Supplementary Materials for " Subspace Estimation from Unbalanced and Incomplete Data Matrices: $\ell_{2,\infty}$ Statistical Guarantees "

Changxiao Cai* Gen Li† Yuejie Chi‡ H. Vincent Poor* Yuxin Chen*

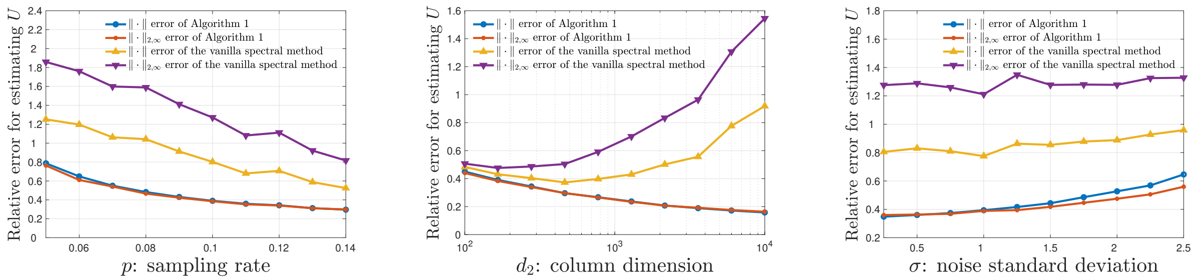
June 9, 2020

Abstract

This supplementary shows details that supports the main text " Subspace Estimation from Unbalanced and Incomplete Data Matrices: $\ell_{2,\infty}$ Statistical Guarantees " submitted to Annals of Statistics. One can find numerical experiments and proofs of our main theorem, corollaries and technical lemmas.

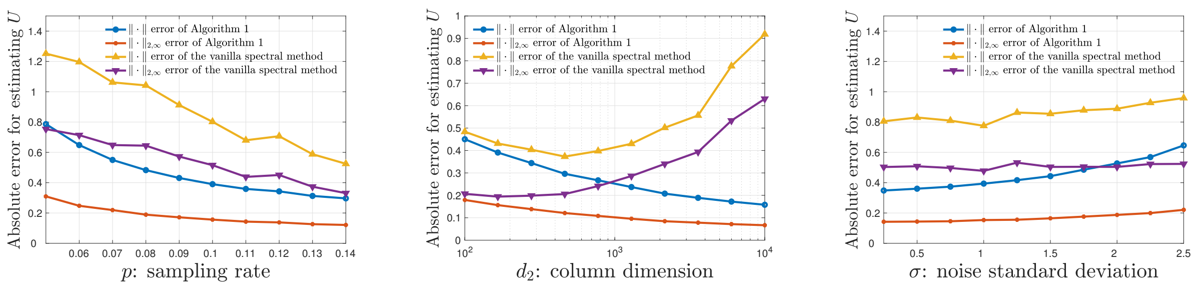
Contents

Corresponding author: Yuxin Chen.
*Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA; email: {ccai,poor,yuxin.chen}@princeton.edu.
†Department of Electronic Engineering, Tsinghua University, Beijing, China 100086, USA; email: g-li16@mails.tsinghua.edu.cn.
‡Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA; email: yuejiechi@cmu.edu.



(a) (b) (c)

Figure 1: Relative estimation errors of the subspace estimate $\boldsymbol{U}$ for both Algorithm 1 and the vanilla spectral method. The results are reported for (a) relative error vs. sampling rate p (where $d_1 = 100$, $d_2 = 1000$, $r = 4$, $\sigma = 1$), (b) relative error vs. column dimension d_2 (where $d_1 = 100$, $r = 4$, $\sigma = 1$, $p = \frac{2r\log(d_1+d_2)}{\sqrt{d_1 d_2}}$), and (c) relative error vs. noise standard deviation σ (where $d_1 = 100$, $d_2 = 1000$, $r = 4$, $p = 0.1$).



(a) (b) (c)

Figure 2: Absolute estimation errors of the subspace estimate $\boldsymbol{U}$ for both Algorithm 1 and the vanilla spectral method. The results are reported for (a) absolute error vs. sampling rate p (where $d_1 = 100$, $d_2 = 1000$, $r = 4$, $\sigma = 1$), (b) absolute error vs. column dimension d_2 (where $d_1 = 100$, $r = 4$, $\sigma = 1$, $p = \frac{2r\log(d_1+d_2)}{\sqrt{d_1 d_2}}$), and (c) absolute error vs. noise standard deviation σ (where $d_1 = 100$, $d_2 = 1000$, $r = 4$, $p = 0.1$).

7 Numerical experiments

To confirm the applicability of our algorithm and the theoretical findings, we conduct a series of numerical experiments. All results reported in this subsection are averaged over 100 independent Monte Carlo trials. For the sake of comparisons, we also report the numerical performance of the vanilla spectral method (namely, returning the r-dimensional principal column subspace of $\boldsymbol{A}$ directly without proper diagonal deletion).

Subspace estimation for random low-rank data matrices. We start with subspace estimation for a randomly generated matrix $\boldsymbol{A}^\star$. Specifically, generate $\boldsymbol{A}^\star = \boldsymbol{Z}_1 \boldsymbol{Z}_2^\top$, where $\boldsymbol{Z}_1 \in \mathbb{R}^{d_1 \times r}$, $\boldsymbol{Z}_2 \in \mathbb{R}^{d_2 \times r}$ consist of i.i.d. standard Gaussian entries. The noise matrix contains i.i.d. Gaussian entries, namely, $N_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right)$ for each $(i,j) \in [d_1] \times [d_2]$. Figures 1 and 2 plot respectively the numerical estimation errors of the estimate $\boldsymbol{U}$ vs. the sampling rate p, the column dimension d_2, and the standard deviation σ of noise. Two types of estimation errors are reported: (1) the absolute spectral norm error $\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|$ and the relative spectral norm error $\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\| \,/\, \|\boldsymbol{U}^\star\|$; (2) the absolute $\ell_{2,\infty}$ norm error $\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|_{2,\infty}$ and the relative $\ell_{2,\infty}$ norm error $\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|_{2,\infty} \,/\, \|\boldsymbol{U}^\star\|_{2,\infty}$, where $\boldsymbol{R} := \arg\min_{\boldsymbol{Q} \in \mathcal{O}_{r \times r}} \|\boldsymbol{U}\boldsymbol{Q} - \boldsymbol{U}^\star\|_{\mathrm{F}}$. As can be seen from the plots, Algorithm 1 yields reasonably good estimates in terms of both the spectral norm and the $\ell_{2,\infty}$ norm, outperforming the vanilla spectral method in all experiments.

Tensor completion from noise data. Next, we consider numerically the problem of tensor completion from noisy observations of its entries. Recall the notations in Section 4.1. We generate $\boldsymbol{W}^\star \in \mathbb{R}^{d \times r}$ with

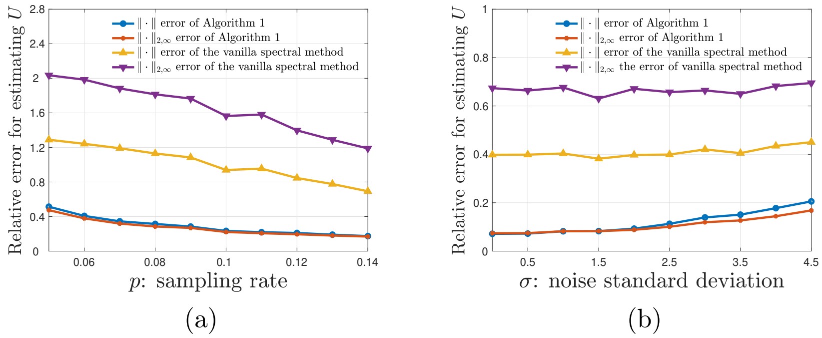


Figure 3: Relative estimation errors of the subspace $\boldsymbol{U}$ spanned by tensor components in tensor completion. The results are plotted for (a) relative error vs. sampling rate p (where $d = 100$, $r = 4$, $\sigma = 2$), and (b) relative error vs. noise standard deviation σ (where $d = 100$, $r = 4$, $p = 0.1$).

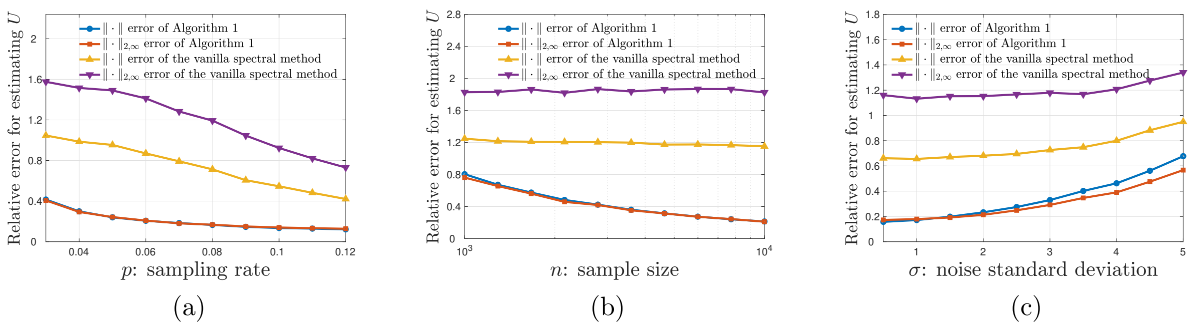


Figure 4: Relative error of the estimate $\boldsymbol{U}$ for covariance estimation with missing data. The results are shown for (a) relative error vs. sampling rate p (where $d = 100$, $n = 5000$, $r = 4$, $\sigma = 1$), (b) relative error vs. sample size n (where $d = 100$, $r = 4$, $\sigma = 1$, $p = 0.05$), and (c) relative error vs. noise standard deviation σ (where $d = 100$, $n = 5000$, $r = 4$, $p = 0.1$).

i.i.d. standard Gaussian entries, and generate $N_{i,j,k} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right)$ independently for each $(i, j, k) \in [d]^3$. Figure 3(a) and Figure 3(b) illustrate the relative estimation errors of the subspace estimate $\boldsymbol{U}$ vs. the sampling rate p and noise standard deviation σ, respectively. Encouragingly, Figure 3 shows that Algorithm 2 accurately recovers the subspace spanned by the tensor factors of interest (with respect to both the spectral norm and the $\ell_{2,\infty}$ norm); in particular, it is capable of producing faithful subspace estimates even when the vanilla spectral method fails.

Covariance estimation with missing data. The next series of experiments is concerned with covariance estimation with missing data. Recall the notations in Section 4.2. We draw $\boldsymbol{x}_i \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}^\star\right)$ independently with $\boldsymbol{\Sigma}^\star = \boldsymbol{U}^\star \boldsymbol{U}^{\star\top}$, where $\boldsymbol{U}^\star \in \mathbb{R}^{d\times r}$ is a i.i.d. standard Gaussian random matrix in $\mathbb{R}^{d\times r}$, and $\boldsymbol{\varepsilon}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_d\right)$ for each $1 \le i \le n$. We first consider the estimation error of the subspace. The numerical estimation errors of the estimate $\boldsymbol{U}$ vs. the sampling rate p, the sample size n and the noise standard deviation σ are plotted in Figure 4(a) – Figure 4(c), respectively. We then turn to the estimation accuracy of the covariance matrix. The numerical estimation errors of the estimate $\boldsymbol{S}$ of Algorithm 3 and the vanilla spectral method vs. the sampling rate p, the sample size n and the noise standard deviation σ are plotted in Figure 5, respectively. Similar to previous experiments, Algorithm 3 produces reliable estimates both in terms of the spectral norm, the $\ell_{2,\infty}$ norm and the ℓ_∞ norm accuracy.

Community recovery in bipartite stochastic block model. Finally, we conduct numerical experiments for community recovery in bipartite stochastic block models. The parameters are chosen to be $q_{\mathsf{in}} = \frac{a \log(n_u + n_v)}{\sqrt{n_u n_v}}$ and $q_{\mathsf{out}} = \frac{b \log(n_u + n_v)}{\sqrt{n_u n_v}}$ for some constants $a > b > 0$. Figure 6(a) reveals a phase transition

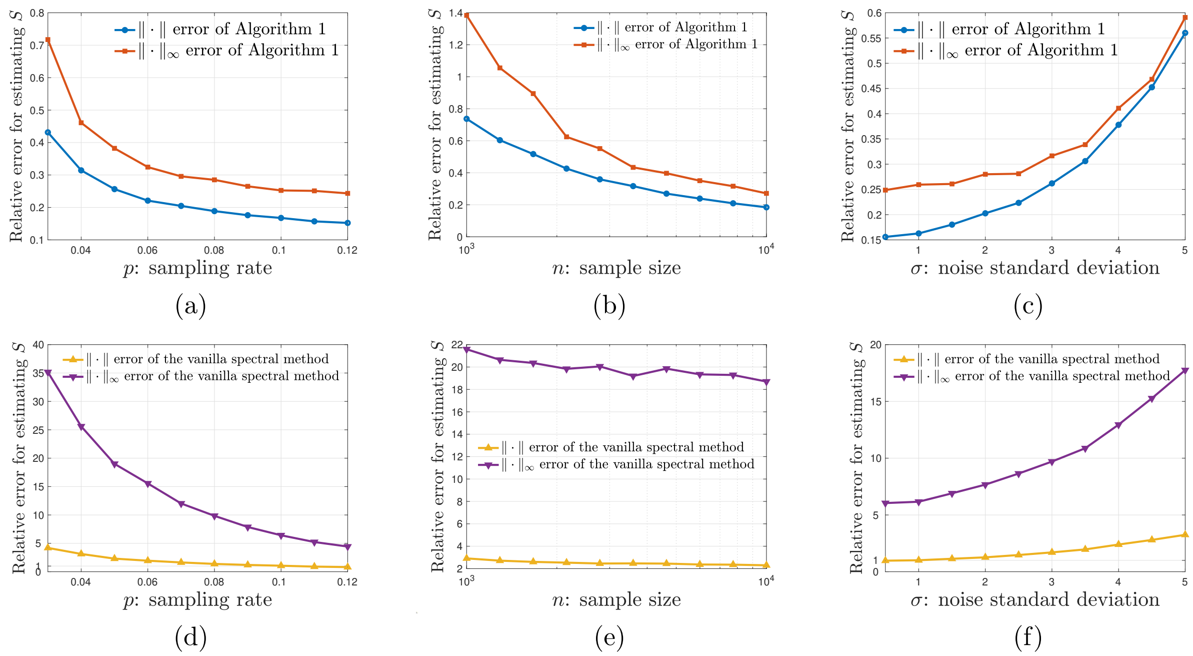


Figure 5: Relative error of the estimate $\boldsymbol{S}$ of Algorithm 3 (top row) and the vanilla spectral method (bottom row) for covariance estimation with missing data. The results are shown for (a,d) relative error vs. sampling rate p (where $d = 100$, $n = 5000$, $r = 4$, $\sigma = 1$), (b,e) relative error vs. sample size n (where $d = 100$, $r = 4$, $\sigma = 1$, $p = 0.05$), and (c,f) relative error vs. noise standard deviation σ (where $d = 100$, $n = 5000$, $r = 4$, $p = 0.1$). It is worthnoting the different scales of the y-axis when plotting the errors of the two algorithms.

phenomenon concerned with exact community recovery. As can be seen, Algorithm 4 always succeeds in achieving exact recovery once a — or equivalently q_{in} — exceeds a certain threshold, which outperforms the vanilla spectral method. In Figure 6(b), we vary the number n_v nodes in $\mathcal{V}$ and plot the empirical success rates for exact recovery. The advantage of Algorithm 4 compared to the vanilla spectral method can be clearly seen from the plot.

8 Analysis

In this section, we discuss in detail the analysis techniques employed to establish Theorem 3.1. This is built upon a leave-one-out (as well as a leave-two-out) analysis strategy that is particularly effective in controlling entrywise and $\ell_{2,\infty}$ estimation errors [EKBB+13, EK15, ZB18, CFMW19, AFWZ17, SCC17, CCFM19, CFMY19, CLL19, LBEK18, PW19].

8.1 Leave-one-out and leave-two-out estimates

In order to facilitate the analysis when bounding $\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|_{2,\infty}$, we introduce a set of auxiliary leave-one-out matrices — a powerful analysis technique that has been employed to decouple complicated statistical dependency. It is worth emphasizing that these procedures are never executed in practice. Specifically, for each $1 \le m \le d_1$, we introduce an auxiliary matrix

$$\boldsymbol{A}^{(m)} = \mathcal{P}_{-m,:}(\boldsymbol{A}) + p\mathcal{P}_{m,:}(\boldsymbol{A}^\star), \tag{1}$$

where $\mathcal{P}_{-m,:}$ (resp. $\mathcal{P}_{m,:}$) represents the projection onto the subspace of matrices supported on the index subset $\{[d_1] \setminus \{m\}\} \times [d_2]$ (resp. $\{m\} \times [d_2]$). In other words, $\boldsymbol{A}^{(m)}$ is obtained by replacing all entries in the m-th row by their expected values (taking into account the sampling rate). By construction, (1) $\boldsymbol{A}^{(m)}$ is statistically independent of the data in the m-th row of $\boldsymbol{A}$, and (2) $\boldsymbol{A}^{(m)}$ is expected to be quite close to $\boldsymbol{A}$,

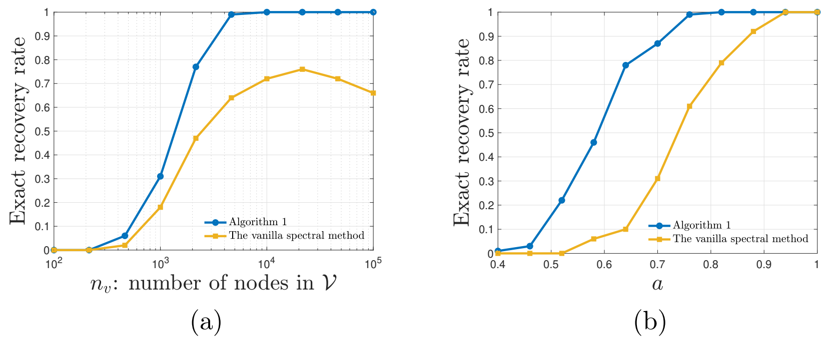


Figure 6: Empirical success rates for exact community recovery in bipartite stochastic block models, where $q_{\mathsf{in}} = \frac{a\log(n_u+n_v)}{\sqrt{n_u n_v}}$ and $q_{\mathsf{out}} = \frac{b\log(n_u+n_v)}{\sqrt{n_u n_v}}$. The results are shown for (a) empirical success rate vs. the number n_v of nodes in $\mathcal{V}$ (where $n_u = 100$, $a = 0.8$, $b = 0.01$), and (b) empirical exact recovery rate vs. a (where $n_u = 100$, $n_v = 10000$, $b = 0.01$).

as we only discard a small fraction of data when constructing $\boldsymbol{A}^{(m)}$. These two observations taken together allow for optimal control of the estimation error in the m-th row of $\boldsymbol{U}$.

Armed with the leave-one-out matrices, we are ready to introduce auxiliary leave-one-out procedures for subspace estimation. Similar to the matrix $\boldsymbol{G}$ in Algorithm 1 (whose eigenspace serves as an estimate of the column space of $\boldsymbol{U}^\star$), we define an auxiliary matrix $\boldsymbol{G}^{(m)} \in \mathbb{R}^{d_1 \times d_1}$ as follows:

$$\boldsymbol{G}^{(m)} = \mathcal{P}_{\mathsf{off\text{-}diag}}\Big(\tfrac{1}{p^2}\boldsymbol{A}^{(m)}\boldsymbol{A}^{(m)\top}\Big), \tag{2}$$

where $\mathcal{P}_{\mathsf{off\text{-}diag}}(\cdot)$ (as already defined in Section 2.2) extracts out all off-diagonal entries from a matrix. The auxiliary procedure, which is summarized in Algorithm 1, is very similar to Algorithm 1 except that it operates upon $\boldsymbol{G}^{(m)}$.

Algorithm 1 The m-th leave-one-out sequence

1: **Input:** sampling set Ω, observed entries $\{A_{i,j} \mid (i,j) \in \Omega\}$, true entries $\{A^\star_{m,j} \mid j \in [d_2]\}$, sampling rate p, rank r.
2: Let $\boldsymbol{U}^{(m)}\boldsymbol{\Lambda}^{(m)}\boldsymbol{U}^{(m)\top}$ be the (truncated) rank-r eigen-decomposition of $\boldsymbol{G}^{(m)}$. Here, $\boldsymbol{G}^{(m)}$ and $\boldsymbol{A}^{(m)}$ are defined respectively in (2) and (1).
3: **Output** $\boldsymbol{U}^{(m)}$ as the subspace estimate and $\boldsymbol{\Sigma}^{(m)} = \big(\boldsymbol{\Lambda}^{(m)}\big)^{1/2}$ as the spectrum estimate.

Given that $\boldsymbol{A}^{(m)}$ (resp. $\boldsymbol{G}^{(m)}$) is very close to $\boldsymbol{A}$ (resp. $\boldsymbol{G}$), one would naturally expect $\boldsymbol{U}^{(m)}$ — the r-dimensional principal eigenspace of $\boldsymbol{G}^{(m)}$ — to stay extremely close to the original estimate $\boldsymbol{U}$. This fact will be formalized shortly.

As it turns out, given that the spectral method is applied to the Gram matrix (which is a quadratic form of the original data matrix), introducing the leave-one-out sequences alone is not yet sufficient for our purpose; we still need to introduce an additional set of "leave-two-out" matrices, in the hope of simultaneously handling the row-wise and the column-wise statistical dependency. Specifically, for each $1 \le m \le d_1$ and each $1 \le l \le d_2$, define the following auxiliary matrices:

$$\boldsymbol{A}^{(m,l)} := \mathcal{P}_{-m,-l}(\boldsymbol{A}) + p\mathcal{P}_{m,l}(\boldsymbol{A}^\star), \tag{3a}$$

$$\boldsymbol{G}^{(m,l)} := \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\tfrac{1}{p^2}\boldsymbol{A}^{(m,l)}\boldsymbol{A}^{(m,l)\top}\right), \tag{3b}$$

where $\mathcal{P}_{-m,-l}$ (resp. $\mathcal{P}_{m,l}$) denotes the projection onto the subspace of matrices supported on $\{[d_1]\setminus\{m\}\} \times \{[d_2]\setminus\{l\}\}$ (resp. $\{m\}\times\{l\}$). Similar to $\boldsymbol{A}^{(m)}$, $\boldsymbol{A}^{(m,l)}$ is generated by replacing all data lying on the m-th row and the l-th column of $\boldsymbol{A}$ by their expected values (taking into account the sampling rate). The precise procedure is summarized in Algorithm 2. Similar to the leave-one-out estimates, one expects the new leave-two-out estimates $\boldsymbol{U}^{(m,l)}$ to be extremely close to $\boldsymbol{U}^{(m)}$ (and hence $\boldsymbol{U}$).

Algorithm 2 The (m,l)-th leave-two-out sequence

1: **Input:** sampling set Ω, observed entries $\{A_{i,j} \mid (i,j) \in \Omega\}$, true entries $\{A^{\star}_{m,j} \mid j \in [d_2]\} \cup \{A^{\star}_{i,l} \mid i \in [d_1]\}$, sampling rate p, rank r.

2: Let $\boldsymbol{U}^{(m,l)}\boldsymbol{\Lambda}^{(m,l)}\boldsymbol{U}^{(m,l)\top}$ be the (truncated) rank-r eigen-decomposition of $\boldsymbol{G}^{(m,l)}$. Here, $\boldsymbol{G}^{(m,l)}$ and $\boldsymbol{A}^{(m,l)}$ are defined respectively in (3b) and (3a).

3: **Output** $\boldsymbol{U}^{(m,l)}$ as the subspace estimate and $\boldsymbol{\Sigma}^{(m,l)} = (\boldsymbol{\Lambda}^{(m,l)})^{1/2}$ as the spectrum estimate.

8.2 Key lemmas

In this subsection, we provide several lemmas that play a crucial role in establishing our main theorem. These lemmas are primarily concerned with the proximity between the original estimate, the leave-one-out estimates, and the ground truth. Throughout this section, we let

$$\boldsymbol{G}^{\star} := \boldsymbol{A}^{\star}\boldsymbol{A}^{\star\top} = \boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star 2}\boldsymbol{U}^{\star\top}. \tag{4}$$

To begin with, we demonstrate that $\boldsymbol{G}$ is sufficiently close to $\boldsymbol{G}^{\star}$ when the difference is measured by the spectral norm. In view of standard matrix perturbation theory (which we shall make precise later), the proximity of $\boldsymbol{G}$ and $\boldsymbol{G}^{\star}$ is crucial in bounding the difference between $\boldsymbol{U}$ and $\boldsymbol{U}^{\star}$. The proof is deferred to Appendix 10.2.

Lemma 1. *Instate the assumptions of Theorem 3.1. With probability at least $1 - O\left(d^{-10}\right)$, one has*

$$\|\boldsymbol{G} - \boldsymbol{G}^{\star}\| \lesssim \underbrace{\frac{\mu r \sigma_1^{\star 2}\log d}{\sqrt{d_1 d_2}\,p} + \sqrt{\frac{\mu r \sigma_1^{\star 4}\log d}{d_2 p}} + \frac{\sigma^2\sqrt{d_1 d_2}\,\log d}{p} + \sigma\sigma_1^{\star}\sqrt{\frac{d_1 \log d}{p}}}_{=:\zeta_{\mathsf{op}}} + \|\boldsymbol{A}^{\star}\|_{2,\infty}^2. \tag{5}$$

In order to get a better sense of the term ζ_{op} appearing above, we make note of a straightforward yet useful fact, which reveals that ζ_{op} is much smaller than any nonzero eigenvalue of $\boldsymbol{G}^{\star}$.

Fact 1. *Instate the assumptions of Theorem 3.1. Then the quantity ζ_{op} as defined in (5) obeys*

$$\begin{aligned}\zeta_{\mathsf{op}} + \|\boldsymbol{A}^{\star}\|_{2,\infty}^2 &\le \frac{\mu r \sigma_1^{\star 2}\log d}{\sqrt{d_1 d_2}\,p} + \sqrt{\frac{\mu r \sigma_1^{\star 4}\log d}{d_2 p}} + \frac{\sigma^2}{\sigma_r^{\star 2}}\frac{\sqrt{d_1 d_2}\,\log d}{p} + \sigma\sigma_1^{\star}\sqrt{\frac{d_1 \log d}{p}} + \frac{\mu_1 r \sigma_1^{\star 2}}{d_1} \\ &\ll \frac{\sigma_r^{\star 2}}{\kappa^2},\end{aligned}$$

where $\|\boldsymbol{A}^{\star}\|_{2,\infty}^2 \le \frac{\mu_1 r \sigma_1^{\star 2}}{d_1}$ (cf. Lemma 11).

Further, the following lemma upper bounds the difference between $\boldsymbol{G}$ and $\boldsymbol{G}^{\star}$ in the m-th row, when projected onto the subspace represented by $\boldsymbol{U}^{\star}$; the proof is postponed to Appendix 10.3. This result gives a more refined control of the difference between $\boldsymbol{G}$ and $\boldsymbol{G}^{\star}$.

Lemma 2. *Instate the assumptions of Theorem 3.1. With probability at least $1 - O\left(d^{-10}\right)$, the following holds simultaneously for all $1 \le m \le d_1$:*

$$\left\| \left(\boldsymbol{G} - \boldsymbol{G}^{\star}\right)_{m,:} \boldsymbol{U}^{\star} \right\|_2 \lesssim \left(\zeta_{\mathsf{op}} + \|\boldsymbol{A}^{\star}\|_{2,\infty}^2\right)\sqrt{\frac{\mu r}{d_1}},$$

where ζ_{op} is defined in (5).

The next step, which is also the most challenging and crucial step, lies in showing that: every row of $\boldsymbol{U}$, under certain global linear transformation, serves as a good approximation of the corresponding row of $\boldsymbol{U}^{\star}$. Towards this end, we begin with the following preparations:

- We first introduce the following matrix $\boldsymbol{H}$ to represent the linear transformation we have in mind:

$$\boldsymbol{H} := \boldsymbol{U}^{\top}\boldsymbol{U}^{\star}. \tag{6}$$

  While this is not a rotation matrix, it is quite close to the rotation matrix $\boldsymbol{R}$ defined in (3.4).

- In addition, we find it convenient to express

$$\boldsymbol{U}^{\star} = \boldsymbol{G}^{\star}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}.$$

  Combining this with Lemma 2, one would expect $\boldsymbol{U}^{\star}$ and $\boldsymbol{G}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}$ to be reasonably close, namely,

$$\boldsymbol{U}^{\star} \approx \boldsymbol{G}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}. \tag{7}$$

  With these in hand, the following lemma (together with Lemma 2) asserts that

$$\boldsymbol{U}\boldsymbol{H} \approx \boldsymbol{G}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2} \approx \boldsymbol{U}^{\star}$$

in an $\ell_{2,\infty}$ sense.

Lemma 3. *Instate the assumptions of Theorem 3.1, and recall the definition of ζ_{op} in (5). With probability at least $1 - O(d^{-10})$, one has*

$$\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{G}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\right\|_{2,\infty} \lesssim \frac{\kappa^{2}\left(\zeta_{\mathsf{op}} + \|\boldsymbol{A}^{\star}\|_{2,\infty}^{2}\right)}{\sigma_{r}^{\star 2}}\left(\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty} + \sqrt{\frac{\mu r}{d_{1}}}\right).$$

The proof of this lemma, however, goes far beyond conventional matrix perturbation theory, and requires delicate decoupling of statistical dependencies. This is accomplished via leave-one-out and leave-two-out analysis arguments. In what follows, we take a moment to explain the high-level idea.

To establish Lemma 3, we first learn from standard matrix perturbation theory [AFWZ17, Lemma 1] that: for each $1 \leq m \leq d_1$,

$$\left\|\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{G}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\right)_{m,:}\right\|_{2} \lesssim \frac{1}{\lambda_{r}\left(\boldsymbol{G}^{\star}\right)^{2}}\left\|\boldsymbol{G} - \boldsymbol{G}^{\star}\right\|\left\|\boldsymbol{G}_{m,:}\boldsymbol{U}^{\star}\right\|_{2} + \frac{1}{\lambda_{r}\left(\boldsymbol{G}^{\star}\right)}\left\|\boldsymbol{G}_{m,:}\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\star}\right)\right\|_{2} \tag{8}$$

holds, provided that $\boldsymbol{G}$ and $\boldsymbol{G}^{\star}$ are sufficiently close.

- The first term on the right-hand side of (8) can already be controlled by Lemma 1 and Lemma 2.

- The second term on the right-hand side of (8), however, is considerably more difficult to analyze, due to the complicated statistical dependence between $\boldsymbol{G}_{m,:}$ and $\boldsymbol{U}\boldsymbol{H}$. In order to decouple statistical dependency, we resort to the leave-one-out sequence $\boldsymbol{U}^{(m)}$ introduced in Algorithm 1 and use the triangle inequality to bound

$$\left\|\boldsymbol{G}_{m,:}\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\star}\right)\right\|_{2} \leq \left\|\boldsymbol{G}_{m,:}\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right)\right\|_{2} + \left\|\boldsymbol{G}_{m,:}\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^{\star}\right)\right\|_{2}, \tag{9}$$

  where $\boldsymbol{H}^{(m)} := \boldsymbol{U}^{(m)\top}\boldsymbol{U}^{\star}$. As mentioned before, the leave-one-out estimate $\boldsymbol{U}^{(m)}$ enjoys two nice properties.

  (1) The true estimate $\boldsymbol{U}$ and the leave-one-out estimate $\boldsymbol{U}^{(m)}$ are exceedingly close, as asserted by the following lemma (to be established in Appendix 10.5).

  Lemma 4. *Instate the assumptions of Theorem 3.1, and recall the definition of $\boldsymbol{H}$ in (6). With probability at least $1 - O\left(d^{-10}\right)$, the following holds simultaneously for all $1 \leq m \leq d_1$:*

$$\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top} - \boldsymbol{U}\boldsymbol{U}^{\top}\right\|_{\mathrm{F}} \lesssim \frac{\zeta_{\mathsf{op}}}{\sigma_{r}^{\star 2}}\left(\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty} + \sqrt{\frac{\mu r}{d_{1}}}\right),$$

  where ζ_{op} is defined in (5).

This result in turn allows us to control the first term on the right-hand side of (9).

(2) Due to the statistical independence between $\boldsymbol{A}_{m,:}$ and $\boldsymbol{U}^{(m)}$, the matrices $\boldsymbol{G}_{m,:}$ and $\boldsymbol{U}^{(m)}$ turn out to be nearly independent. This allows one to invoke simple concentration inequalities to develop tight bounds for the second term on the right-hand side of (9). The detailed proof can be found in Appendix 10.4.

Finally, we make a remark on a technical issue encountered in the proof of Lemma 4. Recall that $\boldsymbol{U}^{(m)}$ is obtained by simply replacing the m-th row of $\boldsymbol{A}$ with its population version, which indicates the statistical dependency between $\boldsymbol{U}^{(m)}$ and the m-th row of $\boldsymbol{A}$. However, there is still some delicate statistical dependency between $\boldsymbol{U}^{(m)}$ and the columns of $\boldsymbol{A}$ that need to be carefully coped with. Fortunately, the leave-two-out estimate $\boldsymbol{U}^{(m,l)}$ — which is obtained by dropping not only the m-th row of $\boldsymbol{A}$ but also the l-th of its columns — allows us to decouple the dependency between $\boldsymbol{U}^{(m,l)}$ (and hence $\boldsymbol{U}$ and $\boldsymbol{U}^{(m)}$) and the l-th column of $\boldsymbol{A}$. This is precisely the main reason why we introduce additional leave-two-out estimates.

8.3 Proof of Theorem 3.1

We are now positioned to establish our main theorem. The proof is split into two parts.

8.3.1 Statistical accuracy measured by $\|\cdot\|$

We begin by establishing the spectral norm bound (3.6c). Let λ_i and $\lambda_i^\star$ be the i-th largest eigenvalue of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}^\star$, respectively. From Lemma 1 and Weyl's inequality, one finds that

$$\max_{1\le i\le r} |\lambda_i - \lambda_i^\star| = \|\boldsymbol{\Lambda} - \boldsymbol{\Lambda}^\star\| \le \|\boldsymbol{G} - \boldsymbol{G}^\star\| \lesssim \zeta_{\mathsf{op}} + \|\boldsymbol{A}^\star\|_{2,\infty}^2 \le \sigma_r^{\star 2} \cdot \mathcal{E}_{\mathsf{general}}, \tag{10}$$

where ζ_{op} and $\mathcal{E}_{\mathsf{general}}$ are defined in (5) and (3.7), respectively. Here, the last inequality arises from the simple fact that $\|\boldsymbol{A}^\star\|_{2,\infty}^2 \le \frac{\mu_1 r \sigma_1^{\star 2}}{d_1}$ (cf. Lemma 11). By virtue of Fact 1, we know that $\|\boldsymbol{\Lambda} - \boldsymbol{\Lambda}^\star\| \ll \sigma_r^{\star 2}$. Given that $\boldsymbol{\Lambda}^\star = \boldsymbol{\Sigma}^{\star 2}$ and $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^2$, this implies that for each $1 \le i \le r$,

$$\frac{1}{4}\sigma_i^{\star 2} = \frac{1}{4}\lambda_i^\star \le \lambda_i^\star - |\lambda_i - \lambda_i^\star| \le \lambda_i \le \lambda_i^\star + |\lambda_i - \lambda_i^\star| \le 4\lambda_i^\star = 4\sigma_i^{\star 2},$$

thus indicating that

$$\frac{1}{2}\sigma_i^\star \le \sigma_i \le 2\sigma_i^\star. \tag{11}$$

In conclusion,

$$\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^\star\| = \max_{1\le i\le r} |\sigma_i - \sigma_i^\star| = \max_{1\le i\le r} \frac{\left|\sigma_i^2 - \sigma_i^{\star 2}\right|}{\sigma_i + \sigma_i^\star} \overset{\text{(a)}}{\le} \max_{1\le i\le r} \frac{\|\boldsymbol{\Lambda} - \boldsymbol{\Lambda}^\star\|}{\frac{3}{2}\sigma_i^\star} \overset{\text{(b)}}{\lesssim} \sigma_r^\star \cdot \mathcal{E}_{\mathsf{general}}$$

as claimed. Here, (a) comes from (11), whereas (b) follows from (10).

Next, we turn attention to (3.6a). First, it is seen that

$$\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\| \le \sqrt{2}\left\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right\| = \|\sin\boldsymbol{\Theta}(\boldsymbol{U}, \boldsymbol{U}^\star)\| \tag{12}$$

where $\boldsymbol{R}$ is defined in (3.4), and $\boldsymbol{\Theta}(\boldsymbol{U}, \boldsymbol{U}^\star)$ denotes a diagonal matrix whose i-th diagonal entry is the i-th principal angle between the two subspaces represented by $\boldsymbol{U}$ and $\boldsymbol{U}^\star$. Here, the first inequality follows from a well-known inequality connecting two different subspace distance metrics [SS90, Chapter II], while the last identity follows from [SS90, Chapter II]. In addition, Lemma 1 and Fact 1 tell us that

$$\|\boldsymbol{G} - \boldsymbol{G}^\star\| \lesssim \zeta_{\mathsf{op}} + \|\boldsymbol{A}^\star\|_{2,\infty}^2 \ll \sigma_r^{\star 2} \tag{13}$$

with probability at least $1 - O(d^{-10})$, which together with Weyl's inequality gives

$$\lambda_{r+1}(\boldsymbol{G}) \le \lambda_{r+1}(\boldsymbol{G}^\star) + \|\boldsymbol{G} - \boldsymbol{G}^\star\| = \|\boldsymbol{G} - \boldsymbol{G}^\star\| \le \sigma_r^{\star 2}/2. \tag{14}$$

Therefore, [SS90, Chapter V, Theorem 3.6] (which is a version of the celebrated Davis-Kahan $\sin\boldsymbol{\Theta}$ Theorem [DK70]) reveals that

$$\|\sin\boldsymbol{\Theta}(\boldsymbol{U},\boldsymbol{U}^{\star})\| \leq \frac{\|\boldsymbol{G}-\boldsymbol{G}^{\star}\|}{\lambda_{r}\left(\boldsymbol{G}^{\star}\right)-\lambda_{r+1}\left(\boldsymbol{G}\right)} \leq \frac{\|\boldsymbol{G}-\boldsymbol{G}^{\star}\|}{\sigma_{r}^{\star2}-\sigma_{r}^{\star2}/2} = \frac{2\left\|\boldsymbol{G}-\boldsymbol{G}^{\star}\right\|}{\sigma_{r}^{\star2}},$$

where we have used the fact $\lambda_r(\boldsymbol{G}^{\star}) = \sigma_r^{\star2}$. The above bounds taken collectively imply that, with probability at least $1-O(d^{-10})$,

$$\|\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}\| \leq \frac{2\sqrt{2}\,\|\boldsymbol{G}-\boldsymbol{G}^{\star}\|}{\sigma_{r}^{\star2}} \lesssim \mathcal{E}_{\mathsf{general}}. \tag{15}$$

8.3.2 Statistical accuracy measured by $\|\cdot\|_{2,\infty}$

Before continuing to the proof, we find it convenient to introduce a few more notations. In addition to the rotation matrix $\boldsymbol{R}$ defined in (3.4) and the linear transformation $\boldsymbol{H}$ defined in (6), we define

$$\mathsf{sgn}\left(\boldsymbol{H}\right) := \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{V}}^{\top} \in \mathbb{R}^{d_1\times d_1}, \tag{16}$$

where the columns of $\widetilde{\boldsymbol{U}} \in \mathbb{R}^{d_1\times d_1}$ (resp. $\widetilde{\boldsymbol{V}} \in \mathbb{R}^{d_1\times d_1}$) are the left (resp. right) singular vectors of $\boldsymbol{H}$. It is well-known that [TB77, Theorem 2]

$$\boldsymbol{R} = \mathsf{sgn}\left(\boldsymbol{H}\right). \tag{17}$$

We now move on to establishing the advertised bound (3.6b).

1. To begin with, we claim that $\boldsymbol{U}\boldsymbol{H}$ is extremely close to $\boldsymbol{U}\boldsymbol{R}$, provided that $\|\boldsymbol{G}-\boldsymbol{G}^{\star}\|$ is sufficiently small. To this end, recognizing that $\|\boldsymbol{G}-\boldsymbol{G}^{\star}\| \lesssim \zeta_{\mathsf{op}} \ll \sigma_r^{\star2}$ (according to Lemma 1 and Fact 1), we can apply [AFWZ17, Lemma 3] to show that

$$\left\|\boldsymbol{H}^{-1}\right\| \lesssim 1 \qquad \text{and} \qquad \sqrt{\|\boldsymbol{H}-\mathsf{sgn}\left(\boldsymbol{H}\right)\|} \leq \left\|\boldsymbol{U}\boldsymbol{U}^{\top}-\boldsymbol{U}^{\star}\boldsymbol{U}^{\star\top}\right\| \lesssim \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\star2}},$$

where the last inequality follows from (15). Thus, invoke the identity (17) to arrive at

$$\begin{aligned}
\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}\boldsymbol{R}\|_{2,\infty} &= \|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}\mathsf{sgn}\left(\boldsymbol{H}\right)\|_{2,\infty} = \left\|\boldsymbol{U}\boldsymbol{H}\boldsymbol{H}^{-1}\big(\boldsymbol{H}-\mathsf{sgn}\left(\boldsymbol{H}\right)\big)\right\|_{2,\infty} \\
&\leq \|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty}\left\|\boldsymbol{H}^{-1}\right\|\|\boldsymbol{H}-\mathsf{sgn}\left(\boldsymbol{H}\right)\| \\
&\lesssim \left(\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\star2}}\right)^2\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty} \ll \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\star2}}\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty}.
\end{aligned} \tag{18}$$

This in turn allows us to focus attention on bounding $\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\star}\|_{2,\infty}$ (instead of $\|\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}\|_{2,\infty}$).

2. Next, recall that $\boldsymbol{G}^{\star} = \boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star2}\boldsymbol{U}^{\star\top}$ and hence $\boldsymbol{G}^{\star}\boldsymbol{U}^{\star}(\boldsymbol{\Sigma}^{\star})^{-2} = \boldsymbol{U}^{\star}$. Invoke the triangle inequality to reach

$$\begin{aligned}
\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\star}\|_{2,\infty} &= \left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{G}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}+\boldsymbol{G}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}-\boldsymbol{G}^{\star}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\right\|_{2,\infty} \\
&\leq \left\|\left(\boldsymbol{G}-\boldsymbol{G}^{\star}\right)\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\right\|_{2,\infty} + \left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{G}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\right\|_{2,\infty} \\
&\leq \|\left(\boldsymbol{G}-\boldsymbol{G}^{\star}\right)\boldsymbol{U}^{\star}\|_{2,\infty}\left\|\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\right\| + \left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{G}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\right\|_{2,\infty} \\
&\leq \frac{1}{\sigma_r^{\star2}}\|\left(\boldsymbol{G}-\boldsymbol{G}^{\star}\right)\boldsymbol{U}^{\star}\|_{2,\infty} + \left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{G}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\right\|_{2,\infty}.
\end{aligned} \tag{19}$$

Regarding the first term of (19), Lemma 2 reveals that with probability at least $1-O\left(d^{-10}\right)$,

$$\frac{1}{\sigma_r^{\star2}}\|\left(\boldsymbol{G}-\boldsymbol{G}^{\star}\right)\boldsymbol{U}^{\star}\|_{2,\infty} \lesssim \frac{\zeta_{\mathsf{op}}+\|\boldsymbol{A}^{\star}\|_{2,\infty}^2}{\sigma_r^{\star2}}\sqrt{\frac{\mu r}{d_1}}. \tag{20}$$

With regards to the second term of (19), Lemma 3 demonstrates that

$$\left\| \boldsymbol{U}\boldsymbol{H} - \boldsymbol{G}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2} \right\|_{2,\infty} \lesssim \frac{\kappa^2\left(\zeta_{\mathsf{op}} + \|\boldsymbol{A}^{\star}\|_{2,\infty}^2\right)}{\sigma_r^{\star 2}} \left(\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty} + \sqrt{\frac{\mu r}{d_1}}\right) \tag{21}$$

with probability at least $1 - O\left(d^{-10}\right)$. Combine (20) and (21) to arrive at

$$\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\star}\|_{2,\infty} \lesssim \frac{\kappa^2\left(\zeta_{\mathsf{op}} + \|\boldsymbol{A}^{\star}\|_{2,\infty}^2\right)}{\sigma_r^{\star 2}} \left(\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty} + \sqrt{\frac{\mu r}{d_1}}\right). \tag{22}$$

3. As a byproduct of (22) and Fact 1, we see that

$$\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\star}\|_{2,\infty} \ll \|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty} + \sqrt{\frac{\mu r}{d_1}}.$$

It then follows from the triangle inequality that

$$\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty} \le \|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\star}\|_{2,\infty} + \|\boldsymbol{U}^{\star}\|_{2,\infty} \lesssim o\left(1\right)\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty} + \sqrt{\frac{\mu r}{d_1}},$$

thus indicating that

$$\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty} \le 2\sqrt{\frac{\mu r}{d_1}}. \tag{23}$$

Substitution into (18) and (22) gives

$$\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}\boldsymbol{R}\|_{2,\infty} \ll \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\star 2}}\sqrt{\frac{\mu r}{d_1}} \quad \text{and} \quad \|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\star}\|_{2,\infty} \lesssim \frac{\kappa^2\left(\zeta_{\mathsf{op}} + \|\boldsymbol{A}^{\star}\|_{2,\infty}^2\right)}{\sigma_r^{\star 2}}\sqrt{\frac{\mu r}{d_1}}.$$

Combining the above results yields

$$\begin{aligned}\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star}\|_{2,\infty} &\le \|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\star}\|_{2,\infty} + \|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}\boldsymbol{R}\|_{2,\infty} \\ &\lesssim \frac{\kappa^2\left(\zeta_{\mathsf{op}} + \|\boldsymbol{A}^{\star}\|_{2,\infty}^2\right)}{\sigma_r^{\star 2}}\sqrt{\frac{\mu r}{d_1}}.\end{aligned}$$

Substituting the value of ζ_{op} into the above inequality and using the upper bound $\|\boldsymbol{A}^{\star}\|_{2,\infty}^2 \le \frac{\mu_1 r \sigma_1^{\star 2}}{d_1}$ (cf. Lemma 11), we conclude the proof.

9 Proofs for corollaries

9.1 Proof of Corollary 4.2

Recall that the spectral algorithm considered herein (cf. Section 4.1) operates upon the noisy copy of the mode-1 matricization of the tensor $\boldsymbol{T}^{\star}$, namely,

$$\boldsymbol{A}^{\star} = \sum_{s=1}^{r} \boldsymbol{w}_s^{\star}\left(\boldsymbol{w}_s^{\star} \otimes \boldsymbol{w}_s^{\star}\right)^{\top}. \tag{24}$$

Consequently, in order to apply Theorem 3.1, the main step boils down to estimating the spectrum and the incoherence parameters of $\boldsymbol{A}^{\star}$. Specifically, we need to upper bound the condition number κ, as well as the incoherence parameters μ_0, μ_1 and μ_2 as introduced in Definition 2.1.

Before proceeding, we introduce a few notations that simplify the presentation. Define

$$\lambda_i^{\star} := \|\boldsymbol{w}_i^{\star}\|_2^3, \qquad 1 \le i \le r,$$

and let $\lambda^\star_{(i)}$ denote the i-th largest value in $\{\lambda^\star_i\}_{i=1}^r$. We also recall that

$$\lambda^\star_{\min} := \min_{1\le i\le r} \|\boldsymbol{w}^\star_i\|_2^3 \qquad \text{and} \qquad \lambda^\star_{\max} := \max_{1\le i\le r} \|\boldsymbol{w}^\star_i\|_2^3 .$$

In addition, we define two matrices of interest

$$\overline{\boldsymbol{W}}^\star := [\overline{\boldsymbol{w}}^\star_1, \cdots, \overline{\boldsymbol{w}}^\star_r] \in \mathbb{R}^{d\times r}, \qquad \widetilde{\boldsymbol{W}}^\star := [\overline{\boldsymbol{w}}^\star_1 \otimes \overline{\boldsymbol{w}}^\star_1, \cdots, \overline{\boldsymbol{w}}^\star_r \otimes \overline{\boldsymbol{w}}^\star_r] \in \mathbb{R}^{d^2\times r},$$

where $\overline{\boldsymbol{w}}^\star_s := \boldsymbol{w}^\star_s / \|\boldsymbol{w}^\star_s\|_2$, and $\boldsymbol{a}\otimes\boldsymbol{b} := \begin{bmatrix} a_1\boldsymbol{b} \\ \vdots \\ a_d\boldsymbol{b}\end{bmatrix}$. In addition, let $\boldsymbol{D}^\star \in \mathbb{R}^{r\times r}$ be a diagonal matrix with diagonal entries

$$D^\star_{s,s} = \|\boldsymbol{w}^\star_s\|_2 , \qquad 1\le s\le r.$$

These allow us to express

$$\boldsymbol{G}^\star = \boldsymbol{A}^\star\boldsymbol{A}^{\star\top} = \overline{\boldsymbol{W}}^\star \boldsymbol{D}^{\star 3}\widetilde{\boldsymbol{W}}^{\star\top}\widetilde{\boldsymbol{W}}^\star \boldsymbol{D}^{\star 3}\overline{\boldsymbol{W}}^{\star\top}.$$

In the sequel, we begin by quantifying the spectrum of $\boldsymbol{G}^\star$, which in turn allows us to understand the spectrum of $\boldsymbol{A}^\star$.

- We first look at the eigenvalues of the matrices $\overline{\boldsymbol{W}}^{\star\top}\overline{\boldsymbol{W}}^\star$ and $\widetilde{\boldsymbol{W}}^{\star\top}\widetilde{\boldsymbol{W}}^\star$. Towards this, let us write

$$\overline{\boldsymbol{W}}^{\star\top}\overline{\boldsymbol{W}}^\star = \boldsymbol{I}_r + \boldsymbol{C}, \qquad \text{and} \qquad \widetilde{\boldsymbol{W}}^{\star\top}\widetilde{\boldsymbol{W}}^\star = \boldsymbol{I}_r + \widetilde{\boldsymbol{C}} \tag{25}$$

for some matrices $\boldsymbol{C}, \widetilde{\boldsymbol{C}} \in \mathbb{R}^{r\times r}$. It follows immediately from the incoherence assumption (4.4) that

$$\|\boldsymbol{C}\|_\infty \le \sqrt{\mu_5/d} \qquad \text{and} \qquad \|\widetilde{\boldsymbol{C}}\|_\infty \le \mu_5/d,$$

thus leading to the simple bounds

$$\|\boldsymbol{C}\| \le r\|\boldsymbol{C}\|_\infty \le r\sqrt{\mu_5/d}, \qquad \|\widetilde{\boldsymbol{C}}\| \le r\|\widetilde{\boldsymbol{C}}\|_\infty \le \mu_5 r/d. \tag{26}$$

These taken collectively with (25) and Weyl's inequality yield

$$\max_{i\in[r]}\left|\lambda_i\big(\overline{\boldsymbol{W}}^{\star\top}\overline{\boldsymbol{W}}^\star\big) - 1\right| \le \|\boldsymbol{C}\| \le r\sqrt{\mu_5/d} \qquad \text{and} \qquad \max_{i\in[r]}\left|\lambda_i\big(\widetilde{\boldsymbol{W}}^{\star\top}\widetilde{\boldsymbol{W}}^\star\big) - 1\right| \le \|\widetilde{\boldsymbol{C}}\| \le \mu_5 r/d,$$

which essentially tell us that

$$\|\overline{\boldsymbol{W}}^\star\| = \sqrt{\lambda_1\big(\overline{\boldsymbol{W}}^{\star\top}\overline{\boldsymbol{W}}^\star\big)} \le \sqrt{1 + r\sqrt{\mu_5/d}} \qquad \text{and} \qquad \|\widetilde{\boldsymbol{W}}^\star\| = \sqrt{\lambda_1\big(\widetilde{\boldsymbol{W}}^{\star\top}\widetilde{\boldsymbol{W}}^\star\big)} \le \sqrt{1 + \mu_5 r/d}. \tag{27}$$

- Returning to $\boldsymbol{G}^\star$, one invokes the definition (25) to deduce that

$$\boldsymbol{G}^\star = \overline{\boldsymbol{W}}^\star\boldsymbol{D}^{\star 6}\overline{\boldsymbol{W}}^{\star\top} + \overline{\boldsymbol{W}}^\star\boldsymbol{D}^{\star 3}\widetilde{\boldsymbol{C}}\boldsymbol{D}^{\star 3}\overline{\boldsymbol{W}}^{\star\top}.$$

Observe that the eigenvalues of $\overline{\boldsymbol{W}}^\star\boldsymbol{D}^{\star 3}\big(\overline{\boldsymbol{W}}^\star\boldsymbol{D}^{\star 3}\big)^\top$ are identical to those of $\big(\overline{\boldsymbol{W}}^\star\boldsymbol{D}^{\star 3}\big)^\top\overline{\boldsymbol{W}}^\star\boldsymbol{D}^{\star 3}$, where the latter can be further decomposed as follows (in view of (25))

$$\big(\overline{\boldsymbol{W}}^\star\boldsymbol{D}^{\star 3}\big)^\top\overline{\boldsymbol{W}}^\star\boldsymbol{D}^{\star 3} = \boldsymbol{D}^{\star 3}\overline{\boldsymbol{W}}^{\star\top}\overline{\boldsymbol{W}}^\star\boldsymbol{D}^{\star 3} = \boldsymbol{D}^{\star 6} + \boldsymbol{D}^{\star 3}\boldsymbol{C}\boldsymbol{D}^{\star 3}.$$

This taken together with Weyl's inequality, (26) and (27) shows that

$$\left|\lambda_i\big(\overline{\boldsymbol{W}}^\star\boldsymbol{D}^{\star 6}\overline{\boldsymbol{W}}^{\star\top}\big) - \lambda_i\big(\boldsymbol{D}^{\star 6}\big)\right| \le \|\boldsymbol{D}^{\star 3}\boldsymbol{C}\boldsymbol{D}^{\star 3}\| \le \|\boldsymbol{D}^\star\|^6\|\boldsymbol{C}\| \le r\sqrt{\frac{\mu_5}{d}}\,\lambda^{\star 2}_{\max}$$

for each $1\le i\le r$. In addition,

$$\left|\lambda_i(\boldsymbol{G}^\star) - \lambda_i\big(\overline{\boldsymbol{W}}^\star\boldsymbol{D}^{\star 6}\overline{\boldsymbol{W}}^{\star\top}\big)\right| \le \|\overline{\boldsymbol{W}}^\star\boldsymbol{D}^{\star 3}\widetilde{\boldsymbol{C}}\boldsymbol{D}^{\star 3}\overline{\boldsymbol{W}}^{\star\top}\| \le \|\overline{\boldsymbol{W}}^\star\|^2\|\boldsymbol{D}^\star\|^6\|\widetilde{\boldsymbol{C}}\| \le \frac{\mu_5 r}{d}\left(1 + r\sqrt{\frac{\mu_5}{d}}\right)\lambda^{\star 2}_{\max}.$$

As a result, invoke the triangle inequality to see that

$$\begin{aligned}\left|\lambda_i\left(\boldsymbol{G}^\star\right)-\lambda_i\left(\boldsymbol{D}^{\star 6}\right)\right| &\le \left|\lambda_i(\boldsymbol{G}^\star)-\lambda_i(\overline{\boldsymbol{W}}^\star\boldsymbol{D}^{\star 6}\overline{\boldsymbol{W}}^{\star\top})\right|+\left|\lambda_i(\overline{\boldsymbol{W}}^\star\boldsymbol{D}^{\star 6}\overline{\boldsymbol{W}}^{\star\top})-\lambda_i(\boldsymbol{D}^{\star 6})\right| \\ &\le \frac{\mu_5 r}{d}\left(1+r\sqrt{\frac{\mu_5}{d}}\right)\lambda_{\max}^{\star 2}+r\sqrt{\frac{\mu_5}{d}}\,\lambda_{\max}^{\star 2}\le 3r\sqrt{\frac{\mu_5}{d}}\,\lambda_{\max}^{\star 2}\end{aligned}$$

for each $1\le i\le r$, where the last inequality holds under the assumption that $r\sqrt{\mu_5/d}\le 1$. This means

$$\left|\lambda_i\left(\boldsymbol{G}^\star\right)-\lambda_{(i)}^{\star 2}\right|\le 3r\sqrt{\frac{\mu_5}{d}}\,\lambda_{\max}^{\star 2},$$

where $\lambda_{(i)}^\star$ denotes the i-th largest value in $\{\lambda_i^\star\}_{i=1}^r$.

- Recalling that $\mu_{\mathsf{tc}}:=\max\left\{\mu_3,\mu_4^2\right\}$, $\kappa_{\mathsf{tc}}:=\lambda_{\max}^\star/\lambda_{\min}^\star$ and the rank assumption $r\ll\kappa_{\mathsf{tc}}^{-2}\sqrt{d/\mu_5}$, we find that

$$\lambda_i\left(\boldsymbol{G}^\star\right)=\lambda_{(i)}^{\star 2}+O\left(r\sqrt{\frac{\mu_5}{d}}\right)\lambda_{\max}^{\star 2}\qquad\text{and}\qquad\sigma_i\left(\boldsymbol{A}^\star\right)=\lambda_{(i)}^\star\left(1+o\left(1\right)\right).\tag{28}$$

As a result, we immediately arrive at

$$\sigma_1\left(\boldsymbol{A}^\star\right)=\lambda_{\max}^\star\left(1+o\left(1\right)\right),\qquad\sigma_r\left(\boldsymbol{A}^\star\right)=\lambda_{\min}^\star\left(1+o\left(1\right)\right),\qquad\text{and}\qquad\kappa=\frac{\sigma_1\left(\boldsymbol{A}^\star\right)}{\sigma_r\left(\boldsymbol{A}^\star\right)}\lesssim\kappa_{\mathsf{tc}}.$$

Next, we turn attention to bounding the incoherence parameters of $\boldsymbol{A}^\star$. Let $\boldsymbol{A}^\star=\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\boldsymbol{V}^{\star\top}$ be the (compact) SVD of $\boldsymbol{A}^\star$. It is seen from (24) that the column space of $\boldsymbol{U}^\star$ (resp. $\boldsymbol{V}^\star$) coincides with the column space of $\overline{\boldsymbol{W}}^\star$ (resp. $\widetilde{\boldsymbol{W}}^\star$). Therefore, there exist orthonormal matrices $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ such that

$$\boldsymbol{U}^\star\boldsymbol{H}_1=\overline{\boldsymbol{W}}^\star\big(\overline{\boldsymbol{W}}^{\star\top}\overline{\boldsymbol{W}}^\star\big)^{-1/2}\qquad\text{and}\qquad\boldsymbol{V}^\star\boldsymbol{H}_2=\widetilde{\boldsymbol{W}}^\star\big(\widetilde{\boldsymbol{W}}^{\star\top}\widetilde{\boldsymbol{W}}^\star\big)^{-1/2}.$$

These allow us to bound

$$\|\boldsymbol{U}^\star\|_{2,\infty}=\|\boldsymbol{U}^\star\boldsymbol{H}_1\|_{2,\infty}\le\|\overline{\boldsymbol{W}}^\star\|_{2,\infty}\big\|\big(\overline{\boldsymbol{W}}^{\star\top}\overline{\boldsymbol{W}}^\star\big)^{-1/2}\big\|\le\sqrt{\frac{\mu_4 r}{d}}\sqrt{\frac{1}{\lambda_r\big(\overline{\boldsymbol{W}}^{\star\top}\overline{\boldsymbol{W}}^\star\big)}}\lesssim\sqrt{\frac{\mu_4 r}{d}}\sqrt{\frac{1}{1-1/3}}\le\sqrt{\frac{2\mu_4 r}{d}},$$

$$\|\boldsymbol{V}^\star\|_{2,\infty}=\|\boldsymbol{V}^\star\boldsymbol{H}_2\|_{2,\infty}\le\|\widetilde{\boldsymbol{W}}^\star\|_{2,\infty}\big\|\big(\widetilde{\boldsymbol{W}}^{\star\top}\widetilde{\boldsymbol{W}}^\star\big)^{-1/2}\big\|\le\sqrt{\frac{\mu_4^2 r}{d^2}}\sqrt{\frac{1}{\lambda_r\big(\widetilde{\boldsymbol{W}}^{\star\top}\widetilde{\boldsymbol{W}}^\star\big)}}\le\sqrt{\frac{\mu_4^2 r}{d^2}}\sqrt{\frac{1}{1-1/3}}\le\sqrt{\frac{2\mu_4^2 r}{d^2}},$$

which follow from (27) and the assumption that $r\ll\sqrt{d/\mu_5}$. Moreover, the incoherence assumption (4.4) gives that

$$\mu_0=\frac{d^3\left\|\boldsymbol{A}^\star\right\|_\infty^2}{\left\|\boldsymbol{A}^\star\right\|_{\mathrm{F}}^2}=\frac{d^3\left\|\boldsymbol{T}^\star\right\|_\infty^2}{\left\|\boldsymbol{T}^\star\right\|_{\mathrm{F}}^2}\le\mu_3.$$

To conclude, the above analysis reveals that

$$\mu_0\le\mu_3,\qquad\mu_1\lesssim\mu_4,\qquad\mu_2\lesssim\mu_4^2,\qquad\mu\lesssim\max\left\{\mu_3,\mu_4^2\right\}=\mu_{\mathsf{tc}}\qquad\text{and}\qquad\kappa\lesssim\kappa_{\mathsf{tc}},$$

where $\mu=\max\left\{\mu_0,\mu_1,\mu_2\right\}$ and $\kappa=\sigma_1\left(\boldsymbol{A}^\star\right)/\sigma_r\left(\boldsymbol{A}^\star\right)$. With these estimates in place, Corollary 4.2 follows immediately from Theorem 3.1.

9.2 Proof of Corollary 4.3

In the problem of covariance estimation with missing data, the ground truth $\boldsymbol{A}^\star$ is effectively given by $\boldsymbol{B}^\star\boldsymbol{F}^\star$, which obeys

$$\boldsymbol{A}^\star=\boldsymbol{B}^\star\boldsymbol{F}^\star=\boldsymbol{U}^\star\boldsymbol{\Lambda}^{\star 1/2}\boldsymbol{F}^\star\in\mathbb{R}^{d\times n},\qquad\boldsymbol{F}^\star=[\boldsymbol{f}_1^\star,\cdots,\boldsymbol{f}_n^\star]\in\mathbb{R}^{r\times n}$$

with $\boldsymbol{f}_i^\star\overset{\text{i.i.d.}}{\sim}\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_r)$.We note that by our assumption on the sample size, one has $n\gg\kappa_{\mathsf{ce}}^2\left(r+\log d\right)$, where $\kappa_{\mathsf{ce}}=\lambda_1^\star/\lambda_r^\star$. In addition, we note that under the assumption of Corollary 4.3, one has

$$\mathcal{E}_{\mathsf{ce}}\ll\kappa_{\mathsf{ce}}^{-1}\le 1,\tag{29}$$

where $\mathcal{E}_{\mathsf{ce}}$ and κ_{ce} are defined in (4.16) and (4.13), respectively.

9.2.1 Estimation error of the principle subspace

In this section, we will prove (4.15a) and (4.15b). To begin the proof, we verify the condition of the random noise (cf. (2.6)). From standard Gaussian concentration results, one is allowed to choose $R \asymp \sigma\sqrt{\log(n+d)}$, so that $|\eta_{i,j}| \leq R$ for all i and j with probability $1-O((n+d)^{-12})$. Under our sample size condition that $n \gg \max\left\{\frac{\log^6(n+d)}{dp^2}, \frac{\log^3(n+d)}{p}\right\}$, the requirement (2.6) is satisfied, namely,

$$\frac{R^2}{\sigma^2} \asymp \log(n+d) \lesssim \min\left\{\frac{p\sqrt{dn}}{\log(n+d)}, \frac{pn}{\log(n+d)}\right\}.$$

Next, we turn to the properties of $\boldsymbol{B}^{\star}\boldsymbol{F}^{\star}$ and start by looking at its spectrum. Define

$$\boldsymbol{C} := \boldsymbol{F}^{\star}\boldsymbol{F}^{\star\top} - n\boldsymbol{I}_r,$$

which allows us to write

$$\boldsymbol{G}^{\star} = \boldsymbol{B}^{\star}\boldsymbol{F}^{\star}\left(\boldsymbol{B}^{\star}\boldsymbol{F}^{\star}\right)^{\top} = \boldsymbol{U}^{\star}\boldsymbol{\Lambda}^{\star 1/2}\boldsymbol{F}^{\star}\boldsymbol{F}^{\star\top}\boldsymbol{\Lambda}^{\star 1/2}\boldsymbol{U}^{\star\top} = n\underbrace{\boldsymbol{U}^{\star}\boldsymbol{\Lambda}^{\star}\boldsymbol{U}^{\star\top}}_{=\boldsymbol{S}^{\star}} + \underbrace{\boldsymbol{U}^{\star}\boldsymbol{\Lambda}^{\star 1/2}\boldsymbol{C}\boldsymbol{\Lambda}^{\star 1/2}\boldsymbol{U}^{\star\top}}_{=:\boldsymbol{\Delta}}. \tag{30}$$

Using standard results on Gaussian random matrices [Ver12], one obtains

$$\begin{aligned}
\|\boldsymbol{C}\| &\lesssim \max\left\{\sqrt{n}\left(\sqrt{r}+\sqrt{\log(n+d)}\right), r+\log(n+d)\right\} \asymp \sqrt{n}\left(\sqrt{r}+\sqrt{\log(n+d)}\right),\\
\left|\sigma_i(\boldsymbol{F}^{\star})-\sqrt{n}\right| &\lesssim \sqrt{r}+\sqrt{\log(n+d)}
\end{aligned} \tag{31}$$

with probability at least $1-O\big((n+d)^{-10}\big)$, provided that $n \gg r+\log(n+d)$. It then follows from Weyl's inequality that

$$\left|\lambda_i(\boldsymbol{G}^{\star})-\lambda_i(n\boldsymbol{S}^{\star})\right| = \left|\lambda_i(\boldsymbol{G}^{\star})-\lambda_i^{\star}n\right| \leq \left\|\boldsymbol{\Delta}\right\| \leq \|\boldsymbol{C}\|\,\|\boldsymbol{U}^{\star}\|^2\,\|\boldsymbol{\Lambda}^{\star}\| \lesssim \lambda_1^{\star}\sqrt{n}\left(\sqrt{r}+\sqrt{\log(n+d)}\right). \tag{32}$$

Under the sample size assumption $n \gg \kappa_{\mathsf{ce}}^2\left(r+\log(n+d)\right)$, we conclude that

$$\lambda_i(\boldsymbol{G}^{\star}) = \lambda_i^{\star}n\left(1+o(1)\right) \qquad \text{and} \qquad \sigma_i(\boldsymbol{B}^{\star}\boldsymbol{F}^{\star}) = \sqrt{\lambda_i^{\star}n}\left(1+o(1)\right), \tag{33}$$

and hence

$$\kappa(\boldsymbol{G}^{\star}) = \frac{\lambda_1(\boldsymbol{G}^{\star})}{\lambda_r(\boldsymbol{G}^{\star})} \asymp \kappa_{\mathsf{ce}} \qquad \text{and} \qquad \kappa(\boldsymbol{B}^{\star}\boldsymbol{F}^{\star}) = \frac{\sigma_1(\boldsymbol{B}^{\star}\boldsymbol{F}^{\star})}{\sigma_r(\boldsymbol{B}^{\star}\boldsymbol{F}^{\star})} \asymp \sqrt{\kappa_{\mathsf{ce}}}. \tag{34}$$

Further, we look at the entrywise infinity norm of $\boldsymbol{B}^{\star}\boldsymbol{F}^{\star}$. From standard Gaussian concentration inequalities,

$$\begin{aligned}
\|\boldsymbol{B}^{\star}\boldsymbol{F}^{\star}\|_{\infty} &= \max_{i,j}\left|\left\langle\left(\boldsymbol{U}^{\star}\boldsymbol{\Lambda}^{\star 1/2}\right)_{i,:}, \boldsymbol{f}_j^{\star}\right\rangle\right| \lesssim \left\|\boldsymbol{U}^{\star}\boldsymbol{\Lambda}^{\star 1/2}\right\|_{2,\infty}\sqrt{\log(n+d)}\\
&\leq \|\boldsymbol{U}^{\star}\|_{2,\infty}\left\|\boldsymbol{\Lambda}^{\star}\right\|^{1/2}\sqrt{\log(n+d)} \leq \sqrt{\frac{\lambda_1^{\star}\mu_{\mathsf{ce}}r\log(n+d)}{d}}
\end{aligned}$$

holds with probability at least $1-O\big((n+d)^{-10}\big)$. Meanwhile, one has

$$\|\boldsymbol{B}^{\star}\boldsymbol{F}^{\star}\|_{\mathrm{F}} \geq \|\boldsymbol{B}^{\star}\|_{\mathrm{F}}\,\sigma_r(\boldsymbol{F}^{\star}) = \left\|\boldsymbol{\Lambda}^{\star 1/2}\right\|_{\mathrm{F}}\sigma_r(\boldsymbol{F}^{\star}) \gtrsim \sqrt{\lambda_r^{\star}rn} = \sqrt{\frac{1}{\kappa_{\mathsf{ce}}}\lambda_1^{\star}rn},$$

where the last step follows from (33) and (34). As a result,

$$\|\boldsymbol{B}^{\star}\boldsymbol{F}^{\star}\|_{\infty} \leq \sqrt{\frac{\mu_{\mathsf{ce}}\kappa_{\mathsf{ce}}\log(n+d)}{nd}}\,\|\boldsymbol{B}^{\star}\boldsymbol{F}^{\star}\|_{\mathrm{F}} \tag{35}$$

Recalling the definition of μ_0 in (2.2a), one obtains

$$\mu_0 \lesssim \mu_{\mathsf{ce}}\kappa_{\mathsf{ce}}\log(n+d). \tag{36}$$

When it comes to the incoherence parameters μ_1 and μ_2 (cf. (2.2b)), it can be easily verified that

$$\mu_1 = \frac{d}{r} \|\boldsymbol{U}^\star\|_{2,\infty}^2 = \mu_{\mathsf{ce}}.$$

In addition, recognizing the existence of an orthonormal matrix $\boldsymbol{H}_2$ such that $\boldsymbol{V}^\star \boldsymbol{H}_2 = \boldsymbol{F}^{\star\top}(\boldsymbol{F}^\star \boldsymbol{F}^{\star\top})^{-1/2}$, we can bound

$$\begin{aligned}\|\boldsymbol{V}^\star\|_{2,\infty} = \|\boldsymbol{V}^\star \boldsymbol{H}_2\|_{2,\infty} \le \left\|\boldsymbol{F}^{\star\top}\right\|_{2,\infty} \left\| \left(\boldsymbol{F}^\star \boldsymbol{F}^{\star\top}\right)^{-1/2} \right\| &\overset{\text{(i)}}{\lesssim} \frac{\sqrt{r} + \sqrt{\log(n+d)}}{\sigma_r(\boldsymbol{F}^\star)} \\ &\overset{\text{(ii)}}{\lesssim} \frac{\sqrt{r} + \sqrt{\log(n+d)}}{\sqrt{n} - \sqrt{r} - \sqrt{\log(n+d)}} \overset{\text{(iii)}}{\lesssim} \frac{\sqrt{r\log(n+d)}}{\sqrt{n}},\end{aligned}$$

where (i) follows from the standard Gaussian concentration result that $\left| \left\|\boldsymbol{F}^{\star\top}\right\|_{2,\infty} - \sqrt{r}\right| \lesssim \sqrt{\log(n+d)}$ with probability $1 - O\big((n+d)^{-20}\big)$, (ii) arises from (31), and (iii) holds true under our sample size assumption. Consequently, we obtain

$$\mu_2 = \frac{n}{r} \|\boldsymbol{V}^\star\|_{2,\infty}^2 \lesssim \log(n+d).$$

Thus far, we have shown that

$$\sigma_r^\star \asymp \sqrt{\lambda_r^\star n}, \quad \mu_0 \lesssim \mu_{\mathsf{ce}}\kappa_{\mathsf{ce}} \log(n+d), \quad \mu_1 = \mu_{\mathsf{ce}}, \quad \mu_2 \lesssim \log(n+d), \quad \mu \lesssim \mu_{\mathsf{ce}}\kappa_{\mathsf{ce}} \log(n+d) \quad \text{and} \quad \kappa \asymp \sqrt{\kappa_{\mathsf{ce}}},$$

where $\sigma_r^\star = \sigma_r^\star(\boldsymbol{B}^\star \boldsymbol{F}^\star)$, $\mu = \max\{\mu_0, \mu_1, \mu_2\}$ and $\kappa = \kappa(\boldsymbol{B}^\star \boldsymbol{F}^\star)$. Under the sample size assumption (4.14) and the rank condition $r \ll \frac{d}{\mu_{\mathsf{ce}}\kappa_{\mathsf{ce}}^2}$, it is straightforward to verify the condition (3.5) is satisfied, i.e.

$$\begin{aligned} &p \gg \max\left\{\frac{\mu_{\mathsf{ce}}\kappa_{\mathsf{ce}}^3 r \log^3(n+d)}{\sqrt{dn}}, \frac{\mu_{\mathsf{ce}}\kappa_{\mathsf{ce}}^5 r \log^3(n+d)}{n}\right\} \gtrsim \max\left\{\frac{\mu\kappa^4 r \log^2(n+d)}{\sqrt{dn}}, \frac{\mu\kappa^8 r\log^2(n+d)}{n}\right\}, \\ &\frac{\sigma}{\sigma_r^\star} \asymp \frac{\sigma}{\sqrt{\lambda_r^\star n}} \ll \min\left\{\frac{\sqrt{p}}{\sqrt{\kappa_{\mathsf{ce}}}\sqrt[4]{dn}\sqrt{\log(n+d)}}, \sqrt{\frac{p}{\kappa_{\mathsf{ce}}^3 d\log(n+d)}}\right\} \lesssim \min\left\{\frac{\sqrt{p}}{\kappa\sqrt[4]{d_1 d_2}\sqrt{\log d}}, \frac{1}{\kappa^3}\sqrt{\frac{p}{d_1 \log d}}\right\}, \\ &r \ll \frac{d}{\mu_{\mathsf{ce}}\kappa_{\mathsf{ce}}^2} \asymp \frac{d}{\mu_1 \kappa^4}. \end{aligned}$$

Applying Theorem 3.1 immediately establishes the claims (4.15a) and (4.15b) in Corollary 4.3. Along the way, we have also established the following upper bound (see Lemma 1), which will be useful in the sequel:

$$\|\boldsymbol{G} - \boldsymbol{G}^\star\| \lesssim \lambda_r(\boldsymbol{G}^\star) \cdot \mathcal{E}_{\mathsf{ce}} \asymp \lambda_r^\star n \cdot \mathcal{E}_{\mathsf{ce}}. \tag{37}$$

Here, we recall that $\boldsymbol{G} = \frac{1}{p^2}\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\mathcal{P}_\Omega(\boldsymbol{X})\mathcal{P}_\Omega(\boldsymbol{X})^\top\right)$.

9.2.2 Estimation error of the covariance matrix

It remains to prove (4.15c) and (4.15d). Before proceeding, we first recall that $\boldsymbol{U\Lambda U}^\top$ is the top-r eigendecomposition of $\boldsymbol{G}$,

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}^{1/2}, \quad \boldsymbol{B} = \frac{1}{\sqrt{n}}\boldsymbol{U\Sigma}, \quad \boldsymbol{B}^\star = \boldsymbol{U}^\star\boldsymbol{\Lambda}^{\star 1/2} \quad \text{and} \quad \boldsymbol{R} = \mathop{\arg\min}_{\boldsymbol{Q}\in\mathcal{O}^{r\times r}} \|\boldsymbol{UQ} - \boldsymbol{U}^\star\|_{\mathrm{F}}. \tag{38}$$

Let us also define

$$\boldsymbol{K} := \mathop{\arg\min}_{\boldsymbol{Q}\in\mathcal{O}^{r\times r}} \|\boldsymbol{BQ} - \boldsymbol{B}^\star\|_{\mathrm{F}}.$$

It is well known that the minimizer $\boldsymbol{K}$ is given by [TB77]

$$\boldsymbol{K} = \mathsf{sgn}\left(\boldsymbol{B}^\top \boldsymbol{B}^\star\right),$$

where the $\mathsf{sgn}(\cdot)$ function is defined in (16). Since $\boldsymbol{K}$ is an orthonormal matrix, one can express

$$\boldsymbol{S} - \boldsymbol{S}^\star = (\boldsymbol{BK})(\boldsymbol{BK})^\top - \boldsymbol{B}^\star\boldsymbol{B}^{\star\top} = (\boldsymbol{BK} - \boldsymbol{B}^\star)(\boldsymbol{BK})^\top + \boldsymbol{B}^\star(\boldsymbol{BK} - \boldsymbol{B}^\star)^\top. \tag{39}$$

As a result, everything boils down to controlling $\|\boldsymbol{B}\boldsymbol{K}-\boldsymbol{B}^{\star}\|$ and $\|\boldsymbol{B}\boldsymbol{K}-\boldsymbol{B}^{\star}\|_{2,\infty}$. To this end, we use (38) to reach the following useful decomposition

$$\boldsymbol{B}\boldsymbol{K}-\boldsymbol{B}^{\star}=\frac{1}{\sqrt{n}}\boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\left(\boldsymbol{K}-\boldsymbol{R}\right)+\boldsymbol{U}\left(\frac{1}{\sqrt{n}}\boldsymbol{\Lambda}^{1/2}\boldsymbol{R}-\boldsymbol{R}\boldsymbol{\Lambda}^{\star 1/2}\right)+\left(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}\right)\boldsymbol{\Lambda}^{\star 1/2}. \tag{40}$$

Given that $\boldsymbol{U}\frac{1}{n}\boldsymbol{\Lambda}\boldsymbol{U}^{\top}$ is the top-r eigendecomposition of $\frac{1}{n}\boldsymbol{G}$, an important step lies in controlling the difference between $\frac{1}{n}\boldsymbol{G}$ and $\boldsymbol{S}^{\star}$. Recalling the matrix $\boldsymbol{\Delta}$ as defined in (30), one can use (32), (37) as well as the definition of $\mathcal{E}_{\mathsf{ce}}$ (cf. (4.16)) to obtain

$$\begin{aligned}\left\|\frac{1}{n}\boldsymbol{G}-\boldsymbol{S}^{\star}\right\| &\leq \frac{1}{n}\left\|\boldsymbol{G}-\boldsymbol{G}^{\star}\right\|+\left\|\frac{1}{n}\boldsymbol{G}^{\star}-\boldsymbol{S}^{\star}\right\|=\frac{1}{n}\left\|\boldsymbol{G}-\boldsymbol{G}^{\star}\right\|+\frac{1}{n}\left\|\boldsymbol{\Delta}\right\|\\ &\lesssim \frac{1}{n}\lambda_{r}\left(\boldsymbol{G}^{\star}\right)\cdot\mathcal{E}_{\mathsf{ce}}+\frac{\lambda_{1}^{\star}}{\sqrt{n}}\left(\sqrt{r}+\sqrt{\log\left(n+d\right)}\right)\asymp\lambda_{r}^{\star}\cdot\mathcal{E}_{\mathsf{ce}},\end{aligned}$$

where the last inequality makes use of the identity $\lambda_{r}(\boldsymbol{G}^{\star})\asymp n\lambda_{r}^{\star}$. Hence, apply [MWCC17, Lemma 46, Lemma 47] (with slight modification on κ) and Weyl's inequality to show that

$$\left\|\frac{1}{\sqrt{n}}\boldsymbol{\Lambda}^{1/2}\boldsymbol{R}-\boldsymbol{R}\boldsymbol{\Lambda}^{\star 1/2}\right\|\lesssim\frac{\kappa\left(\boldsymbol{S}^{\star}\right)}{\sqrt{\lambda_{r}\left(\boldsymbol{S}^{\star}\right)}}\left\|\frac{1}{n}\boldsymbol{G}-\boldsymbol{S}^{\star}\right\|\lesssim\kappa_{\mathsf{ce}}\sqrt{\lambda_{r}^{\star}}\cdot\mathcal{E}_{\mathsf{ce}}; \tag{41}$$

$$\left\|\boldsymbol{K}-\boldsymbol{R}\right\|\lesssim\frac{\sqrt{\kappa\left(\boldsymbol{S}^{\star}\right)}}{\lambda_{r}\left(\boldsymbol{S}^{\star}\right)}\left\|\frac{1}{n}\boldsymbol{G}-\boldsymbol{S}^{\star}\right\|\lesssim\sqrt{\kappa_{\mathsf{ce}}}\cdot\mathcal{E}_{\mathsf{ce}}. \tag{42}$$

In addition, it follows from Weyl's inequality that

$$\left\|\frac{1}{n}\boldsymbol{\Lambda}-\boldsymbol{\Lambda}^{\star}\right\|\leq\left\|\frac{1}{n}\boldsymbol{G}-\boldsymbol{S}^{\star}\right\|\lesssim\lambda_{r}^{\star}\cdot\mathcal{E}_{\mathsf{ce}},$$

which combined with (29) gives

$$\frac{1}{n}\left\|\boldsymbol{\Lambda}\right\|\leq\left\|\frac{1}{n}\boldsymbol{\Lambda}-\boldsymbol{\Lambda}^{\star}\right\|+\left\|\boldsymbol{\Lambda}^{\star}\right\|\lesssim\lambda_{r}^{\star}\cdot\mathcal{E}_{\mathsf{ce}}+\lambda_{1}^{\star}\asymp\lambda_{1}^{\star} \tag{43}$$

under our assumptions.

We are ready to upper bound the difference between $\boldsymbol{B}\boldsymbol{K}-\boldsymbol{B}^{\star}$. Plugging (4.15a), (41) (42) and (43) into (40) shows that

$$\begin{aligned}\left\|\boldsymbol{B}\boldsymbol{K}-\boldsymbol{B}^{\star}\right\| &\leq\frac{1}{\sqrt{n}}\left\|\boldsymbol{\Lambda}\right\|^{1/2}\left\|\boldsymbol{K}-\boldsymbol{R}\right\|+\left\|\frac{1}{\sqrt{n}}\boldsymbol{\Lambda}^{1/2}\boldsymbol{R}-\boldsymbol{R}\boldsymbol{\Lambda}^{\star 1/2}\right\|+\left\|\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}\right\|\left\|\boldsymbol{\Lambda}^{\star}\right\|^{1/2}\\ &\lesssim\sqrt{\kappa_{\mathsf{ce}}\lambda_{1}^{\star}}\cdot\mathcal{E}_{\mathsf{ce}}+\kappa_{\mathsf{ce}}\sqrt{\lambda_{r}^{\star}}\cdot\mathcal{E}_{\mathsf{ce}}+\sqrt{\lambda_{1}^{\star}}\cdot\mathcal{E}_{\mathsf{ce}}\\ &\lesssim\kappa_{\mathsf{ce}}\sqrt{\lambda_{1}^{\star}}\cdot\mathcal{E}_{\mathsf{ce}}.\end{aligned} \tag{44}$$

Since $\boldsymbol{K}\in\mathcal{O}^{r\times r}$, this also implies that

$$\left\|\boldsymbol{B}\right\|=\left\|\boldsymbol{B}\boldsymbol{K}\right\|\leq\left\|\boldsymbol{B}\boldsymbol{K}-\boldsymbol{B}^{\star}\right\|+\left\|\boldsymbol{B}^{\star}\right\|\lesssim\kappa_{\mathsf{ce}}\sqrt{\lambda_{1}^{\star}}\cdot\mathcal{E}_{\mathsf{ce}}+\sqrt{\lambda_{1}^{\star}}\asymp\sqrt{\lambda_{1}^{\star}}, \tag{45}$$

where the last step results from (29). In addition, (4.15b), (29) and the fact that $\boldsymbol{R}\in\mathcal{O}^{r\times r}$ guarantees that

$$\left\|\boldsymbol{U}\right\|_{2,\infty}=\left\|\boldsymbol{U}\boldsymbol{R}\right\|_{2,\infty}\leq\left\|\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}\right\|_{2,\infty}+\left\|\boldsymbol{U}^{\star}\right\|_{2,\infty}\lesssim\kappa_{\mathsf{ce}}^{3/2}\mathcal{E}_{\mathsf{ce}}\sqrt{\frac{\mu_{\mathsf{ce}}r\log\left(n+d\right)}{d}}+\sqrt{\frac{\mu_{\mathsf{ce}}r}{d}}\lesssim\sqrt{\frac{\kappa_{\mathsf{ce}}\mu_{\mathsf{ce}}r\log\left(n+d\right)}{d}}.$$

Consequently, it follows from the decomposition (40) that

$$\left\|\boldsymbol{B}\boldsymbol{K}-\boldsymbol{B}^{\star}\right\|_{2,\infty}\leq\left\|\boldsymbol{U}\right\|_{2,\infty}\frac{1}{\sqrt{n}}\left\|\boldsymbol{\Lambda}\right\|^{1/2}\left\|\boldsymbol{K}-\boldsymbol{R}\right\|+\left\|\boldsymbol{U}\right\|_{2,\infty}\left\|\frac{1}{\sqrt{n}}\boldsymbol{\Lambda}^{1/2}\boldsymbol{R}-\boldsymbol{R}\boldsymbol{\Lambda}^{\star 1/2}\right\|+\left\|\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}\right\|_{2,\infty}\left\|\boldsymbol{\Lambda}^{\star}\right\|^{1/2}$$

$$
\begin{aligned}
&\lesssim \sqrt{\frac{\kappa_{\mathsf{ce}}^2 \mu_{\mathsf{ce}} r \lambda_1^\star \log(n+d)}{d}}\, \mathcal{E}_{\mathsf{ce}} + \sqrt{\frac{\kappa_{\mathsf{ce}}^3 \mu_{\mathsf{ce}} r \lambda_r^\star \log(n+d)}{d}}\, \mathcal{E}_{\mathsf{ce}} + \sqrt{\frac{\kappa_{\mathsf{ce}}^3 \mu_{\mathsf{ce}} r \lambda_1^\star \log(n+d)}{d}}\, \mathcal{E}_{\mathsf{ce}} \\
&\asymp \sqrt{\frac{\kappa_{\mathsf{ce}}^3 \mu_{\mathsf{ce}} r \lambda_1^\star \log(n+d)}{d}}\, \mathcal{E}_{\mathsf{ce}}. \qquad (46)
\end{aligned}
$$

Combining (46) and (29) gives that

$$
\begin{aligned}
\|\boldsymbol{B}\|_{2,\infty} = \|\boldsymbol{B}\boldsymbol{K}\|_{2,\infty} \le \|\boldsymbol{B}\boldsymbol{K} - \boldsymbol{B}^\star\|_{2,\infty} + \|\boldsymbol{B}^\star\|_{2,\infty} &\lesssim \sqrt{\frac{\kappa_{\mathsf{ce}}^3 \mu_{\mathsf{ce}} r \lambda_1^\star \log(n+d)}{d}}\, \mathcal{E}_{\mathsf{ce}} + \sqrt{\frac{\mu_{\mathsf{ce}} r \lambda_1^\star}{d}} \\
&\lesssim \sqrt{\frac{\mu_{\mathsf{ce}} r \lambda_1^\star \log(n+d)}{d}}, \qquad (47)
\end{aligned}
$$

where we use the fact that $\|\boldsymbol{B}^\star\|_{2,\infty} = \|\boldsymbol{U}^\star \boldsymbol{\Lambda}^{\star 1/2}\|_{2,\infty} \lesssim \|\boldsymbol{U}^\star\|_{2,\infty} \|\boldsymbol{\Lambda}^\star\|^{1/2} \lesssim \sqrt{\mu_{\mathsf{ce}} r \lambda_1^\star / d}$ and $\mathcal{E}_{\mathsf{ce}} \ll \kappa_{\mathsf{ce}}^{-1}$.

To finish up, we substitute (44) and (45) into (39) to find that

$$
\begin{aligned}
\|\boldsymbol{S} - \boldsymbol{S}^\star\| \le \|\boldsymbol{B}\boldsymbol{K} - \boldsymbol{B}^\star\| \left(\|\boldsymbol{B}^\star\| + \|\boldsymbol{B}\boldsymbol{K}\|\right) &\le \|\boldsymbol{B}\boldsymbol{K} - \boldsymbol{B}^\star\| \left(\|\boldsymbol{B}^\star\| + \|\boldsymbol{B}\|\right) \\
&\lesssim \kappa_{\mathsf{ce}} \lambda_1^\star \cdot \mathcal{E}_{\mathsf{ce}}.
\end{aligned}
$$

Combining (46) and (47) reveals that

$$
\begin{aligned}
\|\boldsymbol{S} - \boldsymbol{S}^\star\|_\infty \le \|\boldsymbol{B}\boldsymbol{K} - \boldsymbol{B}^\star\|_{2,\infty} \left(\|\boldsymbol{B}\boldsymbol{K}\|_{2,\infty} + \|\boldsymbol{B}^\star\|_{2,\infty}\right) &\le \|\boldsymbol{B}\boldsymbol{K} - \boldsymbol{B}^\star\|_{2,\infty} \left(\|\boldsymbol{B}\|_{2,\infty} + \|\boldsymbol{B}^\star\|_{2,\infty}\right) \\
&\lesssim \frac{\kappa_{\mathsf{ce}}^2 \mu_{\mathsf{ce}} r \lambda_1^\star \log(n+d)}{d} \cdot \mathcal{E}_{\mathsf{ce}}.
\end{aligned}
$$

We have therefore established all claims.

9.3 Proof of Corollary 4.6

Recall from our calculation (4.20) that

$$
\boldsymbol{A}^\star := \mathbb{E}[\boldsymbol{A}] = \frac{(q_{\mathsf{in}} - q_{\mathsf{out}}) \sqrt{n_u n_v}}{2} \boldsymbol{u}^\star \boldsymbol{v}^{\star\top}
$$

is a rank-1 matrix, where

$$
\boldsymbol{u}^\star := \frac{1}{\sqrt{n_u}} \begin{bmatrix} \boldsymbol{1}_{n_u/2} \\ -\boldsymbol{1}_{n_u/2} \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{v}^\star := \frac{1}{\sqrt{n_v}} \begin{bmatrix} \boldsymbol{1}_{n_v/2} \\ -\boldsymbol{1}_{n_v/2} \end{bmatrix}.
$$

Let $\boldsymbol{u} \in \mathbb{R}^{n_u}$ be the leading eigenvector of $\boldsymbol{G}$ (cf. (3.2) and Algorithm 4). To establish Corollary 4.6, the main step boils down to showing that, under the conditions of Corollary 4.6,

$$
\min\left\{\|\boldsymbol{u} - \boldsymbol{u}^\star\|_\infty, \|\boldsymbol{u} + \boldsymbol{u}^\star\|_\infty\right\} \lesssim \frac{1}{\sqrt{n_u}} \mathcal{E}_{\mathsf{bsbm}}, \qquad (48)
$$

holds with probability exceeding $1 - O\left(n^{-10}\right)$, where

$$
\mathcal{E}_{\mathsf{bsbm}} := \frac{q_{\mathsf{in}}}{(q_{\mathsf{in}} - q_{\mathsf{out}})^2} \frac{\log n}{\sqrt{n_u n_v}} + \frac{\sqrt{q_{\mathsf{in}}}}{q_{\mathsf{in}} - q_{\mathsf{out}}} \sqrt{\frac{\log n}{n_v}} + \frac{1}{\sqrt{n_u}}. \qquad (49)
$$

If this claim (48) holds, then under our condition (4.21) one has $\mathcal{E}_{\mathsf{bsbm}} \ll 1$, and hence

$$
\min\left\{\|\boldsymbol{u} - \boldsymbol{u}^\star\|_\infty, \|\boldsymbol{u} + \boldsymbol{u}^\star\|_\infty\right\} \lesssim \frac{1}{\sqrt{n_u}} \mathcal{E}_{\mathsf{bsbm}} < \frac{1}{\sqrt{n_u}}.
$$

In other words, one has either $\mathsf{sign}(u_i) = \mathsf{sign}(u_i^\star)$ for all $1 \le i \le n_u$, or $\mathsf{sign}(u_i) = -\mathsf{sign}(u_i^\star)$ for all $1 \le i \le n_u$. This tells us that the entrywise rounding operation applied to $\boldsymbol{u}$ is sufficient to recover exactly the community memberships of all nodes in $\mathcal{U}$.

The rest of the proof is devoted to establishing the claim (48). In order to apply Theorem 3.1, it suffices to estimate the spectrum and the incoherence parameters of $\boldsymbol{A}^\star$, as well as some simple statistical properties of $\boldsymbol{N} := \boldsymbol{A} - \boldsymbol{A}^\star$.

- We begin by looking at $\boldsymbol{A}^{\star}$, which has rank 1 and satisfies

$$\sigma_1\left(\boldsymbol{A}^{\star}\right)=\frac{\left(q_{\mathsf{in}}-q_{\mathsf{out}}\right)\sqrt{n_u n_v}}{2},\quad \left\|\boldsymbol{A}^{\star}\right\|_{\infty}=\frac{q_{\mathsf{in}}-q_{\mathsf{out}}}{2},\quad \left\|\boldsymbol{u}^{\star}\right\|_{\infty}=\frac{1}{\sqrt{n_u}},\quad \left\|\boldsymbol{v}^{\star}\right\|_{\infty}=\frac{1}{\sqrt{n_v}}.$$

  Recalling the definition of μ_0,μ_1,μ_2 in (2.2a) and (2.2b), we obtain

$$\mu_0=\frac{n_u n_v}{\left\|\boldsymbol{A}^{\star}\right\|_{\mathrm{F}}^2}\left\|\boldsymbol{A}^{\star}\right\|_{\infty}^2=1,\qquad \mu_1=n_u\left\|\boldsymbol{u}^{\star}\right\|_{2,\infty}^2=1,\qquad \mu_2=n_v\left\|\boldsymbol{v}^{\star}\right\|_{2,\infty}^2=1,\qquad \kappa=1.$$

- Next, we consider the maximum magnitude R and the maximum variance σ^2 of all entries of $\boldsymbol{N}$ (see Assumption 2.3). Clearly, one has

$$\begin{aligned}
R&=\max_{i,j}\left|N_{i,j}\right|\leq 1,\\
\sigma^2&=\max_{i,j}\mathsf{Var}(N_{i,j})=\max\left\{q_{\mathsf{in}}(1-q_{\mathsf{in}}),q_{\mathsf{out}}(1-q_{\mathsf{out}})\right\}\leq\max\left\{q_{\mathsf{in}},q_{\mathsf{out}}\right\}\asymp q_{\mathsf{in}},
\end{aligned}$$

  which follows since $N_{i,j}$ is a centered Bernoulli random variable with parameter either q_{in} or q_{out}. From the assumption (4.21) and the fact $q_{\mathsf{in}}^2\geq(q_{\mathsf{in}}-q_{\mathsf{out}})^2$, we know that

$$q_{\mathsf{in}}\geq\frac{\left(q_{\mathsf{in}}-q_{\mathsf{out}}\right)^2}{q_{\mathsf{in}}}\gg\frac{\log n}{\sqrt{n_u n_v}}+\frac{\log n}{n_v}.$$

  Putting the above estimates together, we can straightforwardly verify the random noise requirement (2.6), namely,

$$\frac{R^2}{\sigma^2}\lesssim\frac{1}{q_{\mathsf{in}}}\lesssim\frac{\min\left\{\sqrt{n_u n_v},n_v\right\}}{\log n}.$$

With the preceding bounds in place, Corollary 4.6 is an immediate consequence of Theorem 3.1.

10 Proofs for key lemmas

This section aims to establish the key lemmas listed in Section 8.2.

10.1 Auxiliary quantities, notation, and preliminary facts

To simplify our treatment, the proofs shall consider the influence of missing data and that of noise altogether. Specifically, throughout this section, we shall define a rescaled version of $\boldsymbol{A}$ as follows

$$\boldsymbol{A}^{\mathsf{s}}:=\tfrac{1}{p}\boldsymbol{A}=\boldsymbol{A}^{\star}+\boldsymbol{E}\in\mathbb{R}^{d_1\times d_2},\tag{50}$$

where the matrix $\boldsymbol{E}$ represents the aggregate perturbation

$$\boldsymbol{E}:=\tfrac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{A}^{\star}\right)-\boldsymbol{A}^{\star}+\tfrac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{N}\right).\tag{51}$$

Clearly, $\boldsymbol{E}\in\mathbb{R}^{d_1\times d_2}$ is a random matrix with independent zero-mean entries and $\mathbb{E}\left[\boldsymbol{A}^{\mathsf{s}}\right]=\boldsymbol{A}^{\star}$. In addition, we define the corresponding leave-one-out and leave-two-out versions

$$\boldsymbol{A}^{\mathsf{s},(m)}:=\tfrac{1}{p}\boldsymbol{A}^{(m)},\tag{52}$$

$$\boldsymbol{A}^{\mathsf{s},(m,l)}:=\tfrac{1}{p}\boldsymbol{A}^{(m,l)},\tag{53}$$

for each $1\leq m\leq d_1,1\leq l\leq d_2$.

As we shall see momentarily, it is convenient to introduce the following quantities regarding the above perturbation matrix $\boldsymbol{E}$: (1) $\max_{i\in[d_1],j\in[d_2]}\left|E_{i,j}\right|$; (2) $\max_{i\in[d_1],j\in[d_2]}\sqrt{\mathbb{E}\left[E_{i,j}^2\right]}$; (3) $\max_{i\in[d_1]}\sqrt{\sum_{j\in[d_2]}\mathbb{E}\left[E_{i,j}^2\right]}$; (4) $\max_{j\in[d_2]}\sqrt{\sum_{i\in[d_1]}\mathbb{E}\left[E_{i,j}^2\right]}$. In our settings, it is easy to verify — using the definition of incoherence parameters (cf. Definition 2.1), the assumptions of the random noise (cf. Assumption 2.3), and Lemma 11 — that the quantities defined above admit the following upper bounds

$$\max_{i\in[d_1],j\in[d_2]}|E_{i,j}| \le \frac{\|\boldsymbol{A}^\star\|_\infty + R}{p} \lesssim \frac{\sqrt{\mu r}\,\sigma_1^\star}{\sqrt{d_1 d_2\, p}} + \frac{\sigma \min\left\{\sqrt[4]{d_1 d_2},\, \sqrt{d_2}\right\}}{\sqrt{p \log d}} =: B, \tag{54a}$$

$$\max_{i\in[d_1],j\in[d_2]}\sqrt{\mathbb{E}\left[E_{i,j}^2\right]} \le \frac{\|\boldsymbol{A}^\star\|_\infty + \sigma}{\sqrt{p}} \le \sigma_1^\star\sqrt{\frac{\mu r}{d_1 d_2 p}} + \frac{\sigma}{\sqrt{p}} =: \sigma_\infty, \tag{54b}$$

$$\max_{i\in[d_1]}\sqrt{\sum_{j\in[d_2]}\mathbb{E}\left[E_{i,j}^2\right]} \le \frac{\|\boldsymbol{A}^\star\|_{2,\infty} + \sigma\sqrt{d_2}}{\sqrt{p}} \le \sigma_1^\star\sqrt{\frac{\mu r}{d_1 p}} + \sigma\sqrt{\frac{d_2}{p}} =: \sigma_{\mathsf{row}}, \tag{54c}$$

$$\max_{j\in[d_2]}\sqrt{\sum_{i\in[d_1]}\mathbb{E}\left[E_{i,j}^2\right]} \le \frac{\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty} + \sigma\sqrt{d_1}}{\sqrt{p}} \le \sigma_1^\star\sqrt{\frac{\mu r}{d_2 p}} + \sigma\sqrt{\frac{d_1}{p}} =: \sigma_{\mathsf{col}}, \tag{54d}$$

with probability exceeding $1-O(d^{-12})$. Further, the following lemma singles out a few other useful properties about these quantities (to be established in Appendix 10.6), which will be useful throughout the proof.

Lemma 5. *Instate the assumptions of Theorem 3.1. Then with probability at least $1 - O\left(d^{-12}\right)$, we have*

$$B \lesssim \frac{\min\left\{\sqrt{\sigma_{\mathsf{row}}\sigma_{\mathsf{col}}},\, \sigma_{\mathsf{row}}\right\}}{\sqrt{\log d}}; \tag{55a}$$

$$\sigma_\infty^2 \lesssim B\log d\,\|\boldsymbol{A}^\star\|\sqrt{\frac{\mu r}{d_2}} \lesssim \sigma_{\mathsf{col}}\sqrt{\log d}\,\|\boldsymbol{A}^\star\|\sqrt{\frac{\mu r}{d_1}}; \tag{55b}$$

$$\sqrt{\frac{\mu r}{d_1}} \gtrsim \frac{B\log^{3/2} d\,\|\boldsymbol{A}^\star\|_\infty}{\|\boldsymbol{A}^\star\|^2}; \tag{55c}$$

$$\sigma_r^{\star 2} \gg \max\left\{\kappa^2\sigma_{\mathsf{col}}\sigma_{\mathsf{row}}\log d,\, \kappa^2\sigma_{\mathsf{col}}\sqrt{\log d}\,\|\boldsymbol{A}^\star\|,\, \kappa^2\|\boldsymbol{A}^\star\|_{2,\infty}^2,\, \sigma_{\mathsf{row}}\sqrt{\frac{d_1\log d}{\mu r}}\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty},\, B\log d\,\|\boldsymbol{A}^\star\|_\infty\right\}. \tag{55d}$$

10.2 Proof of Lemma 1

The main component of the proof is to demonstrate that

$$\|\boldsymbol{G}-\boldsymbol{G}^\star\| \lesssim \underbrace{\left(\sigma_{\mathsf{row}}+\sigma_{\mathsf{col}}\right)\left(\sigma_{\mathsf{col}} + \left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\right)\log d + \sigma_{\mathsf{col}}\sqrt{\log d}\,\|\boldsymbol{A}^\star\| + \|\boldsymbol{A}^\star\|_{2,\infty}^2}_{=:\delta_{\mathsf{op}}}. \tag{56}$$

By substituting the values of σ_{row} and σ_{col} (cf. (54)) into the above expression, one derives

$$\delta_{\mathsf{op}} \lesssim \zeta_{\mathsf{op}} + \|\boldsymbol{A}^\star\|_{2,\infty}^2, \tag{57}$$

where ζ_{op} is defined in (5). Therefore, Lemma 1 is an immediate consequence of (56) and (57). The remainder of the proof amounts to justifying (56).

Recall the definitions of $\boldsymbol{G}$ and $\boldsymbol{G}^\star$ in (3.3) and (4), respectively. Given that $\boldsymbol{A}^{\mathsf{s}} = \boldsymbol{A}^\star + \boldsymbol{E}$, we can expand

$$\begin{aligned}\boldsymbol{G}-\boldsymbol{G}^\star &= \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{A}^{\mathsf{s}}\boldsymbol{A}^{\mathsf{s}\top}\right) - \boldsymbol{A}^\star\boldsymbol{A}^{\star\top} = \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{A}^{\mathsf{s}}\boldsymbol{A}^{\mathsf{s}\top} - \boldsymbol{A}^\star\boldsymbol{A}^{\star\top}\right) - \mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{A}^\star\boldsymbol{A}^{\star\top}\right)\\ &= \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right) + \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{A}^\star\boldsymbol{E}^\top + \boldsymbol{E}\boldsymbol{A}^{\star\top}\right) - \mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{A}^\star\boldsymbol{A}^{\star\top}\right),\end{aligned} \tag{58}$$

where $\mathcal{P}_{\mathsf{off\text{-}diag}}$ and $\mathcal{P}_{\mathsf{diag}}$ are defined in Section 2.2. In what follows, we control these three terms separately.

10.2.1 Step 1: bounding the term $\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)$

We first consider the term $\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)$. Since $\{E_{i,j}\}_{i\in[d_1],j\in[d_2]}$ are independent zero-mean random variables, we can express

$$\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right) = \sum\nolimits_{1\leq l\leq d_2}\left(\boldsymbol{E}_{:,l}\boldsymbol{E}_{:,l}^\top - \boldsymbol{D}_l\right) \tag{59}$$

as a sum of independent zero-mean random matrices, where $\boldsymbol{D}_l$ is a random diagonal matrix in $\mathbb{R}^{d_1\times d_1}$ with entries

$$\left(\boldsymbol{D}_l\right)_{i,i} = E_{i,l}^2. \tag{60}$$

We intend to invoke the truncated matrix Bernstein inequality [HSSS16, Proposition A.7] to control the spectral norm of (59). To this end, we need to look at a few quantities.

- We first bound the spectral norm of the following covariance matrix

  $$\boldsymbol{\Sigma}_{\textsf{ns}} := \sum\nolimits_{1\leq l\leq d_2}\mathbb{E}\left[\left(\boldsymbol{E}_{:,l}\boldsymbol{E}_{:,l}^\top - \boldsymbol{D}_l\right)^2\right] \in \mathbb{R}^{d_1\times d_1}.$$

  Straightforward computation reveals that $\boldsymbol{\Sigma}_{\textsf{ns}}$ is a diagonal matrix with entries

  $$(\boldsymbol{\Sigma}_{\textsf{ns}})_{i,i} = \sum\nolimits_{1\leq l\leq d_2}\mathbb{E}\left[E_{i,l}^2\right]\sum\nolimits_{m:m\neq i}\mathbb{E}\left[E_{m,l}^2\right] \leq \sigma_{\textsf{row}}^2\sigma_{\textsf{col}}^2$$

  for each $1\leq i\leq d_1$. This immediately reveals that

  $$V_{\textsf{ns}} := \|\boldsymbol{\Sigma}_{\textsf{ns}}\| \leq \sigma_{\textsf{row}}^2\sigma_{\textsf{col}}^2. \tag{61}$$

- Next, we turn to upper bounding the spectral norm of each summand $\boldsymbol{E}_{:,l}\boldsymbol{E}_{:,l}^\top - \boldsymbol{D}_l$. As shown in the proof of Lemma 12, one has

  $$\mathbb{P}\left\{\left|\|\boldsymbol{E}_{:,l}\|_2^2 - M_1\right| \geq t\right\} \leq 2\exp\left(-\frac{3}{8}\min\left\{\frac{t^2}{V_1},\frac{t}{L_1}\right\}\right), \quad t>0,$$

  where M_1, L_1 and V_1 are given respectively by

  $$\begin{aligned} M_1 &:= \mathbb{E}\left[\|\boldsymbol{E}_{:,l}\|_2^2\right] \leq \sigma_{\textsf{col}}^2, \\ L_1 &:= \max_{1\leq i\leq d_1}\left|E_{i,l}^2 - \mathbb{E}\left[E_{i,l}^2\right]\right| \leq 2B^2, \\ V_1 &:= \sum\nolimits_{1\leq i\leq d_1}\mathsf{Var}\left(E_{i,l}^2\right) \leq B^2\sigma_{\textsf{col}}^2. \end{aligned}$$

  In addition, with probability exceeding $1 - O\left(d^{-20}\right)$,

  $$\begin{aligned} \|\boldsymbol{E}_{:,l}\|_2^2 &\lesssim M_1 + L_1\log d + \sqrt{V_1\log d} \lesssim \sigma_{\textsf{col}}^2 + B^2\log d + \sqrt{B^2\sigma_{\textsf{col}}^2\log d} \\ &\asymp B^2\log d + \sigma_{\textsf{col}}^2, \end{aligned} \tag{62}$$

  where the last line comes from the AM-GM inequality $2\sqrt{B^2\sigma_{\textsf{col}}^2\log d} \leq B^2\log d + \sigma_{\textsf{col}}^2$. This together with the definition $\boldsymbol{D}_l := \mathsf{diag}\left(E_{i,l}^2,\ldots,E_{d_1,l}^2\right)$ gives

  $$\left\|\boldsymbol{E}_{:,l}\boldsymbol{E}_{:,l}^\top - \boldsymbol{D}_l\right\| \leq \|\boldsymbol{E}_{:,l}\|_2^2 + \|\boldsymbol{D}_l\| \leq 2\|\boldsymbol{E}_{:,l}\|_2^2 \lesssim B^2\log d + \sigma_{\textsf{col}}^2.$$

  Therefore, if we set

  $$L_{\textsf{ns}} := C\left(B^2\log d + \sigma_{\textsf{col}}^2\right) \tag{63}$$

  for some sufficiently large constant $C>0$, then the above argument reveals that

  $$L_{\textsf{ns}} \geq \frac{C}{3}\left(M_1 + L_1\log d + \sqrt{V_1\log d}\right) \geq \frac{C}{3}\max\left\{\sqrt{V_1\log d}, L_1\log d\right\}.$$

- In addition, one can easily bound that

$$\mathbb{E}\left[\|\boldsymbol{E}_{:,l}\|_2^2\,\mathbb{1}\left\{\|\boldsymbol{E}_{:,l}\|_2^2\geq L_{\mathsf{ns}}\right\}\right]\leq L_{\mathsf{ns}}\mathbb{P}\left\{\|\boldsymbol{E}_{:,l}\|_2^2\geq L_{\mathsf{ns}}\right\}+\int_{L_{\mathsf{ns}}}^{\infty}\mathbb{P}\left\{\|\boldsymbol{E}_{:,l}\|_2^2\geq t\right\}\mathrm{d}t$$
$$\leq O\left(d^{-20}\right)L_{\mathsf{ns}}+\int_{L_{\mathsf{ns}}}^{\infty}\mathbb{P}\left\{\|\boldsymbol{E}_{:,l}\|_2^2\geq t\right\}\mathrm{d}t.$$

Moreover, we know that $\min\left\{t^2/V_1,t/L_1\right\}\geq t/\max\left\{\sqrt{V_1/\log d},L_1\right\}$ for any $t\geq L_{\mathsf{ns}}/2$. As a result, for sufficiently large d, we have

$$\begin{aligned}\int_{L_{\mathsf{ns}}}^{\infty}\mathbb{P}\left\{\|\boldsymbol{E}_{:,l}\|_2^2\geq t\right\}\mathrm{d}t&\leq\int_{L_{\mathsf{ns}}}^{\infty}\mathbb{P}\left\{\left|\|\boldsymbol{E}_{:,l}\|_2^2-M_1\right|>\frac{t}{2}\right\}\mathrm{d}t\\&\leq 4\int_{\frac{1}{2}L_{\mathsf{ns}}}^{\infty}\exp\left(-\frac{3}{8}\min\left\{\frac{t^2}{V_1},\frac{t}{L_1}\right\}\right)\mathrm{d}t\\&\leq 4\int_{\frac{1}{2}L_{\mathsf{ns}}}^{\infty}\exp\left(-\frac{3}{8}\frac{t}{\max\left\{\sqrt{V_1/\log d},L_1\right\}}\right)\mathrm{d}t\\&\lesssim\max\left\{\sqrt{V_1/\log d},L_1\right\}\exp\left(-\frac{3}{16}\frac{L_{\mathsf{ns}}}{\max\left\{\sqrt{V_1/\log d},L_1\right\}}\right)\\&\leq\max\left\{\sqrt{V_1/\log d},L_1\right\}\exp\left(-\frac{3C}{32}\log d\right)\\&\ll\frac{L_{\mathsf{ns}}}{d^2},\end{aligned}$$

provided that $C>0$ is sufficiently large. Consequently, we have

$$\begin{aligned}R_{\mathsf{ns}}&:=\mathbb{E}\left[\left\|\boldsymbol{E}_{:,l}\boldsymbol{E}_{:,l}^{\top}-\boldsymbol{D}_l\right\|\mathbb{1}\left\{\left\|\boldsymbol{E}_{:,l}\boldsymbol{E}_{:,l}^{\top}-\boldsymbol{D}_l\right\|\geq L_{\mathsf{ns}}\right\}\right]\\&\leq\mathbb{E}\left[2\left\|\boldsymbol{E}_{:,l}\right\|_2^2\mathbb{1}\left\{2\left\|\boldsymbol{E}_{:,l}\right\|_2^2\geq L_{\mathsf{ns}}\right\}\right]\ll\frac{L_{\mathsf{ns}}}{d^2}.\end{aligned}\tag{64}$$

With estimates (61), (63) and (64) in place, we are ready to apply the truncated matrix Bernstein inequality [HSSS16, Proposition A.7] to obtain that, with probability at least $1-O\left(d^{-10}\right)$,

$$\begin{aligned}\left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^{\top}\right)\right\|&=\left\|\sum\nolimits_{1\leq l\leq d_2}\boldsymbol{E}_{:,l}\boldsymbol{E}_{:,l}^{\top}-\boldsymbol{D}_l\right\|\lesssim d_2R_{\mathsf{ns}}+L_{\mathsf{ns}}\log d+\sqrt{V_{\mathsf{ns}}\log d}\\&\asymp L_{\mathsf{ns}}\log d+\sqrt{V_{\mathsf{ns}}\log d}\\&\lesssim B^2\log^2 d+\sigma_{\mathsf{col}}^2\log d+\sigma_{\mathsf{row}}\sigma_{\mathsf{col}}\sqrt{\log d}\\&\lesssim\sigma_{\mathsf{col}}\left(\sigma_{\mathsf{row}}+\sigma_{\mathsf{col}}\right)\log d,\end{aligned}\tag{65}$$

where the last line results from the identity $B^2\log d\lesssim\sigma_{\mathsf{row}}\sigma_{\mathsf{col}}$ (See (55a)).

10.2.2 Step 2: bounding the term $\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{A}^{\star}\boldsymbol{E}^{\top}+\boldsymbol{E}\boldsymbol{A}^{\star\top}\right)$

Next, we turn attention to $\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{A}^{\star}\boldsymbol{E}^{\top}+\boldsymbol{E}\boldsymbol{A}^{\star\top}\right)$. By symmetry, it suffices to control to the spectral norm of $\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{A}^{\star}\boldsymbol{E}^{\top}\right)$. To this end, we first express

$$\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{A}^{\star}\boldsymbol{E}^{\top}\right)=\sum\nolimits_{1\leq l\leq d_2}\left(\boldsymbol{A}_{:,l}^{\star}\boldsymbol{E}_{:,l}^{\top}-\widehat{\boldsymbol{D}}_l\right)\tag{66}$$

as a sum of independent zero-mean random matrices, where $\widehat{\boldsymbol{D}}_l$ is a diagonal matrix obeying

$$\left(\widehat{\boldsymbol{D}}_l\right)_{i,i}=A_{i,l}^{\star}E_{i,l}.\tag{67}$$

To control (66), we need to first look at two matrices defined as follows

$$\widehat{\boldsymbol{\Sigma}}_{\mathsf{crs}} := \sum_{1\le l\le d_2} \mathbb{E}\Big[\big(\boldsymbol{A}^{\star}_{:,l}\boldsymbol{E}^{\top}_{:,l} - \widehat{\boldsymbol{D}}_l\big)\big(\boldsymbol{A}^{\star}_{:,l}\boldsymbol{E}^{\top}_{:,l} - \widehat{\boldsymbol{D}}_l\big)^{\top}\Big];$$
$$\widetilde{\boldsymbol{\Sigma}}_{\mathsf{crs}} := \sum_{1\le l\le d_2} \mathbb{E}\Big[\big(\boldsymbol{A}^{\star}_{:,l}\boldsymbol{E}^{\top}_{:,l} - \widehat{\boldsymbol{D}}_l\big)^{\top}\big(\boldsymbol{A}^{\star}_{:,l}\boldsymbol{E}^{\top}_{:,l} - \widehat{\boldsymbol{D}}_l\big)\Big].$$

Straightforward computation shows that

$$\big(\widehat{\boldsymbol{\Sigma}}_{\mathsf{crs}}\big)_{i,i} = \sum_{1\le l\le d_2} A^{\star 2}_{i,l}\mathbb{E}\left[\|\boldsymbol{E}_{:,l}\|_2^2 - E^2_{i,l}\right], \qquad i\in[d_1],$$
$$\big(\widehat{\boldsymbol{\Sigma}}_{\mathsf{crs}}\big)_{i,j} = \sum_{1\le l\le d_2} A^{\star}_{i,l}A^{\star}_{j,l}\mathbb{E}\left[\|\boldsymbol{E}_{:,l}\|_2^2 - E^2_{i,l} - E^2_{j,l}\right], \qquad i\neq j,$$

and $\widetilde{\boldsymbol{\Sigma}}_{\mathsf{crs}} \in \mathbb{R}^{d_1\times d_1}$ is a diagonal matrix with entries

$$\big(\widetilde{\boldsymbol{\Sigma}}_{\mathsf{crs}}\big)_{i,i} = \sum_{1\le l\le d_2} \left(\left\|\boldsymbol{A}^{\star}_{:,l}\right\|_2^2 - A^{\star 2}_{i,l}\right)\mathbb{E}\left[E^2_{i,l}\right], \qquad i\in[d_1].$$

Hence we have

$$\big\|\widetilde{\boldsymbol{\Sigma}}_{\mathsf{crs}}\big\| \le \max_{1\le i\le d_1}\big|\big(\widetilde{\boldsymbol{\Sigma}}_{\mathsf{crs}}\big)_{i,i}\big| \lesssim \sigma^2_{\mathsf{row}}\left\|\boldsymbol{A}^{\star\top}\right\|^2_{2,\infty}. \tag{68}$$

To control the spectral norm of $\widehat{\boldsymbol{\Sigma}}_{\mathsf{crs}}$, we further decompose it as $\widehat{\boldsymbol{\Sigma}}_{\mathsf{crs}} = \widehat{\boldsymbol{\Sigma}}'_{\mathsf{crs}} - \widehat{\boldsymbol{\Sigma}}''_{\mathsf{crs}}$, where

$$\big(\widehat{\boldsymbol{\Sigma}}'_{\mathsf{crs}}\big)_{i,i} = \sum_{1\le l\le d_2} A^{\star 2}_{i,l}\mathbb{E}\left[\|\boldsymbol{E}_{:,l}\|_2^2\right], \quad i\in[d_1],$$
$$\big(\widehat{\boldsymbol{\Sigma}}'_{\mathsf{crs}}\big)_{i,j} = \sum_{1\le l\le d_2} A^{\star}_{i,l}A^{\star}_{j,l}\mathbb{E}\left[\|\boldsymbol{E}_{:,l}\|_2^2\right], \quad i\neq j,$$

and

$$\big(\widehat{\boldsymbol{\Sigma}}''_{\mathsf{crs}}\big)_{i,i} = \sum_{1\le l\le d_2} A^{\star 2}_{i,l}\mathbb{E}\left[E^2_{i,l}\right], \quad i\in[d_1],$$
$$\big(\widehat{\boldsymbol{\Sigma}}''_{\mathsf{crs}}\big)_{i,j} = \sum_{1\le l\le d_2} A^{\star}_{i,l}A^{\star}_{j,l}\mathbb{E}\left[E^2_{i,l} + E^2_{j,l}\right], \quad i\neq j.$$

- The spectral norm of $\widehat{\boldsymbol{\Sigma}}'_{\mathsf{crs}}$ can be easily upper bounded by

$$\big\|\widehat{\boldsymbol{\Sigma}}'_{\mathsf{crs}}\big\| \le \max_{1\le l\le d_2}\mathbb{E}\left[\|\boldsymbol{E}_{:,l}\|_2^2\right]\left\|\boldsymbol{A}^{\star}\boldsymbol{A}^{\star\top}\right\| \le \sigma^2_{\mathsf{col}}\left\|\boldsymbol{A}^{\star}\right\|^2. \tag{69}$$

- Regarding $\widehat{\boldsymbol{\Sigma}}''_{\mathsf{crs}}$, we first decompose $\widehat{\boldsymbol{\Sigma}}''_{\mathsf{crs}} + \mathcal{P}_{\mathsf{diag}}\big(\widehat{\boldsymbol{\Sigma}}''_{\mathsf{crs}}\big) = \boldsymbol{B}_1 + \boldsymbol{B}_2$, where the diagonal entries of $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are identical and equal to $\sum_{1\le l\le d_2} A^{\star 2}_{i,l}\mathbb{E}[E^2_{i,l}], (1\le i\le d_2)$ while their off-diagonal parts are given by

$$(\boldsymbol{B}_1)_{i,j} = \sum_{1\le l\le d_2} A^{\star}_{i,l}\mathbb{E}\left[E^2_{i,l}\right]A^{\star}_{j,l} \qquad \text{and} \qquad (\boldsymbol{B}_2)_{i,j} = \sum_{1\le l\le d_2} A^{\star}_{i,l}A^{\star}_{j,l}\mathbb{E}\left[E^2_{j,l}\right], \qquad i\neq j.$$

  Let $\boldsymbol{C}$ be a matrix in $\mathbb{R}^{d_1\times d_2}$ with entries $C_{i,j} = A^{\star}_{i,j}\mathbb{E}\left[E^2_{i,j}\right]$. One can easily check that $\boldsymbol{B}_1 = \sum_{1\le l\le d_2}\boldsymbol{C}_{:,l}\boldsymbol{A}^{\star\top}_{:,l} = \boldsymbol{C}\boldsymbol{A}^{\star\top}$ and develop an upper bound

$$\|\boldsymbol{B}_1\| \le \|\boldsymbol{C}\|\,\|\boldsymbol{A}^{\star}\| \le \|\boldsymbol{C}\|_{\mathrm{F}}\,\|\boldsymbol{A}^{\star}\| \le \sigma^2_{\infty}\,\|\boldsymbol{A}^{\star}\|_{\mathrm{F}}\,\|\boldsymbol{A}^{\star}\|.$$

  Note that the same bound also holds for $\boldsymbol{B}_2$. Therefore, we arrive at

$$\begin{aligned}
\big\|\widehat{\boldsymbol{\Sigma}}''_{\mathsf{crs}}\big\| &\le \big\|\mathcal{P}_{\mathsf{diag}}\big(\widehat{\boldsymbol{\Sigma}}''_{\mathsf{crs}}\big)\big\| + \big\|\mathcal{P}_{\mathsf{diag}}\big(\widehat{\boldsymbol{\Sigma}}''_{\mathsf{crs}}\big) + \widehat{\boldsymbol{\Sigma}}''_{\mathsf{crs}}\big\| \le \big\|\mathcal{P}_{\mathsf{diag}}\big(\widehat{\boldsymbol{\Sigma}}''_{\mathsf{crs}}\big)\big\| + \|\boldsymbol{B}_1\| + \|\boldsymbol{B}_2\| \\
&\le \sigma^2_{\infty}\,\|\boldsymbol{A}^{\star}\|^2_{2,\infty} + \sigma^2_{\infty}\,\|\boldsymbol{A}^{\star}\|_{\mathrm{F}}\,\|\boldsymbol{A}^{\star}\| \\
&\le \sigma^2_{\infty}\,\|\boldsymbol{A}^{\star}\|^2_{2,\infty} + \sigma^2_{\infty}\sqrt{r}\,\|\boldsymbol{A}^{\star}\|^2 \\
&\le 2\sigma^2_{\infty}\sqrt{r}\,\|\boldsymbol{A}^{\star}\|^2,
\end{aligned}$$

where we have used the facts that $\|\boldsymbol{A}^\star\|_{2,\infty} \le \|\boldsymbol{A}^\star\|$ and $\|\boldsymbol{A}^\star\|_{\mathrm{F}} \le \sqrt{r}\,\|\boldsymbol{A}^\star\|$. Consequently, the above bounds taken collectively yield

$$\big\|\widehat{\boldsymbol{\Sigma}}_{\mathsf{crs}}\big\| \le \big\|\widehat{\boldsymbol{\Sigma}}_{\mathsf{crs}}'\big\| + \big\|\widehat{\boldsymbol{\Sigma}}_{\mathsf{crs}}''\big\| \lesssim \left(\sigma_{\mathsf{col}}^2 + \sigma_\infty^2\sqrt{r}\right)\|\boldsymbol{A}^\star\|^2 \asymp \sigma_{\mathsf{col}}^2\,\|\boldsymbol{A}^\star\|^2\,, \tag{70}$$

where the last step uses (54b) and (54d).

Putting (68), (69) and (70) together yields

$$V_{\mathsf{crs}} := \max\left\{\big\|\widehat{\boldsymbol{\Sigma}}_{\mathsf{crs}}\big\|, \big\|\widetilde{\boldsymbol{\Sigma}}_{\mathsf{crs}}\big\|\right\} \lesssim \sigma_{\mathsf{col}}^2\,\|\boldsymbol{A}^\star\|^2 + \sigma_{\mathsf{row}}^2\,\big\|\boldsymbol{A}^{\star\top}\big\|_{2,\infty}^2\,. \tag{71}$$

Second, we turn to the spectral norm of each summand $\boldsymbol{A}^\star_{:,l}\boldsymbol{E}_{:,l}^\top - \widehat{\boldsymbol{D}}_l$. Recalling the definition that $\widehat{\boldsymbol{D}}_l = \mathsf{diag}\big(A^\star_{1,l}E_{1,l},\dots,A^\star_{d_1,l}E_{d_1,l}\big)$, we can obtain

$$\big\|\boldsymbol{A}^\star_{:,l}\boldsymbol{E}_{:,l}^\top - \widehat{\boldsymbol{D}}_l\big\| \le \big\|\boldsymbol{A}^\star_{:,l}\big\|_2\,\|\boldsymbol{E}_{:,l}\|_2 + \big\|\widehat{\boldsymbol{D}}_l\big\| \le 2\,\big\|\boldsymbol{A}^\star_{:,l}\big\|_2\,\|\boldsymbol{E}_{:,l}\|_2 \le 2\,\big\|\boldsymbol{A}^{\star\top}\big\|_{2,\infty}\|\boldsymbol{E}_{:,l}\|_2\,.$$

Set

$$L_{\mathsf{crs}} := C\sqrt{L_{\mathsf{ns}}}\,\big\|\boldsymbol{A}^{\star\top}\big\|_{2,\infty} \asymp \left(\sigma_{\mathsf{col}} + B\sqrt{\log d}\right)\big\|\boldsymbol{A}^{\star\top}\big\|_{2,\infty}\,, \tag{72}$$

where L_{ns} is defined in (63) and $C>0$ is some sufficiently large universal constant. Then with probability at least $1 - O\left(d^{-20}\right)$, one has

$$\big\|\boldsymbol{A}^\star_{:,l}\boldsymbol{E}_{:,l}^\top - \widehat{\boldsymbol{D}}_l\big\| \le 2\,\big\|\boldsymbol{A}^{\star\top}\big\|_{2,\infty}\|\boldsymbol{E}_{:,l}\|_2 \lesssim L_{\mathsf{crs}},$$

where the last inequality comes from (62).

Third, we need to control

$$R_{\mathsf{crs}} := \mathbb{E}\left[\big\|\boldsymbol{A}^\star_{:,l}\boldsymbol{E}_{:,l}^\top - \widehat{\boldsymbol{D}}_l\big\|\,\mathbb{1}\big\{\big\|\boldsymbol{A}^\star_{:,l}\boldsymbol{E}_{:,l}^\top - \widehat{\boldsymbol{D}}_l\big\| \ge L_{\mathsf{crs}}\big\}\right].$$

From Jensen's inequality and (64), we know that

$$\mathbb{E}\left[\|\boldsymbol{E}_{:,l}\|_2\,\mathbb{1}\big\{\,\|\boldsymbol{E}_{:,l}\|_2 \ge \sqrt{L_{\mathsf{ns}}}\big\}\right] \le \sqrt{\mathbb{E}\left[\|\boldsymbol{E}_{:,l}\|_2^2\,\mathbb{1}\big\{\,\|\boldsymbol{E}_{:,l}\|_2^2 \ge L_{\mathsf{ns}}\big\}\right]} \ll \frac{\sqrt{L_{\mathsf{ns}}}}{d}.$$

By the definition of L_{crs} in (72) and the fact that

$$\left\{\big\|\boldsymbol{A}^\star_{:,l}\big\|_2\,\|\boldsymbol{E}_{:,l}\|_2 \ge L_{\mathsf{crs}}\right\} \subset \left\{\big\|\boldsymbol{A}^{\star\top}\big\|_{2,\infty}\|\boldsymbol{E}_{:,l}\|_2 \ge L_{\mathsf{crs}}\right\} = \left\{\|\boldsymbol{E}_{:,l}\|_2 \ge C\sqrt{L_{\mathsf{ns}}}\right\},$$

one has

$$\begin{aligned}
R_{\mathsf{crs}} &\le \mathbb{E}\left[2\,\big\|\boldsymbol{A}^\star_{:,l}\big\|_2\,\|\boldsymbol{E}_{:,l}\|_2\,\mathbb{1}\big\{2\,\big\|\boldsymbol{A}^\star_{:,l}\big\|_2\,\|\boldsymbol{E}_{:,l}\|_2 \ge L_{\mathsf{crs}}\big\}\right] \\
&\lesssim \big\|\boldsymbol{A}^{\star\top}\big\|_{2,\infty}\,\mathbb{E}\Big[\,\|\boldsymbol{E}_{:,l}\|_2\,\mathbb{1}\Big\{\,\|\boldsymbol{E}_{:,l}\|_2 \ge \frac{C}{2}\sqrt{L_{\mathsf{ns}}}\Big\}\Big] \\
&\ll \frac{1}{d}\,\big\|\boldsymbol{A}^{\star\top}\big\|_{2,\infty}\sqrt{L_{\mathsf{ns}}} \asymp \frac{L_{\mathsf{crs}}}{d}.
\end{aligned} \tag{73}$$

With (71), (72) and (73) in place, we can apply the truncated matrix Bernstein inequality to obtain that, with with probability at least $1 - O\left(d^{-10}\right)$,

$$\begin{aligned}
\big\|\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{A}^\star\boldsymbol{E}^\top\right)\big\| &\le \Big\|\sum\nolimits_{1\le l\le d_2}\big(\boldsymbol{A}^\star_{:,l}\boldsymbol{E}_{:,l}^\top - \widehat{\boldsymbol{D}}_l\big)\Big\| \lesssim d_2R_{\mathsf{crs}} + L_{\mathsf{crs}}\log d + \sqrt{V_{\mathsf{crs}}\log d} \\
&\asymp L_{\mathsf{crs}}\log d + \sqrt{V_{\mathsf{crs}}\log d} \\
&\lesssim \left(\sigma_{\mathsf{col}}\log d + B\log^{3/2}d\right)\big\|\boldsymbol{A}^{\star\top}\big\|_{2,\infty} + \sigma_{\mathsf{row}}\sqrt{\log d}\,\big\|\boldsymbol{A}^{\star\top}\big\|_{2,\infty} + \sigma_{\mathsf{col}}\sqrt{\log d}\,\|\boldsymbol{A}^\star\| \\
&\lesssim \left(\sigma_{\mathsf{row}} + \sigma_{\mathsf{col}}\right)\log d\,\big\|\boldsymbol{A}^{\star\top}\big\|_{2,\infty} + \sigma_{\mathsf{col}}\sqrt{\log d}\,\|\boldsymbol{A}^\star\|
\end{aligned} \tag{74}$$

under the condition (55a) that $B\sqrt{\log d} \lesssim \sqrt{\sigma_{\mathsf{row}}\sigma_{\mathsf{col}}} \le \max\{\sigma_{\mathsf{row}},\sigma_{\mathsf{col}}\}$.

10.2.3 Step 3: combining Step 1 and Step 2

Taking together (65), (74) and (58), we conclude that

$$\begin{aligned}\|\boldsymbol{G}-\boldsymbol{G}^{\star}\| &\lesssim (\sigma_{\mathsf{row}}+\sigma_{\mathsf{col}})\left(\sigma_{\mathsf{col}}+\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\right)\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\,\|\boldsymbol{A}^{\star}\|+\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{A}^{\star}\boldsymbol{A}^{\star\top}\right)\right\| \\ &\lesssim (\sigma_{\mathsf{row}}+\sigma_{\mathsf{col}})\left(\sigma_{\mathsf{col}}+\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\right)\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\,\|\boldsymbol{A}^{\star}\|+\|\boldsymbol{A}^{\star}\|_{2,\infty}^{2},\end{aligned}$$

where we have used the basic property $\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{A}^{\star}\boldsymbol{A}^{\star\top}\right)\right\|=\|\boldsymbol{A}^{\star}\|_{2,\infty}^{2}$.

10.3 Proof of Lemma 2

We first claim that, for any fixed matrix $\boldsymbol{W}$, with probability at least $1-O(d^{-10})$, the following holds for any $1\leq m\leq d_1$:

$$\begin{aligned}\left\|(\boldsymbol{G}-\boldsymbol{G}^{\star})_{m,:}\boldsymbol{W}\right\|_2 \lesssim &\left(\sigma_{\mathsf{col}}\left(\sigma_{\mathsf{row}}+\|\boldsymbol{A}^{\star}\|_{2,\infty}\right)\sqrt{\log d}+B\log d\,\|\boldsymbol{A}^{\star}\|_{\infty}+\|\boldsymbol{A}^{\star}\|_{2,\infty}^{2}\right)\|\boldsymbol{W}\|_{2,\infty} \\ &+\sigma_{\mathsf{row}}\sqrt{\log d}\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\|\boldsymbol{W}\|.\end{aligned} \tag{75}$$

In particular, taking $\boldsymbol{W}=\boldsymbol{U}^{\star}$ gives

$$\left\|(\boldsymbol{G}-\boldsymbol{G}^{\star})_{m,:}\boldsymbol{U}^{\star}\right\|_2\lesssim\delta_{\mathsf{row}}\sqrt{\frac{\mu r}{d_1}},$$

where

$$\delta_{\mathsf{row}}:=\sigma_{\mathsf{col}}\left(\sigma_{\mathsf{row}}+\|\boldsymbol{A}^{\star}\|_{2,\infty}\right)\sqrt{\log d}+B\log d\,\|\boldsymbol{A}^{\star}\|_{\infty}+\sqrt{\frac{d_1\log d}{\mu r}}\,\sigma_{\mathsf{row}}\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}+\|\boldsymbol{A}^{\star}\|_{2,\infty}^{2}. \tag{76}$$

Using the values of $B,\sigma_{\infty},\sigma_{\mathsf{row}}$ and σ_{col} specified in (54), one can easily verify that

$$\delta_{\mathsf{row}}\lesssim\zeta_{\mathsf{op}}+\|\boldsymbol{A}^{\star}\|_{2,\infty}^{2}, \tag{77}$$

where ζ_{op} is defined in (5). This leads to the advertised bound.

The rest of the proof is thus devoted to proving the claim (75). Recall the definitions of $\boldsymbol{G}$ and $\boldsymbol{G}^{\star}$ in (3.3) and (4). For any $m,i\in[d_1]$, we can expand

$$\begin{aligned}(\boldsymbol{G}-\boldsymbol{G}^{\star})_{m,i}&=\left\langle\boldsymbol{A}^{\mathsf{s}}_{m,:},\boldsymbol{A}^{\mathsf{s}}_{i,:}\right\rangle-\left\langle\boldsymbol{A}^{\star}_{m,:},\boldsymbol{A}^{\star}_{i,:}\right\rangle=\langle\boldsymbol{E}_{m,:},\boldsymbol{E}_{i,:}\rangle+\left\langle\boldsymbol{A}^{\star}_{m,:},\boldsymbol{E}_{i,:}\right\rangle+\left\langle\boldsymbol{E}_{m,:},\boldsymbol{A}^{\star}_{i,:}\right\rangle,\quad i\neq m;\\ (\boldsymbol{G}-\boldsymbol{G}^{\star})_{m,m}&=-G^{\star}_{m,m}=-\left\|\boldsymbol{A}^{\star}_{m,:}\right\|_2^2.\end{aligned}$$

This allows us to derive

$$\begin{aligned}\left\|(\boldsymbol{G}-\boldsymbol{G}^{\star})_{m,:}\boldsymbol{W}\right\|_2\leq&\left\|\sum\nolimits_{i:i\neq m}\langle\boldsymbol{E}_{m,:},\boldsymbol{E}_{i,:}\rangle\boldsymbol{W}_{i,:}\right\|_2+\left\|\sum\nolimits_{i:i\neq m}\left\langle\boldsymbol{A}^{\star}_{m,:},\boldsymbol{E}_{i,:}\right\rangle\boldsymbol{W}_{i,:}\right\|_2\\ &+\left\|\sum\nolimits_{i:i\neq m}\left\langle\boldsymbol{E}_{m,:},\boldsymbol{A}^{\star}_{i,:}\right\rangle\boldsymbol{W}_{i,:}\right\|_2+\left\|G^{\star}_{m,m}\boldsymbol{W}_{m,:}\right\|_2.\end{aligned} \tag{78}$$

We shall control each of these four terms separately.

- For the first term on the right-hand side of (78), we know that

$$\sum\nolimits_{i:i\neq m}\langle\boldsymbol{E}_{m,:},\boldsymbol{E}_{i,:}\rangle\boldsymbol{W}_{i,:}=\sum\nolimits_{(i,j):i\neq m}E_{m,j}E_{i,j}\boldsymbol{W}_{i,:}$$

  is a sum of independent zero-mean random vectors conditional on $\{E_{m,j}\}_{j\in[d_2]}$. In view of the matrix Bernstein inequality, it suffices to control the following two quantities

$$L_1:=\max_{(i,j):i\neq m}\|E_{m,j}E_{i,j}\boldsymbol{W}_{i,:}\|_2\leq\max_{i,j}|E_{i,j}|^2\|\boldsymbol{W}\|_{2,\infty}\leq B^2\|\boldsymbol{W}\|_{2,\infty},$$

$$V_1 := \sum_{(i,j):i\neq m} E_{m,j}^2 \mathbb{E}\left[E_{i,j}^2\right] \|\boldsymbol{W}_{i,:}\|_2^2 \le \|\boldsymbol{W}\|_{2,\infty}^2 \sum_j E_{m,j}^2 \sum_i \mathbb{E}\left[E_{i,j}^2\right] \le \|\boldsymbol{W}\|_{2,\infty}^2 \sigma_{\mathsf{col}}^2 \sum_j E_{m,j}^2,$$

where B and σ_{col} are defined in (54). According to Lemma 12, the following holds with probability at least $1 - O\left(d^{-12}\right)$,

$$V_1 \lesssim \sigma_{\mathsf{col}}^2 \left(\sigma_{\mathsf{row}}^2 + B^2 \log d\right) \|\boldsymbol{W}\|_{2,\infty}^2 \asymp \sigma_{\mathsf{col}}^2 \sigma_{\mathsf{row}}^2 \|\boldsymbol{W}\|_{2,\infty}^2 ,$$

where we use the condition (55a) (namely, $B \lesssim \sigma_{\mathsf{row}}/\sqrt{\log d}$). Apply the matrix Bernstein inequality to demonstrate that with probability exceeding $1 - O(d^{-12})$,

$$\begin{aligned} \Big\| \sum_{i:i\neq m} \langle \boldsymbol{E}_{m,:}, \boldsymbol{E}_{i,:}\rangle \boldsymbol{W}_{i,:} \Big\|_2 &\lesssim L_1 \log d + \sqrt{V_1 \log d} \lesssim \left(B^2 \log d + \sigma_{\mathsf{col}}\sigma_{\mathsf{row}}\sqrt{\log d}\right) \|\boldsymbol{W}\|_{2,\infty} \\ &\asymp \sigma_{\mathsf{col}}\sigma_{\mathsf{row}}\sqrt{\log d}\, \|\boldsymbol{W}\|_{2,\infty} , \end{aligned} \tag{79}$$

where the last line follows from (55a) (i.e. $B \lesssim \sqrt{\sigma_{\mathsf{row}}\sigma_{\mathsf{col}}/\log d}$).

- Regarding the second term on the right-hand side of (78), apply the same argument as above to show that

$$\Big\| \sum_{i:i\neq m} \langle \boldsymbol{A}^\star_{m,:}, \boldsymbol{E}_{i,:}\rangle \boldsymbol{W}_{i,:} \Big\|_2 \lesssim \left(\sigma_{\mathsf{col}}\sqrt{\log d}\, \|\boldsymbol{A}^\star\|_{2,\infty} + B \log d\, \|\boldsymbol{A}^\star\|_\infty \right) \|\boldsymbol{W}\|_{2,\infty} \tag{80}$$

  holds with probability at least $1 - O\left(d^{-12}\right)$.

- Turning to the third term on the right-hand side of (78), we have

$$\Big\| \sum_{i:i\neq m} \langle \boldsymbol{E}_{m,:}, \boldsymbol{A}^\star_{i,:}\rangle \boldsymbol{W}_{i,:} \Big\|_2 \le \Big\| \sum_{1\le j\le d_2} E_{m,j} \left(\boldsymbol{A}^{\star\top}\boldsymbol{W}\right)_{j,:} \Big\|_2,$$

  where the summands are independent zero-mean random vectors. Let us compute that

$$\begin{aligned} L_2 &:= \max_{j\in[d_2]} \left\| E_{m,j} \left(\boldsymbol{A}^{\star\top}\boldsymbol{W}\right)_{j,:} \right\|_2 \le B \left\|\boldsymbol{A}^{\star\top}\boldsymbol{W}\right\|_{2,\infty} \le B \left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty} \|\boldsymbol{W}\| ; \\ V_2 &= \sum_j \mathbb{E}\left[E_{m,j}^2\right] \left\| \left(\boldsymbol{A}^{\star\top}\boldsymbol{W}\right)_{j,:} \right\|_2^2 \le \sigma_{\mathsf{row}}^2 \left\|\boldsymbol{A}^{\star\top}\boldsymbol{W}\right\|_{2,\infty}^2 \le \sigma_{\mathsf{row}}^2 \left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}^2 \|\boldsymbol{W}\|^2 . \end{aligned}$$

  Then the matrix Bernstein inequality reveals that with probability exceeding $1 - O\left(d^{-12}\right)$,

$$\begin{aligned} \Big\| \sum_j E_{m,j} \left(\boldsymbol{A}^{\star\top}\boldsymbol{W}\right)_{j,:} \Big\|_2 &\lesssim L_2 \log d + \sqrt{V_2 \log d} \lesssim \left(B \log d + \sigma_{\mathsf{row}}\sqrt{\log d}\right) \left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty} \|\boldsymbol{W}\| \\ &\asymp \sigma_{\mathsf{row}}\sqrt{\log d} \left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty} \|\boldsymbol{W}\| , \end{aligned} \tag{81}$$

  where the last line follows from the condition (55a) (i.e. $B \lesssim \sigma_{\mathsf{row}}/\sqrt{\log d}$).

- The last term on the right-hand side of (78) can simply be upper bounded by

$$\left\| G^\star_{m,m} \boldsymbol{W}_{m,:} \right\|_2 \le \|\boldsymbol{A}^\star\|_{2,\infty}^2 \|\boldsymbol{W}\|_{2,\infty} . \tag{82}$$

Putting (79), (80), (81) and (82) together yields

$$\begin{aligned} \left\| \left(\boldsymbol{G} - \boldsymbol{G}^\star\right)_{m,:} \boldsymbol{W} \right\|_2 \lesssim & \left(\sigma_{\mathsf{col}}\left(\sigma_{\mathsf{row}} + \|\boldsymbol{A}^\star\|_{2,\infty}\right)\sqrt{\log d} + B \log d\, \|\boldsymbol{A}^\star\|_\infty + \|\boldsymbol{A}^\star\|_{2,\infty}^2\right) \|\boldsymbol{W}\|_{2,\infty} \\ & + \sigma_{\mathsf{row}}\sqrt{\log d} \left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty} \|\boldsymbol{W}\| \end{aligned}$$

as claimed.

10.4 Proof of Lemma 3

We claim for the moment that

$$\left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{G}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\right\|_{2,\infty}\lesssim\frac{\left(\delta_{\mathsf{op}}+\delta_{\mathsf{loo}}\right)\kappa^{2}}{\sigma_{r}^{\star2}}\left(\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty}+\sqrt{\frac{\mu r}{d_{1}}}\right),\tag{83}$$

where δ_{op} is defined in (56), and δ_{loo} is defined as follows:

$$\delta_{\mathsf{loo}}:=\sigma_{\mathsf{col}}\sigma_{\mathsf{row}}\log d+\sigma_{\mathsf{col}}\left\|\boldsymbol{A}^{\star}\right\|\sqrt{\log d}.\tag{84}$$

Using the values of σ_{row} and σ_{col} specified in (54), one can easily see that $\delta_{\mathsf{loo}}\lesssim\zeta_{\mathsf{op}}$, where ζ_{op} is defined in (5). In addition, recall that we have already shown that $\delta_{\mathsf{op}}\lesssim\zeta_{\mathsf{op}}+\|\boldsymbol{A}^{\star}\|_{2,\infty}^{2}$. Putting these together establishes the lemma.

We now start to prove the claim (83). To this end, consider an arbitrary $m\in[d_1]$. In view of [AFWZ17, Lemma 1], we can decompose

$$\left\|\left(\boldsymbol{U}\boldsymbol{H}-\boldsymbol{G}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\right)_{m,:}\right\|_{2}\lesssim\frac{1}{\sigma_{r}^{\star4}}\left\|\boldsymbol{G}-\boldsymbol{G}^{\star}\right\|\left\|\boldsymbol{G}_{m,:}\boldsymbol{U}^{\star}\right\|_{2}+\frac{1}{\sigma_{r}^{\star2}}\left\|\boldsymbol{G}_{m,:}\left(\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\star}\right)\right\|_{2}.\tag{85}$$

- To bound the first term of (85), we apply Lemma 2, (77) and Fact 1 to reach

$$\left\|\left(\boldsymbol{G}-\boldsymbol{G}^{\star}\right)\boldsymbol{U}^{\star}\right\|_{2,\infty}\lesssim\delta_{\mathsf{row}}\sqrt{\frac{\mu r}{d_{1}}}\ll\sigma_{r}^{\star2}\sqrt{\frac{\mu r}{d_{1}}}.$$

  The triangle inequality then gives

$$\left\|\boldsymbol{G}\boldsymbol{U}^{\star}\right\|_{2,\infty}\leq\left\|\left(\boldsymbol{G}-\boldsymbol{G}^{\star}\right)\boldsymbol{U}^{\star}\right\|_{2,\infty}+\left\|\boldsymbol{G}^{\star}\right\|\left\|\boldsymbol{U}^{\star}\right\|_{2,\infty}\lesssim\sigma_{1}^{\star2}\sqrt{\frac{\mu r}{d_{1}}}.\tag{86}$$

  This taken collectively with the upper bound on $\|\boldsymbol{G}-\boldsymbol{G}^{\star}\|$ (cf. (56)) gives

$$\frac{1}{\sigma_{r}^{\star4}}\left\|\boldsymbol{G}-\boldsymbol{G}^{\star}\right\|\left\|\boldsymbol{G}_{m,:}\boldsymbol{U}^{\star}\right\|_{2}\leq\frac{1}{\sigma_{r}^{\star4}}\left\|\boldsymbol{G}-\boldsymbol{G}^{\star}\right\|\left\|\boldsymbol{G}\boldsymbol{U}^{\star}\right\|_{2,\infty}\lesssim\frac{\delta_{\mathsf{op}}\kappa^{2}}{\sigma_{r}^{\star2}}\sqrt{\frac{\mu r}{d_{1}}}.\tag{87}$$

- Turning to the second term of (85), we start with the following bound

$$\left\|\boldsymbol{G}_{m,:}\left(\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\star}\right)\right\|_{2}\leq\left\|\boldsymbol{G}_{m,:}\big(\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big)\right\|_{2}+\left\|\boldsymbol{G}_{m,:}\big(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^{\star}\big)\right\|_{2}.$$

  Lemma 2 tells us that

$$\left\|\boldsymbol{G}_{m,:}\right\|_{2}\leq\left\|\left(\boldsymbol{G}-\boldsymbol{G}^{\star}\right)_{m,:}\right\|_{2}+\left\|\boldsymbol{G}^{\star}\right\|_{2,\infty}\lesssim\sigma_{1}^{\star2}\left\|\boldsymbol{U}^{\star}\right\|_{2,\infty}\leq\sigma_{1}^{\star2}\left\|\boldsymbol{U}^{\star}\right\|=\sigma_{1}^{\star2},$$

  which makes use of the fact that $\|\boldsymbol{G}^{\star}\|_{2,\infty}\leq\|\boldsymbol{U}^{\star}\|_{2,\infty}\|\boldsymbol{\Sigma}^{\star}\|^{2}\|\boldsymbol{U}^{\star\top}\|=\sigma_{1}^{\star2}\|\boldsymbol{U}^{\star}\|_{2,\infty}$. This combined with (90) (to be established shortly in the proof of Lemma 4) and the definitions of $\boldsymbol{H}$ and $\boldsymbol{H}^{(m)}$ gives

$$\begin{aligned}\left\|\boldsymbol{G}_{m,:}\big(\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big)\right\|_{2}&\leq\left\|\boldsymbol{G}_{m,:}\right\|_{2}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}\boldsymbol{H}\right\|\\&=\left\|\boldsymbol{G}_{m,:}\right\|_{2}\left\|\big(\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}\boldsymbol{U}^{\top}\big)\boldsymbol{U}^{\star}\right\|\\&\leq\left\|\boldsymbol{G}_{m,:}\right\|_{2}\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}\boldsymbol{U}^{\top}\right\|\\&\lesssim\frac{\sigma_{1}^{\star2}}{\sigma_{r}^{\star2}}\delta_{\mathsf{loo}}\left(\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty}+\sqrt{\frac{\mu r}{d_{1}}}\right)\\&\leq\delta_{\mathsf{loo}}\kappa^{2}\left(\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty}+\sqrt{\frac{\mu r}{d_{1}}}\right).\end{aligned}\tag{88}$$

  In addition, Lemma 10 shows that with probability at least $1-O\left(d^{-11}\right)$,

$$\left\|\boldsymbol{G}_{m,:}\big(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^{\star}\big)\right\|_{2}\lesssim\delta_{\mathsf{loo}}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^{\star}\right\|_{2,\infty}+\delta_{\mathsf{op}}\kappa^{2}\sqrt{\frac{\mu r}{d_{1}}}$$

$$
\begin{aligned}
&\le \delta_{\mathsf{loo}} \left\| \boldsymbol{U}^{(m)} \boldsymbol{H}^{(m)} \right\|_{2,\infty} + \delta_{\mathsf{loo}} \left\| \boldsymbol{U}^{\star} \right\|_{2,\infty} + \delta_{\mathsf{op}} \kappa^2 \sqrt{\frac{\mu r}{d_1}} \\
&\lesssim \delta_{\mathsf{loo}} \left\| \boldsymbol{U} \boldsymbol{H} \right\|_{2,\infty} + \left(\delta_{\mathsf{loo}} + \delta_{\mathsf{op}} \kappa^2 \right) \sqrt{\frac{\mu r}{d_1}}, \qquad (89)
\end{aligned}
$$

where the inequality (89) results from (100) (also established shortly in the proof of Lemma 4).

Then claim immediately follows from (87), (88), (89) and the union bound.

10.5 Proof of Lemma 4

To begin with, recalling the definition of δ_{loo} (cf. (84)), we claim that

$$
\left\| \boldsymbol{U}^{(m)} \boldsymbol{U}^{(m)\top} - \boldsymbol{U}\boldsymbol{U}^{\top} \right\|_{\mathrm{F}} = \frac{1}{\sigma_r^{\star 2}} \underbrace{\left(\sigma_{\mathsf{col}} \sigma_{\mathsf{row}} \log d + \sigma_{\mathsf{col}} \left\| \boldsymbol{A}^{\star} \right\| \sqrt{\log d} \right)}_{= \delta_{\mathsf{loo}}} \left(\left\| \boldsymbol{U}\boldsymbol{H} \right\|_{2,\infty} + \sqrt{\frac{\mu r}{d_1}} \right). \qquad (90)
$$

As mentioned before, one has $\delta_{\mathsf{loo}} \lesssim \zeta_{\mathsf{op}}$, from which the lemma follows immediately. The rest of the proof thus boils down to proving the claim (90).

We shall apply the Davis-Kahan $\sin \boldsymbol{\Theta}$ theorem [DK70] to derive

$$
\left\| \boldsymbol{U}\boldsymbol{U}^{\top} - \boldsymbol{U}^{(m)} \boldsymbol{U}^{(m)\top} \right\|_{\mathrm{F}} \le \frac{\left\| \left(\boldsymbol{G} - \boldsymbol{G}^{(m)} \right) \boldsymbol{U}^{(m)} \right\|_{\mathrm{F}}}{\lambda_r\left(\boldsymbol{G}^{(m)} \right) - \lambda_{r+1}\left(\boldsymbol{G} \right)} \le \frac{2 \left\| \left(\boldsymbol{G} - \boldsymbol{G}^{(m)} \right) \boldsymbol{U}^{(m)} \right\|_{\mathrm{F}}}{\sigma_r^{\star 2}}. \qquad (91)
$$

Here, the last inequality follows since, by Weyl's inequality,

$$
\begin{aligned}
\lambda_r\left(\boldsymbol{G}^{(m)} \right) - \lambda_{r+1}\left(\boldsymbol{G} \right) &\ge \lambda_r\left(\boldsymbol{G}^{\star} \right) - \left\| \boldsymbol{G}^{(m)} - \boldsymbol{G}^{\star} \right\| - \lambda_{r+1}\left(\boldsymbol{G}^{\star} \right) - \left\| \boldsymbol{G} - \boldsymbol{G}^{\star} \right\| \\
&= \sigma_r^{\star 2} - \left\| \boldsymbol{G}^{(m)} - \boldsymbol{G}^{\star} \right\| - \left\| \boldsymbol{G} - \boldsymbol{G}^{\star} \right\| \\
&\ge \sigma_r^{\star 2}/2, \qquad (92)
\end{aligned}
$$

where the last line follows since $\left\| \boldsymbol{G}^{(m)} - \boldsymbol{G}^{\star} \right\| \lesssim \delta_{\mathsf{op}} \ll \sigma_r^{\star 2}$ — an immediate consequence of Lemma 6 and Condition (55d). As a side note, the fact $\left\| \boldsymbol{G}^{(m)} - \boldsymbol{G}^{\star} \right\| \ll \sigma_r^{\star 2}$ also implies (according to [AFWZ17, Lemma 3])

$$
\left\| \left(\boldsymbol{H}^{(m)} \right)^{-1} \right\| \lesssim 1, \qquad (93)
$$

which will be useful later.

It remains to control the term $\left\| \left(\boldsymbol{G} - \boldsymbol{G}^{(m)} \right) \boldsymbol{U}^{(m)} \right\|_{\mathrm{F}}$ in (91). Recall the definitions of $\boldsymbol{G}$ and $\boldsymbol{G}^{(m)}$ in (3.3) and (2), respectively. It is straightforward to see that $\boldsymbol{G} - \boldsymbol{G}^{(m)}$ is a rank-2 symmetric matrix with nonzero entries located only in the m-th row and the m-th column. Simple calculation reveals that

$$
\begin{aligned}
\left(\boldsymbol{G} - \boldsymbol{G}^{(m)} \right)_{m,i} &= \left\langle \boldsymbol{E}_{m,:}, \boldsymbol{A}^{\mathsf{s}}_{i,:} \right\rangle, \quad i \neq m; \qquad (94) \\
\left(\boldsymbol{G} - \boldsymbol{G}^{(m)} \right)_{m,m} &= 0. \qquad (95)
\end{aligned}
$$

We can then derive

$$
\begin{aligned}
\left\| \left(\boldsymbol{G} - \boldsymbol{G}^{(m)} \right) \boldsymbol{U}^{(m)} \right\|_{\mathrm{F}} &= \left\| \left(\boldsymbol{G} - \boldsymbol{G}^{(m)} \right) \boldsymbol{U}^{(m)} \boldsymbol{H}^{(m)} \left(\boldsymbol{H}^{(m)} \right)^{-1} \right\|_{\mathrm{F}} \le \left\| \left(\boldsymbol{G} - \boldsymbol{G}^{(m)} \right) \boldsymbol{U}^{(m)} \boldsymbol{H}^{(m)} \right\|_{\mathrm{F}} \left\| \left(\boldsymbol{H}^{(m)} \right)^{-1} \right\| \\
&\lesssim \left\| \left(\boldsymbol{G} - \boldsymbol{G}^{(m)} \right) \boldsymbol{U}^{(m)} \boldsymbol{H}^{(m)} \right\|_{\mathrm{F}} \\
&\le \left\| \mathcal{P}_{m,:} \left(\boldsymbol{G} - \boldsymbol{G}^{(m)} \right) \boldsymbol{U}^{(m)} \boldsymbol{H}^{(m)} \right\|_{\mathrm{F}} + \left\| \mathcal{P}_{:,m} \left(\boldsymbol{G} - \boldsymbol{G}^{(m)} \right) \boldsymbol{U}^{(m)} \boldsymbol{H}^{(m)} \right\|_{\mathrm{F}}, \qquad (96)
\end{aligned}
$$

where the second line arises due to (93), and $\mathcal{P}_{m,:}$ (resp. $\mathcal{P}_{:,m}$) is the projection onto the subspace of matrix supported on $\{m\} \times [d_2]$ (resp. $[d_1] \times \{m\}$).

To bound the first term of (96), we make the observation (using (94) and (95)) that

$$
\left\| \mathcal{P}_{m,:} \left(\boldsymbol{G} - \boldsymbol{G}^{(m)} \right) \boldsymbol{U}^{(m)} \boldsymbol{H}^{(m)} \right\|_{\mathrm{F}} = \left\| \left(\boldsymbol{A}^{\mathsf{s}} - \boldsymbol{A}^{\star} \right)_{m,:} \left[\mathcal{P}_{-m,:} \left(\boldsymbol{A}^{\mathsf{s}} \right) \right]^{\top} \boldsymbol{U}^{(m)} \boldsymbol{H}^{(m)} \right\|_{2}.
$$

Controlling this quantity requires the assistance of leave-two-out matrices. Here, we only state our bound: with probability at least $1-O\left(d^{-11}\right)$, one has

$$\begin{aligned}&\big\|\mathcal{P}_{m,:}\big(\boldsymbol{G}-\boldsymbol{G}^{(m)}\big)\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{\mathrm{F}}\\&\quad\lesssim\sigma_{\mathsf{col}}\big(\sigma_{\mathsf{row}}\log d+\|\boldsymbol{A}^{\star}\|\sqrt{\log d}\big)\left(\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{2,\infty}+\sqrt{\frac{\mu r}{d_1}}\right).\end{aligned}\tag{97}$$

This bound will be restated in Lemma 7 and established in Appendix 11. Turning to the second term of (96), we apply Lemma 6 (also established in Appendix 11) to obtain

$$\begin{aligned}\big\|\mathcal{P}_{:,m}\big(\boldsymbol{G}-\boldsymbol{G}^{(m)}\big)\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{\mathrm{F}}&=\Big(\sum\nolimits_{1\le i\le d_1}\big(\boldsymbol{G}-\boldsymbol{G}^{(m)}\big)_{i,m}^2\Big)^{1/2}\Big\|\big(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big)_{m,:}\Big\|_2\\&\le\big\|\boldsymbol{G}-\boldsymbol{G}^{(m)}\big\|\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{2,\infty}\\&\lesssim\big(\sigma_{\mathsf{col}}\big(\sigma_{\mathsf{row}}+\|\boldsymbol{A}^{\star}\|_{2,\infty}\big)\sqrt{\log d}\big)\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{2,\infty}\end{aligned}\tag{98}$$

with probability at least $1-O\left(d^{-11}\right)$. Hence, we can combine (97), (98) and (91) to yield

$$\begin{aligned}&\big\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}\boldsymbol{U}^{\top}\big\|_{\mathrm{F}}\\&\quad\lesssim\left(\sigma_{\mathsf{col}}\sigma_{\mathsf{row}}\log d+\sigma_{\mathsf{col}}\|\boldsymbol{A}^{\star}\|\sqrt{\log d}\right)\left(\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{2,\infty}+\sqrt{\frac{\mu r}{d_1}}\right)\\&\quad=\frac{\delta_{\mathsf{loo}}}{\sigma_r^{\star2}}\left(\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{2,\infty}+\sqrt{\frac{\mu r}{d_1}}\right),\end{aligned}\tag{99}$$

where δ_{loo} is defined in (84). As a result, the proof is complete as long as we can show that

$$\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{2,\infty}\lesssim\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty}+\sqrt{\frac{\mu r}{d_1}}.\tag{100}$$

To finish up, it remains to justify this inequality (100). To this end, from the definitions of $\boldsymbol{H}^{(m)}$ and $\boldsymbol{H}$ we have

$$\begin{aligned}\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{2,\infty}&\le\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}\boldsymbol{H}\big\|_{2,\infty}+\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty}\\&=\big\|\big(\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}\boldsymbol{U}^{\top}\big)\boldsymbol{U}^{\star}\big\|_{2,\infty}+\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty}\\&\le\big\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}\boldsymbol{U}^{\top}\big\|_{\mathrm{F}}\|\boldsymbol{U}^{\star}\|+\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty}\\&=\big\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}\boldsymbol{U}^{\top}\big\|_{\mathrm{F}}+\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty}.\end{aligned}\tag{101}$$

Under the condition (55d), it is easily seen that $\delta_{\mathsf{loo}}\ll\sigma_r^{\star2}$. This together with (99) gives

$$\big\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}\boldsymbol{U}^{\top}\big\|_{\mathrm{F}}\le0.5\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{2,\infty}+0.5\sqrt{\frac{\mu r}{d_1}},\tag{102}$$

which combined with (101) yields

$$\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{2,\infty}\le2\,\|\boldsymbol{U}\boldsymbol{H}\|_{2,\infty}+\sqrt{\frac{\mu r}{d_1}}\tag{103}$$

as claimed.

10.6 Proof of Lemma 5

We start with (55a). In view of the definitions of $B,\sigma_\infty,\sigma_{\mathsf{row}}$ and σ_{col} in (54), we have with probability at least $1-O\left(d^{-12}\right)$,

$$B^2=\frac{\mu r\sigma_1^{\star2}}{d_1d_2p^2}+\frac{\sigma^2\min\left\{\sqrt{d_1d_2},d_2\right\}}{p\log d},$$

$$\sigma_{\mathsf{row}}^2 = \frac{\mu r \sigma_1^{\star 2}}{d_1 p} + \frac{\sigma^2 d_2}{p},$$

$$\sigma_{\mathsf{row}}\sigma_{\mathsf{col}} = \frac{\mu r \sigma_1^{\star 2}}{\sqrt{d_1 d_2}\, p} + \frac{2\sigma\sigma_1^\star\sqrt{\mu r}}{p} + \frac{\sigma^2\sqrt{d_1 d_2}}{p} \asymp \frac{\mu r \sigma_1^{\star 2}}{\sqrt{d_1 d_2}\, p} + \frac{\sigma^2\sqrt{d_1 d_2}}{p},$$

where we have used the AM-GM inequality (i.e. $2\sigma\sigma_1^\star\sqrt{\mu r} \leq \frac{\mu r \sigma_1^{\star 2}}{\sqrt{d_1 d_2}} + \sigma^2\sqrt{d_1 d_2}$) in the last line. Therefore,

$$B^2 \log d \lesssim \sigma_{\mathsf{row}}\sigma_{\mathsf{col}} \qquad \text{and} \qquad B^2 \log d \lesssim \sigma_{\mathsf{row}}^2$$

hold as long as $p \gtrsim (d_1 d_2)^{-1/2}\log d$ and $p \gtrsim d_2^{-1}\log d$.

The next step is to establish (55b). Let us consider the first inequality. By (54), it is easily seen that

$$\sigma_\infty^2 = \frac{\mu r \sigma_1^{\star 2}}{d_1 d_2 p} + \frac{\sigma^2}{p};$$

$$B \log d\, \|\boldsymbol{A}^\star\| \sqrt{\frac{\mu r}{d_2}} = \frac{\mu r \sigma_1^{\star 2}\log d}{\sqrt{d_1}\, d_2 p} + \sigma\sigma_1^\star \min\left\{\sqrt[4]{d_1 d_2},\, \sqrt{d_2}\right\}\sqrt{\frac{\mu r \log d}{d_2 p}}.$$

As a consequence, the first inequality holds as long as $\frac{\sigma}{\sigma_1^\star} \lesssim \min\left\{\sqrt[4]{d_1 d_2},\, \sqrt{d_2}\right\}\sqrt{\frac{\mu r p \log d}{d_2}}$, which is satisfied by our noise assumption that $\frac{\sigma}{\sigma_r^\star} \ll \frac{\sqrt{p}}{\kappa\sqrt[4]{d_1 d_2}\sqrt{\log d}}$. To show the second inequality, we note that

$$B\sqrt{\log d}\, \|\boldsymbol{A}^\star\| \sqrt{\frac{\mu r}{d_2}} \leq \left\{\frac{\sigma_1^{\star 2}\sqrt{\mu r \log d}}{d_2 p} + \sigma\sigma_1^\star\sqrt{\frac{d_1}{p}}\right\}\sqrt{\frac{\mu r}{d_1}}.$$

Recognizing that $\sigma_{\mathsf{col}}\|\boldsymbol{A}^\star\| = \sigma_1^{\star 2}\sqrt{\frac{\mu r}{d_2 p}} + \sigma\sigma_1^\star\sqrt{\frac{d_1}{p}}$, we prove the second inequality provided that $p \gtrsim d_2^{-1}\log d$.

When it comes to (55c): by virtue of Lemma 11 and (54), one has

$$\frac{B\log^{3/2} d\, \|\boldsymbol{A}^\star\|_\infty}{\|\boldsymbol{A}^\star\|^2} \leq \frac{\mu r \log^{3/2} d}{d_1 d_2 p} + \frac{\sigma\sqrt{\mu r}\log d}{\sigma_1^\star\sqrt[4]{d_1 d_2}\sqrt{p}}.$$

Consequently, (55c) holds provided $p \gtrsim \frac{\sqrt{\mu r}\log^{3/2} d}{\sqrt{d_1}\, d_2}$ and $\frac{\sigma}{\sigma_1^\star} \lesssim \sqrt[4]{\frac{d_2}{d_1}}\frac{\sqrt{p}}{\log d}$, which holds under our assumptions that $p \gg \frac{\mu\kappa^4 r \log^2 d}{\sqrt{d_1 d_2}}$ and $\frac{\sigma}{\sigma_r^\star} \ll \frac{\sqrt{p}}{\kappa^3\sqrt[4]{d_1 d_2}\sqrt{\log d}}$.

Finally, using the definitions in (54) and Lemma 11, one obtains the following bounds:

$$\sigma_{\mathsf{row}}\sigma_{\mathsf{col}}\log d \asymp \frac{\mu r \sigma_1^{\star 2}\log d}{\sqrt{d_1 d_2}\, p} + \frac{\sigma^2\sqrt{d_1 d_2}\log d}{p},$$

$$\sigma_{\mathsf{col}}\sqrt{\log d}\, \|\boldsymbol{A}^\star\| = \sigma_1^{\star 2}\sqrt{\frac{\mu r \log d}{d_2 p}} + \sigma\sigma_1^\star\sqrt{\frac{d_1 \log d}{p}};$$

$$\|\boldsymbol{A}^\star\|_{2,\infty}^2 \lesssim \frac{\mu_1 r \sigma_1^{\star 2}}{d_1};$$

$$\sigma_{\mathsf{row}}\sqrt{\frac{d_1 \log d}{\mu r}}\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty} \leq \sigma_1^{\star 2}\sqrt{\frac{\mu r \log d}{d_2 p}} + \sigma\sigma_1^\star\sqrt{\frac{d_1 \log d}{p}};$$

$$B\,\|\boldsymbol{A}^\star\|_\infty \log d \leq \frac{\mu r \sigma_1^{\star 2}\log d}{d_1 d_2 p} + \frac{\sigma\sigma_1^\star\sqrt{\mu r \log d}}{\sqrt[4]{d_1 d_2}\sqrt{p}}.$$

Therefore, it is easy to verify (55d) under the assumptions of Theorem 3.1.

11 Proofs for auxiliary lemmas

This section establishes several useful technical lemmas useful for proving our main theorems. Throughout this section, we shall frequently use the quantities $B, \sigma_\infty, \sigma_{\mathsf{row}}$ and σ_{col} defined in (54). In fact, it suffices to bear in mind the following bounds

$$\begin{aligned} B &\geq \max_{i,j} |E_{i,j}|; & \sigma_\infty &\geq \max_{i,j} \sqrt{\mathbb{E}[E_{i,j}^2]}; \\ \sigma_{\mathsf{row}} &\geq \max_i \sqrt{\textstyle\sum_j \mathbb{E}[E_{i,j}^2]}; & \sigma_{\mathsf{col}} &\geq \max_j \sqrt{\textstyle\sum_i \mathbb{E}[E_{i,j}^2]}. \end{aligned} \tag{104}$$

11.1 Auxiliary technical lemmas

We first gather all technical lemmas to be established in this section, and begin with the following lemma, which shows that the leave-one-out sequence $\boldsymbol{G}^{(m)}$ is close to $\boldsymbol{G}$ and $\boldsymbol{G}^\star$ when measured by the spectral norm.

Lemma 6. *Instate the assumptions of Theorem 3.1. With probability at least $1 - O\left(d^{-11}\right)$, one has*

$$\left\|\boldsymbol{G}^{(m)} - \boldsymbol{G}\right\| \lesssim \sigma_{\mathsf{row}}\big(\sigma_{\mathsf{col}} + \left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\big)\sqrt{\log d}, \tag{105}$$

$$\left\|\boldsymbol{G}^{(m)} - \boldsymbol{G}^\star\right\| \lesssim \delta_{\mathsf{op}} = (\sigma_{\mathsf{row}} + \sigma_{\mathsf{col}})\left(\sigma_{\mathsf{col}} + \left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\right)\log d + \sigma_{\mathsf{col}}\sqrt{\log d}\,\|\boldsymbol{A}^\star\| + \|\boldsymbol{A}^\star\|_{2,\infty}^2, \tag{106}$$

where σ_{row} and σ_{col} are defined in (54).

Proof. See Appendix 11.2. ∎

Similar to $\boldsymbol{H}$ (defined in (6)), we also introduce the following matrices for each $(m,l) \in [d_1] \times [d_2]$:

$$\boldsymbol{H}^{(m)} := \boldsymbol{U}^{(m)\top}\boldsymbol{U}^\star, \tag{107a}$$

$$\boldsymbol{H}^{(m,l)} := \boldsymbol{U}^{(m,l)\top}\boldsymbol{U}^\star, \tag{107b}$$

where $\boldsymbol{U}^{(m)}$ and $\boldsymbol{U}^{(m,l)}$ are defined in Algorithm 1 and Algorithm 2, respectively.

Lemma 7 serves a crucial step towards proving Lemma 4.

Lemma 7. *Instate the assumptions of Theorem 3.1. For any fixed $1 \leq m \leq d_1$, with probability at least $1 - O(d^{-11})$, one has*

$$\begin{aligned} &\left\|\left(\boldsymbol{A}^{\mathsf{s}} - \boldsymbol{A}^\star\right)_{m,:}\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)^\top\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_2 \\ &\qquad \lesssim \sigma_{\mathsf{col}}\big(\sigma_{\mathsf{row}}\log d + \|\boldsymbol{A}^\star\|\sqrt{\log d}\big)\left(\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} + \sqrt{\frac{\mu r}{d_1}}\right), \end{aligned}$$

where $\boldsymbol{A}^{\mathsf{s}}$ is defined in (50), and σ_{row} and σ_{col} are both defined in (54).

Proof. See Appendix 11.3. ∎

The proof of Lemma 7 relies on an upper bound on the $\ell_{2,\infty}$ norm of $\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)^\top\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}$, which is formalized below in Lemma 8. This is built upon a leave-two-out argument.

Lemma 8. *Instate the assumptions of Theorem 3.1. With probability at least $1 - O\big(d^{-10}\big)$, the following holds simultaneously for all $m \in [d_1]$,*

$$\left\|\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)^\top\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} \lesssim \big(B\log d + \sigma_{\mathsf{col}}\sqrt{\log d}\big)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} + \|\boldsymbol{A}^\star\|\sqrt{\frac{\mu r}{d_2}},$$

where $\boldsymbol{A}^{\mathsf{s}}$ is defined in (50), and B and σ_{col} are defined in (54).

Proof. See Appendix 11.4. ∎

The proof of Lemma 8 requires the proximity between $\boldsymbol{U}^{(m)}$ and $\boldsymbol{U}^{(m,l)}$, which is demonstrated below in Lemma 9.

Lemma 9. *Instate the assumptions of Theorem 3.1. With probability at least $1-O\left(d^{-10}\right)$, the following holds simultaneously for any $m\in[d_1]$ and $l\in[d_2]$,*

$$\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\|\lesssim\frac{1}{\sigma_r^{\star 2}}\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)^2\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}$$
$$+\frac{\sigma_\infty^2}{\sigma_r^{\star 2}}+\frac{1}{\sigma_r^{\star 2}}\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}, \tag{108}$$

where B,σ_∞ and σ_{col} are defined in (54).

Proof. See Appendix 11.5. ☐

Finally, Lemma 10 stated below constitutes the main part of Lemma 3 (recalling the decomposition in (8) and (9)).

Lemma 10. *Instate the assumptions of Theorem 3.1. For each fixed $m\in[d_1]$, the following holds with probability exceeding $1-O\left(d^{-11}\right)$,*

$$\left\|\boldsymbol{G}_{m,:}\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right)\right\|_2\lesssim\delta_{\mathsf{loo}}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right\|_{2,\infty}+\delta_{\mathsf{op}}\kappa^2\sqrt{\frac{\mu r}{d_1}},$$

where δ_{op} and δ_{loo} are defined in (5) and (84), respectively.

Proof. See Appendix 11.6. ☐

11.2 Proof of Lemma 6

Recall the definitions of $\boldsymbol{G}$ and $\boldsymbol{G}^{(m)}$ in (3.3) and (2). As shown in (94) in the proof of Lemma 4 in Appendix 10.5, we know that $\boldsymbol{G}-\boldsymbol{G}^{(m)}$ is a rank-2 symmetric matrix with nonzero entries located only in the m-th row and the m-th column. In particular, one has

$$\left(\boldsymbol{G}-\boldsymbol{G}^{(m)}\right)_{m,i}=\left\langle\boldsymbol{E}_{m,:},\boldsymbol{A}^{\mathsf{s}}_{i,:}\right\rangle,\qquad i\neq m,$$
$$\left(\boldsymbol{G}-\boldsymbol{G}^{(m)}\right)_{m,m}=0,$$

thus indicating that

$$\left(\boldsymbol{G}-\boldsymbol{G}^{(m)}\right)_{m,:}=\boldsymbol{E}_{m,:}\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^\top.$$

This allows us to upper bound

$$\begin{aligned}\left\|\boldsymbol{G}-\boldsymbol{G}^{(m)}\right\|\leq\left\|\boldsymbol{G}-\boldsymbol{G}^{(m)}\right\|_{\mathrm{F}}&\lesssim\left\|\left(\boldsymbol{G}-\boldsymbol{G}^{(m)}\right)_{m,:}\right\|_2\\&=\left\|\boldsymbol{E}_{m,:}\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^\top\right\|_2\\&\lesssim\sigma_{\mathsf{row}}\left(\sigma_{\mathsf{col}}+\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\right)\sqrt{\log d}.\end{aligned}$$

Here, the last line follows the following. First, notice $\boldsymbol{A}^{\mathsf{s}}=\boldsymbol{A}^\star+\boldsymbol{E}$ and

$$\boldsymbol{E}_{m,:}\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^\star+\boldsymbol{E}\right)\right]^\top=\sum\nolimits_{i:i\neq m}\left\langle\boldsymbol{E}_{m,:},\boldsymbol{E}_{i,:}\right\rangle\boldsymbol{e}_i^\top+\sum\nolimits_{i:i\neq m}\left\langle\boldsymbol{E}_{m,:},\boldsymbol{A}^\star_{i,:}\right\rangle\boldsymbol{e}_i^\top,$$

where $\boldsymbol{e}_i$ is the i-th standard basis in $\mathbb{R}^{d_1}$. It follows from (79) and (81) shown in the proof of Lemma 2 (cf. Appendix 10.3) that with probability at least $1-O(d^{-11})$,

$$\begin{aligned}\left\|\boldsymbol{E}_{m,:}\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^\star+\boldsymbol{E}\right)\right]^\top\right\|_2&\leq\left\|\sum\nolimits_{i:i\neq m}\left\langle\boldsymbol{E}_{m,:},\boldsymbol{E}_{i,:}\right\rangle\boldsymbol{e}_i^\top\right\|_2+\left\|\sum\nolimits_{i:i\neq m}\left\langle\boldsymbol{E}_{m,:},\boldsymbol{A}^\star_{i,:}\right\rangle\boldsymbol{e}_i^\top\right\|_2\\&\lesssim\sigma_{\mathsf{col}}\sigma_{\mathsf{row}}\sqrt{\log d}+\sigma_{\mathsf{row}}\sqrt{\log d}\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}.\end{aligned}$$

In addition, the above bound combined with Lemma 1 immediately yields (106). The proof is complete by taking the union bound over $1\leq m\leq d_1$.

11.3 Proof of Lemma 7

By construction, the m-th row of $\boldsymbol{A}^{\mathsf{s}}-\boldsymbol{A}^{\star}$ is independent of $\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^{\top}\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}$. As a result,

$$\left(\boldsymbol{A}^{\mathsf{s}}-\boldsymbol{A}^{\star}\right)_{m,:}\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^{\top}\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}=\sum_{j\in[d_2]}E_{m,j}\left(\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^{\top}\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right)_{j,:}$$

can be viewed as a sum of independent zero-mean random vectors (where the randomness comes from $\{E_{m,j}\}_{j\in[d_2]}$). It is straightforward to calculate that

$$\begin{aligned}L &:= \max_{j\in[d_2]}\left\|E_{m,j}\left(\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^{\top}\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right)_{j,:}\right\|_2 \leq B\left\|\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)^{\top}\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty},\\ V &:= \sum_{j\in[d_2]}\mathbb{E}\left[E_{m,j}^2\right]\left\|\left(\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^{\top}\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right)_{j,:}\right\|_2^2 \leq \sigma_\infty^2\left\|\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^{\top}\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{\mathrm{F}}^2.\end{aligned}$$

In view of the matrix Bernstein inequality, it boils down to controlling L and V. To this end, let us first bound L. From Lemma 8, one has that with probability at least $1-O\left(d^{-11}\right)$,

$$L \lesssim B\big(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\big)\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{2,\infty}+B\left\|\boldsymbol{A}^{\star}\right\|\sqrt{\frac{\mu r}{d_2}}. \tag{109}$$

Regarding V, Lemma 13 guarantees the following upper bound with probability exceeding $1-O\left(d^{-11}\right)$,

$$\begin{aligned}\left\|\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^{\top}\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{\mathrm{F}} &\leq \left\|\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right\|\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{\mathrm{F}}\\ &\leq \sqrt{d_1}\left\|\boldsymbol{A}^{\mathsf{s}}\right\|\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{2,\infty}\\ &\lesssim \sqrt{d_1}\left(B\log d+\left(\sigma_{\mathsf{row}}+\sigma_{\mathsf{col}}\right)\sqrt{\log d}+\left\|\boldsymbol{A}^{\star}\right\|\right)\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{2,\infty}\\ &\lesssim \sqrt{d_1}\left(\sigma_{\mathsf{row}}\sqrt{\log d}+\left\|\boldsymbol{A}^{\star}\right\|\right)\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{2,\infty},\end{aligned} \tag{110}$$

where the last inequality follows from the condition (55d) that $B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\ll\sigma_r^{\star}$. Applying the matrix Bernstein inequality yields that with probability at least $1-O\left(d^{-11}\right)$: one has

$$\begin{aligned}&\left\|\left(\boldsymbol{A}^{\mathsf{s}}-\boldsymbol{A}^{\star}\right)_{m,:}\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^{\top}\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_2 \lesssim L\log d+\sqrt{V\log d}\\ &\quad\lesssim \left(B^2\log^2 d+B\sigma_{\mathsf{col}}\log^{3/2}d\right)\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{2,\infty}+B\log d\left\|\boldsymbol{A}^{\star}\right\|\sqrt{\frac{\mu r}{d_2}}\\ &\qquad+\sqrt{d_1}\,\sigma_\infty\left(\sigma_{\mathsf{row}}\log d+\left\|\boldsymbol{A}^{\star}\right\|\sqrt{\log d}\right)\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{2,\infty}\\ &\quad\overset{\text{(i)}}{\lesssim} \sigma_{\mathsf{col}}\left(\sigma_{\mathsf{row}}\log d+\left\|\boldsymbol{A}^{\star}\right\|\sqrt{\log d}\right)\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{2,\infty}+\sigma_{\mathsf{col}}\left\|\boldsymbol{A}^{\star}\right\|\sqrt{\log d}\sqrt{\frac{\mu r}{d_1}}\\ &\quad\lesssim \sigma_{\mathsf{col}}\left(\sigma_{\mathsf{row}}\log d+\left\|\boldsymbol{A}^{\star}\right\|\sqrt{\log d}\right)\left(\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_{2,\infty}+\sqrt{\frac{\mu r}{d_1}}\right),\end{aligned} \tag{111}$$

where (i) uses (54) that $\sigma_{\mathsf{col}}^2\asymp d_1\sigma_\infty^2$ as well as the conditions (55a), (55d) and (55b) (namely, $B\lesssim\sqrt{\sigma_{\mathsf{row}}\sigma_{\mathsf{col}}/\log d}$, $B\log d\ll\left\|\boldsymbol{A}^{\star}\right\|$ and $B\log d\sqrt{\mu r/d_2}\lesssim\sigma_{\mathsf{col}}\sqrt{\log d}\sqrt{\mu r/d_1}$).

11.4 Proof of Lemma 8

For notational convenience, we denote

$$\boldsymbol{A}^{\mathsf{s},(m),0}:=\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right).$$

Fix an arbitrary $l\in[d_2]$, and we would like to upper bound $\big\|\boldsymbol{A}_{:,l}^{\mathsf{s},(m),0\top}\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\big\|_2$. The main difficulty here lies in the complicated statistical dependence between $\boldsymbol{A}_{:,l}^{\mathsf{s},(m),0}$ and $\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}$. Recall the definitions of the auxiliary matrices $\boldsymbol{U}^{(m,l)}$ and $\boldsymbol{H}^{(m,l)}$ in Algorithm 2 and (107b), respectively. By construction, $\boldsymbol{A}_{:,l}^{\mathsf{s},(m),0}$ is independent of $\boldsymbol{U}^{(m,l)}$ and $\boldsymbol{H}^{(m,l)}$. Moreover, Lemma 9 guarantees that $\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}$ is extremely close to $\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}$. Thus, invoke the triangle inequality to upper bound

$$\begin{aligned}\left\|\boldsymbol{A}_{:,l}^{\mathsf{s},(m),0\top}\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2} \leq & \underbrace{\left\|\left(\boldsymbol{A}_{:,l}^{\mathsf{s},(m),0}-\mathbb{E}\left[\boldsymbol{A}_{:,l}^{\mathsf{s},(m),0}\right]\right)^{\top}\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2}}_{=:\alpha_{1}}+\underbrace{\left\|\mathbb{E}\left[\boldsymbol{A}_{:,l}^{\mathsf{s},(m),0}\right]^{\top}\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2}}_{=:\alpha_{2}}\\ &+\underbrace{\left\|\boldsymbol{A}_{:,l}^{\mathsf{s},(m),0}\right\|_{2}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|}_{=:\alpha_{3}}.\end{aligned}$$

Before moving on, we make note of the following two useful upper bounds on $\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\|$ based on Lemma 9. On the one hand, one has with probability at least $1-O\left(d^{-13}\right)$,

$$\begin{aligned}\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\| &\lesssim \frac{1}{\sigma_{r}^{\star2}}\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)^{2}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}\\ &\quad+\frac{\sigma_{\infty}^{2}}{\sigma_{r}^{\star2}}+\frac{1}{\sigma_{r}^{\star2}}\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\\ &\overset{\text{(i)}}{\lesssim}\frac{1}{\sigma_{r}^{\star2}}\left(B^{2}\log^{2}d+\sigma_{\mathsf{col}}^{2}\log d+B\log d\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}+\sigma_{\mathsf{col}}\sqrt{\log d}\left\|\boldsymbol{A}^{\star}\right\|\right)\\ &\lesssim\frac{\delta_{\mathsf{op}}}{\sigma_{r}^{\star2}}\ll1. \end{aligned}\tag{112}$$

where (i) follows from the facts that $\sigma_{\infty}\leq\sigma_{\mathsf{col}}$, $\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}\leq\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|\leq1$ and $\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\leq\left\|\boldsymbol{A}^{\star}\right\|$; (112) is due to the definition of σ_{col} in (54d), the definition of δ_{op} (cf. (5)) as well as conditions (55a) and (55d). On the other hand, we can also bound

$$\begin{aligned}\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\| &\overset{\text{(i)}}{\lesssim}\frac{1}{\sigma_{r}^{\star2}}\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)^{2}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}\\ &\quad+\frac{1}{\sigma_{r}^{\star2}}\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{A}^{\star}\right\|\sqrt{\frac{\mu r}{d_{2}}}\\ &=o\left(1\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}+\frac{1}{\sigma_{r}^{\star2}}\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{A}^{\star}\right\|\sqrt{\frac{\mu r}{d_{2}}}\end{aligned}\tag{113}$$

where (i) arises from (55b) and the inequality that $\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\leq\left\|\boldsymbol{A}^{\star}\right\|\left\|\boldsymbol{V}^{\star}\right\|_{2,\infty}\leq\left\|\boldsymbol{A}^{\star}\right\|\sqrt{\mu r/d_{2}}$, and (113) is due to conditions (55a) and (55d). In the sequel, we control the α_i's separately.

- For α_1, it is easy to see that

$$\left(\boldsymbol{A}_{:,l}^{\mathsf{s},(m),0}-\mathbb{E}\left[\boldsymbol{A}_{:,l}^{\mathsf{s},(m),0}\right]\right)^{\top}\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}=\sum_{i:i\neq m}E_{i,l}\left(\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right)_{i,:}$$

  is a sum of independent zero-mean random vectors conditional on $\left\{E_{i,j}\right\}_{i\in[d_1]\setminus\{m\},j\in[d_2]\setminus\{l\}}$. Straightforward calculation gives that

$$\begin{aligned}L&:=\max_{i:i\neq m}\left\|E_{i,l}\left(\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right)_{i,:}\right\|_{2}\leq B\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty},\\ V&:=\sum_{i:i\neq m}\mathbb{E}\left[E_{i,l}^{2}\right]\left\|\left(\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right)_{i,:}\right\|_{2}^{2}\leq\sigma_{\mathsf{col}}^{2}\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}^{2}.\end{aligned}$$

  Then we apply the matrix Bernstein inequality to obtain that with probability at least $1-O\big(d^{-13}\big)$,

$$\left\|\left(\boldsymbol{A}_{:,l}^{\mathsf{s},(m),0}-\mathbb{E}\left[\boldsymbol{A}_{:,l}^{\mathsf{s},(m),0}\right]\right)^{\top}\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2}\lesssim L\log d+\sqrt{V\log d}$$

$$
\begin{aligned}
&\le \left(B\log d + \sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty} \\
&\le \left(B\log d + \sigma_{\mathsf{col}}\sqrt{\log d}\right)\Big(\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} + \left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top} - \boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\|\Big),
\end{aligned} \tag{114}
$$

where the last line results from the following observation:

$$
\begin{aligned}
\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty} &\le \left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} + \left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top} - \boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\|_{2,\infty} \\
&\le \left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} + \left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top} - \boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\|.
\end{aligned} \tag{115}
$$

- Turning to α_2, we obtain the simple upper bound

$$
\left\|\mathbb{E}\left[\boldsymbol{A}_{:,l}^{\mathsf{s},(m),0}\right]^{\top}\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_2 \le \left\|\boldsymbol{A}_{:,l}^{\star}\right\|_2\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\| \le \left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}. \tag{116}
$$

- With regards to α_3, Lemma 12 reveals that with probability at least $1 - O\left(d^{-13}\right)$,

$$
\begin{aligned}
\left\|\boldsymbol{A}_{:,l}^{\mathsf{s},(m),0}\right\|_2\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\| &\le \left\|\boldsymbol{A}_{:,l}\right\|_2\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top} - \boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\| \\
&\lesssim \left(\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty} + B\sqrt{\log d} + \sigma_{\mathsf{col}}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top} - \boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\| \\
&\lesssim \left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty} + \left(B\sqrt{\log d} + \sigma_{\mathsf{col}}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top} - \boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\|,
\end{aligned} \tag{117}
$$

where we use (112) in the last step.

Combining (114), (116), (117) implies that with probability greater than $1 - O\left(d^{-13}\right)$,

$$
\begin{aligned}
&\left\|\boldsymbol{A}_{:,l}^{\mathsf{s},(m),0\top}\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_2 \\
&\quad\lesssim \left(B\log d + \sigma_{\mathsf{col}}\sqrt{\log d}\right)\Big(\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} + \left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top} - \boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\|\Big) + \left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty} \\
&\quad\overset{\text{(i)}}{\lesssim} \left(B\log d + \sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} + \frac{1}{\sigma_r^{\star 2}}\left(B\log d + \sigma_{\mathsf{col}}\sqrt{\log d}\right)^2\left\|\boldsymbol{A}^{\star}\right\|\sqrt{\frac{\mu r}{d_2}} + \left\|\boldsymbol{A}^{\star}\right\|\sqrt{\frac{\mu r}{d_2}} \\
&\quad\overset{\text{(ii)}}{\lesssim} \left(B\log d + \sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} + \left\|\boldsymbol{A}^{\star}\right\|\sqrt{\frac{\mu r}{d_2}},
\end{aligned}
$$

where (i) is by (113) and $\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty} \le \left\|\boldsymbol{A}^{\star}\right\|\sqrt{\mu r/d_2}$, and (ii) follows from conditions (55a) and (55d). The proof is complete by taking the union bound over $1 \le l \le d_2$.

11.5 Proof of Lemma 9

Fix arbitrary $m \in [d_1]$ and $l \in [d_2]$. Recalling the definitions of $\boldsymbol{G}^{(m)}$ and $\boldsymbol{G}^{(m,l)}$ in (2) and (3b), we see that $\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}$ is symmetric with entries

$$
\begin{aligned}
&\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right)_{i,j} = E_{i,l}E_{j,l} + A_{i,l}^{\star}E_{j,l} + E_{i,l}A_{j,l}^{\star}, \qquad i \ne m, j \ne m, i \ne j, \\
&\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right)_{i,m} = A_{m,l}^{\star}E_{i,l}, \qquad i \ne m, \\
&\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right)_{i,i} = 0, \qquad 1 \le i \le d_1.
\end{aligned}
$$

Note that $\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}$ depends only on $\{E_{i,l}\}_{i \in [d_1]\setminus\{m\}}$ and is hence statistically independent of $\boldsymbol{U}^{(m,l)}$ and $\boldsymbol{H}^{(m,l)}$. In particular, we can express

$$
\mathcal{P}_{-m}\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right) = \mathcal{P}_{\mathsf{off\text{-}diag}}\mathcal{P}_{-m}\left(\boldsymbol{E}_{:,l}\boldsymbol{E}_{:,l}^{\top} + \boldsymbol{E}_{:,l}\boldsymbol{A}_{:,l}^{\star\top} + \boldsymbol{A}_{:,l}^{\star}\boldsymbol{E}_{:,l}^{\top}\right),
$$

where $\mathcal{P}_{-m}$ is the projection onto the subspace of matrices supported on $\{(i,j) \in [d_1] \times [d_2] : i \neq m \text{ and } j \neq m\}$ and $\mathcal{P}_{\text{off-diag}}$ extracts the off-diagonal part. In addition,

$$\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right)_{m,:} = A^{\star}_{m,l} \boldsymbol{E}_{:,l}^{\top} - A^{\star}_{m,l} E_{m,l} \boldsymbol{e}_m^{\top},$$

where $\boldsymbol{e}_m$ stands for the m-th standard basis in $\mathbb{R}^{d_1}$.

In the sequel, we shall apply the Davis-Kahan $\sin \boldsymbol{\Theta}$ theorem to prove the claim. Towards this end, we need to control $\|\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\|$ and $\|(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)})\boldsymbol{U}^{(m,l)}\|$.

11.5.1 Step 1: controlling $\|\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\|$

Recall that we have already dealt with $\mathcal{P}_{-m}(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)})$ in the proof of Lemma 1 in Appendix 10.2. Straightforward computation gives

$$\begin{aligned}
\left\|\mathcal{P}_{-m}\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right)\right\| &\lesssim \|\boldsymbol{E}_{:,l}\|_2^2 + \|\boldsymbol{E}_{:,l}\|_2 \left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty} \leq \|\boldsymbol{E}_{:,l}\|_2^2 + \|\boldsymbol{E}_{:,l}\|_2 \|\boldsymbol{A}^{\star}\|, \\
\left\|\mathcal{P}_{m}\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right)\right\| &\lesssim \left\|\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right)_{m,:}\right\|_2 \leq \|\boldsymbol{E}_{:,l}\|_2 \|\boldsymbol{A}^{\star}\|_{\infty},
\end{aligned}$$

where $\mathcal{P}_m$ is the projection onto the subspace of matrices supported on $\{(i,j) \in [d_1] \times [d_2] : i = m \text{ or } j = m\}$. In view of (62) (shown in the proof of Lemma 1 in Appendix 10.2), we know that

$$\begin{aligned}
\left\|\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right\| &\leq \left\|\mathcal{P}_{-m}\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right)\right\| + \left\|\mathcal{P}_{m}\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right)\right\| \\
&\lesssim \|\boldsymbol{E}_{:,l}\|_2^2 + \|\boldsymbol{E}_{:,l}\|_2 \|\boldsymbol{A}^{\star}\| \\
&\lesssim B^2 \log d + \sigma_{\mathsf{col}}^2 + \left(B\sqrt{\log d} + \sigma_{\mathsf{col}}\right) \|\boldsymbol{A}^{\star}\| \\
&\ll \sigma_r^{\star 2},
\end{aligned}$$

where the last step results from the conditions (55a) and (55d). Since $\|\boldsymbol{G}^{(m)} - \boldsymbol{G}^{\star}\| \lesssim \delta_{\mathsf{op}} \ll \sigma_r^{\star 2}$ by Lemma 6 and the condition (55d), this also implies

$$\left\|\boldsymbol{G}^{(m,l)} - \boldsymbol{G}^{\star}\right\| \ll \sigma_r^{\star 2} \quad \text{and} \quad \left\|\left(\boldsymbol{H}^{(m,l)}\right)^{-1}\right\| \lesssim 1, \tag{118}$$

according to [AFWZ17, Lemma 2]. Moreover, it follows from Weyl's inequality that

$$\begin{aligned}
\lambda_r\left(\boldsymbol{G}^{(m)}\right) - \lambda_{r+1}\left(\boldsymbol{G}^{(m)}\right) - \left\|\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right\| &\geq \lambda_r\left(\boldsymbol{G}^{\star}\right) - \lambda_{r+1}\left(\boldsymbol{G}^{\star}\right) - 2\left\|\boldsymbol{G}^{(m)} - \boldsymbol{G}^{\star}\right\| - \left\|\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right\| \\
&\gtrsim \sigma_r^{\star 2}.
\end{aligned} \tag{119}$$

11.5.2 Step 2: controlling $\|(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)})\boldsymbol{U}^{(m,l)}\|$

In view of (118), we can obtain

$$\begin{aligned}
\left\|\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right)\boldsymbol{U}^{(m,l)}\right\| &\leq \left\|\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right)\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|\left\|\left(\boldsymbol{H}^{(m,l)}\right)^{-1}\right\| \\
&\lesssim \left\|\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right)\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\| \\
&\leq \underbrace{\left\|\mathcal{P}_m\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right)\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{\mathrm{F}}}_{=:\alpha_1} + \underbrace{\left\|\mathcal{P}_{-m}\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right)\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|}_{=:\alpha_2}.
\end{aligned} \tag{120}$$

Therefore, it suffices to control α_1 and α_2 separately.

- Regarding α_1, Lemma 12 reveals that, with probability at least $1 - O(d^{-13})$,

$$\begin{aligned}
&\left\|\mathcal{P}_m\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right)\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{\mathrm{F}} \\
&\quad \leq \left\|\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right)_{m,:}\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_2 + \left\|\left(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\right)_{m,:}\right\|_2 \left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty} \\
&\quad \leq \|\boldsymbol{A}^{\star}\|_{\infty} \left\|\left(\boldsymbol{E}_{:,l}\right)^{\top}\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_2 + \|\boldsymbol{A}^{\star}\|_{\infty} \|\boldsymbol{E}_{:,l}\|_2 \left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty} \\
&\quad \lesssim \left(B\log d + \sigma_{\mathsf{col}}\sqrt{\log d}\right) \|\boldsymbol{A}^{\star}\|_{\infty} \left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}.
\end{aligned} \tag{121}$$

- When it comes to α_2, since the spectral norm of a submatrix is always less than that of its original matrix, we can further upper bound

$$\left\|\mathcal{P}_{-m}\big(\boldsymbol{G}^{(m)}-\boldsymbol{G}^{(m,l)}\big)\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\| \leq \underbrace{\left\|\left(\boldsymbol{E}_{:,l}\boldsymbol{E}_{:,l}^{\top}-\boldsymbol{D}_l\right)\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|}_{=:\beta_1} + \underbrace{\left\|\left(\boldsymbol{A}^{\star}_{:,l}\boldsymbol{E}_{:,l}^{\top}+\boldsymbol{E}_{:,l}\boldsymbol{A}^{\star\top}_{:,l}-2\widehat{\boldsymbol{D}}_l\right)\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|}_{=:\beta_2},$$

where $\boldsymbol{D}_l$ and $\widehat{\boldsymbol{D}}_l$ are defined in (60) and (67) in Appendix 10.2. In what follows, let us bound β_1 and β_2.

  - To bound β_1, we have

$$\left\|\left(\boldsymbol{E}_{:,l}\boldsymbol{E}_{:,l}^{\top}-\boldsymbol{D}_l\right)\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\| \leq \left\|\boldsymbol{E}_{:,l}\right\|_2\left\|\left(\boldsymbol{E}_{:,l}\right)^{\top}\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_2+\left\|\boldsymbol{D}_l\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|.$$

By Lemma 12, one has with probability at least $1-O(d^{-13})$,

$$\begin{aligned}\left\|\boldsymbol{E}_{:,l}\right\|_2\left\|\left(\boldsymbol{E}_{:,l}\right)^{\top}\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_2 &\lesssim \left(B\sqrt{\log d}+\sigma_{\mathsf{col}}\right)\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}\\ &\leq \left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)^2\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}.\end{aligned}$$

As for the second term $\left\|\boldsymbol{D}_l\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|$, we first observe that

$$\left\|\mathbb{E}\left[\boldsymbol{D}_l\right]\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\| \leq \left\|\mathbb{E}\left[\boldsymbol{D}_l\right]\right\|\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\| \leq \max_{i\in[d_1]}\mathbb{E}\left[E_{i,l}^2\right]\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\| \leq \sigma_\infty^2.$$

Additionally, the deviation $\big(\boldsymbol{D}_l-\mathbb{E}\left[\boldsymbol{D}_l\right]\big)\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}=\sum_{i\in[d_1]}\big(E_{i,l}^2-\mathbb{E}[E_{i,l}^2]\big)\boldsymbol{e}_i\boldsymbol{e}_i^{\top}\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}$ is a sum of independent zero-mean random matrices. By the matrix Bernstein inequality, we have with probability at least $1-O(d^{-13})$,

$$\begin{aligned}&\left\|\left(\boldsymbol{D}_l-\mathbb{E}\left[\boldsymbol{D}_l\right]\right)\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|\\ &\qquad\lesssim \left(\max_{i\in[d_1]}\left|E_{i,l}^2-\mathbb{E}[E_{i,l}^2]\right|\log d+\sqrt{\sum\nolimits_{i\in[d_1]}\mathbb{E}\left[E_{i,l}^4\right]\log d}\right)\max_{i\in[d_1]}\left\|\boldsymbol{e}_i\boldsymbol{e}_i^{\top}\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}\\ &\qquad\lesssim \left(\max_{i\in[d_1]}\left|E_{i,l}^2-\mathbb{E}[E_{i,l}^2]\right|\log d+\sqrt{\sum\nolimits_{i\in[d_1]}\mathbb{E}\left[E_{i,l}^4\right]\log d}\right)\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}\\ &\qquad\lesssim \left(B^2\log d+\sqrt{\left(B^2\sigma_{\mathsf{col}}^2\right)\log d}\right)\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}\\ &\qquad\overset{\text{(i)}}{\lesssim} \left(B^2\log d+\left(B^2+\sigma_{\mathsf{col}}^2\right)\sqrt{\log d}\right)\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}\\ &\qquad\asymp \left(B^2\log d+\sigma_{\mathsf{col}}^2\sqrt{\log d}\right)\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty},\end{aligned}$$

where we have used the AM-GM inequality in (i). Combining the estimates above yields

$$\beta_1=\left\|\left(\boldsymbol{E}_{:,l}\boldsymbol{E}_{:,l}^{\top}-\boldsymbol{D}_l\right)\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\| \lesssim \left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)^2\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}+\sigma_\infty^2. \tag{122}$$

  - Turning to β_2, we see from Lemma 12 that with probability at least $1-O\left(d^{-13}\right)$, one has

$$\begin{aligned}\left\|\boldsymbol{A}^{\star}_{:,l}\boldsymbol{E}_{:,l}^{\top}\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\| &\leq \left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\left\|\boldsymbol{E}_{:,l}^{\top}\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|\\ &\lesssim \left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}\end{aligned}$$

and

$$\left\|\boldsymbol{E}_{:,l}\boldsymbol{A}^{\star\top}_{:,l}\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\| \leq \left\|\boldsymbol{E}_{:,l}\right\|_2\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|$$

$$\lesssim \left(B\sqrt{\log d}+\sigma_{\mathsf{col}}\right)\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}.$$

In addition, $\widehat{\boldsymbol{D}}_l\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}=\sum_{i\in[d_1]}A^{\star}_{i,l}E_{i,l}\boldsymbol{e}_i\boldsymbol{e}_i^{\top}\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}$ is a sum of independent zero-mean random matrices. It then follows from the matrix Bernstein inequality that with probability at least $1-O\left(d^{-13}\right)$,

$$\begin{aligned}
\left\|\widehat{\boldsymbol{D}}_l\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\| &\lesssim \left(\max_{i\in[d_1]}\left|A^{\star}_{i,l}E_{i,l}\right|\log d+\sqrt{\sum\nolimits_{i\in[d_1]}A^{\star 2}_{i,l}\mathbb{E}\left[E^2_{i,l}\right]\log d}\right)\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}\\
&\lesssim \left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{A}^{\star}\right\|_{\infty}\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}.
\end{aligned}$$

Hence, we know that

$$\begin{aligned}
\beta_2 &= \left\|\left(\boldsymbol{A}^{\star}_{:,l}\boldsymbol{E}_{:,l}^{\top}+\boldsymbol{E}_{:,l}\boldsymbol{A}^{\star\top}_{:,l}-2\widehat{\boldsymbol{D}}_l\right)\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|\\
&\lesssim \left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}+\left(B\sqrt{\log d}+\sigma_{\mathsf{col}}\right)\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\\
&\lesssim \left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}, \qquad (123)
\end{aligned}$$

which results from the facts that $\left\|\boldsymbol{A}^{\star}\right\|_{\infty}\leq\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}$ and $\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}\leq\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|\leq 1$.

Putting (122) and (123) together yields that

$$\begin{aligned}
\left\|\mathcal{P}_{-m}\left(\boldsymbol{G}^{(m)}-\boldsymbol{G}^{(m,l)}\right)\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\| &\lesssim \left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)^2\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}\\
&\quad+\sigma_{\infty}^2+\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}. \qquad (124)
\end{aligned}$$

This combined with (121) and (120) implies

$$\begin{aligned}
\left\|\left(\boldsymbol{G}^{(m)}-\boldsymbol{G}^{(m,l)}\right)\boldsymbol{U}^{(m,l)}\right\| &\lesssim \left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)^2\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}+\sigma_{\infty}^2\\
&\quad+\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}+\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{A}^{\star}\right\|_{\infty}\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}\\
&\overset{\text{(i)}}{\leq} \left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)^2\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}+\sigma_{\infty}^2\\
&\quad+\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}+\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{A}^{\star}\right\|_{\infty}\\
&\overset{\text{(ii)}}{\asymp} \left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)^2\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}+\sigma_{\infty}^2\\
&\quad+\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}, \qquad (125)
\end{aligned}$$

where (i) is due to the facts that $\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|_{2,\infty}\leq\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\|\leq 1$, and (125) arises from the inequality $\left\|\boldsymbol{A}^{\star}\right\|_{\infty}\leq\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}$.

11.5.3 Step 3: combining Step 1 and Step 2

From (119) and (125), we apply the Davis-Kahan $\sin\boldsymbol{\Theta}$ theorem to obtain that with probability exceeding $1-O\left(d^{-13}\right)$,

$$\begin{aligned}
\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\| &\leq \frac{\left\|\left(\boldsymbol{G}^{(m)}-\boldsymbol{G}^{(m,l)}\right)\boldsymbol{U}^{(m,l)}\right\|}{\lambda_r\left(\boldsymbol{G}^{(m)}\right)-\lambda_{r+1}\left(\boldsymbol{G}^{(m)}\right)-\left\|\boldsymbol{G}^{(m)}-\boldsymbol{G}^{(m,l)}\right\|}\\
&\lesssim \frac{1}{\sigma_r^{\star 2}}\left\|\left(\boldsymbol{G}^{(m)}-\boldsymbol{G}^{(m,l)}\right)\boldsymbol{U}^{(m,l)}\right\|\\
&\overset{\text{(i)}}{\lesssim} \frac{1}{\sigma_r^{\star 2}}\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)^2\left(\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}+\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\|\right)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{\sigma_\infty^2}{\sigma_r^{\star 2}}+\frac{1}{\sigma_r^{\star 2}}\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\\
&\overset{\text{(ii)}}{\lesssim}\frac{1}{\sigma_r^{\star 2}}\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)^2\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}+o\left(1\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\|\\
&+\frac{\sigma_\infty^2}{\sigma_r^{\star 2}}+\frac{1}{\sigma_r^{\star 2}}\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}.
\end{aligned}
$$

Here, we have used (115) in (i), and the condition (55d) (i.e. $\max\left\{B\log d,\sigma_{\mathsf{col}}\sqrt{\log d}\right\}\ll\sigma_r^\star$) in (ii). Rearrange the inequalities and taking the union bound over $m\in[d_1]$ and $l\in[d_2]$ complete the proof.

11.6 Proof of Lemma 10

Recall the definition of $\boldsymbol{G}$ in (3.2). We can express

$$
\boldsymbol{G}_{m,:}\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right)=\boldsymbol{A}^{\mathsf{s}}_{m,:}\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right).
$$

Consequently, one can upper bound

$$
\begin{aligned}
\left\|\boldsymbol{G}_{m,:}\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right)\right\|_2\le&\underbrace{\left\|\boldsymbol{A}^\star_{m,:}\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right)\right\|_2}_{=:\beta_1}\\
&+\underbrace{\left\|\left(\boldsymbol{A}^{\mathsf{s}}-\boldsymbol{A}^\star\right)\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right)\right\|_2}_{=:\beta_2}.
\end{aligned}
$$

In what follows, we shall control β_1 and β_2 separately.

- To upper bound β_1, we have

$$
\left\|\boldsymbol{A}^\star_{m,:}\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right)\right\|_2\le\left\|\boldsymbol{A}^\star_{m,:}\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^\top\right\|_2\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right\|.
$$

It is straightforward to derive

$$
\left\|\boldsymbol{A}^\star_{m,:}\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^\top\right\|_2\le\left\|\boldsymbol{A}^\star_{m,:}\boldsymbol{A}^{\mathsf{s}\top}\right\|_2\le\left\|\boldsymbol{A}^\star_{m,:}\boldsymbol{A}^{\star\top}\right\|_2+\left\|\boldsymbol{A}^\star_{m,:}\boldsymbol{E}^\top\right\|_2,
$$

whose first term can be bounded by

$$
\left\|\boldsymbol{A}^\star_{m,:}\boldsymbol{A}^{\star\top}\right\|_2\le\left\|\boldsymbol{A}^\star_{m,:}\right\|_2\left\|\boldsymbol{A}^\star\right\|\le\sigma_1^\star\left\|\boldsymbol{A}^\star\right\|_{2,\infty}.
$$

In addition, Lemma 14 indicates that

$$
\begin{aligned}
\left\|\boldsymbol{A}^\star_{m,:}\boldsymbol{E}^\top\right\|_2^2=\sum\nolimits_i\left(\sum\nolimits_j A^\star_{m,j}E_{i,j}\right)^2&\lesssim\left(\sigma_{\mathsf{col}}^2+\sigma_\infty^2\log^2 d\right)\left\|\boldsymbol{A}^\star\right\|_{2,\infty}^2+B^2\left\|\boldsymbol{A}^\star\right\|_\infty^2\log^3 d\\
&\le\left(\sigma_{\mathsf{col}}^2+B^2\log^2 d\right)\left\|\boldsymbol{A}^\star\right\|_{2,\infty}^2+B^2\left\|\boldsymbol{A}^\star\right\|_\infty^2\log^3 d
\end{aligned}
$$

holds with probability at least $1-O\left(d^{-11}\right)$. Hence, we have

$$
\begin{aligned}
\left\|\boldsymbol{A}^\star_{m,:}\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^\top\right\|_2&\le\left\|\boldsymbol{A}^\star_{m,:}\boldsymbol{A}^{\star\top}\right\|_2+\left\|\boldsymbol{A}^\star_{m,:}\boldsymbol{E}^\top\right\|_2\\
&\lesssim\sigma_1^\star\left\|\boldsymbol{A}^\star\right\|_{2,\infty}+\left(\sigma_{\mathsf{col}}+B\log d\right)\left\|\boldsymbol{A}^\star\right\|_{2,\infty}+B\log^{3/2}d\left\|\boldsymbol{A}^\star\right\|_\infty\\
&\lesssim\sigma_1^\star\left\|\boldsymbol{A}^\star\right\|_{2,\infty}+\sigma_1^{\star 2}\sqrt{\frac{\mu r}{d_1}}\lesssim\sigma_1^{\star 2}\sqrt{\frac{\mu r}{d_1}},
\end{aligned}\tag{126}
$$

using conditions (55a), (55c) and (55d). Moreover, from Lemma 1 and Lemma 6, we know that

$$
\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right\|\lesssim\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right\|\le\frac{\left\|\boldsymbol{G}^{(m)}-\boldsymbol{G}^\star\right\|}{\lambda_r\left(\boldsymbol{G}^\star\right)-\lambda_{r+1}\left(\boldsymbol{G}^{(m)}\right)}
$$

$$
\begin{aligned}
&\le \frac{\left\|\boldsymbol{G}^{(m)}-\boldsymbol{G}^\star\right\|}{\lambda_r\left(\boldsymbol{G}^\star\right)-\lambda_{r+1}\left(\boldsymbol{G}^\star\right)-\left\|\boldsymbol{G}^{(m)}-\boldsymbol{G}^\star\right\|}\\
&\lesssim \frac{1}{\sigma_r^{\star 2}}\left\|\boldsymbol{G}^{(m)}-\boldsymbol{G}^\star\right\| \lesssim \frac{\delta_{\mathsf{op}}}{\sigma_r^{\star 2}},
\end{aligned}
\tag{127}
$$

where δ_{op} is defined in (5). Combining (126) and (127) yields

$$
\left\|\boldsymbol{A}_{m,:}^\star\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right)\right\|_2 \lesssim \left\|\boldsymbol{A}_{m,:}^\star\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)^\top\right\|_2\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right\| \lesssim \delta_{\mathsf{op}}\kappa^2\sqrt{\frac{\mu r}{d_1}}.
\tag{128}
$$

- Next, we look at β_2. Before we start, we pause to note that by (113), one has

$$
\begin{aligned}
\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\| &\lesssim \frac{1}{\sigma_r^{\star 2}}\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)^2\left(\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right\|_{2,\infty}+\left\|\boldsymbol{U}^\star\right\|_{2,\infty}\right)\\
&\quad+\frac{1}{\sigma_r^{\star 2}}\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{A}^\star\right\|\sqrt{\frac{\mu r}{d_2}}.
\end{aligned}
\tag{129}
$$

We now ready to control $\left(\boldsymbol{A}^{\mathsf{s}}-\boldsymbol{A}^\star\right)_{m,:}\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right)$, which can be accomplished in the same way as in the proof of Lemma 7 in Appendix 11.3. We omit the proof details for conciseness here and only give the proof sketch. First, we can use $\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}-\boldsymbol{U}^\star$ as the surrogate for $\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star$ to deal with the statistical dependence issue, and apply the Bernstein inequality to show that with probability at least $1-O\left(d^{-11}\right)$,

$$
\begin{aligned}
&\left\|\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right)\right\|_{2,\infty}\\
&\quad\lesssim \left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right\|_{2,\infty}\\
&\qquad+\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right\|\\
&\qquad+\left(\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}+B\sqrt{\log d}+\sigma_{\mathsf{col}}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\|\\
&\quad\overset{\text{(i)}}{\lesssim} \left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right\|_{2,\infty}+\frac{\delta_{\mathsf{op}}}{\sigma_r^{\star 2}}\left\|\boldsymbol{A}^\star\right\|\sqrt{\frac{\mu r}{d_2}}\\
&\qquad+\frac{1}{\sigma_r^{\star 2}}\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)^3\left(\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right\|_{2,\infty}+\left\|\boldsymbol{U}^\star\right\|_{2,\infty}\right)\\
&\qquad+\frac{1}{\sigma_r^{\star 2}}\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)^2\left\|\boldsymbol{A}^\star\right\|\sqrt{\frac{\mu r}{d_2}}\\
&\quad\lesssim \left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right\|_{2,\infty}+\frac{\delta_{\mathsf{op}}}{\sigma_r^\star}\left(\left\|\boldsymbol{U}^\star\right\|_{2,\infty}+\kappa\sqrt{\frac{\mu r}{d_2}}\right),
\end{aligned}
\tag{130}
$$

where (i) follows from (112), (127) and (129) and the inquality $\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty}\le\left\|\boldsymbol{A}^\star\right\|\sqrt{\mu r/d_2}$; (130) arises from the definition of δ_{op} in (56) and conditions (55a) and (55d) (namely, $B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\ll\sigma_r^\star/\kappa$ and $\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)^2\lesssim\delta_{\mathsf{op}}\ll\sigma_r^{\star 2}$). Applying the matrix Bernstein inequality yields that with probability at least $1-O\left(d^{-11}\right)$,

$$
\begin{aligned}
&\left\|\left(\boldsymbol{A}^{\mathsf{s}}-\boldsymbol{A}^\star\right)_{m,:}\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right)\right\|_2\\
&\quad\lesssim B\log d\left\|\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right)\right\|_{2,\infty}\\
&\qquad+\sqrt{d_1}\,\sigma_\infty\left(\sigma_{\mathsf{row}}\log d+\left\|\boldsymbol{A}^\star\right\|\sqrt{\log d}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right\|_{2,\infty}\\
&\quad\overset{\text{(i)}}{\lesssim} B\log d\left(B\log d+\sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^\star\right\|_{2,\infty}\\
&\qquad+\delta_{\mathsf{op}}\frac{B\log d}{\sigma_r^\star}\left(\left\|\boldsymbol{U}^\star\right\|_{2,\infty}+\kappa\sqrt{\frac{\mu r}{d_2}}\right)
\end{aligned}
$$

$$
\begin{aligned}
&+ \sigma_{\mathsf{col}}\big(\sigma_{\mathsf{row}} \log d + \|\boldsymbol{A}^\star\| \sqrt{\log d}\big) \big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\star\big\|_{2,\infty} \\
&\overset{\text{(ii)}}{\lesssim} \sigma_{\mathsf{col}}\big(\sigma_{\mathsf{row}} \log d + \|\boldsymbol{A}^\star\| \sqrt{\log d}\big) \big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\star\big\|_{2,\infty} \\
&\quad + \delta_{\mathsf{op}} \frac{B\log d}{\sigma_r^\star}\left(\|\boldsymbol{U}^\star\|_{2,\infty} + \kappa\sqrt{\frac{\mu r}{d_2}}\right) \\
&\overset{\text{(iii)}}{\lesssim} \sigma_{\mathsf{col}}\big(\sigma_{\mathsf{row}} \log d + \|\boldsymbol{A}^\star\| \sqrt{\log d}\big) \big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\star\big\|_{2,\infty} + o\left(1\right)\delta_{\mathsf{op}}\|\boldsymbol{U}^\star\|_{2,\infty} + \delta_{\mathsf{op}}\frac{\kappa\sigma_{\mathsf{col}}\sqrt{\log d}}{\sigma_r^\star}\sqrt{\frac{\mu r}{d_1}} \\
&\overset{\text{(iv)}}{\lesssim} \sigma_{\mathsf{col}}\big(\sigma_{\mathsf{row}} \log d + \|\boldsymbol{A}^\star\| \sqrt{\log d}\big) \big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\star\big\|_{2,\infty} + o\left(1\right)\delta_{\mathsf{op}}\sqrt{\frac{\mu r}{d_1}}.
\end{aligned}
$$

Here, (i) follows from (54) and (130); (ii) is due to conditions (55a) and (55d) that $B^2 \log d \lesssim \sigma_{\mathsf{col}}\sigma_{\mathsf{row}}$, $B\log d \ll \sigma_r^\star$ and $B\log d\left(B\log d + \sigma_{\mathsf{col}}\sqrt{\log d}\right) \lesssim B\log^2 d + \sigma_{\mathsf{col}}^2 \log d \leq \delta_{\mathsf{op}}$; (iii) holds true because of (55b) and (55d) that $B\log d \ll \sigma_r^\star$; and (iv) arises from (55d) that $\sigma_{\mathsf{col}}\sqrt{\log d} \ll \sigma_r^\star/\kappa$. Recalling the definition of δ_{loo} in (84), we obtain that

$$
\left\|\left(\boldsymbol{A}^{\mathsf{s}} - \boldsymbol{A}^\star\right)_{m,:}\left[\mathcal{P}_{-m,:}\left(\boldsymbol{A}^{\mathsf{s}}\right)\right]^\top \left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\star\right)\right\|_2 \lesssim \delta_{\mathsf{loo}}\big\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\star\big\|_{2,\infty} + o\left(1\right)\delta_{\mathsf{op}}\sqrt{\frac{\mu r}{d_1}}. \tag{131}
$$

Putting (128) and (131) together, we arrive at the advertised bound.

12 Proofs for lower bounds

12.1 Proof of Theorem 3.4

Without loss of generality, it suffices to focus on the set of matrices with $\sigma_r(\boldsymbol{A}^\star) \in [0.9, 1.1]$; otherwise one can always rescale the matrices $\boldsymbol{A}^\star$ and $\boldsymbol{N}$ by the same factor $1/\sigma_r(\boldsymbol{A}^\star)$ simultaneously.

Let us start with the minimax spectral norm bound (3.11a). Recognizing the elementary fact that

$$
\big\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\big\| \asymp \min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|,
$$

we have

$$
\inf_{\widehat{\boldsymbol{U}}} \sup_{\boldsymbol{A}^\star\in\mathcal{M}^\star} \mathbb{E}\Big[\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \big\|\widehat{\boldsymbol{U}}\boldsymbol{R} - \boldsymbol{U}(\boldsymbol{A}^\star)\big\|\Big] \asymp \inf_{\widehat{\boldsymbol{U}}} \sup_{\boldsymbol{A}^\star\in\mathcal{M}^\star} \mathbb{E}\Big[\big\|\widehat{\boldsymbol{U}}\widehat{\boldsymbol{U}}^\top - \boldsymbol{U}(\boldsymbol{A}^\star)\big(\boldsymbol{U}(\boldsymbol{A}^\star)\big)^\top\big\|\Big].
$$

In light of this, we shall focus attention on bounding $\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}(\boldsymbol{A}^\star)\big(\boldsymbol{U}(\boldsymbol{A}^\star)\big)^\top\|$ in the remainder of the proof. In addition, it can be easily seen (which we omit for brevity) that it is sufficient to establish the lower bounds for the rank-1 case (i.e. $r = 1$).[1] In what follows, we assume that

$$
\boldsymbol{A} = \mathcal{P}_\Omega(\underbrace{\boldsymbol{u}^\star\boldsymbol{v}^{\star\top}}_{=\boldsymbol{A}^\star} + \boldsymbol{N}),
$$

where $\boldsymbol{v}^\star \sim \mathcal{N}(\boldsymbol{0}, \frac{1}{d_2}\boldsymbol{I}_{d_2})$ and $N_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$. Without loss of generality, we assume throughout that $d_1/2$ is an integer.

Step 1: constructing a collection of hypotheses. Let us begin by constructing a family of well-separated unit vectors $\{\boldsymbol{u}^i\}_{1\leq i\leq M} \subseteq \mathbb{R}^{d_1}$. In view of the celebrated Varshamov-Gilbert bound [Mas07, Lemma 4.7], one can find a set of vectors $\left\{\boldsymbol{w}^i\right\}_{i=1}^M \subseteq \{-1,1\}^{d_1/2}$ obeying

$$
\log M \geq d_1/32 \qquad \text{and} \qquad \min\left\{\big\|\boldsymbol{w}^i \pm \boldsymbol{w}^j\big\|_2\right\} \geq \sqrt{d_1}/2, \quad \forall i \neq j, \tag{132}
$$

[1] Consider the true matrix $\boldsymbol{U}^\star \in \mathbb{R}^{d_1\times r}$ with $\boldsymbol{U}^\star = \begin{bmatrix}\boldsymbol{u}^\star & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{Q}^\star\end{bmatrix}$ for some vector $\boldsymbol{u}^\star \in \mathbb{R}^{d_1/2}, \|\boldsymbol{u}^\star\|_2 = 1$ and $\boldsymbol{Q}^\star \in \mathcal{O}^{d_1/2\times(r-1)}$. Suppose that there is an oracle informing us of $\boldsymbol{Q}^\star$, then the problem of estimating $\boldsymbol{U}^\star$ is reduced to the rank-1 case. This suggests that we can focus on the rank-1 case to derive the lower bound.

where we denote $\min \|\boldsymbol{a} \pm \boldsymbol{b}\|_2 = \min\{\|\boldsymbol{a}-\boldsymbol{b}\|_2, \|\boldsymbol{a}+\boldsymbol{b}\|_2\}$. For some $\delta \in (0,1)$ to be chosen later, we generate the d_1-dimensional vectors

$$\boldsymbol{u}^i := \frac{\delta}{\sqrt{d_1/2}} \begin{bmatrix} \boldsymbol{w}^i \\ \boldsymbol{0} \end{bmatrix} + \sqrt{\frac{1-\delta^2}{d_1/2}} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{1} \end{bmatrix} \in \mathbb{R}^{d_1}, \qquad 1 \le i \le M, \tag{133}$$

where $\boldsymbol{0}$ (resp. $\boldsymbol{1}$) denotes the all-zero (resp. all-one) vector. By construction, it is easily seen that $\|\boldsymbol{u}^i\|_2 = 1$ for all $1 \le i \le M$, and that

$$\begin{aligned} \left\|\boldsymbol{u}^i\boldsymbol{u}^{i\top} - \boldsymbol{u}^j\boldsymbol{u}^{j\top}\right\| &\ge \frac{1}{\sqrt{2}} \left\|\boldsymbol{u}^i\boldsymbol{u}^{i\top} - \boldsymbol{u}^j\boldsymbol{u}^{j\top}\right\|_{\mathrm{F}} = \frac{1}{\sqrt{2}} \sqrt{\mathsf{tr}\big(\boldsymbol{u}^i\boldsymbol{u}^{i\top} - \boldsymbol{u}^j\boldsymbol{u}^{j\top}\big)\big(\boldsymbol{u}^i\boldsymbol{u}^{i\top} - \boldsymbol{u}^j\boldsymbol{u}^{j\top}\big)} \\ &= \frac{1}{\sqrt{2}}\sqrt{2 - 2\langle \boldsymbol{u}^i, \boldsymbol{u}^j\rangle^2} \ge \frac{1}{\sqrt{2}}\sqrt{2 - 2|\langle \boldsymbol{u}^i, \boldsymbol{u}^j\rangle|} \\ &= \frac{1}{\sqrt{2}}\sqrt{\|\boldsymbol{u}^i\|_2^2 + \|\boldsymbol{u}^j\|_2^2 - 2|\langle \boldsymbol{u}^i, \boldsymbol{u}^j\rangle|} = \frac{1}{\sqrt{2}} \min \|\boldsymbol{u}^i \pm \boldsymbol{u}^j\|_2 \\ &= \frac{1}{\sqrt{2}} \cdot \frac{\delta}{\sqrt{d_1/2}} \min\left\{\left\|\boldsymbol{w}^i \pm \boldsymbol{w}^j\right\|_2\right\} \ge \frac{\delta}{2}, \end{aligned} \tag{134}$$

where the last inequality arises from (132). We shall then associate each vector $\boldsymbol{u}^i$ $(1 \le i \le M)$ with a hypothesis as follows:

$$\mathcal{H}_i: \quad \boldsymbol{A} = \mathcal{P}_\Omega(\boldsymbol{u}^i\boldsymbol{v}^{\star\top} + \boldsymbol{N}), \qquad 1 \le i \le M.$$

In the sequel, for each $1 \le i \le M$ and $1 \le k \le d_2$, we denote

- $\mathbb{P}^i$: the distribution of $\boldsymbol{A}$ under the hypothesis $\mathcal{H}_i$;
- $\mathbb{P}^i_\Omega$: the distribution of $\boldsymbol{A}$ under the hypothesis $\mathcal{H}_i$, conditional on Ω;
- $\mathbb{P}^i_{\Omega,k}$: the distribution of the k-th column of $\boldsymbol{A}$ under the hypothesis $\mathcal{H}_i$, conditional on Ω.

Additionally, standard Gaussian concentration inequalities imply that: with high probability, one has

$$\|\boldsymbol{u}^i\boldsymbol{v}^{\star\top}\| = \left\|\boldsymbol{u}^i\right\|_2 \|\boldsymbol{v}^\star\|_2 = 1 + o(1),$$

and hence $\boldsymbol{u}^i\boldsymbol{v}^{\star\top} \in \mathcal{M}^\star$ for all $1 \le i \le M$.

Step 2: bounding the KL divergence between each pair of hypotheses. Fix any $1 \le i \ne j \le M$. The next step lies in upper bounding the KL divergence of $\mathbb{P}^j$ from $\mathbb{P}^i$. Towards this, we observe that

$$\begin{aligned} \mathsf{KL}\big(\mathbb{P}^i \,\|\, \mathbb{P}^j\big) &= \mathsf{KL}\big(\mathbb{E}_\Omega\big[\mathbb{P}^i_\Omega\big] \,\|\, \mathbb{E}_\Omega\big[\mathbb{P}^j_\Omega\big]\big) \le \mathbb{E}_\Omega\big[\mathsf{KL}\big(\mathbb{P}^i_\Omega \,\|\, \mathbb{P}^j_\Omega\big)\big] \\ &= \mathbb{E}_\Omega\Big[\sum_{1 \le k \le d_2} \mathsf{KL}\big(\mathbb{P}^i_{\Omega,k} \,\|\, \mathbb{P}^j_{\Omega,k}\big)\Big]. \end{aligned} \tag{135}$$

Here, the penultimate inequality arises from the convexity of KL divergence and Jensen's inequality, whereas the last line follows since the noise components are independently generated and KL divergence is additive for independent distributions.

Before moving on, we find it convenient to introduce additional notation to simplify presentation. For any vector $\boldsymbol{u} := [u_i]_{1 \le i \le d_1}$ and any index set $\mathcal{A} \subseteq [d_1]$, we define

$$\boldsymbol{u}_\mathcal{A} := [u_i]_{i \in \mathcal{A}} \in \mathbb{R}^{|\mathcal{A}|},$$

which is obtained by maintaining only those entries of $\boldsymbol{u}$ lying within $\mathcal{A}$. In addition, define

$$\begin{aligned} \Omega_k &:= \{m \in [d_1] \colon (m,k) \in \Omega\}; \\ \widehat{\Omega}_k &:= \{m \in [d_1/2] \colon (m,k) \in \Omega\}; \\ \widetilde{\Omega}_k &:= \{d_1/2 < m \le d_1 : (m,k) \in \Omega\}; \end{aligned}$$

$$\widehat{\Omega}_k^{(i,j),\mathsf{diff}} := \{m \in [d_1/2] \colon (m,k) \in \Omega \text{ and } u_m^i \neq u_m^j\};$$
$$\widehat{\Omega}_k^{(i,j),\mathsf{same}} := \{m \in [d_1/2] \colon (m,k) \in \Omega \text{ and } u_m^i = u_m^j\}.$$

By construction, one clearly has $\Omega_k = \widehat{\Omega}_k \cup \widetilde{\Omega}_k = \widehat{\Omega}_k^{(i,j),\mathsf{diff}} \cup \widehat{\Omega}_k^{(i,j),\mathsf{same}} \cup \widetilde{\Omega}_k$, and

$$\|\boldsymbol{u}_{\Omega_k}^i\|_2^2 = \frac{2\delta^2}{d_1}|\widehat{\Omega}_k| + \frac{2(1-\delta^2)}{d_1}|\widetilde{\Omega}_k|, \qquad 1 \le i \le M. \tag{137}$$

With these in place, we are in a position to control the KL divergence. We first make the observation that: conditional on the sampling set Ω_k and under the hypothesis $\mathcal{H}_i$, the entries of $\boldsymbol{A}_{:,k}$ within Ω_k follow a multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\Omega_k}^i)$, where

$$\boldsymbol{\Sigma}_{\Omega_k}^i := \sigma^2 \boldsymbol{I}_{|\Omega_k|} + \frac{1}{d_2}\boldsymbol{u}_{\Omega_k}^i \boldsymbol{u}_{\Omega_k}^{i\top}.$$

As a result, invoking the KL divergence for multivariate Gaussians, we can deduce that

$$\begin{aligned}
\mathsf{KL}\big(\mathbb{P}_{\Omega,k}^i \,\|\, \mathbb{P}_{\Omega,k}^j\big) &= \mathsf{KL}\big(\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\Omega_k}^i) \,\|\, \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\Omega_k}^j)\big) = \frac{1}{2}\Big(\mathsf{tr}\big((\boldsymbol{\Sigma}_{\Omega_k}^j)^{-1}\boldsymbol{\Sigma}_{\Omega_k}^i\big) - |\Omega_k|\Big). \\
&= \frac{1}{2}\mathsf{tr}\Big(\big(\boldsymbol{I}_{|\Omega_k|} + \frac{1}{\sigma^2 d_2}\boldsymbol{u}_{\Omega_k}^j \boldsymbol{u}_{\Omega_k}^{j\top}\big)^{-1}\big(\boldsymbol{I}_{|\Omega_k|} + \frac{1}{\sigma^2 d_2}\boldsymbol{u}_{\Omega_k}^i \boldsymbol{u}_{\Omega_k}^{i\top}\big)\Big) - \frac{1}{2}|\Omega_k| \\
&\overset{\text{(i)}}{=} \frac{1}{2}\mathsf{tr}\Big(\Big(\boldsymbol{I}_{|\Omega_k|} - \frac{1}{\sigma^2 d_2 + \big\|\boldsymbol{u}_{\Omega_k}^j\big\|_2^2}\boldsymbol{u}_{\Omega_k}^j \boldsymbol{u}_{\Omega_k}^{j\top}\Big)\big(\boldsymbol{I}_{|\Omega_k|} + \frac{1}{\sigma^2 d_2}\boldsymbol{u}_{\Omega_k}^i \boldsymbol{u}_{\Omega_k}^{i\top}\big)\Big) - \frac{1}{2}|\Omega_k| \\
&= \frac{1}{2\sigma^2 d_2}\big\|\boldsymbol{u}_{\Omega_k}^i\big\|_2^2 - \frac{1}{2\big(\sigma^2 d_2 + \big\|\boldsymbol{u}_{\Omega_k}^j\big\|_2^2\big)}\big\|\boldsymbol{u}_{\Omega_k}^j\big\|_2^2 - \frac{\big\langle \boldsymbol{u}_{\Omega_k}^i, \boldsymbol{u}_{\Omega_k}^j\big\rangle^2}{2\sigma^2 d_2\big(\sigma^2 d_2 + \big\|\boldsymbol{u}_{\Omega_k}^j\big\|_2^2\big)} \\
&\overset{\text{(ii)}}{=} \frac{\big\|\boldsymbol{u}_{\Omega_k}^i\big\|_2^4 - \big\langle \boldsymbol{u}_{\Omega_k}^i, \boldsymbol{u}_{\Omega_k}^j\big\rangle^2}{2\sigma^2 d_2\big(\sigma^2 d_2 + \big\|\boldsymbol{u}_{\Omega_k}^i\big\|_2^2\big)} = \frac{\big\langle \boldsymbol{u}_{\Omega_k}^i, \boldsymbol{u}_{\Omega_k}^i + \boldsymbol{u}_{\Omega_k}^j\big\rangle\big\langle \boldsymbol{u}_{\Omega_k}^i, \boldsymbol{u}_{\Omega_k}^i - \boldsymbol{u}_{\Omega_k}^j\big\rangle}{2\sigma^2 d_2\big(\sigma^2 d_2 + \big\|\boldsymbol{u}_{\Omega_k}^i\big\|_2^2\big)},
\end{aligned} \tag{138}$$

where (i) follows from the Woodbury matrix identity, (ii) arises since $\big\|\boldsymbol{u}_{\Omega_k}^i\big\|_2 = \big\|\boldsymbol{u}_{\Omega_k}^j\big\|_2$ (cf. (137)). Next, straightforward calculations yield

$$\begin{aligned}
\big\langle \boldsymbol{u}_{\Omega_k}^i, \boldsymbol{u}_{\Omega_k}^i - \boldsymbol{u}_{\Omega_k}^j\big\rangle &= \frac{4\delta^2}{d_1}\big|\widehat{\Omega}_k^{(i,j),\mathsf{diff}}\big|, \\
\big\langle \boldsymbol{u}_{\Omega_k}^i, \boldsymbol{u}_{\Omega_k}^i + \boldsymbol{u}_{\Omega_k}^j\big\rangle &= \frac{4\delta^2}{d_1}\big|\widehat{\Omega}_k^{(i,j),\mathsf{same}}\big| + \frac{4(1-\delta^2)}{d_1}\big|\widetilde{\Omega}_k\big|.
\end{aligned}$$

Substituting the above identities and the identity (137) into (138) gives

$$\begin{aligned}
\mathsf{KL}\big(\mathbb{P}_{\Omega,k}^i \,\|\, \mathbb{P}_{\Omega,k}^j\big) &\le \frac{\big\langle \boldsymbol{u}_{\Omega_k}^i, \boldsymbol{u}_{\Omega_k}^i + \boldsymbol{u}_{\Omega_k}^j\big\rangle\big\langle \boldsymbol{u}_{\Omega_k}^i, \boldsymbol{u}_{\Omega_k}^i - \boldsymbol{u}_{\Omega_k}^j\big\rangle}{2\sigma^2 d_2 \cdot \sigma^2 d_2} \\
&\le \frac{16\delta^2\big|\widehat{\Omega}_k^{(i,j),\mathsf{diff}}\big| \cdot \big(\delta^2\big|\widehat{\Omega}_k^{(i,j),\mathsf{same}}\big| + (1-\delta^2)\big|\widetilde{\Omega}_k\big|\big)}{\sigma^4 d_2^2 d_1^2} \\
&\le \frac{16\delta^2\big|\widehat{\Omega}_k^{(i,j),\mathsf{diff}}\big| \cdot \big(\big|\widehat{\Omega}_k^{(i,j),\mathsf{same}}\big| + \big|\widetilde{\Omega}_k\big|\big)}{\sigma^4 d_2^2 d_1^2},
\end{aligned}$$

where we have used the fact that $\delta \in (0,1)$. In addition, the elementary inequality $2\,\big\|\boldsymbol{u}_{\Omega_k}^i\big\|_2^2 \ge \big\langle \boldsymbol{u}_{\Omega_k}^i, \boldsymbol{u}_{\Omega_k}^i + \boldsymbol{u}_{\Omega_k}^j\big\rangle$ together with the preceding identities yields

$$\begin{aligned}
\mathsf{KL}\big(\mathbb{P}_{\Omega,k}^i \,\|\, \mathbb{P}_{\Omega,k}^j\big) &= \frac{\big\langle \boldsymbol{u}_{\Omega_k}^i, \boldsymbol{u}_{\Omega_k}^i + \boldsymbol{u}_{\Omega_k}^j\big\rangle\big\langle \boldsymbol{u}_{\Omega_k}^i, \boldsymbol{u}_{\Omega_k}^i - \boldsymbol{u}_{\Omega_k}^j\big\rangle}{2\sigma^2 d_2 \|\boldsymbol{u}_{\Omega_k}^i\|_2^2} \le \frac{\big\langle \boldsymbol{u}_{\Omega_k}^i, \boldsymbol{u}_{\Omega_k}^i - \boldsymbol{u}_{\Omega_k}^j\big\rangle}{\sigma^2 d_2} \\
&= \frac{4\delta^2\big|\widehat{\Omega}_k^{(i,j),\mathsf{diff}}\big|}{\sigma^2 d_2 d_1}.
\end{aligned}$$

Putting the above bounds and the inequality (135) together leads to

$$\mathsf{KL}\big(\mathbb{P}^i \,\|\, \mathbb{P}^j\big) \le \mathbb{E}\left[\sum_{k=1}^{d_2} \frac{16\delta^2 \big|\widehat{\Omega}_k^{(i,j),\mathsf{diff}}\big| \cdot \big(\big|\widehat{\Omega}_k^{(i,j),\mathsf{same}}\big| + \big|\widetilde{\Omega}_k\big|\big)}{\sigma^4 d_2^2 d_1^2}\right] \lesssim \frac{\delta^2 p^2}{\sigma^4 d_2}; \tag{139a}$$

$$\mathsf{KL}\big(\mathbb{P}^i \,\|\, \mathbb{P}^j\big) \le \sum_{k=1}^{d_2} \frac{4\delta^2 \big|\widehat{\Omega}_k^{(i,j),\mathsf{diff}}\big|}{\sigma^2 d_2 d_1} \lesssim \frac{\delta^2 p}{\sigma^2}. \tag{139b}$$

Step 3: invoking Fano's inequality. Fano's inequality [Tsy08, Corollary 2.6] asserts that if

$$\frac{1}{M}\sum_{i=2}^{M} \mathsf{KL}\big(\mathbb{P}^i \,\|\, \mathbb{P}^1\big) \le \frac{\log M}{32}, \tag{140}$$

then the minimax probability of testing error necessarily obeys

$$p_{\mathrm{e},M} := \inf_{\psi} \max_{1\le j\le M} \mathbb{P}\left\{\psi \ne j \mid \mathcal{H}_j\right\} \ge 0.2,$$

where the infimum is taken over all tests. In view of (132) and the upper bounds (139), we observe that the bound (140) would hold by taking

$$\delta = c_1 \min\left\{\frac{\sigma^2\sqrt{d_1 d_2}}{p} + \sigma\sqrt{\frac{d_1}{p}},\, 1\right\} \tag{141}$$

for some sufficiently small constant $c_1 > 0$. Therefore, adopting the standard reduction scheme as introduced in [Tsy08, Chapter 2.2], we arrive at

$$\begin{aligned}\inf_{\widehat{\boldsymbol{U}}} \sup_{\boldsymbol{A}^\star\in\mathcal{M}^\star} \mathbb{E}\Big[\big\|\widehat{\boldsymbol{U}}\widehat{\boldsymbol{U}}^\top - \boldsymbol{U}(\boldsymbol{A}^\star)\big(\boldsymbol{U}(\boldsymbol{A}^\star)\big)^\top\big\|\Big] &\gtrsim \min_{i\ne j} \big\|\boldsymbol{u}^i\boldsymbol{u}^{i\top} - \boldsymbol{u}^j\boldsymbol{u}^{j\top}\big\| \gtrsim \delta \\ &\asymp \min\left\{\frac{\sigma^2\sqrt{d_1 d_2}}{p} + \sigma\sqrt{\frac{d_1}{p}},\, 1\right\},\end{aligned}$$

where the penultimate inequality comes from (134), and the last line makes use of our choice (141). Combined with the high-probability fact that $\|\boldsymbol{u}^i\boldsymbol{v}^{\star\top}\| \in [0.9, 1.1]$, we establish the minimax spectral norm bound (3.11a).

Given that $\|\boldsymbol{Z}\|_{2,\infty} \ge \frac{1}{\sqrt{d_1}}\|\boldsymbol{Z}\|$ holds for any $\boldsymbol{Z} \in \mathbb{R}^{d_1\times r}$, the advertised $\ell_{2,\infty}$ lower bound (3.11b) follows immediately from the spectral norm lower bound (3.11a).

12.2 Proof of Theorem 3.5

To begin with, the sampling set Ω can be equivalently viewed as the edge set of a random bipartite graph $\mathcal{G}(d_1, d_2, p)$. Here, we recall that $\mathcal{G}(d_1, d_2, p)$ is generated by (i) taking the complete bipartite graph connecting two disjoint vertex sets $\mathcal{U}$ and $\mathcal{V}$, where $|\mathcal{U}| = d_1$ and $|\mathcal{V}| = d_2$, and (ii) removing each edge independently with probability $1-p$. As shown in [Joh12, Theorem 6], if $p < \frac{1-\epsilon}{\sqrt{d_1 d_2}}$ for some constant $0 < \epsilon < 1$ and if $d_1 \le d_2$, then with probability $1 - o(1)$, there is no connected component in $\mathcal{G}(d_1, d_2, p)$ containing more than $O(\log d_1)$ (resp. $O(\sqrt{d_1 d_2}\log d_1)$) vertices in $\mathcal{U}$ (resp. $\mathcal{V}$). In what follows, we let $\mathcal{C}_1, \cdots, \mathcal{C}_K$ denote the collection of connected components in $\mathcal{G}(d_1, d_2, p)$, and denote by $\mathcal{U}_i$ (resp. $\mathcal{V}_i$) the set of vertices in $\mathcal{U}$ (resp. $\mathcal{V}$) that reside within $\mathcal{C}_i$.

Generate $\boldsymbol{u}^\star$ and $\boldsymbol{v}^\star$ such that

$$u_i^\star = \begin{cases} 1/\sqrt{d_1}, & \text{with prob. } 0.5 \\ -1/\sqrt{d_1}, & \text{else}\end{cases} \qquad \text{and} \qquad v_j^\star = \begin{cases} 1/\sqrt{d_2}, & \text{with prob. } 0.5 \\ -1/\sqrt{d_2}, & \text{else}\end{cases}$$

for each $1 \le i \le d_1$ and $1 \le j \le d_2$. Letting $\boldsymbol{u}_{\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}|}$ represent a vector comprising the entries of $\boldsymbol{u}$ whose indices come from $\mathcal{S}$, we generate

$$\widetilde{\boldsymbol{u}}^{\star}_{\mathcal{C}_i} = z_i \boldsymbol{u}^{\star}_{\mathcal{C}_i} \qquad \text{and} \qquad \widetilde{\boldsymbol{v}}^{\star}_{\mathcal{C}_i} = z_i \boldsymbol{v}^{\star}_{\mathcal{C}_i};$$

here, z_i is a set of independent Bernoulli variables with $z_i = 1$ with probability 0.5 and $z_i = -1$ otherwise. As one can easily verify (which we omit for brevity),

- $\boldsymbol{u}^{\star}_{\mathcal{C}_i} \boldsymbol{v}^{\star\top}_{\mathcal{C}_i} = \widetilde{\boldsymbol{u}}^{\star}_{\mathcal{C}_i} \widetilde{\boldsymbol{v}}^{\star\top}_{\mathcal{C}_i}$ for each i, and hence $\mathcal{P}_{\Omega}(\boldsymbol{u}^{\star}\boldsymbol{v}^{\star\top}) = \mathcal{P}_{\Omega}(\widetilde{\boldsymbol{u}}^{\star}\widetilde{\boldsymbol{v}}^{\star\top})$;
- with probability $1 - o(1)$, one has $\min \|\boldsymbol{u}^{\star} \pm \widetilde{\boldsymbol{u}}^{\star}\|_2 \asymp 1$ and $\|\boldsymbol{u}^{\star}\boldsymbol{v}^{\star\top} - \widetilde{\boldsymbol{u}}^{\star}\widetilde{\boldsymbol{v}}^{\star\top}\|_{\mathrm{F}} \asymp 1$.

This concludes the proof.

13 A few more auxiliary lemmas

In this section, we establish a few auxiliary facts that are useful throughout the proof of the main theorem. We begin with some basic properties about the truth $\boldsymbol{A}^{\star}$ and $\boldsymbol{G}^{\star}$.

Lemma 11. *Recall the definition of the incoherence parameters in Definition 2.1. Then one has*

$$\begin{aligned}
&\|\boldsymbol{A}^{\star}\|_{2,\infty} \le \sqrt{\frac{\mu_1 r \sigma_1^{\star 2}}{d_1}}, \quad \left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty} \le \sqrt{\frac{\mu_2 r \sigma_1^{\star 2}}{d_2}}, \\
&\|\boldsymbol{G}^{\star}\|_{2,\infty} \le \sqrt{\frac{\mu_1 r \sigma_1^{\star 4}}{d_1}}, \quad \|\boldsymbol{A}^{\star}\|_{\infty} \le \min\left\{\sqrt{\frac{\mu_1\mu_2 r^2}{d_1 d_2}}, \sigma_1^{\star}\left\|\boldsymbol{U}^{\star}\right\|_{2,\infty}, \sigma_1^{\star}\left\|\boldsymbol{V}^{\star}\right\|_{2,\infty}\right\}.
\end{aligned}$$

Next, we summarize several facts related to the matrix $\boldsymbol{E}$ defined in (51), which contains independent zero-mean entries.

Lemma 12. *Fix any matrices $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$. With probability greater than $1 - O\left(d^{-20}\right)$, the following holds*

$$\begin{aligned}
\max_{i\in[d_1]} \sum_{j\in[d_2]} E_{i,j}^2 &\lesssim B^2 \log d + \sigma_{\mathsf{row}}^2, \\
\max_{i\in[d_2]} \|\boldsymbol{E}_{i,:}\boldsymbol{W}_1\|_2 &\lesssim \left(B\log d + \sigma_{\mathsf{row}}\sqrt{\log d}\right)\|\boldsymbol{W}_1\|_{2,\infty}, \\
\max_{j\in[d_2]} \sum_{i\in[d_1]} E_{i,j}^2 &\lesssim B^2 \log d + \sigma_{\mathsf{col}}^2, \\
\max_{j\in[d_2]} \left\|\left(\boldsymbol{E}_{:,j}\right)^{\top}\boldsymbol{W}_2\right\|_2 &\lesssim \left(B\log d + \sigma_{\mathsf{col}}\sqrt{\log d}\right)\|\boldsymbol{W}_2\|_{2,\infty},
\end{aligned}$$

where σ_{row}, σ_{col}, and B are respectively upper bounds on $\max_{i\in[d_1]}\sqrt{\sum_{j\in[d_2]}\mathbb{E}\left[E_{i,j}^2\right]}$, $\max_{j\in[d_2]}\sqrt{\sum_{i\in[d_1]}\mathbb{E}\left[E_{i,j}^2\right]}$, and $\max_{i\in[d_1],j\in[d_2]}|E_{i,j}|$; see (54) for precise definitions. As a result, one has

$$\begin{aligned}
\|\boldsymbol{E}\|_{2,\infty} &\lesssim B\sqrt{\log d} + \sigma_{\mathsf{row}}, \\
\|\boldsymbol{A}^{\mathsf{s}}\|_{2,\infty} &\lesssim \|\boldsymbol{A}^{\star}\|_{2,\infty} + B\sqrt{\log d} + \sigma_{\mathsf{row}}, \\
\left\|\boldsymbol{E}^{\top}\right\|_{2,\infty} &\lesssim B\sqrt{\log d} + \sigma_{\mathsf{col}}, \\
\left\|\boldsymbol{A}^{\mathsf{s}\top}\right\|_{2,\infty} &\lesssim \left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty} + B\sqrt{\log d} + \sigma_{\mathsf{col}},
\end{aligned}$$

where $\boldsymbol{A}^{\mathsf{s}} = \boldsymbol{A}^{\star} + \boldsymbol{E}$ is defined in (50).

Lemma 13. *With probability greater than $1 - O\left(d^{-20}\right)$, one has*

$$\|\boldsymbol{E}\| \lesssim B\log d + (\sigma_{\mathsf{row}} + \sigma_{\mathsf{col}})\sqrt{\log d},$$

where B, σ_{row} and σ_{col} are defined in (54).

Lemma 14. *Fix any vector $\boldsymbol{w} \in \mathbb{R}^{d_2}$. With probability at least $1 - O\left(d^{-20}\right)$, one has*

$$\sum_{i\in[d_1]}\left(\sum_{j\in[d_2]} w_j E_{i,j}\right)^2 \lesssim \|\boldsymbol{w}\|_2^2\left(\sigma_{\mathsf{col}}^2 + \sigma_\infty^2 \log^2 d\right) + \|\boldsymbol{w}\|_\infty^2 B^2 \log^3 d,$$

where B, σ_∞ and σ_{col} are defined in (54).

13.1 Proof of Lemma 11

Given the SVD of $\boldsymbol{A}^\star = \boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\boldsymbol{V}^{\star\top}$, one has $\boldsymbol{G}^\star = \boldsymbol{A}^\star\boldsymbol{A}^{\star\top} = \boldsymbol{U}^\star\boldsymbol{\Sigma}^{\star 2}\boldsymbol{U}^{\star\top}$. Using the definition of the incoherence parameters, one can derive

$$\begin{aligned}
\|\boldsymbol{A}^\star\|_{2,\infty} &= \max_{i\in[d_1]}\left\|\boldsymbol{U}_{i,:}^\star\boldsymbol{\Sigma}^\star\boldsymbol{V}^{\star\top}\right\|_2 \le \max_{i\in[d_1]}\left\|\boldsymbol{U}_{i,:}^\star\right\|_2\|\boldsymbol{\Sigma}^\star\|\,\|\boldsymbol{V}^\star\| \le \sigma_1^\star\|\boldsymbol{U}^\star\|_{2,\infty} \le \sqrt{\frac{\mu_1 r\sigma_1^{\star 2}}{d_1}};\\
\left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty} &= \max_{j\in[d_2]}\left\|\boldsymbol{V}_{i,:}^\star\boldsymbol{\Sigma}^\star\boldsymbol{U}^{\star\top}\right\|_2 \le \max_{j\in[d_2]}\left\|\boldsymbol{V}_{j,:}^\star\right\|_2\|\boldsymbol{\Sigma}^\star\|\,\|\boldsymbol{U}^\star\| \le \sigma_1^\star\|\boldsymbol{V}^\star\|_{2,\infty} \le \sqrt{\frac{\mu_2 r\sigma_1^{\star 2}}{d_2}};\\
\|\boldsymbol{G}^\star\|_{2,\infty} &= \max_{i\in[d_1]}\left\|\boldsymbol{U}_{i,:}^\star\boldsymbol{\Sigma}^{\star 2}\boldsymbol{U}^{\star\top}\right\|_2 \le \max_{i\in[d_1]}\left\|\boldsymbol{U}_{i,:}^\star\right\|_2\left\|\boldsymbol{\Sigma}^{\star 2}\right\|\|\boldsymbol{U}^\star\| \le \sigma_1^{\star 2}\|\boldsymbol{U}^\star\|_{2,\infty} \le \sqrt{\frac{\mu_1 r\sigma_1^{\star 4}}{d_1}}.
\end{aligned}$$

Moreover, the Cauchy-Schwartz inequality allows one to upper bound

$$\|\boldsymbol{A}^\star\|_\infty = \max_{(i,j)\in[d_1]\times[d_2]}\left|\boldsymbol{U}_{i,:}^\star\boldsymbol{\Sigma}^\star\left(\boldsymbol{V}_{j,:}^\star\right)^\top\right| \le \|\boldsymbol{U}^\star\|_{2,\infty}\|\boldsymbol{\Sigma}^\star\|\,\|\boldsymbol{V}^\star\|_{2,\infty} \le \sigma_1^\star\|\boldsymbol{U}^\star\|_{2,\infty}\|\boldsymbol{V}^\star\|_{2,\infty}.$$

In view of the simple bounds $\|\boldsymbol{U}^\star\|_{2,\infty} \le \|\boldsymbol{U}^\star\| \le 1$ and $\|\boldsymbol{V}^\star\|_{2,\infty} \le \|\boldsymbol{V}^\star\| \le 1$, we conclude that

$$\|\boldsymbol{A}^\star\|_\infty \le \sigma_1^\star\|\boldsymbol{U}^\star\|_{2,\infty} \qquad \text{and} \qquad \|\boldsymbol{A}^\star\|_\infty \le \sigma_1^\star\|\boldsymbol{V}^\star\|_{2,\infty}.$$

13.2 Proof of Lemma 12

We shall only prove the results concerning σ_{col}; the results concerning σ_{row} follow immediately via nearly identical arguments.

In view of the Bernstein inequality, we have

$$\mathbb{P}\left\{\left|\sum\nolimits_{i\in[d_1]} E_{i,j}^2 - M_1\right| \ge t\right\} \le 2\exp\left(-\frac{3}{8}\min\left\{\frac{t^2}{V_1}, \frac{t}{L_1}\right\}\right), \quad t > 0,$$

where M_1, L_1 and S_1 are given respectively by

$$\begin{aligned}
M_1 &:= \sum\nolimits_{i\in[d_1]}\mathbb{E}\left[E_{i,l}^2\right] \le \sigma_{\mathsf{col}}^2,\\
L_1 &:= \max_{i\in[d_1]}\left|E_{i,l}^2 - \mathbb{E}\left[E_{i,l}^2\right]\right| \le B^2 + \sigma_\infty^2 \le 2B^2,\\
V_1 &:= \sum\nolimits_{i\in[d_1]}\mathsf{Var}\left(E_{i,l}^2\right) \le \sum\nolimits_{i\in[d_1]}\mathbb{E}\left[E_{i,l}^4\right] \le B^2\sigma_{\mathsf{col}}^2.
\end{aligned}$$

Here, we have made use of the fact that $\sigma_\infty \le B$. As a result, one has

$$\begin{aligned}
\sum\nolimits_{i\in[d_1]} E_{i,j}^2 &\lesssim M_1 + L_1\log d + \sqrt{V_1\log d} \lesssim \sigma_{\mathsf{col}}^2 + B^2\log d + B\sigma_{\mathsf{col}}\sqrt{\log d}\\
&\asymp \sigma_{\mathsf{col}}^2 + B^2\log d
\end{aligned}$$

with probability exceeding $1 - O\left(d^{-20}\right)$, where the last line arises from the AM-GM inequality (namely, $2B\sigma_{\mathsf{col}}\sqrt{\log d} \le \sigma_{\mathsf{col}}^2 + B^2\log d$). As an immediate consequence, with probability at least $1 - O\left(d^{-20}\right)$,

$$\|\boldsymbol{E}_{:,j}\|_2 = \sqrt{\sum\nolimits_{i\in[d_1]} E_{i,j}^2} \lesssim \sigma_{\mathsf{col}} + B\sqrt{\log d},$$

$$\left\|\boldsymbol{A}^{\mathsf{s}}_{:,j}\right\|_2 \leq \left\|\boldsymbol{A}^{\star}_{:,j}\right\|_2 + \left\|\boldsymbol{E}_{:,j}\right\|_2 \lesssim \left\|\boldsymbol{A}^{\star\top}\right\|_{2,\infty} + \sigma_{\mathsf{col}} + B\sqrt{\log d}.$$

Next, we turn to the claim concerning a fixed matrix $\boldsymbol{W}_2$. Observe that $\left(\boldsymbol{E}_{:,l}\right)^\top \boldsymbol{W}_2 = \sum_{i\in[d_1]} E_{i,l}(\boldsymbol{W}_2)_{i,:}$ is a sum of independent zero-mean random vectors. In order to invoke standard concentration inequalities, we compute

$$\begin{aligned}
L_2 &:= \max_{i\in[d_1]} \left\|E_{i,l}(\boldsymbol{W}_2)_{i,:}\right\|_2 \leq B\left\|\boldsymbol{W}_2\right\|_{2,\infty}, \\
V_2 &:= \sum\nolimits_{i\in[d_1]} \mathbb{E}\left[E_{i,l}^2\right]\left\|(\boldsymbol{W}_2)_{i,:}\right\|_2^2 \leq \sigma_{\mathsf{col}}^2 \left\|\boldsymbol{W}_2\right\|_{2,\infty}^2.
\end{aligned}$$

Invoking the matrix Bernstein inequality yields that with probability exceeding $1-O\left(d^{-20}\right)$,

$$\left\|\left(\boldsymbol{E}_{:,l}\right)^\top \boldsymbol{W}_2\right\|_2 \lesssim L_2\log d + \sqrt{V_2\log d} \lesssim \left(B\log d + \sigma_{\mathsf{col}}\sqrt{\log d}\right)\left\|\boldsymbol{W}_2\right\|_{2,\infty}.$$

13.3 Proof of Lemma 13

First, we can write

$$\boldsymbol{E} = \sum_{i\in[d_1],j\in[d_2]} E_{i,j}\boldsymbol{e}_i\boldsymbol{e}_j^\top$$

as a sum of independent zero-mean random matrices (since $\mathbb{E}[E_{i,j}]=0$). We make the observation that

$$\begin{aligned}
L &:= \max_{i\in[d_1],j\in[d_2]} \left\|E_{i,j}\boldsymbol{e}_i\boldsymbol{e}_j^\top\right\| \leq B; \\
V &:= \max\left\{\left\|\sum_{i\in[d_1],j\in[d_2]} \mathbb{E}\left[E_{i,j}^2\right]\boldsymbol{e}_i\boldsymbol{e}_i^\top\right\|, \left\|\sum_{i\in[d_1],j\in[d_2]} \mathbb{E}\left[E_{i,j}^2\right]\boldsymbol{e}_j\boldsymbol{e}_j^\top\right\|\right\} \leq \sigma_{\mathsf{row}}^2 + \sigma_{\mathsf{col}}^2.
\end{aligned}$$

It then follows from the matrix Bernstein inequality that, with probability at least $1-O\left(d^{-10}\right)$,

$$\|\boldsymbol{E}\| \lesssim L\log d + \sqrt{V\log d} \lesssim B\log d + \left(\sigma_{\mathsf{row}} + \sigma_{\mathsf{col}}\right)\sqrt{\log d}.$$

13.4 Proof of Lemma 14

Let us define a sequence of independent zero-mean random variables $\{X_i\}_{1\leq i\leq d_1}$ as follows

$$X_i := \sum_{j\in[d_2]} w_j E_{i,j}.$$

It is easily seen that

$$\max_{j\in[d_2]} |w_j E_{i,j}| \leq \|\boldsymbol{w}\|_\infty B; \quad \mathbb{E}\left[X_i^2\right] = \sum_{j\in[d_2]} w_j^2\sigma_{i,j}^2 \leq \|\boldsymbol{w}\|_2^2\sigma_\infty^2.$$

We can therefore apply the Bernstein inequality to show that, with probability at least $1-O\left(d^{-11}\right)$,

$$|X_i| \lesssim \left(\|\boldsymbol{w}\|_\infty B\right)\log d + \sqrt{\left(\|\boldsymbol{w}\|_2^2\sigma_\infty^2\right)\log d} =: R. \tag{142}$$

Next, let us introduce a sequence of independent random variables $\{Y_i\}_{1\leq i\leq d_1}$, obtained by truncating X_i

$$Y_i \triangleq X_i\,\mathbb{1}\left\{|X_i| \leq CR\right\}$$

for some sufficiently large absolute constant $C>0$. From (142) and the union bound, we know that $Y_i = X_i$ holds simultaneously for all $1\leq i\leq d_1$ with probability at least $1-O\left(d^{-10}\right)$.

Further, it is straightforward to compute that

$$M_2 := \sum_{i\in[d_1]} \mathbb{E}\left[Y_i^2\right] \le \sum_{i\in[d_1]} \mathbb{E}\left[X_i^2\right] \le \sum_{i\in[d_1],j\in[d_2]} w_j^2 \sigma_{i,j}^2 \le \|\boldsymbol{w}\|_2^2 \sigma_{\mathsf{col}}^2;$$

$$L_2 := \max_{i\in[d_1]} \left|Y_i^2 - \mathbb{E}\left[Y_i^2\right]\right| \lesssim R^2 \lesssim \|\boldsymbol{w}\|_\infty^2 B^2 \log^2 d + \|\boldsymbol{w}\|_2^2 \sigma_\infty^2 \log d;$$

$$V_2 := \sum_{i\in[d_1]} \mathsf{Var}\left(Y_i^2\right) \le \sum_{i\in[d_1]} \mathbb{E}\left[Y_i^4\right] \le \sum_{i\in[d_1]} \mathbb{E}\left[X_i^4\right] \lesssim \sum_{i\in[d_1]} \sum_{j\in[d_2]} w_j^4 \mathbb{E}\left[E_{i,j}^4\right] + \sum_{i\in[d_1]} \sum_{j_1\neq j_2} w_{j_1}^2 w_{j_2}^2 \mathbb{E}\left[E_{i,j_1}^2\right] \mathbb{E}\left[E_{i,j_2}^2\right]$$

$$\lesssim \|\boldsymbol{w}\|_\infty^2 \|\boldsymbol{w}\|_2^2 B^2 \sigma_{\mathsf{col}}^2 + \|\boldsymbol{w}\|_2^4 \sigma_\infty^2 \sigma_{\mathsf{col}}^2.$$

We then apply the Bernstein inequality to conclude that with probability at least $1 - O\left(d^{-10}\right)$:

$$\sum_{i\in[d_1]} Y_i^2 \lesssim M_2 + L_2 \log d + \sqrt{V_2 \log d}$$

$$\lesssim \|\boldsymbol{w}\|_2^2 \sigma_{\mathsf{col}}^2 + \|\boldsymbol{w}\|_\infty^2 B^2 \log^3 d + \|\boldsymbol{w}\|_2^2 \sigma_\infty^2 \log^2 d + \left(\|\boldsymbol{w}\|_\infty \|\boldsymbol{w}\|_2 B\sigma_{\mathsf{col}} + \|\boldsymbol{w}\|_2^2 \sigma_\infty \sigma_{\mathsf{col}}\right) \sqrt{\log d}$$

$$\asymp \|\boldsymbol{w}\|_2^2 \left(\sigma_{\mathsf{col}}^2 + \sigma_\infty^2 \log^2 d\right) + \|\boldsymbol{w}\|_\infty^2 B^2 \log^3 d,$$

where the last line arises from the AM-GM inequality (namely, $2\|\boldsymbol{w}\|_\infty \|\boldsymbol{w}\|_2 B\sigma_{\mathsf{col}}\sqrt{\log d} \le \|\boldsymbol{w}\|_\infty^2 B^2 \log d + \|\boldsymbol{w}\|_2^2 \sigma_{\mathsf{col}}^2$ and $2\|\boldsymbol{w}\|_2^2 \sigma_\infty \sigma_{\mathsf{col}} \sqrt{\log d} \le \|\boldsymbol{w}\|_2^2 \sigma_\infty^2 \log d + \|\boldsymbol{w}\|_2^2 \sigma_{\mathsf{col}}^2$).

References

[AFWZ17] E. Abbe, J. Fan, K. Wang, and Y. Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *arXiv:1709.09565*, 2017.

[CCFM19] Y. Chen, Y. Chi, J. Fan, and C. Ma. Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval. *Mathematical Programming*, 176(1-2):5–37, July 2019.

[CFMW19] Y. Chen, J. Fan, C. Ma, and K. Wang. Spectral method and regularized MLE are both optimal for top-K ranking. *Annals of Statistics*, 47(4):2204–2235, August 2019.

[CFMY19] Y. Chen, J. Fan, C. Ma, and Y. Yan. Inference and uncertainty quantification for noisy matrix completion. *Proceedings of the National Academy of Sciences*, 116(46):22931–22937, 2019.

[CLL19] J. Chen, D. Liu, and X. Li. Nonconvex rectangular matrix completion via gradient descent without $\ell_{2,\infty}$ regularization. *arXiv preprint arXiv:1901.06116*, 2019.

[DK70] C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

[EK15] N. El Karoui. On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators. *Probability Theory and Related Fields*, pages 1–81, 2015.

[EKBB+13] N. El Karoui, D. Bean, P. J. Bickel, C. Lim, and B. Yu. On robust regression with high-dimensional predictors. *Proceedings of the National Academy of Sciences*, 110(36):14557–14562, 2013.

[HSSS16] S. B. Hopkins, T. Schramm, J. Shi, and D. Steurer. Fast spectral algorithms from sum-of-squares proofs: tensor decomposition and planted sparse vectors. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 178–191. ACM, 2016.

[Joh12] T. Johansson. The giant component of the random bipartite graph. 2012.

[LBEK18] L. Lei, P. J. Bickel, and N. El Karoui. Asymptotics for high dimensional regression m-estimates: fixed design results. *Probability Theory and Related Fields*, 172(3-4):983–1079, 2018.

[Mas07] P. Massart. *Concentration inequalities and model selection*, volume 6. Springer, 2007.

[MWCC17] C. Ma, K. Wang, Y. Chi, and Y. Chen. Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion and blind deconvolution. *arXiv preprint arXiv:1711.10467*, 2017.

[PW19] A. Pananjady and M. J. Wainwright. Value function estimation in Markov reward processes: Instance-dependent ℓ_∞-bounds for policy evaluation. *arXiv preprint arXiv:1909.08749*, 2019.

[SCC17] P. Sur, Y. Chen, and E. J. Candès. The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square. *Probability Theory and Related Fields*, pages 1–72, 2017.

[SS90] G. W. Stewart and J.-g. Sun. Matrix perturbation theory. 1990.

[TB77] J. M. Ten Berge. Orthogonal procrustes rotation for two or more matrices. *Psychometrika*, 42(2):267–276, 1977.

[Tsy08] A. B. Tsybakov. *Introduction to nonparametric estimation.* Springer Science & Business Media, 2008.

[Ver12] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. *Compressed Sensing, Theory and Applications*, pages 210 – 268, 2012.

[ZB18] Y. Zhong and N. Boumal. Near-optimal bounds for phase synchronization. *SIAM Journal on Optimization*, 28(2):989–1016, 2018.