Submitted to the Annals of Statistics

SINGULAR VECTOR AND SINGULAR SUBSPACE DISTRIBUTION FOR THE MATRIX DENOISING MODEL

By Zhigang Bao*, Xiucai Ding†, Ke Wang‡

Hong Kong University of Science and Technology
University of Toronto and Duke University
Hong Kong University of Science and Technology

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 assumptions 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

$$Y = S + X, \tag{1.1}$$

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

$$S = UDV^* = \sum_{i=1}^{r} d_i \boldsymbol{u}_i \boldsymbol{v}_i^*, \tag{1.2}$$

*Z.G. Bao was partially supported by Hong Kong RGC grant ECS 26301517, GRF 16300618, GRF 16301519, and NSFC 11871425.

†X.C. Ding was partially supported by NSERC of Canada (Jeremy Quastel).

‡Ke Wang was partially supported by Hong Kong RGC grant GRF 16301618 and GRF 16308219.

MSC 2010 subject classifications: Primary 60B20, 62G10; secondary 62H10, 15B52, 62H25

Keywords and phrases: random matrix, singular vector, singular subspace, matrix denoising model, signal-plus-noise model, non-universality

where $D = \mathrm{diag}(d_1, \ldots, d_r)$ consists of the singular values of S and we assume $d_1 > \ldots > d_r > 0$; $U = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r) \in \mathbb{R}^{M\times r}$ and $V = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r) \in \mathbb{R}^{n\times r}$ are the matrices consisting of the ℓ^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

$$\mathbb{E}x_{ij} = 0, \quad \mathbb{E}|x_{ij}|^2 = \frac{1}{n}. \tag{1.3}$$

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

$$\mathbb{E}|\sqrt{n}x_{ij}|^q \leq C_q < \infty. \tag{1.4}$$

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 interested in the inference of S or its left and right singular spaces, which are the subspaces spanned by $\boldsymbol{u}_i$'s or $\boldsymbol{v}_i$'s, respectively. Such a problem arises in many scientific applications such as matrix denoising [3, 29], multiple signal classification (MUSIC) [38, 65] and multidimensional scaling [32, 56]. We call the model in (1.1) the matrix denoising model, which is also known as the signal-plus-noise model in the literature. We refer to Section 1.2 for more introduction on the application aspects.

We denote the singular value decomposition of Y by

$$Y = \widehat{U}\Lambda\widehat{V}^* = \sum_{i=1}^{M\wedge n} \sqrt{\mu_i}\widehat{\boldsymbol{u}}_i\widehat{\boldsymbol{v}}_i^*, \tag{1.5}$$

where $\mu_1 \geq \cdots \geq \mu_{M\wedge n}$ are the squares of the non-trivial singular values, and $\widehat{\boldsymbol{u}}_i$'s and $\widehat{\boldsymbol{v}}_i$'s are the ℓ^2-normalized sample singular vectors. Here $\widehat{U} = (\widehat{\boldsymbol{u}}_1, \ldots, \widehat{\boldsymbol{u}}_M)$ and $\widehat{V} = (\widehat{\boldsymbol{v}}_1, \ldots, \widehat{\boldsymbol{v}}_n)$ and Λ is $M \times n$ with singular values on its main diagonal.

In this paper, we are interested in the distributions of the principal left and right singular vectors of Y and the subspaces spanned by them. On singular vectors, a natural quantity to look into is the projection of a sample principal singular vector onto its deterministic counterpart, i.e., $|\langle \widehat{\boldsymbol{u}}_i, \boldsymbol{u}_i\rangle|$ and $|\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_i\rangle|$, which characterizes the deviation of an original signal from the noisy one. On singular spaces, the natural estimators for U and V are their noisy counterparts

$$\widehat{U}_r = (\widehat{\boldsymbol{u}}_1, \ldots, \widehat{\boldsymbol{u}}_r) \quad \text{and} \quad \widehat{V}_r = (\widehat{\boldsymbol{v}}_1, \ldots, \widehat{\boldsymbol{v}}_r),$$

respectively, i.e., the matrices consisting of the first r left and right singular vectors of Y, respectively. To measure the distance between $\widehat{U}_r$ and U, or $\widehat{V}_r$ and V, we consider the following matrix of the cosine principal angles between two subspaces (see [37, Section 6.4.3] for instance):

$$\cos\Theta(\widehat{V}_r, V) = \operatorname{diag}(\sigma_1^V, \ldots, \sigma_r^V), \qquad \cos\Theta(\widehat{U}_r, U) = \operatorname{diag}(\sigma_1^U, \ldots, \sigma_r^U),$$

where σ_i^V's and σ_i^U's are the singular values of the matrices $\widehat{V}_r^* V$ and $\widehat{U}_r^* U$, respectively. Therefore, an appropriate measure of the distance between the subspaces is $L := \|\cos\Theta(U, \widehat{U}_r)\|_F^2$ for the left singular subspace or $R := \|\cos\Theta(V, \widehat{V}_r)\|_F^2$ for the right singular subspace, where $\|\cdot\|_F^2$ stands for the Frobenius norm. Note that L and R can also be written as

$$L := \sum_{i,j=1}^{r} |\langle \widehat{\boldsymbol{u}}_i, \boldsymbol{u}_j \rangle|^2 = \frac{1}{2}\Big(2r - \|\widehat{U}_r \widehat{U}_r^* - UU^*\|_F^2\Big), \tag{1.6}$$

$$R := \sum_{i,j=1}^{r} |\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_j \rangle|^2 = \frac{1}{2}\Big(2r - \|\widehat{V}_r \widehat{V}_r^* - VV^*\|_F^2\Big). \tag{1.7}$$

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

$$M \equiv M(n), \qquad y \equiv y_n := \frac{M}{n} \to c \in [\tau, \tau^{-1}], \quad \text{as } n \to \infty. \tag{1.8}$$

Our main results are on the limiting distributions of individual $|\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_i \rangle|^2$ (resp. $|\langle \widehat{\boldsymbol{u}}_i, \boldsymbol{u}_i \rangle|^2$) and R (resp. L) when the signal strength, d_i's, are supercritical (c.f. Assumption 2.1). They are detailed in Theorems 2.3, 2.9, 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 [5], 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 [5] for the complex spiked covariance matrix. We also refer to [6, 12, 13, 4, 23, 3, 28, 7, 54] 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 [3, 4, 9, 10, 11, 25, 26, 30, 16, 17, 7, 54, 36, 55, 58, 44]. Especially, the fluctuations of the outliers are shown to be non-universal for the deformed Wigner matrices, first in [25] under certain special assumptions on the structure of the deformation and the distribution of the matrix entries, and then in [7] in full generality.

The study on the behavior of the extreme eigenvectors has been mainly focused on the level of the first order limit [12, 13, 22, 3, 35, 54]. 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 [15] 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 [15] 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 (i.e. $\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{w} \rangle$ for any deterministic vector $\boldsymbol{w}$) are χ^2 distributed. For spiked Gaussian sample covariance matrices, in the supercritical regime, the fluctuation of a fixed-dimensional normalized subvector of the outlier eigenvector is proved to be Gaussian in [54], but this result cannot tell the distribution of $\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_i \rangle$. Under some special assumptions on the structure of the deformation and the distribution of the random matrix entries, it is shown in [24] 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 [7].

1.2. *On singular subspace inference.* From the applied perspective, our model (1.1) appears prominently in the study of signal processing [41, 52], machine learning [61, 64] and statistics [19, 20, 29, 34]. 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 underlying true mean vectors of samples [19]. 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 (1.7) can be used for the inference of the structure of the singular subspace of S. We remark that these statistics have been used extensively to explore the properties of singular subspace. To name a few, in [40], the authors studied the problem of testing whether the sample singular subspace is equal to some given subspace; in [21], the authors studied the eigenvector inference problems for the correlated stochastic block model; in [39], the authors analyzed the impact of dimensionality reduction for subspace clustering algorithms; and in [19], the authors studied the high-dimensional clustering problem and the canonical correlation analysis. In the high-dimensional regime (1.8), 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 [53], the authors improve the perturbation bounds of Davis-Kahan theorem to be nearly optimal. In [19], the authors study similar problems and their related statistical applications. Most recently, in the papers [33, 34, 66], the authors derive the ℓ^∞ pertubation 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 $\boldsymbol{x}_i \simeq \mathcal{N}(\mathbf{0}, \Sigma), i = 1, 2, \cdots, n$ and $X = (\boldsymbol{x}_i)$, under the assumption that the order of $\frac{\mathrm{Tr}\Sigma}{\|\Sigma\|}$ 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 Σ. Furthermore, in [62], 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 $r^4 K \log^3 m = o(m),\ K = \max\{M, n\}$, the author shows that the principal angles of the subspace estimated using NNPLS are asymptotically normal.

1.3. *Organization.* The rest of the paper is organized as follows. In Section 2, we state our main results and summarize our method for the proofs. In Section 3, we design Monte Carlo simulations to demonstrate the accuracy of our main results and briefly illustrate their applications through a hypothesis testing problem. 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, based on the recursive estimate in Proposition 5.2. We state more simulation results, further discussions of statistical applications, the proofs of Theorem 2.9 and some technical lemmas in the supplementary material [7].

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

2.1. *Main results.* In this paper, the singular values of S are assumed to satisfy the following *supercritical* condition.

ASSUMPTION 2.1 (Supercritical condition). *There exist a constant* $C > 0$ *and a (small) constant* $\delta > 0$, *such that*

$$y^{1/4} + \delta \le d_r < \cdots < d_2 < d_1 \le C, \qquad \min_{1 \le j \ne i \le r} |d_i - d_j| \ge \delta.$$

REMARK 2.2. The first inequality above ensures that the first r singular values of Y are outliers, and the threshold $y^{1/4}$ is the analogous BBP transition point in [5]. The second inequality guarantees that the outliers of Y are well separated from each other. We also assume that $d_1, \cdots, d_r$ are bounded by some constant C. All these 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. Also, the separation of d_i's by an order 1 distance δ is not necessary. In [15], a much weaker separation of order $n^{-1/2+\epsilon}$ is enough for the discussion of the eigenvalues. Moreover, we can also extend our results to the case when $d_1, \cdots, d_r$ diverge with n. But we do not pursue these directions in the current paper.

In the sequel, we will only state the results for the right singular vectors and the right singular subspace. The results for the left ones can be obtained from the right ones by simply considering the transpose (with a rescaling) of our matrix model in (1.1). To state our results, we need more notations. First, we define

$$p(d) := \frac{(d^2+1)(d^2+y)}{d^2}. \tag{2.1}$$

For each $i \in [r]$, we will write $p_i \equiv p(d_i)$ for short. Recall (1.5). In [3, Theorem 3.4], it has been shown that p_i is the limit of μ_i. Further, we set

$$a(d) := \frac{d^4-y}{d^2(d^2+1)}. \tag{2.2}$$

It has been proved in [3] that $a(d_i)$ are the limits of $|\langle \boldsymbol{v}_i, \widehat{\boldsymbol{v}}_i \rangle|^2$ respectively (see Lemma D.1 in [7]). We also denote by κ_l the l-th cumulant of the random variables $\sqrt{n}x_{ij}$. For a vector $\boldsymbol{w} = (w(1), \ldots, w(m))^T$ and $l \in \mathbb{N}$, we introduce the notation

$$\boldsymbol{s}_l(\boldsymbol{w}) := \sum_{i=1}^m w(i)^l.$$

Set

$$\theta(d) := \frac{d^4+2yd^2+y}{d^3(d^2+1)^2}, \qquad \psi(d) := \frac{d^6-3yd^2-2y}{d^3(d^2+1)^2}, \tag{2.3}$$

and

$$\begin{aligned} \mathcal{V}^E(d) := & \frac{2}{d^4-y}\Bigg(2y(y+1)\theta(d)^2 - \frac{y(y-1)(5y+1)}{d(d^2+1)^2}\theta(d) \\ & + \frac{(d^4+y)(d^2+y)^2}{d^3(d^2+1)^2}\psi(d) + \frac{2y^2(y-1)^2}{d^2(d^2+1)^4}\Bigg). \end{aligned} \tag{2.4}$$

For the right singular vectors, we have the following theorem.

THEOREM 2.3 (Right singular vectors). *Assume that (1.3), (1.4), (1.8) and Assumption 2.1 hold. For $i \in [r]$, define the random variable*

$$\Delta_i := -2\sqrt{n}\theta(d_i)\boldsymbol{u}_i^* X \boldsymbol{v}_i - \frac{2\psi(d_i)}{d_i^2}\big(\frac{\kappa_3}{n}\boldsymbol{s}_1(\boldsymbol{u}_i)\boldsymbol{s}_1(\boldsymbol{v}_i)\big), \tag{2.5}$$

and let $\mathcal{Z}_i \sim \mathcal{N}(0, \mathcal{V}_i)$ *be a random variable independent of* Δ_i*, where*

$$\mathcal{V}_i := \mathcal{V}^E(d_i) - \frac{4}{d_i}\theta(d_i)\psi(d_i)\big(\frac{\kappa_3}{\sqrt{n}}\boldsymbol{s}_3(\boldsymbol{u}_i)\boldsymbol{s}_1(\boldsymbol{v}_i)\big) + \frac{4}{d_i}\theta(d_i)^2\big(\frac{\kappa_3}{\sqrt{n}}\boldsymbol{s}_1(\boldsymbol{u}_i)\boldsymbol{s}_3(\boldsymbol{v}_i)\big) + \frac{\psi(d_i)^2}{d_i^2}\kappa_4\boldsymbol{s}_4(\boldsymbol{u}_i) + \frac{y\theta(d_i)^2}{d_i^2}\kappa_4\boldsymbol{s}_4(\boldsymbol{v}_i).$$

Then for any $i \in [r]$ *and any bounded continuous function* f*, we have*

$$\lim_{n\to\infty}\Big(\mathbb{E}f\left(\sqrt{n}\left(|\langle \boldsymbol{v}_i, \widehat{\boldsymbol{v}}_i\rangle|^2 - a(d_i)\right)\right) - \mathbb{E}f(\Delta_i + \mathcal{Z}_i)\Big) = 0.$$

REMARK 2.4. In [7], 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 distribution 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 [7]. 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 $\Delta \simeq \mathcal{N}(0, 4\theta(d)^2)$ (see Definition 4.9 for the meaning of $\simeq$). Hence, we find $\Delta + \mathcal{Z}$ is asymptotically distributed as

$$\mathcal{N}\left(0, 4\theta(d)^2 + \mathcal{V}^E(d)\right).$$

REMARK 2.6. If both $\boldsymbol{u}$ and $\boldsymbol{v}$ are delocalized in the sense that $\|\boldsymbol{u}\|_\infty = o(1)$ and $\|\boldsymbol{v}\|_\infty = o(1)$. Then $\boldsymbol{s}_l(\boldsymbol{u}) = o(1)$ and $\boldsymbol{s}_l(\boldsymbol{v}) = o(1)$ for $l = 3, 4$. By (1.3), (1.4) and the fact $\|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = 1$, we find that $\mathbb{E}(\boldsymbol{u}^* X\boldsymbol{v}) = 0$ and $\mathbb{E}(\boldsymbol{u}^* X\boldsymbol{v})^2 = n^{-1}$. Then we conclude from Lyapunov's CLT for triangular array that

$$\Delta \simeq \mathcal{N}\left(-\frac{2\psi(d)}{d^2}\big(\frac{\kappa_3}{n}\boldsymbol{s}_1(\boldsymbol{u})\boldsymbol{s}_1(\boldsymbol{v})\big), 4\theta(d)^2\right), \tag{2.6}$$

and therefore $\Delta + \mathcal{Z}$ has asymptotically the same distribution as

$$\mathcal{N}\left(-\frac{2\psi(d)}{d^2}\big(\frac{\kappa_3}{n}\boldsymbol{s}_1(\boldsymbol{u})\boldsymbol{s}_1(\boldsymbol{v})\big), 4\theta(d)^2 + \mathcal{V}^E(d)\right).$$

The only difference from the Gaussian case is a shift caused by the non-vanishing third cumulant.

REMARK 2.7. If one of $\boldsymbol{u}$ and $\boldsymbol{v}$ is delocalized, say $\|\boldsymbol{u}\|_\infty = o(1)$, then Δ still has the limiting distribution in (2.6). Therefore $\Delta+\mathcal{Z}$ has asymptotically the same distribution as a Gaussian random variable with mean

$$-\frac{2\psi(d)}{d^2}\big(\frac{\kappa_3}{n}\boldsymbol{s}_1(\boldsymbol{u})\boldsymbol{s}_1(\boldsymbol{v})\big)$$

and variance

$$4\theta(d)^2 + \mathcal{V}^E(d) + \frac{4}{d}\theta(d)^2\big(\frac{\kappa_3}{\sqrt{n}}\boldsymbol{s}_1(\boldsymbol{u})\boldsymbol{s}_3(\boldsymbol{v})\big) + y\frac{\theta(d)^2}{d^2}\kappa_4\boldsymbol{s}_4(\boldsymbol{v}).$$

REMARK 2.8. If neither $\boldsymbol{u}$ nor $\boldsymbol{v}$ is delocalized, then $\Delta+\mathcal{Z}$ is no longer Gaussian in general. For example, if $\boldsymbol{u}=\boldsymbol{e}_1$ and $\boldsymbol{v}=\boldsymbol{f}_1$ where $\boldsymbol{e}_1$ and $\boldsymbol{f}_1$ are the canonical basis vectors in $\mathbb{R}^M$ and $\mathbb{R}^n$ respectively, then $\Delta+\mathcal{Z}$ is asymptotically distributed as

$$-2\theta(d)\sqrt{n}X_{11} + \mathcal{N}\left(0, \mathcal{V}^E(d) + \kappa_4\frac{\psi(d)^2+y\theta(d)^2}{d^2}\right),$$

which depends on the distribution of X_{11} and thus is non-universal.

If the assumptions of Theorem 2.3 hold, we conclude from Remarks 2.6–2.9 that $|\langle \boldsymbol{v}_i, \widehat{\boldsymbol{v}}_i\rangle|^2$ always has a Gaussian fluctuation if either the entries of X are Gaussian or one of $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ is delocalized in the sense $\|\boldsymbol{u}_i\|_\infty = o(1)$ or $\|\boldsymbol{v}_i\|_\infty = o(1)$. In the general setting when the noise matrix is non-Gaussian, the detailed distribution will rely on both the structure of the singular vectors and the noise matrix X.

Next, we study the distributions of the right singular space. For two vectors $\boldsymbol{w}_a = (w_a(1), \dots, w_a(m))^T, a=1,2$, we denote

$$\boldsymbol{s}_{k,l}(\boldsymbol{w}_1, \boldsymbol{w}_2) := \sum_{i=1}^m w_1(i)^k w_2(i)^l.$$

Recall R from (1.7). We have the following theorem.

THEOREM 2.9 (Right singular subspace). *Assume that (1.3), (1.4), (1.8) and Assumption 2.1 hold. Let $\Delta = \sum_{i=1}^r \Delta_i$, where Δ_i is defined in* (2.5). *Let $\mathcal{Z}$ be a random variable independent of Δ with law $\mathcal{Z}\sim\mathcal{N}(0,\mathcal{V})$, where*

$$\mathcal{V} := \sum_{i=1}^r \mathcal{V}^E(d_i) + \kappa_4 \sum_{i,j=1}^r \left(\frac{\psi(d_i)\psi(d_j)}{d_i d_j}\boldsymbol{s}_{2,2}(\boldsymbol{u}_i,\boldsymbol{u}_j) + y\frac{\theta(d_i)\theta(d_j)}{d_i d_j}\boldsymbol{s}_{2,2}(\boldsymbol{v}_i,\boldsymbol{v}_j)\right)$$
$$+ \frac{\kappa_3}{\sqrt{n}}\sum_{i,j=1}^r \frac{4}{d_i}\theta(d_j)\Big(\theta(d_i)\boldsymbol{s}_{2,1}(\boldsymbol{v}_i,\boldsymbol{v}_j)\boldsymbol{s}_1(\boldsymbol{u}_j) - \psi(d_i)\boldsymbol{s}_{2,1}(\boldsymbol{u}_i,\boldsymbol{u}_j)\boldsymbol{s}_1(\boldsymbol{v}_j)\Big).$$

Then for any bounded continuous function f, we have that

$$\lim_{n\to\infty}\Big(\mathbb{E}f\big(\sqrt{n}\big(R-\sum_{i=1}^{r}a(d_i)\big)\big)-\mathbb{E}f(\Delta+\mathcal{Z})\Big)=0.$$

2.2. *Proof strategy.* In this subsection, we briefly describe our proof strategy. We first review the method used in a related work [7], and then we highlight the novelty of our strategy.

As we mentioned, in [7], 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 $\langle \boldsymbol{u},(W-z)^{-1}\boldsymbol{v}\rangle$ for any deterministic vectors $\boldsymbol{u},\boldsymbol{v}$. 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 $\langle \boldsymbol{u},(W-z)^{-1}\boldsymbol{u}\rangle$. So one can turn to establish the law of the quadratic form of the Green function instead. In [7], 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 $\boldsymbol{u}$ 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 [7] 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 [14, 43] as a main technical input. It turns out that for the singular vectors, we can approximately represent $\sqrt{n}|\langle\widehat{\boldsymbol{u}}_i,\boldsymbol{u}_i\rangle|$ (after appropriate centering) in terms of a quantity of the form

$$\mathcal{Q}_i=\sqrt{n}\Big(\mathrm{Tr}(G(p_i))-\Pi_1(p_i))A_i+\mathrm{Tr}(G'(p_i)-\Pi_1'(p_i))B_i\Big), \tag{2.7}$$

where G is the Green function of the linearization of the sample covariance matrix and Π_1 is the deterministic approximation of G; see (4.5) and (4.10) for the definitions. Here both A_i and B_i 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 [7], here we derive the law of the above $\mathcal{Q}_i$ directly for generally distributed matrix. Recall Δ_i defined in (2.5), whose random part is proportional to $\boldsymbol{u}_i^* X \boldsymbol{v}_i$, which is simply a linear combination of the entries of X. Inspired by [7], we decompose Δ_i into two parts, say $\widetilde{\Delta}_i$ and $\widehat{\Delta}_i$. The former contains the linear combination of $x_{k\ell}$'s for those indices k, ℓ corresponding to the large components u_{ik} and $v_{i\ell}$ in $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$. The latter contains the linear combinations of the rest of $x_{k\ell}$'s. Note that $\widehat{\Delta}_i$ is asymptotically normal by CLT since the coefficients of $x_{k\ell}$'s are small. However, $\widetilde{\Delta}_i$ 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

$$\mathbb{E}(\mathcal{Q}_i - \widetilde{\Delta}_i)^k \mathrm{e}^{\mathrm{i}t\widetilde{\Delta}_i} = (k-1)\widetilde{\mathcal{V}}_i \mathbb{E}(\mathcal{Q}_i - \widetilde{\Delta}_i)^{k-2} \mathrm{e}^{\mathrm{i}t\widetilde{\Delta}_i} + o(1), \tag{2.8}$$

for some positive number $\widetilde{\mathcal{V}}_i$. Choosing $t = 0$, we can derive the asymptotic normality of $\mathcal{Q}_i - \widetilde{\Delta}_i$ for (2.8) by the recursive moment estimate. Choosing t to be arbitrary, we can further deduce from (2.8) that

$$\mathbb{E}\mathrm{e}^{\mathrm{i}s(\mathcal{Q}_i - \widetilde{\Delta}_i) + \mathrm{i}t\widetilde{\Delta}_i} = \mathbb{E}\mathrm{e}^{\mathrm{i}s(\mathcal{Q}_i - \widetilde{\Delta}_i)} \mathbb{E}\mathrm{e}^{\mathrm{i}t\widetilde{\Delta}_i} + o(1).$$

Then asymptotic independence between $\mathcal{Q}_i - \widetilde{\Delta}_i$ and $\widetilde{\Delta}_i$ follows. Hence, we prove both the asymptotic normality and asymptotic independence from (2.8). 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.

Moreover, 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 [2] and the deformed Wigner matrix.

Finally, we briefly compare the methods used in this paper and the related work [24]. In [24], the authors study the distribution of $|\langle \widehat{\boldsymbol{v}}, \boldsymbol{e}_1 \rangle|^2$ of a deformed Wigner matrix whose deformation is a block diagonal deterministic Hermitian matrix containing one large spike $\theta \boldsymbol{e}_1 \boldsymbol{e}_1^*$ which creates one outlier of the deformed Wigner matrix. Here $\widehat{\boldsymbol{v}}$ is the random outlier eigenvector. By Helffer-Sjöstrand formula, they represent $|\langle \widehat{\boldsymbol{v}}, \boldsymbol{e}_1 \rangle|^2$ in terms of an integral (over z) of $\boldsymbol{e}_1^*(W-z)^{-1}\boldsymbol{e}_1$. In contrast to our work, the major difference in [24] is that they establish the limiting distribution for the whole process $\boldsymbol{e}_1^*(W-z)^{-1}\boldsymbol{e}_1$ in z, and then use functional limit theorem to conclude the limit of the integral. In our work, relying on the isotropic law, we first integrate out the contour integral approximately. This results in the linear combination in (2.7), and then we only need to consider the joint distribution of the quadratic form of G and G' at a single point $p(d_i)$. Moreover, in [24], the authors decompose the quadratic form $\boldsymbol{e}_1^*(W-z)^{-1}\boldsymbol{e}_1$ into two parts using Schur's complement, where one of them can be proved to be Gaussian using an extension of the CLT for quadratic forms as in the previous work [25]. It is worth noticing that the independence between the Gaussian and non-Gaussian parts follows directly from the special structure of the model in [24]. However, in [7] and our work, since we do not have structural assumptions on S, we need to make more dedicated efforts for the independence (see [7, Proposition 7.12] and Proposition 5.1).

3. Simulations and statistical applications.

3.1. *Numerical simulations.* In this section, we present some numerical simulations for our results stated in Section 2.1. 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 $\frac{1}{3}\delta_{\sqrt{2}}+\frac{2}{3}\delta_{-\frac{1}{\sqrt{2}}}$. 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 $\frac{1}{3}\delta_{\sqrt{2}}+\frac{2}{3}\delta_{-\frac{1}{\sqrt{2}}}$ are $\kappa_3=\frac{1}{\sqrt{2}}$ and $\kappa_4=-\frac{3}{2}$. In the sequel, let $\{\boldsymbol{e}_i\}_{i=1}^M$ and $\{\boldsymbol{f}_j\}_{j=1}^n$ be the canonical basis of $\mathbb{R}^M$ and $\mathbb{R}^n$, respectively. Denote by $\mathbf{1}_m$ the all-one vector in $\mathbb{R}^m$.

Assume that S has rank $r=1$ and admits the singular value decomposition $S=d\boldsymbol{u}^T\boldsymbol{v}$. Set the dimension ratio $y=M/n=0.5$. We present the simulations corresponding to the special cases discussed in Remarks 2.5 - 2.8. Specifically, we consider following four cases: 1. Gaussian noise, $\boldsymbol{u}=\boldsymbol{e}_1$ and $\boldsymbol{v}=\boldsymbol{f}_1$; 2. Two-point noise, $\boldsymbol{u}=\mathbf{1}_M/\sqrt{M}$ and $\boldsymbol{v}=\mathbf{1}/\sqrt{n}$; 3. Two-point noise, $\boldsymbol{u}=\mathbf{1}_M/\sqrt{M}$ and $\boldsymbol{v}=\boldsymbol{f}_1$; 4. Two-point noise, $\boldsymbol{u}=\boldsymbol{e}_1$ and $\boldsymbol{v}=\boldsymbol{f}_1$. The normalization of $\sqrt{n}(|\langle\widehat{\boldsymbol{v}},\boldsymbol{v}\rangle|^2-a(d))$ listed in the above cases are chosen according to the calculations in Remarks 2.5 - 2.8. For case 4, we further subtract the non-Gaussian part $-2\theta(d)\sqrt{n}X_{11}$ from the statistic. Hence, in all four cases, we expect that the asymptotic distributions are normal. We denote the normalized statistics of the above four cases as $\mathcal{R}_g,\mathcal{R}_{dt},\mathcal{R}_{pt}$ and $\mathcal{R}_{st}$,, respectively, and we refer to the supplementary material [7, Section A] for more details on the definitions.

In Figure S1 of [7], we plot the ECDFs of of $\mathcal{R}_g, \mathcal{R}_{dt}, \mathcal{R}_{pt}, \mathcal{R}_{st}$ 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. In [7, Section A], we also record the probabilities for different quantiles of the empirical cumulative distributions (ECDFs) of the above statistics, they are fairly close to standard Gaussian even for a small sample size $n = 200$.

3.2. *Statistical applications.* In this section, we will briefly discuss the applications of our main results to the singular vector and singular subspace estimation and inference, and leave more details to the supplementary material [7].

We start with the estimation part and focus on the right singular vector and subspace. The estimation of singular vector and subspace is important in the recovery of low-rank matrix based on noisy observations (see for instance [19, 21, 29] and reference therein). It is clear that (see Lemma D.1 in [7]) 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 $a(d)$ defined in (2.2). Further, when d increases, the sample singular vector will get closer to the true singular vector in ℓ^2 norm. It can be seen from the result in Theorem 2.3 that the variance of the fluctuation also decays when d increases. This phenomenon is recorded in Figure S2 in the supplementary material [7].

Empirically, it can be seen from Figure S2 in [7] that for a sequence of $y \in [\frac{1}{10}, 10]$, 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 κ_3 and κ_4, we can ignore the fluctuation for a smaller value of d. Once the singular vectors are estimated, the estimation of the singular subspace follows.

Next, we consider the inference of the singular vectors and subspace of S. Recall the decomposition in (1.2). For brevity, here we focus our discussion on the inference of V, assuming that U, D and the necessary parameters of X (e.g. cumulants of the entries of X) are known. In the supplementary material [7], we will also briefly discuss the possible extension of our results to adapt to the situation when D and the parameters of X are not known. Especially, using Theorem 2.3 we can test whether a singular vector $\boldsymbol{v}_i$ is equal to a given vector $\boldsymbol{v}_{i0}$, which can be formulated as

$$\text{(T0)} \qquad \boldsymbol{H}_0 : \boldsymbol{v}_i = \boldsymbol{v}_{i0}, \qquad \boldsymbol{H}_a : \boldsymbol{v}_i \neq \boldsymbol{v}_{i0},$$

and we can choose the testing statistic to be

$$\boldsymbol{S}_0 := \sqrt{n}(|\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_{i0}\rangle|^2 - a(d_i)).$$

Further, using Theorem 2.9, one can test if the matrix V is equal to a given matrix, which can be formulated as

$$\text{(T1)} \qquad \boldsymbol{H}_0 : V = V_0, \quad \boldsymbol{H}_a : V \neq V_0,$$

where $V_0 = (\boldsymbol{v}_{10}, \dots, \boldsymbol{v}_{i0})$ is a given matrix consisting of orthonormal columns. We can choose the testing statistic to be

$$\begin{aligned} \boldsymbol{S}_1 &= \sqrt{n}\Big(\sum_{i,j=1}^{r} |\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_{j0}\rangle|^2 - \sum_{i=1}^{r} a(d_i)\Big) \\ &= \sqrt{n}\Big(\frac{1}{2}\Big(2r - \|\widehat{V}_r\widehat{V}_r^* - V_0V_0^*\|_F^2\Big) - \sum_{i=1}^{r} a(d_i)\Big). \end{aligned} \tag{3.1}$$

We remark here that in some cases like X is Gaussian, we can see from Theorem 2.9 that $\boldsymbol{S}_1$ is not a good statistic to distinguish V_0 from V_0O for some deterministic $r \times r$ orthogonal matrix O. Specifically, one cannot tell if $\widehat{V}_r$ is the matrix of the singular vectors of the model $X + UDV_0^*$ or $X + UD(V_0O)^*$, since $V_0V_0^* = (V_0O)(V_0O)^*$ in (3.1) and the limiting distribution of $\boldsymbol{S}_1$ does not depend on V when X is Gaussian. Hence, we do not expect the statistic $\boldsymbol{S}_1$ to be powerful for the test (T1) when the alternative is of the form V_0O in some cases like Gaussian noise. In other words, in this case, what one can test is if $VV^* = V_0V_0^*$. Nevertheless, one can still do the test (T1) by using the testing statistic of the diagonal parts of $\boldsymbol{S}_1$ only, i.e., $\boldsymbol{S}_{1d} = \sqrt{n}\Big(\sum_i^r |\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_{i0}\rangle|^2 - \sum_{i=1}^r a(d_i)\Big)$. Under the null hypothesis, $\boldsymbol{S}_{1d}$ has the same distribution as $\boldsymbol{S}_1$ since it will be clear that $|\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_{j0}\rangle|^2$ is negligible if $i \neq j$, in the null case. But note that the limiting distribution of $\boldsymbol{S}_{1d}$ is no longer invariant under taking right orthogonal transformation for V_0. Hence, it can be used to test if $V = V_0$.

We mention that both (T0) and (T1) could be useful in many scientific disciplines, especially when the singular vectors of S are sparse and have practical meanings. For instance, an important goal of the study of gene expression data for cancer is to simultaneously identify related genes and subjects grouped together according to the cancer types [49, Section 2]. For this purpose, the right singular vectors are used to visualize the gene grouping (see Figure 1 of [49]) and the left singular vectors are used to represent the subject grouping (see Figure 2 of [49]). Other examples include the study of the nutrition content data of different foods [49] and the mortality rate data after expanding on suitable basis functions [63, Section 3]. In the literature, various algorithms have been proposed to estimate the sparse singular vectors, for instance see [3, 49, 63, 64]. From the statistical perspective, with the above estimates, it is natural to do inference on the singular vectors. For instance, for the gene expression data of lung caner, researchers may be interested in testing whether a certain type of cancer is determined by a subset of genes and this is related to doing inference on the right singular vectors and right singular subspace.

Since we assume that U, D and the necessary parameters of X (e.g. cumulants of the entries of X) are known, we can carry out the z-score test to test $\boldsymbol{H}_0$ in both (T0) and (T1). Due the similarity of (T0) and (T1), we focus on (T1) and leave the detailed discussions and simulations to the supplementary material [7].

4. Techincal 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 [14, 43]. It will be stated in Section 4.2. In subsection 4.3, we represent (asymptotically) $|\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_i \rangle|^2$'s and R (c.f. (1.7)) in terms of the Green function. The discussion is based on [3], where the limits for $|\langle \widehat{\boldsymbol{u}}_i, \boldsymbol{u}_j \rangle|^2$ and $|\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_j \rangle|^2$ are studied. We then collect a few auxiliary definitions in Section 4.4.

4.1. *Basic notions.*

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

$$\mathcal{Y}(z) := \mathcal{U}\mathcal{D}(z)\mathcal{U}^* + H(z), \qquad z = E + \mathrm{i}\eta \in \mathbb{C}^+, \tag{4.1}$$

where

$$\mathcal{U} := \begin{pmatrix} U & \\ & V \end{pmatrix}, \quad \mathcal{D}(z) := \sqrt{z} \begin{pmatrix} & D \\ D & \end{pmatrix}, \quad H(z) := \sqrt{z} \begin{pmatrix} & X \\ X^* & \end{pmatrix}. \tag{4.2}$$

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

$$F_1(x) := \frac{1}{M}\sum_{i=1}^{M} \mathbf{1}_{\{\lambda_i(XX^*)\le x\}}, \qquad F_2(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{\lambda_i(X^*X)\le x\}}.$$

$F_1(x)$ and $F_2(x)$ are known to satisfy the Marchenko-Pastur (MP) law [51]. More precisely, almost surely, $F_1(x)$ converges weakly to a non-random limit $F_{1y}(x)$ which has a density function given by

$$\rho_1(x) := \begin{cases} \frac{1}{2\pi x y}\sqrt{(\lambda_+ - x)(x-\lambda_-)}, & \text{if } \lambda_- \le x \le \lambda_+, \\ 0, & \text{otherwise,} \end{cases}$$

and has a point mass $1-1/y$ at the origin if $y>1$, where $\lambda_+ = (1+\sqrt{y})^2$ and $\lambda_- = (1-\sqrt{y})^2$. Further, the Stieltjes's transform of F_{1y} is given by

(4.3)
$$m_1(z) := \int \frac{1}{x-z}\,\mathrm{d}F_{1y}(x) = \frac{1-y-z+\mathrm{i}\sqrt{(\lambda_+ - z)(z-\lambda_-)}}{2zy} \quad \text{for } z\in\mathbb{C}^+,$$

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

$$\rho_2(x) := \begin{cases} \frac{1}{2\pi x}\sqrt{(\lambda_+ - x)(x-\lambda_-)}, & \text{if } \lambda_- \le x \le \lambda_+, \\ 0, & \text{otherwise,} \end{cases}$$

and a point mass $1-y$ at the origin if $y<1$. The corresponding Stieltjes's transform is

$$m_2(z) := \int \frac{1}{x-z}\,\mathrm{d}F_{2y}(x) = \frac{y-1-z+\mathrm{i}\sqrt{(\lambda_+ - z)(z-\lambda_-)}}{2z}. \tag{4.4}$$

Our estimation relies on the local MP law [57] and its isotropic version [14, 43], which provide sharp large deviation estimates for the Green functions

$$G(z) = (H-z)^{-1}, \quad \mathcal{G}_1(z) = (XX^*-z)^{-1}, \qquad \mathcal{G}_2(z) = (X^*X-z)^{-1}.$$

Here we recall the definition in (4.2). By Schur complement, one can derive

$$G(z) = \begin{pmatrix} \mathcal{G}_1(z) & z^{-1/2}\mathcal{G}_1(z)X \\ z^{-1/2}X^*\mathcal{G}_1(z) & \mathcal{G}_2(z) \end{pmatrix}. \tag{4.5}$$

The Stieltjes transforms for the ESD of XX^* and X^*X are defined by
(4.6)

$$m_{1n}(z) = \frac{1}{M}\mathrm{Tr}\mathcal{G}_1(z) = \frac{1}{M}\sum_{i=1}^{M} G_{ii}(z), \quad m_{2n}(z) = \frac{1}{n}\mathrm{Tr}\mathcal{G}_2(z) = \frac{1}{n}\sum_{\mu=M+1}^{M+n} G_{\mu\mu}(z).$$

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 (4.3) and (4.4). Specifically, for any fixed $z \in \mathbb{C}^+$, the following hold,

$$m_{1n}(z) - m_1(z) \xrightarrow{a.s.} 0, \quad m_{2n}(z) - m_2(z) \xrightarrow{a.s.} 0.$$

Furthermore, one can easily check that $m_1(z)$ and $m_2(z)$ satisfy the following self-consistent equations (see [2] for instance)

$$m_1(z) + \frac{1}{z - (1-y) + zym_1(z)} = 0, \tag{4.7}$$

$$m_2(z) + \frac{1}{z + (1-y) + zm_2(z)} = 0. \tag{4.8}$$

We can also derive the following simple relation from the definitions

$$m_1(z) = \frac{y^{-1} - 1}{z} + y^{-1}m_2(z). \tag{4.9}$$

Next we summarize some basic identities in the following lemma without proof. They can be checked from (4.3) and (4.4) via elementary calculations.

LEMMA 4.1. *Denote $p \equiv p(x)$ in (2.1). For any $x > y^{1/4}$, we have*

$$m_1(p) = \frac{-1}{x^2 + y}, \qquad m_2(p) = \frac{-1}{x^2 + 1},$$
$$m_1'(p) = \frac{x^4}{(x^2+y)^2(x^4-y)}, \qquad m_2'(p) = \frac{x^4}{(x^2+1)^2(x^4-y)}.$$

Furthermore, denote by $\mathcal{T}(t) = tm_1(t)m_2(t)$. We have

$$\mathcal{T}(p) = x^{-2}, \qquad \mathcal{T}'(p) = (y - x^4)^{-1}.$$

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 [31].

DEFINITION 4.3 (Stochastic domination). *Let*

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

be two families of nonnegative random variables, where $\mathsf{U}^{(n)}$ *is a possibly* n*-dependent parameter set. We say that* X *is stochastically dominated by* Y*, uniformly in* u*, if for all small* ϵ *and large* φ*, we have*

$$\sup_{u \in \mathsf{U}^{(n)}} \mathbb{P}\Big(\mathsf{X}^{(n)}(u) > n^{\epsilon}\mathsf{Y}^{(n)}(u)\Big) \leq n^{-\varphi},$$

for large enough $n \geq n_0(\epsilon, \varphi)$*. In addition, we use the notation* $\mathsf{X} = O_{\prec}(\mathsf{Y})$ *if* $|\mathsf{X}|$ *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 [14]. Let $\oplus$ be the direct sum of two matrices. Set

$$\Pi_1(z) := m_1(z) I_M \oplus m_2(z) I_n. \tag{4.10}$$

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

$$\boldsymbol{S}_o \equiv \boldsymbol{S}_o(\tau) := \{z = E + \mathrm{i}\eta \in \mathbb{C}^+ : \lambda_+ + \tau \leq E \leq \tau^{-1},\ 0 \leq \eta \leq \tau^{-1}\}, \tag{4.11}$$

where $\tau > 0$ is a fixed small constant. Recall m_{1n} and m_{2n} defined in (4.6).

LEMMA 4.4 (Theorem 3.7 of [43], Theorem 3.12 of [14] and Theorem 3.1 of [57]). *Fix* $\tau > 0$*, for any unit deterministic* $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{M+n}$*, we have*

$$\langle \boldsymbol{u}, (G(z) - \Pi_1(z))\boldsymbol{v} \rangle = O_{\prec}\Big(\sqrt{\frac{\operatorname{Im} m_2(z)}{n\eta}}\Big), \tag{4.12}$$

$$|m_{1n}(z) - m_1(z)| = O_{\prec}(\frac{1}{n}), \qquad |m_{2n}(z) - m_2(z)| = O_{\prec}(\frac{1}{n}), \tag{4.13}$$

uniformly in $z \in \boldsymbol{S}_o$*.*

REMARK 4.5. The bounds in (4.13) cannot be directly read from any of Theorem 3.7 of [43], Theorem 3.12 of [14] or Theorem 3.1 of [57]. In all these theorems, a weaker bound $O_{\prec}(\frac{1}{n\eta})$ is stated for z both inside and 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 $\frac{1}{n\eta}$. For $z \in \boldsymbol{S}_o$, such a bound follows from the rigidity estimates of eigenvalues in [57] and the definition of the Stieltjes transform easily. Specifically, by (3.7) in [57], we know that for $a = 1, 2$, $\sup_{t\in\mathbb{R}} |F_a(t) - F_{ay}(t)| \prec \frac{1}{n}$, and further by (3.6) of [57] we know that $\sup_{t\in\mathbb{R}:|t|\geq 2+n^{-\frac{2}{3}+\varepsilon}} |F_a(t) - F_{ay}(t)| = 0$ with high probability. Then using the integration by parts to $m_{an}(z) - m_a(z) = \int (t-z)^{-1} \mathrm{d}(F_a(t) - F_{ay}(t))$, one can easily conclude the bounds in (4.13).

Following from Lemma 4.4, by further using Cauchy's integral formula for derivatives, we have the following uniformly in $z \in \boldsymbol{S}_o$, for any given $l \in \mathbb{N}$,

$$\langle \boldsymbol{u}, (G^{(l)}(z) - \Pi_1^{(l)}(z))\boldsymbol{v}\rangle = O_{\prec}\Big(\sqrt{\frac{\operatorname{Im} m_2(z)}{n\eta}}\Big). \tag{4.14}$$

Denote by $\kappa = |E - \lambda_+|$. We summarize some basic estimates of $m_{1,2}(z)$ without proof. For any two numbers a_n and b_n (might be n-dependent), we write $a_n \sim b_n$ if there exist two positive constants C_1 and C_2 (independent of n) such that $C_1|b_n| \leq |a_n| \leq C_2|b_n|$.

LEMMA 4.6. *The following estimates hold uniformly in $z \in \boldsymbol{S}_o$,*

$$|m_{1,2}'(z)| \sim |m_{1,2}(z)| \sim 1, \tag{4.15}$$

$$\operatorname{Im} m_1(z) \sim \operatorname{Im} m_2(z) \sim \frac{\eta}{\sqrt{\kappa+\eta}}. \tag{4.16}$$

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 (4.14) that the following estimates hold uniformly in $z \in \boldsymbol{S}_o$: For any fixed $k, \ell \in \mathbb{N}$,

$$\begin{aligned}
&\max_{\mu,\nu}\Big|(G^{(l)}(z)A)_{\mu\nu} - (\Pi_1^{(l)}(z)A)_{\mu\nu}\Big| = O_{\prec}\Big(\frac{1}{\sqrt{n}}\Big),\\
&\mathrm{Tr}G^{(l)}(z)A - \mathrm{Tr}\Pi_1^{(l)}(z)A = O_{\prec}\Big(\frac{1}{\sqrt{n}}\Big),\\
&\max_{\mu,\nu}\Big|(G^{(k)}(z)AG^{(l)}(z))_{\mu\nu} - (\Pi_1^{(k)}A\Pi_1^{(l)})_{\mu\nu}\Big| = O_{\prec}(\frac{1}{\sqrt{n}}).
\end{aligned} \tag{4.17}$$

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 whose proof is stated in [7].

LEMMA 4.7. *We have the following recursive relation*

$$G^2 = 2G' + \frac{G}{z},\ G^3 = (G^2)' + \frac{G^2}{z},\ G^4 = \frac{2}{3}(G^3)' + \frac{G^3}{z}. \tag{4.18}$$

Recall Π_1 defined in (4.10) and further define

$$\Pi_2 := 2\Pi_1' + \frac{1}{z}\Pi_1, \qquad \Pi_3 := \Pi_2' + \frac{1}{z}\Pi_2, \qquad \Pi_4 := \frac{2}{3}\Pi_3' + \frac{1}{z}\Pi_3. \tag{4.19}$$

With Lemma 4.7, similarly to (4.12) and (4.14), we can get the following estimates for $l = 1, 2, 3, 4$,

$$\langle \boldsymbol{u}, (G^l - \Pi_l)\boldsymbol{v} \rangle = O_{\prec}\Big(\frac{1}{\sqrt{n}}\Big), \tag{4.20}$$

uniformly in $z \in \boldsymbol{S}_o$. For brevity, in the sequel, we will use the notation

$$\Xi_l \equiv \Xi_l(z) := G^l(z) - \Pi_l(z), \qquad l \in \mathbb{N}. \tag{4.21}$$

4.3. *Green function representation.* In this section, we represent (asymptotically) $|\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_i \rangle|^2$'s and R (c.f (1.7)) in terms of the Green function. The derivation relies on the results obtained in [3]. Recall $p(d)$ in (2.1) and $a(d)$ in (2.2). For $i \in [r]$, define

$$h_i(x) = \frac{x^4 p'(x) p(x)}{(x + d_i)^2}, \tag{4.22}$$

and we use the shorthand notation $\bar{i} = i + r$. 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

$$\begin{aligned} \big(W_i(x)\big)_{ii} &= m_2^2(x), \qquad \big(W_i(x)\big)_{\bar{i}\bar{i}} = \frac{1}{d_i^2 x}, \\ \big(W_i(x)\big)_{i\bar{i}} &= \big(W_i(x)\big)_{\bar{i}i} = -\frac{m_2(x)}{d_i\sqrt{x}}. \end{aligned} \tag{4.23}$$

We further denote the matrix function

$$M_i(x) = \mathcal{U} W_i(x) \mathcal{U}^*. \tag{4.24}$$

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

$$\begin{aligned} A_i^R &= -d_i^2\Big(h_i'(d_i) M_i(p_i) + h_i(d_i) p'(d_i) M_i'(p_i)\Big), \\ B_i^R &= -d_i^2 h(d_i) p'(d_i) M_i(p_i). \end{aligned} \tag{4.25}$$

In light of the definition of $\mathcal{U}$ in (4.2), we have

$$A_i^R = \begin{pmatrix} \omega_{i1}\boldsymbol{u}_i\boldsymbol{u}_i^T & \omega_{i2}\boldsymbol{u}_i\boldsymbol{v}_i^T \\ \omega_{i3}\boldsymbol{v}_i\boldsymbol{u}_i^T & \omega_{i4}\boldsymbol{v}_i\boldsymbol{v}_i^T \end{pmatrix}, \qquad B_i^R = \begin{pmatrix} \varpi_{i1}\boldsymbol{u}_i\boldsymbol{u}_i^T & \varpi_{i2}\boldsymbol{u}_i\boldsymbol{v}_i^T \\ \varpi_{i3}\boldsymbol{v}_i\boldsymbol{u}_i^T & \varpi_{i4}\boldsymbol{v}_i\boldsymbol{v}_i^T \end{pmatrix}. \tag{4.26}$$

Here we used the notations

$$\begin{aligned}
&\omega_{i1} := -d_i^2\big(h_i'(d_i)(W_i(p_i))_{ii} + h_i(d_i)p'(d_i)(W_i'(p_i))_{ii}\big),\\
&\omega_{i4} := -d_i^2\big(h_i'(d_i)(W_i(p_i))_{\bar{i}\bar{i}} + h_i(d_i)p'(d_i)(W_i'(p_i))_{\bar{i}\bar{i}}\big),\\
&\omega_{i2} = \omega_{i3} := -d_i^2\big(h_i'(d_i)(W_i(p_i))_{i\bar{i}} + h_i(d_i)p'(d_i)(W_i'(p_i))_{i\bar{i}}\big),\\
&\varpi_{i1} := -d_i^2 h_i(d_i)p'(d_i)(W_i(p_i))_{ii},\\
&\varpi_{i4} := -d_i^2 h_i(d_i)p'(d_i)(W_i(p_i))_{\bar{i}\bar{i}},\\
&\varpi_{i2} = \varpi_{i3} := -d_i^2 h_i(d_i)p'(d_i)(W_i(p_i))_{i\bar{i}}.
\end{aligned}$$

Recall the notation introduced in (4.21). We have the following lemma whose proof is stated in [7].

Lemma 4.8. *Under assumptions of (1.3), (1.4), (1.8) and Assumption 2.1, we have*

$$|\langle \boldsymbol{v}_i, \widehat{\boldsymbol{v}}_i\rangle|^2 = a(d_i) + \operatorname{Tr}\left(\Xi_1(p_i)A_i^R\right) + \operatorname{Tr}\left(\Xi_1'(p_i)B_i^R\right) + O_\prec(\frac{1}{n}),$$

Furthermore, we have

$$R = \sum_{i=1}^r a(d_i) + \sum_{i=1}^r \left(\operatorname{Tr}\big(\Xi_1(p_i)A_i^R\big) + \operatorname{Tr}\big(\Xi_1'(p_i)B_i^R\big)\right) + O_\prec(\frac{1}{n}). \tag{4.27}$$

4.4. *Auxiliary definitions.* It is convenient to introduce the following notion of convergence in distribution.

Definition 4.9 ([7, Definition 7.3]). *Two sequences of random variables,* $\{\mathsf{X}_n\}$ *and* $\{\mathsf{Y}_n\}$*, are* asymptotically equal in distribution*, denoted as* $\mathsf{X}_n \simeq \mathsf{Y}_n$*, if they are tight and satisfy*

$$\lim_{n\to\infty}\left(\mathbb{E}f(\mathsf{X}_n) - \mathbb{E}f(\mathsf{Y}_n)\right) = 0$$

for any bounded continuous function f*.*

We also collect some basic results on convergence and equivalence in distribution in the supplementary material [7], Lemma C.3.

The following notation from [7, Definition 7.11] will be convenient for us when we replace random variables with their i.i.d copies.

DEFINITION 4.10. *Let $\{\sigma_n\}$ be a sequence of bounded positive numbers. If X_n and Y_n are independent random variables with $\mathsf{Y}_n \simeq \mathcal{N}(0, \sigma_n^2)$, and if $\mathsf{S}_n \simeq \mathsf{X}_n + \mathsf{Y}_n$, we write $\mathsf{S}_n \simeq \mathsf{X}_n + \mathcal{N}(0, \sigma_n^2)$.*

5. Proof of Theorems 2.3. For brevity, in this section, we omit the subindices of $d_i, \boldsymbol{u}_i, \boldsymbol{v}_i, \widehat{\boldsymbol{u}}_i, \widehat{\boldsymbol{v}}_i$ and write $d, \boldsymbol{u}, \boldsymbol{v}, \widehat{\boldsymbol{u}}, \widehat{\boldsymbol{v}}$ instead. Similarly, we write the matrices A_i^R and B_i^R (c.f. (4.25)) 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

$$\mathcal{Q} \equiv \mathcal{Q}(z) := \sqrt{n}\Big(\mathrm{Tr}\big(\Xi_1(z)A\big) + \mathrm{Tr}\big(\Xi_1'(z)B\big)\Big), \tag{5.1}$$

at $z = p(d)$ (c.f.(2.1)).

In the sequel, we will prove the limiting distribution of $\mathcal{Q}(z)$ 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 $j \in [n]$. Denote the shorthand notation

$$j' = j + M. \tag{5.2}$$

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

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

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

$$\mathcal{S}(\nu) = ([M] \times [n]) \setminus \mathcal{B}(\nu). \tag{5.3}$$

For brevity, we introduce the notation

$$\mathcal{P}(\alpha_1, \dots, \alpha_m), \tag{5.4}$$

to represent the set of all the permutations of $(\alpha_1, \dots, \alpha_m)$, where α_i's can be alike. Recall (4.10) and (4.19). We set the deterministic quantity

$$\begin{aligned}\Delta_d \equiv \Delta_d(z) := &-\frac{\kappa_3 z^{3/2}}{n} \sum_{i,j} \Big((\Pi_1)_{ii} (\Pi_1)_{j'j'} \big(2(\Pi_1 A \Pi_1)_{ij'} + (\Pi_1 B \Pi_1')_{ij'} + (\Pi_1' B \Pi_1)_{ij'}\big) \\ &+ \frac{1}{2} \sum_{(a_1,a_2,a_3)\in\mathcal{P}(2,1,1)} (\Pi_{a_1})_{ii} (\Pi_{a_2})_{j'j'} \big((\Pi_1 B \Pi_{a_3})_{ij'} + (\Pi_{a_3} B \Pi_1)_{ij'}\big)\Big),\end{aligned} \tag{5.5}$$

and the random variable

$$\Delta_r \equiv \Delta_r(z) := \sqrt{nz} \sum_{(i,j)\in\mathcal{B}(\nu)} x_{ij} c_{ij}, \tag{5.6}$$

where

$$\begin{aligned} c_{ij} \equiv c_{ij}(z) := - \sum_{\substack{l_1,l_2\in\{i,j'\} \\ l_1\neq l_2}} \Big(& (\Pi_1 A \Pi_1)_{l_1 l_2} - \frac{1}{2z}(\Pi_1 B \Pi_1)_{l_1 l_2} \\ & + \frac{1}{2}(\Pi_1 B \Pi_2)_{l_1 l_2} + \frac{1}{2}(\Pi_2 B \Pi_1)_{l_1 l_2}\Big). \end{aligned} \tag{5.7}$$

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

$$\begin{aligned} s_{ij} \equiv s_{ij}(z) := \sum_{\substack{l_1,\cdots,l_4\in\{i,j'\} \\ l_1\neq l_4, l_2\neq l_3}} \Big(& (\Pi_1 A \Pi_1)_{l_1 l_2}(\Pi_1)_{l_3 l_4} - \frac{1}{2z}(\Pi_1 B \Pi_1)_{l_1 l_2}(\Pi_1)_{l_3 l_4} \\ & + \frac{1}{2} \sum_{(a_1,a_2,a_3)\in\mathcal{P}(2,1,1)} (\Pi_{a_1} B \Pi_{a_2})_{l_1 l_2}(\Pi_{a_3})_{l_3 l_4}\Big). \end{aligned} \tag{5.8}$$

Further, we define the function

$$V \equiv V(z) := \mathcal{V}^E(z) + 2\frac{\kappa_3 z^{\frac{3}{2}}}{\sqrt{n}} \sum_{(i,j)\in\mathcal{S}(\nu)} c_{ij} s_{ij} + \frac{\kappa_4 z^2}{n} \sum_{i,j} s_{ij}^2 + z \sum_{(i,j)\in\mathcal{S}(\nu)} c_{ij}^2, \tag{5.9}$$

where

$$\mathcal{V}^E \equiv \mathcal{V}^E(z) := -\sqrt{z} \sum_{\alpha=1,2} \big(m_\alpha \mathfrak{a}_{1\alpha} + \frac{m_\alpha}{2}\tilde{\mathfrak{b}}_{1\alpha} + m'_\alpha \mathfrak{b}_{1\alpha}\big). \tag{5.10}$$

Here we refer to (S9) in [7] for the definitions of $\mathfrak{a}_{1\alpha}$, $\mathfrak{b}_{1\alpha}$ and $\tilde{\mathfrak{b}}_{1\alpha}$ for $\alpha = 1, 2$.
With Δ_d and Δ_r defined in (5.5) and (5.6), we introduce the notation

$$\Delta \equiv \Delta(z) := \Delta_r(z) + \Delta_d(z) \tag{5.11}$$

and define

$$Q \equiv Q(z) := \mathcal{Q}(z) - \Delta(z). \tag{5.12}$$

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

$$Q(p_i) \simeq \mathcal{N}(0, V(p_i)). \tag{5.13}$$

We first show how Proposition 5.1 implies Theorem 2.3.

PROOF OF THEOREM 2.3. By Lemma 4.8 and (5.1),

$$\sqrt{n}\big(|\langle \boldsymbol{v}_i, \widehat{\boldsymbol{v}}_i\rangle|^2 - a(d_i)\big) = \mathcal{Q}(p_i) + O_{\prec}(n^{-\frac{1}{2}}).$$

Here $\mathcal{Q}(p_i)$ is defined in (5.1) with $(A, B) = (A_i^R, B_i^R)$ (c.f.(4.25)). By Proposition 5.1, we have that at $z = p_i$,

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

Next, by Central Limit Theorem and Lemma C.3 in [7], one has

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

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 (5.7) and recall $S(z)$ from (5.8). Using Lemma C.3 in [7], we conclude that

$$\mathcal{Q}(p_i) \simeq \Delta_d(p_i) + \sqrt{np_i}\mathrm{Tr}(X^*C(p_i)) + \mathcal{N}(0, \mathcal{V}(p_i)),$$

where

$$\mathcal{V}(p_i) = \mathcal{V}^E(p_i) + 2\frac{\kappa_3 {p_i}^{3/2}}{\sqrt{n}}\mathrm{Tr}\big(C(p_i)^*S(p_i)\big) + \frac{\kappa_4 {p_i}^2}{n}\mathrm{Tr}\big(S(p_i)^*S(p_i)\big).$$

Denote

$$\Delta_i = \sqrt{np_i}\mathrm{Tr}\big(X^*C(p_i)\big) + \Delta_d(p_i)$$

and $\mathcal{Z}_i \sim \mathcal{N}(0, \mathcal{V}(p_i))$, which is independent of Δ_i. Next, plugging $z = p_i$ into (5.5), (5.7), (5.8), using Lemma 4.1 and taking into account the definitions of A_i^R, B_i^R in (4.25), we find that

$$\Delta_i = -\sqrt{n}\frac{2(d_i^4 + 2yd_i^2 + y)}{d_i^3(d_i^2+1)^2}\boldsymbol{u}_i^* X\boldsymbol{v}_i - \frac{2(d_i^6 - 3yd_i^2 - 2y)}{d_i^5(d_i^2+1)^2}\Big(\frac{\kappa_3}{n}\sum_{k,l}\boldsymbol{u}_i(k)\boldsymbol{v}_i(l)\Big).$$

The variance $\mathcal{V}(p_i)$ is the sum of

$$
\begin{aligned}
&2\frac{\kappa_3}{\sqrt{n}}p_i^{3/2}\mathrm{Tr}\big(C(p_i)^*S(p_i)\big)+\frac{\kappa_4}{n}p_i^2\mathrm{Tr}\big(S(p_i)^*S(p_i)\big)\\
&=-\frac{4(d_i^4+2yd_i^2+y)(d_i^6-3yd_i^2-2y)}{d_i^7(d_i^2+1)^4}\Big(\frac{\kappa_3}{\sqrt{n}}\sum_{k,l}\boldsymbol{u}_i(k)^3\boldsymbol{v}_i(l)\Big)\\
&\quad+\frac{4(d_i^4+2yd_i^2+y)^2}{d_i^7(d_i^2+1)^4}\Big(\frac{\kappa_3}{\sqrt{n}}\sum_{k,l}\boldsymbol{u}_i(k)\boldsymbol{v}_i(l)^3\Big)\\
&\quad+\frac{(d_i^6-3yd_i^2-2y)^2}{d_i^8(d_i^2+1)^4}\Big(\kappa_4\sum_{k}\boldsymbol{u}_i(k)^4\Big)+\frac{(d_i^4+2yd_i^2+y)^2}{d_i^8(d_i^2+1)^4}\Big(\kappa_4 y_n\sum_{l}\boldsymbol{v}_i(l)^4\Big)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{V}^E(p_i)=\frac{2}{d_i^4-y}\Big(&2y(y+1)\big(\frac{d^4+2yd^2+y}{d^3(d^2+1)^2}\big)^2-\frac{y(y-1)(5y+1)}{d_i(d_i^2+1)^2}\big(\frac{d^4+2yd^2+y}{d^3(d^2+1)^2}\big)\\
&+\frac{(d_i^4+y)(d_i^2+y)^2}{d_i^3(d_i^2+1)^2}\big(\frac{d^6-3yd^2-2y}{d^3(d^2+1)^2}\big)+\frac{2y^2(y-1)^2}{d_i^2(d_i^2+1)^4}\Big).
\end{aligned}
$$

The last expression is obtained by using the definitions of $\mathfrak{a}_{1\alpha}$, $\mathfrak{b}_{1\alpha}$ and $\tilde{\mathfrak{b}}_{1\alpha}$ for $\alpha=1,2$ in (S9) of [7] and performing tedious yet elementary calculations. Recall (2.3). The conclusion of Theorem 2.3 follows immediately by rewriting Δ_i and $\mathcal{V}(p_i)$ in terms of $\theta(d_i)$ and $\psi(d_i)$. □

The rest of this section is devoted to the proof of Proposition 5.1. Our proof relies on the cumulant expansion in Lemma C.1 of [7], where we need to control the expectation. Throughout the proof, we will frequently use the estimates in (4.17). 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

$$
z=p(d)+\mathrm{i}n^{-C}, \tag{5.14}
$$

for a large constant C. On the bad event, we will use the naive bound of the Green function $\|G\|\le N^C$, which will be compensated by the tiny probability of the bad event. At the end, by the continuity of $G(\tilde{z})$ at $\tilde{z}$ away from the support of the MP law, it is (asymptotically) equivalent to work with (5.14), for the proof of Proposition 5.1. We first claim that it suffices to establish the following recursive estimate.

PROPOSITION 5.2. *Suppose that the assumptions of Theorem 2.3 hold. Let $z_0 = p(d)$ and z be defined in (5.14). We have*

$$\mathbb{E}Q(z)e^{\mathrm{i}t\Delta(z_0)} = O_{\prec}(n^{-\frac{1}{2}+4\nu}), \tag{5.15}$$

and for any fixed integer $k \geq 2$,

$$\mathbb{E}Q^k(z)e^{\mathrm{i}t\Delta(z_0)} = (k-1)V\mathbb{E}Q^{k-2}(z)e^{\mathrm{i}t\Delta(z_0)} + O_{\prec}(n^{-\frac{1}{2}+4\nu}). \tag{5.16}$$

The proof of Proposition 5.2 is our main technical task, which will be stated in Section E of [7]. 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

$$\left| e^{\mathrm{i}x} - \sum_{k=0}^{N} \frac{(\mathrm{i}x)^k}{k!} \right| \leq \min\left\{ \frac{|x|^{N+1}}{(N+1)!}, \frac{2|x|^N}{N!} \right\}. \tag{5.17}$$

First, we write $Q(z) = Q_R(z) + \mathrm{i}Q_I(z)$, 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 (5.14), 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 (5.15) and (5.16) that

$$\mathbb{E}Q_R(z)e^{\mathrm{i}t\Delta(z_0)} = O_{\prec}(n^{-\frac{1}{2}+4\nu}), \tag{5.18}$$

$$\mathbb{E}Q_R^k(z)e^{\mathrm{i}t\Delta(z_0)} = (k-1)V\mathbb{E}Q_R^{k-2}(z)e^{\mathrm{i}t\Delta(z_0)} + O_{\prec}(n^{-\frac{1}{2}+4\nu}). \tag{5.19}$$

For any $s, t \in \mathbb{R}$, by (5.17), we have

(5.20)
$$\mathbb{E}e^{\mathrm{i}sQ_R(z)+\mathrm{i}t\Delta(z_0)} = \sum_{k=0}^{2N-1} \frac{(\mathrm{i}s)^k}{k!}\mathbb{E}Q_R^k(z)e^{\mathrm{i}t\Delta(z_0)} + O\left(\frac{s^{2N}}{(2N)!}\mathbb{E}Q_R^{2N}(z) \right).$$

For the error term on the right side of (5.20), using (5.19) recursively for $t = 0$, we first find

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

Thus, for arbitrarily small $\epsilon > 0$, by taking N sufficiently large, we have $\frac{(2N-1)!!V^N}{(2N)!} < \epsilon$ and it follows that

$$\left| \mathbb{E} e^{\mathrm{i}sQ_R(z)+\mathrm{i}t\Delta(z_0)} - \sum_{k=0}^{2N-1} \frac{(\mathrm{i}s)^k}{k!} \mathbb{E} Q_R^k(z) e^{\mathrm{i}t\Delta(z_0)} \right| < \epsilon + O_{\prec}(n^{-\frac{1}{2}+4\nu}). \tag{5.21}$$

Using (5.19), we get the following estimate

$$\sum_{k=0}^{2N-1} \frac{(\mathrm{i}s)^k}{k!} \mathbb{E} Q_R^k(z) e^{\mathrm{i}t\Delta(z_0)} = \sum_{k=0}^{N-1} \frac{(\mathrm{i}s)^{2k}}{(2k)!!} V^k \mathbb{E} e^{\mathrm{i}t\Delta(z_0)} + O_{\prec}(n^{-\frac{1}{2}+4\nu}). \tag{5.22}$$

Next, combing (5.22) with the fact

$$\exp(\frac{x^2}{2}) = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!!},$$

together with (5.21), we conclude that

$$\left| \mathbb{E} e^{\mathrm{i}sQ_R(z)+\mathrm{i}t\Delta(z_0)} - e^{-\frac{1}{2}Vs^2} \mathbb{E} e^{\mathrm{i}t\Delta(z_0)} \right| < 2\epsilon + O_{\prec}(n^{-\frac{1}{2}+4\nu}). \tag{5.23}$$

The asymptotic independence of $Q_R(z)$ and $\Delta(z_0)$ is a consequence of (5.23) and the fact ϵ is arbitrarily small. (5.13) can be proved by setting $s = 0$. Although Proposition 5.2 is proved under the choice (5.14), 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. □

Acknowledgements. X.C. Ding would like to thank Wei Q. Deng for many helpful discussions on the applications in statistical genetics. The authors also would like to thank two anonymous referees, the associated editor and editor for their suggestions and comments, which have significantly improved the paper. The authors also want to thank Jiang Hu for many helpful discussions.

References.

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

[2] Z. Bai and J. Silverstein. Spectral Analysis of Large Dimensional Random Matrices. 2nd Edition, Springer Series in Statistics, Springer, 2010.

[3] 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.

[4] Z. Bai and J. Yao. On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis*, 106: 167–177, 2012.
[5] 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.
[6] 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.
[7] Z. Bao, X. Ding, and K. Wang. Supplementary material to "Singular vector and singular subspace distribution for the matrix denoising model", 2019.
[8] Z. Bao, X. Ding, J. Wang, and K. Wang. Principal components of spiked covariance matrices in the supercritical regime. *arXiv: 1907.12251*, 2019.
[9] Z. Bao, J. Hu, G. Pan, and W. Zhou. Canonical correlation coefficients of high-dimensional Gaussian vectors: finite rank case. *Ann. Statist.* , 47(1): 612-640, 2019.
[10] 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.
[11] 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.
[12] F. Benaych-Georges and R. R. Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494 – 521, 2011.
[13] 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.
[14] 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.
[15] 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.
[16] A. Bloemendal, and B. Virág. Limits of spiked random matrices I. *Probability Theory and Related Fields*, 156(3-4): 795–825, 2013.
[17] A. Bloemendal, and B. Virág. Limits of spiked random matrices II. *The Annals of Probability*, 44(4): 2726–2769, 2016.
[18] 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).
[19] 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.
[20] J. Cape, M. Tang, and C. E. Priebe. Signal-plus-noise matrix models: eigenvector deviations and fluctuations. *arXiv:1802.00381*, 2018.
[21] 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.
[22] M. Capitaine. Limiting eigenvectors of outliers for Spiked Information-Plus-Noise type matrices. *arXiv:1701.08069*, 2017.
[23] M. Capitaine and C. Donati-Martin. Spectrum of deformed random matrices and free probability. *arXiv:1607.05560*, 2016.
[24] M. Capitaine and C. Donati-Martin. Non universality of fluctuations of outlier eigenvectors for block diagonal deformations of Wigner matrices. *arXiv:1807.07773*, 2018.
[25] 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.

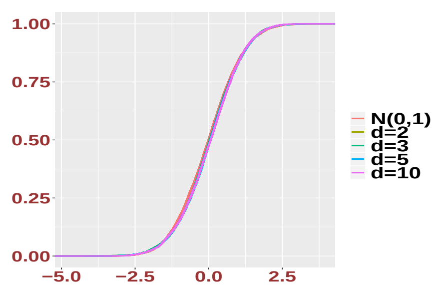
[26] M. Capitaine, C. Donati-Martin, and D. Féral. Central limit theorems for eigenvalues of deformations of Wigner matrices. *Ann. Inst. H. Poincaré, Probab. Statist.*, 48(1): 107–133, 2012.
[27] X. Ding. High dimensional deformed rectangular matrices with applications in matrix denoising. *Bernoulli (in press)*, 2019.
[28] X. Ding, and F. Yang. Spiked separable covariance matrices and principal components. *arXiv: 1905.13060*, 2019.
[29] D. Donoho and M. Gavish. Minimax risk of matrix denoising by singular value thresholding. *Ann. Statist.*, 42(6):2413–2440, 2014.
[30] 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.
[31] L. Erdős, A. Knowles, and H.-T. Yau. Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré*, 14(8):1837–1926, 2013.
[32] J. Fan, Q. Sun, W.-X. Zhou, and Z. Zhu. Principal component analysis for big data. *arXiv:1801.01602*, 2018.
[33] J. Fan, W. Wang, and Y. Zhong. An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *Journal of Machine Learning Research*, 18(207):1–42, 2018.
[34] J. Fan and Y. Zhong. Optimal Subspace Estimation Using Overidentifying Vectors via Generalized Method of Moments. *arXiv:1805.02826*, 2018.
[35] Z. Fan, I. Johnstone and Y. Sun. Spiked covariances and principal components analysis in high-dimensional random effects models. *arXiv: 1806.09529*, 2018.
[36] D. Féral, and S. Péché. The largest eigenvalue of rank one deformation of large Wigner matrices. *Communications in mathematical physics*, 272(1): 185–228 ,2007.
[37] G. H. Golub and C. F. Van Loan. *Matrix computations*, volume 3. JHU Press, 2012.
[38] W. Hachem, P. Loubaton, X. Mestre, J. Najim, and P. Vallet. A subspace estimator for fixed rank perturbations of large random matrices. *Journal of Multivariate Analysis*, 114:427 – 447, 2013.
[39] R. Heckel, M. Tschannen, and H. Bolcskei. Dimensionality-reduced subspace clustering, *Information and Inference: A Journal of the IMA*, 6:246-283, 2017.
[40] J. Huang, Q. Qiu, and R. Calderbank. The Role of Principal Angles in Subspace Classification. *IEEE Trans. Signal Process*, 64(8):1933 - 1945, 2016.
[41] S. M. Kay. *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*. Prentice-Hall, 1998.
[42] A. Knowles and J. Yin. The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.*, 66(11): 1663–1750, 2013.
[43] A. Knowles and J. Yin. Anisotropic local laws for random matrices. *Probab. Theory Related Fields*, 169(1-2): 257–352, 2017.
[44] A. Knowles and J. Yin. The outliers of a deformed Wigner matrix. *The Annals of Probability*, 42(5): 1980–2031, 2014.
[45] V. Koltchinskii, M. Löffler, and R. Nickl. Efficient estimation of linear functionals of principal components. *arXiv:1708.07642*, 2017.
[46] V. Koltchinskii and K. Lounici. Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Ann. Inst. H. Poincaré Probab. Statist.*, 52(4):1976–2013, 2016.
[47] V. Koltchinskii and K. Lounici. Normal approximation and concentration of spectral projectors of sample covariance. *Ann. Statist.*, 45(1):121–157, 2017.
[48] J. Lee and K. Schnelli. Local law and Tracy-Widom limit for sparse random matrices. *arXiv:1605.08767*, 2016.
[49] M. Lee, H. Shen, J. Huang, and J. Marron. Biclustering via Sparse Singular Value

Decomposition. *Biometrics*, 66: 1087–1095, 2010.
[50] A. Levin and B. Nadler. Natural image denoising: Optimality and inherent bounds. In *2011 IEEE Conference On Computer Vision And Pattern Recognition (CVPR). 40.* 2011.
[51] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, 1967.
[52] 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.
[53] 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.
[54] D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17(4):1617–1642, 2007.
[55] S. Péché. The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probability Theory and Related Fields*, 134(1):127–173, 2006.
[56] E. Peterfreund and M. Gavish. Multidimensional Scaling of Noisy High Dimensional Data. *arXiv:1801.10229*, 2018.
[57] N. S. Pillai and J. Yin. Universality of covariance matrices. *Ann. Appl. Probab.*, 24(3):935–1001, 2014.
[58] A. Pizzo, D. Renfrew, and A. Soshnikov. On finite rank deformations of Wigner matrices. *Ann. Inst. H. Poincaré Probab. Statist.*, 49(1): 64–94, 2013.
[59] G. Raskutti, M. Yuan, and H. Chen. Convex Regularization for High-Dimensional Multi-Response Tensor Regression. *Ann. Statist. (to appear)*, 2018.
[60] J. R. Schott. Some tests for common principal component subspaces in several groups. *Biometrika*, 78(4):771–777, 1991.
[61] W. Wang, A. C.-P. Miguel, and Z. Lu. A denoising view of matrix completion. In *Advances in Neural Information Processing Systems 24*, pages 334–342. 2011.
[62] D. Xia. Confidence interval of singular vectors for high-dimensional and low-rank matrix regression. *arXiv:1805.09871*, 2018.
[63] A Sparse Singular Value Decomposition Method for High-Dimensional Data. *Journal of Computational and Graphical Statistics*, 23(4):923-942, 2014.
[64] D. Yang, Z. Ma, and A. Buja. Rate optimal denoising of simultaneously sparse and low rank matrices. *Journal of Machine Learning Research*, 17(92):1–27, 2016.
[65] A. Zahernia, M. J. Dehghani, and R. Javidan. MUSIC algorithm for DOA estimation using MIMO arrays. In *2011 6th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, pages 149–153, 2011.
[66] Y. Zhong and N. Boumal. Near-optimal bounds for phase synchronization. *SIAM Journal on Optimization*, 28(2):989–1016, 2018.

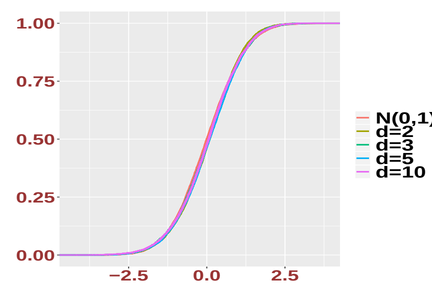
Supplementary material to "Singular vector and singular subspace distribution for the matrix denoising model"

This file contains detailed simulation results, further discussions on statistical applications, auxiliary lemmas, the proofs of Theorem 2.9 and some technical lemmas of the paper [1].

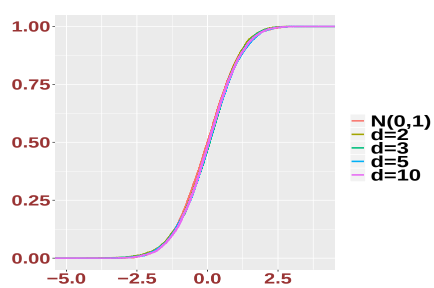
A. Detailed simulation results. In this section, we state detailed simulation results for Section 3.1 of [S1].



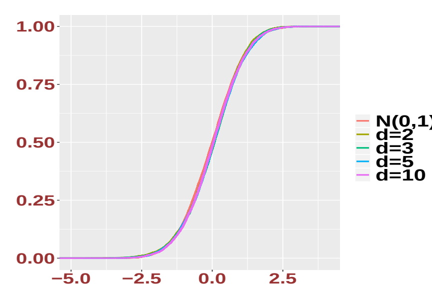
(a) ECDF of $\mathcal{R}_g$ with Gaussian noise.



(b) ECDF of $\mathcal{R}_{dt}$ with Two-Point noise and delocalized singular vectors.



(c) ECDF of $\mathcal{R}_{pt}$ with Two-Point noise and delocalized left singular vector and sparse right singular vector.



(d) ECDF of $\mathcal{R}_{st}$ with Two-Point noise and sparse singular vectors.

Fig S1: Plots of the ECDFs of $\mathcal{R}_g, \mathcal{R}_{dt}, \mathcal{R}_{pt}, \mathcal{R}_{st}$.

Case 1. Gaussian noise. Recall the discussion in Remark 2.5. In this case, the structure of the singular vectors does not play a role. We choose $\boldsymbol{u} = \boldsymbol{e}_1$

and $\boldsymbol{v} = \boldsymbol{f}_1$. Denote by

$$\mathcal{R}_g := \frac{\sqrt{n}}{\sigma}\left(|\langle \widehat{\boldsymbol{v}}, \boldsymbol{v}\rangle|^2 - a(d)\right),$$

where

$$\text{(S1)}\quad \sigma^2 = (8d^{12}+24d^{10}+26d^8+20d^6+15d^4+8d^2+2)/(2d^4(2d^4-1)(d^2+1)^4).$$

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 the case that both $\boldsymbol{u}$ and $\boldsymbol{v}$ are delocalized, corresponding to the discussion in Remark 2.6. Let $\boldsymbol{u} = \mathbf{1}_M/\sqrt{M}$ and $\boldsymbol{v} = \mathbf{1}_n/\sqrt{n}$. Then

$$\mathcal{R}_{dt} := \frac{1}{\sigma}\left(\sqrt{n}(|\langle \widehat{\boldsymbol{v}}, \boldsymbol{v}\rangle|^2 - a(d)) + \frac{d^6 - 1.5d^2 - 1}{d^5(d^2+1)^2}\right)$$

is asymptotically $\mathcal{N}(0,1)$, where σ is defined in (S1).
Case 3. Two-point noise and one of the singular vectors of S is delocalized. We set $\boldsymbol{u} = \mathbf{1}_M/\sqrt{M}$ and $\boldsymbol{v} = \boldsymbol{f}_1$. From Remark 2.7, we know that the random variable

$$\mathcal{R}_{pt} := \frac{\sqrt{n}}{\sigma_t}\left(|\langle \widehat{\boldsymbol{v}}, \boldsymbol{v}\rangle|^2 - a(d)\right)$$

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

$$\sigma_t^2 = \sigma^2 + 2(d^4+d^2+0.5)^2/(d^7(d^2+1)^4) - 0.75(d^4+d^2+0.5)^2/(d^8(d^2+1)^4).$$

Case 4. Two-point noise and both singular vectors of S are sparse (localized). Let $\boldsymbol{u} = \boldsymbol{e}_1$ and $\boldsymbol{v} = \boldsymbol{f}_1$. From the proof of Proposition 5.1, especially the decomposition in (5.11), by setting

$$\mathcal{R}_{st} := \frac{1}{\sigma_s}\left(\sqrt{n}\left(|\langle \widehat{\boldsymbol{v}}, \boldsymbol{v}\rangle|^2 - a(d)\right) - \frac{2\sqrt{n}}{d^3}X_{11}\right),$$

with

$$\sigma_s^2 = (d^{16}+4d^{14}+6d^{12}+d^{10}-6d^8-2d^6+6.5d^4+6.25d^2+1.6875)/(d^8(d^2+1)^4(2d^4-1)),$$

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 $\mathcal{R}_g$ is fairly close to standard Gaussian even for a small sample size $n = 200$. (The same is also observed for $\mathcal{R}_{dt}, \mathcal{R}_{pt}, \mathcal{R}_{st}$.)

	$n = 200$					$n = 500$				
Normal	$d = 2$	$d = 3$	$d = 5$	$d = 10$	SE	$d = 2$	$d = 3$	$d = 5$	$d = 10$	SE
0.01	0.012	0.0134	0.0106	0.0128	0.003	0.0128	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 $\mathcal{R}_g$: Gaussian noise.

	$n = 200$					$n = 500$				
Normal	$d = 2$	$d = 3$	$d = 5$	$d = 10$	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	0.0455	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 $\mathcal{R}_{dt}$: Two-Point noise and delocalized singular vectors.

	$n=200$					$n=500$				
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	0.279	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	0.68	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 3

Distribution of $\mathcal{R}_{pt}$: Two-Point noise and delocalized left singular vector and sparse right singular vector.

	$n=200$					$n=500$				
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	0.009	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	0.889	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

TABLE 4

Distribution of $\mathcal{R}_{st}$: Two-Point noise and both singular vectors sparse.

Further, in Figure S1, we plot the ECDFs of of $\mathcal{R}_g, \mathcal{R}_{dt}, \mathcal{R}_{pt}, \mathcal{R}_{st}$ in subfigures (A), (B), (C), (D) respectively, for $n = 500$ and various values of $d = 2, 3, 5, 10$.

B. Discussions on statistical applications. In this section, we provide simulation results of Section 3.2 and some further discussions on the statistical applications. We first provide the results of the mean-variance discussion of the estimation of singular vectors, which is illustrated in Figure S2. 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 M is even. Assume the left singular vectors of S are $\boldsymbol{u}_1 = \frac{1}{\sqrt{M}}\mathbf{1}_M$ and $\boldsymbol{u}_2 = \frac{1}{\sqrt{M}}(\mathbf{1}_{M/2}^T, -\mathbf{1}_{M/2}^T)^T$, a vector with the first half entries $1/\sqrt{M}$ and remaining entries $-1/\sqrt{M}$. Set $V_0 = (\boldsymbol{f}_1, \boldsymbol{f}_2)$.

Recall the definitions in (2.3) and (2.4). When the noise is Gaussian, we use the statistic

$$\boldsymbol{T}_{1g} = \frac{\sqrt{n}}{\sigma}\Big(\sum_{i,j=1}^{2} |\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_j \rangle|^2 - a(d_1) - a(d_2) \Big), \tag{S1}$$

where

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

Note that $\boldsymbol{T}_{1g}$ is a scaled version of the proposed statistic $\boldsymbol{S}_1$ in (3.1), i.e. $\boldsymbol{T}_{1g} = \boldsymbol{S}_1/\sigma$. When the noise is Two-point type, we use the statistic

$$\boldsymbol{T}_{1t} := \frac{\sqrt{n}}{\sigma_t}\Big(\sum_{i,j=1}^{2} |\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_j \rangle|^2 - a(d_1) - a(d_2) \Big), \tag{S2}$$

where

$$\sigma_t^2 = \sum_{i=1}^{2} \big(4\theta(d_i)^2 + \mathcal{V}^E(d_i)\big) - \frac{3y}{2} \sum_{i=1}^{2} \frac{\theta(d_i)^2}{d_i^2} + \frac{4\sqrt{y}}{\sqrt{2}} \frac{\theta(d_1)^2}{d_1}.$$

$\boldsymbol{T}_{1t}$ is also a scaled version of $\boldsymbol{S}_1$.

Under the nominal level α, we will reject $\boldsymbol{H}_0$ when

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

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantitle 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			
	$\alpha = 0.05$		$\alpha = 0.1$		$\alpha = 0.05$		$\alpha = 0.1$	
	$n = 200$	$n = 500$	$n = 200$	$n = 500$	$n = 200$	$n = 500$	$n = 200$	$n = 500$
$y = 0.5$	0.047	0.0482	0.098	0.0967	0.0501	0.0496	0.105	0.0945
$y = 1$	0.057	0.046	0.092	0.096	0.0488	0.0491	0.097	0.099
$y = 2$	0.0494	0.052	0.0984	0.0955	0.0474	0.049	0.091	0.094

TABLE 5
Type I error under $\boldsymbol{H}_0$ for (T1) using z-score test.

Finally, to study the power of our test against the alternatives, we consider the matrix $V_a = (\boldsymbol{f}_1, \sqrt{1-\delta^2}\boldsymbol{f}_2 + \delta \boldsymbol{f}_3)$ for a parameter $\delta \in (0,1)$. In Figure

S3, we record the simulated power for different values of δ under the nominal level $\alpha = 0.05$ when X is a Two-point noise matrix. We find that the power of our tests increases when δ increases. Furthermore, at the same level of δ, the power is improved when n increases.

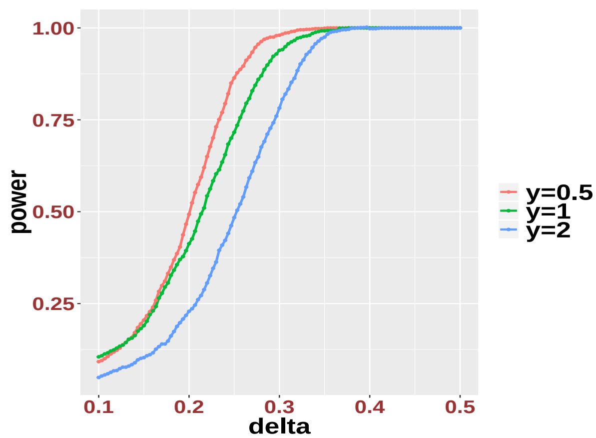


Fig S3: Power vs δ under the nominal level $\alpha = 0.05$ for $y = 0.5, 1, 2$ respectively.

As we mentioned in the article [S1], we assume that D, U and the necessary parameters of X are known and we do the hypothesis testing about V. Although in general we cannot drop all the a priori information about D, U, and X, some efforts can be made along this direction. In the sequel, for instance, we discuss some possible extension to the case when D and the necessary parameters of X are unknown. More specifically, recall that d_i and $\sqrt{\mu_i}$ are the singular values of S and Y, respectively. We know that $\sqrt{\mu_i}$ converges to $p(d_i)$ in probability. In our theorems for singular vectors, we can use $p^{-1}(\sqrt{\mu_i})$ to replace d_i. Such a replacement will change the distribution of our statistics. For instance, if we change $a(d_i)$ by $a(p^{-1}(\sqrt{\mu_i}))$ in Theorem 2.3, such a change will bring additional fluctuation of the statistic. However, one can still use our method to derive the limiting distribution for such a modified statistic where d_i's are replaced by their estimates, i.e., $p^{-1}(\sqrt{\mu_i})$'s. It is simply because the fluctuation of $\sqrt{\mu_i}$ can also be written as a quadratic form of the Green function. We anyway have the joint distribution of the quadratic forms of the Green function and its derivative. So we can also derive the joint distribution of the singular vectors and the singular values. Replacing d_i in quantities like Δ_i and $\mathcal{V}_i$ by the estimator $p^{-1}(\sqrt{\mu_i})$ is completely harmless since the error is of order $\frac{1}{\sqrt{n}}$. Further, in some simple case, we can also estimate the cumulants of the noise of X. For instance, suppose we use the $p^{-1}(\sqrt{\mu}_i)$ to replace d_i in our theorems, and further we assume that the left singular vectors are known and we want to test whether the right singular subspace V is identical to some given matrix V_0. In this case, we can estimate the parameters of X by considering $\widehat{X} := Y - \widehat{S}$. Here $\widehat{S} = U\widehat{D}V^*$ with $\widehat{D} = \operatorname{diag}(p^{-1}(\sqrt{\mu_i}))_{i=1}^r$. Since the entries of X are assumed to be i.i.d., we can estimate the the second moment of $\sqrt{n}x_{ij}$'s as the following

$$\frac{1}{M}\sum_{i,j}|x_{ij}|^2 = \frac{1}{M}\mathrm{Tr}XX^* = \frac{1}{M}\mathrm{Tr}\widehat{X}\widehat{X}^* - \frac{1}{M}\mathrm{Tr}(S-\widehat{S})\widehat{X}^* - \frac{1}{M}\mathrm{Tr}\widehat{X}(S-\widehat{S})^* + \frac{1}{M}\mathrm{Tr}(S-\widehat{S})(S-\widehat{S})^*.$$

Using the facts $\|\widehat{X}\| = O(1)$, $\|S - \widehat{S}\| = O(n^{-\frac{1}{2}})$ and $\operatorname{rank}(S - \widehat{S}) = r$ which is fixed, it is easy to see that the last three terms are of order n^{-1} in probability. Further, it is easy to see that $\frac{1}{M}\mathrm{Tr}XX^*$ can estimate $\mathbb{E}(\sqrt{n}x_{ij})^2$ up to an error of order $\frac{1}{n}$ in probability. Hence, we can estimate $\mathbb{E}(\sqrt{n}x_{ij})^2$ by $\frac{1}{M}\mathrm{Tr}\widehat{X}\widehat{X}^*$ which can be computed from the data, if the null hypothesis holds. The other cumulants can be estimated in a similar way.

In practice, some of the extension above could be quite important. For instance, the joint distribution of the singular values and vectors allow us to consider the inference on statistics involving both of them. In [S8], to test whether the community memberships of the two networks are the same in the stochastic block model, the authors proposed a statistic involving the scaled principal angles, where the scalings are the singular values (see [S8, Section 4.2] for details). Further, in [S4], the authors derived the formulas for the optimal shrinkers of the singular values under various norms. These shrinkers are essentially combinations of products of the singular values and inner products of $|\langle \boldsymbol{u}_i, \widehat{\boldsymbol{u}}_i\rangle|, |\langle \boldsymbol{v}_i, \widehat{\boldsymbol{v}}_i\rangle|$. Finally, there exist a lot works on estimating the low rank matrix S, to name but a few [S5, S10, S11]. With the results on the joint distribution, it is possible for us to do inference on the estimation of the low-rank matrix S. Nevertheless, we leave all the extensions to the future work.

C. Preliminary results. In this section, we list some preliminary results which will be used in the technical proof.

C.1. *Auxiliary lemmas.* A key tool for our computation is the following cumulant expansion formula, whose proof can be found in [S9, Proposition 3.1] and [S6, Section II], for instance.

LEMMA C.1. *Let $\ell \in \mathbb{N}$ be fixed and let $f \in C^{\ell+1}(\mathbb{R})$. Let ξ be a centered random variable with finite first $\ell+2$ moments. Let $\kappa_k(\xi)$ be the k-th cumulant of ξ. Then we have the expansion*

$$\mathbb{E}(\xi f(\xi)) = \sum_{k=1}^{\ell} \frac{\kappa_{k+1}(\xi)}{k!} \mathbb{E}(f^{(k)}(\xi)) + \mathbb{E}(\epsilon_\ell(\xi f(\xi))), \tag{S1}$$

where $\epsilon_\ell(\xi f(\xi))$ satisfies

$$|\mathbb{E}(\epsilon_\ell(\xi f(\xi)))| \le C_\ell \mathbb{E}(|\xi|^{\ell+2}) \sup_{|t|\le\chi} |f^{(\ell+1)}(t)| + C_\ell \mathbb{E}(|\xi|^{\ell+2}\mathbf{1}(|\xi|>\chi)) \sup_{t\in\mathbb{R}} |f^{(\ell+1)}(t)|$$

for any $\chi > 0$.

Note that when ξ is a standard Gaussian random variable (i.e. $\kappa_i = 0, i \ge 3$), (S1) boils down to the celebrated Stein's lemma [S12]. Next we introduce the identities on the derivatives of the Green functions in (4.5). These can be verified by elementary calculus so we omit the proofs. 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 C.2. *Let $\mathcal{E}_{ij} = E_{ij'} + E_{j'i}$ and $k \in \mathbb{N}$. We have*

$$\frac{\partial^k G}{\partial x_{ij}^k} = (-1)^k k! z^{\frac{k}{2}} (G\mathcal{E}_{ij})^k G,$$
$$\frac{\partial^k (G^2)}{\partial x_{ij}^k} = (-1)^k k! z^{\frac{k}{2}} \sum_{s=0}^{k} (G\mathcal{E}_{ij})^s G (G\mathcal{E}_{ij})^{k-s} G.$$

Below we also collect some basic results on convergence and equivalence in distribution for sum of random variables. They can be found in [S7, Lemma 7.7, 7.8 and 7.10].

LEMMA C.3. *(1). Let $\mathsf{X}_n \simeq \mathsf{Y}_n$ and R_n satisfy $\lim_{n\to\infty} \mathbb{P}\Big(|\mathsf{R}_n| \le \epsilon_n\Big) = 1$, where $\{\epsilon_n\}$ is a positive null sequence. Then*

$$\mathsf{X}_n \simeq \mathsf{Y}_n + \mathsf{R}_n.$$

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

$$\mathsf{X}_n + \mathsf{Y}_n \simeq \mathsf{X}'_n + \mathsf{Y}'_n.$$

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

$$\mathbb{E}f(\mathsf{Z}_n\mathsf{X}_n) - \mathbb{E}f(\mathsf{Z}_n\mathsf{X}) \to 0.$$

C.2. *Collection of derivatives.* In this part, we summarize some basic identities on the derivatives of G and Q defined in (5.12) without proof. Recall the notation introduced in (5.4).

Using Lemma C.2, it is easy to check

$$\left(\frac{\partial^2 G}{\partial x_{ij}^2} W\right)_{ab} = 2z \sum_{\substack{l_1,\cdots,l_4 \in \{i,j'\} \\ l_1 \neq l_2, l_3 \neq l_4}} G_{al_1} G_{l_2 l_3} (GW)_{l_4 b}, \tag{S2}$$

$$\left(\frac{\partial^3 G}{\partial x_{ij}^3} W\right)_{ab} = -6z^{\frac{3}{2}} \sum_{\substack{l_1,\cdots,l_6 \in \{i,j'\} \\ l_1 \neq l_2, l_3 \neq l_4, l_5 \neq l_6}} G_{al_1} G_{l_2 l_3} G_{l_4 l_5} (GW)_{l_6 b}, \tag{S3}$$

$$\left(\frac{\partial^4 G}{\partial x_{ij}^4} W\right)_{ab} = 24z^2 \sum_{\substack{l_1,\cdots,l_8 \in \{i,j'\} \\ l_1 \neq l_2, l_3 \neq l_4, l_5 \neq l_6, l_7 \neq l_8}} G_{al_1} G_{l_2 l_3} G_{l_4 l_5} G_{l_6 l_7} (GW)_{l_8 b}. \tag{S4}$$

and also the following identities

$$\left(\frac{\partial^2 G^2}{\partial x_{ij}^2} W\right)_{ab} = 2z \sum_{(a_1,a_2,a_3) \in \mathcal{P}(2,1,1)} \sum_{\substack{l_1,\cdots,l_4 \in \{i,j'\} \\ l_1 \neq l_2, l_3 \neq l_4}} G^{a_1}_{al_1} G^{a_2}_{l_2 l_3} (G^{a_3} W)_{l_4 b},$$

$$\left(\frac{\partial^3 G^2}{\partial x_{ij}^3} W\right)_{ab} = -6z^{\frac{3}{2}} \sum_{(a_1,\ldots,a_4) \in \mathcal{P}(2,1,1,1)} \sum_{\substack{l_1,\cdots,l_6 \in \{i,j'\} \\ l_1 \neq l_2, l_3 \neq l_4, l_5 \neq l_6}} G^{a_1}_{al_1} G^{a_2}_{l_2 l_3} G^{a_3}_{l_4 l_5} (G^{a_4} W)_{l_6 b},$$

(S5)

$$\left(\frac{\partial^4 G^2}{\partial x_{ij}^4} W\right)_{ab} = 24z^2 \sum_{(a_1,\ldots,a_5) \in \mathcal{P}(2,1,1,1,1)} \sum_{\substack{l_1,\cdots,l_8 \in \{i,j'\} \\ l_1 \neq l_2, l_3 \neq l_4, l_5 \neq l_6, l_7 \neq l_8}} G^{a_1}_{al_1} G^{a_2}_{l_2 l_3} G^{a_3}_{l_4 l_5} G^{a_4}_{l_6 l_7} (G^{a_5} W)_{l_8 b}.$$

Similarly, using Lemma C.2 and a discussion similar to (S1), we can also derive

$$\frac{\partial^2 Q}{\partial x_{ij}^2} = 2z\sqrt{n} \sum_{\substack{l_1,\cdots,l_4 \in \{i,j'\} \\ l_1 \neq l_4, l_2 \neq l_3}} \Big((GAG)_{l_1 l_2} G_{l_3 l_4} - \frac{1}{2z} (GBG)_{l_1 l_2} G_{l_3 l_4}$$

(S6)
$$+\frac{1}{2}\sum_{(a_1,a_2,a_3)\in\mathcal{P}(2,1,1)}(G^{a_1}BG^{a_2})_{l_1l_2}G^{a_3}_{l_3l_4}\Big),$$

$$\frac{\partial^3 Q}{\partial x_{ij}^3}=-6z^{\frac{3}{2}}\sqrt{n}\sum_{\substack{l_1,\cdots,l_6\in\{i,j'\}\\ l_1\neq l_6,l_2\neq l_3,l_4\neq l_5}}\Big((GAG)_{l_1l_2}G_{l_3l_4}G_{l_5l_6}-\frac{1}{2z}(GBG)_{l_1l_2}G_{l_3l_4}G_{l_5l_6}$$

(S7)
$$+\frac{1}{2}\sum_{(a_1,\ldots,a_4)\in\mathcal{P}(2,1,1,1)}(G^{a_1}BG^{a_2})_{l_1l_2}G^{a_3}_{l_3l_4}G^{a_4}_{l_5l_6}\Big),$$

$$\frac{\partial^4 Q}{\partial x_{ij}^4}=24z^2\sqrt{n}\sum_{\substack{l_1,\cdots,l_8\in\{i,j'\}\\ l_1\neq l_8,l_2\neq l_3,\\ l_4\neq l_5,l_6\neq l_7}}\Big((GAG)_{l_1l_2}G_{l_3l_4}G_{l_5l_6}G_{l_7l_8}-\frac{1}{2z}(GBG)_{l_1l_2}G_{l_3l_4}G_{l_5l_6}G_{l_7l_8}$$

(S8)
$$+\frac{1}{2}\sum_{(a_1,\ldots,a_5)\in\mathcal{P}(2,1,1,1,1)}(G^{a_1}BG^{a_2})_{l_1l_2}G^{a_3}_{l_3l_4}G^{a_4}_{l_5l_6}G^{a_5}_{l_7l_8}\Big).$$

D. Proof of Lemmas 4.7 and 4.8.

Proof of Lemma 4.7. We focus our discussion on the first identity (4.18). 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

$$G^2(H-z)=G,\ G^3(H-z)=G^2.$$

This completes the proof. □

Proof of Lemma 4.8. To prove Lemma 4.8, we first need the following result from [S3].

Lemma D.1 (Theorem 3.3 and 3.4 of [S3]). *Under assumptions of (1.3), (1.4), (1.8) 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

$$|\langle \boldsymbol{u}_i, \widehat{\boldsymbol{u}}_i \rangle^2 - a_1(d_i)| = O_\prec(n^{-\frac{1}{2}}), \qquad |\langle \boldsymbol{v}_i, \widehat{\boldsymbol{v}}_i \rangle^2 - a(d_i)| = O_\prec(n^{-\frac{1}{2}}),$$

and for $1 \leq i \neq j \leq r$,

$$|\langle \boldsymbol{u}_i, \widehat{\boldsymbol{u}}_j \rangle|^2 = O_\prec(\frac{1}{n}), \qquad |\langle \boldsymbol{v}_i, \widehat{\boldsymbol{v}}_j \rangle|^2 = O_\prec(\frac{1}{n}).$$

With Lemma D.1, we can rewrite (1.7) as

$$L = \sum_{i=1}^r |\langle \widehat{\boldsymbol{u}}_i, \boldsymbol{u}_i \rangle|^2 + O_\prec(\frac{1}{n}) \quad \text{and} \quad R = \sum_{i=1}^r |\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_i \rangle|^2 + O_\prec(\frac{1}{n}). \tag{S1}$$

We next write the above quantities in terms of the Green functions. Recall from (4.1) $\mathcal{Y} \equiv \mathcal{Y}(z)$ and denote by $\widehat{G}(z) = (\mathcal{Y} - z)^{-1}$. By spectral decomposition, we write

$$\begin{aligned} \widehat{G}(z) &= \sum_{i=1}^{M \wedge n} \frac{1}{\mu_i - z} \begin{pmatrix} \widehat{\boldsymbol{u}}_i \widehat{\boldsymbol{u}}_i^* & z^{-1/2} \sqrt{\mu_i} \widehat{\boldsymbol{u}}_i \widehat{\boldsymbol{v}}_i^* \\ z^{-1/2} \sqrt{\mu_i} \widehat{\boldsymbol{v}}_i \widehat{\boldsymbol{u}}_i^* & \widehat{\boldsymbol{v}}_i \widehat{\boldsymbol{v}}_i^* \end{pmatrix} \\ &\quad - \frac{1}{z} \sum_{i=M \wedge n}^{M \vee n} \begin{pmatrix} \mathbf{1}_{M>n} \widehat{\boldsymbol{u}}_i \widehat{\boldsymbol{u}}_i^* & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{M<n} \widehat{\boldsymbol{v}}_i \widehat{\boldsymbol{v}}_i^* \end{pmatrix}. \end{aligned} \tag{S2}$$

For any $i \in [r]$, denote $\Gamma_i := \partial B_\rho(d_i)$, where $B_\rho(d_i)$ is the open disc of radius ρ around d_i. Here ρ 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. Denote

$$\widehat{\mathcal{G}}_1(z) = (YY^* - z)^{-1}, \ \widehat{\mathcal{G}}_2(z) = (Y^*Y - z)^{-1}.$$

Note that on one hand, we have for $i \leq r$,

$$\langle \boldsymbol{v}_i, \widehat{\mathcal{G}}_2(z) \boldsymbol{v}_i \rangle = \langle \mathbf{v}_i, \widehat{G}(z) \mathbf{v}_i \rangle, \quad \mathbf{v}_i = (\mathbf{0}, \boldsymbol{v}_i^*)^*.$$

One the other hand, by Lemma D.1 and Cauchy's integral formula, with high probability, we have

$$\widehat{\boldsymbol{v}}_i \widehat{\boldsymbol{v}}_i^* = -\frac{1}{2\pi \mathrm{i}} \oint_{p(\Gamma_i)} \widehat{\mathcal{G}}_2(z) dz.$$

Together with (S2), with high probability, we have the following integral representation

$$|\langle \boldsymbol{v}_i, \widehat{\boldsymbol{v}}_i \rangle|^2 = \frac{1}{2 d_i^2 \pi \mathrm{i}} \oint_{p(\Gamma_i)} \left((\mathcal{D}^{-1} + \mathcal{U}^* G(z) \mathcal{U})^{-1} \right)_{ii} \frac{\mathrm{d}z}{z},$$

where we used the fact that

$$\mathcal{U}^*\widehat{G}(z)\mathcal{U} = \mathcal{D}^{-1} - \mathcal{D}^{-1}(\mathcal{D}^{-1} + \mathcal{U}^*G(z)\mathcal{U})^{-1}\mathcal{D}^{-1}.$$

Recall (4.10) and denote

$$\Psi(z) = -\mathcal{U}^*(\Pi_1(z) - G(z))\mathcal{U}. \tag{S3}$$

Using Lemma 4.4, we have

$$\|\Psi(z)\|_{\mathrm{op}} = O_{\prec}(n^{-\frac{1}{2}}), \quad z \in \boldsymbol{S}_o. \tag{S4}$$

We can decompose $\mathcal{D}^{-1}(\mathcal{D}^{-1} + \mathcal{U}^*G(z)\mathcal{U})^{-1}\mathcal{D}^{-1}$ as

$$\mathcal{D}^{-1} + \mathcal{U}^*G(z)\mathcal{U} = \mathcal{D}^{-1} + \mathcal{U}^*\Pi_1(z)\mathcal{U} + \Psi(z).$$

We further employ the resolvent expansion for $(\mathcal{D}^{-1} + \mathcal{U}^*G(z)\mathcal{U})^{-1}$ to write

$$|\langle \boldsymbol{v}_i, \widehat{\boldsymbol{v}}_i\rangle|^2 = \frac{1}{d_i^2}(S_0 + S_1) + O_{\prec}(\frac{1}{n}),$$

where

$$S_0 = \frac{1}{2\pi \mathrm{i}} \oint_{p(\Gamma_i)} \big(\big(\mathcal{D}^{-1} + \mathcal{U}^*\Pi_1(z)\mathcal{U}\big)^{-1}\big)_{ii} \frac{\mathrm{d}z}{z},$$

(S5)

$$S_1 = \frac{1}{2\pi \mathrm{i}} \oint_{p(\Gamma_i)} \Big(\big(\mathcal{D}^{-1} + \mathcal{U}^*\Pi_1(z)\mathcal{U}\big)^{-1}\Psi(z)\big(\mathcal{D}^{-1} + \mathcal{U}^*\Pi_1(z)\mathcal{U}\big)^{-1}\Big)_{ii} \frac{\mathrm{d}z}{z}.$$

Here we used a discussion similar to Eq. (5.19) and Lemma 5.5 of [S3] and omit further details. By the residual theorem, we have $S_0 = d_i^2 a(d_i)$. Recall (4.23) and denote

$$f_i(z) := -\mathrm{Tr}\big(\Xi_1(z)\mathcal{U}W_i(z)\mathcal{U}^*\big).$$

We can then write

$$S_1 = \frac{1}{2\pi \mathrm{i}} \oint_{p(\Gamma_i)} \frac{z f_i(z)}{(z m_{1c}(z) m_{2c}(z) - d_i^{-2})^2} \mathrm{d}z.$$

As $p(d)$ is a monotone function when $d > y^{1/4}$ and by Lemma 4.1, we find that

$$S_1 = \frac{d_i^4}{2\pi \mathrm{i}} \oint_{\Gamma_i} \frac{p(\zeta) f_i(p(\zeta))\zeta^4 p'(\zeta)}{(d_i - \zeta)^2 (d_i + \zeta)^2} \mathrm{d}\zeta.$$

Then, by residue theorem, we obtain

(S6)
$$S_1 = d_i^4\Big(f_i(p(\zeta))\frac{\zeta^4 p'(\zeta)p(\zeta)}{(d_i+\zeta)^2}\Big)'\Big|_{\zeta=d_i} = d_i^2\mathrm{Tr}\big(\Xi_1(p_i)A_i^R\big) + d_i^2\mathrm{Tr}\big(\Xi_1'(p_i)B_i^R\big),$$

where we recall (4.22) and the definitions of A_i^R and B_i^R in (4.25). The conclusion for $|\langle \boldsymbol{v}_i, \widehat{\boldsymbol{v}}_i\rangle|^2$ follows immediately.

The above discussion holds for all $i \in [r]$. Rearranging the terms of (S6) and using Lemma 4.1, we can conclude our proof for R using (S1). Similar discussion yields the conclusion of $|\langle \boldsymbol{u}_i, \widehat{\boldsymbol{u}}_i\rangle|^2$ for each $i \in [r]$ and L. This completes the proof of Lemma 4.8. □

E. 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 Δ, separately. However, for brevity, we will omit both two parameters for simplicity in the sequel.

First of all, applying (4.17) to the definition in (5.1), we have

$$\mathcal{Q} = O_{\prec}(1). \tag{S1}$$

Denote $(M+n)\times(M+n)$ diagonal matrices

$$\mathtt{I}^{\mathrm{u}} := \begin{pmatrix} I_M & \\ & 0 \end{pmatrix} \quad \text{and} \quad \mathtt{I}^{\mathrm{l}} := \begin{pmatrix} 0 & \\ & I_n \end{pmatrix}. \tag{S2}$$

We further define $A_1 = A\mathtt{I}^{\mathrm{u}}$, $A_2 = A\mathtt{I}^{\mathrm{l}}$ and define B_1, B_2 analogously. In addition, we set

$$\begin{aligned} f_\alpha &:= -m_\alpha \mathrm{Tr} H\Xi_1 A_\alpha + (1+zm_\alpha)\mathrm{Tr} G A_\alpha, \\ g_\alpha &:= -\frac{m_\alpha}{2}\mathrm{Tr}H\Xi_2 B_\alpha + \frac{1+zm_\alpha}{2}\mathrm{Tr}G^2B_\alpha + \frac{zm_\alpha - 1}{2z}\mathrm{Tr}GB_\alpha \\ &\quad - m_\alpha'\mathrm{Tr}B_\alpha + m_\alpha'\mathrm{Tr}H\Pi_1 B_\alpha, \qquad \alpha = 1,2. \end{aligned} \tag{S3}$$

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

Lemma E.1. *Recall (5.5) and (5.6). For z defined in (5.14), we have*

$$Q = \sqrt{n}\big(f_1+f_2+g_1+g_2\big) + \sqrt{nz}\sum_{(i,j)\in\mathcal{S}(\nu)} c_{ij}x_{ij} - \Delta_d. \tag{S4}$$

To state the second crucial lemma, Lemma E.2. We first introduce some notations. Recall that Π_a $(1\le a\le 4)$ in (4.10) and (4.19) approximates

G^a. We introduce the following matrices to approximate the powers of G interacting with block diagonal matrices $\mathtt{I}^{\mathrm{u}}$ and $\mathtt{I}^{\mathrm{l}}$. For $1 \leq a_1, a_2 \leq 2$, define

$$\Pi^{\mathrm{u}}_{a_1,a_2} := \Pi_{a_1}\mathtt{I}^{\mathrm{u}}\Pi_{a_2} \quad \text{and} \quad \Pi^{\mathrm{l}}_{a_1,a_2} := \Pi_{a_1}\mathtt{I}^{\mathrm{l}}\Pi_{a_2}. \tag{S5}$$

Note that they approximate $G^{a_1}\mathtt{I}^{\mathrm{u}}\Pi_{a_2}$ and $G^{a_2}\mathtt{I}^{\mathrm{l}}\Pi_{a_2}$ respectively. We further define

$$\Pi^{\mathrm{u}}_2 := m_1' I_M \oplus (m_2' + \frac{1}{z}m_2)I_n \quad \text{and} \quad \Pi^{\mathrm{l}}_2 := (m_1' + \frac{1}{z}m_1)I_M \oplus m_2' I_n, \tag{S6}$$

which approximate $G\mathtt{I}^{\mathrm{u}}G$ and $G\mathtt{I}^{\mathrm{l}}G$.

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

$$\begin{aligned}
&\mathfrak{d}_1^a := \frac{2z}{n}\sum_{i,j}(\Pi_1)_{ii}(\Pi_1)_{j'j'}(\Pi_1 A_1)_{j'i}, \quad \mathfrak{d}_2^a := \frac{2z}{n}\sum_{i,j}(\Pi_1)_{ii}(\Pi_1)_{j'j'}(\Pi_1 A_2)_{ij'},\\
&\tilde{\mathfrak{d}}_1 := \frac{2z}{n}\sum_{(a_1,a_2,a_2)\in\mathcal{P}(2,1,1)}\sum_{i,j}(\Pi_{a_1})_{ii}(\Pi_{a_2})_{j'j'}(\Pi_{a_3}B_1)_{j'i},\\
&\tilde{\mathfrak{d}}_2 := \frac{2z}{n}\sum_{(a_1,a_2,a_2)\in\mathcal{P}(2,1,1)}\sum_{i,j}(\Pi_{a_1})_{ii}(\Pi_{a_2})_{j'j'}(\Pi_{a_3}B_2)_{ij'}.
\end{aligned} \tag{S7}$$

And $\mathfrak{d}_1^b$ (resp. $\mathfrak{d}_2^b$) is defined by replacing A_1 (resp. A_2) to B_1 (resp. B_2) in the expression of $\mathfrak{d}_1^a$ (resp. $\mathfrak{d}_2^a$). Using (S2), we further set

$$\begin{aligned}
\Pi^{\mathrm{u}}_3 &:= (m_1'' + \frac{1}{z}m_1')I_M \oplus (m_2'' + \frac{2}{z}m_2')I_n,\\
\Pi^{\mathrm{l}}_3 &:= (m_1'' + \frac{2}{z}m_1')I_M \oplus (m_2'' + \frac{1}{z}m_2')I_n,\\
\Pi^{\mathrm{u}}_4 &:= (\frac{2}{3}m_1^{(3)} + \frac{2}{z}m_1'' + \frac{1}{z^2}m_1')I_M \oplus (\frac{2}{3}m_2^{(3)} + \frac{2}{z}m_2'')I_n,\\
\Pi^{\mathrm{l}}_4 &:= (\frac{2}{3}m_1^{(3)} + \frac{2}{z}m_1'')I_M \oplus (\frac{2}{3}m_2^{(3)} + \frac{2}{z}m_2'' + \frac{1}{z^2}m_2')I_n.
\end{aligned} \tag{S8}$$

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

$$\begin{aligned}
\mathfrak{a}_{11} := -(k-1)\sqrt{z}\Big(&2\mathrm{Tr}(\Pi^{\mathrm{l}}_2 - \Pi^{\mathrm{l}}_{1,1})A_1\Pi_1 A - \frac{1}{z}\mathrm{Tr}(\Pi^{\mathrm{l}}_2 - \Pi^{\mathrm{l}}_{1,1})A_1\Pi_1 B\\
&+ \mathrm{Tr}(\Pi^{\mathrm{l}}_2 - \Pi^{\mathrm{l}}_{1,1})A_1\Pi_2 B + \mathrm{Tr}(\Pi^{\mathrm{l}}_3 - \Pi^{\mathrm{l}}_{2,1})A_1\Pi_1 B\Big),
\end{aligned}$$

$$\tilde{\mathfrak{b}}_{11} := -(k-1)\sqrt{z}\Big(2\mathrm{Tr}(\Pi^{\mathrm{l}}_3 - \Pi^{\mathrm{l}}_{1,2})B_1\Pi_1 A - \frac{1}{z}\mathrm{Tr}(\Pi^{\mathrm{l}}_3 - \Pi^{\mathrm{l}}_{1,2})B_1\Pi_1 B$$
$$+ \mathrm{Tr}(\Pi^{\mathrm{l}}_3 - \Pi^{\mathrm{l}}_{1,2})B_1\Pi_2 B + \mathrm{Tr}(\Pi^{\mathrm{l}}_4 - \Pi^{\mathrm{l}}_{2,2})B_1\Pi_1 B\Big). \tag{S9}$$

In addition, $\mathfrak{a}_{12}$ is defined via replacing A_1 with A_2 and $\Pi^{\mathrm{l}}_a, \Pi^{\mathrm{l}}_{a_1,a_2}$ with $\Pi^{\mathrm{u}}_a, \Pi^{\mathrm{u}}_{a_1,a_2}$ in the definition of $\mathfrak{a}_{11}$. We further define $\mathfrak{b}_{11}$ (resp. $\mathfrak{b}_{12}$) via replacing A_1 (resp. A_2) with B_1 (resp. B_2) in the definition of $\mathfrak{a}_{11}$ (resp. $\mathfrak{a}_{12}$). Similarly, $\tilde{\mathfrak{b}}_{12}$ is obtained by replacing B_1 with B_2 and $\Pi^{\mathrm{l}}_a, \Pi^{\mathrm{l}}_{a_1,a_2}$ with $\Pi^{\mathrm{u}}_a, \Pi^{\mathrm{u}}_{a_1,a_2}$ in the definition of $\tilde{\mathfrak{b}}_{11}$.

Next, recall c_{ij} defined in (5.7) and set

$$\mathfrak{a}_{21} := -\frac{(k-1)z}{\sqrt{n}} \sum_{(i,j)\in\mathcal{S}(\nu)} (\Pi_1)_{j'j'}(\Pi_1 A_1)_{ii} c_{ij},$$
$$\tilde{\mathfrak{b}}_{21} := -\frac{(k-1)z}{\sqrt{n}} \sum_{(i,j)\in\mathcal{S}(\nu)} \big((\Pi_1)_{j'j'}(\Pi_2 B_1)_{ii} + (\Pi_2)_{j'j'}(\Pi_1 B_1)_{ii}\big) c_{ij}. \tag{S10}$$

Further, $\mathfrak{a}_{22}$ (resp. $\tilde{\mathfrak{b}}_{22}$) 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 $\mathfrak{a}_{21}$ (resp. $\tilde{\mathfrak{b}}_{21}$). Then we recall s_{ij} in (5.8) and set

$$\mathfrak{a}_{31} := -\frac{2(k-1)z^{3/2}}{n} \sum_{i,j} (\Pi_1)_{j'j'}(\Pi_1 A_1)_{ii} s_{ij},$$
$$\tilde{\mathfrak{b}}_{31} = -\frac{2(k-1)z^{3/2}}{n} \sum_{i,j} \big((\Pi_1)_{j'j'}(\Pi_2 B_1)_{ii} + (\Pi_2)_{j'j'}(\Pi_1 B_1)_{ii}\big) s_{ij}. \tag{S11}$$

Further, $\mathfrak{a}_{32}$ (resp. $\tilde{\mathfrak{b}}_{32}$) 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 $\mathfrak{a}_{31}$ (resp. $\tilde{\mathfrak{b}}_{31}$). Also, $\mathfrak{b}_{31}$ (resp. $\mathfrak{b}_{32}$) is defined by replacing A_1 (resp. A_2) with B_1 (resp. B_2) in the definition of $\mathfrak{a}_{31}$ (resp. $\mathfrak{a}_{32}$).

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

$$\mathfrak{a}_{0\alpha} := \mathfrak{a}_{1\alpha} + \kappa_3 \mathfrak{a}_{2\alpha} + \frac{\kappa_4}{2}\mathfrak{a}_{3\alpha},$$
$$\mathfrak{b}_{0\alpha} := \frac{m_\alpha}{2}\tilde{\mathfrak{b}}_{1\alpha} + m'_\alpha \mathfrak{b}_{1\alpha} + \frac{\kappa_3 m_\alpha}{2}\tilde{\mathfrak{b}}_{2\alpha} + \kappa_3 m'_\alpha \mathfrak{b}_{2\alpha} + \frac{\kappa_4 m_\alpha}{4}\tilde{\mathfrak{b}}_{3\alpha} + \frac{\kappa_4 m'_\alpha}{2}\mathfrak{b}_{3\alpha}. \tag{S12}$$

For brevity, we also adopt the notation

$$\mathfrak{q}^{(l)} = Q^l(z) e^{\mathrm{i}t\Delta(z_0)}.$$

Recall the notations in (S3). With the above notations, we now state the following lemma.

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

$$\sqrt{n}\mathbb{E} f_\alpha \mathfrak{q}^{(k-1)} = -\sqrt{z} m_\alpha \mathbb{E}\Big(\frac{\kappa_3}{2}\mathfrak{d}_\alpha^a \mathfrak{q}^{(k-1)} + \mathfrak{a}_{0\alpha}\mathfrak{q}^{(k-2)}\Big) + O_\prec(n^{-\frac{1}{2}+4\nu}), \tag{S13}$$

$$\sqrt{n}\mathbb{E} g_\alpha \mathfrak{q}^{(k-1)} = -\sqrt{z}\mathbb{E}\Big(\frac{\kappa_3}{4}\Big(m_\alpha \tilde{\mathfrak{d}}_\alpha + 2m'_\alpha \mathfrak{d}_\alpha^b\Big)\mathfrak{q}^{(k-1)} + \mathfrak{b}_{0\alpha}\mathfrak{q}^{(k-2)}\Big) + O_\prec(n^{-\frac{1}{2}+4\nu}), \tag{S14}$$

In addition, we also have

$$\begin{aligned}\sqrt{nz}\sum_{(i,j)\in\mathcal{S}(\nu)} c_{ij}\mathbb{E} x_{ij}\mathfrak{q}^{(k-1)} = (k-1)\Big(z\sum_{(i,j)\in\mathcal{S}(\nu)} c_{ij}^2 + \frac{z^{\frac{3}{2}}\kappa_3}{\sqrt{n}}\sum_{(i,j)\in\mathcal{S}(\nu)} s_{ij}c_{ij}\Big)\mathbb{E}\mathfrak{q}^{(k-2)} \\ + O_\prec(n^{-\frac{1}{2}+4\nu}).\end{aligned} \tag{S15}$$

With Lemmas E.1 and E.2, we can now prove Proposition 5.2.

PROOF OF PROPOSITION 5.2. By simply combining Lemma E.1 and E.2, we can write

$$\mathbb{E}\mathfrak{q}^{(k)} = \mathfrak{c}_1\mathbb{E}\mathfrak{q}^{(k-1)} + \mathfrak{c}_2\mathbb{E}\mathfrak{q}^{(k-2)} - \Delta_d\mathbb{E}\mathfrak{q}^{(k-1)} + O_\prec(n^{-\frac{1}{2}+4\nu}),$$

where

$$\begin{aligned}\mathfrak{c}_1 &= -\sqrt{z}\kappa_3\sum_{\alpha=1,2}\Big(\frac{1}{2}m_\alpha\mathfrak{d}_\alpha^a + \frac{1}{4}m_\alpha\tilde{\mathfrak{d}}_\alpha + \frac{1}{2}m'_\alpha\mathfrak{d}_\alpha^b\Big),\\ \mathfrak{c}_2 &= -\sqrt{z}\sum_{\alpha=1,2}\Big(m_\alpha\mathfrak{a}_{1\alpha} + \kappa_3 m_\alpha\mathfrak{a}_{2\alpha} + \frac{\kappa_4 m_\alpha}{2}\mathfrak{a}_{3\alpha} + \frac{m_\alpha}{2}\tilde{\mathfrak{b}}_{1\alpha} + m'_\alpha\mathfrak{b}_{1\alpha} \\ &\qquad + \frac{\kappa_3 m_\alpha}{2}\tilde{\mathfrak{b}}_{2\alpha} + \kappa_3 m'_\alpha\mathfrak{b}_{2\alpha} + \frac{\kappa_4 m_\alpha}{4}\tilde{\mathfrak{b}}_{3\alpha} + \frac{\kappa_4 m'_\alpha}{2}\mathfrak{b}_{3\alpha}\Big)\\ &\quad + z\sum_{(i,j)\in\mathcal{S}(\nu)} c_{ij}^2 + \frac{z^{\frac{3}{2}}\kappa_3}{\sqrt{n}}\sum_{(i,j)\in\mathcal{S}(\nu)} s_{ij}c_{ij}.\end{aligned}$$

Also recall Δ_d from (5.5) and V from (5.9). By substituting the definitions of the notations in (S7), (S9), (S10), (S11), and also their analogues, it is elementary to check

$$\mathfrak{c}_1 = \Delta_d, \qquad \mathfrak{c}_2 = V. \tag{S16}$$

This completes the proof of (5.16). Further we can regard (5.15) as a degenerate case of (5.16). 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 E.1 and E.2. We prove Lemma E.1 in the rest of this section, and state the proof of Lemma E.2 in Section F.

PROOF OF LEMMA E.1. Recall from (5.11) and (5.12) that

$$Q = \mathcal{Q} - \Delta_r - \Delta_d. \tag{S17}$$

For brevity, we also write

$$F_1 = 1 + zm_1,\ F_2 = 1 + zm_2. \tag{S18}$$

By (4.7) and (4.8), it is easy to check that

$$F_1 = -zm_1m_2,\ F_2 = -zym_1m_2. \tag{S19}$$

Note that by definition $\mathrm{Tr}GA = \mathrm{Tr}GA_1 + \mathrm{Tr}GA_2$ and $\mathrm{Tr}\Pi_1 A = m_1\mathrm{Tr}A_1 + m_2\mathrm{Tr}A_2$. Thus using (S18), we have

$$\begin{aligned}\mathrm{Tr}\Xi_1 A &= \mathrm{Tr}GA_1 + \mathrm{Tr}GA_2 - m_1\mathrm{Tr}A_1 - m_2\mathrm{Tr}A_2\\ &= -m_1\mathrm{Tr}HGA_1 - m_2\mathrm{Tr}HGA_2 + F_1\mathrm{Tr}GA_1 + F_2\mathrm{Tr}GA_2,\end{aligned} \tag{S20}$$

where in the last step, we used the fact $zG = HG - I$.

Using (4.18) and (4.19), one can write

$$\mathrm{Tr}\Xi_1' B = \frac{1}{2}\mathrm{Tr}G^2B_1 + \frac{1}{2}\mathrm{Tr}G^2B_2 - \frac{1}{2z}\mathrm{Tr}GB_1 - \frac{1}{2z}\mathrm{Tr}GB_2 - m_1'\mathrm{Tr}B_1 - m_2'\mathrm{Tr}B_2.$$

By further using the identity $zG^2 = HG^2 - G$, it is not difficult to check

$$\begin{aligned}\mathrm{Tr}\Xi_1' B = &-\frac{m_1}{2}\mathrm{Tr}HG^2B_1 + \frac{F_1}{2}\mathrm{Tr}G^2B_1 + \frac{1}{2}(m_1 - \frac{1}{z})\mathrm{Tr}GB_1 - m_1'\mathrm{Tr}B_1\\ &-\frac{m_2}{2}\mathrm{Tr}HG^2B_2 + \frac{F_2}{2}\mathrm{Tr}G^2B_2 + \frac{1}{2}(m_2 - \frac{1}{z})\mathrm{Tr}GB_2 - m_2'\mathrm{Tr}B_2.\end{aligned} \tag{S21}$$

Recall the definition (5.1). Putting (S20) and (S21) together, we get

$$\begin{aligned}\mathcal{Q} = \sqrt{n}\Big(&- m_1\mathrm{Tr}HGA_1 + F_1\mathrm{Tr}GA_1 - m_2\mathrm{Tr}HGA_2 + F_2\mathrm{Tr}GA_2\\ &-\frac{m_1}{2}\mathrm{Tr}HG^2B_1 + \frac{F_1}{2}\mathrm{Tr}G^2B_1 + \frac{1}{2}(m_1 - \frac{1}{z})\mathrm{Tr}GB_1 - m_1'\mathrm{Tr}B_1\\ &-\frac{m_2}{2}\mathrm{Tr}HG^2B_2 + \frac{F_2}{2}\mathrm{Tr}G^2B_2 + \frac{1}{2}(m_2 - \frac{1}{z})\mathrm{Tr}GB_2 - m_2'\mathrm{Tr}B_2\Big).\end{aligned} \tag{S22}$$

Recall the definition of Δ_r from (5.6). We write

$$\Delta_r = \sqrt{nz}\sum_{i,j} x_{ij}c_{ij} - \sqrt{nz}\sum_{(i,j)\in\mathcal{S}(\nu)} x_{ij}c_{ij}.$$

Further recall the definition of c_{ij} from (5.7). It is elementary to check that

$$\sqrt{nz}\sum_{i,j} x_{ij}c_{ij} = -\sqrt{n}\Big(m_1\mathrm{Tr}H\Pi_1A_1 + m_2\mathrm{Tr}H\Pi_1A_2 + \frac{m_1}{2}\mathrm{Tr}H\Pi_2B_1$$
$$+\frac{m_2}{2}\mathrm{Tr}H\Pi_2B_2 + m_1'\mathrm{Tr}H\Pi_1B_1 + m_2'\mathrm{Tr}H\Pi_1B_2\Big). \tag{S23}$$

Using (S22) and (S23), with the notations defined in (S3), we can write

$$\mathcal{Q} - \sqrt{nz}\sum_{i,j} x_{ij}c_{ij} = \sqrt{n}\big(f_1 + f_2 + g_1 + g_2\big). \tag{S24}$$

Combining (5.6), (S17) and (S24) we can conclude the proof. □

F. Proof of Lemma E.2. To prove Lemma E.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 (5.12) that

$$Q = \sqrt{n}\Big(\mathrm{Tr}\big(\Xi_1 A\big) + \mathrm{Tr}\big(\Xi_1' B\big)\Big) - \sqrt{nz}\sum_{(i,j)\in\mathcal{B}(\nu)} x_{ij}c_{ij} - \Delta_d,$$

where $\Xi_1 = G - \Pi_1$ and Δ_d is a deterministic quantity in (5.5). Using $G' = \frac{1}{2}(G^2 - z^{-1}G)$ in Lemma 4.7, we find that

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

By Lemma C.2, it can be further seen that

$$\frac{\partial Q}{\partial x_{ij}} = -\sqrt{nz}\sum_{\substack{l_1,l_2\in\{i,j'\}\\ l_1\neq l_2}}\Big((GAG)_{l_1l_2} - \frac{1}{2z}(GBG)_{l_1l_2} + \frac{1}{2}(GBG^2)_{l_1l_2} + \frac{1}{2}(G^2BG)_{l_1l_2}\Big)$$
$$- \mathbf{1}\Big((i,j)\in\mathcal{B}(\nu)\Big)\sqrt{nz}c_{ij}. \tag{S1}$$

LEMMA F.1. *Under the assumptions of Proposition 5.1, we have*

$$\frac{\partial Q}{\partial x_{ij}} = \sqrt{nz}\mathbf{1}\Big((i,j)\in\mathcal{S}(\nu)\Big)c_{ij} + O_{\prec}(1). \tag{S2}$$

Consequently, we have the bounds

$$\frac{\partial Q}{\partial x_{ij}} = \begin{cases} O_{\prec}(1), & \forall\ (i,j)\in\mathcal{B}(\nu) \\ O_{\prec}(n^{\frac{1}{2}-\nu}), & \forall (i,j)\in\mathcal{S}(\nu). \end{cases} \tag{S3}$$

PROOF OF LEMMA F.1. First, recall the definitions in (4.19) and (4.18). By (4.17), we have that for $a_1, a_2 = 1, 2$,

$$(G^{a_1}AG^{a_2})_{l_1l_2} = (\Pi_{a_1}A\Pi_{a_2})_{l_1l_2} + O_{\prec}(n^{-\frac{1}{2}}).$$

Applying the above estimates to (S1), we find that

$$\begin{aligned}\frac{\partial Q}{\partial x_{ij}} = -\sqrt{zn}\sum_{\substack{l_1,l_2\in\{i,j'\}\\ l_1\neq l_2}}\Big((\Pi_1A\Pi_1)_{l_1l_2} - \frac{1}{2z}(\Pi_1B\Pi_1)_{l_1l_2} + \frac{1}{2}(\Pi_1B\Pi_2)_{l_1l_2} + \frac{1}{2}(\Pi_2B\Pi_1)_{l_1l_2}\Big) \\ - \mathbf{1}\Big((i,j)\in\mathcal{B}(\nu)\Big)\sqrt{nz}c_{ij} + O_{\prec}(1).\end{aligned} \tag{S4}$$

Comparing (S4) with the definition of c_{ij} in (5.7), we prove (S2) and the first case of (S3).

Next, by the definitions of A, B in (4.26) and the set $\mathcal{S}(\nu)$ in (5.3), it follows immediately that there exists some constant $C > 0$, such that

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

By the estimates in (4.15), we get the second case of (S3). This concludes the proof of Lemma F.1. □

The remaining of the section is devoted to the proof of Lemma E.2.

PROOF OF LEMMA E.2. We will focus on the proof of (S13). Since the proof of (S14) is analogous, we shall only outline the main steps. Recall from the definition in (S3) and (S18) that

$$\sqrt{n}\mathbb{E}f_1\mathfrak{q}^{(k-1)} = \mathbb{E}\Big(-m_1\sqrt{zn}\sum_{i,j}x_{ij}(\Xi_1A_1)_{j'i} + \sqrt{n}F_1\mathrm{Tr}GA_1\Big)\mathfrak{q}^{(k-1)}. \tag{S5}$$

For brevity, we use the notations

$$h_1 = \big(\Xi_1 A_1\big)_{j'i}, \quad h_2 = Q^{k-1}, \quad h_3 = e^{\mathrm{i}t\Delta}. \tag{S6}$$

Note that h_1 actually depends on the index (j', i). However, we drop this dependence from notation for brevity. By Lemma C.1, one has

$$\sqrt{n}\sum_{i,j}\mathbb{E}x_{ij}(\Xi_1 A_1)_{j'i}\mathfrak{q}^{(k-1)} = \sqrt{n}\sum_{i,j}\mathbb{E}x_{ij}(h_1h_2h_3)$$
$$= \sum_{l=1}^{3}\frac{\kappa_{l+1}}{l!n^{l/2}}\sum_{i,j}\mathbb{E}\Big(\frac{\partial^l}{\partial x_{ij}^l}(h_1h_2h_3)\Big) + \mathbb{E}\mathcal{R}_1, \tag{S7}$$

where $\mathcal{R}_1$ satisfies that, for any sufficiently small $\epsilon > 0$ and sufficiently large $K > 0$,

$$|\mathbb{E}\mathcal{R}_1| \le \sum_{i,j}\mathbb{E}\left(n^{-2}\sup_{|x_{ij}|\le n^{-\frac{1}{2}+\epsilon}}\left|\frac{\partial^4}{\partial x_{ij}^4}(h_1h_2h_3)\right| + n^{-K}\sup_{x_{ij}\in\mathbb{R}}\left|\frac{\partial^4}{\partial x_{ij}^4}(h_1h_2h_3)\right|\right). \tag{S8}$$

Here we used the assumption that $\mathbb{E}|\sqrt{n}x_{ij}|^p \le C_p$ for all $p \ge 3$. Therefore, the main technical estimates are the first four derivatives of $h_1h_2h_3$. By product rule, for each $l \in \mathbb{N}$, we have

$$\frac{\partial^l}{\partial x_{ij}^l}(h_1h_2h_3) = \sum_{l_1+l_2+l_3=l}\binom{l}{l_1, l_2, l_3}\frac{\partial^{l_1}h_1}{\partial x_{ij}^{l_1}}\frac{\partial^{l_2}h_2}{\partial x_{ij}^{l_2}}\frac{\partial^{l_3}h_3}{\partial x_{ij}^{l_3}}. \tag{S9}$$

First, it is elementary to verify

$$\frac{\partial^l h_3}{\partial x_{ij}^l} = \mathbf{1}((i,j)\in\mathcal{B}(\nu))\left(\mathrm{i}t\sqrt{nz}c_{ij}\right)^l e^{\mathrm{i}t\Delta}, \tag{S10}$$

and

$$\frac{\partial^l h_1}{\partial x_{ij}^l} = \Big(\frac{\partial^l G}{\partial x_{ij}^l}A_1\Big)_{j'i}.$$

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

$$\frac{\partial^2 h_2}{\partial x_{ij}^2} = \frac{(k-1)!}{(k-3)!}Q^{k-3}\Big(\frac{\partial Q}{\partial x_{ij}}\Big)^2 + (k-1)Q^{k-2}\frac{\partial^2 Q}{\partial x_{ij}^2}.$$

The third derivative of h_2 is

$$\frac{\partial^3 h_2}{\partial x_{ij}^3} = \frac{(k-1)!}{(k-4)!}Q^{k-4}\Big(\frac{\partial Q}{\partial x_{ij}}\Big)^3 + 3\frac{(k-1)!}{(k-3)!}Q^{k-3}\frac{\partial Q}{\partial x_{ij}}\frac{\partial^2 Q}{\partial x_{ij}^2} + (k-1)Q^{k-2}\frac{\partial^3 Q}{\partial x_{ij}^3}.$$

The fourth derivative of h_2 is

$$\begin{aligned}\frac{\partial^4 h_2}{\partial x_{ij}^4} &= \frac{(k-1)!}{(k-5)!}Q^{k-5}\Big(\frac{\partial Q}{\partial x_{ij}}\Big)^4 + 6\frac{(k-1)!}{(k-4)!}Q^{k-4}\Big(\frac{\partial Q}{\partial x_{ij}}\Big)^2\frac{\partial^2 Q}{\partial x_{ij}^2} \\ &+ \frac{(k-1)!}{(k-3)!}Q^{k-3}\Big(4\frac{\partial Q}{\partial x_{ij}}\frac{\partial^3 Q}{\partial x_{ij}^3} + 3\Big(\frac{\partial^2 Q}{\partial x_{ij}^2}\Big)^2\Big) + (k-1)Q^{k-2}\frac{\partial Q^4}{\partial x_{ij}^4}.\end{aligned}$$

As we can see from the above identities, the key ingredients are the partial derivatives of Q and GA_1.

For brevity, we introduce the notation

$$\hbar(l_1,l_2,l_3) := n^{-\frac{l_1+l_2+l_3}{2}}\sum_{i,j}\frac{\partial^{l_1}h_1}{\partial x_{ij}^{l_1}}\frac{\partial^{l_2}h_2}{\partial x_{ij}^{l_2}}\frac{\partial^{l_3}h_3}{\partial x_{ij}^{l_3}}. \tag{S11}$$

In the following two lemmas, we summarize the estimates of $\hbar(l_1,l_2,l_3)$ for $l_1+l_2+l_3 \le 4$. The proofs of the two lemmas will be given in Sections F.1 and F.2.

Lemma F.2. *For the first derivative of $h_1h_2h_3$, we have that*

$$\hbar(1,0,0) = -\sqrt{nz}m_2\mathrm{Tr}(GA_1)\mathfrak{q}^{(k-1)} + O_\prec(n^{-\frac12}), \tag{S12}$$

$$\hbar(0,1,0) = \mathfrak{a}_{11}\mathfrak{q}^{(k-2)} + O_\prec(n^{-\frac12+4\nu}), \tag{S13}$$

$$\hbar(0,0,1) = O_\prec(n^{-\frac12+4\nu}). \tag{S14}$$

Lemma F.3. *On higher order derivatives of $h_1h_2h_3$, we have the following estimates.*
(1). For the second derivative, we have

$$\hbar(2,0,0) = \mathfrak{d}_1^a\mathfrak{q}^{(k-1)} + O_\prec(n^{-\frac12}), \tag{S15}$$

$$\begin{aligned}&\hbar(1,1,0) = \mathfrak{a}_{21}\mathfrak{q}^{(k-2)} + O_\prec(n^{-\frac12+4\nu}), \qquad\qquad \text{(S16)}\\ &\hbar(1,0,1) = O_\prec(n^{-\frac12+4\nu}), \quad \hbar(0,2,0) = O_\prec(n^{-\frac12}),\\ &\hbar(0,1,1) = O_\prec(n^{-\frac12+4\nu}), \quad \hbar(0,0,2) = O_\prec(n^{-1+4\nu}).\end{aligned}$$

(2). For the third derivative, we have

$$\hbar(1,2,0) = \mathfrak{a}_{31}\mathfrak{q}^{(k-2)} + O_\prec(n^{-\frac12}), \tag{S17}$$

$$\hbar(3,0,0) = O_\prec(n^{-\frac{1}{2}}), \qquad \hbar(0,3,0) = O_\prec(n^{-\frac{1}{2}}),$$
$$\hbar(2,1,0) = O_\prec(n^{-1}), \qquad \hbar(2,0,1) = O_\prec(n^{-\frac{3}{2}+4\nu}),$$
$$\hbar(1,1,1) = O_\prec(n^{-1+4\nu}), \qquad \hbar(1,0,2) = O_\prec(n^{-\frac{3}{2}+4\nu}),$$
$$\hbar(0,2,1) = O_\prec(n^{-\frac{1}{2}+4\nu}).$$

(3). For the fourth derivative, all the terms in the RHS of (S9) can be bounded by $O_\prec(n^{-\frac{1}{2}+4\nu})$.

By Lemma F.2 and Lemma F.3, the first term in (S7) is estimated by

$$\sum_{l=1}^{3} \frac{\kappa_{l+1}}{l! n^{l/2}} \sum_{i,j} \mathbb{E}\Big(\frac{\partial^l}{\partial x_{ij}^l} (h_1 h_2 h_3) \Big) = \sum_{l=1}^{3} \sum_{i,j} \sum_{l_1+l_2+l_3=l} \frac{\kappa_{l+1}}{l_1! l_2! l_3!} \hbar(l_1, l_2, l_3)$$
$$= -\sqrt{nz} m_2 \mathrm{Tr}(GA_1) \mathfrak{q}^{(k-1)} + \frac{\kappa_3}{2} \mathfrak{d}_1^a \mathfrak{q}^{(k-1)} + \Big(\mathfrak{a}_{11} + \kappa_3 \mathfrak{a}_{21} + \frac{\kappa_4}{2} \mathfrak{a}_{31} \Big) \mathfrak{q}^{(k-2)}.$$

For the second term in (S7), we claim that

$$|\mathbb{E}\mathcal{R}_1| \le n^{-1/2+4\nu}. \tag{S18}$$

To prove (S18), it is enough to bound the two terms on the right hand side of (S8). We apply Lemma F.3 to the first term on the right hand side of (S8) to get

$$\sum_{i,j} \mathbb{E} n^{-2} \sup_{|x_{ij}| \le n^{-\frac{1}{2}+\epsilon}} \left| \frac{\partial^4}{\partial x_{ij}^4} (h_1 h_2 h_3) \right| \le n^{-1/2+4\nu}.$$

A minor issue with the above step is that Lemma F.3 is proved for the matrix X with all entries random variables. In our application of Lemma F.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 F.3 holds for such random matrix X.

For the second term on the right hand side of (S8), we use the trivial bounds for G and its derivatives to obtain

$$\sum_{i,j} \mathbb{E} n^{-K} \sup_{x_{ij} \in \mathbb{R}} \left| \frac{\partial^4}{\partial x_{ij}^4} (h_1 h_2 h_3) \right| \le n^{-K+2+C}$$

for a positive constant C. By taking K sufficiently large, we conclude (S18).

Plugging (S7) into (S5), we finally get

$$\sqrt{n}\mathbb{E}f_1\mathfrak{q}^{(k-1)} = -m_1\sqrt{z}\mathbb{E}\Big(\frac{\kappa_3}{2}\mathfrak{d}_1^a\mathfrak{q}^{(k-1)} + \Big(\mathfrak{a}_{11} + \kappa_3\mathfrak{a}_{21} + \frac{\kappa_4}{2}\mathfrak{a}_{31}\Big)\mathfrak{q}^{(k-2)}\Big) + O_\prec(n^{-\frac{1}{2}+4\nu}).$$

Note that by (S19), the term $\sqrt{n}zm_1m_2\mathrm{Tr}GA_1\mathfrak{q}^{(k-1)}$ is cancelled with $F_1\mathrm{Tr}GA_1\mathfrak{q}^{(k-1)}$ in (S5). This verifies (S13) in case of $\alpha = 1$ by recalling the definition in (S12).

Next, we turn to (S14) for $\alpha = 1$. Recall the definition of g_1 in (S3). We have

$$\sqrt{n}\mathbb{E}g_1\mathfrak{q}^{(k-1)} = \sqrt{n}\mathbb{E}\Big(-\frac{m_1}{2}\sqrt{z}\sum_{i,j}x_{ij}(\Xi_2B_1)_{j'i} + \frac{F_1}{2}\mathrm{Tr}G^2B_1 + \frac{zm_1-1}{2z}\mathrm{Tr}GB_1 - m_1'\mathrm{Tr}B_1 + m_1'\mathrm{Tr}H\Pi_1B_1\Big)\mathfrak{q}^{(k-1)}. \tag{S19}$$

The main task is to estimate the cumulant expansion of the term

$$\sqrt{n}\sum_{i,j}\mathbb{E}x_{ij}(\Xi_2B_1)_{j'i}\mathfrak{q}^{(k-1)},$$

which is analogous to (S7). Recall h_2 and h_3 in (S6) and denote

$$\tilde{h}_1 = \big(\Xi_2B_1\big)_{j'i}. \tag{S20}$$

Note that $\tilde{h}_1$ depends on the indices i, j. However, we drop these dependence from the notation for brevity. Similarly to (S11), we introduce the notation

$$\tilde{\hbar}(l_1,l_2,l_3) := n^{-\frac{l_1+l_2+l_3}{2}}\sum_{i,j}\frac{\partial^{l_1}\tilde{h}_1}{\partial x_{ij}^{l_1}}\frac{\partial^{l_2}h_2}{\partial x_{ij}^{l_2}}\frac{\partial^{l_3}h_3}{\partial x_{ij}^{l_3}}. \tag{S21}$$

We collect the estimates of $\tilde{\hbar}(l_1,l_2,l_3)$ for $l_1+l_2+l_3 \le 4$ in the following two lemmas, whose proofs are postponed to Section F.3.

LEMMA F.4. *For the first derivative of $\tilde{h}_1h_2h_3$, we have*

$$\tilde{\hbar}(1,0,0) = -\sqrt{nz}\big((2m_2' + \frac{m_2}{z})\,\mathrm{Tr}(GB_1) + m_2\,\mathrm{Tr}(G^2B_1)\big)\mathfrak{q}^{(k-1)} + O_\prec(n^{-\frac{1}{2}}),$$
$$\tilde{\hbar}(0,1,0) = \tilde{\mathfrak{b}}_{11}\mathfrak{q}^{(k-2)} + O_\prec(n^{-\frac{1}{2}+4\nu}),$$
$$\tilde{\hbar}(0,0,1) = O_\prec(n^{-\frac{1}{2}+4\nu}).$$

Lemma F.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

$$\begin{aligned}
\tilde{\hbar}(2,0,0) &= \tilde{\mathfrak{d}}_1 \mathfrak{q}^{(k-1)} + O_\prec(n^{-\frac{1}{2}}), \\
\tilde{\hbar}(1,1,0) &= \tilde{\mathfrak{b}}_{21} \mathfrak{q}^{(k-2)} + O_\prec(n^{-\frac{1}{2}}), \\
\tilde{\hbar}(0,2,0) &= O_\prec(n^{-\frac{1}{2}}).
\end{aligned}$$

All the other terms with $l_3 \geq 1$ can be bounded by $O_\prec(n^{-\frac{1}{2}+4\nu})$.
(2). For the third derivative, we have

$$\begin{aligned}
&\tilde{\hbar}(1,2,0) = \tilde{\mathfrak{b}}_{31} \mathfrak{q}^{(k-2)} + O_\prec(n^{-\frac{1}{2}}), \\
&\tilde{\hbar}(3,0,0) = O_\prec(n^{-\frac{1}{2}}), \quad \tilde{\hbar}(0,3,0) = O_\prec(n^{-\frac{1}{2}}), \\
&\tilde{\hbar}(2,1,0) = O_\prec(n^{-1}).
\end{aligned}$$

All the other terms with $l_3 \geq 1$ can be bounded by $O_\prec(n^{-\frac{1}{2}+4\nu})$.
(3). For the fourth derivative, all the terms can be bounded by $O_\prec(n^{-\frac{1}{2}+4\nu})$.

With these preparations, using arguments similar to those of (S5), we find that

$$\begin{aligned}
\sqrt{n}\mathbb{E} g_1 \mathfrak{q}^{(k-1)} = &-\frac{m_1\sqrt{z}}{2}\mathbb{E}\Big(\frac{\kappa_3}{2}\tilde{\mathfrak{d}}_1\mathfrak{q}^{(k-1)} + \Big(\tilde{\mathfrak{b}}_{11} + \kappa_3\tilde{\mathfrak{b}}_{21} + \frac{\kappa_4}{2}\tilde{\mathfrak{b}}_{31}\Big)\mathfrak{q}^{(k-2)}\Big) \\
&+ \sqrt{n}\mathbb{E}\Big(\frac{m_1'}{m_1}\mathrm{Tr} G B_1 - m_1'\mathrm{Tr} B_1 + m_1'\mathrm{Tr} H\Pi_1 B_1\Big)\mathfrak{q}^{(k-1)} + O_\prec(n^{-\frac{1}{2}+4\nu}).
\end{aligned}$$

In the above, we use (S19) and an identity

$$\frac{m_1}{2}\Big(z m_1 (2m_2' + \frac{m_2}{z}) + m_1 - \frac{1}{z}\Big) = m_1', \tag{S22}$$

which can be checked from (4.7) and (4.8). Next, observe that

$$\begin{aligned}
&\sqrt{n}\mathbb{E}\Big(\frac{m_1'}{m_1}\mathrm{Tr} G B_1 + m_1'\mathrm{Tr} H\Pi_1 B_1 - m_1'\mathrm{Tr} B_1\Big)\mathfrak{q}^{(k-1)} \\
&= \sqrt{n}\mathbb{E}\Big(-z m_1'\mathrm{Tr} G B_1 + \frac{m_1' F_1}{m_1}\mathrm{Tr} G B_1 + m_1'\mathrm{Tr} H\Pi_1 B_1 - m_1'\mathrm{Tr} B_1\Big)\mathfrak{q}^{(k-1)}, \\
&= \frac{m_1'}{m_1}\sqrt{n}\mathbb{E}\Big(-m_1\mathrm{Tr} H\Xi_1 B_1 + F_1\mathrm{Tr} G B_1\Big)\mathfrak{q}^{(k-1)}.
\end{aligned}$$

In the first step above, we simply use the definition of F_1 in (S18). 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

$$\sqrt{n}\mathbb{E}g_1\mathfrak{q}^{(k-1)} = -\sqrt{z}\mathbb{E}\Big(\frac{m_1\kappa_3}{4}\Big(\tilde{\mathfrak{d}}_1 + 2\frac{m_1'}{m_1}\mathfrak{d}_1^b\Big)\mathfrak{q}^{(k-1)} + \Big(\frac{m_1}{2}\tilde{\mathfrak{b}}_{11} + m_1'\mathfrak{b}_{11}\Big)\mathfrak{q}^{(k-2)} + \Big(\frac{m_1\kappa_3}{2}\tilde{\mathfrak{b}}_{21} + \kappa_3 m_1'\mathfrak{b}_{21} + \frac{m_1\kappa_4}{4}\tilde{\mathfrak{b}}_{31} + \frac{\kappa_4 m_1'}{2}\mathfrak{b}_{31}\Big)\mathfrak{q}^{(k-2)}\Big) + O_\prec(n^{-\frac{1}{2}+4\nu}).$$

This verifies (S14) in case of $\alpha = 1$ by recalling the definition in (S12).

The proofs of (S13) and (S14) in case of $\alpha = 2$ are analogous to those of (S5) and (S19). We outline the main steps. First observe that

$$\sqrt{n}\mathbb{E}f_2\mathfrak{q}^{(k-1)} = \mathbb{E}\Big(- m_2\sqrt{nz}\sum_{i,j} x_{ij}(\Xi_1 A_2)_{ij'} + F_2 \mathrm{Tr} G A_2\Big)\mathfrak{q}^{(k-1)},$$

$$\sqrt{n}\mathbb{E}g_2\mathfrak{q}^{(k-1)} = \mathbb{E}\Big(- \frac{m_2}{2}\sqrt{nz}\sum_{i,j} x_{ij}(\Xi_2 B_2)_{ij'} + \frac{F_2}{2}\mathrm{Tr}G^2 B_2 + \frac{zm_2 - 1}{2z}\mathrm{Tr}GB_2 - m_1'\mathrm{Tr}B_2 + m_1'\mathrm{Tr}H\Pi_1 B_2\Big)\mathfrak{q}^{(k-1)}.$$

Recall h_2 and h_3 in (S6) and denote

$$\mathbb{h}_1 = (\Xi_1 A_2)_{ij'}, \quad \tilde{\mathbb{h}}_1 = (\Xi_2 B_2)_{ij'}.$$

Analogously to (S11) and (S21), we introduce the notations

$$\mathfrak{h}(l_1, l_2, l_3) := n^{-\frac{l_1+l_2+l_3}{2}} \sum_{i,j} \frac{\partial^{l_1}\mathbb{h}_1}{\partial x_{ij}^{l_1}} \frac{\partial^{l_2} h_2}{\partial x_{ij}^{l_2}} \frac{\partial^{l_3} h_3}{\partial x_{ij}^{l_3}},$$

and $\tilde{\mathfrak{h}}(l_1, l_2, l_3)$ which is defined via replacing $\mathbb{h}_1$ by $\tilde{\mathbb{h}}_1$ in the above definition.

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

Lemma F.6. *For $\mathfrak{h}$, we have*

$$\begin{aligned} &\mathfrak{h}(1,0,0) = -\sqrt{nz}ym_1(\mathrm{Tr}\, GA_2)\mathfrak{q}^{(k-1)} + O_\prec(n^{-\frac{1}{2}}), \\ &\mathfrak{h}(0,1,0) = \mathfrak{a}_{12}\mathfrak{q}^{(k-2)} + O_\prec(n^{-\frac{1}{2}+4\nu}), \\ &\mathfrak{h}(0,0,1) = O_\prec(n^{-\frac{1}{2}+4\nu}). \end{aligned} \tag{S23}$$

Similarly, for $\tilde{\mathfrak{h}}$, we have

$$\tilde{\mathfrak{h}}(1,0,0) = -\sqrt{nz}y\Big((2m_1' + \frac{m_1}{z})\mathrm{Tr}GB_2 + m_1\,\mathrm{Tr}\,G^2B_2\Big)\mathfrak{q}^{(k-1)} + O_\prec(n^{-\frac{1}{2}}),$$

$$\tilde{\mathfrak{h}}(0,1,0) = \tilde{\mathfrak{b}}_{12}\mathfrak{q}^{(k-2)} + O_\prec(n^{-\frac{1}{2}+4\nu}),$$

$$\tilde{\mathfrak{h}}(0,0,1) = O_\prec(n^{-\frac{1}{2}+4\nu}).$$

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

LEMMA F.7. *We have the following estimates in case $l_1+l_2+l_3 \geq 2$.*
(1). For $\mathfrak{h}(l_1,l_2,l_3)$, we have

$$\mathfrak{h}(2,0,0) = \mathfrak{d}_2^a \mathfrak{q}^{(k-1)} + O_\prec(n^{-\frac{1}{2}}),$$
$$\mathfrak{h}(1,1,0) = \mathfrak{a}_{22} \mathfrak{q}^{(k-2)} + O_\prec(n^{-\frac{1}{2}+4\nu}),$$
$$\mathfrak{h}(1,2,0) = \mathfrak{a}_{32} \mathfrak{q}^{(k-2)} + O_\prec(n^{-\frac{1}{2}}).$$

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 $\tilde{\mathfrak{h}}(l_1,l_2,l_3)$ we have

$$\tilde{\mathfrak{h}}(2,0,0) = \tilde{\mathfrak{d}}_2 \mathfrak{q}^{(k-1)} + O_\prec(n^{-\frac{1}{2}}),$$
$$\tilde{\mathfrak{h}}(1,1,0) = \tilde{\mathfrak{b}}_{22} \mathfrak{q}^{(k-2)} + O_\prec(n^{-\frac{1}{2}}),$$
$$\tilde{\mathfrak{h}}(1,2,0) = \tilde{\mathfrak{b}}_{32} \mathfrak{q}^{(k-2)} + O_\prec(n^{-\frac{1}{2}}).$$

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 F.3. The remaining estimates for $\sqrt{n}\mathbb{E} f_2 \mathfrak{q}^{(k-1)}$ and $\sqrt{n}\mathbb{E} g_2 \mathfrak{q}^{(k-1)}$ follow the same arguments as those of (S5) and (S19), and are therefore omitted. As a side note, we mention an identity (comparable to (S22))

$$\frac{m_2}{2}\Big(zym_2(2m_1' + \frac{m_1}{z}) + m_2 - \frac{1}{z}\Big) = m_2'$$

used in the derivation of the g_2 term.

Lastly, we prove (S15). Recall $h_2 = Q^{k-1}$ and $h_3 = e^{\mathrm{i}t\Delta}$. By Lemma C.1, we have

(S24)
$$\sqrt{nz} \sum_{(i,j)\in\mathcal{S}(\nu)} c_{ij}\mathbb{E} x_{ij}\mathfrak{q}^{(k-1)} = \sqrt{z} \sum_{(i,j)\in\mathcal{S}(\nu)} c_{ij}\mathbb{E}\Big(\frac{1}{\sqrt{n}}\frac{\partial(h_2h_3)}{\partial x_{ij}} + \frac{\kappa_3}{2n}\frac{\partial^2(h_2h_3)}{\partial x_{ij}^2}\Big) + \mathbb{E}\mathcal{R},$$

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

$$|\mathbb{E}\mathcal{R}| \leq \sum_{i,j}\mathbb{E}\Big(n^{-\frac{3}{2}} \sup_{|x_{ij}|\leq n^{-\frac{3}{2}+\epsilon}}\Big|c_{ij}\frac{\partial^3(h_2h_3)}{\partial x_{ij}^3}\Big| + n^{-K}\sup_{|x_{ij}|\in\mathbb{R}}\Big|c_{ij}\frac{\partial^3(h_2h_3)}{\partial x_{ij}^3}\Big|\Big).$$

We first show that

$$|\mathbb{E}\mathcal{R}| = O_{\prec}(n^{-\frac{1}{2}+4\nu}). \tag{S25}$$

Similar to the discussion of (S18), the proof boils down to estimate the third order derivative of h_2h_3. Using the same proof as (S14) in Lemma F.3 (given in Section F.1), we observe that in the derivatives of h_2h_3, any term containing the derivatives of h_3 can bounded by $O_{\prec}(n^{-\frac{1}{2}+4\nu})$. Thus, by product rule,

$$\frac{\partial^3(h_2h_3)}{\partial x_{ij}^3} = \frac{\partial^3 h_2}{\partial x_{ij}^3}h_3 + O_{\prec}(n^{-\frac{1}{2}+4\nu}) = O_{\prec}\Big(\boldsymbol{u}(i)\boldsymbol{v}(j) + n^{-\frac{1}{2}+4\nu}\Big).$$

The last step is obtained analogously to (S17). We omit the details. To conclude (S25), we also use $c_{ij} = O_{\prec}(\boldsymbol{u}(i)\boldsymbol{v}(j))$ by recalling its definition (5.7) and the fact that $\boldsymbol{u}$, $\boldsymbol{v}$ are both unit vectors.

Next, using arguments similar to (S16) and (S17), we get

$$\frac{1}{\sqrt{n}}\frac{\partial(h_2h_3)}{\partial x_{ij}} = \frac{1}{\sqrt{n}}\frac{\partial h_2}{\partial x_{ij}}h_3 + O_{\prec}(n^{-\frac{1}{2}+4\nu}) = (k-1)\sqrt{z}c_{ij}\mathfrak{q}^{(k-2)} + O_{\prec}(n^{-\frac{1}{2}+4\nu}), \tag{S26}$$

and

$$\frac{1}{n}\frac{\partial^2(h_2h_3)}{\partial x_{ij}^2} = \frac{1}{n}\frac{\partial^2 h_2}{\partial x_{ij}^2}h_3 + O_{\prec}(n^{-\frac{1}{2}+4\nu}) = 2\frac{(k-1)z}{\sqrt{n}}s_{ij}\mathfrak{q}^{(k-2)} + O_{\prec}(n^{-\frac{1}{2}+4\nu}). \tag{S27}$$

Plugging (S25)-(S27) into (S24), we obtain (S15). The proof of Lemma E.2 is now complete. ☐

F.1. *Proof of Lemma F.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

$$\Big(\frac{\partial G}{\partial x_{ij}}W\Big)_{ab} = -\sqrt{z}\Big(G_{aj'}(GW)_{ib} + G_{ai}(GW)_{j'b}\Big). \tag{S28}$$

We emphasize that both (4.15) and a basic fact (as a consequence of (S1))

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

will be applied to bound the error terms throughout the proofs of Lemma F.2-Lemma F.7.

For convenience, we denote the blocks of A and B (c.f. (4.26)) by $\mathcal{A}_k$'s and $\mathcal{B}_k$'s, i.e.,

$$A = \begin{pmatrix} \omega_1 \boldsymbol{u}\boldsymbol{u}^* & \omega_2 \boldsymbol{u}\boldsymbol{v}^* \\ \omega_3 \boldsymbol{v}\boldsymbol{u}^* & \omega_4 \boldsymbol{v}\boldsymbol{v}^* \end{pmatrix} := \begin{pmatrix} \mathcal{A}_1 & \mathcal{A}_2 \\ \mathcal{A}_3 & \mathcal{A}_4 \end{pmatrix}, \quad B = \begin{pmatrix} \varpi_1 \boldsymbol{u}\boldsymbol{u}^* & \varpi_2 \boldsymbol{u}\boldsymbol{v}^* \\ \varpi_3 \boldsymbol{v}\boldsymbol{u}^* & \varpi_4 \boldsymbol{v}\boldsymbol{v}^* \end{pmatrix} := \begin{pmatrix} \mathcal{B}_1 & \mathcal{B}_2 \\ \mathcal{B}_3 & \mathcal{B}_4 \end{pmatrix}. \tag{S29}$$

With the above preparation, we now prove Lemma F.2.

Proof of Lemma F.2. First, by recalling the notations in (S6) and (S11), and using (S28), we have

$$\begin{aligned} \hbar(1,0,0) &= \frac{1}{\sqrt{n}} \sum_{i,j} \Big(\frac{\partial G}{\partial x_{ij}} A_1 \Big)_{j'i} \mathfrak{q}^{(k-1)} \\ &= -\sqrt{nz} \frac{1}{n} \sum_{i,j} \Big(G_{j'j'} (GA_1)_{ii} + G_{j'i} (GA_1)_{j'i} \Big) \mathfrak{q}^{(k-1)}. \end{aligned}$$

Moreover, by (4.6) and (4.13), we further get

$$\begin{aligned} \hbar(1,0,0) &= -\sqrt{nz} m_{2n} (\mathrm{Tr} G A_1) \mathfrak{q}^{(k-1)} - \sqrt{\frac{z}{n}} (\mathrm{Tr} G A_1 G) \mathfrak{q}^{(k-1)} \\ &= -\sqrt{nz} m_2 (\mathrm{Tr} G A_1) \mathfrak{q}^{(k-1)} + O_{\prec}(n^{-\frac{1}{2}}), \end{aligned} \tag{S30}$$

where the last step follows from the property of trace and (4.20).

Next, using the fact $|\mathcal{B}(\nu)| \le C n^{4\nu}$ together with the definition of c_{ij} in (5.7) and (4.17), we obtain

$$\hbar(0,0,1) = \sqrt{z} \sum_{(i,j) \in \mathcal{B}(\nu)} \mathrm{i} t c_{ij} (\Xi_1)_{j'i} \mathfrak{q}^{(k-1)} = O_{\prec}(n^{-\frac{1}{2} + 4\nu}). \tag{S31}$$

The main task is the estimate of

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

In light of the expression of $\partial Q / \partial x_{ij}$ in (S1), by symmetry, we get

$$\begin{aligned} \hbar(0,1,0) = &- (k-1)\sqrt{z} \sum_{i,j} (\Xi_1 A_1)_{j'i} \big[2(GAG)_{j'i} - \frac{1}{z} (GBG)_{j'i} + (GBG^2)_{j'i} + (G^2 BG)_{j'i} \big] \mathfrak{q}^{(k-2)} \\ &- (k-1)\sqrt{z} \sum_{(i,j) \in \mathcal{B}(\nu)} (\Xi_1 A_1)_{j'i} c_{ij} \mathfrak{q}^{(k-2)}. \end{aligned} \tag{S32}$$

The last term on the right hand side of (S32) is bounded by $O_\prec(n^{-\frac{1}{2}+4\nu})$, by exactly the same estimate of (S31). Now we turn towards the first term on the right hand side of (S32). We first claim that

$$\sum_{i,j}(\Xi_1 A_1)_{j'i}(GAG)_{j'i} = \mathrm{Tr}(\Pi_2^{\mathrm{l}} - \Pi_{1,1}^{\mathrm{l}})A_1\Pi_1 A + O_\prec(n^{-\frac{1}{2}}). \tag{S33}$$

To derive the above statement, a key observation is that the summation on the left hand side of (S33) can be written in terms of a trace, with the aid of the block diagonal matrices $\mathtt{I}^{\mathrm{u}}$ and $\mathtt{I}^{\mathrm{l}}$ in (S2). Indeed, we find

$$\sum_{i,j}(\Xi_1 A_1)_{j'i}(GAG)_{j'i} = \mathrm{Tr}(\mathtt{I}^{\mathrm{l}}\Xi_1 A_1 \mathtt{I}^{\mathrm{u}} GAG\mathtt{I}^{\mathrm{l}}) = \mathrm{Tr}(G\mathtt{I}^{\mathrm{l}}G - G\mathtt{I}^{\mathrm{l}}\Pi_1)A_1 GA.$$

Thus $\Pi_{1,1}^{\mathrm{l}}$ and Π_2^{l} (c.f. (S5) and (S6)) appear naturally in (S33).

To prove (S33), using the expressions of G in (4.5) and A in (S29), we have that

$$(\Xi_1 A)_{j'i} = (\frac{1}{\sqrt{z}}X^*\mathcal{G}_1\mathcal{A}_1 + (\mathcal{G}_2 - m_2)\mathcal{A}_3)_{ji}, \tag{S34}$$

$$(GAG)_{j'i} = (\frac{1}{\sqrt{z}}X^*\mathcal{G}_1\mathcal{A}_1\mathcal{G}_1 + \mathcal{G}_2\mathcal{A}_3\mathcal{G}_1 + \frac{1}{z}X^*\mathcal{G}_1\mathcal{A}_2X^*\mathcal{G}_1 + \frac{1}{\sqrt{z}}\mathcal{G}_2\mathcal{A}_4X^*\mathcal{G}_1)_{ji}. \tag{S35}$$

Expanding the left hand side of (S33) with the above expressions, we shall show that there are two main terms and all others are negligible.

The first contributing term is

$$\begin{aligned}\sum_{i,j}((\mathcal{G}_2 - m_2)\mathcal{A}_3)_{ji}(\mathcal{G}_2\mathcal{A}_3\mathcal{G}_1)_{ji} &= \omega_3^2\mathrm{Tr}((\mathcal{G}_2 - m_2)\boldsymbol{v}\boldsymbol{u}^*\mathcal{G}_1\boldsymbol{u}\boldsymbol{v}^*\mathcal{G}_2)\\ = \omega_3^2(\boldsymbol{u}^*\mathcal{G}_1\boldsymbol{u})(\boldsymbol{v}^*\mathcal{G}_2(\mathcal{G}_2 - m_2)\boldsymbol{v}) &= \sum_{i,j}((\Pi_2^{\mathrm{l}} - \Pi_{1,1}^{\mathrm{l}})A_1\Pi_1)_{j'i}A_{ij'} + O_\prec(n^{-\frac{1}{2}}),\end{aligned}$$

where in the last step we use $\mathcal{G}_2^2 = \mathcal{G}_2'$ and the definition of A in (S29), followed by (4.17) and (4.20).

The second contributing term is

$$\frac{1}{z}\sum_{i,j}(X^*\mathcal{G}_1\mathcal{A}_1)_{ji}(X^*\mathcal{G}_1\mathcal{A}_1\mathcal{G}_1)_{ji} = \omega_1^2(\boldsymbol{u}^*\mathcal{G}_1\boldsymbol{u})\big(\frac{1}{z}\boldsymbol{u}^*\mathcal{G}_1XX^*\mathcal{G}_1\boldsymbol{u}\big).$$

Let $\bar{\boldsymbol{v}} = (\mathbf{0}, \boldsymbol{v})^*$ and $\bar{\boldsymbol{u}} = (\boldsymbol{u}, \mathbf{0})^*$ denote the augmented vectors in $\mathbb{R}^{M+n}$. Note that by (4.20), we first have

$$\bar{\boldsymbol{u}}^*G^2\bar{\boldsymbol{u}} = \boldsymbol{u}^*\mathcal{G}_1^2\boldsymbol{u} + \frac{1}{z}\boldsymbol{u}^*\mathcal{G}_1XX^*\mathcal{G}_1\boldsymbol{u} = 2m_1' + \frac{m_1}{z} + O_\prec(n^{-\frac{1}{2}}).$$

Further observe that

$$\boldsymbol{u}^*\mathcal{G}_1^2\boldsymbol{u} = \bar{\boldsymbol{u}}^* G' \bar{\boldsymbol{u}} = m_1' + O_\prec(n^{-\frac{1}{2}}),$$

where the last equation follows from (4.14). Putting them together, we conclude that

$$\frac{1}{z}\boldsymbol{u}^*\mathcal{G}_1 XX^*\mathcal{G}_1\boldsymbol{u} = m_1' + \frac{m_1}{z} + O_\prec(n^{-\frac{1}{2}}).$$

As a consequence,

$$\frac{1}{z}\sum_{i,j}(X^*\mathcal{G}_1\mathcal{A}_1)_{ji}(X^*\mathcal{G}_1\mathcal{A}_1\mathcal{G}_1)_{ji} = \sum_{i,j}((\Pi_2^{\mathfrak{l}} - \Pi_{1,1}^{\mathfrak{l}})A_1\Pi_1)_{ij'}A_{j'i} + O_\prec(n^{-\frac{1}{2}}).$$

Note that

$$\sum_{i,j}((\Pi_2^{\mathfrak{l}} - \Pi_{1,1}^{\mathfrak{l}})A_1\Pi_1)_{ij'}A_{j'i} + \sum_{i,j}((\Pi_2^{\mathfrak{l}} - \Pi_{1,1}^{\mathfrak{l}})A_1\Pi_1)_{j'i}A_{ij'} = \operatorname{Tr}(\Pi_2^{\mathfrak{l}} - \Pi_{1,1}^{\mathfrak{l}})A_1\Pi_1 A.$$

What remains is to show all other terms in the expansion of the left hand side of (S33) with (S34) and (S35) are negligible. Let us concentrate on the following term. All other remaining terms are estimated similarly; we omit the details.

$$\begin{aligned}\frac{1}{\sqrt{z}}\sum_{i,j}(X^*\mathcal{G}_1\mathcal{A}_1)_{j'i}(\mathcal{G}_2\mathcal{A}_3\mathcal{G}_1)_{j'i} &= \frac{\omega_1\omega_3}{\sqrt{z}}\operatorname{Tr}(X^*\mathcal{G}_1\boldsymbol{u}\boldsymbol{u}^*\mathcal{G}_1\boldsymbol{u}\boldsymbol{v}^*\mathcal{G}_2)\\ &= \frac{\omega_1\omega_3}{\sqrt{z}}\operatorname{Tr}(\boldsymbol{v}^*\mathcal{G}_2^2X^*\boldsymbol{u}\boldsymbol{u}^*\mathcal{G}_1\boldsymbol{u}) = \frac{\omega_1\omega_3}{\sqrt{z}}(\boldsymbol{u}^*\mathcal{G}_1\boldsymbol{u})(\boldsymbol{v}^*\mathcal{G}_2^2X^*\boldsymbol{u}).\end{aligned}$$

In the second step above, we use the fact $X^*\mathcal{G}_1 = \mathcal{G}_2X^*$ which can be checked easily via the singular value decomposition. Therefore, using $\mathcal{G}_2^2 = \mathcal{G}_2'$ and $G' = (G^2 - z^{-1}G)/2$, together with (4.12) and (4.14), we get that

$$\boldsymbol{v}^*\mathcal{G}_2^2X^*\boldsymbol{u} = (\bar{\boldsymbol{v}}^*\sqrt{z}G\bar{\boldsymbol{u}})' = \frac{1}{2\sqrt{z}}\bar{\boldsymbol{v}}^*G\bar{\boldsymbol{u}} + \sqrt{z}\bar{\boldsymbol{v}}^*G'\bar{\boldsymbol{u}} = O_\prec(n^{-\frac{1}{2}}).$$

Hence, we conclude that

$$\frac{1}{\sqrt{z}}\sum_{i,j}(X^*\mathcal{G}_1\mathcal{A}_1)_{j'i}(\mathcal{G}_2\mathcal{A}_3\mathcal{G}_1)_{j'i} = O_\prec(n^{-\frac{1}{2}}).$$

The proof of (S33) is complete.

Next, analogously, we shall show that

$$\sum_{i,j}(\Xi_1A_1)_{j'i}(G^2BG)_{j'i} = \operatorname{Tr}(\Pi_3^{\mathfrak{l}} - \Pi_{2,1}^{\mathfrak{l}})A_1\Pi_1B + O_\prec(n^{-\frac{1}{2}}). \tag{S36}$$

A simple calculation using (4.5) and (4.26) yields

$$(G^2BG)_{j'i} =(\frac{1}{\sqrt{z}}X^*\mathcal{G}_1^2\mathcal{B}_1\mathcal{G}_1 + \frac{1}{\sqrt{z}}\mathcal{G}_2X^*\mathcal{G}_1\mathcal{B}_1\mathcal{G}_1 + \frac{1}{z}X^*\mathcal{G}_1^2X\mathcal{B}_3\mathcal{G}_1 + \mathcal{G}_2^2\mathcal{B}_3\mathcal{G}_1 + \frac{1}{z}X^*\mathcal{G}_1^2\mathcal{B}_2X^*\mathcal{G}_1$$

(S37)

$$+ \frac{1}{z}\mathcal{G}_2X^*\mathcal{G}_1\mathcal{B}_2X^*\mathcal{G}_1 + \frac{1}{z^{\frac{3}{2}}}X^*\mathcal{G}_1^2X\mathcal{B}_4X^*\mathcal{G}_1 + \frac{1}{\sqrt{z}}\mathcal{G}_2^2\mathcal{B}_4X^*\mathcal{G}_1)_{ji}.$$

In a similar way to the discussion of (S33), we expand $(\Xi_1 A)_{j'i}(G^2BG)_{j'i}$ using (S34) and (S37). There are only four non-negligible terms in the expansion.

Recall $\mathcal{A}_1$ and $\mathcal{B}_1$ in (4.26). The first non-negligible term is

$$\frac{1}{z}\sum_{i,j}(X^*\mathcal{G}_1\mathcal{A}_1)_{j'i}(X^*\mathcal{G}_1^2\mathcal{B}_1\mathcal{G}_1)_{j'i} = \frac{\omega_1\varpi_1}{z}(\boldsymbol{u}^*\mathcal{G}_1\boldsymbol{u})(\boldsymbol{u}^*\mathcal{G}_1^2XX^*\mathcal{G}_1\boldsymbol{u}).$$

To estimate $\boldsymbol{u}^*\mathcal{G}_1^2XX^*\mathcal{G}_1\boldsymbol{u}$ in the above, we observe that (via elementary calculations and the fact $\mathcal{G}_2X^* = X^*\mathcal{G}_1$)

$$\bar{\boldsymbol{u}}^*G^3\bar{\boldsymbol{u}} = \boldsymbol{u}^*\mathcal{G}_1^3\boldsymbol{u} + \frac{3}{z}(\boldsymbol{u}^*\mathcal{G}_1^2XX^*\mathcal{G}_1\boldsymbol{u}).$$

Moreover, by $\mathcal{G}_1^3 = \frac{1}{2}\mathcal{G}_1''$, (4.14) and (4.20), we find

$$\boldsymbol{u}^*\mathcal{G}_1^3\boldsymbol{u} = \frac{1}{2}\bar{\boldsymbol{u}}^*G''\bar{\boldsymbol{u}} = \frac{1}{2}m_1'' + O_\prec(n^{-\frac{1}{2}}),$$
$$\bar{\boldsymbol{u}}^*G^3\bar{\boldsymbol{u}} = 2m_1'' + \frac{3}{z}m_1' + O_\prec(n^{-\frac{1}{2}}).$$

Hence,

$$\boldsymbol{u}^*\mathcal{G}_1^2XX^*\mathcal{G}_1\boldsymbol{u} = \frac{z}{2}m_1'' + m_1' + O_\prec(n^{-\frac{1}{2}}). \tag{S38}$$

We conclude that

$$\frac{1}{z}\sum_{i,j}(X^*\mathcal{G}_1\mathcal{A}_1)_{ji}(X^*\mathcal{G}_1^2\mathcal{B}_1\mathcal{G}_1)_{ji} = \frac{1}{2}\sum_{i,j}((\Pi_3^{\mathfrak{l}} - \Pi_{2,1}^{\mathfrak{l}})A_1\Pi_1)_{ij'}B_{ij'} + O_\prec(n^{-\frac{1}{2}}).$$

Using the fact $X\mathcal{G}_2 = \mathcal{G}_1X$ and the same arguments as above, we can show the second non-negligible term is

$$\frac{1}{z}\sum_{i,j}(X^*\mathcal{G}_1\mathcal{A}_1)_{ji}(\mathcal{G}_2X^*\mathcal{G}_1\mathcal{B}_1\mathcal{G}_1)_{ji} = \frac{1}{2}\sum_{i,j}((\Pi_3^{\mathfrak{l}} - \Pi_{2,1}^{\mathfrak{l}})A_1\Pi_1)_{ij'}B_{ij'} + O_\prec(n^{-\frac{1}{2}}),$$

The third non-negligible term is

$$\frac{1}{z}\sum_{i,j}((\mathcal{G}_2-m_2)\mathcal{A}_3)_{j'i}(X^*\mathcal{G}_1^2X\mathcal{B}_3\mathcal{G}_1)_{j'i}=\frac{1}{z}(\boldsymbol{u}^*\mathcal{G}_1\boldsymbol{u})(\boldsymbol{v}^*X^*\mathcal{G}_1^2X(\mathcal{G}_2-m_2)\boldsymbol{v})$$
$$=\omega_3\varpi_3 m_1(\frac{m_2''}{2}+\frac{m_2'}{z}-\frac{m_2^2+zm_2m_2'}{z})+O_\prec(n^{-\frac{1}{2}}),$$

where we used the facts $X^*\mathcal{G}_1^2X\mathcal{G}_2=X^*\mathcal{G}_1^3X$ and $\mathcal{G}_2^3=\frac{1}{2}\mathcal{G}_2''$, as well as

$$\bar{\boldsymbol{v}}^*G^3\bar{\boldsymbol{v}}=\frac{3}{z}\boldsymbol{v}^*X^*\mathcal{G}_1^3X\boldsymbol{v}+\boldsymbol{v}^*\mathcal{G}_2^3\boldsymbol{v}=2m_2''+\frac{3m_2'}{z}+O_\prec(n^{-\frac{1}{2}}).$$

The last non-negligible term can be estimated similarly as

$$\sum_{i,j}((\mathcal{G}_2-m_2)\mathcal{A}_3)_{ji}(\mathcal{G}_2^2\mathcal{B}_3\mathcal{G}_1)_{ji}=\omega_3\varpi_3m_1(\frac{m_2''}{2}-m_2m_2')+O_\prec(n^{-\frac{1}{2}}).$$

Consequently, we have

$$\frac{1}{z}\sum_{i,j}((\mathcal{G}_2-m_2)\mathcal{A}_3)_{ji}(X^*\mathcal{G}_1^2X\mathcal{B}_3\mathcal{G}_1)_{ji}+\sum_{i,j}((\mathcal{G}_2-m_2)\mathcal{A}_3)_{ji}(\mathcal{G}_2^2\mathcal{B}_3\mathcal{G}_1)_{ji}$$
$$=\sum_{i,j}((\Pi_3^{\mathrm{l}}-\Pi_{2,1}^{\mathrm{l}})A_1\Pi_1)_{j'i}B_{j'i}+O_\prec(n^{-1/2}).$$

Note that the sum of the four contributing terms is extactly

$$\mathrm{Tr}(\Pi_3^{\mathrm{l}}-\Pi_{2,1}^{\mathrm{l}})A_1\Pi_1B+O_\prec(n^{-\frac{1}{2}}).$$

To wrap up the proof of (S36), it suffices to show all the other terms in the expansion of $\sum_{i,j}(\Xi_1A)_{j'i}(G^2BG)_{j'i}$ can be bounded by $O_\prec(n^{-\frac{1}{2}})$. To see that, for instance, we focus on

$$z^{-3/2}\sum_{i,j}(X^*\mathcal{G}_1\mathcal{A}_1)_{j'i}(X^*\mathcal{G}_1^2X\mathcal{B}_3\mathcal{G}_1)_{j'i}=\omega_1\varpi_1(z^{-3/2}\boldsymbol{v}^*X^*\mathcal{G}_1^2XX^*\mathcal{G}_1\boldsymbol{u})(\boldsymbol{u}^*\mathcal{G}_1\boldsymbol{u}).$$

Note that

$$z^{-3/2}\boldsymbol{v}^*X^*\mathcal{G}_1^2XX^*\mathcal{G}_1\boldsymbol{u}=\bar{\boldsymbol{u}}^*G^3\bar{\boldsymbol{v}}-\boldsymbol{u}^*(\frac{1}{\sqrt{z}}\mathcal{G}_1^3X+\frac{1}{\sqrt{z}}\mathcal{G}_1^2X\mathcal{G}_2+\frac{1}{\sqrt{z}}\mathcal{G}_1X\mathcal{G}_2^2)\boldsymbol{v},$$
$$=\bar{\boldsymbol{u}}G^3\bar{\boldsymbol{v}}-\frac{3}{\sqrt{z}}\boldsymbol{u}^*\mathcal{G}_1^3X\boldsymbol{v}=\bar{\boldsymbol{u}}G^3\bar{\boldsymbol{v}}-\frac{3}{2\sqrt{z}}\bar{\boldsymbol{u}}^*(\sqrt{z}G)''\bar{\boldsymbol{v}}=O_\prec(n^{-\frac{1}{2}}),$$

where in the third step we use $\mathcal{G}_1^3=\frac{1}{2}\mathcal{G}_1''$ and in the last step we use (4.20). Consequently,

$$z^{-3/2}\sum_{i,j}(X^*\mathcal{G}_1\mathcal{A}_1)_{j'i}(X^*\mathcal{G}_1^2X\mathcal{B}_3\mathcal{G}_1)_{j'i}=O_\prec(n^{-\frac{1}{2}}).$$

All the rest terms can be bounded by $O_\prec(n^{-\frac{1}{2}})$ analogously; we omit the details. The proof of (S36) is now complete.

The remaining two terms in (S32) can be estimated the same way as (S33) and (S36); the details are omitted. We get

$$\sum_{i,j}(\Xi_1 A_1)_{j'i}(GBG^2)_{j'i} = \mathrm{Tr}(\Pi_2^{\mathfrak{l}} - \Pi_{1,1}^{\mathfrak{l}})A_1\Pi_2 B + O_\prec(n^{-\frac{1}{2}}),$$

$$\sum_{i,j}(\Xi_1 A_1)_{j'i}(GBG)_{ij'} = \mathrm{Tr}(\Pi_2^{\mathfrak{l}} - \Pi_{1,1}^{\mathfrak{l}})A_1\Pi_1 B + O_\prec(n^{-\frac{1}{2}}). \tag{S39}$$

Plugging (S33), (S36) and (S39) into (S32), recalling the definition of $\mathfrak{a}_{11}$ in (S9), we conclude that

$$\hbar(0,1,0) = \mathfrak{a}_{11}\mathfrak{q}^{(k-2)} + O_\prec(n^{-\frac{1}{2}+4\nu}). \tag{S40}$$

This completes the proof. □

F.2. *Proof of Lemma F.3.* We use this subsection to prove Lemma F.3.

Proof of Lemma F.3. We first study the second derivatives. By (S11) and (S2), we have

$$\hbar(2,0,0) = \frac{1}{n}\sum_{i,j}\Big(\frac{\partial^2 G}{\partial x_{ij}^2}A_1\Big)_{j'i}\mathfrak{q}^{(k-1)}$$
$$= \frac{2z}{n}\sum_{i,j}\Big(\big(G_{j'j'}G_{ij'} + G_{j'i}G_{j'j'}\big)(GA_1)_{ii} + \big(G_{j'j'}G_{ii} + G_{j'i}G_{j'i}\big)(GA_1)_{j'i}\Big)\mathfrak{q}^{(k-1)}.$$

First of all, by (4.17) and (4.20), we find that

$$\frac{1}{n}\sum_{i,j}G_{j'j'}G_{ii}(GA_1)_{j'i} = \frac{1}{n}\sum_{i,j}(\Pi_1)_{ii}(\Pi_1)_{j'j'}(\Pi_1 A_1)_{j'i} + O_\prec(n^{-\frac{1}{2}}).$$

It is simple to check that $(GA)_{ii} = (\mathcal{G}_1\mathcal{A}_1 + z^{-1/2}\mathcal{G}_1 X\mathcal{A}_3)_{ii}$. By (4.15) and (4.20), we get

$$\frac{1}{n}\sum_{i,j}G_{j'j'}G_{ij'}(GA_1)_{ii} = O_\prec\Big(n^{-\frac{3}{2}}\sum_{i,j}(\mathcal{G}_1\boldsymbol{u}\boldsymbol{u}^* + \mathcal{G}_1 X\boldsymbol{v}\boldsymbol{u}^*)_{ii}\Big) = O_\prec(n^{-\frac{1}{2}}). \tag{S41}$$

Similarly, we also have

$$\frac{1}{n}\sum_{i,j} G_{j'j'}G_{j'i}(GA_1)_{ii} = O_\prec(n^{-\frac{1}{2}}),$$
$$\frac{1}{n}\sum_{i,j} G_{j'i}G_{j'i}(GA_1)_{j'i} = O_\prec(n^{-1}).$$

Putting the above estimates together, and recalling $\mathfrak{d}_1^a$ in (S7), we conclude that

$$\hbar(2,0,0) = \mathfrak{d}_1^a \mathfrak{q}^{(k-1)} + O_\prec(n^{-\frac{1}{2}}).$$

Next, the estimation of

$$\hbar(1,1,0) = \frac{k-1}{n}\sum_{i,j}\Big(\frac{\partial G}{\partial x_{ij}}A_1\Big)_{j'i}\frac{\partial Q}{\partial x_{ij}}\mathfrak{q}^{(k-2)}$$

follows closely the same steps as the derivation of (S13). By (S28),

$$\hbar(1,1,0) = -\frac{(k-1)\sqrt{z}}{n}\sum_{i,j}\Big(G_{j'j'}(GA_1)_{ii} + G_{j'i}(GA_1)_{j'i}\Big)\frac{\partial Q}{\partial x_{ij}}\mathfrak{q}^{(k-2)}. \tag{S42}$$

We shall prove that

$$\frac{1}{n}\sum_{i,j} G_{j'j'}(GA_1)_{ii}\frac{\partial Q}{\partial x_{ij}} = \sqrt{\frac{z}{n}}\sum_{\sum_{(i,j)\in\mathcal{S}(\nu)}}(\Pi_1 A)_{ii}(\Pi_1)_{j'j'}c_{ij} + O_\prec(n^{-\frac{1}{2}+4\nu}), \tag{S43}$$

which will be used several times later. We postpone the proof of (S43) till the end of this subsection.

Again by (4.20), recalling the definitions of c_{ij} in (5.7) and A in (4.26), we have that

$$\frac{1}{n}\sum_{i,j} G_{j'i}(GA_1)_{j'i}\frac{\partial Q}{\partial x_{ij}} = O_\prec(n^{-3/2}\sum_{i,j} A_{ij'}c_{ij}) = O_\prec(n^{-\frac{1}{2}}). \tag{S44}$$

Inserting (S43) and (S44) back into (S42), by recalling $\mathfrak{a}_{21}$ in (S10), we conclude that

$$\hbar(1,1,0) = \mathfrak{a}_{21}\mathfrak{q}^{(k-2)} + O_\prec(n^{-\frac{1}{2}+4\nu}).$$

Using a discussion similar to (S14), we also have

$$\hbar(1,0,1)=\sqrt{\frac{z}{n}}\sum_{(i,j)\in\mathcal{B}(\nu)} \mathrm{i}tc_{ij}\Big(\frac{\partial G}{\partial x_{ij}}A_1\Big)_{j'i}\mathfrak{q}^{(k-1)}=O_{\prec}(n^{-\frac{1}{2}+4\nu}).$$

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

It remains to estimate

$$\hbar(0,2,0)=\frac{1}{n}\sum_{i,j}(\Xi_1A_1)_{j'i}\Big((k-1)\frac{\partial^2 Q}{\partial x_{ij}^2}\mathfrak{q}^{(k-2)}+(k-1)(k-2)\Big(\frac{\partial Q}{\partial x_{ij}}\Big)^2\mathfrak{q}^{(k-3)}\Big). \tag{S45}$$

The calculation of (S45) is similar to that of (S32) and due to an extra factor $n^{-1/2}$ in front, we shall show that $\hbar(0,2,0)$ 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 (S45) using the expressions of $(\Xi_1A_1)_{j'i}$ in (S34), $\partial Q/\partial x_{ij}$ in (S1) and $\partial^2 Q/\partial x_{ij}^2$ in (S6).

Most derivations of the items in (S32) can be directly applied to those in (S45) except three items, which are discussed below. Denote $\boldsymbol{e}_i$ with $i\in[M]$ as the standard basis in $\mathbb{R}^M$ and $\boldsymbol{f}_j$ with $j\in[N]$ as those in $\mathbb{R}^N$,

First, by (4.17) and (4.20),

$$\begin{aligned}\sum_{i,j}(X^*\mathcal{G}_1\mathcal{A}_1)_{j'i}(X^*\mathcal{G}_1\mathcal{A}_1\mathcal{G}_1)_{j'i}(X^*\mathcal{G}_1\mathcal{A}_1\mathcal{G}_1)_{j'i}&=\sum_{i,j}(\boldsymbol{e}_j^*X^*\mathcal{G}_1\mathcal{A}_1\boldsymbol{e}_i)(\boldsymbol{e}_j^*X^*\mathcal{G}_1\mathcal{A}_1\mathcal{G}_1\boldsymbol{e}_i)^2\\ &=O_{\prec}\Big(n^{-\frac{3}{2}}\sum_{i,j}\boldsymbol{u}^3(i)\Big)=O_{\prec}(n^{-\frac{1}{2}}).\end{aligned} \tag{S46}$$

Second, using (4.17) and the fact that $\boldsymbol{u},\boldsymbol{v}$ are unit vectors, we get

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i,j}(X^*\mathcal{G}_1\mathcal{A}_1)_{ji}(GAG)_{j'j'}G_{ii}&=\frac{m_1m_2^2}{\sqrt{n}}\sum_{i,j}(X^*\mathcal{G}_1\mathcal{A}_1)_{ji}A_{j'j'}+O_{\prec}(n^{-\frac{1}{2}}),\\ &=\frac{m_1m_2^2\omega_1^2}{\sqrt{n}}\sum_{i,j}\boldsymbol{f}_j^*X^*\mathcal{G}_1\boldsymbol{u}\boldsymbol{u}(i)\boldsymbol{v}^2(j)+O_{\prec}(n^{-\frac{1}{2}})\\ &=\frac{m_1m_2^2\omega_1^2}{n}\sum_{i,j}\boldsymbol{u}(i)\boldsymbol{v}^2(j)+O_{\prec}(n^{-\frac{1}{2}})\\ &=O_{\prec}(n^{-\frac{1}{2}}).\end{aligned}$$

Third, we invoke (4.17) to get that $\sum_j \boldsymbol{f}_j^* X^* \mathcal{G}_1 \boldsymbol{u} = \sqrt{n} \boldsymbol{f} X^* \mathcal{G}_1 \boldsymbol{u} = O_\prec(1)$, where $\boldsymbol{f} = \frac{1}{\sqrt{n}} \mathbf{1}$. Then it follows that

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i,j} (X^* \mathcal{G}_1 \mathcal{A}_1)_{ji} (GAG)_{ii} G_{j'j'} &= \frac{m_1^2 m_2}{\sqrt{n}} \sum_{i,j} (X^* \mathcal{G}_1 \mathcal{A}_1)_{ji} A_{ii} + O_\prec(n^{-\frac{1}{2}}), \\
&= \frac{m_1^2 m_2 \omega_1^2}{\sqrt{n}} \sum_{i,j} \boldsymbol{f}_j^* X^* \mathcal{G}_1 \boldsymbol{u} \boldsymbol{u}^3(i) + O_\prec(n^{-\frac{1}{2}}) \\
&= \frac{m_1^2 m_2 \omega_1^2}{\sqrt{n}} \Big(\sum_i \boldsymbol{u}^3(i) \Big) \Big(\sum_j \boldsymbol{f}_j^* X^* \mathcal{G}_1 \boldsymbol{u} \Big) + O_\prec(n^{-\frac{1}{2}}) \\
&= O_\prec(n^{-\frac{1}{2}}). \qquad \text{(S47)}
\end{aligned}$$

Finally, we can conclude that

$$\hbar(0,2,0) = O_\prec(n^{-\frac{1}{2}}).$$

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

(S48)
$$\hbar(1,2,0) = n^{-\frac{3}{2}} \sum_{i,j} (\frac{\partial G}{\partial x_{ij}} A_1)_{j'i} \Big((k-1) \frac{\partial^2 Q}{\partial x_{ij}^2} \mathfrak{q}^{(k-2)} + (k-1)(k-2) \Big(\frac{\partial Q}{\partial x_{ij}} \Big)^2 \mathfrak{q}^{(k-3)} \Big).$$

Recalling (S28) and (S6), by (4.17) and (4.20), the first term on the right hand side of (S48) is estimated by

$$\begin{aligned}
& n^{-\frac{3}{2}} \sum_{i,j} (\frac{\partial G}{\partial x_{ij}} A_1)_{j'i} \frac{\partial^2 Q}{\partial x_{ij}^2} \\
&= \frac{-2z^{\frac{3}{2}}}{n} \sum_{i,j} G_{j'j'} (GA_1)_{ii} \Big((GAG)_{ii} G_{j'j'} - \frac{1}{2z} (GBG)_{ii} G_{j'j'} + \frac{1}{2} (GBG)_{ii} G_{j'j'}^2 \\
&\qquad + \frac{1}{2} (G^2BG)_{ii} G_{j'j'} + \frac{1}{2} (GBG^2)_{ii} G_{j'j'} \Big) + O_\prec(n^{-\frac{1}{2}}) \\
&= -\frac{2z^{\frac{3}{2}}}{n} \sum_{i,j} (\Pi_1 A)_{ii} (\Pi_1)_{j'j'} s_{ij} + O_\prec(n^{-1/2}). \qquad \text{(S49)}
\end{aligned}$$

In the last equation above, we recall the definition of s_{ij} in (5.8). Furthermore, recalling (S2) and (S28), by (4.20), it is easy to see that the second term on the right hand side of (S48) is

$$n^{-3/2}\sum_{i,j}\Big(\frac{\partial G}{\partial x_{ij}}A_1\Big)_{j'i}\Big(\frac{\partial Q}{\partial x_{ij}}\Big)^2\mathfrak{q}^{(k-3)}=O_\prec\Big(n^{-3/2}\sum_{i,j}(A_1)_{ii}c_{ij}^2\Big)=O_\prec(n^{-1}).$$

For the last equation above, we refer to the definition of c_{ij} in (5.7). Using $\mathfrak{a}_{31}$ defined in (S11), we hence conclude that

$$\hbar(1,2,0)=\mathfrak{a}_{31}\mathfrak{q}^{(k-2)}+O_\prec(n^{-1/2}).$$

Next we study

$$\hbar(0,3,0)=n^{-\frac{3}{2}}\sum_{ij}(\Xi_1A_1)_{j'i}\Big((k-1)\frac{\partial^3Q}{\partial x_{ij}^3}\mathfrak{q}^{(k-2)}+3\frac{(k-1)!}{(k-3)!}\frac{\partial^2Q}{\partial x_{ij}^2}\frac{\partial Q}{\partial x_{ij}}\mathfrak{q}^{(k-3)}+\frac{(k-1)!}{(k-4)!}\Big(\frac{\partial Q}{\partial x_{ij}}\Big)^3\mathfrak{q}^{(k-4)}\Big).$$

We briefly argue that $\hbar(0,3,0)$ is bounded by $O_\prec(n^{-\frac{1}{2}})$, using a discussion similar to those of $\hbar(0,1,0)$ in (S32) and $\hbar(0,2,0)$ in (S45).

Recalling (S1) and (S6), it is easy to see that

$$n^{-\frac{3}{2}}\sum_{i,j}(\Xi_1A_1)_{j'i}\frac{\partial^2Q}{\partial x_{ij}^2}\frac{\partial Q}{\partial x_{ij}}=O_\prec(n^{-1}\sum_{i,j}\boldsymbol{u}(i)\boldsymbol{v}^3(j))=O_\prec(n^{-\frac{1}{2}}).$$

Similarly, by (S1) and (S7), it can also be shown that

$$n^{-\frac{3}{2}}\sum_{i,j}(\Xi_1A_1)_{j'i}\Big(\frac{\partial Q}{\partial x_{ij}}\Big)^3=O_\prec(n^{-\frac{1}{2}}\sum_{i,j}\boldsymbol{u}^3(i)\boldsymbol{v}^3(j))=O_\prec(n^{-\frac{1}{2}}),$$
$$n^{-\frac{3}{2}}\sum_{i,j}(\Xi_1A_1)_{j'i}\frac{\partial^3Q}{\partial x_{ij}^3}=O_\prec(n^{-\frac{1}{2}}).$$

This completes the discussion of $\hbar(0,3,0)$. The same arguments can be applied to show that

$$\hbar(2,1,0)=n^{-\frac{3}{2}}\sum_{i,j}\Big(\frac{\partial^2G}{\partial x_{ij}^2}A_1\Big)_{j'i}\frac{\partial Q}{\partial x_{ij}}\mathfrak{q}^{(k-2)}=O_\prec(n^{-1}\sum_{i,j}\boldsymbol{u}^2(i)\boldsymbol{v}^2(j))=O_\prec(n^{-1}).$$

and

$$\hbar(3,0,0)=n^{-\frac{3}{2}}\sum_{i,j}\Big(\frac{\partial^3G}{\partial x_{ij}^3}A_1\Big)_{j'i}\mathfrak{q}^{(k-1)}=O_\prec(n^{-3/2}\sum_{i,j}(A_1)_{j'i})=O_\prec(n^{-\frac{1}{2}})$$

by using the expressions (S2) and (S3) respectively.

For all the rest of the items containing the derivatives of h_3, they can be easily estimated using a discussion similar to (S14).

Finally, using (S2)-(S8), (S1), (S28) and (4.17), all the fourth order derivatives can bounded by $O_\prec(n^{-\frac{1}{2}})$. The discussion is similar to that of (S45); we omit further details here. This concludes our proof. □

PROOF OF (S43). We split the left hand side of (S43) as the sum of the following three items

$$\frac{1}{n}\sum_{i,j}(G_{j'j'}-m_2)(GA_1)_{ii}\frac{\partial Q}{\partial x_{ij}},\ \ \frac{1}{n}\sum_{i,j}m_2(\Xi_1 A_1)_{ii}\frac{\partial Q}{\partial x_{ij}},\ \ \frac{1}{n}\sum_{i,j}m_1m_2(A_1)_{ii}\Big(\frac{\partial Q}{\partial x_{ij}}-\sqrt{nz}c_{ij}\Big).$$

First of all, by (4.20) and (S1), we have

$$\frac{1}{n}\sum_{i,j}(G_{j'j'}-m_2)(GA_1)_{ii}\frac{\partial Q}{\partial x_{ij}}=O_\prec\Big(n^{-1}\sum_{(i,j)\in\mathcal{S}(\nu)}\boldsymbol{u}^3(i)\boldsymbol{v}(j)\Big)=O_\prec(n^{-\frac{1}{2}}).$$

Similarly, we also have

$$\begin{aligned}\frac{1}{n}\sum_{i,j}m_2(\Xi_1 A_1)_{ii}\frac{\partial Q}{\partial x_{ij}}&=\frac{m_2\omega_1}{n}\sum_{i,j}\boldsymbol{e}_i^*\Xi_1\boldsymbol{u}\boldsymbol{u}(i)\frac{\partial Q}{\partial x_{ij}}\\&=O_\prec(n^{-1}\sum_{i,j}\boldsymbol{u}(i)^2\boldsymbol{v}(j))=O_\prec(n^{-\frac{1}{2}}).\end{aligned}$$

Furthermore, using a discussion similar to that of (S47), we get

$$\frac{1}{n}\sum_{i,j}m_1m_2(A_1)_{ii}\Big(\frac{\partial Q}{\partial x_{ij}}-\sqrt{nz}c_{ij}\Big)=O_\prec(n^{-\frac{1}{2}}\sum_i\boldsymbol{u}^3(i))=O_\prec(n^{-\frac{1}{2}}),$$

where we apply the fact that

$$(GAG)_{j'i}-m_1m_2A_{j'i}=O_\prec\Big(\frac{\boldsymbol{u}(i)}{\sqrt{n}}\Big).$$

Summing up the above three estimates, we can conclude the proof of (S43). □

F.3. *Proofs of Lemmas F.4-F.7.* In this subsection, we will prove Lemmas F.4-F.7. The proofs are analogous to those of Lemma F.2 and Lemma F.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

(S50)
$$\Big(\frac{\partial G^2}{\partial x_{ij}} W\Big)_{ab} = -\sqrt{z}(G^2_{aj'}(GW)_{ib}+G^2_{ai}(GM)_{j'b}+G_{aj'}(G^2W)_{ib}+G_{ai}(G^2W)_{j'b}).$$

Proof of Lemma F.4. Recalling $\tilde{h}_1$ in (S20), by a discussion similar to (S30), we get

$$\begin{aligned}\tilde{\hbar}(1,0,0) &= \frac{1}{\sqrt{n}}\sum_{i,j}\frac{\partial \tilde{h}_1}{\partial x_{ij}}h_2h_3 = \frac{1}{\sqrt{n}}\sum_{i,j}\Big(\frac{\partial G^2}{\partial x_{ij}}B_1\Big)_{j'i}\mathfrak{q}^{(k-1)} \\ &= -\sqrt{nz}\big((2m_2'+\frac{m_2}{z})\mathrm{Tr}(GB_1)+m_2\mathrm{Tr}(G^2B_1)\big)\mathfrak{q}^{(k-1)}+O_\prec(n^{-\frac{1}{2}}).\end{aligned}$$

In the last step above, we use (S50), (4.17) and (4.20). Next, we turn towards to the term

$$\tilde{\hbar}(0,1,0) = \frac{1}{\sqrt{n}}\sum_{i,j}\tilde{h}_1\frac{\partial h_2}{\partial x_{ij}}h_3 = \frac{(k-1)}{n}\sum_{i,j}(\Xi_2 B_1)_{j'i}\frac{\partial Q}{\partial x_{ij}}\mathfrak{q}^{(k-2)}, \tag{S51}$$

which will be estimated following exactly the same steps as those of (S13). Observe that

$$(\Xi_2 B)_{j'i} = \Big((\frac{1}{\sqrt{z}}X^*\mathcal{G}_1^2+\frac{1}{\sqrt{z}}\mathcal{G}_2X^*\mathcal{G}_1)\mathcal{B}_1+(\frac{1}{z}X^*\mathcal{G}_1^2X+\mathcal{G}_2^2-2m_2'-\frac{m_2}{z})\mathcal{B}_3\Big)_{j'i}.$$

By (S1), after expanding the product on the right hand side of (S51), we find the following estimates.

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i,j}(\Xi_2B_1)_{j'i}(GAG)_{j'i} &= \mathrm{Tr}(\Pi_3^{\mathfrak{l}}-\Pi_{1,2}^{\mathfrak{l}})B_1\Pi_1A+O_\prec(n^{-\frac{1}{2}}),\\ \frac{1}{\sqrt{n}}\sum_{i,j}(\Xi_2B_1)_{j'i}(GBG)_{j'i} &= \mathrm{Tr}(\Pi_3^{\mathfrak{l}}-\Pi_{1,2}^{\mathfrak{l}})B_1\Pi_1B+O_\prec(n^{-\frac{1}{2}}),\\ \frac{1}{\sqrt{n}}\sum_{i,j}(\Xi_2B_1)_{j'i}(G^2BG)_{j'i} &= \mathrm{Tr}(\Pi_4^{\mathfrak{l}}-\Pi_{2,2}^{\mathfrak{l}})B_1\Pi_1B+O_\prec(n^{-\frac{1}{2}}),\\ \frac{1}{\sqrt{n}}\sum_{i,j}(\Xi_2B_1)_{j'i}(GBG^2)_{j'i} &= \mathrm{Tr}(\Pi_3^{\mathfrak{l}}-\Pi_{1,2}^{\mathfrak{l}})B_1\Pi_2B+O_\prec(n^{-\frac{1}{2}}).\end{aligned}$$

Putting these estimates together and invoking $\tilde{\mathfrak{b}}_{11}$ in (S9), we have

$$\tilde{\hbar}(0,1,0) = \tilde{\mathfrak{b}}_{11}\mathfrak{q}^{(k-2)}+O_\prec(n^{-\frac{1}{2}+4\nu}).$$

Lastly, $\tilde{\hbar}(0,0,1)$ can be estimated using a discussion similar to (S14). We can therefore conclude our proof. ☐

The proof of Lemma F.5 follows along the exact lines of Lemma F.3 with minor changes. We only sketch it below.

PROOF OF LEMMA F.5. First of all, by (4.20) and (S5), using a discussion similar to (S15), we have that

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

Second, following the same steps in (S42), together with (S50) and (S1), we find that

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

Third, the same arguments of (S17) using (S50) and (S6) yield

$$\tilde{\hbar}(1,2,0) = n^{-\frac{3}{2}} \sum_{i,j} \frac{\partial \tilde{h}_1}{\partial x_{ij}} \frac{\partial^2 h_2}{\partial x_{ij}^2} h_3 = \tilde{\mathfrak{b}}_{31} \mathfrak{q}^{k-2} + O_{\prec}(n^{-1/2}).$$

For the rest of the items, we can apply discussions similar to those of the corresponding items from Lemma F.3. We omit the details here.

□

Lemma F.6 is an analogue of Lemma F.2 and F.4 for the matrices A_2 and B_2; the proof is analogous.

PROOF OF LEMMA F.6. Recall (S28). Using a discussion similar to that of (S30), by (4.17) and (4.20), we find that

$$\begin{aligned}
\mathfrak{h}(1,0,0) &= \frac{1}{\sqrt{n}}\sum_{i,j} \frac{\partial \mathbb{h}_1}{\partial x_{ij}} h_2 h_3 = \frac{1}{\sqrt{n}}\sum_{i,j}\left(\frac{\partial G}{\partial x_{ij}} A_2\right)_{ij'} \mathfrak{q}^{(k-1)} \\
&= -\sqrt{nz}\frac{1}{n}\sum_{i,j}\Big(G_{ii}(GA_2)_{j'j'} + G_{ij'}(GA_2)_{ij'}\Big)\mathfrak{q}^{(k-1)} \\
&= -\sqrt{nz} y m_{1n} (\mathrm{Tr} G A_2) \mathfrak{q}^{(k-1)} - \frac{\sqrt{z}}{\sqrt{n}} (\mathrm{Tr} G A_2^* G) \mathfrak{q}^{(k-1)} \\
&= -\sqrt{nz} y m_1 (\mathrm{Tr} G A_2) \mathfrak{q}^{(k-1)} + O_{\prec}(n^{-\frac{1}{2}}),
\end{aligned}$$

where we recall that $y = \frac{M}{n}$. Similarly, using (S50), we also have

$$\begin{aligned}\tilde{\mathfrak{h}}(1,0,0) &= \frac{1}{\sqrt{n}}\sum_{i,j}\frac{\partial\tilde{\mathbb{h}}_1}{\partial x_{ij}}h_2h_3 = \frac{1}{\sqrt{n}}\sum_{i,j}\Big(\frac{\partial G^2}{\partial x_{ij}}B_2\Big)_{ij'}\mathfrak{q}^{(k-1)}\\ &= -\sqrt{n}z\frac{1}{n}\sum_{i,j}\Big(G_{ii}^2(GB_2)_{j'j'}+G_{ii}(G^2B_2)_{j'j'}\Big)\mathfrak{q}^{(k-1)}+O_\prec(n^{-\frac12})\\ &= -\sqrt{n}zy\Big((2m_1'+\frac{m_1}{z})(\mathrm{Tr}GB_2)+m_1\mathrm{Tr}G^2B_2\Big)\mathfrak{q}^{(k-1)}+O_\prec(n^{-\frac12}),\end{aligned}$$

Next, we estimate

$$\mathfrak{h}(0,1,0) = \frac{k-1}{\sqrt{n}}\sum_{i,j}(\Xi_1A_2)_{ij'}\frac{\partial Q}{\partial x_{ij}}\mathfrak{q}^{(k-2)}.$$

Recall the expression of $\partial Q/\partial x_{ij}$ in (S1). As seen in the discussion below (S32), the key observation is that

$$\sum_{i,j}(\Xi_1A_2)_{ij'}(GAG)_{ij'} = \mathrm{Tr}(G\mathtt{I}^{\mathrm{u}}G - G\mathtt{I}^{\mathrm{u}}\Pi_1)A_2GA$$

Thus we shall prove

$$\sum_{i,j}(\Xi_1A_2)_{ij'}(GAG)_{ij'} = \mathrm{Tr}(\Pi_2^{\mathrm{u}}-\Pi_{1,1}^{\mathrm{u}})A_2\Pi_1A+O_\prec(n^{-\frac12}).$$

The proof follows from

$$(\Xi_1A_2)_{ij'} = ((\mathcal{G}_1-m_1)\mathcal{A}_2+z^{-1/2}\mathcal{G}_1X\mathcal{A}_4)_{ij}$$

and exactly the same arguments as (S33). Likewise, we also get

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i,j}(\Xi_1A_2)_{ij'}(GBG)_{ij'} &= \mathrm{Tr}(\Pi_2^{\mathrm{u}}-\Pi_{1,1}^{\mathrm{u}})A_2\Pi_1B+O_\prec(n^{-\frac12}),\\ \frac{1}{\sqrt{n}}\sum_{i,j}(\Xi_1A_2)_{ij'}(G^2BG)_{ij'} &= \mathrm{Tr}(\Pi_3^{\mathrm{u}}-\Pi_{2,1}^{\mathrm{u}})A_2\Pi_1B+O_\prec(n^{-\frac12}),\\ \frac{1}{\sqrt{n}}\sum_{i,j}(\Xi_1A_2)_{ij'}(GBG^2)_{ij'} &= \mathrm{Tr}(\Pi_2^{\mathrm{u}}-\Pi_{1,1}^{\mathrm{u}})A_2\Pi_2B+O_\prec(n^{-\frac12}).\end{aligned}$$

Putting the above estimates together and recalling $\mathfrak{a}_{12}$ below (S9), we finish the computation for (S23). The other term $\tilde{\mathfrak{h}}(0,1,0)$ can be estimated analogously by noting

$$(\Xi_2B_2)_{ij'} = ((\mathcal{G}_1^2+\frac{1}{z}\mathcal{G}_1XX^*\mathcal{G}_1-2m_1'-\frac{m_1}{z})\mathcal{B}_2+(\frac{1}{\sqrt{z}}\mathcal{G}_1^2X+\frac{1}{\sqrt{z}}\mathcal{G}_1X\mathcal{G}_2)\mathcal{B}_4)_{ij}.$$

Finally, $\mathfrak{h}(0,0,1)$ and $\tilde{\mathfrak{h}}(0,0,1)$ can be estimated using a discussion similar to that of (S14). The details are omitted. □

The remaining part of this section is the proof of Lemma F.7, which is an analogue of Lemma F.3 and F.5 for the matrices A_2 and B_2.

PROOF OF LEMMA F.7. We shall outline our computation on the dominating terms. The discussions of the negligible terms are similar to those in Lemma F.3 and F.5, and are therefore omitted.

Recall (S2). Using the same proof of (S15), we first get

$$
\begin{aligned}
\mathfrak{h}(2,0,0) &= \frac{1}{n}\sum_{i,j}\frac{\partial^2 \mathbb{h}_1}{\partial x_{ij}^2}h_2h_3 = \frac{1}{n}\sum_{i,j}\Big(\frac{\partial^2 G}{\partial x_{ij}^2}A_2\Big)_{ij'}\mathfrak{q}^{(k-1)} \\
&= \frac{2z}{n}\sum_{i,j}G_{ii}G_{j'j'}(GA_2)_{ij'}\mathfrak{q}^{(k-1)} + O_{\prec}(n^{-\frac{1}{2}}) \\
&= \frac{2z}{n}\sum_{i,j}(\Pi_1)_{ii}(\Pi_1)_{j'j'}(\Pi_1A_2)_{ij'}\mathfrak{q}^{(k-1)} + O_{\prec}(n^{-\frac{1}{2}}) \\
&= \mathfrak{d}_2^a\mathfrak{q}^{(k-1)} + O_{\prec}(n^{-\frac{1}{2}}).
\end{aligned}
$$

Likewise, applying (S5), we find that

$$
\tilde{\mathfrak{h}}(2,0,0) = \frac{1}{n}\sum_{i,j}\frac{\partial^2 \tilde{\mathbb{h}}_1}{\partial x_{ij}^2}h_2h_3 = \frac{1}{n}\sum_{i,j}\Big(\frac{\partial^2 G^2}{\partial x_{ij}^2}B_2\Big)_{ij'}\mathfrak{q}^{(k-1)} = \tilde{\mathfrak{d}}_2\mathfrak{q}^{(k-1)} + O_{\prec}(n^{-\frac{1}{2}}).
$$

Next, recall (S28) and (S1). By a discussion similar to that of (S42), we conclude that

$$
\begin{aligned}
\mathfrak{h}(1,1,0) &= \frac{1}{n}\sum_{i,j}\frac{\partial \mathbb{h}_1}{\partial x_{ij}}\frac{\partial h_2}{\partial x_{ij}}h_3 \\
&= -\frac{(k-1)z}{\sqrt{n}}\sum_{i,j}(\Pi_1)_{ii}(\Pi_1A_2)_{j'j'}c_{ij}\mathfrak{q}^{(k-2)} + O_{\prec}(n^{-\frac{1}{2}+4\nu}) \\
&= \mathfrak{a}_{22}\mathfrak{q}^{(k-2)} + O_{\prec}(n^{-\frac{1}{2}+4\nu}).
\end{aligned}
$$

Similarly, recalling (S50), we have

$$
\tilde{\mathfrak{h}}(1,1,0) = \frac{1}{n}\sum_{i,j}\frac{\partial \tilde{\mathbb{h}}_1}{\partial x_{ij}}\frac{\partial h_2}{\partial x_{ij}}h_3 = \tilde{\mathfrak{b}}_{22}\mathfrak{q}^{(k-2)} + O_{\prec}(n^{-\frac{1}{2}}).
$$

Finally, recall (S4) and (S6). The same arguments as (S49) yield

$$\mathfrak{h}(1,2,0) = n^{-\frac{3}{2}} \sum_{i,j} \frac{\partial \mathbb{h}_1}{\partial x_{ij}} \frac{\partial^2 h_2}{\partial x_{ij}^2} h_3 = \mathfrak{a}_{32}\mathfrak{q}^{(k-2)} + O_{\prec}(n^{-\frac{1}{2}}),$$

$$\tilde{\mathfrak{h}}(1,2,0) = n^{-\frac{3}{2}} \sum_{i,j} \frac{\partial \tilde{\mathbb{h}}_1}{\partial x_{ij}} \frac{\partial^2 h_2}{\partial x_{ij}^2} h_3 = \tilde{\mathfrak{b}}_{32}\mathfrak{q}^{(k-2)} + O_{\prec}(n^{-\frac{1}{2}}).$$

This concludes our proof. □

G. Proof of Theorems 2.9. 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 $\mathcal{Q}$ defined below. After necessary notations are introduced, as done in the beginning of Section 5, it suffices to prove Proposition G.1, which is an analogue of Proposition 5.1. The proof of Proposition G.1 essentially relies on a recursive estimate presented in Proposition G.2. Thus the main goal of this section is to prove Proposition G.2.

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

$$\mathcal{Q} \equiv \mathcal{Q}(\boldsymbol{z}) := \sqrt{n} \sum_{\beta=1}^{r} \Big(\mathrm{Tr}(\Xi_1(z_\beta)) A_\beta^R + \mathrm{Tr}(\Xi_1'(z_\beta)) B_\beta^R \Big).$$

Denote the index set as

$$\mathcal{B}(\nu) := \bigcup_{\beta=1}^{r} \mathcal{B}_\beta(\nu),$$

where $\mathcal{B}_\beta(\nu)$ is defined as

$$\mathcal{B}_\beta(\nu) := \Big\{ (i,j) \in [M] \times [n] : |\boldsymbol{u}_\beta(i)| > n^{-\nu},\ |\boldsymbol{v}_\beta(j)| > n^{-\nu} \Big\}.$$

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

For $\beta \in [r]$, invoke $\Delta_d(z_\beta)$ by plugging z_β in (5.5). We also introduce the random variable

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

where $(c_\beta)_{ij} \equiv (c_\beta(z_\beta))_{ij}$ is defined by inserting z_β into c_{ij} in (5.7). Similarly, we denote $(s_\beta)_{ij} \equiv (s_\beta(z_\beta))_{ij}$ by plugging z_β into s_{ij} in (5.8). Let C_β and S_β be $M \times n$ matrices with entries $(c_\beta)_{ij}$ and $(s_\beta)_{ij}$ respectively. Denote

$$\Delta_r \equiv \Delta_r(\boldsymbol{z}) := \sum_{\beta=1}^{r} \Delta_r(z_\beta), \quad \Delta_d \equiv \Delta_d(\boldsymbol{z}) := \sum_{\beta=1}^{r} \Delta_d(z_\beta),$$

and $\Delta = \Delta_d + \Delta_r$. Furthermore, we denote

$$Q = \mathcal{Q} - \Delta. \tag{S1}$$

Note $\mathcal{V}^E(z_\beta)$ is defined in (5.10) by plugging z_β. Set

$$\mathcal{V}^E(\boldsymbol{z}) = \sum_{\beta=1}^{r} \mathcal{V}^E(z_\beta).$$

Then we define the function

$$\begin{aligned} V &\equiv V(\boldsymbol{z}) \\ &= \mathcal{V}^E(\boldsymbol{z}) + 2\frac{\kappa_3}{\sqrt{n}} \mathrm{Tr}\Big((\sum_{\beta=1}^{r} z_\beta S_\beta)(\sum_{\beta=1}^{r} \sqrt{z_\beta} C_\beta)^* \Big) + \frac{\kappa_4}{n} \mathrm{Tr}\Big((\sum_{\beta=1}^{r} z_\beta S_\beta)(\sum_{\beta=1}^{r} z_\beta S_\beta)^* \Big) \\ &\quad + \sum_{(i,j)\in \mathcal{S}(\nu)} \Big(\sum_{\beta=1}^{r} \sqrt{z_\beta} (c_\beta)_{ij} \Big)^2. \end{aligned}$$

Recall $p_\beta = p(d_\beta)$ in (2.1). Let $\boldsymbol{z}_0 := (p_1, \dots, p_r)$.

Proposition G.1. *Under the assumptions of Theorem 2.9, we have that $Q(\boldsymbol{z}_0)$ and $\Delta(\boldsymbol{z}_0)$ are asymptotically independent. Furthermore,*

$$Q(\boldsymbol{z}_0) \simeq \mathcal{N}\left(0, V(\boldsymbol{z}_0)\right).$$

Theorem 2.9 follows from Proposition G.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_β for $1 \le \beta \le r$ 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 G.1, it suffices to establish the following recursive estimates. It is an analogue of Proposition 5.2.

Proposition G.2. *Suppose the assumptions of Theorem 2.9 hold. Let $z_\beta = p_\beta + \mathrm{i}n^{-C}$ and $z_{0\beta} = p_\beta$ for all $\beta \in [r]$. We have*

$$\mathbb{E}Q(z_\beta)e^{\mathrm{i}t\Delta(z_{0\beta})} = O_\prec(n^{-1/2+\nu}),$$

and for any fixed integer $k \geq 2$,

$$\mathbb{E}Q^k(z_\beta)e^{\mathrm{i}t\Delta(z_{0\beta})} = (k-1)V\mathbb{E}Q^{k-2}(z_\beta)e^{\mathrm{i}t\Delta(z_{0\beta})} + O_\prec(n^{-1/2+\nu}).$$

H. Proof of Proposition G.2. Proposition G.2 can be proved in a way similar to Proposition 5.2. Recall from Section E that the proof of Proposition 5.2 is based on Lemma E.1 and Lemma E.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 (S1). Recall (S2) and for each $\beta \in [r]$, denote

$$A_{\beta,1} := A_\beta^R \mathtt{I}^{\mathrm{u}}, \quad A_{\beta,1} := A_\beta^R \mathtt{I}^{\mathrm{l}},$$
$$B_{\beta,1} := B_\beta^R \mathtt{I}^{\mathrm{u}}, \quad B_{\beta,1} := B_\beta^R \mathtt{I}^{\mathrm{l}}.$$

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

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

and

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

where $F_{\beta,\alpha}$ is defined in (S18) with $z = z_\beta$. Finally, for $\beta \in [r]$, we denote

$$Q_\beta := \sqrt{n}\sum_{\alpha=1,2}(f_{\beta,\alpha} + g_{\beta,\alpha}) + \sqrt{nz_\beta}\sum_{(i,j)\in\mathcal{S}(\nu)} x_{ij}(c_\beta)_{ij} - \Delta_d(z_\beta). \tag{S1}$$

Lemma H.1. *Under the assumptions of Proposition G.1, we have*

$$Q = \sum_{\beta=1}^r Q_\beta. \tag{S2}$$

Indeed, Lemma H.1 is the analogue of Lemma E.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 G.2, it suffices to study the following

$$
\begin{aligned}
\mathbb{E}Q^k e^{it\Delta} &= \sqrt{n}\sum_{\beta=1}^{r}\sum_{\alpha=1}^{2}\mathbb{E}\Big(f_{\beta,\alpha}+g_{\beta,\alpha}\Big)Q^{k-1}e^{it\Delta} \\
&\quad + \sqrt{n}\sum_{(i,j)\in\mathcal{S}(\nu)}\big(\sum_{\beta=1}^{r}\sqrt{z_\beta}(c_\beta)_{ij}\big)\mathbb{E}x_{ij}Q^{k-1}e^{it\Delta} - \Delta_d\mathbb{E}Q^{k-1}e^{it\Delta}. \qquad \text{(S3)}
\end{aligned}
$$

Step 2. In the second step, we will use the cumulant expansion to estimate the items on the right hand side of (S3) and prove the analogue of Lemma E.2.

Observe that for the rank r case, we have

$$
\frac{\partial Q}{\partial x_{ij}} = \sum_{\beta=1}^{r}\frac{\partial Q_\beta}{\partial x_{ij}}. \qquad \text{(S4)}
$$

The estimates of the cumulant expansion for the terms in (S3) follow along the exact lines of Lemma F.2-F.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^R_β and B^R_β in (4.26)) 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 (S13), we shall encounter an term of the following form

$$
\frac{1}{\sqrt{n}}\sum_{i,j}\Big(\Xi_1(z_\beta)A_{\beta,1}\Big)_{j'i}\frac{\partial Q}{\partial x_{ij}}Q^{k-2}e^{\mathrm{i}t\Delta} = \frac{1}{\sqrt{n}}\sum_{\gamma=1}^{r}\sum_{i,j}\Big(\Xi_1(z_\beta)A_{\beta,1}\Big)_{j'i}\frac{\partial Q_\gamma}{\partial x_{ij}}Q^{k-2}e^{\mathrm{i}t\Delta}.
$$

Applying (S1) for each $\partial Q_\gamma/\partial x_{ij}$, by (4.17) and orthogonality of the singular vectors, we find the only contributing part is

$$
\frac{1}{\sqrt{n}}\sum_{i,j}\Big(\Xi_1(z_\beta)A_{\beta,1}\Big)_{j'i}\frac{\partial Q_\beta}{\partial x_{ij}}Q^{k-2}e^{\mathrm{i}t\Delta}
$$

and what remains is exactly the same as the proof of (S13). 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 (S7). For $\beta \in [r]$, we denote $\mathfrak{d}^a_{\beta,\alpha}, \mathfrak{d}^b_{\beta,\alpha}, \tilde{\mathfrak{d}}_{\beta,\alpha}$ by replacing z with z_β and A_α, B_α with $A_{\beta,\alpha}, B_{\beta,\alpha}$ $(\alpha = 1, 2)$ correspondingly. We also define $\mathfrak{a}_{\beta,1\alpha}, \mathfrak{b}_{\beta,1\alpha}, \tilde{\mathfrak{b}}_{\beta,1\alpha}$ for $\alpha = 1, 2$ in the same fashion using (S9). Next, we denote

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

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

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

$$\mathfrak{a}_{\beta,31} := -\frac{2(k-1)z_\beta^{3/2}}{n} \sum_{i,j} \big(\Pi_1(z_\beta)\big)_{j'j'} \big(\Pi_1(z_\beta)A_{\beta,1}\big)_{ii} (\sum_{\gamma=1}^r z_\gamma S_\gamma)_{ij},$$

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

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

$$\mathfrak{a}_{\beta,0\alpha} := \mathfrak{a}_{\beta,1\alpha} + \kappa_3 \mathfrak{a}_{\beta,2\alpha} + \frac{\kappa_4}{2}\mathfrak{a}_{\beta,3\alpha},$$

$$\mathfrak{b}_{\beta,0\alpha} := \frac{m_\alpha(z_\beta)}{2}\tilde{\mathfrak{b}}_{\beta,1\alpha} + m'_\alpha(z_\beta)\mathfrak{b}_{\beta,1\alpha} + \frac{\kappa_3 m_\alpha(z_\beta)}{2}\tilde{\mathfrak{b}}_{\beta,2\alpha} + \kappa_3 m'_\alpha(z_\beta)\mathfrak{b}_{\beta,2\alpha} + \frac{\kappa_4 m_\alpha(z_\beta)}{4}\tilde{\mathfrak{b}}_{\beta,3\alpha} + \frac{\kappa_4 m'_\alpha(z_\beta)}{2}\mathfrak{b}_{\beta,3\alpha}.$$

We adopt the notation

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

With these preparations, we present the following analogue of Lemma E.2.

LEMMA H.2. *Under the assumptions of Proposition G.2, for each $\beta \in [r]$ and $\alpha = 1, 2$, we have*

(S5)
$$\sqrt{n}\mathbb{E} f_{\beta,\alpha}\mathfrak{q}^{(k-1)} = -\sqrt{z_\beta} m_\alpha(z_\beta)\mathbb{E}\big(\frac{\kappa_3}{2}\mathfrak{d}^a_{\beta,\alpha}\mathfrak{q}^{(k-1)} + \mathfrak{a}_{\beta,0\alpha}\mathfrak{q}^{(k-2)}\big) + O_\prec(n^{-\frac{1}{2}+4\nu}),$$
$$\sqrt{n}\mathbb{E} g_{\beta,\alpha}\mathfrak{q}^{(k-1)} = -\sqrt{z_\beta}\,\mathbb{E}\Big(\frac{\kappa_3}{4}\big(m_\alpha(z_\beta)\tilde{\mathfrak{d}}_{\beta,\alpha} + 2m'_\alpha(z_\beta)\mathfrak{d}^b_{\beta,\alpha}\big)\mathfrak{q}^{(k-1)} + \mathfrak{d}_{\beta,0\alpha}\mathfrak{q}^{(k-2)}\Big) + O_\prec(n^{-\frac{1}{2}+4\nu}).$$

In addition, we have

$$\sqrt{n}\sum_{(i,j)\in\mathcal{S}(\nu)}\big(\sum_{\beta=1}^r \sqrt{z_\beta}(c_\beta)_{ij}\mathbb{E} x_{ij}\mathfrak{q}^{(k-1)} = (k-1)\Big[\sum_{(i,j)\in\mathcal{S}(\nu)}\big(\sum_{\beta=1}^r\sqrt{z_\beta}(c_\beta)_{ij}\big)^2 + \frac{\kappa_3}{\sqrt{n}}\sum_{(i,j)\in\mathcal{S}(\nu)}\big(\sum_{\beta=1}^r (z_\beta s_\beta)_{ij}\big)\big(\sum_{\beta=1}^r\sqrt{z_\beta}(c_\beta)_{ij}\big)\Big]\mathbb{E}\mathfrak{q}^{(k-2)} + O_\prec(n^{-\frac{1}{2}+4\nu}).$$

Similar to the proof of Proposition 5.2, Proposition G.2 follows immediately from Lemma H.1 and H.2. We omit the details here.

Next, we turn to the proof of Lemma H.2. We will only focus our discussion on the term $\sqrt{n}\mathbb{E} f_{\beta,1}\mathfrak{q}^{(k-1)}$ and the other terms can be estimated likewise. Using a discussion similar to (S5), for each fixed $\beta \in [r]$, we have

$$\sqrt{n}\mathbb{E} f_{\beta,1}\mathfrak{q}^{(k-1)} = \mathbb{E}\Big(-m_1\sqrt{n z_\beta}\sum_{i,j} x_{ij}\big(\Xi_1(z_\beta)A_{\beta,1}\big)_{j'i} + \sqrt{n}F_1\operatorname{Tr}\big(G(z_\beta)A_{\beta,1}\big)\Big)\mathfrak{q}^{(k-1)}.$$

As seen in the proof of (S13), we need the following estimates which are analogues of those in Lemma F.2 and F.3. We adopt the notations in (S11) by denoting

$$h_1 = \big(\Xi_1(z_\beta)A_{\beta,1}\big)_{j'i}, \quad h_2 = Q^{k-1}, \quad h_3 = e^{\mathrm{i}t\Delta}.$$

LEMMA H.3. *For the derivatives of $h_1h_2h_3$, we have*

(S6) $$\hbar(1,0,0) = -\sqrt{n z_\beta} m_2(z_\beta)\operatorname{Tr}\big(G(z_\beta)A_{\beta,1}\big)\mathfrak{q}^{(k-1)} + O_\prec(n^{-1/2}),$$

(S7) $$\hbar(0,1,0) = \mathfrak{a}_{\beta,11}\mathfrak{q}^{(k-2)} + O_\prec(n^{-1/2}),$$
$$\hbar(2,0,0) = \mathfrak{d}^a_{\beta,1}\mathfrak{q}^{(k-1)} + O_\prec(n^{-1/2}),$$
$$\hbar(1,1,0) = \mathfrak{a}_{\beta,21}\mathfrak{q}^{(k-2)} + O_\prec(n^{-1/2}),$$
$$\hbar(1,2,0) = \mathfrak{a}_{\beta,31}\mathfrak{q}^{(k-2)} + O_\prec(n^{-1/2}).$$

Furthermore, all the other terms $\hbar(l_1,l_2,l_3)$ for $l_1+l_2+l_3 \le 4$ can be bounded by $O_\prec(n^{-1/4+4\nu})$.

It is easy to see that (S5) follows directly from Lemma H.3. Thus the final task is to prove Lemma H.3. In the proof, we will use the orthogonality of the singular vectors, that is, for $\beta_1 \neq \beta_2$,

$$\langle \boldsymbol{u}_{\beta_1}, \boldsymbol{u}_{\beta_2} \rangle = 0, \quad \langle \boldsymbol{v}_{\beta_1}, \boldsymbol{v}_{\beta_2} \rangle = 0. \tag{S8}$$

PROOF OF LEMMA H.3. First of all, (S6) can be estimated similarly as (S30). The other four dominating terms can be analyzed analogously and we shall only focus on the estimate of (S7). Observe that

$$\hbar(0,1,0) = \frac{1}{\sqrt{n}} \sum_{i,j} h_1 \frac{\partial h_2}{\partial x_{ij}} h_3 = \frac{(k-1)}{\sqrt{n}} \sum_{i,j} \big(\Xi_1(z_\beta) A_{\beta,1}\big)_{j'i} \frac{\partial Q}{\partial x_{ij}} \mathfrak{q}^{k-2}. \tag{S9}$$

Plugging in (S4), we have

$$\hbar(0,1,0) = \frac{(k-1)}{\sqrt{n}} \sum_{\gamma=1}^{r} \sum_{i,j} \big(\Xi_1(z_\beta) A_{\beta,1}\big)_{j'i} \frac{\partial Q_\gamma}{\partial x_{ij}} Q^{k-2} e^{\mathrm{i}t\Delta},$$

where by (S1),

$$\begin{aligned} \frac{\partial Q_\gamma}{\partial x_{ij}} = -\sqrt{n z_\gamma} \sum_{\substack{l_1, l_2 \in \{i,j'\} \\ l_1 \neq l_2}} \Big[\big(G(z_\beta) A_\gamma^R G(z_\beta)\big)_{l_1 l_2} - \frac{1}{2 z_\gamma} \big(G(z_\beta) B_\gamma^R G(z_\beta)\big)_{l_1 l_2} \\ + \frac{1}{2} \sum_{(a_1,a_2) \in \mathcal{P}(2,1)} \big(G^{a_1}(z_\beta) B_\gamma^R G^{a_2}(z_\beta)\big)_{l_1 l_2} \Big] - \mathbf{1}\Big((i,j) \in \mathcal{B}(\nu)\Big) \sqrt{n z_\gamma} (C_\gamma)_{ij}. \end{aligned} \tag{S10}$$

Using a discussion similar to (S13), we have that

$$\frac{(k-1)}{\sqrt{n}} \sum_{i,j} \big(\Xi_1(z_\beta) A_{\beta,1}\big)_{j'i} \frac{\partial Q_\beta}{\partial x_{ij}} \mathfrak{q}^{(k-2)} = \mathfrak{a}_{\beta,11} \mathfrak{q}^{(k-2)} + O_\prec(n^{-1/2}).$$

Therefore, it suffices to show that for $\gamma \neq \beta$,

$$\frac{1}{\sqrt{n}} \sum_{i,j} \big(\Xi_1(z_\beta) A_{\beta,1}\big)_{j'i} \frac{\partial Q_\gamma}{\partial x_{ij}} \mathfrak{q}^{(k-2)} = O_\prec(n^{-\frac{1}{2}}). \tag{S11}$$

To prove this, we shall argue in a similar way to (S13) by expanding the product above using (S10). We start with

$$\sum_{i,j} \big(\Xi_1(z_\beta) A_{\beta,1}\big)_{j'i} \big(G(z_\beta) A_\gamma^R G(z_\beta)\big)_{j'i}.$$

Recall (S34) and (S35). By (4.20) and (S8), we have

$$\begin{aligned}&\sum_{i,j}\big((\mathcal{G}_2(z_\beta)-m_2(z_\beta))\mathcal{A}_{\beta,3})_{j'i}(\mathcal{G}_2\mathcal{A}_{\gamma,3}\mathcal{G}_1)_{j'i}\\&=\omega_{\beta,3}\omega_{\gamma,3}\mathrm{Tr}\Big(\big(\mathcal{G}_2(z_\beta)-m_2(z_\beta)\big)\boldsymbol{v}_\beta\boldsymbol{u}_\beta^*\mathcal{G}_1(z_\beta)\boldsymbol{u}_\gamma\boldsymbol{v}_\gamma^*\mathcal{G}_2(z_\beta)\Big)\\&=\omega_{\beta,3}\omega_{\gamma,3}\Big(\boldsymbol{u}_\beta^*\mathcal{G}_1(z_\beta)\boldsymbol{u}_\gamma\Big)\Big(\boldsymbol{v}_\gamma^*\mathcal{G}_2\big(\mathcal{G}_2(z_\beta)-m_2(z_\beta)\big)\boldsymbol{v}_\beta\Big)=O_\prec(n^{-\frac{1}{2}}),\end{aligned}$$

where the coefficients $\omega_{\beta,3}$ are defined using the block decomposition of A_β^R as in (S29). We can estimate the other terms in the expansion (in light of (S34) and (S35)) using similar discussions. Hence, we conclude that

$$\sum_{i,j}\big(\Xi_1(z_\beta)A_{\beta,1}\big)_{j'i}\big(G(z_\beta)A_\gamma^RG(z_\beta)\big)_{j'i}=O_\prec(n^{-1/2}).$$

Likewise, we can show that each of the following terms

$$\sum_{i,j}\big(\Xi_1(z_\beta)A_{\beta,1}\big)_{j'i}\big(G(z_\beta)A_\gamma^RG(z_\beta)\big)_{ij'},\quad \sum_{i,j}\big(\Xi_1(z_\beta)A_{\beta,1}\big)_{j'i}\big(G(z_\beta)^2B_\gamma^RG(z_\beta)\big)_{j'i},$$
$$\sum_{i,j}\big(\Xi_1(z_\beta)A_{\beta,1}\big)_{j'i}\big(G^2(z_\beta)B_\gamma^RG(z_\beta)\big)_{ij'},\quad \sum_{i,j}\big(\Xi_1(z_\beta)A_{\beta,1}\big)_{j'i}\big(G(z_\beta)B_\gamma^RG^2(z_\beta)\big)_{j'i},$$
$$\sum_{i,j}\big(\Xi_1(z_\beta)A_{\beta,1}\big)_{j'i}\big(G(z_\beta)B_\gamma^RG^2(z_\beta)\big)_{ij'}$$

can be bounded by $O_\prec(n^{-1/2})$. In view of (S10), we conclude the proof of (S11). This completes our proof. □

References.

- [S1] Z. Bao, X. Ding, and K. Wang. Singular vector and singular subspace distribution for the matrix denoising model. *arXiv:1809.10476*, 2018.
- [S2] Z. Bao, X. Ding, J. Wang, and K. Wang. Principal components of spiked covariance matrices in the supercritical regime. *arXiv: 1907.12251*, 2019.
- [S3] X. Ding. High dimensional deformed rectangular matrices with applications in matrix denoising. *Bernoulli (in press)*, 2019.
- [S4] M. Gavish and D. Donoho. Optimal Shrinkage of Singular Values. *IEEE Trans. Inform. Theory*, 63(4): 2137–2152, 2017.
- [S5] J. Josse and S. Wager. Bootstrap-Based Regularization for Low-Rank Matrix Estimation. *J. Mach. Learn. Res.*, 17:1-29,2016.
- [S6] A. M. Khorunzhy, B. A. Khoruzhenko, and L. A. Pastur. Asymptotic properties of large random matrices with independent entries. *Journal of Mathematical Physics*, 37(10):5033–5060, 1996.
- [S7] A. Knowles and J. Yin. The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.*, 66(11): 1663–1750, 2013.

[S8] Y. Li and H. Li. Two-sample Test of Community Memberships of Weighted Stochastic Block Models. *arXiv preprint arXiv 1811.12593*, 2018.

[S9] A. Lytova and L. Pastur. Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.*, 37(5):1778–1840, 2009.

[S10] R. Nadakuditi. OptShrink: An Algorithm for Improved Low-Rank Signal Matrix Denoising by Optimal, Data-Driven Singular Value Shrinkage. *IEEE Trans. Inform. Theory*, 60(5):3002–3017, 2014.

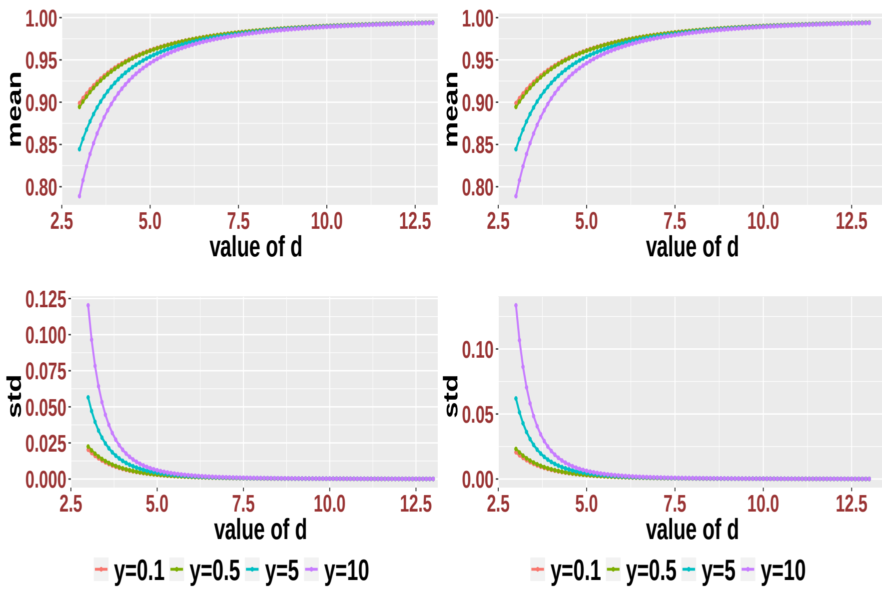
[S11] E. Richard, P. Savalle and N. Vayatis. Estimation of Simultaneously Sparse and Low Rank Matrices. *ICML'12 Proceedings of the 29th International Coference on International Conference on Machine Learning*, 51–58, 2012.

[S12] C. M. Stein. Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, 9(6):1135–1151, 1981.

Department of Mathematics,
Hong Kong University of Science and Technology,
Hong Kong
E-mail: mazgbao@ust.hk

Department of Statistical Sciences
University of Toronto
Canada
E-mail: xiucai.ding@mail.toronto.edu

Department of Mathematics,
Hong Kong University of Science and Technology,
Hong Kong
E-mail: kewang@ust.hk



(a) Gaussian noise (b) Two-Point noise

Fig S2: Mean-Variance Discussion. In both of the figures, we plot the mean function $a(d)$ 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 (2.3) and (2.4). The standard deviation is $\sqrt{4\theta(d)^2 + \mathcal{V}^E(d)}$ for the Gaussian noise and $\sqrt{4\theta(d)^2 + \mathcal{V}^E(d) + 4\sqrt{y}\theta(d)^2/(\sqrt{2}d) - 3y\theta(d)^2/(2d^2)}$ for the Two-Point noise. We choose the true right singular vector to be $\boldsymbol{f}_1$ and left singular vector to be $\mathbf{1}_M/\sqrt{M}$. Hence for the Two-Point noise, we need to add a part depending on $\kappa_3 = 1/\sqrt{2}$ and $\kappa_4 = -3/2$.