
Coherence-free Entrywise Estimation of Eigenvectors in Low-rank Signal-plus-noise Matrix Models

Hao Yan

Department of Statistics
University of Wisconsin–Madison
Madison, WI 53706
United States of America
hyan84@wisc.edu

Keith Levin

Department of Statistics
University of Wisconsin–Madison
Madison, WI 53706
United States of America
kdlevin@wisc.edu

Abstract

Spectral methods are widely used to estimate eigenvectors of a low-rank signal matrix subject to noise. These methods use the leading eigenspace of an observed matrix to estimate this low-rank signal. Typically, the entrywise estimation error of these methods depends on the coherence of the low-rank signal matrix with respect to the standard basis. In this work, we present a novel method for eigenvector estimation that avoids this dependence on coherence. Assuming a rank-one signal matrix, under mild technical conditions, the entrywise estimation error of our method provably has no dependence on the coherence under Gaussian noise (i.e., in the spiked Wigner model), and achieves the optimal estimation rate up to logarithmic factors. Simulations demonstrate that our method performs well under non-Gaussian noise and that an extension of our method to the case of a rank- r signal matrix has little to no dependence on the coherence. In addition, we derive new metric entropy bounds for rank- r singular subspaces under $\ell_{2,\infty}$ distance, which may be of independent interest. We use these new bounds to improve the best known lower bound for rank- r eigenspace estimation under $\ell_{2,\infty}$ distance.

1 Introduction

Spectral methods are extensively used in contemporary data science and engineering [27]. The fundamental idea underlying these methods is that the eigenspace or singular subspace of an observed matrix reflects important structure present in the data from which it is derived. Spectral methods have been deployed successfully in a variety of tasks, including low-rank matrix denoising [30, 10], factor analysis [21, 32, 3, 61, 11, 62], community detection [49, 52, 1, 40, 50], pairwise ranking [43, 25] and matrix completion [48, 24, 7]. The widespread use of spectral methods has driven extensive research into the theoretical properties of eigenspaces and singular subspaces, yielding normal approximation results [10, 11, 33, 58, 5] as well as perturbation bounds [60, 20, 22, 45]. For a more comprehensive recent review of spectral methods, see [27].

1.1 Eigenspace estimation in low-rank matrix models

Consider an unknown symmetric matrix $M^* = U^* \Lambda^* U^{*\top} \in \mathbb{R}^{n \times n}$, where $U^* \in \mathbb{R}^{n \times r}$ has orthonormal columns and $\Lambda^* \in \mathbb{R}^{r \times r}$ is diagonal, containing the nonzero eigenvalues of M^* ordered so that $|\lambda_1^*| \geq |\lambda_2^*| \geq \dots \geq |\lambda_r^*| > 0$. Our goal is to estimate U^* from a noisy observation

$$Y = M^* + W \in \mathbb{R}^{n \times n}, \quad (1)$$

where $\mathbf{W} = [W_{ij}]_{1 \leq i, j \leq n}$ is a symmetric random noise matrix with mean zero. We restrict our attention here to the symmetric case for the sake of simplicity, but we expect that our results can be extended to the asymmetric case using standard dilation arguments [53].

Throughout this paper, we assume that the entries of $\mathbf{W}$ are subgaussian.

Assumption 1. *The entries of $\mathbf{W}$ on and above the diagonal are independent and symmetric about zero with common variance σ^2 and common subgaussian parameter ν_W .*

We remind the reader that the subgaussian parameter ν_W serves as a “proxy” for the variance. Indeed, in the Gaussian case, we have $\sigma^2 = \nu_W$, while $\sigma^2 \leq \nu_W$ more generally [54, 56].

Spectral methods often estimate $\mathbf{U}^*$ directly using the r leading eigenvectors $\mathbf{U}$ of $\mathbf{Y}$. As a result, the entrywise and row-wise behavior of $\mathbf{U}$ has attracted considerable attention [31, 22, 41, 2, 4, 14]. Given an estimator $\hat{\mathbf{U}} \in \mathbb{R}^{n \times r}$, the estimation error is measured in terms of the $\ell_{2,\infty}$ distance

$$d_{2,\infty}(\hat{\mathbf{U}}, \mathbf{U}^*) = \min_{\mathbf{\Gamma} \in \mathbb{O}_r} \|\mathbf{U}^* - \hat{\mathbf{U}}\mathbf{\Gamma}\|_{2,\infty}, \quad (2)$$

where the presence of $\mathbf{\Gamma}$ is to resolve rotational non-identifiability. In the rank-one case, this reduces to the ℓ_∞ distance,

$$d_\infty(\hat{\mathbf{u}}, \mathbf{u}^*) = \min \{ \|\mathbf{u}^* - \hat{\mathbf{u}}\|_\infty, \|\mathbf{u}^* + \hat{\mathbf{u}}\|_\infty \} \quad (3)$$

for $\mathbf{u}^* \in \mathbb{R}^n$ and a given estimator $\hat{\mathbf{u}} \in \mathbb{R}^n$. Estimation error bounds in ℓ_∞ or $\ell_{2,\infty}$ distance typically rely on the incoherence parameter μ of $\mathbf{U}^*$, defined as

$$\mu = \frac{n}{r} \|\mathbf{U}^*\|_{2,\infty}^2 \in [1, n/r].$$

In the rank-one case with Gaussian noise, if the leading eigenvalue λ^* of $\mathbf{M}^*$ satisfies $|\lambda^*| = \Omega(\sigma\sqrt{n})$, Theorem 4.1 of [27] shows that with probability at least $1 - O(n^{-8})$, the leading eigenvector $\mathbf{u}$ of $\mathbf{Y}$ satisfies

$$d_\infty(\mathbf{u}, \mathbf{u}^*) \lesssim \frac{\sigma(\sqrt{\log n} + \sqrt{n}\|\mathbf{u}^*\|_\infty)}{|\lambda^*|} = \frac{\sigma\sqrt{\log n} + \sigma\sqrt{\mu}}{|\lambda^*|}. \quad (4)$$

When $\sigma\sqrt{n} \lesssim |\lambda^*| \lesssim \sigma\sqrt{n \log n}$, a regime of most interest (no polynomial-time algorithm is known when $|\lambda^*| \ll \sigma\sqrt{n}$ [6], and estimation is easy when $|\lambda^*| \gg \sigma\sqrt{n \log n}$), we show in Lemma 1 that Equation (4) is not improvable up to log-factors, as a result of the Baik-Ben Arous-Péché (BBP) phase transition [8] (see [12, 38, 34] for BBP-style phase transitions in the setting of this paper).

Lemma 1. *Under Equation (1) with Gaussian noise, let $\mathbf{M}^* = \lambda^* \mathbf{u}^* \mathbf{u}^{*\top}$. If both limits $\lim_{n \rightarrow \infty} \lambda^*/(\sigma\sqrt{n}) > 1$ and $\lim_{n \rightarrow \infty} \mu/n$ exist, then for any $\mu \in [1, n]$, there exists $\mathbf{u}^* \in \mathbb{S}^{n-1}$ such that almost surely,*

$$\liminf_{n \rightarrow \infty} d_\infty(\mathbf{u}, \mathbf{u}^*) \geq \lim_{n \rightarrow \infty} \frac{\sigma^2 \sqrt{n\mu}}{2\sqrt{2}|\lambda^*|^2}. \quad (5)$$

Equation (5) shows that the spectral estimate $\mathbf{u}$ has an intrinsic dependence on μ , with especially bad performance when μ is large and $|\lambda^*| = \Theta(\sigma\sqrt{n \log n})$. In this large- μ regime, beyond low-rankedness, $\mathbf{M}^*$ exhibits additional structure (e.g., sparsity) that is not fully utilized by the spectral estimator. This suggests that the dependence on μ in Equations (4) and (5) is a shortcoming of the spectral estimator. In Algorithm 1, we present a new estimator designed to remove this dependence on μ . Theorem 1 shows that up to log-factors, it matches the minimax lower bound discussed below (see Equation (9) in Section 1.2). Experiments in Section 5 further support our theoretical results.

In the rank- r case with Gaussian noise, Theorem 4.2 in [27] shows that when $|\lambda_r^*| \gtrsim \sigma\sqrt{n \log n}$,

$$d_{2,\infty}(\mathbf{U}, \mathbf{U}^*) \lesssim \frac{\sigma(\kappa\sqrt{\mu r} + \sqrt{r \log n})}{|\lambda_r^*|} \quad (6)$$

with probability at least $1 - O(n^{-8})$, where $\kappa = |\lambda_1^*|/|\lambda_r^*|$ is the condition number of $\mathbf{M}^*$. Here again, the estimation error depends on μ , and we conjecture that this dependence is also sub-optimal. Algorithm 2 extends our rank-1 estimation algorithm to this more general rank- r case. Experiments in Section 5 show that Algorithm 2 outperforms the naïve spectral method in the general rank- r case, with little to no sensitivity to the coherence μ . We note in passing that the dependence on κ in Equation (6) can likely be removed [3, 62, 57], though we do not pursue this here.

1.2 Minimax lower bounds for subspace estimation

Minimax lower bounds have been established for a variety of subspace estimation problems, including sparse PCA [21, 55], matrix denoising [20], structural matrix estimation [19], network estimation [35, 63] and estimating linear functions of eigenvectors [42, 28]. Most of these studies focus on minimax lower bounds under the Frobenius or operator norm, derived using the packing numbers of Grassmann manifolds [46, 13]. In the matrix denoising literature, it is well-known that for $r \geq 1$,

$$\min_{\Gamma \in \mathbb{O}_r} \left\| \widehat{\mathbf{U}}\Gamma - \mathbf{U}^* \right\|_F \gtrsim \min \left\{ \frac{\sigma\sqrt{nr}}{|\lambda_r^*|}, \sqrt{r} \right\} \quad (7)$$

holds in a minimax sense (see Theorem 3 in [28] or Theorem 4 in [63]). Far fewer papers have considered lower bounds under $d_{2,\infty}$ [18, 4], and these results are derived via the trivial lower bound

$$d_{2,\infty}(\widehat{\mathbf{U}}, \mathbf{U}^*) \geq \frac{1}{\sqrt{n}} \min_{\Gamma \in \mathbb{O}_r} \left\| \widehat{\mathbf{U}}\Gamma - \mathbf{U}^* \right\|_F, \quad (8)$$

which holds for any $\widehat{\mathbf{U}}, \mathbf{U}^* \in \mathbb{R}^{n \times r}$. Applying the lower bounds in Equations (7) and (8), we have

$$\sup_{\mathbf{U}^* \in \mathbb{R}^{n \times r}: \mathbf{U}^{*\top} \mathbf{U}^* = \mathbf{I}_r} d_{2,\infty}(\widehat{\mathbf{U}}, \mathbf{U}^*) \gtrsim \min \left\{ \frac{\sigma\sqrt{r}}{|\lambda_r^*|}, \sqrt{\frac{r}{n}} \right\} \quad (9)$$

holds for any estimator $\widehat{\mathbf{U}} \in \mathbb{R}^{n \times r}$. When $\mathbf{U}^*$ is incoherent, meaning that $\mu = O(1)$, this lower bound is achieved by the spectral estimation rate in Equation (6) when $|\lambda_r^*| = \Omega(\sigma\sqrt{n \log n})$, and is achieved trivially by $\widehat{\mathbf{U}} = \mathbf{0}_{n,r}$ when $|\lambda_r^*| = o(\sigma\sqrt{n})$. On the other hand, when $\mathbf{U}^*$ is coherent, in the sense that $\mu = \omega(1)$, and $|\lambda_r^*| = \Omega(\sigma\sqrt{n \log n})$, our discussion above in Section (1.1) (including our new results in Theorem 1) suggests that the rate in Equation (9) can be achieved up to log-factors.

This leaves open the question of the minimax rate when $\mu = \omega(1)$ and $|\lambda_r^*| = o(\sigma\sqrt{n})$. In this case, the lower bound in Equation (9) cannot exceed $\sqrt{r/n}$. This seems suboptimal, as we expect some dependence on $\|\mathbf{U}^*\|_{2,\infty}$ (for example, consider the extreme case when λ^* is very near zero). This suboptimality arises from the naïve lower bound in Equation (8). In Theorem 2, we improve this lower bound, removing the $\sqrt{r/n}$ dependence in Equation (9). This improved lower bound makes use of novel metric entropy bounds for singular subspaces, which may be of independent interest.

1.3 Notation and roadmap

We use C to denote a constant whose precise values may change from line to line. For a positive integer n , we write $[n] = \{1, 2, \dots, n\}$. $|\mathcal{A}|$ denotes the cardinality of a set $\mathcal{A}$. For real numbers a and b , we write $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. For a vector $\mathbf{v} = (v_1, v_2, \dots, v_n)^\top \in \mathbb{R}^n$, we use the norms $\|\mathbf{v}\|_2 = \sqrt{\sum_{i=1}^n v_i^2}$ and $\|\mathbf{v}\|_\infty = \max_i |v_i|$. We let $\mathbf{e}_i \in \mathbb{R}^n, i \in [n]$ denote the standard basis vectors of $\mathbb{R}^n$. $\mathbb{S}^{n-1} = \{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u}\|_2 = 1\}$ denotes the unit sphere. For a matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$, $\mathbf{M}_{i,\cdot}$ denotes its i -th row as a row vector, $\|\mathbf{M}\|$ denotes its operator norm and $\|\mathbf{M}\|_{2,\infty} = \max_{i \in [n]} \|\mathbf{M}_{i,\cdot}\|_2$ indicates the maximum row-wise ℓ_2 norm. $\mathbf{I}_n \in \mathbb{R}^n$ denotes the n -by- n identity matrix. $\mathbb{O}_r$ denotes the r -dimensional orthogonal group. We use both standard Landau notation and asymptotic notation: for positive functions $f(n)$ and $g(n)$, we write $f(n) \gg g(n)$, $f(n) = \omega(g(n))$ or $g(n) = o(f(n))$ if $f(n)/g(n) \rightarrow \infty$ as $n \rightarrow \infty$. We write $f(n) \gtrsim g(n)$, $f(n) = \Omega(g(n))$ or $g(n) = O(f(n))$ if for some constant $C > 0$, we have $f(n)/g(n) \geq C$ for all sufficiently large n . We write $f(n) = \Theta(g(n))$ if both $f(n) = O(g(n))$ and $g(n) = O(f(n))$.

The remainder of the paper is organized as follows. In Section 2, we study the eigenspace estimation problem for rank-one matrices and propose a new algorithm that achieves the minimax optimal error rate up to logarithmic factors in the growth regime where $|\lambda^*| = \Omega(\sigma\sqrt{n \log n})$ (Theorem 1). In Section 3, we extend this algorithm to rank- r eigenspace estimation. In Section 4, we present theoretical results for the metric entropy of subspaces under d_∞ and $d_{2,\infty}$ and improve Equation (9) under the growth regime where $|\lambda^*| = O(\sigma\sqrt{n})$ (Theorem 2). Numerical results are provided in Section 5. We conclude in Section 6 with a discussion of the limitations of our study and directions for future work. Detailed proofs of all lemmas and theorems can be found in the appendix.

2 Rank-one matrix eigenspace estimation

In this section, we study the model in Equation (1) when M^* is rank-one with eigendecomposition $M^* = \lambda^* \mathbf{u}^* \mathbf{u}^{*\top}$, where $\lambda^* \in \mathbb{R}$ and $\mathbf{u}^* \in \mathbb{S}^{n-1}$. As discussed in Section 1, spectral methods may have sub-optimal dependence on μ compared to the lower bound in Equation (9). We show that a better estimator is possible by working with a carefully selected subset of entries of $\mathbf{Y}$. We start with a key observation in Lemma 2, which states that any unit vector contains a subset of large entries, and this subset has a sufficiently large cardinality. This subset is the key to our new estimator.

Lemma 2. *Let $\mathcal{A} \subset [0, 1]$ be the set*

$$\mathcal{A} = \left\{ \log^{-\frac{1}{2}} n, \dots, \log^{-\frac{\lceil L \rceil - 1}{2}} n, \log^{-\frac{L}{2}} n \right\}, \quad (10)$$

where L is given by

$$L = \frac{\log(2n)}{\log \log n}. \quad (11)$$

For n sufficiently large, for every $\mathbf{v} \in \mathbb{S}^{n-1}$, there exists $\alpha_0 \in \mathcal{A}$ such that

$$\frac{1}{\alpha_0^2} \geq |\{i : |v_i| \geq \alpha_0\}| > \frac{1}{\alpha_0^2 \log^2 n}. \quad (12)$$

To motivate Algorithm 1, suppose $\mathbf{u}^*$ is entrywise positive. For $\alpha_0 \in \mathcal{A}$, denote the set in Equation (12) by I_{α_0} . By Equation (12), the sum of entries M_{ij}^* with $i, j \in I_{\alpha_0}$ grows as $\Omega(|\lambda^*| |I_{\alpha_0}| \log^{-2} n)$, while the sum of the corresponding entries of $\mathbf{W}$ grows as $O(\sigma |I_{\alpha_0}| \sqrt{\log n})$. That is, when $|\lambda^*|$ is sufficiently large, the signal contained in the entries $i, j \in I_{\alpha_0}$ dominates the noise. If we knew I_{α_0} , utilizing the entries in I_{α_0} would reduce the estimation error incurred by small entries of $\mathbf{u}^*$. In practice, we do not know I_{α_0} and must estimate such a subset. To ensure that this is possible, we impose a technical assumption on $\mathbf{u}^*$. We discuss this assumption below in Remark 2.

Assumption 2. *There exists an $\alpha_0 \in \mathcal{A}$ satisfying Equation (12) and a constant $0 < \epsilon_0 \leq 1$ such that for all sufficiently large n ,*

$$\{i : |u_i^*| \in [(1 - \epsilon_0)\alpha_0, (1 + \epsilon_0)\alpha_0]\} = \emptyset.$$

We pause to give a few examples to illustrate Assumption 2. First, consider $\mathbf{u}^* = c_1 \mathbf{e}_1 + c_2 n^{-1/2} \mathbf{1}_n$, with $c_1, c_2 = \Theta(1)$ chosen so that $\|\mathbf{u}^*\|_2 = 1$. We note that c_1, c_2 both depend on n , but are bounded away from zero as n grows, and one can verify that Assumption 2 holds with $\alpha_0 = \log^{-1/2} n$ and $\epsilon_0 = 1/2$. As another example, consider $\mathbf{u}^* = n^{-1/2} \mathbf{1}_n \in \mathbb{R}^n$. One may verify that taking $\alpha_0 = (\log n)^{-L/2} = 1/\sqrt{2n}$ and $\epsilon_0 = 0.4$ satisfies the conditions in Assumption 2.

As an example of a setting that violates Assumption 2, consider $\mathbf{u}^*$ obtained by renormalizing a vector of i.i.d. Gaussians. This results in $\mathbf{u}^*$ being Haar-distributed on $\mathbb{S}^{n-1}$ and Assumption 2 is violated with high probability. To see this, note that $\mathbf{u}^* \approx \mathbf{g}/\sqrt{n}$ where $\mathbf{g} \sim N(0, \mathbf{I}_n)$ (see Theorem 3.4.6 in [54]). Since with high probability $\|\mathbf{g}\|_\infty = O(\sqrt{\log n})$, Equation (12) holds only when $\alpha_0 \approx 1/\sqrt{n}$. For Assumption 2 to hold, there must be a gap of $\Theta(1)$ between the entries of $\mathbf{g}$, which fails with high probability.

It is tempting to conclude from the counter-example just given that renormalizing a vector of i.i.d. entries must necessarily result in a $\mathbf{u}^*$ that violates Assumption 2, but this is not always the case. If the entrywise distribution has suitable structure, $\mathbf{u}^*$ may still obey Assumption 2. As an illustration, suppose that $\mathbf{u}^*$ is obtained by renormalizing a vector $\mathbf{g} = (g_1, g_2, \dots, g_n)^\top \in \mathbb{R}^n$ with i.i.d. entries from a distribution with variance 1, so that $\mathbf{u}^* \approx \mathbf{g}/\sqrt{n}$. If the g_i are drawn by taking $g_i = a$ with probability p and $g_i = b$ with probability $1 - p$, then each entry of $\mathbf{u}^*$ is either approximately $ap/\sqrt{n}$ or approximately $b(1 - p)/\sqrt{n}$. Choosing a, b and p appropriately, we can ensure a gap between the entries of $\mathbf{u}^*$ of size $O(n^{-1/2})$, and we can take $\alpha_0 = (\log n)^{-L} \approx n^{-1/2}$.

Our new estimator of $\mathbf{u}^*$ is a refinement based on the leading eigenvector and eigenvalue of $\mathbf{Y}$. For the spectral estimator to provide a useful initialization, we make Assumption 3 on λ^* .

Assumption 3. *There exists a constant $C_1 > 2400/\epsilon_0$, where ϵ_0 is as in Assumption 2, such that the leading eigenvalue λ^* of M^* satisfies*

$$|\lambda^*| \geq C_1 \sqrt{\nu_W n \log n}. \quad (13)$$

Remark 1. The dependence on ϵ_0 in Assumption 3 is for technical reasons discussed in Remark 2. In our proofs, we do not optimize the dependence on C_1 and assume that $C_1 > 2400/\epsilon_0$. As demonstrated in our experiments in Section 5, $|\lambda^*| \geq \sqrt{\nu_W n \log n}$ appears sufficient in practice. When $|\lambda^*| \leq \sqrt{\nu_W n}$, the spectral estimator fails to provide any useful initial estimate and it is believed that no polynomial-time algorithm can succeed. We provide more discussion on this matter in Remark 6.

The final ingredient required for Algorithm 1 is a leading eigenvalue estimate $\hat{\lambda}$ that recovers the true signal eigenvalue λ^* suitably well.

Assumption 4. Under Assumption 3, $\hat{\lambda}$ is such that with probability at least $1 - O(n^{-8} \log n)$,

$$|\hat{\lambda} - \lambda^*| \leq C_2 \sqrt{\nu_W} \log^{5/2} n.$$

Assumption 4 seems stringent at first. The top eigenvalue λ of $\mathbf{Y}$ achieves only a $O(\sqrt{\nu_W n})$ error rate (see Lemma 2.2 and Equation (3.12) in [27]). This is because $\lambda = \lambda^* + n\sigma^2/\lambda^* + O(\sqrt{\nu_W \log n})$ (see [47] or Theorem 2.3 in [23]; see also [26, 51, 17]). Luckily, in our setting, the bias-corrected estimate

$$\hat{\lambda}_c = \frac{1}{2} \left(\lambda + \sqrt{\lambda^2 - 4n\sigma^2} \right), \quad (14)$$

does satisfy Assumption 4 [26]. Another estimator, which falls naturally out of Algorithm 1, also satisfies Assumption 4. We find that it performs similarly to the debiased estimator $\hat{\lambda}_c$ empirically, and so we do not explore it here. Theoretical results for this estimator are in the appendix.

With the above assumptions in hand, we propose a new method given in Algorithm 1. We note that the main computational bottleneck of Algorithm 1 is to find the leading eigenvector $\mathbf{u}$ of $\mathbf{Y}$, and thus the runtime is essentially the same as for standard spectral methods (see, e.g., [37]).

Algorithm 1 Coherence-free eigenvector estimation algorithm

Input: Observed matrix $\mathbf{Y} \in \mathbb{R}^{n \times n}$; leading eigenvalue estimate $\hat{\lambda}$; parameter $\beta > 0$.

Output: $\hat{\mathbf{u}} \in \mathbb{R}^n$

- 1: If $\hat{\lambda} < 0$, set $\mathbf{Y} = -\mathbf{Y}$. Obtain the top eigenvector $\mathbf{u} \in \mathbb{S}^{n-1}$ of $\mathbf{Y}$.
- 2: Pick any $\hat{\alpha} \in \mathcal{A}$ such that the set $\hat{I} = \{i : |u_i| \geq \hat{\alpha}\}$ satisfies

$$|\hat{I}| \geq \frac{1}{\hat{\alpha}^2 \log^2 n}, \quad \text{and} \quad (15)$$

$$\{i : (1 - \beta)\hat{\alpha} < |u_i| < (1 + \beta)\hat{\alpha}\} = \emptyset, \quad (16)$$

- 3: Let $\mathbf{Q} \in \mathbb{R}^{n \times n}$ be diagonal with $Q_{kk} = \begin{cases} \text{sgn}(u_k) & \text{if } k \in \hat{I} \\ 1 & \text{if } k \in \hat{I}^c \end{cases}$ and let $\tilde{\mathbf{Y}} = \mathbf{Q}\mathbf{Y}\mathbf{Q}$.
 - 4: Set $\hat{S} = \sqrt{\sum_{j,k \in \hat{I}} \tilde{Y}_{jk}}$ and let $\hat{v}_j = \left(\sum_{k \in \hat{I}} \tilde{Y}_{jk} \right) / \left(\hat{S} \sqrt{\hat{\lambda}} \right)$ for $j \in [n]$.
 - 5: For each $j \in [n]$, set $\hat{u}_j = u_j$ if $|u_j| \leq \left(\sigma / \sqrt{\hat{\lambda}} \right) \log n$, and $\hat{u}_j = Q_{jj} \hat{v}_j$ otherwise.
-

Remark 2. In our proofs, we set $\beta = \epsilon_0/2$. Equation (16), Assumption 2 and the ϵ_0 -dependence in Assumption 3 are technical requirements to ensure that with high probability, $\hat{I}$ is one of a few deterministic sets, avoiding the complicated dependence between $\hat{I}$ and $\mathbf{W}$. Empirically, Algorithm 1 works well even without these technical conditions. We conjecture that Assumption 2 as well as the ϵ_0 -dependence in Assumption 3 can be removed. See Section 5 for further discussion.

As alluded to above, the intuition behind Algorithm 1 is that we aim to concentrate our efforts on estimating the large entries of $\mathbf{u}^*$. Consider an entry of $\mathbf{Y}$ given by $\lambda^* u_i^* u_j^* + W_{ij}$. Intuitively, locations corresponding to small entries of $\mathbf{u}^*$ produce small $u_i^* u_j^*$. These entries of $\mathbf{Y}$ have a small signal to noise ratio compared to those arising from products of large entries of $\mathbf{u}^*$. If we knew the locations of the large entries of $\mathbf{u}^*$, we could use them to obtain more accurate estimates of $\mathbf{u}^*$. Essentially, both Algorithm 1 above and Algorithm 2 presented below consist of two parts: finding the large locations, and using those locations to improve our initial spectral estimate of $\mathbf{u}^*$.

Algorithm 1 assumes that the entrywise variance σ^2 of $\mathbf{W}$ is known. Of course, in practice, this is not the case, and we must estimate σ^2 . There are several well-established methods for this estimation task. For example, when $\mathbf{W}$ is asymmetric, [36] introduces an estimator based on the median singular value of $\mathbf{Y}$. In our case, it suffices to estimate σ^2 using a simple plug-in estimator

$$\hat{\sigma}^2 = \frac{2}{n(n+1)} \sum_{1 \leq i \leq j \leq n} (Y_{ij} - \widehat{M}_{ij})^2, \quad (17)$$

where $\widehat{\mathbf{M}} = \lambda \mathbf{u} \mathbf{u}^\top \in \mathbb{R}^{n \times n}$. In general, if $\mathbf{M}^*$ has rank r , then we set $\widehat{\mathbf{M}} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^\top$, where $\mathbf{\Lambda}$ is the leading r eigenvalues of $\mathbf{Y}$ (sorted by non-increasing magnitude) and $\mathbf{U} \in \mathbb{R}^{n \times r}$ contains the corresponding r leading orthonormal eigenvectors as its columns. Lemma 3 controls the estimation error of the plug-in estimator $\hat{\sigma}^2$ for a general rank- r signal matrix.

Lemma 3. *Under the model given in Equation (1), let $\mathbf{M}^* = \mathbf{U}^* \mathbf{\Lambda}^* \mathbf{U}^{*\top}$ be a rank- r matrix with $r \geq 1$, where $\mathbf{\Lambda}^* = \text{diag}(\lambda_1^*, \lambda_2^*, \dots, \lambda_r^*)$ such that $|\lambda_1^*| \geq \dots \geq |\lambda_r^*|$. Suppose that Assumption 1 holds and that $|\lambda_r^*| \geq 20\sqrt{\nu_W n}$, then the estimator $\hat{\sigma}^2$ given in Equation (17) is such that with probability at least $1 - O(n^{-8})$,*

$$|\hat{\sigma}^2 - \sigma^2| \leq \frac{400\nu_W r}{n} + \frac{4\nu_W \sqrt{\log n}}{cn} + \frac{200\nu_W r \sqrt{\log n}}{n^{3/2}} \quad (18)$$

where $c > 0$ is a universal constant.

Remark 3. We use the plug-in estimator $\hat{\sigma}$ in the debiased estimator $\hat{\lambda}_c$ in Equation (14) and to construct $\hat{\mathbf{u}}$ in Step 5 of Algorithm 1. Lemma 3 shows that this only introduces an extra log-factor to the error bound, which does not affect the estimation error rate of either λ^* or $\mathbf{u}^*$.

Our main result, Theorem 1, controls the estimation error of Algorithm 1, as measured under ℓ_∞ .

Theorem 1. *Under the model in Equation (1), suppose that Assumptions 1, 2, 3 and 4 hold. Then for n sufficiently large, the estimate $\hat{\mathbf{u}} \in \mathbb{R}^n$ produced by Algorithm 1 satisfies*

$$d_\infty(\hat{\mathbf{u}}, \mathbf{u}^*) \leq \frac{C\sqrt{\nu_W}(\log n)^{5/2}}{|\lambda^*|}$$

with probability at least $1 - O(n^{-8} \log n)$, where $C > 0$ is a universal constant.

Remark 4. Under Gaussian noise, in the regime $|\lambda^*| = \Omega(\sigma\sqrt{n \log n})$, our upper bound is minimax rate-optimal up to log-factors compared to the lower bound in Equation (9). In particular, the rate obtained in Theorem 1 does not depend on the coherence parameter μ . A more careful analysis might be able to remove some of the log-factors in Theorem 1, but we leave this matter for future work.

3 Rank- r matrix eigenspace estimation

To handle the more general case in which the signal matrix $\mathbf{M}^*$ is rank r , we propose Algorithm 2, which yields an estimate of $\mathbf{U}^*$. This is achieved by estimating the r leading eigenvectors separately, then combining them into an estimate $\widehat{\mathbf{U}} \in \mathbb{R}^{n \times r}$. We explore the empirical performance of Algorithm 2 via simulation in Section 5.2 and leave its theoretical analysis to future work. Algorithm 2 requires the observed matrix $\mathbf{Y}$ and an estimate of the k -th leading eigenvalue λ_k^* of $\mathbf{M}^*$ as input. Similar to the rank-one case, the top- r leading eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_r$ of $\mathbf{Y}$ are biased [47]. We again use a debiased estimator $\hat{\lambda}_{k,c}$ for $k \in [r]$ as input to Algorithm 2, given by

$$\hat{\lambda}_{k,c} = \frac{1}{2} \left(\lambda_k + \sqrt{\lambda_k^2 - 4n\sigma^2} \right). \quad (19)$$

Algorithm 2 is a natural extension of Algorithm 1. Essentially, it converts the problem of estimating $\mathbf{u}_k^*$ into an eigenvector estimation problem under a rank-one signal-plus-noise model given by

$$\mathbf{Y} = \lambda_k^* \mathbf{u}_k^* \mathbf{u}_k^{*\top} + (\mathbf{M}_{-k}^* + \mathbf{W}) = \lambda_k^* \mathbf{u}_k^* \mathbf{u}_k^{*\top} + (\mathbf{M}^* - \lambda_k^* \mathbf{u}_k^* \mathbf{u}_k^{*\top} + \mathbf{W}),$$

where $\mathbf{M}_{-k}^* = \mathbf{M}^* - \lambda_k^* \mathbf{u}_k^* \mathbf{u}_k^{*\top}$. There are two differences between Algorithms 1 and 2. First, we remove Equation (16) from Algorithm 1, as this is mainly a technical requirement (see Remark 2).

More importantly, in Algorithm 2, we conjugate $\mathbf{Y}$ by a random orthogonal matrix $\mathbf{H} \in \mathbb{O}_n$. The $(r-1)$ leading eigenvectors of $\mathbf{H}\mathbf{M}_{-k}^*\mathbf{H}^\top$ form a random subspace of $\mathbb{R}^{n \times (r-1)}$, which allows us to treat $\mathbf{H}\mathbf{M}_{-k}^*\mathbf{H}^\top$ as a noise matrix. By way of illustration, consider the rank-2 case with $\mathbf{M}_{-1}^* = \lambda_2^* \mathbf{u}_2^* \mathbf{u}_2^{*\top}$. $\mathbf{H}\mathbf{u}_2^*$ behaves similarly to a random vector drawn uniformly from $\mathbb{S}^{n-1}$. Therefore, one would expect $\mathbf{H}\mathbf{u}_2^* \mathbf{u}_2^{*\top} \mathbf{H}$ to behave similarly to a noise matrix $n^{-1} \mathbf{g} \mathbf{g}^\top$, where $\mathbf{g} \sim N(0, \mathbf{I}_n)$ (see Chapter 3 of [54]). As in the discussion after Lemma 2, we can find a set of indices I_{α_0} such that the signal in the corresponding entries of $\mathbf{H}\mathbf{u}_1^*$ dominates the noise.

Algorithm 2 Coherent optimal eigenvector estimation algorithm

Input: Observed matrix $\mathbf{Y} \in \mathbb{R}^{n \times n}$; k -th leading eigenvalue estimate $\hat{\lambda}_k$. If $\hat{\lambda}_k < 0$, set $\mathbf{Y} = -\mathbf{Y}$.

Output: $\hat{\mathbf{u}}_k \in \mathbb{R}^n$

- 1: Obtain the top- r eigenvectors $\tilde{\mathbf{U}}$ of $\mathbf{H}\mathbf{Y}\mathbf{H}^\top$, where $\mathbf{H} \in \mathbb{O}_n$ is Haar-distributed.
 - 2: Set $\mathbf{Q} = \text{diag}(\text{sgn}(\tilde{\mathbf{U}}_{:,k}))$ and set $\tilde{\mathbf{Y}} = \mathbf{Q}\mathbf{H}\mathbf{Y}\mathbf{H}^\top\mathbf{Q}$.
 - 3: Pick an $\alpha_0 \in \mathcal{A}$ such that for $\hat{I} := \{i : |\tilde{U}_{i,k}| \geq \alpha_0\}$, $|\hat{I}| \geq 1/\alpha_0^2 \log^2 n$.
 - 4: Set $\hat{\mathbf{S}} = \left(\sum_{j,\ell \in \hat{I}} \tilde{Y}_{j\ell}\right)^{1/2}$ and set $\hat{v}_j = \sum_{\ell \in \hat{I}} \tilde{Y}_{j\ell} / \left(\hat{\mathbf{S}} \sqrt{\hat{\lambda}_k}\right)$ for $j \in [n]$.
 - 5: Let $\mathbf{U} = \mathbf{H}^\top \tilde{\mathbf{U}}$. For $j \in [n]$, set $\hat{u}_{k,j} = \begin{cases} U_{k,j} & \text{if } |U_{k,j}| \leq (\sigma/|\hat{\lambda}_k|) \log n \\ (\mathbf{H}^\top \mathbf{Q} \hat{\mathbf{v}})_j & \text{otherwise.} \end{cases}$
-

As mentioned above, our experiments in Section 5.2 indicate that Algorithm 2 performs well. A proof of its performance, however, is more complicated than Theorem 1. The main difficulty arises from the fact that in Algorithm 2, we conjugate by a random orthogonal transformation. This ensures that when considering the large entries of one signal eigenvector, the other signal eigenvectors are ignorable. Unfortunately, this random orthogonal transformation breaks Assumption 2 and introduces complicated dependency structure, requiring a more careful analysis that we leave for future work.

4 Estimation lower bounds under $\ell_{2,\infty}$ distance

Theorem 1 demonstrates that for a rank-one signal, dependence of the estimation rate on μ can be removed when $|\lambda^*| = \Omega(\sigma\sqrt{n \log n})$. In this regime, our result matches the lower bound in Equation (9) up to log-factors, making this the minimax lower bound under Assumption 2. As discussed in Remark 2, we expect Assumption 2 can be removed via a more careful analysis, rendering the lower bound in Equation (9) optimal under $|\lambda^*| = \Omega(\sigma\sqrt{n \log n})$. On the other hand, as discussed in Section 1.2, the lower bound in Equation (9) is sub-optimal in the regime where $|\lambda^*| = O(\sigma\sqrt{n})$. In what follows, we aim to improve upon Equation (9) by deriving metric entropy bounds [54] for rank- r singular subspaces under the $\ell_{2,\infty}$ distance when $|\lambda^*| = O(\sigma\sqrt{n})$.

Recall that for a semi-metric ρ defined on a set $\mathbb{K}$, we may define the δ -packing number $\mathcal{M}(\mathbb{K}, \rho, \delta)$ of $\mathbb{K}$ under ρ (see Chapter 15 in [56]). The packing δ -entropy, $\log \mathcal{M}(\mathbb{K}, \rho, \delta)$, captures the complexity of the space $\mathbb{K}$, and a lower bound on the δ -entropy can be translated into a lower bound on the minimax estimation error rate. For a given $r \in [n]$ and $1 \leq \mu \leq n/r$, we consider the parameter set

$$\mathbb{K}(n, r, \sqrt{\mu r/n}) = \left\{ \mathbf{U} \in \mathbb{R}^{n \times r} : \mathbf{U}^\top \mathbf{U} = \mathbf{I}_r, \|\mathbf{U}\|_{2,\infty} \leq \sqrt{\frac{r\mu}{n}} \right\}. \quad (20)$$

Below, we write $\mathbb{K}_{r,\mu}$ for $\mathbb{K}(n, r, \sqrt{\mu r/n})$. Lemma 4 lower bounds $\log \mathcal{M}(\mathbb{K}_{r,\mu}, d_{2,\infty}, \delta)$. We focus on the range $r/n \lesssim \delta^2 \lesssim \mu r/n$, as the lower bound in Equation (9) shows that $\delta^2 \ll r/n$ is not achievable when $|\lambda^*| = O(\sigma\sqrt{n})$ and $\delta^2 \gg \mu r/n$ is achieved by the trivial all-zeros estimate.

Lemma 4. Suppose that $n/\mu \geq \max\{4, r\}$ and $\mu \geq 12 \log(12n)$, and let $\delta > 0$ be such that

$$\frac{c_0^2 r}{8e^2 n} \leq \delta^2 \leq \frac{c_0^2 \mu r}{96e^2 n \log(12n/\mu)}, \quad (21)$$

where $c_0 > 0$ is a universal constant. Then when n is sufficiently large,

$$\log \mathcal{M}(\mathbb{K}_{r,\mu}, d_{2,\infty}, \delta) \gtrsim \frac{r^2}{\delta^2} \quad (22)$$

Remark 5. Lemma 4 applies when $\log n \lesssim \mu \lesssim n/r$. A more careful analysis might relax the lower bound, but when $\mu \lesssim \log n$, any $\mathbf{U} \in \mathbb{K}_{r,\mu}$ is nearly incoherent and Equation (9) is nearly optimal. An upper bound matching Lemma 4 up to log-factors can be found in the appendix.

Lemma 4 implies an improved lower bound compared to Equation (9). Consider the parameter space

$$\Omega(\lambda^*, \mu, r) = \{(\mathbf{\Lambda}^*, \mathbf{U}^*) : \mathbf{\Lambda}^* = \lambda^* \mathbf{I}_r, \mathbf{U}^* \in \mathbb{K}_{r,\mu}\}.$$

Using the Yang-Barron method [59], we obtain a lower bound for eigenspace estimation in the rank- r signal-plus-noise model for $|\lambda^*| = O(\sigma\sqrt{n})$. A detailed proof can be found in the appendix.

Theorem 2. Under Assumption 1 and the conditions of Lemma 4, for any $0 < \lambda^* \leq (6\sqrt{C_0})^{-1}\sigma\sqrt{n}$, where $C_0 > 0$ is a universal constant related to covering numbers of Grassmann manifolds, there is a universal constant $c > 0$ such that for all sufficiently large n ,

$$\inf_{\hat{\mathbf{U}} \in \mathbb{R}^{n \times r}} \sup_{(\mathbf{\Lambda}^*, \mathbf{U}^*) \in \Omega(\lambda^*, \mu, r)} \mathbb{E}_{\mathbf{\Lambda}^*, \mathbf{U}^*} d_{2,\infty}(\hat{\mathbf{U}}, \mathbf{U}^*) \geq c \left(\frac{\sigma\sqrt{r}}{\lambda^*} \wedge \sqrt{\frac{\mu r}{n \log(n/\mu)}} \right).$$

Remark 6. Theorem 2 removes the upper limit $\sqrt{r/n}$ from Equation (9), suggesting that μ only comes to bear when $\lambda^* \leq \sigma\sqrt{n}$. This regime is not well studied, as the BBP transition [8] implies that spectral methods fail, but other algorithms might achieve our lower bound. For example, signal detection is possible if structure is present [6]. Unfortunately, any such algorithm is likely to be computationally expensive, given the general belief that no polynomial-time algorithm can succeed when $\lambda^* \leq \sigma\sqrt{n}$ [39, 9]. We leave further exploration of this small- λ^* regime to future work.

Remark 7. The parameter space $\Omega(\lambda^*, \mu, r)$ considered in Theorem 2 contains only signal matrices with condition number $\kappa = 1$. In recent work, the authors have established lower bounds akin to Theorem 2 that show the role of condition number. A full accounting of the interplay between condition number and coherence is a promising area for future work.

5 Numerical experiments

We turn to a brief experimental exploration of our theoretical results. All experiments were run in a distributed environment on commodity hardware without GPUs. In total, the experiments reported below used 3425 compute-hours. Mean memory usage was 3.5 GB, with a maximum of 11 GB.

5.1 Simulations for rank-one eigenspace estimation

We begin with the rank-one setting considered in Algorithm 1, in which we observe

$$\mathbf{Y} = \mathbf{M}^* + \mathbf{W} = \lambda^* \mathbf{u}^* \mathbf{u}^{*\top} + \mathbf{W},$$

and wish to recover $\mathbf{u}^* \in \mathbb{S}^{n-1}$. We take $\lambda^* = \sqrt{n \log n}$ in all experiments, matching the rate in Remark 1. We consider three distributions for the entries of $\mathbf{W}$: Gaussian, Laplacian and Rademacher, all scaled to have variance $\sigma^2 = 1$. We consider two approaches to generating $\mathbf{u}^*$. In either case, we set a random entry of $\mathbf{u}^*$ to be $a \in \{0.3, 0.55, 0.8\}$, then generate the remaining entries by either

1. drawing uniformly from $\sqrt{1 - a^2} \mathbb{S}^{n-2}$, or
2. drawing uniformly from $\{\pm 1\}^{n-1}$ then normalizing these to have ℓ_2 norm $\sqrt{1 - a^2}$.

In both cases, $\|\mathbf{u}^*\|_\infty = a$ and $\mu = a^2 n$ with high probability. We take $a = \Theta(1)$, since in finite samples, $C n^{-1/2} \log n$ (which is nearly incoherent) is hard to discern from a constant (e.g., when $n = 20000$, $4n^{-1/2} \log n \approx 0.28$, nearly matching $a = 0.3$). Having generated $\mathbf{Y} = \mathbf{M}^* + \mathbf{W}$, we estimate $\mathbf{u}^*$ using both the spectral estimate $\hat{\mathbf{u}}$ and Algorithm 1 and measure their estimation error under d_∞ . We report the mean of 20 independent trials for each combination of problem size n , magnitude a and methods for generating $\mathbf{u}^*$ and $\mathbf{W}$. We vary n from 100 to 15100 in increments of 1000.

When running Algorithm 1, we use the debiased estimate $\hat{\lambda}_c$ from Equation (14). This requires an estimate of σ , for which we use the plug-in estimator in Equation (17). We set $\beta = 0$, eliminating Equation (16). Similar to Algorithm 2, we conjugate $\mathbf{Y}$ by a random orthogonal matrix $\mathbf{H} \in \mathbb{O}_n$. We expect the top eigenvector $\tilde{\mathbf{u}}$ of $\mathbf{H}\mathbf{Y}\mathbf{H}^\top$ to be approximately uniformly distributed on $\mathbb{S}^{n-1}$, as it is close to $\mathbf{H}\mathbf{u}^*$ (see Theorem 2.1 in [44]). Thus, the median of the absolute values of $\tilde{\mathbf{u}}$ should be $\Theta(n^{-1/2})$. Instead of selecting α_0 according to Equation (15), we set α_0 to be this median. The

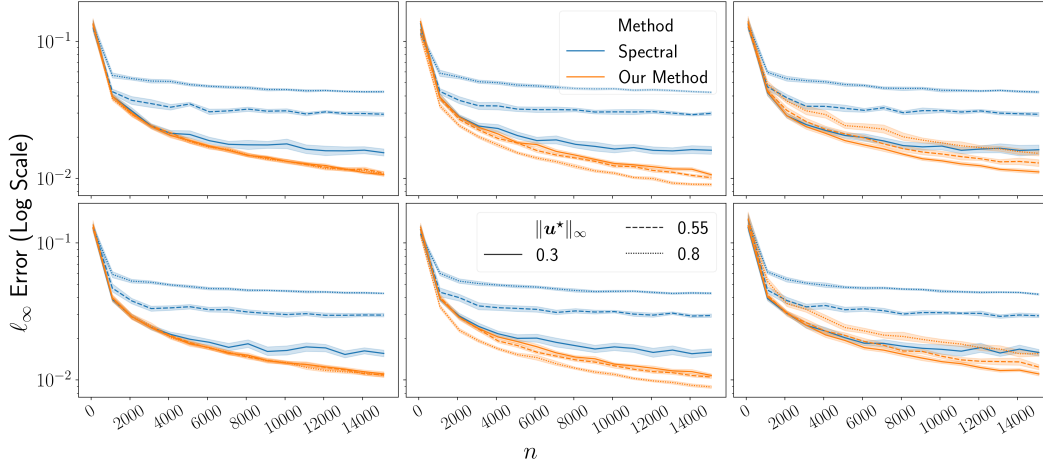


Figure 1: Estimation error measured in d_∞ as a function of dimension n , by the leading eigenvector (blue) and Algorithm 1 (orange), for $\|u^*\|_\infty$ equal to 0.8, 0.55 and 0.3 (dotted, dashed and solid lines, respectively). We consider u^* generated from the Bernoulli (top row) and Haar (bottom row) schemes, and we consider Gaussian (left), Rademacher (middle) and Laplacian (right) noise.

requirement in Equation (15) is then fulfilled, since there are roughly $n/2$ entries larger than the median. After obtaining $\hat{u}$ from Algorithm 1, we return $H^\top \hat{u}$ as our estimate of u^* . We note that this random rotation further serves to illustrate that Assumption 2 is merely a technical requirement: after a random rotation, with high probability, Hu^* does not satisfy Assumption 2.

Figure 1 compares the accuracy in estimating u^* using the leading eigenvector of Y (blue) and Algorithm 1 (orange) under the three noise settings and two generating procedures for u^* . Shaded bands indicate 95% bootstrap confidence intervals (CIs). Across settings, Algorithm 1 recovers u^* with a much smaller estimation error under d_∞ compared to the naïve spectral estimate, especially when $\|u^*\|_\infty$ (i.e., the coherence μ) is large. The spectral method degrades noticeably as coherence increases, while Algorithm 1 has far less dependence on μ . Indeed, under Gaussian noise (the first column of Figure 1), it has no visible dependence on μ . Under Rademacher noise (middle column of Figure 1), the dependence of Algorithm 1 on μ appears slightly reversed from that of the spectral estimator. Under Laplacian noise (right column of Figure 1), there seems to be a slight dependence on μ . Further examination in the appendix suggests that this is due to estimating the entries *other* than the largest element of u^* , and is likely asymptotically smaller than the rate in Theorem 1.

5.2 Simulations for rank- r eigenvector estimation

For the rank- r setting, we have a signal matrix with eigenvalues $|\lambda_1^*| \geq \dots \geq |\lambda_r^*|$. We observe

$$Y = M^* + W = U^* \Lambda^* U^{*\top} + W,$$

where $\Lambda^* = \text{diag}(\lambda_1^*, \dots, \lambda_r^*)$, and we wish to recover $U^* \in \mathbb{R}^{n \times r}$. We take $\sigma = 1$ and $\lambda_r^* = \sqrt{n \log n}$. Recovering each column of U^* separately requires an eigengap, defined for $k \in [r]$ as $\Delta_k := |\lambda_k^* - \lambda_{k+1}^*|$ and $\Delta_0 = \infty$. Typically, the estimation error of the k -th eigenvector has a $O(\min\{\Delta_k, \Delta_{k-1}\}^{-1})$ dependence on the eigengaps [57]. Here, we set $\Delta_k = 0.5\sqrt{n \log n}$, so $\lambda_k^* = 0.5(r - k + 2)\sqrt{n \log n}$ for all $k \in [r]$. As in the rank-one case, we generate entries of W from Gaussian, Laplacian and Rademacher distributions, all scaled to have unit variance. The true eigenvectors U^* are generated by repeating the following procedure for $k \in [r]$:

1. Randomly select an element of $v \in \mathbb{R}^n$ and set it to be $a \in \{0.3, 0.55, 0.8\}$.
2. Generate the rest of v by drawing uniformly from $\sqrt{1 - a^2} \mathbb{S}^{n-2}$.
3. Set $u_k^* = (I_n - U_{:,1:(k-1)}^* U_{:,1:(k-1)}^{*\top}) v$, normalize u_k^* to have unit ℓ_2 norm, and set $U_{:,k}^* = u_k^*$. If $k = 1$, then we take $U_{:,1:(k-1)}^* U_{:,1:(k-1)}^{*\top}$ to be the zero matrix.

Under this procedure, each u_k^* has coherence approximately $a^2 n$. Since the large entries of the u_k^* are unlikely to appear in the same rows, U^* also has coherence $a^2 n$. We take $r = 2$ here.

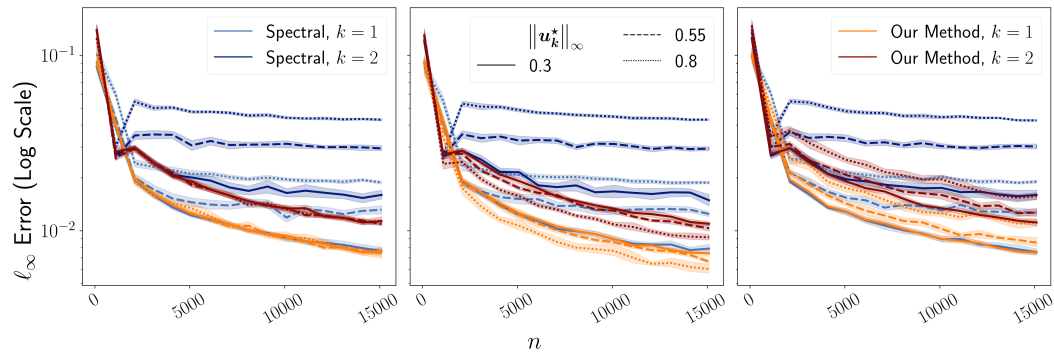


Figure 2: Estimation error under d_∞ as a function of size n , by the k -th eigenvector (blue/purple) and the estimator in Algorithm 2 (orange/red) for $\|u^*\|_\infty$ equal to 0.8 (dotted lines), 0.55 (dashed lines) or 0.3 (solid lines) with Gaussian (left), Rademacher (center) or Laplacian (right) noise.

Having generated $\mathbf{Y} = \mathbf{M}^* + \mathbf{W}$, we obtain estimates via the naïve spectral method and Algorithm 2, and measure their estimation error under d_∞ . We vary n from 100 to 15100 in increments of 1000 and report the mean of 20 independent trials for each combination of n , a and noise distribution. The results are summarized in Figure 2, showing the error for u_k^* , $k \in \{1, 2\}$ using the spectral estimate (blue/purple) and Algorithm 2 (orange/red), under Gaussian (left), Rademacher (middle) and Laplacian (right) noise. Shaded bands indicate 95% bootstrap CIs. In all settings, Algorithm 2 improves markedly on the spectral estimator. Algorithm 2 shows no visible dependence on μ under Gaussian noise. Under Rademacher noise, its μ -dependence is the reverse of the spectral estimator. Under Laplacian noise, it shows slight dependence on μ , which we again expect to be asymptotically smaller than the rate in Theorem 1. In the appendix, we consider $r = 3$ and measure error under $d_{2,\infty}$, and we again find that Algorithm 2 outperforms spectral estimators and is far less sensitive to μ .

5.3 Comparison with other methods

To the best of our knowledge, we are the first paper to consider the task of non-spectral entrywise eigenvector estimation. The nearest obvious competing method might be based on approximate message passing (AMP; see [34] for an overview). Such a comparison is included in the appendix, where our experiments indicate that while AMP performs well in recovering the signal eigenvectors as measured by ℓ_2 error, our method as specified in Algorithms 1 and 2 perform better under entrywise and $\ell_{2,\infty}$ error. Experimental details and further discussion can be found in the appendix.

6 Discussion, limitations and conclusion

We have presented new methods for eigenvector estimation in signal-plus-noise matrix models and new lower bounds for estimation rates in these models. The entrywise estimation error of our method has no dependence on the coherence μ for rank-one signal matrices, and achieves the optimal estimation rate up to log-factors. Simulations show that our method tolerates non-Gaussian noise and its extension to rank- r signal matrices has little dependence on μ . One limitation of our method is that it assumes homoscedastic noise. Future work will aim to relax this assumption and the technical condition in Assumption 2. In the rank- r case, Algorithm 2 estimates each eigenvector separately, requiring an eigengap. Future work will avoid this by simultaneously or iteratively estimating multiple eigenvectors. We note in closing that inequitable social impacts from abuse or misuse of models and methods are common, but we see no particular such impacts in the present work.

Acknowledgments and Disclosure of Funding

The authors gratefully acknowledge the support of the United States National Science Foundation via grants NSF DMS 2052918 and NSF DMS 2023239. KL was additionally supported by the University of Wisconsin–Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation.

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, K. Wang, and Y. Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *Annals of Statistics*, 48(3):1452, 2020.
- [3] J. Agterberg and J. Sulam. Entrywise recovery guarantees for sparse PCA via sparsistent algorithms. In *International Conference on Artificial Intelligence and Statistics*, pages 6591–6629. PMLR, 2022.
- [4] J. Agterberg and A. Zhang. Estimating higher-order mixed memberships via the $\ell_{2,\infty}$ tensor perturbation bound. *arXiv preprint arXiv:2212.08642*, 2022.
- [5] J. Agterberg, Z. Lubberts, and C. E. Priebe. Entrywise estimation of singular vectors of low-rank matrices with heteroskedasticity and dependence. *IEEE Transactions on Information Theory*, 68(7):4618–4650, 2022.
- [6] A. E. Alaoui, F. Krzakala, and M. Jordan. Fundamental limits of detection in the spiked Wigner model. *Annals of Statistics*, 48(2):863 – 885, 2020.
- [7] S. Athey, M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi. Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116(536):1716–1730, 2021.
- [8] J. Baik, G. B. 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.
- [9] J. Banks, C. Moore, R. Vershynin, N. Verzelen, and J. Xu. Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization. *IEEE Transactions on Information Theory*, 64(7):4872–4894, 2018.
- [10] Z. Bao, X. Ding, , and K. Wang. Singular vector and singular subspace distribution for the matrix denoising model. *Annals of Statistics*, 49(1):370 – 392, 2021.
- [11] Z. Bao, X. Ding, J. Wang, and K. Wang. Statistical inference for principal components of spiked covariance matrices. *Annals of Statistics*, 50(2):1144 – 1169, 2022.
- [12] F. Benaych-Georges and R. R. Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011. ISSN 0001-8708.
- [13] T. Bendokat, R. Zimmermann, and P.-A. Absil. A Grassmann manifold handbook: Basic geometry and computational aspects. *Advances in Computational Mathematics*, 50(1):1–51, 2024.
- [14] 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.
- [15] R. Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer Science & Business Media, 2013.
- [16] N. Boumal. *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, 2023.
- [17] W. Bryc and J. W. Silverstein. Singular values of large non-central random matrices. *Random Matrices: Theory and Applications*, 9(04):2050012, 2020.
- [18] 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. *Annals of Statistics*, 49(2):944 – 967, 2021.
- [19] T. Cai, H. Li, and R. Ma. Optimal structured principal subspace estimation: Metric entropy and minimax rates. *Journal of Machine Learning Research*, 22(46):1–45, 2021.
- [20] T. T. Cai and A. Zhang. Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *Annals of Statistics*, 46(1):60 – 89, 2018.
- [21] T. T. Cai, Z. Ma, and Y. Wu. Sparse PCA: Optimal rates and adaptive estimation. *Annals of Statistics*, 41(6):3074 – 3110, 2013.
- [22] J. Cape, M. Tang, and C. E. Priebe. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Annals of Statistics*, 47(5):2405 – 2439, 2019.

- [23] M. Capitaine, C. Donati-Martin, and D. Féral. The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *The Annals of Probability*, 37(1):1 – 47, 2009.
- [24] S. Chatterjee. Matrix estimation by universal singular value thresholding. *Annals of Statistics*, 43(1):177 – 214, 2015.
- [25] P. Chen, C. Gao, and A. Y. Zhang. Partial recovery for top-k ranking: optimality of MLE and suboptimality of the spectral method. *Annals of Statistics*, 50(3):1618–1652, 2022.
- [26] Y. Chen, C. Cheng, and J. Fan. Asymmetry helps: Eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices. *Annals of Statistics*, 49(1):435, 2021.
- [27] 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.
- [28] 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.
- [29] Y. Chikuse. *Statistics on Special Manifolds*, volume 174 of *Lecture Notes in Statistics*. Springer Science & Business Media, 2012.
- [30] D. Donoho and M. Gavish. Minimax risk of matrix denoising by singular value thresholding. *Annals of Statistics*, 42(6):2413–2440, 2014.
- [31] 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.
- [32] J. Fan, K. Wang, Y. Zhong, and Z. Zhu. Robust high dimensional factor models with applications to statistical machine learning. *Statistical Science: a review journal of the Institute of Mathematical Statistics*, 36(2):303, 2021.
- [33] J. Fan, Y. Fan, X. Han, and J. Lv. Asymptotic theory of eigenvectors for random matrices with diverging spikes. *Journal of the American Statistical Association*, 117(538):996–1009, 2022.
- [34] O. Y. Feng, R. Venkataramanan, C. Rush, and R. J. Samworth. A unifying tutorial on approximate message passing. *Foundations and Trends in Machine Learning*, 15(4):335–536, 2022.
- [35] C. Gao and Z. Ma. Minimax rates in network analysis: Graphon estimation, community detection and hypothesis testing. *Statistical Science*, 36(1):16 – 33, 2021. doi: 10.1214/19-STS736.
- [36] M. Gavish and D. L. Donoho. The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory*, 60(8):5040–5053, 2014.
- [37] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 4 edition, 2013.
- [38] F. Haddadi and A. Amini. Eigenvectors of deformed Wigner random matrices. *IEEE Transactions on Information Theory*, 67(2):1069–1079, 2021.
- [39] F. Krzakala, J. Xu, and L. Zdeborová. Mutual information in rank-one matrix estimation. In *2016 IEEE Information Theory Workshop (ITW)*, pages 71–75, 2016.
- [40] J. Lei and K. Z. Lin. Bias-adjusted spectral clustering in multi-layer stochastic block models. *Journal of the American Statistical Association*, pages 1–13, 2022.
- [41] L. Lei. Unified $\ell_{2 \rightarrow \infty}$ eigenspace perturbation theory for symmetric random matrices. *arXiv preprint arXiv:1909.04798*, 2019.
- [42] G. Li, C. Cai, H. V. Poor, and Y. Chen. Minimax estimation of linear functions of eigenvectors in the face of small eigen-gaps. *arXiv preprint arXiv:2104.03298*, 2021.
- [43] S. Negahban, S. Oh, and D. Shah. Iterative ranking from pair-wise comparisons. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.
- [44] S. O’Rourke, V. Vu, and K. Wang. Eigenvectors of random matrices: a survey. *Journal of Combinatorial Theory, Series A*, 144:361–442, 2016.

- [45] S. O’Rourke, V. Vu, and K. Wang. Matrices with gaussian noise: optimal estimates for singular subspace perturbation. *IEEE Transactions on Information Theory*, 2023.
- [46] A. Pajor. Metric entropy of the Grassmann manifold. In *Convex Geometric Analysis*, pages 181–188. Cambridge University Press, 1998.
- [47] M. Peng. Eigenvalues of deformed random matrices. *arXiv preprint arXiv:1205.0572*, 2012.
- [48] B. Recht. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12(12), 2011.
- [49] K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *Annals of Statistics*, 39(4):1878 – 1915, 2011.
- [50] 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.
- [51] J. W. Silverstein. The spectral radii and norms of large dimensional non-central random matrices. *Stochastic Models*, 10(3):525–532, 1994.
- [52] D. L. Sussman, M. Tang, and C. E. Priebe. Consistent latent position estimation and vertex classification for random dot product graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1): 48–57, 2013.
- [53] J. A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8(1–2):1–230, 2015.
- [54] R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.
- [55] V. Vu and J. Lei. Minimax sparse principal subspace estimation in high dimensions. *Annals of Statistics*, 41(6):2905 – 2947, 2013.
- [56] M. J. Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.
- [57] K. Wang. Analysis of singular subspaces under random perturbations. *arXiv preprint arXiv:2403.09170*, 2024.
- [58] D. Xia. Normal approximation and confidence region of singular subspaces. *Electronic Journal of Statistics*, 15(2):3798–3851, 2021.
- [59] Y. Yang and A. Barron. Information-theoretic determination of minimax rates of convergence. *Annals of Statistics*, pages 1564–1599, 1999.
- [60] Y. Yu, T. Wang, and R. J. Samworth. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2014.
- [61] A. R. Zhang, T. T. Cai, and Y. Wu. Heteroskedastic PCA: Algorithm, optimality, and applications. *Annals of Statistics*, 50(1):53–80, 2022.
- [62] Y. Zhou and Y. Chen. Deflated HeteroPCA: Overcoming the curse of ill-conditioning in heteroskedastic PCA. *arXiv preprint arXiv:2303.06198*, 2023.
- [63] Z. Zhou, F. Zhou, P. Li, and C.-H. Zhang. Rate-optimal subspace estimation on random graphs. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 20283–20294. Curran Associates, Inc., 2021.

A Proof of Lemma 1

Proof. Let $\theta = \lim_{n \rightarrow \infty} (\sigma\sqrt{n})^{-1}\lambda^*$. For any $\mathbf{u}^* \in \mathbb{S}^{n-1}$, by the Baik-Ben Arous-Péché (BBP) phase transition ([8]; see also Theorem 2 in [38]), the top eigenvector $\mathbf{u}$ of $\mathbf{Y}$ obeys

$$\langle \mathbf{u}, \mathbf{u}^* \rangle^2 \xrightarrow{\text{a.s.}} \begin{cases} 1 - \theta^{-2} & \text{if } \theta > 1 \\ 0 & \text{if } \theta \leq 1 \end{cases}. \quad (23)$$

Let $s = \lceil n/\mu \rceil$, and consider

$$\mathbf{u}^* = \left(\frac{1}{\sqrt{s}} \mathbf{1}_s^\top, \mathbf{0}_{n-s}^\top \right)^\top \in \mathbb{R}^n. \quad (24)$$

Without loss of generality, we assume that $d_\infty(\mathbf{u}, \mathbf{u}^*) = \|\mathbf{u} - \mathbf{u}^*\|_\infty$, since otherwise we can repeat the following argument with $-\mathbf{u}$ instead of $\mathbf{u}$. We have

$$d_\infty(\mathbf{u}, \mathbf{u}^*) \geq \max_{i \in [s]} |u_i - u_i^*| \geq \frac{1}{\sqrt{s}} \left(\sum_{i=1}^s |u_i - u_i^*|^2 \right)^{1/2}. \quad (25)$$

Expanding the term inside the square root,

$$\sum_{i=1}^s |u_i - u_i^*|^2 = \sum_{i=1}^s u_i^2 + \sum_{i=1}^s u_i^{*2} - 2 \sum_{i=1}^s u_i^* u_i = \sum_{i=1}^s u_i^2 + 1 - 2\langle \mathbf{u}^*, \mathbf{u} \rangle, \quad (26)$$

where the last equality follows from the construction of $\mathbf{u}^*$ in Equation (24). By the Cauchy-Schwarz inequality, we have

$$\sum_{i=1}^s u_i^2 \geq \frac{1}{s} \left(\sum_{i=1}^s u_i \right)^2 = \left(\sum_{i=1}^s \frac{1}{\sqrt{s}} u_i \right)^2 = \langle \mathbf{u}^*, \mathbf{u} \rangle^2, \quad (27)$$

where the last equality follows from the construction of $\mathbf{u}^*$ in Equation (24). Plugging Equation (27) into Equation (26), we obtain

$$\sum_{i=1}^s |u_i - u_i^*|^2 \geq \langle \mathbf{u}^*, \mathbf{u} \rangle^2 + 1 - 2\langle \mathbf{u}^*, \mathbf{u} \rangle = (1 - \langle \mathbf{u}^*, \mathbf{u} \rangle)^2.$$

When $\theta > 1$, applying Equation (23) yields that

$$\liminf_{n \rightarrow \infty} \sum_{i=1}^s |u_i - u_i^*|^2 \geq \liminf_{n \rightarrow \infty} (1 - \langle \mathbf{u}^*, \mathbf{u} \rangle)^2 = \left(1 - \sqrt{1 - \frac{1}{\theta^2}} \right)^2 \quad (28)$$

holds almost surely. Again using the fact that $\theta > 1$, we have

$$1 - \sqrt{1 - \frac{1}{\theta^2}} = \left(1 + \sqrt{1 - \frac{1}{\theta^2}} \right)^{-1} \left[1 - \left(1 - \frac{1}{\theta^2} \right) \right] \geq \frac{1}{2\theta^2}. \quad (29)$$

Combining Equations (25), (28) and (29), it holds almost surely that

$$\liminf_{n \rightarrow \infty} d_\infty(\mathbf{u}, \mathbf{u}^*) \geq \liminf_{n \rightarrow \infty} \frac{1}{\sqrt{s}} \left(|u_i - u_i^*|^2 \right)^{1/2} \geq \lim_{n \rightarrow \infty} \frac{1}{2\theta^2} \cdot \frac{1}{\sqrt{1 + n/\mu}},$$

where the last inequality follows from $s = \lceil n/\mu \rceil$. Since $\mu \leq n$, it follows that

$$\liminf_{n \rightarrow \infty} d_\infty(\mathbf{u}, \mathbf{u}^*) \geq \lim_{n \rightarrow \infty} \frac{1}{2\theta^2} \cdot \sqrt{\frac{\mu}{\mu + n}} \geq \lim_{n \rightarrow \infty} \frac{1}{2\sqrt{2}\theta^2} \cdot \sqrt{\frac{\mu}{n}}$$

almost surely. By the definition of θ , we have

$$\liminf_{n \rightarrow \infty} d_\infty(\mathbf{u}, \mathbf{u}^*) \geq \lim_{n \rightarrow \infty} \frac{\sigma^2 n}{2\sqrt{2}|\lambda^*|^2} \sqrt{\frac{\mu}{n}} = \lim_{n \rightarrow \infty} \frac{\sigma^2 \sqrt{n\mu}}{2\sqrt{2}|\lambda^*|^2},$$

completing the proof. $\square$

B Proof of Lemma 2

Proof. The upper bound

$$|\{i : |v_i| \geq \alpha_0\}| \leq \frac{1}{\alpha_0^2}$$

follows trivially from the fact that $\|\mathbf{v}\|_2 = 1$.

To prove the lower bound, we consider the disjoint intervals

$$\mathcal{J}_\ell = \begin{cases} [\log^{-1/2} n, 1] & \text{if } \ell = 0 \\ [\log^{-(\ell+1)/2} n, \log^{-\ell/2} n) & \text{if } \ell \in \{1, 2, \dots, \lceil L \rceil - 2\} \\ [\log^{-L/2} n, \log^{-(\lceil L \rceil - 1)/2} n) & \text{if } \ell = \lceil L \rceil - 1. \end{cases} \quad (30)$$

Recall the definition of L from Equation (11),

$$L = \frac{\log(2n)}{\log \log n}.$$

Rearranging terms, we have

$$-\frac{L}{2} \log \log n = -\frac{1}{2} \log(2n).$$

Exponentiating both sides of the above display,

$$\log^{-L/2} n = \frac{1}{\sqrt{2n}}.$$

Noting that by Equation (30),

$$\bigcup_{\ell=0}^{\lceil L \rceil - 1} \mathcal{J}_\ell = [\log^{-L/2} n, 1] = \left[\frac{1}{\sqrt{2n}}, 1 \right],$$

for any $i \in [n]$ such that $|v_i| \notin \bigcup_{\ell=0}^{\lceil L \rceil - 1} \mathcal{J}_\ell$, we have

$$|v_i| < \log^{-L/2} n = \frac{1}{\sqrt{2n}},$$

and therefore, summing over all $i \in [n]$ such that $|v_i| \notin \bigcup_{\ell=0}^{\lceil L \rceil - 1} \mathcal{J}_\ell$ yields that

$$\sum_{i: |v_i| \notin \bigcup_{\ell=0}^{\lceil L \rceil - 1} \mathcal{J}_\ell} v_i^2 < \sum_{i: |v_i| \notin \bigcup_{\ell=0}^{\lceil L \rceil - 1} \mathcal{J}_\ell} \frac{1}{2n} \leq n \cdot \frac{1}{2n} = \frac{1}{2}. \quad (31)$$

Let $x_\ell = |\{i : |v_i| \in \mathcal{J}_\ell\}|$ for all $\ell \in \{0, 1, \dots, \lceil L \rceil - 1\}$. We have

$$\sum_{i: |v_i| \in \mathcal{J}_\ell} v_i^2 \leq x_\ell \max_{i: |v_i| \in \mathcal{J}_\ell} v_i^2 \leq \frac{x_\ell}{\log^\ell n},$$

where the last inequality follows from the definition of $\mathcal{J}_\ell$ in Equation (30). Summing over all $\ell \in \{0, 1, \dots, \lceil L \rceil - 1\}$ on both sides, we have

$$\sum_{\ell=0}^{\lceil L \rceil - 1} \frac{x_\ell}{\log^\ell n} \geq \sum_{i: |v_i| \in \bigcup_{\ell=0}^{\lceil L \rceil - 1} \mathcal{J}_\ell} v_i^2 = 1 - \sum_{i: |v_i| \notin \bigcup_{\ell=0}^{\lceil L \rceil - 1} \mathcal{J}_\ell} v_i^2 > \frac{1}{2},$$

where the last inequality follows from Equation (31). The above further implies that

$$\lceil L \rceil \cdot \max_{\ell \in \{0, 1, \dots, \lceil L \rceil - 1\}} \left\{ \frac{x_\ell}{\log^\ell n} \right\} \geq \sum_{\ell=0}^{\lceil L \rceil - 1} \frac{x_\ell}{\log^\ell n} > \frac{1}{2}.$$

Hence, rearranging, there exists an $\ell_0 \in \{0, 1, \dots, \lceil L \rceil - 1\}$ such that

$$x_{\ell_0} > \frac{\log^{\ell_0} n}{2^{\lceil L \rceil}} > \log^{\ell_0-1} n = \frac{\log^{\ell_0+1} n}{\log^2 n},$$

where the second inequality holds for suitably large n by the definition of L given in Equation (11).

Finally, note that for all $i \in [n]$ such that $|v_i| \in \mathcal{J}_{\ell_0}$, we have $|v_i| \geq \log^{-(\ell_0+1)/2} n$ by the definition of $\mathcal{J}_{\ell_0}$ in Equation (30). Taking

$$\alpha_0 = \begin{cases} \log^{-(\ell_0+1)/2} n & \text{if } \ell_0 \in \{0, 1, \dots, \lceil L \rceil - 2\} \\ \log^{-L/2} n & \text{if } \ell_0 = \lceil L \rceil - 1 \end{cases}$$

yields that

$$|\{i : |v_i| \geq \alpha_0\}| \geq x_{\ell_0} > \frac{1}{\alpha_0^2 \log^2 n}.$$

We complete the proof by noting that α_0 is in $\mathcal{A}$, where $\mathcal{A}$ is defined in Equation (10). $\square$

C Proof of Lemma 3

Proof. Expanding the right hand side of Equation (17), we have

$$\begin{aligned} \hat{\sigma}^2 &= \frac{2}{n(n+1)} \sum_{1 \leq i \leq j \leq n} \left(M_{ij}^* - \widehat{M}_{ij} + W_{ij} \right)^2 \\ &= \frac{2}{n(n+1)} \sum_{1 \leq i \leq j \leq n} \left(M_{ij}^* - \widehat{M}_{ij} \right)^2 + \frac{2}{n(n+1)} \sum_{1 \leq i \leq j \leq n} W_{ij}^2 \\ &\quad + \frac{4}{n(n+1)} \sum_{1 \leq i \leq j \leq n} W_{ij} \left(M_{ij}^* - \widehat{M}_{ij} \right). \end{aligned}$$

Subtracting σ^2 on both sides and applying the triangle inequality, we have

$$\begin{aligned} |\hat{\sigma}^2 - \sigma^2| &\leq \frac{2}{n(n+1)} \sum_{1 \leq i \leq j \leq n} \left(M_{ij}^* - \widehat{M}_{ij} \right)^2 + \left| \frac{2}{n(n+1)} \sum_{1 \leq i \leq j \leq n} W_{ij}^2 - \sigma^2 \right| \\ &\quad + \left| \frac{4}{n(n+1)} \sum_{1 \leq i \leq j \leq n} W_{ij} \left(M_{ij}^* - \widehat{M}_{ij} \right) \right| \\ &\leq \frac{2}{n(n+1)} \left\| \mathbf{M}^* - \widehat{\mathbf{M}} \right\|_{\text{F}}^2 + \left| \frac{2}{n(n+1)} \sum_{1 \leq i \leq j \leq n} W_{ij}^2 - \sigma^2 \right| \\ &\quad + \left| \frac{2}{n(n+1)} \text{tr } \mathbf{W} \left(\mathbf{M}^* - \widehat{\mathbf{M}} \right) \right| + \frac{2}{n(n+1)} \left| \sum_{i=1}^n W_{ii} \left(M_{ii}^* - \widehat{M}_{ii} \right) \right|. \end{aligned} \tag{32}$$

We proceed to bound each term in the last inequality of Equation (32) separately.

For the first term in Equation (32), by an argument analogous to that of Equation (3.15) in [27],

$$\frac{2}{n(n+1)} \left\| \widehat{\mathbf{M}} - \mathbf{M}^* \right\|_{\text{F}}^2 \leq \frac{8r}{n(n+1)} \left\| \mathbf{W} \right\|^2.$$

By standard matrix concentration inequalities [53, 54], with probability at least $1 - O(n^{-8})$,

$$\left\| \mathbf{W} \right\| \leq 5\sqrt{\nu_W n \log n}. \tag{33}$$

Combining the above two displays, we obtain an analogue of Equation (3.16) in [27], namely that with probability at least $1 - O(n^{-8})$,

$$\frac{2}{n(n+1)} \left\| \widehat{\mathbf{M}} - \mathbf{M}^* \right\|_{\text{F}}^2 \leq \frac{200\nu_W r}{n+1}. \tag{34}$$

For the second term in Equation (32), applying standard concentration inequalities for subexponential random variables (see, e.g., Chapter 2 in [54]),

$$\mathbb{P} \left[\left| \sum_{1 \leq i \leq j \leq n} (W_{ij}^2 - \sigma^2) \right| \geq t \right] \leq 2 \exp \left\{ -c^2 \min \left\{ \frac{2t^2}{n(n+1)\nu_W^2}, \frac{t}{\nu_W} \right\} \right\},$$

where $c > 0$ is an absolute constant. Taking $t = 2\nu_W \sqrt{n(n+1) \log n}/c$ and dividing through by $n(n+1)/2$, it holds with probability at least $1 - O(n^{-8})$ that

$$\left| \frac{2}{n(n+1)} \sum_{1 \leq i \leq j \leq n} W_{ij}^2 - \sigma^2 \right| \leq \frac{4\nu_W \sqrt{\log n}}{cn}. \quad (35)$$

For the third term in Equation (32), using the matrix Hölder inequality (see Corollary IV.2.6 in [15]), we have

$$\left| \text{tr} \mathbf{W} (\mathbf{M}^* - \widehat{\mathbf{M}}) \right| \leq \|\mathbf{W}\| \|\mathbf{M}^* - \widehat{\mathbf{M}}\|_*,$$

where $\|\cdot\|_*$ denotes the nuclear norm, defined to be the sum of the singular values. Using the fact that $\mathbf{M}^* - \widehat{\mathbf{M}}$ has at most rank $2r$, we have

$$\left| \text{tr} \mathbf{W} (\mathbf{M}^* - \widehat{\mathbf{M}}) \right| \leq 2r \|\mathbf{W}\| \|\mathbf{M}^* - \widehat{\mathbf{M}}\|.$$

Applying Equation (33) and following an argument analogous to that in Equations (3.12) and (3.15) in [27], we have that with probability at least $1 - O(n^{-8})$,

$$\|\mathbf{W}\| \leq 5\sqrt{\nu_W n} \quad \text{and} \quad \|\mathbf{M}^* - \widehat{\mathbf{M}}\| \leq 10\sqrt{\nu_W n}.$$

It follows that with probability at least $1 - O(n^{-8})$,

$$\frac{2}{n(n+1)} \left| \text{tr} \mathbf{W} (\mathbf{M}^* - \widehat{\mathbf{M}}) \right| \leq \frac{200\nu_W r}{n+1}. \quad (36)$$

For the fourth term in Equation (32), we have

$$\begin{aligned} \left| \sum_{i=1}^n W_{ii} (M_{ii}^* - \widehat{M}_{ii}) \right| &\leq \max_{i \in [n]} \{|W_{ii}|\} \|\widehat{\mathbf{M}} - \mathbf{M}^*\|_* \\ \frac{2}{n(n+1)} \left| \sum_{i=1}^n W_{ii} (M_{ii}^* - \widehat{M}_{ii}) \right| &\leq \frac{200\nu_W r \sqrt{\log n}}{n^{3/2}}. \end{aligned} \quad (37)$$

Applying Equations (34), (35), (36) and (37) to Equation (32), with probability at least $1 - O(n^{-8})$,

$$|\widehat{\sigma}^2 - \sigma^2| \leq \frac{400\nu_W r}{n} + \frac{4\nu_W \sqrt{\log n}}{cn} + \frac{200\nu_W r \sqrt{\log n}}{n^{3/2}},$$

as we set out to show. $\square$

D Proof of Theorem 1

Without loss of generality, we derive our results under the assumption that

$$d_\infty(\mathbf{u}, \mathbf{u}^*) = \|\mathbf{u} - \mathbf{u}^*\|_\infty.$$

Otherwise, we can replace $\mathbf{u}$ with $-\mathbf{u}$ and obtain the same results. We remind the reader that we use c and C to denote constants with respect to n , whose precise value might change from line to line. To prove Theorem 1, we first state and prove a few technical lemmas.

D.1 Technical lemmas

Lemma 5. If $|\lambda^*| \geq 80\sqrt{\nu_W n}$, then under Assumption 1, it holds with probability at least $1 - O(n^{-8})$ that for all $\ell \in [n]$,

$$\min \{|u_\ell - u_\ell^*|, |u_\ell + u_\ell^*|\} \leq \frac{80\sqrt{\nu_W \log n} + 120\sqrt{\nu_W n}|u_\ell^*|}{|\lambda^*|}. \quad (38)$$

Proof. We largely follow the proof of Theorem 4.1 in [27], albeit with a slightly more careful analysis. In particular, we note that the proof of Theorem 4.1 in [27] is written for the case where $\mathbf{W}$ has Gaussian entries. It is straightforward to extend this argument to subgaussian entries by applying standard subgaussian concentration inequalities and truncation arguments [54, 56], in essence replacing all appearances of σ^2 in their results with the subgaussian parameter (i.e., “variance proxy”) ν_W . Here, we only sketch out a few key points and refer the reader to Section 4.1.4 of [27] for a full argument. For each $\ell \in [n]$, we construct a leave-one-out copy $\mathbf{Y}^{(\ell)}$ as

$$\mathbf{Y}^{(\ell)} = \lambda^* \mathbf{u}^* \mathbf{u}^{*\top} + \mathbf{W}^{(\ell)},$$

where

$$W_{ij}^{(\ell)} = \begin{cases} W_{ij}, & \text{if } \ell \notin \{i, j\} \\ 0 & \text{otherwise.} \end{cases}$$

For each $\ell \in [n]$, let $\lambda^{(\ell)}$ and $\mathbf{u}^{(\ell)}$ denote, respectively, the largest-magnitude eigenvalue of $\mathbf{Y}^{(\ell)}$ and its corresponding eigenvector. Without loss of generality, flipping signs if necessary, we assume that

$$d_\infty(\mathbf{u}, \mathbf{u}^*) = \|\mathbf{u} - \mathbf{u}^*\|_\infty.$$

By Equation (4.20) in [27], with probability at least $1 - O(n^{-8})$ we have

$$|u_\ell^{(\ell)} - u_\ell^*| \leq \frac{20\sqrt{\nu_W n}}{|\lambda^*|} |u_\ell^*|. \quad (39)$$

By Equation (4.15) in [27], we have

$$\|\mathbf{u} - \mathbf{u}^{(\ell)}\|_2 \leq \frac{4 \left\| (\mathbf{Y} - \mathbf{Y}^{(\ell)}) \mathbf{u}^{(\ell)} \right\|_2}{|\lambda^*|}.$$

Following the argument just after Equation (4.18) in [27], replacing the Gaussian concentration inequalities with subgaussian concentration inequalities, one can show that

$$\left\| (\mathbf{Y} - \mathbf{Y}^{(\ell)}) \mathbf{u}^{(\ell)} \right\|_2 \leq 5\sqrt{\nu_W \log n} + 5\sqrt{\nu_W n} |u_\ell| + 5\sqrt{\nu_W n} \|\mathbf{u} - \mathbf{u}^{(\ell)}\|_2,$$

with probability at least $1 - O(n^{-8})$. Combining the above two displays and rearranging slightly,

$$\left(1 - \frac{5\sqrt{\nu_W n}}{|\lambda^*|}\right) \|\mathbf{u} - \mathbf{u}^{(\ell)}\|_2 \leq \frac{5\sqrt{\nu_W \log n} + 5\sqrt{\nu_W n} |u_\ell|}{|\lambda^*|}.$$

Since $\lambda^* \geq 40\sqrt{\nu_W n}$ by assumption, it follows that

$$\|\mathbf{u} - \mathbf{u}^{(\ell)}\|_2 \leq \frac{40\sqrt{\nu_W \log n} + 40\sqrt{\nu_W n} |u_\ell|}{|\lambda^*|}. \quad (40)$$

Combining the triangle inequality with the trivial upper bound $\|\mathbf{u} - \mathbf{u}^{(\ell)}\|_\infty \leq \|\mathbf{u} - \mathbf{u}^{(\ell)}\|_2$, we have for any $\ell \in [n]$,

$$|u_\ell - u_\ell^*| \leq |u_\ell^* - u_\ell^{(\ell)}| + \|\mathbf{u} - \mathbf{u}^{(\ell)}\|_2.$$

Applying Equations (39) and (40), it holds with probability at least $1 - O(n^{-8})$ that

$$\begin{aligned} |u_\ell - u_\ell^*| &\leq \frac{20\sqrt{\nu_W n}}{|\lambda^*|} |u_\ell^*| + \frac{40\sqrt{\nu_W \log n} + 40\sqrt{\nu_W n} |u_\ell|}{|\lambda^*|} \\ &\leq \frac{20\sqrt{\nu_W n}}{|\lambda^*|} |u_\ell^*| + \frac{40\sqrt{\nu_W \log n} + 40\sqrt{\nu_W n} |u_\ell^*| + 40\sqrt{\nu_W n} |u_\ell - u_\ell^*|}{|\lambda^*|}, \end{aligned}$$

where the second inequality follows from the triangle inequality. After rearranging terms and using the fact that $|\lambda^*| \geq 80\sqrt{\nu_W n}$, we obtain the desired bound

$$|u_\ell - u_\ell^*| \leq \frac{80\sqrt{\nu_W \log n} + 120\sqrt{\nu_W n}|u_\ell^*|}{|\lambda^*|},$$

for all $\ell \in [n]$, completing the proof. $\square$

Lemma 6, which controls the behavior of the index set $\hat{I}$ in Algorithm 1, follows from an application of Lemma 5.

Lemma 6. *Under the model in Equation (1) with $M^* = \lambda^* u^* u^{*\top}$, suppose that Assumption 1 holds and Assumption 3 holds with constant $C_1 > 2400/\epsilon_0$, where $\epsilon_0 \in (0, 1)$ is fixed. For $\alpha \in \mathcal{A}$, define*

$$\hat{I}_\alpha = \{i : |u_i| \geq \alpha\},$$

where u is the leading eigenvector of Y , as in Step 1 of Algorithm 1. For all sufficiently large n , the following all hold with probability at least $1 - O(n^{-8} \log n)$:

$$\{i : |u_i^*| \geq (1 + \epsilon_0/2)\alpha\} \subseteq \hat{I}_\alpha \subseteq \{i : |u_i^*| \geq (1 - \epsilon_0/2)\alpha\}, \quad (41)$$

$$\{i : (1 - \epsilon_0/2)\alpha < |u_i| < (1 + \epsilon_0/2)\alpha\} \subseteq \{i : (1 - \epsilon_0)\alpha < |u_i^*| < (1 + \epsilon_0)\alpha\}, \quad (42)$$

$$\{i : |u_i^*| \geq \alpha\} \subseteq \{i : |u_i| > (1 - \epsilon_0/2)\alpha\} \quad (43)$$

and

$$\{i : |u_i| \geq (1 + \epsilon_0/2)\alpha\} \subseteq \{i : |u_i^*| \geq \alpha\}. \quad (44)$$

Proof. Fix $\alpha \in \mathcal{A}$. For any $i \in \hat{I}_\alpha$, we have $|u_i| \geq \alpha$ by definition. We have

$$\begin{aligned} & \mathbb{P} \left(\bigcup_{i \in \hat{I}_\alpha} \left\{ |u_i^*| < \alpha - \frac{80\sqrt{\nu_W \log n} + 120\sqrt{\nu_W n}|u_i^*|}{|\lambda^*|} \right\} \right) \\ & \leq \mathbb{P} \left(\bigcup_{i \in \hat{I}_\alpha} \left\{ |u_i^*| < |u_i| - \frac{80\sqrt{\nu_W \log n} + 120\sqrt{\nu_W n}|u_i^*|}{|\lambda^*|} \right\} \right) \\ & \leq \mathbb{P} \left(\bigcup_{i \in [n]} \left\{ |u_i^*| < |u_i| - \frac{80\sqrt{\nu_W \log n} + 120\sqrt{\nu_W n}|u_i^*|}{|\lambda^*|} \right\} \right) \leq O(n^{-8}). \end{aligned} \quad (45)$$

where the first inequality follows from the fact that $|u_i| \geq \alpha$ for $i \in \hat{I}_\alpha$, the second inequality follows from set inclusion, and the last inequality follows from combining the triangle inequality with Lemma 5. If

$$|u_i^*| \geq \alpha - \frac{80\sqrt{\nu_W \log n} + 120\sqrt{\nu_W n}|u_i^*|}{|\lambda^*|} \quad (46)$$

and we take $C_1 > 2400/\epsilon_0$ in Assumption 3, then we have

$$|u_i^*| \geq \alpha - \frac{80\epsilon_0\sqrt{\nu_W \log n}}{2400\sqrt{\nu_W n \log n}} - \frac{120\epsilon_0\sqrt{\nu_W n}}{2400\sqrt{\nu_W n \log n}}|u_i^*| \geq \alpha - \frac{\epsilon_0}{30}\sqrt{\frac{1}{n}} - \frac{\epsilon_0}{20\sqrt{\log n}}|u_i^*|.$$

Since $\alpha \geq \sqrt{1/2n}$, after rearranging terms,

$$\begin{aligned} |u_i^*| & \geq \left(1 + \frac{\epsilon_0}{20\sqrt{\log n}}\right)^{-1} \left(\alpha - \frac{\epsilon_0}{30}\sqrt{\frac{1}{n}}\right) \geq \left(1 + \frac{\epsilon_0}{20\sqrt{\log n}}\right)^{-1} \left(1 - \frac{\epsilon_0}{15}\right)\alpha \\ & > (1 - \epsilon_0/2)\alpha, \end{aligned} \quad (47)$$

where the last inequality holds for all $i \in [n]$ satisfying Equation (46) (i.e., all $i \in \hat{I}_\alpha$) when n is sufficiently large. Thus, we conclude that, applying a union bound followed by Equation (45),

$$\begin{aligned} \mathbb{P}\left(\hat{I}_\alpha \subseteq \{i : |u_i^*| \geq (1 - \epsilon_0/2)\alpha\}\right) &= \mathbb{P}\left(\bigcap_{i \in \hat{I}_\alpha} \{|u_i^*| \geq (1 - \epsilon_0/2)\alpha\}\right) \\ &\geq 1 - \mathbb{P}\left(\bigcup_{i \in \hat{I}_\alpha} \left\{|u_i^*| < \alpha - \frac{80\sqrt{\nu_W \log n} + 120\sqrt{\nu_W n}|u_i^*|}{|\lambda^*|}\right\}\right) \\ &\geq 1 - O(n^{-8}). \end{aligned} \quad (48)$$

Following a similar argument, we can show that Equation (44) holds with probability at least $1 - O(n^{-8})$, and that, also with probability at least $1 - O(n^{-8})$,

$$\{i : (1 - \epsilon_0/2)\alpha < |u_i|\} \subseteq \{i : (1 - \epsilon_0)\alpha < |u_i^*|\}. \quad (49)$$

On the other hand, combining the triangle inequality with the bound in Equation (38) in Lemma 5, it holds with probability at least $1 - O(n^{-8})$ that for all $i \in [n]$ satisfying $|u_i^*| \geq (1 + \epsilon_0/2)\alpha$,

$$|u_i| \geq |u_i^*| - |u_i - u_i^*| \geq \left(1 + \frac{\epsilon_0}{2}\right)\alpha - \frac{\epsilon_0}{30\sqrt{n}} - \frac{\epsilon_0(1 + \epsilon_0/2)\alpha}{20\sqrt{\log n}} > \alpha,$$

where the last inequality holds when n is sufficiently large. It follows that

$$\begin{aligned} \mathbb{P}\left(\hat{I}_\alpha \supseteq \{i : |u_i^*| \geq (1 + \epsilon_0/2)\alpha\}\right) &= \mathbb{P}\left(\bigcap_{i \in \{i : |u_i^*| \geq (1 + \epsilon_0/2)\alpha\}} \{|u_i| \geq \alpha\}\right) \\ &\geq 1 - O(n^{-8}). \end{aligned} \quad (50)$$

Following the same argument, we have that Equation (43) holds with probability at least $1 - O(n^{-8})$ and that

$$\{i : |u_i^*| \geq (1 + \epsilon_0)\alpha\} \subseteq \{i : |u_i| \geq (1 + \epsilon_0/2)\alpha\} \quad (51)$$

also with probability at least $1 - O(n^{-8})$.

Combining Equation (48) and Equation (50) yields that Equation (41) holds with probability at least $1 - O(n^{-8})$. Similarly, combining Equations (49) and (51) implies that Equation (42) holds with probability at least $1 - O(n^{-8})$. A union bound over all $\alpha \in \mathcal{A}$ yields that Equations (41), (42) (43) and (44) hold simultaneously for all $\alpha \in \mathcal{A}$ with probability at least $1 - O(|\mathcal{A}|n^{-8})$. Recalling from Equation (10) that $|\mathcal{A}| = O(\log n)$ completes the proof. $\square$

The results in Lemma 6 lead to Lemma 7.

Lemma 7. Consider the model in Equation (1) with $\mathbf{M}^* = \lambda^* \mathbf{u}^* \mathbf{u}^{*\top}$ and suppose that Assumptions 1, 2 and 3 hold. For each $\alpha \in \mathcal{A}$, let

$$I_\alpha := \{i : |u_i^*| \geq \alpha\} \quad (52)$$

and define the set

$$\mathcal{A}_0 = \{\alpha \in \mathcal{A} : |I_\alpha| \geq (\alpha^2 \log^2 n)^{-1}\}, \quad (53)$$

where $\mathcal{A}$ is as defined in Equation (10). Let $\hat{I}$ be the set constructed in Step 2 of Algorithm 1 with $\beta = \epsilon_0/2$. With probability at least $1 - O(n^{-8} \log n)$, there exists an $\alpha \in \mathcal{A}_0$ such that $\hat{I} = I_\alpha$.

Proof. We first show that with probability at least $1 - O(n^{-8} \log n)$, for any $\alpha \in \mathcal{A}$ satisfying Equations (15) and (16), we have $\alpha \in \mathcal{A}_0$ and $\hat{I}_\alpha = I_\alpha$, where $\hat{I}_\alpha$ is as defined in Lemma 6. We then prove that there exists at least one $\alpha \in \mathcal{A}_0 \subseteq \mathcal{A}$ such that both Equations (15) and (16) hold. Having established this, the value $\hat{\alpha} \in \mathcal{A}$ chosen in Step 2 of Algorithm 1, which satisfies Equations (15) and (16) by definition, must be an element of $\mathcal{A}_0 \subseteq \mathcal{A}$. Since $\hat{I} = \hat{I}_{\hat{\alpha}}$ by definition, we must have $\hat{I} = I_{\hat{\alpha}}$ with probability at least $1 - O(n^{-8} \log n)$, which will complete the proof.

Suppose that $\alpha \in \mathcal{A}$ satisfies Equations (15) and (16). Trivially, by the definition of $\mathcal{A}_0$ in Equation (53), Equation (15) implies that $\alpha \in \mathcal{A}_0$. By Lemma 6, with probability at least $1 - O(n^{-8} \log n)$, Equations (43) and (44) hold for all elements of $\mathcal{A}$. Thus, in particular,

$$\{i : |u_i| \geq (1 + \epsilon_0/2)\hat{\alpha}\} \subseteq \{i : |u_i^*| \geq \hat{\alpha}\} \subseteq \{i : |u_i| > (1 - \epsilon_0/2)\hat{\alpha}\}.$$

By Equation (16) and the choice $\beta = \epsilon_0/2$ in Algorithm 1,

$$\hat{I}_\alpha = \{i : |u_i| > (1 - \epsilon_0/2)\alpha\} = \{i : |u_i| \geq (1 + \epsilon_0/2)\alpha\},$$

from which it follows that $\hat{I}_\alpha = I_\alpha$ with probability at least $1 - O(n^{-8} \log n)$.

We now establish the existence of at least one $\alpha \in \mathcal{A}_0 \subseteq \mathcal{A}$ satisfying Equations (15) and (16). By Assumption 2, for sufficiently large n , there exists $\alpha \in \mathcal{A}$ such that Equation (12) holds, from which $\alpha \in \mathcal{A}_0$ immediately. Also by Assumption 2, there are no $i \in [n]$ for which

$$(1 - \epsilon_0)\alpha \leq |u_i^*| \leq (1 + \epsilon_0)\alpha.$$

By Lemma 6, Equation (41) holds with probability at least $1 - O(n^{-8} \log n)$ for all elements of $\mathcal{A}_0 \subseteq \mathcal{A}$, and thus $\hat{I}_\alpha = I_\alpha$. Therefore, we have $|\hat{I}_\alpha| = |I_\alpha| \geq (\alpha^2 \log^2 n)^{-1}$, where the lower-bound follows from Equation (12). As a result, any such α guaranteed by Assumption 2 satisfies Equation (15).

By Equation (42) in Lemma 6 and the fact that α satisfies Assumption 2, with probability at least $1 - O(n^{-8} \log n)$ we have

$$\{i : (1 - \epsilon_0/2)\alpha < |u_i| < (1 + \epsilon_0/2)\alpha\} = \emptyset.$$

Thus, with $\beta = \epsilon_0/2$, this particular α also satisfies Equation (16) and we conclude that with probability at least $1 - O(n^{-8} \log n)$, there exists at least one $\alpha \in \mathcal{A}_0$ such that both Equations (15) and (16) hold.

Our argument above ensures that with probability at least $1 - O(n^{-8} \log n)$, there exists $\alpha \in \mathcal{A}_0$ satisfying Equations (15) and (16). Thus, $\hat{\alpha} \in \mathcal{A}_0$ as constructed in Step 2 exists, and our argument above ensures that $\hat{I}$ as constructed in Step 2 satisfies $\hat{I} = \hat{I}_{\hat{\alpha}} = I_{\hat{\alpha}}$, completing the proof. $\square$

Lemma 8. *Under Assumptions 1 and 3, with probability at least $1 - O(n^{-8})$ it holds for all $i \in [n]$ that*

$$|u_i^*| \geq \frac{1}{24\sqrt{n}} \quad \text{implies} \quad \text{sgn}(u_i) = \text{sgn}(u_i^*).$$

Proof of Lemma 8. Let $i \in [n]$ and suppose that $|u_i^*| \geq 1/24\sqrt{n}$ with $u_i^* \geq 0$. The case for $u_i^* < 0$ follows by an analogous argument and details are omitted. By Equation (38) in Lemma 5, it holds with probability at least $1 - O(n^{-8})$ that for all $i \in [n]$ such that $u_i^* \geq 0$,

$$u_i \geq u_i^* - |u_i - u_i^*| \geq \left(1 - \frac{120\sqrt{\nu_W n}}{|\lambda^*|}\right) u_i^* - \frac{80\sqrt{\nu_W \log n}}{|\lambda^*|}.$$

By Assumption 3, it follows that when $|\lambda^*| \geq 2400\sqrt{\nu_W n}$

$$u_i \geq \frac{19}{20}u_i^* - \frac{80\sqrt{\nu_W \log n}}{|\lambda^*|}.$$

Since $u_i^* \geq 1/24\sqrt{n}$ by assumption, it follows that when $|\lambda^*| \geq 2400\sqrt{\nu_W n}$, we have

$$u_i^* \geq \frac{100\sqrt{\nu_W \log n}}{|\lambda^*|},$$

from which we have $u_i > 0$ and therefore $\text{sgn}(u_i) = \text{sgn}(u_i^*)$. $\square$

Lemma 9. *Fix $\alpha \in \mathcal{A}$ and $\mathbf{s} \in \{\pm 1\}^n$. For a random noise matrix $\mathbf{W} \in \mathbb{R}^{n \times n}$ satisfying Assumption 1, it holds with probability at least $1 - O(n^{-8})$ that*

$$\left| \sum_{j,k \in I_\alpha} s_j s_k W_{jk} \right| \leq C |I_\alpha| \sqrt{\nu_W \log n} \quad (54)$$

and

$$\max_{j \in [n]} \left| \sum_{k \in I_\alpha} s_j s_k W_{jk} \right| \leq C \sqrt{|I_\alpha| \nu_W \log n}. \quad (55)$$

Proof. Since $\mathbf{s} \in \{\pm 1\}^n$ is fixed and the entries of $\mathbf{W}$ are symmetric about zero, we assume without loss of generality that $\mathbf{s}$ is a vector of all 1's. By standard concentration inequalities [54], we have that for any $t > 0$,

$$\mathbb{P} \left(\left| \sum_{j,k \in I_\alpha} W_{jk} \right| \geq t \right) \leq 2 \exp \left(\frac{-t^2}{2|I_\alpha|^2 \nu_W} \right).$$

Setting $t = C|I_\alpha| \sqrt{\nu_W \log n}$ for suitably-chosen $C > 0$, Equation (54) holds with probability at least $1 - O(n^{-8})$. Similarly, by standard subgaussian concentration inequalities [54] and a union bound over $j \in [n]$, it holds for all $t > 0$ that

$$\mathbb{P} \left(\max_{j \in [n]} \left| \sum_{k \in I_\alpha} W_{jk} \right| > t \right) \leq 2n \exp \left(\frac{-t^2}{2|I_\alpha| \nu_W} \right).$$

Taking $t^2 = C|I_\alpha| \nu_W \log n$ for $C > 0$ chosen suitably large, Equation (55) holds with probability at least $1 - O(n^{-8})$, completing the proof. $\square$

D.2 Proof of Theorem 1

Proof. In this proof, we control the estimation error of $\hat{\mathbf{u}}$ under d_∞ . There are two types of entries of $\hat{\mathbf{u}}$, as stated in Step 5 of Algorithm 1. One type is derived from $\hat{\mathbf{v}}$ constructed in Step 4 of Algorithm 1, and corresponds to large entries of $\mathbf{u}^*$. The other type is given by the top eigenvector $\mathbf{u}$ of $\mathbf{Y}$, corresponding to small entries of $\mathbf{u}^*$. We handle these two types of entries separately. In **Part I** of the proof, we control the estimation error corresponding to the first type of entries, derived from $\hat{\mathbf{v}}$. In **Part II** of the proof, we control the estimation error corresponding to the second type of entries, derived from $\mathbf{u}$. Combining these two parts, we obtain a high probability bound for the estimation error $d_\infty(\hat{\mathbf{u}}, \mathbf{u}^*)$.

Part I. Estimation error related to $\hat{\mathbf{v}}$.

We define the event $\mathcal{E}_1$ according to

$$\mathcal{E}_1 = \{|\hat{\lambda} - \lambda^*| \leq C_2 \sqrt{\nu_W} \log^{5/2} n\}. \quad (56)$$

On event $\mathcal{E}_1$, if $\lambda^* > 0$, then by the triangle inequality, we have

$$\hat{\lambda} \geq \lambda^* - |\lambda^* - \hat{\lambda}| \geq \lambda^* - C_2 \sqrt{\nu_W} \log^{5/2} n > 0, \quad (57)$$

where the second inequality holds for all sufficiently large n by Assumption 3. Similarly, for all sufficiently large n we have $\hat{\lambda} < 0$ if $\lambda^* < 0$. In other words, for all sufficiently large n , we have

$$\text{sgn}(\hat{\lambda}) = \text{sgn}(\lambda^*)$$

on event $\mathcal{E}_1$. Hence, Step 1 in Algorithm 1 guarantees that we work with the original $\mathbf{Y}$ when $\lambda^* > 0$, or with $-\mathbf{Y}$ when $\lambda^* < 0$. In either case, the algorithm proceeds with a signal matrix that has a positive leading eigenvalue on $\mathcal{E}_1$. Without loss of generality, we assume that $\lambda^* > 0$.

On event $\mathcal{E}_1$, we have

$$\left| \frac{\lambda^*}{\hat{\lambda}} - 1 \right| \leq \frac{C_2 \sqrt{\nu_W} (\log n)^{5/2}}{\lambda^* - |\lambda^* - \hat{\lambda}|} \leq \frac{C \sqrt{\nu_W} (\log n)^{5/2}}{\lambda^*}. \quad (58)$$

where the second inequality holds for all sufficiently large n by Assumption 3. Similarly, for n sufficiently large, following Equation (58), event $\mathcal{E}_1$ also implies

$$1/2 \leq |\lambda^* / \hat{\lambda}| \leq 2. \quad (59)$$

Recalling the set $\mathcal{A}_0$ from Equation (53) and I_α from Equation (52) above, for each $\alpha \in \mathcal{A}_0$, define the events $\mathcal{E}_{2,\alpha}$ and $\mathcal{E}_{3,\alpha}$ according to

$$\mathcal{E}_{2,\alpha} := \{\hat{I} = I_\alpha\}, \quad (60)$$

where $\hat{I}$ is as constructed in Step 2 of Algorithm 1, and

$$\mathcal{E}_{3,\alpha} := \{Q_{kk} = \text{sgn}(u_k^*) \text{ for all } k \in I_\alpha, \text{ and } Q_{kk} = 1 \text{ for all } k \in I_\alpha^c\}. \quad (61)$$

Define $s \in \{\pm 1\}^n$ according to

$$s_k := \begin{cases} \text{sgn}(u_k^*) & \text{if } k \in I_\alpha, \\ 1 & \text{otherwise,} \end{cases} \quad (62)$$

and for each $\alpha \in \mathcal{A}_0$, define the event $\mathcal{E}_{4,\alpha}$ as

$$\mathcal{E}_{4,\alpha} = \left\{ \left| \sum_{j,k \in I_\alpha} s_j s_k W_{jk} \right| \leq C |I_\alpha| \sqrt{\nu_W \log n} \right\}. \quad (63)$$

For $\hat{S}$ given in Step 4 of Algorithm 1, plugging in $\tilde{Y} = QYQ$ and expanding, it follows that on the event $\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha}$,

$$\hat{S}^2 = \sum_{j,k \in \hat{I}} Q_{jj} Q_{kk} Y_{jk} = \lambda^* \left(\sum_{k \in I_\alpha} |u_k^*| \right)^2 + \sum_{j,k \in I_\alpha} s_j s_k W_{jk}. \quad (64)$$

On the event $\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha}$, we can lower-bound this right-hand side, obtaining

$$\hat{S}^2 \geq \lambda^* \left(\sum_{k \in I_\alpha} |u_k^*| \right)^2 - C |I_\alpha| \sqrt{\nu_W \log n} \geq \lambda^* |I_\alpha|^2 \alpha^2 - C |I_\alpha| \sqrt{\nu_W \log n}, \quad (65)$$

where the first inequality follows from the fact that event $\mathcal{E}_{4,\alpha}$ holds and the second inequality follows from the definition of I_α in Equation (52). Using Equation (53) and the fact that $\alpha \in \mathcal{A}_0$, we may further bound this by

$$\hat{S}^2 \geq \frac{|I_\alpha|}{\log^2 n} \left(\lambda^* - C \sqrt{\nu_W \log^5 n} \right) > 0,$$

where the second inequality holds under Assumption 3 when n is sufficiently large. Thus, on the event $\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha}$, we have $\sum_{j,k \in \hat{I}} \tilde{Y}_{jk} > 0$ and thus $\hat{S}$ is well-defined (i.e., it is the square root of a positive number).

We now show that $\hat{S}$ is close to its population target. Rearranging the terms in Equation (64), we have that on the event $\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha}$,

$$\left| \hat{S}^2 - \lambda^* \left(\sum_{k \in I_\alpha} |u_k^*| \right)^2 \right| \leq C |I_\alpha| \sqrt{\nu_W \log n}. \quad (66)$$

Dividing both sides by $\hat{S} + \sqrt{\lambda^*} \sum_{k \in I_\alpha} |u_k^*|$, when the event $\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha}$ holds, we have

$$\left| \hat{S} - \sqrt{\lambda^*} \sum_{k \in I_\alpha} |u_k^*| \right| \leq \frac{C |I_\alpha| \sqrt{\nu_W \log n}}{\left(\hat{S} + \sqrt{\lambda^*} \sum_{k \in I_\alpha} |u_k^*| \right)} \leq \frac{C |I_\alpha| \sqrt{\nu_W \log n}}{\sqrt{\lambda^*} \sum_{k \in I_\alpha} |u_k^*|} \leq \frac{C \sqrt{\nu_W \log n}}{\sqrt{\lambda^*} \alpha}, \quad (67)$$

where the second inequality follows from the fact that $\hat{S} > 0$ and the last inequality follows from the definition of I_α in Equation (52). Dividing by $\sqrt{\lambda^*} \sum_{k \in I_\alpha} |u_k^*|$ on both sides (note that this quantity is positive by definition of I_α and the fact that $\alpha \in \mathcal{A}_0$), it follows that on the event $\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha}$,

$$\left| \frac{\hat{S}}{\sqrt{\lambda^*} \sum_{k \in I_\alpha} |u_k^*|} - 1 \right| \leq \frac{C \sqrt{\nu_W \log n}}{\lambda^* \alpha \sum_{k \in I_\alpha} |u_k^*|} \leq \frac{C \sqrt{\nu_W \log n}}{\lambda^* |I_\alpha| \alpha^2} \leq \frac{C \sqrt{\nu_W} (\log n)^{5/2}}{\lambda^*}, \quad (68)$$

where the second inequality follows from the definition of I_α in Equation (52) and the last inequality follows from Equation (53).

We move on to Step 4 of Algorithm 1, from which we recall, for ease of reference, that

$$\hat{v}_j = \frac{\sum_{k \in \hat{I}} \tilde{Y}_{jk}}{\sqrt{\hat{\lambda}} \hat{S}} \quad \text{for } j \in [n]. \quad (69)$$

Note that by Equation (57), the square root of $\hat{\lambda}$ is well-defined on $\mathcal{E}_1$. Rearranging the definition of $\hat{v}_j$ in Equation (69), plugging in $\tilde{Y} = \mathbf{QYQ}$ and expanding, we have for $j \in [n]$,

$$\frac{\hat{v}_j \hat{S}}{\sqrt{\lambda^*}} = \sqrt{\frac{\lambda^*}{\hat{\lambda}}} Q_{jj} u_j^* \sum_{k \in \hat{I}} Q_{kk} u_k^* + \frac{\sum_{k \in \hat{I}} Q_{jj} Q_{kk} W_{jk}}{\sqrt{\hat{\lambda}} \lambda^*}.$$

Rearranging terms and adding $\hat{v}_j \sum_{k \in \hat{I}} Q_{kk} u_k^*$ to both sides, we have for any $j \in [n]$

$$\begin{aligned} (\hat{v}_j - Q_{jj} u_j^*) \sum_{k \in \hat{I}} Q_{kk} u_k^* &= \hat{v}_j \left(\sum_{k \in \hat{I}} Q_{kk} u_k^* - \frac{\hat{S}}{\sqrt{\lambda^*}} \right) \\ &\quad + \left(\sqrt{\frac{\lambda^*}{\hat{\lambda}}} - 1 \right) Q_{jj} u_j^* \sum_{k \in \hat{I}} Q_{kk} u_k^* + \frac{\sum_{k \in \hat{I}} Q_{jj} Q_{kk} W_{jk}}{\sqrt{\hat{\lambda}} \lambda^*}. \end{aligned}$$

Dividing by $\sum_{k \in \hat{I}} Q_{kk} u_k^*$ on both sides, again noting that this quantity is positive on the event $\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha}$,

$$\hat{v}_j - Q_{jj} u_j^* = \hat{v}_j \left(1 - \frac{\hat{S}}{\sqrt{\lambda^*} \sum_{k \in \hat{I}} Q_{kk} u_k^*} \right) + \left(\sqrt{\frac{\lambda^*}{\hat{\lambda}}} - 1 \right) Q_{jj} u_j^* + \frac{\sum_{k \in \hat{I}} Q_{jj} Q_{kk} W_{jk}}{\sqrt{\hat{\lambda}} \lambda^* \sum_{k \in \hat{I}} Q_{kk} u_k^*}. \quad (70)$$

Taking absolute values in Equation (70) and applying the triangle inequality,

$$|\hat{v}_j - Q_{jj} u_j^*| \leq |\hat{v}_j| \left| 1 - \frac{\hat{S}}{\sqrt{\lambda^*} \sum_{k \in \hat{I}} Q_{kk} u_k^*} \right| + \left| \sqrt{\frac{\lambda^*}{\hat{\lambda}}} - 1 \right| |Q_{jj} u_j^*| + \left| \frac{\sum_{k \in \hat{I}} Q_{jj} Q_{kk} W_{jk}}{\sqrt{\hat{\lambda}} \lambda^* \sum_{k \in \hat{I}} Q_{kk} u_k^*} \right|.$$

Trivially, $|Q_{jj} u_j^*| \leq 1$. Further, on the event $\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha}$, we have $\hat{I} = I_\alpha$ and $Q_{jj} = \text{sgn}(u_j^*) = s_j$ for all $j \in I_\alpha$. Thus, we may write

$$|\hat{v}_j - Q_{jj} u_j^*| \leq |\hat{v}_j| \left| 1 - \frac{\hat{S}}{\sqrt{\lambda^*} \sum_{k \in I_\alpha} |u_k^*|} \right| + \left| \sqrt{\frac{\lambda^*}{\hat{\lambda}}} - 1 \right| + \left| \frac{\sum_{k \in I_\alpha} s_j s_k W_{jk}}{\lambda^* \sum_{k \in I_\alpha} |u_k^*|} \right| \sqrt{\frac{\lambda^*}{\hat{\lambda}}}. \quad (71)$$

On the event $\mathcal{E}_1$, applying Equations (58) and (59) yields that

$$\left| \sqrt{\frac{\lambda^*}{\hat{\lambda}}} - 1 \right| = \left| \frac{\frac{\lambda^*}{\hat{\lambda}} - 1}{\sqrt{\frac{\lambda^*}{\hat{\lambda}}} + 1} \right| \leq \frac{C \sqrt{\nu_W} \log^{5/2} n}{\lambda^*}. \quad (72)$$

Applying this bound in Equation (71), on the event $\mathcal{E}_1 \cap \mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha}$, we have

$$|\hat{v}_j - Q_{jj} u_j^*| \leq |\hat{v}_j| \left| 1 - \frac{\hat{S}}{\sqrt{\lambda^*} \sum_{k \in I_\alpha} |u_k^*|} \right| + \frac{C \sqrt{\nu_W} \log^{5/2} n}{\lambda^*} + \left| \frac{\sum_{k \in I_\alpha} s_j s_k W_{jk}}{\lambda^* \sum_{k \in I_\alpha} |u_k^*|} \right| \sqrt{\frac{\lambda^*}{\hat{\lambda}}}.$$

We remind the reader that in the proof, constant C may change its precise value from line to line. Another application of Equation (72) and using Assumption 3 yields

$$|\hat{v}_j - Q_{jj} u_j^*| \leq |\hat{v}_j| \left| 1 - \frac{\hat{S}}{\sqrt{\lambda^*} \sum_{k \in I_\alpha} |u_k^*|} \right| + \frac{C \sqrt{\nu_W} \log^{5/2} n}{\lambda^*} + C \left| \frac{\sum_{k \in I_\alpha} s_j s_k W_{jk}}{\lambda^* \sum_{k \in I_\alpha} |u_k^*|} \right|.$$

On the event $\mathcal{E}_1 \cap \mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha}$, Equation (68) holds and we have

$$|\hat{v}_j - Q_{jj} u_j^*| \leq |\hat{v}_j| \frac{C \sqrt{\nu_W} \log^{5/2} n}{\lambda^*} + \frac{C \sqrt{\nu_W} \log^{5/2} n}{\lambda^*} + C \left| \frac{\sum_{k \in I_\alpha} s_j s_k W_{jk}}{\lambda^* \sum_{k \in I_\alpha} |u_k^*|} \right|. \quad (73)$$

Define the event

$$\mathcal{E}_{5,\alpha} = \left\{ \max_{j \in [n]} \left| \sum_{k \in I_\alpha} s_j s_k W_{jk} \right| \leq C \sqrt{|I_\alpha| \nu_W \log n} \right\}. \quad (74)$$

Under the event $\mathcal{E}_1 \cap \mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha} \cap \mathcal{E}_{5,\alpha}$, we can further bound Equation (73) by

$$|\hat{v}_j - Q_{jj} u_j^*| \leq |\hat{v}_j| \frac{C \sqrt{\nu_W} \log^{5/2} n}{\lambda^*} + \frac{C \sqrt{\nu_W} \log^{5/2} n}{\lambda^*} + \frac{C \sqrt{|I_\alpha| \nu_W \log n}}{\lambda^* |I_\alpha| \alpha}.$$

Applying the definition of $\mathcal{A}_0$ in Equation (53), on the event $\mathcal{E}_1 \cap \mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha} \cap \mathcal{E}_{5,\alpha}$, it holds for all $j \in [n]$ that

$$|\hat{v}_j - Q_{jj} u_j^*| \leq |\hat{v}_j| \frac{C \sqrt{\nu_W} \log^{5/2} n}{\lambda^*} + \frac{C \sqrt{\nu_W} \log^{5/2} n}{\lambda^*}. \quad (75)$$

Applying the triangle inequality and using the fact that $|Q_{jj} u_j^*| \leq 1$ by definition,

$$\begin{aligned} (|\hat{v}_j| + 1) \frac{C \sqrt{\nu_W} \log^{5/2} n}{\lambda^*} &\leq |\hat{v}_j - Q_{jj} u_j^*| \frac{C \sqrt{\nu_W} \log^{5/2} n}{\lambda^*} + (|Q_{jj} u_j^*| + 1) \frac{C \sqrt{\nu_W} \log^{5/2} n}{\lambda^*} \\ &\leq |\hat{v}_j - Q_{jj} u_j^*| \frac{C \sqrt{\nu_W} \log^{5/2} n}{\lambda^*} + \frac{C \sqrt{\nu_W} \log^{5/2} n}{\lambda^*}. \end{aligned}$$

Applying this bound to Equation (75),

$$|\hat{v}_j - Q_{jj} u_j^*| \leq |\hat{v}_j - Q_{jj} u_j^*| \frac{C \sqrt{\nu_W} (\log n)^{5/2}}{\lambda^*} + \frac{C \sqrt{\nu_W} (\log n)^{5/2}}{\lambda^*}.$$

Rearranging the terms, under Assumption 3 for n sufficiently large and when the event $\mathcal{E}_1 \cap \mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha} \cap \mathcal{E}_{5,\alpha}$ holds, we have for all $j \in [n]$,

$$|\hat{v}_j - Q_{jj} u_j^*| \leq \frac{C \sqrt{\nu_W} (\log n)^{5/2}}{\lambda^*}. \quad (76)$$

Recall from Step 5 of Algorithm 1 that we set $\hat{u}_j = Q_{jj} \hat{v}_j$ for all $j \in [n]$ such that $|u_j| > (\sigma/\hat{\lambda}) \log n$. For any such j , on the event $\mathcal{E}_1 \cap \mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha} \cap \mathcal{E}_{5,\alpha}$, Equation (76) holds, and we have

$$|\hat{u}_j - u_j^*| = |Q_{jj} \hat{v}_j - Q_{jj} u_j^*| = |Q_{jj}| |\hat{v}_j - Q_{jj} u_j^*| \leq \frac{C \sqrt{\nu_W} \log^{5/2} n}{\lambda^*}. \quad (77)$$

Consider the event

$$\mathcal{E} = \left\{ \|\mathbf{Q} \hat{\mathbf{v}} - \mathbf{u}^*\|_\infty \leq \frac{C \sqrt{\nu_W} \log^{5/2} n}{\lambda^*} \right\}.$$

Recalling the events $\mathcal{E}_1, \mathcal{E}_{2,\alpha}, \mathcal{E}_{3,\alpha}, \mathcal{E}_{4,\alpha}$ and $\mathcal{E}_{5,\alpha}$ defined in Equations (56) (60) (61) (63) and (74), respectively, we have

$$\begin{aligned} \mathbb{P}(\mathcal{E}) &\geq \mathbb{P} \left(\mathcal{E} \cap \mathcal{E}_1 \cap \left(\bigcup_{\alpha \in \mathcal{A}_0} \mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha} \cap \mathcal{E}_{5,\alpha} \right) \right) \\ &= \mathbb{P} \left(\bigcup_{\alpha \in \mathcal{A}_0} \mathcal{E}_1 \cap \mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha} \cap \mathcal{E}_{5,\alpha} \right), \end{aligned} \quad (78)$$

where the last equality holds because for any $\alpha \in \mathcal{A}_0$, the event $\mathcal{E}$ is implied by the event $\mathcal{E}_1 \cap \mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha} \cap \mathcal{E}_{5,\alpha}$, which is a fact established in the proof of Equations (76) and (77).

Since by the definition in Equation (60), $\{\mathcal{E}_{2,\alpha}\}_{\alpha \in \mathcal{A}_0}$ are disjoint, it follows from Equation (78) that

$$\begin{aligned} \mathbb{P}(\mathcal{E}) &\geq \sum_{\alpha \in \mathcal{A}_0} \mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha} \cap \mathcal{E}_{5,\alpha}) \\ &= \sum_{\alpha \in \mathcal{A}_0} [\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_{2,\alpha}) - \mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_{2,\alpha} \cap (\mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha} \cap \mathcal{E}_{5,\alpha})^c)]. \end{aligned} \quad (79)$$

By the union bound and basic set inclusions,

$$\begin{aligned}\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_{2,\alpha} \cap (\mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha} \cap \mathcal{E}_{5,\alpha})^c) &= \mathbb{P}[\mathcal{E}_1 \cap \mathcal{E}_{2,\alpha} \cap (\mathcal{E}_{3,\alpha}^c \cup \mathcal{E}_{4,\alpha}^c \cup \mathcal{E}_{5,\alpha}^c)] \\ &\leq \mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha}^c) + \mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_{2,\alpha} \cap \mathcal{E}_{4,\alpha}^c) \\ &\quad + \mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_{2,\alpha} \cap \mathcal{E}_{5,\alpha}^c) \\ &\leq \mathbb{P}(\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha}^c) + \mathbb{P}(\mathcal{E}_{4,\alpha}^c) + \mathbb{P}(\mathcal{E}_{5,\alpha}^c).\end{aligned}$$

Applying this bound in Equation (79),

$$\mathbb{P}(\mathcal{E}) \geq \sum_{\alpha \in \mathcal{A}_0} [\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_{2,\alpha}) - (\mathbb{P}(\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha}^c) + \mathbb{P}(\mathcal{E}_{4,\alpha}^c) + \mathbb{P}(\mathcal{E}_{5,\alpha}^c))] \quad (80)$$

The terms $\mathbb{P}(\mathcal{E}_{4,\alpha}^c)$ and $\mathbb{P}(\mathcal{E}_{5,\alpha}^c)$ are bounded in Lemma 9 as

$$\mathbb{P}(\mathcal{E}_{4,\alpha}^c) + \mathbb{P}(\mathcal{E}_{5,\alpha}^c) \leq O(n^{-8}). \quad (81)$$

For the term $\mathbb{P}(\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha}^c)$, we have

$$\mathbb{P}(\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha}^c) = \mathbb{P}\left(\bigcup_{k \in I_\alpha} \{Q_{kk} \neq \text{sgn}(u_k^*), \hat{I} = I_\alpha\} \cup \bigcup_{k \in I_\alpha^c} \{Q_{kk} \neq 1, \hat{I} = I_\alpha\}\right).$$

By Step 3 in Algorithm 1, we must have

$$\bigcup_{k \in I_\alpha^c} \{Q_{kk} \neq 1, \hat{I} = I_\alpha\} = \bigcup_{k \in \hat{I}^c} \{Q_{kk} \neq 1, \hat{I} = I_\alpha\} = \emptyset.$$

It follows that

$$\begin{aligned}\mathbb{P}(\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha}^c) &= \mathbb{P}\left(\bigcup_{k \in I_\alpha} \{\text{sgn}(u_k) \neq \text{sgn}(u_k^*), \hat{I} = I_\alpha\}\right) \\ &\leq \mathbb{P}\left(\bigcup_{k \in I_\alpha} \{\text{sgn}(u_k) \neq \text{sgn}(u_k^*)\}\right).\end{aligned} \quad (82)$$

For any $k \in I_\alpha$, by the construction of $\mathcal{A}$ in Equation (10), we must have $|u_k^*| \geq \alpha \geq 1/\sqrt{2n}$. Thus, by Lemma 8, we can further bound Equation (82) as

$$\mathbb{P}(\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha}^c) \leq O(n^{-8}). \quad (83)$$

Applying Equations (81) and (83) to Equation (80), we have

$$\begin{aligned}\mathbb{P}(\mathcal{E}) &\geq \sum_{\alpha \in \mathcal{A}_0} \mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_{2,\alpha}) - O(|\mathcal{A}_0|n^{-8}) \\ &= 1 - \mathbb{P}\left(\mathcal{E}_1^c \cup \left(\bigcap_{\alpha \in \mathcal{A}_0} \mathcal{E}_{2,\alpha}^c\right)\right) - O(|\mathcal{A}_0|n^{-8}) \\ &\geq 1 - \mathbb{P}(\mathcal{E}_1^c) - \mathbb{P}\left(\bigcap_{\alpha \in \mathcal{A}_0} \{\hat{I} \neq I_\alpha\}\right) - O(|\mathcal{A}_0|n^{-8}).\end{aligned}$$

Under Assumption 4, we have $\mathbb{P}(\mathcal{E}_1^c) \leq O(n^{-8} \log n)$, and it follows that

$$\mathbb{P}(\mathcal{E}) \geq 1 - O(n^{-8} \log n) - \mathbb{P}\left(\bigcap_{\alpha \in \mathcal{A}_0} \{\hat{I} \neq I_\alpha\}\right) - O(|\mathcal{A}_0|n^{-8}).$$

Applying Lemma 7 to the right hand side yields that

$$\mathbb{P}(\mathcal{E}) \geq 1 - O(n^{-8} \log n) - O(n^{-8}) - O(|\mathcal{A}_0|n^{-8}) \geq 1 - O(n^{-8} \log n),$$

where the last inequality follows from the fact that $|\mathcal{A}_0| \leq |\mathcal{A}|$ and $|\mathcal{A}| = O(\log n)$ by Equation (10).

At this stage, we have shown that with probability at least $1 - O(n^{-8} \log n)$,

$$\|Q\hat{v} - \mathbf{u}^*\|_\infty \leq \frac{C\sqrt{\nu_W} \log^{5/2} n}{\lambda^*}. \quad (84)$$

Part II. Estimation error related to $\mathbf{u}$.

Note that Step 5 in Algorithm 1 further refines $\hat{v}$ to adjust the estimates of small entries of $\mathbf{u}^*$. In the remainder of the proof, we show that this refinement does not affect our bound in Equation (84). By Lemma 5, it holds with probability at least $1 - O(n^{-8})$ that for all $j \in [n]$,

$$\begin{aligned} |u_j - u_j^*| &\leq \frac{80\sqrt{\nu_W \log n} + 120\sqrt{\nu_W n} |u_j^*|}{\lambda^*} \\ &\leq \frac{80\sqrt{\nu_W \log n}}{\lambda^*} + \frac{120\sqrt{\nu_W n}}{\lambda^*} (|u_j| + |u_j - u_j^*|), \end{aligned}$$

where the second line follows from the triangle inequality. Rearranging terms and noting that $\lambda^* > 240\sqrt{\nu_W n \log n}$ under Assumption 3, we have that with probability at least $1 - O(n^{-8})$, it holds for all $j \in [n]$ that

$$\begin{aligned} |u_j - u_j^*| &\leq \frac{1}{1 - 120\sqrt{\nu_W n}/\lambda^*} \left(\frac{80\sqrt{\nu_W \log n}}{\lambda^*} + \frac{120\sqrt{\nu_W n}}{\lambda^*} |u_j| \right) \\ &\leq \frac{160\sqrt{\nu_W \log n}}{\lambda^*} + \frac{|u_j|}{\sqrt{\log n}}. \end{aligned}$$

That is, defining the event

$$\mathcal{E}_6 = \bigcap_{j=1}^n \left\{ |u_j - u_j^*| \leq \frac{160\sqrt{\nu_W \log n}}{\lambda^*} + \frac{|u_j|}{\sqrt{\log n}} \right\}, \quad (85)$$

we have

$$\mathbb{P}(\mathcal{E}_6^c) \leq O(n^{-8}). \quad (86)$$

Note that by Equation (58) and Assumption 3, when $\mathcal{E}_1$ holds, we have $|\lambda^*/\hat{\lambda}| \leq 2$ for sufficiently large n . It follows that when $\mathcal{E}_1$ holds, for all $j \in [n]$ such that $|u_j| \leq (\sqrt{\nu_W}/\hat{\lambda}) \log n$, we have

$$|u_j| = \frac{\sqrt{\nu_W} \log n}{\lambda^*} \cdot \frac{\lambda^*}{\hat{\lambda}} \leq \frac{2\sqrt{\nu_W} \log n}{\lambda^*}, \quad (87)$$

where the inequality follows from Equation (58). Define the set $\hat{J} = \{j : |u_j| \leq (\sqrt{\nu_W}/\hat{\lambda}) \log n\}$. We have

$$\begin{aligned} &\mathbb{P} \left(\bigcup_{j \in \hat{J}} \left\{ |u_j - u_j^*| > \frac{162\sqrt{\nu_W \log n}}{\lambda^*} \right\} \right) \\ &\leq \mathbb{P} \left(\bigcup_{j \in \hat{J}} \left\{ |u_j - u_j^*| > \frac{162\sqrt{\nu_W \log n}}{\lambda^*} \right\} \cap \mathcal{E}_1 \right) + \mathbb{P}(\mathcal{E}_1^c) \\ &\leq \mathbb{P} \left(\bigcup_{j \in \hat{J}} \left\{ |u_j - u_j^*| > \frac{160\sqrt{\nu_W \log n}}{\lambda^*} + \frac{|u_j|}{\sqrt{\log n}} \right\} \cap \mathcal{E}_1 \right) + O(n^{-8}), \end{aligned}$$

where the second inequality follows from Equation (87) and Assumption 4. Recalling the definition of $\mathcal{E}_6$ in Equation (85) and using the fact that $\hat{J} \subseteq [n]$, we have that the set in the last inequality of the above display is a subset of $\mathcal{E}_6^c$. Following the above bound and basic set inclusions, we have

$$\mathbb{P} \left(\bigcup_{j \in \hat{J}} \left\{ |u_j - u_j^*| > \frac{162\sqrt{\nu_W \log n}}{\lambda^*} \right\} \right) \leq \mathbb{P}(\mathcal{E}_6^c) + O(n^{-8}) \leq O(n^{-8}).$$

Combining this with Equation (84) and recalling the definition of $\hat{\mathbf{u}}$ in Step 5, we have that with probability at least $1 - O(n^{-8})$,

$$d_\infty(\hat{\mathbf{u}}, \mathbf{u}^*) \leq \frac{C\sqrt{\nu_W} \log^{5/2} n}{\lambda^*},$$

as we set out to show. $\square$

E A new estimator for λ^*

Consider $\tilde{Y}$ and $\hat{S}$ as constructed in Steps 3 and 4, respectively, of Algorithm 1. We can estimate λ^* by

$$\hat{\lambda} = \frac{\sum_{j=1}^n \left(\sum_{k \in \hat{I}} \tilde{Y}_{jk} \right)^2 - |\hat{I}| n \sigma^2}{\hat{S}^2}. \quad (88)$$

Proposition 1 controls the estimation error of $\hat{\lambda}$ given in Equation (88).

Proposition 1. *Under the model in Equation (1), suppose that Assumptions 1, 2 and 3 hold and consider the eigenvalue estimate $\hat{\lambda}$ as defined in Equation (88). For n sufficiently large, we have*

$$|\hat{\lambda} - \lambda^*| \lesssim \sqrt{\nu_W} \log^{5/2} n$$

with probability at least $1 - O(n^{-8} \log n)$.

E.1 Technical lemmas for Proposition 1

Our proof of Proposition 1, which appears in Section E.2 below, makes use of two technical lemmas, which we establish first.

Lemma 10. *Let $W \in \mathbb{R}^{n \times n}$ be a random matrix satisfying Assumption 1. For a fixed $s \in \{\pm 1\}^n$, for any $\alpha \in \mathcal{A}_0$, with probability at least $1 - O(n^{-8})$,*

$$\left| \sum_{j=1}^n \left(\sum_{k \in I_\alpha} s_k W_{jk} \right)^2 - n |I_\alpha| \sigma^2 \right| \leq C \nu_W |I_\alpha| \sqrt{n} \log n. \quad (89)$$

Proof. Since $s \in \{\pm 1\}^n$ is fixed and the entries of W are symmetric about zero, without loss of generality, we assume that s is a vector of all 1's. Reindexing if necessary, we assume without loss of generality that $I_\alpha = [K]$, where $K = |I_\alpha| > 0$. We have $K > 0$ since $\alpha \in \mathcal{A}_0$ and by the definition of $\mathcal{A}_0$ in Equation (53), $|I_\alpha| \geq (\alpha^2 \log^2 n)^{-1} > 0$. It follows that

$$\begin{aligned} \sum_{j=1}^n \left(\sum_{k \in I_\alpha} W_{jk} \right)^2 - n |I_\alpha| \sigma^2 &= \sum_{j=1}^n \sum_{k=1}^K W_{jk}^2 - n K \sigma^2 + 2 \sum_{j=1}^n \sum_{1 \leq k < \ell \leq K} W_{jk} W_{j\ell} \\ &= \gamma_1 + \gamma_2 + \gamma_3, \end{aligned} \quad (90)$$

where

$$\begin{aligned} \gamma_1 &:= \sum_{j=K+1}^n \sum_{k=1}^K W_{jk}^2 + \sum_{k=1}^K W_{kk}^2 + 2 \sum_{1 \leq j < k \leq K} W_{jk}^2 - n K \sigma^2, \\ \gamma_2 &:= 2 \sum_{j=K+1}^n \sum_{1 \leq k < \ell \leq K} W_{jk} W_{j\ell} \text{ and} \\ \gamma_3 &:= \sum_{j=1}^K \sum_{1 \leq k \neq \ell \leq K} W_{jk} W_{j\ell}. \end{aligned}$$

We will bound these three quantities separately.

We begin by bounding γ_1 in Equation (90). For ease of notation, define $N = K(n - K/2 + 1/2)$. γ_1 is a sum of N independent sub-exponential random variables. By Equation (2.18) in [56], for any $t \in (0, 2)$,

$$\mathbb{P} \left(\frac{|\gamma_1|}{N} \geq \nu_W t \right) \leq 2e^{-Nt^2/32}.$$

Taking $t = 16\sqrt{N^{-1} \log n}$ yields that for all suitably large n , with probability at least $1 - O(n^{-8})$,

$$|\gamma_1| \leq 16\nu_W \sqrt{N \log n} \leq 16\nu_W \sqrt{Kn \log n}, \quad (91)$$

where the second inequality follows from the trivial bound $N \leq Kn$.

Turning to γ_2 in Equation (90), define

$$\mathbf{A} = \mathbf{1}_K \mathbf{1}_K^\top - \mathbf{I}_K \in \mathbb{R}^{K \times K},$$

where $\mathbf{1}_K \in \mathbb{R}^K$ is a vector of all 1's and $\mathbf{I}_K$ is the $K \times K$ identity matrix. It follows that

$$\gamma_2 = \sum_{j=K+1}^n \sum_{1 \leq k \neq \ell \leq K} W_{jk} W_{j\ell} = \sum_{j=K+1}^n \sum_{1 \leq k, \ell \leq K} W_{jk} A_{k\ell} W_{j\ell}.$$

By the Hanson-Wright inequality (see Theorem 6.2.1 in [54]), for $t \geq 0$, we have

$$\begin{aligned} & \mathbb{P} \left(\left| \sum_{j=K+1}^n \sum_{1 \leq k, \ell \leq K} W_{jk} A_{k\ell} W_{j\ell} \right| \geq t \right) \\ & \leq 2 \exp \left\{ -c \min \left(\frac{t^2}{(n-K) \nu_W^2 \|\mathbf{A}\|_F^2}, \frac{t}{\sqrt{n-K} \nu_W \|\mathbf{A}\|_F} \right) \right\}, \end{aligned}$$

where $c > 0$ is a universal constant. We note that $\|\mathbf{A}\|_F \leq K^2$ by construction. Taking $t = (8\nu_W/c)K\sqrt{n-K} \log n$, with probability at least $1 - O(n^{-8})$ we have

$$|\gamma_2| = \left| \sum_{j=K+1}^n \sum_{1 \leq k, \ell \leq K} W_{jk} A_{k\ell} W_{j\ell} \right| \leq \frac{8\nu_W}{c} K \sqrt{n-K} \log n. \quad (92)$$

Finally, to bound γ_3 in Equation (90), we define

$$\mathbf{w} = \text{vech}([W_{ij}]_{1 \leq i, j \leq K}) \in \mathbb{R}^{K(K+1)/2},$$

where $\text{vech}(\cdot)$ is the half-vectorization operator that vectorizes the upper triangular part (including the diagonal) of a given matrix.

We identify the elements of $\mathbf{w}$ with the elements of

$$\mathcal{J}_K = \{(i_1, i_2) : 1 \leq i_1 \leq i_2 \leq K\},$$

that is, pairs (i_1, i_2) satisfying $1 \leq i_1 \leq i_2 \leq K$. We note that γ_3 is a sum of products of the form $w_{(i_1, i_2)} w_{(j_1, j_2)}$ such that $(i_1, i_2), (j_1, j_2) \in \mathcal{J}_K$ and one element of (i_1, i_2) agrees with one element of (j_1, j_2) , while the others disagree. For $(i_1, i_2) \in \mathcal{J}_K$ with $i_1 < i_2$, there are $2(K-1)$ other pairs $(j_1, j_2) \in \mathcal{J}_K$ satisfying this requirement, while for $(i_1, i_2) \in \mathcal{J}_K$ with $i_1 = i_2$, there are $(K-1)$ other elements of $\mathcal{J}_K$ satisfying the requirement. In total, there are $K^2(K-1)$ such pairs, which agrees with the number of terms in γ_3 . We define matrix $\mathbf{B} \in \mathbb{R}^{K(K+1)/2 \times K(K+1)/2}$ by identifying its rows and columns with elements of $\mathcal{J}_K$ and setting

$$B_{(i_1, i_2), (j_1, j_2)} = \max \{ \delta_{i_1 j_1} (1 - \delta_{i_2 j_2}), \delta_{i_1 j_2} (1 - \delta_{i_2 j_1}), \delta_{i_2 j_1} (1 - \delta_{i_1 j_2}), \delta_{i_2 j_2} (1 - \delta_{i_1 j_1}) \}$$

for $(i_1, i_2), (j_1, j_2) \in \mathcal{J}_K$, where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

There are $K^2(K-1)$ entries in $\mathbf{B}$ that are equal to 1, while all others are equal to zero. Thus, we have $\|\mathbf{B}\|_F = K\sqrt{K-1} \leq K^{3/2}$. By construction, one can verify that

$$\mathbf{w}^\top \mathbf{B} \mathbf{w} = \sum_{(i_1, i_2) \in \mathcal{J}} \sum_{(j_1, j_2) \in \mathcal{J}} w_{(i_1, i_2)} B_{(i_1, i_2), (j_1, j_2)} w_{(j_1, j_2)} = \sum_{j=1}^K \sum_{1 \leq k \neq \ell \leq K} W_{jk} W_{j\ell} = \gamma_3.$$

Again, by the Hanson-Wright inequality, for every $t \geq 0$, we have

$$\mathbb{P} (|\mathbf{w}^\top \mathbf{B} \mathbf{w}| \geq t) \leq 2 \exp \left\{ -c \min \left(\frac{t^2}{\nu_W^2 \|\mathbf{B}\|_F^2}, \frac{t}{\nu_W \|\mathbf{B}\|_F} \right) \right\},$$

where $c > 0$ is a universal constant. Taking $t = (8\nu_W/c)\|\mathbf{B}\|_F \log n$ yields that with probability at least $1 - O(n^{-8})$,

$$|\gamma_3| = |\mathbf{w}^\top \mathbf{B} \mathbf{w}| \leq \frac{8\nu_W}{c} \|\mathbf{B}\|_F \log n \leq \frac{8\nu_W}{c} K^{3/2} \log n, \quad (93)$$

where the second inequality follows from our bound on $\|\mathbf{B}\|_F$.

Finally, plugging the bounds in Equations (91)–(92) and (93) into Equation (90), we have that with probability at least $1 - O(n^{-8})$, Equation (89) holds, as we set out to show. $\square$

Lemma 11. *Under Assumption 1, for any fixed $\mathbf{s} \in \{\pm 1\}^n$ and any fixed $\mathbf{u}^* \in \mathbb{S}^{n-1}$, for any $\alpha \in \mathcal{A}$, it holds with probability at least $1 - O(n^{-8})$ that*

$$\left| \sum_{k \in I_\alpha} \sum_{j=1}^n u_j^* s_k W_{jk} \right| \leq 4\sqrt{2\nu_W |I_\alpha| \log n}.$$

Proof. Since $\mathbf{s} \in \{\pm 1\}^n$ is fixed and the entries of $\mathbf{W}$ are symmetric about zero, without loss of generality, we assume that $\mathbf{s}$ is a vector of all 1's. Reindexing if necessary, we again assume that $I_\alpha = [K]$ without loss of generality, where $K = |I_\alpha|$. Rearranging sums, we have

$$\begin{aligned} \sum_{k=1}^K \sum_{j=1}^n u_j^* s_k W_{jk} &= \sum_{k=1}^K \sum_{j=1}^n u_j^* W_{jk} \\ &= \sum_{j=1}^K \sum_{k=j+1}^K (u_j^* + u_k^*) W_{jk} + \sum_{j=1}^K u_j^* W_{jj} + \sum_{k=1}^K \sum_{j=K+1}^n u_j^* W_{jk}. \end{aligned}$$

By standard subgaussian concentration inequalities [54, 56], using the fact that $\mathbf{u}^*$ is unit-norm,

$$\mathbb{P} \left(\left| \sum_{k=1}^K \sum_{j=1}^n u_j^* s_k W_{jk} \right| \geq 4\sqrt{\nu_W 2K \log n} \right) \leq O(n^{-8}),$$

completing the proof. $\square$

E.2 Proof of Proposition 1

Proof. Similar to the proof of Theorem 1, we assume that $\lambda^* > 0$ without loss of generality. Recalling the quantity $\hat{S}$ as defined in Step 4 of Algorithm 1 and the fact that $\hat{S} > 0$ from Equation (65), we define

$$R_{\hat{I}} = \frac{\sqrt{\lambda^*} \left(\sum_{k \in \hat{I}} Q_{kk} u_k^* \right)}{\hat{S}}, \quad (94)$$

where we recall that $\hat{I}$ is the set given in Step 2 of Algorithm 1 and $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is a diagonal matrix given in Step 3 of Algorithm 1.

Plugging in $\tilde{\mathbf{Y}} = \mathbf{Q} \mathbf{Y} \mathbf{Q}$ and expanding the right hand side of Equation (88), we have

$$\begin{aligned} \hat{\lambda} &= \frac{1}{\hat{S}^2} \left(\sum_{j=1}^n \left(\lambda^* Q_{jj} u_j^* \sum_{k \in \hat{I}} Q_{kk} u_k^* + \sum_{k \in \hat{I}} Q_{jj} Q_{kk} W_{jk} \right)^2 \right) - \frac{n|\hat{I}| \sigma^2}{\hat{S}^2} \\ &= \frac{1}{\hat{S}^2} \sum_{j=1}^n (\lambda^* u_j^*)^2 \left(\sum_{k \in \hat{I}} Q_{kk} u_k^* \right)^2 + \frac{1}{\hat{S}^2} \sum_{j=1}^n \left(\sum_{k \in \hat{I}} Q_{kk} W_{jk} \right)^2 \\ &\quad + \frac{2\lambda^*}{\hat{S}^2} \sum_{j=1}^n \left(u_j^* \sum_{k \in \hat{I}} Q_{kk} u_k^* \sum_{\ell \in \hat{I}} Q_{\ell\ell} W_{j\ell} \right) - \frac{n|\hat{I}| \sigma^2}{\hat{S}^2}. \end{aligned}$$

Applying the definition of $R_{\hat{I}}$ in Equation (94) and rearranging terms,

$$\hat{\lambda} = \lambda^* R_{\hat{I}}^2 + R_{\hat{I}}^2 \frac{\sum_{j=1}^n (\sum_{k \in \hat{I}} Q_{kk} W_{jk})^2 - n|\hat{I}|\sigma^2}{(\sqrt{\lambda^*} \sum_{k \in \hat{I}} Q_{kk} u_k^*)^2} + 2R_{\hat{I}}^2 \sum_{j=1}^n u_j^* \frac{\sum_{k \in \hat{I}} Q_{kk} W_{jk}}{\sum_{k \in \hat{I}} Q_{kk} u_k^*}. \quad (95)$$

To control the term $R_{\hat{I}}$ in Equation (95), we proceed to bound a related term R_α for all $\alpha \in \mathcal{A}_0$, defined as

$$R_\alpha = \frac{\sqrt{\lambda^*} (\sum_{k \in I_\alpha} |u_k^*|)}{\hat{S}}.$$

We recall the events $\mathcal{E}_{2,\alpha}$, $\mathcal{E}_{3,\alpha}$ and $\mathcal{E}_{4,\alpha}$ for all $\alpha \in \mathcal{A}_0$ as defined in Equations (60), (61) and (63) in the proof of Theorem 1. For any $\alpha \in \mathcal{A}_0$, to control the term R_α , we divide $\hat{S}$ on both sides of Equation (67) stated in the proof of Theorem 1. It follows from Equation (67) that

$$|R_\alpha - 1| = \left| \frac{\sqrt{\lambda^*} \sum_{k \in I_\alpha} |u_k^*|}{\hat{S}} - 1 \right| \leq \frac{C\sqrt{\nu_W} \log n}{\alpha \sqrt{\lambda^*} \hat{S}}$$

holds on the event $\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha}$. Following the previous bound, on the event $\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha}$, we have that

$$\begin{aligned} |R_\alpha - 1| &\leq \frac{C\sqrt{\nu_W} \log n}{\alpha \sqrt{\lambda^*} \hat{S}} \leq \frac{C\sqrt{\nu_W} \log n}{\alpha \lambda^* (\sum_{k \in I_\alpha} |u_k^*|) - \alpha \sqrt{\lambda^*} |\hat{S} - \sqrt{\lambda^*} \sum_{k \in I_\alpha} |u_k^*|}} \\ &\leq \frac{C\sqrt{\nu_W} \log n}{\lambda^* |I_\alpha| \alpha^2 - C\sqrt{\nu_W} \log n} \leq \frac{C\sqrt{\nu_W} \log^{5/2} n}{\lambda^* - C\sqrt{\nu_W} \log^{5/2} n}, \end{aligned}$$

where the second inequality holds by the triangle inequality, the third inequality follows from Equations (66) and (52), the last inequality holds by Equation (53). We remind the reader that we allow constant C to change its precise value from line to line in the proof. Under Assumption 3, it follows from the previous bound that for n sufficiently large, we have on the event $\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha}$ that

$$|R_\alpha - 1| \leq \frac{C\sqrt{\nu_W} \log^{5/2} n}{\lambda^*}. \quad (96)$$

By Equation (96) and Assumption 3, for n sufficiently large, we have

$$1/2 \leq R_\alpha \leq 2 \quad (97)$$

on the event $\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha}$.

Recall $\mathbf{s} \in \{\pm 1\}^n$ as defined in Equation (62). On the event $\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha}$, we find that, using Equation (95) to substitute for $\hat{\lambda}$ and applying the triangle inequality,

$$\begin{aligned} |\hat{\lambda} - \lambda^*| &\leq \lambda^* |R_\alpha^2 - 1| + R_\alpha^2 \frac{\left| \sum_{j=1}^n (\sum_{k \in I_\alpha} s_k W_{jk})^2 - n|I_\alpha|\sigma^2 \right|}{(\sqrt{\lambda^*} \sum_{k \in I_\alpha} |u_k^*|)^2} \\ &\quad + 2R_\alpha^2 \left| \sum_{j=1}^n u_j^* \frac{\sum_{k \in I_\alpha} s_k W_{jk}}{\sum_{k \in I_\alpha} |u_k^*|} \right|, \end{aligned}$$

where we substitute $R_{\hat{I}}$ by R_α using the fact that on $\mathcal{E}_{2,\alpha}$, we have $\hat{I} = I_\alpha$. Continuing from the previous bound, it follows from Equation (96) that

$$\begin{aligned} |\hat{\lambda} - \lambda^*| &\leq \lambda^* \cdot \frac{C\sqrt{\nu_W} \log^{5/2} n}{\lambda^*} \cdot |R_\alpha + 1| + R_\alpha^2 \frac{\left| \sum_{j=1}^n (\sum_{k \in I_\alpha} s_k W_{jk})^2 - n|I_\alpha|\sigma^2 \right|}{(\sqrt{\lambda^*} \sum_{k \in I_\alpha} |u_k^*|)^2} \\ &\quad + 2R_\alpha^2 \left| \sum_{j=1}^n u_j^* \frac{\sum_{k \in I_\alpha} s_k W_{jk}}{\sum_{k \in I_\alpha} |u_k^*|} \right|. \end{aligned}$$

Applying Equation (97) to R_α in the above display, we have

$$\begin{aligned} |\hat{\lambda} - \lambda^*| &\leq C\sqrt{\nu_W} \log^{5/2} n + \frac{C \left| \sum_{j=1}^n (\sum_{k \in I_\alpha} s_k W_{jk})^2 - n|I_\alpha| \sigma^2 \right|}{\left(\sqrt{\lambda^*} \sum_{k \in I_\alpha} |u_k^*| \right)^2} \\ &\quad + C \left| \frac{\sum_{k \in I_\alpha} \sum_{j=1}^n u_j^* s_k W_{jk}}{\sum_{k \in I_\alpha} |u_k^*|} \right|. \end{aligned} \quad (98)$$

For all $\alpha \in \mathcal{A}_0$, define the events

$$\begin{aligned} \mathcal{G}_{1,\alpha} &= \left\{ \left| \sum_{j=1}^n \left(\sum_{k \in I_\alpha} s_k W_{jk} \right)^2 - n|I_\alpha| \sigma^2 \right| \leq C\nu_W |I_\alpha| \sqrt{n} \log n \right\} \quad \text{and} \\ \mathcal{G}_{2,\alpha} &= \left\{ \left| \sum_{k \in I_\alpha} \sum_{j=1}^n u_j^* s_k W_{jk} \right| \leq 4\sqrt{2\nu_W} |I_\alpha| \log n \right\}. \end{aligned} \quad (99)$$

On the event $\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha} \cap \mathcal{G}_{1,\alpha} \cap \mathcal{G}_{2,\alpha}$, we can further bound Equation (98) according to

$$|\hat{\lambda} - \lambda^*| \leq C\sqrt{\nu_W} \log^{5/2} n + \frac{C\nu_W |I_\alpha| \sqrt{n} \log n}{\left(\sqrt{\lambda^*} \sum_{k \in I_\alpha} |u_k^*| \right)^2} + \frac{C\sqrt{\nu_W} |I_\alpha| \log n}{\sum_{k \in I_\alpha} |u_k^*|},$$

following the definition of $\mathcal{G}_{1,\alpha}$ and $\mathcal{G}_{2,\alpha}$ in Equation (99). By the definition of I_α in Equation (52), it follows that

$$|\hat{\lambda} - \lambda^*| \leq C\sqrt{\nu_W} \log^{5/2} n + \frac{C\nu_W |I_\alpha| \sqrt{n} \log n}{\lambda^* |I_\alpha|^2 \alpha^2} + \frac{C\sqrt{\nu_W} |I_\alpha| \log n}{|I_\alpha| \alpha}$$

holds on the event $\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha} \cap \mathcal{G}_{1,\alpha} \cap \mathcal{G}_{2,\alpha}$. Applying Equation (53) to lower bound $|I_\alpha|$, we have

$$|\hat{\lambda} - \lambda^*| \leq C\sqrt{\nu_W} \log^{5/2} n + \frac{C\nu_W \sqrt{n} \log^3 n}{\lambda^*} + C\sqrt{\nu_W} \log^{3/2} n \leq C\sqrt{\nu_W} \log^{5/2} n, \quad (100)$$

holds on the event $\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha} \cap \mathcal{G}_{1,\alpha} \cap \mathcal{G}_{2,\alpha}$, where the last inequality holds under Assumption 3 when n is sufficiently large.

Consider the event

$$\mathcal{G} = \left\{ |\hat{\lambda} - \lambda^*| \leq C\sqrt{\nu_W} \log^{5/2} n \right\}.$$

By the proof leading to Equation (100), we have for all $\alpha \in \mathcal{A}_0$,

$$\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha} \cap \mathcal{G}_{1,\alpha} \cap \mathcal{G}_{2,\alpha} \subseteq \mathcal{G}.$$

It follows that

$$\begin{aligned} \mathbb{P}(\mathcal{G}) &\geq \mathbb{P} \left(\bigcup_{\alpha \in \mathcal{A}_0} \mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha} \cap \mathcal{G}_{1,\alpha} \cap \mathcal{G}_{2,\alpha} \right) \\ &= \sum_{\alpha \in \mathcal{A}_0} \mathbb{P}(\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha} \cap \mathcal{E}_{4,\alpha} \cap \mathcal{G}_{1,\alpha} \cap \mathcal{G}_{2,\alpha}), \end{aligned}$$

where the first inequality follows from set inclusion and the last equality follows from the fact that $\{\mathcal{E}_{2,\alpha}\}_{\alpha \in \mathcal{A}_0}$ are disjoint events according to Equation (60). By basic set inclusions, it follows that

$$\mathbb{P}(\mathcal{G}) \geq \sum_{\alpha \in \mathcal{A}_0} \left(\mathbb{P}(\mathcal{E}_{2,\alpha}) - \mathbb{P}(\mathcal{E}_{2,\alpha} \cap \mathcal{E}_{3,\alpha}^c) - \mathbb{P}(\mathcal{E}_{4,\alpha}^c) - \sum_{j=1}^2 \mathbb{P}(\mathcal{G}_{j,\alpha}^c) \right).$$

Applying Lemma 9, Equations (83), Lemma 10 and Lemma 11 to the above display yields that

$$\mathbb{P}(\mathcal{G}) \geq \sum_{\alpha \in \mathcal{A}_0} \mathbb{P}(\mathcal{E}_{2,\alpha}) - O(|\mathcal{A}_0| n^{-8}) = 1 - \mathbb{P} \left(\bigcap_{\alpha \in \mathcal{A}_0} \{ \hat{I} \neq I_\alpha \} \right) - O(|\mathcal{A}_0| n^{-8}).$$

Applying Lemma 7 and using the fact that $|\mathcal{A}_0| \leq |\mathcal{A}|$ and $|\mathcal{A}| = O(\log n)$ by its definition in Equation (10), we have

$$\mathbb{P}(\mathcal{G}) \geq 1 - O(n^{-8}) - O(|\mathcal{A}_0| n^{-8}) \geq 1 - O(n^{-8} \log n),$$

completing the proof. $\square$

F Proofs for lower bounds of metric entropy under $d_{2,\infty}$

To prove Lemma 4, we first state a few technical lemmas.

F.1 Technical Lemmas

For a semi-metric ρ defined over a set $\mathbb{K}$, we let $\mathcal{N}(\mathbb{K}, \rho, \delta)$ denote the δ -covering number of $\mathbb{K}$ under ρ (see Chapter 15 in [56]). The collection of all linear subspaces of fixed dimension r of the Euclidean space $\mathbb{R}^n$ forms the Grassmann manifold $\mathbb{G}_{n,r}$, also termed the Grassmannian. For points on the Grassmannian, we adopt the projector perspective [13]: a subspace $\mathcal{U} \in \mathbb{G}_{n,r}$ is identified with the (unique) orthogonal projector $\mathbf{P} \in \mathbb{R}^{n \times n}$ onto $\mathcal{U}$, which in turn is uniquely represented by $\mathbf{P} = \mathbf{U}\mathbf{U}^\top$, where $\mathbf{U} \in \mathbb{R}^{n \times r}$ whose columns form an orthonormal basis of $\mathcal{U}$. For any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ with singular values denoted by $\sigma_i(\mathbf{A})$ for $1 \leq i \leq \min\{m, n\}$, the Schatten- q norm $\|\cdot\|_{S_q}$ is defined for any $1 \leq q \leq \infty$

$$\|\mathbf{A}\|_{S_q} := \left(\sum_{i=1}^{\min\{m,n\}} \sigma_i^q(\mathbf{A}) \right)^{1/q}.$$

For a pair of subspaces $\mathcal{U}_1, \mathcal{U}_2 \in \mathbb{G}_{n,r}$ identified with projectors $\mathbf{P}_1 = \mathbf{U}_1\mathbf{U}_1^\top, \mathbf{P}_2 = \mathbf{U}_2\mathbf{U}_2^\top \in \mathbb{R}^{n \times n}$, respectively, we consider the distance $d_{S_q}(\cdot, \cdot)$ induced by the Schatten- q norm

$$d_{S_q}(\mathcal{U}_1, \mathcal{U}_2) := \|\mathbf{P}_1 - \mathbf{P}_2\|_{S_q} = \|\mathbf{U}_1\mathbf{U}_1^\top - \mathbf{U}_2\mathbf{U}_2^\top\|_{S_q}. \quad (101)$$

Lemma 12 controls the covering number of $\mathbb{G}_{n,r}$ under d_{S_q} .

Lemma 12 ([46] Proposition 8). *For any integer $1 \leq r \leq n$ such that $2r \leq n$, any q such that $1 \leq q \leq \infty$ and any $\delta > 0$, we have*

$$\left(\frac{c_0}{\delta}\right)^{r(n-r)} \leq \mathcal{N}(\mathbb{G}_{n,r}, d_{S_q}, \delta r^{1/q}) \leq \left(\frac{C_0}{\delta}\right)^{r(n-r)}$$

where $c_0, C_0 > 0$ are universal constants, d_{S_q} is the distance defined in Equation (101) induced by the Schatten- q norm.

Let $\mathbb{V}_{n,r} := \{\mathbf{U} \in \mathbb{R}^{n \times r} : \mathbf{U}^\top \mathbf{U} = \mathbf{I}_r\}$ be the $n \times r$ Stiefel manifold [16]. The distance $d_{S_q}(\cdot, \cdot)$ can also be viewed as a distance defined on $\mathbb{V}_{n,r}$. For a pair of orthogonal matrices $\mathbf{U}_1, \mathbf{U}_2 \in \mathbb{V}_{n,r}$, we let

$$d_{S_q}(\mathbf{U}_1, \mathbf{U}_2) := \|\mathbf{U}_1\mathbf{U}_1^\top - \mathbf{U}_2\mathbf{U}_2^\top\|_{S_q}.$$

When $q = 2$, the Schatten- q norm coincides with the Frobenius norm, and we have

$$d_{S_2}(\mathbf{U}_1, \mathbf{U}_2) = \left\| \mathbf{U}_1\mathbf{U}_1^\top - \mathbf{U}_2\mathbf{U}_2^\top \right\|_{\text{F}}.$$

Define a distance d_{F} over $\mathbb{V}_{n,r}$ as

$$d_{\text{F}}(\mathbf{U}_1, \mathbf{U}_2) := \min_{\mathbf{\Gamma} \in \mathbb{O}_r} \|\mathbf{U}_1 - \mathbf{U}_2\mathbf{\Gamma}\|_{\text{F}}. \quad (102)$$

Lemma 13 controls the covering number of $\mathbb{V}_{n,r}$ under d_{F} , which follows immediately from Lemma 12.

Lemma 13. *For any integer $1 \leq r \leq n/2$ and for every $\delta > 0$, we have*

$$\left(\frac{c_0\sqrt{r}}{\delta}\right)^{r(n-r)} \leq \mathcal{N}(\mathbb{V}_{n,r}, d_{\text{F}}, \delta) \leq \left(\frac{C_0\sqrt{2r}}{\delta}\right)^{r(n-r)},$$

where $c_0, C_0 > 0$ are universal constants and d_{F} is the

Proof of Lemma 13. We identify each element $\mathbf{U}$ of $\mathbb{V}_{n,r}$ with the element $\mathbf{U}\mathbf{U}^\top$ in $\mathbb{G}_{n,r}$ and apply Lemma 12 to $\mathbb{V}_{n,r}$ under d_{F} . By the fact that (see Lemma 2.6 in [27])

$$\frac{1}{\sqrt{2}} \left\| \mathbf{U}_1\mathbf{U}_1^\top - \mathbf{U}_2\mathbf{U}_2^\top \right\|_{\text{F}} \leq d_{\text{F}}(\mathbf{U}_1, \mathbf{U}_2) \leq \left\| \mathbf{U}_1\mathbf{U}_1^\top - \mathbf{U}_2\mathbf{U}_2^\top \right\|_{\text{F}},$$

we have

$$\mathcal{N}(\mathbb{V}_{n,r}, d_{S_2}, \delta) \leq \mathcal{N}(\mathbb{V}_{n,r}, d_F, \delta) \leq \mathcal{N}\left(\mathbb{V}_{n,r}, d_{S_2}, \frac{\delta}{\sqrt{2}}\right)$$

and it follows from Lemma 12 that

$$\left(\frac{c_0\sqrt{r}}{\delta}\right)^{r(n-r)} \leq \mathcal{N}(\mathbb{V}_{s,r}, d_F, \delta) \leq \left(\frac{C_0\sqrt{2r}}{\delta}\right)^{r(n-r)},$$

completing the proof. $\square$

For any $s \geq r$, Lemma 14 relates the packing number of any subset $\mathbb{K} \subseteq \mathbb{V}_{s,r}$ to its Haar measure and the covering number of $\mathbb{V}_{s,r}$. Recall that the Haar measure on $\mathbb{V}_{s,r}$ is the invariant measure under both left- and right-orthogonal transformation [29]. In other words, for any subset $\mathbb{K} \subseteq \mathbb{V}_{s,r}$ and orthogonal matrices $\mathbf{\Gamma}_1 \in \mathbb{O}_s$ and $\mathbf{\Gamma}_2 \in \mathbb{O}_r$, we have

$$\xi_H(\mathbf{\Gamma}_1 \cdot \mathbb{K}) = \xi_H(\mathbb{K}) \quad \text{and} \quad \xi_H(\mathbb{K} \cdot \mathbf{\Gamma}_2) = \xi_H(\mathbb{K}),$$

where

$$\mathbf{\Gamma}_1 \cdot \mathbb{K} := \{\mathbf{\Gamma}_1 \mathbf{U} : \mathbf{U} \in \mathbb{K}\} \quad \text{and} \quad \mathbb{K} \cdot \mathbf{\Gamma}_2 := \{\mathbf{U} \mathbf{\Gamma}_2 : \mathbf{U} \in \mathbb{K}\}.$$

For any norm $\|\cdot\|$ defined over a set $\mathbb{K}$, we also use the notation $\mathcal{M}(\mathbb{K}, \|\cdot\|, \delta)$ and $\mathcal{N}(\mathbb{K}, \|\cdot\|, \delta)$ to denote the δ -packing number and δ -covering number of $\mathbb{K}$ under the distance induced by $\|\cdot\|$, respectively. With the above setup, we state Lemma 14 below.

Lemma 14. For $1 \leq r \leq s$, let ξ_H denote the Haar measure on $\mathbb{V}_{s,r}$. For $1 \leq r \leq s$ and any $\mathbb{K} \subseteq \mathbb{V}_{s,r}$ such that $\xi_H(\mathbb{K}) \geq \gamma > 0$, for any unitarily invariant norm $\|\cdot\|$ on $\mathbb{R}^{s \times r}$, we have

$$\mathcal{M}\left(\mathbb{K}, \|\cdot\|, \frac{\delta}{2}\right) \geq \gamma \mathcal{N}(\mathbb{V}_{s,r}, \|\cdot\|, \delta), \quad (103)$$

and

$$\mathcal{M}\left(\mathbb{K}, \|\cdot\|, \frac{\delta}{2}\right) \leq \mathcal{N}\left(\mathbb{V}_{s,r}, \|\cdot\|, \frac{\delta}{8}\right). \quad (104)$$

Proof of Lemma 14. The bound in Equation (104) follows from the fact that $\mathbb{K} \subseteq \mathbb{V}_{s,r}$, and (see Exercise 4.2.10 in [54]),

$$\mathcal{M}\left(\mathbb{K}, \|\cdot\|, \frac{\delta}{2}\right) \leq \mathcal{N}\left(\mathbb{K}, \|\cdot\|, \frac{\delta}{4}\right) \leq \mathcal{N}\left(\mathbb{V}_{s,r}, \|\cdot\|, \frac{\delta}{8}\right).$$

To establish the bound in Equation (103), suppose that $\mathbb{V}_{s,r}$ has a δ -packing set $\mathcal{P} = \{\mathbf{U}^{(i)}\}_{i=1}^M$ under $\|\cdot\|$, then

$$\bigcup_{i=1}^M \mathbb{B}\left(\mathbf{U}^{(i)}, \|\cdot\|, \frac{\delta}{2}\right) \subseteq \mathbb{V}_{s,r}$$

where

$$\mathbb{B}\left(\mathbf{U}^{(i)}, \|\cdot\|, \frac{\delta}{2}\right) := \left\{\mathbf{U} \in \mathbb{V}_{s,r} : \|\mathbf{U} - \mathbf{U}^{(i)}\| \leq \delta/2\right\}.$$

Since $\mathcal{P}$ is a δ -packing set, the sets $\mathbb{B}\left(\mathbf{U}^{(i)}, \|\cdot\|, \delta/2\right)$ for $i = 1, 2, \dots, M$ are disjoint and we have

$$\sum_{i=1}^M \xi_H\left(\mathbb{B}\left(\mathbf{U}^{(i)}, \|\cdot\|, \frac{\delta}{2}\right)\right) = \xi_H\left(\bigcup_{i=1}^M \mathbb{B}\left(\mathbf{U}^{(i)}, \|\cdot\|, \frac{\delta}{2}\right)\right) \leq \xi_H(\mathbb{V}_{s,r}) = 1. \quad (105)$$

For any $\mathbf{U}, \tilde{\mathbf{U}} \in \mathbb{V}_{s,r}$, there exists $\mathbf{\Gamma} \in \mathbb{O}_s$ such that $\mathbf{\Gamma}\mathbf{U} = \tilde{\mathbf{U}}$. Since $\|\cdot\|$ is invariant under $\mathbb{O}_s$, we have

$$\mathbf{\Gamma} \cdot \mathbb{B}\left(\mathbf{U}, \|\cdot\|, \frac{\delta}{2}\right) = \mathbb{B}\left(\mathbf{\Gamma}\mathbf{U}, \|\cdot\|, \frac{\delta}{2}\right) = \mathbb{B}\left(\tilde{\mathbf{U}}, \|\cdot\|, \frac{\delta}{2}\right).$$

Since ξ_H is invariant under left-orthogonal transformation, it follows that

$$\xi_H\left(\mathbb{B}\left(\mathbf{U}, \|\cdot\|, \frac{\delta}{2}\right)\right) = \xi_H\left(\mathbf{\Gamma} \cdot \mathbb{B}\left(\mathbf{U}, \|\cdot\|, \frac{\delta}{2}\right)\right) = \xi_H\left(\mathbb{B}\left(\tilde{\mathbf{U}}, \|\cdot\|, \frac{\delta}{2}\right)\right).$$

Therefore, all $\|\cdot\|$ -balls with the same radius have the same Haar measure, which does not depend on the particular choice of $\mathbf{U}$ due to the invariance properties of $\|\cdot\|$ and ξ_H . We denote the Haar measure of a $\|\cdot\|$ -balls with radius $\delta/2$ as

$$\psi(\delta/2) := \xi_H \left(\mathbb{B} \left(\mathbf{U}, \|\cdot\|, \frac{\delta}{2} \right) \right).$$

It then follows from Equation (105) that

$$M\psi(\delta/2) = \sum_{i=1}^M \xi_H \left(\mathbb{B} \left(\mathbf{U}^{(i)}, \|\cdot\|, \frac{\delta}{2} \right) \right) \leq 1,$$

from which we conclude that $M \leq 1/\psi(\delta/2)$. Taking the maximal such δ -packing set yields that

$$\mathcal{M}(\mathbb{V}_{s,r}, \|\cdot\|, \delta) \leq \frac{1}{\psi(\delta/2)}.$$

Recall that $\mathbb{K}$ is a subset of $\mathbb{V}_{s,r}$ and $\xi_H(\mathbb{K}) \geq \gamma$. Suppose that $\mathcal{C}_{\mathbb{K}} = \{\mathbf{V}^{(i)}\}_{i=1}^N$ is a $\delta/2$ -covering set of $\mathbb{K}$ under $\|\cdot\|$, then

$$\mathbb{K} \subseteq \bigcup_{i=1}^N \mathbb{B} \left(\mathbf{V}^{(i)}, \|\cdot\|, \frac{\delta}{2} \right)$$

and by the subadditivity of measures,

$$\sum_{i=1}^N \xi_H \left(\mathbb{B} \left(\mathbf{V}^{(i)}, \|\cdot\|, \frac{\delta}{2} \right) \right) \geq \xi_H \left(\bigcup_{i=1}^N \mathbb{B} \left(\mathbf{V}^{(i)}, \|\cdot\|, \frac{\delta}{2} \right) \right) \geq \xi_H(\mathbb{K}) \geq \gamma.$$

Thus, it follows that

$$N\psi(\delta/2) \geq \gamma \implies N \geq \gamma \mathcal{M}(\mathbb{V}_{s,r}, \|\cdot\|, \delta).$$

Taking the maximal $\delta/2$ -covering set yields that

$$\mathcal{N} \left(\mathbb{K}, \|\cdot\|, \frac{\delta}{2} \right) \geq \gamma \mathcal{M}(\mathbb{V}_{s,r}, \|\cdot\|, \delta).$$

Finally, by Lemma 4.2.8 in [54], we have

$$\mathcal{M} \left(\mathbb{K}, \|\cdot\|, \frac{\delta}{2} \right) \geq \mathcal{N} \left(\mathbb{K}, \|\cdot\|, \frac{\delta}{2} \right).$$

and

$$\mathcal{M}(\mathbb{V}_{s,r}, \|\cdot\|, \delta) \geq \mathcal{N}(\mathbb{V}_{s,r}, \|\cdot\|, \delta)$$

Combining the three displays above yields Equation (103), completing the proof. $\square$

Lemma 15 controls the Haar measure of $\mathbb{K}(s, r, \sqrt{\mu r/n})$. We remind the reader that $\mathbb{K}(s, r, \sqrt{\mu r/n})$ is defined in Equation (20).

Lemma 15. For $n/\mu \geq \max\{4, r\}$ and $s \geq (12n/\mu) \log(12n/\mu)$ we have

$$\xi_H(\mathbb{K}(s, r, \sqrt{\mu r/n})) \geq \frac{1}{2}.$$

Proof of Lemma 15. Define the set

$$\mathbb{L} \left(s, r, \sqrt{\frac{\mu}{n}} \right) = \left\{ \mathbf{U} \in \mathbb{V}_{s,r} : \|\mathbf{U}\|_{\infty} \geq \sqrt{\frac{\mu}{n}} \right\},$$

where $\|\mathbf{U}\|_{\infty}$ denotes the entrywise ℓ_{∞} norm (i.e., $\|\mathbf{U}\|_{\infty} := \max_{i,j} |U_{i,j}|$). In what follows, we omit the parameters and abbreviate $\mathbb{L}(s, r, \sqrt{\mu/n})$ to $\mathbb{L}$. Noting that for any $\mathbf{U} \in \mathbb{L}^c$, we have $\|\mathbf{U}\|_{\infty} \leq \sqrt{\mu/n}$, which implies that $\|\mathbf{U}\|_{2,\infty} \leq \sqrt{\mu r/n}$. Therefore, $\mathbb{L}^c \subseteq \mathbb{K}(s, r, \sqrt{\mu r/n})$ and

$$\xi_H \left(\mathbb{K}(s, r, \sqrt{\mu r/n}) \right) \geq 1 - \xi_H(\mathbb{L}). \quad (106)$$

Since

$$\mathbb{L} = \bigcup_{i=1}^r \left\{ \mathbf{U} \in \mathbb{V}_{s,r} : \|\mathbf{U}_{\cdot,i}\|_{\infty} \geq \sqrt{\frac{\mu}{n}} \right\},$$

it follows that

$$\xi_H(\mathbb{L}) \leq r \xi_H \left(\left\{ \mathbf{U} \in \mathbb{V}_{s,r} : \|\mathbf{U}_{\cdot,1}\|_{\infty} \geq \sqrt{\frac{\mu}{n}} \right\} \right).$$

Since $\mathbf{U}$ is distributed according to the Haar measure on $\mathbb{V}_{s,r}$, the column $\mathbf{U}_{\cdot,1} \in \mathbb{R}^s$ is uniformly distributed on $\mathbb{S}^{s-1}$. As a result, we have

$$\xi_H \left(\left\{ \mathbf{U} \in \mathbb{V}_{s,r} : \|\mathbf{U}_{\cdot,1}\|_{\infty} \geq \sqrt{\frac{\mu}{n}} \right\} \right) = \mathbb{P} \left(\frac{\|\mathbf{g}\|_{\infty}}{\|\mathbf{g}\|_2} \geq \sqrt{\frac{\mu}{n}} \right),$$

where $\mathbf{g} \sim N(0, \mathbf{I}_s)$ (see Lemma 10.1 in [44]). Combining the above two displays,

$$\xi_H(\mathbb{L}) \leq r \mathbb{P} \left(\frac{\|\mathbf{g}\|_{\infty}}{\|\mathbf{g}\|_2} \geq \sqrt{\frac{\mu}{n}} \right). \quad (107)$$

We introduce two events

$$\mathcal{E}_{0,1} := \left\{ \|\mathbf{g}\|_2^2 \geq \frac{3n}{\mu} \log s \right\} \quad \text{and} \quad \mathcal{E}_{0,2} := \left\{ \|\mathbf{g}\|_{\infty} \leq \sqrt{3 \log s} \right\}.$$

By Lemma 10.2 in [44], we have for all $t > 0$

$$\mathbb{P} \left(\|\mathbf{g}\|_2^2 \leq s - 2\sqrt{st} \right) \leq \exp(-t).$$

Taking $t = 2 \log s$ yields that

$$\mathbb{P} \left(\|\mathbf{g}\|_2^2 \leq s - 2\sqrt{2s \log s} \right) \leq \frac{1}{s^2}.$$

When $n \geq 4\mu$ and $s \geq (12n/\mu) \log(12n/\mu)$, one can verify that

$$s - 2\sqrt{2s \log s} \geq s/2 \geq \frac{3n}{\mu} \log s.$$

Thus, it follows that

$$\mathbb{P}(\mathcal{E}_{0,1}^c) \leq \mathbb{P} \left(\|\mathbf{g}\|_2^2 \leq s - 2\sqrt{2s \log s} \right) \leq \frac{1}{s^2}. \quad (108)$$

By standard Gaussian concentration inequalities and the union bound, we have

$$\mathbb{P}(\mathcal{E}_{0,2}^c) = \mathbb{P} \left(\|\mathbf{g}\|_{\infty} \geq \sqrt{3 \log s} \right) \leq 2s \exp(-3 \log s) = \frac{2}{s^2}. \quad (109)$$

On the event $\mathcal{E}_{0,1} \cap \mathcal{E}_{0,2}$, one has

$$\frac{\|\mathbf{g}\|_{\infty}}{\|\mathbf{g}\|_2} \leq \frac{\sqrt{3 \log s}}{\sqrt{3n \log s / \mu}} = \sqrt{\frac{\mu}{n}}.$$

Combining Equations (108) and Equations (109), it follows that

$$\mathbb{P} \left(\frac{\|\mathbf{g}\|_{\infty}}{\|\mathbf{g}\|_2} \geq \sqrt{\frac{\mu}{n}} \right) \leq \mathbb{P}(\mathcal{E}_{0,1}^c \cup \mathcal{E}_{0,2}^c) \leq \frac{3}{s^2}.$$

Applying this bound to Equation (107), we obtain

$$\xi_H(\mathbb{L}) \leq r \mathbb{P} \left(\frac{\|\mathbf{g}\|_{\infty}}{\|\mathbf{g}\|_2} \geq \sqrt{\frac{\mu}{n}} \right) \leq \frac{3r}{s^2}. \quad (110)$$

By the assumption that $n/\mu \geq \max\{4, r\}$, for any $r \geq 1$ we have

$$s \geq (12n/\mu) \log(12n/\mu) > 12r > \sqrt{6r}.$$

Combining the above bound with Equation (110), we have

$$\xi_H(\mathbb{L}) \leq \frac{1}{2}.$$

Finally, applying this upper bound to Equation (106), we have

$$\xi_H(\mathbb{K}(s, r, \sqrt{\mu r/n})) \geq 1 - \xi_H(\mathbb{L}) \geq \frac{1}{2},$$

as desired. $\square$

F.2 Proof of Lemma 4

Proof. Set $s = \lceil c_0^2 r / 8e^2 \delta^2 \rceil$. Under our upper bound assumption on δ in Equation (21) and $n/\mu \geq \max\{4, r\}$, we have

$$s \geq \frac{c_0^2 r}{8e^2} \cdot \frac{96e^2 n \log(12n/\mu)}{c_0^2 \mu r} = (12n/\mu) \log(12n/\mu) > 12r, \quad (111)$$

so that we always have

$$s \geq \max \left\{ \frac{12n}{\mu} \log \left(\frac{12n}{\mu} \right), \frac{c_0^2 r}{8e^2 \delta^2} \right\}. \quad (112)$$

Note that owing to our assumptions that $\mu \geq 12 \log(12n)$ and $\delta^2 \geq c_0^2 r / 8e^2 n$, we have $s \leq n$. Consider the subset

$$\mathbb{K}_s(n, r, \sqrt{\mu r/n}) := \{ \mathbf{U} \in \mathbb{K}_{r,\mu} : \mathbf{U}_{i,\cdot} = 0, \text{ for } s+1 \leq i \leq n \},$$

which is a subset related (but different) to $\mathbb{K}(s, r, \sqrt{\mu r/n})$ previously considered in Lemma 15. Our lower bound relies on the key observation that

$$\mathcal{M}(\mathbb{K}_{r,\mu}, d_{2,\infty}, \delta) \geq \mathcal{M}(\mathbb{K}_s(n, r, \sqrt{\mu r/n}), d_{2,\infty}, 2\delta) \geq \mathcal{M}(\mathbb{K}_s(n, r, \sqrt{\mu r/n}), d_F, 2\sqrt{s}\delta), \quad (113)$$

where the first inequality follows from Exercise 4.2.10 in [54], and the second inequality follows from the fact that for any $\mathbf{U}_1, \mathbf{U}_2 \in \mathbb{K}_s(n, r, \sqrt{\mu r/n})$, we have

$$d_{2,\infty}(\mathbf{U}_1, \mathbf{U}_2) \geq \frac{1}{\sqrt{s}} d_F(\mathbf{U}_1, \mathbf{U}_2).$$

Starting from Equation (113), we proceed to obtain a $\sqrt{s}\delta$ -packing for $\mathbb{K}_s(n, r, \sqrt{\mu r/n})$ under d_F . Since any $\mathbf{U} \in \mathbb{K}_s(n, r, \sqrt{\mu r/n})$ can be uniquely identified with an element in $\mathbb{K}(s, r, \sqrt{\mu r/n})$ by restricting it to its first s rows. By the assumption that $n/\mu \geq \max\{4, r\}$ and the lower bound on s in Equation (111), we verify that the conditions of Lemma 15 are all satisfied. Thus, following directly from the lower bound of $\xi_H(\mathbb{K}(s, r, \sqrt{\mu r/n}))$ in Lemma 15 and Lemma 14, we have

$$\mathcal{M}(\mathbb{K}_s(n, r, \sqrt{\mu r/n}), d_F, 2\sqrt{s}\delta) \geq \frac{1}{2} \mathcal{N}(\mathbb{V}_{s,r}, d_F, 4\sqrt{s}\delta).$$

Applying the lower bound in Lemma 13 to the right hand side of the above bound, we have

$$\mathcal{M}(\mathbb{K}_s(n, r, \sqrt{\mu r/n}), d_F, 2\sqrt{s}\delta) \geq \frac{1}{2} \left(\frac{c_0 \sqrt{r}}{4\sqrt{s}\delta} \right)^{r(s-r)}. \quad (114)$$

Under our upper bound assumption on δ in Equation (21) and $n/\mu \geq \max\{4, r\}$,

$$\frac{c_0^2 r}{384e^2 \delta^2} \geq \frac{c_0^2 r}{384e^2} \cdot \frac{96e^2 n \log(12n/\mu)}{c_0^2 \mu r} = \frac{n}{4\mu} \log(12n/\mu) \geq 1,$$

from our choice of s , it follows that

$$s \leq \frac{c_0^2 r}{8e^2 \delta^2} + 1 \leq \frac{c_0^2 r}{8e^2 \delta^2} + \frac{c_0^2 r}{384e^2 \delta^2} \leq \frac{c_0^2 r}{7e^2 \delta^2},$$

which implies $\sqrt{s}\delta \leq (\sqrt{7}e)^{-1} c_0 \sqrt{r}$. Thus, it follows from Equation (114) that

$$\mathcal{M}(\mathbb{K}_s(n, r, \sqrt{\mu r/n}), d_F, 2\sqrt{s}\delta) \geq \frac{1}{2} \left(\frac{\sqrt{7}e}{4} \right)^{r(s-r)} \geq \frac{1}{2} \left(\frac{\sqrt{7}e}{4} \right)^{\frac{rs}{2}},$$

where the last inequality follows from Equation (111). Noting that $\sqrt{7}e/4 \geq \sqrt{e}$, we have

$$\mathcal{M}(\mathbb{K}_s(n, r, \sqrt{\mu r/n}), d_F, 2\sqrt{s}\delta) \geq \frac{1}{2} \exp\left(\frac{rs}{4}\right)$$

Applying Equation (113) followed by Equation (112), we obtain

$$\begin{aligned} \mathcal{M}(\mathbb{K}_{r,\mu}, d_{2,\infty}, \delta) &\geq \mathcal{M}(\mathbb{K}_s(n, r, \sqrt{\mu r/n}), d_F, 2\sqrt{s}\delta) \\ &\geq \frac{1}{2} \exp\left(\max\left\{\frac{3nr}{\mu} \log\left(\frac{12n}{\mu}\right), \frac{c_0^2 r^2}{32e^2 \delta^2}\right\}\right) \geq \frac{1}{2} \exp\left(\frac{c_0^2 r^2}{32e^2 \delta^2}\right). \end{aligned} \quad (115)$$

Taking the logarithm on both sides of the above display yields Equation (22). $\square$

G Upper bounds of metric entropy under $d_{2,\infty}$

To provide a complete picture, Lemma 16 establishes an upper bound on the packing δ -entropy of $\mathbb{K}_{r,\mu}$ under $d_{2,\infty}$ that matches the lower bound in Lemma 4 up to log-factors when δ satisfies Equation (21).

Lemma 16. Assume that $n \geq \mu r$ for $r \in [n]$ and $\mu > 0$. For all n sufficiently large and any $\sqrt{4/(n-1)} < \delta < \sqrt{\mu r/n}$, we have

$$\log \mathcal{M}(\mathbb{K}_{r,\mu}, d_{2,\infty}, \delta) \lesssim \frac{r^2}{\delta^2} \log n. \quad (116)$$

Proof of Lemma 16. We obtain an upper bound of $\mathcal{M}(\mathbb{K}_{r,\mu}, d_{2,\infty}, \delta)$ by finding an upper bound of $\mathcal{N}(\mathbb{K}_{r,\mu}, d_{2,\infty}, \delta)$. Since for any $\mathbf{U}_1$ and $\mathbf{U}_2 \in \mathbb{R}^{n \times r}$,

$$d_{2,\infty}(\mathbf{U}_1, \mathbf{U}_2) \leq \|\mathbf{U}_1 - \mathbf{U}_2\|_{2,\infty},$$

we have

$$\mathcal{N}(\mathbb{K}_{r,\mu}, d_{2,\infty}, \delta) \leq \mathcal{N}(\mathbb{K}_{r,\mu}, \|\cdot\|_{2,\infty}, \delta). \quad (117)$$

Thus, it suffices to obtain an upper bound for $\mathcal{N}(\mathbb{K}_{r,\mu}, \|\cdot\|_{2,\infty}, \delta)$ instead. Let

$$\mathbb{T} := \{\mathbf{u} \geq 0 : \mathbf{u} \in \sqrt{r}\mathbb{S}^{n-1}, \|\mathbf{u}\|_\infty \leq 1\}. \quad (118)$$

For any $\mathbf{U} \in \mathbb{K}_{r,\mu}$, noting that $\sqrt{\mu r/n} \leq 1$, we define a mapping $h : \mathbb{K}_{r,\mu} \rightarrow \mathbb{T}$ given by

$$h(\mathbf{U}) := (\|\mathbf{U}_{1,\cdot}\|_2, \|\mathbf{U}_{2,\cdot}\|_2, \dots, \|\mathbf{U}_{n,\cdot}\|_2)^\top.$$

Indeed, $h(\mathbf{U}) \in \mathbb{T}$ since by definition, we have $h(\mathbf{U}) \geq 0$, $\|h(\mathbf{U})\|_2 = \sqrt{r}$ and $\|h(\mathbf{U})\|_\infty \leq 1$.

We pause to give a roadmap of our proof. Recall that for a set $\mathbb{K}$, its exterior δ -covering is a δ -covering, except that the exterior δ -covering allows elements not in $\mathbb{K}$ to form the covering (see Exercise 4.2.9 in [54]). To obtain $\mathcal{N}(\mathbb{K}_{r,\mu}, \|\cdot\|_{2,\infty}, \delta)$, we will explicitly construct an exterior δ -covering of $\mathbb{K}_{r,\mu}$ under $\|\cdot\|_{2,\infty}$. To do so, we proceed in three steps. In the first step, we construct an exterior δ -covering $\mathcal{C}$ of $\mathbb{T}$ under $\|\cdot\|_\infty$. In the second step, we divide the set $\mathbb{K}_{r,\mu}$ into separate subsets $\{\mathbb{U}_v\}_{v \in \mathbb{T}}$ via the mapping $h : \mathbb{K}_{r,\mu} \rightarrow \mathbb{T}$. In this way, each element v in $\mathbb{T}$ is associated with a subset $\mathbb{U}_v \subseteq \mathbb{K}_{r,\mu}$. For every $v \in \mathcal{C}$ (the exterior $(\delta/2)$ -covering set of $\mathbb{T}$), we construct an exterior $(\delta/2)$ -covering $\mathcal{C}_v$ of $\mathbb{U}_v$ under $\|\cdot\|_{2,\infty}$. In the last step, we take the union of $\mathcal{C}_v$ over $v \in \mathcal{C}$ to form a set $\mathcal{C}_{\mathbb{K}}$, and show this set is an exterior δ -covering set for $\mathbb{K}_{r,\mu}$ under $\|\cdot\|_{2,\infty}$. Finally, we control the cardinality of $\mathcal{C}_{\mathbb{K}}$ to obtain an upper bound of $\mathcal{N}(\mathbb{K}_{r,\mu}, \|\cdot\|_{2,\infty}, \delta)$, which in turn will yield our desired upper bound on $\mathcal{M}(\mathbb{K}_{r,\mu}, d_{2,\infty}, \delta)$.

Step 1. Construct an exterior $(\delta/2)$ -covering set of $\mathbb{T}$ under $\|\cdot\|_\infty$.

Recall that $\mathbb{T}$ is given in Equation (118). Let $s = \lceil 4/\delta^2 \rceil$. Under the assumption that $\delta > \sqrt{4/(n-1)}$, we have $s \leq n$. We consider the nonnegative integer solutions to the indeterminate equation

$$z_1 + z_2 + \dots + z_n = rs, \quad \|\mathbf{z}\|_\infty \leq s + 1 \quad (119)$$

and let

$$\mathcal{C} := \left\{ \sqrt{\frac{1}{s}} (\sqrt{z_1}, \dots, \sqrt{z_n}) : \mathbf{z} \text{ is a solution of Equation (119)} \right\}. \quad (120)$$

By Lemma 17, we have $\mathcal{C}$ forms an exterior $\delta/2$ -covering set of $\mathbb{T}$ under $\|\cdot\|_\infty$. Since without the constraint $\|\mathbf{z}\|_\infty \leq s + 1$, Equation (119) has $\binom{n+rs-1}{rs}$ solutions, we have

$$|\mathcal{C}| \leq \binom{n+rs-1}{rs} \leq \left(\frac{e(n+rs)}{rs} \right)^{rs} = \left(\frac{e(n+r \lceil 4/\delta^2 \rceil)}{r \lceil 4/\delta^2 \rceil} \right)^{r \lceil 4/\delta^2 \rceil}. \quad (121)$$

Step 2. Construct an exterior $(\delta/2)$ -covering set of $\mathbb{U}_v$ under $\|\cdot\|_{2,\infty}$.

For a given $v \in \mathcal{C}$, consider the subset

$$\mathbb{U}_v := \{\mathbf{U} \in \mathbb{K}_{r,\mu} : h(\mathbf{U}) = v\}, \quad (122)$$

our next step is to construct an exterior $(\delta/2)$ -covering set under $\|\cdot\|_{2,\infty}$ for $\mathbb{U}_v$. For every $i \in [n]$, consider a $(\delta/2)$ -covering set $\mathcal{C}_{v_i}$ of $v_i \mathbb{S}^{r-1}$ under the ℓ_2 norm. By Corollary 4.2.13 in [54],

$$|\mathcal{C}_{v_i}| \leq \left(\frac{6v_i}{\delta}\right)^r \quad \text{for all } i \in [n]. \quad (123)$$

Consider the set

$$\mathcal{C}_v := \{\Theta \in \mathbb{R}^{n \times r} : \Theta = (\theta_1, \theta_2, \dots, \theta_n)^\top, \theta_i \in \mathcal{C}_{v_i} \text{ for } i \in [n]\}. \quad (124)$$

We claim that $\mathcal{C}_v$ is an exterior $(\delta/2)$ -covering set for $\mathbb{U}_v$ under $\|\cdot\|_{2,\infty}$. To see this, note that for any $U \in \mathbb{U}_v$ and any $i \in [n]$, since $U_{i,\cdot}^\top \in v_i \mathbb{S}^{r-1}$ and $\mathcal{C}_{v_i}$ is a $(\delta/2)$ -covering of $v_i \mathbb{S}^{r-1}$ under the ℓ_2 norm, there exists a $\theta_i \in \mathcal{C}_{v_i}$ such that

$$\|\theta_i - U_{i,\cdot}^\top\|_2 \leq \frac{\delta}{2}, \quad \text{for all } i \in [n].$$

Let $\Theta = (\theta_1, \theta_2, \dots, \theta_n)^\top \in \mathcal{C}_v$. It follows that

$$\|\Theta - U\|_{2,\infty} = \max_{i \in [n]} \|\theta_i - U_{i,\cdot}^\top\|_2 \leq \frac{\delta}{2}.$$

Since $v \in \mathcal{C}$, from Equation (119), we have

$$\|v\|_0 \leq rs \quad (125)$$

as any solution z to Equation (119) has at most rs nonzero entries. By Equations (119) and (120), we also have

$$\|v\|_\infty \leq \frac{\sqrt{s+1}}{\sqrt{s}} \leq 2. \quad (126)$$

By Equations (123) and (124),

$$|\mathcal{C}_v| \leq \prod_{i=1}^n \left(\frac{6v_i}{\delta}\right)^r = \prod_{i:v_i>0} \left(\frac{6v_i}{\delta}\right)^r \leq \left(\frac{12}{\delta}\right)^{sr^2}.$$

where the last inequality follows from Equations (125) and (126). By the choice of $s = \lceil 4/\delta^2 \rceil$, we have

$$|\mathcal{C}_v| \leq \left(\frac{12}{\delta}\right)^{r^2 \lceil 4/\delta^2 \rceil}. \quad (127)$$

Step 3. Construct an exterior δ -covering set for $\mathbb{K}_{r,\mu}$ under $\|\cdot\|_{2,\infty}$.

Consider the set

$$\mathcal{C}_\mathbb{K} := \bigcup_{v \in \mathcal{C}} \mathcal{C}_v, \quad (128)$$

we claim that this set forms an exterior δ -covering set for $\mathbb{K}_{r,\mu}$ under $\|\cdot\|_{2,\infty}$.

For any $U \in \mathbb{K}_{r,\mu}$, since $\mathcal{C}$ is an exterior $(\delta/2)$ -covering of $\mathbb{T}$ under $\|\cdot\|_\infty$, we can find $v \in \mathcal{C}$ such that $\|h(U) - v\|_\infty \leq \delta/2$. Consider a matrix $\tilde{U} \in \mathbb{R}^{n \times r}$ with rows given by

$$\tilde{U}_{i,\cdot} = \begin{cases} v_i U_{i,\cdot} / h_i(U) & \text{if } h_i(U) > 0, \\ \theta_i^\top & \text{if } h_i(U) = 0, \end{cases}$$

for all $i \in [n]$, where θ_i is an arbitrary element of $\mathcal{C}_{v_i}$. Recall the definition of $\mathbb{U}_v$ from Equation (122). By construction, $\tilde{U} \in \mathbb{U}_v$, which implies that there exists $\Theta \in \mathcal{C}_v \subseteq \mathcal{C}_\mathbb{K}$ such that

$$\|\Theta - \tilde{U}\|_{2,\infty} \leq \frac{\delta}{2},$$

where we use the fact that $\mathcal{C}_v$ defined in Equation (124) is an exterior $(\delta/2)$ -covering of $\mathbb{U}_v$. For $i \in [n]$ such that $h_i(U) > 0$, we have

$$\|\tilde{U}_{i,\cdot} - U_{i,\cdot}\|_2 = \left| \frac{v_i}{h_i(U)} - 1 \right| \|U_{i,\cdot}\|_2 = |v_i - h_i(U)| \leq \frac{\delta}{2}$$

and for $i \in [n]$ such that $h_i(\mathbf{U}) = 0$, we have

$$\left\| \tilde{\mathbf{U}}_{i,\cdot} - \mathbf{U}_{i,\cdot} \right\|_2 = \|\boldsymbol{\theta}_i\|_2 = v_i = |v_i - h_i(\mathbf{U})| \leq \frac{\delta}{2}.$$

Combining the above two displays, it follows that

$$\left\| \tilde{\mathbf{U}} - \mathbf{U} \right\|_{2,\infty} = \max_{i \in [n]} \left\| \tilde{\mathbf{U}}_{i,\cdot} - \mathbf{U}_{i,\cdot} \right\|_2 \leq \frac{\delta}{2}.$$

Thus, we have found $\boldsymbol{\Theta} \in \mathcal{C}_{\mathbb{K}}$ such that

$$\|\boldsymbol{\Theta} - \mathbf{U}\|_{2,\infty} \leq \|\boldsymbol{\Theta} - \tilde{\mathbf{U}}\|_{2,\infty} + \|\tilde{\mathbf{U}} - \mathbf{U}\|_{2,\infty} \leq \delta,$$

and it follows that $\mathcal{C}_{\mathbb{K}}$ is an exterior δ -covering set for $\mathbb{K}_{r,\mu}$ under $\|\cdot\|_{2,\infty}$.

Step 4. An upper bound on $\mathcal{M}(\mathbb{K}_{r,\mu}, \|\cdot\|_{2,\infty}, \delta)$.

Recalling the definition of $\mathcal{C}_{\mathbb{K}}$ from Equation (128), Equations (121) and (127) imply that

$$|\mathcal{C}_{\mathbb{K}}| \leq \sum_{\mathbf{v} \in \mathcal{C}} |\mathcal{C}_{\mathbf{v}}| \leq \left(\frac{e(n + r \lceil 4/\delta^2 \rceil)}{r \lceil 4/\delta^2 \rceil} \right)^{r \lceil 4/\delta^2 \rceil} \cdot \left(\frac{12}{\delta} \right)^{r^2 \lceil 4/\delta^2 \rceil}.$$

Using this bound and the fact that $\mathcal{C}_{\mathbb{K}}$ is an exterior δ -covering set for $\mathbb{K}_{r,\mu}$ under $\|\cdot\|_{2,\infty}$, Exercise 4.2.9 in [54] implies that

$$\begin{aligned} \mathcal{N}(\mathbb{K}_{r,\mu}, \|\cdot\|_{2,\infty}, 2\delta) &\leq |\mathcal{C}_{\mathbb{K}}| \leq \left(\frac{e(n + r \lceil 4/\delta^2 \rceil)}{r \lceil 4/\delta^2 \rceil} \right)^{r \lceil 4/\delta^2 \rceil} \left(\frac{12}{\delta} \right)^{r^2 \lceil 4/\delta^2 \rceil} \\ &\leq \left(\frac{en\delta^2}{4r} + e \right)^{\frac{4r}{\delta^2} + r} \left(\frac{12}{\delta} \right)^{\frac{4r^2}{\delta^2} + r^2}. \end{aligned}$$

Combining this with Equation (117) and Lemma 4.2.8 in [54], we conclude that

$$\mathcal{M}(\mathbb{K}_{r,\mu}, d_{2,\infty}, \delta) \leq \mathcal{N}(\mathbb{K}_{r,\mu}, d_{2,\infty}, \delta/2) \leq \left(\frac{en\delta^2}{64r} + e \right)^{\frac{64r}{\delta^2} + r} \left(\frac{48}{\delta} \right)^{\frac{64r^2}{\delta^2} + r^2}. \quad (129)$$

Taking the logarithm on both sides of Equation (129), we have

$$\begin{aligned} \log \mathcal{M}(\mathbb{K}_{r,\mu}, d_{2,\infty}, \delta) &\leq r \left(\frac{64 + \delta^2}{\delta^2} \right) \log \left(\frac{en\delta^2}{64r} + e \right) + r^2 \left(\frac{64 + \delta^2}{\delta^2} \right) \log \left(\frac{48}{\delta} \right) \\ &\leq \frac{65r}{\delta^2} \log \left(\frac{e\mu}{64} + e \right) + \frac{65r^2}{\delta^2} \log \left(\frac{48}{\delta} \right) \end{aligned}$$

where the last inequality follows from $\delta < \sqrt{\mu r/n} \leq 1$. Under the assumption that $\delta > \sqrt{4/n}$ and $\mu r \leq n$, the right hand side of the above display can be further written as

$$\log \mathcal{M}(\mathbb{K}_{r,\mu}, d_{2,\infty}, \delta) \leq \frac{65r}{\delta^2} \log \left(\frac{en}{64} + e \right) + \frac{65r^2}{\delta^2} \log(24\sqrt{n}),$$

yields Equation (116) for all n sufficiently large, completing the proof. $\square$

Recall the definition of $\mathbb{T}$ in Equation (118) and $\mathcal{C}$ defined in Equation (120), we have Lemma 17.

Lemma 17. *Let the set $\mathbb{T} \subseteq \sqrt{r}\mathbb{S}^{n-1}$ be as defined in Equation (118), and let $\mathbf{h}$ be an arbitrary element in $\mathbb{T}$. Then there exists $\mathbf{v} \in \mathcal{C}$ such that*

$$\|\mathbf{h} - \mathbf{v}\|_{\infty} \leq \frac{1}{\sqrt{s}} \leq \frac{\delta}{2}, \quad (130)$$

where δ is given in Lemma 16 and $s = \lceil 4/\delta^2 \rceil$.

Proof. The inequality $1/\sqrt{s} \leq \delta/2$ in Equation (130) follows directly from $s = \lceil 4/\delta^2 \rceil$. To construct a $\mathbf{v} \in \mathcal{C}$ satisfying Equation (130), we first construct a $\boldsymbol{\zeta} \in \mathbb{R}^n$ and then make a slight modification to $\boldsymbol{\zeta}$ to obtain $\mathbf{v}$.

To construct $\boldsymbol{\zeta}$, we set ζ_1^2 to be either $\lfloor h_1^2 s \rfloor / s$ or $\lceil h_1^2 s \rceil / s$, and follow the recursive procedure to obtain an entrywise positive $\boldsymbol{\zeta} \geq 0$:

$$\zeta_i^2 = \begin{cases} \lfloor h_i^2 s \rfloor / s & \text{if } \sum_{j=1}^{i-1} h_j^2 - \sum_{j=1}^{i-1} \zeta_j^2 < 0, \\ \lceil h_i^2 s \rceil / s & \text{if } \sum_{j=1}^{i-1} h_j^2 - \sum_{j=1}^{i-1} \zeta_j^2 \geq 0 \end{cases} \quad (131)$$

for $2 \leq i \leq n$. The construction given in Equation (131) guarantees that

$$\|\mathbf{h} - \boldsymbol{\zeta}\|_\infty \leq \max_{i \in [n]} \left\{ h_i - \sqrt{\lfloor h_i^2 s \rfloor / s}, \sqrt{\lceil h_i^2 s \rceil / s} - h_i \right\} \leq \frac{1}{\sqrt{s}}, \quad (132)$$

where the last inequality follows from the fact that $\sqrt{a} - \sqrt{b} \leq \sqrt{a-b}$ for any $a \geq b > 0$. Now we show that following the procedure in Equation (131), we have

$$\left| \sum_{i=1}^k \zeta_i^2 - \sum_{i=1}^k h_i^2 \right| \leq \frac{1}{s}, \quad (133)$$

for all $k \in [n]$. We prove the above bound by induction on $k \in [n]$. When $k = 1$, we trivially have

$$|\zeta_1^2 - h_1^2| \leq \frac{1}{s}.$$

For the inductive step, suppose that for $k > 1$, we have

$$\left| \sum_{i=1}^{k-1} h_i^2 - \sum_{i=1}^{k-1} \zeta_i^2 \right| \leq \frac{1}{s}.$$

By definition in Equation (131), if

$$0 \leq \sum_{i=1}^{k-1} h_i^2 - \sum_{i=1}^{k-1} \zeta_i^2 \leq \frac{1}{s}, \quad (134)$$

then

$$-\frac{1}{s} \leq \sum_{i=1}^{k-1} h_i^2 - \sum_{i=1}^{k-1} \zeta_i^2 + h_k^2 - \frac{\lfloor h_k^2 s \rfloor}{s} \leq \frac{1}{s}.$$

Thus, $\zeta_k^2 = \lfloor h_k^2 s \rfloor / s$ satisfies

$$\left| \sum_{i=1}^k h_i^2 - \sum_{i=1}^k \zeta_i^2 \right| \leq \frac{1}{s}.$$

If, contrary to Equation (134), we have

$$-\frac{1}{s} \leq \sum_{i=1}^{k-1} h_i^2 - \sum_{i=1}^{k-1} \zeta_i^2 < 0,$$

then

$$-\frac{1}{s} \leq \sum_{i=1}^{k-1} h_i^2 - \sum_{i=1}^{k-1} \zeta_i^2 + h_k^2 - \frac{\lceil h_k^2 s \rceil}{s} \leq \frac{1}{s},$$

and thus $\zeta_k^2 = \lceil h_k^2 s \rceil / s$ again satisfies

$$\left| \sum_{i=1}^k h_i^2 - \sum_{i=1}^k \zeta_i^2 \right| \leq \frac{1}{s}.$$

Therefore, Equation (133) holds for all $k \in [n]$. Taking $k = n$ in Equation (133) and by the fact that $\|\mathbf{h}\|_2 = r$, we obtain that

$$|s\|\boldsymbol{\zeta}\|_2^2 - sr| \leq 1. \quad (135)$$

Following Equation (131), we also have that $s\zeta_i^2$ is an integer for every $i \in [n]$. Combined with Equation (135), we have a stronger statement that

$$(s\|\zeta\|_2^2 - sr) \in \{-1, 0, 1\}.$$

If $(s\|\zeta\|_2^2 - sr) = 0$, setting $\mathbf{v} = \zeta$ already guarantees that $\|\mathbf{v}\|_2^2 = r$ and $\|\mathbf{v} - \mathbf{h}\|_\infty \leq 1/\sqrt{s}$. Otherwise, if $(s\|\zeta\|_2^2 - sr) = -1$, then by Equation (131), there must be an $i_0 \in [n]$ such that

$$s\zeta_{i_0}^2 = \lfloor h_{i_0}^2 s \rfloor < h_{i_0}^2 s < \lceil h_{i_0}^2 s \rceil = \lfloor h_{i_0}^2 s \rfloor + 1.$$

Setting

$$v_i = \begin{cases} \sqrt{\zeta_{i_0}^2 + 1/s} = \sqrt{\lceil h_{i_0}^2 s \rceil / s}, & \text{if } i = i_0 \\ \zeta_i, & \text{otherwise} \end{cases}$$

guarantees that $\|\mathbf{v}\|_2^2 = r$ and

$$\|\mathbf{v} - \mathbf{h}\|_\infty = \max \left\{ \max_{i \neq i_0} |\zeta_i - h_i|, \left| \sqrt{\lceil h_{i_0}^2 s \rceil / s} - h_{i_0} \right| \right\} \leq \frac{1}{\sqrt{s}},$$

where the last inequality follows from Equation (132). Similarly, if $(s\|\zeta\|_2^2 - sr) = 1$, then we can alter one element of ζ to obtain $\mathbf{v}$, such that $\|\mathbf{v}\|_2^2 = r$ and $\|\mathbf{v} - \mathbf{h}\|_\infty \leq 1/\sqrt{s}$.

Since sv_i^2 are integers for all $i \in [n]$, $s\|\mathbf{v}\|_2^2 = sr$, and the fact that

$$s\|\mathbf{v}\|_\infty^2 \leq \max_{i \in [n]} \lceil h_i^2 s \rceil \leq s + 1,$$

where the last inequality follows from $\|\mathbf{h}\|_\infty \leq 1$ for all $\mathbf{h} \in \mathbb{T}$, we see that $\mathbf{v} \in \mathcal{C}$. $\square$

H Proof of Theorem 2

In this section, we use the Yang-Barron method [59] along with a Fano lower bound (see Proposition 15.12 and Lemma 15.21 in [56]) to prove the minimax lower bound for eigenspace estimation stated in Theorem 2. We follow the notation used in [56]. Given a class of distributions $\mathcal{P}$, we let θ denote a functional on the space $\mathcal{P}$, that is, a mapping from a distribution $\mathbb{P}$ to a parameter $\theta(\mathbb{P})$ taking values in some space Ω . We let $\rho : \Omega \times \Omega \rightarrow [0, \infty)$ be a semi-metric. Proposition 2 states the Fano lower bound (Proposition 15.12 in [56]).

Proposition 2 (Fano lower bound). *Let $\{\theta_1, \theta_2, \dots, \theta_M\} \subseteq \Omega$ be a 2δ -separated set under the ρ semi-metric, and suppose that J is uniformly distributed over the index set $[M]$, and $(Z \mid J = j) \sim \mathbb{P}_{\theta_j}$. Then for any increasing function $\Phi : [0, \infty) \rightarrow [0, \infty)$, the minimax risk is lower bounded as*

$$\inf_{\hat{\theta}} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \Phi \left(\rho \left(\hat{\theta}, \theta(\mathbb{P}) \right) \right) \geq \Phi(\delta) \left(1 - \frac{I(Z; J) + \log 2}{\log M} \right),$$

where the infimum is over all estimators $\hat{\theta}$ and $I(Z; J)$ is the mutual information between Z and J .

The Yang-Barron method gives an upper bound for the mutual information $I(Z; J)$.

Lemma 18 (Yang-Barron method [59]). *Let $\mathcal{N}_{\text{KL}}(\varepsilon; \mathcal{P})$ denote the ε -covering number of $\mathcal{P}$ in the square-root KL-divergence. Then the mutual information is upper bounded as*

$$I(Z; J) \leq \inf_{\varepsilon > 0} \log \mathcal{N}_{\text{KL}}(\varepsilon; \mathcal{P}) + \varepsilon^2.$$

Proof of Theorem 2. To each $(\mathbf{\Lambda}^*, \mathbf{U}^*) \in \Omega(\lambda^*, \mu, r)$, we associate a probability distribution $\mathbb{P}_{\mathbf{\Lambda}^*, \mathbf{U}^*}$ on $\mathbb{R}^{n \times n}$ with density given by

$$f_{\mathbf{\Lambda}^*, \mathbf{U}^*}(\mathbf{W}) = \prod_{1 \leq i \leq j \leq n} g \left(\frac{(\mathbf{W} - \mathbf{U}^* \mathbf{\Lambda}^* \mathbf{U}^{*\top})_{ij}}{\sigma} \right),$$

where $g(\cdot)$ denotes the density of a standard normal. We define the class of distributions

$$\mathcal{P} = \{\mathbb{P}_{\mathbf{\Lambda}^*, \mathbf{U}^*} : (\mathbf{\Lambda}^*, \mathbf{U}^*) \in \Omega(\lambda^*, \mu, r)\}.$$

For $\mu_1, \mu_2 \in \mathbb{R}^n$ and a pair of normal distributions $N(\mu_1, \sigma^2 \mathbf{I}_n)$ and $N(\mu_2, \sigma^2 \mathbf{I}_n)$, their KL-divergence is given by (see Example 15.13 in [56])

$$\text{KL}\left(N(\mu_1, \sigma^2 \mathbf{I}_n) \parallel N(\mu_2, \sigma^2 \mathbf{I}_n)\right) = \frac{1}{2\sigma^2} \|\mu_2 - \mu_1\|_2^2.$$

It follows that for any $\mathbb{P}_{\Lambda^*, U_1^*}, \mathbb{P}_{\Lambda^*, U_2^*} \in \mathcal{P}$,

$$\text{KL}\left(\mathbb{P}_{\Lambda^*, U_1^*} \parallel \mathbb{P}_{\Lambda^*, U_2^*}\right) \leq \frac{\lambda^{*2}}{2\sigma^2} \left\| U_1^* U_1^{*\top} - U_2^* U_2^{*\top} \right\|_F^2$$

and thus, taking square roots,

$$\sqrt{\text{KL}(\mathbb{P}_{\Lambda^*, U_1^*} \parallel \mathbb{P}_{\Lambda^*, U_2^*})} \leq \frac{\sqrt{2}\lambda^*}{2\sigma} \left\| U_1^* U_1^{*\top} - U_2^* U_2^{*\top} \right\|_F \leq \frac{\lambda^*}{\sigma} d_F(U_1^*, U_2^*), \quad (136)$$

where the second inequality holds from Lemma 2.6 in [27].

To apply the Yang-Barron method, we follow a two-step procedure (see Chapter 15 of [56]):

1. Pick the smallest ε such that $\varepsilon^2 \geq \log \mathcal{N}_{\text{KL}}(\mathcal{P}, \varepsilon)$.
2. Choose the largest δ that we can find a δ -packing that satisfies the lower bound

$$\log \mathcal{M}(\mathbb{K}_{r, \mu}, d_{2, \infty}, \delta) \geq 4\varepsilon^2 + 2 \log 2.$$

Then it follows from Lemma 18 and Proposition 2 that the minimax risk is lower bounded by $\delta/2$.

For the first step, noting that by Equation (136), a $(\sigma\varepsilon/\lambda^*)$ -covering set for $\mathbb{V}_{n, r}$ under d_F yields an ε -covering set for the $\sqrt{\text{KL}}$ -divergence, we have

$$\mathcal{N}_{\text{KL}}(\mathcal{P}, \varepsilon) \leq \mathcal{N}(\mathbb{V}_{n, r}, d_F, \varepsilon/\lambda^*) \leq \left(\frac{C_0 \lambda^* \sqrt{2r}}{\sigma \varepsilon} \right)^{r(n-r)},$$

where the last inequality follows from Lemma 13. Thus, in order to have $\varepsilon^2 \geq \log \mathcal{N}_{\text{KL}}(\mathcal{P}, \varepsilon)$, it suffices to have

$$\varepsilon^2 \geq r(n-r) \log \left(\frac{C_0 \lambda^* \sqrt{2r}}{\sigma \varepsilon} \right)$$

which holds if we set $\varepsilon = C_0 \lambda^* \sqrt{2r}/\sigma$. With this choice of ε , we then follow the second step of the Yang-Barron method and pick a δ to satisfy

$$\log \mathcal{M}(\mathbb{K}_{\mu, r}, d_{2, \infty}, \delta) \geq \frac{8C_0 r \lambda^{*2}}{\sigma^2} + 2 \log 2. \quad (137)$$

By Equation (115) in the proof of Lemma 4, we have

$$\log \mathcal{M}(\mathbb{K}_{r, \mu}, d_{2, \infty}, \delta) \geq \frac{c_0^2 r^2}{32e^2 \delta^2} - \log 2 \quad (138)$$

for δ satisfying Equation (21). Combining Equations (137) and (138), our goal is to find a δ , such that

$$\frac{c_0^2 r^2}{32e^2 \delta^2} \geq \frac{8C_0 r \lambda^{*2}}{\sigma^2} + 3 \log 2. \quad (139)$$

We pick δ to be

$$\delta = \frac{c_0 \sigma \sqrt{r}}{12e \lambda^* \sqrt{2C_0}} \wedge \frac{c_0}{4e} \sqrt{\frac{\mu r}{6n \log(12n/\mu)}}. \quad (140)$$

One can verify that δ satisfies Equation (21) under the assumption $\lambda^* \leq (6\sqrt{C_0})^{-1} \sigma \sqrt{n}$. To see that the δ given by Equation (140) satisfies Equation (139), we separate our discussion into two cases, one for large λ^* , and the other for small λ^* . When

$$\lambda^* \geq \sigma \sqrt{\frac{n \log(12n/\mu)}{3C_0 \mu}} \quad (141)$$

by Equation (140) we have

$$\delta = \frac{c_0 \sigma \sqrt{r}}{12e\lambda^* \sqrt{2C_0}} \quad (142)$$

and therefore, Equation (139) follows from

$$\frac{c_0^2 r^2}{32e^2 \delta^2} = \frac{9C_0 r \lambda^{*2}}{\sigma^2} \geq \frac{8C_0 r \lambda^{*2}}{\sigma^2} + 3 \log 2,$$

where the first equality holds from Equation (142) and the last inequality follows from Equation (141) for all n sufficiently large.

On the other hand, when

$$\lambda^* < \sigma \sqrt{\frac{n \log(12n/\mu)}{3C_0 \mu}}, \quad (143)$$

by Equation (140) we have

$$\delta = \frac{c_0}{4e} \sqrt{\frac{\mu r}{6n \log(12n/\mu)}}, \quad (144)$$

and therefore, Equation (139) follows from

$$\frac{c_0^2 r^2}{32e^2 \delta^2} = \frac{3rn \log(12n/\mu)}{\mu} \geq \frac{8rn \log(12n/\mu)}{\mu} + 3 \log 2 \geq \frac{8C_0 r \lambda^{*2}}{\sigma^2} + 3 \log 2,$$

where the first equality follows from Equation (144), the next inequality holds for all n sufficiently large, and the last inequality holds from Equation (143). Thus, we conclude that the choice of δ given in Equation (140) satisfies Equation (139), completing the second step of the Yang-Barron method.

Finally, combining Proposition 2 and Lemma 18, we conclude that for a universal constant $c > 0$,

$$\inf_{\hat{U}} \sup_{(\Lambda^*, U^*) \in \Omega(\lambda^*, \mu, r)} \mathbb{E}_{\Lambda^*, U^*} d_{2, \infty}(\hat{U}, U^*) \geq \frac{\delta}{2} \geq c \left(\frac{\sigma \sqrt{r}}{\lambda^*} \wedge \sqrt{\frac{\mu r}{n \log(n/\mu)}} \right),$$

completing the proof. $\square$

I Additional experiments

Here we collect additional experiments to complement our simulations in Section 5.

I.1 Additional numerical results on rank-one eigenvector estimation

We provide additional details for the experiments previously discussed in Section 5.1. Recall that we observe

$$\mathbf{Y} = \mathbf{M}^* + \mathbf{W} = \lambda^* \mathbf{u}^* \mathbf{u}^{*\top} + \mathbf{W},$$

and wish to recover $\mathbf{u}^* \in \mathbb{S}^{n-1}$. We refer the reader to Section 5.1 for the details of how λ^* , $\mathbf{u}^*$ and $\mathbf{W}$ are generated and how we run Algorithm 1.

In Section 5.1 we mentioned that under Laplacian noise, as shown in the third column of Figure 1, the estimator $\hat{\mathbf{u}}$ given by Algorithm 1 seems to have a slight dependence on the coherence parameter μ . We give a close examination of the estimation error of the largest entry of $\mathbf{u}^*$ in Figure 3, with shaded bands indicating 95% bootstrap confidence intervals. Figure 3 shows that the estimation error of the largest entry of $\mathbf{u}^*$ for $\hat{\mathbf{u}}$ is seen to be much smaller than 10^{-2} , while in the third column of Figures 1, the estimation error $d_\infty(\hat{\mathbf{u}}, \mathbf{u}^*)$ is well above 10^{-2} . This indicates that the dependence on μ observed in the right-hand subplot of Figure 1 comes not from estimating the largest entry of $\mathbf{u}^*$, but rather from estimating the other entries. In our experiment, the other entries are all nearly incoherent (that is, they have a magnitude at most $O(\sqrt{\log n/n})$), whence their estimation errors are expected to have no dependence on μ asymptotically. Thus, the slight dependence on μ exhibited Figure 1 is most likely to be a small order dependence compared to the error rate stated in Theorem 1, and should not affect the asymptotic error rate.

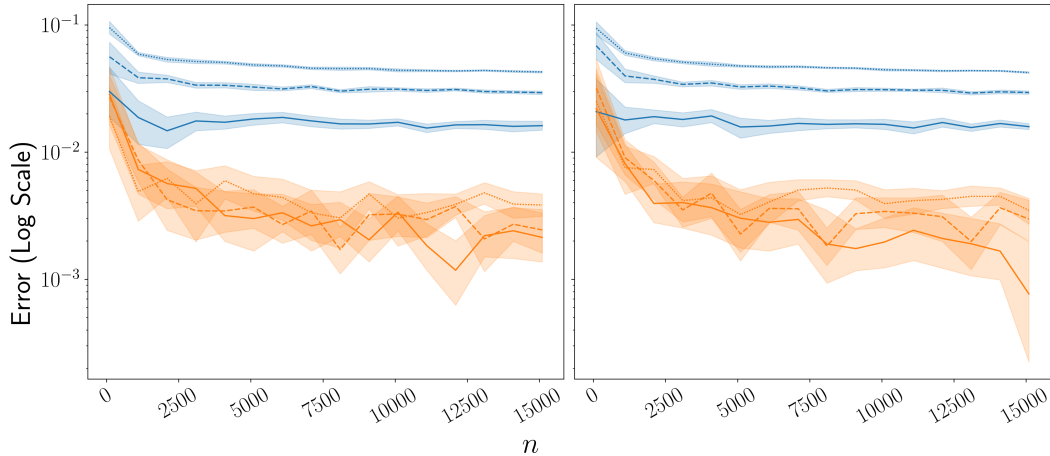


Figure 3: Numerical error in recovering the largest entry of $\mathbf{u}^*$ as a function of matrix dimension n , by the leading eigenvector (blue line) or the estimator given in Algorithm 1 (orange line) for three different choices of $\|\mathbf{u}^*\|_\infty$: 0.8 (dotted lines), 0.55 (dashed lines) and 0.3 (solid lines). The plot on the left corresponds to $\mathbf{u}^*$ generated via the Bernoulli scheme, while the plot on the right corresponds to the Haar scheme.

I.2 Additional numerical results on rank- r eigenvector estimation

We provide additional details for the experiments previously discussed in Section 5.2. Recall that we observe

$$\mathbf{Y} = \mathbf{M}^* + \mathbf{W} = \mathbf{U}^* \mathbf{\Lambda}^* \mathbf{U}^{*\top} + \mathbf{W},$$

and wish to recover $\mathbf{U}^* \in \mathbb{R}^{n \times r}$. In addition to the rank-2 setting discussed in Section 5.2, we include experimental results for the case when $\mathbf{M}^*$ has rank 3. We also provide estimation error under $d_{2,\infty}$ for our estimate $\hat{\mathbf{U}}$ in Algorithm 2 and the spectral estimate $\mathbf{U}$, to complement our results under d_∞ reported in Section 5.2. The experiments follow the same setup outlined Section 5.2, and we refer the readers there for details regarding running Algorithm 2 the generation of $\mathbf{\Lambda}^*$, $\mathbf{U}^*$ and $\mathbf{W}$.

We first provide more details as to how we obtain the estimation error under $d_{2,\infty}$. Normally, since

$$d_{2,\infty}(\hat{\mathbf{U}}, \mathbf{U}^*) = \min_{\mathbf{\Gamma} \in \mathbb{O}_r} \left\| \hat{\mathbf{U}} \mathbf{\Gamma} - \mathbf{U}^* \right\|_{2,\infty},$$

obtaining the $d_{2,\infty}$ estimation error requires finding an orthogonal matrix that minimizes a nonsmooth function, which is hard to achieve. In our simulation, however, since we have sufficiently large eigengaps between the eigenvalues, we know how each column of $\hat{\mathbf{U}}$ corresponds to the columns of $\mathbf{U}^*$. Thus, we merely need to resolve the ambiguity in the sign of each column of $\hat{\mathbf{U}}$, rather than searching over all $\mathbf{\Gamma} \in \mathbb{O}_r$. By assigning each column of $\hat{\mathbf{U}}$ the sign of the corresponding column in $\mathbf{U}^*$, we resolve this ambiguity and can obtain the $d_{2,\infty}$ estimation error.

Figure 4 compares the empirical accuracy, measured by $d_{2,\infty}$, of estimating $\mathbf{U}^*$ via the r leading eigenvectors $\mathbf{U}$ of $\mathbf{Y}$ (blue lines) and via $\hat{\mathbf{U}}$ produced by Algorithm 2 (orange lines). Shaded bands are generated from point-wise 95% confidence intervals using bootstrap approximation. Similar to the rank-one setting, Algorithm 2 recovers $\mathbf{U}^*$ with a much smaller estimation error under $d_{2,\infty}$ compared to the naïve spectral estimate, especially when the coherence μ is large. The figure also shows that Algorithm 2 performs well under different noise distributions.

Figure 5 corresponds to $\mathbf{M}^*$ having rank-3. It compares the empirical accuracy, measured under d_∞ , of estimating each $\mathbf{u}_k^*$, for $k = 1, 2, 3$ via the leading eigenvectors $\mathbf{u}_k$ of $\mathbf{Y}$ (blue/purple lines) and via $\hat{\mathbf{u}}_k$ given by Algorithm 2 (orange/red lines). Shaded bands are generated from point-wise 95% confidence intervals using bootstrap approximation. As in the first plot of Figure 2, under Gaussian noise, the estimation error of $\hat{\mathbf{u}}_k$ shows little to no visible dependence on μ , and is much smaller compared to the leading eigenvectors of $\mathbf{Y}$. Under Rademacher noise, as in the second plot

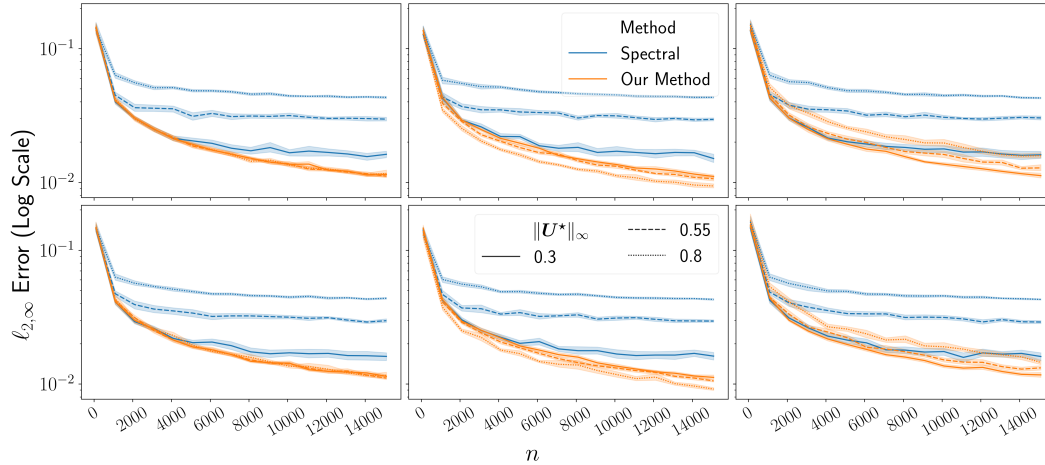


Figure 4: Error as measured in $d_{2,\infty}$ as a function of matrix dimension n , by spectral estimate (blue) and the estimator in Algorithm 2 (orange) for three different choices of $\|U^*\|_\infty$: 0.8 (dotted lines), 0.55 (dashed lines) and 0.3 (solid lines). Columns correspond to W being Gaussian (left), Rademacher (center) and Laplacian (right). The rows correspond to the signal matrix having rank-2 (top) and rank-3 (bottom).

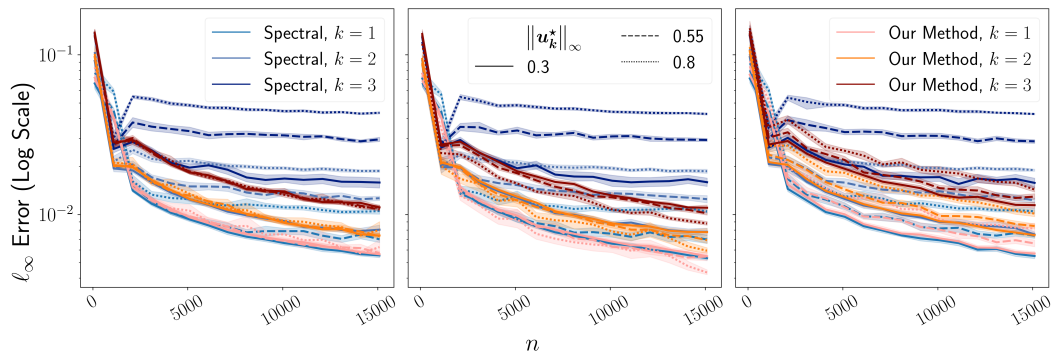


Figure 5: Error measured in d_∞ as a function of matrix dimension n , for the three leading signal eigenvectors u_k^* , $k = 1, 2, 3$ (line width) by the spectral estimate (blue/purple) or the estimator given in Algorithm 2 (orange/red) for three different choices of $\|u_k^*\|_\infty$: 0.8 (dotted lines), 0.55 (dashed lines) and 0.3 (solid lines). The plots correspond to W being Gaussian (left), Rademacher (center) and Laplacian (right).

of Figure 2, the dependence on μ again appears slightly reversed from that of the spectral estimator. For Laplacian noise, as in the third plot of Figure 2, there again seems to be a slight dependence on μ . For the same reason discussed in Sections 5.1 and 5.2, we expect such dependence to be of smaller order than the rate in Theorem 1 and should not affect the asymptotic error rate.

I.3 Comparison with AMP-based eigenvector estimation

As mentioned in the main text, to the best of our knowledge, we are the first paper to consider the problem of non-spectral entrywise eigenvector estimation. The nearest obvious method to serve as a comparison point, if we were to insist upon one, would likely be one based on approximate message passing (AMP; see [34] for an overview). AMP methods make no explicit coherence assumptions, but the underlying mechanism essentially requires incoherence: inherent to AMP-based eigenvector recovery methods is that the eigenvector is modeled as having its entries drawn i.i.d. according to a common distribution. Specifically, AMP methods typically make a mean field assumption whereby the empirical distribution of the entries of u^* converges in ℓ_2 to some distribution π . Since

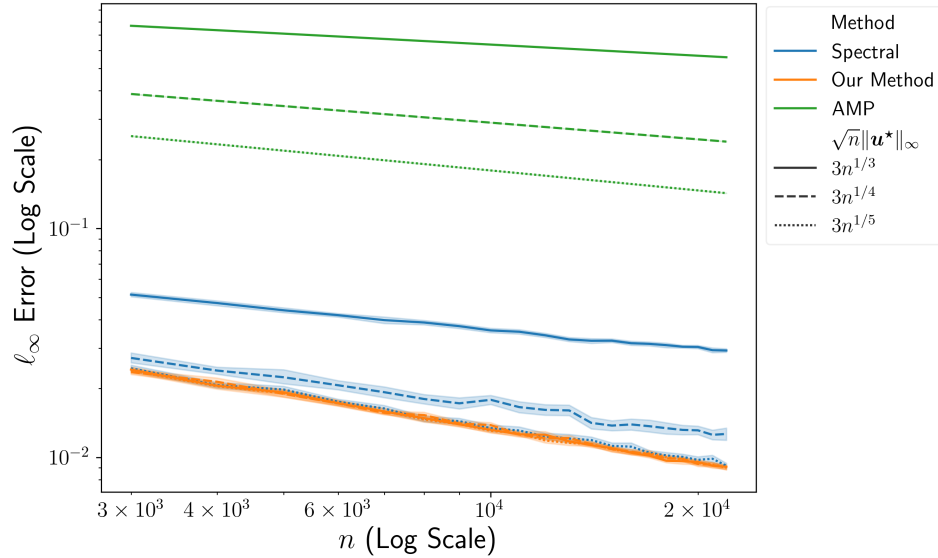


Figure 6: Estimation error under d_∞ as a function of size n , by approximate message passing (AMP; green), the leading eigenvector (blue) and Algorithm 1 (orange) for $\sqrt{n}\|\mathbf{u}^*\|_\infty$ equal to $3n^{1/3}$ (solid lines), $3n^{1/4}$ (dashed lines) or $3n^{1/5}$ (dotted lines) under Gaussian noise. Each data point is the mean of 30 independent trials.

AMP-based methods are tailored to ℓ_2 -recovery, they are suboptimal for entrywise recovery problems: small ℓ_2 error does not necessary imply small entrywise error.

The unsuitability of typical AMP methods notwithstanding, we include here a comparison against our Algorithm 1 for the sake of completeness. The experimental setup mirrors that of Figure 1, but now includes estimation error for an AMP-based method, as well. We generate $\mathbf{u}^*$ according to the following procedure: set a random entry of $\mathbf{u}^*$ to be $a \in \{3n^{1/3}/\sqrt{n}, 3n^{1/4}/\sqrt{n}, 3n^{1/5}/\sqrt{n}\}$, then generate the remaining entries by drawing uniformly from $\{\pm 1\}^{n-1}$ and normalizing these to have ℓ_2 norm $\sqrt{1 - a^2}$. This way, $\|\mathbf{u}^*\|_\infty = a$ and the coherence is $\mu = a^2 n$. We choose this setting to ensure that the limiting prior distribution satisfies the assumptions required by AMP, and thus we can apply the update procedures using Equations (22) (34) and the example after Equation (35) in [34].

Having generated $\mathbf{Y} = \mathbf{M}^* + \mathbf{W}$, we estimate $\mathbf{u}^*$ using the spectral estimate $\mathbf{u}$, the AMP method and our method as described in Section 5 and measure the estimation error under d_∞ . We report the mean of 30 independent trials for each combination of conditions (i.e., each combination of problem size n and magnitude a). We vary the matrix size n from 3000 to 22000 in increments of 1000.

Figure 6 compares the entrywise estimation error of the AMP method (green), the leading eigenvector of $\mathbf{Y}$ (blue) and our Algorithm 1 (orange). Across settings, Algorithm 1 recovers $\mathbf{u}^*$ with a much smaller error under d_∞ compared to the other two estimates, and shows an error rate with no visible dependence on the coherence. Both the spectral method and the AMP-based method exhibit different estimation error rates as the coherence increases. In particular, the AMP method performs much worse due to the fact that it is a mean field approximation method and not well-suited for our setting. Adapting AMP-based methods to target entrywise recovery rather than ℓ_2 recovery, in hopes of rectifying the poor performance exhibited in Figure 6, is a promising direction for future work.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: the abstract and introduction give overviews of our two main results: 1) a new method for estimating the eigenvectors in signal-plus-noise matrix models and 2) a new lower bound for eigenvector estimation in the small-eigenvalue regime. These results are expanded upon (including a thorough discussion of assumptions and limitations) in Sections 2, 3 4 and 6. Our experiments in Section 5 illustrate which of our assumptions can likely be relaxed or removed entirely.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: Our main assumptions, the reasons for needing them, and their implications are discussed in Sections 1 through 4. In particular, see Remarks 1, 2 and 6. Our experiments in Section 5 explore which of our assumptions might be relaxed or removed. A broad, albeit short, overview of limitations and areas for future work is also given in Section 6, titled "Discussion, limitations and conclusion". Computational costs and efficiency are addressed just before Algorithm 1.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.

- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory Assumptions and Proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [\[Yes\]](#)

Justification: Our main technical assumptions, Assumptions 1, 2, 3 and 4 are prominently laid out and explained in Sections 1 and 2. The necessity of these assumptions and the potential to relax or remove them are discussed in those sections, as well as in Section 6. All theorems and lemmas are stated with their necessary additional assumptions (e.g., growth rates and relations between parameters), and all are numbered, cross-referenced and have working hyperlinks. Full proofs are given in the supplementary materials. We have provided thorough background and intuition surrounding these proofs, though proof sketches have largely been removed due to space constraints. We are committed to including proof sketches in a camera-ready version using the allotted extra page, should the reviewers request them.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental Result Reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [\[Yes\]](#)

Justification: Algorithms are described in the text, including details regarding how hyperparameters were selected and minor changes made to Algorithms 1 and 2 “as written” as opposed to “as run on the cluster”.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [\[Yes\]](#)

Justification: in addition to the algorithmic descriptions in Sections 2 and 3 and the details in Section 5, we have included code for running all reported experiments in our supplemental materials.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental Setting/Details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [\[Yes\]](#)

Justification: the main such concern in this paper surrounds the selection of the parameters α_0 and β in Algorithm 1. Selection of these parameters is discussed at length in Remark 2 and in our description of the experiments in Section 5. We have also provided code that reflects the discussion in the paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment Statistical Significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [\[Yes\]](#)

Justification: all experiment plots include error bars denoting 95% bootstrap confidence intervals. This is discussed in Section 5. It is clearly stated in the text that the randomness in the error bars comes from independent repetitions of the same experimental setup.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments Compute Resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [\[Yes\]](#)

Justification: This is addressed at the beginning of Section 5.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code Of Ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [\[Yes\]](#)

Justification: This paper concerns a theoretical question surrounding minimax estimation rates. It involves neither human subjects nor sensitive data. Societal impacts, either positive or negative, are of course possible (e.g., statistical methods and models are frequently abused or used toward malicious ends), but our work creates no risks that warrant mitigation, to the best of our knowledge. All experimental code uses only open source software tools.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [\[Yes\]](#)

Justification: This paper concerns a theoretical/methodological question regarding estimation of eigenvectors from noisy matrices. Given the theoretical nature of this work, it is hard to foresee direct social impacts, either positive or negative, from this work. Positive impacts of eigenvector estimation methods (e.g., PCA, matrix denoising, etc) are ubiquitous (see, for example, medical image processing, to name just one application area). These positive impacts are alluded to in the introduction. Of course, the abuse or misuse of statistical methods and models has frequently resulted in negative societal impacts and inequitable outcomes for underrepresented groups. We acknowledge these possible negative outcomes, albeit briefly, in Section 6.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [\[NA\]](#)

Justification: This paper concerns a theoretical question surrounding minimax estimation rates and includes no models or data with risk of misuse, to the best of our knowledge.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [NA]

Justification: the experimental code, included in the supplementary materials, uses only open source software tools.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. New Assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: the experimental code, included in the supplementary materials, uses only open source software tools.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. Crowdsourcing and Research with Human Subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: This paper involves neither crowdsourcing nor human subject research.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: This paper involves neither crowdsourcing nor human subject research.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.