in Lemma 7.2 and 7.3. We adopt the notations in (7.11) by denoting
\[
h_1 = (\Xi_1 (z_\beta) A_{\beta,1})_{i,j}, \quad h_2 = Q^{k-1}, \quad h_3 = e^{it\Delta}.
\]

Lemma 8.5. For the derivatives of $h_1 h_2 h_3$, we have
\[
h(1,0,0) = -\sqrt{\pi/2} m_2 (z_\beta) \text{Tr}(G(z_\beta) A_{\beta,1}) q^{(k-1)} + O_\prec (n^{-1/2}),
\]
(8.7) \\
\[
h(0,1,0) = a_{\beta,11} q^{(k-2)} + O_\prec (n^{-1/2}),
\]
(8.8) \\
\[
h(2,0,0) = d_{\beta,1} q^{(k-1)} + O_\prec (n^{-1/2}),
\]
\[
h(1,1,0) = a_{\beta,21} q^{(k-2)} + O_\prec (n^{-1/2}),
\]
\[
h(1,2,0) = a_{\beta,31} q^{(k-2)} + O_\prec (n^{-1/2}).
\]
Furthermore, all the other terms $h(l_1, l_2, l_3)$ for $l_1 + l_2 + l_3 \leq 4$ can be bounded by $O_\prec (n^{-1/4+4\nu})$.

It is easy to see that (8.6) follows directly from Lemma 8.5. Thus the final task is to prove Lemma 8.5. In the proof, we will use the orthogonality of the singular vectors, that is, for $\beta_1 \neq \beta_2$,
\[
\{u_{\beta_1}, u_{\beta_2}\} = 0, \quad \{v_{\beta_1}, v_{\beta_2}\} = 0.
\]
(8.9) \\

Proof of Lemma 8.5. First of all, (8.7) can be estimated similarly as (7.30). The other four dominating terms can be analyzed analogously and we shall only focus on the estimate of (8.8). Observe that
\[
h(0,1,0) = \frac{1}{\sqrt{n}} \sum_{i,j} h_1 \frac{\partial h_2}{\partial x_{ij}} h_3 = \frac{k-1}{\sqrt{n}} \sum_{i,j} (\Xi_1 (z_\beta) A_{\beta,1})_{i,j} \frac{\partial Q}{\partial x_{ij}} q^{k-2}.
\]
(8.10)
Plugging in (8.5), we have
\[
\hat{h}(0, 1, 0) = \frac{(k - 1)}{\sqrt{n}} \sum_{\gamma = 1}^{r} \sum_{i,j} \left( \Xi_1(z_\beta) A_{\beta,1} \right)_{j'} \partial Q_{\gamma}' \partial x_{ij} Q^{k-2} e^{itA},
\]
where by (7.1),
\[
\frac{\partial Q_{\gamma}}{\partial x_{ij}} = -\sqrt{n} \sum_{l_1, l_2 \in \{i,j\}} \left[ \left( (G(z_\beta) A_{\gamma}^R G(z_\beta))_{l_1 l_2} - \frac{1}{2z_\gamma} (G(z_\beta) B_{\gamma}^R G(z_\beta))_{l_1 l_2} \right) + \frac{1}{2} \sum_{(a_1, a_2) \in \mathcal{P}(2,1)} \left( G^{a_1}(z_\beta) B_{\gamma}^R G^{a_2}(z_\beta) \right)_{l_1 l_2} \right] - 1 \left( (i, j) \in B(\nu) \right) \sqrt{n} z_\gamma (C_\gamma)_{ij}.
\] (8.11)

Using a discussion similar to (7.13), we have that
\[
\frac{(k - 1)}{\sqrt{n}} \sum_{i,j} \left( \Xi_1(z_\beta) A_{\beta,1} \right)_{j'} \partial Q_{\gamma}' \partial x_{ij} q^{(k-2)} = a_{\beta,11} q^{(k-2)} + O_{\prec}(n^{-1/2}).
\]

Therefore, it suffices to show that for \( \gamma \neq \beta \),
\[
\frac{1}{\sqrt{n}} \sum_{i,j} \left( \Xi_1(z_\beta) A_{\beta,1} \right)_{j'} \partial Q_{\gamma}' \partial x_{ij} q^{(k-2)} = O_{\prec}(n^{-\frac{1}{2}}). \] (8.12)

To prove this, we shall argue in a similar way to (7.13) by expanding the product above using (8.11). We start with
\[
\sum_{i,j} \left( \Xi_1(z_\beta) A_{\beta,1} \right)_{j'} (G(z_\beta) A_{\gamma}^R G(z_\beta))_{j}. \]

Recall (7.34) and (7.35). By (4.18) and (8.9), we have
\[
\sum_{i,j} \left( (G(z_\beta) m_2(z_\beta)) A_{\beta,2} \right)_{j'} (G(z_\beta) A_{\gamma}^R G(z_\beta))_{j} = \omega_{\beta, \omega_{\beta, \gamma}} Tr \left( (G(z_\beta) m_2(z_\beta)) v_\gamma u_\gamma^* G(z_\beta) u_\gamma v_\gamma^* G(z_\beta) \right) = \omega_{\beta, \omega_{\beta, \gamma}} (u_\gamma^* G(z_\beta) m_2(z_\beta) v_\gamma) = O_{\prec}(n^{-\frac{1}{2}}),
\]
where the coefficients \( \omega_{\beta, \beta} \) are defined using the block decomposition of \( A_{\gamma}^R \) as in (7.29). We can estimate the other terms in the expansion (in light of (7.34) and (7.35)) using similar discussions. Hence, we conclude that
\[
\sum_{i,j} \left( \Xi_1(z_\beta) A_{\beta,1} \right)_{j'} (G(z_\beta) A_{\gamma}^R G(z_\beta))_{j} = O_{\prec}(n^{-1/2}).
\]

Likewise, we can show that each of the following terms
\[
\sum_{i,j} \left( \Xi_1(z_\beta) A_{\beta,1} \right)_{j'} (G(z_\beta) A_{\gamma}^R G(z_\beta))_{j}, \quad \sum_{i,j} \left( \Xi_1(z_\beta) A_{\beta,1} \right)_{j'} (G(z_\beta)^2 B_{\gamma}^R G(z_\beta))_{j},
\]
\[
\sum_{i,j} \left( \Xi_1(z_\beta) A_{\beta,1} \right)_{j'} (G(z_\beta) B_{\gamma}^R G(z_\beta))_{j}, \quad \sum_{i,j} \left( \Xi_1(z_\beta) A_{\beta,1} \right)_{j'} (G(z_\beta) B_{\gamma}^R G^2(z_\beta))_{j},
\]
\[
\sum_{i,j} \left( \Xi_1(z_\beta) A_{\beta,1} \right)_{j'} (G(z_\beta) B_{\gamma}^R G(z_\beta))_{j}
\]

can be bounded by $O_{\star}(n^{-1/2})$. In view of (8.11), we conclude the proof of (8.12). This completes our proof.

At the end, we claim that the proof of Theorem 2.11 is analogous.

**Proof of Theorem 2.11.** By considering $Y^*$ instead of $Y$, the proof of Theorem 2.9 applies to the right singular vectors of $Y^*$, which are the left singular vectors of $Y$. Hence, we conclude the proof of Theorem 2.11.

---

**Appendix A. Collection of derivatives**

In this appendix, we summarize some basic identities on the derivatives of $G$ and $Q$ defined in (5.11) without proof. Recall the notation introduced in (5.3).

Using Lemma 4.11, it is easy to check

$$
\left( \frac{\partial^2 G}{\partial x_{ij}^2} \right)_{ab} = 2z \sum_{l_1, l_2 \in \mathcal{P}(2,1,1)} \sum_{l_1 \neq l_2, l_3 \neq l_4} G_{a_1} G_{l_2 l_3} (GW)_{l_4 b},
$$

(A.1)

$$
\left( \frac{\partial^3 G}{\partial x_{ij}^3} \right)_{ab} = -6z \sum_{l_1, \ldots, l_6 \in \{i, j\}} \sum_{l_1 \neq l_2, l_3 \neq l_4} G_{a_1} G_{l_2 l_3} G_{l_4 l_5} (GW)_{l_6 b},
$$

(A.2)

$$
\left( \frac{\partial^4 G}{\partial x_{ij}^4} \right)_{ab} = 24z^2 \sum_{l_1, \ldots, l_8 \in \{i, j\}} \sum_{l_1 \neq l_2, l_3 \neq l_4, l_5 \neq l_6} G_{a_1} G_{l_2 l_3} G_{l_4 l_5} G_{l_6 l_7} (GW)_{l_8 b}.
$$

(A.3)

and also the following identities

$$
\left( \frac{\partial^2 G^2}{\partial x_{ij}^2} \right)_{ab} = 2z \sum_{(a_1, a_2, a_3) \in \mathcal{P}(2,1,1)} \sum_{l_1 \neq l_2, l_3 \neq l_4} G_{a_1}^2 G_{l_2 l_3} (G^a W)_{l_4 b},
$$

(A.4)

Similarly, using Lemma 4.11 and a discussion similar to (7.1), we can also derive

$$
\frac{\partial^2 Q}{\partial x_{ij}} = 2z \sqrt{n} \sum_{l_1, \ldots, l_4 \in \{i, j\}} \left( (GAG)_{l_1 l_2} G_{l_3 l_4} - \frac{1}{2z} (GBG)_{l_1 l_2} G_{l_3 l_4} \right) + \frac{1}{2} \sum_{(a_1, a_2, a_3) \in \mathcal{P}(2,1,1)} (G_{a_1} B G_{a_2})_{l_1 l_2} G_{l_3 l_4}.
$$

(A.5)

$$
\frac{\partial^3 Q}{\partial x_{ij}^3} = -6z \sqrt{n} \sum_{l_1, \ldots, l_6 \in \{i, j\}} \left( (GAG)_{l_1 l_2} G_{l_3 l_4} G_{l_5 l_6} - \frac{1}{2z} (GBG)_{l_1 l_2} G_{l_3 l_4} G_{l_5 l_6} \right) + \frac{1}{2} \sum_{(a_1, \ldots, a_3) \in \mathcal{P}(2,1,1)} (G_{a_1} B G_{a_2})_{l_1 l_2} G_{l_3 l_4} G_{l_5 l_6}.
$$

(A.6)
\[
\frac{\partial^2 Q}{\partial x_{ij}} = 2az^2 \sqrt{n} \sum_{l_1, \ldots, l_5 \in \{1, \ldots, k\}} \sum_{l_1 \neq l_2, l_3 \neq l_4, l_5 \neq l_7} \left( (GAG)_{l_1l_2} G_{l_2l_3} G_{l_3l_4} G_{l_4l_5} G_{l_5l_7} - \frac{1}{2z} (GBG)_{l_1l_2} G_{l_2l_3} G_{l_3l_4} G_{l_4l_5} G_{l_5l_7} \right) + \frac{1}{2} \sum_{(a_1, \ldots, a_5) \in \mathcal{P}(2,1,1,1,1)} (G^{a_1} B G^{a_2})_{l_1l_2} G_{l_2l_3} G_{l_3l_4} G_{l_4l_5} G_{l_5l_7}. \tag{A.7}
\]

**References**

[1] T. W. Anderson. Asymptotic theory for principal component analysis. *Ann. Math. Statist.*, 34(1):122–148, 1963.

[2] Z. Bai and J. Yao. Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. H. Poincaré, Probab. Statist.*, 44(3):447–474, 2008.

[3] Z. Bai and J. Yao. On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis*, 106: 167–177, 2012.

[4] J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005.

[5] J. Baik, and J.W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6): 1382–1408, 2006.

[6] Z. Bao, X. Ding, J. Wang, and K. Wang. Outlier eigenvector distribution of spiked covariance matrix. *In preparation*.

[7] Z. Bao, J. Hu, G. Pan, and W. Zhou. Canonical correlation coefficients of high-dimensional Gaussian vectors: finite rank case. *Ann. Statist. (to appear)*, 2018.

[8] Z. Bao, G. Pan, and W. Zhou. Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.*, 43(1):382–421, 2015.

[9] F. Benaych-Georges, A. Guionnet, and M. Maida. Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Probab.*, 16:1621–1662, 2011.

[10] F. Benaych-Georges and R. R. Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large rectangular random matrices. *Advances in Mathematics*, 227(1):494 – 521, 2011.

[11] F. Benaych-Georges and R. R. Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120 – 135, 2012.

[12] A. Bloemendal, L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.*, 19:no. 33, 53, 2014.

[13] A. Bloemendal, A. Knowles, H.-T. Yau, and J. Yin. On the principal components of sample covariance matrices. *Probab. Theory Related Fields*, 164(1-2): 459–552, 2016.

[14] A. Bloemendal, and B. Virág. Limits of spiked random matrices I. *Probability Theory and Related Fields*, 156(3-4): 795–825, 2013.

[15] A. Bloemendal, and B. Virág. Limits of spiked random matrices II. *The Annals of Probability*, 44(4): 2726–2769, 2016.

[16] E. Bura, and R. Pfeiffer. On the distribution of the left singular vectors of a random matrix and its applications. *Statistics & Probability Letters*, 78(15): 2275–2280 (2008).

[17] T. T. Cai and A. Zhang. Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *Ann. Statist.*, 46(1):60–89, 2018.

[18] J. Cape, M. Tang, and C. E. Priebe. Signal-plus-noise matrix models: eigenvector deviations and fluctuations. *arXiv:1802.00381*, 2018.

[19] J. Cape, M. Tang, and C. E. Priebe. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Ann. Statist (to appear)*, 2018.

[20] M. Capitaine. Limiting eigenvectors of outliers for Spiked Information-Plus-Noise type matrices. *arXiv:1701.08069*, 2017.

[21] M. Capitaine and C. Donati-Martin. Spectrum of deformed random matrices and free probability. *arXiv:1607.05560*, 2016.

[22] M. Capitaine and C. Donati-Martin. Non universality of fluctuations of outlier eigenvectors for block diagonal deformations of Wigner matrices. *arXiv:1807.07775*, 2018.

[23] M. Capitaine, C. Donati-Martin, and D. Féral. The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.*, 37(1):1–47, 2009.
[24] M. Capitaine, C. Donati-Martin, and D. Féral. Central limit theorems for eigenvalues of deformations of Wigner matrices. *Ann. Inst. H. Poincaré, Probab. Statist.*, 48(1): 107–133, 2012.

[25] X. Ding. High dimensional deformed rectangle matrices with applications in matrix denoising. *arXiv:1702.06975*, 2017.

[26] X. Ding, and Q. Sun. Modified Multidimensional Scaling and High Dimensional Clustering. *arXiv:1810.10172*, 2018.

[27] D. Donoho and M. Gavish. Minimax risk of matrix denoising by singular value thresholding. *Ann. Statist.*, 42(6):2413–2440, 2014.

[28] N. El Karoui. Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.*, 35(2):663–714, 2007.

[29] L. Erdős, A. Knowles, and H.-T. Yau. Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré*, 14(8):1837–1926, 2013.

[30] J. Fan, Q. Sun, W.-X. Zhou, and Z. Zhu. Principal component analysis for big data. *arXiv:1801.01602*, 2018.

[31] J. Fan, W. Wang, and Y. Zhong. An \( \ell_\infty \) eigenvector perturbation bound and its application to robust covariance estimation. *Journal of Machine Learning Research*, 18(207):1–42, 2018.

[32] A. Knowles and J. Yin. The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.*, 66(11):2784–2883, 2013.

[33] V. A. Marchenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, 1967.
B. Nadler and I. M. Johnstone. On the distribution of Roy’s Largest Root Test in MANOVA and in signal detection in noise. Technical report, Department of Statistics, Stanford University, 2011.

S. O’Rourke, V. Vu, and K. Wang. Random perturbation of low rank matrices: Improving classical bounds. *Linear Algebra and its Applications*, 540:26 – 59, 2018.

D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistics Sinica*, 17(4):1617–1642, 2007.

S. Péché. The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probability Theory and Related Fields*, 134(1):127–173, 2006.

E. Peterfreund and M. Gavish. Multidimensional Scaling of Noisy High Dimensional Data. *arXiv:1801.10229*, 2018.

S. O’Rourke, V. Vu, and K. Wang. Random perturbation of low rank matrices: Improving classical bounds. *Linear Algebra and its Applications*, 540:26 – 59, 2018.

D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistics Sinica*, 17(4):1617–1642, 2007.

S. Péché. The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probability Theory and Related Fields*, 134(1):127–173, 2006.

E. Peterfreund and M. Gavish. Multidimensional Scaling of Noisy High Dimensional Data. *arXiv:1801.10229*, 2018.

S. O’Rourke, V. Vu, and K. Wang. Random perturbation of low rank matrices: Improving classical bounds. *Linear Algebra and its Applications*, 540:26 – 59, 2018.

D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistics Sinica*, 17(4):1617–1642, 2007.