Geometric bias in eigenspace perturbation under random heterogeneous noise

Fengkai Liu[*] Ke Wang[†] Wanjie Wang[‡]

September 1, 2026

Abstract

Spectral methods rely fundamentally on the stability of principal eigenspaces under random perturbations. Classically, this stability is quantified by the Davis-Kahan and Wedin theorems, which bound the eigenspace error using the operator norm of the noise and the relevant spectral gaps. While these worst-case bounds are sharp for arbitrary deterministic perturbations, they can be wasteful in the low-rank signal-plus-random-noise setting, as they fail to capture the fine-grained interaction between the signal geometry and the noise distribution. In this paper, we study the spectral perturbation of signal-plus-noise matrices corrupted by sparse, random noise with an arbitrary, inhomogeneous variance profile. We demonstrate that under heterogeneous noise variances, the empirical eigenvectors suffer a systematic, deterministic geometric bias that is entirely invisible to classical perturbation bounds. By leveraging the Quadratic Vector Equation (QVE) and establishing fine-grained isotropic local laws, we derive near-optimal, non-asymptotic perturbation bounds for the leading eigenspaces in the operator and $2 \to \infty$ norms. The bounds separate the usual signal-to-noise contribution, stochastic fluctuations, and structured geometric bias terms determined by the alignment between the signal eigenspaces and the row-wise variance profile. We further develop refined rowwise bounds that adapt to the variance-weighted leverage of the signal space, leading to sharper guarantees in delocalized regimes.

As applications, we establish strong consistency of adjacency spectral clustering for degree-corrected stochastic block models with heterogeneous degree parameters and unbalanced communities, recovering the logarithmic expected-degree scale in the regular balanced case. We also study spectral embedding for generalized random dot product graphs, showing that the full signal embedding admits sharp rowwise control, whereas spectral truncation can retain a systematic geometric bias determined by the omitted signal directions and the heterogeneous variance profile.

Contents

1 Introduction and main results **3**
1.1 Top-k eigenspace perturbation 5
1.2 Refined perturbation bounds and consequences 9

[*]Department of Mathematics, Hong Kong University of Science and Technology. Email: fliuar@connect.ust.hk
[†]Department of Mathematics, Hong Kong University of Science and Technology. Email: kewang@ust.hk
[‡]Department of Statistics and Data Science, National University of Singapore. Email: wanjie.wang@nus.edu.sg

2 Applications **11**
- 2.1 Strong consistency for the degree-corrected stochastic block model 11
- 2.2 Generalized random dot product graphs 16

3 Geometric bias, sharpness, and debiasing **20**
- 3.1 Interpretation of the geometric bias 20
- 3.2 Sharpness and numerical experiments 22
- 3.3 Oracle de-biasing 23

4 Related literature **24**

5 Deterministic QVE estimates and preliminary bounds **26**

6 Proof strategy and technical inputs **28**
- 6.1 Isotropic local laws 28
- 6.2 Outlier eigenvalue location 30
- 6.3 Proof roadmap 31

7 Proof of the main perturbation bounds: Theorem 1 **31**

8 Proof of the isotropic local law: Theorem 4 **38**

9 Proofs of the outlier eigenvalue locations: Theorems 5 and 6 **40**
- 9.1 Proof of Theorem 5 40
  - 9.1.1 Proof of Lemma 9.3 45
  - 9.1.2 Proof of Lemma 9.2 49
- 9.2 Proof of Theorem 6 54

A Proofs of Lemmas 5.1, Lemma 5.2, Proposition 5.1, and Lemma 5.4 **55**
- A.1 Proof of Lemma 5.1 55
- A.2 Proof of Lemma 5.2 58
- A.3 Proof of Proposition 5.1 61
- A.4 Proof of Lemma 5.4 64

B Proofs of Proposition 7.1 and Proposition 7.2 **65**
- B.1 Proof of Proposition 7.1 65
- B.2 Proof of Proposition 7.2 67

C Proof of Proposition 8.1 **70**
- C.1 Proof of Lemma C.2 76

D Proof of Proposition 8.2 **95**

E Proofs of Corollary 3 and Corollary 4 **97**
- E.1 Proof of Corollary 3 97
- E.2 Proof of Corollary 4 99

F Extensions to general signal eigenspaces **103**

G Extension to rectangular matrices: singular subspace perturbations **105**

H Proofs of Lemma 3.1, Proposition 3.1, (J.1) and (J.2) **107**
H.1 Proof of Lemma 3.1 107
H.2 Proof of Proposition 3.1 108
H.3 Proof of (J.1) and (J.2) 111

I Proofs of Propositions 1.1 and 1.2 **112**
I.1 Proof of Proposition 1.1 112
I.2 Proof of Proposition 1.2 113

J Plug-in de-biasing implementation and limitations **117**

K Proofs for the applications **118**
K.1 Proofs for the DCSBM application 118
K.2 Proofs for the GRDPG application 121

1 Introduction and main results

Spectral methods provide a simple and powerful way to extract low-dimensional structure from high-dimensional data. They underlie principal component methods, low-rank matrix denoising and approximation, spectral clustering and community detection, latent-position estimation in networks, and many other procedures in modern statistics and data science. In these problems, the statistical object of interest is often encoded by a small number of dominant eigenvalues and eigenvectors of a population matrix, while the available data produce only a noisy version of that matrix. Understanding the stability of the leading spectral structure is therefore a basic step in understanding the performance of the resulting statistical procedure.

We study the signal-plus-noise model

$$\widetilde{A} = A + E,$$

where $A \in \mathbb{R}^{n \times n}$ is a deterministic symmetric matrix of low rank and E is a symmetric random perturbation with independent centered entries up to symmetry. Our main tasks are to understand the locations of the empirical outlier eigenvalues and the perturbation of the corresponding signal eigenspaces, both in the usual operator-norm geometry and in the $2 \to \infty$ norm that measures the largest rowwise error. We also study a closely related object, the weighted spectral embeddings, that arises naturally in applications.

Classical perturbation results such as Weyl's inequality and the Davis–Kahan–Wedin $\sin \Theta$ theorem control these quantities through the operator norm of the perturbation and the relevant spectral gaps. Such bounds are sharp for arbitrary deterministic perturbations, but they can be conservative when the perturbation is random. A substantial literature has therefore developed stochastic perturbation bounds that exploit independence, concentration, leave-one-out arguments, and resolvent methods; see Section 4 for a detailed discussion. The present paper addresses a feature that becomes essential when the noise is heterogeneous: the variance profile does not merely change the size of the random fluctuations, but can also change the geometry of the population signal space.

More specifically, heterogeneous noise produces a systematic second-order effect governed by its row-variance profile. As a consequence, two population signal directions that are orthogonal in the usual Euclidean geometry need not remain orthogonal in the variance-weighted geometry induced by the perturbation. This creates a systematic mixing of signal directions, which we call the *geometric bias*. The phenomenon is invisible to a perturbation bound based only on $\|E\|$. It also disappears in the homogeneous case, where all row variances are equal, but can be of leading order for a genuinely inhomogeneous variance profile.

Our first contribution is a non-asymptotic perturbation theory that retains the geometry induced by the heterogeneous variance profile. Using the Quadratic Vector Equation (QVE) together with isotropic and compressed local laws, we separate the eigenspace error into a deterministic geometric bias, stochastic mixing within the signal space, and leakage into the population null space. For truncated eigenspaces, the omitted signal directions interact with the variance profile and can produce a systematic error amplified by a small within-signal spectral gap.

Our second contribution is a row-adaptive refinement of the $2 \to \infty$ bound. A leave-one-out local-law argument replaces the global null-space fluctuation scale by a variance-weighted leverage scale tailored to the signal eigenvectors, while leaving the signal-space mixing terms unchanged. In the full signal eigenspace, the geometric-bias and within-signal mixing terms disappear. The same refinement also yields sharper bounds for weighted spectral embeddings, which are used in our network applications.

We illustrate the two sides of the theory through two statistical applications: for the degree-corrected stochastic block model, where the full signal eigenspace is retained, the row-adaptive bound yields strong consistency under heterogeneous degree parameters and unbalanced communities, reducing to the logarithmic expected-degree scale in the regular balanced case; for generalized random dot product graphs, we instead focus on the choice of embedding dimension and show that truncation introduces the geometric bias predicted by our theory through the interaction of the omitted signal directions, signed spectral gaps, and heterogeneous variance profile.

Roadmap. The remainder of this section states the precise top-k eigenspace perturbation theorem and its row-adaptive refinements and consequences. Section 2 develops the two statistical applications. Section 3 explains the geometric-bias mechanism through a rank-two model and the QVE expansion, and includes numerical experiments and an oracle debiasing procedure. Section 4 discusses the related literature. The technical arguments are developed in Sections 5–9: we first derive the QVE expansion responsible for the variance-profile bias, then establish the isotropic, compressed, and row-adaptive local-law inputs and the outlier eigenvalue locations, and finally combine these estimates with blockwise eigenvector equations. Additional proofs and extensions are collected in the appendices.

Notation. For a vector $v = (v_1, \dots, v_n)^\top \in \mathbb{R}^n$, we let $\|v\|$ denote its Euclidean ℓ_2 norm and $\|v\|_\infty := \max_i |v_i|$ denote its maximum absolute entry. The oscillation of a vector v is denoted by $\operatorname{osc}(v) := \max_i v_i - \min_i v_i$. For a matrix B, we define its off-diagonal part as $\operatorname{Off}(B) := B - \operatorname{diag}(B)$, where $\operatorname{diag}(B)$ is the diagonal matrix formed by the diagonal entries of B. We use various matrix norms throughout the paper. Let $\|B\|$ denote the operator norm, and let $\|B\|_F$ denote the Frobenius norm. The $2 \to \infty$ norm of a matrix is defined as $\|B\|_{2,\infty} := \max_i \|e_i^\top B\|$, which is the maximum Euclidean norm of the rows of B. The infinity operator norm is denoted by $\|B\|_{\infty\to\infty} := \max_i \sum_j |B_{ij}|$. We let $\mathbb{O}(k)$ denote the set of $k \times k$ orthogonal matrices. For two non-negative sequences X_n and Y_n, we write $X_n \lesssim Y_n$ (or $Y_n \gtrsim X_n$) if there exists an absolute constant $c > 0$ such that $X_n \leqslant cY_n$ for all sufficiently large n. We write $X_n \asymp Y_n$ if both $X_n \lesssim Y_n$ and $X_n \gtrsim Y_n$ hold. Throughout the paper, C, C' denote positive absolute constants whose exact values may change from line to line.

1.1 Top-k eigenspace perturbation

We now introduce the precise model notation and assumptions. Write the rank-r eigendecomposition of the signal as

$$A = U\Lambda U^\top = \sum_{i=1}^{r} \lambda_i u_i u_i^\top,$$

and order the nonzero eigenvalues algebraically. Let

$$r_+ := \#\{j : \lambda_j > 0\}, \qquad r_- := r - r_+,$$

so that

$$\lambda_1 \geqslant \cdots \geqslant \lambda_{r_+} > 0 > \lambda_{r_+ + 1} \geqslant \cdots \geqslant \lambda_r.$$

The ordered eigenvalues and their corresponding eigenvectors of $\widetilde{A}$ are denoted as $\widetilde{\lambda}_i, \widetilde{u}_i$ for $1 \leqslant i \leqslant n$. In the main text we focus on the right-edge outliers and the top-k ($1 \leqslant k \leqslant r_+$) positive signal eigenspace and its perturbed counterpart

$$U_k := (u_1, \ldots, u_k), \qquad \widetilde{U}_k := (\widetilde{u}_1, \ldots, \widetilde{u}_k).$$

We write $\Lambda_k := \operatorname{diag}(\lambda_1, \ldots, \lambda_k)$ and define $\widetilde{\Lambda}_k$ analogously. For an index set $J \subseteq [r]$, U_J denotes the submatrix of U formed by the columns indexed by J. The negative-spike case follows by applying the same argument to $-A$; see Remark 1.3.

Throughout the main results, fix $D > 0$. We impose the following assumptions on random noise E:

ASSUMPTION 1.1 (Structural assumption on the noise)**.** *Let E be an $n \times n$ symmetric random matrix with independent centered entries (up to symmetry) and variance profile $\Sigma = (\sigma_{ij}^2)_{1 \leqslant i,j \leqslant n}$ where $\mathbb{E}(E_{ij}^2) = \sigma_{ij}^2$. Assume there exists a parameter $K \geqslant 1$ such that the sub-Gaussian norm* of each entry satisfies:*

$$\|E_{ij}\|_{\psi_2} \leqslant K\sigma_{ij}.$$

Equivalently, we can write $E_{ij} = \sigma_{ij}\xi_{ij}$ where the random variables ξ_{ij} have mean 0, variance 1, and satisfy $\max_{i,j} \|\xi_{ij}\|_{\psi_2} \leqslant K$.

Let $\sigma_{\max} := \max_{i,j} \sigma_{ij} > 0$. We further assume that there exists a parameter $\beta \geqslant 1$ such that

$$\mathbb{E}|E_{ij}|^3 \leqslant \beta K \sigma_{\max}\, \sigma_{ij}^2, \qquad \mathbb{E}|E_{ij}|^4 \leqslant \beta^2 K^2 \sigma_{\max}^2\, \sigma_{ij}^2. \tag{1.1}$$

*The sub-Gaussian norm of a random variable X is defined as $\|X\|_{\psi_2} := \inf\{t > 0 : \mathbb{E}(\exp(X^2/t^2)) \leqslant 2\}$. In particular, $\mathbb{P}(|X| \geqslant t) \leqslant 2\exp(-t^2/\|X\|_{\psi_2}^2)$. See [93] for details.

REMARK 1.1 (The role of β)**.** *The parameter β separates the third- and fourth-moment controls from the sub-Gaussian norm K and is useful in sparse regimes. Under the sub-Gaussian assumption alone, one may always take $\beta \lesssim \sqrt{\log(eK)}$, but this can introduce an artificial logarithmic loss. For example, for centered Bernoulli noise with edge probability p, one has $K \lesssim p^{-1/2}$, whereas the third and fourth moments satisfy the sharper bounds in* (1.1) *with $\beta = O(1)$. Thus keeping β explicit allows our results to retain the correct logarithmic scale in sparse network applications.*

Let us denote $R_i := \sum_{j=1}^n \sigma_{ij}^2$. Define the diagonal matrix

$$\mathcal{R} := \operatorname{diag}(R_1, \dots, R_n)$$

and the *geometric bias matrix*

$$\mathcal{V} := U^\top \mathcal{R} U = (u_i^\top \mathcal{R} u_j)_{1 \leqslant i,j \leqslant r}. \tag{1.2}$$

To capture the heterogeneity of the noise profile, we define the *variance oscillation* parameter:

$$\operatorname{osc}(\mathcal{R}) = R_{\max} - R_{\min},$$

where

$$R_{\max} := \max_i R_i \qquad \text{and} \qquad R_{\min} := \min_i R_i.$$

Assumption 1.1 is imposed throughout. In addition, we work under one of the following two auxiliary regimes.

ASSUMPTION 1.2 (Auxiliary regularity conditions)**.** *Fix $D > 0$. We assume that one of the following conditions holds.*

(i) General sub-Gaussian regime:

$$\sqrt{R_{\max}} \geqslant C_0(D+8)K\sigma_{\max} \log n.$$

(ii) Bounded-entry refinement: *If there exists an absolute constant $c_0 \geqslant 1$ such that* $\max_{i,j} |E_{ij}| \leqslant c_0 K \sigma_{\max}$ *almost surely, then we only require*

$$\sqrt{R_{\max}} \geqslant C_0(D+8)K\sigma_{\max}\sqrt{\log n}.$$

Here $C_0 > 0$ is a sufficiently large absolute constant.

The second regime is a refinement of the first noise model rather than a separate distributional assumption: in both cases Assumption 1.1 holds, while the bounded-entry condition in (ii) allows the variance condition to be weakened by a factor $\sqrt{\log n}$. Note that the bounded-entry condition implies (1.1) with $\beta \leqslant c_0$.

REMARK 1.2 (On the Assumption 1.2)**.** *The regularity condition prevents any single maximal entry from dominating its row sum. Under this condition, our estimate (see Lemma 5.4) yields $\|E\| \leqslant 3\sqrt{R_{\max}}$ with high probability, thereby guaranteeing a macroscopic signal-to-noise separation ($\lambda_k \geqslant 3\|E\|$) for any supercritical spike $\lambda_k \geqslant 9\sqrt{R_{\max}}$. Both assumptions naturally accommodate sparse regimes. For example, for centered Bernoulli noise with parameter p, Assumption 1.2 requires $p \gtrsim \log n/n$.*

For technical convenience, throughout the main results we impose

$$\lambda_1 \leqslant R_{\max}^3, \qquad R_{\max} \leqslant n^{100}, \qquad \sigma_{\max}^{-1} \leqslant n^{100}.$$

These conditions only restrict the range in which we carry out the local-law analysis and are not essential to the perturbation phenomenon. The polynomial bounds are extremely mild; the exponent 100 is arbitrary and may be replaced by any sufficiently large fixed constant. In particular, they are automatic in the network models considered below, where $R_{\max} \leqslant n$.

The upper bound $\lambda_1 \leqslant R_{\max}^3$ merely excludes an ultra-strong signal regime. When $|z| \gg \|E\|$, the resolvent admits a direct Neumann-series expansion, and the corresponding perturbation bounds follow by simpler arguments. We therefore exclude this regime to avoid a separate treatment; see [98] for related arguments without such an upper restriction.

Define the *spectral gap*

$$\delta_k := \begin{cases} \lambda_k - \lambda_{k+1}, & \text{if } k < r_+; \\ \lambda_k - 3\sqrt{R_{\max}}, & \text{if } k = r_+. \end{cases}$$

With the convention that $\delta_0 := +\infty$.

We define the regime-dependent fluctuation scale

$$M_D := \begin{cases} C'(D+6)^{3/2} r\beta K \sigma_{\max} (\log n)^2, & \text{in the sparse sub-Gaussian regime,} \\ C'(D+6) r c_0 K \sigma_{\max} \log n, & \text{in the sparse bounded-entry regime,} \end{cases} \tag{1.3}$$

where $C' > 0$ is a sufficiently large absolute constant.

Furthermore, for fixed $k \in [r_+]$, define index sets

$$\mathcal{J} := \{k+1, \ldots, r_+\}, \quad \mathcal{I} := [r] \backslash \mathcal{J},$$
$$\mathcal{N} := \{r_+ + 1, \ldots, r\}, \quad \mathcal{K} := [r] \backslash \mathcal{N},$$

and denote

$$\mathcal{B}_k := \frac{10\sqrt{k}}{\delta_k \lambda_k} \|\mathcal{V}_{\mathcal{J}\mathcal{I}}\| + \frac{4\sqrt{k}}{\lambda_k^2} \|\mathcal{V}_{\mathcal{N}\mathcal{K}}\| + \frac{80\sqrt{k}}{\delta_k \lambda_k} \cdot \frac{R_{\max}}{\lambda_k^2} \operatorname{osc}(\mathcal{R}), \tag{1.4}$$

where $\mathcal{V}_{\mathcal{J}\mathcal{I}} = U_{\mathcal{J}}^\top \mathcal{R} U_{\mathcal{I}}$ and $\mathcal{V}_{\mathcal{N}\mathcal{K}} = U_{\mathcal{N}}^\top \mathcal{R} U_{\mathcal{K}}$. If $k = r_+$ (thus $\mathcal{J} = \varnothing$), then $\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\| = 0$.

We also set

$$\rho_k \equiv \rho_{k,D} := 10 M_D + 15 \frac{\operatorname{osc}(\mathcal{R})}{\lambda_k}.$$

THEOREM 1 (Top-k eigenspace perturbations)**.** *Fix $D > 0$. Under Assumptions 1.1 and 1.2, for any $k \in [r_+]$ such that $\lambda_k \geqslant 9\sqrt{R_{\max}} + 100\rho_{k,D}$ and*

$$\delta_k \geqslant \sqrt{k} \left(65 \rho_{k,D} + 20 \frac{\operatorname{osc}(\mathcal{R})}{\lambda_k} \right),$$

we have the following bounds with probability at least $1 - n^{-D}$:

(i) *For the operator norm bound, we have*

$$\|\sin\angle(U_k,\widetilde{U}_k)\| \leqslant \mathcal{B}_k + 10\sqrt{k}\frac{M_D}{\delta_k} + 2\frac{\|E\|}{\lambda_k}.$$

(ii) *For the* $2\to\infty$ *norm bound, we first have*

$$\|\widetilde{U}_k - P_{U_k}\widetilde{U}_k\|_{2,\infty} \leqslant \|U\|_{2,\infty}\left(\mathcal{B}_k + 10\sqrt{k}\frac{M_D}{\delta_k} + 3\sqrt{k}\frac{\operatorname{osc}(\mathcal{R})}{\lambda_k^2}\right) + 3\sqrt{k}\frac{M_D}{\lambda_k}.$$

Consequently,

$$\begin{aligned}\min_{O\in\mathbb{O}(k)}\|\widetilde{U}_k - U_kO\|_{2,\infty} &\leqslant \|\widetilde{U}_k - P_{U_k}\widetilde{U}_k\|_{2,\infty} + \|U_k\|_{2,\infty}\|\sin\angle(U_k,\widetilde{U}_k)\|^2,\\ &\leqslant 4\|U\|_{2,\infty}\left(\mathcal{B}_k + 20\sqrt{k}\frac{M_D}{\delta_k} + 3\sqrt{k}\frac{\operatorname{osc}(\mathcal{R})}{\lambda_k^2} + 27\frac{R_{\max}}{\lambda_k^2}\right) + 3\sqrt{k}\frac{M_D}{\lambda_k}.\end{aligned}$$

Theorem 1 separates three distinct mechanisms in the eigenspace perturbation. The first is the geometric-bias term $\mathcal{B}_k$, which comes from mixing between the target eigenspace and the non-target signal directions. Its leading components are governed by

$$\mathcal{V}_{\mathcal{J}\mathcal{I}} = U_{\mathcal{J}}^\top\mathcal{R}\,U_{\mathcal{I}}, \qquad \mathcal{V}_{\mathcal{N}\mathcal{K}} = U_{\mathcal{N}}^\top\mathcal{R}\,U_{\mathcal{K}},$$

and quantify how the signal eigenspaces interact with the row-variance profile $\mathcal{R} = \operatorname{diag}(R_i)$. These couplings vanish when $\mathcal{R} = cI$. Indeed, by orthogonality,

$$\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\| = \|U_{\mathcal{J}}^\top(\mathcal{R} - cI)\,U_{\mathcal{I}}\| \leqslant \frac{1}{2}\operatorname{osc}(\mathcal{R})$$

by choosing $c = \frac{1}{2}(R_{\max} + R_{\min})$. Likewise, $\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\| \leqslant \frac{1}{2}\operatorname{osc}(\mathcal{R})$.

The term M_D/δ_k is the stochastic contribution associated with mixing between the target eigenspace and the remaining signal directions. It therefore appears in both the principal-angle and $2\to\infty$ norm bounds. By contrast, the standalone term M_D/λ_k in the $2\to\infty$ norm bound comes from the component of the empirical eigenspace in the null space of A. For the principal-angle loss, this rowwise null-space leakage can be controlled at the coarser operator-norm scale $\|E\|/\lambda_k$.

REMARK 1.3 (Bounds for general eigenspaces)**.** *Theorem 1 is stated for the top positive eigenspace only to simplify notation. Analogous bounds hold for bottom and intermediate signal eigenspaces by symmetry and elementary projector decompositions. See Appendix F for details.*

REMARK 1.4 (Extension to rectangular matrices)**.** *Our eigenspace perturbation results extend to singular subspace perturbation for rectangular matrices via the standard Hermitian dilation technique. See Appendix G for a brief discussion.*

1.2 Refined perturbation bounds and consequences

Although Theorem 1 focuses on the operator-norm principal angle and the $2 \to \infty$ loss, its proof actually controls the empirical eigenspace in both the non-target signal directions and the null space of A. Following the framework of [98], these estimates can be extended to other error metrics, including Frobenius and unitarily invariant norms, linear forms, max norms, and weighted rowwise losses. We only record the extensions needed for the applications below.

In our heterogeneous setting, the same decomposition also allows us to refine one term in the $2 \to \infty$ bound of Theorem 1. The term M_D/δ_k comes from stochastic mixing with the non-target signal directions and is unchanged by the refinement below. In contrast, the standalone term M_D/λ_k arises from the null-space action and can be sharpened using the row-adaptive local law from Section 6.1, at the cost of a slightly stronger signal-strength condition. This refinement is important in the applications, where it allows us to reach the logarithmic-degree regime.

Before stating the refined bounds, we record a stronger variance condition required by the leave-one-out argument.

ASSUMPTION 1.3 (Row-adaptive variance condition)**.** *For fixed $D > 0$, we assume*

$$\sqrt{R_{\max}} \geqslant \begin{cases} C_1 K \sigma_{\max}(D+r)\log n, & \textit{in the sparse sub-Gaussian regime}, \\ C_1 K \sigma_{\max}\sqrt{(D+r)\log n}, & \textit{in the sparse bounded-entry regime}, \end{cases}$$

where $C_1 > 0$ is a sufficiently large absolute constant.

We further define the regime-dependent entry scale

$$L_{D,r}^{\mathrm{ra}} := \begin{cases} K\sigma_{\max}\sqrt{(D+r)\log n}, & \text{in the sparse sub-Gaussian regime}, \\ K\sigma_{\max}, & \text{in the sparse bounded-entry regime}. \end{cases} \tag{1.5}$$

In the sub-Gaussian regime, $L_{D,r}^{\mathrm{ra}}$ is the truncation level used in the leave-one-out argument, whereas in the bounded-entry regime the deterministic entry bound is used directly.

Denote

$$\mathfrak{R}_D(U) := \sqrt{(D+r)\log n}\,\Gamma(U) + \left((D+r)\log n\, L_{D,r}^{\mathrm{ra}} + M_D\right)\|U\|_{2,\infty}, \tag{1.6}$$

where

$$\Gamma(U) := \max_i \Big(\sum_j \sigma_{ij}^2 \|e_j^\top U\|^2\Big)^{1/2}.$$

PROPOSITION 1.1 (Refined rowwise top-k perturbation)**.** *Fix $D > 0$ and $1 \leqslant k < r_+$. Suppose Assumptions 1.1 and 1.3 hold. Assume*

$$\lambda_k \geqslant C\sqrt{(D+r)R_{\max}\log n} \quad \textit{and} \quad \delta_k \geqslant C\sqrt{k}\left(M_D + \frac{\operatorname{osc}(\mathcal{R})}{\lambda_k}\right).$$

Then, with probability at least $1 - n^{-D}$,

$$\|(I - P_{U_k})\widetilde{U}_k\|_{2,\infty} \leqslant \|U\|_{2,\infty}\left(\mathcal{B}_k + 10\sqrt{k}\frac{M_D}{\delta_k} + 3\sqrt{k}\frac{\operatorname{osc}(\mathcal{R})}{\lambda_k^2}\right) + C'\sqrt{k}\frac{\mathfrak{R}_D(U)}{\lambda_k}.$$

Consequently,

$$\min_{O\in\mathbb{O}(k)} \|\widetilde{U}_k - U_k O\|_{2,\infty} \leqslant C'\|U\|_{2,\infty}\left(\mathcal{B}_k + \sqrt{k}\frac{M_D}{\delta_k} + \sqrt{k}\frac{\operatorname{osc}(\mathcal{R})}{\lambda_k^2} + \frac{R_{\max}}{\lambda_k^2}\right) + C'\sqrt{k}\frac{\mathfrak{R}_D(U)}{\lambda_k}.$$

Compared with Theorem 1, the only change is that the null-space term of order M_D/λ_k is replaced by $\mathfrak{R}_D(U)/\lambda_k$. The latter can be far smaller for delocalized signal spaces: unlike the global scale M_D, $\mathfrak{R}_D(U)$ depends on the variance-weighted row leverage $\Gamma(U)$, with its remaining terms scaled by $\|U\|_{2,\infty}$. For instance, for a fixed-rank incoherent signal with $\|U\|_{2,\infty} \asymp n^{-1/2}$ and $R_i \asymp d$, in the sparse bounded-entry regime,

$$\mathfrak{R}_D(U) \lesssim \sqrt{\frac{d\log n}{n}} + \frac{\log n}{\sqrt{n}},$$

whereas M_D is of order $\log n$. Thus the refinement can be much sharper in the delocalized regimes relevant to the network applications below.

Next, we present a clean refinement for the full signal eigenspace, where the signal-space mixing terms disappear. This is a direct consequence of Proposition 1.1.

COROLLARY 1 (Full signal eigenspace)**.** *Suppose $A \succeq 0$ has rank $r \geqslant 1$. Fix $D > 0$. Suppose Assumptions 1.1 and 1.3 hold. Assume*

$$\lambda_r \geqslant C\left(\sqrt{(D+r)R_{\max}\log n} + M_D\right).$$

Then, with probability at least $1 - n^{-D}$,

$$\min_{O\in\mathbb{O}(r)} \|\widetilde{U}_r - UO\|_{2,\infty} \leqslant C'\sqrt{r}\,\frac{\mathfrak{R}_D(U)}{\lambda_r} + C'\|U\|_{2,\infty}\frac{\sqrt{r}\operatorname{osc}(\mathcal{R}) + R_{\max}}{\lambda_r^2}.$$

The same row-adaptive refinement carries over to the weighted spectral embedding given below. For $k \in [r_+]$, define the population and empirical rank-k spectral truncations

$$A_k := U_k\Lambda_k U_k^\top, \qquad \widetilde{A}_k := \widetilde{U}_k\widetilde{\Lambda}_k\widetilde{U}_k^\top.$$

We write

$$A = A_k + A_{>k}, \qquad A_{>k} := U_{>k}\Lambda_{>k}U_{>k}^\top = \sum_{s=k+1}^{r}\lambda_s u_s u_s^\top,$$

where $U_{>k} := (u_{k+1},\dots,u_r)$ and $\Lambda_{>k} := \operatorname{diag}(\lambda_{k+1},\dots,\lambda_r)$, with the convention $A_{>r} = 0$. We also introduce the weighted spectral embeddings

$$Y_k := U_k\Lambda_k^{1/2}, \qquad \widetilde{Y}_k := \widetilde{U}_k\widetilde{\Lambda}_k^{1/2}.$$

For later use, set

$$\tau_k := \left(\sum_{s=1}^{k}\lambda_s\right)^{1/2}$$

and

$$\eta_k := \|U_k^\top E U_k\| + \left(\|\Lambda_{>k}\| + \|E\|\right)\left(\mathcal{B}_k + 10\sqrt{k}\frac{M_D}{\delta_k} + 2\frac{\|E\|}{\lambda_k}\right)^2.$$

PROPOSITION 1.2 (Refined weighted spectral embedding). *Under the assumptions of Proposition 1.1, with probability at least* $1 - n^{-D}$,

$$\min_{O \in \mathbb{O}(k)} \|\widetilde{Y}_k O - Y_k\|_{2,\infty} \leqslant \|U\|_{2,\infty} \left(\sqrt{\lambda_k}\, \mathcal{B}_k \mathbf{1}_{\{k<r\}} + 10 \frac{M_D \tau_k}{\delta_k} \mathbf{1}_{\{k<r\}} + 3\sqrt{k} \frac{\operatorname{osc}(\mathcal{R})}{\lambda_k^{3/2}} \right) + C\sqrt{k} \frac{\mathfrak{R}_D(U)}{\sqrt{\lambda_k}} + \frac{\|U_k\|_{2,\infty}}{\sqrt{\lambda_k}} \eta_k.$$

For the full signal embedding, the corresponding simplification is obtained directly by setting the non-target signal space to be empty; we use this form in the proof of the network application below. The proofs of the above results are given in Appendix I.

2 Applications

2.1 Strong consistency for the degree-corrected stochastic block model

The degree-corrected stochastic block model (DCSBM) extends the stochastic block model with vertex-specific degree parameters [70]. Let $r \geqslant 2$ be the number of communities, let $g : [n] \to [r]$ be the community assignment, and let $Z \in \{0,1\}^{n \times r}$ be the corresponding membership matrix. We allow $r = r_n$.

Let $\Theta := \operatorname{diag}(\theta_1, \ldots, \theta_n)$ with $\theta_i > 0$ contain the vertex-specific degree parameters, and let $B \in \mathbb{R}^{r \times r}$ be a symmetric block-connectivity matrix. Conditional on the deterministic probability matrix

$$P := \Theta Z B Z^\top \Theta,$$

the observed adjacency matrix $\widetilde{A} = (\widetilde{A}_{ij})_{i,j=1}^n$ satisfies $\widetilde{A}_{ij} \sim \operatorname{Bernoulli}(P_{ij})$ for $i \leqslant j$ independently up to symmetry. Throughout, every community is nonempty, $0 \leqslant P_{ij} \leqslant 1$, and B has full rank r. Write

$$\widetilde{A} = P + E, \qquad P = U \Lambda U^\top, \qquad E := \widetilde{A} - P.$$

We next introduce the quantities that describe community imbalance and degree heterogeneity. For $a \in [r]$, set the community size and community weighted mass parameters as

$$n_a := \#\{i : g(i) = a\}, \qquad s_a := \sum_{i : g(i) = a} \theta_i^2, \qquad s_- := \min_{a \in [r]} s_a.$$

Define the diagonal matrix

$$S := Z^\top \Theta^2 Z = \operatorname{diag}(s_1, \ldots, s_r)$$

and the population row scales

$$\mu_i := \frac{\theta_i}{\sqrt{s_{g(i)}}}, \qquad \mu_- := \min_{i \in [n]} \mu_i, \qquad \mu_+ := \max_{i \in [n]} \mu_i.$$

Let

$$d_i := \sum_{j=1}^n P_{ij}, \qquad d_{\max} := \max_{i \in [n]} d_i, \qquad p_{\max} := \max_{i,j} P_{ij}.$$

In our setting,

$$R_i := \sum_{j=1}^{n} P_{ij}(1-P_{ij}), \qquad R_{\max} := \max_i R_i, \qquad \mathcal{R} := \operatorname{diag}(R_1,\dots,R_n),$$

and define the variance-weighted leverage scale

$$\Gamma_{\mathrm{dc}} := \max_{i\in[n]} \left(\sum_{j=1}^{n} P_{ij}(1-P_{ij})\mu_j^2\right)^{1/2}.$$

Finally, denote the extreme singular values of B as $\sigma_{\min}(B) = \min_{1\leqslant j\leqslant r} |\lambda_j(B)|$ and $\sigma_{\max}(B) = \|B\|$. Note that the nonzero eigenvalues of P coincide with those of $S^{1/2}BS^{1/2}$. To see this, set $X_0 := \Theta Z S^{-1/2}$. Thus $X_0^\top X_0 = I_r$ and $P = X_0(S^{1/2}BS^{1/2})X_0^\top$. Hence, the smallest population signal of P is denoted by

$$\tau_n := \min_{1\leqslant j\leqslant r} |\lambda_j(P)| = \sigma_{\min}\left(S^{1/2}BS^{1/2}\right).$$

In particular, we have

$$\tau_n \geqslant \sigma_{\min}(B)s_-. \tag{2.1}$$

ASSUMPTION 2.1 (Heterogeneous DCSBM). *The matrix $B = B_n$ is symmetric, entrywise nonnegative, and nonsingular. Moreover, the graph is sparse in the sense that $p_{\max} = o(1)$.*

Assumption 2.1 places no lower bound on the community proportions and no comparability condition on the degree parameters. It also allows $\sigma_{\min}(B)$ and $\sigma_{\max}(B)$ to vary with n and r. In particular, $\sigma_{\max}(B)$ may grow with r.

Assumption 2.1 allows the number of communities r_n to increase with n. As n increases, it is natural to have more communities. However, the discussion about theoretical guarantee of community detection with an increasing r_n is quite limited. On SBM, [17] provides the average degree requirement for nontrivial community recovery, and [109] establishes a minimax theory for both strong consistency and weak consistency. However, both works do not involve degree heterogeneity and unbalanced community sizes.

Spectral clustering algorithm. Our goal is to recover the unknown labels $g(1),\dots,g(n)$ from $\widetilde{A}$. We use the same raw-adjacency eigenspace and spherical normalization as in Lei and Rinaldo [74]. To accommodate highly unbalanced communities, Algorithm 1 combines this embedding with the farthest-first traversal of Gonzalez [58], followed by nearest-center assignment. Unlike a global r-means objective, the final step depends only on geometric separation and does not weight a community in proportion to its size.

To see why the algorithm works, the population eigenvector matrix satisfies

$$e_i^\top U = \mu_i e_{g(i)}^\top V$$

for some $V \in \mathbb{O}(r)$; see (K.1) in Appendix K.1. Row normalization therefore removes the vertex-specific factor μ_i and maps all vertices in community a to the same population center $e_a^\top V$. These r centers have pairwise distance $\sqrt{2}$ (see Lemma K.1).

Algorithm 1 Farthest-first spectral clustering for the DCSBM

Input: adjacency matrix $\widetilde{A} \in \mathbb{R}^{n\times n}$ and number of communities r.
Output: estimated labels $\widetilde{\mathbf{g}} \in [r]^n$.
Step 1. Compute the r eigenvectors of $\widetilde{A}$ corresponding to its r eigenvalues with largest absolute values, and collect them as $\widetilde{U}_r \in \mathbb{R}^{n\times r}$.
Step 2. Normalize the rows: $\widetilde{z}_i := e_i^\top \widetilde{U}_r / \|e_i^\top \widetilde{U}_r\|$, $i \in [n]$.
Step 3. Choose any $i_1 \in [n]$. For $\ell = 2, \ldots, r$, choose $i_\ell \in \operatorname{argmax}_{i\in[n]} \min_{1\leqslant s<\ell} \|\widetilde{z}_i - \widetilde{z}_{i_s}\|$.
Step 4. Assign each vertex to its nearest selected center: $\widetilde{g}(i) \in \operatorname{argmin}_{1\leqslant \ell\leqslant r} \|\widetilde{z}_i - \widetilde{z}_{i_\ell}\|$.

A uniform rowwise perturbation of size at most a sufficiently small multiple of μ_- preserves this separation after row normalization. The farthest-first step then selects one representative from each community, regardless of the community sizes, and nearest-center assignment recovers every label exactly.

For fixed $D > 0$, set

$$t_{D,r} := (D + r)\log n$$

and define the error scale

$$\varepsilon_{D,n}(\tau_n) := \sqrt{r}\,\frac{\sqrt{t_{D,r}}\,\Gamma_{\mathrm{dc}} + t_{D,r}\mu_+}{\tau_n} + (\sqrt{r}+1)\mu_+\frac{d_{\max}}{\tau_n^2}. \tag{2.2}$$

Since $\tau_n \geqslant \sigma_{\min}(B)s_-$ as in (2.1), we get $\varepsilon_{D,n}(\tau_n) \leqslant \varepsilon_{D,n}(\sigma_{\min}(B)s_-)$.

THEOREM 2 (Strong consistency for heterogeneous DCSBMs)**.** *Suppose Assumption 2.1 holds. For every fixed $D > 0$, there exist constants $C_{1,D}, C_{2,D}, c_D > 0$, independent of B, such that, if*

$$d_{\max} \geqslant C_{1,D} t_{D,r} \qquad \text{and} \qquad \tau_n \geqslant C_{1,D}\sqrt{t_{D,r} d_{\max}}, \tag{2.3}$$

then, with probability at least $1 - n^{-D}$,

$$\min_{O\in\mathbb{O}(r)} \|\widetilde{U}_r O - U\|_{2,\infty} \leqslant C_{2,D}\cdot \varepsilon_{D,n}(\tau_n). \tag{2.4}$$

If, in addition,

$$\varepsilon_{D,n}(\tau_n) \leqslant c_D \mu_-, \tag{2.5}$$

then Algorithm 1 exactly recovers the communities up to a permutation of their labels.

REMARK 2.1. *Since $\tau_n \geqslant \sigma_{\min}(B)s_-$, the signal condition in* (2.3) *may be replaced by*

$$\sigma_{\min}(B)s_- \geqslant C_{1,D}\sqrt{t_{D,r} d_{\max}},$$

which is stronger but directly verifiable. Accordingly, the right-hand side of (2.4) *may be replaced by $C_{2,D}\cdot\varepsilon_{D,n}(\sigma_{\min}(B)s_-)$, and exact recovery follows whenever $\varepsilon_{D,n}(\sigma_{\min}(B)s_-) \leqslant c_D\mu_-$.*

Theorem 2 does not impose uniform bounds on the singular values of B, which is useful when the number of communities $r = r_n$ grows and the spectral scale of B may vary with r. The dependence on r is retained explicitly through $t_{D,r} = (D + r)\log n$ and the quantities τ_n, $d_{\max}$, and Γ_{dc}. Thus the admissible growth of r depends on the accompanying signal and sparsity scales. For example, if $r \asymp \log n$, then $t_{D,r} \asymp (\log n)^2$ and the signal condition becomes

$$d_{\max} \gtrsim_D \log^2 n, \qquad \tau_n \gtrsim_D \sqrt{d_{\max}}\, \log n + (\log n)^2.$$

When r is fixed, [90] establishes strong consistency when the expected degree exceeds $C \log n$. When r increases, the required degree also grows with r, reflecting the need for a denser network to achieve exact community recovery. We next consider fixed r, where these quantities admit a simpler interpretation in terms of community imbalance and the degree scale.

The following corollary gives an explicit singular-value formulation when the degree parameters remain comparable but the community sizes may be highly unbalanced. Assume that r is fixed and that, for constants $c_\theta, C_\theta > 0$ and a scale $\theta_0 = \theta_0(n) > 0$,

$$c_\theta \theta_0 \leqslant \theta_i \leqslant C_\theta \theta_0 \qquad \text{for all } i \in [n]. \tag{2.6}$$

We also assume $\theta_0^2 \sigma_{\max}(B) = o(1)$. Since

$$p_{\max} = \max_{i,j} P_{ij} = \max_{i,j} \theta_i \theta_j B_{g(i),g(j)} \leqslant C_\theta^2 \theta_0^2 \max_{a,b} B_{ab} \leqslant C_\theta^2 \theta_0^2 \sigma_{\max}(B),$$

this implies that the condition $p_{\max} = o(1)$ in Assumption 2.1 is satisfied. Let

$$m_n := \min_{a \in [r]} n_a, \qquad N_n := \max_{a \in [r]} n_a, \qquad \kappa_n := \frac{N_n}{m_n}, \qquad d_- := m_n \theta_0^2.$$

Thus, $\mu_- \asymp 1/\sqrt{N_n}$ and $s_- \asymp m_n \theta_0^2 = d_-$.

COROLLARY 2 (Unbalanced communities with a varying block matrix)**.** *Suppose Assumption 2.1 and* (2.6) *hold, and suppose r is fixed. For every fixed $D > 0$, there exist constants $C_{1,D}, C_{2,D} > 0$, depending only on D, r, c_θ, and C_θ, such that, if*

$$d_- = m_n \theta_0^2 \geqslant C_{1,D} \frac{\sigma_{\max}(B)}{\sigma_{\min}(B)^2} \left(\kappa_n \log n + \kappa_n^{3/2} \right), \tag{2.7}$$

then, with probability at least $1 - n^{-D}$,

$$\min_{O \in \mathbb{O}(r)} \|\widetilde{U}_r O - U\|_{2,\infty} \leqslant \frac{C_{2,D}}{\sqrt{N_n}} \left(\frac{\sqrt{\sigma_{\max}(B)}}{\sigma_{\min}(B)} \sqrt{\frac{\kappa_n \log n}{d_-}} + \frac{\sigma_{\max}(B)}{\sigma_{\min}(B)^2} \frac{\kappa_n^{3/2}}{d_-} \right). \tag{2.8}$$

Moreover, $C_{1,D}$ can be chosen sufficiently large so that Algorithm 1 exactly recovers the communities up to a permutation of their labels.

The corollary is particularly clear in the regular case. If the communities are balanced, so that $m_n \asymp N_n \asymp n$ and $\kappa_n \asymp 1$, and $\sigma_{\max}(B) \asymp \sigma_{\min}(B) \asymp 1$, then Corollary 2 reduces to the logarithmic degree condition

$$n\theta_0^2 \gtrsim_D \log n.$$

Thus the spectral algorithm achieves exact recovery down to the optimal logarithmic expected-degree regime (see [1]). More generally, if $B = c_n B_0$ for a fixed well-conditioned matrix B_0, then $\sigma_{\max}(B) \asymp \sigma_{\min}(B) c_n$, and the condition becomes

$$c_n n \theta_0^2 \gtrsim_D \log n,$$

again corresponding to logarithmic expected degree. In this case, the leading term in (2.8) is

$$n^{-1/2}\sqrt{\frac{\log n}{c_n n\theta_0^2}}.$$

For unbalanced communities, condition (2.7) further quantifies the losses due to the weakest block direction, through $\sigma_{\min}(B)$, and the increased noise associated with $\sigma_{\max}(B)$ and community imbalance. The proofs of Theorem 2 and Corollary 2 are deferred to Appendix K.1.

Numerical simulations. We consider the DCSBM

$$P = \theta_0^2 \Theta Z B Z^\top \Theta, \qquad \theta_i \overset{\text{iid}}{\sim} \mathrm{Unif}(0.8, 1.2),$$

where θ_0^2 denotes the overall network density. We consider $r = 5$ communities with a severe imbalance level. For each setting, we report mean Adjusted Rand Index (ARI) [68] with one-standard-deviation bands over 100 repetitions. A larger ARI indicates a better community detection result. We apply our Algorithm 1 and the classical spectral community detection algorithm in [74], where the only difference is at the k-means step. To avoid extreme values in empirical analysis, we take the center of 10 farthest data points in Step 3. The results are shown in Figure 1.

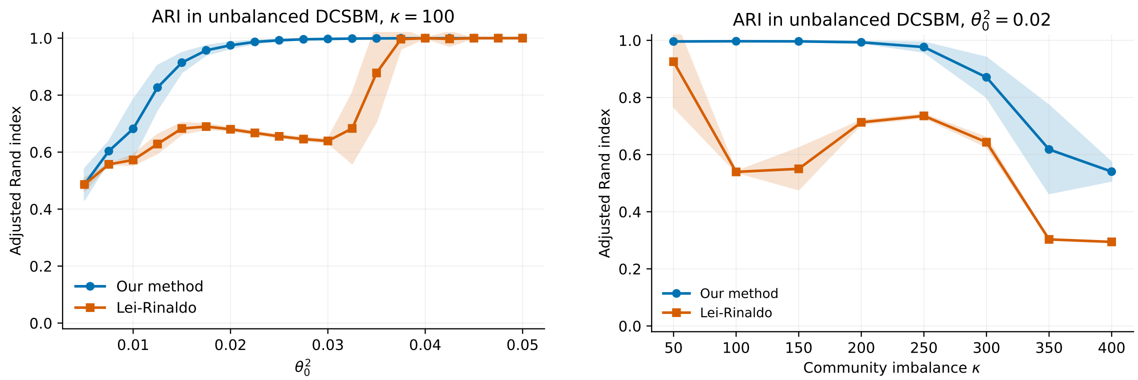
In the first experiment, the community sizes are $(8000, 3000, 800, 300, 80)$, so $\kappa = 100$, with $\mathrm{diag}(B) = (1, 1, 3, 8, 20)$ and off-diagonals of B as 0.2. The diagonal entries mean that smaller communities usually have denser connections. As θ_0^2 increases, our proposed method recovers the communities substantially earlier than the standard spectral community detection method in [74], where the main difference is the k-means step. In sufficiently dense networks, both methods perform well.

In the second experiment, $n = 12000$, $\theta_0^2 = 0.02$, and the community sizes follow $n_k = n\frac{\kappa^{-(k-1)/4}}{\sum_{j=1}^5 \kappa^{-(j-1)/4}}$, $k = 1, 2, 3, 4, 5$, so that $n_1/n_5 = \kappa$. We take $\kappa = 50, \ldots, 400$. For each κ, we set the diagonals of B matrix as

$$B_{kk} = C_k \frac{n_1}{n_k}, \qquad C = \left(1, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right).$$

Hence, as κ increases, the smallest community n_5 enjoys that the same order of $n_5 B_{55}$. Our proposed method remains nearly perfect through $\kappa = 250$ and consistently outperforms the classical spectral community detection method as imbalance increases. When κ exceeds 250, B_{55} cannot increase since the probability must be cut at 1. Hence, the performance declines due to the shrunk signal strength.

Relation to earlier DCSBM results. Lei and Rinaldo [74] established weak consistency for spherical adjacency spectral clustering under the DCSBM. Related spectral procedures include regularized clustering [88], SCORE normalization [69], and normalized-adjacency methods [60]. Lei and Zhu [75] subsequently obtained exact recovery at logarithmic degree by combining a consistent initializer with sample splitting and cross-clustering; their main presentation assumes community sizes proportional to n. Su, Wang, and Zhang [90] establish strong consistency for regularized spectral clustering under a different normalized-Laplacian framework for balanced



(a) Fixed $\kappa = 100$, varying θ_0^2.

(b) Fixed $\theta_0^2 = 0.02$, varying κ.

Figure 1: ARI under unbalanced DCSBMs. The curves report the mean over 100 Monte Carlo repetitions, with shaded regions representing one standard deviation. The proposed method is more robust to community-size imbalance than classical spectral clustering.

communities. When $r \to \infty$, existing theory is much more limited and is primarily available for balanced SBMs; see, for example, [17, 109]. Theorem 2 instead analyzes the raw adjacency eigenspace directly and allows the number of communities $r = r_n$ to increase with n. Its guarantees retain arbitrary imbalance and degree heterogeneity through s_-, $\mu_\pm$, and Γ_{dc}. Corollary 2 provides an explicit imbalance-dependent guarantee without sample splitting or a local refinement step.

2.2 Generalized random dot product graphs

The generalized random dot product graph (GRDPG) provides a broad latent-position framework for network models, including stochastic block models and their variants; see [89, 11]. In particular, degree heterogeneity can be incorporated through vertex-specific scalings of the latent positions, so the DCSBM considered in Section 2.1 may be viewed as a structured special case. Since the preceding application already treats heterogeneous degree parameters and unbalanced communities, here we impose a regular latent design and focus on a different question: the effect of the embedding dimension.

When the full signal dimension is retained, adjacency spectral embedding estimates the latent positions up to the natural non-identifiability of the GRDPG. In practice, however, one may embed into a smaller dimension $k < r$, either because the signal rank is unknown or because a lower-dimensional representation is desired. Our perturbation results show that such truncation can introduce a systematic geometric bias when the variance profile is heterogeneous.

Consider a GRDPG with deterministic latent positions $y_1, \dots, y_n \in \mathbb{R}^r$. Let

$$J := \mathrm{diag}(I_p, -I_q), \qquad p + q = r,$$

and write $Y = (y_1, \dots, y_n)^\top \in \mathbb{R}^{n \times r}$. Conditional on the latent positions, the edges are independent with

$$\widetilde{A}_{ij} \sim \mathrm{Bernoulli}(P_{ij}), \qquad P_{ij} = \rho_n y_i^\top J y_j, \qquad i < j,$$

where ρ_n controls the network density. We allow independent self-loops for theoretical simplicity. The population adjacency matrix is

$$P := \mathbb{E}\widetilde{A} = \rho_n Y J Y^\top.$$

The matrix J specifies the signature of the latent inner product, allowing P to have both positive and negative eigenvalues; the ordinary random dot product graph corresponds to $q = 0$. The primary inferential goal is to recover the latent positions Y from $\widetilde{A}$.

Define

$$d_n := n\rho_n, \qquad \Delta_n := \frac{1}{n} Y^\top Y, \qquad \bar{y}_n := \frac{1}{n}\sum_{i=1}^n y_i.$$

We work first under a regular-design condition.

ASSUMPTION 2.2 (Regular GRDPG). *There are fixed constants $c, C > 0$ such that*

$$\max_i \|y_i\| \leqslant C, \qquad cI_r \preceq \Delta_n \preceq CI_r, \qquad c \leqslant y_i^\top J \bar{y}_n \leqslant C, \quad i \in [n],$$

and

$$0 \leqslant P_{ij} = \rho_n y_i^\top J y_j \leqslant 1 - c, \qquad i, j \in [n]. \tag{2.9}$$

In Assumption 2.2, the bound $\max_i \|y_i\| \leqslant C$ keeps individual latent positions bounded, while $cI_r \preceq \Delta_n \preceq CI_r$ ensures they span all r directions at a nondegenerate scale. Since $\sum_{j=1}^n P_{ij} = d_n\, y_i^\top J \bar{y}_n$, the quantity $y_i^\top J \bar{y}_n$ is proportional to the expected degree of vertex i, so the bounds $c \leqslant y_i^\top J \bar{y}_n \leqslant C$ force all expected degrees to be of order $d_n = n\rho_n$. Finally, (2.9) keeps all edge probabilities valid and uniformly below one. Together, these conditions make the nonzero eigenvalues of P of order d_n while allowing the model to remain sparse as $\rho_n \to 0$.

Let

$$P = U \Lambda U^\top, \qquad \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_r)$$

be its rank-r eigendecomposition, with the positive eigenvalues listed first and the negative eigenvalues last, each ordered by decreasing magnitude. Set

$$X := U|\Lambda|^{1/2}, \qquad |\Lambda| := \operatorname{diag}(|\lambda_1|, \ldots, |\lambda_r|).$$

Then $XJX^\top = P$. Since $P = \rho_n Y J Y^\top$, there exists $Q_0 \in \mathbb{O}(p, q) := \{Q \in \mathbb{R}^{r \times r} : Q^\top J Q = J\}$ such that

$$X Q_0 = \sqrt{\rho_n}\, Y. \tag{2.10}$$

Thus X represents the latent positions Y up to a scaling and an indefinite orthogonal transformation.

For the observed adjacency matrix $\widetilde{A}$, select its k eigenvalues of largest magnitude and reorder the selected eigenpairs so that the positive eigenvalues precede the negative ones, with decreasing magnitude within each sign class. Denote the resulting eigenvalue matrix by

$$\widetilde{\Lambda}_k := \operatorname{diag}(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_k).$$

Denote $|\widetilde{\Lambda}_k| = \operatorname{diag}(|\widetilde{\lambda}_1|, \ldots, |\widetilde{\lambda}_k|)$. The k-dimensional adjacency spectral embedding is

$$\widetilde{X}_k := \widetilde{U}_k |\widetilde{\Lambda}_k|^{1/2}.$$

When $k = r$, the selected eigenpairs are the r empirical eigenvalues of largest magnitude, ordered by the same sign convention as the population eigenvalues. In this case, $\widetilde{X}_r$ estimates X up to a block-orthogonal transformation in $\mathbb{O}(p) \oplus \mathbb{O}(q)$, and hence the scaled latent positions $\sqrt{\rho_n} Y$ up to the intrinsic $\mathbb{O}(p, q)$ nonidentifiability.

In practice, however, the latent dimension r is typically unknown, so the embedding dimension k must be selected from the data. The statistical behaviour depends qualitatively on whether $k = r$, $k < r$, or $k > r$. The choice of the embedding dimension is itself a model-selection problem. Existing methods include network cross-validation [43] and spectral procedures that retain eigenvalues above an estimated noise floor [65]. Our purpose here is complementary. Conditional on a chosen embedding dimension, we quantify the rowwise error of the resulting embedding and determine how the heterogeneous Bernoulli variances distort its population target.

Full embedding. We first consider the correctly specified case $k = r$, where the adjacency spectral embedding targets the full latent position matrix, omitting no signal directions. The next theorem gives a finite-sample $2 \to \infty$ error bound for this embedding and, consequently, a uniform latent-position recovery guarantee down to the logarithmic expected-degree regime. The logarithmic-degree threshold follows from Proposition 1.2; we defer the proof to Appendix K.2.

THEOREM 3 (Full GRDPG spectral embedding)**.** *Suppose Assumption 2.2 holds. For every fixed $D > 0$, there are constants $C_{1,D}, C_{2,D} > 0$ such that, if*

$$d_n \geqslant C_{1,D} \log n,$$

then, with probability at least $1 - n^{-D}$, the r empirical eigenvalues of largest magnitude consist of p positive and q negative signal outliers, and

$$\min_{W \in \mathbb{O}(p) \oplus \mathbb{O}(q)} \|\widetilde{X}_r W - X\|_{2,\infty} \leqslant C_{2,D} \sqrt{\frac{\log n}{n}}, \tag{2.11}$$

$$\inf_{Q \in \mathbb{O}(p,q)} \|\rho_n^{-1/2} \widetilde{X}_r Q - Y\|_{2,\infty} \leqslant C_{2,D} \sqrt{\frac{\log n}{d_n}}. \tag{2.12}$$

Consequently, the rescaled adjacency spectral embedding is uniformly consistent whenever $d_n / \log n \to \infty$.

Theorem 3 strengthens existing uniform recovery guarantees for the GRDPG. Rubin-Delanchy, Cape, Tang, and Priebe [89] established uniform consistency and asymptotic normality under an i.i.d. latent-position model and a polylogarithmic expected-degree condition, i.e., $d_n = \omega\big(\log^{4c} n\big)$ for $c > 1$. In contrast, Theorem 3 is finite-sample, allows a deterministic latent design, and is valid down to the logarithmic degree regime $d_n \gtrsim \log n$. Moreover, the bound $\sqrt{\log n / n}$ in (2.11) has the same order as the minimax lower bound of Yan and Levin [103] in the bounded-condition-number regime. Thus the usual adjacency spectral embedding attains the optimal rowwise scale, up to constants, at essentially the weakest degree scale at which uniform recovery can be expected.

Truncated embeddings and geometric bias. The situation is qualitatively different when $k < r$. Under-embedding necessarily loses information about the omitted signal directions, but it can also distort the retained coordinates themselves. Our interest is in this second effect: heterogeneous noise variances induce a systematic geometric bias through the interaction between retained and omitted signal directions. This is complementary to the recent work of Taing and Levin [91], which studies dimension misspecification relative to the full latent-position target. Here we instead compare $\widetilde{X}_k$ with the truncated population spectral factor X_k and quantify the additional bias within the retained signal space.

Write

$$P = U_+\Lambda_+U_+^\top - U_-\Lambda_-U_-^\top,$$

where

$$\Lambda_+ = \operatorname{diag}(\lambda_1^+, \dots, \lambda_p^+), \qquad \Lambda_- = \operatorname{diag}(\lambda_1^-, \dots, \lambda_q^-),$$

and both sequences are decreasing and positive.

Fix $0 \leqslant k_+ \leqslant p$ and $0 \leqslant k_- \leqslant q$ with $k := k_+ + k_- < r$, and define

$$U_k := (U_{+,k_+}, U_{-,k_-}), \qquad \Lambda_k := \operatorname{diag}(\Lambda_{+,k_+}, -\Lambda_{-,k_-}), \qquad X_k := U_k|\Lambda_k|^{1/2},$$

where

$$\Lambda_{+,k_+} := \operatorname{diag}(\lambda_1^+, \dots, \lambda_{k_+}^+), \qquad \Lambda_{-,k_-} := \operatorname{diag}(\lambda_1^-, \dots, \lambda_{k_-}^-),$$

and empty blocks are omitted. The empirical embedding $\widetilde{X}_k$ is defined analogously from the selected positive and negative empirical outliers.

For the centered adjacency noise $E := \widetilde{A} - P$, we have $R_i := \sum_{j=1}^n P_{ij}(1 - P_{ij})$ and $\mathcal{R} := \operatorname{diag}(R_1, \dots, R_n)$. Define the within-sign boundary gaps

$$\delta_{+,n} := \lambda_{k_+}^+ - \lambda_{k_++1}^+, \qquad \delta_{-,n} := \lambda_{k_-}^- - \lambda_{k_-+1}^-,$$

whenever $0 < k_+ < p$ and $0 < k_- < q$, respectively. We adopt the convention that $\delta_{+,n} = \infty$ if $k_+ \in \{0, p\}$ and $\delta_{-,n} = \infty$ if $k_- \in \{0, q\}$.

For $k_+ > 0$, set

$$\lambda_{+,*} := \lambda_{k_+}^+, \qquad \mathcal{V}_{+,1} := \left\|U_{+,>k_+}^\top \mathcal{R}(U_{+,k_+}, U_-)\right\|, \qquad \mathcal{V}_\pm := \|U_-^\top \mathcal{R} U_+\|,$$

where $\mathcal{V}_{+,1} := 0$ if $k_+ = p$, and define

$$\mathcal{B}_{k_+}^+ := \frac{10\sqrt{k_+}}{\delta_{+,n}\lambda_{+,*}} \left(\mathcal{V}_{+,1} + 8\frac{R_{\max}}{\lambda_{+,*}^2}\operatorname{osc}(\mathcal{R})\right) + 4\sqrt{k_+}\frac{\mathcal{V}_\pm}{\lambda_{+,*}^2}.$$

Set $\mathcal{B}_0^+ := 0$. Define $\lambda_{-,*}$ and $\mathcal{B}_{k_-}^-$ analogously by interchanging the positive and negative signal blocks, and set $\mathcal{B}_0^- := 0$.

Finally, define the geometric-bias contribution to the truncated embedding by

$$\operatorname{GBias}_{k_+,k_-} := \|U\|_{2,\infty}\left(\sqrt{\lambda_{+,*}}\,\mathcal{B}_{k_+}^+ + \sqrt{\lambda_{-,*}}\,\mathcal{B}_{k_-}^-\right), \tag{2.13}$$

where a term is omitted when the corresponding retained sign block is empty.

The leading components of (2.13) are

$$\|U\|_{2,\infty}\frac{\sqrt{k_+}\left\|U_{+,>k_+}^\top \mathcal{R} U_{+,k_+}\right\|}{\delta_{+,n}\sqrt{\lambda_{+,*}}}, \qquad \|U\|_{2,\infty}\frac{\sqrt{k_-}\left\|U_{-,>k_-}^\top \mathcal{R} U_{-,k_-}\right\|}{\delta_{-,n}\sqrt{\lambda_{-,*}}}.$$

They measure the failure of the retained and omitted signal directions to remain orthogonal under the variance-weighted inner product induced by $\mathcal{R}$. Now we present the truncated embedding result. Denote

$$\delta_n := \min\{\delta_{+,n}, \delta_{-,n}\},$$

PROPOSITION 2.1 (Truncated GRDPG embedding)**.** *Suppose Assumption 2.2 holds. For every fixed $D > 0$,there exist constants $C_{1,D}, C_{2,D} > 0$ such that , if*

$$d_n \geqslant C_{1,D} \log n, \quad \textit{and} \quad \delta_n \geqslant C_{1,D} \log n,$$

then, with probability at least $1 - n^{-D}$,

$$\min_{W \in \mathbb{O}(k_+) \oplus \mathbb{O}(k_-)} \|\widetilde{X}_k W - X_k\|_{2,\infty} \leqslant \mathrm{GBias}_{k_+,k_-} + C_{2,D} \left(\sqrt{\frac{\log n}{n}} + \sqrt{\frac{d_n}{n}} \frac{\log n}{\delta_n} \right). \tag{2.14}$$

Proposition 2.1 separates the systematic geometric bias from the stochastic rowwise error. The leading bias is governed by the variance-weighted couplings

$$U_{+,>k_+}^\top \mathcal{R} U_{+,1:k_+} \quad \text{and} \quad U_{-,>k_-}^\top \mathcal{R} U_{-,1:k_-},$$

and is amplified when the truncation boundary lies near another same-sign signal eigenvalue. Thus networks with comparable signal eigenvalues and noise levels may nevertheless exhibit different truncated-embedding errors because their variance profiles interact differently with the signal eigenspaces. In particular, the leading bias vanishes when $\mathcal{R} = cI$. This provides information complementary to methods that select eigenvalues above an estimated noise floor [65]: such methods determine which directions are distinguishable from noise, whereas our bound also quantifies the stability of truncating within the signal spectrum. In the well-separated regime $\delta_n \geqslant c d_n$, the geometric bias is of smaller order, and Proposition 2.1 simplifies to

$$\min_{W \in \mathbb{O}(k_+) \oplus \mathbb{O}(k_-)} \|\widetilde{X}_k W - X_k\|_{2,\infty} \lesssim \sqrt{\frac{\log n}{n}}. \tag{2.15}$$

A short verification of (2.15) is given in the appendix. The over-embedded case $k > r$ involves noise-bulk eigenvectors rather than signal outliers and is outside the scope of the present perturbation theory.

3 Geometric bias, sharpness, and debiasing

3.1 Interpretation of the geometric bias

The present variance-profile model introduces a new deterministic geometric bias term

$$\mathcal{B}_k = \frac{10\sqrt{k}}{\delta_k \lambda_k} \left(\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\| + 8 \frac{R_{\max}}{\lambda_k^2} \operatorname{osc}(\mathcal{R}) \right) + 4\sqrt{k} \frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\lambda_k^2},$$

which captures the interaction between the signal eigenspaces and the inhomogeneous row-variance profile $\mathcal{R} = \operatorname{diag}(R_i)$. In the homogeneous noise setting where $\mathcal{R} = cI$, the geometric bias term

$\mathcal{B}_k$ vanishes; our current bounds in Theorem 1 match the two-term bounds studied in [98], up to logarithmic factors and rank factors.

The two quantities $\mathcal{V}_{\mathcal{J}\mathcal{I}}, \mathcal{V}_{\mathcal{N}\mathcal{K}}$ play different roles in the bias term. The term $\mathcal{V}_{\mathcal{J}\mathcal{I}}$ captures how the target top-k cluster interacts with the lower positive cluster $\{k+1, \dots, r_+\}$, appearing at the gap-dependent scale $(\delta_k \lambda_k)^{-1}$. By contrast, $\mathcal{V}_{\mathcal{N}\mathcal{K}}$ captures interactions between negative and nonnegative signal directions, but at the weaker scale λ_k^{-2}. In practice, the contribution from $\mathcal{V}_{\mathcal{N}\mathcal{K}}$ is often negligible. For instance, when all spikes are nonnegative, we have $\mathcal{N} = \varnothing$ and this term vanishes. Even when negative spikes exist, the term involving $\mathcal{V}_{\mathcal{N}\mathcal{K}}$ is usually dominated by $\mathcal{V}_{\mathcal{J}\mathcal{I}}$ as it has a stronger suppression factor λ_k^{-2}.

We believe that the leading geometric bias terms

$$\frac{\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\|}{\delta_k \lambda_k}, \quad \frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\lambda_k^2}$$

are intrinsic to the variance-profile model.

A rank-two model illustration. To explain where these terms come from, consider a simple rank-2 model $A = \lambda_1 u_1 u_1^\top + \lambda_2 u_2 u_2^\top$ with $\lambda_1 > \lambda_2 > 0$ and the leading empirical eigenvector $\widetilde{u}_1$. Let Q be the orthogonal projection onto the null space of A. From the decomposition $\widetilde{u}_1 = u_1 u_1^\top \widetilde{u}_1 + u_2 u_2^\top \widetilde{u}_1 + Q\widetilde{u}_1$, taking the ℓ_2 norm on both sides and rearranging terms, we have

$$\sin^2 \angle(u_1, \widetilde{u}_1) = (u_2^\top \widetilde{u}_1)^2 + \|Q\widetilde{u}_1\|^2.$$

The term $u_2^\top \widetilde{u}_1$ captures the interaction between signal directions. In our proofs, this quantity is controlled by projecting the eigenvector equation $\widetilde{u}_1 = G(\widetilde{\lambda}_1)A\widetilde{u}_1$ onto u_2, where $G(z) := (zI - E)^{-1}$ is the resolvent of E. To precisely estimate this term, from the eigenvector equation $(A+E)\widetilde{u}_1 = \widetilde{\lambda}_1 \widetilde{u}_1$, we solve for

$$\widetilde{u}_1 = (\widetilde{\lambda}_1 - E)^{-1} A\widetilde{u}_1 = G(\widetilde{\lambda}_1)A\widetilde{u}_1.$$

A key technical input in our proof is to show that the random resolvent G can be precisely approximated by its QVE approximation, a deterministic matrix Φ (defined (5.1)). This Φ isolates the deterministic contribution of the variance profile. Indeed, from

$$u_2^\top \widetilde{u}_1 \approx u_2^\top \Phi(\widetilde{\lambda}_1) A\widetilde{u}_1 = \lambda_1 (u_2^\top \Phi(\widetilde{\lambda}_1) u_1) u_1^\top \widetilde{u}_1 + \lambda_2 (u_2^\top \Phi(\widetilde{\lambda}_1) u_2) u_2^\top \widetilde{u}_1,$$

plugging in the expansion of $\Phi \approx z^{-1} I + z^{-3} \mathcal{R}$ in (5.3), the leading off-diagonal term is

$$u_2^\top \Phi(\widetilde{\lambda}_1) u_1 \approx \frac{u_2^\top \mathcal{R} u_1}{\widetilde{\lambda}_1^3}.$$

Substituting this back and approximating $\widetilde{\lambda}_1 \approx \lambda_1$ and $u_1^\top \widetilde{u}_1 \approx 1$, we obtain

$$u_2^\top \widetilde{u}_1 \approx \frac{1}{\lambda_1^2}(u_2^\top \mathcal{R} u_1) + \frac{\lambda_2}{\lambda_1} u_2^\top \widetilde{u}_1.$$

Rearranging this equation to solve $u_2^\top \widetilde{u}_1$ yields

$$|u_2^\top \widetilde{u}_1| \approx \frac{|u_2^\top \mathcal{R} u_1|}{\delta_1 \lambda_1},$$

which mirrors the $\frac{\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\|}{\delta_k\lambda_k}$ term in our theorem. If instead $\lambda_1 > 0 > \lambda_2$, the same calculation reveals

$$|u_2^\top \widetilde{u}_1| \approx \frac{|u_2^\top \mathcal{R} u_1|}{\lambda_1(\lambda_1 - \lambda_2)} = \frac{|u_2^\top \mathcal{R} u_1|}{\lambda_1(\lambda_1 + |\lambda_2|)}.$$

We bound $\frac{|u_2^\top \mathcal{R} u_1|}{\lambda_1(\lambda_1+|\lambda_2|)} \leqslant \frac{|u_2^\top \mathcal{R} u_1|}{\lambda_1^2}$ in our perturbation results for simplicity.

The preceding calculation explains the leading bias terms $\frac{\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\|}{\delta_k\lambda_k}, \frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\lambda_k^2}$. By contrast, the higher-order term in $\mathcal{B}_k$ involving

$$\frac{R_{\max}}{\delta_k\lambda_k^3}\operatorname{osc}(\mathcal{R})$$

comes from bounding the diagonal remainder $\varepsilon(z)$ in the second-order expansion of $\Phi(z)$ in (5.1):

$$\Phi(z) = \frac{1}{z} I_n + \frac{1}{z^3}\mathcal{R} + \varepsilon(z).$$

We control this remainder using a uniform oscillation estimate, which may not be sharp. A higher-order expansion of $\Phi(z)$ would reveal additional structured variance-profile terms beyond $\mathcal{R}$ and replace the uniform $\operatorname{osc}(\mathcal{R})$ bound with finer geometric couplings, yielding a sharper higher-order term. However, isolating these higher-order structures requires substantially more delicate analysis to control the deterministic remainders and extract the finer stochastic fluctuations. This is beyond the scope of the current paper.

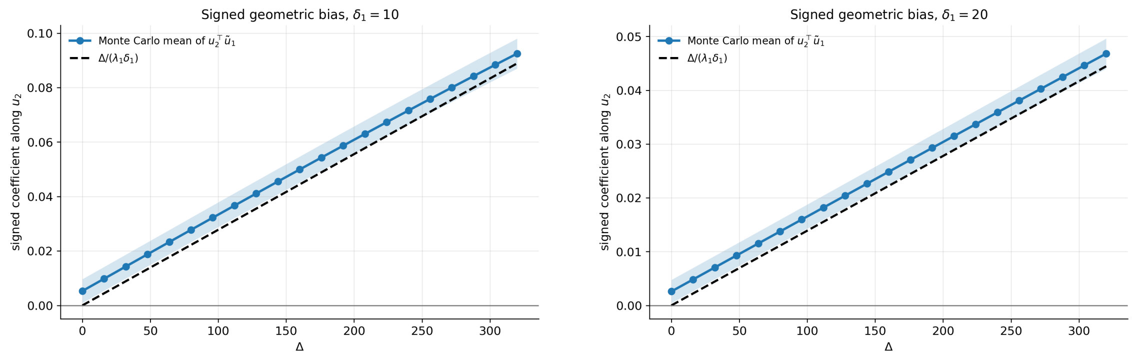
3.2 Sharpness and numerical experiments

Before presenting the numerical experiments, we briefly discuss the sharpness and necessity of the different terms in our perturbation bounds. In the homogeneous-noise setting studied in [98], eigenspace perturbation is governed by two mechanisms: stochastic mixing between distinct signal directions, of order $\sqrt{k}/\delta_k$ times the relevant fluctuation scale, and leakage into the noise subspace, of order $\|E\|/\lambda_k$. As discussed in [98, Section 4.1], these contributions are near-optimal in the i.i.d. Gaussian noise case.

Under a heterogeneous variance profile, Theorem 1 contains an additional contribution, namely the geometric-bias term $\mathcal{B}_k$. Its leading components are determined by the variance-weighted couplings between the target and non-target signal eigenspaces, and vanish in the homogeneous case $\mathcal{R} = cI$. Thus the first two mechanisms are inherited from the homogeneous theory, whereas $\mathcal{B}_k$ is specific to variance heterogeneity.

The rank-two example above shows explicitly that this geometric contribution appears at the predicted scale and therefore cannot in general be removed from the perturbation bound. The simulations below are designed to isolate this effect and illustrate that the deterministic bias persists under random perturbations and is visible at finite sample sizes.

Numerical simulations. We consider a rank-2 model $A = \lambda_1 u_1 u_1^\top + \lambda_2 u_2 u_2^\top$ with spikes $\lambda_1 > \lambda_2 > 0$ and eigenvectors $u_1 = \frac{1}{\sqrt{2}}(e_1 + e_2)$ and $u_2 = \frac{1}{\sqrt{2}}(e_1 - e_2)$. The noise variance profile has row sums $\mathcal{R}_\Delta = \operatorname{diag}(\rho + \Delta, \rho - \Delta, \rho, \ldots, \rho)$, where $\rho > 0$ is the baseline variance and $0 \leqslant \Delta \leqslant 0.8\rho$ is the heterogeneity parameter. This gives the geometric coupling $\mathcal{V}_{12} = u_2^\top \mathcal{R}_\Delta u_1 = \Delta$.



(a) $\delta_1 = 10$. (b) $\delta_1 = 20$.

Figure 2: Geometric bias under varying spectral gaps. Setup: dimension $n = 1200$, baseline noise $\rho = 400$, top spike $\lambda_1 = 360$, heterogeneity parameter $\Delta \in \{0, 16, 32, \ldots, 320\}$, with 300 Monte Carlo samples. Each panel shows the empirical alignment $\mathbb{E}[u_2^\top \widetilde{u}_1]$ versus Δ, compared with the theoretical prediction $\Delta/(\lambda_1\delta_1)$.

To show that geometric bias is deterministic rather than random, we compute the Monte Carlo average of $u_2^\top \widetilde{u}_1$, where in each trial we align the empirical eigenvector so that $u_1^\top \widetilde{u}_1 \geqslant 0$. Since the noise is zero-mean, averaging removes random fluctuations while preserving the deterministic bias. Figure 2 compares these empirical averages with our theoretical prediction $\frac{\Delta}{\delta_1 \lambda_1}$ for two different spectral gaps $\delta_1 = \lambda_1 - \lambda_2$.

3.3 Oracle de-biasing

The goal of this subsection is modest. We show, at the oracle level, that the leading $\mathcal{V}_{\mathcal{J}\mathcal{I}}$-bias in $\mathcal{B}_k$ can be removed. As discussed earlier, $\mathcal{V}_{\mathcal{J}\mathcal{I}}$ causes the main deviation from previous homogeneous bounds. When $k = r_+$, $\mathcal{J} = \varnothing$ and this bias vanishes. For intermediate eigenspaces ($k < r_+$), correction may help when the deterministic bias dominates the stochastic fluctuations, i.e.,

$$\frac{\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\|}{\lambda_k} \gg M.$$

For $s \in [k]$, consider the block matrices $U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{I}}$ and $U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{J}}$. Define

$$\mathcal{T}_s := (I - U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{J}} \Lambda_{\mathcal{J}})^{-1} U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{I}} \Lambda_{\mathcal{I}}. \tag{3.1}$$

The next result guarantees that $\mathcal{T}_s$ is well-defined on the event considered in Theorem 1. Its proof is provided in Appendix H.1.

LEMMA 3.1. *Assume $\mathcal{J} = \{k+1, \ldots, r_+\} \neq \varnothing$. Under the assumptions and with the same probability as in Theorem 1, for every $s \in [k]$,*

$$\left\| \left(I - U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{J}} \Lambda_{\mathcal{J}} \right)^{-1} \right\| \leqslant 60 \frac{\widetilde{\lambda}_s}{\delta_k}.$$

Define the oracle corrected vectors

$$w_s^{\mathrm{orc}} := P_U \widetilde{u}_s - U_{\mathcal{J}} \mathcal{T}_s U_{\mathcal{I}}^\top \widetilde{u}_s, \qquad s = 1, \dots, k.$$

Let

$$W_k^{\mathrm{orc}} := (w_1^{\mathrm{orc}}, \dots, w_k^{\mathrm{orc}}), \qquad U_k^{\mathrm{orc}} := \mathrm{orth}(W_k^{\mathrm{orc}}).$$

We denote by $\mathrm{orth}(W)$ the orthonormalization of the matrix W, which consists of orthonormal columns spanning the same subspace as W.

Define the oracle error term

$$\mathcal{E}_k^{\mathrm{orc}} := 4\sqrt{k}\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\lambda_k^2} + 24\sqrt{k}\frac{R_{\max}}{\lambda_k^4}\mathrm{osc}(\mathcal{R}) + 124\sqrt{k}\frac{M_D}{\delta_k}.$$

PROPOSITION 3.1 (Oracle blockwise de-biasing)**.** *Assume that the conditions of Theorem 1 hold. The following holds with the same probability as Theorem 1:*

$$\|\sin\angle(U_k^{\mathrm{orc}}, U_k)\| \leqslant 2\mathcal{E}_k^{\mathrm{orc}},$$
$$\min_{O \in \mathbb{O}(k)} \|U_k^{\mathrm{orc}} - U_k O\|_{2,\infty} \leqslant 3\|U\|_{2,\infty}\mathcal{E}_k^{\mathrm{orc}}.$$

The proof of Proposition 3.1 is deferred to Appendix H.2. The key point is that the oracle correction subtracts the deterministic contribution from the $\mathcal{J}$-block predicted by the QVE. This removes the leading $\mathcal{V}_{\mathcal{J}\mathcal{I}}$-term in $\mathcal{B}_k$. Additionally, the projection P_U removes the null-space component, which explains why the signal-noise term $\|E\|/\lambda_k$ in Theorem 1 is absent from Proposition 3.1.

The oracle estimator uses the unknown population eigenspace. A natural plug-in version replaces the population blocks by empirical outlier blocks. We provide a plug-in de-biasing implementation and explain its limitations in the Appendix J. Developing optimal data-driven bias correction methods is left for future work.

4 Related literature

This section highlights perturbation results relevant to heterogeneous random noise. A broader review of matrix perturbation, rowwise eigenspace analysis, and statistical applications can be found in [98]. Here we focus on work most closely related to variance heterogeneity and random matrix perturbation with non-identically distributed noise.

Perturbation bounds beyond worst-case theory. Classical perturbation results, including Weyl's inequality [101] and the Davis-Kahan-Wedin sin Θ theorem [50, 100], control eigenvalue and eigenspace errors through the operator norm of the perturbation and the relevant spectral separation. While sharp for arbitrary deterministic perturbations, these bounds can be conservative for structured random perturbations. Modern refinements include population-gap variants, Schatten and unitarily invariant norm bounds, perturbation projection error bounds, and other deterministic or stochastic improvements; see, for example, [107, 97, 36, 79, 110, 99, 8, 22, 112, 95, 84, 27].

Extensive work has also addressed entrywise and rowwise perturbation bounds. Representative works include [54, 3, 2, 34, 39, 45, 76, 113, 27, 105, 106, 104]. Perturbation of linear and bilinear forms of eigenvectors and singular vectors has been studied in [73, 102, 52, 77, 47, 4, 5]. These results play an important role in high-dimensional statistics, including matrix completion, spectral clustering, ranking, community detection, submatrix localization, and Gaussian mixture models; see, for example, [38, 37, 71, 9, 10, 35, 46, 81, 94, 96, 78, 15, 72, 32, 33, 80, 51, 55, 61, 82, 31, 49]. For a statistical survey of spectral methods and perturbation bounds, see [44].

In the homogeneous Gaussian-noise setting, [85, 98] obtained sharp singular-subspace perturbation estimates showing that the error is governed by a signal-to-noise term and a stochastic signal-space fluctuation term. Recent work of Tran and Vu [**?** 92] develops combinatorial contour-expansion and relative-norm perturbation bounds that exploit structural interaction between the signal and the perturbation. Our work shares the goal of moving beyond worst-case perturbation theory, but through a different mechanism: we use variance-profile local laws and QVE expansions to identify a deterministic geometric bias induced by heterogeneous noise.

Heterogeneous noise in statistical spectral problems. There is a substantial statistical literature on spectral estimation under heteroskedastic or heterogeneous noise. Zhang, Cai and Wu [108] introduced HeteroPCA, based on diagonal deletion and imputation, to correct the bias caused by heterogeneous diagonal noise in spiked covariance models. Yan, Chen and Fan [105] developed inference for heteroskedastic PCA with missing data, obtaining distributional guarantees for the estimated principal subspace and entrywise inference for the spiked covariance matrix. Related work on high-dimensional heteroskedastic PCA and weighted PCA includes [64, 66, 67]. Zhang and Mondelli [111] recently studied rank-one matrix denoising with doubly heteroscedastic noise and derived asymptotic fundamental limits and optimal spectral methods.

These works are close in motivation but differ from ours in both model and mechanism. Much of this literature concerns sample covariance, missing-data, or rectangular denoising models, where the main issue is often diagonal-noise bias correction, whitening, optimal shrinkage, or statistical inference. By contrast, we study a symmetric additive signal-plus-noise model with arbitrary entrywise variance profile and possible sparsity. The bias identified here is not merely a diagonal variance bias; it is a geometric bias governed by $\mathcal{V} = U^\top \mathcal{R} U$, which records how the row-wise variance profile changes the geometry of the signal eigenspaces.

We also cite related work on rowwise, entrywise, and inferential perturbation for completeness, including [54, 39, 3, 2, 34, 76, 27, 105, 106, 5, 4, 47, 42]; these works concern fine-grained perturbation and distributional inference in related spectral models, but they do not address the variance-profile geometric bias studied here.

Related literature in random matrix theory. Our work is also connected to the random matrix literature on finite-rank deformations and outlier eigenvectors. The BBP transition was introduced by Baik, Ben Arous and Péché [13], with subsequent developments for spiked covariance models, matrix denoising, and finite-rank deformations in [14, 87, 12, 24, 25, 23, 30, 40, 41, 19, 20, 21, 26, 53, 83]. The variance-profile setting is closely related to the theory of general Wigner-type matrices and the quadratic vector equation. Ajanki, Erdős and Krüger [7] developed the QVE/Dyson-equation framework and proved local laws and universality for matrices with general variance profiles. Our use of the QVE is more elementary and takes place in the large-z outlier domain, where the solution is stable and admits a direct expansion.

Recent asymptotic work has begun to study outliers, detection, and inference limits in spiked models with variance-profile, inhomogeneous, or structured noise. Bhattacharya, Chakrabarty and Hazra [29] analyze outlier eigenvalues and eigenvectors of generalized Wigner matrices with finite-rank deterministic perturbations. Guionnet, Ko, Krzakala and Zdeborová [59] study low-rank matrix estimation with inhomogeneous output channels and detection thresholds. Bao, Cheong, Lee and Li [18] study signal detection from spiked noise via asymmetrization. We also mention the physics work of Ferreira and Metz [56] on the BBP transition in an inhomogeneous rank-one spiked Wigner model. These results are complementary to ours: they are mainly asymptotic and focus on limiting distributions, information limits, detection thresholds, or BBP transitions, whereas our results are finite-sample perturbation bounds for deterministic low-rank signals under sparse independent variance-profile noise.

5 Deterministic QVE estimates and preliminary bounds

Consider the resolvent (Green function) of E:

$$G(z) = (zI - E)^{-1}, \qquad z \in \mathbb{C}^+.$$

For $z \in \mathbb{C}^+$, define a deterministic diagonal matrix

$$\Phi \equiv \Phi(z) = \mathrm{diag}(\phi_1(z), \dots, \phi_n(z))$$

where $\phi(z) = (\phi_1(z), \dots, \phi_n(z))$ is the solution to the Quadratic Vector Equations (QVE) given by:

$$\frac{1}{\phi_i(z)} = z - \sum_{j=1}^{n} \sigma_{ij}^2 \phi_j(z), \qquad i \in [n], \tag{5.1}$$

with $\operatorname{Im} \phi_i(z) < 0$ for $z \in \mathbb{C}^+$. Each $\phi_i(z)$ is analytic on $\mathbb{C}^+$. It is shown in [7, Theorem 2.1] that such a solution exists, is unique, and extends continuously to the real axis outside the spectrum (so in particular $\phi_i(z) > 0$ for z sufficiently large).

We collect deterministic estimates for the QVE solution $\Phi(z)$ and elementary bounds for resolvents $G(z)$. These estimates are used in two places: to identify the deterministic locations of outlier eigenvalues, and to extract the geometric bias from the expansion of $\Phi(z)$.

Recall that

$$R_i = \sum_{j=1}^{n} \sigma_{ij}^2, \qquad R_{\max} = \max_i R_i, \qquad \mathcal{R} = \mathrm{diag}(R_1, \dots, R_n).$$

The next result provides basic bounds on $\phi_i(z)$ and the second-order expansion of these ϕ_i's for large z. The proof is provided in Appendix A.1.

LEMMA 5.1. *Assume $z \in \mathbb{C}$ satisfying $|z| \geqslant \sqrt{6R_{\max}}$.*
(i) *For every $i \in [n]$, we have*

$$\frac{1}{2|z|} \leqslant |\phi_i(z)| \leqslant \frac{3}{2|z|}$$

and

$$\frac{3}{4}|z| \leqslant \big|z - \sum_j \sigma_{ij}^2 \phi_j(z)\big| \leqslant \frac{5}{4}|z|.$$

Furthermore, we have

$$\|\Phi'(z)\| = \max_i |\phi_i'(z)| \leqslant \frac{4}{|z|^2} \quad \textit{and} \quad \|\Phi''(z)\| = \max_i |\phi_i''(z)| \leqslant \frac{28}{|z|^3}.$$

(ii) *For $i \in [n]$, we have the second-order expansion:*

$$\phi_i(z) = \frac{1}{z} + \frac{R_i}{z^3} + \varepsilon_i(z) \quad \textit{with} \quad |\varepsilon_i(z)| \leqslant \frac{45}{8}\frac{R_{\max}^2}{|z|^5}. \tag{5.2}$$

In particular,

$$\Phi(z) = \frac{1}{z} I_n + \frac{1}{z^3}\mathcal{R} + \varepsilon(z), \tag{5.3}$$

where

$$\varepsilon(z) = \operatorname{diag}(\varepsilon_1(z), \cdots, \varepsilon_n(z)) \quad \textit{and} \quad \|\varepsilon(z)\| \leqslant \frac{45}{8}\frac{R_{\max}^2}{|z|^5}.$$

Furthermore, we have

$$\Phi'(z) = -\frac{1}{z^2} I_n - \frac{3}{z^4}\mathcal{R} + \varepsilon'(z), \tag{5.4}$$

where

$$\varepsilon'(z) = \operatorname{diag}(\varepsilon_1'(z), \cdots, \varepsilon_n'(z)) \quad \textit{and} \quad \|\varepsilon'(z)\| \leqslant \frac{243}{5}\frac{R_{\max}^2}{|z|^6}.$$

The entrywise bound in Lemma 5.1 is not sufficient for the eigenspace perturbation theorem. For the error term $\varepsilon(z) = \operatorname{diag}(\varepsilon_1(z), \ldots, \varepsilon_n(z))$ in the second-order expansion of $\Phi(z)$ in (5.2), we control its *oscillation*. For $z \in \mathbb{C}$ satisfying $|z| \geqslant \sqrt{6R_{\max}}$, denote

$$\operatorname{osc}(\varepsilon(z)) := \max_{i,j} |\varepsilon_i(z) - \varepsilon_j(z)|.$$

If z is a real number, then $\operatorname{osc}(\varepsilon(z)) = \max_i \varepsilon_i(z) - \min_i \varepsilon_i(z)$. Recall that

$$\operatorname{osc}(\mathcal{R}) = R_{\max} - R_{\min}.$$

The bound on $\operatorname{osc}(\varepsilon(z))$ is given below; its proof is deferred to Appendix A.2.

LEMMA 5.2. *For any complex number z with $|z| \geqslant \sqrt{6R_{\max}}$, we have*

$$\operatorname{osc}(\varepsilon(z)) \leqslant \frac{15R_{\max}}{|z|^5}\operatorname{osc}(\mathcal{R}).$$

The same QVE expansion also determines the typical locations around which the outlier eigenvalues concentrate. We prove the next result in Appendix A.3.

PROPOSITION 5.1 (Deterministic outlier location)**.** *Let $j \leqslant r_+$ and assume $\lambda_j \geqslant 6\sqrt{R_{\max}}$. Define*

$$\widehat{\alpha}_j(z) := 1 - \lambda_j u_j^\top \Phi(z) u_j \quad \textit{for} \quad z \geqslant 3\sqrt{R_{\max}}.$$

Then $\widehat{\alpha}_j$ has a unique real zero $\vartheta_j \in [3\sqrt{R_{\max}}, \infty)$. Moreover, this zero is simple and satisfies

$$\left|\vartheta_j - \lambda_j - \frac{\mathcal{V}_{jj}}{\lambda_j}\right| \leqslant 145\frac{R_{\max}}{\lambda_j^3}\operatorname{osc}(\mathcal{R}).$$

Since $\widehat{\alpha}_j(\lambda_j)$ is increasing and $\widehat{\alpha}_j(\lambda_j) \leqslant 0$, we also have

$$\vartheta_j \geqslant \lambda_j. \tag{5.5}$$

The following result provides crude bounds on the operator norms of Green functions $G(z) = (zI - E)^{-1}$ and $G^{(k)} = (zI - E^{(k)})^{-1}$, where $E^{(k)}$ is obtained from E with the k-th row and column replaced by zero.

LEMMA 5.3 (Lemma 16 from [85]). *For $z \in \mathbb{C}$ with $|z| \geqslant 2\|E\|$, we have*

$$\|G(z)\| \leqslant \frac{2}{|z|}, \quad \|G^{(k)}\| \leqslant \frac{2}{|z|}. \tag{5.6}$$

LEMMA 5.4 (Bound on the noise matrix). *Let $E = (E_{ij})_{n\times n}$ be a symmetric matrix whose entries are centered and independent (up to symmetry) random variables. Assume the entries E_{ij}'s are sub-Gaussian with $\|E_{ij}\|_{\psi_2} \leqslant K\sigma_{ij}$. Denote $R_{\max} = \max_i \sum_j \sigma_{ij}^2$. Then for any $D > 0$, we have*

$$\mathbb{P}\left(\|E\| \leqslant 2.9\sqrt{R_{\max}} + cK\sigma_{\max}(D+2)\log n\right) \geqslant 1 - 3n^{-D}$$

for an absolute constant $c > 0$.

The result is a direct consequence of the sharp non-asymptotic norm bounds established in Bandeira and van Handel [16]. We provide a short derivation in Appendix A.4.

6 Proof strategy and technical inputs

This section states the probabilistic inputs used in the proof of Theorem 1 and explains how they combine with the deterministic QVE estimates from Section 5. The two main probabilistic ingredients are the isotropic local laws for the resolvent $G(z)$ and the location theorem for the outlier eigenvalues of $A + E$.

6.1 Isotropic local laws

Consider $G(z) = (zI - E)^{-1}$ for $z \in \mathbb{C}^+$. Recall the QVE approximation $\Phi(z)$ defined in (5.1). Our first main technical step is to establish an isotropic local law for $G(z)$. Denote

$$\mathfrak{m}_D := \begin{cases} C'_{\mathrm{gen}}(D+6)^{3/2}\beta K\sigma_{\max}(\log n)^2, & \text{in the sparse sub-Gaussian regime,} \\ C'_{\mathrm{bd}}(D+6)c_0 K\sigma_{\max}\log n, & \text{in the sparse bounded-entry regime.} \end{cases}$$

Here, C'_{gen} and C'_{bd} are sufficiently large absolute constants.

THEOREM 4 (Isotropic local law). *Fix $D > 0$. Suppose Assumptions 1.1 and 1.2 hold. For any deterministic unit vectors $v, w \in \mathbb{R}^n$ and every fixed $z \in \mathbb{C}$ with $|z| \geqslant 6\sqrt{R_{\max}}$, we have*

$$\left|v^\top(G(z) - \Phi(z))w\right| \leqslant \frac{\mathfrak{m}_D}{|z|^2}$$

with probability at least $1 - n^{-D-4}$.

Theorem 4 is proved in Section 8. We also need a version that is uniform over the spectral domain where the outlier eigenvalues are located. Define

$$\mathcal{S}_{\text{out}} := \left\{ z \in \mathbb{C} : 6\sqrt{R_{\max}} \leqslant |z| \leqslant 2R_{\max}^3 \right\}.$$

This result is a direct consequence of Theorem 4; its proof is provided in Appendix E. Recall that $A = U\Lambda U^\top$.

COROLLARY 3 (Uniform compressed local law)**.** *Fix $D > 0$. Suppose the assumptions of Theorem 4 hold. Then, with probability at least $1 - n^{-D}$,*

$$\sup_{z\in\mathcal{S}_{\text{out}}} |z|^2 \|U^\top(G(z) - \Phi(z))U\| \leqslant C'r \cdot \mathfrak{m}_D = M_D, \tag{6.1}$$

$$\sup_{z\in\mathcal{S}_{\text{out}}} |z|^2 \max_{1\leqslant i\leqslant n} \|e_i^\top(G(z) - \Phi(z))U\| \leqslant M_D. \tag{6.2}$$

Consequently,

$$\sup_{z\in\mathcal{S}_{\text{out}}} |z|^2 \max_{1\leqslant i\leqslant n} \left\| e_i^\top Q(G(z) - \Phi(z))U \right\| \leqslant 2M_D, \tag{6.3}$$

where $Q = I - UU^\top$ and M_D is defined in (1.3)*.*

We next establish a row-adaptive local law that retains the rowwise variation of the deterministic test matrix. This refinement will be important in the applications, where it allows us to reach the sharp degree regime. For flexibility, we formulate the result for a deterministic matrix $B \in \mathbb{R}^{n\times r}$. The proof combines a leave-one-out argument with the isotropic local law in Theorem 4. The leave-one-out step requires a slightly stronger variance-dominance condition.

Fix $D > 0$. For simplicity, define

$$t_{D,r} := C(D + r)\log n, \tag{6.4}$$

where $C > 0$ is a sufficiently large absolute constant. For the deterministic matrix $B \in \mathbb{R}^{n\times r}$, define

$$\gamma_i(B)^2 := \sum_{j=1}^n \sigma_{ij}^2 \|e_j^\top B\|^2, \qquad \Gamma(B) := \max_{1\leqslant i\leqslant n} \gamma_i(B).$$

Note that $\gamma_i(B) = \left(\mathbb{E}\|e_i^\top EB\|^2\right)^{1/2}$ is the standard-deviation scale for the ith row of EB, while $\Gamma(B)$ is the largest such rowwise fluctuation scale.

Finally, define the row-adaptive spectral domain

$$\mathcal{S}_{\text{ra}} := \left\{ z \in \mathbb{C} : C_{\text{ra}}\sqrt{t_{D,r}R_{\max}} \leqslant |z| \leqslant 2R_{\max}^3 \right\}, \tag{6.5}$$

where $C_{\text{ra}} > 0$ is a sufficiently large absolute constant.

We are now ready to state the row-adaptive local law. Recall that $L_{D,r}$ is defined in (1.5).

COROLLARY 4 (Uniform row-adaptive local law)**.** *Fix $D > 0$. Suppose Assumption 1.1 and Assumption 1.3 hold. Let $B \in \mathbb{R}^{n\times r}$ be deterministic with $\|B\| \leqslant n^{100}$. Then, with probability at least $1 - n^{-D}$,*

$$\sup_{z\in\mathcal{S}_{\text{ra}}} |z|^2 \|(G(z) - \Phi(z))B\|_{2,\infty} \leqslant C\left[\sqrt{t_{D,r}}\,\Gamma(B) + \left(t_{D,r}L_{D,r}^{\text{ra}} + \mathfrak{m}_D\right)\|B\|_{2,\infty}\right]. \tag{6.6}$$

Here $C > 0$ is an absolute constant.

The proof of Corollary 4 is given in Appendix E.

6.2 Outlier eigenvalue location

We next state the outlier-location results used in the proof of the eigenspace perturbation results. For any $\lambda_j \geqslant 6\sqrt{R_{\max}}$, Proposition 5.1 guarantees that there exists a unique solution ϑ_j to the equation $\lambda_j^{-1} - u_j^\top \Phi(z) u_j = 0$ for $z \geqslant 3\sqrt{R_{\max}}$. Besides,

$$\vartheta_j \approx \lambda_j + \frac{\mathcal{V}_{jj}}{\lambda_j}.$$

For a simple supercritical eigenvalue λ_k, we will show that ϑ_k is the typical location of $\widetilde{\lambda}_k$.

Recall that

$$\rho_k \equiv \rho_{k,D} := 10M_D + 15\frac{\operatorname{osc}(\mathcal{R})}{\lambda_k}.$$

Denote the index set for the supercritical eigenvalues as

$$\mathcal{O}_+ := \{l \in [r_+] : \lambda_l \geqslant 6\sqrt{R_{\max}}\}.$$

THEOREM 5 (Individual outlier eigenvalue location)**.** *Fix $D > 0$ and $k \in [r_+]$. If $\lambda_k \geqslant 6\sqrt{R_{\max}} + 100\rho_k$ is a simple eigenvalue and the gap condition holds:*

$$\min_{l<k}(\vartheta_l - \vartheta_k) \geqslant 50\rho_k \quad \textit{and} \quad \min_{l\in\mathcal{O}_+,\, l>k}(\vartheta_k - \vartheta_l) \geqslant 50\rho_k$$

with the convention that the minimum over an empty set is $+\infty$, then, with probability at least $1 - n^{-D}$,

$$|\widetilde{\lambda}_k - \vartheta_k| \leqslant \rho_k.$$

The proof of Theorem 5 is given in Section 9. Furthermore, we provide the following result on the outlier eigenvalue clusters, which will be used in the proof of the top-k eigenspace perturbations.

THEOREM 6 (Lower edge of the top-k outlier clusters)**.** *Fix $D > 0$ and $k \in [r_+]$. Define*

$$\underline{\vartheta}_k := \min_{1\leqslant t\leqslant k} \vartheta_t.$$

If $\lambda_k \geqslant 6\sqrt{R_{\max}} + 100\rho_k$ and the gap condition holds:

$$\min_{l\in\mathcal{O}_+,\, l>k}(\underline{\vartheta}_k - \vartheta_l) \geqslant 50\rho_k \tag{6.7}$$

with the convention that the minimum over an empty set is $+\infty$, then, with probability at least $1 - n^{-D}$,

$$\widetilde{\lambda}_k \geqslant \underline{\vartheta}_k - \rho_k.$$

In particular, for every $s \in [k]$,

$$\widetilde{\lambda}_s \geqslant \underline{\vartheta}_k - \rho_k.$$

REMARK 6.1. *Theorem 6 holds under a stronger but more conventional assumption that the k-th deterministic root lower-bounds the preceding roots (i.e., $\vartheta_l \geqslant \vartheta_k$ for all $l < k$). Under that assumption, $\underline{\vartheta}_k = \vartheta_k$. However, we introduce $\underline{\vartheta}_k$ to avoid the internal ordering of these ϑ_s's for $s \leqslant k$. Indeed, the deterministic roots can physically cross and permute, even when the underlying λ_s's are strictly ordered. For instance, even if $\lambda_l > \lambda_s$ for $l < s \in [k]$, as given in Proposition 5.1, $\vartheta_l \approx \lambda_l + \frac{\mathcal{V}_{ll}}{\lambda_l}$ could be dominated by $\vartheta_s \approx \lambda_s + \frac{\mathcal{V}_{ss}}{\lambda_s}$. This is because the shift $\mathcal{V}_{ll} = u_l^\top \mathcal{R} u_l$ depends on the alignment of the l-th population eigenvector with the variance profile $\mathcal{R}$. In the presence of anisotropic noise, it is entirely possible for a lower-ranked spike's variance projection (e.g., $\mathcal{V}_{ss}$) to be significantly larger than that of a higher-ranked spike (e.g., $\mathcal{V}_{ll}$).*

The proof of Theorem 6 parallels that of Theorem 5 and can be found in Section 9.

6.3 Proof roadmap

We briefly explain how the technical inputs above are used to prove Theorem 1. The proof is carried out on a high-probability event where $\|E\| \leqslant 3\sqrt{R_{\max}}$ and the isotropic local laws Theorem 4 and Corollary 3 hold.

The signal condition $\lambda_k \geqslant 9\sqrt{R_{\max}}$ ensures that the outlier eigenvalues satisfy

$$\widetilde{\lambda}_s \geqslant \lambda_k - \|E\| \geqslant 6\sqrt{R_{\max}} \tag{6.8}$$

for all $s \in [k]$ by Weyl's inequalty. Thus the uniform local law Corollary 3 applies at the random eigenvalues $z = \widetilde{\lambda}_s$.

The deterministic part of the proof starts from the eigenvector equation

$$\widetilde{u}_s = G(\widetilde{\lambda}_s) A \widetilde{u}_s, \qquad s \in [k].$$

Projecting this identity onto the non-target signal directions and the null space gives separate bounds for how much $\widetilde{u}_s$ extends outside the target space. The null-space projection produces the usual signal-to-noise contribution $\|E\|/\lambda_k$. For the signal-space terms, we decompose

$$G(z) = \Phi(z) + \Xi(z), \qquad \Xi(z) := G(z) - \Phi(z).$$

The compressed local law controls the contribution of $\Xi(z)$, while the large-z expansion of $\Phi(z)$ identifies the deterministic variance-profile contribution. This yields the stochastic term involving M_D and the geometric bias term $\mathcal{B}_k$. The final operator-norm and $2 \to \infty$ estimates follow from these blockwise bounds through deterministic reductions.

7 Proof of the main perturbation bounds: Theorem 1

Set

$$\Xi(z) := G(z) - \Phi(z).$$

Fix $D > 0$. Define the events

$$\mathcal{E}_{\mathrm{op}} := \{\|E\| \leqslant 3\sqrt{R_{\max}}\}$$

and

$$\mathcal{E}_{\mathrm{loc}}(D) := \left\{ \sup_{z \in \mathcal{S}} |z|^2 \|U^\top \Xi(z) U\| \leqslant M_D \right\} \cap \left\{ \sup_{z \in \mathcal{S}_{\mathrm{out}}} |z|^2 \max_{1 \leqslant i \leqslant n} \left\| e_i^\top Q \Xi(z) U \right\| \leqslant 2M_D \right\},$$

where $\mathcal{S}_{\mathrm{out}} := \left\{ z \in \mathbb{C} : 6\sqrt{R_{\max}} \leqslant |z| \leqslant 2R_{\max}^3 \right\}$ and $Q = I - UU^\top$. Set

$$\mathcal{E}_D := \mathcal{E}_{\mathrm{op}} \cap \mathcal{E}_{\mathrm{loc}}(D).$$

By Lemma 5.4 and Corollary 3, we have

$$\mathbb{P}(\mathcal{E}_D) \geqslant 1 - n^{-D}.$$

All estimates below are deterministic on $\mathcal{E}_D$, except when the outlier-location theorem is invoked. In that step we intersect with the corresponding outlier-location event, which still has probability at least $1 - n^{-D}$ after adjusting constants.

On the event $\mathcal{E}_D$, note that our signal assumption $9\sqrt{R_{\max}} \leqslant \lambda_k \leqslant R_{\max}^3$ ensures $\widetilde{\lambda}_s \in \mathcal{S}_{\mathrm{out}}$ for all $s \leqslant k$ by Weyl's inequality (see (6.8)).

For simplicity, we write

$$M \equiv M_D.$$

We first present deterministic bounds on the difference between the eigenspaces U_k and $\widetilde{U}_k$. Let $J = [r]\backslash[k]$ denote the complement index set and $U_J = (u_{k+1}, \ldots, u_r)$. Let P_W be the orthogonal projection onto the subspace W.

PROPOSITION 7.1. *For $1 \leqslant k \leqslant r+$, assume $\lambda_k \geqslant 2\|E\|$. Then the following deterministic bounds hold.*

(i) *For the operator norm bound, we have*

$$\|\sin \angle(U_k, \widetilde{U}_k)\| \leqslant \|U_J^\top \widetilde{U}_k\|_F + 2\frac{\|E\|}{\lambda_k}. \tag{7.1}$$

(ii) *For the $2 \to \infty$ norm bound, we have*

$$\min_{O \in \mathbb{O}(k)} \|\widetilde{U}_k - U_k O\|_{2,\infty} \leqslant \|\widetilde{U}_k - P_{U_k}\widetilde{U}_k\|_{2,\infty} + \|U_k\|_{2,\infty} \|\sin \angle(U_k, \widetilde{U}_k)\|^2, \tag{7.2}$$

where

$$\begin{aligned}\|\widetilde{U}_k - P_{U_k}\widetilde{U}_k\|_{2,\infty} \leqslant{}& \|U\|_{2,\infty} \cdot \|U_J^\top \widetilde{U}_k\|_F + \frac{8}{3}\sqrt{k}\|U\|_{2,\infty}\frac{\mathrm{osc}(\mathcal{R})}{\widetilde{\lambda}_k^2} \\ &+ 2\sqrt{\sum_{s=1}^k \widetilde{\lambda}_s^2 \max_{1 \leqslant i \leqslant n} \left\| e_i^\top Q \Xi(\widetilde{\lambda}_s) U \right\|^2}.\end{aligned} \tag{7.3}$$

Given these deterministic bounds, the main technical contribution of our perturbation analysis is the estimation of

$$\|U_J^\top \widetilde{U}_k\|_F = \sqrt{\sum_{s=1}^k \|U_J^\top \widetilde{u}_s\|^2}.$$

We first present a deterministic bound for each $\|U_J^\top \widetilde{u}_s\|$. The geometric bias matrix $\mathcal{V} = U^\top \mathcal{R} U$ is defined in (1.2). For any index sets $\mathtt{I}, \mathtt{J} \in [r]$ with $U_{\mathtt{I}} = (u_i)_{i \in \mathtt{I}}$ and $U_{\mathtt{J}} = (u_j)_{j \in \mathtt{J}}$, we denote

$$\mathcal{V}_{\mathtt{IJ}} := U_{\mathtt{I}}^\top \mathcal{R}\, U_{\mathtt{J}}.$$

For a matrix B, we denote by $\mathrm{Off}(B)$ the matrix obtained by setting all diagonal entries to zero.

PROPOSITION 7.2. *Define index sets*

$$\begin{aligned} \mathcal{J} &:= \{k+1, \ldots, r_+\}, \quad \mathcal{I} := [r] \backslash \mathcal{J}, \\ \mathcal{N} &:= \{r_+ + 1, \ldots, r\}, \quad \mathcal{K} := [r] \backslash \mathcal{N}. \end{aligned}$$

Fix $s \in [k]$. Define the gap

$$\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s) := \widetilde{\lambda}_s \min_{j \in \mathcal{J}} |1 - \lambda_j u_j^\top \Phi(\widetilde{\lambda}_s) u_j|$$

with the convention that the minimum over an empty set is $+\infty$. If the following gap condition holds:

$$\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s) \geqslant \frac{2\lambda_{k+1}}{\widetilde{\lambda}_s^2} \|\mathrm{Off}(\mathcal{V}_{\mathcal{J}\mathcal{J}})\| + \frac{13 \lambda_{k+1} R_{\max}}{\widetilde{\lambda}_s^4} \mathrm{osc}(\mathcal{R}), \tag{7.4}$$

then we have

$$\begin{aligned} \|U_J^\top \widetilde{u}_s\| \leqslant & \frac{4}{\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s)} \left(\frac{\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\|}{\widetilde{\lambda}_s} + \frac{13}{2} \frac{R_{\max}}{\widetilde{\lambda}_s^3} \mathrm{osc}(\mathcal{R}) + \widetilde{\lambda}_s^2 \|U^\top \Xi(\widetilde{\lambda}_s) U\| \right) \\ & + 2\sqrt{3} \left(\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\widetilde{\lambda}_s^2} + \frac{13}{2} \frac{R_{\max}}{\widetilde{\lambda}_s^4} \mathrm{osc}(\mathcal{R}) + \widetilde{\lambda}_s \|U^\top \Xi(\widetilde{\lambda}_s) U\| \right). \end{aligned} \tag{7.5}$$

Note that when $\mathcal{J} = \varnothing$, we have $\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s) = \infty$, so the corresponding term on the right-hand side of (7.5) vanishes. The proofs of Propositions 7.1 and 7.2 are deferred to Appendix B. On the event $\mathcal{E}_D$, the condition $\lambda_k \geqslant 9\sqrt{R_{\max}}$ implies $\lambda_k \geqslant 3\|E\| \geqslant 2\|E\|$, so Proposition 7.1 applies.

Now we continue with the estimation of $\|U_J^\top \widetilde{u}_s\|$ and split the discussion into two cases: $k = r_+$ or $k < r_+$.

<u>Case 1.</u> Assume $k = r_+$. Then $\mathcal{J} = \varnothing$ and $\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s) = +\infty$. In this case,

$$J = \mathcal{N} = \{r_+ + 1, \ldots, r\}.$$

We just bound

$$\|U_J^\top \widetilde{u}_s\| \leqslant 2\sqrt{3} \left(\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\widetilde{\lambda}_s^2} + \frac{13}{2} \frac{R_{\max}}{\widetilde{\lambda}_s^4} \mathrm{osc}(\mathcal{R}) + \widetilde{\lambda}_s \|U^\top \Xi(\widetilde{\lambda}_s) U\| \right).$$

By Theorem 6 (note that the gap condition (6.7) holds trivially), for every $s \in [k]$,

$$\widetilde{\lambda}_s \geqslant \underline{\vartheta}_k - \rho_k \geqslant \lambda_k - \rho_k \geqslant \frac{99}{100} \lambda_k.$$

Thus,

$$\|U_J^\top \widetilde{u}_s\| = \|U_{\mathcal{N}}^\top \widetilde{u}_s\| < 4\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\lambda_k^2} + 24\frac{R_{\max}}{\lambda_k^4}\mathrm{osc}(\mathcal{R}) + 4\frac{M}{\lambda_k} \tag{7.6}$$

and

$$\|U_J^\top \widetilde{U}_k\|_F = \sqrt{\sum_{s=1}^{k} \|U_J^\top \widetilde{u}_s\|^2} \leqslant \sqrt{r_+}\left(4\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\lambda_{r_+}^2} + 24\frac{R_{\max}}{\lambda_{r_+}^4}\mathrm{osc}(\mathcal{R}) + 4\frac{M}{\lambda_{r_+}}\right). \tag{7.7}$$

<u>Case 2.</u> Assume $k < r_+$. Under our gap assumption

$$\delta_k \geqslant 65\rho_k + 20\frac{\mathrm{osc}(\mathcal{R})}{\lambda_k} \geqslant 65\rho_k + 20\frac{\lambda_{k+1}}{\lambda_k^2}\mathrm{osc}(\mathcal{R}), \tag{7.8}$$

we first show that (6.7) holds, that is,

$$\min_{l\in\mathcal{O}_+,\, l>k} (\underline{\vartheta}_k - \vartheta_l) \geqslant 50\rho_k \quad \text{where} \quad \underline{\vartheta}_k = \min_{1\leqslant t\leqslant k} \vartheta_t.$$

If $\mathcal{O}_+\backslash[k] = \varnothing$, then the inequality holds because the left-hand side is $+\infty$ by our notation. Assume $\mathcal{O}_+\backslash[k] \neq \varnothing$.

Suppose

$$\underline{\vartheta}_k = \vartheta_t$$

for some $1 \leqslant t \leqslant k$. Observe from (5.5) that

$$\vartheta_t \geqslant \lambda_t \geqslant \lambda_k > \lambda_{k+1} \geqslant \lambda_j \tag{7.9}$$

for any $j \in \mathcal{J} = \{k+1, \ldots, r_+\}$.

Define two auxiliary functions

$$\mathcal{L}(x) := x + \frac{R_{\min}}{x} - 145\frac{R_{\max}}{x^3}\mathrm{osc}(\mathcal{R}), \quad \mathcal{U}(x) := x + \frac{R_{\max}}{x} + 145\frac{R_{\max}}{x^3}\mathrm{osc}(\mathcal{R})$$

for $x \geqslant 6\sqrt{R_{\max}}$. Note that both $\mathcal{L}(x)$ and $\mathcal{U}(x)$ are increasing functions. Note that $\lambda_t \geqslant 9\sqrt{R_{\max}}$. Thus,

$$\mathcal{L}(\lambda_t) \leqslant \vartheta_t \leqslant \mathcal{U}(\lambda_t).$$

For any $l \in \mathcal{O}_+$ satisfying $l > k$, since $\lambda_l \leqslant \lambda_{k+1}$, by the definition of ϑ_l for and Proposition 5.1, we have

$$\mathcal{L}(\lambda_l) \leqslant \vartheta_l \leqslant \mathcal{U}(\lambda_l) \leqslant \mathcal{U}(\lambda_{k+1}).$$

In particular, for any $l \in \mathcal{O}_+\backslash[k]$ (note that $t \in \mathcal{O}_+$), we have

$$\begin{aligned}\vartheta_t - \vartheta_l &\geqslant \mathcal{L}(\lambda_t) - \mathcal{U}(\lambda_{k+1}) \\ &= (\lambda_t - \lambda_{k+1}) + \frac{R_{\min}}{\lambda_t} - \frac{R_{\max}}{\lambda_{k+1}} - 145\frac{R_{\max}}{\lambda_t^3}\mathrm{osc}(\mathcal{R}) - 145\frac{R_{\max}}{\lambda_{k+1}^3}\mathrm{osc}(\mathcal{R}).\end{aligned}$$

Rewriting

$$\frac{R_{\min}}{\lambda_t} - \frac{R_{\max}}{\lambda_{k+1}} = \frac{\lambda_{k+1} - \lambda_t}{\lambda_t \lambda_{k+1}} R_{\min} + \frac{R_{\min} - R_{\max}}{\lambda_{k+1}}$$

and using the assumption $\lambda_t, \lambda_{k+1} \geqslant 9\sqrt{R_{\max}} \geqslant 6\sqrt{R_{\min}}$, we further get

$$\begin{aligned}
\vartheta_t - \vartheta_l &\geqslant \left(1 - \frac{R_{\min}}{\lambda_t \lambda_{k+1}}\right)(\lambda_t - \lambda_{k+1}) + \frac{R_{\min} - R_{\max}}{\lambda_{k+1}} - \frac{145}{36}\frac{\operatorname{osc}(\mathcal{R})}{\lambda_t} - \frac{145}{36}\frac{\operatorname{osc}(\mathcal{R})}{\lambda_{k+1}} \\
&\geqslant \frac{35}{36}\delta_k - \left(1 + \frac{145}{36}\right)\frac{\operatorname{osc}(\mathcal{R})}{\lambda_{k+1}} - \frac{145}{36}\frac{\operatorname{osc}(\mathcal{R})}{\lambda_t} \\
&\geqslant \frac{35}{36}\delta_k - \frac{181}{36}\frac{\operatorname{osc}(\mathcal{R})}{\lambda_{k+1}} - \frac{145}{36}\frac{\operatorname{osc}(\mathcal{R})}{\lambda_k},
\end{aligned}$$

where we used $\lambda_t \geqslant \lambda_k$ in the last inequality. To proceed, note that

$$\frac{\operatorname{osc}(\mathcal{R})}{\lambda_{k+1}} = \frac{\operatorname{osc}(\mathcal{R})}{\lambda_k} + \delta_k \frac{\operatorname{osc}(\mathcal{R})}{\lambda_k \lambda_{k+1}} \leqslant \frac{\operatorname{osc}(\mathcal{R})}{\lambda_k} + \frac{1}{36}\delta_k.$$

We get

$$\vartheta_t - \vartheta_l \geqslant \left(\frac{35}{36} - \frac{181}{36^2}\right)\delta_k - \frac{326}{36}\frac{\operatorname{osc}(\mathcal{R})}{\lambda_k} \geqslant 50\rho_k$$

by our assumption (7.8) on δ_k. Hence, the conclusion of Theorem 6 holds. In particular, for every $s \in [k]$,

$$\widetilde{\lambda}_s \geqslant \underline{\vartheta}_k - \rho_k = \vartheta_t - \rho_k. \tag{7.10}$$

Recall $\Delta_{\mathcal{J}}^{\Phi}$ defined in Proposition 7.2. Next, we show that for any $s \in [k]$,

$$\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s) = \widetilde{\lambda}_s \min_{j \in \mathcal{J}} |1 - \lambda_j u_j^\top \Phi(\widetilde{\lambda}_s) u_j| \geqslant \frac{1}{2}\delta_k.$$

Denote $\alpha_j(z) = \lambda_j^{-1} - u_j^\top \Phi(z) u_j$ for each $j \in [r]$. Then

$$\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s) = \min_{j \in \mathcal{J}} \left|\widetilde{\lambda}_s \lambda_j \alpha_j(\widetilde{\lambda}_s)\right|.$$

For each $j \in \mathcal{J}$, by the expansion of $\Phi(z)$ in (5.3), we have

$$\alpha_j(z) = \frac{1}{\lambda_j} - u_j^\top \Phi(z) u_j = \frac{1}{\lambda_j} - \frac{1}{z} - \frac{\mathcal{V}_{jj}}{z^3} - \varepsilon_{jj}(z),$$

where $\varepsilon_{jj}(z) := u_j^\top \varepsilon(z) u_j$. Similarly,

$$\alpha_t(z) = \frac{1}{\lambda_t} - \frac{1}{z} - \frac{\mathcal{V}_{tt}}{z^3} - \varepsilon_{tt}(z).$$

From the difference $\alpha_j(z) - \alpha_t(z)$, we find

$$\alpha_j(z) = \left(\frac{1}{\lambda_j} - \frac{1}{\lambda_t}\right) + \alpha_t(z) - \frac{\mathcal{V}_{jj} - \mathcal{V}_{tt}}{z^3} - (\varepsilon_{jj}(z) - \varepsilon_{tt}(z)).$$

Therefore,

$$\begin{aligned}\widetilde{\lambda}_s\lambda_j\alpha_j(\widetilde{\lambda}_s) &= \widetilde{\lambda}_s\left(1-\frac{\lambda_j}{\lambda_t}\right)+\widetilde{\lambda}_s\lambda_j\alpha_t(\widetilde{\lambda}_s)-\frac{\lambda_j}{\widetilde{\lambda}_s^2}\left(\mathcal{V}_{jj}-\mathcal{V}_{tt}\right)-\widetilde{\lambda}_s\lambda_j(\varepsilon_{jj}(\widetilde{\lambda}_s)-\varepsilon_{tt}(\widetilde{\lambda}_s))\\ &:= T_1+T_2+T_3+T_4.\end{aligned}$$

It follows that

$$\left|\widetilde{\lambda}_s\lambda_j\alpha_j(\widetilde{\lambda}_s)\right| \geqslant T_1+T_2-|T_3|-|T_4|.$$

We estimate T_i's respectively. For T_1, by (7.9), we have

$$T_1=\widetilde{\lambda}_s\left(1-\frac{\lambda_j}{\lambda_t}\right)\geqslant \widetilde{\lambda}_s\left(1-\frac{\lambda_{k+1}}{\lambda_k}\right)=\frac{\widetilde{\lambda}_s}{\lambda_k}\delta_k.$$

Then, by (7.10), we get

$$T_1 \geqslant (\vartheta_t-\rho_k)\frac{\delta_k}{\lambda_k}=\frac{\vartheta_t}{\lambda_k}\delta_k-\rho_k\frac{\delta_k}{\lambda_k}\geqslant \delta_k-\rho_k$$

since $\frac{\vartheta_t}{\lambda_k}\geqslant 1$ by (7.9) and $\frac{\delta_k}{\lambda_k}\leqslant 1$.

Next, we estimate $T_2=\widetilde{\lambda}_s\lambda_j\alpha_t(\widetilde{\lambda}_s)$. By the same estimates as (9.20) and (9.21), we have

$$\frac{5}{2x^2}\leqslant \alpha_t'(x)\leqslant \frac{2}{x^2}\quad \text{for } x\geqslant 3\sqrt{R_{\max}}.$$

In particular, $\alpha_t(x)$ is increasing for $x \geqslant 3\sqrt{R_{\max}}$. By our definition, $\alpha_t(\vartheta_t)$=0. Since $\widetilde{\lambda}_s \geqslant \vartheta_t-\rho_k$ by (7.10),

$$\alpha_t(\widetilde{\lambda}_s)\geqslant \alpha_t(\vartheta_t-\rho_k)=\alpha_t(\vartheta_t-\rho_k)-\alpha_t(\vartheta_t)=-\rho_k\alpha_t'(\zeta),$$

where the last equation is due to the mean value theorem and ζ is a value between $\vartheta_t-\rho_k$ and ϑ_t. Then, by (7.9),

$$\widetilde{\lambda}_s\lambda_j\alpha_t(\widetilde{\lambda}_s)\geqslant -\widetilde{\lambda}_s\lambda_j\frac{2}{\zeta^2}\rho_k\geqslant -2\frac{\widetilde{\lambda}_s\lambda_{k+1}}{(\vartheta_t-\rho_k)^2}\rho_k.$$

Since $\vartheta_t-\rho_k\geqslant \frac{99}{100}\vartheta_t$ and $\widetilde{\lambda}_s\leqslant \frac{3}{2}\lambda_s\leqslant\frac{3}{2}\vartheta_s$ by Weyl's inequality and (5.5), we have

$$T_2\geqslant -2(\frac{100}{99})^2\frac{\frac{3}{2}\vartheta_s\lambda_{k+1}}{\vartheta_t^2}\rho_k>-4\rho_k.$$

Finally, for T_3 and T_4, by Lemma 5.2, the bounds $\widetilde{\lambda}_s\leqslant\frac{3}{2}\lambda_s$ and

$$\widetilde{\lambda}_s\geqslant\vartheta_t-\rho_k\geqslant\frac{99}{100}\vartheta_t\geqslant\frac{99}{100}\lambda_k,$$

we have

$$\begin{aligned}|T_3|+|T_4| &\leqslant \frac{\lambda_j}{\widetilde{\lambda}_s^2}\operatorname{osc}(\mathcal{R})+\widetilde{\lambda}_s\lambda_j\cdot\operatorname{osc}(\varepsilon(\widetilde{\lambda}_s))\\ &\leqslant 4\frac{\lambda_{k+1}}{\lambda_k^2}\operatorname{osc}(\mathcal{R})+\frac{15\cdot 2^4}{36}\frac{\lambda_{k+1}}{\lambda_k^2}\operatorname{osc}(\mathcal{R})=\frac{32}{3}\frac{\lambda_{k+1}}{\lambda_k^2}\operatorname{osc}(\mathcal{R}).\end{aligned}$$

Combining the above estimates, we arrive at

$$\widetilde{\lambda}_s \lambda_j \alpha_j(\widetilde{\lambda}_s) \geqslant \delta_k - 5\rho_k - \frac{32}{3}\frac{\lambda_{k+1}}{\lambda_k^2}\mathrm{osc}(\mathcal{R}) \geqslant \frac{1}{2}\delta_k$$

by our gap assumption (7.8). Hence,

$$\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s) = \min_{j\in\mathcal{J}} \left|\widetilde{\lambda}_s \lambda_j \alpha_j(\widetilde{\lambda}_s)\right| \geqslant \frac{1}{2}\delta_k.$$

Furthermore, (7.4) in Proposition 7.2 holds. To see this, note that the right-hand side of (7.2)

$$\begin{aligned}
&\frac{2\lambda_{k+1}}{\widetilde{\lambda}_s^2}\|\mathrm{Off}(\mathcal{V}_{\mathcal{J}\mathcal{J}})\| + \frac{13\lambda_{k+1}R_{\max}}{\widetilde{\lambda}_s^4}\mathrm{osc}(\mathcal{R}) \\
&\leqslant 2\left(\frac{100}{99}\right)^4 \frac{\lambda_{k+1}}{\lambda_k^2}\mathrm{osc}(\mathcal{R}) + \frac{13}{36}\left(\frac{100}{99}\right)^4 \frac{\lambda_{k+1}}{\lambda_k^2}\mathrm{osc}(\mathcal{R}) \\
&< 3\frac{\lambda_{k+1}}{\lambda_k^2}\mathrm{osc}(\mathcal{R}) < \frac{1}{2}\delta_k \leqslant \Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s).
\end{aligned}$$

We continue from the bound established in Proposition 7.2. On the event $\mathcal{E}_D$, substituting the uniform local law bound $\|U^\top \Xi(\widetilde{\lambda}_s)U\| \leqslant M/\widetilde{\lambda}_s^2$ and using $\widetilde{\lambda}_s \geqslant \frac{99}{100}\lambda_k$,

$$\begin{aligned}
\|U_J^\top \widetilde{u}_s\| \leqslant& \frac{4}{\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s)}\left(\frac{\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\|}{\widetilde{\lambda}_s} + \frac{13}{2}\frac{R_{\max}}{\widetilde{\lambda}_s^3}\mathrm{osc}(\mathcal{R}) + \widetilde{\lambda}_s^2\|U^\top \Xi(\widetilde{\lambda}_s)U\|\right) \\
&+ 2\sqrt{3}\left(\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\widetilde{\lambda}_s^2} + \frac{13}{2}\frac{R_{\max}}{\widetilde{\lambda}_s^4}\mathrm{osc}(\mathcal{R}) + \widetilde{\lambda}_s\|U^\top \Xi(\widetilde{\lambda}_s)U\|\right) \\
\leqslant& \frac{8}{\delta_k}\left(\frac{100}{99}\frac{\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\|}{\lambda_k} + \frac{13}{2}\left(\frac{100}{99}\right)^3 \frac{\mathrm{osc}(\mathcal{R})R_{\max}}{\lambda_k^3} + M\right) \\
&+ 2\sqrt{3}\left(\left(\frac{100}{99}\right)^2 \frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\lambda_k^2} + \frac{13}{2}\left(\frac{100}{99}\right)\frac{\mathrm{osc}(\mathcal{R})R_{\max}}{\lambda_k^2} + \frac{100}{99}\frac{M}{\lambda_k}\right).
\end{aligned}$$

Now the bound above simplifies to

$$\|U_J^\top \widetilde{u}_s\| < \frac{10}{\delta_k \lambda_k}\left(\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\| + 8\frac{R_{\max}}{\lambda_k^2}\mathrm{osc}(\mathcal{R})\right) + 3\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\lambda_k^2} + 10\frac{M}{\delta_k}.$$

Finally, we establish the bound

$$\begin{aligned}
\|U_J^\top \widetilde{U}_k\|_F &= \sqrt{\sum_{s=1}^k \|U_J^\top \widetilde{u}_s\|^2} \\
&\leqslant \frac{10\sqrt{k}}{\delta_k \lambda_k}\left(\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\| + 8\frac{R_{\max}}{\lambda_k^2}\mathrm{osc}(\mathcal{R})\right) + 3\sqrt{k}\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\lambda_k^2} + 10\sqrt{k}\frac{M}{\delta_k}.
\end{aligned}$$

Combining (7.7), for any $k \in [r_+]$, we always have

$$\|U_J^\top \widetilde{U}_k\|_F \leqslant \mathcal{B}_k + 10\sqrt{k}\frac{M}{\delta_k}, \tag{7.11}$$

where

$$\mathcal{B}_k := \frac{10\sqrt{k}}{\delta_k \lambda_k}\left(\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\| + 8\frac{R_{\max}}{\lambda_k^2}\operatorname{osc}(\mathcal{R})\right) + 4\sqrt{k}\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\lambda_k^2}.$$

Consequently, by (7.1),

$$\|\sin\angle(U_k, \widetilde{U}_k)\| \leqslant \mathcal{B}_k + 10\sqrt{k}\frac{M}{\delta_k} + 2\frac{\|E\|}{\lambda_k}.$$

From (7.3), using (7.11), on the event $\mathcal{E}_D$, we also have

$$\begin{aligned}
&\|\widetilde{U}_k - P_{U_k}\widetilde{U}_k\|_{2,\infty} \\
&\leqslant \|U\|_{2,\infty} \cdot \|U_J^\top \widetilde{U}_k\|_F + \frac{8}{3}\sqrt{k}\|U\|_{2,\infty}\frac{\operatorname{osc}(\mathcal{R})}{\widetilde{\lambda}_k^2} + 2\sqrt{\sum_{s=1}^k \widetilde{\lambda}_s^2 \max_{1\leqslant i\leqslant n}\left\|e_i^\top Q\Xi(\widetilde{\lambda}_s)U\right\|^2} \\
&\leqslant \|U\|_{2,\infty}\left(\mathcal{B}_k + 10\sqrt{k}\frac{M}{\delta_k} + 3\sqrt{k}\frac{\operatorname{osc}(\mathcal{R})}{\lambda_k^2}\right) + 2M\sqrt{\sum_{s=1}^k \frac{1}{\widetilde{\lambda}_s^2}} \\
&\leqslant \|U\|_{2,\infty}\left(\mathcal{B}_k + 10\sqrt{k}\frac{M}{\delta_k} + 3\sqrt{k}\frac{\operatorname{osc}(\mathcal{R})}{\lambda_k^2}\right) + 3\sqrt{k}\frac{M}{\lambda_k}
\end{aligned}$$

using $\widetilde{\lambda}_k \geqslant \frac{99}{100}\lambda_k$. Now, from (7.2), we obtain that

$$\begin{aligned}
&\min_{O\in\mathbb{O}(k)} \|\widetilde{U}_k - U_k O\|_{2,\infty} \\
&\leqslant \|\widetilde{U}_k - P_{U_k}\widetilde{U}_k\|_{2,\infty} + \|U_k\|_{2,\infty}\|\sin\angle(U_k, \widetilde{U}_k)\|^2 \\
&\leqslant 4\|U\|_{2,\infty}\left(\mathcal{B}_k + 20\sqrt{k}\frac{M}{\delta_k} + 3\sqrt{k}\frac{\operatorname{osc}(\mathcal{R})}{\lambda_k^2} + 27\frac{R_{\max}}{\lambda_k^2}\right) + 3\sqrt{k}\frac{M}{\lambda_k}
\end{aligned}$$

by our assumptions on λ_k and δ_k.

8 Proof of the isotropic local law: Theorem 4

Let $z \in \mathbb{C}$ satisfy $|z| \geqslant 6\sqrt{R_{\max}}$. Recall that

$$G(z) = (zI - E)^{-1}.$$

Denote

$$H \equiv H(z) = \operatorname{diag}(h_1(z), \ldots, h_n(z)), \qquad h_i \equiv h_i(z) = \sum_{j=1}^n \sigma_{ij}^2 \phi_j(z).$$

Then the QVE can be written as

$$\Phi(z) = (zI - H)^{-1}.$$

Since $(zI - E)G = I$ and $(zI - H)\Phi = I$, subtracting these identities, we have

$$z(G - \Phi) = EG - H\Phi = (E - H)G + H(G - \Phi).$$

Rearranging the terms, we further obtain $(zI - H)(G - \Phi) = (E - H)G$. Hence, for deterministic unit vectors $v, w \in \mathbb{R}^n$,

$$v^\top (G - \Phi)w = v^\top (z - H)^{-1}(E - H)Gw. \tag{8.1}$$

Set

$$\widehat{v}^\top := v^\top (z - H)^{-1}.$$

Then

$$v^\top (G - \Phi)w = \widehat{v}^\top (E - H)Gw.$$

We also define

$$L \equiv L(z) = \operatorname{diag}(l_1(z), \ldots, l_n(z)), \qquad l_i \equiv l_i(z) = \sum_{j=1}^{n} \sigma_{ij}^2 G_{jj}(z).$$

Then (8.1) gives the decomposition

$$v^\top (G - \Phi)w = \widehat{v}^\top (E - L)Gw + \widehat{v}^\top (L - H)Gw. \tag{8.2}$$

We start with a bound on $\|\widehat{v}\|$. By Lemma 5.1,

$$\|H\| \leqslant R_{\max} \max_i |\phi_i(z)| \leqslant \frac{3R_{\max}}{2|z|} \leqslant \frac{|z|}{4},$$

where the last inequality follows from $|z|^2 \geqslant 36R_{\max}$. Hence

$$\|\widehat{v}\| \leqslant \|(zI - H)^{-1}\| \leqslant \frac{1}{|z| - \|H\|} \leqslant \frac{4}{3|z|}. \tag{8.3}$$

With this $\|\widehat{v}\|$ bound, the control of the first error term in (8.2) reduces to the following proposition, which we prove in Appendix C. The proof follows the cumulant expansion strategy of He, Knowles and Rosenthal [62], which compares the random resolvent with its deterministic self-consistent approximation. Our application is simpler in two ways. First, we work only in the large-z outlier regime, where resolvent derivatives have elementary bounds. Second, we need only a second-order cumulant expansion with a controlled third-order remainder. The main challenge is handling highly inhomogeneous and sparse variance profiles.

PROPOSITION 8.1. *Fix $D > 0$. Let $v, w \in \mathbb{R}^n$ be deterministic unit vectors, and let $z \in \mathbb{C}_+$ satisfy $|z| \geqslant 6\sqrt{R_{\max}}$. Under Assumptions 1.1 and 1.2, with probability at least $1 - n^{-D-4}$,*

$$\left|v^\top (E - L)Gw\right| \leqslant \frac{\mathfrak{m}_D}{|z|} + Cn^{-500}.$$

Here, the Cn^{-500} term appears only in the general sub-Gaussian regime; in the bounded sparse-entry regime, it may be omitted.

The second error term in (8.2) is controlled using Proposition 8.1 and a deterministic stability argument for the diagonal self-consistent equation. Its proof is given in Appendix D.

PROPOSITION 8.2. *Fix $D > 0$. Under the assumption of Propositon 8.1, with probability at least $1 - n^{-D-4}$:*

$$\left|v^\top (L-H)Gw\right| \leqslant C\left(\frac{\mathfrak{m}_D}{|z|} + n^{-500}\right).$$

In the bounded sparse-entry regime, the n^{-500} term may be omitted.

Proof of Theorem 4. We use the decomposition (8.2). For the first error term in (8.2), Proposition 8.1 and (8.3) yield

$$\left|\widehat{v}^\top (E-L)Gw\right| \leqslant \|\widehat{v}\|\left(\frac{\mathfrak{m}_D}{|z|} + Cn^{-500}\right) \leqslant C\frac{\mathfrak{m}_D}{|z|^2} + C\frac{n^{-500}}{|z|}. \tag{8.4}$$

Similarly, for the second error term in (8.2), from Proposition 8.2 and (8.3), we get

$$\left|\widehat{v}^\top (L-H)Gw\right| \leqslant C\frac{\mathfrak{m}_D}{|z|^2} + C\frac{n^{-500}}{|z|}. \tag{8.5}$$

Combining (8.4) and (8.5) with (8.2), we obtain

$$\left|v^\top (G-\Phi)w\right| \leqslant C\frac{\mathfrak{m}_D}{|z|^2} + C\frac{n^{-500}}{|z|}.$$

Since $R_{\max} \leqslant n^{100}$ and $\sigma_{\max}^{-1} \leqslant n^{100}$, we have $\mathfrak{m}_D \geqslant cn^{-100}$ for an absolute constant $c > 0$. Absorbing the n^{-500}-term, we get

$$\left|v^\top (G-\Phi)w\right| \leqslant \frac{\mathfrak{m}_D}{|z|^2}.$$

This completes the proof. □

9 Proofs of the outlier eigenvalue locations: Theorems 5 and 6

9.1 Proof of Theorem 5

For notational convenience throughout this section, we replace the target index k with j. We first consider the case that $\lambda_j \geqslant 6\sqrt{R_{\max}}$ for $j \in [r_+]$ is a simple eigenvalue. In this section, we use the normalized version

$$\alpha_j(z) := \lambda_j^{-1} - u_j^\top \Phi(z) u_j,$$

so that $\widehat{\alpha}_j(z) = \lambda_j \alpha_j(z)$. The deterministic root ϑ_j defined in Proposition 5.1 is equivalently characterized by $\alpha_j(\vartheta_j) = 0$.

We still work on the event $\mathcal{E}_D$. However, we only need a weaker signal assumption $\lambda_j \geqslant 6\sqrt{R_{\max}} + 100\rho_j$ since the local law is applied on deterministic contours around λ_j, whereas in the eigenspace proof, it is applied at the random points $z = \widetilde{\lambda}_s$.

By Proposition 5.1, there exists a unique deterministic location ϑ_j satisfying

$$1 - \lambda_j u_j^\top \Phi(\vartheta_j) u_j = 0$$

with

$$\left| \vartheta_j - \lambda_j - \frac{\mathcal{V}_{jj}}{\lambda_j} \right| \leqslant 145 \frac{R_{\max}^2}{\lambda_j^3}.$$

In particular, since $\mathcal{V}_{jj} \geqslant 0$, we have

$$\vartheta_j \leqslant \lambda_j + \frac{\mathcal{V}_{jj}}{\lambda_j} + 145 \frac{R_{\max}^2}{\lambda_j^3} \leqslant \lambda_j + \frac{1}{36}\lambda_j + \frac{145}{36^2}\lambda_j < 1.16\lambda_j$$

and directly by Proposition 5.1,

$$\vartheta_j \geqslant \lambda_j \tag{9.1}$$

We will show that on the event $\mathcal{E}_D$, there exists exactly one eigenvalue $\widetilde{\lambda}_j$ lying inside the disk

$$\mathtt{D}(\vartheta_j, \rho) := \{z \in \mathbb{C} : |z - \vartheta_j| \leqslant \rho\}$$

for $\rho \equiv \rho_j$ defined in (9.2). Denote

$$K(z) := \Lambda^{-1} - U^\top \Phi(z) U \quad \text{and} \quad B(z) := U^\top (G(z) - \Phi(z)) U.$$

Set

$$g(z) := \det(K(z)) \quad \text{and} \quad f(z) = \det(\Lambda^{-1} - U^\top G(z) U) = \det(K(z) - B(z)).$$

By Lemma 21 from [85], the eigenvalues of $\widetilde{A}$ outside the $[-\|E\|, \|E\|]$ are the zeros of $f(z)$. Besides, the algebraic multiplicity of each eigenvalue matches the corresponding multiplicity of each zero. Next, we show that the zeros of $f(z)$ are close to the zeros of $g(z)$.

The following result is a special case of the generalized Rouché's theorem for operator functions (see [57]). We include a short proof for completeness.

LEMMA 9.1. *Let $\Gamma \subseteq \mathbb{C}$ be a simple closed contour. Assume $K(z)$ and $B(z)$ are analytic in a neighborhood of Γ and its interior. If*

$$\sup_{z \in \Gamma} \|K^{-1}(z) B(z)\| < 1,$$

then f and g have the same number of zeros inside Γ, counting multiplicity.

Proof. For $t \in [0,1]$, define $F_t(z) := \det(K(z) - tB(z))$. For $z \in \Gamma$,

$$K(z) - tB(z) = K(z)(I - tK^{-1}(z)B(z)).$$

Since $\|tK^{-1}(z)B(z)\| < 1$, $I - tK^{-1}(z)B(z)$ is invertible. Hence, $F_t(z) \neq 0$ on Γ for all $t \in [0,1]$. By the argument principle ([6, Section 5.2]), the number of zeros of F_t inside Γ is a constant in t. Taking $t = 0$ and $t = 1$ proves the claim. $\square$

We will take a circle

$$\Gamma_j(\rho) := \{z \in \mathbb{C} : |z - \vartheta_j| = \rho\}$$

with the radius

$$\rho \equiv \rho_j := 10M + 15\frac{\operatorname{osc}(\mathcal{R})}{\lambda_j}. \tag{9.2}$$

To apply Lemma 9.1 on $\Gamma_j(\rho)$, we bound

$$\sup_{z\in\Gamma_j(\rho)} \|K^{-1}(z)B(z)\| \leqslant \sup_{z\in\Gamma_j(\rho)} \|K^{-1}(z)\| \cdot \sup_{z\in\Gamma_j(\rho)} \|B(z)\|.$$

To bound $\sup_{z\in\Gamma_j(\rho)} \|B(z)\|$, first note that $\Gamma_j(\rho) \subseteq \mathcal{S}_{\mathrm{out}}$. To see this, since $\lambda_j \geqslant 6\sqrt{R_{\max}}+100\rho_j$ and $\rho_j \leqslant \lambda_j/100$, for every $z \in \Gamma_j(\rho_j)$,

$$\begin{aligned} |z| &\geqslant \vartheta_j - \rho_j \geqslant \lambda_j - \rho_j \geqslant 6\sqrt{R_{\max}}, \\ |z| &\leqslant \vartheta_j + \rho_j \leqslant 1.16\lambda_j + \frac{\lambda_j}{100} \leqslant 2R_{\max}^3 \end{aligned}$$

Hence, on the event $\mathcal{E}_D$, we get

$$\sup_{z\in\Gamma_j(\rho)} \|B(z)\| \leqslant \left(\frac{100}{99}\right)^2 \frac{M}{\vartheta_j^2}. \tag{9.3}$$

To bound $\sup_{z\in\Gamma_j(\rho)} \|K^{-1}(z)\|$, we rewrite $K(z)$ for $z \in \Gamma_j(\rho)$ in block form. Without loss of generality, by relabeling the columns of U and the diagonal entries of Λ, we may assume that the index j is the first one. Namely, write

$$U = (u_j, U_{-j}) \quad \text{and} \quad \Lambda = \begin{pmatrix} \lambda_j & 0 \\ 0 & \Lambda_{-j} \end{pmatrix}.$$

Thus,

$$K(z) = \Lambda^{-1} - U^\top \Phi(z) U = \begin{pmatrix} \alpha_j(z) & -b_j(z)^\top \\ -b_j(z) & D_j(z) \end{pmatrix},$$

where

$$\alpha_j(z) := \lambda_j^{-1} - u_j^\top \Phi(z) u_j, \quad b_j(z) := U_{-j}^\top \Phi(z) u_j, \quad D_j(z) = \Lambda_{-j}^{-1} - U_{-j}^\top \Phi(z) U_{-j}.$$

The following lemma provides the estimates of $|\alpha_j(z)|$, $\|b_j(z)\|$, and the smallest singular value $s_{\min}(D_j(z))$ on the circle $\Gamma_j(\rho)$. Its proof is deferred to Section 9.1.2.

LEMMA 9.2. *Fix $j \in [r_+]$. Assume $\lambda_j \geqslant 6\sqrt{R_{\max}} + 100\rho$ and λ_j is a simple eigenvalue of A. Assume the following gap condition holds:*

$$\min_{l<j}(\vartheta_l - \vartheta_j) \geqslant 50\rho \quad \textit{and} \quad \min_{l\in\mathcal{O}_+,\, l>j}(\vartheta_j - \vartheta_l) \geqslant 50\rho. \tag{9.4}$$

where $\mathcal{O}_+ := \{l \in [r_+] : \lambda_l \geqslant 6\sqrt{R_{\max}}\}$ is the index set for the supercritical eigenvalues. Then for $z \in \Gamma_j(\rho)$, we have

$$\frac{\rho}{4\vartheta_j^2} \leqslant |\alpha_j(z)| \leqslant \frac{3\rho}{\vartheta_j^2}, \tag{9.5}$$

$$\|b_j(z)\| \leqslant \frac{1}{10}\frac{\rho}{\vartheta_j^2}, \tag{9.6}$$

$$s_{\min}(D_j(z)) \geqslant 3\frac{\rho}{\vartheta_j^2}. \tag{9.7}$$

In particular, (9.7) suggests that $D_j(z)$ is invertible on $\Gamma_j(z)$ and

$$\|D_j^{-1}(z)\| \leqslant \frac{1}{3}\frac{\vartheta_j^2}{\rho}. \tag{9.8}$$

Define

$$h_j(z) := \alpha_j(z) - b_j(z)^\top D_j^{-1}(z) b_j(z)$$

for $z \in \Gamma_j(\rho)$. By Lemma 9.2, $h_j(z)$ is well-defined on $\Gamma_j(\rho)$. Besides,

$$|h_j(z)| \geqslant |\alpha_j(z)| - \|b_j(z)\|^2 \|D_j^{-1}(z)\| \geqslant \frac{\rho}{4\vartheta_j^2} - \frac{1}{300}\frac{\rho}{\vartheta_j^2} = \frac{37}{150}\frac{\rho}{\vartheta_j^2} > 0$$

and

$$|h_j^{-1}(z)| \leqslant \frac{150}{37}\frac{\vartheta_j^2}{\rho}. \tag{9.9}$$

Now we are ready to estimate $\|K^{-1}(z)\|$ on $\Gamma_j(\rho)$. By Schur's complement,

$$K(z)^{-1} = \begin{pmatrix} h_j^{-1}(z) & h_j^{-1}(z) b_j(z)^\top D_j^{-1}(z) \\ h_j^{-1}(z) D_j^{-1}(z) b_j(z) & D_j^{-1}(z) + h_j^{-1}(z) D_j^{-1}(z) b_j(z) b_j(z)^\top D_j^{-1}(z) \end{pmatrix}$$

We decompose $K(z)^{-1}$ as

$$K(z)^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & D_j^{-1}(z) \end{pmatrix} + h_j^{-1}(z) \begin{pmatrix} 1 & b_j(z)^\top D_j^{-1}(z) \\ D_j^{-1}(z) b_j(z) & D_j^{-1}(z) b_j(z) b_j(z)^\top D_j^{-1}(z) \end{pmatrix}.$$

Note that by setting $\mathtt{v} := b_j(z)^\top D_j^{-1}(z)$, we have

$$\mathtt{H} := \begin{pmatrix} 1 & b_j(z)^\top D_j^{-1}(z) \\ D_j^{-1}(z) b_j(z) & D_j^{-1}(z) b_j(z) b_j(z)^\top D_j^{-1}(z) \end{pmatrix} = \begin{pmatrix} 1 & \mathtt{v}^\top \\ \mathtt{v} & \mathtt{v}\mathtt{v}^\top \end{pmatrix} = \begin{pmatrix} 1 \\ \mathtt{v} \end{pmatrix} \begin{pmatrix} 1 & \mathtt{v}^\top \end{pmatrix}.$$

Hence,

$$\|\mathtt{H}\| = 1 + \|\mathtt{v}\|^2 = 1 + \|b_j(z)^\top D_j^{-1}(z)\|^2 \leqslant 1 + \|b_j(z)\|^2 \cdot \|D_j^{-1}(z)\|^2.$$

It follows that

$$\begin{aligned} \|K(z)^{-1}\| &\leqslant \|D_j^{-1}(z)\| + |h_j^{-1}(z)| \cdot \|\mathtt{H}\| \\ &\leqslant \|D_j^{-1}(z)\| + |h_j^{-1}(z)| \left(1 + \|b_j(z)\|^2 \cdot \|D_j^{-1}(z)\|^2\right). \end{aligned} \tag{9.10}$$

Plugging the bounds in Lemma 9.2, (9.8), and (9.9), we get

$$\|K(z)^{-1}\| \leqslant \frac{1}{3}\frac{\vartheta_j^2}{\rho} + \frac{150}{37}\frac{\vartheta_j^2}{\rho}\left(1 + \frac{1}{900}\right) < \frac{9}{2}\frac{\vartheta_j^2}{\rho}.$$

Together with (9.3), we arrive at

$$\sup_{z \in \Gamma_j(\rho)} \|K^{-1}(z)B(z)\| \leqslant \sup_{z \in \Gamma_j(\rho)} \|K^{-1}(z)\| \cdot \sup_{z \in \Gamma_j(\rho)} \|B(z)\| \leqslant \frac{9}{2}\frac{\vartheta_j^2}{\rho} \cdot \left(\frac{100}{99}\right)^2 \frac{M}{\vartheta_j^2} < 1,$$

where we used $\rho \geqslant 5M$ in the last inequality. Hence, by Lemma 9.1, the functions $f(z) = \det(K(z) - B(z))$ and $g(z) = \det(K(z))$ have the same number of zeros inside $\Gamma_j(\rho)$, counting multiplicities. This comparison alone, however, does not yet identify that number, since $K(z)$ is not diagonal. We therefore perform the zero counting by comparing $K(z)$ with its diagonal approximation.

Recall that $\mathtt{D}(\vartheta_l, \rho) := \{z \in \mathbb{C} : |z - \vartheta_l| \leqslant \rho\}$ denotes the disk centered as ϑ_l with radius ρ. Set

$$D_j^+ := \bigcup_{l<j} \mathtt{D}(\vartheta_l, \rho), \qquad D_j^- := D_j^+ \cup \mathtt{D}(\vartheta_j, \rho), \tag{9.11}$$

and let $\mathcal{C}_j^\pm := \partial D_j^\pm$ be positively oriented on each connected component. By our gap assumption (9.4), the disk $\mathtt{D}(\vartheta_j, \rho)$ is disjoint from D_j^+. Moreover, the disks corresponding to roots ϑ_l with $l < j$ lie strictly to the right of $\mathtt{D}(\vartheta_j, \rho)$, while the supercritical roots ϑ_l with $l > j$ lie strictly to its left.

Define the diagonal matrix of $K(z)$ as

$$A_0(z) := \operatorname{diag}(\alpha_1(z), \ldots, \alpha_r(z)) \quad \text{with} \quad \alpha_l(z) := \lambda_l^{-1} - u_l^\top \Phi(z) u_l. \tag{9.12}$$

We show the next result whose proof is deferred to Section 9.1.1

LEMMA 9.3. *Under the assumption of Lemma 9.2, we have*

$$\sup_{z \in \mathcal{C}_j^\pm} \|A_0(z)^{-1}(K(z) - A_0(z))\| < 1, \qquad \sup_{z \in \mathcal{C}_j^\pm} \|K(z)^{-1}B(z)\| < 1.$$

Applying Lemma 9.1 twice on the contours $\mathcal{C}_j^\pm$, first to $A_0(z)$ and $K(z)$, and then to $K(z)$ and $K(z) - B(z)$, shows that

$$\det(A_0(z)), \qquad g(z) = \det(K(z)), \qquad f(z) = \det(K(z) - B(z))$$

have the same number of zeros inside each of the regions D_j^+ and D_j^-, counting multiplicities.

Now

$$\det(A_0(z)) = \prod_{l=1}^r \alpha_l(z),$$

and among these factors on the right-hand side, only $\alpha_l(z)$ with $l \in \mathcal{O}_+$ have zeros in the supercritical regime, namely at $z = \vartheta_l$ (see Proposition 5.1). Since λ_j is supercritical, all λ_l with $l < j$

are also supercritical. Hence, by construction, $\det(A_0(z))$ has exactly $j-1$ zeros inside D_j^+ and exactly j zeros inside D_j^-. Consequently, $f(z)$ has exactly $j-1$ zeros inside D_j^+ and exactly j zeros inside D_j^-.

Since $D_j^- \backslash D_j^+ = \mathtt{D}(\vartheta_j, \rho)$ and $\Gamma_j(\rho) = \partial\, \mathtt{D}(\vartheta_j, \rho)$, it follows that $f(z) = \det(K(z) - B(z))$ has exactly one zero in $\mathtt{D}(\vartheta_j, \rho)$. This zero is the same zero enclosed by $\Gamma_j(\rho) = \partial \mathtt{D}(\vartheta_j, \rho)$. Denote it by $\widehat{\lambda}$. As noted above, the zeros of f outside $[-\|E\|, \|E\|]$ coincide with the eigenvalues of $\widetilde{A} = A + E$, with matching algebraic multiplicities. Moreover, by (9.1) and $\rho \leqslant \lambda_j/100$, the disk $\mathtt{D}(\vartheta_j, \rho)$ lies strictly to the right of the noise bulk $[-\|E\|, \|E\|]$. Therefore $\widehat{\lambda}$ is an eigenvalue of $\widetilde{A}$.

It remains only to identify its index. The zero count above shows that exactly $j-1$ zeros of f lie in D_j^+, which is strictly to the right of $\mathtt{D}(\vartheta_j, \rho)$. The gap condition also places all remaining supercritical roots ϑ_l with $l > j$ strictly to the left of $\mathtt{D}(\vartheta_j, \rho)$; applying the same comparison on the corresponding separating contours rules out additional zeros of f in the gaps between these root neighborhoods. Finally, all non-outlier eigenvalues of $\widetilde{A}$ lie in $[-\|E\|, \|E\|]$, which is also strictly to the left of $\mathtt{D}(\vartheta_j, \rho)$. Hence there are exactly $j-1$ eigenvalues of $\widetilde{A}$ larger than $\widehat{\lambda}$. Hence,

$$\widehat{\lambda} = \widetilde{\lambda}_j.$$

9.1.1 Proof of Lemma 9.3

We estimate $\|A_0(z)^{-1}\|$, $\|K(z) - A_0(z)\|$, and $\|K^{-1}(z)\|$ respectively for $z \in \mathcal{C}_j^{\pm}$. Since

$$K(z) - A_0(z) = -\mathrm{Off}(U^\top \Phi(z) U),$$

we have

$$\|K(z) - A_0(z)\| \leqslant \|\mathrm{Off}(U^\top \Phi(z) U)\|.$$

Claim. For $z \in \mathcal{C}_j^{\pm}$, the following estimates hold:

$$|\alpha_l(z)| \geqslant \frac{1}{2} \frac{\rho}{|z|^2}, \tag{9.13}$$

$$\|b_j(z)\| \leqslant \frac{2}{25} \frac{\rho}{|z|^2}, \tag{9.14}$$

$$\|\mathrm{Off}(U^\top \Phi(z) U)\| \leqslant \frac{1}{5} \frac{\rho}{|z|^2}, \tag{9.15}$$

$$\|\mathrm{Off}(U^\top \varepsilon(z) U)\| \leqslant \frac{1}{10} \frac{\rho}{|z|^2}. \tag{9.16}$$

We defer the proof of this claim to the end of this section. Using these bounds, we first have

$$\|A_0(z)^{-1}\| = \max_{l \in [r]} |\alpha_l^{-1}(z) = \frac{1}{|\min_{l \in [r]} \alpha_l(z)|} \leqslant 2 \frac{|z|^2}{\rho}$$

Hence,

$$\begin{aligned}\sup_{z \in \mathcal{C}_j^{\pm}} \|A_0(z)^{-1}(K(z) - A_0(z))\| &\leqslant \sup_{z \in \mathcal{C}_j^{\pm}} \|A_0(z)^{-1}\| \sup_{z \in \mathcal{C}_j^{\pm}} \|K(z) - A_0(z))\| \\ &\leqslant 2 \frac{|z|^2}{\rho} \cdot \frac{1}{5} \frac{\rho}{|z|^2} < 1.\end{aligned}$$

Next, we estimate $\|K(z)^{-1}\|$ for $z \in \mathcal{C}_j^{\pm}$, which is similar to the previous estimates on $\Gamma_j(\rho)$. Since $\left\|\mathrm{Off}\left(U_{-j}^\top \Phi(z) U_{-j}\right)\right\| \leqslant \left\|\mathrm{Off}\left(U^\top \Phi(z) U\right)\right\|$, from (9.22) and by the **Claim**, we get

$$
\begin{aligned}
s_{\min}(D_j(z)) &\geqslant \min_{l \in [r]\backslash\{j\}} |\alpha_l(z)| - \left\|\mathrm{Off}\left(U_{-j}^\top \Phi(z) U_{-j}\right)\right\| \\
&\geqslant \frac{1}{2}\frac{\rho}{|z|^2} - \frac{1}{5}\frac{\rho}{|z|^2} = \frac{3}{10}\frac{\rho}{|z|^2}.
\end{aligned}
$$

and thus

$$
\|D_j^{-1}(z)\| \leqslant \frac{10}{3}\frac{|z|^2}{\rho}.
$$

Recall that $h_j(z) = \alpha_j(z) - b_j(z)^\top D_j^{-1}(z) b_j(z)$. By the claim,

$$
|h_j(z)| \geqslant |\alpha_j(z)| - \|b_j(z)\|^2 \|D_j^{-1}(z)\| \geqslant \frac{1}{2}\frac{\rho}{|z|^2} - \left(\frac{2}{25}\frac{\rho}{|z|^2}\right)^2 \frac{3}{10}\frac{|z|^2}{\rho} = \frac{359}{750}\frac{\rho}{|z|^2}.
$$

Hence, $|h_j^{-1}(z)| \leqslant \frac{750}{359}\frac{|z|^2}{\rho}$ and by (9.10),

$$
\begin{aligned}
\|K(z)^{-1}\| &\leqslant \|D_j^{-1}(z)\| + |h_j^{-1}(z)| \left(1 + \|b_j(z)\|^2 \cdot \|D_j^{-1}(z)\|^2\right) \\
&\leqslant \frac{10}{3}\frac{|z|^2}{\rho} + \frac{750}{359}\frac{|z|^2}{\rho}\left(1 + \left(\frac{2}{25}\frac{\rho}{|z|^2}\right)^2 \left(\frac{3}{10}\frac{|z|^2}{\rho}\right)^2\right) \\
&< 5.5\frac{|z|^2}{\rho}.
\end{aligned}
$$

Finally,

$$
\sup_{z \in \mathcal{C}_j^{\pm}} \|K^{-1}(z) B(z)\| \leqslant \sup_{z \in \mathcal{C}_j^{\pm}} \|K^{-1}(z)\| \cdot \sup_{z \in \mathcal{C}_j^{\pm}} \|B(z)\| \leqslant 5.5\frac{|z|^2}{\rho}\frac{M}{|z|^2} < 1,
$$

where we used $\rho \geqslant 10M$ in the last inequality.

To finish the proof, it remains to prove the claim.

Proof of (9.13) We bound $|\alpha_l(z)|$ for $z \in \mathcal{C}_j^{\pm}$. The proof is similar to that of (9.29). We split the discussion into three cases: $l \in \mathcal{O}_+$, $l \in [r_+]\backslash\mathcal{O}_+$, and $l \in [r]\backslash[r_+]$.

Case 1: $l \in \mathcal{O}_+$. We have $\lambda_l \geqslant 6\sqrt{R_{\max}}$. By Proposition 5.1, define ϑ_l as the unique solution to $\alpha_l(z) = 0$. Also, $\vartheta_l \geqslant 5\sqrt{R_{\max}}$.

For $z = x + \sqrt{-1}y \in \mathcal{C}_j^{\pm}$, we have $|y| \leqslant \rho$ and $x \geqslant \vartheta_j - \rho \geqslant \frac{99}{100}\vartheta_j > 4\sqrt{R_{\max}}$. Note that using the same estimates as (9.20) and (9.21), we have

$$
\frac{4}{5}\frac{1}{x^2} \leqslant \alpha_l'(x) = -u_l^\top \Phi'(x) u_l \leqslant \frac{3}{x^2}. \tag{9.17}
$$

Indeed, we refined the constant for the lower bound using $x \geqslant 4\sqrt{R_{\max}}$:

$$\alpha_l'(x) = -u_l^\top \Phi'(x) u_l \geqslant \frac{1}{x^2} - \frac{243}{5}\frac{R_{\max}^2}{x^6} \geqslant \frac{1}{x^2}\left(1 - \frac{243}{5 \cdot 256}\right) > \frac{4}{5x^2}.$$

Since $\alpha_l(\vartheta_l) = 0$, from $|\alpha_l(x)| = |\int_{\vartheta_l}^x \alpha_l'(t)\, dt|$, we get

$$\frac{4}{5}\frac{|x - \vartheta_l|}{x\vartheta_l} \leqslant |\alpha_l(x)| \leqslant 3\frac{|x - \vartheta_l|}{x\vartheta_l}. \tag{9.18}$$

Now the Taylor expansion of $\alpha_l(z)$ at x yields:

$$\alpha_l(z) = \alpha_l(x) + \sqrt{-1}y\alpha_l'(x) + \frac{1}{2}\alpha_l''(\zeta)(\sqrt{-1}y)^2,$$

where ζ lies in a vertical segment connecting x and z. In particular, $|\zeta| \geqslant x \geqslant 4\sqrt{R_{\max}}$. It follows from part (i) of Lemma 5.1 that

$$|\alpha_l''(\zeta)| = |u_l^\top \Phi''(\zeta) u_l| \leqslant \|\Phi''(\zeta)\| \leqslant \frac{28}{|\zeta|^3} \leqslant \frac{28}{x^3}.$$

Next, plugging in (9.17) and (9.18) yields

$$\begin{aligned}|\alpha_l(x) + \sqrt{-1}y\alpha_l'(x)| &= \sqrt{\alpha_l(x)^2 + y^2\alpha_l'(x)^2} \\ &\geqslant \sqrt{\frac{16}{25}\frac{(x-\vartheta_l)^2}{x^2\vartheta_l^2} + \frac{16}{25}\frac{y^2}{x^4}} = \frac{4}{5}\frac{1}{x^2}\sqrt{(x-\vartheta_l)^2\frac{x^2}{\vartheta_l^2} + y^2}.\end{aligned}$$

If $\vartheta_l \leqslant \frac{100}{99}x$, then $x/\vartheta_l \geqslant 0.99$ and

$$|\alpha_l(x) + \sqrt{-1}y\alpha_l'(x)| \geqslant \frac{4}{5x^2}0.99\sqrt{(x-\vartheta_l)^2 + y^2} = 0.792\frac{|z - \vartheta_l|}{x^2} \geqslant 0.792\frac{\rho}{x^2}.$$

If $\vartheta_l > \frac{100}{99}x$, we use the fact that the function $f(t) = \frac{t-x}{xt}$ is increasing for $t > x$. Its minimum value occurs at the boundary $t = \frac{100}{99}x$. By (9.18),

$$|\alpha_l(x) + \sqrt{-1}y\alpha_l'(x)| \geqslant |\alpha_l(x)| \geqslant \frac{4}{5}\frac{\vartheta_l - x}{x\vartheta_l} \geqslant \frac{4}{5}\frac{\frac{100}{99}x - x}{x \cdot \frac{100}{99}x} = \frac{4x}{500x^2}.$$

Since $x \geqslant 99\rho$, we have $\frac{4x}{500x^2} \geqslant \frac{4 \cdot 99\rho}{500x^2} = 0.792\frac{\rho}{x^2}$.

Hence, we obtain

$$\begin{aligned}|\alpha_l(z)| &\geqslant |\alpha_l(x) + \sqrt{-1}y\alpha_l'(x)| - 14\frac{y^2}{x^3} \\ &\geqslant 0.792\frac{\rho}{x^2} - \frac{14}{99}\frac{\rho}{x^2} > \frac{\rho}{2x^2} \geqslant \frac{\rho}{2|z|^2},\end{aligned}$$

where we used $|z| \geqslant x \geqslant 99\rho \geqslant 99|y|$.

Case 2: $l \in [r_+]\backslash\mathcal{O}_+$. The lower bound on $|\alpha_l(z)|$ follows the same approach as (9.27). We skip the proof. In this case, we have

$$|\alpha_l(z)| \geqslant 70\frac{\rho}{\vartheta_j^2} \geqslant 70\left(\frac{99}{100}\right)^2\frac{\rho}{|z|^2} > 60\frac{\rho}{|z|^2}.$$

Case 3: $l \in [r]\backslash[r_+]$. We have $\lambda_l < 0$. For $z = x + \sqrt{-1}y \in \mathcal{C}_j^{\pm}$, we have $x \geqslant \vartheta_j - \rho \geqslant \frac{99}{100}\vartheta_j 4\sqrt{R_{\max}}$. From (5.3), we expand

$$\alpha_l(z) = \frac{1}{\lambda_l} - u_l^\top\Phi(z)u_l = \frac{1}{\lambda_l} - \frac{1}{z} - \frac{1}{z^3}u_l^\top\mathcal{R}u_l - u_l^\top\varepsilon(z)u_l.$$

Note that

$$\left|\frac{1}{\lambda_l} - \frac{1}{z}\right| \geqslant \left|\mathrm{Re}\left(\frac{1}{\lambda_l} - \frac{1}{z}\right)\right| = \frac{1}{|\lambda_l|} + \frac{x}{|z|^2} > \frac{x}{|z|^2}.$$

Furthermore, we have

$$\left|\frac{1}{z^3}u_l^\top\mathcal{R}u_l + u_l^\top\varepsilon(z)u_l\right| \leqslant \frac{R_{\max}}{|z|^3} + \frac{45}{8}\frac{R_{\max}^2}{|z|^5} \leqslant \frac{173}{128}\frac{R_{\max}}{|z|^3} \leqslant \frac{173}{128\cdot 16}\frac{x}{|z|^2}$$

using the bound $|z| \geqslant x \geqslant 4\sqrt{R_{\max}}$. Hence,

$$|\alpha_l(z)| \geqslant \frac{x}{|z|^2} - \frac{173}{128\cdot 16}\frac{x}{|z|^2} > 90\frac{\rho}{|z|^2}.$$

Combining the above estimates, we have proved (9.13).

Proof of (9.14) The proof is almost identical to that of (9.6). We briefly it here. Following the same approach as (9.6), we first get

$$\|b_j(z)\| \leqslant \frac{\mathrm{osc}(\mathcal{R})}{2|z|^3} + \frac{15R_{\max}}{2|z|^5}\mathrm{osc}(\mathcal{R}) < \frac{\mathrm{osc}(\mathcal{R})}{|z|^3}.$$

Since $|z| \geqslant \vartheta_j - \rho \geqslant \frac{99}{100}\vartheta_j$ and

$$\rho \geqslant 13\frac{\mathrm{osc}(\mathcal{R})}{\vartheta_j},$$

we further get

$$\frac{\mathrm{osc}(\mathcal{R})}{|z|^3} \leqslant \frac{100}{99}\frac{\mathrm{osc}(\mathcal{R})}{\vartheta_j}\frac{1}{|z|^2} \leqslant \frac{100}{99\cdot 13}\frac{\rho}{|z|^2}, \tag{9.19}$$

and hence

$$\|b_j(z)\| \leqslant \frac{2}{25}\frac{\rho}{|z|^2}.$$

Proofs of (9.15) **and** (9.16) The bound for $\mathrm{Off}\big(U^\top \Phi(z) U\big)$ are almost identical to the bound for $\mathrm{Off}\big(U_{-j}^\top \Phi(z) U_{-j}\big)$ derived in (9.7). We briefly sketch it here. It follows from the same steps that

$$\|\mathrm{Off}(U_{-j}^\top \varepsilon(z) U_{-j})\| \leqslant \frac{15 R_{\max}}{\sqrt{2}|z|^5}\mathrm{osc}(\mathcal{R}) \leqslant \frac{15}{9\sqrt{2}}\frac{\mathrm{osc}(\mathcal{R})}{|z|^3} \leqslant \frac{1}{10}\frac{\rho}{|z|^2}$$

and

$$\begin{aligned}\big\|\mathrm{Off}\big(U_{-j}^\top \Phi(z) U_{-j}\big)\big\| &\leqslant \frac{\mathrm{osc}(\mathcal{R})}{2|z|^3} + \frac{15 R_{\max}}{\sqrt{2}|z|^5}\mathrm{osc}(\mathcal{R}) \\ &\leqslant \frac{\mathrm{osc}(\mathcal{R})}{|z|^3}\left(\frac{1}{2} + \frac{15}{9\sqrt{2}}\right) \leqslant \frac{1}{5}\frac{\rho}{|z|^2}\end{aligned}$$

where we used (9.19).

9.1.2 Proof of Lemma 9.2

Recall from (9.2) that

$$\rho = 5M + 15\frac{\mathrm{osc}(\mathcal{R})}{\lambda_j}.$$

By our assumption, $\rho \leqslant \lambda_j/100$. From the estimates in (9.1), we also have

$$\rho \geqslant 15 \cdot 0.87\frac{\mathrm{osc}(\mathcal{R})}{\vartheta_j} > 13\frac{\mathrm{osc}(\mathcal{R})}{\vartheta_j}.$$

For $z \in \Gamma_j(\rho)$ satisfying $|z - \vartheta_j| = \rho$, we have

$$\frac{99}{100}\vartheta_j \leqslant \vartheta_j - \rho \leqslant |z| \leqslant \vartheta_j - \rho \leqslant \frac{101}{100}\vartheta_j.$$

Proof of (9.5)**.** We first estimate $|\alpha_j(z)|$. Recall that $\alpha_j(\vartheta_j) = 0$. By Taylor's theorem,

$$\alpha_j(z) = \alpha_j'(\vartheta_j)(z - \vartheta_j) + \mathtt{E}_2,$$

where

$$|\mathtt{E}_2| \leqslant \frac{1}{2}|z - \vartheta_j|^2 \max_{|\zeta - \vartheta_j| \leqslant \rho} |\alpha_j''(\zeta)| = \frac{1}{2}\rho^2 \max_{|\zeta - \vartheta_j| \leqslant \rho} |\alpha_j''(\zeta)|.$$

By part (ii) of Lemma 5.1 (see also (A.15)),

$$\alpha_j'(\vartheta_j) = -u_j^\top \Phi'(\vartheta_j) u_j \geqslant \frac{1}{\vartheta_j^2} - \frac{243}{5}\frac{R_{\max}^2}{\vartheta_j^6} \geqslant \frac{1}{\vartheta_j^2}\left(1 - \frac{243}{5 \cdot 81}\right) = \frac{2}{5\vartheta_j^2}. \tag{9.20}$$

For ζ satisfying $|\zeta - \vartheta_j| \leqslant \rho$, we have $|\zeta| \geqslant \vartheta_j - \rho \geqslant \frac{99}{100}\vartheta_j > 4\sqrt{R_{\max}}$. It follows from part (i) of Lemma 5.1 that

$$|\alpha_j''(\zeta)| = |u_j^\top \Phi''(\zeta) u_j| \leqslant \|\Phi''(\zeta)\| \leqslant \frac{28}{|\zeta|^3} \leqslant 28\left(\frac{100}{99}\right)^3 \frac{1}{\vartheta_j^3} < \frac{29}{\vartheta_j^3}.$$

Thus,

$$|\mathrm{E}_2| \leqslant \frac{29}{2}\frac{\rho^2}{\vartheta_j^3} \leqslant \frac{29}{200}\frac{\rho}{\vartheta_j^2}$$

and

$$|\alpha_j(z)| \geqslant \frac{2\rho}{5\vartheta_j^2} - \frac{29}{200}\frac{\rho}{\vartheta_j^2} > \frac{\rho}{4\vartheta_j^2}.$$

For the upper bound, observe that

$$|\alpha_j(z)| = |\alpha_j(z) - \alpha_j(\vartheta_j)| \leqslant \max_{|\zeta - \vartheta_j| \leqslant \rho} |\alpha_j'(\zeta)| \cdot \rho.$$

By part (ii) of Lemma 5.1 (see also (A.15)), we have

$$|\alpha_j'(\zeta)| \leqslant \frac{1}{|\zeta|^2} + \frac{3\mathcal{V}_{jj}}{|\zeta|^4} + \frac{243}{5}\frac{R_{\max}^2}{|\zeta|^6} \leqslant \frac{2}{|\zeta|^2} \leqslant \frac{3}{\vartheta_j^2} \tag{9.21}$$

using $|\zeta| \geqslant \frac{99}{100}\vartheta_j$ and $\vartheta_j > 5\sqrt{R_{\max}}$. Hence,

$$|\alpha_j(z)| \leqslant \frac{3\rho}{\vartheta_j^2}.$$

Proof of (9.6)**.** Next, we estimate $\|b_j(z)\| = \|U_{-j}^\top \Phi(z) u_j\|$. Since $U_{-j}^\top u_j = 0$, using the expansion of $\Phi(z)$ in (5.3), we immediately get

$$b_j(z) = \frac{1}{z^3} U_{-j}^\top \mathcal{R} u_j + U_{-j}^\top \varepsilon(z) u_j.$$

Thus,

$$\|b_j(z)\| \leqslant \frac{\|U_{-j}^\top \mathcal{R} u_j\|}{|z|^3} + \|U_{-j}^\top \varepsilon(z) u_j\|.$$

Since $U_{-j}^\top u_j = 0$,

$$\|U_{-j}^\top \mathcal{R} u_j\| = \|U_{-j}^\top (\mathcal{R} - cI) u_j\| \leqslant \|\mathcal{R} - cI\| \leqslant \frac{1}{2}\operatorname{osc}(\mathcal{R})$$

by selecting $c = \frac{1}{2}(R_{\max} + R_{\min})$. Similarly, we also have

$$\|U_{-j}^\top \varepsilon(z) u_j\| \leqslant \frac{1}{2}\operatorname{osc}(\varepsilon(z)) \leqslant \frac{15 R_{\max}}{2|z|^5}\operatorname{osc}(\mathcal{R}),$$

where the last inequality follows from Lemma 5.2. Thus

$$\|b_j(z)\| \leqslant \frac{\operatorname{osc}(\mathcal{R})}{2|z|^3} + \frac{15R_{\max}}{2|z|^5}\operatorname{osc}(\mathcal{R}) < \frac{\operatorname{osc}(\mathcal{R})}{|z|^3} \leqslant \left(\frac{100}{99}\right)^2 \frac{\operatorname{osc}(\mathcal{R})}{\vartheta_j^3} < \frac{1}{10}\frac{\rho}{\vartheta_j^2}$$

by our assumption $\rho \geqslant 15\frac{\operatorname{osc}(\mathcal{R})}{\lambda_j}$.

Proof of (9.7)**.** Finally, we show the lower bound on $s_{\min}(D_j(z))$. For each $l \in [r]$, set

$$\alpha_l(z) := \lambda_l^{-1} - u_l^\top \Phi(z) u_l.$$

We rewrite

$$\begin{aligned} D_j(z) &= \Lambda_{-j}^{-1} - U_{-j}^\top \Phi(z) U_{-j} \\ &= \Lambda_{-j}^{-1} - \operatorname{diag}(u_l^\top \Phi(z) u_l)_{l \in [r]\setminus\{j\}} - \operatorname{Off}\big(U_{-j}^\top \Phi(z) U_{-j}\big) \\ &= \operatorname{diag}(\alpha_l(z))_{l \in [r]\setminus\{j\}} - \operatorname{Off}\big(U_{-j}^\top \Phi(z) U_{-j}\big). \end{aligned}$$

Thus,

$$s_{\min}(D_j(z)) \geqslant \min_{l \in [r]\setminus\{j\}} |\alpha_l(z)| - \Big\|\operatorname{Off}\big(U_{-j}^\top \Phi(z) U_{-j}\big)\Big\|. \tag{9.22}$$

We first bound $\operatorname{Off}\big(U_{-j}^\top \Phi(z) U_{-j}\big)$. By (5.3),

$$\begin{aligned} \operatorname{Off}\big(U_{-j}^\top \Phi(z) U_{-j}\big) &= \operatorname{Off}\big(\frac{1}{z} I + \frac{1}{z^3} U_{-j}^\top \mathcal{R} U_{-j} + U_{-j}^\top \varepsilon(z) U_{-j}\big) \\ &= \operatorname{Off}\big(\frac{1}{z^3} U_{-j}^\top \mathcal{R} U_{-j}\big) + \operatorname{Off}\big(U_{-j}^\top \varepsilon(z) U_{-j}\big). \end{aligned} \tag{9.23}$$

Note that

$$\operatorname{Off}(U_{-j}^\top (\mathcal{R} - c_1 I) U_{-j}) = \operatorname{Off}(U_{-j}^\top \mathcal{R} U_{-j}).$$

Choosing $c_1 = \frac{1}{2}(R_{\max} + R_{\min})$, we get

$$\Big\|\operatorname{Off}\big(U_{-j}^\top \mathcal{R} U_{-j}\big)\Big\| \leqslant \|\mathcal{R} - c_1 I\| \leqslant \frac{1}{2}\operatorname{osc}(\mathcal{R}).$$

Similarly, we also have

$$\|\operatorname{Off}(U_{-j}^\top \varepsilon(z) U_{-j})\| = \|\operatorname{Off}(U_{-j}^\top (\varepsilon(z) - c_2 I) U_{-j})\| \leqslant \|\varepsilon(z) - c_2 I\|.$$

Decompose $\varepsilon(z) = \operatorname{diag}(\varepsilon_k(z)) = \operatorname{diag}(x_k) + \sqrt{-1}\operatorname{diag}(y_k)$. Set $c_2 = c_x + \sqrt{-1} c_y = \frac{1}{2}(\max_k x_k + \min_k x_k) + \sqrt{-1}(\max_k y_k + \min_k y_k)$. Then

$$|\varepsilon_k(z) - c_2|^2 = |x_k - c_x|^2 + |y_k - c_y|^2| \leqslant \frac{1}{4}\operatorname{osc}(x)^2 + \frac{1}{4}\operatorname{osc}(y)^2 \leqslant \frac{1}{2}\operatorname{osc}(\varepsilon(z))^2.$$

It follows that

$$\|\operatorname{Off}(U_{-j}^\top \varepsilon(z) U_{-j})\| \leqslant \frac{1}{\sqrt{2}}\operatorname{osc}(\varepsilon(z)) \leqslant \frac{15 R_{\max}}{\sqrt{2}|z|^5}\operatorname{osc}(\mathcal{R}).$$

The last inequality is by Lemma 5.2. Combining these estimates with (9.23) yields

$$\begin{aligned} \big\|\operatorname{Off}\big(U_{-j}^\top \Phi(z) U_{-j}\big)\big\| &\leqslant \frac{\operatorname{osc}(\mathcal{R})}{2|z|^3} + \frac{15 R_{\max}}{\sqrt{2}|z|^5}\operatorname{osc}(\mathcal{R}) \\ &\leqslant \frac{\operatorname{osc}(\mathcal{R})}{|z|^3}\left(\frac{1}{2} + \frac{15}{9\sqrt{2}}\right) \\ &< 1.8\frac{\operatorname{osc}(\mathcal{R})}{\vartheta_j^3} \leqslant \frac{1.8}{13}\frac{\rho}{\vartheta_j^2} \end{aligned} \tag{9.24}$$

where we used $|z| \geqslant \frac{99}{100}\vartheta_j$ and $\rho > 13\frac{\operatorname{osc}(\mathcal{R})}{\vartheta_j}$.

It remains to bound $\min_{l\in[r]\backslash\{j\}} |\alpha_l(z)|$ where

$$\alpha_l(z) = \lambda_l^{-1} - u_l^\top \Phi(z) u_l.$$

Define

$$\mathcal{O}_+ := \{l \in [r_+] : \lambda_l \geqslant 6\sqrt{R_{\max}}\}.$$

We split the discussion into three cases: $l \in \mathcal{O}_+$, $l \in [r_+]\backslash\mathcal{O}_+$ and $l \in [r]\backslash[r_+]$.

Case 1: $l \in \mathcal{O}_+$. We have $\lambda_l \geqslant 6\sqrt{R_{\max}}$. By Proposition 5.1, define ϑ_l as the unique solution to $\alpha_l(z) = 0$. Also, $\vartheta_l \geqslant 5\sqrt{R_{\max}}$. Using similar estimates as (9.20), we get, for any real $t \geqslant 3\sqrt{R_{\max}}$,

$$\alpha_l'(t) = -u_l^\top \Phi'(t) u_l \geqslant \frac{2}{5t^2}.$$

Since $\alpha_l(\vartheta_l) = 0$,

$$|\alpha_l(\vartheta_j)| = \left|\int_{\vartheta_l}^{\vartheta_j} \alpha_l'(t)\,dt\right| \geqslant \left|\int_{\vartheta_l}^{\vartheta_j} \frac{2}{5t^2}\,dt\right| = \frac{2}{5}\frac{|\vartheta_l - \vartheta_j|}{\vartheta_l\vartheta_j} \geqslant \frac{2}{5}\frac{|\vartheta_l - \vartheta_j|}{\vartheta_j(\vartheta_j + |\vartheta_l - \vartheta_j|)}.$$

The last step is from $\vartheta_l \leqslant \vartheta_j + |\vartheta_l - \vartheta_j|$. Since $|\vartheta_l - \vartheta_j| \geqslant 50\rho$ by our assumption (9.4), together with $\rho \leqslant \vartheta_j/100$, we get

$$|\alpha_l(\vartheta_j)| \geqslant \frac{40}{3}\frac{\rho}{\vartheta_j^2}.$$

Next, using the same estimates as in (9.21), we also have

$$|\alpha_l(\vartheta_j) - \alpha_l(z)| \leqslant \max_{|\zeta - \vartheta_j| \leqslant \rho} |\alpha_l'(\zeta)| \cdot \rho \leqslant 3\frac{\rho}{\vartheta_j^3}.$$

Hence,

$$|\alpha_l(z)| \geqslant |\alpha_l(\vartheta_j)| - |\alpha_l(\vartheta_j) - \alpha_l(z)| \geqslant \left(\frac{40}{3} - 3\right)\frac{\rho}{\vartheta_j^2} = \frac{10}{3}\frac{\rho}{\vartheta_j^2}. \tag{9.25}$$

Case 2: $l \in [r_+]\backslash\mathcal{O}_+$. We have $0 < \lambda_l < 6\sqrt{R_{\max}} \leqslant \lambda_j$. From

$$\alpha_l(z) - \alpha_j(z) = \left(\lambda_l^{-1} - \lambda_j^{-1}\right) - \left(u_l^\top \Phi(z) u_l - u_j^\top \Phi(z) u_j\right),$$

observe that

$$|\alpha_l(z) - \alpha_j(z)| \geqslant \frac{\lambda_j - \lambda_l}{\lambda_j\lambda_l} - \left|u_l^\top \Phi(z) u_l - u_j^\top \Phi(z) u_j\right|. \tag{9.26}$$

Furthermore, by (5.3),

$$\begin{aligned}
\left|u_l^\top \Phi(z) u_l - u_j^\top \Phi(z) u_j\right| &= \left|\frac{1}{z^3} u_l^\top \mathcal{R} u_l - \frac{1}{z^3} u_j^\top \mathcal{R} u_j - (u_l^\top \varepsilon(z) u_l - u_j^\top \varepsilon(z) u_j)\right| \\
&\leqslant \frac{1}{|z|^3}|u_l^\top \mathcal{R} u_l - u_j^\top \mathcal{R} u_j| + |u_l^\top \varepsilon(z) u_l - u_j^\top \varepsilon(z) u_j| \\
&\leqslant \frac{\operatorname{osc}(\mathcal{R})}{|z|^3} + \operatorname{osc}(\varepsilon(z)).
\end{aligned}$$

For the last inequality above, since u_l, u_j are unit vectors, their quadratic forms represent convex combinations of the diagonal entries. Therefore, the difference between them is strictly bounded by the maximum pairwise distance in the complex plane, which is the oscillation.

By Lemma 5.2,

$$\left|u_l^\top \Phi(z) u_l - u_j^\top \Phi(z) u_j\right| \leqslant \frac{\operatorname{osc}(\mathcal{R})}{|z|^3}\left(1+\frac{15 R_{\max}}{|z|^2}\right) \leqslant \frac{8}{3} \frac{\operatorname{osc}(\mathcal{R})}{|z|^3}.$$

Continuing from (9.26), together with $|\alpha_j(z)| \leqslant \frac{3\rho}{\vartheta_j^2}$ from (9.5) and the triangle inequality, we obtain

$$|\alpha_l(z)| \geqslant |\alpha_l(z)-\alpha_j(z)|-|\alpha_j(z)| \geqslant \frac{\lambda_j-\lambda_l}{\lambda_j^2}-\frac{8}{3} \frac{\operatorname{osc}(\mathcal{R})}{|z|^3}-\frac{3\rho}{\vartheta_j^2}.$$

Since $\lambda_l < 6\sqrt{R_{\max}}$ and $\lambda_j \geqslant 6\sqrt{R_{\max}}+100\rho$, we have $\lambda_j-\lambda_l \geqslant 100\rho$. Using our assumptions, $|z| \geqslant \frac{99}{100}\vartheta_j$, $0.87\lambda_j \leqslant \vartheta_j \leqslant 1.16\lambda_j$ from (9.1), and $\rho > 13\frac{\operatorname{osc}(\mathcal{R})}{\vartheta_j}$. Thus we get

$$|\alpha_l(z)| \geqslant (100 \cdot 0.87^2-3) \frac{\rho}{\vartheta_j^2}-\frac{8}{3}\left(\frac{100}{99}\right)^3 \frac{\operatorname{osc}(\mathcal{R})}{\vartheta_j^3}>70 \frac{\rho}{\vartheta_j^2}. \tag{9.27}$$

Case 3: $l \in [r] \backslash [r_+]$. We have $\lambda_l < 0$. Using the expansion (5.3), we get

$$\begin{aligned} u_l^\top \Phi(\vartheta_j) u_l &= \frac{1}{\vartheta_j}+\frac{\mathcal{V}_{ll}}{\vartheta_j^3}+\epsilon_l(\vartheta_j) \\ &\geqslant \frac{1}{\vartheta_j}-\|\epsilon_l(\vartheta_j)\| \geqslant \frac{1}{\vartheta_j}-\frac{45}{8} \frac{R_{\max}^2}{\vartheta_j^5} \\ &\geqslant \frac{1}{\vartheta_j}-\frac{5}{72} \frac{1}{\vartheta_j}=\frac{67}{72} \frac{1}{\vartheta_j}. \end{aligned}$$

Thus, $\alpha_l(\vartheta_j)=\lambda_l^{-1}-u_l^\top \Phi(\vartheta_j) u_l<0$ and

$$|\alpha_l(\vartheta_j)| \geqslant \frac{67}{72} \frac{1}{\vartheta_j} \geqslant \frac{6700}{72} \frac{\rho}{\vartheta_j^2}.$$

Using similar estimates as (9.20), we have $|\alpha_j'(\zeta)| \leqslant \frac{3}{\vartheta_j^2}$ for any ζ satisfying $|\zeta-\vartheta_j| \leqslant \rho$. Consequently,

$$|\alpha_l(z)-\alpha_l(\vartheta_j)| \leqslant \rho \cdot \max_{|\zeta-\vartheta_j| \leqslant \rho}|\alpha_j'(\zeta)| \leqslant 3 \frac{\rho}{\vartheta_j^2}$$

and

$$|\alpha_l(z)| \geqslant |\alpha_l(\vartheta_j)|-|\alpha_l(z)-\alpha_l(\vartheta_j)| \geqslant\left(\frac{6700}{72}-3\right) \frac{\rho}{\vartheta_j^2}>30 \frac{\rho}{\vartheta_j^2}. \tag{9.28}$$

Finally, combining (9.25), (9.27), and (9.28), we obtain

$$\min_{l\in[r]\backslash\{j\}} |\alpha_l(z)| \geqslant \frac{10}{3}\frac{\rho}{\vartheta_j^2}. \tag{9.29}$$

Continuing from (9.22), together with (9.24), we arrive at

$$s_{\min}(D_j(z)) \geqslant \min_{l\in[r]\backslash\{j\}} |\alpha_l(z)| - \left\|\mathrm{Off}\left(U_{-j}^\top \Phi(z) U_{-j}\right)\right\| \geqslant \left(\frac{10}{3} - \frac{1.8}{13}\right)\frac{\rho}{\vartheta_j^2} > 3\frac{\rho}{\vartheta_j^2}.$$

This completes the proof.

9.2 Proof of Theorem 6

Theorem 6 will be proved using similar techniques and estimates as Theorem 5. We briefly sketch it here. Recall that $\rho \equiv \rho_j = 10M + 15\frac{\mathrm{osc}(\mathcal{R})}{\lambda_j}$.

To establish the lower bound for the top-j outlier cluster, we define the block contour enclosing the deterministic roots $\vartheta_1, \ldots, \vartheta_j$. Following our previous notation (9.11), let

$$D_j^- := \bigcup_{l=1}^{j} \mathtt{D}(\vartheta_l, \rho_j),$$

and let $\mathcal{C}_j^- := \partial D_j^-$ be its positively oriented boundary. By our gap assumption $\min_{l>k}(\underline{\vartheta}_k - \vartheta_l) \geqslant 50\rho_k$, every supercritical root excluded from D_k^- lies at a distance 49ρ from $\mathcal{C}_k^-$, ensuring the contour is strictly isolated.

Recall from (9.12) that

$$A_0(z) := \mathrm{diag}(\alpha_1(z), \ldots, \alpha_r(z)) \quad \text{with} \quad \alpha_l(z) := \lambda_l^{-1} - u_l^\top \Phi(z) u_l.$$

The main goal is to establish an analogous result to Lemma 9.3:

LEMMA 9.4. *Under the assumption of Theorem 6, we have*

$$\sup_{z\in\mathcal{C}_j^-} \|A_0(z)^{-1}(K(z) - A_0(z))\| < 1, \qquad \sup_{z\in\mathcal{C}_j^-} \|K(z)^{-1}B(z)\| < 1.$$

Because the boundary $\mathcal{C}_j^-$ maintains a distance of at least ρ_j from any enclosed root ϑ_l, the uniform estimates established in the **Claim** of Section 9.1.1 hold identically for all $z \in \mathcal{C}_j^-$. Consequently, Lemma 9.4 holds.

Hence, by Lemma 9.4 and Lemma 9.1, $\det(A_0(z))$, $g(z) = \det(K(z))$, and $f(z) = \det(K(z) - B(z))$ have the same number of zeros inside the region D_j^-.

Since $\det(A_0(z)) = \prod_{l=1}^r \alpha_l(z)$, its zeros in the supercritical regime occur at ϑ_l's. By construction, D_j^- contains exactly j such roots: $\vartheta_1, \ldots, \vartheta_j$. Therefore, $f(z)$ has exactly j zeros inside D_j^-. As noted previously, the zeros of $f(z)$ outside $[-\|E\|, \|E\|]$ correspond exactly to the eigenvalues of the perturbed matrix $\widetilde{A}$. Since the eigenvalues are strictly ordered (i.e., $\widetilde{\lambda}_1 \geqslant \cdots \geqslant \widetilde{\lambda}_r$), these j enclosed roots inside D_j^- must be exactly the top j perturbed eigenvalues $\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_j$. Because all j of these eigenvalues lie strictly inside D_j^-, their locations are bounded from below by the leftmost boundary of the region.

Recalling that $\underline{\vartheta}_j := \min_{1 \leqslant l \leqslant j} \vartheta_l$, the real part of any $z \in D_k^-$ satisfies $\mathrm{Re}(z) \geqslant \underline{\vartheta}_k - \rho_k$. Thus, we conclude that for all $s \in [k]$,

$$\widetilde{\lambda}_s \geqslant \underline{\vartheta}_k - \rho_k.$$

The proof is complete.

A Proofs of Lemmas 5.1, Lemma 5.2, Proposition 5.1, and Lemma 5.4

A.1 Proof of Lemma 5.1

Fix z with $|z|^2 \geqslant 6R_{\max}$.

Proof of Part (i): Define a mapping $F : \mathbb{C}^n \to \mathbb{C}^n$ componentwise as

$$(F(y))_i = \frac{1}{z - \sum_j \sigma_{ij}^2 y_j} = \frac{1}{z - (\Sigma y)_i}.$$

Consider the metric $\|y\|_\infty = \max_i |y_i|$ on $\mathbb{C}^n$. Note that ϕ is a fixed point of F, i.e., $F(\phi) = \phi$. Define a closed set

$$\mathcal{B} := \left\{ y \in \mathbb{C}^n : \|y\|_\infty \leqslant \frac{3}{2|z|} \right\}.$$

We first show $F(\mathcal{B}) \subseteq \mathcal{B}$. For $y \in \mathcal{B}$ and every $i \in [n]$,

$$|(\Sigma y)_i| = \Big|\sum_j \sigma_{ij}^2 y_j\Big| \leqslant \|y\|_\infty R_{\max} \leqslant \frac{2}{|z|} R_{\max} \leqslant \frac{|z|}{3}$$

by our supposition $|z|^2 \geqslant 6R_{\max}$. Hence,

$$\big|z - (\Sigma y)_i\big| \geqslant |z| - \big|(\Sigma y)_i\big| \geqslant \frac{2}{3}|z| \tag{A.1}$$

and

$$\big|\,(F(y))_i\,\big| = \frac{1}{|z - (\Sigma y)_i|} \leqslant \frac{3}{2|z|}.$$

This shows that $F(y) \in \mathcal{B}$.

Next, we show that $F : \mathcal{B} \to \mathcal{B}$ is a contraction. Take $y, y' \in \mathcal{B}$. For each $i \in [n]$, using (A.1), we have

$$\begin{aligned} \big|\,(F(y))_i - \big(F(y')\big)_i\,\big| &= \left| \frac{\sum_j \sigma_{ij}^2 (y_j - y'_j)}{(z - (\Sigma y)_i)(z - (\Sigma y')_i)} \right| \\ &\leqslant \frac{R_{\max}\|y - y'\|_\infty}{(2|z|/3)^2} = \frac{9}{4}\frac{R_{\max}}{|z|^2}\|y - y'\|_\infty \leqslant \frac{3}{8}\|y - y'\|_\infty. \end{aligned}$$

Since $(\mathcal{B}, \|\cdot\|_\infty)$ is a complete metric space, by Banach's fixed point theorem, F has a unique fixed point in $y^* \in \mathcal{B}$. Since the solution ϕ to the QVE (5.1) is unique and $\phi(z) \to 0$ as $|z| \to \infty$. Therefore, $\phi = y^* \in \mathcal{B}$. In particular,

$$\max_i |\phi_i(z)| \leqslant \frac{3}{2|z|}. \tag{A.2}$$

We have

$$\Big|\big(\Sigma\phi(z)\big)_i\Big| = \Big|\sum_j \sigma_{ij}^2 \phi_j(z)\Big| \leqslant \frac{3}{2|z|} R_{\max}. \tag{A.3}$$

Hence, $\Big|\sum_j \sigma_{ij}^2 \phi_j(z)\Big| \leqslant \frac{1}{4}|z|$ and

$$\frac{3}{4}|z| \leqslant \big|z - \sum_j \sigma_{ij}^2 \phi_j(z)\big| \leqslant \frac{5}{4}|z|.$$

To show the bounds on $\phi'(z)$, differentiating the equation $\phi_i(z)^{-1} = z - \sum_{j=1}^n \sigma_{ij}^2 \phi_j(z)$ gives

$$-\phi_i(z)^{-2}\phi_i'(z) = 1 - \sum_{j=1}^n \sigma_{ij}^2 \phi_j'(z),$$

hence

$$\phi_i'(z) = -\phi_i(z)^2 \Big(1 - \sum_{j=1}^n \sigma_{ij}^2 \phi_j'(z)\Big)$$

and

$$|\phi_i'(z)| \leqslant |\phi_i(z)|^2 \Big(1 + \sum_{j=1}^n \sigma_{ij}^2 |\phi_j'(z)|\Big).$$

Let $M_1 := \max_i |\phi_i'(z)|$. By (A.2),

$$M_1 \leqslant \frac{9}{4|z|^2}(1 + R_{\max} M_1).$$

Since $R_{\max}/|z|^2 \leqslant 1/6$, we solve $M_1 \leqslant \frac{9}{4|z|^2} + \frac{3}{8}M_1$ to obtain

$$M_1 = \max_i |\phi_i'(z)| \leqslant \frac{18}{5|z|^2}.$$

To bound $\phi''(z)$, similarly, differentiating the equation one more time yields

$$\phi_i''(z) = -2\phi_i(z)\phi_i'(z)\Big(1 - \sum_{j=1}^n \sigma_{ij}^2 \phi_j'(z)\Big) + \phi_i(z)^2 \sum_{j=1}^n \sigma_{ij}^2 \phi_j''(z).$$

Hence,

$$|\phi_i''(z)| \leqslant 2|\phi_i(z)|\,|\phi_i'(z)|\Big(1 + \sum_{j=1}^n \sigma_{ij}^2 |\phi_j'(z)|\Big) + |\phi_i(z)|^2 \sum_{j=1}^n \sigma_{ij}^2 |\phi_j''(z)|.$$

Let $M_2 := \max_i |\phi_i''(z)|$. Using the previous bounds on M_1 and (A.2), we have

$$M_2 \leqslant 2 \cdot \frac{3}{2|z|} \cdot \frac{18}{5|z|^2} \cdot \left(1 + \frac{3}{5}\right) + \frac{9R_{\max}}{4|z|^2} M_2.$$

Thus,

$$M_2 = \max_i |\phi_i''(z)| \leqslant \frac{3456}{125|z|^3} \leqslant \frac{28}{|z|^3}.$$

Proof of Part (ii): In the following, we prove (5.2). From $\phi_i(z) = \frac{1}{z-(\Sigma\phi)_i}$ and the exact identity $\frac{1}{z-x} = \frac{1}{z} + \frac{x}{z(z-x)}$, treating $x = (\Sigma\phi)_i$, we get

$$\phi_i(z) = \frac{1}{z} + \frac{(\Sigma\phi)_i}{z(z-(\Sigma\phi)_i)} = \frac{1}{z} + \frac{(\Sigma\phi)_i}{z}\phi_i(z) := \frac{1}{z} + \delta_i(z), \tag{A.4}$$

where

$$|\delta_i(z)| \leqslant \frac{|(\Sigma\phi)_i|}{|z|}|\phi_i(z)| \leqslant \frac{9}{4}\frac{R_{\max}}{|z|^3}$$

by (A.2) and (A.3). Next, from the algebraic identity $\frac{1}{z-x} = \frac{1}{z} + \frac{x}{z^2} + \frac{x^2}{z^2(z-x)}$, we also have

$$\phi_i(z) = \frac{1}{z} + \frac{(\Sigma\phi)_i}{z^2} + \frac{(\Sigma\phi)_i^2}{z^2}\phi_i(z) := \frac{1}{z} + \frac{(\Sigma\phi)_i}{z^2} + \tilde{\delta}_i(z), \tag{A.5}$$

where, by (A.2) and (A.3), we have the estimate

$$|\tilde{\delta}_i(z)| \leqslant \frac{|(\Sigma\phi)_i|^2}{|z|^2}\big|\phi_i(z)\big| \leqslant \frac{27}{8}\frac{R_{\max}^2}{|z|^5}.$$

Now we plug (A.4) into the term $\frac{(\Sigma\phi)_i}{z^2}$ in (A.5):

$$\frac{(\Sigma\phi)_i}{z^2} = \frac{1}{z^2}\sum_j \sigma_{ij}^2\left(\frac{1}{z} + \delta_j(z)\right) = \frac{R_i}{z^3} + \frac{\sum_j \sigma_{ij}^2\delta_j(z)}{z^2}.$$

Observe that

$$\left|\frac{\sum_j \sigma_{ij}^2\delta_j(z)}{z^2}\right| \leqslant \frac{R_{\max}}{|z|^2}\max_j|\delta_j(z)| \leqslant \frac{9}{4}\frac{R_{\max}^2}{|z|^5}.$$

Combining the above estimates, we arrive at the final expansion:

$$\phi_i(z) = \frac{1}{z} + \frac{R_i}{z^3} + \left(\tilde{\delta}_i(z) + \frac{\sum_j \sigma_{ij}^2\delta_j(z)}{z^2}\right) := \frac{1}{z} + \frac{R_i}{z^3} + \varepsilon_i(z)$$

with the error term $\varepsilon_i(z)$ satisfying

$$|\varepsilon_i(z)| \leqslant |\tilde{\delta}_i(z)| + \left|\frac{\sum_j \sigma_{ij}^2\delta_j(z)}{z^2}\right| \leqslant \frac{27}{8}\frac{R_{\max}^2}{|z|^5} + \frac{9}{4}\frac{R_{\max}^2}{|z|^5} = \frac{45}{8}\frac{R_{\max}^2}{|z|^5}.$$

This proves (5.2).

To prove (5.4), we start with

$$\phi_i(z) = \left(\frac{1}{z} + \frac{R_i}{z^3} + \varepsilon_i(z)\right)' = -\frac{1}{z^2} - \frac{3R_i}{z^4} + \varepsilon_i'(z).$$

Recall that

$$\varepsilon_i(z) = \tilde{\delta}_i(z) + \frac{1}{z^2}\sum_j \sigma_{ij}^2 \delta_j(z),$$

where

$$\delta_i(z) = \frac{(\Sigma\phi)_i}{z}\phi_i(z) \quad \text{and} \quad \tilde{\delta}_i(z) = \frac{(\Sigma\phi)_i^2}{z^2}\phi_i(z).$$

It remains to bound $|\varepsilon_i'(z)|$. Note that

$$\begin{aligned}
\delta_i'(z) &= -\frac{(\Sigma\phi)_i}{z^2}\phi_i(z) + \frac{(\Sigma\phi')_i}{z}\phi_i(z) + \frac{(\Sigma\phi)_i}{z}\phi_i'(z), \\
\tilde{\delta}_i'(z) &= -2\frac{(\Sigma\phi)_i^2}{z^3}\phi_i(z) + 2\frac{(\Sigma\phi)_i(\Sigma\phi')_i}{z^2}\phi_i(z) + \frac{(\Sigma\phi)_i^2}{z^2}\phi_i'(z).
\end{aligned}$$

Plugging in the previous estimates

$$|\phi_i(z)| \leqslant \frac{3}{2|z|}, \quad |\phi_i'(z)| \leqslant \frac{18}{5|z|^2}, \quad |(\Sigma\phi)_i| \leqslant \frac{3R_{\max}}{2|z|}, \quad |(\Sigma\phi')_i| \leqslant \frac{18R_{\max}}{5|z|^2}$$

yields

$$|\delta_i'(z)| \leqslant 13.05\frac{R_{\max}}{|z|^4} \quad \text{and} \quad |\tilde{\delta}_i'(z)| \leqslant 31.05\frac{R_{\max}^2}{|z|^6}.$$

Together with $|\delta_i(z)| \leqslant \frac{9}{4}\frac{R_{\max}}{|z|^3}$, we obtain that

$$|\varepsilon_i'(z)| \leqslant |\tilde{\delta}_i'(z)| + \left|-\frac{2}{z^3}\sum_j \sigma_{ij}^2\delta_j(z) + \frac{1}{z^2}\sum_j \sigma_{ij}^2\delta_j'(z)\right| \leqslant 48.6\frac{R_{\max}^2}{|z|^6}.$$

We have proved (5.4).

A.2 Proof of Lemma 5.2

Fix $z \in \mathbb{C}$ with $|z| \geqslant \sqrt{6R_{\max}} > 0$. Let $\sqrt{\cdot}$ denote the principal branch of the complex square root on $\mathbb{C}\backslash(-\infty, 0]$. Recall that in (5.1), $\phi_i(z) = \frac{1}{z - \sum_j \sigma_{ij}^2\phi_j(z)}$. For comparison, for each $i \in [k]$, if $R_i > 0$, define an auxiliary function

$$m_i(z) = \frac{z - \sqrt{z^2 - 4R_i}}{2R_i} = \frac{z}{2R_i}\left(1 - \sqrt{1 - \frac{4R_i}{z^2}}\right)$$

and if $R_i = 0$, define $m_i(z) = \frac{1}{z}$. Then $m_i(z)$ satisfies the quadratic equation

$$m_i(z) = \frac{1}{z - R_i m_i(z)}. \tag{A.6}$$

Set

$$d_i(z) = \phi_i(z) - m_i(z) \quad \text{and} \quad d(z) = (d_1(z), \ldots, d_n(z)).$$

Then

$$\Phi(z) = \operatorname{diag}(m_1(z), \ldots, m_n(z)) + d(z).$$

We can rewrite

$$\varepsilon_i(z) = \phi_i(z) - \frac{1}{z} - \frac{R_i}{z^3} = m_i(z) - \frac{1}{z} - \frac{R_i}{z^3} + d_i(z),$$

which yields

$$\operatorname{osc}(\varepsilon_i(z)) \leqslant \operatorname{osc}\left(m_i(z) - \frac{1}{z} - \frac{R_i}{z^3}\right) + 2\|d(z)\|_\infty. \tag{A.7}$$

We now bound each term on the right-hand side separately.

We start with the first term on the right-hand side of (A.7). Since $|\frac{4R_i}{z^2}| \leqslant \frac{2}{3} < 1$, the binomial series for $\sqrt{1-w}$ is absolutely convergent at $w = \frac{4R_i}{z^2}$. Hence,

$$\sqrt{1-w} = 1 - 2\sum_{k=0}^{\infty} \frac{1}{k+1}\binom{2k}{k}\left(\frac{w}{4}\right)^{k+1} = 1 - 2\sum_{k=0}^{\infty} C_k \left(\frac{w}{4}\right)^{k+1},$$

where $C_k = \frac{1}{k+1}\binom{2k}{k}$ denotes the Catalan number. Applying this expansion

$$m_i(z) = \frac{z}{2R_i}\left(1 - \sqrt{1 - \frac{4R_i}{z^2}}\right),$$

we get

$$\begin{aligned} m_i(z) &= \frac{z}{2R_i}\left[1 - \left(1 - 2\sum_{k=0}^{\infty} C_k \left(\frac{4R_i/z^2}{4}\right)^{k+1}\right)\right] = \frac{z}{R_i}\left(\sum_{k=0}^{\infty} C_k \frac{R_i^{k+1}}{z^{2k+2}}\right) \\ &= \frac{1}{z} + \frac{R_i}{z^3} + \sum_{k=2}^{\infty} C_k \frac{R_i^k}{z^{2k+1}} := \frac{1}{z} + \frac{R_i}{z^3} + v_i(z) \end{aligned} \tag{A.8}$$

and thus

$$\Phi(z) = \frac{1}{z} I + \frac{1}{z^3}\mathcal{R} + \operatorname{diag}(v_i(z))_{1 \leqslant i \leqslant n} + d(z). \tag{A.9}$$

It follows that

$$\begin{aligned} \operatorname{osc}\left(m_i(z) - \frac{1}{z} - \frac{R_i}{z^3}\right) &= \max_{i,j} |v_i(z) - v_j(z)| \\ &= \max_{i,j} \left|\sum_{k=2}^{\infty} C_k \frac{R_i^k - R_j^k}{z^{2k+1}}\right| \\ &\leqslant \left(\sum_{k=2}^{\infty} C_k \frac{k R_{\max}^{k-1}}{|z|^{2k+1}}\right) \operatorname{osc}(\mathcal{R}), \end{aligned}$$

where we used

$$|R_i^k - R_j^k| = |(R_i - R_j)(R_i^{k-1} + R_i^{k-2}R_j + \cdots + R_j^{k-1})| \leqslant \operatorname{osc}(\mathcal{R}) \cdot kR_{\max}^{k-1}.$$

Since $R_{\max}/|z|^2 \leqslant 1/6$ by our assumption, we continue to have

$$\begin{aligned}\operatorname{osc}\left(m_i(z) - \frac{1}{z} - \frac{R_i}{z^3}\right) &\leqslant \frac{R_{\max}}{|z|^5}\left(\sum_{k=2}^{\infty} kC_k \frac{R_{\max}^{k-2}}{|z|^{2(k-2)}}\right)\operatorname{osc}(\mathcal{R}) \\ &\leqslant \frac{R_{\max}}{|z|^5}\operatorname{osc}(\mathcal{R})\left(\sum_{k=2}^{\infty}\frac{kC_k}{6^{k-2}}\right) \\ &\leqslant \frac{11R_{\max}}{|z|^5}\operatorname{osc}(\mathcal{R}). \end{aligned} \tag{A.10}$$

In the last step, we used $\sum_{k=2}^{\infty}\frac{kC_k}{6^{k-2}} \leqslant 11$. This can be derived by taking the generating function of the Catalan numbers

$$c(x) = \sum_{k=0}^{\infty} C_k x^k = \frac{1-\sqrt{1-4x}}{2x}. \tag{A.11}$$

The sum is $S = \sum_{k=2}^{\infty} kC_k x^{k-2}$ evaluated at $x = 1/6$. Differentiating the generating function gives $c'(x) = \sum_{k=1}^{\infty} kC_k x^{k-1} = 1 + xS$. Therefore, the sum is exactly given by $S = \frac{c'(x)-1}{x}$. Taking the derivative yields $c'(x) = \frac{1-2x-\sqrt{1-4x}}{2x^2\sqrt{1-4x}}$. Evaluating this at $x = 1/6$ gives $c'(1/6) = 12\sqrt{3} - 18$. Substituting this back into the sum formula yields the exact value:

$$\sum_{k=2}^{\infty}\frac{kC_k}{6^{k-2}} = 6(12\sqrt{3} - 19) \leqslant 11.$$

Furthermore, we obtain the following bound for later estimates:

$$\begin{aligned}|m_i(z) - m_j(z)| &\leqslant \frac{|R_i - R_j|}{|z|^3} + \operatorname{osc}\left(m_i(z) - \frac{1}{z} - \frac{R_i}{z^3}\right) \\ &\leqslant \frac{\operatorname{osc}(\mathcal{R})}{|z|^3} + \frac{11R_{\max}}{|z|^5}\operatorname{osc}(\mathcal{R}) \\ &\leqslant \left(1 + \frac{1}{6}\right)\frac{\operatorname{osc}(\mathcal{R})}{|z|^3} = \frac{7}{6}\frac{\operatorname{osc}(\mathcal{R})}{|z|^3}. \end{aligned} \tag{A.12}$$

In the last inequality, we applied $R_{\max}/|z|^2 \leqslant 1/6$.

Next, we turn to the second term on the right-hand side of (A.7). For each i, we have

$$\begin{aligned}|d_i(z)| = |\phi_i(z) - m_i(z)| &= \left|\frac{1}{z - \sum_j \sigma_{ij}^2\phi_j(z)} - \frac{1}{z - R_i m_i(z)}\right| \\ &= \frac{|R_i m_i(z) - \sum_j \sigma_{ij}^2\phi_j(z)|}{|z - \sum_j \sigma_{ij}^2\phi_j(z)| \cdot |z - R_i m_i(z)|} \\ &\leqslant \frac{2}{|z|^2}\left|R_i m_i(z) - \sum_j \sigma_{ij}^2\phi_j(z)\right|. \end{aligned} \tag{A.13}$$

For the last step above, observe that $|z - \sum_j \sigma_{ij}^2 \phi_j(z)| \geqslant \frac{3}{4}|z|$ by Lemma 5.1 (i). In addition, we claim that

$$|z - R_i m_i(z)| \geqslant \frac{2|z|}{3}.$$

To see this, from (A.6), we have $|z - R_i m_i(z)| = \frac{1}{|m_i(z)|}$. Using the series expansion of $m_i(z)$ in (A.8) and $R_{\max}/|z|^2 \leqslant 1/6$, we get

$$|m_i(z)| \leqslant \frac{1}{|z|} \sum_{k=0}^{\infty} C_k \Big(\frac{R_i}{|z|^2} \Big)^k \leqslant \frac{1}{|z|} \sum_{k=0}^{\infty} \frac{C_k}{6^k} = \frac{c(1/6)}{|z|} < \frac{1.5}{|z|},$$

where $c(x)$ is defined in (A.11). The claim is proved.

It remains to estimate $|R_i m_i(z) - \sum_j \sigma_{ij}^2 \phi_j(z)|$. We write

$$\begin{aligned} \Big| R_i m_i(z) - \sum_j \sigma_{ij}^2 \phi_j(z) \Big| &= |\sum_j \sigma_{ij}^2 (m_i(z) - m_j(z)) + \sum_j \sigma_{ij}^2 (m_j(z) - \phi_j(z))| \\ &\leqslant \sum_j \sigma_{ij}^2 |m_i(z) - m_j(z)| + R_{\max} \|d(z)\|_\infty . \end{aligned}$$

Plugging in (A.12), we further have

$$|R_i m_i(z) - \sum_j \sigma_{ij}^2 \phi_j(z)| \leqslant \frac{7}{6} \frac{R_{\max}}{|z|^3} \operatorname{osc}(\mathcal{R}) + R_{\max} \|d(z)\|_\infty$$

and thus from (A.13),

$$\begin{aligned} \|d(z)\|_\infty = \max_i |d_i(z)| &\leqslant \frac{2}{|z|^2} \left(\frac{7}{6} \frac{R_{\max}}{|z|^3} \operatorname{osc}(\mathcal{R}) + R_{\max} \|d(z)\|_\infty \right) \\ &\leqslant \frac{7}{3} \frac{R_{\max}}{|z|^5} \operatorname{osc}(\mathcal{R}) + \frac{1}{3} \|d(z)\|_\infty . \end{aligned}$$

Rearranging terms, we arrive at

$$\|d(z)\|_\infty \leqslant \frac{7 R_{\max}}{2|z|^5} \operatorname{osc}(\mathcal{R}). \tag{A.14}$$

Finally, we conclude that

$$\operatorname{osc}(\varepsilon_i(z)) \leqslant \operatorname{osc} \left(m_i(z) - \frac{1}{z} - \frac{R_i}{z^3} \right) + 2\|d(z)\|_\infty \leqslant 14.5 \frac{R_{\max}}{z^5} \operatorname{osc}(\mathcal{R}).$$

A.3 Proof of Proposition 5.1

By part (ii) of Lemma 5.1,

$$u_j^\top \Phi(z) u_j = \frac{1}{z} + \frac{\mathcal{V}_{jj}}{z^3} + \epsilon_j(z), \qquad |\epsilon_j(z)| = |u_j^\top \varepsilon(z) u_j| \leqslant \frac{45}{8} \frac{R_{\max}^2}{z^5},$$

for all real $z \geqslant 3\sqrt{R_{\max}}$. For brevity, denote $f(z) = u_j^\top \Phi(z) u_j$ and thus $\widehat{\alpha}_j(z) = 1 - \lambda_j f(z)$.

We first show that $\widehat{\alpha}_j$ has a unique zero. Note that $\mathcal{V}_{jj} \geqslant 0$. At $z_0 := 3\sqrt{R_{\max}}$,

$$f(z_0) = u_j^\top \Phi(z_0) u_j \geqslant \frac{1}{z_0} - \frac{45}{8}\frac{R_{\max}^2}{z_0^5} = \left(\frac{1}{3} - \frac{45}{8 \cdot 3^5}\right)\frac{1}{\sqrt{R_{\max}}} = \frac{67}{216}\frac{1}{\sqrt{R_{\max}}},$$

hence

$$\widehat{\alpha}_j(z_0) \leqslant 1 - \frac{67}{216}\frac{\lambda_j}{\sqrt{R_{\max}}} < 0$$

since $\lambda_j \geqslant 6\sqrt{R_{\max}}$. On the other hand, $\widehat{\alpha}_j(z) \to 1$ as $z \to \infty$. Next, by part (ii) of Lemma 5.1,

$$f'(z) = u_j^\top \Phi'(z) u_j = -\frac{1}{z^2} - \frac{3\mathcal{V}_{jj}}{z^4} + \epsilon_j'(z), \qquad |\epsilon_j'(z)| = |u_j^\top \varepsilon'(z) u_j| \leqslant \frac{243}{5}\frac{R_{\max}^2}{z^6} \tag{A.15}$$

Therefore, on the interval $[3\sqrt{R_{\max}}, \infty)$,

$$\widehat{\alpha}_j'(z) = -f'(z) \geqslant \lambda_j \left(\frac{1}{z^2} - \frac{243}{5}\frac{R_{\max}^2}{z^6}\right) > 0$$

since

$$\frac{243}{5}\frac{R_{\max}^2}{z^6} \leqslant \frac{243}{5}\frac{1}{81 z^2} = \frac{243}{405}\frac{1}{z^2} < \frac{1}{z^2}.$$

Therefore, $\widehat{\alpha}_j$ is strictly increasing, and has a unique and simple real zero $\vartheta_j \in [3\sqrt{R_{\max}}, \infty)$.

Observe that at $z = \lambda_j/2 \geqslant 3\sqrt{R_{\max}}$ and $z = 2\lambda_j$, we have

$$\begin{aligned}\widehat{\alpha}_j(\lambda_j/2) = 1 - \lambda_j u_j^\top \Phi(\lambda_j/2) u_j &\leqslant 1 - \lambda_j \left(\frac{2}{\lambda_j} - \frac{45}{8}\frac{2^5 R_{\max}^2}{\lambda_j^5}\right) \\ &\leqslant -1 + \frac{5}{36} < 0\end{aligned}$$

and

$$\begin{aligned}\widehat{\alpha}_j(2\lambda_j) = 1 - \lambda_j u_j^\top \Phi(2\lambda_j) u_j &\geqslant 1 - \lambda_j \left(\frac{1}{2\lambda_j} + \frac{45}{8}\frac{R_{\max}^2}{2^5 \lambda_j^5}\right) \\ &\geqslant \frac{1}{2} - \frac{45}{9216} > 0.\end{aligned}$$

Since $\widehat{\alpha}_j$ is strictly increasing, we have

$$\frac{\lambda_j}{2} \leqslant \vartheta_j \leqslant 2\lambda_j.$$

Now we locate ϑ_j. Note that $f(\vartheta_j) = \lambda_j^{-1}$. We introduce a scalar comparison function

$$m_{\mathcal{V},j}(z) = \frac{z - \sqrt{z^2 - 4\mathcal{V}_{jj}}}{2\mathcal{V}_{jj}}$$

which satisfies the quadratic equation

$$m(z) = \frac{1}{z - \mathcal{V}_{jj} m(z)}. \tag{A.16}$$

This function approximates $f(z) = u_j^\top \Phi(z) u_j$ by noting $\mathcal{V}_{jj} = \sum_i u_j(i)^2 R_i$ and the definition of $\phi_i(z)$ in (5.1).

Denote
$$x_j := \lambda_j + \frac{\mathcal{V}_{jj}}{\lambda_j}.$$

Then x_j is a zero of $1 - \lambda_j m_{\mathcal{V},j}(z) = 0$. Indeed, plugging $m_{\mathcal{V},j}(z) = \frac{1}{\lambda_j}$ into (A.16), we get $\lambda_j = z - \mathcal{V}_{jj} m_{\mathcal{V},j}(z) = z - \frac{\mathcal{V}_{jj}}{\lambda_j}$ and thus $z = \lambda_j + \frac{\mathcal{V}_{jj}}{\lambda_j}$.

Next, we expand $m_{\mathcal{V},j}(z)$ as a Catalan series the same as (A.8) at $z = x_j$:
$$\frac{1}{\lambda_j} = m_{\mathcal{V},j}(x_j) = \frac{1}{x_j} + \frac{\mathcal{V}_{jj}}{x_j^3} + \sum_{k=2}^{\infty} C_k \frac{\mathcal{V}_{jj}^k}{x_j^{2k+1}}. \tag{A.17}$$

To estimate $|\vartheta_j - x_j|$, we first show that $f(x_j) = u_j^\top \Phi(x_j) u_j \approx \frac{1}{\lambda_j} = f(\vartheta_j)$. Then we apply the mean value theorem to bound $|\vartheta_j - x_j|$.

In the proof of Lemma 5.2, we have the expansion of $\Phi(z)$ given in (A.9). Hence,
$$f(x_j) = u_j^\top \Phi(x_j) u_j = \frac{1}{x_j} + \frac{\mathcal{V}_{jj}}{x_j^3} + \sum_{k=2}^{\infty} C_k \frac{\sum_i u_j(i)^2 R_i^k}{x_j^{2k+1}} + u_j^\top d(x_j) u_j.$$

Subtracting with (A.17), we obtain
$$|f(x_j) - f(\vartheta_j)| = |f(x_j) - \frac{1}{\lambda_j}| \leqslant \sum_{k=2}^{\infty} \frac{C_k}{x_j^{2k+1}} \left| \mathcal{V}_{jj}^k - \sum_i u_j(i)^2 R_i^k \right| + \|d(x_j)\|_\infty.$$

Note that
$$\begin{aligned}\left| \mathcal{V}_{jj}^k - \sum_i u_j(i)^2 R_i^k \right| \leqslant \sum_i u_j(i)^2 \left| \mathcal{V}_{jj}^k - R_i^k \right| &\leqslant k R_{\max}^k \sum_i u_j(i)^2 \left| \mathcal{V}_{jj} - R_i \right| \\ &\leqslant k R_{\max}^k \operatorname{osc}(\mathcal{R}).\end{aligned}$$

Following the same step as (A.10) and combining the bound on $\|d\|_\infty$ in (A.14), we arrive at
$$|f(x_j) - f(\vartheta_j)| \leqslant 11 \frac{R_{\max}}{x_j^5} \operatorname{osc}(\mathcal{R}) + \frac{7}{2} \frac{R_{\max}}{x_j^5} \operatorname{osc}(\mathcal{R}) \leqslant 14.5 \frac{R_{\max}}{\lambda_j^5} \operatorname{osc}(\mathcal{R}).$$

Finally, by mean value theorem, $|f(x_j) - f(\vartheta_j)| = |x_j - \vartheta_j| \cdot |f'(w)|$ for some w between x_j and ϑ_j. Since $\lambda_j/2 \leqslant \vartheta_j \leqslant 2\lambda_j$ and $x_j \leqslant 2\lambda_j$, we have $w \leqslant 2\lambda_j$. By (A.15), we get
$$f'(w) \geqslant \frac{1}{w^2} - \frac{243}{5} \frac{R_{\max}^2}{w^6} \geqslant \frac{1}{w^2} \left(1 - \frac{243}{5 \cdot 81} \right) = \frac{2}{5w^2} \geqslant \frac{1}{10\lambda_j^2}.$$

Rearranging yields that
$$\left| \lambda_j + \frac{\mathcal{V}_{jj}}{\lambda_j} - \vartheta_j \right| = |x_j - \vartheta_j| \leqslant 145 \frac{R_{\max}}{\lambda_j^3} \operatorname{osc}(\mathcal{R}).$$

The proof is now complete.

A.4 Proof of Lemma 5.4

We apply Corollary 3.12 from [16] with a standard truncation technique. Denote

$$L_n := K\sigma_{\max}((D+4)\log n)^{1/2}.$$

By a union bound, we have

$$\mathbb{P}(\max_{i,j}|E_{ij}| > L_n) \leqslant n^2 \cdot 2\exp(-(L_n/K)^2) = 2n^{-(D+2)}.$$

We split E into big and small parts according to L_n. Namely, we write $E = Z + M + B$, where we define these symmetric matrices via

$$Z_{ij} = E_{ij}\mathbf{1}_{\{|E_{ij}|\leqslant L_n\}} - \mathbb{E}(E_{ij}\mathbf{1}_{\{|E_{ij}|\leqslant L_n\}}),\; M_{ij} = \mathbb{E}(E_{ij}\mathbf{1}_{\{|E_{ij}|\leqslant L_n\}}),\; B_{ij} = E_{ij}\mathbf{1}_{\{|E_{ij}|>L_n\}}.$$

Note that if $\max_{i,j}|E_{ij}| \leqslant L_n$, then $B = 0$. We work on the event $\mathcal{E} = \{\max_{i,j}|E_{ij}| \leqslant L_n\}$ below and previous calculations suggest that $\mathbb{P}(\mathcal{E}) \geqslant 1 - 2n^{-(D+2)}$.

Next, we bound $\|M\|$. Note that since E_{ij} has mean 0, $M_{ij} = -\mathbb{E}(E_{ij}\mathbf{1}_{\{|E_{ij}|>L_n\}})$. Therefore,

$$|M_{ij}| \leqslant \mathbb{E}(|E_{ij}|\mathbf{1}_{\{|E_{ij}|>L_n\}}) = \int_{L_n}^{\infty} \mathbb{P}(|E_{ij}| \geqslant t)\,dt \leqslant 2L_n n^{-(D+4)}$$

since $\mathbb{P}(|E_{ij}| > L_n) \leqslant 2n^{-(D+4)}$. Then we have

$$\|M\| \leqslant \max_i \sum_j |M_{ij}| \leqslant 2L_n n^{-(D+3)}.$$

Finally, for the bound on $\|Z\|$, we apply Corollary 3.12 from [16]. Note that Z_{ij}'s have mean 0 and $|Z_{ij}| \leqslant 2L_n$. Also, $\mathbb{E}(Z_{ij}^2) \leqslant \mathbb{E}(E_{ij}^2) = \sigma_{ij}^2$. Applying of [16, Corollary 3.12], together with a standard symmetrization (as in the proof of [16, Corollary 3.3]), yields that for $t > 0$,

$$\mathbb{P}\left(\|Z\| \geqslant 2\sqrt{2}(1+\epsilon)\tilde{\sigma} + t\right) \leqslant n\exp(-t^2/(c_\epsilon L_n)^2).$$

We choose ϵ such that $2\sqrt{2}(1+\epsilon) = 2.9$ and denote $c_\epsilon = c$. The parameter $\tilde{\sigma} = \max_i \sqrt{\sum_j \mathbb{E}(Z_{ij}^2)} = \sqrt{\max_i \sigma_{ij}^2} = \sqrt{R_{\max}}$ by our supposition. Choosing $t = cL_n\sqrt{(D+3)\log n}$, we obtain that

$$\mathbb{P}\left(\|Z\| \geqslant 2.9\sqrt{R_{\max}} + cL_n\sqrt{(D+3)\log n}\right) \leqslant n^{-(D+2)}.$$

Combining the above estimates, we have that with probability at least $1 - 3n^{-(D+2)}$,

$$\begin{aligned}\|E\| \leqslant \|Z\| + \|M\| + \|B\| &\leqslant 2L_n n^{-(D+3)} + 1.5\sqrt{R_{\max}} + cL_n\sqrt{(D+3)\log n} \\ &\leqslant 2.9\sqrt{R_{\max}} + cK\sigma_{\max}(D+4)\log n.\end{aligned}$$

B Proofs of Proposition 7.1 and Proposition 7.2

B.1 Proof of Proposition 7.1

Let $Q = I - UU^\top$ be the orthogonal projection onto the null space of A. Set $P_J = U_J U_J^\top$ where $J = [r]\backslash[k]$ and $U_J = (u_{k+1}, \dots, u_r)$. Then

$$P_{U_k^\perp} = P_J + Q.$$

We first prove (7.1) in part (i). Starting from $\|\sin\angle(U_k, \widetilde{U}_k)\| = \|P_{U_k^\perp} P_{\widetilde{U}_k}\|$ (see, for instance, [28, Exercises VII. 1. 9-1.11]), we have

$$\begin{aligned}\|\sin\angle(U_k, \widetilde{U}_k)\| = \|(P_J + Q)P_{\widetilde{U}_k}\| &\leqslant \|P_J P_{\widetilde{U}_k}\| + \|QP_{\widetilde{U}_k}\| \\ &\leqslant \|U_J^\top \widetilde{U}_k\|_F + \|QP_{\widetilde{U}_k}\|.\end{aligned}$$

It remains to bound $\|QP_{\widetilde{U}_k}\| = \|Q\widetilde{U}_k\|$. From the spectral decomposition of $\widetilde{A}$, we have $(A + E)\widetilde{U}_k = \widetilde{U}_k \tilde{\Lambda}_k$. Multiplying by Q on the left and noting that $QA = 0$, we obtain $QE\widetilde{U}_k = Q\widetilde{U}_k\widetilde{\Lambda}_k$. Since $\widetilde{\Lambda}_k$ is invertible, we have

$$Q\widetilde{U}_k = QE\widetilde{U}_k\widetilde{\Lambda}_k^{-1}.$$

To see that $\widetilde{\Lambda}_k$ is indeed invertible, note that by Weyl's inequality, for every $1 \leqslant s \leqslant k$,

$$\frac{3}{2}\lambda_s \geqslant \lambda_s + \|E\| \geqslant \widetilde{\lambda}_s \geqslant \lambda_s - \|E\| \geqslant \frac{1}{2}\lambda_s, \tag{B.1}$$

where the last inequality follows from our assumption $\lambda_s \geqslant \lambda_k \geqslant 2\|E\|$. Therefore,

$$\|QP_{\widetilde{U}_k}\| = \|QE\widetilde{U}_k\widetilde{\Lambda}_k^{-1}\| \leqslant \|QE\widetilde{U}_k\| \cdot \|\widetilde{\Lambda}_k^{-1}\| \leqslant \frac{\|E\|}{\widetilde{\lambda}_k} \leqslant 2\frac{\|E\|}{\lambda_k}. \tag{B.2}$$

Now we turn to part (ii). Note that (7.2) is given in Eq. (10) of [98]. It remains to prove (7.3). It follows from the decomposition $\widetilde{u}_s = P_{U_k}\widetilde{u}_s + P_J\widetilde{u}_s + Q\widetilde{u}_s$ for each $1 \leqslant s \leqslant k$ that

$$\widetilde{U}_k = P_{U_k}\widetilde{U}_k + P_J\widetilde{U}_k + Q\widetilde{U}_k$$

and thus

$$\begin{aligned}\|\widetilde{U}_k - P_{U_k}\widetilde{U}_k\|_{2,\infty} = \|P_J\widetilde{U}_k + Q\widetilde{U}_k\|_{2,\infty} &\leqslant \|U_J U_J^\top \widetilde{U}_k\|_{2,\infty} + \|Q\widetilde{U}_k\|_{2,\infty} \\ &\leqslant \|U\|_{2,\infty} \cdot \|U_J^\top \widetilde{U}_k\|_F + \|Q\widetilde{U}_k\|_{2,\infty}.\end{aligned}$$

To finish the proof, we show that

$$\|Q\widetilde{U}_k\|_{2,\infty} \leqslant \frac{13}{2}\sqrt{k}\|U\|_{2,\infty}\frac{R_{\max}}{\widetilde{\lambda}_k^2} + 2\sqrt{\sum_{s=1}^{k} \widetilde{\lambda}_s^2 \max_{1\leqslant i\leqslant n} \left\|e_i^\top Q\Xi(z)U\right\|^2}.$$

For $1 \leqslant s \leqslant k$, from the $(A+E)\widetilde{u}_s = \widetilde{\lambda}_s \widetilde{u}_s$, we have $A\widetilde{u}_s = (\widetilde{\lambda}_s - E)\widetilde{u}_s$. Therefore,

$$\widetilde{u}_s = G(\widetilde{\lambda}_s)A\widetilde{u}_s = \Phi(\widetilde{\lambda}_s)A\widetilde{u}_s + \Xi(\widetilde{\lambda}_s)A\widetilde{u}_s \tag{B.3}$$

and for any canonical vector e_l, we obtain

$$e_l^\top Q\widetilde{U}_k = \left(e_l^\top Q\Phi(\widetilde{\lambda}_s)A\widetilde{u}_s\right)_{s\in[k]} + \left(e_l^\top Q\Xi(\widetilde{\lambda}_s)A\widetilde{u}_s\right)_{s\in[k]} := \mathsf{q}_l^{(1)} + \mathsf{q}_l^{(2)}.$$

It follows that

$$\|Q\widetilde{U}_k\|_{2,\infty} = \max_{1\leqslant l\leqslant n} \|e_l^\top Q\widetilde{U}_k\| \leqslant \max_{1\leqslant l\leqslant n} \left(\|\mathsf{q}_l^{(1)}\| + \|\mathsf{q}_l^{(2)}\|\right).$$

We bound $\|\mathsf{q}_l^{(1)}\|$ and $\|\mathsf{q}_l^{(2)}\|$ on the right hand side. We start from

$$\|\mathsf{q}_l^{(1)}\| = \sqrt{\sum_{s=1}^k \left(e_l^\top Q\Phi(\widetilde{\lambda}_s)A\widetilde{u}_s\right)^2}.$$

For each $l \in [n]$, observe that

$$|e_l^\top Q\Phi(\widetilde{\lambda}_s)A\widetilde{u}_s| = |e_l^\top Q\Phi(\widetilde{\lambda}_s)U \cdot \Lambda U^\top \widetilde{u}_s| \leqslant \|e_l^\top Q\Phi(\widetilde{\lambda}_s)U\| \cdot \|\Lambda U^\top \widetilde{u}_s\|. \tag{B.4}$$

Note that

$$\|\Lambda U^\top \widetilde{u}_s\| \leqslant 2\widetilde{\lambda}_s.$$

This is because from $(A+E)\widetilde{u}_s = \widetilde{\lambda}_s\widetilde{u}_s$, we have $A\widetilde{u}_s = U\Lambda U^\top \widetilde{u}_s = (\widetilde{\lambda}_s I - E)\widetilde{u}_s$. Multiplying $U^\top$ on each side, we further get $\Lambda U^\top \widetilde{u}_s = U^\top(\widetilde{\lambda}_s - E)\widetilde{u}_s$ and

$$\|\Lambda U^\top \widetilde{u}_s\| \leqslant (\widetilde{\lambda}_s + \|E\|) \leqslant 2\widetilde{\lambda}_s$$

by (B.1) and our assumption $\lambda_s \geqslant \lambda_k \geqslant 2\|E\|$.

Next, we estimate $\|e_l^\top Q\Phi(\widetilde{\lambda}_s)U\|$. Since $\widetilde{\lambda}_s \geqslant \frac{1}{2}\lambda_s \geqslant 3\sqrt{R_{\max}}$, Note that $\Phi(\widetilde{\lambda}_s)$ is a diagonal matrix. Indeed, for a general real diagonal matrix $D = \operatorname{diag}(d_1, \cdots, d_n)$,

$$\begin{aligned} e_l^\top QDU = e_l^\top(I - UU^\top)DU &= e_l^\top DU - e_l^\top U(U^\top DU) \\ &= e_l^\top U(d_l I - U^\top DU) = e_l^\top U \cdot U^\top(d_l I - D)U \end{aligned}$$

due to $e_l^\top D = d_l e_l^\top$ and hence,

$$\|e_l^\top QDU\| \leqslant \|e_l^\top U\| \max_i |d_i - d_l| \leqslant \|U\|_{2,\infty}\operatorname{osc}(D).$$

We apply this to $D = \Phi(\widetilde{\lambda}_s)$. Since $\Phi(z) = \frac{1}{z}I_n + \frac{1}{z^3}\mathcal{R} + \varepsilon(z)$ from (5.3), by Lemma 5.2, we get

$$\operatorname{osc}(\Phi(\widetilde{\lambda}_s)) \leqslant \left(1 + 15\frac{R_{\max}}{\widetilde{\lambda}_s^2}\right)\frac{\operatorname{osc}(\mathcal{R})}{\widetilde{\lambda}_s^3} \leqslant \frac{8}{3}\frac{\operatorname{osc}(\mathcal{R})}{\widetilde{\lambda}_s^3},$$

where we used $\widetilde{\lambda}_s \geqslant \frac{1}{2}\lambda_s \geqslant 3\sqrt{R_{\max}}$ from (B.1) and our assumption $\lambda_k \geqslant 6\sqrt{R_{\max}}$. Thus,

$$\|e_l^\top Q\Phi(\widetilde{\lambda}_s)U\| \leqslant \|U\|_{2,\infty}\operatorname{osc}(\Phi(\widetilde{\lambda}_s)) \leqslant \frac{8}{3}\|U\|_{2,\infty}\frac{\operatorname{osc}(\mathcal{R})}{\widetilde{\lambda}_s^3}.$$

Combining the above estimates, we continue from (B.4) to achieve

$$|e_l^\top Q\Phi(\widetilde{\lambda}_s)A\widetilde{u}_s| \leqslant \frac{8}{3}\|U\|_{2,\infty}\frac{\mathrm{osc}(\mathcal{R})}{\widetilde{\lambda}_s^2} \leqslant \frac{8}{3}\|U\|_{2,\infty}\frac{\mathrm{osc}(\mathcal{R})}{\widetilde{\lambda}_k^2}.$$

Thus

$$\|\mathsf{q}_l^{(1)}\| = \sqrt{\sum_{s=1}^k \left(e_l^\top Q\Phi(\widetilde{\lambda}_s)A\widetilde{u}_s\right)^2} \leqslant \frac{8}{3}\sqrt{k}\|U\|_{2,\infty}\frac{\mathrm{osc}(\mathcal{R})}{\widetilde{\lambda}_k^2}.$$

The estimate for

$$\|\mathsf{q}_l^{(2)}\| = \sqrt{\sum_{s=1}^k \left(e_l^\top Q\Xi(\widetilde{\lambda}_s)A\widetilde{u}_s\right)^2}$$

is simpler. Using an estimate similar to (B.4), we have

$$\begin{aligned}|e_l^\top Q\Xi(\widetilde{\lambda}_s)A\widetilde{u}_s| = |e_l^\top Q\Xi(\widetilde{\lambda}_s)U\cdot\Lambda U^\top\widetilde{u}_s| &\leqslant \|e_l^\top Q\Xi(\widetilde{\lambda}_s)U\|\cdot\|\Lambda U^\top\widetilde{u}_s\| \\ &\leqslant 2\widetilde{\lambda}_s\|e_l^\top Q\Xi(\widetilde{\lambda}_s)U\|.\end{aligned}$$

Thus

$$\|\mathsf{q}_l^{(2)}\| \leqslant 4\sqrt{\sum_{s=1}^k \widetilde{\lambda}_s^2 \max_{1\leqslant i\leqslant n}\left\|e_i^\top Q\Xi(\widetilde{\lambda}_s)U\right\|^2}.$$

Combining all the above estimates, we complete the proof.

B.2 Proof of Proposition 7.2

Fix $s \in [k]$. Recall that $J = [r]\backslash[k]$. We partition J into two parts

$$\mathcal{J} := \{k+1,\ldots,r_+\} \quad \text{and} \quad \mathcal{N} := \{r_+ + 1,\ldots,r\}.$$

Then

$$\|U_J^\top\widetilde{u}_s\|^2 = \|U_{\mathcal{J}}^\top\widetilde{u}_s\|^2 + \|U_{\mathcal{N}}^\top\widetilde{u}_s\|^2.$$

We estimate each term separately.

Step 1. Bound on $\|U_{\mathcal{J}}^\top\widetilde{u}_s\|$. Set $\mathcal{I} = [k]\cup\{r_+ + 1,\ldots,r\} = [r]\backslash\mathcal{J}$. We then have the decomposition

$$A = U\Lambda U^\top = U_{\mathcal{J}}\Lambda_{\mathcal{J}}U_{\mathcal{J}}^\top + U_{\mathcal{I}}\Lambda_{\mathcal{I}}U_{\mathcal{I}}^\top.$$

Multiplying both sides of (B.3) by $U_{\mathcal{J}}^\top$ from the left yields

$$\begin{aligned}U_{\mathcal{J}}^\top\widetilde{u}_s &= U_{\mathcal{J}}^\top\Phi(\widetilde{\lambda}_s)A\widetilde{u}_s + U_{\mathcal{J}}^\top\Xi(\widetilde{\lambda}_s)A\widetilde{u}_s \\ &= U_{\mathcal{J}}^\top\Phi(\widetilde{\lambda}_s)U_{\mathcal{J}}\Lambda_{\mathcal{J}}\cdot U_{\mathcal{J}}^\top\widetilde{u}_s + U_{\mathcal{J}}^\top\Phi(\widetilde{\lambda}_s)U_{\mathcal{I}}\cdot\Lambda_{\mathcal{I}}U_{\mathcal{I}}^\top\widetilde{u}_s + U_{\mathcal{J}}^\top\Xi(\widetilde{\lambda}_s)U\cdot\Lambda U^\top\widetilde{u}_s,\end{aligned}$$

where we have used the decomposition of A. Isolating the term $U_{\mathcal{J}}^\top\widetilde{u}_s$ gives

$$\left(I - U_{\mathcal{J}}^\top\Phi(\widetilde{\lambda}_s)U_{\mathcal{J}}\Lambda_{\mathcal{J}}\right)U_{\mathcal{J}}^\top\widetilde{u}_s = U_{\mathcal{J}}^\top\Phi(\widetilde{\lambda}_s)U_{\mathcal{I}}\cdot\Lambda_{\mathcal{I}}U_{\mathcal{I}}^\top\widetilde{u}_s + U_{\mathcal{J}}^\top\Xi(\widetilde{\lambda}_s)U\cdot\Lambda U^\top\widetilde{u}_s. \quad \text{(B.5)}$$

Denote

$$L_{\mathcal{J}}(\widetilde{\lambda}_s) := I - U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{J}} \Lambda_{\mathcal{J}}.$$

We first bound the smallest singular value $s_{\min}(L_{\mathcal{J}}(\widetilde{\lambda}_s))$. We split

$$\begin{aligned} U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{J}} &= \operatorname{diag}(u_j^\top \Phi(\widetilde{\lambda}_s) u_j)_{j\in\mathcal{J}} + \operatorname{Off}(U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{J}}) \\ &:= \operatorname{diag}(u_j^\top \Phi(\widetilde{\lambda}_s) u_j)_{j\in\mathcal{J}} + Q_{\mathcal{J}}(\widetilde{\lambda}_s) \end{aligned}$$

and rewrite

$$\begin{aligned} L_{\mathcal{J}}(\widetilde{\lambda}_s) &= I - \operatorname{diag}(u_j^\top \Phi(\widetilde{\lambda}_s) u_j)_{j\in\mathcal{J}} \Lambda_{\mathcal{J}} - Q_{\mathcal{J}}(\widetilde{\lambda}_s)\Lambda_{\mathcal{J}} \\ &= \operatorname{diag}\left(1 - \lambda_j u_j^\top \Phi(\widetilde{\lambda}_s) u_j\right)_{j\in\mathcal{J}} - Q_{\mathcal{J}}(\widetilde{\lambda}_s)\Lambda_{\mathcal{J}}. \end{aligned}$$

Then

$$s_{\min}(L_{\mathcal{J}}(\widetilde{\lambda}_s)) \geqslant \frac{\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s)}{\widetilde{\lambda}_s} - \|Q_{\mathcal{J}}(\widetilde{\lambda}_s)\Lambda_{\mathcal{J}}\| \tag{B.6}$$

with $\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s) = \widetilde{\lambda}_s \min_{j\in\mathcal{J}} |1 - \lambda_j u_j^\top \Phi(\widetilde{\lambda}_s) u_j|$ defined in (7.4).

Now we bound $\|Q_{\mathcal{J}}(\widetilde{\lambda}_s)\Lambda_{\mathcal{J}}\|$ on the right-hand side of (B.6). Since $\widetilde{\lambda}_s \geqslant 3\sqrt{R_{\max}}$, by the second-order expansion of Φ in (A.5), we have

$$Q_{\mathcal{J}}(\widetilde{\lambda}_s) = \operatorname{Off}(U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{J}}) = \operatorname{Off}\left(\frac{\mathcal{V}_{\mathcal{J}\mathcal{J}}}{\widetilde{\lambda}_s^3}\right) + \operatorname{Off}\left(U_{\mathcal{J}}^\top \varepsilon(\widetilde{\lambda}_s) U_{\mathcal{J}}\right).$$

Note, by orthogonality, that

$$\left\|\operatorname{Off}\left(U_{\mathcal{J}}^\top \varepsilon(\widetilde{\lambda}_s) U_{\mathcal{J}}\right)\right\| = \left\|\operatorname{Off}\left(U_{\mathcal{J}}^\top \big(\varepsilon(\widetilde{\lambda}_s) - c^* I\big) U_{\mathcal{J}}\right)\right\| \leqslant \|\varepsilon(\widetilde{\lambda}_s) - c^* I\| = \frac{1}{2}\operatorname{osc}(\varepsilon(\widetilde{\lambda}_s)),$$

where we select $c^* = \frac{1}{2}\left(\max_i \varepsilon_i(\widetilde{\lambda}_s) + \min_i \varepsilon_i(\widetilde{\lambda}_s)\right)$. It follows, by applying Proposition 5.2, that

$$\|Q_{\mathcal{J}}(\widetilde{\lambda}_s)\Lambda_{\mathcal{L}}\| \leqslant \frac{\lambda_{k+1}}{\widetilde{\lambda}_s^3}\|\operatorname{Off}(\mathcal{V}_{\mathcal{J}\mathcal{J}})\| + \frac{13}{2}\frac{R_{\max}\lambda_{k+1}}{\widetilde{\lambda}_s^5}\operatorname{osc}(\mathcal{R}).$$

Our assumption implies that

$$\frac{\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s)}{\widetilde{\lambda}_s} \geqslant 2\|Q_{\mathcal{J}}(\widetilde{\lambda}_s)\Lambda_{\mathcal{L}}\|$$

and hence, from (B.6), we get

$$s_{\min}(L_{\mathcal{J}}(\widetilde{\lambda}_s)) \geqslant \frac{1}{2}\frac{\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s)}{\widetilde{\lambda}_s}.$$

Next, we turn to the term on the right-hand side of (B.5). Applying the triangle inequality yields

$$\begin{aligned} &\|U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{I}} \cdot \Lambda_{\mathcal{I}} U_{\mathcal{I}}^\top \widetilde{u}_s + U_{\mathcal{J}}^\top \Xi(\widetilde{\lambda}_s) U \cdot \Lambda U^\top \widetilde{u}_s\| \\ &\leqslant \|U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{I}}\| \cdot \|\Lambda_{\mathcal{I}} U_{\mathcal{I}}^\top \widetilde{u}_s\| + \|U_{\mathcal{J}}^\top \Xi(\widetilde{\lambda}_s) U\| \cdot \|\Lambda U^\top \widetilde{u}_s\|. \end{aligned}$$

For the bound on $\|\Lambda U^\top \widetilde{u}_s\|$, starting from $(A+E)\widetilde{u}_s = \widetilde{\lambda}_s \widetilde{u}_s$, we have $U\Lambda U^\top \widetilde{u}_s = (\widetilde{\lambda}_s I - E)\widetilde{u}_s$. Multiplying $U^\top$ from the left on both sides and taking the norms yield

$$\|\Lambda U^\top \widetilde{u}_s\| = \|U^\top (\widetilde{\lambda}_s I - E)\widetilde{u}_s\| \leqslant \widetilde{\lambda}_s + \|E\| \leqslant 2\widetilde{\lambda}_s.$$

The last step follows from (B.1) and our assumption $\lambda_s \geqslant 2\|E\|$. Similarly, we also have

$$\|\Lambda_{\mathcal{I}} U_{\mathcal{I}}^\top \widetilde{u}_s\| \leqslant 2\widetilde{\lambda}_s.$$

The right-hand side of (B.5) is bounded by

$$2\widetilde{\lambda}_s \left(\|U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{I}}\| + \|U^\top \Xi(\widetilde{\lambda}_s) U\| \right).$$

It remains to estimate $\|U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{I}}\|$. We plug in the second-order expansion of $\Phi(\widetilde{\lambda}_s)$ in (A.5). By the orthogonality of eigenvectors, we get

$$\begin{aligned} U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{I}} &= \frac{U_{\mathcal{J}}^\top \mathcal{R} U_{\mathcal{I}}}{\widetilde{\lambda}_s^3} + U_{\mathcal{J}}^\top \varepsilon(\widetilde{\lambda}_s) U_{\mathcal{I}} \\ &= \frac{\mathcal{V}_{\mathcal{J}\mathcal{I}}}{\widetilde{\lambda}_s^3} + U_{\mathcal{J}}^\top \left(\varepsilon(\widetilde{\lambda}_s) - c^* I \right) U_{\mathcal{I}}, \end{aligned}$$

where $c^* = \frac{1}{2}\left(\max_i \varepsilon_i(\widetilde{\lambda}_s) + \min_i \varepsilon_i(\widetilde{\lambda}_s) \right)$. By Proposition 5.2, we obtain

$$\begin{aligned} \|U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{I}}\| &\leqslant \frac{\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\|}{\widetilde{\lambda}_s^3} + \|\varepsilon(\widetilde{\lambda}_s) - c^* I\| \\ &\leqslant \frac{\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\|}{\widetilde{\lambda}_s^3} + \frac{1}{2}\mathrm{osc}(\varepsilon(\widetilde{\lambda}_s)) \leqslant \frac{\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\|}{\widetilde{\lambda}_s^3} + \frac{13}{2}\frac{R_{\max}}{\widetilde{\lambda}_s^5}\mathrm{osc}(\mathcal{R}). \end{aligned}$$

Therefore, the right-hand side of can be bounded by

$$2\widetilde{\lambda}_s \left(\frac{\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\|}{\widetilde{\lambda}_s^3} + \frac{13}{2}\frac{R_{\max}}{\widetilde{\lambda}_s^5}\mathrm{osc}(\mathcal{R}) + \|U^\top \Xi(\widetilde{\lambda}_s) U\| \right).$$

Finally, combining the above estimates, we continue from (B.5) to achieve the bound

$$\begin{aligned} \|U_{\mathcal{J}}^\top \widetilde{u}_s\| &\leqslant \frac{2\widetilde{\lambda}_s \left(\frac{\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\|}{\widetilde{\lambda}_s^3} + \frac{13}{2}\frac{R_{\max}}{\widetilde{\lambda}_s^5}\mathrm{osc}(\mathcal{R}) + \|U^\top \Xi(\widetilde{\lambda}_s) U\| \right)}{s_{\min}(L_{\mathcal{J}}(\widetilde{\lambda}_s))} \\ &\leqslant \frac{4}{\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s)} \left(\frac{\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\|}{\widetilde{\lambda}_s} + \frac{13}{2}\frac{R_{\max}}{\widetilde{\lambda}_s^3}\mathrm{osc}(\mathcal{R}) + \widetilde{\lambda}_s^2 \|U^\top \Xi(\widetilde{\lambda}_s) U\| \right). \end{aligned}$$

Step 2. Bound on $\|U_{\mathcal{N}}^\top \widetilde{u}_s\|$. Denote the index set $\mathcal{K} := [r] \backslash \mathcal{N}$. We decompose

$$A = U_{\mathcal{N}} \Lambda_{\mathcal{N}} U_{\mathcal{N}}^\top + U_{\mathcal{K}} \Lambda_{\mathcal{K}} U_{\mathcal{K}}^\top.$$

Following the same derivation of (B.5), we have

$$\left(I - U_{\mathcal{N}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{N}} \Lambda_{\mathcal{N}}\right) U_{\mathcal{N}}^\top \widetilde{u}_s = U_{\mathcal{N}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{K}} \cdot \Lambda_{\mathcal{K}} U_{\mathcal{K}}^\top \widetilde{u}_s + U_{\mathcal{N}}^\top \Xi(\widetilde{\lambda}_s) U \cdot \Lambda U^\top \widetilde{u}_s. \quad \text{(B.7)}$$

Denote

$$\widehat{L}_{\mathcal{N}}(\widetilde{\lambda}_s) := I - U_{\mathcal{N}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{N}} \Lambda_{\mathcal{N}}.$$

We first bound $s_{\min}\big(\widehat{L}_{\mathcal{N}}(\widetilde{\lambda}_s)\big)$. Since $\Lambda_{\mathcal{N}}$ contains only negative eigenvalues, set $D_{\mathcal{N}} := \operatorname{diag}(|\lambda_l|)_{l \in \mathcal{N}} = -\Lambda_{\mathcal{N}} \succ 0$. Denote $Q_{\mathcal{N}} := U_{\mathcal{N}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{N}}$ for simplicity. Since $\phi_i(\widetilde{\lambda}_s) > 0$, $Q_{\mathcal{N}} \succ 0$. Rewrite

$$\widehat{L}_{\mathcal{N}}(\widetilde{\lambda}_s) = I + Q_{\mathcal{N}} D_{\mathcal{N}} = Q_{\mathcal{N}}^{1/2} \left(I + Q_{\mathcal{N}}^{1/2} D_{\mathcal{N}} Q_{\mathcal{N}}^{-1/2}\right) Q_{\mathcal{N}}^{1/2}.$$

Observe that $I + Q_{\mathcal{N}}^{1/2} D_{\mathcal{N}} Q_{\mathcal{N}}^{-1/2} \succeq I$. We get

$$s_{\min}\big(\widehat{L}_{\mathcal{N}}(\widetilde{\lambda}_s)\big) \geqslant \sqrt{\frac{\lambda_{\min}(Q_{\mathcal{N}})}{\lambda_{\max}(Q_{\mathcal{N}})}}.$$

To bound the eigenvalues of $Q_{\mathcal{N}}$, note that for any unit vector v, $v^\top Q_{\mathcal{N}} v = (U_{\mathcal{N}} v)^\top \Phi(\widetilde{\lambda}_s)(U_{\mathcal{N}} v)$. Combining Lemma 5.1 (i), we get

$$\frac{1}{2\widetilde{\lambda}_s} \leqslant \min_i \phi(\widetilde{\lambda}_s) \leqslant \lambda_{\min}(Q_{\mathcal{N}}) \leqslant \lambda_{\max}(Q_{\mathcal{N}}) \leqslant \max_i \phi(\widetilde{\lambda}_s) \leqslant \frac{3}{2\widetilde{\lambda}_s}$$

and thus

$$s_{\min}\big(\widehat{L}_{\mathcal{N}}(\widetilde{\lambda}_s)\big) \geqslant \sqrt{\frac{1}{3}}.$$

By the same estimation as in **Step 1**, we bound the right-hand side of (B.7) by

$$2\widetilde{\lambda}_s \left(\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\widetilde{\lambda}_s^3} + \frac{13}{2}\frac{R_{\max}}{\widetilde{\lambda}_s^5}\operatorname{osc}(\mathcal{R}) + \|U^\top \Xi(\widetilde{\lambda}_s) U\|\right).$$

Hence,

$$\|U_{\mathcal{N}}^\top \widetilde{u}_s\| \leqslant 2\sqrt{3}\left(\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\widetilde{\lambda}_s^2} + \frac{13}{2}\frac{R_{\max}}{\widetilde{\lambda}_s^4}\operatorname{osc}(\mathcal{R}) + \widetilde{\lambda}_s \|U^\top \Xi(\widetilde{\lambda}_s) U\|\right).$$

This completes the proof.

C Proof of Proposition 8.1

Before we prove Proposition 8.1, we introduce the necessary notation and a cumulant expansion formula to be used in the proof (adapted from [62]).

For any matrix $F \in \mathbb{R}^{n \times n}$ and vectors $v, w \in \mathbb{R}^n$, we write

$$F_{vw} := v^\top F w, \qquad F_{iw} := e_i^\top F w, \qquad F_{vj} := v^\top F e_j, \qquad F_{ij} := e_i^\top F e_j,$$

where $\{e_1, \dots, e_n\}$ denotes the standard basis of $\mathbb{R}^n$. Moreover, for $1 \leqslant i, j \leqslant n$ define

$$\Delta^{ij} := (1 + \delta_{ij})^{-1}\big(e_i e_j^\top + e_j e_i^\top\big), \tag{C.1}$$

and for a differentiable function $f = f(H)$ of a symmetric matrix $H = (H_{ij})$ (with $\{H_{ij} : i \leqslant j\}$ treated as independent variables) set

$$\partial_{ij} f \;:=\; \frac{\partial}{\partial H_{ij}} f(H) \;=\; \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} f\big(H + t\,\Delta_{ij}\big).$$

Since there is potential ambiguity between $\big|\partial_{ij}^r(G_{vw})\big|$ and $\big|(\partial_{ij}^r G)_{vw}\big|$, we explicitly note that

$$\partial_{ij}^r\big(v^\top G w\big) \;=\; v^\top\big(\partial_{ij}^r G\big) w,$$

and throughout the proof we therefore write $\big|\partial_{ij}^r G_{vw}\big|$ with no confusion.

LEMMA C.1 (Cumulant expansion, Lemma 2.4 from [62])**.** *Let h be a real random variable with finite moments of all orders. For $k \geqslant 1$ define the k-th cumulant by*

$$\mathcal{C}_k(h) := (-\mathrm{i})^k \, \frac{\mathrm{d}^k}{\mathrm{d}t^k} \log \mathbb{E}\big[e^{\mathrm{i}th}\big]\bigg|_{t=0}.$$

Let $f : \mathbb{R} \to \mathbb{C}$ be smooth, and write $f^{(k)}$ for its k-th derivative. Fix $\ell \in \mathbb{N}$. Then

$$\mathbb{E}\big[hf(h)\big] = \sum_{k=0}^{\ell} \frac{1}{k!}\,\mathcal{C}_{k+1}(h)\,\mathbb{E}\big[f^{(k)}(h)\big] + \mathtt{R}_{\ell+1},$$

where the remainder $R_{\ell+1}$ satisfies, for any $t > 0$,

$$\begin{aligned}|\mathtt{R}_{\ell+1}| \leqslant 2^\ell(\ell+2) \cdot \left(\mathbb{E}\Big[\sup_{|x| \leqslant |h|}\big|f^{(\ell+1)}(x)\big|^2\Big] \cdot \mathbb{E}\big[|h|^{2\ell+4}\mathbf{1}(|h| > t)\big]\right)^{1/2} \\ + 2^\ell(\ell+2) \cdot \mathbb{E}|h|^{\ell+2} \cdot \sup_{|x| \leqslant t}\big|f^{(\ell+1)}(x)\big|.\end{aligned}$$

Now we provide the proof. Define

$$L := \operatorname{diag}(l_1, \dots, l_n) \quad \text{with} \quad l_k := \sum_{s=1}^{n} \sigma_{ks}^2 G_{ss}. \tag{C.2}$$

The main technical task is to establish:

PROPOSITION C.1. *Fix $D > 0$. Let $v, w \in \mathbb{R}^n$ be deterministic unit vectors, and let $z \in \mathbb{C}$ satisfy $|z| \geqslant 6\sqrt{R_{\max}}$.*

(i) *Under the general sub-Gaussian regime, with probability at least* $1 - 6n^{-D-4}$,

$$\big|v^\top(E - L)Gw\big| \leqslant C(D+6)^{3/2}\frac{\beta K \sigma_{\max}\log^2 n}{|z|} + Cn^{-500},$$

where $C \geqslant 1$ is an absolute constant.

(ii) *Under the bounded sparse-entry regime, namely if* $\max_{i,j} |E_{ij}| \leqslant c_0 K\sigma_{\max}$ *almost surely, then, with probability at least* $1 - 6n^{-D-4}$,

$$\left|v^\top (E-L)Gw\right| \leqslant C(D+6)c_0 \frac{K\sigma_{\max}\log n}{|z|},$$

where $C > 0$ is an absolute constant.

The key to proving Proposition C.1 is the following moment estimates, whose proof is deferred to Section C.1. For brevity, we denote

$$\mathcal{D} := (E-L)G.$$

and bound the moments of

$$\mathcal{D}_{vw} = v^\top (E-L)Gw.$$

We define a regime-dependent truncation parameter

$$\mathtt{L}_n := \begin{cases} 100\sqrt{D+6}K\sigma_{\max}\log n, & \text{in the general sub-Gaussian regime,} \\ c_0 K\sigma_{\max}, & \text{in the bounded sparse-entry regime.} \end{cases} \tag{C.3}$$

Define

$$\Omega_n^0 := \left\{ \|E\| \leqslant 3\sqrt{R_{\max}}, \quad \max_{a,b} |E_{ab}| \leqslant \mathtt{L}_n \right\}$$

and a slightly larger event

$$\Omega_n := \left\{ \|E\| \leqslant 3.5\sqrt{R_{\max}}, \quad \max_{a,b} |E_{ab}| \leqslant 2\mathtt{L}_n \right\}. \tag{C.4}$$

Let $\chi = \chi(E) \in [0,1]$ be a smooth cutoff such that

$$\chi(E) = 1 \quad \text{on } \Omega_n^0, \qquad \chi(E) = 0 \quad \text{outside } \Omega_n.$$

Our main technical estimate is the following bound:

LEMMA C.2. *For any integer $p \geqslant 3$,*

$$\begin{aligned}\mathbb{E}(|\mathcal{D}_{vw}|^{2p}\chi(E)) \leqslant & CK\beta \sum_{m=1}^{3} p^{m-1} \frac{\sigma_{\max}^m}{|z|^m} \mathbb{E}\left[\left(|\mathcal{D}_{vw}| + 22\frac{\mathtt{L}_n}{|z|} \right)^{2p-m} \chi(E) \right] \\ & + CK^2\beta^2 \sum_{m=1}^{4} p^{m-1} \frac{\sigma_{\max}^{m+1}}{|z|^{m+1}} \mathbb{E}\left[\left(|\mathcal{D}_{vw}| + 22\frac{\mathtt{L}_n}{|z|} \right)^{2p-m} \chi(E) \right] + \varepsilon_p\end{aligned}$$

for an absolute constant $C \geqslant 1$. Here, we define

$$\varepsilon_p := C^p p^3 \exp\bigl(-5000(D+6)(\log n)^2\bigr) \frac{(\beta K\sigma_{\max})^2}{|z|^2} \tag{C.5}$$

in the general sub-Gaussian regime and $\varepsilon_p := 0$ in the bounded sparse-entry regime.

Using this lemma, we first obtain a bound on

$$\mathtt{Z}_p := \left(\mathbb{E}(|\mathcal{D}_{vw}|^{2p}\chi(E))\right)^{1/2p}.$$

By the Hölder inequality, for any $q \leqslant 2p$,

$$\mathbb{E}[|\mathcal{D}_{vw}|^q\chi(E)] \leqslant \left(\mathbb{E}|\mathcal{D}_{vw}|^{2p}\chi(E)\right)^{\frac{q}{2p}} = \mathtt{Z}_p^q.$$

The same argument gives

$$\mathbb{E}\left[\left(|\mathcal{D}_{vw}| + 22\frac{\mathtt{L}_n}{|z|}\right)^q\chi(E)\right] \leqslant \left(\mathbb{E}\left[\left(|\mathcal{D}_{vw}| + 22\frac{\mathtt{L}_\mathtt{n}}{|z|}\right)^{2p}\chi(E)\right]\right)^{\frac{q}{2p}}.$$

Note that by the Minkowski inequality,

$$\left(\mathbb{E}\left[\left(|\mathcal{D}_{vw}| + 22\frac{\mathtt{L}_\mathtt{n}}{|z|}\right)^{2p}\chi(E)\right]\right)^{\frac{1}{2p}} \leqslant \left(\mathbb{E}|\mathcal{D}_{vw}|^{2p}\chi(E)\right)^{\frac{1}{2p}} + 22\frac{\mathtt{L}_\mathtt{n}}{|z|} = \mathtt{Z}_p + 22\frac{\mathtt{L}_\mathtt{n}}{|z|}.$$

We get

$$\mathbb{E}\left[\left(|\mathcal{D}_{vw}| + 22\frac{\mathtt{L}_\mathtt{n}}{|z|}\right)^q\chi(E)\right] \leqslant \left(\mathtt{Z}_p + 22\frac{\mathtt{L}_n}{|z|}\right)^q.$$

Therefore, Lemma C.2 implies

$$\begin{aligned}\mathtt{Z}_p^{2p} \leqslant & CK\beta\sum_{m=1}^{3} p^{m-1}\frac{\sigma_{\max}^m}{|z|^m}\left(\mathtt{Z}_p + 22\frac{\mathtt{L}_n}{|z|}\right)^{2p-m} \\ & + CK^2\beta^2\sum_{m=1}^{4} p^{m-1}\frac{\sigma_{\max}^{m+1}}{|z|^{m+1}}\left(\mathtt{Z}_p + 22\frac{\mathtt{L}_n}{|z|}\right)^{2p-m} + \varepsilon_p,\end{aligned} \tag{C.6}$$

where ε_p is defined in (C.5).

We claim that

$$\mathtt{Z}_p \leqslant C\left(p\frac{\mathtt{L}_\mathtt{n}}{|z|} + p\frac{\beta K\sigma_{\max}}{|z|} + \left(\frac{\beta K\sigma_{\max}}{|z|}\right)^2 + \varepsilon_p^{1/(2p)}\right). \tag{C.7}$$

To see this, consider the case where $\mathtt{Z}_p \leqslant 88p\frac{\mathtt{L}_\mathtt{n}}{|z|} + 2\varepsilon_p^{1/(2p)}$. In this regime, (C.7) holds trivially. Now, let us assume $\mathtt{Z}_p > 88p\frac{\mathtt{L}_\mathtt{n}}{|z|}$ and $\mathtt{Z}_p > 2\varepsilon_p^{1/(2p)}$. Thus, $\mathtt{Z}_p + 22\frac{\mathtt{L}_n}{|z|} \leqslant (1+\frac{1}{4p})\mathtt{Z}_p$ and for $1 \leqslant m \leqslant 4$,

$$\left(\mathtt{Z}_p + 22\frac{\mathtt{L}_\mathtt{n}}{|z|}\right)^{2p-m} \leqslant \left(1 + \frac{1}{4p}\right)^{2p}\mathtt{Z}_p^{2p-m} \leqslant e^{1/2}\mathtt{Z}_p^{2p-m}.$$

Moreover, the last term in (C.6) satisfies $\varepsilon_p \leqslant 2^{-2p}\mathtt{Z}_p^{2p}$ and can be absorbed into the left-hand side. Hence, (C.6) simplifies to

$$\mathtt{Z}_p^{2p} \leqslant \sum_{m=1}^{3} CK\beta p^{m-1}\frac{\sigma_{\max}^m}{|z|^m}\mathtt{Z}_p^{2p-m} + \sum_{m=1}^{4} CK^2\beta^2 p^{m-1}\frac{\sigma_{\max}^{m+1}}{|z|^{m+1}}\mathtt{Z}_p^{2p-m} \tag{C.8}$$

for an absolute constant $C \geqslant 1$. Next, we apply Young's inequality for $\mathtt{B} \cdot \mathtt{Z}_p^{2p-m}$. Let $\delta > 0$ be a small constant to be specified later. Since

$$\begin{aligned}\mathtt{B} \cdot \mathtt{Z}_p^{2p-m} = (\delta \mathtt{Z}_p^{2p-m}) \cdot \frac{\mathtt{B}}{\delta} &\leqslant \frac{(\delta \mathtt{Z}_p^{2p-m})^{2p/(2p-m)}}{2p/(2p-m)} + \frac{(\mathtt{B}/\delta)^{2p/m}}{2p/m} \\ &= \frac{\delta^{2p/(2p-m)}}{2p/(2p-m)} \mathtt{Z}_p^{2p} + \frac{(1/\delta)^{2p/m}}{2p/m} \mathtt{B}^{\frac{2p}{m}}.\end{aligned}$$

Choose $\delta = 1/100$. Then $\frac{\delta^{2p/(2p-m)}}{2p/(2p-m)} < 1/10$ for $1 \leqslant m \leqslant 4$. We absorb the $\mathtt{Z}_p^{2p}$ terms to the left-hand side of (C.8). Continuing from (C.8), we have

$$\mathtt{Z}_p^{2p} \leqslant \sum_{m=1}^{3} \left(CK\beta p^{m-1} \frac{\sigma_{\max}^m}{|z|^m} \right)^{\frac{2p}{m}} + \sum_{m=1}^{4} \left(CK^2\beta^2 p^{m-1} \frac{\sigma_{\max}^{m+1}}{|z|^{m+1}} \right)^{\frac{2p}{m}}.$$

Taking the $2p$-th roots on both sides yields

$$\mathtt{Z}_p \leqslant C \max_{1 \leqslant m \leqslant 3} \left(K\beta p^{m-1} \frac{\sigma_{\max}^m}{|z|^m} \right)^{\frac{1}{m}} + C \max_{1 \leqslant m \leqslant 4} \left(K^2\beta^2 p^{m-1} \frac{\sigma_{\max}^{m+1}}{|z|^{m+1}} \right)^{\frac{1}{m}}.$$

Since $K\beta \geqslant 1$, $p \geqslant 3$, and $|z| \geqslant \sigma_{\max}$, we have for $1 \leqslant m \leqslant 3$,

$$\left(K\beta p^{m-1} \frac{\sigma_{\max}^m}{|z|^m} \right)^{\frac{1}{m}} = (K\beta)^{\frac{1}{m}} p^{\frac{m-1}{m}} \frac{\sigma_{\max}}{|z|} \leqslant p \frac{\beta K \sigma_{\max}}{|z|}$$

and for $2 \leqslant m \leqslant 4$,

$$\begin{aligned}(K\beta)^{\frac{2}{m}} p^{\frac{m-1}{m}} \left(\frac{\sigma_{\max}}{|z|} \right)^{1+\frac{1}{m}} &= \left(\frac{\beta K \sigma_{\max}}{|z|} \right) \cdot (K\beta)^{\frac{2-m}{m}} \cdot p^{\frac{m-1}{m}} \cdot \left(\frac{\sigma_{\max}}{|z|} \right)^{\frac{1}{m}} \\ &\leqslant p \frac{\beta K \sigma_{\max}}{|z|}.\end{aligned}$$

The above inequality simplifies to

$$\mathtt{Z}_p \leqslant C \left(p \frac{\beta K \sigma_{\max}}{|z|} + \left(\frac{\beta K \sigma_{\max}}{|z|} \right)^2 \right).$$

Combining this bound with the exceptional case proves (C.7).

Now we prove Proposition 8.1. Let us first consider the general sub-Gaussian regime. By the union bound and our sub-Gaussian tail assumption,

$$\mathbb{P}\left(\max_{a,b} |E_{ab}| > 2\mathtt{L}_n \right) \leqslant \sum_{a,b} \mathbb{P}(|E_{ab}| > \mathtt{L}_n) \leqslant 2n^2 \exp\left(-10^4(D+6)(\log n)^2\right) \leqslant 2n^{-D-4}.$$

Moreover, by Lemma 5.4, we have

$$\mathbb{P}(\|E\| > 3\sqrt{R_{\max}}) \leqslant 3n^{-D-4}$$

by selecting the constant C_0 in Assumption 1.2 sufficiently large. Hence,

$$\mathbb{P}\big((\Omega_n^0)^c\big) \leqslant 5n^{-D-4}.$$

From the decomposition $|\mathcal{D}_{vw}| = |\mathcal{D}_{vw}|\mathbf{1}_{\Omega_n^0} + |\mathcal{D}_{vw}|\mathbf{1}_{(\Omega_n^0)^c}$ and $\chi(E) = 1$ on Ω_n^0, we have $|\mathcal{D}_{vw}|\mathbf{1}_{\Omega_n^0} \leqslant |\mathcal{D}_{vw}|\chi(E)$. Thus

$$|\mathcal{D}_{vw}| \leqslant |\mathcal{D}_{vw}|\chi(E) + |\mathcal{D}_{vw}|\mathbf{1}_{(\Omega_n^0)^c}.$$

It follows that

$$\mathbb{P}(|\mathcal{D}_{vw}| \geqslant t) \leqslant \mathbb{P}(|\mathcal{D}_{vw}|\chi(E) \geqslant t) + \mathbb{P}\big((\Omega_n^0)^c\big).$$

Then by Markov's inequality and $0 \leqslant \chi \leqslant 1$, for any integer $p \geqslant 3$,

$$\begin{aligned}\mathbb{P}(|\mathcal{D}_{vw}| \geqslant t) &\leqslant \frac{\mathbb{E}(|\mathcal{D}_{vw}|^{2p}\chi(E))}{t^{2p}} + 5n^{-D-4}\\ &= \left(\frac{\mathtt{Z}_p}{t}\right)^{2p} + 5n^{-D-4}.\end{aligned}$$

We choose $p = \lceil (D+4)\log n \rceil$ and in view of (C.7),

$$t = e \cdot C\left(p\frac{\mathtt{L}_n}{|z|} + p\frac{\beta K\sigma_{\max}}{|z|} + \left(\frac{\beta K\sigma_{\max}}{|z|}\right)^2 + \varepsilon_p^{1/(2p)}\right).$$

With this choice of t,

$$\left(\frac{\mathtt{Z}_p}{t}\right)^{2p} \leqslant e^{-2p} \leqslant \exp\big(-2(D+4)\log n\big) = n^{-2D-8} \leqslant n^{-D-4}.$$

It follows that

$$\mathbb{P}(|\mathcal{D}_{vw}| \geqslant t) \leqslant n^{-D-4} + 5n^{-D-4} = 6n^{-D-4}.$$

Finally, let us simplify the terms in t under the choice of p. For the term $\varepsilon_p^{1/(2p)}$, note that by Assumption 1.2, $R_{\max} \leqslant n\sigma_{\max}^2$ and $\sqrt{R_{\max}} \geqslant C_0(D+8)K\sigma_{\max}\log n$. We have

$$K \leqslant \frac{\sqrt{n}}{C_0(D+8)\log n}.$$

As discussed in Remark 1.1, $\beta \leqslant C\sqrt{\log(eK)} \leqslant C\sqrt{\log n}$. Under our assumption $|z| \geqslant 6\sqrt{R_{\max}}$, $\frac{\beta K\sigma_{\max}}{|z|} \leqslant C'$. Thus

$$\begin{aligned}\varepsilon_p^{1/(2p)} &= C^{1/2}p^{3/(2p)}\exp\left(-\frac{10(D+6)\log^2 n}{2p}\right)\left(\frac{\beta K\sigma_{\max}}{|z|}\right)^{1/p}\\ &\leqslant C\exp\left(-\frac{5000(D+6)\log^2 n}{2(D+5)\log n}\right) \leqslant C\exp(-500\log n) = Cn^{-500}.\end{aligned}$$

Next, substituting $p \leqslant (D+5)\log n$ and $\mathtt{L}_n = 5\sqrt{D+6}K\sigma_{\max}\log n$, we also have

$$p\frac{\mathtt{L}_n}{|z|} \leqslant 5(D+6)^{3/2}\frac{K\sigma_{\max}\log^2 n}{|z|} \quad \text{and} \quad p\frac{\beta K\sigma_{\max}}{|z|} \leqslant (D+5)\frac{\beta K\sigma_{\max}\log n}{|z|}.$$

Note that

$$p\frac{\beta K\sigma_{\max}}{|z|} \geqslant \left(\frac{\beta K\sigma_{\max}}{|z|}\right)^2 \tag{C.9}$$

by selecting the constant C_0 sufficiently large in Assumption 1.2. Thus, we conclude that with probability at least $1-6n^{-D-4}$,

$$\begin{aligned}|\mathcal{D}_{vw}| &\leqslant C\left(5(D+6)^{3/2}\frac{K\sigma_{\max}\log^2 n}{|z|} + (D+5)\frac{\beta K\sigma_{\max}\log n}{|z|} + n^{-500}\right)\\ &\leqslant C(D+6)^{3/2}\frac{K\sigma_{\max}\log^2 n}{|z|} + Cn^{-500}.\end{aligned}$$

We now prove the bounded sparse-entry case. In this regime, $\mathtt{L}_n = c_0 K\sigma_{\max}$ and $\varepsilon_p = 0$. Thus from (C.7), together with (C.9), we obtain

$$\mathtt{Z}_p \leqslant C\left(p\frac{c_0 K\sigma_{\max}}{|z|} + p\frac{\beta K\sigma_{\max}}{|z|}\right).$$

Moreover, the entrywise truncation event is deterministic, i.e., $\max_{i,j}|E_{ij}| \leqslant \mathtt{L}_n$ almost surely. Also, $\beta \leqslant c_0$. Taking $p = \lceil (D+4)\log n\rceil$ and applying the same Markov argument as above, together with the operator norm event, yields

$$|\mathcal{D}_{vw}| \leqslant C(D+6)\frac{c_0 K\sigma_{\max}\log n}{|z|}$$

with probability at least $1-6n^{-D-4}$. This proves part (ii) and completes the proof of Proposition C.1.

C.1 Proof of Lemma C.2

We prove Lemma C.2 in this section. Let

$$Q := \mathcal{D}_{vw}^{p-1}\overline{\mathcal{D}}_{vw}^{p}.$$

We rewrite

$$\mathbb{E}[|\mathcal{D}_{vw}|^{2p}\chi(E)] = \mathbb{E}\left[(EG)_{vw}Q\chi(E)\right] - \mathbb{E}\left[(LG)_{vw}\,Q\chi(E)\right]. \tag{C.10}$$

We analyze the first term on the right-hand side of (C.10) using the cumulant expansion. For $1 \leqslant i,j \leqslant n$, define the auxiliary functions

$$h_{ij}(E) := G_{jw}Q, \qquad f_{ij}^{\chi}(E) := h_{ij}(E)\chi(E).$$

Thus

$$\mathbb{E}\left[(EG)_{vw}Q\,\chi(E)\right] = \sum_{i,j} v_i\,\mathbb{E}\left[E_{ij}f_{ij}^{\chi}(E)\right].$$

Let $\chi \equiv \chi(E)$. For $l \leqslant 3$, the product rule gives

$$\partial_{ij}^{l} f_{ij}^{\chi} = \chi\cdot\partial_{ij}^{l}h_{ij} + \sum_{a=1}^{l}\binom{l}{a}(\partial_{ij}^{l-a}h_{ij})(\partial_{ij}^{a}\chi).$$

The first term on the right-hand side is the main term estimated below. The remaining terms contain at least one derivative of the cutoff function. We choose χ so that, for $1 \leqslant a \leqslant 3$,

$$|\partial_{ij}^a \chi(E)| \leqslant C_a \mathtt{L}_n^{-a}.$$

These derivative terms are supported in $\Omega_n \backslash \Omega_n^0$. Using the derivative bounds proved below, together with Cauchy-Schwarz inequality and the high-probability estimate for $(\Omega_n^0)^c$, the total contribution of the terms containing $\partial_{ij}^a \chi$, $a \geqslant 1$, is absorbed into an error $O(n^{-D-5})$, after increasing C_0 if necessary.

In the estimates below, we write only the terms in which no derivative falls on χ. The terms involving derivatives of χ are supported on $\Omega_n \backslash \Omega_n^0$. By the same derivative estimates used below and the probability bound for this transition region, their total contribution is negligible compared with the final high-probability bound. We therefore record only the main terms in the moment estimate. The remainder ε_p below refers only to the tail part of the cumulant remainder; in the bounded sparse-entry regime this tail part is zero. Therefore, to simplify notation, we suppress the cutoff $\chi(E)$ from the main estimates. From this point on, we simply work with

$$f_{ij}(E) := G_{jw} Q$$

Since E is symmetric, the independent random variables are E_{ij} $(1 \leqslant i \leqslant j \leqslant n)$. For $i < j$, we write ∂_{ij} for differentiation with respect to $E_{ij} = E_{ji}$, and we set $\partial_{ji} := \partial_{ij}$. For the diagonal entries, ∂_{ii} denotes differentiation with respect to E_{ii}. With this convention, for $x, y \in \mathbb{R}^n$,

$$\partial_{ij} G_{xy} = (1 + \delta_{ij})^{-1} (G_{xi} G_{jy} + G_{xj} G_{iy}), \tag{C.11}$$

Now, from

$$(EG)_{vw} Q = \sum_i v_i E_{ii} G_{iw} Q + \sum_{i<j} E_{ij} \big(v_i G_{jw} + v_j G_{iw}\big) Q = \sum_{i,j} v_i E_{ij} f_{ij}(E),$$

we apply the cumulant expansion Lemma C.1 to each independent entry E_{ij} $(1 \leqslant i \leqslant j \leqslant n)$ with $l = 2$ to get

$$\mathbb{E}[(EG)_{vw} Q] := X_1 + X_2 + \mathtt{R}_3,$$

where for $k = 1, 2$,

$$X_k := \sum_{i,j} v_i \frac{1}{k!} \mathcal{C}_{k+1}(E_{ij}) \mathbb{E}\left[\partial_{ij}^k f_{ij}(E)\right]. \tag{C.12}$$

Here the $k = 0$ cumulant term vanishes because the entries of E are centered. The remainder is written as

$$\mathtt{R}_3 := \sum_{i,j} v_i \mathtt{R}_3^{ij} \tag{C.13}$$

where

$$\mathtt{R}_3^{ij} := \mathbb{E}\big[E_{ij} f_{ij}(E)\big] - \sum_{k=1}^{2} \frac{1}{k!} \mathcal{C}_{k+1}(E_{ij}) \mathbb{E}\big[\partial_{ij}^k f_{ij}(E)\big].$$

With these notations, we have

$$\mathbb{E}[|\mathcal{D}_{vw}|^{2p}] = (X_1 - \mathbb{E}\left[(LG)_{vw} Q\right]) + X_2 + \mathtt{R}_3. \tag{C.14}$$

The next lemma bounds each term on the right-hand side of (C.14).

Lemma C.3. *(i) For X_1,*

$$|X_1 - \mathbb{E}\left[(LG)_{vw}\, Q\right]| \leqslant \frac{2}{3}\sigma_{\max}\frac{\mathbb{E}\left[|\mathcal{D}_{vw}|^{2p-1}\right]}{|z|} + 36p\sigma_{\max}^2\frac{\mathbb{E}[|\mathcal{D}_{vw}|^{2p-2}]}{|z|^2}.$$

(ii) For X_2,

$$|X_2| \leqslant \frac{4}{9}\beta K\sigma_{\max}\frac{\mathbb{E}\left[|\mathcal{D}_{vw}|^{2p-1}\right]}{|z|} + 178p\beta K\sigma_{\max}^2\frac{\mathbb{E}[|\mathcal{D}_{vw}|^{2p-2}]}{|z|^2} + 700p^2\beta K\sigma_{\max}^3\frac{\mathbb{E}[|\mathcal{D}_{vw}|^{2p-3}]}{|z|^3}.$$

(iii) For $\mathtt{R}_3$, recalling $\mathtt{L}_n$ from (C.3)*, we have*

$$\begin{aligned}|\mathtt{R}_3| \leqslant & CK^2\beta^2\sum_{m=1}^{4} p^{m-1}\frac{\sigma_{\max}^{m+1}}{|z|^{m+1}}\mathbb{E}\left[\left(|\mathcal{D}_{vw}| + 22\frac{\mathtt{L}_n}{|z|}\right)^{2p-m}\right] \\ & + C^p p^3\exp\big(-10(D+6)\log^2 n\big)\frac{(\beta K\sigma_{\max})^2}{|z|^2}\end{aligned}$$

for an absolute constant $C \geqslant 1$.

Again, we emphasize that in Lemma C.3, we suppress $\chi(E)$ from the notation. More precisely, every expectation in the bounds for X_1, X_2, and $\mathtt{R}_3$ is understood to include the factor $\chi(E)$, except for the explicitly separated cutoff-derivative terms.

With this lemma, the conclusion of Lemma C.2 follows from (C.14):

$$\begin{aligned}\mathbb{E}\big[|\mathcal{D}_{vw}|^{2p}\chi(E)\big] &\leqslant |X_1 - \mathbb{E}\left[(LG)_{vw}\, Q\right]| + |X_2| + |\mathtt{R}_3| \\ &\leqslant CK\beta\sum_{m=1}^{3} p^{m-1}\frac{\sigma_{\max}^m}{|z|^m}\mathbb{E}\left[|\mathcal{D}_{vw}|^{2p-m}\chi(E)\right] \\ &\quad + CK^2\beta^2\sum_{m=1}^{4} p^{m-1}\frac{\sigma_{\max}^{m+1}}{|z|^{m+1}}\mathbb{E}\left[\left(|\mathcal{D}_{vw}| + 22\frac{\mathtt{L}_n}{|z|}\right)^{2p-m}\chi(E)\right] \\ &\quad + C^p p^3\exp\big(-10(D+6)\log^2 n\big)\frac{(\beta K\sigma_{\max})^2}{|z|^2}.\end{aligned}$$

In the remainder of this section, we prove Lemma C.3.

Proof of Lemma C.3 (i). From the definition of $L = \operatorname{diag}(\sum_{j=1}^n \sigma_{kj}^2 G_{jj})_{1\leqslant k\leqslant n}$ in (C.2), we have

$$\mathbb{E}[(LG)_{vw}Q] = \mathbb{E}[\sum_{i,j} v_i L_{ij}G_{jw}Q] = \sum_{i,j} v_i\sigma_{ij}^2\mathbb{E}[G_{jj}G_{iw}Q]. \tag{C.15}$$

From (C.12), by product rule, we get

$$\begin{aligned}X_1 &= \sum_{i,j} v_i\mathcal{C}_2(E_{ij})\mathbb{E}[\partial_{ij}f_{ij}(E)] \\ &= \sum_{i,j} v_i\sigma_{ij}^2\mathbb{E}[\partial_{ij}(G_{jw}Q)] \\ &= \sum_{i,j} v_i\sigma_{ij}^2\mathbb{E}\left[(\partial_{ij}G_{jw})Q\right] + \sum_{i,j} v_i\sigma_{ij}^2\mathbb{E}\left[G_{jw}(\partial_{ij}Q)\right].\end{aligned}$$

Plugging in $\partial_{ij}G_{jw} = (1+\delta_{ij})^{-1}\big(G_{ji}G_{jw} + G_{jj}G_{iw}\big)$ by (C.11), we further obtain

$$X_1 = \sum_{i,j} v_i\sigma_{ij}^2\mathbb{E}\left[G_{jj}G_{iw}Q\right] + \sum_{i\neq j} v_i\sigma_{ij}^2\mathbb{E}\left[G_{ji}G_{jw}Q\right] + \sum_{i,j} v_i\sigma_{ij}^2\mathbb{E}\left[G_{jw}(\partial_{ij}Q)\right].$$

Comparing with (C.15), we arrive at

$$X_1 - \mathbb{E}\left[(LG)_{vw}\,Q\right] = T^{(1)} + T^{(2)},$$

where

$$T^{(1)} := \mathbb{E}\Big[\sum_{i\neq j} v_i\sigma_{ij}^2 G_{ji}G_{jw}Q\Big], \qquad T^{(2)} := \mathbb{E}\Big[\sum_{i,j} v_i\sigma_{ij}^2 G_{jw}(\partial_{ij}Q)\Big].$$

We now provide upper bounds for $T^{(1)}$ and $T^{(2)}$.

Step 1. *Estimates for* $T^{(1)}$. Applying Cauchy-Schwarz, we have

$$\Big|\sum_{i\neq j} v_i\sigma_{ij}^2 G_{ji}G_{jw}\Big| \leqslant \Big(\sum_{i\neq j} v_i^2\sigma_{ij}^2|G_{ij}|^2\Big)^{1/2}\Big(\sum_{i\neq j}\sigma_{ij}^2|G_{jw}|^2\Big)^{1/2}.$$

Note that

$$\sum_{i\neq j} v_i^2\sigma_{ij}^2|G_{ij}|^2 \leqslant \sum_i v_i^2\big(\sum_j \sigma_{ij}^2|G_{ij}|^2\big) \leqslant \sigma_{\max}^2\sum_i v_i^2\|Ge_i\|^2 \leqslant \sigma_{\max}^2\|G\|^2$$

and

$$\sum_{i\neq j}\sigma_{ij}^2|G_{jw}|^2 \leqslant \sum_j |G_{jw}|^2\big(\sum_i \sigma_{ij}^2\big) \leqslant R_{\max}\|Gw\|^2 \leqslant R_{\max}\|G\|^2.$$

We get

$$\Big|\sum_{i,j} v_i\sigma_{ij}^2 G_{ji}G_{jw}\Big| \leqslant \sigma_{\max}\sqrt{R_{\max}}\|G\|^2 \leqslant \frac{4\,\sigma_{\max}\sqrt{R_{\max}}}{|z|^2} \leqslant \frac{2}{3}\frac{\sigma_{\max}}{|z|}.$$

by applying (5.6) and the assumption $|z| \geqslant 6\sqrt{R_{\max}}$. Consequently,

$$|T^{(1)}| \leqslant \frac{2}{3}\frac{\sigma_{\max}}{|z|}\mathbb{E}\left[|\mathcal{D}_{vw}|^{2p-1}\right].$$

Step 2. *Estimates for* $T^{(2)}$. Recall that

$$T^{(2)} := \sum_{i,j} v_i\sigma_{ij}^2\mathbb{E}\left[G_{jw}\partial_{ij}(\mathcal{D}_{vw}^{p-1}\overline{\mathcal{D}}_{vw}^{p})\right].$$

For $p \geqslant 2$, by the product rule,

$$\partial_{ij}(\mathcal{D}_{vw}^{p-1}\overline{\mathcal{D}}_{vw}^{p}) = (p-1)(\partial_{ij}\mathcal{D}_{vw})\mathcal{D}_{vw}^{p-2}\overline{\mathcal{D}}_{vw}^{p} + p(\partial_{ij}\overline{\mathcal{D}}_{vw})|\mathcal{D}_{vw}|^{2(p-1)}.$$

For $p = 1$, the first term is absent, and the same final bound follows. Hence, it suffices to bound the two quantities

$$\mathtt{A} := \sum_{i,j} v_i \sigma_{ij}^2 G_{jw} (\partial_{ij} \mathcal{D}_{vw}), \qquad \bar{\mathtt{A}} := \sum_{i,j} v_i \sigma_{ij}^2 G_{jw} (\partial_{ij} \overline{\mathcal{D}}_{vw}).$$

We first estimate $\mathtt{A}$. From the identity $(zI - E)G = I$, we have $EG = zG - I$. Plugging into $\mathcal{D} = (E - L)G = EG - LG$ yields

$$\mathcal{D} = (zI - L)G - I.$$

Thus $\partial_{ij}\mathcal{D} = (zI - L)\partial_{ij} G - (\partial_{ij} L)G$. Recall Δ_{ij} from (C.1). We also have

$$\partial_{ij}\mathcal{D}_{vw} = v^\top (zI - L) G \Delta^{ij} G w - v^\top (\partial_{ij} L) G w.$$

Denote $\widetilde{v}^\top := v^\top (zI - L)$. From the definition of L in (C.2) and the partial derivative formula (C.11), we compute

$$\partial_{ij} l_s = \sum_k \sigma_{sk}^2 \partial_{ij} G_{kk} = 2(1 + \delta_{ij})^{-1} \sum_k \sigma_{sk}^2 G_{ki} G_{kj}.$$

Hence,

$$v^\top (\partial_{ij} L) G w = 2(1 + \delta_{ij})^{-1} \sum_{s,k} v_s \sigma_{s,k}^2 G_{ki} G_{kj} G_{sw}.$$

It follows that

$$\partial_{ij}\mathcal{D}_{vw} = (1 + \delta_{ij})^{-1} \Big(G_{\widetilde{v}i} G_{jw} + G_{\widetilde{v}j} G_{iw} - 2 \sum_{s,k} v_s \sigma_{sk}^2 G_{ki} G_{kj} G_{sw} \Big). \tag{C.16}$$

We are ready to bound

$$\begin{aligned} |\mathtt{A}| = \Big| \sum_{i,j} v_i \sigma_{ij}^2 G_{jw} (\partial_{ij} \mathcal{D}_{vw}) \Big| \leqslant & \Big| \sum_{i,j} v_i \sigma_{ij}^2 G_{jw}^2 G_{\widetilde{v}i} \Big| + \Big| \sum_{i,j} v_i \sigma_{ij}^2 G_{jw} G_{\widetilde{v}j} G_{iw} \Big| \\ & + 2 \Big| \sum_{i,j} v_i \sigma_{ij}^2 G_{jw} \sum_{s,k} v_s \sigma_{s,k}^2 G_{ki} G_{kj} G_{sw} \Big| \\ & := T_1^{(2)} + T_2^{(2)} + T_3^{(2)}. \end{aligned}$$

Note that

$$\|L\| \leqslant R_{\max} \|G\| \leqslant 2 \frac{R_{\max}}{|z|} \leqslant \frac{1}{18} |z|$$

by (5.6) and our supposition $|z| > 6\sqrt{R_{\max}}$. Thus $\|\widetilde{v}\| \leqslant \|zI - L\| \leqslant |z| + \|L\| \leqslant \frac{19}{18}|z|$. We will repeatedly use $\sum_i |G_{iw}|^2 = \|Gw\|^2 \leqslant \|G\|^2$.

Now we estimate $T_1^{(2)}, T_2^{(2)}$ and $T_3^{(2)}$ respectively. For $T_1^{(2)}$, since $\max_{i,j} \sigma_{ij}^2 \leqslant \sigma_{\max}^2$,

$$\begin{aligned} T_1^{(2)} &\leqslant \sigma_{\max}^2 \sum_i |v_i| |G_{\widetilde{v}i}| \cdot \sum_j |G_{jw}|^2 \leqslant \sigma_{\max}^2 \|Gw\|^2 \sqrt{\sum_i |v_i|^2 \sum_i |G_{\widetilde{v}i}|^2} \\ &\leqslant \sigma_{\max}^2 \|G\|^2 \|\widetilde{v}^\top G\| \leqslant \sigma_{\max}^2 \|G\|^3 \|\widetilde{v}\| \leqslant \frac{76 \sigma_{\max}^2}{9|z|^2}. \end{aligned}$$

Similarly,

$$\begin{aligned}T_2^{(2)} &\leqslant \sigma_{\max}^2 \sum_i |v_i||G_{iw}| \cdot \sum_j |G_{jw}||G_{\widetilde{v}j}| \leqslant \sigma_{\max}^2 \sqrt{\sum_i |v_i|^2 \sum_i |G_{iw}|^2} \sqrt{\sum_j |G_{jw}|^2 \sum_j |G_{\widetilde{v}j}|^2} \\ &\leqslant \sigma_{\max}^2 \|G\|^3 \|\widetilde{v}\| \leqslant \frac{76\sigma_{\max}^2}{9|z|^2}.\end{aligned}$$

For $T_3^{(2)}$, observe that

$$\sum_s |v_s \sigma_{sk}^2||G_{sw}| \leqslant \sigma_{\max}^2 \sum_s |v_s||G_{sw}| \leqslant \sigma_{\max}^2 \|G\|$$

and

$$\sum_k |G_{ki}G_{kj}| \leqslant \sqrt{\sum_k |G_{ki}|^2 \sum_k |G_{kj}|^2} \leqslant \|G\|^2$$

by Cauchy-Schwarz inequality. We further have

$$\begin{aligned}T_3^{(2)} &\leqslant 2\sigma_{\max}^2 \|G\|^3 \sum_{i,j} |v_i \sigma_{ij}| \cdot |\sigma_{ij} G_{jw}| \leqslant 2\sigma_{\max}^2 \|G\|^3 \sqrt{\sum_{i,j} |v_i|^2 \sigma_{ij}^2} \sqrt{\sum_{i,j} \sigma_{ij}^2 |G_{jw}|^2} \\ &\leqslant 2\sigma_{\max}^2 \|G\|^3 \sqrt{R_{\max}} \sqrt{R_{\max} \|Gw\|^2} \leqslant 2\sigma_{\max}^2 R_{\max} \|G\|^4 \leqslant \frac{8\sigma_{\max}^2}{9|z|^2}.\end{aligned}$$

Combining the above estimates, we obtain

$$|\mathtt{A}| \leqslant \frac{76\sigma_{\max}^2}{9|z|^2} + \frac{76\sigma_{\max}^2}{9|z|^2} + \frac{8\sigma_{\max}^2}{9|z|^2} = \frac{160\sigma_{\max}^2}{9|z|^2} \leqslant \frac{18\sigma_{\max}^2}{|z|^2}.$$

The same estimate holds for $\bar{\mathtt{A}}$. Indeed, since $\partial_{ij}\overline{\mathcal{D}}_{vw} = \overline{\partial_{ij}\mathcal{D}_{vw}}$, the preceding argument applies with G, L, z replaced by $\overline{G}, \overline{L}, \overline{z}$. Hence,

$$|\bar{\mathtt{A}}| \leqslant \frac{18\sigma_{\max}^2}{|z|^2}.$$

Finally, we bound

$$\begin{aligned}|T^{(2)}| &\leqslant (p-1)|\mathbb{E}\big[|\mathtt{A}| \cdot \mathcal{D}_{vw}^{p-2}\overline{\mathcal{D}}_{vw}^{p}\big]| + p|\mathbb{E}\big[|\bar{\mathtt{A}}| \cdot |\mathcal{D}_{vw}|^{2(p-1)}\big]| \\ &\leqslant \frac{36p\sigma_{\max}^2}{|z|^2} \mathbb{E}[|\mathcal{D}_{vw}|^{2p-2}].\end{aligned}$$

This completes the proof of Lemma C.3 (i).

Proof of Lemma C.3 (ii). For brevity, we suppress the harmless distinction between $\mathcal{D}_{vw}^{p-1}\overline{\mathcal{D}}_{vw}^{p}$ and $\mathcal{D}_{vw}^{2p-1}$. The argument with complex conjugates is identical, using $\partial_{ij}\overline{\mathcal{D}}_{vw} = \overline{\partial_{ij}\mathcal{D}_{vw}}$.

Recall that

$$X_2 = \sum_{i,j} v_i \, \frac{1}{2} \, \mathcal{C}_3(E_{ij}) \, \mathbb{E}\big[\partial_{ij}^2 \big(G_{jw} \, \mathcal{D}_{vw}^{2p-1}\big)\big] \, .$$

Using the Leibniz rule and our assumption $|\mathcal{C}_3(E_{ij})| = |\mathbb{E}(E_{ij}^3)| \leqslant \beta K \sigma_{\max}\sigma_{ij}^2$, we get

$$|X_2| \leqslant \beta K \sigma_{\max} \sum_{s=0}^{2} \mathbb{E}\Big[\sum_{i,j} |v_i|\sigma_{ij}^2 |\partial_{ij}^{2-s}(G_{jw})| \cdot |\partial_{ij}^{s}(\mathcal{D}_{vw}^{2p-1})|\Big]. \tag{C.17}$$

We estimate the three values $s = 0, 1, 2$ separately.

• **Case 1:** $s = 0$**.** We estimate

$$\sum_{i,j} |v_i|\sigma_{ij}^2 |\partial_{ij}^2(G_{jw})|.$$

Using

$$\partial_{ij}^2 G = 2! G(\Delta^{ij} G)^2,$$

one obtains

$$\sum_{i,j} |v_i|\sigma_{ij}^2 |\partial_{ij}^2(G_{jw})| \leqslant 2\sum_{i,j} |v_i|\sigma_{ij}^2 |e_j^\top G(\Delta^{ij}G)^2 w| = 2\sum_j \big(\sum_i |v_i|\sigma_{ij}^2\big)\cdot |e_j^\top G(\Delta^{ij}G)^2 w|.$$

Let $\mathrm{v} = (|v_i|)$ and v is still a unit vector. By Cauchy-Schwartz, together with $\|\Delta^{ij}\| \leqslant 1$, we get

$$\begin{aligned}\sum_{i,j} |v_i|\sigma_{ij}^2 |\partial_{ij}^2(G_{jw})| &\leqslant 2\sqrt{\sum_j \big(\sum_i \sigma_{ij}^2 \mathrm{v}_i\big)^2}\sqrt{\sum_j |e_j^\top G(\Delta^{ij}G)^2 w|^2} \\ &= 2\|\Sigma \mathrm{v}\| \cdot \|G(\Delta^{ij}G)^2 w\| \leqslant 2\|\Sigma\|\cdot\|G\|^3.\end{aligned}$$

Since $\|\Sigma\| \leqslant \sqrt{\|\Sigma\|_1 \cdot \|\Sigma\|_\infty} = R_{\max}$, $\|G\| \leqslant 2/|z|$ and $|z| \geqslant 6\sqrt{R_{\max}}$, this gives

$$\sum_{i,j} |v_i|\sigma_{ij}^2 |\partial_{ij}^2(G_{jw})| \leqslant 16\frac{R_{\max}}{|z|^3} \leqslant \frac{4}{9}\frac{1}{|z|}.$$

Therefore, the contribution from the $s = 0$ term in (C.17) is bounded by

$$\frac{4}{9}\beta K \sigma_{\max} \frac{\mathbb{E}|\mathcal{D}_{vw}|^{2p-1}}{|z|}.$$

• **Case 2:** $s = 1$**.** We estimate

$$\mathbb{E}\Big[\sum_{i,j} |v_i|\sigma_{ij}^2 |\partial_{ij} G_{jw}| \cdot |\partial_{ij}(\mathcal{D}_{vw}^{2p-1})|\Big].$$

From (C.11), we first have

$$\partial_{ij} G_{jw} = (1+\delta_{ij})^{-1}(G_{ij}G_{jw} + G_{jj}G_{iw}).$$

By the chain rule,

$$|\partial_{ij}(\mathcal{D}_{vw}^{2p-1})| \leqslant 2p|\mathcal{D}_{vw}|^{2p-2}|\partial_{ij}\mathcal{D}_{vw}|.$$

Recall from (C.16) that

$$\partial_{ij}\mathcal{D}_{vw} = (1+\delta_{ij})^{-1}\Big(G_{\widetilde{v}i}G_{jw} + G_{\widetilde{v}j}G_{iw} - \mathcal{K}\Big),$$

where for brevity we define

$$\mathcal{K} := 2\sum_{s,k} v_s \sigma_{sk}^2 G_{ki} G_{kj} G_{sw}.$$

We also recall that $\widetilde{v}^\top := v^\top(zI - L)$ and $\|\widetilde{v}\| \leqslant \frac{19}{18}|z|$.

For vectors v, w, we will repeatedly use the following estimates:

$$\sum_i |G_{i\mathsf{w}}|^2 = \|G\mathsf{w}\|^2 \leqslant \|G\|^2 \|\mathsf{w}\|^2 \tag{C.18}$$

Consequently, by the Cauchy-Schwarz inequality,

$$\sum_i |\mathsf{v}_i G_{i\mathsf{w}}| \leqslant \|G\|\|\mathsf{v}\|\|\mathsf{w}\| \quad \text{and} \quad \sum_j |G_{\mathsf{v}j} G_{j\mathsf{w}}| \leqslant \|G\mathsf{v}\|\|G\mathsf{w}\| \leqslant \|G\|^2 \|\mathsf{v}\|\|\mathsf{w}\|. \tag{C.19}$$

It follows that

$$|\mathcal{K}| \leqslant 2\sigma_{\max}^2 \Big|\sum_s v_s G_{sw}\Big| \cdot \Big|\sum_k G_{ik} G_{kj}\Big| \leqslant 2\sigma_{\max}^2 \|G\|^3. \tag{C.20}$$

Now we are ready to estimate

$$\begin{aligned}\sum_{i,j} \big|v_i \sigma_{ij}^2 (\partial_{ij} G_{jw})(\partial_{ij} \mathcal{D}_{vw})\big| \leqslant{}& \sum_{i,j} \big|v_i \sigma_{ij}^2 (G_{ij} G_{jw} + G_{jj} G_{iw})(G_{\widetilde{v}i} G_{jw} + G_{\widetilde{v}j} G_{iw})\big| \\ &+ \sum_{i,j} |v_i \sigma_{ij}^2 (G_{ij} G_{jw} + G_{jj} G_{iw}) \mathcal{K}|. \end{aligned} \tag{C.21}$$

For the first term on the right-hand side of (C.21), expanding the product yields four terms. We bound each respectively using (C.18) and (C.19):

$$\sum_{i,j} |v_i \sigma_{ij}^2||G_{ij} G_{jw} G_{\widetilde{v}i} G_{jw}| \leqslant \sigma_{\max}^2 \|G\| \sum_i |v_i G_{\widetilde{v}i}| \sum_j |G_{jw}|^2 \leqslant \sigma_{\max}^2 \|G\|^4 \|\widetilde{v}\|.$$

$$\sum_{i,j} |v_i \sigma_{ij}^2||G_{ij} G_{jw} G_{\widetilde{v}j} G_{iw}| \leqslant \sigma_{\max}^2 \|G\| \sum_i |v_i G_{iw}| \sum_j |G_{jw}||G_{\widetilde{v}j}| \leqslant \sigma_{\max}^2 \|G\|^4 \|\widetilde{v}\|.$$

$$\sum_{i,j} |v_i \sigma_{ij}^2||G_{jj} G_{iw} G_{\widetilde{v}i} G_{jw}| \leqslant \sigma_{\max} \|G\|^2 \sum_i |v_i G_{\widetilde{v}i}| \sum_j |\sigma_{ij} G_{jw}| \leqslant \sigma_{\max} \sqrt{R_{\max}}\, \|G\|^4 \|\widetilde{v}\|.$$

$$\sum_{i,j} |v_i \sigma_{ij}^2||G_{jj} G_{iw} G_{\widetilde{v}j} G_{iw}| \leqslant \sigma_{\max} \|G\|^2 \sum_i |v_i G_{iw}| \sum_j |\sigma_{ij} G_{\widetilde{v}j}| \leqslant \sigma_{\max} \sqrt{R_{\max}}\, \|G\|^4 \|\widetilde{v}\|.$$

Summing these estimates and using the fact that $\sigma_{\max}^2 \leqslant \sigma_{\max}\sqrt{R_{\max}}$, we obtain

$$\sum_{i,j} \big|v_i \sigma_{ij}^2 (G_{ij} G_{jw} + G_{jj} G_{iw})(G_{\widetilde{v}i} G_{jw} + G_{\widetilde{v}j} G_{iw})\big| \leqslant 4\sigma_{\max} \sqrt{R_{\max}}\, \|G\|^4 \|\widetilde{v}\|.$$

For the second term on the right-hand side of (C.21), note that by (C.20)

$$\sum_{i,j}|v_i\sigma_{ij}^2(G_{ij}G_{jw}+G_{jj}G_{iw})\mathcal{K}| \leqslant 2\sigma_{\max}^2\|G\|^3\sum_{i,j}|v_i\sigma_{ij}^2|(|G_{ij}G_{jw}|+|G_{jj}G_{iw}|)$$
$$\leqslant 4\sigma_{\max}^2 R_{\max}\|G\|^5,$$

where in the last inequality, we apply (C.19) to bound

$$\sum_{i,j}|v_i\sigma_{ij}^2||G_{ij}||G_{jw}| \leqslant \sqrt{\sum_{i,j}v_i^2\sigma_{ij}^2|G_{ij}|^2}\sqrt{\sum_{i,j}\sigma_{ij}^2|G_{jw}|^2} \leqslant \sigma_{\max}\sqrt{R_{\max}}\|G\|^2$$

and

$$\sum_{i,j}|v_i\sigma_{ij}^2||G_{jj}G_{iw}| \leqslant \|G\|\sum_i|v_i||G_{iw}|\sum_j\sigma_{ij}^2 \leqslant R_{\max}\|G\|^2.$$

Combining the above estimates, we have

$$\sum_{i,j}|v_i\sigma_{ij}^2(\partial_{ij}G_{jw})(\partial_{ij}\mathcal{D}_{vw})| \leqslant 4\sigma_{\max}\sqrt{R_{\max}}\|G\|^4\|\widetilde{v}\| + 4\sigma_{\max}^2 R_{\max}\|G\|^5.$$

Using (5.6) and our assumption $|z| \geqslant 6\sqrt{R_{\max}}$, we further have

$$\sum_{i,j}\left|v_i\sigma_{ij}^2(\partial_{ij}G_{jw})(\partial_{ij}\mathcal{D}_{vw})\right| \leqslant 4\sigma_{\max}\frac{|z|}{6}\frac{2^4}{|z|^4}\frac{19|z|}{18} + 4\sigma_{\max}\frac{|z|^2}{36}\frac{2^5}{|z|^5} \leqslant \frac{64\sigma_{\max}}{|z|^2}.$$

Consequently, the contribution from the $s=1$ term in (C.17) is bounded by

$$\beta K\sigma_{\max}\mathbb{E}\Big[\sum_{i,j}|v_i|\sigma_{ij}^2|\partial_{ij}G_{jw}|\cdot|\partial_{ij}(\mathcal{D}_{vw}^{2p-1})|\Big] \leqslant 128p\beta K\sigma_{\max}^2\frac{\mathbb{E}|\mathcal{D}_{vw}|^{2p-2}}{|z|^2}.$$

• **Case 3:** $s=2$**.** We estimate

$$\mathbb{E}\Big[\sum_{i,j}|v_i|\sigma_{ij}^2|G_{jw}||\partial_{ij}^2(\mathcal{D}_{vw}^{2p-1})|\Big].$$

The second derivative gives

$$|\partial_{ij}^2\left(\mathcal{D}_{vw}^{2p-1}\right)| \leqslant 2p|\mathcal{D}_{vw}|^{2p-2}|\partial_{ij}^2\mathcal{D}_{vw}| + 4p^2|\mathcal{D}_{vw}|^{2p-3}|\partial_{ij}\mathcal{D}_{vw}|^2.$$

Hence,

$$\mathbb{E}\Big[\sum_{i,j}|v_i|\sigma_{ij}^2|G_{jw}||\partial_{ij}^2(\mathcal{D}_{vw}^{2p-1})|\Big]$$
$$\leqslant 2p\mathbb{E}\Big[\sum_{i,j}|v_i|\sigma_{ij}^2|G_{jw}||\partial_{ij}^2\mathcal{D}_{vw}||\mathcal{D}_{vw}|^{2p-2}\Big] + 4p^2\mathbb{E}\Big[\sum_{i,j}|v_i|\sigma_{ij}^2|G_{jw}||\partial_{ij}\mathcal{D}_{vw}|^2|\mathcal{D}_{vw}|^{2p-3}\Big] \tag{C.22}$$

We first estimate

$$\sum_{i,j}|v_i|\sigma_{ij}^2|G_{jw}||\partial_{ij}^2\mathcal{D}_{vw}|.$$

From (C.16), we have

$$\partial_{ij}^2 \mathcal{D}_{vw} = (1+\delta_{ij})^{-1}\Big[\partial_{ij}(G_{\widetilde{v}i}G_{jw}) + \partial_{ij}(G_{\widetilde{v}j}G_{iw}) - \partial_{ij}\mathcal{K}\Big]$$

and thus

$$\begin{aligned}\sum_{i,j}|v_i\sigma_{ij}^2 G_{jw}(\partial_{ij}^2\mathcal{D}_{vw})| \leqslant& \sum_{i,j}|v_i\sigma_{ij}^2 G_{jw}\partial_{ij}(G_{\widetilde{v}i}G_{jw})| + \sum_{i,j}|v_i\sigma_{ij}^2 G_{jw}\partial_{ij}(G_{\widetilde{v}j}G_{iw})| \\ &+ \sum_{i,j}|v_i\sigma_{ij}^2 G_{jw}\partial_{ij}\mathcal{K}|. \end{aligned} \tag{C.23}$$

The first two terms on the right-hand side of (C.23) can be estimated similarly. We provide the calculation for the first term. By (C.11),

$$\begin{aligned}&\sum_{i,j}|v_i\sigma_{ij}^2 G_{jw}\partial_{ij}(G_{\widetilde{v}i}G_{jw})| \\ &\leqslant \sum_{i,j}\Big[\big|v_i\sigma_{ij}^2 G_{jw}^2(G_{\widetilde{v}i}G_{ij} + G_{\widetilde{v}j}G_{ii})\big| + \big|v_i\sigma_{ij}^2 G_{jw}G_{\widetilde{v}i}(G_{ij}G_{jw} + G_{jj}G_{iw})\big|\Big].\end{aligned}$$

Note that each term on the right-hand side is bounded by $(\sigma_{\max}^2 + \sigma_{\max}\sqrt{R_{\max}})\|G\|^4\|\widetilde{v}\|$ by applying (C.18) and (C.19). For instance,

$$\begin{aligned}&\sum_{i,j}\big|v_i\sigma_{ij}^2 G_{jw}^2(G_{\widetilde{v}i}G_{ij} + G_{\widetilde{v}j}G_{ii})\big| \\ &\leqslant \sigma_{\max}^2\|G\|\sum_i|v_iG_{\widetilde{v}i}|\sum_j|G_{jw}|^2 + \sigma_{\max}\|G\|^2\|\widetilde{v}\|\sum_j|G_{jw}|^2\sum_i|v_i\sigma_{ij}| \\ &\leqslant (\sigma_{\max}^2 + \sigma_{\max}\sqrt{R_{\max}})\|G\|^4\|\widetilde{v}\|.\end{aligned}$$

Continuing from (C.23), we have

$$\sum_{i,j}|v_i\sigma_{ij}^2 G_{jw}(\partial_{ij}^2\mathcal{D}_{vw})| \leqslant 8\sigma_{\max}\sqrt{R_{\max}}\,\|G\|^4\|\widetilde{v}\| + \sum_{i,j}|v_iG_{jw}\partial_{ij}\mathcal{K}|.$$

It remains to estimate $\sum_{i,j}|v_iG_{jw}\partial_{ij}\mathcal{K}|$, where

$$\partial_{ij}\mathcal{K} = 2\sum_{s,k}v_s\sigma_{sk}^2\partial_{ij}(G_{ki}G_{kj}G_{sw}).$$

We first estimate $|\partial_{ij}\mathcal{K}|$. Expanding $\partial_{ij}(G_{ki}G_{kj}G_{sw})$ using the product rule and (C.11) yields several terms, each of which admits a similar treatment. We present the calculation for one representative term. For instance,

$$\begin{aligned}\sum_{s,k}|v_s\sigma_{sk}^2(\partial_{ij}G_{ki})G_{kj}G_{sw}| &\leqslant \sigma_{\max}^2\sum_{s,k}|v_s(G_{ki}G_{jj} + G_{kj}G_{ij})G_{kj}G_{sw}| \\ &\leqslant \sigma_{\max}^2\|G\|\sum_s|v_sG_{sw}|\sum_k(|G_{ki}| + |G_{kj}|)|G_{kj}| \\ &\leqslant 2\sigma_{\max}^2\|G\|^4\end{aligned}$$

where we applied (C.18) and (C.19) in the last inequality. Hence,

$$|\partial_{ij}\mathcal{K}| \leqslant 12\,\sigma_{\max}^2\|G\|^4$$

and by Cauchy-Schwarz and (C.18),

$$\sum_{i,j}|v_i\sigma_{ij}^2 G_{jw}\partial_{ij}\mathcal{K}| \leqslant 12\sigma_{\max}^2\|G\|^4\sum_{i,j}|v_i\sigma_{ij}^2 G_{jw}| \leqslant 12\sigma_{\max}^2\|G\|^4\cdot R_{\max}\|G\| = 12\sigma_{\max}^2 R_{\max}\|G\|^5.$$

Finally, using $|z| \geqslant 6\sqrt{R_{\max}}$ and $\|G\| \leqslant 2/|z|$, we have

$$\begin{aligned}\sum_{i,j}|v_i\sigma_{ij}^2 G_{jw}(\partial_{ij}^2\mathcal{D}_{vw})| &\leqslant 8\sigma_{\max}\sqrt{R_{\max}}\|G\|^4\|\widetilde{v}\| + 12\sigma_{\max}^2 R_{\max}\|G\|^5 \\ &\leqslant 8\sigma_{\max}\frac{|z|}{6}\frac{2^4}{|z|^4}\frac{19|z|}{18} + 12\sigma_{\max}\frac{|z|^3}{6^3}\frac{2^5}{|z|^5} < 25\frac{\sigma_{\max}}{|z|^2}. \qquad \text{(C.24)}\end{aligned}$$

Next, we bound

$$\sum_{i,j}|v_i|\sigma_{ij}^2|G_{jw}||\partial_{ij}\mathcal{D}_{vw}|^2.$$

Continuing from (C.16), we get

$$(\partial_{ij}\mathcal{D}_{vw})^2 = (1+\delta_{ij})^{-2}\Big[(G_{\widetilde{v}i}G_{jw} + G_{\widetilde{v}j}G_{iw})^2 + \mathcal{K}^2 - 2\mathcal{K}(G_{\widetilde{v}i}G_{jw} + G_{\widetilde{v}j}G_{iw})\Big],$$

where

$$\mathcal{K} = 2\sum_{s,k} v_s\sigma_{sk}^2 G_{ki}G_{kj}G_{sw}.$$

Plugging this expansion into $\sum_{i,j}|v_i G_{jw}(\partial_{ij}\mathcal{D}_{vw})^2|$, we first observe from (C.20) that

$$\sum_{i,j}|v_i\sigma_{ij}^2 G_{jw}||\mathcal{K}|^2 \leqslant (2\sigma_{\max}^2\|G\|^3)^2\sum_{i,j}|v_i\sigma_{ij}^2||G_{jw}| \leqslant 4\sigma_{\max}^4\|G\|^6\cdot R_{\max}\|G\|.$$

Next, using (C.18) and (C.19), we also have

$$\begin{aligned}&2\sum_{i,j}|v_i\sigma_{ij}^2 G_{jw}(G_{\widetilde{v}i}G_{jw} + G_{\widetilde{v}j}G_{iw})||\mathcal{K}| \\ &\leqslant 4\sigma_{\max}^4\|G\|^3\left(\sum_{i,j}|v_i G_{\widetilde{v}i}|\cdot|G_{jw}|^2 + \sum_i|v_i G_{iw}|\sum_j|G_{jw}G_{\widetilde{v}j}|\right) \\ &\leqslant 8\sigma_{\max}^4\|G\|^6\|\widetilde{v}\|.\end{aligned}$$

Finally, we claim that

$$\sum_{i,j}|v_i\sigma_{ij}^2 G_{jw}(G_{\widetilde{v}i}G_{jw} + G_{\widetilde{v}j}G_{iw})^2| \leqslant 4\sigma_{\max}^2\|G\|^5\|\widetilde{v}\|^2.$$

To see this, we expand the square in the summation and show that each term is bounded by $\|G\|^5\|\widetilde{v}\|^2$, using a similar calculation with (C.18) and (C.19). For instance,

$$\sum_{i,j}|v_i G_{jw}G_{\widetilde{v}i}^2 G_{jw}^2| \leqslant \|G\|^2\|\widetilde{v}\|\sum_i|v_i G_{\widetilde{v}i}|\sum_j|G_{jw}|^2 \leqslant \|G\|^5\|\widetilde{v}\|^2,$$

and the remaining terms are estimated in the same manner.

Hence, together with $\|\widetilde{v}\| \leqslant \frac{19}{18}|z|$, we get

$$\sum_{i,j} |v_i \sigma_{ij}^2 G_{jw} (\partial_{ij} \mathcal{D}_{vw})^2| \leqslant 4\sigma_{\max}^4 R_{\max} \|G\|^7 + 8\sigma_{\max}^4 \|G\|^6 \|\widetilde{v}\| + 4\sigma_{\max}^2 \|G\|^5 \|\widetilde{v}\|^2$$
$$\leqslant \frac{2^3 \sigma_{\max}^2}{|z|^3} (4 \frac{|z|^4}{6^4} \frac{2^4}{|z|^4} + 8 \frac{|z|^2}{6^2} \frac{2^3}{|z|^3} \frac{19|z|}{18} + 4 \frac{2^2}{|z|^2} \frac{19^2 |z|^2}{18^2}) < 175 \frac{\sigma_{\max}^2}{|z|^3}. \tag{C.25}$$

We continue from (C.22) to get

$$\mathbb{E}\Big[\sum_{i,j} |v_i| \sigma_{ij}^2 |G_{jw}| |\partial_{ij}^2 (\mathcal{D}_{vw}^{2p-1})|\Big] \leqslant 50 p \sigma_{\max} \frac{\mathbb{E}|\mathcal{D}_{vw}|^{2p-2}}{|z|^2} + 700 p^2 \sigma_{\max}^2 \frac{\mathbb{E}|\mathcal{D}_{vw}|^{2p-3}}{|z|^3}.$$

Thus, the contribution from the $s = 2$ term in (C.17) is bounded by

$$\beta K \sigma_{\max} \mathbb{E}\Big[\sum_{i,j} |v_i| \sigma_{ij}^2 |G_{jw}| |\partial_{ij}^2 (\mathcal{D}_{vw}^{2p-1})|\Big]$$
$$\leqslant 50 p \beta K \sigma_{\max}^2 \frac{\mathbb{E}|D_{vw}|^{2p-2}}{|z|^2} + 700 p^2 \beta K \sigma_{\max}^3 \frac{\mathbb{E}|\mathcal{D}_{vw}|^{2p-3}}{|z|^3}.$$

Combining all the above estimates completes the proof of Lemma C.3 (ii).

A third-order bound for the remainder term. Before turning to the remainder term, we record an additional derivative estimate required by the cumulant remainder formula in Lemma C.1. This estimate serves as a third-order analogue of those used in bounding X_2 above, corresponding to the third-order derivatives of

$$G_{jw} \mathcal{D}_{vw}^{2p-1}.$$

The argument follows the structure of the $s = 0, 1, 2$ cases in the proof of Lemma C.3 (ii). We isolate this estimate here for use in the remainder analysis.

Recall from the truncation parameter $\mathtt{L}_n$ from (C.3). For $1 \leqslant i, j \leqslant n$ and $|x| \leqslant \mathtt{L}_n$, let

$$E^{ij(x)} := E + (x - E_{ij}) \Delta^{ij}, \qquad G^{ij(x)}(z) := (zI - E^{ij(x)})^{-1}.$$

Define

$$L^{ij(x)} := \operatorname{diag}(l_1^{ij(x)}, \ldots, l_n^{ij(x)}) \qquad \text{with} \quad l_s^{ij(x)} := \sum_{k=1}^{n} \sigma_{sk}^2 G_{kk}^{ij(x)},$$

and

$$\mathcal{D}^{ij(x)} := (E^{ij(x)} - L^{ij(x)}) G^{ij(x)}(z).$$

We use the notation $\mathcal{D}_{vw}^{ij(x)} = v^\top \mathcal{D}^{ij(x)} w$ and $G_{vw}^{ij(x)} = v^\top G^{ij(x)} w$. Set

$$f_{ij}(E^{ij(x)}) := G_{jw}^{ij(x)} \left(\mathcal{D}_{vw}^{ij(x)}\right)^{2p-1}.$$

Recall from (C.4) the event

$$\Omega_n := \left\{ \|E\| \leqslant 3.5\sqrt{R_{\max}}, \quad \max_{a,b} |E_{ab}| \leqslant 2\mathtt{L}_n \right\}.$$

The goal is to prove the following estimate on the event Ω_n (the support of $\chi(E)$):

$$\sum_{i,j} |v_i| \sigma_{ij}^2 \sup_{|x| \leqslant \mathtt{L}_n} \left| \partial_{ij}^3 f_{ij}(E^{ij(x)}) \right| \leqslant C \sum_{m=1}^{4} p^{m-1} \frac{\sigma_{\max}^{m-1}}{|z|^{m+1}} \left(|\mathcal{D}_{vw}| + 22 \frac{\mathtt{L}_n}{|z|} \right)^{2p-m} \tag{C.26}$$

for an absolute constant $C > 0$.

We start with some preliminary estimates. On the event Ω_n, we have, for $|x| \leqslant \mathtt{L}_n$,

$$\|E^{ij(x)}\| \leqslant \|E\| + (|x| + |E_{ij}|)\|\Delta_{ij}\| \leqslant 3.5\sqrt{R_{\max}} + 3\mathtt{L}_n \leqslant 4\sqrt{R_{\max}} \tag{C.27}$$

provided the constant C_0 in Assumption 1.2 is chosen sufficiently large. Hence, for $|z| \geqslant 6\sqrt{R_{\max}}$, by the Neumann expansion, we have

$$\|G^{ij(x)}(z)\| \leqslant \frac{3}{|z|}. \tag{C.28}$$

Next, we show that

$$\sup_{|x| \leqslant \mathtt{L}_n} |\mathcal{D}_{vw}^{ij(x)} - \mathcal{D}_{vw}| \leqslant 22 \frac{\mathtt{L}_n}{|z|}.$$

To see this, for $|x| \leqslant \mathtt{L}_n$, we bound

$$|\mathcal{D}_{vw}^{ij(x)} - \mathcal{D}_{vw}| = |v^\top (\mathcal{D}^{ij(x)} - \mathcal{D}) w| \leqslant \|\mathcal{D}^{ij(x)} - \mathcal{D}\|.$$

From the definitions of $\mathcal{D}^{ij(x)}$ and $\mathcal{D}$, we further have

$$\begin{aligned}
|\mathcal{D}_{vw}^{ij(x)} - \mathcal{D}_{vw}| &\leqslant \|(E^{ij(x)} - L^{ij(x)}) G^{ij(x)}(z) - (E - L) G(z)\| \\
&= \|[(E^{ij(x)} - E) - (L^{ij(x)} - L)] G^{ij(x)}(z) + (E - L)(G^{ij(x)}(z) - G(z))\| \\
&\leqslant (\|E^{ij(x)} - E\| + \|L^{ij(x)} - L\|) \|G^{ij(x)}(z)\| + (\|E\| + \|L\|) \|G^{ij(x)}(z) - G(z)\|.
\end{aligned}$$

For simplicity, denote $P^{ij} := E^{ij(x)} - E = (x - E_{ij})\Delta^{ij}$. Similar to (C.27), we have

$$\|E^{ij(x)} - E\| = \|P^{ij}\| \leqslant (|x| + |E_{ij}|)\|\Delta_{ij}\| \leqslant 3\mathtt{L}_n.$$

Next, by the resolvent identity, we have

$$G^{ij(x)}(z) - G(z) = G^{ij(x)}(z) P^{ij} G(z)$$

Thus, by (C.28) and (5.6),

$$\|G^{ij(x)}(z) - G(z)\| = \|G^{ij(x)}(z)\| \cdot \|P^{ij}\| \cdot \|G(z)\| \leqslant 18 \frac{\mathtt{L}_n}{|z|^2}.$$

Then

$$\|L^{ij(x)} - L\| = \max_s |l_s^{ij(x)} - l_s| = \max_s \sum_{k=1}^n \sigma_{sk}^2 |G_{kk}^{ij(x)}(z) - G_{kk}(z)|$$
$$\leqslant R_{\max}\|G^{ij(x)}(z) - G(z)\| \leqslant 18\mathtt{L}_n \frac{R_{\max}}{|z|^2} \leqslant \frac{1}{2}\mathtt{L}_n$$

by the assumption $|z| \geqslant 6\sqrt{R_{\max}}$. Similarly,

$$\|L\| = \max_s |l_s| \leqslant \max_s \sum_{k=1}^n \sigma_{sk}^2 |G_{kk}(z)| \leqslant R_{\max}\|G(z)\| \leqslant 2\frac{R_{\max}}{|z|}.$$

It follows that

$$\|E\| + \|L\| \leqslant 3.5\sqrt{R_{\max}} + 2\frac{R_{\max}}{|z|} \leqslant \left(\frac{7}{12} + \frac{1}{18}\right)|z|.$$

Combining the above estimates, we conclude that

$$|\mathcal{D}_{vw}^{ij(x)} - \mathcal{D}_{vw}| \leqslant 10.5\frac{\mathtt{L}_n}{|z|} + 18\left(\frac{7}{12} + \frac{1}{18}\right)\frac{\mathtt{L}_n}{|z|} = 22\frac{\mathtt{L}_n}{|z|}.$$

Thus

$$\sup_{|x| \leqslant \mathtt{L}_n} |\mathcal{D}_{vw}^{ij}(x)| \leqslant |\mathcal{D}_{vw}| + 22\frac{\mathtt{L}_n}{|z|}. \tag{C.29}$$

Now, we proceed to the proof of (C.26). By the Leibniz rule,

$$\partial_{ij}^3 f_{ij}(E^{ij(x)}) = \sum_{s=0}^3 \binom{3}{s}\left(\partial_{ij}^{3-s} G_{jw}^{ij(x)}\right)\left(\partial_{ij}^s (\mathcal{D}_{vw}^{ij(x)})^{2p-1}\right).$$

We estimate the four cases $s = 0, 1, 2, 3$ separately. The cases $s = 0, 1, 2$ follow similar arguments to those used for X_2 in (C.17).

First, for $s = 0$, using

$$\partial_{ij}^3 G^{ij(x)} = 3! G^{ij(x)}(\Delta^{ij} G^{ij(x)})^3$$

and the same estimate used for the $s = 0$ term in the X_2 bound, we get

$$\sum_{i,j} |v_i|\sigma_{ij}^2 |\partial_{ij}^3 G_{jw}^{ij(x)}| \leqslant 6R_{\max}\|G^{ij(x)}\|^4 \leqslant 6R_{\max}\frac{3^4}{|z|^4} \leqslant \frac{27}{2|z|^2}.$$

Therefore,

$$\sum_{i,j} |v_i|\sigma_{ij}^2 |\partial_{ij}^3 G_{jw}^{ij(x)}| \cdot |(\mathcal{D}_{vw}^{ij(x)})^{2p-1}| \leqslant \frac{27}{2|z|^2}|\mathcal{D}_{vw}^{ij(x)}|^{2p-1}.$$

Next, for $s = 1$, since

$$|\partial_{ij}(\mathcal{D}_{vw}^{ij(x)})^{2p-1}| \leqslant 2p|\mathcal{D}_{vw}^{ij(x)}|^{2p-2}|\partial_{ij}\mathcal{D}_{vw}^{ij(x)}|,$$

we aim to bound

$$\sum_{i,j} |v_i|\sigma_{ij}^2|\partial_{ij}^2 G_{jw}^{ij}(x)||\partial_{ij}\mathcal{D}_{vw}^{ij(x)}|.$$

This is proved exactly as the $s = 1$ estimate for X_2, but with an extra factor of G. We sketch the key steps here. Expand

$$\partial_{ij}\mathcal{D}_{vw}^{ij(x)} = (1+\delta_{ij})^{-1}\left(G_{\widetilde{v}_x i}^{ij(x)}G_{jw}^{ij(x)} + G_{\widetilde{v}_x j}^{ij(x)}G_{iw}^{ij(x)} - \mathcal{K}_x\right),$$

where $\widetilde{v}_x^\top := v^\top(zI - L^{ij(x)})$ and

$$\mathcal{K}_x := 2\sum_{s,k} v_s\sigma_{sk}^2 G_{ki}^{ij(x)}G_{kj}^{ij(x)}G_{sw}^{ij(x)}.$$

Since

$$\|L^{ij(x)}\| \leqslant \max_s \sum_{k=1}^n \sigma_{sk}^2|G^{ij(x)}(z)| \leqslant R_{\max}\|G^{ij(x)}(z)\| \leqslant 3\frac{R_{\max}}{|z|} \leqslant \frac{1}{12}|z|,$$

we have $\|\widetilde{v}_x^\top\| \leqslant \frac{13}{12}|z|$. Similar to (C.20), we obtain

$$|\mathcal{K}_x| \leqslant 2\sigma_{\max}^2\|G^{ij(x)}\|^3.$$

For the second derivative of the resolvent

$$\partial_{ij}^2 G_{jw}^{ij(x)} = 2(1+\delta_{ij})^{-2}\Big[\big((G_{ij}^{ij(x)})^2 + G_{ii}^{ij(x)}G_{jj}^{ij(x)}\big)G_{jw}^{ij(x)} + 2G_{ij}^{ij(x)}G_{jj}^{ij(x)}G_{iw}^{ij(x)}\Big],$$

we apply the crude upper bound

$$|\partial_{ij}^2 G_{jw}^{ij(x)}| \leqslant 4\|G^{ij(x)}\|^2\big(|G_{jw}^{ij(x)}| + |G_{iw}^{ij(x)}|\big).$$

Multiplying this by the expansion of $\partial_{ij}\mathcal{D}_{vw}^{ij(x)}$ and applying the identical Cauchy-Schwarz estimates used for X_2, we obtain

$$\begin{aligned}
\sum_{i,j} |v_i|\sigma_{ij}^2|\partial_{ij}^2 G_{jw}^{ij(x)}||\partial_{ij}\mathcal{D}_{vw}^{ij(x)}| &\leqslant 16\sigma_{\max}\sqrt{R_{\max}}\|G^{ij(x)}\|^5\|\widetilde{v}_x\| + 16\sigma_{\max}^2 R_{\max}\|G^{ij(x)}\|^6 \\
&\leqslant 16\sigma_{\max}\left(\frac{|z|}{6}\right)\frac{3^5}{|z|^5}\left(\frac{13|z|}{12}\right) + 16\sigma_{\max}\left(\frac{|z|^3}{216}\right)\frac{3^6}{|z|^6} \\
&\leqslant 702\frac{\sigma_{\max}}{|z|^3} + 54\frac{\sigma_{\max}}{|z|^3} = 756\frac{\sigma_{\max}}{|z|^3}.
\end{aligned}$$

Hence the $s = 1$ contribution is bounded by

$$1512p\frac{\sigma_{\max}}{|z|^3}|\mathcal{D}_{vw}^{ij(x)}|^{2p-2}.$$

For $s = 2$, we use

$$|\partial_{ij}^2\left((\mathcal{D}_{vw}^{ij(x)})^{2p-1}\right)| \leqslant 2p|\mathcal{D}_{vw}^{ij(x)}|^{2p-2}|\partial_{ij}^2\mathcal{D}_{vw}^{ij(x)}| + 4p^2|\mathcal{D}_{vw}^{ij(x)}|^{2p-3}|\partial_{ij}\mathcal{D}_{vw}^{ij(x)}|^2.$$

The first estimate needed is

$$\sum_{i,j} |v_i|\sigma_{ij}^2 |\partial_{ij} G_{jw}^{ij}(x)||\partial_{ij}^2 \mathcal{D}_{vw}^{ij}(x)|.$$

Using $|\partial_{ij} G_{jw}^{ij(x)}| \leqslant 2\|G^{ij(x)}\|(|G_{jw}^{ij(x)}| + |G_{iw}^{ij(x)}|)$ and multiplying by the bounds established in (C.24), we pull out an extra factor of $2\|G^{ij(x)}\|$ and double the sum to account for $|G_{iw}|$:

$$\begin{aligned}\sum_{i,j} |v_i|\sigma_{ij}^2 |\partial_{ij} G_{jw}^{ij(x)}||\partial_{ij}^2 \mathcal{D}_{vw}^{ij(x)}| &\leqslant 32\sigma_{\max}\sqrt{R_{\max}}\|G^{ij(x)}\|^5\|\widetilde{v}_x\| + 48\sigma_{\max}^2 R_{\max}\|G^{ij(x)}\|^6 \\ &\leqslant 1404\frac{\sigma_{\max}}{|z|^3} + 162\frac{\sigma_{\max}}{|z|^3} = 1566\frac{\sigma_{\max}}{|z|^3}.\end{aligned}$$

We also require the estimate for

$$\sum_{i,j} |v_i|\sigma_{ij}^2 |\partial_{ij} G_{jw}^{ij}(x)||\partial_{ij} \mathcal{D}_{vw}^{ij(x)}|^2.$$

Again, using $|\partial_{ij} G_{jw}^{ij(x)}| \leqslant 2\|G^{ij(x)}\|(|G_{jw}^{ij(x)}| + |G_{iw}^{ij(x)}|)$ to the estimates from (C.25) yields:

$$\begin{aligned}\sum_{i,j} |v_i|\sigma_{ij}^2 |\partial_{ij} G_{jw}^{ij(x)}||\partial_{ij} \mathcal{D}_{vw}^{ij(x)}|^2 &\leqslant 24\sigma_{\max}^2\|G^{ij(x)}\|^6\|\widetilde{v}_x\|^2 + 16\sigma_{\max}^4 R_{\max}\|G^{ij(x)}\|^8 \\ &\leqslant 24\sigma_{\max}^2\frac{3^6}{|z|^6}\left(\frac{169|z|^2}{144}\right) + 16\sigma_{\max}^2\left(\frac{|z|^4}{1296}\right)\frac{3^8}{|z|^8} \\ &\leqslant C\frac{\sigma_{\max}^2}{|z|^4}.\end{aligned}$$

Therefore, the total $s = 2$ contribution is bounded by

$$Cp\frac{\sigma_{\max}}{|z|^3}|\mathcal{D}_{vw}^{ij(x)}|^{2p-2} + Cp^2\frac{\sigma_{\max}^2}{|z|^4}|\mathcal{D}_{vw}^{ij(x)}|^{2p-3}.$$

Finally, for $s = 3$, applying the generalized chain rule (Faà di Bruno's formula [48, p. 137]) to the third derivative of $(\mathcal{D}_{vw}^{ij(x)})^{2p-1}$ yields:

$$\begin{aligned}\left|\partial_{ij}^3\left((\mathcal{D}_{vw}^{ij(x)})^{2p-1}\right)\right| \leqslant\ & 2p|\mathcal{D}_{vw}^{ij(x)}|^{2p-2}|\partial_{ij}^3\mathcal{D}_{vw}^{ij(x)}| \\ &+ 12p^2|\mathcal{D}_{vw}^{ij(x)}|^{2p-3}|\partial_{ij}\mathcal{D}_{vw}^{ij(x)}||\partial_{ij}^2\mathcal{D}_{vw}^{ij(x)}| \\ &+ 8p^3|\mathcal{D}_{vw}^{ij(x)}|^{2p-4}|\partial_{ij}\mathcal{D}_{vw}^{ij(x)}|^3.\end{aligned}$$

Bounding this requires three separate estimates:

$$\sum_{i,j} |v_i|\sigma_{ij}^2 |G_{jw}^{ij(x)}||\partial_{ij}^3\mathcal{D}_{vw}^{ij(x)}| \leqslant C_1\frac{\sigma_{\max}}{|z|^3}, \tag{C.30}$$

$$\sum_{i,j} |v_i|\sigma_{ij}^2 |G_{jw}^{ij(x)}||\partial_{ij}\mathcal{D}_{vw}^{ij(x)}||\partial_{ij}^2\mathcal{D}_{vw}^{ij(x)}| \leqslant C_2\frac{\sigma_{\max}^2}{|z|^4}, \tag{C.31}$$

$$\sum_{i,j} |v_i|\sigma_{ij}^2 |G_{jw}^{ij(x)}||\partial_{ij}\mathcal{D}_{vw}^{ij(x)}|^3 \leqslant C_3\frac{\sigma_{\max}^3}{|z|^5}, \tag{C.32}$$

where $C_1, C_2, C_3 > 0$ are absolute constants (which are all bounded 10^5).

We briefly sketch the proof of these estimates. For (C.30), from the identity $\mathcal{D} = (zI - L)G - I$, we have

$$\partial_{ij}^3 \mathcal{D} = (zI - L)\partial_{ij}^3 G - 3(\partial_{ij} L)\partial_{ij}^2 G - 3(\partial_{ij}^2 L)\partial_{ij} G - (\partial_{ij}^3 L)G.$$

Then we multiply this expansion with $|v_i|\sigma_{ij}^2|G_{jw}^{ij(x)}|$ and sum over i, j. The first term is bounded by expanding $\partial_{ij}^3 G = 3! G(\Delta^{ij} G)^3$. The factor $|G_{jw}^{ij(x)}|$ combined with the variance weight generates one factor of $\sigma_{\max}$, while the remaining Green functions are controlled by (C.18) and (C.19). This yields the $C_1 \sigma_{\max}|z|^{-3}$ bound. The terms containing $\partial_{ij}^a L$ are handled by expanding

$$\partial_{ij}^a l_s = \sum_k \sigma_{sk}^2 \partial_{ij}^a G_{kk}^{ij(x)},$$

together with

$$|\partial_{ij}^a G_{kk}^{ij(x)}| \leqslant \widetilde{C}_a \|G^{ij(x)}\|^{a-1} \sum_{s \in \{i,j\}} |G_{ks}^{ij(x)}|^2.$$

and the Cauchy-Schwarz inequality. This proves (C.30).

The estimates (C.31) and (C.32) follow by combining the already proved bounds for $\partial_{ij}\mathcal{D}_{vw}^{ij(x)}$ and $\partial_{ij}^2 \mathcal{D}_{vw}^{ij(x)}$, and the factor $|G_{jw}^{ij(x)}|$. Note that each additional $\partial_{ij}\mathcal{D}$ contributes one factor $\sigma_{\max}/|z|$ in the estimate, exactly as in the $s = 2$ case of the X_2 estimate.

Consequently, the $s = 3$ contribution is bounded by

$$2p\, C_1 \frac{\sigma_{\max}}{|z|^3}|\mathcal{D}_{vw}^{ij(x)}|^{2p-2} + 12p^2\, C_2 \frac{\sigma_{\max}^2}{|z|^4}|\mathcal{D}_{vw}^{ij(x)}|^{2p-3} + 8p^3\, C_3 \frac{\sigma_{\max}^3}{|z|^5}|\mathcal{D}_{vw}^{ij(x)}|^{2p-4}.$$

Combining the four cases $s = 0, 1, 2, 3$ yields

$$\sum_{i,j} |v_i|\sigma_{ij}^2 \sup_{|x| \leqslant \mathtt{L}_n} \left|\partial_{ij}^3 \left(G_{jw}^{ij(x)} (\mathcal{D}_{vw}^{ij(x)})^{2p-1}\right)\right| \leqslant C \sum_{m=1}^{4} p^{m-1} \frac{\sigma_{\max}^{m-1}}{|z|^{m+1}} |\mathcal{D}_{vw}^{ij(x)}|^{2p-m}. \quad \text{(C.33)}$$

Using $|\mathcal{D}_{vw}^{ij(x)}| \leqslant |\mathcal{D}_{vw}| + 22\mathtt{L}_n/|z|$ from (C.29), we finally obtain

$$\sum_{i,j} |v_i|\sigma_{ij}^2 \sup_{|x| \leqslant \mathtt{L}_n} \left|\partial_{ij}^3 \left(G_{jw}^{ij(x)} (\mathcal{D}_{vw}^{ij(x)})^{2p-1}\right)\right| \leqslant C \sum_{m=1}^{4} p^{m-1} \frac{\sigma_{\max}^{m-1}}{|z|^{m+1}} \left(|\mathcal{D}_{vw}| + 22\frac{\mathtt{L}_n}{|z|}\right)^{2p-m}.$$

This proves (C.26).

Proof of Lemma C.3 (iii). We estimate the remainder term in the cumulant expansion with $l = 2$. Recall from (C.13) that

$$\mathtt{R}_3 = \sum_{i,j} v_i \mathtt{R}_3^{ij}.$$

By Lemma C.1,

$$
\begin{aligned}
|\mathtt{R}_3| &\leqslant \sum_{i,j} |v_i \mathtt{R}_3^{ij}| \\
&\leqslant 16 \sum_{i,j} |v_i| \mathbb{E}\left[|E_{ij}|^4\right] \mathbb{E}\left[\sup_{|x|\leqslant t} |\partial_{ij}^3 f_{ij}(E^{ij} + x\Delta^{ij})|\right] && \text{(Term 1)} \\
&+ 16 \sum_{i,j} |v_i| \left(\mathbb{E}\left[\sup_{|x|\leqslant |E_{ij}|} |\partial_{ij}^3 f_{ij}(E^{ij(x)})|^2\right]\right)^{1/2} \left(\mathbb{E}\left[|E_{ij}|^8 \mathbf{1}_{|E_{ij}|>t}\right]\right)^{1/2} && \text{(Term 2)}
\end{aligned}
$$

We choose $t = \mathtt{L}_n$. In the general sub-Gaussian regime, $\mathtt{L}_n = 5\sqrt{D+6}K\sigma_{\max}\log n$, , while in the bounded sparse-entry regime, it is equal to zero. Here, as before, the cutoff χ is suppressed from the notation. Also, as explained in the beginning of Section C.1, terms with derivatives applied to χ are negligible.

Let us consider the general sub-Gaussian regime. We first estimate Term 1. Set

$$\theta := \beta K \sigma_{\max}.$$

By Assumption 1.1,

$$\mathbb{E}|E_{ij}|^4 \leqslant \beta^2 K^2 \sigma_{\max}^2 \sigma_{ij}^2 = \theta^2 \sigma_{ij}^2.$$

Hence, by (C.26), the contribution from Term 1 is bounded by

$$
\begin{aligned}
16 \sum_{i,j} &|v_i| \mathbb{E}\left[|E_{ij}|^4\right] \mathbb{E}\left[\sup_{|x|\leqslant \mathtt{L}_n} |\partial_{ij}^3 f_{ij}(E^{ij} + x\Delta^{ij})|\right] \\
&\leqslant 16\theta^2 \sum_{i,j} |v_i| \sigma_{ij}^2 \mathbb{E}\left[\sup_{|x|\leqslant \mathtt{L}_n} \left|\partial_{ij}^3 f_{ij}(E^{ij(x)})\right|\right] \\
&\leqslant C\theta^2 \sum_{m=1}^{4} p^{m-1} \frac{\sigma_{\max}^{m-1}}{|z|^{m+1}} \mathbb{E}\left[\left(|\mathcal{D}_{vw}| + 22\frac{\mathtt{L}_n}{|z|}\right)^{2p-m}\right].
\end{aligned}
$$

It remains to estimate Term 2. By the sub-Gaussian tail assumption, we have

$$\left(\mathbb{E}\left[|E_{ij}|^8 \mathbf{1}_{\{|E_{ij}|>\mathtt{L}_n\}}\right]\right)^{1/2} \leqslant \theta^2 \sigma_{ij}^2 \exp(-10(D+6)\log^2 n). \tag{C.34}$$

To see this, note that the case $\sigma_{ij} = 0$ is trivial. For $\sigma_{ij} > 0$, write $E_{ij} = \sigma_{ij}\xi_{ij}$ with $\|\xi_{ij}\|_{\psi_2} \leqslant K$. A standard sub-Gaussian tail integration yields

$$
\begin{aligned}
\mathbb{E}\left[|E_{ij}|^8 \mathbf{1}_{\{|E_{ij}|>\mathtt{L}_n\}}\right] &\leqslant C(K\sigma_{ij})^8 \left(1 + \frac{\mathtt{L}_n}{K\sigma_{ij}}\right)^8 \exp\left(-\frac{\mathtt{L}_n^2}{K^2\sigma_{ij}^2}\right) \\
&= C\left(K\sigma_{ij} + \mathtt{L}_n\right)^8 \exp\left(-\frac{\mathtt{L}_n^2}{K^2\sigma_{ij}^2}\right).
\end{aligned} \tag{C.35}
$$

Because $\sigma_{ij} \leqslant \sigma_{\max}$ and $\mathtt{L}_n = 5\sqrt{D+6}K\sigma_{\max}\log n$, we bound

$$(K\sigma_{ij} + \mathtt{L}_n)^8 \leqslant CD^4 K^8 \sigma_{\max}^8 (\log n)^8.$$

Moreover, we write

$$\exp\left(-\frac{\mathrm{L}_n^2}{K^2\sigma_{ij}^2}\right) = \frac{\sigma_{ij}^4}{\sigma_{\max}^4}\left[\frac{\sigma_{\max}^4}{\sigma_{ij}^4}\exp\left(-25(D+6)\frac{\sigma_{\max}^2}{\sigma_{ij}^2}\log^2 n\right)\right]$$
$$\leqslant \frac{\sigma_{ij}^4}{\sigma_{\max}^4}\exp\big(-25(D+6)\log^2 n\big),$$

where we used the fact that the function $y \mapsto y^2\exp(-cy)$ is strictly decreasing for $y \geqslant 1$ and large $c > 0$, and then $y^2\exp(-cy) \leqslant \exp(-c)$.

Thus, the right-hand side of (C.35) is bounded by $CD^4K^8\sigma_{\max}^4\sigma_{ij}^4(\log n)^8\exp(-25(D+6)\log^2 n)$. To get the desired bound (C.34), we only need this quantity

$$D^4K^8\sigma_{\max}^4\sigma_{ij}^4(\log n)^8\exp(-25(D+6)\log^2 n) \leqslant \theta^4\sigma_{ij}^4\exp(-20(D+6)\log^2 n),$$

where $\theta = \beta K\sigma_{\max}$. Since $\beta \geqslant 1$, we require

$$\exp(25(D+6)\log^2 n) \geqslant D^4K^4\exp(20(D+6)\log^2 n). \tag{C.36}$$

We now use Assumption 1.2. Since $R_{\max} \leqslant n\sigma_{\max}^2$ and $\sqrt{R_{\max}} \geqslant C_0(D+8)K\sigma_{\max}\log n$, we have

$$K \leqslant \frac{\sqrt{n}}{C_0(D+8)\log n}.$$

Hence, (C.36) holds and (C.34) is proved.

Now, Term 2 is bounded by

$$16\sum_{i,j}|v_i|\left(\mathbb{E}\left[\sup_{|x|\leqslant|E_{ij}|}|\partial_{ij}^3 f_{ij}(E^{ij(x)})|^2\right]\right)^{1/2}\left(\mathbb{E}\left[|E_{ij}|^8\mathbf{1}_{|E_{ij}|>t}\right]\right)^{1/2}$$
$$\leqslant C\exp(-10(D+6)\log^2 n)\theta^2\sum_{i,j}\left(\mathbb{E}\left[\sup_{|x|\leqslant|E_{ij}|}\Big||v_i|\sigma_{ij}^2\partial_{ij}^3 f_{ij}(E^{ij(x)})\Big|^2\right]\right)^{1/2}.$$

Recall that the cutoff χ is suppressed from the notation. On the support Ω_n of χ, (C.33) holds. Moreover, by (C.27) and (C.28), we also have this crude bound

$$\|\mathcal{D}^{ij(x)}\| = \|(E^{ij(x)} - L^{ij(x)})G^{ij(x)}\| \leqslant C.$$

Therefore,

$$|\mathcal{D}_{vw}^{ij(x)}|^{2p-m} \leqslant C^p, \qquad 1 \leqslant m \leqslant 4.$$

It follows from (C.33) that

$$\left(\mathbb{E}\left[\sup_{|x|\leqslant|E_{ij}|}\Big||v_i|\sigma_{ij}^2\partial_{ij}^3 f_{ij}(E^{ij(x)})\Big|^2\right]\right)^{1/2} \leqslant C^p\sum_{m=1}^{4}p^{m-1}\frac{\sigma_{\max}^{m-1}}{|z|^{m+1}}.$$

Since $\sigma_{\max} \leqslant \sqrt{R_{\max}} \leqslant |z|/6$, we simplify this bound to

$$\left(\mathbb{E}\left[\sup_{|x|\leqslant|E_{ij}|}\Big||v_i|\sigma_{ij}^2\partial_{ij}^3 f_{ij}(E^{ij(x)})\Big|^2\right]\right)^{1/2} \leqslant C^p p^3\frac{1}{|z|^2}.$$

Combining the above bound and (C.34), the total contribution of Term 2 is bounded by

$$\begin{aligned}&16\sum_{i,j}|v_i|\left(\mathbb{E}\left[\sup_{|x|\leqslant|E_{ij}|}|\partial_{ij}^3 f_{ij}(E^{ij(x)})|^2\right]\right)^{1/2}\left(\mathbb{E}\left[|E_{ij}|^8\mathbf{1}_{|E_{ij}|>t}\right]\right)^{1/2}\\&\leqslant C^p p^3\exp\big(-10(D+6)\log^2 n\big)\frac{\theta^2}{|z|^2}.\end{aligned}$$

Finally, we arrive at

$$\begin{aligned}|\mathtt{R}_3|&\leqslant C\theta^2\sum_{m=1}^4 p^{m-1}\frac{\sigma_{\max}^{m-1}}{|z|^{m+1}}\mathbb{E}\left[\left(|\mathcal{D}_{vw}|+22\frac{\mathtt{L}_n}{|z|}\right)^{2p-m}\right]+C^p p^3\exp\big(-10(D+6)\log^2 n\big)\frac{\theta^2}{|z|^2}\\&=C(K\beta)^2\sum_{m=1}^4 p^{m-1}\frac{\sigma_{\max}^{m+1}}{|z|^{m+1}}\mathbb{E}\left[\left(|\mathcal{D}_{vw}|+22\frac{\mathtt{L}_n}{|z|}\right)^{2p-m}\right]+C^p p^3\exp\big(-10(D+6)\log^2 n\big)\frac{\theta^2}{|z|^2}.\end{aligned}$$

The bounded sparse-entry regime follows the same approach. Indeed, the Term 2 in the bound for $|\mathtt{R}_3|$ is equal to zero because $\max_{i,j}|E_{ij}|\leqslant\mathtt{L}_n$ almost surely. This simplifies the derivation. We skip the details. The proof of Lemma C.3 is now complete.

D Proof of Proposition 8.2

Recall that

$$H=\operatorname{diag}(h_1,\dots,h_n),\qquad h_i=\sum_{j=1}^n\sigma_{ij}^2\phi_j;$$
$$L=\operatorname{diag}(l_1,\dots,l_n),\qquad l_i=\sum_{j=1}^n\sigma_{ij}^2 G_{jj}.$$

Set

$$\mathcal{D}:=(E-L)G,\qquad \mathcal{D}_{ii}:=e_i^\top\mathcal{D}e_i.$$

Note that $\max_i|\phi_i(z)|\leqslant\frac{3}{2|z|}$ by Lemma 5.1. Define

$$d_i:=\phi_i-G_{ii},\qquad d:=(d_1,\dots,d_n)^\top$$

We first bound $\|d\|_\infty$. From $(zI-E)G=I$, we have $EG=zG-I$. Therefore, taking the (i,i)-entry of $\mathcal{D}=(E-L)G$ yields

$$\mathcal{D}_{ii}=(EG)_{ii}-l_iG_{ii}=(z-l_i)G_{ii}-1.$$

Equivalently,

$$(z-l_i)G_{ii}=1+\mathcal{D}_{ii}.$$

On the other hand, we have from the QVE that $(z-h_i)\phi_i=1$. Thus

$$\begin{aligned}(z-h_i)(\phi_i-G_{ii})&=(z-h_i)\phi_i-(z-h_i)G_{ii}\\&=1-(z-h_i)G_{ii}\\&=(z-l_i)G_{ii}-\mathcal{D}_{ii}-(z-h_i)G_{ii}\\&=(h_i-l_i)G_{ii}-\mathcal{D}_{ii}.\end{aligned}$$

Multiplying by ϕ_i, since $\phi_i = (z - h_i)^{-1}$, yields

$$d_i = \phi_i G_{ii}(h_i - l_i) - \phi_i \mathcal{D}_{ii}. \tag{D.1}$$

Moreover,

$$h_i - l_i = \sum_{j=1}^{n} \sigma_{ij}^2 (\phi_j - G_{jj}) = \sum_{j=1}^{n} \sigma_{ij}^2 d_j.$$

Substituting this into (D.1), we obtain

$$d_i = \phi_i G_{ii} \Big(\sum_{j=1}^{n} \sigma_{ij}^2 \Big) d_j - \phi_i \mathcal{D}_{ii}.$$

Thus, by taking the ℓ_∞ norm, we get

$$\|d\|_\infty \leqslant \max_i |\phi_i G_{ii}| \cdot R_{\max} \|d\|_\infty + \max_i |\phi_i| \cdot \max_i |\mathcal{D}_{ii}|.$$

Since

$$\max_i |\phi_i G_{ii}| \leqslant \frac{3}{2|z|} \cdot \frac{2}{|z|} = \frac{3}{|z|^2},$$

we bound

$$\|d\|_\infty \leqslant \frac{3R_{\max}}{|z|^2} \|d\|_\infty + \frac{3}{2|z|} \max_i |\mathcal{D}_{ii}|.$$

Since $|z|^2 \geqslant 36 R_{\max}$ by our assumption, we further obtain

$$\|d\|_\infty \leqslant \frac{2}{|z|} \max_i |\mathcal{D}_{ii}|. \tag{D.2}$$

We now return to $v^\top (L - H) G w$. Since v, w are unit vectors,

$$|v^\top (L - H) G w| \leqslant \|L - H\| \cdot \|G\|,$$

where

$$\|L - H\| = \max_i \Big| \sum_{j=1}^{n} \sigma_{ij}^2 (G_{jj} - \phi_j) \Big| \leqslant R_{\max} \|d\|_\infty.$$

Combining this with $\|G\| \leqslant 2/|z|$ and (D.2), we arrive at

$$|v^\top (L - H) G w| \leqslant \frac{4R_{\max}}{|z|^2} \max_i |\mathcal{D}_{ii}| \leqslant \frac{1}{9} \max_i |\mathcal{D}_{ii}|. \tag{D.3}$$

To bound $\mathcal{D}_{ii}$, we apply Proposition 8.1 with $v = w = e_i$ and take a union bound to obtain

$$\max_{1 \leqslant i \leqslant n} |\mathcal{D}_{ii}| \leqslant C \left(\frac{\mathfrak{m}_D}{|z|} + n^{-500} \right)$$

with probability at least $1 - n^{-D-4}$. In the bounded sparse-entry regime, the analogous bound follows without the n^{-500} term.

Combining the above estimate with (D.3) proves Proposition 8.2.

E Proofs of Corollary 3 and Corollary 4

E.1 Proof of Corollary 3

Write
$$\Xi(z) = G(z) - \Phi(z).$$
We start with the following consequence of the isotropic local law:

COROLLARY 5 (Compressed local law on the signal subspace)**.** *Fix $D > 0$. Suppose the assumptions of Theorem 4 hold. Then, for every fixed $z \in \mathbb{C}_+$ with $|z| \geqslant 6\sqrt{R_{\max}}$, with probability at least $1 - n^{-D-2}$,*
$$\left\| U^\top (G(z) - \Phi(z))U \right\| \leqslant Cr \frac{\mathfrak{m}_D}{|z|^2}.$$

Proof of Corollary 5. Since
$$\begin{aligned}\left\| U^\top (G(z) - \Phi(z))U \right\| \leqslant \left\| U^\top (G(z) - \Phi(z))U \right\|_F &= \sqrt{\sum_{i,j=1}^{r} \left| e_i{}^\top U^\top (G(z) - \Phi(z)) U e_j \right|^2} \\ &\leqslant r \max_{1 \leqslant i,j \leqslant r} \left| e_i{}^\top U^\top (G(z) - \Phi(z)) U e_j \right|,\end{aligned}$$
the conclusion of Corollary 5 follows directly from Theorem 4 and the union bounds. □

To extend the pointwise bound Corollary 5 to the spectral domain
$$\mathcal{S}_{\text{out}} := \{ z \in \mathbb{C} : 6\sqrt{R_{\max}} \leqslant |z| \leqslant 2R_{\max}^3 \},$$
we use a standard ε-net argument. Set
$$\eta := \min\left\{1, \frac{r\mathfrak{m}_D}{20}\right\}.$$
Let $\mathcal{N}_\eta$ be a η-net of $\mathcal{S}_{\text{out}}$. A simple volume argument (see for instance [86, Lemma 3.3]) shows that
$$|\mathcal{N}_\eta| \leqslant \left(1 + \frac{8R_{\max}^3}{\eta}\right)^2.$$
Since $R_{\max} \leqslant n^{100}$ and $\sigma_{\max}^{-1} \leqslant n^{100}$, we have $\mathfrak{m}_D \geqslant cn^{-100}$ for an absolute constant $c > 0$. Thus $R_{\max}^3/\eta \leqslant C' n^{400}$ and $|\mathcal{N}_\eta| \leqslant Cn^{800}$.

Applying Corollary 5 with $D_1 = D + 810$ and taking a union bound over $z \in \mathcal{N}_\eta$, we obtain
$$\mathbb{P}\left(\max_{z \in \mathcal{N}_\eta} |z|^2 \left\| U^\top \Xi(z) U \right\| \geqslant Cr\mathfrak{m}_{D_1} \right) \leqslant |\mathcal{N}_\eta| n^{-D-814} < n^{-D-10}.$$
Using $\mathfrak{m}_{D_1} = \mathfrak{m}_{D+810} \leqslant C'\mathfrak{m}_D$, we establish that with probability at least $1 - n^{-D-10}$,
$$\max_{z \in \mathcal{N}_\eta} |z|^2 \left\| U^\top \Xi(z) U \right\| \leqslant Cr\mathfrak{m}_D.$$

Now we extend the estimate from $\mathcal{N}_\eta$ to the whole domain $\mathcal{S}_{\rm out}$. Define

$$f(z) := z^2 U^\top G(z) U, \qquad g(z) := z^2 U^\top \Phi(z) U.$$

We first show that f and g are Lipschitz on $\mathcal{S}$.

Let $z, w \in \mathcal{S}_{\rm out}$ and assume without loss of generality that $|z| \geqslant |w| \geqslant 6\sqrt{R_{\max}}$. By the resolvent identity,

$$G(z) - G(w) = (w - z) G(z) G(w).$$

Using $\|G(z)\| \leqslant 2/|z|$ and $\|G(w)\| \leqslant 2/|w|$, we get

$$\begin{aligned}
\|f(z) - f(w)\| &\leqslant |z||z-w|\|G(z)\| + |w||z-w|\|G(z)\| \\
&\quad + |w|^2 \|G(z) - G(w)\| \\
&\leqslant 2|z-w| + 2\frac{|w|}{|z|}|z-w| + 4\frac{|w|}{|z|}|z-w| \\
&\leqslant 8|z-w|.
\end{aligned}$$

Hence f is 8-Lipschitz on $\mathcal{S}$.

Next we prove the Lipschitz bound for g. For each $i \in [n]$, the QVE gives

$$\frac{1}{\phi_i(z)} = z - \sum_j \sigma_{ij}^2 \phi_j(z), \qquad \frac{1}{\phi_i(w)} = w - \sum_j \sigma_{ij}^2 \phi_j(w).$$

Subtracting the two identities and multiplying by $\phi_i(z)\phi_i(w)$, we obtain

$$\phi_i(z) - \phi_i(w) = \phi_i(z)\phi_i(w) \left[w - z + \sum_j \sigma_{ij}^2 (\phi_j(z) - \phi_j(w)) \right].$$

Therefore, by Lemma 5.1,

$$\|\Phi(z) - \Phi(w)\| \leqslant \frac{9}{4|z||w|} \left(|z-w| + R_{\max} \|\Phi(z) - \Phi(w)\| \right).$$

Since $|z||w| \geqslant 36 R_{\max}$, we have

$$\frac{9R_{\max}}{4|z||w|} \leqslant \frac{1}{16}.$$

Thus,

$$\|\Phi(z) - \Phi(w)\| \leqslant \frac{12}{5} \frac{|z-w|}{|z||w|}.$$

Using again Lemma 5.1, we have $\|\Phi(z)\| \leqslant 3/(2|z|)$. Hence

$$\begin{aligned}
\|g(z) - g(w)\| &\leqslant |z||z-w|\|\Phi(z)\| + |w||z-w|\|\Phi(z)\| \\
&\quad + |w|^2 \|\Phi(z) - \Phi(w)\| \\
&\leqslant \frac{3}{2}|z-w| + \frac{3}{2}\frac{|w|}{|z|}|z-w| + \frac{12}{5}\frac{|w|}{|z|}|z-w| \\
&< 6|z-w|.
\end{aligned}$$

Therefore g is 6-Lipschitz on $\mathcal{S}$.

Now take any $z \in \mathcal{S}_{\text{out}}$. By the definition of the η-net, there exists $w \in \mathcal{N}_\eta$ such that $|z-w| \leqslant \eta$. Then

$$\begin{aligned} |z|^2 \left\| U^\top \Xi(z) U \right\| &= \|f(z) - g(z)\| \\ &\leqslant \|f(z) - f(w)\| + \|f(w) - g(w)\| + \|g(w) - g(z)\| \\ &\leqslant 14\eta + |w|^2 \left\| U^\top \Xi(w) U \right\|. \end{aligned}$$

Since $\eta \leqslant \frac{r\mathfrak{m}_D}{20}$, we get

$$\sup_{z \in \mathcal{S}_{\text{out}}} |z|^2 \left\| U^\top \Xi(z) U \right\| \leqslant \frac{14 r\mathfrak{m}_D}{20} + C r\mathfrak{m}_D \leqslant C' r\mathfrak{m}_D$$

with probability at least $1 - n^{-D-10}$. This completes the proof of (6.1).

The proof of (6.2) is identical to the proof of (6.1). One applies Theorem 4 to the pairs (e_i, u_a) for $i \in [n]$ and $a \in [r]$, takes a union bound over i, a, and uses the same net argument over $\mathcal{S}_{\text{out}}$.

Finally,

$$e_i^\top Q \Xi(z) U = e_i^\top \Xi(z) U - e_i^\top U \cdot U^\top \Xi(z) U,$$

so (6.3) follows from (6.2) and (6.1).

E.2 Proof of Corollary 4

We first collect two preliminary estimates used in the proof. We begin with the concentration estimate used in the leave-one-out argument. It follows from the scalar Bernstein inequality [93, Theorem 2.8.4], combined with a standard net argument; in the sub-Gaussian regime, we first truncate the variables at the appropriate high-probability level.

LEMMA E.1 (Vector Bernstein inequality)**.** *Let $X_1, \ldots, X_n$ be independent centered real random variables and $c_1, \ldots, c_n \in \mathbb{R}^r$ be deterministic. Set $v^2 := \sum_{j=1}^n \mathbb{E}(X_j^2)\|c_j\|^2$ and $c_\infty := \max_{1 \leqslant j \leqslant n} \|c_j\|$. There exists an absolute constant $C > 0$ such that the following hold.*

(a) *If $|X_j| \leqslant L$ almost surely for every j, then, for $t \geqslant Cr$,*

$$\Big\| \sum_{j=1}^n X_j c_j \Big\| \leqslant C\Big(\sqrt{t}\, v + tLc_\infty\Big)$$

with probability at least $1 - 2e^{-ct}$.

(b) *Suppose $\|X_j\|_{\psi_2} \leqslant K\tau_j$, where $\tau_j^2 = \mathbb{E}(X_j^2) \leqslant \sigma_*^2$. Then, for $t \geqslant C(r + \log n)$,*

$$\Big\| \sum_{j=1}^n X_j c_j \Big\| \leqslant C\Big(\sqrt{t}\, v + tK\sigma_* \sqrt{t}\, c_\infty\Big)$$

with probability at least $1 - 4e^{-ct}$.

The same conclusions hold conditionally if the vectors c_j are measurable with respect to a σ-field independent of $X_1, \ldots, X_n$.

In particular, taking $X_j = E_{ij}$ and $t = t_{D,r} = C(D + r)\log n$, in either regime we have

$$\Big\| \sum_j E_{ij} c_j \Big\| \leqslant C\Big[\sqrt{t_{D,r}}\big(\sum_j \sigma_{ij}^2 \|c_j\|_2^2\big)^{1/2} + t_{D,r} L_{D,r}^{\mathrm{ra}} \max_j \|c_j\|_2\Big] \tag{E.1}$$

with probability at least $1 - 4e^{-ct_{D,r}}$.

Next, we collect several resolvent identities needed for the leave-one-out approach. For $i \in [n]$, let $E^{(i)}$ be the principal minor obtained from E by deleting its ith row and column, and write

$$h_i := (E_{ji})_{j \neq i} \in \mathbb{R}^{n-1}, \qquad G^{(i)}(z) := (zI - E^{(i)})^{-1}.$$

We also write $B^{(i)} \in \mathbb{R}^{(n-1)\times r}$ for the matrix obtained from B by deleting its ith row. We defer a short proof of these resolvent identities to the end of this section.

LEMMA E.2. *For every $i \in [n]$,*

$$G_{ii} = \Big(z - E_{ii} - h_i^\top G^{(i)} h_i\Big)^{-1}, \tag{E.2}$$

$$e_i^\top G B = G_{ii}\Big(e_i^\top B + h_i^\top G^{(i)} B^{(i)}\Big), \tag{E.3}$$

$$G^{(i)} B^{(i)} = (GB)^{(i)} - G^{(i)} h_i (e_i^\top G B), \tag{E.4}$$

where $(GB)^{(i)}$ denotes the matrix obtained from GB by deleting its ith row.

We first prove the row-adaptive local law for a fixed spectral parameter. Recall the definition of $L_{D,r}$ from (1.5) and from (6.4) that

$$t_{D,r} := C(D + r)\log n.$$

PROPOSITION E.1. *Let $B \in \mathbb{R}^{n\times r}$ be deterministic and suppose $|z| \geqslant C'\sqrt{t_{D,r} R_{\max}}$ for a sufficiently large absolute constant C'. Then, with probability at least $1 - Cn^{-D-2}$,*

$$|z|^2 \|(G(z) - \Phi(z))B\|_{2,\infty} \leqslant C\left[\sqrt{t_{D,r}}\,\Gamma(B) + \big(t_{D,r} L_{D,r}^{\mathrm{ra}} + \mathfrak{m}_D\big)\|B\|_{2,\infty}\right]. \tag{E.5}$$

Proof. Our goal is to control $\|(G - \Phi)B\|_{2,\infty}$ by a leave-one-out argument. We first note that $\|\Phi\| = \max_i |\phi_i| \leqslant \frac{3}{2|z|}$ by Lemma 5.1 (i) and thus

$$\|GB\|_{2,\infty} \leqslant \|\Phi B\|_{2,\infty} + \|(G - \Phi)B\|_{2,\infty} \leqslant \frac{3}{2|z|}\|B\|_{2,\infty} + \|(G - \Phi)B\|_{2,\infty}. \tag{E.6}$$

This inequality will be used to close the bootstrap below.

Fix $i \in [n]$ and condition on $E^{(i)}$ and E_{ii}. Then h_i is independent of the conditioning σ-field, whereas $G^{(i)} B^{(i)}$ is measurable with respect to it. Hence, combining (E.1) with a union bound over i yields

$$\|h_i^\top G^{(i)} B^{(i)}\| \leqslant C\Big[\sqrt{t_{D,r}}\big(\sum_{j\neq i} \sigma_{ij}^2 \|e_j^\top G^{(i)} B^{(i)}\|^2\big)^{1/2} + t_{D,r} L_{D,r}^{\mathrm{ra}} \|G^{(i)} B^{(i)}\|_{2,\infty}\Big] \tag{E.7}$$

for all i, with probability at least $1 - 4ne^{-ct_{D,r}}$.

We now estimate the two terms on the right-hand side of (E.7). By (E.4), for $j \neq i$,

$$e_j^\top G^{(i)} B^{(i)} = e_j^\top GB - e_j^\top G^{(i)} h_i \cdot e_i^\top GB.$$

Decomposing

$$e_j^\top GB = e_j^\top \Phi B + e_j^\top (G - \Phi)B = \phi_j e_j^\top B + e_j^\top (G - \Phi)B$$

and applying Minkowski's inequality, we obtain

$$\begin{aligned}\big(\sum_{j\neq i} \sigma_{ij}^2 \|e_j^\top G^{(i)} B^{(i)}\|^2\big)^{1/2} \leqslant & \big(\sum_{j\neq i} \sigma_{ij}^2 |\phi_j|^2 \|e_j^\top B\|^2\big)^{1/2} + \sqrt{R_i}\, \|(G-\Phi)B\|_{2,\infty} \\ & + \big(\sum_{j\neq i} \sigma_{ij}^2 |e_j^\top G^{(i)} h_i|^2\big)^{1/2} \|GB\|_{2,\infty}.\end{aligned}$$

Using $\max_j |\phi_j| \leqslant 3/(2|z|)$ and $\sigma_{ij} \leqslant \sigma_{\max}$, we further get

$$\big(\sum_{j\neq i} \sigma_{ij}^2 \|e_j^\top G^{(i)} B^{(i)}\|^2\big)^{1/2} \leqslant \frac{3}{2|z|}\gamma_i(B) + \sqrt{R_i}\, \|(G-\Phi)B\|_{2,\infty} + \sigma_{\max} \|G^{(i)} h_i\| \cdot \|GB\|_{2,\infty}. \tag{E.8}$$

Likewise, (E.4) yields

$$\|G^{(i)} B^{(i)}\|_{2,\infty} \leqslant \big(1 + \|G^{(i)} h_i\|\big) \|GB\|_{2,\infty}. \tag{E.9}$$

Combining (E.7), (E.8), and (E.9), and using $\sigma_{\max} \leqslant L_{D,r}^{\mathrm{ra}}$ with

$$\|G^{(i)} h_i\| \leqslant \|G^{(i)}\|\, \|E e_i\| \leqslant \|G^{(i)}\|\, \|E\| \leqslant \frac{6\sqrt{R_{\max}}}{|z|},$$

we obtain

$$\begin{aligned}\|h_i^\top G^{(i)} B^{(i)}\| \leqslant C \Bigg[& \frac{\sqrt{t_{D,r}}}{|z|}\gamma_i(B) + \sqrt{t_{D,r} R_i}\, \|(G-\Phi)B\|_{2,\infty} \\ & + t_{D,r} L_{D,r}^{\mathrm{ra}} \left(1 + \frac{\sqrt{R_{\max}}}{|z|}\right) \|GB\|_{2,\infty} \Bigg].\end{aligned}$$

Since $|z| \geqslant C' \sqrt{t_{D,r} R_{\max}}$ and by (E.6),

$$\|GB\|_{2,\infty} \leqslant \frac{3}{2|z|} \|B\|_{2,\infty} + \|(G-\Phi)B\|_{2,\infty},$$

this simplifies to

$$\begin{aligned}\|h_i^\top G^{(i)} B^{(i)}\| \leqslant C \Bigg[& \frac{\sqrt{t_{D,r}}}{|z|}\gamma_i(B) + \frac{t_{D,r} L_{D,r}^{\mathrm{ra}}}{|z|} \|B\|_{2,\infty} \\ & + \big(\sqrt{t_{D,r} R_i} + t_{D,r} L_{D,r}^{\mathrm{ra}}\big) \|(G-\Phi)B\|_{2,\infty} \Bigg].\end{aligned} \tag{E.10}$$

To proceed, note that, by (E.3),

$$e_i^\top (G - \Phi)B = (G_{ii} - \phi_i)e_i^\top B + G_{ii} h_i^\top G^{(i)} B^{(i)}.$$

Using (E.10) and taking the maximum over i, together with $\|G\| \leqslant \frac{2}{|z|}$, we get

$$\begin{aligned} \|(G - \Phi)B\|_{2,\infty} \leqslant\ & \max_i |G_{ii} - \phi_i| \, \|B\|_{2,\infty} \\ & + \frac{C}{|z|^2} \left[\sqrt{t_{D,r}}\, \Gamma(B) + t_{D,r} L_{D,r}^{\mathrm{ra}} \|B\|_{2,\infty} \right] \\ & + C \frac{\sqrt{t_{D,r} R_{\max}} + t_{D,r} L_{D,r}^{\mathrm{ra}}}{|z|} \|(G - \Phi)B\|_{2,\infty}. \end{aligned} \tag{E.11}$$

By Assumption 1.3,

$$t_{D,r} L_{D,r}^{\mathrm{ra}} \lesssim \sqrt{t_{D,r} R_{\max}} \leqslant (1/C')|z|,$$

so the last coefficient in (E.11) is at most $1/2$ when C' is sufficiently large. Therefore,

$$\|(G - \Phi)B\|_{2,\infty} \leqslant 2 \max_i |G_{ii} - \phi_i| \, \|B\|_{2,\infty} + \frac{C}{|z|^2} \left[\sqrt{t_{D,r}}\, \Gamma(B) + t_{D,r} L_{D,r}^{\mathrm{ra}} \|B\|_{2,\infty} \right].$$

Finally, it follows by applying Theorem 4 with $v = w = e_i$ and taking a union bound over i that

$$\max_i |G_{ii} - \phi_i| \leqslant \frac{\mathfrak{m}_D}{|z|^2},$$

which proves (E.5). $\square$

We now pass from the fixed-z estimate to the uniform row-adaptive local law. The argument follows the proof of Corollary 3, so we only sketch it.

Proof of Corollary 4. It remains to make Proposition E.1 uniform over $z \in \mathcal{S}_{\mathrm{ra}}$ defined in (6.5). On the event $\{\|E\| \leqslant 3\sqrt{R_{\max}}\}$, the same resolvent calculation as in the proof of Corollary 3 shows that the maps rowwise$z \mapsto z^2 G(z)B$ and $z \mapsto z^2 \Phi(z)B$ are both $8\|B\|$-Lipschitz.

We now apply the same ε-net argument as in the proof of Corollary 3. By the polynomial bounds in Assumption 1.1, a suitable net of $\mathcal{S}_{\mathrm{ra}}$ has cardinality polynomial in n. Applying Proposition E.1 on this net and taking a union bound, the choice

$$t_{D,r} = C(D + r) \log n$$

absorbs the net cardinality and yields failure probability at most n^{-D}. The Lipschitz bounds above extend the estimate from the net to all $z \in \mathcal{S}_{\mathrm{ra}}$. This proves (6.6). $\square$

Finally, we provide the proof of Lemma E.2.

Proof of Lemma E.2. Without loss of generality, denote

$$zI - E = \begin{pmatrix} z - E_{ii} & -h_i^\top \\ -h_i & zI - E^{(i)} \end{pmatrix}.$$

Write accordingly

$$G = \begin{pmatrix} G_{ii} & G_{i,-i} \\ G_{-i,i} & G_{-i,-i} \end{pmatrix},$$

where the subscript $-i$ means that the ith coordinate is deleted. The block inverse formula (the Schur complement) yields (E.2) and

$$G_{i,-i} = G_{ii} h_i^\top G^{(i)}, \qquad G_{-i,-i} = G^{(i)} + G^{(i)} h_i G_{ii} h_i^\top G^{(i)},$$

where $-i$ denotes deletion of the ith coordinate. Since

$$B = \begin{pmatrix} e_i^\top B \\ B^{(i)} \end{pmatrix},$$

multiplying the first block row by B leads to

$$e_i^\top GB = G_{ii} \left(e_i^\top B + h_i^\top G^{(i)} B^{(i)} \right),$$

which is (E.3). Similarly, the lower block row gives

$$(GB)^{(i)} = G^{(i)} B^{(i)} + G^{(i)} h_i G_{ii} \left(e_i^\top B + h_i^\top G^{(i)} B^{(i)} \right) = G^{(i)} B^{(i)} + G^{(i)} h_i (e_i^\top GB),$$

which is equivalent to (E.4). □

F Extensions to general signal eigenspaces

Theorem 1 is stated for the top positive eigenspace only to simplify notation. The proof extends directly to general eigenspaces $U_{l:s} = (u_l, \dots, u_s)$ for $1 \leqslant l < s \leqslant r$. We also provide a simple alternative approach using the bounds for top eigenspaces.

For negative population spikes, we apply our framework to $-A$ and $-(A+E)$. In our ordering $\lambda_1 \geqslant \cdots \geqslant \lambda_{r_+} > 0 > \lambda_{r_+ + 1} \geqslant \dots \geqslant \lambda_r$, the eigenspaces corresponding to the most negative eigenvalues become the top eigenspaces. This symmetry yields the same perturbation guarantees for the bottom eigenspaces without additional analysis.

For intermediate blocks $U_{l:s}$ and $\widetilde{U}_{l:s}$ where $1 \leqslant l < s \leqslant r_+$. We use the decomposition $P_{U_{l:s}} = P_{U_s} - P_{U_{l-1}}$ to obtain

$$\| \sin \angle (U_{l:s}, \widetilde{U}_{l:s}) \| \leqslant \| \sin \angle (U_{l-1}, \widetilde{U}_{l-1}) \| + \| \sin \angle (U_s, \widetilde{U}_s) \|, \tag{F.1}$$

and

$$\min_O \| \widetilde{U}_{l:s} - U_{l:s} O \|_{2,\infty} \leqslant \| (I - P_{U_{l:s}}) \widetilde{U}_{l:s} \|_{2,\infty} + \| U_{l:s} \|_{2,\infty} \| \sin \angle (U_{l:s}, \widetilde{U}_{l:s}) \|^2, \tag{F.2}$$

where

$$\| (I - P_{U_{l:s}}) \widetilde{U}_{l:s} \|_{2,\infty} \leqslant \| (I - P_{U_s}) \widetilde{U}_s \|_{2,\infty} + \| U_{l-1} \|_{2,\infty} \| \sin \angle (U_{l-1}, \widetilde{U}_{l-1}) \|.$$

Thus, the perturbation bounds for $U_{l:s}, \widetilde{U}_{l:s}$ follow by substituting the estimates for the top-s and top-$(l-1)$ eigenspaces. By symmetry, analogous decompositions hold for intermediate blocks of negative spikes.

We provide the proofs of (F.1) and (F.2) below.

Proof of (F.1): First, by [28, Exercises VII. 1. 9– 1.11],

$$\|\sin\angle(U_{l:s},\widetilde{U}_{l:s})\| = \|(I-P_{U_{l:s}})\widetilde{U}_{l:s}\| = \|P_{U_{l:s}^{\perp}}\widetilde{U}_{l:s}\|.$$

From the decomposition $P_{U_{l:s}} = P_{U_s} - P_{U_{l-1}}$, we can express

$$P_{U_{l:s}^{\perp}} = I - P_{U_{l:s}} = (I-P_{U_s}) + P_{U_{l-1}} = P_{U_s^{\perp}} + P_{U_{l-1}}.$$

It follows that

$$\|P_{U_{l:s}^{\perp}}\widetilde{U}_{l:s}\| \leqslant \|P_{U_s^{\perp}}\widetilde{U}_{l:s}\| + \|P_{U_{l-1}}\widetilde{U}_{l:s}\|. \tag{F.3}$$

For the first term, observe that the columns of $\widetilde{U}_{l:s}$ are a subset of the columns of $\widetilde{U}_s$. Thus

$$\|P_{U_s^{\perp}}\widetilde{U}_{l:s}\| \leqslant \|P_{U_s^{\perp}}\widetilde{U}_s\| = \|\sin\angle(U_s,\widetilde{U}_s)\|.$$

For the second term in (F.3), since $\widetilde{U}_{l:s}$ is orthogonal to $\widetilde{U}_{l-1}$, we have $P_{\widetilde{U}_{l-1}^{\perp}}\widetilde{U}_{l:s} = \widetilde{U}_{l:s}$. Then

$$P_{U_{l-1}}\widetilde{U}_{l:s} = P_{U_{l-1}}P_{\widetilde{U}_{l-1}^{\perp}}\widetilde{U}_{l:s},$$

and therefore,

$$\|P_{U_{l-1}}P_{\widetilde{U}_{l-1}^{\perp}}\widetilde{U}_{l:s}\| \leqslant \|P_{U_{l-1}}P_{\widetilde{U}_{l-1}^{\perp}}\|\cdot\|\widetilde{U}_{l:s}\| = \|\sin\angle(U_{l-1},\widetilde{U}_{l-1})\|.$$

Substituting these two bounds back into (F.3) completes the proof of (F.1).

Proof of (F.2): Let $U_{l:s}^{\top}\widetilde{U}_{l:s} = WDV^{\top}$ be the singular value decomposition. We choose $O = WV^{\top}$ and then decompose

$$\begin{aligned}\widetilde{U}_{l:s} - U_{l:s}O &= (I-P_{U_{l:s}})\widetilde{U}_{l:s} + P_{U_{l:s}}(\widetilde{U}_{l:s} - U_{l:s}O)\\ &= (I-P_{U_{l:s}})\widetilde{U}_{l:s} + U_{l:s}(U_{l:s}^{\top}\widetilde{U}_{l:s} - O).\end{aligned}$$

Taking the $2\to\infty$ norm and applying the triangle inequality yields

$$\|\widetilde{U}_{l:s} - U_{l:s}O\|_{2,\infty} \leqslant \|(I-P_{U_{l:s}})\widetilde{U}_{l:s}\|_{2,\infty} + \|U_{l:s}(U_{l:s}^{T}\widetilde{U}_{l:s} - O)\|_{2,\infty}. \tag{F.4}$$

By our choice of O, the second term in (F.4) is bounded by

$$\|U_{l:s}W(D-I)V^{\top}\|_{2,\infty} \leqslant \|U_{l:s}\|_{2,\infty}\|W(I-D)V^{\top}\| = \|U_{l:s}\|_{2,\infty}\|I-D\|.$$

Note that the diagonal entries of D are the cosines of the principal angles θ_i. The entries of $I-D^2$ are exactly $\sin^2\theta_i$. Because $D \succeq 0$, we have $I - D \preceq I - D^2$, which implies

$$\|I-D\| \leqslant \|I-D^2\| = \|\sin\angle(U_{l:s},\widetilde{U}_{l:s})\|^2.$$

This proves the main inequality in (F.2). It remains to bound $\|(I-P_{U_{l:s}})\widetilde{U}_{l:s}\|_{2,\infty}$. Using $I - P_{U_{l:s}} = (I-P_{U_s}) + P_{U_{l-1}}$, we expand

$$(I-P_{U_{l:s}})\widetilde{U}_{l:s} = (I-P_{U_s})\widetilde{U}_{l:s} + P_{U_{l-1}}\widetilde{U}_{l:s}.$$

Thus

$$\|(I - P_{U_{l:s}})\widetilde{U}_{l:s}\|_{2,\infty} \leqslant \|(I - P_{U_s})\widetilde{U}_{l:s}\|_{2,\infty} + \|P_{U_{l-1}}\widetilde{U}_{l:s}\|_{2,\infty}. \tag{F.5}$$

For the first term, since $(I - P_{U_s})\widetilde{U}_{l:s}$ consists of a subset of the columns of $(I - P_{U_s})\widetilde{U}_s$, we get

$$\|(I - P_{U_s})\widetilde{U}_{l:s}\|_{2,\infty} \leqslant \|(I - P_{U_s})\widetilde{U}_s\|_{2,\infty}.$$

For the second term, plugging in $P_{U_{l-1}} = U_{l-1}U_{l-1}^\top$, we find

$$\|P_{U_{l-1}}\widetilde{U}_{l:s}\|_{2,\infty} = \|U_{l-1}(U_{l-1}^\top\widetilde{U}_{l:s})\|_{2,\infty} \leqslant \|U_{l-1}\|_{2,\infty}\|U_{l-1}^\top\widetilde{U}_{l:s}\|.$$

Finally, because $\widetilde{U}_{l:s}$ is orthogonal to $\widetilde{U}_{l-1}$, we obtain

$$\|U_{l-1}^\top\widetilde{U}_{l:s}\| = \|U_{l-1}^\top P_{\widetilde{U}_{l-1}^\perp}\widetilde{U}_{l:s}\| \leqslant \|U_{l-1}^\top P_{\widetilde{U}_{l-1}^\perp}\|\|\widetilde{U}_{l:s}\| = \|\sin\angle(U_{l-1}, \widetilde{U}_{l-1})\|.$$

Substituting these bounds into (F.5) finishes the proof.

G Extension to rectangular matrices: singular subspace perturbations

We briefly discuss the extension of current results to the rectangular model:

$$\widetilde{X} = X + E \in \mathbb{R}^{n_1 \times n_2},$$

where $X = U\Sigma V^\top = \sum_{i=1}^r \sigma_i u_i v_i^\top$ is a rank r matrix with singular values $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$. The random noise matrix E has i.i.d. entries with $\mathbb{E}(E_{ij}) = 0$ and $\mathbb{E}(E_{ij}^2) = \sigma_{ij}^2$. Denote $\Sigma = (\sigma_{ij}^2) \in \mathbb{R}^{n_1 \times n_2}$. The entries of E are sub-Gaussian with a variance profile: $\|E_{ij}\|_{\psi_2} \leqslant K\sigma_{ij}$.

Using the standard Hermitian dilation (or linearization), we embed the rectangular matrices into larger symmetric block matrices by considering the $(n_1 + n_2) \times (n_1 + n_2)$ augmented symmetric matrices:

$$\mathbf{A} = \begin{pmatrix} 0 & X \\ X^\top & 0 \end{pmatrix}, \qquad \mathbf{E} = \begin{pmatrix} 0 & E \\ E^\top & 0 \end{pmatrix}, \quad \widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{E}.$$

The nonzero eigenvalues of $\mathbf{A}$ are

$$\lambda_1 = \sigma_1 \geqslant \cdots \geqslant \lambda_r = \sigma_r > 0 > \lambda_{r+1} = -\sigma_1 \geqslant \cdots \geqslant \lambda_{2r} = -\sigma_r$$

with corresponding eigenvectors

$$w_j := \frac{1}{\sqrt{2}}\begin{pmatrix} u_j \\ v_j \end{pmatrix}, \qquad w_{j+r} := \frac{1}{\sqrt{2}}\begin{pmatrix} u_j \\ -v_j \end{pmatrix}, \qquad 1 \leqslant j \leqslant r.$$

Let $\mathbf{A} = WDW^\top$, where

$$W = \frac{1}{\sqrt{2}}\begin{pmatrix} U & U \\ V & -V \end{pmatrix}.$$

For $I \subseteq [2r]$, we denote by W_I the submatrix of W formed by columns $\{w_i : i \in I\}$. All information about the top-k singular subspaces U_k and V_k is encoded in the eigenspace W_{I_k}, where

$$I_k := \{1, \ldots, k\} \cup \{k+1, \ldots, k+r\}.$$

Hence, we can directly translate our eigenspace perturbation results to singular subspaces. We now introduce the relevant parameters. Define

$$R_i^{(L)} := \sum_{j=1}^{n_1} \sigma_{ij}^2, \qquad R_j^{(R)} := \sum_{i=1}^{n_2} \sigma_{ij}^2,$$

and let $\mathcal{R}_L := \mathrm{diag}(R_i^{(L)})_{1 \leqslant i \leqslant n_1}$, $\mathcal{R}_R := \mathrm{diag}(R_i^{(R)})_{1 \leqslant i \leqslant n_2}$,

$$\mathcal{R}_{\mathrm{aug}} := \mathrm{diag}(\mathcal{R}_L, \mathcal{R}_R), \qquad R_{\max}^{\mathrm{aug}} := \max\Big\{\max_i R_i^{(L)}, \max_j R_j^{(R)}\Big\}.$$

Define $M_{\mathrm{aug}} \equiv M_{\mathrm{aug,D}}$ as in (1.3) with the dimension factor $n_1 + n_2$ and set

$$\rho_k^{\mathrm{aug}} := 10 M_{\mathrm{aug}} + 15 \frac{\mathrm{osc}(\mathcal{R}_{\mathrm{aug}})}{\sigma_k},$$

and

$$\delta_k := \sigma_k - \sigma_{k+1}.$$

Also, we denote $\mathcal{V} = W^\top \mathcal{R}_{\mathrm{aug}} W$. Note that

$$\mathcal{V}_{\mathcal{J}\mathcal{I}} = W_{\mathcal{J}}^\top \mathcal{R}_{\mathrm{aug}} W_{\mathcal{I}} = \frac{1}{2}\left(U_{\mathcal{J}}^\top \mathcal{R}_L U_{\mathcal{I}} + V_{\mathcal{J}}^\top \mathcal{R}_R V_{\mathcal{I}}\right).$$

We define $\mathcal{B}_k^{\mathrm{aug}}$ similar to $\mathcal{B}_k$ (1.4):

$$\mathcal{B}_k^{\mathrm{aug}} := \frac{10\sqrt{k}}{\delta_k \sigma_k}\left(\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\| + 8\frac{R_{\max}^{\mathrm{aug}}}{\sigma_k^2}\mathrm{osc}(\mathcal{R}_{\mathrm{aug}})\right) + 4\sqrt{k}\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\sigma_k^2}.$$

We can then directly translate eigenspace perturbation results from the symmetric model $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{E}$ to the rectangular model $\widetilde{X} = X + E$ using the revised parameters above and the following key relationships (see [98] for details): for any unitarily invariant norm $\|\!|\cdot|\!\|$,

$$\max\{\|\!|\sin\angle(U_k, \widetilde{U}_k)|\!\|, \|\!|\sin\angle(V_k, \widetilde{V}_k)|\!\|\} \leqslant \|\!|\sin\angle(W_I, \widetilde{W}_k)|\!\|$$

and

$$\max\{\|\widetilde{U}_k - P_{U_k}\widetilde{U}_k\|_{2,\infty}, \|\widetilde{V}_k - P_{V_k}\widetilde{V}_k\|_{2,\infty}\} \leqslant \|\widetilde{W}_I - P_{W_I}\widetilde{W}_I\|_{2,\infty}.$$

Specifically, this implies that the following $2 \to \infty$ bound holds with probability at least $1 - n^{-D}$:

$$\begin{aligned} &\min_{O \in \mathbb{O}(k)} \|\widetilde{U}_k - U_k O\|_{2,\infty} \\ &\qquad \leqslant 4\max\{\|U\|_{2,\infty}, \|V\|_{2,\infty}\}\left(\mathcal{B}_k^{\mathrm{aug}} + 30\frac{R_{\max}^{\mathrm{aug}}}{\sigma_k^2}\right) + 120\sqrt{k}\frac{M_{\mathrm{aug}}}{\sigma_k}. \end{aligned} \tag{G.1}$$

The same bound also holds for $\min_{O\in\mathbb{O}(k)}\|\widetilde{V}_k - V_kO\|_{2,\infty}$.

We conclude by clarifying the nature of the coupled terms in (G.1). While the dependence on both variance profiles ($\mathcal{R}_L$ and $\mathcal{R}_R$) reflects the intrinsic statistical coupling in the asymmetric noise matrix, the dependence on the maximum joint incoherence $\max\{\|U\|_{2,\infty},\|V\|_{2,\infty}\}$ and the total dimension n_1+n_2 is an artifact of the symmetric embedding framework. We plan to address this limitation in future work.

H Proofs of Lemma 3.1, Proposition 3.1, (J.1) and (J.2)

H.1 Proof of Lemma 3.1

For brevity, write

$$L_{\mathcal{J},s} := I - U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s)U_{\mathcal{J}}\Lambda_{\mathcal{J}}.$$

For $j\in\mathcal{J}$, set

$$\alpha_j(z) := \lambda_j^{-1} - u_j^\top\Phi(z)u_j.$$

Recall the quantity defined in Proposition 7.2

$$\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s) := \widetilde{\lambda}_s\min_{j\in\mathcal{J}}\left|1-\lambda_j u_j^\top\Phi(\widetilde{\lambda}_s)u_j\right| = \min_{j\in\mathcal{J}}\left|\widetilde{\lambda}_s\lambda_j\alpha_j(\widetilde{\lambda}_s)\right|.$$

We first relate $s_{\min}(L_{\mathcal{J},s})$ to $\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s)$. Decompose

$$U_{\mathcal{J}}^\top\Phi(\widetilde{\lambda}_s)U_{\mathcal{J}} = \operatorname{diag}\big(u_j^\top\Phi(\widetilde{\lambda}_s)u_j\big)_{j\in\mathcal{J}} + Q_{\mathcal{J}}(\widetilde{\lambda}_s),$$

where

$$Q_{\mathcal{J}}(\widetilde{\lambda}_s) := \operatorname{Off}\big(U_{\mathcal{J}}^\top\Phi(\widetilde{\lambda}_s)U_{\mathcal{J}}\big).$$

Then

$$L_{\mathcal{J},s} = \operatorname{diag}\Big(1-\lambda_j u_j^\top\Phi(\widetilde{\lambda}_s)u_j\Big)_{j\in\mathcal{J}} - Q_{\mathcal{J}}(\widetilde{\lambda}_s)\Lambda_{\mathcal{J}}.$$

Hence

$$s_{\min}(L_{\mathcal{J},s}) \geqslant \frac{\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s)}{\widetilde{\lambda}_s} - \|Q_{\mathcal{J}}(\widetilde{\lambda}_s)\Lambda_{\mathcal{J}}\|. \tag{H.1}$$

We next estimate the off-diagonal term $\|Q_{\mathcal{J}}(\widetilde{\lambda}_s)\Lambda_{\mathcal{J}}\|$. Using the expansion

$$\Phi(z) = \frac{1}{z}I + \frac{1}{z^3}R + \varepsilon(z),$$

we have

$$Q_{\mathcal{J}}(\widetilde{\lambda}_s) = \frac{1}{\widetilde{\lambda}_s^3}\operatorname{Off}(\mathcal{V}_{\mathcal{J}\mathcal{J}}) + \operatorname{Off}\big(U_{\mathcal{J}}^\top\varepsilon(\widetilde{\lambda}_s)U_{\mathcal{J}}\big).$$

Since the eigenvalues in $\mathcal{J}$ are positive and bounded above by λ_{k+1},

$$\|Q_{\mathcal{J}}(\widetilde{\lambda}_s)\Lambda_{\mathcal{J}}\| \leqslant \frac{\lambda_{k+1}}{\widetilde{\lambda}_s^3}\|\operatorname{Off}(\mathcal{V}_{\mathcal{J}\mathcal{J}})\| + \lambda_{k+1}\left\|\operatorname{Off}\big(U_{\mathcal{J}}^\top\varepsilon(\widetilde{\lambda}_s)U_{\mathcal{J}}\big)\right\|.$$

By Lemma 5.2,

$$\left\|\mathrm{Off}\left(U_{\mathcal{J}}^{\top}\varepsilon(\widetilde{\lambda}_s)U_{\mathcal{J}}\right)\right\| \leqslant \frac{1}{2}\mathrm{osc}(\varepsilon(\widetilde{\lambda}_s)) \leqslant \frac{15}{2}\frac{R_{\max}}{\widetilde{\lambda}_s^5}\mathrm{osc}(\mathcal{R}).$$

Therefore

$$\|Q_{\mathcal{J}}(\widetilde{\lambda}_s)\Lambda_{\mathcal{J}}\| \leqslant \frac{\lambda_{k+1}}{\widetilde{\lambda}_s^3}\|\mathrm{Off}(V_{\mathcal{J}\mathcal{J}})\| + \frac{15}{2}\frac{\lambda_{k+1}R_{\max}}{\widetilde{\lambda}_s^5}\mathrm{osc}(\mathcal{R}). \tag{H.2}$$

We now verify that the off-diagonal term in (H.1) is small compared with the diagonal part. Using $\widetilde{\lambda}_s \geqslant \frac{1}{2}\lambda_k$, $R_{\max} \leqslant 36\lambda_k^2$, and $\|\mathrm{Off}(\mathcal{V}_{\mathcal{J}\mathcal{J}})\| \leqslant \mathrm{osc}(\mathcal{R})$, from (H.2), we get

$$\widetilde{\lambda}_s\|Q_{\mathcal{J}}(\widetilde{\lambda}_s)\Lambda_{\mathcal{J}}\| \leqslant \left(8+\frac{5}{3}\right)\frac{\lambda_{k+1}}{\lambda_k^2}\mathrm{osc}(\mathcal{R}) \leqslant \frac{1}{20}\left(8+\frac{5}{3}\right)\delta_k,$$

where the last inequality again follows from the gap assumption. Since $\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s) \geqslant \delta_k/2$, we get

$$\|Q_{\mathcal{J}}(\widetilde{\lambda}_s)\Lambda_{\mathcal{J}}\| \leqslant \frac{1}{10}\left(8+\frac{5}{3}\right)\frac{\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s)}{\widetilde{\lambda}_s}.$$

Returning to (H.1), we obtain

$$s_{\min}(L_{\mathcal{J},s}) \geqslant \frac{1}{30}\frac{\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s)}{\widetilde{\lambda}_s} \geqslant \frac{1}{60}\frac{\delta_k}{\widetilde{\lambda}_s}.$$

Thus $\|L_{\mathcal{J},s}^{-1}\| \leqslant 60\frac{\widetilde{\lambda}_s}{\delta_k}$. This proves the desired bound.

H.2 Proof of Proposition 3.1

For $s \in [k]$, the eigen-equation gives

$$\widetilde{u}_s = G(\widetilde{\lambda}_s)A\widetilde{u}_s = \Phi(\widetilde{\lambda}_s)A\widetilde{u}_s + \Xi(\widetilde{\lambda}_s)A\widetilde{u}_s,$$

where $\Xi(z) := G(z) - \Phi(z)$. It follows that

$$U^{\top}\widetilde{u}_s = U^{\top}\Phi(\widetilde{\lambda}_s)U\Lambda U^{\top}\widetilde{u}_s + U^{\top}\Xi(\widetilde{\lambda}_s)U\Lambda U^{\top}\widetilde{u}_s.$$

Taking the $\mathcal{J}$-block yields

$$U_{\mathcal{J}}^{\top}\widetilde{u}_s = U_{\mathcal{J}}^{\top}\Phi(\widetilde{\lambda}_s)U_{\mathcal{I}}\Lambda_{\mathcal{I}}\cdot U_{\mathcal{I}}^{\top}\widetilde{u}_s + U_{\mathcal{J}}^{\top}\Phi(\widetilde{\lambda}_s)U_{\mathcal{J}}\Lambda_{\mathcal{J}}\cdot U_{\mathcal{J}}^{\top}\widetilde{u}_s + U_{\mathcal{J}}^{\top}\Xi(\widetilde{\lambda}_s)U\Lambda U^{\top}\widetilde{u}_s.$$

Collecting the $U_{\mathcal{J}}^{\top}\widetilde{u}_s$ terms gives

$$(I - U_{\mathcal{J}}^{\top}\Phi(\widetilde{\lambda}_s)U_{\mathcal{J}}\Lambda_{\mathcal{J}})U_{\mathcal{J}}^{\top}\widetilde{u}_s = U_{\mathcal{J}}^{\top}\Phi(\widetilde{\lambda}_s)U_{\mathcal{I}}\Lambda_{\mathcal{I}}\cdot U_{\mathcal{I}}^{\top}\widetilde{u}_s + U_{\mathcal{J}}^{\top}\Xi(\widetilde{\lambda}_s)U\Lambda U^{\top}\widetilde{u}_s.$$

Recall $\mathcal{T}_s$ from (3.1). We further get

$$U_{\mathcal{J}}^{\top}\widetilde{u}_s - \mathcal{T}_s U_{\mathcal{I}}^{\top}\widetilde{u}_s = (I - U_{\mathcal{J}}^{\top}\Phi(\widetilde{\lambda}_s)U_{\mathcal{J}}\Lambda_{\mathcal{J}})^{-1}U_{\mathcal{J}}^{\top}\Xi(\widetilde{\lambda}_s)U\Lambda U^{\top}\widetilde{u}_s. \tag{H.3}$$

By definition and $\mathcal{N} = \{r_+ + 1, \dots, r\}$,

$$P_U \widetilde{u}_s = U_k U_k^\top \widetilde{u}_s + U_\mathcal{J} U_\mathcal{J}^\top \widetilde{u}_s + U_\mathcal{N} U_\mathcal{N}^\top \widetilde{u}_s.$$

Combining this with (H.3), we get

$$\begin{aligned} w_s^{\text{orc}} &= P_U \widetilde{u}_s - U_\mathcal{J} \mathcal{T}_s U_\mathcal{I}^\top \widetilde{u}_s \\ &= U_k U_k^\top \widetilde{u}_s + U_\mathcal{N} U_\mathcal{N}^\top \widetilde{u}_s + U_\mathcal{J} (I - U_\mathcal{J}^\top \Phi(\widetilde{\lambda}_s) U_\mathcal{J} \Lambda_\mathcal{J})^{-1} U_\mathcal{J}^\top \Xi(\widetilde{\lambda}_s) U \cdot \Lambda U^\top \widetilde{u}_s. \end{aligned} \tag{H.4}$$

Note that we have established in (7.6) that

$$\|U_\mathcal{N}^\top \widetilde{u}_s\| < 4\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\lambda_k^2} + 24\frac{R_{\max}}{\lambda_k^4}\operatorname{osc}(\mathcal{R}) + 4\frac{M}{\lambda_k}. \tag{H.5}$$

Now we estimate the last term on the right-hand side of (H.4). We first have

$$\|\Lambda U^\top \widetilde{u}_s\| \leqslant 2|\widetilde{\lambda}_s|.$$

To see this, from $(A + E)\widetilde{u}_s = \widetilde{\lambda}_s \widetilde{u}_s$, we have

$$A\widetilde{u}_s = (\widetilde{\lambda}_s I - E)\widetilde{u}_s.$$

Multiplying by $U^\top$ gives $\Lambda U^\top \widetilde{u}_s = U^\top A \widetilde{u}_s = U^\top(\widetilde{\lambda}_s I - E)\widetilde{u}_s$. Hence,

$$\|\Lambda U^\top \widetilde{u}_s\| \leqslant |\widetilde{\lambda}_s| + \|E\| \leqslant 2|\widetilde{\lambda}_s|$$

by our assumption $\lambda_k \geqslant 9\sqrt{R_{\max}}$ and Weyl's inequality.

On the event $\mathcal{E}_D$, we have

$$\|U_\mathcal{J}^\top \Xi(\widetilde{\lambda}_s) U\| \leqslant \frac{M}{\widetilde{\lambda}_s^2}.$$

Combining these estimates with Lemma 3.1 yields

$$\begin{aligned} &\|U_\mathcal{J} (I - U_\mathcal{J}^\top \Phi(\widetilde{\lambda}_s) U_\mathcal{J} \Lambda_\mathcal{J})^{-1} U_\mathcal{J}^\top \Xi(\widetilde{\lambda}_s) U \cdot \Lambda U^\top \widetilde{u}_s\| \\ &\quad \leqslant \|(I - U_\mathcal{J}^\top \Phi(\widetilde{\lambda}_s) U_\mathcal{J} \Lambda_\mathcal{J})^{-1} U_\mathcal{J}^\top \Xi(\widetilde{\lambda}_s) U \cdot \Lambda U^\top \widetilde{u}_s\| \leqslant 120\frac{M}{\delta_k}. \end{aligned} \tag{H.6}$$

From (H.4) and the orthogonality of the signal subspaces, $w_s^{\text{orc}} - U_k U_k^\top \widetilde{u}_s$ lies entirely in the span of $U_k^\perp$. Combining the above bounds, we have

$$\begin{aligned} \|w_s^{\text{orc}} - U_k U_k^\top \widetilde{u}_s\| &\leqslant \|U_\mathcal{N}^\top \widetilde{u}_s\| + \|U_\mathcal{J} (I - U_\mathcal{J}^\top \Phi(\widetilde{\lambda}_s) U_\mathcal{J} \Lambda_\mathcal{J})^{-1} U_\mathcal{J}^\top \Xi(\widetilde{\lambda}_s) U \cdot \Lambda U^\top \widetilde{u}_s\| \\ &\leqslant 4\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\lambda_k^2} + 24\frac{R_{\max}}{\lambda_k^4}\operatorname{osc}(\mathcal{R}) + 124\frac{M}{\delta_k}. \end{aligned}$$

Denote

$$\Delta^{\text{orc}} := (w_1^{\text{orc}} - U_k U_k^\top \widetilde{u}_1, \dots, w_k^{\text{orc}} - U_k U_k^\top \widetilde{u}_k).$$

It follows immediately that

$$\|\Delta^{\text{orc}}\| \leqslant \|\Delta^{\text{orc}}\|_F \leqslant \sqrt{k}\left(4\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\lambda_k^2} + 24\frac{R_{\max}}{\lambda_k^4}\operatorname{osc}(\mathcal{R}) + 124\frac{M}{\delta_k}\right) = \mathcal{E}_k^{\text{osc}}.$$

We rewrite

$$W_k^{\mathrm{orc}} = (w_1^{\mathrm{orc}}, \dots, w_k^{\mathrm{orc}}) = U_k U_k^\top \widetilde{U}_k + \Delta^{\mathrm{orc}}.$$

Note that

$$U_k^{\mathrm{orc}} = \mathrm{orth}(W_k^{\mathrm{orc}}) = W_k^{\mathrm{orc}}((W_k^{\mathrm{orc}})^\top W_k^{\mathrm{orc}})^{-1/2}. \tag{H.7}$$

We aim to bound

$$\|\sin\angle(U_k, U_k^{\mathrm{orc}})\| = \|P_{U_k^\perp} U_k^{\mathrm{orc}}\| = \|P_{U_k^\perp} W_k^{\mathrm{orc}}((W_k^{\mathrm{orc}})^\top W_k^{\mathrm{orc}})^{-1/2}\| \leqslant \frac{\|P_{U_k^\perp} W_k^{\mathrm{orc}}\|}{s_{\min}(W_k^{\mathrm{orc}})}.$$

First,

$$\|P_{U_k^\perp} W_k^{\mathrm{orc}}\| = \|P_{U_k^\perp}(U_k U_k^\top \widetilde{U}_k + \Delta^{\mathrm{orc}})\| = \|P_{U_k^\perp} \Delta^{\mathrm{orc}}\| \leqslant \|\Delta^{\mathrm{orc}}\|.$$

Next, since $W_k^{\mathrm{orc}} = U_k U_k^\top \widetilde{U}_k + \Delta^{\mathrm{orc}}$, and Δ^{orc} is orthogonal to U_k, we have

$$\sigma_{\min}(W_k^{\mathrm{orc}}) \geqslant \sigma_{\min}(U_k U_k^\top \widetilde{U}_k) = \sigma_{\min}(U_k^\top \widetilde{U}_k) = \sqrt{1 - \|\sin\angle(U_k, \widetilde{U}_k)\|^2}.$$

On the event that Theorem 1 holds, $\|\sin\angle(U_k, \widetilde{U}_k)\| \leqslant 1/100$. Thus,

$$\sigma_{\min}(W_k^{\mathrm{orc}}) \geqslant \frac{1}{2}.$$

We conclude that

$$\|\sin\angle(U_k, U_k^{\mathrm{orc}})\| \leqslant 2\|\Delta^{\mathrm{orc}}\| \leqslant 2\mathcal{E}_k^{\mathrm{osc}}.$$

For the $2 \to \infty$ norm bound, we start with the standard decomposition

$$\min_{O \in \mathbb{O}(k)} \|U_k^{\mathrm{orc}} - U_k O\|_{2,\infty} \leqslant \|(I - P_{U_k}) U_k^{\mathrm{orc}}\|_{2,\infty} + \|U_k\|_{2,\infty} \|\sin\angle(U_k, U_k^{\mathrm{orc}})\|^2.$$

It remains to bound $\|(I - P_{U_k}) U_k^{\mathrm{orc}}\|_{2,\infty}$. Since $P_{U_k^\perp} W_k^{\mathrm{orc}} = \Delta^{\mathrm{orc}}$, from (H.7), we get

$$(I - P_{U_k}) U_k^{\mathrm{orc}} = \Delta^{\mathrm{orc}}((W_k^{\mathrm{orc}})^\top W_k^{\mathrm{orc}})^{-1/2}.$$

Thus

$$\begin{aligned}
\|(I - P_{U_k}) U_k^{\mathrm{orc}}\|_{2,\infty} &\leqslant \|\Delta^{\mathrm{orc}}\|_{2,\infty} \|((W_k^{\mathrm{orc}})^\top W_k^{\mathrm{orc}})^{-1/2}\| \\
&\leqslant \frac{\|\Delta^{\mathrm{orc}}\|_{2,\infty}}{\sigma_{\min}(W_k^{\mathrm{orc}})} \leqslant 2\|\Delta^{\mathrm{orc}}\|_{2,\infty}.
\end{aligned}$$

We bound $\|\Delta^{\mathrm{orc}}\|_{2,\infty}$. From (H.4),

$$\begin{aligned}
\Delta^{\mathrm{orc}} &= (w_1^{\mathrm{orc}} - U_k U_k^\top \widetilde{u}_1, \dots, w_k^{\mathrm{orc}} - U_k U_k^\top \widetilde{u}_k) \\
&= U_{\mathcal{N}} \cdot U_{\mathcal{N}}^\top \widetilde{U}_k + U_{\mathcal{J}} \cdot \mathtt{B}
\end{aligned}$$

where $\mathtt{B} := [\mathtt{b}_1, \cdots, \mathtt{b}_k]$ with columns

$$\mathtt{b}_s = (I - U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{J}} \Lambda_{\mathcal{J}})^{-1} U_{\mathcal{J}}^\top \Xi(\widetilde{\lambda}_s) U \cdot \Lambda U^\top \widetilde{u}_s.$$

In (H.6), we have

$$\|\mathtt{b}_s\| \leqslant 120\frac{M}{\delta_k}$$

and thus

$$\|\mathtt{B}\| \leqslant \|\mathtt{B}\|_F \leqslant 120\sqrt{k}\frac{M}{\delta_k}.$$

Also, by (H.5), we get

$$\begin{aligned}\|U_{\mathcal{N}}^\top \widetilde{U}_k\| \leqslant \|U_{\mathcal{N}}^\top \widetilde{U}_k\|_F &\leqslant \sqrt{k}\max_{s\in k}\|U_{\mathcal{N}}^\top \widetilde{u}_s\| \\ &\leqslant 4\sqrt{k}\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\lambda_k^2} + 24\sqrt{k}\frac{R_{\max}}{\lambda_k^4}\mathrm{osc}(\mathcal{R}) + 4\sqrt{k}\frac{M}{\lambda_k}.\end{aligned}$$

Now we get

$$\begin{aligned}\|(I-P_{U_k})U_k^{\mathrm{orc}}\|_{2,\infty} &\leqslant 2\|\Delta^{\mathrm{orc}}\|_{2,\infty} = 2\max_{1\leqslant i\leqslant n}\|e_i^\top(U_{\mathcal{N}}\cdot U_{\mathcal{N}}^\top \widetilde{U}_k + U_{\mathcal{J}}\cdot \mathtt{B})\| \\ &\leqslant 2\|U\|_{2,\infty}(\|U_{\mathcal{N}}^\top \widetilde{U}_k\| + \|\mathtt{B}\|) \\ &\leqslant \|U\|_{2,\infty}\left(8\sqrt{k}\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\lambda_k^2} + 48\sqrt{k}\frac{R_{\max}}{\lambda_k^4}\mathrm{osc}(\mathcal{R}) + 248\sqrt{k}\frac{M}{\delta_k}\right) \\ &= 2\|U\|_{2,\infty}\mathcal{E}_k^{\mathrm{osc}}.\end{aligned}$$

Finally, combining the bound on $\|\sin\angle(U_k, U_k^{\mathrm{orc}})\|$, we obtain

$$\begin{aligned}\min_{O\in\mathbb{O}(k)}\|U_k^{\mathrm{orc}} - U_k O\|_{2,\infty} &\leqslant \|(I-P_{U_k})U_k^{\mathrm{orc}}\|_{2,\infty} + \|U_k\|_{2,\infty}\|\sin\angle(U_k, U_k^{\mathrm{orc}})\|^2 \\ &\leqslant 3\|U\|_{2,\infty}\mathcal{E}_k^{\mathrm{osc}}.\end{aligned}$$

The proof is complete.

H.3 Proof of (J.1) and (J.2)

By the definitions of $\widehat{w}_s$ and w_s^{orc}, their difference is

$$\widehat{w}_s - w_s^{\mathrm{orc}} = P_{\widetilde{U}}\widetilde{u}_s - P_U\widetilde{u}_s - \left(\widetilde{U}_{\mathcal{J}}\widehat{\mathcal{T}}_s\widetilde{U}_{\mathcal{I}}^\top\widetilde{u}_s - U_{\mathcal{J}}\mathcal{T}_s U_{\mathcal{I}}^\top\widetilde{u}_s\right).$$

Since $P_{\widetilde{U}}\widetilde{u}_s = \widetilde{u}_s$, we get

$$\widehat{W}_k - W_k^{\mathrm{orc}} = (I-P_U)\widetilde{U}_k - \mathfrak{D}_k.$$

Let $\Delta := \|\widehat{W}_k - W_k^{\mathrm{orc}}\|$. Applying the triangle inequality, we bound

$$\Delta \leqslant \|(I-P_U)\widetilde{U}_k\| + \|\mathfrak{D}_k\| \leqslant 2\frac{\|E\|}{\lambda_k} + \|\mathfrak{D}_k\|.$$

The bound in the second inequality was established in (B.2).

Similarly, taking the $2\to\infty$ norm yields

$$\Delta_{2,\infty} := \|\widehat{W}_k - W_k^{\mathrm{orc}}\|_{2,\infty} \leqslant \|(I-P_U)\widetilde{U}_k\|_{2,\infty} + \|\mathfrak{D}_k\|_{2,\infty}.$$

Note that the assumption $\mathcal{E}_k^{\mathrm{orc}} + \frac{\|E\|}{\lambda_k} + \|\mathfrak{D}_k\| \leqslant c$ guarantees that $\mathcal{E}_k^{\mathrm{orc}} + \Delta \leqslant c \leqslant 1/10$ for a sufficiently small absolute constant c.

Recall from the proof of Proposition 3.1 that $W_k^{\mathrm{orc}} = U_k U_k^\top \widetilde{U}_k + \Delta^{\mathrm{orc}}$, where $\|P_{U_k^\perp} W_k^{\mathrm{orc}}\| \leqslant \|\Delta^{\mathrm{orc}}\| \leqslant \mathcal{E}_k^{\mathrm{orc}}$ and $\|P_{U_k^\perp} W_k^{\mathrm{orc}}\|_{2,\infty} \leqslant \|U\|_{2,\infty}\mathcal{E}_k^{\mathrm{orc}}$.

From the decomposition $\widehat{W}_k = W_k^{\mathrm{orc}} + (\widehat{W}_k - W_k^{\mathrm{orc}})$, we obtain

$$\|P_{U_k^\perp}\widehat{W}_k\| \leqslant \|P_{U_k^\perp} W_k^{\mathrm{orc}}\| + \|\widehat{W}_k - W_k^{\mathrm{orc}}\| \leqslant \mathcal{E}_k^{\mathrm{orc}} + \Delta.$$

Similarly, the $2 \to \infty$ norm satisfies

$$\|P_{U_k^\perp}\widehat{W}_k\|_{2,\infty} \leqslant \|P_{U_k^\perp} W_k^{\mathrm{orc}}\|_{2,\infty} + \|\widehat{W}_k - W_k^{\mathrm{orc}}\|_{2,\infty} \leqslant \|U\|_{2,\infty}\mathcal{E}_k^{\mathrm{orc}} + \Delta_{2,\infty}. \tag{H.8}$$

Next, we lower-bound $s_{\min}(\widehat{W}_k)$. We established previously that $s_{\min}(W_k^{\mathrm{orc}}) \geqslant 1/2$. By Weyl's inequality,

$$s_{\min}(\widehat{W}_k) \geqslant s_{\min}(W_k^{\mathrm{orc}}) - \|\widehat{W}_k - W_k^{\mathrm{orc}}\| \geqslant \frac{1}{2} - \Delta \geqslant \frac{1}{2} - \frac{1}{10} = \frac{2}{5}$$

since $\Delta \leqslant 1/10$.

Similar to the proof of Proposition 3.1, we get

$$\|\sin\angle(U_k, \widehat{U}_k^{\mathrm{db}})\| = \|P_{U_k^\perp}\widehat{U}_k^{\mathrm{db}}\| \leqslant \frac{\|P_{U_k^\perp}\widehat{W}_k\|}{\sigma_{\min}(\widehat{W}_k)} \leqslant \frac{5}{2}(\mathcal{E}_k^{\mathrm{orc}} + \Delta).$$

For the $2 \to \infty$ bound, we start with

$$\min_{O \in \mathbb{O}(k)} \|\widehat{U}_k^{\mathrm{db}} - U_k O\|_{2,\infty} \leqslant \|(I - P_{U_k})\widehat{U}_k^{\mathrm{db}}\|_{2,\infty} + \|U_k\|_{2,\infty}\|\sin\angle(U_k, \widehat{U}_k^{\mathrm{db}})\|^2.$$

Since $(I - P_{U_k})\widehat{U}_k^{\mathrm{db}} = P_{U_k^\perp}\widehat{W}_k(\widehat{W}_k^\top\widehat{W}_k)^{-1/2}$, we bound the first term using (H.8):

$$\|(I - P_{U_k})\widehat{U}_k^{\mathrm{db}}\|_{2,\infty} \leqslant \frac{\|P_{U_k^\perp}\widehat{W}_k\|_{2,\infty}}{\sigma_{\min}(\widehat{W}_k)} \leqslant \frac{5}{2}\left(\|U\|_{2,\infty}\mathcal{E}_k^{\mathrm{orc}} + \Delta_{2,\infty}\right).$$

Substituting this and the bound on $\|\sin\angle(U_k, \widehat{U}_k^{\mathrm{db}})\|^2$ back into the decomposition yields the desired bound.

I Proofs of Propositions 1.1 and 1.2

All proofs are carried out on the event $\mathcal{E}_D$, as mentioned at the beginning of Section 7.

I.1 Proof of Proposition 1.1

We first record the projected form of Corollary 4. Let $Q = I - UU^\top$. Since

$$Q(G - \Phi)U = (G - \Phi)U - U\big[U^\top(G - \Phi)U\big],$$

we have

$$\|Q(G-\Phi)U\|_{2,\infty} \leqslant \|(G-\Phi)U\|_{2,\infty} + \|U\|_{2,\infty}\|U^\top(G-\Phi)U\|.$$

Denote $\Xi(z) = G(z) - \Phi(z)$. Applying Corollary 4 with $B = U$ to the first term and Corollary 3 to the second term, we obtain that with probability at least $1 - n^{-D}$,

$$\sup_{z \in \mathcal{S}_{\mathrm{ra}}} |z|^2 \|Q\Xi(z)U\|_{2,\infty} \leqslant C\mathfrak{R}_D(U). \tag{I.1}$$

Here we used that $\mathcal{S}_{\mathrm{ra}} \subseteq \mathcal{S}_{\mathrm{out}}$ and $\mathfrak{m}_D \leqslant M_D$.

On the event $\|E\| \leqslant 3\sqrt{R_{\max}}$, by Weyl's inequality and our assumption on λ_k, we have

$$\widetilde{\lambda}_s \geqslant \lambda_k - 3\sqrt{R_{\max}} \geqslant C_{\mathrm{ra}}\sqrt{(D+r)R_{\max}\log n}.$$

Thus, $\widetilde{\lambda}_s \in \mathcal{S}_{\mathrm{ra}}$ for all $s \leqslant k$. Hence (I.1) applies at $z = \widetilde{\lambda}_s$ for $s \in [k]$.

We now follow the proof of Theorem 1. The only estimate that changes is the stochastic null-space term

$$\sqrt{\sum_{s=1}^{k} \widetilde{\lambda}_s^2 \left\|Q\Xi(\widetilde{\lambda}_s)U\right\|_{2,\infty}^2}$$

on the right-hand side of (7.3). By (I.1), we bound it by

$$\sqrt{\sum_{s=1}^{k} \widetilde{\lambda}_s^2 \left\|Q\Xi(\widetilde{\lambda}_s)U\right\|_{2,\infty}^2} \leqslant \sqrt{\sum_{s=1}^{k} \frac{C^2\mathfrak{R}_D(U)^2}{\widetilde{\lambda}_s^2}} \leqslant C\sqrt{k}\frac{\mathfrak{R}_D(U)}{\lambda_k},$$

where the last inequality follows from $\widetilde{\lambda}_s \geqslant c\lambda_k$. All remaining estimates are unchanged from the proof of Theorem 1, which proves (1.1).

I.2 Proof of Proposition 1.2

Recall that

$$\widetilde{Y}_k := \widetilde{U}_k\widetilde{\Lambda}_k^{1/2}, \qquad Y_k := U_k\Lambda_k^{1/2}.$$

We start with a deterministic decomposition. By inserting the projection operators $P_{U_k} = U_kU_k^\top$ and $P_{U_k^\perp} = I - P_{U_k}$, we have for any $O \in \mathbb{O}(k)$,

$$\widetilde{Y}_kO - Y_k = P_{U_k^\perp}\widetilde{Y}_kO + U_k(U_k^\top\widetilde{Y}_kO - \Lambda_k^{1/2}).$$

Thus,

$$\|\widetilde{U}_k\widetilde{\Lambda}_k^{1/2}O - U_k\Lambda_k^{1/2}\|_{2,\infty} \leqslant \|P_{U_k^\perp}\widetilde{Y}_k\|_{2,\infty} + \|U_k\|_{2,\infty}\|U_k^\top\widetilde{Y}_kO - \Lambda_k^{1/2}\|. \tag{I.2}$$

We first control the second term on the right-hand side of (I.2) by choosing a specific $O \in \mathbb{O}(k)$. For simplicity, denote $Z := U_k^\top\widetilde{U}_k\widetilde{\Lambda}_k^{1/2} = U_k^\top\widetilde{Y}_k$. Consider the polar decomposition $Z = HR$, where $H = (ZZ^\top)^{1/2}$ is a symmetric positive semi-definite matrix and $R \in \mathbb{O}(k)$. We set $O = R^\top$. Thus,

$$ZO = HRR^\top = H.$$

Consequently,

$$\|U_k^\top \widetilde{Y}_k O - \Lambda_k^{1/2}\| = \|ZO - \Lambda_k^{1/2}\| = \|H - \Lambda_k^{1/2}\|.$$

To bound this difference, noting that both H and Λ_k are positive semi-definite, we apply Theorem 6.2 from [63] to get

$$\|H - \Lambda_k^{1/2}\| \leqslant \frac{\|H^2 - \Lambda_k\|}{\lambda_{\min}(H) + \lambda_{\min}(\Lambda_k^{1/2})} \leqslant \frac{\|H^2 - \Lambda_k\|}{\sqrt{\lambda_k}},$$

where the final inequality follows from the fact that $\lambda_{\min}(H) \geqslant 0$.

Now we turn to bound $\|H^2 - \Lambda_k\|$. It follows from the definition of H that

$$H^2 = U_k^\top (\widetilde{U}_k \widetilde{\Lambda}_k \widetilde{U}_k^\top) U_k.$$

Since $\widetilde{A}_k = \widetilde{U}_k \widetilde{\Lambda}_k \widetilde{U}_k^\top$, we also have $H^2 = U_k^\top \widetilde{A}_k U_k$.

Plugging in $\Lambda_k = U_k^\top A U_k$, the difference can be written as

$$H^2 - \Lambda_k = U_k^\top (\widetilde{A}_k - A) U_k.$$

Let $\widetilde{U}_{>k} \in \mathbb{R}^{n \times (n-k)}$ be an orthonormal basis of $\mathrm{Span}(\widetilde{U}_k)^\perp$, formed by the remaining eigenvectors of $\widetilde{A}$, and let $\widetilde{\Lambda}_{>k}$ be the corresponding diagonal matrix of eigenvalues. Then

$$\widetilde{A} = \widetilde{U}_k \widetilde{\Lambda}_k \widetilde{U}_k^\top + \widetilde{U}_{>k} \widetilde{\Lambda}_{>k} \widetilde{U}_{>k}^\top := \widetilde{A}_k + \widetilde{A}_{>k}.$$

From $\widetilde{A} = A + E$ and the decomposition $\widetilde{A} = \widetilde{A}_k + \widetilde{A}_{>k}$, we substitute $\widetilde{A}_k - A = E - \widetilde{A}_{>k}$ to obtain

$$H^2 - \Lambda_k = U_k^\top E U_k - U_k^\top \widetilde{A}_{>k} U_k.$$

Applying the triangle inequality yields

$$\|H^2 - \Lambda_k\| \leqslant \|U_k^\top E U_k\| + \|U_k^\top \widetilde{A}_{>k} U_k\|.$$

For the second term, observe that $U_k^T \widetilde{A}_{>k} U_k = (U_k^\top \widetilde{U}_{>k}) \widetilde{\Lambda}_{>k} (\widetilde{U}_{>k}^\top U_k)$. Taking the operator norm, we get

$$\|U_k^\top \widetilde{A}_{>k} U_k\| \leqslant \|\widetilde{\Lambda}_{>k}\| \cdot \|U_k^\top \widetilde{U}_{>k}\|^2.$$

By Weyl's inequality, $\|\widetilde{\Lambda}_{>k}\| \leqslant \|\Lambda_{>k}\| + \|E\|$. Furthermore, $\|U_k^\top \widetilde{U}_{>k}\| = \|\sin\angle(U_k, \widetilde{U}_k)\|$. Substituting these bounds back into the triangle inequality yields

$$\|H^2 - \Lambda_k\| \leqslant \|U_k^\top E U_k\| + (\|\Lambda_{>k}\| + \|E\|)\|\sin\angle(U_k, \widetilde{U}_k)\|^2.$$

Finally, we obtain

$$\|U_k^\top \widetilde{U}_k \widetilde{\Lambda}_k^{1/2} O - \Lambda_k^{1/2}\| \leqslant \frac{1}{\sqrt{\lambda_k}} \left(\|U_k^\top E U_k\| + (\|\Lambda_{>k}\| + \|E\|)\|\sin\angle(U_k, \widetilde{U}_k)\|^2 \right).$$

Next, we bound $\|P_{U_k^\perp}\widetilde{Y}_k\|_{2,\infty} = \|P_{U_k^\perp}\widetilde{U}_k\widetilde{\Lambda}_k^{1/2}\|_{2,\infty}$. First, we obtain a deterministic bound by directly applying the component-wise estimates established in the proof of Proposition 7.1. Recall the decomposition $P_{U_k^\perp} = P_J + Q$, where $P_J = U_J U_J^\top$ with $U_J = (u_{k+1}, \ldots, u_r)$, and $Q = I - UU^\top$. By triangle inequality, we get

$$\begin{aligned}
\|P_{U_k^\perp}\widetilde{Y}_k\|_{2,\infty} &\leqslant \|P_J\widetilde{U}_k\widetilde{\Lambda}_k^{1/2}\|_{2,\infty} + \|Q\widetilde{U}_k\widetilde{\Lambda}_k^{1/2}\|_{2,\infty} \\
&\leqslant \|U\|_{2,\infty} \cdot \|U_J^\top\widetilde{U}_k\widetilde{\Lambda}_k^{1/2}\| + \|Q\widetilde{U}_k\widetilde{\Lambda}_k^{1/2}\|_{2,\infty} \\
&\leqslant \|U\|_{2,\infty} \cdot \|U_J^\top\widetilde{U}_k\widetilde{\Lambda}_k^{1/2}\|_F + \|Q\widetilde{U}_k\widetilde{\Lambda}_k^{1/2}\|_{2,\infty}.
\end{aligned}$$

Now we evaluate the rows of $Q\widetilde{U}_k\widetilde{\Lambda}_k^{1/2}$. The squared Euclidean norm of the l-th row is given by $\sum_{s=1}^k \widetilde{\lambda}_s(e_l^\top Q\widetilde{u}_s)^2$. In the proof of Proposition 7.1, we established the resolvent decomposition $e_l^\top Q\widetilde{u}_s = \mathsf{q}_{ls}^{(1)} + \mathsf{q}_{ls}^{(2)}$, where the individual terms are deterministically bounded by

$$|\mathsf{q}_{ls}^{(1)}| \leqslant \frac{8}{3}\|U\|_{2,\infty}\frac{\operatorname{osc}(\mathcal{R})}{\widetilde{\lambda}_s^2} \qquad \text{and} \qquad |\mathsf{q}_{ls}^{(2)}| \leqslant 2\widetilde{\lambda}_s\|e_l^\top Q\Xi(\widetilde{\lambda}_s)U\|_2.$$

Next, we multiply these squared bounds by the weight $\widetilde{\lambda}_s$ and sum over $s \in [k]$ to get

$$\sqrt{\sum_{s=1}^k \widetilde{\lambda}_s|\mathsf{q}_{ls}^{(1)}|^2} \leqslant \frac{8}{3}\|U\|_{2,\infty}\operatorname{osc}(\mathcal{R})\sqrt{\sum_{s=1}^k \frac{1}{\widetilde{\lambda}_s^3}} \leqslant 3\sqrt{k}\|U\|_{2,\infty}\frac{\operatorname{osc}(\mathcal{R})}{\lambda_k^{3/2}},$$

and

$$\sqrt{\sum_{s=1}^k \widetilde{\lambda}_s|\mathsf{q}_{ls}^{(2)}|^2} \leqslant 2\sqrt{\sum_{s=1}^k \widetilde{\lambda}_s^3\|e_l^\top Q\Xi(\widetilde{\lambda}_s)U\|_2^2}.$$

Combining these estimates via the triangle inequality yields the final deterministic bound for the weighted orthogonal projection:

$$\begin{aligned}
\|P_{U_k^\perp}\widetilde{Y}_k\|_{2,\infty} \leqslant{}& \|U\|_{2,\infty}\|U_J^\top\widetilde{U}_k\widetilde{\Lambda}_k^{1/2}\|_F + 3\sqrt{k}\|U\|_{2,\infty}\frac{\operatorname{osc}(\mathcal{R})}{\lambda_k^{3/2}} \\
&+ 2\sqrt{\sum_{s=1}^k \widetilde{\lambda}_s^3 \max_{1\leqslant i\leqslant n}\|e_i^\top Q\Xi(\widetilde{\lambda}_s)U\|_2^2}.
\end{aligned} \tag{I.3}$$

It remains to bound

$$\|U_J^\top\widetilde{U}_k\widetilde{\Lambda}_k^{1/2}\|_F = \sqrt{\sum_{s=1}^k \widetilde{\lambda}_s\|U_J^\top\widetilde{u}_s\|^2}.$$

For $J \neq \varnothing$, from the bound (7.5), we get

$$\begin{aligned}
\sqrt{\widetilde{\lambda}_s}\|U_J^\top\widetilde{u}_s\| \leqslant{}& \frac{4}{\Delta_{\mathcal{J}}^{\Phi}(\widetilde{\lambda}_s)}\left(\frac{\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\|}{\widetilde{\lambda}_s^{1/2}} + \frac{13}{2}\frac{R_{\max}}{\widetilde{\lambda}_s^{5/2}}\operatorname{osc}(\mathcal{R}) + \widetilde{\lambda}_s^{5/2}\|U^\top\Xi(\widetilde{\lambda}_s)U\|\right) \\
&+ 2\sqrt{3}\left(\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\widetilde{\lambda}_s^{3/2}} + \frac{13}{2}\frac{R_{\max}}{\widetilde{\lambda}_s^{7/2}}\operatorname{osc}(\mathcal{R}) + \widetilde{\lambda}_s^{3/2}\|U^\top\Xi(\widetilde{\lambda}_s)U\|\right)
\end{aligned}$$

Substituting $\Delta^{\Phi}_{\mathcal{J}}(\widetilde{\lambda}_s) \geqslant \frac{1}{2}\delta_k$ and $\|U^\top \Xi(\widetilde{\lambda}_s) U\| \leqslant M/\widetilde{\lambda}_s^2$ into the above bound, we obtain

$$\begin{aligned}\sqrt{\widetilde{\lambda}_s}\|U_J^\top \widetilde{u}_s\| \leqslant \frac{8}{\delta_k}\frac{\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\|}{\widetilde{\lambda}_s^{1/2}} + \frac{52}{\delta_k}\frac{R_{\max}}{\widetilde{\lambda}_s^{5/2}}\mathrm{osc}(\mathcal{R}) + \frac{8M\widetilde{\lambda}_s^{1/2}}{\delta_k} \\ + 2\sqrt{3}\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\widetilde{\lambda}_s^{3/2}} + 13\sqrt{3}\frac{R_{\max}}{\widetilde{\lambda}_s^{7/2}}\mathrm{osc}(\mathcal{R}) + 2\sqrt{3}\frac{M}{\widetilde{\lambda}_s^{1/2}}.\end{aligned}$$

To compute the Frobenius norm $\|U_J^\top \widetilde{U}_k \widetilde{\Lambda}_k^{1/2}\|_F = \big\|\big(\sqrt{\widetilde{\lambda}_s}\|U_J^\top \widetilde{u}_s\|\big)_{s=1}^k\big\|_2$, we sum over all k components and use $\frac{3}{2}\lambda_s \geqslant \widetilde{\lambda}_s \geqslant \frac{99}{100}\lambda_k$, which yields

$$\begin{aligned}&\|U_J^\top \widetilde{U}_k \widetilde{\Lambda}_k^{1/2}\|_F \\ &\quad\leqslant \frac{8.1\sqrt{k}}{\delta_k \lambda_k^{1/2}}\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\| + \frac{3.6\sqrt{k}}{\lambda_k^{3/2}}\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\| + 78\sqrt{k}\frac{R_{\max}\mathrm{osc}(\mathcal{R})}{\delta_k \lambda_k^{5/2}} + \frac{10M}{\delta_k}\sqrt{\sum_{s=1}^k \lambda_s} + \frac{3.5\sqrt{k}}{\lambda_k^{1/2}}M.\end{aligned}$$

By (I.1), for every $s \leqslant k$,

$$\max_{1\leqslant i\leqslant n}\|e_i^\top Q\Xi(\widetilde{\lambda}_s)U\|_2 \leqslant C\frac{\mathfrak{R}_D(U)}{\widetilde{\lambda}_s^2}.$$

Thus,

$$2\sqrt{\sum_{s=1}^k \widetilde{\lambda}_s^3 \max_{1\leqslant i\leqslant n}\|e_i^\top Q\Xi(\widetilde{\lambda}_s)U\|_2^2} \leqslant C\mathfrak{R}_D(U)\sqrt{\sum_{s=1}^k \frac{1}{\widetilde{\lambda}_s}} \leqslant C\sqrt{k}\frac{\mathfrak{R}_D(U)}{\sqrt{\widetilde{\lambda}_k}} < C\sqrt{k}\frac{\mathfrak{R}_D(U)}{\sqrt{\lambda_k}}.$$

Finally, combining these estimates into (I.3) and noting that $\frac{13}{2}\widetilde{\lambda}_k^{-3/2} \leqslant 7\lambda_k^{-3/2}$, we obtain

$$\begin{aligned}\|P_{U_k^\perp}\widetilde{Y}_k\|_{2,\infty} &\leqslant \|U\|_{2,\infty}\left(\frac{8.1\sqrt{k}}{\delta_k\lambda_k^{1/2}}\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\| + \frac{3.6\sqrt{k}}{\lambda_k^{3/2}}\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\| + 78\sqrt{k}\frac{R_{\max}\mathrm{osc}(\mathcal{R})}{\delta_k\lambda_k^{5/2}}\right)\mathbf{1}_{\{k<r\}} \\ &\quad + \|U\|_{2,\infty}\left(\frac{10M}{\delta_k}\sqrt{\sum_{s=1}^k \lambda_s} + \frac{3.5\sqrt{k}}{\lambda_k^{1/2}}M\right)\mathbf{1}_{\{k<r\}} + 3\sqrt{k}\|U\|_{2,\infty}\frac{\mathrm{osc}(\mathcal{R})}{\lambda_k^{3/2}} + C\sqrt{k}\frac{\mathfrak{R}_D(U)}{\lambda_k^{1/2}}.\end{aligned}$$

Recall that

$$\mathcal{B}_k := \frac{10\sqrt{k}}{\delta_k\lambda_k}\left(\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\| + 8\frac{R_{\max}}{\lambda_k^2}\mathrm{osc}(\mathcal{R})\right) + 4\sqrt{k}\frac{\|\mathcal{V}_{\mathcal{N}\mathcal{K}}\|}{\lambda_k^2}.$$

We simplify the bound to

$$\begin{aligned}\|P_{U_k^\perp}\widetilde{Y}_k\|_{2,\infty} \leqslant &\|U\|_{2,\infty}\sqrt{\lambda_k}\mathcal{B}_k\mathbf{1}_{\{k<r\}} + \|U\|_{2,\infty}\frac{10M}{\delta_k}\sqrt{\sum_{s=1}^k \lambda_s \mathbf{1}_{\{k<r\}}} \\ &+ C\sqrt{k}\frac{\mathfrak{R}_D(U)}{\lambda_k^{1/2}} + 3\sqrt{k}\|U\|_{2,\infty}\frac{\mathrm{osc}(\mathcal{R})}{\lambda_k^{3/2}}.\end{aligned} \tag{I.4}$$

Now, from (I.2), we get

$$
\begin{aligned}
&\|\widetilde{U}_k\widetilde{\Lambda}_k^{1/2}O - U_k\Lambda_k^{1/2}\|_{2,\infty} \\
&\qquad \leqslant \|P_{U_k^\perp}\widetilde{Y}_k\|_{2,\infty} + \frac{\|U_k\|_{2,\infty}}{\sqrt{\lambda_k}}\left[\|U_k^\top E U_k\| + (\|\Lambda_{>k}\| + \|E\|)\|\sin\angle(U_k,\widetilde{U}_k)\|^2\right],
\end{aligned}
$$

where $\|P_{U_k^\perp}\widetilde{Y}_k\|_{2,\infty}$ is bounded by (I.4). Applying the bound for $\|\sin\angle(U_k,\widetilde{U}_k)\|$ from Theorem 1 completes the proof.

J Plug-in de-biasing implementation and limitations

To keep the discussion simple, assume in this paragraph that r and r_+ are known and the variance profile Σ is known, so that $\Phi(z)$ is computable from the QVE..

Define

$$
\widetilde{U}_{\mathcal{J}} := (\widetilde{u}_{k+1},\ldots,\widetilde{u}_{r_+}), \qquad \widetilde{U}_{\mathcal{I}} := (\widetilde{u}_1,\ldots,\widetilde{u}_k,\widetilde{u}_{r_++1},\ldots,\widetilde{u}_r),
$$

with corresponding diagonal eigenvalue matrices $\widetilde{\Lambda}_{\mathcal{J}}$ and $\widetilde{\Lambda}_{\mathcal{I}}$. Also write

$$
\widetilde{U} := (\widetilde{u}_1,\ldots,\widetilde{u}_r), \qquad P_{\widetilde{U}} := \widetilde{U}\widetilde{U}^\top.
$$

If $\mathcal{J} = \varnothing$, no correction is needed and we set $\widehat{U}_k^{\mathrm{db}} := \widetilde{U}_k$. Otherwise, for $s \in [k]$, define

$$
\widehat{\mathcal{T}}_s := \left(I - \widetilde{U}_{\mathcal{J}}^\top\Phi(\widetilde{\lambda}_s)\widetilde{U}_{\mathcal{J}}\widetilde{\Lambda}_{\mathcal{J}}\right)^{-1}\widetilde{U}_{\mathcal{J}}^\top\Phi(\widetilde{\lambda}_s)\widetilde{U}_{\mathcal{I}}\widetilde{\Lambda}_{\mathcal{I}},
$$

whenever the inverse exists. The plug-in corrected vector is

$$
\widehat{w}_s := P_{\widetilde{U}}\widetilde{u}_s - \widetilde{U}_{\mathcal{J}}\widehat{\mathcal{T}}_s\widetilde{U}_{\mathcal{I}}^\top\widetilde{u}_s, \qquad s = 1,\ldots,k.
$$

Note that for $s \leqslant k$, $P_{\widetilde{U}}\widetilde{u}_s = \widetilde{u}_s$. Finally, set

$$
\widehat{W}_k := (\widehat{w}_1,\ldots,\widehat{w}_k), \qquad \widehat{U}_k^{\mathrm{db}} := \mathrm{orth}(\widehat{W}_k).
$$

To compare this estimator with the oracle estimator, define the correction discrepancy

$$
\mathfrak{D}_k := \left(\widetilde{U}_{\mathcal{J}}\widehat{\mathcal{T}}_s\widetilde{U}_{\mathcal{I}}^\top\widetilde{u}_s - U_{\mathcal{J}}\mathcal{T}_sU_{\mathcal{I}}^\top\widetilde{u}_s\right)_{s=1}^k.
$$

Then the difference between the plug-in and oracle matrices is given by

$$
\widehat{W}_k - W_k^{\mathrm{orc}} = (I - P_U)\widetilde{U}_k - \mathfrak{D}_k.
$$

Consequently, the accuracy of the plug-in correction is governed by the norms of this difference. As shown in (B.2) below, $\|(I-P_U)\widetilde{U}_k\| \leqslant 2\|E\|/\lambda_k$. On the event that

$$
\mathcal{E}_k^{\mathrm{orc}} + \frac{\|E\|}{\lambda_k} + \|\mathfrak{D}_k\| \leqslant c
$$

for a sufficiently small absolute constant $c > 0$, the same argument as in Proposition 3.1 gives

$$\|\sin\angle(\widehat{U}_k^{\mathrm{db}}, U_k)\| \leqslant C\left(\mathcal{E}_k^{\mathrm{orc}} + \frac{\|E\|}{\lambda_k} + \|\mathfrak{D}_k\|\right). \tag{J.1}$$

Moreover,

$$\begin{aligned}\min_{O\in\mathbb{O}(k)} \|\widehat{U}_k^{\mathrm{db}} - U_k O\|_{2,\infty} \leqslant C\Big[&\|U\|_{2,\infty}\mathcal{E}_k^{\mathrm{orc}} + \|(I-P_U)\widetilde{U}_k\|_{2,\infty} + \|\mathfrak{D}_k\|_{2,\infty} \\ &+ \|U_k\|_{2,\infty}\left(\mathcal{E}_k^{\mathrm{orc}} + \frac{\|E\|}{\lambda_k} + \|\mathfrak{D}_k\|\right)^2\Big].\end{aligned} \tag{J.2}$$

The proofs of (J.1) and (J.2) are given in Appendix H.3 for completeness.

REMARK J.1 (Limitation and useful regimes)**.** *The plug-in correction $\widehat{U}_k^{\mathrm{db}}$ is not a universal replacement for the raw empirical eigenspace $\widetilde{U}_k$. The term $\mathfrak{D}_k$ measures the error in estimating the oracle correction from the empirical lower-positive $\mathcal{J}$-block. This term could be large in two situations. First, if the lower-positive block $\mathcal{J}$ contains weak or poorly separated spikes, then the empirical block $\widetilde{U}_{\mathcal{J}}$ may not estimate $U_{\mathcal{J}}$ accurately. Second, when eigenvalues are tightly clustered (δ_k is small), the matrices $I - U_{\mathcal{J}}^\top \Phi(\widetilde{\lambda}_s) U_{\mathcal{J}} \Lambda_{\mathcal{J}}$ become ill-conditioned. Inverting them will make $\|\mathfrak{D}_k\|$ blow up.*

Thus the plug-in estimator is useful only when

$$\|\mathfrak{D}_k\| \lesssim \mathcal{E}_k^{\mathrm{orc}}$$

and the removed bias is larger than the residual terms

$$\frac{\sqrt{k}}{\delta_k \lambda_k}\left(\|\mathcal{V}_{\mathcal{J}\mathcal{I}}\| + \frac{R_{\max}}{\lambda_k^2}\mathrm{osc}(\mathcal{R})\right) \gg \mathcal{E}_k^{\mathrm{orc}} + \frac{\|E\|}{\lambda_k}.$$

This condition does not require the geometric bias to vanish. It may hold, for instance, when $\mathcal{J}$ contains strong, well-separated spikes that can be accurately estimated, even if the variance profile $\mathcal{R}$ has a large coupling between $U_{\mathcal{J}}$ and $U_{\mathcal{I}}$. A systematic analysis of optimal fully data-driven bias correction is left for future work.

K Proofs for the applications

K.1 Proofs for the DCSBM application

We first record the population geometry. The calculation is standard; see, for example, [74, 69]. Let

$$S := Z^\top \Theta^2 Z = \mathrm{diag}(s_1, \ldots, s_r) \quad \text{and} \quad X_0 := \Theta Z S^{-1/2}.$$

Note that S has rank r and $X_0^\top X_0 = I_r$.

LEMMA K.1 (Population geometry of the DCSBM)**.** *Write the eigen-decomposition $S^{1/2} B S^{1/2} = V \Lambda V^\top$ where $V \in \mathbb{O}(r)$. Then the eigenspace of P may be represented as $U = X_0 V$, and*

$$e_i^\top U = \mu_i e_{g(i)}^\top V, \qquad \frac{e_i^\top U}{\|e_i^\top U\|} = e_{g(i)}^\top V. \tag{K.1}$$

Consequently,

$$\|e_i^\top U\| = \mu_i, \qquad \|U\|_{2,\infty} = \mu_+, \qquad \min_i \|e_i^\top U\|_2 = \mu_-,$$

and the normalized population centers of two distinct communities are separated by distance $\sqrt{2}$.

Proof. Note that

$$P = \Theta Z B Z^\top \Theta = X_0\big(S^{1/2} B S^{1/2}\big) X_0^\top = (X_0 V)\Lambda(X_0 V)^\top.$$

Thus we may take $U = X_0 V$. If $g(i) = a$, then

$$e_i^\top X_0 = \frac{\theta_i}{\sqrt{s_a}} e_a^\top = \mu_i e_a^\top,$$

which proves (K.1). Since the rows of V are orthonormal, $\|e_a^\top V - e_b^\top V\| = \sqrt{2}$ for $a \neq b$. $\square$

We use the following specialization of the refined full-signal perturbation bound from the main results. It is obtained by applying Corollary 1 to the positive and negative eigenspaces separately (see Remark 1.3). Recall that $t_{D,r} = (D + r)\log n$.

PROPOSITION K.1 (Perturbation input)**.** *Let $P = U\Lambda U^\top$ be a symmetric rank-r probability matrix and let $\lambda_\star := \min_{1\leqslant j\leqslant r} |\lambda_j(P)|$. For centered Bernoulli noise, suppose $R_{\max} \geqslant C_D t_{D,r}$ and $\lambda_\star \geqslant C_D\sqrt{t_{D,r} R_{\max}}$. Then, with probability at least $1 - n^{-D}$,*

$$\min_{O\in\mathbb{O}(r)} \big\|\widetilde{U}_r O - U\big\|_{2,\infty} \leqslant C_D\left[\sqrt{r}\,\frac{\sqrt{t_{D,r}}\,\Gamma(U) + t_{D,r}\|U\|_{2,\infty}}{\lambda_\star} + (\sqrt{r}+1)\|U\|_{2,\infty}\frac{R_{\max}}{\lambda_\star^2}\right].$$

Here $\Gamma(U) := \max_i \big(\sum_j P_{ij}(1-P_{ij})\|e_j^\top U\|^2\big)^{1/2}$.

Now we are in the position to prove Theorem 2.

Proof of Theorem 2. By Lemma K.1, $\|U\|_{2,\infty} = \mu_+$ and $\Gamma(U) = \Gamma_{\mathrm{dc}}$. As given in (2.1), Assumption 2.1 suggests

$$\tau_n = \lambda_\star = \min_{1\leqslant j\leqslant r} |\lambda_j(P)| \geqslant \sigma_{\min}(B) s_-. \tag{K.2}$$

Since P is symmetric and entrywise nonnegative,

$$\|P\| \leqslant d_{\max} = \max_i \sum_j P_{ij}.$$

Moreover, $p_{\max} = o(1)$ implies

$$R_i = \sum_j P_{ij}(1-P_{ij}) \asymp d_i \quad \text{uniformly in } i, \qquad R_{\max} \asymp d_{\max}, \qquad \operatorname{osc}(\mathcal{R}) \leqslant R_{\max}. \tag{K.3}$$

For the centered Bernoulli entries and fixed D, we have $L^{\mathrm{ra}}_{D,r} \lesssim 1$ and $M_D \lesssim_D t_{D,r}$, and therefore (recall $\mathfrak{R}_D(U)$ in (1.6))

$$\mathfrak{R}_D(U) \lesssim_D \sqrt{t_{D,r}}\,\Gamma_{\mathrm{dc}} + t_{D,r}\mu_+. \tag{K.4}$$

By (2.3), (K.2) and (K.3), the conditions in Proposition K.1 are satisfied.

Applying Proposition K.1 and using (K.2)–(K.4) yields

$$\min_{O\in\mathbb{O}(r)} \|\widetilde{U}_r O - U\|_{2,\infty} \leqslant C_{2,D}\cdot \varepsilon_{D,n}(\tau_n),$$

which proves (2.4).

It remains to prove exact recovery. Let O_* attain the bound (2.4) and write

$$\widehat{u}_i := e_i^\top \widetilde{U}_r O_*, \qquad u_i := e_i^\top U.$$

If (2.5) holds with c_D sufficiently small, then every $\widehat{u}_i$ is nonzero and

$$\left\| \frac{\widehat{u}_i}{\|\widehat{u}_i\|} - \frac{u_i}{\|u_i\|} \right\| \leqslant \frac{2\|\widehat{u}_i - u_i\|}{\|u_i\|} \leqslant \delta_D < \sqrt{2}/4.$$

Thus two normalized empirical rows from the same community are at distance at most $2\delta_D$, whereas rows from distinct communities are at distance at least $\sqrt{2}-2\delta_D$.

We now verify the farthest-first step. Suppose the selected centers represent a proper subset of the communities. Every point from a represented community lies within $2\delta_D$ of one of the selected centers, while every point from an unrepresented community has distance at least $\sqrt{2}-2\delta_D > 2\delta_D$ from all selected centers. Hence the next farthest point belongs to an unrepresented community. Induction shows that the r selected centers contain exactly one representative from each community. The same distance comparison shows that nearest-center assignment is exact. A common orthogonal transformation preserves row norms and Euclidean distances, so the Procrustes alignment does not affect the algorithm. □

Proof of Corollary 2. Under the assumptions of the corollary,

$$s_a \asymp n_a\theta_0^2, \qquad s_- \asymp d_-, \qquad \mu_+ \asymp m_n^{-1/2}, \qquad \mu_- \asymp N_n^{-1/2} \tag{K.5}$$

where the constants depend only on c_θ and C_θ.

Because r is fixed, B is entrywise nonnegative, and its singular values are bounded above and below, every row and column of B has total mass bounded above and below by positive constants. Consequently,

$$d_{\max} \asymp N_n\theta_0^2 = \kappa_n d_-. \tag{K.6}$$

Also, $p_{\max} \lesssim \theta_0^2 = o(1)$, so $R_{\max} \asymp d_{\max}$.

We next evaluate the Γ_{dc}. If i belongs to community a, then Lemma K.1 gives

$$\sum_j P_{ij}(1-P_{ij})\|e_j^\top U\|_2^2 \leqslant \theta_i \sum_{b=1}^r B_{ab}\frac{\sum_{j:g(j)=b}\theta_j^3}{s_b} \lesssim \theta_0^2\sum_{b=1}^r B_{ab} \lesssim \sigma_{\max}(B)\theta_0^2.$$

Thus

$$\Gamma_{\mathrm{dc}} \lesssim \theta_0\sqrt{\sigma_{\max}(B)}. \tag{K.7}$$

Since r is fixed, $t_{D,r} \asymp_D \log n$. By (K.5),

$$\tau_n \geqslant \sigma_{\min}(B)s_- \asymp \sigma_{\min}(B)d_-. \tag{K.8}$$

The condition (2.7) implies the two conditions in Theorem 2. Indeed, using $\sigma_{\max}(B) \geqslant \sigma_{\min}(B)$, $\kappa_n \geqslant 1$, and the lower bound in (K.6),

$$d_{\max} \gtrsim \sigma_{\min}(B)\kappa_n d_- \gtrsim_D \log n.$$

Using the upper bound in (K.6),

$$\sqrt{d_{\max}\log n} \lesssim \sqrt{\sigma_{\max}(B)\kappa_n d_- \log n} \lesssim_D \sigma_{\min}(B)d_-,$$

and (2.7) also gives $\log n \lesssim_D \sigma_{\max}(B)d_-$. Hence

$$\tau_n \gtrsim \sigma_{\max}(B)d_- \gtrsim_D \sqrt{d_{\max}\log n}.$$

Substituting (K.5), (K.6),(K.7), and (K.8) into (2.2) yields

$$\begin{aligned}
\varepsilon_{D,n}(\tau_n) &\lesssim_D \frac{\theta_0\sqrt{\sigma_{\max}(B)\log n}}{\sigma_{\min}(B)d_-} + \frac{\log n}{\sigma_{\min}(B)\sqrt{m_n}\,d_-} + \frac{\sigma_{\max}(B)\kappa_n}{\sigma_{\min}(B)^2\sqrt{m_n}\,d_-} \\
&\lesssim_D \frac{1}{\sqrt{N_n}}\left[\frac{\sqrt{\sigma_{\max}(B)}}{\sigma_{\min}(B)}\sqrt{\frac{\kappa_n\log n}{d_-}} + \frac{1}{\sigma_{\min}(B)}\frac{\sqrt{\kappa_n}\log n}{d_-} + \frac{\sigma_{\max}(B)}{\sigma_{\min}(B)^2}\frac{\kappa_n^{3/2}}{d_-}\right].
\end{aligned}$$

By our assumption on d_-, we have

$$\frac{1}{\sigma_{\min}(B)}\frac{\sqrt{\kappa_n}\log n}{d_-} \lesssim \frac{\sqrt{\sigma_{\max}(B)}}{\sigma_{\min}(B)}\sqrt{\frac{\kappa_n\log n}{d_-}}.$$

Then (2.8) follows from Theorem 2.

Finally, divide the last display by $\mu_- \asymp N_n^{-1/2}$. The two terms on the right-hand side of (2.8) are made sufficiently small by (2.7). After increasing $C_{1,D}$, we therefore have $\varepsilon_{D,n}(\tau_n) \leqslant c_D\mu_-$. Exact recovery follows from Theorem 2.

□

K.2 Proofs for the GRDPG application

Throughout this section, the rank $r = p + q$ is fixed, and all implicit constants may depend on D, r, and the constants in Assumption 2.2, but not on n.

We first record the estimation of related parameters under Assumption 2.2. Recall $d_n = n\rho_n$.

LEMMA K.2. *Suppose Assumption 2.2 holds. Then $P = U\Lambda U^\top$ has rank r, with exactly p positive and q negative nonzero eigenvalues. Denote $\lambda_{\text{sig}} := \min_{1\leqslant a\leqslant r}|\lambda_a(P)|$. Then*

$$\lambda_{\text{sig}} \asymp d_n, \qquad \|P\| \asymp d_n, \qquad \|U\|_{2,\infty} \lesssim n^{-1/2}. \tag{K.9}$$

The row-variance profile satisfies

$$R_i \asymp d_n \quad \text{for all } i \in [n], \qquad R_{\max} \asymp d_n, \qquad \operatorname{osc}(\mathcal{R}) \lesssim d_n. \tag{K.10}$$

In addition,

$$\Gamma(U)^2 := \max_{i\in[n]}\sum_{j=1}^n P_{ij}(1-P_{ij})\|e_j^\top U\|_2^2 \lesssim \frac{d_n}{n}. \tag{K.11}$$

Finally, the matrix Q_0 in (2.10) *may be chosen so that $\|Q_0\| + \|Q_0^{-1}\| \lesssim 1$.*

Proof. For simplicity, set $S := Y^\top Y = n\Delta_n$. Since Δ_n is positive definite,

$$P = (YS^{-1/2})\big(\rho_n S^{1/2} J S^{1/2}\big)(YS^{-1/2})^\top,$$

and $YS^{-1/2}$ has orthonormal columns. Hence the nonzero eigenvalues of P are the eigenvalues of

$$\rho_n S^{1/2} J S^{1/2} = d_n \Delta_n^{1/2} J \Delta_n^{1/2}.$$

By Sylvester's law of inertia, $\Delta_n^{1/2} J \Delta_n^{1/2}$, and therefore P, has exactly p positive and q negative eigenvalues. Moreover,

$$\|\Delta_n^{1/2} J \Delta_n^{1/2}\| \lesssim 1, \qquad \|(\Delta_n^{1/2} J \Delta_n^{1/2})^{-1}\| = \|\Delta_n^{-1/2} J \Delta_n^{-1/2}\| \lesssim 1.$$

Since $\Delta_n^{1/2} J \Delta_n^{1/2}$ is symmetric, all its eigenvalues have magnitudes bounded above and below by positive constants. This proves the first two relations in (K.9). For the last relation in (K.9), note that if $V \in \mathbb{O}(r)$ diagonalizes $d_n \Delta_n^{1/2} J \Delta_n^{1/2}$, then $U = YS^{-1/2}V$. Consequently,

$$\|e_i^\top U\| \leqslant \|y_i\| \cdot \|S^{-1/2}\| \lesssim n^{-1/2}.$$

Now we prove (K.10). The expected degree of vertex i is $\sum_{j=1}^n P_{ij} = d_n y_i^\top J \bar{y}_n \asymp d_n$. Since $0 \leqslant P_{ij} \leqslant 1 - c$,

$$c \sum_{j=1}^n P_{ij} \leqslant R_i = \sum_{j=1}^n P_{ij}(1 - P_{ij}) \leqslant \sum_{j=1}^n P_{ij}.$$

It also gives (K.11): $\Gamma(U)^2 \leqslant \|U\|_{2,\infty}^2 R_{\max} \lesssim \frac{d_n}{n}$.

Finally, since X and $\sqrt{\rho_n} Y$ have full column rank and the same column space, there is an invertible matrix Q_0 such that

$$XQ_0 = \sqrt{\rho_n} Y.$$

Substituting this identity into $XJX^\top = \rho_n Y J Y^\top$ and using the full column rank of X, we get $Q_0 J Q_0^\top = J$, or equivalently $Q_0^\top J Q_0 = J$. Thus $Q_0 \in \mathcal{O}(p, q)$. Furthermore,

$$Q_0^\top |\Lambda| Q_0 = \rho_n Y^\top Y = d_n \Delta_n.$$

Both $|\Lambda|$ and $d_n \Delta_n$ have eigenvalues comparable to d_n. This suggests that $\|Q_0\| + \|Q_0^{-1}\| \lesssim 1$. $\square$

We now prove the full-embedding Theorem 3. Let $E := \widetilde{A} - P$. Recall $\mathfrak{R}_D(U)$ from (1.6). In bounded-Bernoulli regime, we have, with probability at least $1 - n^{-D}$, that

$$\|E\| \lesssim_D \sqrt{d_n}, \qquad M_D \lesssim_D \log n, \qquad \mathfrak{R}_D(U) \lesssim_D \sqrt{\frac{d_n \log n}{n}} + \frac{\log n}{\sqrt{n}}. \tag{K.12}$$

Here the last relation follows from (K.11) and $\|U\|_{2,\infty} \lesssim n^{-1/2}$.

Proof of Theorem 3. By Lemma K.2, every nonzero signal eigenvalue has magnitude comparable to d_n, whereas $R_{\max} \asymp d_n$. Hence, after choosing $C_{1,D}$ sufficiently large, the condition $d_n \geqslant C_{1,D} \log n$ verifies the signal strength condition in Proposition 1.2.

We apply Proposition 1.2 separately to the complete positive signal block of P and to the complete negative signal block through $-P$. Combining the two orthogonal alignments yields

$$\min_{W \in \mathbb{O}(p) \oplus \mathbb{O}(q)} \|\widetilde{X}_r W - X\|_{2,\infty} \lesssim_D \frac{\mathfrak{R}_D(U)}{\sqrt{d_n}} + \|U\|_{2,\infty} \frac{R_{\max} + \operatorname{osc}(\mathcal{R})}{d_n^{3/2}} + \frac{\|U\|_{2,\infty}}{\sqrt{d_n}} \eta_{\text{full}} \tag{K.13}$$

where $\eta_{\text{full}} := \eta_+ + \eta_-$ with

$$\eta_\pm := \|U_\pm^\top E U_\pm\| + \big(\|\Lambda_\mp\| + \|E\|\big) \|\sin\angle(U_\pm, \widetilde{U}_\pm)\|^2.$$

By Davis-Kahan theorem,

$$\|\sin\angle(U_\pm, \widetilde{U}_\pm)\| \lesssim_D \frac{\|E\|}{d_n} \lesssim_D d_n^{-1/2}.$$

Consequently,

$$\eta_{\text{full}} \lesssim_D \|E\| + d_n \max_{\varsigma \in \{+,-\}} \|\sin\angle(U_\varsigma, \widetilde{U}_\varsigma)\|^2 \lesssim_D \sqrt{d_n}.$$

Substituting this estimate together with (K.9) and (K.12) into (K.13) yields

$$\min_{W \in \mathbb{O}(p) \oplus \mathbb{O}(q)} \|\widetilde{X}_r W - X\|_{2,\infty} \lesssim_D \sqrt{\frac{\log n}{n}} + \frac{\log n}{\sqrt{n d_n}} + \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n d_n}} \lesssim_D \sqrt{\frac{\log n}{n}},$$

where we used $d_n \geqslant C_{1,D} \log n$ to get the last inequality. This proves (2.11).

Moreover, by Weyl's inequality, the r eigenvalues of $\widetilde{A} = P + E$ of largest magnitude are the signal outliers, with p positive and q negative eigenvalues. Indeed, the r nonzero eigenvalues of P have magnitudes of order d_n, whereas $\|E\| \lesssim_D \sqrt{d_n}$.

It remains to pass from X to the latent positions Y. Let $W \in \mathbb{O}(p) \oplus \mathbb{O}(q)$ attaining (2.11), and set $Q := W Q_0$. Both W and Q_0 preserve J, so $Q \in \mathbb{O}(p, q)$. By $X Q_0 = \sqrt{\rho_n}\, Y$ from (2.10),

$$\widetilde{X}_r Q - \sqrt{\rho_n}\, Y = (\widetilde{X}_r W - X) Q_0.$$

Hence, by (2.11) and Lemma K.2,

$$\|\widetilde{X}_r Q - \sqrt{\rho_n}\, Y\|_{2,\infty} \lesssim_D \sqrt{\frac{\log n}{n}}.$$

Dividing by $\sqrt{\rho_n} = \sqrt{\frac{d_n}{n}}$ proves (2.12). The consistency statement follows immediately. □

Finally, we provide the proof of Proposition 2.1.

Proof of Proposition 2.1. By Lemma K.2 and Weyl's inequality, for $\varsigma \in \{+,-\}$ and every $a \leqslant k_\varsigma$, we have $\widetilde{\lambda}_a^\varsigma \asymp \sqrt{d_n}$ and thus,

$$\tau_\varsigma := \Big(\sum_{a \leqslant k_\varsigma} \widetilde{\lambda}_a^\varsigma \Big)^{1/2} \asymp \sqrt{d_n}.$$

The assumptions $d_n \geqslant C_{1,D} \log n$ and $\delta_n \geqslant C_{1,D} \log n$ imply the signal strength and gap conditions in the corresponding signed versions of Proposition 1.2.

We directly apply Proposition 1.2 to the retained positive block of P and to the retained negative block through $-P$, and combine the two blockwise alignments. Note that the deterministic terms are precisely GBias_{k_+,k_-} defined in (2.13). Thus

$$\min_{W\in\mathbb{O}(k_+)\oplus\mathbb{O}(k_-)} \|\widetilde{X}_k W - X_k\|_{2,\infty} \leqslant \mathrm{GBias}_{k_+,k_-} + C_D\left[\frac{\mathfrak{R}_D(U)}{\sqrt{d_n}} + \sqrt{\frac{d_n}{n}}\log n\left(\frac{1}{\delta_{+,n}} + \frac{1}{\delta_{-,n}}\right)\right.$$
$$\left. + \frac{1}{\sqrt{nd_n}} + \sum_{\varsigma\in\{+,-\}} \frac{\|U_{\varsigma,k_\varsigma}\|_{2,\infty}}{\sqrt{d_n}}\eta_\varsigma\right], \quad \text{(K.14)}$$

where, for each $\varsigma \in \{+,-\}$,

$$\eta_\varsigma \lesssim_D \|U_{\varsigma,k_\varsigma}^\top E U_{\varsigma,k_\varsigma}\| + d_n\left(\mathcal{B}_{k_\varsigma}^\varsigma + \frac{\sqrt{k_\varsigma}\log n}{\delta_{\varsigma,n}} + \frac{1}{\sqrt{d_n}}\right)^2. \quad \text{(K.15)}$$

It remains to bound the bias quantities and the last term in (K.14).

We first note that if A and B have orthonormal columns and $A^\top B = 0$, then

$$\|A^\top \mathcal{R} B\| = \|A^\top(\mathcal{R} - c_0 I)B\| \leqslant \frac{1}{2}\mathrm{osc}(\mathcal{R}), \qquad c_0 := \frac{R_{\max} + R_{\min}}{2}.$$

Hence, by (K.10), all variance-profile couplings appearing in $\mathcal{B}_{k_\varsigma}^\varsigma$ (defined in (1.4)) are $O(d_n)$ and thus

$$\mathcal{B}_{k_\varsigma}^\varsigma \lesssim \frac{1}{\delta_{\varsigma,n}} + \frac{1}{d_n} \leqslant \frac{1}{\delta_n} + \frac{1}{d_n}. \quad \text{(K.16)}$$

Since r is fixed and $\delta_{\varsigma,n} \geqslant \delta_n$, (K.15) and (K.16) give

$$\eta_\varsigma \lesssim_D \|U_{\varsigma,k_\varsigma}^\top E U_{\varsigma,k_\varsigma}\| + d_n\left(\frac{\log n}{\delta_n} + \frac{1}{\sqrt{d_n}}\right)^2 \leqslant \|E\| + d_n\left(\frac{\log n}{\delta_n} + \frac{1}{\sqrt{d_n}}\right)^2.$$

Here we used $d_n \gtrsim_D \log n$ to absorb the d_n^{-1} term. Moreover, using $\|E\| \lesssim_D \sqrt{d_n}$ and $\|U_{\varsigma,k_\varsigma}\|_{2,\infty} \lesssim n^{-1/2}$, we get

$$\frac{\|U_{\varsigma,k_\varsigma}\|_{2,\infty}}{\sqrt{d_n}}\eta_\varsigma \lesssim_D \frac{1}{\sqrt{n}} + \sqrt{\frac{d_n}{n}}\left(\frac{\log n}{\delta_n} + \frac{1}{\sqrt{d_n}}\right)^2 \lesssim_D \sqrt{\frac{\log n}{n}} + \sqrt{\frac{d_n}{n}}\frac{\log n}{\delta_n}.$$

Indeed, the assumption $\delta_n \geqslant C_{1,D}\log n$ implies $\log n/\delta_n \lesssim_D 1$, while $d_n \gtrsim_D \log n$. Also, note that $\frac{1}{\delta_{+,n}} + \frac{1}{\delta_{-,n}} \leqslant \frac{2}{\delta_n}$ and, by (K.12),

$$\frac{\mathfrak{R}_D(U)}{\sqrt{d_n}} + \frac{1}{\sqrt{nd_n}} \lesssim_D \sqrt{\frac{\log n}{n}}.$$

Substituting these estimates into (K.14) proves (2.14).

□

We wrap up this section with a short derivation of (2.15).

Proof of (2.15). Suppose $\delta_n \geqslant c d_n$. Then (K.16) gives $\mathcal{B}^{\varsigma}_{k_\varsigma} \lesssim \frac{1}{d_n}$. Since $\lambda_{\varsigma,*} \asymp d_n$ and $\|U\|_{2,\infty} \lesssim n^{-1/2}$, (2.13) yields

$$\mathrm{GBias}_{k_+,k_-} \lesssim \frac{1}{\sqrt{n d_n}}.$$

Moreover,

$$\sqrt{\frac{d_n}{n}} \frac{\log n}{\delta_n} \lesssim \frac{\log n}{\sqrt{n d_n}} \lesssim_D \sqrt{\frac{\log n}{n}},$$

where the last inequality is from $d_n \gtrsim_D \log n$. Substituting these estimates into (2.14) proves (2.15). $\square$

Acknowledgment. Ke Wang is supported by Hong Kong RGC 16304222 and 16305526.

References

[1] E. Abbe. Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.

[2] E. Abbe, J. Fan, and K. Wang. An ℓ_p theory of PCA and spectral clustering. *The Annals of Statistics*, 50(4):2359–2385, 2022.

[3] E. Abbe, J. Fan, K. Wang, and Y. Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *The Annals of Statistics*, 48(3):1452–1474, 2020.

[4] J. Agterberg. Distributional theory and statistical inference for linear functions of eigenvectors with small eigengaps. *arXiv preprint arXiv:2308.02480*, 2023.

[5] J. Agterberg, Z. Lubberts, and C. E. Priebe. Entrywise estimation of singular vectors of low-rank matrices with heteroskedasticity and dependence. *IEEE Trans. Inform. Theory*, 68(7):4618–4650, 2022.

[6] L. V. Ahlfors and L. V. Ahlfors. *Complex analysis*, volume 3. McGraw-Hill New York, 1979.

[7] O. Ajanki, L. Erdős, and T. Krüger. *Quadratic vector equations on complex upper half-plane*, volume 261. American Mathematical Society, 2019.

[8] R. Allez and J.-P. Bouchaud. Eigenvector dynamics under free addition. *Random Matrices Theory Appl.*, 3(3):1450010, 17, 2014.

[9] N. Alon, M. Krivelevich, and B. Sudakov. Finding a large hidden clique in a random graph. *Random Structures & Algorithms*, 13(3-4):457–466, 1998.

[10] E. Arias-Castro and N. Verzelen. Community detection in dense random networks. *The Annals of Statistics*, 42(3):940–969, 2014.

[11] A. Athreya, D. E. Fishkind, M. Tang, C. E. Priebe, Y. Park, J. T. Vogelstein, K. Levin, V. Lyzinski, Y. Qin, and D. L. Sussman. Statistical inference on random dot product graphs: a survey. *Journal of Machine Learning Research*, 18(226):1–92, 2018.

[12] 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.

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

[14] 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.

[15] S. Balakrishnan, M. Kolar, A. Rinaldo, A. Singh, and L. Wasserman. Statistical and computational tradeoffs in biclustering. In *NIPS 2011 Workshop on Computational Trade-offs in Statistical Learning*, volume 4, 2011.

[16] A. S. Bandeira and R. van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability*, 44(4):2479–2506, 2016.

[17] J. Banks, C. Moore, J. Neeman, and P. Netrapalli. Information-theoretic thresholds for community detection in sparse networks. In *Conference on Learning Theory*, pages 383–416. PMLR, 2016.

[18] Z. Bao, K. M. Cheong, J. Lee, and Y. Li. Signal detection from spiked noise via asymmetrization. *arXiv preprint arXiv:2504.19450*, 2025.

[19] Z. Bao, X. Ding, J. Wang, and K. Wang. Statistical inference for principal components of spiked covariance matrices. *The Annals of Statistics*, 50(2):1144–1169, 2022.

[20] Z. Bao, X. Ding, and K. Wang. Singular vector and singular subspace distribution for the matrix denoising model. *The Annals of Statistics*, 49(1):370–392, 2021.

[21] Z. Bao and D. Wang. Eigenvector distribution in the critical regime of BBP transition. *Probability Theory and Related Fields*, 182:399—-479, 2022.

[22] F. Benaych-Georges, N. Enriquez, and A. Michaïl. Eigenvectors of a matrix under random perturbation. *Random Matrices Theory Appl.*, 10(2):2150023, 2021.

[23] 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.

[24] 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.

[25] 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.

[26] L. Benigni. Eigenvectors distribution and quantum unique ergodicity for deformed wigner matrices. *Annales de l'Institut Henri Poincaré-Probabilités et Statistiques*, 56(4):2822–2867, 2020.

[27] A. Bhardwaj and V. Vu. Matrix perturbation: Davis-Kahan in the infinity norm. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 880–934. SIAM, 2024.

[28] R. Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

[29] B. Bhattacharya, A. Chakrabarty, and R. S. Hazra. Outlier eigenvalues and eigenvectors of generalized wigner matrices with finite-rank perturbations. *arXiv preprint arXiv:2601.10204*, 2026.

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

[31] M. Brennan, G. Bresler, and W. Huleihel. Reducibility and computational lower bounds for problems with planted sparse structure. In *Conference On Learning Theory*, pages 48–166. PMLR, 2018.

[32] C. Butucea and Y. I. Ingster. Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli*, 19(5B):2652–2688, 2013.

[33] C. Butucea, Y. I. Ingster, and I. A. Suslina. Sharp variable selection of a sparse submatrix in a high-dimensional noisy matrix. *ESAIM: Probability and Statistics*, 19:115–134, 2015.

[34] C. Cai, G. Li, Y. Chi, H. V. Poor, and Y. Chen. Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees. *The Annals of Statistics*, 49(2):944–967, 2021.

[35] T. T. Cai, T. Liang, and A. Rakhlin. Computational and statistical boundaries for submatrix localization in a large noisy matrix. *The Annals of Statistics*, 45(4):1403–1430, 2017.

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

[37] E. J. Candès and Y. Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98:925–936, 2010.

[38] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009.

[39] J. Cape, M. Tang, and C. E. Priebe. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *The Annals of Statistics*, 47(5):2405–2439, 2019.

[40] M. Capitaine. Limiting eigenvectors of outliers for Spiked Information-Plus-Noise type matrices. *Séminaire de Probabilités XLIX*, pages 119–164, 2018.

[41] M. Capitaine and C. Donati-Martin. Non universality of fluctuations of outlier eigenvectors for block diagonal deformations of wigner matrices. *ALEA: Latin American Journal of Probability and Mathematical Statistics*, 18(1):129–165, 2021.

[42] J. Chang and J. Cape. Extreme value theory for singular subspace estimation in the matrix denoising model. *arXiv preprint arXiv:2507.19978*, 2025.

[43] K. Chen and J. Lei. Network cross-validation for determining the number of communities in network data. *Journal of the American Statistical Association*, 113(521):241–251, 2018.

[44] Y. Chen, Y. Chi, J. Fan, and C. Ma. Spectral methods for data science: A statistical perspective. *Foundations and Trends® in Machine Learning*, 14(5):566–806, 2021.

[45] Y. Chen, J. Fan, C. Ma, and K. Wang. Spectral method and regularized MLE are both optimal for top-K ranking. *The Annals of Statistics*, 47(4):2204–2235, 2019.

[46] Y. Chen and J. Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *The Journal of Machine Learning Research*, 17(1):882–938, 2016.

[47] C. Cheng, Y. Wei, and Y. Chen. Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise. *IEEE Transactions on Information Theory*, 67(11):7380–7419, 2021.

[48] L. Comtet. *Advanced Combinatorics: The art of finite and infinite expansions*. Springer Science & Business Media, 2012.

[49] M. Dadon, W. Huleihel, and T. Bendory. Detection and recovery of hidden submatrices. *IEEE Transactions on Signal and Information Processing over Networks*, 10:69–82, 2024.

[50] C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.*, 7:1–46, 1970.

[51] Y. Dekel, O. Gurel-Gurevich, and Y. Peres. Finding hidden cliques in linear time with high probability. *Combinatorics, Probability and Computing*, 23(1):29–49, 2014.

[52] J. Eldridge, M. Belkin, and Y. Wang. Unperturbed: spectral analysis beyond Davis-Kahan. In *Algorithmic Learning Theory*, pages 321–358. PMLR, 2018.

[53] J. Fan, Y. Fan, X. Han, and J. Lv. Asymptotic theory of eigenvectors for large random matrices. *Journal of the American Statistical Association*, 117(538):996–1009, 2022.

[54] 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.

[55] U. Feige and D. Ron. Finding hidden cliques in linear time. In *DMTCS Proceedings, AofA'10*, volume AM, pages 189–204, 2010.

[56] L. S. Ferreira and F. L. Metz. BBP transition and the leading eigenvector of the spiked wigner model with inhomogeneous noise. *arXiv preprint arXiv:2604.18523*, 2026.

[57] I. C. Gohberg and E. I. Sigal. An operator generalization of the logarithmic residue theorem and the theorem of Rouché. *Mathematics of the USSR-Sbornik*, 13(4):603–625, 1971.

[58] T. F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985.

[59] A. Guionnet, J. Ko, F. Krzakala, and L. Zdeborová. Low-rank matrix estimation with inhomogeneous noise. *Information and Inference: A Journal of the IMA*, 14(2):iaaf010, 2025.

[60] L. Gulikers, M. Lelarge, and L. Massoulié. A spectral method for community detection in moderately sparse degree-corrected stochastic block models. *Advances in Applied Probability*, 49(3):686–721, 2017.

[61] B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62(5):2788–2797, 2016.

[62] Y. He, A. Knowles, and R. Rosenthal. Isotropic self-consistent equations for mean-field random matrices. *Probab. Theory Related Fields*, 171(1-2):203–249, 2018.

[63] N. J. Higham. *Functions of Matrices: Theory and Computation*. SIAM, 2008.

[64] D. Hong, L. Balzano, and J. A. Fessler. Asymptotic performance of PCA for high-dimensional heteroscedastic data. *Journal of Multivariate Analysis*, 167:435–452, 2018.

[65] D. Hong and J. Cape. Network signflip parallel analysis for selecting the embedding dimension. *IEEE Transactions on Information Theory*, 2026.

[66] D. Hong, K. Gilman, L. Balzano, and J. A. Fessler. HePPCAT: Probabilistic PCA for data with heteroscedastic noise. *IEEE Transactions on Signal Processing*, 69:4819–4834, 2021.

[67] D. Hong, F. Yang, J. A. Fessler, and L. Balzano. Optimally weighted PCA for high-dimensional heteroscedastic data. *SIAM Journal on Mathematics of Data Science*, 5(1):222–250, 2023.

[68] L. Hubert and P. Arabie. Comparing partitions. *Journal of classification*, 2(1):193–218, 1985.

[69] J. Jin. Fast community detection by SCORE. *The Annals of Statistics*, 43(1):57–89, 2015.

[70] B. Karrer and M. E. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 83(1):016107, 2011.

[71] R. Keshavan, A. Montanari, and S. Oh. Matrix completion from noisy entries. *Advances in Neural Information Processing Systems*, 22, 2009.

[72] M. Kolar, S. Balakrishnan, A. Rinaldo, and A. Singh. Minimax localization of structural information in large noisy matrices. *Advances in Neural Information Processing Systems*, 24, 2011.

[73] V. Koltchinskii and D. Xia. Perturbation of linear forms of singular vectors under Gaussian noise. In *High dimensional probability VII*, volume 71 of *Progr. Probab.*, pages 397–423. Springer, [Cham], 2016.

[74] J. Lei and A. Rinaldo. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43:215–237, 2015.

[75] J. Lei and L. Zhu. Generic sample splitting for refined community recovery in degree corrected stochastic block models. *Statistica Sinica*, 27(4):1639–1659, 2017.

[76] L. Lei. Unified $\ell_{2\to\infty}$ eigenspace perturbation theory for symmetric random matrices. *arXiv preprint arXiv:1909.04798*, 2019.

[77] G. Li, C. Cai, H. V. Poor, and Y. Chen. Minimax estimation of linear functions of eigenvectors in the face of small eigen-gaps. *IEEE Transactions on Information Theory*, 71(2):1200–1247, 2024.

[78] M. Löffler, A. Y. Zhang, and H. H. Zhou. Optimality of spectral clustering in the gaussian mixture model. *The Annals of Statistics*, 49(5):2506–2530, 2021.

[79] Y. Luo, R. Han, and A. R. Zhang. A Schatten-q low-rank matrix perturbation analysis via perturbation projection error bound. *Linear Algebra and its Applications*, 630:225–240, 2021.

[80] Z. Ma and Y. Wu. Computational barriers in minimax submatrix detection. *The Annals of Statistics*, 43:1089–1116, 2015.

[81] F. McSherry. Spectral partitioning of random graphs. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pages 529–537. IEEE, 2001.

[82] A. Montanari, D. Reichman, and O. Zeitouni. On the limitation of spectral methods: From the gaussian hidden clique problem to rank-one perturbations of gaussian tensors. *Advances in Neural Information Processing Systems*, 28, 2015.

[83] S. O'Rourke, V. Vu, and K. Wang. Eigenvectors of random matrices: A survey. *Journal of Combinatorial Theory, Series A*, 144:361 – 442, 2016. Fifty Years of the Journal of Combinatorial Theory.

[84] S. O'Rourke, V. Vu, and K. Wang. Random perturbation of low rank matrices: improving classical bounds. *Linear Algebra Appl.*, 540:26–59, 2018.

[85] S. O'Rourke, V. Vu, and K. Wang. Matrices with Gaussian noise: optimal estimates for singular subspace perturbation. *IEEE Transactions on Information Theory*, 70(3):1978–2002, 2024.

[86] S. O'Rourke and N. Williams. Pairing between zeros and critical points of random polynomials with independent roots. *Trans. Amer. Math. Soc.*, 371(4):2343–2381, 2019.

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

[88] T. Qin and K. Rohe. Regularized spectral clustering under the degree-corrected stochastic blockmodel. *Advances in neural information processing systems*, 26, 2013.

[89] P. Rubin-Delanchy, J. Cape, M. Tang, and C. E. Priebe. A statistical interpretation of spectral embedding: the generalised random dot product graph. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(4):1446–1473, 2022.

[90] L. Su, W. Wang, and Y. Zhang. Strong consistency of spectral clustering for stochastic block models. *IEEE Transactions on Information Theory*, 66(1):324–338, 2019.

[91] R. Taing and K. Levin. On the effect of misspecifying the embedding dimension in low-rank network models. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026.

[92] P. Tran and V. Vu. New matrix perturbation bounds with relative norm: Perturbation of eigenspaces. *arXiv preprint arXiv:2409.20207*, 2024.

[93] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge university press, 2018.

[94] U. Von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, 36(2):555–586, 2008.

[95] V. Vu. Singular vectors under random perturbation. *Random Structures & Algorithms*, 39(4):526–538, 2011.

[96] V. Vu. A simple SVD algorithm for finding hidden partitions. *Combinatorics, Probability and Computing*, 27(1):124–140, 2018.

[97] V. Q. Vu and J. Lei. Minimax sparse principal subspace estimation in high dimensions. *The Annals of Statistics*, 41(6):2905–2947, 2013.

[98] K. Wang. Analysis of singular subspaces under random perturbations. *The Annals of Statistics*, 54(2):667–691, 2026.

[99] R. Wang. Singular vector perturbation under Gaussian noise. *SIAM J. Matrix Anal. Appl.*, 36(1):158–177, 2015.

[100] P.-Å. Wedin. Perturbation bounds in connection with singular value decomposition. *Nordisk Tidskr. Informationsbehandling (BIT)*, 12:99–111, 1972.

[101] H. Weyl. Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Mathematische Annalen*, 71(4):441–479, 1912.

[102] D. Xia and F. Zhou. The sup-norm perturbation of HOSVD and low rank tensor denoising. *Journal of Machine Learning Research*, 20(61):1–42, 2019.

[103] H. Yan and K. Levin. Minimax rates for latent position estimation in the generalized random dot product graph. *arXiv preprint arXiv:2307.01942*, 2023.

[104] H. Yan and K. Levin. Coherence-free entrywise estimation of eigenvectors in low-rank signal-plus-noise matrix models. *Advances in Neural Information Processing Systems*, 37:126566–126619, 2024.

[105] Y. Yan, Y. Chen, and J. Fan. Inference for heteroskedastic PCA with missing data. *The Annals of Statistics*, 52(2):729–756, 2024.

[106] Y. Yan and M. J. Wainwright. Entrywise inference for causal panel data: A simple and instance-optimal approach. *arXiv preprint arXiv:2401.13665*, 2024.

[107] Y. Yu, T. Wang, and R. J. Samworth. A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015.

[108] A. R. Zhang, T. T. Cai, and Y. Wu. Heteroskedastic PCA: Algorithm, optimality, and applications. *The Annals of Statistics*, 50(1):53–80, 2022.

[109] A. Y. Zhang and H. H. Zhou. Minimax rates of community detection in stochastic block models. *The Annals of Statistics*, 44(5):2252 – 2280, 2016.

[110] A. Y. Zhang and H. H. Zhou. Leave-one-out singular subspace perturbation analysis for spectral clustering. *The Annals of Statistics*, 52(5):2004–2033, 2024.

[111] Y. Zhang and M. Mondelli. Matrix denoising with doubly heteroscedastic noise: Fundamental limits and optimal spectral methods. *Advances in Neural Information Processing Systems*, 37:93060–93117, 2024.

[112] Y. Zhong. Eigenvector under random perturbation: A nonasymptotic Rayleigh-Schrödinger theory. *arXiv preprint arXiv:1702.00139*, 2017.

[113] Y. Zhong and N. Boumal. Near-optimal bounds for phase synchronization. *SIAM J. Optim.*, 28(2):989–1016, 2018.