Inference for Heteroskedastic PCA with Missing Data

Yuling Yan* Yuxin Chen† Jianqing Fan*

July 26, 2021

Abstract

This paper studies how to construct confidence regions for principal component analysis (PCA) in high dimension, a problem that has been vastly under-explored. While computing measures of uncertainty for nonlinear/nonconvex estimators is in general difficult in high dimension, the challenge is further compounded by the prevalent presence of missing data and heteroskedastic noise. We propose a suite of solutions to perform valid inference on the principal subspace based on two estimators: a vanilla SVD-based approach, and a more refined iterative scheme called HeteroPCA (Zhang et al., 2018). We develop non-asymptotic distributional guarantees for both estimators, and demonstrate how these can be invoked to compute both confidence regions for the principal subspace and entrywise confidence intervals for the spiked covariance matrix. Particularly worth highlighting is the inference procedure built on top of HeteroPCA, which is not only valid but also statistically efficient for broader scenarios (e.g., it covers a wider range of missing rates and signal-to-noise ratios). Our solutions are fully data-driven and adaptive to heteroskedastic random noise, without requiring prior knowledge about the noise levels and noise distributions.

Keywords: principal component analysis, confidence regions, missing data, uncertainty quantification, heteroskedastic data, matrix denoising, subspace estimation

Contents

Corresponding author: Yuxin Chen (Email: `yuxin.chen@princeton.edu`).

*Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA; Email: `{yulingy,jqfan}@princeton.edu`.

†Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA; Email: `yuxin.chen@princeton.edu`.

1 Introduction

The applications of modern data science frequently ask for succinct representations of high-dimensional data. At the core of this pursuit lies principal component analysis (PCA), which serves as an effective means of dimension reduction and has been deployed across a broad range of domains (Fan et al., 2021b; Johnstone and Paul, 2018; Jolliffe, 1986; Vaswani et al., 2018). In reality, data collection could often be far from ideal — for instance, the acquired data might be subject to random contamination and contain incomplete observations — which inevitably affects the fidelity of PCA and calls for additional care when interpreting the results. To enable informative assessment of the influence of imperfect data acquisition, it would be desirable to accompany the PCA estimators in use with valid measures of uncertainty or "confidence".

1.1 Problem formulation

To allow for concrete and precise studies, the present paper concentrates on a tractable model that captures the effects of random heteroskedastic noise and missing data in PCA. In what follows, we start by formulating the problem, in the hope of facilitating more precise discussions.

Model. Imagine we are interested in n independent random vectors $\boldsymbol{x}_j = [x_{1,j}, \cdots, x_{d,j}]^\top \in \mathbb{R}^d$ drawn from the following distribution

$$\boldsymbol{x}_j \overset{\text{ind.}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{S}^\star), \qquad 1 \le j \le n, \tag{1.1}$$

where the unknown covariance matrix $\boldsymbol{S}^\star \in \mathbb{R}^{d\times d}$ is assumed to be rank-r ($r < n$) with eigen-decomposition

$$\boldsymbol{S}^\star = \boldsymbol{U}^\star \boldsymbol{\Lambda}^\star \boldsymbol{U}^{\star\top}. \tag{1.2}$$

Here, the orthonormal columns of $\boldsymbol{U}^\star \in \mathbb{R}^{d\times r}$ constitute the r leading eigenvectors of $\boldsymbol{S}^\star$, whereas $\boldsymbol{\Lambda}^\star \in \mathbb{R}^{r\times r}$ is a diagonal matrix whose diagonal entries are composed of the non-zero eigenvalues of $\boldsymbol{S}^\star$. In other words, these vectors $\{\boldsymbol{x}_j\}_{1\le j\le n}$ are randomly drawn from a low-dimensional subspace when r is small. What we have available are partial and randomly corrupted observations of the entries of the above vectors. Specifically, suppose that we only get to observe

$$y_{l,j} = x_{l,j} + \eta_{l,j} \qquad \text{for all } (l,j) \in \Omega \tag{1.3}$$

over a subsampled index set $\Omega \subseteq [d] \times [n]$ (with $[n] \coloneqq \{1, \cdots, n\}$), where $\eta_{l,j}$ represents the noise that contaminates the observation in this location. Throughout this paper, we assume random sampling and random noise as follows.

- *Random sampling*: each index (l, j) is contained in Ω independently with probability p;
- *Heteroskedastic random noise with unknown variance*: the noise components $\{\eta_{l,j}\}$ are independently generated sub-Gaussian random variables obeying

  $$\mathbb{E}[\eta_{l,j}] = 0, \qquad \mathbb{E}[\eta_{l,j}^2] = \omega_l^{\star 2}, \qquad \text{and} \qquad \|\eta_{l,j}\|_{\psi_2} = O(\omega_l^\star),$$

  where $\{\omega_l^\star\}_{1\le l\le d}$ denote the standard deviations that are *a priori* unknown, and $\|\cdot\|_{\psi_2}$ stands for the sub-Gaussian norm of a random variable (Vershynin, 2017). The noise levels $\{\omega_l^\star\}_{1\le l\le d}$ are allowed to vary across locations, so as to model the so-called *heteroskedasticity* of noise.

With the observed data $\{y_{l,j} \mid (l,j) \in \Omega\}$ in hand, can we perform statistical inference on the orthonormal matrix $\boldsymbol{U}^\star$ — which embodies the ground-truth r-dimensional principal subspace underlying the vectors $\{\boldsymbol{x}_j\}_{1\le j\le n}$ — and make inference on the underlying covariance matrix $\boldsymbol{S}^\star$? Mathematically, the task can often be phrased as constructing valid confidence intervals/regions for both $\boldsymbol{U}^\star$ and $\boldsymbol{S}^\star$ based on the incomplete and corrupted observations $\{y_{l,j} \mid (l,j) \in \Omega\}$. Noteworthily, this model is frequently studied in econometrics and financial modeling under the name of factor models (Bai and Wang, 2016; Fan et al., 2021a; Fan and Yao, 2017; Gagliardini et al., 2019), and is closely related to the noisy matrix completion problem (except that the current goal is not to reconstruct all missing data) (Candès and Plan, 2010; Candès and Recht, 2009; Chi et al., 2019; Keshavan et al., 2010b).

Inadequacy of prior works. While methods for estimating principal subspace are certainly not in shortage (e.g., Balzano et al. (2018); Cai et al. (2021); Cai and Zhang (2018); Li et al. (2021); Lounici (2014); Zhang et al. (2018); Zhu et al. (2019)), methods for constructing confidence regions for principal subspace remain vastly under-explored. The fact that the estimators in use for PCA are typically nonlinear and nonconvex presents a substantial challenge in the development of a distributional theory, let alone uncertainty quantification. As some representative recent attempts, Bao et al. (2018); Xia (2019b) established normal approximations of the distance between the true subspace and its estimate for the matrix denoising setting, while Koltchinskii et al. (2020) further established asymptotic normality of some debiased estimator for linear functions of principal components. These distributional guarantees pave the way for the development of statistical inference procedures for PCA. However, it is noteworthy that these results required the noise components to either be i.i.d. Gaussian or at least exhibit matching moments (up to the 4th order), which fell short of accommodating heteroskedastic noise. The challenge is further compounded when statistical inference needs to be conducted in the face of missing data, a scenario that is beyond the reach of these prior works.

1.2 Our contributions

In light of the insufficiency of prior results, this paper takes a step towards developing data-driven inference and uncertainty quantification procedures for PCA, in the hope of accommodating both heteroskedastic noise and missing data. Our inference procedures are built on top of two prior estimation algorithms — a vanilla approach based on singular value decomposition (SVD), and a more refined iterative scheme called HeteroPCA — to be detailed in Section 2. Our main contributions are summarized as follows.

- *Distributional theory for PCA and covariance estimation.* Focusing on the aforementioned two estimators, we derive, in a non-asymptotic manner, row-wise distributional characterizations of the principal subspace estimate, as well as entrywise distributions of the covariance matrix of $\{\boldsymbol{x}_l\}_{1\leq l\leq n}$. These distributional characterizations take the form of certain tractable Gaussian approximations centered at the ground truth.

- *Fine-grained confidence regions and intervals.* Our distributional theory in turns allows for construction of row-wise confidence region for the subspace $\boldsymbol{U}^\star$ (up to global rotation) as well as entrywise confidence intervals for the matrix $\boldsymbol{S}^\star$. The proposed inference procedures are fully data-driven and do not require prior knowledge of the noise levels and noise distributions.

In particular, our findings highlight the advantages of the inference method based on HeteroPCA: it is theoretically appealing for a much broader range of scenarios (e.g., it allows for a higher missing rate and smaller signal-to-noise ratio) when compared with the SVD-based approach, and it uniformly outperforms the SVD-based approach in all our numerical experiments. Further, all of our theory allows the observed data to be highly incomplete and covers heteroskedastic noise, which is previously unavailable.

1.3 Paper organization

The remainder of the paper is organized as follows. In Section 2, we introduce two choices of estimation algorithms available in prior literature. Aimed at conducting statistical inference and uncertainty quantification based on these two estimators, Section 3 develops a suite of distributional theory and demonstrates how to use it to construct fine-grained confidence regions and confidence intervals for the unknowns; the detailed proofs of our theorems are deferred to the appendices. In Section 4, we carry out a series of numerical experiments to confirm the validity and applicability of our theoretical findings. Section 5 gives an overview of several related works. Section 6 takes a detour to analyze two intimately related problems, which will then be utilized to establish our main results. We conclude the paper with a discussion of future directions in Section 7. Most of the proof details are deferred to the appendix.

1.4 Notation

Before proceeding, we introduce several notation that will be useful throughout. We let $f(n) \lesssim g(n)$ or $f(n) = O(g(n))$ represent the condition that $|f(n)| \leq Cg(n)$ for some constant $C > 0$ when n is sufficiently large; we use $f(n) \gtrsim g(n)$ to denote $f(n) \geq C|g(n)|$ for some constant $C > 0$ when n is sufficiently large; and we let $f(n) \asymp g(n)$ indicate that $f(n) \lesssim g(n)$ and $f(n) \gtrsim g(n)$ hold simultaneously. The notation $f(n) \gg g(n)$ (resp. $f(n) \ll g(n)$) means that there exists some sufficiently large (resp. small) constant $c_1 > 0$ (resp. $c_2 > 0$) such that $f(n) \geq c_1 g(n)$ (resp. $f(n) \leq c_2 g(n)$). For any real number $a, b \in \mathbb{R}$, we shall define $a \wedge b \coloneqq \min\{a, b\}$ and $a \vee b \coloneqq \max\{a, b\}$.

For any matrix $\boldsymbol{M} = [M_{i,j}]_{1 \leq i \leq n_1, 1 \leq j \leq n_2}$, we let $\boldsymbol{M}_{i,\cdot}$ and $\boldsymbol{M}_{\cdot,j}$ stand for the i-th row and the j-th column of $\boldsymbol{M}$, respectively. We shall also let $\|\boldsymbol{M}\|$, $\|\boldsymbol{M}\|_{\mathrm{F}}$, $\|\boldsymbol{M}\|_{2,\infty}$ and $\|\boldsymbol{M}\|_{\infty}$ denote the spectral norm, the Frobenius norm, the $\ell_{2,\infty}$ norm (i.e., $\|\boldsymbol{M}\|_{2,\infty} \coloneqq \max_i \|\boldsymbol{M}_{i,\cdot}\|_2$), and the entrywise ℓ_∞ norm ($\|\boldsymbol{M}\|_\infty \coloneqq \max_{i,j} |M_{i,j}|$) of $\boldsymbol{M}$, respectively. For any index set Ω, the notation $\mathcal{P}_\Omega(\boldsymbol{M})$ represents the Euclidean projection of a matrix $\boldsymbol{M}$ onto the subspace of matrices supported on Ω, and define $\mathcal{P}_{\Omega^{\mathrm{c}}}(\boldsymbol{M}) \coloneqq \boldsymbol{M} - \mathcal{P}_\Omega(\boldsymbol{M})$ as well. In addition, we denote by $\mathcal{P}_{\mathsf{diag}}(\boldsymbol{G})$ the Euclidean projection of a square matrix $\boldsymbol{G}$ onto the subspace of matrices that vanish outside the diagonal, and define $\mathcal{P}_{\mathsf{off\text{-}diag}}(\boldsymbol{G}) \coloneqq \boldsymbol{G} - \mathcal{P}_{\mathsf{diag}}(\boldsymbol{G})$. For a non-singular matrix $\boldsymbol{H} \in \mathbb{R}^{k \times k}$ with SVD $\boldsymbol{U}_H \boldsymbol{\Sigma}_H \boldsymbol{V}_H^\top$, we denote by $\mathsf{sgn}(\boldsymbol{H})$ the following orthogonal matrix

$$\mathsf{sgn}(\boldsymbol{H}) \coloneqq \boldsymbol{U}_H \boldsymbol{V}_H^\top. \tag{1.4}$$

Finally, we denote by $\mathscr{C}^d$ the set of all convex sets in $\mathbb{R}^d$. For any Lebesgue measurable set $\mathcal{A} \subseteq \mathbb{R}^d$, we adopt the shorthand notation $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})\{\mathcal{A}\} \coloneqq \mathbb{P}(\boldsymbol{z} \in \mathcal{A})$, where $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Throughout this paper, we let $\Phi(\cdot)$ (resp. $\phi(\cdot)$) represent the cumulative distribution function (resp. probability distribution function) of the standard Gaussian distribution. We also denote by $\chi^2(k)$ the chi-square distribution with k degrees of freedom.

2 Background: two estimation algorithms for PCA

In order to conduct statistical inference for PCA, the first step lies in selecting an algorithm to estimate the principal subspace and the covariance matrix of interest, which we elaborate on in this section. Before continuing, we introduce several useful matrix notation as follows

$$\boldsymbol{X} \coloneqq [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n] \in \mathbb{R}^{d \times n}, \tag{2.1a}$$

$$\boldsymbol{Y} \coloneqq \mathcal{P}_\Omega(\boldsymbol{X} + \boldsymbol{N}) \in \mathbb{R}^{d \times n}, \tag{2.1b}$$

where $\mathcal{P}_\Omega$ has been defined in Section 1.4, and $\boldsymbol{N} \in \mathbb{R}^{d \times n}$ represents the noise matrix such that the (l, j)-th entry of $\boldsymbol{N}$ is given by $\eta_{l,j}$. In other words, $\boldsymbol{Y}$ encapsulates all the observed data $\{y_{l,j} \mid (l, j) \in \Omega\}$, with any entry outside Ω taken to be zero. If one has full access to the noiseless data matrix $\boldsymbol{X}$, then a natural strategy to estimate $\boldsymbol{U}^\star$ would be to return the top-r eigenspace of the sample covariance matrix $n^{-1}\boldsymbol{X}\boldsymbol{X}^\top$, or equivalently, the top-r left singular subspace of $\boldsymbol{X}$. In practice, however, one needs to extract information from the corrupted and incomplete data matrix $\boldsymbol{Y}$.

The first algorithm: a vanilla SVD-based approach. Given that $p^{-1}\boldsymbol{Y} = p^{-1}\mathcal{P}_\Omega(\boldsymbol{X} + \boldsymbol{N})$ is an unbiased estimate of $\boldsymbol{X}$ (conditional on $\boldsymbol{X}$), a natural idea that comes into mind is to resort to the top-r left singular subspace of $p^{-1}\boldsymbol{Y}$ when estimating $\boldsymbol{U}^\star$. This simple procedure is summarized in Algorithm 1.

Algorithm 1 An SVD-based approach.

Input: sampling set Ω, data matrix $\boldsymbol{Y}$ (cf. (2.1b)), sampling rate p, rank r.
Compute the truncated rank-r SVD $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ of $p^{-1}\boldsymbol{Y}/\sqrt{n}$, where $\boldsymbol{U} \in \mathbb{R}^{d \times r}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{r \times r}$ and $\boldsymbol{V} \in \mathbb{R}^{n \times r}$.
Output $\boldsymbol{U}$ as the subspace estimate, $\boldsymbol{\Sigma}$ as an estimate of $\boldsymbol{\Sigma}^\star$, and $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^\top$ as the covariance matrix estimate.

As we shall see momentarily, Algorithm 1 returns reliable estimates of $\boldsymbol{U}^\star$ and $\boldsymbol{S}^\star$ in the regime of moderate-to-high signal-to-noise ratio (SNR), but might fail to be effective if either the missing rate $1 - p$ or the noise levels are too large. To offer a high-level explanation, we find it helpful to compute

$$\frac{1}{p^2}\boldsymbol{E}\left[\boldsymbol{Y}\boldsymbol{Y}^\top \,\middle|\, \boldsymbol{X}\right] = \boldsymbol{X}\boldsymbol{X}^\top + \left(\frac{1}{p} - 1\right)\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{X}\boldsymbol{X}^\top\right) + \frac{n}{p}\mathsf{diag}\left\{\left[\omega_l^{\star 2}\right]_{1 \leq l \leq d}\right\}, \tag{2.2}$$

where for any vector $\boldsymbol{z} = [z_l]_{1\leq l\leq d}$ we denote by $\mathsf{diag}(\boldsymbol{z}) \in \mathbb{R}^{d\times d}$ a diagonal matrix whose (l,l)-th entry equals z_l. If the sampling rate p is overly small and/or if the noise is of large size but heteroskedastic, then the second and the third terms on the right-hand side of (2.2) might result in significant bias on the diagonal of the matrix $\boldsymbol{E}\left[\boldsymbol{Y}\boldsymbol{Y}^\top \,\middle|\, \boldsymbol{X}\right]$, thus hampering the statistical accuracy of the eigenspace of $p^{-2}\boldsymbol{Y}\boldsymbol{Y}^\top$ (or equivalently, the left singular space of $p^{-1}\boldsymbol{Y}$) when employed to estimate $\boldsymbol{U}^\star$. Viewed in this light, a more effective estimation algorithm would need to include procedures that properly handle the diagonal components of $p^{-2}\boldsymbol{Y}\boldsymbol{Y}^\top$. The next algorithm epitomizes this idea.

The second algorithm: HeteroPCA. To remedy the above-mentioned issue, several previous works (e.g., Cai et al. (2021); Florescu and Perkins (2016)) adopted a spectral method with diagonal deletion, which essentially discards any diagonal entry of $p^{-2}\boldsymbol{Y}\boldsymbol{Y}^\top$ before computing its top-r eigenspace. However, diagonal deletion comes at a price: while this operation mitigates the significant bias due to heteroskedasticity and missing data, it introduces another type of bias that might be non-negligible if the goal is to enable efficient fine-grained inference. To address this bias issue, Zhang et al. (2018) proposed an iterative refinement scheme — termed HeteroPCA — that copes with the diagonal entries in a more refined manner. Informally, HeteroPCA starts by computing the rank-r eigenspace of the diagonal-deleted version of $p^{-2}\boldsymbol{Y}\boldsymbol{Y}^\top$, and then alternates between imputing the diagonal entries of $\boldsymbol{X}\boldsymbol{X}^\top$ and estimating the eigenspace of $p^{-2}\boldsymbol{Y}\boldsymbol{Y}^\top$ with the aid of the imputed diagonal. A precise description of this procedure is summarized in Algorithm 2; here, $\mathcal{P}_{\textsf{off-diag}}$ and $\mathcal{P}_{\textsf{diag}}$ have been defined in Section 1.4.

Algorithm 2 HeteroPCA (by Zhang et al. (2018)).

Input: sampling set Ω, data matrix $\boldsymbol{Y}$ (cf. (2.1b)), sampling rate p, rank r, maximum number of iterations t_0.
Initialization: set $\boldsymbol{G}^0 = \frac{1}{np^2}\mathcal{P}_{\textsf{off-diag}}(\boldsymbol{Y}\boldsymbol{Y}^\top)$.
Updates: **for** $t = 0, 1, \ldots, t_0$ **do**

$$\left(\boldsymbol{U}^t, \boldsymbol{\Lambda}^t\right) = \mathsf{eigs}\left(\boldsymbol{G}^t, r\right);$$
$$\boldsymbol{G}^{t+1} = \mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{G}^t\right) + \mathcal{P}_{\textsf{diag}}\left(\boldsymbol{U}^t\boldsymbol{\Lambda}^t\boldsymbol{U}^{t\top}\right) = \frac{1}{np^2}\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{Y}\boldsymbol{Y}^\top\right) + \mathcal{P}_{\textsf{diag}}\left(\boldsymbol{U}^t\boldsymbol{\Lambda}^t\boldsymbol{U}^{t\top}\right).$$

Here, for any symmetric matrix $\boldsymbol{G} \in \mathbb{R}^{d\times d}$ and $1 \leq r \leq d$, $\mathsf{eigs}(\boldsymbol{G}, r)$ returns $(\boldsymbol{U}, \boldsymbol{\Lambda})$, where $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ is the top-r eigen-decomposition of $\boldsymbol{G}$.
Output $\boldsymbol{U} = \boldsymbol{U}^{t_0}$ as the subspace estimate, $\boldsymbol{\Sigma} = (\boldsymbol{\Lambda}^{t_0})^{1/2}$ as an estimate of $\boldsymbol{\Sigma}^\star$, and $\boldsymbol{S} = \boldsymbol{U}^{t_0}\boldsymbol{\Lambda}^{t_0}\boldsymbol{U}^{t_0\top}$ as the covariance matrix estimate.

3 Distributional theory and inference procedures

In this section, we augment the two estimators introduced in Section 2 by a suite of distributional theory, and demonstrate how to employ our distributional characterizations to perform inference on both the principal subspace represented by $\boldsymbol{U}^\star$ and the covariance matrix $\boldsymbol{S}^\star$.

3.1 Key quantities and assumptions

Before continuing, we introduce several additional notation and assumptions that play a key role in our theoretical development. Recall that the eigen-decomposition of the covariance matrix $\boldsymbol{S}^\star \in \mathbb{R}^{d\times d}$ (see (1.2)) is assumed to be $\boldsymbol{U}^\star\boldsymbol{\Lambda}^\star\boldsymbol{U}^{\star\top}$. We assume the diagonal matrix $\boldsymbol{\Lambda}^\star$ to be $\boldsymbol{\Lambda}^\star = \mathsf{diag}\{\lambda_1^\star, \ldots, \lambda_r^\star\}$, where the diagonal entries are given by the non-zero eigenvalues of $\boldsymbol{S}^\star$ obeying

$$\lambda_1^\star \geq \cdots \geq \lambda_r^\star > 0.$$

The condition number of $\boldsymbol{S}^\star$ is denoted by

$$\kappa := \lambda_1^\star / \lambda_r^\star. \tag{3.1}$$

We also find it helpful to introduce the square root of $\boldsymbol{\Lambda}^\star$ as follows

$$\boldsymbol{\Sigma}^\star = \mathsf{diag}\{\sigma_1^\star, \dots, \sigma_r^\star\} = (\boldsymbol{\Lambda}^\star)^{1/2}, \qquad \text{where } \sigma_i^\star = (\lambda_i^\star)^{1/2},\ 1 \le i \le r. \tag{3.2}$$

Furthermore, we introduce an incoherence parameter commonly employed in prior literature (Candès, 2014; Chi et al., 2019).

Definition 1 (Incoherence). *The rank-r matrix $\boldsymbol{S}^\star \in \mathbb{R}^{d\times d}$ defined in (1.2) is said to be μ-incoherent if the following condition holds:*

$$\|\boldsymbol{U}^\star\|_{2,\infty} \le \sqrt{\frac{\mu r}{d}}. \tag{3.3}$$

Here, we recall that $\|\boldsymbol{U}^\star\|_{2,\infty}$ denotes the largest ℓ_2 norm of all rows of the matrix $\boldsymbol{U}^\star$.

Remark 1. When μ is small (e.g., $\mu \asymp 1$), this condition essentially ensures that the energy of $\boldsymbol{U}^\star$ is more or less dispersed across all of its rows. As a worthy note, the theory developed herein allows the incoherence parameter μ to grow with the problem dimension.

In light of a global rotational ambiguity issue (i.e., for any $r\times r$ rotation matrix $\boldsymbol{R}$, the matrices $\boldsymbol{U} \in \mathbb{R}^{d\times r}$ and $\boldsymbol{U}\boldsymbol{R} \in \mathbb{R}^{d\times r}$ share the same column space), in general we can only hope to estimate $\boldsymbol{U}^\star$ up to global rotation (unless additional eigenvalue separation conditions are imposed). Consequently, our theoretical development focuses on characterizing the error distribution $\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star$ of an estimator $\boldsymbol{U}$ when accounting for a proper rotation matrix $\boldsymbol{R}$. In particular, we shall pay particular attention to a specific way of rotation as follows

$$\boldsymbol{U}\mathsf{sgn}\left(\boldsymbol{U}^\top \boldsymbol{U}^\star\right) - \boldsymbol{U}^\star,$$

where we recall that for any non-singular matrix $\boldsymbol{H} \in \mathbb{R}^{k\times k}$ with SVD $\boldsymbol{U}_H\boldsymbol{\Sigma}_H\boldsymbol{V}_H^\top$, the matrix $\mathsf{sgn}(\boldsymbol{H})$ is defined to be the rotation matrix $\boldsymbol{U}_H\boldsymbol{V}_H^\top$. This particular choice aligns $\boldsymbol{U}$ and $\boldsymbol{U}^\star$ in the following sense

$$\mathsf{sgn}\left(\boldsymbol{U}^\top \boldsymbol{U}^\star\right) = \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|_{\mathrm{F}},$$

where $\mathcal{O}^{r\times r}$ indicates the set of all $r \times r$ rotation matrices; see (Ma et al., 2020, Appendix D.2.1).

The last assumption we would like to impose guarantees that the noise levels — while being possibly all different — are roughly on the same order, thereby simplifying the presentation of our main findings. This assumption will only be invoked when justifying the validity of confidence region construction, with the main purpose of simplifying presentation.

Assumption 1 (Noise levels). *The noise levels $\{\omega_i^\star\}_{1\le i\le d}$ obey*

$$\omega_{\max} \asymp \omega_{\min} \qquad \textit{with} \quad \omega_{\max} := \max_{1\le i\le d} \omega_i^\star \quad \textit{and} \quad \omega_{\min} := \min_{1\le i\le d} \omega_i^\star. \tag{3.4}$$

3.2 Inferential procedure and theory for the SVD-based approach

Focusing on Algorithm 1 as the estimator for the principal subspace, we present our inference procedure and accompanying theory, followed by a discussion of the inadequacy of this approach.

3.2.1 Distributional theory and inference for the principal subspace $\boldsymbol{U}^\star$

Distributional guarantees. The subspace estimate returned by Algorithm 1 turns out to be approximately unbiased and Gaussian in a row-wise manner, as formalized below. Here, we focus on the scenario with $\kappa, \mu, r \asymp 1$ for simplicity of presentation, with the general case deferred to Theorem 11 in Appendix B.

Theorem 1. *Suppose that $p < 1-\delta$ for some arbitrary constant $0 < \delta < 1$, and that $\kappa, \mu, r \asymp 1$. In addition, suppose that Assumption 1 holds and that*

$$\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{n+d}{np}} \lesssim \frac{1}{\log^{3/2}(n+d)} \wedge \frac{\sqrt{1 \wedge (d/n)}}{\log^{1/2}(n+d)}, \tag{3.5a}$$

$$np \gtrsim \log^3(n+d), \quad dp \gg \log^2(n+d), \quad n \gtrsim \log^4(n+d). \tag{3.5b}$$

Let $\boldsymbol{R}$ be the $r \times r$ rotation matrix $\boldsymbol{R} = \mathsf{sgn}(\boldsymbol{U}^\top \boldsymbol{U}^\star)$.

(a) Assume that $\boldsymbol{U}^\star$ is μ-incoherent and satisfies the following condition

$$\left\|\boldsymbol{U}^\star_{l,\cdot}\right\|_2 \gtrsim \left[\frac{\log^{5/2}(n+d)}{\sqrt{(n\wedge d)\,p}} + \frac{\log^{3/2}(n+d)}{\sqrt{np}\wedge\sqrt{d^2p/n}} + \frac{\log^{3/2}(n+d)}{\sqrt{d}}\right]\sqrt{\frac{1}{d}} \tag{3.5c}$$

for some $1 \le l \le d$. Then the estimate $\boldsymbol{U}$ returned by Algorithm 1 obeys

$$\sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left(\left[\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^\star\right]_{l,\cdot}\in\mathcal{C}\right) - \mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}^\star_{U,l}\right)\{\mathcal{C}\}\right| = o(1), \tag{3.6}$$

where $\mathscr{C}^r$ is the set of all convex sets in $\mathbb{R}^r$ and

$$\boldsymbol{\Sigma}^\star_{U,l} \coloneqq \left(\frac{1-p}{np}\left\|\boldsymbol{U}^\star_{l,\cdot}\boldsymbol{\Sigma}^\star\right\|_2^2 + \frac{\omega_l^{\star 2}}{np}\right)(\boldsymbol{\Sigma}^\star)^{-2} + \frac{2(1-p)}{np}\left(\boldsymbol{U}^\star_{l,\cdot}\right)^\top\boldsymbol{U}^\star_{l,\cdot}. \tag{3.7}$$

(b) If in addition Condition (3.5c) holds simultaneously for all $1 \le l \le d$, then one further has

$$\sup_{1\le l\le d}\sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left(\left[\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^\star\right]_{l,\cdot}\in\mathcal{C}\right) - \mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}^\star_{U,l}\right)\{\mathcal{C}\}\right| = o(1). \tag{3.8}$$

Remark 2. When there is no missing data (i.e., $p = 1$), the above theorem continues to hold even without assuming (3.5c). Specifically, assume that $p = 1$ and that Conditions (3.5a) and (3.5b) hold, then the above distributional guarantee is valid with $\boldsymbol{\Sigma}^\star_{U,l} = \frac{\omega_l^{\star 2}}{n}(\boldsymbol{\Sigma}^\star)^{-2}$.

Construction of confidence regions for the principal subspace. The distributional theory presented in Theorem 1 paves the path for confidence region construction, as long as the covariance matrix $\boldsymbol{\Sigma}^\star_{U,l}$ can be reliably estimated. We describe in Algorithm 3 a plausible procedure aimed at estimating $\boldsymbol{\Sigma}^\star_{U,l}$ and building confidence regions. A little explanation is as follows.

- Observe that $\{y_{l,j} : (l,j)\in\Omega\}$ is a set of independent zero-mean random variables with common variance $S^\star_{l,l} + \omega_l^{\star 2}$. This motivates us to employ the difference between the empirical average of $\{y^2_{l,j} : (l,j)\in\Omega\}$ and $S_{l,l}$ (which is an estimate of $S^\star_{l,l}$) to estimate the noise level $\omega_l^{\star 2}$.
- We then invoke a sort of "plug-in" approach to estimate $\boldsymbol{\Sigma}^\star_{U,l}$ (i.e., replacing $\boldsymbol{U}^\star$, $\boldsymbol{\Sigma}^\star$ and $\omega_l^\star$ by $\boldsymbol{U}$, $\boldsymbol{\Sigma}$ and ω_l, respectively), as inspired by the closed-form expression (3.7).
- The confidence region is then constructed based on the Gaussian approximation (3.6).

Our performance guarantees for the inference procedure in Algorithm 3 are stated below, again focusing on the simple scenario with $\kappa, \mu, r \asymp 1$. The general case can be found in Theorem 12 in Section B.

Theorem 2. *Instate the assumptions in Theorem 1(b). Additionally, assume that $np \gtrsim \log^5(n+d)$. Then there exists a $r\times r$ rotation matrix $\boldsymbol{R} = \mathsf{sgn}\left(\boldsymbol{U}^\top\boldsymbol{U}^\star\right)$ such that the confidence regions $\left\{\mathsf{CR}^{1-\alpha}_{U,l}\right\}$ computed in Algorithm 3 obey*

$$\sup_{1\le l\le d}\left|\mathbb{P}\left(\boldsymbol{U}^\star_{l,\cdot}\boldsymbol{R}^\top\in\mathsf{CR}^{1-\alpha}_{U,l}\right) - (1-\alpha)\right| = o(1).$$

Note that the subspace represented by $\boldsymbol{U}^\star$ and the one represented by $\boldsymbol{U}^\star\boldsymbol{R}^\top$ are essentially identical. As a result, the constructed confidence regions $\left\{\mathsf{CR}^{1-\alpha}_{U,l}\right\}$ taken together form a valid set that contains a ground-truth subspace representation in a row-wise valid fashion. There are many applications — e.g., community detection (Rohe et al., 2011), Gaussian mixture models (Löffler et al., 2019) — in which the rows of $\boldsymbol{U}^\star$ might contain crucial operational information. When $r = 1$, the above theorem provides entrywise confidence intervals for the principal component of interest.

Algorithm 3 Confidence regions for $\boldsymbol{U}^{\star}_{l,\cdot}$ base on the SVD-based approach.

Input: output $(\boldsymbol{U}, \boldsymbol{\Sigma}, \boldsymbol{S})$ of Algorithm 1, sampling rate p, coverage level $1-\alpha$.
Compute an estimate of the noise level $\omega_l^{\star}$ as

$$\omega_l^2 \coloneqq \frac{\sum_{j=1}^n y_{l,j}^2 \, \mathbb{1}_{(l,j)\in\Omega}}{\sum_{j=1}^n \mathbb{1}_{(l,j)\in\Omega}} - S_{l,l}.$$

Compute an estimate of $\boldsymbol{\Sigma}^{\star}_{U,l}$ (cf. (3.7)) as

$$\boldsymbol{\Sigma}_{U,l} \coloneqq \left(\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\right\|_2^2 + \frac{\omega_l^2}{np}\right)(\boldsymbol{\Sigma})^{-2} + \frac{2\,(1-p)}{np}\boldsymbol{U}_{l,\cdot}^{\top}\boldsymbol{U}_{l,\cdot}.$$

Compute the $(1-\alpha)$-quantile $\tau_{1-\alpha}$ of $\chi^2(r)$ and construct a Euclidean ball:

$$\mathcal{B}_{1-\alpha} \coloneqq \left\{\boldsymbol{z}\in\mathbb{R}^r : \|\boldsymbol{z}\|_2^2 \le \tau_{1-\alpha}\right\}.$$

Output the $(1-\alpha)$-confidence region

$$\mathsf{CR}_{U,l}^{1-\alpha} \coloneqq \boldsymbol{U}_{l,\cdot} + \left(\boldsymbol{\Sigma}_{U,l}\right)^{1/2}\mathcal{B}_{1-\alpha} = \left\{\boldsymbol{U}_{l,\cdot} + \left(\boldsymbol{\Sigma}_{U,l}\right)^{1/2}\boldsymbol{z} : \boldsymbol{z}\in\mathcal{B}_{1-\alpha}\right\}.$$

Interpretations and implications. We now take a moment to interpret the above two theorems and parse the required conditions. We shall first focus on the case when $n \lesssim d$.

- *A key error decomposition.* As a high-level remark for Theorem 1, we make note of the following key decomposition
$$\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star} = \underbrace{n^{-1}\left(p^{-1}\boldsymbol{Y} - \boldsymbol{X}\right)\boldsymbol{X}^{\top}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}}_{=: \boldsymbol{Z} \text{ (first-order approximation)}} + \underbrace{\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star} - \boldsymbol{Z}}_{=: \boldsymbol{\Psi} \text{ (residual term)}},$$
where the first term $\boldsymbol{Z}$ is a linear mapping of the data matrix $\boldsymbol{Y}$. As we shall justify momentarily in the analysis, the l-th row of $\boldsymbol{Z}$ enjoys a Gaussian approximation
$$\boldsymbol{Z}_{l,\cdot} \overset{\mathrm{d}}{\approx} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}^{\star}_{U,l}\right), \qquad 1 \le l \le d, \tag{3.9}$$
while the $\ell_{2,\infty}$ norm of the residual term $\boldsymbol{\Psi}$ (which can be interpreted as a higher-order term) is well controlled. In fact, Condition (3.5c) is imposed in order to guarantee (i) the validity of the Gaussian approximation (3.9), and (ii) the dominance of $\boldsymbol{Z}_{l,\cdot}$ compared to the residual term.

- *Incomplete observations.* Theorem 1 and Theorem 2 accommodate the case where a large fraction of data entries are unseen. Specifically, they only require the sampling rate to exceed $p \ge \widetilde{\Omega}(1/n)$ in order for the distributional characterization to be valid and for constructing confidence regions. The reason why constructing confidence regions requires a larger sample complexity than in our distributional theory is that we need to account for the error of estimating $\boldsymbol{\Sigma}^{\star}_{U,l}$ using $\boldsymbol{\Sigma}_{U,l}$.

- *Tolerable noise levels.* The noise condition required for Theorem 1 and Theorem 2 to hold reads $\omega_{\max} \le \widetilde{O}(\sqrt{np/d}\sigma_r^{\star})$. Note that when $\kappa, \mu, r \asymp 1$, the variance obeys
$$\max_{(l,j)\in\Omega} \mathsf{var}\left(x_{l,j}\right) = \max_{l\in[d]} S_{l,l}^{\star} \asymp \max_{l\in[d]}\left\|\boldsymbol{U}^{\star}_{l,\cdot}\right\|_2^2 \sigma_1^{\star 2} \asymp \frac{1}{d}\sigma_1^{\star 2}.$$
When $p \ge \widetilde{\Omega}(1/n)$, our tolerable entrywise noise level $\omega_{\max}$ is allowed to be significantly (i.e., $\widetilde{\Omega}(\sqrt{np})$ times) larger than the largest standard deviation of $x_{l,j}$ for all $(l,j)\in\Omega$, thereby accommodating a wide range of noise levels.

- *Non-vanishing size of* $\|\boldsymbol{U}^\star_{l,\cdot}\|_2$. In order to perform inference on $\boldsymbol{U}^\star_{l,\cdot}$, Theorem 1 and Theorem 2 impose the following condition

$$\left\|\boldsymbol{U}^\star_{l,\cdot}\right\|_2 \geq \widetilde{\Omega}\left(\sqrt{\frac{1}{np}}\right) \cdot \sqrt{\frac{1}{d}}\left\|\boldsymbol{U}^\star\right\|_{\mathrm{F}}.$$

  Note that the typical ℓ_2 norm of a row of $\boldsymbol{U}^\star$ is $\|\boldsymbol{U}^\star\|_{\mathrm{F}}/\sqrt{d}$ when the energy is uniformly spread out across all rows. Therefore under our conditions on the sampling rate p, we allow $\|\boldsymbol{U}^\star_{l,\cdot}\|_2$ to be significantly smaller than its typical size, which is often a mild condition.

3.2.2 Distributional theory and inference for the covariance matrix $\boldsymbol{S}^\star$

The above distributional characterization for $\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star$ in turn enables entrywise distributional characterization of the estimate $\boldsymbol{S}$ returned by Algorithm 1, which further suggests how to build entrywise confidence interval for the covariance matrix $\boldsymbol{S}^\star$.

Entrywise distributional guarantees. Our next theorem shows that for any $i,j \in [d]$, the (i,j)-th entry of $\boldsymbol{S} - \boldsymbol{S}^\star$ is approximately a zero-mean Gaussian with variance $v^\star_{i,j}$ defined as

$$v^\star_{i,j} \coloneqq \frac{2-p}{np} S^\star_{i,i} S^\star_{j,j} + \frac{4-3p}{np} S^{\star 2}_{i,j} + \frac{1}{np}\left(\omega_i^{\star 2} S^\star_{j,j} + \omega_j^{\star 2} S^\star_{i,i}\right), \qquad \text{if } i \neq j, \tag{3.10a}$$

$$v^\star_{i,i} \coloneqq \frac{12-9p}{np} S^{\star 2}_{i,i} + \frac{4}{np}\omega_i^{\star 2} S^\star_{i,i}. \tag{3.10b}$$

Theorem 3. *Suppose that $p < 1-\delta$ for some arbitrary constant $0<\delta<1$, and that $\kappa, \mu, r \asymp 1$. Consider any $1 \leq i,j \leq d$. Assume that $\boldsymbol{U}^\star$ is μ-incoherent and satisfies the following conditions*

$$\min\left\{\left\|\boldsymbol{U}^\star_{i,\cdot}\right\|_2, \left\|\boldsymbol{U}^\star_{j,\cdot}\right\|_2\right\} \gtrsim \left[\frac{\log^{5/2}(n+d)}{\sqrt{(n\wedge d)\,p}} + \frac{\log^{3/2}(n+d)}{\sqrt{np} \wedge \sqrt{d^2p/n}} + \frac{\log^{3/2}(n+d)}{\sqrt{d}}\right]\sqrt{\frac{1}{d}} \tag{3.11a}$$

$$\text{and} \qquad \left\|\boldsymbol{U}^\star_{i,\cdot}\right\|_2 + \left\|\boldsymbol{U}^\star_{j,\cdot}\right\|_2 \gtrsim \log^{5/2}(n+d)\,\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d}{np}} \cdot \sqrt{\frac{1}{d}}. \tag{3.11b}$$

In addition, suppose that Assumption 1 holds and that $d \gtrsim \log^2(n+d)$,

$$np \gtrsim \log^5(n+d), \qquad \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{n+d}{np}} \lesssim \frac{1}{\log(n+d)} \wedge \sqrt{\frac{1\wedge (d/n)}{\log(n+d)}}.$$

Then the matrix $\boldsymbol{S}$ computed by Algorithm 1 obeys

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\frac{S_{i,j} - S^\star_{i,j}}{\sqrt{v^\star_{i,j}}} \leq t\right) - \Phi(t)\right| = o(1),$$

where $\Phi(\cdot)$ is the CDF of the standard Gaussian distribution, and $v^\star_{i,j}$ is defined in (3.10).

***Remark* 3.** This theorem can also be extended to the scenario with no missing data (i.e., $p=1$), in which case Condition (3.11a) is not required.

***Remark* 4.** Observe that $S^\star_{i,i} = \|\boldsymbol{U}^\star_i\boldsymbol{\Sigma}^\star\|_2^2 \asymp \|\boldsymbol{U}^\star_i\|_2^2\sigma_r^{\star 2}$ and $S^\star_{j,j} = \|\boldsymbol{U}^\star_j\boldsymbol{\Sigma}^\star\|_2^2 \asymp \|\boldsymbol{U}^\star_j\|_2^2\sigma_r^{\star 2}$, the condition (3.11b) is equivalent to

$$\omega_{\max}^2 \leq \widetilde{O}\left(np\big(S^\star_{i,i} + S^\star_{j,j}\big)\right).$$

In words, (3.11b) allow the noise level $\omega_{\max}$ to be significantly (i.e., $\widetilde{\Omega}(\sqrt{np})$ times) larger than the standard deviation of $\{x_{i,l} : (i,l)\in\Omega\}$ and $\{x_{j,l} : (i,l)\in\Omega\}$, which is a mild condition.

Algorithm 4 Confidence intervals for $S_{i,j}^{\star}$ $(1 \leq i,j \leq d)$ base on the SVD-based approach.

Input: output $(\boldsymbol{U}, \boldsymbol{\Sigma}, \boldsymbol{S})$ of Algorithm 1, sampling rate p, coverage level $1-\alpha$.
Compute estimates of the noise level $\omega_l^{\star}$ as follows

$$\omega_l^2 := \frac{\sum_{j=1}^n y_{l,j}^2 \mathbb{1}_{(l,j)\in\Omega}}{\sum_{j=1}^n \mathbb{1}_{(l,j)\in\Omega}} - S_{l,l}.$$

Compute an estimate of $v_{i,j}^{\star}$ (cf. (3.20) or (3.21))as follows: if $i \neq j$ then

$$v_{i,j} := \frac{2-p}{np} S_{i,i} S_{j,j} + \frac{4-3p}{np} S_{i,j}^2 + \frac{1}{np}\left(\omega_i^2 S_{j,j}^{\star} + \omega_j^2 S_{i,i}\right);$$

if $i = j$ then

$$v_{i,i} := \frac{12-9p}{np} S_{i,i}^2 + \frac{4}{np}\omega_i^2 S_{i,i}.$$

Output the $(1-\alpha)$-confidence interval

$$\mathsf{CI}_{i,j}^{1-\alpha} := \left[S_{i,j} \pm \Phi^{-1}\left(1-\alpha/2\right)\sqrt{v_{i,j}}\right].$$

Construction of entrywise confidence intervals. The preceding distributional characterization motivates a plausible scheme to construct a confidence interval for $S_{i,j}^{\star}$, as long as we can estimate $v_{i,j}^{\star}$ accurately. As it turns out, a simple plug-in estimator is already accurate enough for this purpose. We summarize the procedure in Algorithm 4, and justify its validity in Theorem 4.

Theorem 4. *Instate the conditions in Theorem 3. Then the confidence interval computed in Algorithm 4 obeys*

$$\mathbb{P}\left(S_{i,j}^{\star} \in \mathsf{CI}_{i,j}^{1-\alpha}\right) = 1 - \alpha + o\left(1\right).$$

Theorems 3-4 focus on the case with $\kappa, \mu, r \asymp 1$. The more general case which allows κ, μ, r to grow with the problem dimension is postponed to Theorems 13-14 in Appendix B.

3.2.3 Inadequacy of the SVD-based approach

Thus far, we have studied how to quantify the uncertainty of the estimator presented in Algorithm 1 when $n \lesssim d$. However, this approach is inherently suboptimal when $n \gg d$, given that the initial estimator is already far from optimal to start with for this regime. In what follows, we elaborate on the sub-optimality of this approach in the regime where $n \gg d$, by comparing it with the estimation theory in Cai et al. (2021) from several aspects.

- The sampling rate required in the estimation guarantee (Cai et al., 2021, Corollary 1) reads $p \geq \widetilde{\Omega}(1/\sqrt{nd})$. In comparison, our distributional theory in Theorem 1 requires that

$$p \geq \widetilde{\Omega}\left(\frac{1}{d}\right), \tag{3.12}$$

  which is $\sqrt{n/d}$ times more stringent compared to the condition in Cai et al. (2021) (up to logarithmic factor).

- The noise condition required in the estimation theory (Cai et al., 2021, Corollary 1) reads $\omega_{\max}^2 \leq \widetilde{O}\big(\sqrt{n/d}\, p\sigma_r^{\star 2}\big)$. In comparison, Theorem 1 presented above requires that

$$\omega_{\max}^2 \leq \widetilde{O}\left(\frac{d}{n} p\sigma_r^{\star 2}\right). \tag{3.13}$$

Our assumption on $\omega_{\max}^2$ is $(n/d)^{3/2}$ times more stringent than the prior estimation guarantees (up to logarithm factor).

- Theorem 1 requires $\|\boldsymbol{U}_{l,\cdot}^{\star}\|_2$ to exceed

$$\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2 \geq \widetilde{\Omega}\left(\frac{1}{\sqrt{d^2p/n}}\right)\cdot\sqrt{\frac{1}{d}}\left\|\boldsymbol{U}^{\star}\right\|_{\mathrm{F}}, \tag{3.14}$$

which turns out to be very stringent when $n \gg d^2p$.

The above comparisons make clear the inadequacy of the SVD-based approach even for estimation purpose. In fact, our inferential theory tailored to the SVD-based approach fails to cover an important regime where consistent estimation remains feasible. Fortunately, as we shall elucidate in the next subsection, inference based on another estimation algorithm HeteroPCA can be far more effective and adaptive.

3.3 Inferential procedure and theory for HeteroPCA

We now move on to investigate how to assess the uncertainty of the more refined estimator HeteroPCA, and to discuss the superiority of this approach in comparison to the previous SVD-based approach. For simplicity of presentation, we shall abuse some notation (e.g., $\boldsymbol{\Sigma}_{U,l}^{\star}$ and $v_{i,j}^{\star}$) whenever it is clear from the context.

3.3.1 Distributional theory and inference for the principal subspace $\boldsymbol{U}^{\star}$

Distributional guarantees. Encouragingly, the subspace estimate returned by Algorithm 2 turns out to be approximately unbiased and Gaussian under milder conditions, as posited in the following theorem. The general result beyond the case with $\kappa,\mu,r \asymp 1$ is postponed to Theorem 15 in Appendix B.

Theorem 5. *Suppose that $p < 1-\delta$ for some arbitrary constant $0<\delta<1$, and that $\kappa,\mu,r\asymp 1$. Assume, in addition, that Assumption 1 holds and that $d \gtrsim \log^5 n$,*

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\log^3(n+d)}, \qquad \frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\log^2(n+d)}, \tag{3.15a}$$

$$ndp^2 \gtrsim \log^5(n+d), \qquad np \gtrsim \log^4(n+d). \tag{3.15b}$$

Let $\boldsymbol{R}$ be the $r\times r$ rotation matrix $\boldsymbol{R} = \mathsf{sgn}(\boldsymbol{U}^{\top}\boldsymbol{U}^{\star})$.

(a) Assume that $\boldsymbol{U}^{\star}$ is μ-incoherent and satisfies the following condition

$$\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2 \gtrsim \left[\frac{\log^5(n+d)}{\sqrt{ndp^2}} + \frac{\log^3(n+d)}{\sqrt{np}} + \frac{\log^{7/2}(n+d)}{\sqrt{d}}\right]\sqrt{\frac{1}{d}}. \tag{3.15c}$$

Suppose that the number of iterations exceeds

$$t_0 \gtrsim \log\left[\left(\frac{\log^2(n+d)}{\sqrt{ndp}} + \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}\log(n+d) + \frac{\log(n+d)}{\sqrt{np}} + \frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d\log(n+d)}{np}}\right)^{-1}\right]. \tag{3.16}$$

Then the estimate $\boldsymbol{U}$ returned by Algorithm 2 obeys

$$\sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left(\left[\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}\right]_{l,\cdot}\in\mathcal{C}\right) - \mathcal{N}\big(\boldsymbol{0},\boldsymbol{\Sigma}_{U,l}^{\star}\big)\left\{\mathcal{C}\right\}\right| = o\left(1\right),$$

where $\mathscr{C}^r$ represents the set of all convex sets in $\mathbb{R}^r$, and

$$\begin{aligned}\boldsymbol{\Sigma}_{U,l}^{\star} &\coloneqq \left(\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{\omega_l^{\star 2}}{np}\right)\left(\boldsymbol{\Sigma}^{\star}\right)^{-2} + \frac{2\left(1-p\right)}{np}\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^{\star} \\ &\quad + \left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{U}^{\star\top}\mathsf{diag}\left\{\left[d_{l,i}^{\star}\right]_{1\leq i\leq d}\right\}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\end{aligned} \tag{3.17}$$

with

$$d_{l,i}^{\star} := \frac{1}{np^2}\left[\omega_l^{\star 2} + (1-p)\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\right]\left[\omega_i^{\star 2} + (1-p)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\right] + \frac{2(1-p)^2}{np^2}S_{l,i}^{\star 2}.$$

(b) If, in addition, (3.15c) holds simultaneously for all $1 \le l \le d$, then one further has

$$\sup_{1\le l\le d}\sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left([\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}]_{l,\cdot}\in\mathcal{C}\right) - \mathcal{N}\big(\boldsymbol{0},\boldsymbol{\Sigma}_{U,l}^{\star}\big)\{\mathcal{C}\}\right| = o(1).$$

Theorem 5 asserts that each row of the estimate $\boldsymbol{U}$ returned by HeteroPCA is nearly unbiased and admits a nearly tight Gaussian approximation, whose covariance matrix can be determined via the closed-form expression (3.17). Given that $\boldsymbol{U}\boldsymbol{R}$ and $\boldsymbol{U}$ represent the same subspace, this theorem delivers a fine-grained row-wise distributional characterization for the estimator HeteroPCA.

Remark 5. When there is no missing data (i.e., $p = 1$), we do not require the condition (3.15c).

Remark 6. We shall briefly mention the key error decomposition behind this theorem. Letting $\boldsymbol{E} := n^{-1/2}(p^{-1}\boldsymbol{Y} - \boldsymbol{X})$, we can decompose

$$\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star} = \underbrace{\left[\boldsymbol{E}\boldsymbol{X}^{\top} + \mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^{\top}\right)\right]\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}}_{=:\boldsymbol{Z}\text{ (first- and second-order approximation)}} + \underbrace{\left[\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star} - \boldsymbol{Z}\right]}_{=:\boldsymbol{\Psi}\text{ (residual term)}}.$$

Here, $\boldsymbol{Z}$ contains not only linear mapping of $\boldsymbol{E}$ but also a certain second-order term, the latter of which is crucial when coping with the regime $n \gg d$. As will be solidified in the proof, $\boldsymbol{Z}$ admits the following Gaussian approximation

$$\boldsymbol{Z}_{l,\cdot} \overset{\mathrm{d}}{\approx} \mathcal{N}\big(\boldsymbol{0},\boldsymbol{\Sigma}_{U,l}^{\star}\big), \qquad 1 \le l \le d, \tag{3.18}$$

while, at the same time, the $\ell_{2,\infty}$ norm of the residual term $\boldsymbol{\Psi}$ is well controlled. We impose the condition (3.15c) to ensure that (i) the Gaussian approximation (3.18) is valid, and (ii) the typical size of $\boldsymbol{Z}_{l,\cdot}$ dominates the corresponding part in $\boldsymbol{\Psi}$.

Construction of confidence regions for the principal subspace. With the above distributional theory in place, we are well-equipped to construct fine-grained confidence regions for $\boldsymbol{U}^{\star}$, provided that the covariance matrix $\boldsymbol{\Sigma}_{U,l}^{\star}$ can be estimated in a faithful manner. In Algorithm 5, we propose a procedure that allows us to estimate $\boldsymbol{\Sigma}_{U,l}^{\star}$ and, in turn, provide confidence regions. As before, our estimator for $\boldsymbol{\Sigma}_{U,l}^{\star}$ can be viewed as a sort of "plug-in" method in accordance with the expression (3.17).

The following theorem confirms the validity of the proposed inference procedure when $\kappa, \mu, r \asymp 1$. The more general case will be studied in Theorem 16 in Section B.

Theorem 6. *Instate the conditions in Theorem 5(b). Further, assume that*

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\log^4(n+d)}, \qquad \frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\log^4(n+d)}, \tag{3.19a}$$

and

$$ndp^2 \gtrsim \log^{11}(n+d), \qquad np \gtrsim \log^{9/2}(n+d). \tag{3.19b}$$

Then there exists a $r \times r$ rotation matrix $\boldsymbol{R} = \mathsf{sgn}\left(\boldsymbol{U}^{\top}\boldsymbol{U}^{\star}\right)$ such that the confidence regions $\mathsf{CR}_{U,l}^{1-\alpha}$ $(1 \le l \le d)$ computed in Algorithm 5 obey

$$\sup_{1\le l\le d}\left|\mathbb{P}\left(\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{R}^{\top}\in\mathsf{CR}_{U,l}^{1-\alpha}\right) - (1-\alpha)\right| = o(1).$$

In words, Theorem 6 uncovers that: a valid ground-truth subspace representation is contained — in a row-wise reliable manner — within the confidence regions $\mathsf{CR}_{U,l}^{1-\alpha}$ $(1 \le l \le d)$ we construct. In the special case with $r = 1$, this result leads to valid entrywise confidence intervals for the principal component.

Algorithm 5 Confidence regions for $\boldsymbol{U}^{\star}_{l,\cdot}$ $(1 \le l \le d)$ base on $\mathsf{HeteroPCA}$.

Input: output $(\boldsymbol{U}, \boldsymbol{\Sigma}, \boldsymbol{S})$ of Algorithm 2, sampling rate p, coverage level $1-\alpha$.
Compute estimates of the noise levels $\{\omega_l^{\star}\}_{1 \le l \le d}$ as follows

$$\omega_l^2 \coloneqq \frac{\sum_{j=1}^n y_{l,j}^2 \, \mathbb{1}_{(l,j)\in\Omega}}{\sum_{j=1}^n \mathbb{1}_{(l,j)\in\Omega}} - S_{l,l} \qquad \text{for all} \quad 1 \le l \le d.$$

Compute an estimate of $\boldsymbol{\Sigma}_{U,l}^{\star}$ (cf. (3.17)) as follows:

$$\boldsymbol{\Sigma}_{U,l} \coloneqq \left(\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\right\|_2^2 + \frac{\omega_l^2}{np}\right)\boldsymbol{\Sigma}^{-2} + \frac{2(1-p)}{np}\boldsymbol{U}_{l,\cdot}^{\top}\boldsymbol{U}_{l,\cdot} + (\boldsymbol{\Sigma})^{-2}\boldsymbol{U}^{\top}\mathsf{diag}\left\{[d_{l,i}]_{1\le i\le d}\right\}\boldsymbol{U}(\boldsymbol{\Sigma})^{-2},$$

where

$$d_{l,i} \coloneqq \frac{1}{np^2}\left[\omega_l^2 + (1-p)\left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\right\|_2^2\right]\left[\omega_i^2 + (1-p)\left\|\boldsymbol{U}_{i,\cdot}\boldsymbol{\Sigma}\right\|_2^2\right] + \frac{2(1-p)^2}{np^2}S_{l,i}^2.$$

Compute the $(1-\alpha)$-quantile $\tau_{1-\alpha}$ of $\chi^2(r)$ and construct a Euclidean ball:

$$\mathcal{B}_{1-\alpha} \coloneqq \left\{\boldsymbol{z} \in \mathbb{R}^r : \|\boldsymbol{z}\|_2^2 \le \tau_{1-\alpha}\right\}.$$

Output the $(1-\alpha)$-confidence region

$$\mathsf{CR}_{U,l}^{1-\alpha} \coloneqq \boldsymbol{U}_{l,\cdot} + \left(\boldsymbol{\Sigma}_{U,l}\right)^{1/2}\mathcal{B}_{1-\alpha} = \left\{\boldsymbol{U}_{l,\cdot} + \left(\boldsymbol{\Sigma}_{U,l}\right)^{1/2}\boldsymbol{z} : \boldsymbol{z} \in \mathcal{B}_{1-\alpha}\right\}.$$

Interpretations and implications. We now take a moment to interpret the conditions required in Theorem 5 and Theorem 6, and discuss some appealing attributes of our methods. As before, the discussion below focuses on the scenario where $\mu, \kappa, r \asymp 1$ for the sake of simplicity.

- *Missing data.* Both theorems accommodate the case when a large fraction of data are missing, namely, they cover the range
$$p \ge \widetilde{\Omega}\Big(\frac{1}{n \wedge \sqrt{nd}}\Big)$$
for both distributional characterizations and confidence region construction using $\mathsf{HeteroPCA}$. When $n \gg d$, this sampling rate requirement is $\widetilde{O}\big(\sqrt{n/d}\big)$ times less stringent than the one imposed for the SVD-based approach (see (3.12)). When $n \lesssim d$, the sampling rate condition required for $\mathsf{HeteroPCA}$ coincides with the one for the SVD-based method.

- *Noise levels.* The noise level required in both Theorem 5 and Theorem 6 is given by
$$\omega_{\max}^2 \le \widetilde{O}\left(\Big(\frac{n}{d} \wedge \sqrt{\frac{n}{d}}\Big)p\sigma_r^{\star 2}\right).$$
When $n \gg d$, the largest possible noise level $\omega_{\max}^2$ we can tolerate is $(n/d)^{3/2}$ times larger than the one required for the SVD-based approach (see (3.13)). When $n \lesssim d$, the noise conditions required for both SVD-based method and $\mathsf{HeteroPCA}$ are nearly identical.

- *Non-vanishing size of $\|\boldsymbol{U}^{\star}_{l,\cdot}\|_2$.* When it comes to inference for $\boldsymbol{U}^{\star}_{l,\cdot}$, our theorems impose the following condition:
$$\left\|\boldsymbol{U}^{\star}_{l,\cdot}\right\|_2 \ge \widetilde{\Omega}\left(\frac{1}{\sqrt{ndp^2}} + \frac{1}{\sqrt{np}} + \frac{1}{\sqrt{d}}\right) \cdot \sqrt{\frac{1}{d}}\left\|\boldsymbol{U}^{\star}\right\|_{\mathrm{F}}.$$
Note that the typical ℓ_2 norm of a row of $\boldsymbol{U}^{\star}$ is $\|\boldsymbol{U}^{\star}\|_{\mathrm{F}}/\sqrt{d}$ when the energy is uniformly spread out across all rows. This means that under our sampling rate condition, our results allow $\|\boldsymbol{U}^{\star}_{l,\cdot}\|_2$ to be much

smaller than its typical size. When $n \gg d$, this improves upon the condition on $\|\boldsymbol{U}^{\star}_{l,\cdot}\|_2$ required for the SVD-based method (see (3.14)) by a factor of $\min\{\sqrt{(n/d)np}, n/d, \sqrt{n/(dp)}\}$.

- *Adaptivity to heteroskedasticity and unknown noise levels.* Our proposed inferential procedure is fully data-driven: it is automatically adaptive to unknown heteroskedastic noise, without requiring prior knowledge of the noise levels.

Comparison with prior estimation theory. While the main purpose of the current paper is to enable efficient statistical inference for the principal subspace, our theory also enables improved estimation guarantees compared to prior works.

- Recall that the estimation algorithm HeteroPCA was originally proposed and studied by Zhang et al. (2018). Our results broaden the sample size range supported by their theory. More specifically, note that Zhang et al. (2018, Theorem 6 and Remark 10) required the sampling rate p to satisfy

$$ndp \gtrsim \max\left\{d^{1/3}n^{2/3}, d\right\} \mathsf{polylog}\,(n,d)$$

in order to guarantee consistent estimation, while our theoretical guarantees only require

$$ndp \gtrsim \max\left\{\sqrt{nd}, d\right\} \mathsf{polylog}\,(n,d)\,.$$

When $n \gg d$, the sample size requirement in Zhang et al. (2018) is $(n/d)^{1/6}$ times more stringent than the one imposed in our theory.

- We shall also pause to discuss the advantage of HeteroPCA compared to the diagonal-deleted spectral method studied in Cai et al. (2021, Algorithms 1 and 3). Due to diagonal deletion, there is an additional bias term (see the last term $\mu_{\mathsf{ce}}\kappa_{\mathsf{ce}}r/d$ in Equation (4.16) in Cai et al. (2021)), which turns out to negatively affect our capability of performing inference. In contrast, HeteroPCA eliminates this bias term by means of successive refining, thus facilitating the subsequent inference stage.

3.3.2 Distributional theory and inference for the covariance matrix $\boldsymbol{S}^{\star}$

As it turns out, the above distributional theory for $\boldsymbol{U}^{\star}$ further hints at how to perform statistical inference for the covariance matrix $\boldsymbol{S}^{\star}$, which we elaborate on in this subsection.

Entrywise distributional guarantees. We now focus attention on characterizing the distribution of the (i,j)-th entry of $\boldsymbol{S}$ returned by Algorithm 2, which in turn suggests how to construct entrywise confidence intervals for $\boldsymbol{S}^{\star}$. Before proceeding, let us define a set of variance parameters $\{v^{\star}_{i,j}\}_{1\leq i,j\leq d}$ which, as we shall demonstrate momentarily, correspond to the (approximate) variance of the entries of $\boldsymbol{S}$.

- For any $1 \leq i,j \leq d$ obeying $i \neq j$, we define

$$\begin{aligned} v^{\star}_{i,j} &\coloneqq \frac{2-p}{np} S^{\star}_{i,i} S^{\star}_{j,j} + \frac{4-3p}{np} S^{\star 2}_{i,j} + \frac{1}{np}\left(\omega_i^{\star 2} S^{\star}_{j,j} + \omega_j^{\star 2} S^{\star}_{i,i}\right) \\ &\quad + \frac{1}{np^2}\sum_{k=1}^{d}\left\{\left[\omega_i^{\star 2} + (1-p)\, S^{\star}_{i,i}\right]\left[\omega_k^{\star 2} + (1-p)\, S^{\star}_{k,k}\right] + 2\,(1-p)^2\, S^{\star 2}_{i,k}\right\}\left(\boldsymbol{U}^{\star}_{k,\cdot}\boldsymbol{U}^{\star\top}_{j,\cdot}\right)^2 \\ &\quad + \frac{1}{np^2}\sum_{k=1}^{d}\left\{\left[\omega_j^{\star 2} + (1-p)\, S^{\star}_{j,j}\right]\left[\omega_k^{\star 2} + (1-p)\, S^{\star}_{k,k}\right] + 2\,(1-p)^2\, S^{\star 2}_{j,k}\right\}\left(\boldsymbol{U}^{\star}_{k,\cdot}\boldsymbol{U}^{\star\top}_{i,\cdot}\right)^2. \end{aligned} \tag{3.20}$$

- For any $1 \leq i \leq d$, we set

$$\begin{aligned} v^{\star}_{i,i} &\coloneqq \frac{12-9p}{np} S^{\star 2}_{i,i} + \frac{4}{np}\omega_i^{\star 2} S^{\star}_{i,i} \\ &\quad + \frac{4}{np^2}\sum_{k=1}^{d}\left\{\left[\omega_i^{\star 2} + (1-p)\, S^{\star}_{i,i}\right]\left[\omega_k^{\star 2} + (1-p)\, S^{\star}_{k,k}\right] + 2\,(1-p)^2\, S^{\star 2}_{i,k}\right\}\left(\boldsymbol{U}^{\star}_{k,\cdot}\boldsymbol{U}^{\star\top}_{i,\cdot}\right)^2. \end{aligned} \tag{3.21}$$

We are now positioned to present our distributional theory for the scenario where $\kappa, \mu, r \asymp 1$, with the more general version deferred to Theorem 17 in Appendix B. Here and throughout, $S_{i,j}$ (resp. $S^\star_{i,j}$) represents the (i,j)-th entry of the matrix $\boldsymbol{S}$ (resp. $\boldsymbol{S}^\star$).

Theorem 7. *Suppose that $p < 1 - \delta$ for some arbitrary constant $0 < \delta < 1$, and that $\kappa, \mu, r \asymp 1$. Consider any $1 \le i, j \le d$. Assume that $\boldsymbol{U}^\star$ is μ-incoherent and satisfies the following conditions*

$$\left\|\boldsymbol{U}^\star_{i,\cdot}\right\|_2 + \left\|\boldsymbol{U}^\star_{j,\cdot}\right\|_2 \gtrsim \left[\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d\log^5(n+d)}{np}} + \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d\log^5(n+d)}{n}}\right]\sqrt{\frac{1}{d}} \tag{3.22a}$$

and

$$\min\left\{\left\|\boldsymbol{U}^\star_{i,\cdot}\right\|_2, \left\|\boldsymbol{U}^\star_{j,\cdot}\right\|_2\right\} \gtrsim \left[\frac{\log^{7/2}(n+d)}{\sqrt{d}} + \frac{\log^5(n+d)}{\sqrt{ndp^2}} + \frac{\log^3(n+d)}{\sqrt{np}}\right]\sqrt{\frac{1}{d}}. \tag{3.22b}$$

In addition, suppose that Assumption 1 holds and that

$$d \gtrsim \log^5 n, \qquad np \gtrsim \log^5(n+d), \qquad ndp^2 \gtrsim \log^7(n+d)$$

and

$$\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\log^2(n+d)}, \qquad \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\log^3(n+d)}.$$

Assume that the number of iterations satisfies (3.16). Then the matrix $\boldsymbol{S}$ computed by Algorithm 2 obeys

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\frac{S_{i,j} - S^\star_{i,j}}{\sqrt{v^\star_{i,j}}} \le t\right) - \Phi(t)\right| = o(1),$$

where $\Phi(\cdot)$ denotes the CDF of the standard Gaussian distribution.

In words, the above theorem indicates that: if the ℓ_2 norm of the rows $\boldsymbol{U}^\star_{i,\cdot}$ and $\boldsymbol{U}^\star_{j,\cdot}$ are not exceedingly small, then the estimation error $S_{i,j} - S^\star_{i,j}$ of HeteroPCA is approximately a zero-mean Gaussian with variance $v^\star_{i,j}$. The required conditions in Theorem 7 mainly parallel the ones in Theorem 5.

***Remark* 7.** When there is no missing data (i.e., $p = 1$), we do not require the condition (3.22b).

***Remark* 8.** Similar to the SVD-based approach, we can show that the condition (3.22a) is equivalent to

$$\omega_{\max}^2 \le \widetilde{O}\left(\min\left\{np, \sqrt{ndp^2}\cdot\left(\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}\right)^{-1}\right\}\left(S^\star_{i,i} + S^\star_{j,j}\right)\right),$$

which allow the noise level $\omega_{\max}$ to be significantly larger than the standard deviation of $\{x_{i,l} : (i,l) \in \Omega\}$ and $\{x_{j,l} : (i,l) \in \Omega\}$ under our sampling rate condition and noise condition, and is hence very mild.

Construction of entrywise confidence intervals. The distributional characterization in Theorem 7 enables valid construction of entrywise confidence intervals for $\boldsymbol{S}^\star$, as long as we can obtain reliable estimate of the variance $v^\star_{i,j}$. In what follows, we come up with an algorithm — as summarized in Algorithm 6 — that attempts to estimate $v^\star_{i,j}$ and build confidence intervals in a data-driven manner, as confirmed by the following theorem for the scenario with $\kappa, \mu, r \asymp 1$. The more general result is postponed to Theorem 18 in Appendix B.

Theorem 8. *Instate the conditions in Theorem 7. Further assume that*

$$ndp^2 \gtrsim \log^8(n+d), \qquad np \gtrsim \log^6(n+d),$$

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\log^3(n+d)}, \qquad \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\log^{5/2}(n+d)}.$$

Then the confidence interval computed in Algorithm 6 obeys

$$\mathbb{P}\left(S^\star_{i,j} \in \mathsf{CI}^{1-\alpha}_{i,j}\right) = 1 - \alpha + o(1).$$

Algorithm 6 Confidence intervals for $S^\star_{i,j}$ $(1 \le i, j \le d)$ base on HeteroPCA.

Input: output $(\boldsymbol{U}, \boldsymbol{\Sigma}, \boldsymbol{S})$ of Algorithm 2, sampling rate p, coverage level $1-\alpha$.
Compute estimates of the noise level $\omega_l^\star$ as follows

$$\omega_l^2 := \frac{\sum_{j=1}^n y_{l,j}^2 \mathbb{1}_{(l,j)\in\Omega}}{\sum_{j=1}^n \mathbb{1}_{(l,j)\in\Omega}} - S_{l,l}.$$

Compute an estimate of $v^\star_{i,j}$ (cf. (3.20) or (3.21))as follows: if $i \neq j$ then

$$\begin{aligned} v_{i,j} := & \frac{2-p}{np} S_{i,i} S_{j,j} + \frac{4-3p}{np} S_{i,j}^2 + \frac{1}{np}\left(\omega_i^2 S_{j,j}^\star + \omega_j^2 S_{i,i}\right) \\ & + \frac{1}{np^2}\sum_{k=1}^d \left\{\left[\omega_i^2 + (1-p) S_{i,i}\right]\left[\omega_k^2 + (1-p) S_{k,k}\right] + 2(1-p)^2 S_{i,k}^2\right\}\left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^\top\right)^2 \\ & + \frac{1}{np^2}\sum_{k=1}^d \left\{\left[\omega_j^2 + (1-p) S_{j,j}\right]\left[\omega_k^2 + (1-p) S_{k,k}\right] + 2(1-p)^2 S_{j,k}^2\right\}\left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{i,\cdot}^\top\right)^2; \end{aligned}$$

If $i = j$ then

$$\begin{aligned} v_{i,i} := & \frac{12-9p}{np} S_{i,i}^2 + \frac{4}{np}\omega_i^2 S_{i,i} \\ & + \frac{4}{np^2}\sum_{k=1}^d \left\{\left[\omega_i^2 + (1-p) S_{i,i}\right]\left[\omega_k^2 + (1-p) S_{k,k}\right] + 2(1-p)^2 S_{i,k}^2\right\}\left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{i,\cdot}^\top\right)^2. \end{aligned}$$

Output the $(1-\alpha)$-confidence interval

$$\mathsf{CI}_{i,j}^{1-\alpha} := \left[S_{i,j} \pm \Phi^{-1}(1-\alpha/2)\sqrt{v_{i,j}}\right].$$

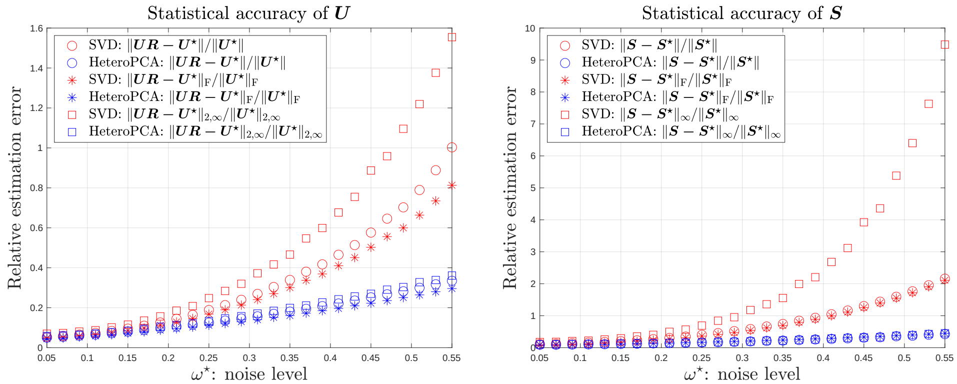
Compared with Cai et al. (2021, Corollary 2), we can see that when consistent estimation is possible — namely, under the sampling rate condition $p \ge \widetilde{\Omega}((n \wedge \sqrt{nd})^{-1})$ and the noise conditions $\omega_{\max} \le \widetilde{\Omega}((\sqrt{n/d} \wedge \sqrt[4]{n/d})\sqrt{p}\sigma_r^\star)$ — it is plausible to construct fine-grained confidence interval for $S^\star_{i,j}$, provided that the sizes of $\boldsymbol{U}^\star_{i,\cdot}$ and $\boldsymbol{U}^\star_{j,\cdot}$ are not exceedingly small in the sense that

$$\min\left\{\left\|\boldsymbol{U}^\star_{i,\cdot}\right\|_2, \left\|\boldsymbol{U}^\star_{j,\cdot}\right\|_2\right\} \gtrsim \widetilde{\Omega}\left(\frac{1}{\sqrt{ndp^2}} + \frac{1}{\sqrt{np}} + \frac{1}{\sqrt{d}} + \frac{\omega_{\max}^2}{\sigma_r^{\star 2}}\sqrt{\frac{d}{np^2}} + \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d}{np}}\right) \cdot \sqrt{\frac{1}{d}}\left\|\boldsymbol{U}^\star\right\|_{\mathrm{F}}.$$

4 Numerical experiments

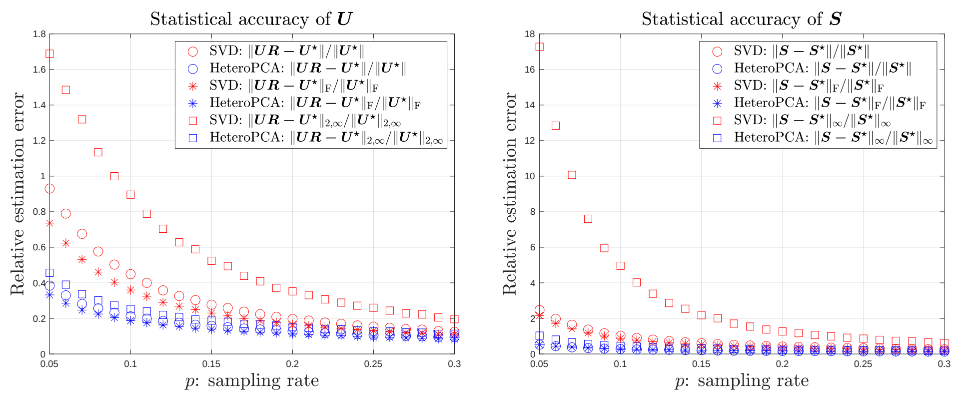
Setup. This section conducts a series of numerical experiments to validate our distributional and inference theory developed in Section 3. Throughout this section, we fix the dimension to be $d = 100$ and the number of sample vectors to be $n = 2000$, and we generate the covariance matrix as $\boldsymbol{S}^\star = \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}$ with $\boldsymbol{U}^\star \in \mathbb{R}^{n\times r}$ being a random orthonormal matrix. In each Monte Carlo trial, the observed data are produced according to the model described in Section 1.1. For the purpose of introducing heteroskedasticity, we will introduce a parameter $\omega^\star$ that controls the noise level: in each independent trial, each noise level $\omega_l^\star$ $(1 \le l \le d)$ is independently drawn from $\mathsf{Uniform}[0.5\omega^\star, 2\omega^\star]$; the random noise component $\eta_{l,j}$ is then drawn from $\mathcal{N}(0, \omega_l^{\star 2})$ independently for every $l \in [d]$ and $j \in [n]$.

Superiority of HeteroPCA to the SVD-based approach in estimation. To begin with, we first compare the empirical estimation accuracy of the SVD approach (cf. Algorithm 1) and HeteroPCA (cf. Algorithm 2). Figure 1 displays the relative estimation errors — including the ones tailored to the principal



(a) (b)

Figure 1: The relative estimation error of $\boldsymbol{U}$ and $\boldsymbol{S}$ returned by both SVD-based approach (cf. Algorithm 1) and HeteroPCA (cf. Algorithm 2). (a) Relative estimation errors of $\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^\star$ measured by $\|\cdot\|$, $\|\cdot\|_{\mathrm{F}}$ and $\|\cdot\|_{2,\infty}$ vs. the noise level $\omega^\star$; (b) Relative estimation errors of $\boldsymbol{S}-\boldsymbol{S}^\star$ measured by $\|\cdot\|$, $\|\cdot\|_{\mathrm{F}}$ and $\|\cdot\|_{\infty}$ vs. the noise level $\omega^\star$. The results are reported over 200 independent trials for $r=3$ and $p=0.6$.

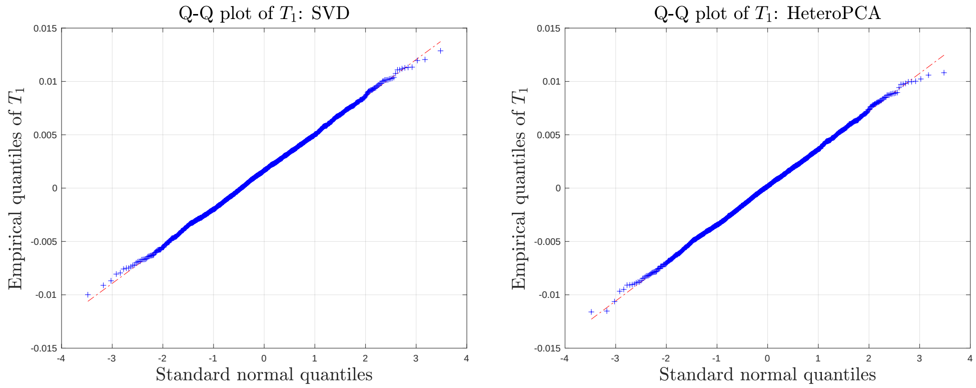


(a) (b)

Figure 2: The relative estimation error of $\boldsymbol{U}$ and $\boldsymbol{S}$ returned by both SVD-based approach (cf. Algorithm 1) and HeteroPCA (cf. Algorithm 2). (a) Relative estimation errors of $\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^\star$ measured by $\|\cdot\|$, $\|\cdot\|_{\mathrm{F}}$ and $\|\cdot\|_{2,\infty}$ vs. the missing rate p; (b) Relative estimation errors of $\boldsymbol{S}-\boldsymbol{S}^\star$ measured by $\|\cdot\|$, $\|\cdot\|_{\mathrm{F}}$ and $\|\cdot\|_{\infty}$ vs. the missing rate p. The results are reported over 200 independent trials for $r=3$ and $\omega^\star=0.05$.

subspace: $\|\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^\star\|/\|\boldsymbol{U}^\star\|$, $\|\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^\star\|_{\mathrm{F}}/\|\boldsymbol{U}^\star\|_{2,\infty}$, $\|\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^\star\|_{2,\infty}/\|\boldsymbol{U}^\star\|_{2,\infty}$, and the ones tailored to the covariance matrix: $\|\boldsymbol{S}-\boldsymbol{S}^\star\|/\|\boldsymbol{S}^\star\|$, $\|\boldsymbol{S}-\boldsymbol{S}^\star\|_{\mathrm{F}}/\|\boldsymbol{S}^\star\|_{\mathrm{F}}$, $\|\boldsymbol{S}-\boldsymbol{S}^\star\|_{\infty}/\|\boldsymbol{S}^\star\|_{\infty}$ — of both algorithms as the noise level $\omega^\star$ varies, with $r=3$ and $p=0.6$. Similarly, Figure 2 displays the relative numerical estimation errors of both algorithms vs. the sampling rate p, with $r=3$ and $\omega^\star=0.05$. As we shall see from both figures, HeteroPCA uniformly outperforms the SVD-based approach in all experiments, and is able to achieve appealing performance for a much wider range of noise levels and sampling rates, both of which match our theoretical findings.

Confidence regions for the principal subspace $\boldsymbol{U}^\star$. Next, we carry out a series of experiments to corroborate the practical validity of the confidence regions constructed using the SVD-based approach (cf. Algorithm 3) and HeteroPCA (cf. Algorithm 5). To this end, we define $\widehat{\mathsf{Cov}}_U(i)$ to be the empirical probability that the constructed confidence interval $\mathsf{CR}_{U,i}^{0.95}$ covers $\boldsymbol{U}_{i,\cdot}^\star\mathsf{sgn}(\boldsymbol{U}^{\star\top}\boldsymbol{U})$ over 200 Monte Carlo trials, where $\boldsymbol{U}$ is the estimate returned by either algorithm. We also let $\mathsf{Mean}(\widehat{\mathsf{Cov}}_U)$ (resp. $\mathsf{std}(\widehat{\mathsf{Cov}}_U)$) be the empirical mean (resp. standard deviation) of $\widehat{\mathsf{Cov}}_U(i)$ over $i\in[d]$. Table 2 gathers $\mathsf{Mean}(\widehat{\mathsf{Cov}}$



(a) (b)

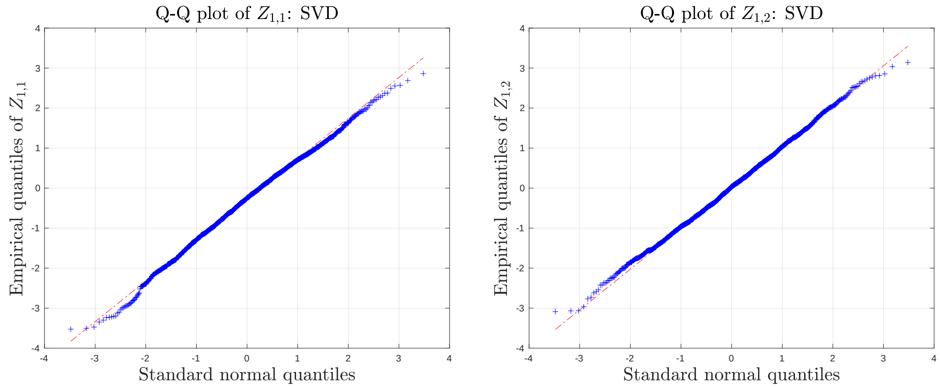
Figure 3: (a) Q-Q (quantile-quantile) plot of T_1 vs. the standard normal distribution for the SVD-based approach; (b) Q-Q (quantile-quantile) plot of T_1 vs. the standard normal distribution for HeteroPCA. The results are reported over 2000 independent trials for $r = 1$, $p = 0.6$ and $\omega^\star = 0.05$.

and $\mathsf{std}(\widehat{\mathsf{Cov}})$ for $r = 3$ and different choices of $(p, \omega^\star)$ for both algorithms. Encouragingly, the empirical coverage rates are all close to 95% for both methods when p is not too small and $\omega^\star$ is not too large. When p becomes smaller or $\omega^\star$ grows larger, HeteroPCA is still capable of performing valid statistical inference, while the SVD-based approach fails. This provides another empirical evidence on the advantage and broader applicability of HeteroPCA compared to the SVD-based approach. In addition, for the rank-1 case ($r = 1$), we define $T_i := [U - \mathsf{sign}(\boldsymbol{U}^\top \boldsymbol{U}^\star) U^\star]_i / \sqrt{\Sigma_{U,i}}$. Figure 3 displays the Q-Q (quantile-quantile) plot of $T_1 := [U - \mathsf{sign}(\boldsymbol{U}^\top \boldsymbol{U}^\star) U^\star]_1 / \sqrt{\Sigma_{U,1}}$ vs. the standard Gaussian random variable over 2000 Monte Carlo simulations for both algorithms (when $p = 0.6$ and $\omega^\star = 0.05$); the near-Gaussian empirical distribution of T_1 also corroborates our distributional guarantees.

Entrywise confidence intervals for $\boldsymbol{S}^\star$. Finally, we provide numerical evidence that confirms the validity of the confidence interval constructed on the basis of the SVD-based approach (cf. Algorithm 4) and HeteroPCA (cf. Algorithm 6). Define $\widehat{\mathsf{Cov}}_S(i,j)$ to be the empirical probability that the 95% confidence interval $[S_{i,j} \pm 1.96\sqrt{v_{i,j}}]$ covers $S^\star_{i,j}$ over 200 Monte Carlo trials, where $S_{i,j}$ is the (i,j)-th entry of the estimate $\boldsymbol{S}$ returned by either algorithm. Let $\mathsf{Mean}(\widehat{\mathsf{Cov}}_S)$ (resp. $\mathsf{std}(\widehat{\mathsf{Cov}}_S)$) be the empirical mean (resp. standard deviation) of $\widehat{\mathsf{Cov}}_S(i,j)$ over all $i, j \in [d]$. Table 2collects $\mathsf{Mean}(\widehat{\mathsf{Cov}})$ and $\mathsf{std}(\widehat{\mathsf{Cov}})$ for $r = 3$ and accounts for different choices of $(p, \omega^\star)$ for both algorithms. Similar to previous experiments, HeteroPCA uniformly outperforms the SVD-based approach, which again suggests that HeteroPCA is the method of choice. In addition, we define $Z_{i,j} := (S_{i,j} - S^\star_{i,j})/\sqrt{v_{i,j}}$. For both algorithms, Figure 4 and Figure 5 depict the Q-Q (quantile-quantile) plot of $Z_{1,1}$ and $Z_{1,2}$ vs. standard Gaussian distributions over 2000 Monte Carlo trials for the case with $r = 3$, $p = 0.6$ and $\omega^\star = 0.05$, which again confirm the practical validity of our distributional theory.

Table 1: Empirical coverage rates of $\boldsymbol{U}^\star \mathsf{sgn}(\boldsymbol{U}^{\star\top} \boldsymbol{U})$ for different $(p, \omega^\star)$'s over 200 Monte Carlo trials

		The SVD-based Approach		HeteroPCA	
p	$\omega^\star$	$\mathsf{Mean}(\widehat{\mathsf{Cov}})$	$\mathsf{Std}(\widehat{\mathsf{Cov}})$	$\mathsf{Mean}(\widehat{\mathsf{Cov}})$	$\mathsf{Std}(\widehat{\mathsf{Cov}})$
0.6	0.05	0.9265	0.0275	0.9500	0.0156
0.6	0.1	0.9016	0.0443	0.9468	0.0149
0.4	0.05	0.8839	0.0489	0.9433	0.0182
0.4	0.1	0.8490	0.0738	0.9395	0.0175
0.2	0.05	0.7382	0.1135	0.9279	0.0206
0.2	0.1	0.6120	0.1627	0.9322	0.0188



(a) (b)

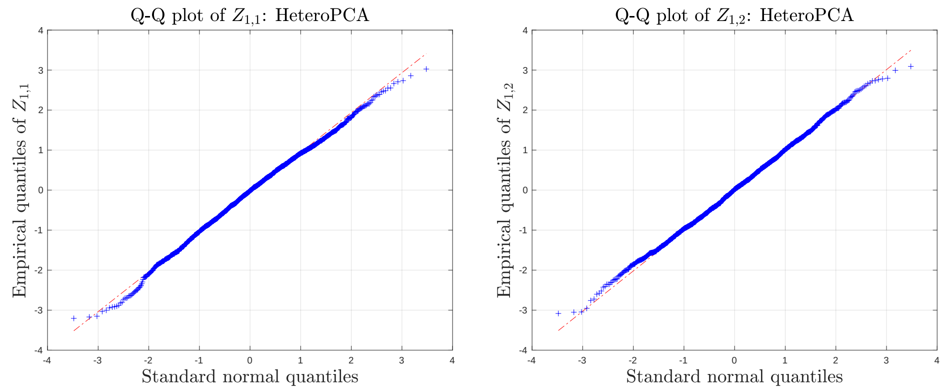
Figure 4: (a) Q-Q (quantile-quantile) plot of $Z_{1,1}$ vs. the standard normal distribution for the SVD-based approach; (b) Q-Q (quantile-quantile) plot of $Z_{1,2}$ vs. a standard Gaussian distribution for the SVD-based approach. The results are reported over 2000 independent trials for $r = 3$, $p = 0.6$, $\omega^\star = 0.05$.

5 Other related works

Low-rank matrix denoising serves as a common model to study the effectiveness of spectral methods (Chen et al., 2020c), and has been the main subject of many prior works including Abbe et al. (2020); Agterberg et al. (2021); Bao et al. (2018); Cai and Zhang (2018); Cape et al. (2019); Chen et al. (2021); Ding (2020); Lei (2019); Montanari et al. (2018); Xia (2019b), among others. Several recent works began to pursue a distributional theory for the eigenvector or singular vectors of the observed data matrix (Bao et al., 2018; Cheng et al., 2020; Fan et al., 2020; Xia, 2019b). To name a few examples, Bao et al. (2020) studied the limiting distribution of the inner product between an empirical singular vector and the corresponding ground truth, assuming that the associated spectral gap is sufficient large and that the noise components are homoskedastic; Xia (2019b) established non-asymptotic Gaussian approximation for certain projection distance in the presence of i.i.d. Gaussian noise. Furthermore, the presence of missing data forms another source of technical challenges, leading to a problem often dubbed as noisy low-rank matrix completion (Candès and Plan, 2010; Chen et al., 2020d; Negahban and Wainwright, 2012). Spectral methods have been successfully applied to tackle noisy matrix completion (Chen et al., 2020a; Chen and Wainwright, 2015; Cho et al., 2017; Keshavan et al., 2010a; Ma et al., 2020; Sun and Luo, 2016; Zheng and Lafferty, 2016), which commonly serve as an effective initialization scheme for nonconvex optimization methods (Chi et al., 2019). While statistical inference for noisy matrix completion has been investigated recently (Chen et al., 2019c; Chernozhukov et al., 2021; Xia and Yuan, 2021), these prior works focused on performing inference based on optimization-based estimators. How to construct fine-grained confidence intervals based on spectral methods remains previously out of reach for noisy matrix completion. The recent work (Xia, 2019a) tackled the confidence regions for spectral estimators tailored to the low-rank matrix regression problem, without accommodating the noisy matrix completion context. Most importantly, while the SVD-based vanilla spectral

Table 2: Empirical coverage rates of $S^\star_{i,j}$ for different $(\omega^\star, p)$'s over 200 Monte Carlo trials

		The SVD-based Approach		HeteroPCA	
p	$\omega^\star$	$\mathsf{Mean}(\widehat{\mathsf{Cov}})$	$\mathsf{Std}(\widehat{\mathsf{Cov}})$	$\mathsf{Mean}(\widehat{\mathsf{Cov}})$	$\mathsf{Std}(\widehat{\mathsf{Cov}})$
0.6	0.05	0.9378	0.0249	0.9474	0.0154
0.6	0.1	0.9228	0.0444	0.9486	0.0153
0.4	0.05	0.9197	0.0517	0.9482	0.0156
0.4	0.1	0.9007	0.0741	0.9487	0.0153
0.2	0.05	0.8670	0.1028	0.9490	0.0164
0.2	0.1	0.8501	0.1197	0.9488	0.0163



(a) (b)

Figure 5: (a) Q-Q (quantile-quantile) plot of $Z_{1,1}$ vs. the standard normal distribution for HeteroPCA; (b) Q-Q (quantile-quantile) plot of $Z_{1,2}$ vs. a standard Gaussian distribution for HeteroPCA. The results are reported over 2000 independent trials for $r = 3$, $p = 0.6$, $\omega^\star = 0.05$.

method often works well for the balanced case (such that the column dimension and the row dimension are on the same order), sub-optimality has been well recognized when estimating the column subspace of interest in the highly unbalanced case (so that the column dimension far exceeds the row dimension); this issue is also present when it comes to existing optimization-based methods like nuclear norm minimization. As a result, all prior schemes mentioned in this paragraph failed to tackle the highly balanced case in an statistically efficient manner.

Turning to PCA or subspace estimation, there has been an enormous literature dedicated to this topic; see Balzano et al. (2018); Johnstone and Paul (2018) for an overview of prior development. Noteworthily, the need to handle the diagonals of the sample covariance matrix in the presence of heteroskedastic noise and/or missing data has been pointed out in many prior works, e.g., Cai et al. (2021); Florescu and Perkins (2016); Loh and Wainwright (2012); Lounici (2014); Montanari and Sun (2018). The iterative refinement scheme proposed by Zhang et al. (2018) turns out to be among the most effective and adaptive schemes in handling the diagonals. Aimed at designing fine-grained estimators for the principal components, Koltchinskii et al. (2020); Li et al. (2021) proposed statistically efficient de-biased estimators for linear functionals of principal components, and moreover, the estimator proposed in Koltchinskii et al. (2020) has also been shown to exhibit asymptotic normality in the presence of i.i.d. Gaussian noise. Bloemendal et al. (2016) also pinned down the asymptotic distributions of certain principal components under a spiked covariance model. However, these papers fell short of presenting valid and data-driven uncertainty quantification methods for the proposed estimators, and their results operates under the assumptions of homoskedastic noise without any missing data, a scenario that is remarkably more restricted than ours. Under the spiked covariance model, Bao et al. (2020) studied the limiting distribution of the angle between the eigenvectors of the sample covariance matrix and any fixed vector, under the "balanced" scenario where the aspect ratio n/d is a constant. In addition, recent years have witnessed much activity in high-dimensional PCA in the face of missing data (Cai et al., 2021; Pavez and Ortega, 2020; Zhang et al., 2018; Zhu et al., 2019); these works, however, focused primarily on developing estimation guarantees, which did not provide either distributional guarantees for the estimators or concrete procedures that allow for confidence region construction.

From a technical viewpoint, it is worth mentioning that the ℓ_∞ and $\ell_{2,\infty}$ perturbation theory has been an active research direction in recent years (Agterberg et al., 2021; Cape et al., 2019; Chen et al., 2021; Eldridge et al., 2018; Fan et al., 2018; Xie, 2021). Among multiple existing technical frameworks, the leave-one-out analysis idea — which has been applied to a variety of statistical estimation problems (Cai et al., 2020a,b; Chen et al., 2020b, 2019a, 2020e; El Karoui, 2015; El Karoui et al., 2013; Ling, 2020; Zhong and Boumal, 2018) — provides a powerful and flexible framework that enables ℓ_∞ and $\ell_{2,\infty}$ statistical guarantees for spectral methods (Abbe et al., 2020; Cai et al., 2021; Chen et al., 2019b); see (Chen et al., 2020c, Chapter 4) for an accessible introduction of this powerful framework. Our theory for the SVD-based approach and matrix denoising is inspired by the analysis of Abbe et al. (2020), while our analysis for the HeteroPCA approach is influenced by the one in Cai et al. (2021). Note, however, that none of these works came with any distributional guarantees for spectral methods, which we seek to accomplish in this paper.

Finally, we note in passing that constructing confidence intervals for sparse regression (based on, say, the Lasso estimator or other sparsity-promoting estimator), has attracted a flurry of research activity in the past few years (Cai and Guo, 2017; Celentano et al., 2020; Javanmard and Montanari, 2014; Ning and Liu, 2017; Ren et al., 2015; van de Geer et al., 2014; Zhang and Zhang, 2014). The methods derived therein, however, are not directly applicable to perform statistical inference for PCA and/or other low-rank models.

6 A detour: matrix denoising and subspace estimation

We now take a detour to look at two intimately related problems called *matrix denoising* and *subspace estimation*, which will play a crucial role in understanding the SVD-based approach and the HeteroPCA approach, respectively. This section primarily focuses on developing fine-grained understanding for these two problems; the connections between these problems and PCA will be illuminated in Appendices D.1 and G.1.

6.1 Matrix denoising

This subsection introduces a general matrix denoising model, and develops some fine-grained understanding that will assist in studying the SVD-based approach for PCA (as we shall detail in Appendix D). In what follows, we shall start by formulating the problem of matrix denoising, followed by a useful spectral perturbation bound.

6.1.1 Model and assumptions: matrix denoising

Suppose that we are interested in a rank-r matrix $\boldsymbol{M}^{\natural} \in \mathbb{R}^{n_1 \times n_2}$, whose SVD is given by

$$\boldsymbol{M}^{\natural} = \boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural}\boldsymbol{V}^{\natural\top} = \sum_{i=1}^{r} \sigma_i^{\natural}\boldsymbol{u}_i^{\natural}\boldsymbol{v}_i^{\natural\top}. \tag{6.1}$$

Here, $\boldsymbol{U}^{\natural} = [\boldsymbol{u}_1^{\natural}, \dots, \boldsymbol{u}_r^{\natural}]$ (resp. $\boldsymbol{V}^{\natural} = [\boldsymbol{v}_1^{\natural}, \dots, \boldsymbol{v}_r^{\natural}]$) consists of orthonormal columns that correspond to the left (resp. right) singular vectors of $\boldsymbol{M}^{\natural}$, and $\boldsymbol{\Sigma}^{\natural} = \mathsf{diag}\{\sigma_1^{\natural}, \dots, \sigma_r^{\natural}\}$ is a diagonal matrix consisting of the singular values of $\boldsymbol{M}^{\natural}$. Without loss of generality, we assume that

$$n = \max\{n_1, n_2\}.$$

It is assumed that the singular values are sorted (in magnitude) in descending order, namely,

$$\sigma_1^{\natural} \geq \cdots \geq \sigma_r^{\natural} \geq 0, \tag{6.2}$$

with the condition number denoted by

$$\kappa^{\natural} := \sigma_1^{\natural}/\sigma_r^{\natural}. \tag{6.3}$$

What we have observed is a noisy copy of $\boldsymbol{M}^{\natural}$, namely,

$$\boldsymbol{M} = \boldsymbol{M}^{\natural} + \boldsymbol{E}, \tag{6.4}$$

where $\boldsymbol{E} = [E_{i,j}]_{1 \leq i,j \leq n}$ stands for a noise matrix.

In addition, we impose the following two assumptions on both the ground-truth matrix $\boldsymbol{M}^{\star}$ and the noise matrix $\boldsymbol{E}$.

Assumption 2 (Incoherence). *A rank-r matrix $\boldsymbol{M}^{\natural} \in \mathbb{R}^{n_1 \times n_2}$ with SVD $\boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural}\boldsymbol{V}^{\natural\top}$ is said to be $\mu^{\natural}$-incoherent if*

$$\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} \leq \sqrt{\frac{\mu^{\natural}}{n_1}}\left\|\boldsymbol{U}^{\natural}\right\|_{\mathrm{F}} = \sqrt{\frac{\mu^{\natural} r}{n_1}} \qquad \textit{and} \qquad \left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} \leq \sqrt{\frac{\mu^{\natural}}{n_2}}\left\|\boldsymbol{V}^{\natural}\right\|_{\mathrm{F}} = \sqrt{\frac{\mu^{\natural} r}{n_2}}.$$

Assumption 3 (**Heteroskedastic random noise**). *Assume that the noise $E_{i,j}$'s are independent random variables and that*

$$\mathbb{E}\left[E_{i,j}\right] = 0, \qquad \mathbb{E}\left[E_{i,j}^2\right] = \sigma_{i,j}^2 \le \sigma^2, \qquad |E_{i,j}| \le B$$

hold for each $(i,j) \in [n_1] \times [n_2]$, where

$$B \lesssim \sigma \sqrt{\frac{\min\{n_1, n_2\}}{\mu^\natural \log n}}. \tag{6.5}$$

To simplify notation, we shall assume, without loss of generality, that

$$n_1 \le n_2. \tag{6.6}$$

6.1.2 $\ell_{2,\infty}$ perturbation bounds for singular subspaces

We would like to bound the perturbation of the singular subspaces in response to the noise matrix. Specifically, letting $\boldsymbol{U} \in \mathbb{R}^{n_1 \times r}$ (resp. $\boldsymbol{V} \in \mathbb{R}^{n_2 \times r}$) denote an orthonormal matrix whose columns are the top-r left (resp. right) singular vectors of $\boldsymbol{M}$, we seek to bound the deviation of $\boldsymbol{U}$ (resp. $\boldsymbol{V}$) from $\boldsymbol{U}^\natural$ (resp. $\boldsymbol{V}^\natural$). Our results are stated in the following theorem. Here and below, we let $n = n_1 \vee n_2$, and introduce the global rotation matrices

$$\boldsymbol{R}_{\boldsymbol{U}} \coloneqq \arg\min_{\boldsymbol{O} \in \mathcal{O}^{r\times r}} \left\|\boldsymbol{U}\boldsymbol{O} - \boldsymbol{U}^\natural\right\|_{\mathrm{F}}^2 \qquad \text{and} \qquad \boldsymbol{R}_{\boldsymbol{V}} \coloneqq \arg\min_{\boldsymbol{O} \in \mathcal{O}^{r\times r}} \left\|\boldsymbol{V}\boldsymbol{O} - \boldsymbol{V}^\natural\right\|_{\mathrm{F}}^2.$$

Theorem 9. *Consider the above matrix denoising setting. Suppose that Assumptions 2-3 hold. In addition, assume that $\min\{n_1, n_2\} \gtrsim \mu^\natural r$ and*

$$\sigma\sqrt{n \log n} \ll \sigma_r^\natural / \kappa^\natural.$$

Then with probability exceeding $1 - O(n^{-10})$, the top-r singular subspaces $\boldsymbol{U}$ and $\boldsymbol{V}$ of $\boldsymbol{M}$ can be written as

$$\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^\natural = \boldsymbol{E}\boldsymbol{V}^\natural \left(\boldsymbol{\Sigma}^\natural\right)^{-1} + \boldsymbol{\Psi}_{\boldsymbol{U}}, \tag{6.7a}$$

$$\boldsymbol{V}\boldsymbol{R}_{\boldsymbol{V}} - \boldsymbol{V}^\natural = \boldsymbol{E}^\top\boldsymbol{U}^\natural \left(\boldsymbol{\Sigma}^\natural\right)^{-1} + \boldsymbol{\Psi}_{\boldsymbol{V}}, \tag{6.7b}$$

where $\boldsymbol{\Psi}_{\boldsymbol{U}}$ and $\boldsymbol{\Psi}_{\boldsymbol{V}}$ are some matrices obeying

$$\left\|\boldsymbol{\Psi}_{\boldsymbol{U}}\right\|_{2,\infty} \lesssim \frac{\sigma^2\sqrt{nr}\log n}{\sigma_r^{\natural 2}} + \left(\frac{\kappa^\natural \sigma^2 n}{\sigma_r^{\natural 2}} + \frac{\sigma\sqrt{r\log n}}{\sigma_r^\natural}\right)\left\|\boldsymbol{U}^\natural\right\|_{2,\infty},$$

$$\left\|\boldsymbol{\Psi}_{\boldsymbol{V}}\right\|_{2,\infty} \lesssim \frac{\sigma^2\sqrt{nr}\log n}{\sigma_r^{\natural 2}} + \left(\frac{\kappa^\natural \sigma^2 n}{\sigma_r^{\natural 2}} + \frac{\sigma\sqrt{r\log n}}{\sigma_r^\natural}\right)\left\|\boldsymbol{V}^\natural\right\|_{2,\infty}.$$

Proof. See Appendix C. □

In a nutshell, Theorem 9 uncovers a sort of error decomposition taking the form of 6.7. Given that $\boldsymbol{\Psi}_{\boldsymbol{U}}$ and $\boldsymbol{\Psi}_{\boldsymbol{V}}$ are sufficiently small in size, we have essentially established the goodness of the following first-order approximation

$$\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^\natural \approx \boldsymbol{E}\boldsymbol{V}^\natural \left(\boldsymbol{\Sigma}^\natural\right)^{-1}, \tag{6.8a}$$

$$\boldsymbol{V}\boldsymbol{R}_{\boldsymbol{V}} - \boldsymbol{V}^\natural \approx \boldsymbol{E}^\top\boldsymbol{U}^\natural \left(\boldsymbol{\Sigma}^\natural\right)^{-1}, \tag{6.8b}$$

which are linear mappings of the noise matrix $\boldsymbol{E}$. As we shall elucidate momentarily, the near-tightness of such linear mappings reveals that the perturbation follows zero-mean Gaussian distributions approximately — a feature that can be justified using, say, non-asymptotic versions of the Berry–Esseen theorem (Chen et al., 2010).

6.2 Subspace estimation

This subsection studies another setting — which we shall refer to as subspace estimation — very similar to matrix denoising. We will set out to develop a fine-grained statistical theory for HeteroPCA when applied to this subspace estimation setting. The resulting theory will be invoked in Appendix G to analyze the PCA context.

6.2.1 A general subspace estimation model

Model and assumptions. Reusing the notation for matrix denoising in Section 6.1, we consider an unknown rank-r matrix $\boldsymbol{M}^{\natural}$ with SVD

$$\boldsymbol{M}^{\natural} = \sum_{i=1}^{r} \sigma_i^{\natural} \boldsymbol{u}_i^{\natural} \boldsymbol{v}_i^{\natural\top} = \boldsymbol{U}^{\natural} \boldsymbol{\Sigma}^{\natural} \boldsymbol{V}^{\natural\top} \in \mathbb{R}^{n_1 \times n_2} \tag{6.9}$$

with condition number $\kappa^{\natural} \coloneqq \sigma_1^{\natural}/\sigma_r^{\natural}$, and as before, we only observe a randomly corrupted copy of $\boldsymbol{M}^{\natural}$:

$$\boldsymbol{M} = \boldsymbol{M}^{\natural} + \boldsymbol{E}. \tag{6.10}$$

In contrast to the matrix denoising setting where one seeks to estimate $\boldsymbol{U}^{\natural}$, $\boldsymbol{\Sigma}^{\natural}$ and $\boldsymbol{V}^{\natural}$ (and hence $\boldsymbol{M}^{\natural}$), this subsection focuses primarily on *estimating the column subspace* represented by $\boldsymbol{U}^{\natural}$ and the singular values encapsulated in $\boldsymbol{\Sigma}^{\natural}$, but not the row space $\boldsymbol{V}^{\natural}$. An important special scenario one should bear in mind is the highly unbalanced case where the column dimension n_2 far exceeds the row dimension n_1; in this case, it is common to encounter situations where reliable estimation of $\boldsymbol{M}^{\natural}$ and $\boldsymbol{V}^{\natural}$ is infeasible but that of $\boldsymbol{U}^{\natural}$ shows promise. For this reason, we refer to this setting as *subspace estimation* in order to differentiate it from matrix denoising, emphasizing that we are only interested in column subspace estimation.

With the new aim in mind, we shall modify our incoherence and noise assumptions accordingly. Here, we abuse the notation with the understanding that the following set of assumptions will be used only when analyzing the approach based on HeteroPCA. We shall also denote $n \coloneqq \max\{n_1, n_2\}$.

Assumption 4 (Incoherence). *The rank-r matrix $\boldsymbol{M}^{\natural} \in \mathbb{R}^{n_1 \times n_2}$ defined in (6.9) is said to be $\mu^{\natural}$-incoherent if the following holds:*

$$\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} \leq \sqrt{\frac{\mu^{\natural} r}{n_1}}, \quad \left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} \leq \sqrt{\frac{\mu^{\natural} r}{n_2}}, \quad \text{and } \left\|\boldsymbol{M}^{\natural}\right\|_{\infty} \leq \sqrt{\frac{\mu^{\natural}}{n_1 n_2}} \left\|\boldsymbol{M}^{\natural}\right\|_{\mathrm{F}}.$$

Assumption 5 (Heteroskedastic random noise). *Assume that the $E_{i,j}$'s are independently generated, and suppose that there exist quantities $\sigma, B \geq 0$ obeying*

$$\forall (i,j) \in [n_1] \times [n_2]: \qquad \mathbb{E}\left[E_{i,j}\right] = 0, \qquad \mathsf{var}\left(E_{i,j}^2\right) = \sigma_{i,j}^2 \leq \sigma^2, \qquad |E_{i,j}| \leq B,$$

where

$$B \lesssim \frac{\sigma \min\left\{\sqrt{n_2}, \sqrt[4]{n_1 n_2}\right\}}{\sqrt{\log n}}. \tag{6.11}$$

The major change compared to the matrix denoising setting lies in Condition 6.11: when $n_1 \ll n_2$, this condition 6.11 is clearly much less stringent than the condition (6.5) assumed previously.

Algorithm: HeteroPCA for subspace estimation. The paradigm HeteroPCA can naturally be applied to tackle the above subspace estimation task. Let us introduce the ground-truth gram matrix as follows

$$\boldsymbol{G}^{\natural} \coloneqq \boldsymbol{M}^{\natural} \boldsymbol{M}^{\natural\top}. \tag{6.12}$$

Given that $\boldsymbol{M} = \boldsymbol{M}^{\natural} + \boldsymbol{E}$ is an unbiased estimate of $\boldsymbol{M}^{\natural}$, one might naturally attempt to estimate the column space of $\boldsymbol{M}$ by looking at the eigenspace of the sample Gram matrix $\boldsymbol{M}\boldsymbol{M}^{\top}$. It can be easily seen that

$$\boldsymbol{E}\left[\boldsymbol{M}\boldsymbol{M}^{\top}\right] = \boldsymbol{M}^{\natural} \boldsymbol{M}^{\natural\top} + \mathsf{diag}\left\{\left[\sum_{j=1}^{n_2} \sigma_{i,j}^2\right]_{1 \leq i \leq n_1}\right\}, \tag{6.13}$$

where the diagonal term on the right-hand side of (6.13) might incur significant bias in the most challenging regime. The HeteroPCA algorithm seeks to handle the diagonal part in an iterative manner, alternating between imputing the values of the diagonal entries and eigen-decomposition of $\boldsymbol{M}\boldsymbol{M}^{\top}$ with the diagonal replaced by the imputed values. The procedure is summarized in Algorithm 7.

Algorithm 7 HeteroPCA for general subspace estimation (HeteroPCA).

Initialization: set $\boldsymbol{G}^0 = \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{M}\boldsymbol{M}^{\top}\right)$.
Updates: **for** $t = 0, 1, \dots, t_0$ **do**

$$\left(\boldsymbol{U}^t, \boldsymbol{\Lambda}^t\right) = \mathsf{eigs}\left(\boldsymbol{G}^t, r\right); \tag{6.14a}$$

$$\boldsymbol{G}^{t+1} = \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}^t\right) = \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{M}\boldsymbol{M}^{\top}\right) + \mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{U}^t\boldsymbol{\Lambda}^t\boldsymbol{U}^{t\top}\right). \tag{6.14b}$$

Here, $\mathsf{eigs}(\boldsymbol{G}, r)$ returns $(\boldsymbol{U}, \boldsymbol{\Lambda})$ where $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\top}$ is the top-r eigen-decomposition of $\boldsymbol{G}$.
Output: $\boldsymbol{U} = \boldsymbol{U}^{t_0}$, $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^{t_0}$, $\boldsymbol{\Sigma} = (\boldsymbol{\Lambda}^{t_0})^{1/2}$, $\boldsymbol{S} = \boldsymbol{U}^{t_0}\boldsymbol{\Lambda}^{t_0}\boldsymbol{U}^{t_0\top}$.

6.2.2 Fine-grained statistical guarantees for HeteroPCA

We are now in a position to present our theoretical guarantees for Algorithm 7. In order to account for the potential global rotational ambiguity, we introduce the following rotation matrix as before

$$\boldsymbol{R}_{\boldsymbol{U}} \coloneqq \arg\min_{\boldsymbol{O}\in\mathcal{O}^{r\times r}} \left\|\boldsymbol{U}\boldsymbol{O} - \boldsymbol{U}^{\natural}\right\|_{\mathrm{F}}^2, \tag{6.15}$$

where we recall that $\mathcal{O}^{r\times r}$ represents the set of $r \times r$ orthonormal matrices. It is also helpful to define the following quantity:

$$\zeta_{\mathsf{op}} \coloneqq \sigma^2\sqrt{n_1 n_2}\log n + \sigma\sigma_1^{\natural}\sqrt{n_1\log n}. \tag{6.16}$$

Our result is as follows, with the proof postponed to Appendix F.

Theorem 10. *Suppose that Assumptions 4-5 hold. Assume that*

$$n_1 \gtrsim \kappa^{\natural 4}\mu^{\natural}r + \mu^{\natural 2}r\log^2 n, \qquad n_2 \gtrsim r\log^4 n, \qquad \text{and} \qquad \zeta_{\mathsf{op}} \ll \frac{\sigma_r^{\natural 2}}{\kappa^{\natural 2}}, \tag{6.17}$$

and that the algorithm is run for $t_0 \geq \log\left(\frac{\sigma_1^{\star 2}}{\zeta_{\mathsf{op}}}\right)$ *iterations. With probability exceeding* $1 - O(n^{-10})$*, there exist two matrices* $\boldsymbol{Z}$ *and* $\boldsymbol{\Psi}$ *such that the estimates returned by HeteroPCA obey*

$$\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^{\natural} = \boldsymbol{Z} + \boldsymbol{\Psi}, \tag{6.18}$$

where

$$\boldsymbol{Z} \coloneqq \boldsymbol{E}\boldsymbol{V}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-1} + \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^{\top}\right)\boldsymbol{U}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-2}, \tag{6.19a}$$

$$\|\boldsymbol{\Psi}\|_{2,\infty} \lesssim \kappa^{\natural 2}\frac{\mu^{\natural}r}{n_1}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}} + \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}}\sqrt{\frac{\mu^{\natural}r}{n_1}}. \tag{6.19b}$$

The expressions (6.18) and (6.19) make apparent a key decomposition of the estimation error. As we shall see, the term $\boldsymbol{Z}$ is often the dominant term, which captures both the first-order and second-order approximation (w.r.t. the noise matrix $\boldsymbol{E}$) of the estimation error. Unless the noise level σ is very small, we cannot simply ignore the second-order term $\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^{\top}\right)\boldsymbol{U}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-2}$, as it is not necessarily dominated in size by the linear mapping term $\boldsymbol{E}\boldsymbol{V}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-1}$. The simple and closed-form expression of $\boldsymbol{Z}$ — in conjunction with the fact that $\boldsymbol{\Psi}$ is well-controlled — plays a crucial role when developing a non-asymptotic distributional theory.

While Theorem 10 is established mainly to help derive distributional characterizations for PCA, we remark that our analysis also delivers $\ell_{2,\infty}$ statistical guarantees in terms of estimating $\boldsymbol{U}^{\natural}$ (see Lemma 28 in the appendix). More specifically, our analysis asserts that

$$\left\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}\lesssim\kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^{\natural}r}{n_1}} \tag{6.20}$$

with high probability, under the conditions of Theorem 10. It is perhaps helpful to compare (6.20) with prior $\ell_{2,\infty}$ theory concerning estimation of $\boldsymbol{U}^{\natural}$.

- We first compare Theorem 10 with the recent work (Agterberg et al., 2021, Theorem 2), which focused on the regime $n_2 \gtrsim n_1$ and showed that

$$\inf_{\boldsymbol{O}\in\mathcal{O}^{r\times r}}\left\|\boldsymbol{U}\boldsymbol{O}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}\lesssim\left(\frac{\sigma^2}{\sigma_r^{\natural 2}}\sqrt{rn_1n_2}\log n+\kappa^{\natural}\frac{\sigma}{\sigma_r^{\natural}}\sqrt{rn_1\log n}\right)\sqrt{\frac{\mu^{\natural}r}{n_1}}\asymp\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^{\natural}r^2}{n_1}}$$

  under the noise condition $\sigma\sqrt{n_2}\ll\sigma_r^{\natural}/(\kappa^{\natural}\sqrt{r\log n})$ (in addition to a few other conditions omitted here). Note that when $\kappa^{\natural},\mu^{\natural},r\asymp 1$, their $\ell_{2,\infty}$ error bound resembles (6.20), but the condition $\sigma\sqrt{n_2}\ll\sigma_r^{\natural}/\sqrt{\log n}$ required therein is much stronger than the noise condition $\zeta_{\mathsf{op}}\ll\sigma_r^{\natural 2}$ — which is equivalent to $\sigma\sqrt[4]{n_1n_2}\ll\sigma_r^{\natural}/\sqrt{\log n}$ when $n_2\gtrsim n_1$ — imposed by our theory (see (6.17)). It is also worth emphasizing that the theory of Agterberg et al. (2021) is capable of accommodating dependent data (i.e. they only require the rows of $\boldsymbol{E}$ to be independent and allow dependence within rows), which is beyond the scope of the present paper.

- Compared with the $\ell_{2,\infty}$ estimation error guarantees for the diagonal-deleted spectral method in Cai et al. (2021, Theorem 1), our bound (6.20) is able to get rid of the bias term incurred by diagonal deletion (see Cai et al. (2021, Equation (17))), thus improving upon this prior result.

7 Discussion

In this paper, we have developed a suite of statistical inference procedures to construct confidence regions for PCA in the presence of missing data and heterogeneous corruption, which should be easy-to-use in practice due to their data-driven nature. The solution developed based on HeteroPCA is particularly appealing, as it enjoys a broadened applicability range without compromising statistical efficiency. The fine-grained distributional characterizations we have developed are non-asymptotic, which naturally lend themselves to high-dimensional settings. Moving forward, there are a variety of directions that are worthy of further investigation. For instance, how to conduct valid inference on individual principal components, particularly when the associated eigengap is vanishingly small? If the observed data are inherently biased with *a priori* unknown means, how to properly compensate for the bias? What if the noise components are inter-dependent, and what if the observed data samples are further corrupted by a non-negligible fraction of adversarial outliers? Uncertainty quantification in the face of heterogeneous missing patterns is also an important topic of practical value.

Acknowledgements

Y. Chen is supported in part by the AFOSR YIP award FA9550-19-1-0030, by the ONR grant N00014-19-1-2120, by the ARO grants W911NF-20-1-0097 and W911NF-18-1-0303, by the NSF grants CCF-1907661, IIS-1900140, IIS-2100158 and DMS-2014279. J. Fan is supported in part by the ONR grant N00014-19-1-2120 and the NSF grants DMS-1662139, DMS-1712591, DMS-2052926, DMS-2053832, and the NIH grant 2R01-GM072611-15.

A Additional notation and organization of the appendix

For any matrix $\boldsymbol{U}$ with orthonormal columns, we denote by $\mathcal{P}_{\boldsymbol{U}}(\boldsymbol{M}) \coloneqq \boldsymbol{U}\boldsymbol{U}^\top \boldsymbol{M}$ the Euclidean projection of a matrix $\boldsymbol{M}$ onto the column space of $\boldsymbol{U}$, and let $\mathcal{P}_{\boldsymbol{U}^\perp}(\boldsymbol{M}) = \boldsymbol{M} - \mathcal{P}_{\boldsymbol{U}}(\boldsymbol{M})$ denote the Euclidean projection of $\boldsymbol{M}$ onto the orthogonal complement of the column space of $\boldsymbol{U}$. For any matrix $\boldsymbol{B} \in \mathbb{R}^{n_1 \times n_2}$ and some $l \in [n_1]$, define $\mathcal{P}_{-l,\cdot}(\boldsymbol{B})$ to be the orthogonal projection of the matrix $\boldsymbol{B}$ onto the subspace of matrix that vanishes outside the l-th row, namely,

$$[\mathcal{P}_{-l,\cdot}(\boldsymbol{B})]_{i,j} = \begin{cases} B_{i,j}, & \text{if } i \neq l, \\ 0, & \text{otherwise,} \end{cases} \qquad \forall (i,j) \in [n_1] \times [n_2]. \tag{A.1}$$

For any point $x \in \mathbb{R}^d$ and any non-empty convex set $\mathcal{C} \in \mathscr{C}^d$ satisfying $\mathcal{C} \neq \mathbb{R}^d$, let us define the signed distance function as follows

$$\delta_{\mathcal{C}}(x) \coloneqq \begin{cases} -\mathsf{dist}\left(x, \mathbb{R}^d \setminus \mathcal{C}\right), & \text{if } x \in \mathcal{C}; \\ \mathsf{dist}\left(x, \mathcal{C}\right), & \text{if } x \notin \mathcal{C}. \end{cases} \tag{A.2}$$

Here, $\mathsf{dist}(x, \mathcal{A})$ is the Euclidean distance between a point $x \in \mathbb{R}^d$ and a non-empty set $\mathcal{A} \subseteq \mathbb{R}^d$. Also, for any $\varepsilon \in \mathbb{R}$, define

$$\mathcal{C}^\varepsilon \coloneqq \left\{x \in \mathbb{R}^d : \delta_{\mathcal{C}}(x) \leq \varepsilon\right\} \tag{A.3}$$

for any non-empty convex set $\mathcal{C} \in \mathscr{C}^d$ satisfying $\mathcal{C} \neq \mathbb{R}^d$, and define $\varnothing^\varepsilon = \varnothing$ and $(\mathbb{R}^d)^\varepsilon = \mathbb{R}^d$.

Organization of the appendix. The rest of the appendix is organized as follows. In Appendix B, we present Theorems 11-18, which are general versions of the theorems in the main text. Appendix C is devoted to proving Theorem 9. Appendices D and E establish our theoretical guarantees for the SVD-based approach (i.e., Theorems 11 and 12). Appendix F is concerned with proving Theorem 10. Appendices G and H are dedicated to establishing our theory for HeteroPCA (i.e., Theorems 15-18).

B A list of general theorems

For simplicity of presentation, the theorems presented in the main text (i.e., Section 3) concentrate on the scenario where $\kappa, \mu, r \asymp 1$. Note, however, that our theoretical framework can certainly allow these parameters to grow with the problem dimension. In this section, we provide a list of theorems accommodating more general scenarios; all subsequent proofs are dedicated to establishing these general theorems.

B.1 General theorems tailored to the SVD-based approach

Let us start with our results tailored to the SVD-based approach. The following theorems generalize Theorem 1, Theorem 2, Theorem 3 and Theorem 4, respectively.

Theorem 11. *Suppose that $p < 1 - \delta$ for some arbitrary constant $0 < \delta < 1$. In addition, suppose that Assumption 1 holds and that*

$$\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{n+d}{np}} \lesssim \frac{1}{\kappa^{1/2} r^{3/4} \log^{3/2}(n+d)} \wedge \frac{\sqrt{1 \wedge (d/n)}}{\kappa \mu^{1/2} r^{3/4} \log^{1/2}(n+d)},$$

$$np \gtrsim \kappa^4 \mu r^{5/2} \log^3(n+d), \quad dp \gg \kappa^2 \mu r \log^2(n+d), \quad n \gtrsim \kappa^5 r^2 \log^4(n+d), \quad d \gtrsim \kappa \mu r^{5/2} \log^2(n+d).$$

Let $\boldsymbol{R}$ be the $r \times r$ rotation matrix $\mathsf{sgn}(\boldsymbol{U}^\top \boldsymbol{U}^\star)$.

(a) Assume that $\boldsymbol{U}^\star$ is μ-incoherent and satisfies the following condition

$$\left\|\boldsymbol{U}_{l,\cdot}^\star\right\|_2 \gtrsim \left[\frac{\kappa^{3/2} \mu r^{3/4} \log^{5/2}(n+d)}{\sqrt{(n \wedge d)\, p}} + \frac{\kappa^2 \mu^{3/2} r^{5/4} \log^{3/2}(n+d)}{\sqrt{np} \wedge \sqrt{d^2 p/n}} + \frac{\kappa \mu r^{5/4} \log^{3/2}(n+d)}{\sqrt{d}}\right] \sqrt{\frac{r}{d}} \tag{B.1}$$

for some $1 \le l \le d$. *Then the estimate* $\boldsymbol{U}$ *returned by Algorithm 1 obeys*

$$\sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left(\left[\boldsymbol{U}\mathsf{sgn}\left(\boldsymbol{U}^\top\boldsymbol{U}^\star\right)-\boldsymbol{U}^\star\right]_{l,\cdot}\in\mathcal{C}\right)-\mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}_{U,l}^\star\right)\{\mathcal{C}\}\right|=o\left(1\right),$$

where $\mathscr{C}^r$ *is the the set of all convex sets in* $\mathbb{R}^r$ *and*

$$\boldsymbol{\Sigma}_{U,l}^\star \coloneqq \left(\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2+\frac{\omega_l^{\star 2}}{np}\right)\left(\boldsymbol{\Sigma}^\star\right)^{-2}+\frac{2\left(1-p\right)}{np}\left(\boldsymbol{U}_{l,\cdot}^\star\right)^\top\boldsymbol{U}_{l,\cdot}^\star.$$

(b) If in addition Condition (B.1) holds simultaneously for all $1 \le l \le d$, *then one further has*

$$\sup_{1\le l\le d}\sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left(\left[\boldsymbol{U}\mathsf{sgn}\left(\boldsymbol{U}^\top\boldsymbol{U}^\star\right)-\boldsymbol{U}^\star\right]_{l,\cdot}\in\mathcal{C}\right)-\mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}_{U,l}^\star\right)\{\mathcal{C}\}\right|=o\left(1\right).$$

Theorem 12. *Instate the assumptions in Theorem 11(a). Additionally, assume that*

$$\left(n\wedge d\right)p\gtrsim\kappa^{7/2}\mu r^{7/4}\log^{5/2}\left(n+d\right),\qquad np\gtrsim\kappa^5 r^{5/2}\log^5\left(n+d\right),$$

and

$$\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{n+d}{np}}\lesssim\frac{1}{\kappa r^{3/8}\log^{5/4}\left(n+d\right)}.$$

Then the confidence region $\mathsf{CR}_{U,l}^{1-\alpha}$ *computed in Algorithm 3 obeys*

$$\mathbb{P}\left(\boldsymbol{U}_{l,\cdot}^\star\mathsf{sgn}\left(\boldsymbol{U}^{\star\top}\boldsymbol{U}\right)\in\mathsf{CR}_{U,l}^{1-\alpha}\right)=1-\alpha+o\left(1\right).$$

Theorem 13. *Suppose that* $p < 1-\delta$ *for some arbitrary constant* $0<\delta<1$. *Consider any* $1 \le i,j \le d$. *Assume that* $\boldsymbol{U}^\star$ *is* μ*-incoherent and satisfies the following conditions*

$$\min\left\{\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2,\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right\}\gtrsim\left[\frac{\kappa^2\mu r\log^{5/2}\left(n+d\right)}{\sqrt{\left(n\wedge d\right)p}}+\frac{\kappa^{5/2}\mu^{3/2}r\log^{3/2}\left(n+d\right)}{\sqrt{np}\wedge\sqrt{d^2p/n}}+\frac{\kappa^{3/2}\mu r\log^{3/2}\left(n+d\right)}{\sqrt{d}}\right]\sqrt{\frac{r}{d}},$$

and

$$\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\gtrsim\kappa r^{1/2}\log^{5/2}\left(n+d\right)\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d}{np}}\cdot\sqrt{\frac{r}{d}}.$$

In addition, suppose that Assumption 1 holds and that $d\gtrsim\kappa^2\mu r^2\log^2(n+d)$,

$$np\gtrsim\kappa^2r^2\log^5\left(n+d\right),\qquad\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{n+d}{np}}\lesssim\frac{1}{\sqrt{\kappa^2r\log^2\left(n+d\right)}}\wedge\sqrt{\frac{1\wedge(d/n)}{\kappa^3\mu r\log\left(n+d\right)}},$$

Then the matrix $\boldsymbol{S}$ *computed by Algorithm 1 obeys*

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\frac{S_{i,j}-S_{i,j}^\star}{\sqrt{v_{i,j}^\star}}\le t\right)-\Phi\left(t\right)\right|=o\left(1\right).$$

Theorem 14. *Instate the conditions in Theorem 13. Further, assume that* $n\gtrsim\kappa^3 r\log^3(n+d)$. *Then the confidence interval computed in Algorithm 4 obeys*

$$\mathbb{P}\left(S_{i,j}^\star\in\mathsf{CI}_{i,j}^{1-\alpha}\right)=1-\alpha+o\left(1\right).$$

B.2 General theorems tailored to HeteroPCA

We now turn attention to the results tailored to the HeteroPCA algorithm. The following theorems generalize Theorem 5, Theorem 6, Theorem 7 and Theorem 8, respectively.

Theorem 15. *Suppose that $p < 1 - \delta$ for some arbitrary constant $0 < \delta < 1$. In addition, suppose that Assumption 1 holds and that $n \gtrsim \kappa^8 \mu^2 r^4 \log^4(n+d)$, $d \gtrsim \kappa^6 \mu^2 r^{5/2} \log^5(n+d)$,*

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\kappa^4 \mu^{3/2} r^{11/4} \log^{5/2}(n+d)}, \qquad \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\kappa^3 \mu^{1/2} r^{3/4} \log^2(n+d)},$$

$$ndp^2 \gtrsim \kappa^8 \mu^4 r^{13/2} \log^6(n+d), \qquad np \gtrsim \kappa^8 \mu^3 r^{11/2} \log^5(n+d).$$

(a) Assume that $\boldsymbol{U}^\star$ is μ-incoherent and satisfies the following condition

$$\left\|\boldsymbol{U}_{l,\cdot}^\star\right\|_2 \gtrsim \left[\frac{\kappa^{9/2}\mu^{5/2} r^{9/4} \log^5(n+d)}{\sqrt{ndp^2}} + \frac{\kappa^4 \mu^{3/2} r^{5/4} \log^3(n+d)}{\sqrt{np}}\sigma_1^\star + \frac{\kappa^4 \mu^2 r^{7/4} \log^{7/2}(n+d)}{\sqrt{d}}\right]\sqrt{\frac{r}{d}}. \tag{B.3}$$

Then the estimate $\boldsymbol{U}$ returned by Algorithm 2 with number of iterations satisfying (3.16) satisfies

$$\sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left(\left[\boldsymbol{U}\mathsf{sgn}\left(\boldsymbol{U}^\top\boldsymbol{U}^\star\right) - \boldsymbol{U}^\star\right]_{l,\cdot} \in \mathcal{C}\right) - \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{U,l}^\star\right)\{\mathcal{C}\}\right| = o(1),$$

where $\mathscr{C}^r$ represents the the set of all convex sets in $\mathbb{R}^r$, and

$$\boldsymbol{\Sigma}_{U,l}^\star := \left(\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2 + \frac{\omega_l^{\star 2}}{np}\right)(\boldsymbol{\Sigma}^\star)^{-2} + \frac{2(1-p)}{np}\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^\star + (\boldsymbol{\Sigma}^\star)^{-2}\boldsymbol{U}^{\star\top}\mathsf{diag}\left\{\left[d_{l,i}^\star\right]_{1\le i\le d}\right\}\boldsymbol{U}^\star(\boldsymbol{\Sigma}^\star)^{-2}$$

where

$$d_{l,i}^\star := \frac{1}{np^2}\left[\omega_l^{\star 2} + (1-p)\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\right]\left[\omega_i^{\star 2} + (1-p)\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\right] + \frac{2(1-p)^2}{np^2}S_{l,i}^{\star 2}.$$

(b) Further, if in addition (B.3) holds simultaneously for all $1 \le l \le d$, then one has

$$\sup_{1\le l\le d}\sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left(\left[\boldsymbol{U}\mathsf{sgn}\left(\boldsymbol{U}^\top\boldsymbol{U}^\star\right) - \boldsymbol{U}^\star\right]_{l,\cdot} \in \mathcal{C}\right) - \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{U,l}^\star\right)\{\mathcal{C}\}\right| = o(1).$$

Theorem 16. *Instate the conditions in Theorem 15. Further assume that $n \gtrsim \kappa^{12}\mu^3 r^{11/2}\log^5(n+d)$,*

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\kappa^{9/2}\mu^{3/2} r^{9/4} \log^{7/2}(n+d)}, \qquad \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\kappa^5 \mu^{3/2} r^{9/4} \log^3(n+d)}.$$

$$ndp^2 \gtrsim \kappa^{11}\mu^5 r^{13/2}\log^9(n+d), \qquad np \gtrsim \kappa^{11}\mu^4 r^{11/2}\log^7(n+d),$$

Then the confidence region $\mathsf{CR}_{U,l}^{1-\alpha}$ computed in Algorithm 5 obeys

$$\mathbb{P}\left(\boldsymbol{U}_{l,\cdot}^\star\mathsf{sgn}\left(\boldsymbol{U}^{\star\top}\boldsymbol{U}\right) \in \mathsf{CR}_{U,l}^{1-\alpha}\right) = 1 - \alpha + o(1).$$

Theorem 17. *Suppose that $p < 1 - \delta$ for some arbitrary constant $0 < \delta < 1$. Consider any $1 \le i, j \le d$. Assume that $\boldsymbol{U}^\star$ is μ-incoherent and satisfies the following conditions*

$$\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2 \gtrsim \kappa r^{1/2}\log^{5/2}(n+d)\left(\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d}{np}} + \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}\right)\sqrt{\frac{r}{d}}$$

and

$$\min\left\{\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2, \left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right\} \gtrsim \left[\frac{\kappa^5\mu^{5/2}r^2\log^5(n+d)}{\sqrt{ndp^2}} + \frac{\kappa^{9/2}\mu^{3/2} r\log^3(n+d)}{\sqrt{np}} + \frac{\kappa^{9/2}\mu^2 r^{3/2}\log^{7/2}(n+d)}{\sqrt{d}}\right]\sqrt{\frac{r}{d}}.$$

In addition, suppose that Assumption 1 holds and that $n \gtrsim r\log^4(n+d)$, $d \gtrsim \kappa^7\mu^2 r^2\log^5(n+d)$,

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\kappa^3\mu^{1/2} r\log^3(n+d)}, \qquad \frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\kappa^{7/2}\mu^{1/2}r^{1/2}\log^2(n+d)},$$

$$ndp^2 \gtrsim \kappa^5\mu^2 r^4\log^7(n+d), \qquad np \gtrsim \kappa^4\mu r^2\log^5(n+d),$$

Then the matrix $\boldsymbol{S}$ computed by Algorithm 2 with number of iterations satisfying (3.16) obeys

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\frac{S_{i,j}-S_{i,j}^{\star}}{\sqrt{v_{i,j}^{\star}}}\leq t\right)-\Phi(t)\right| = o(1).$$

Theorem 18. *Instate the conditions in Theorem 17. Further assume that $n \gtrsim \kappa^9\mu^3 r^4\log^4(n+d)$,*

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\kappa^3\mu^{3/2} r^{5/2}\log^3(n+d)}, \qquad \frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\kappa^{7/2}\mu^{3/2}r^{5/2}\log^{5/2}(n+d)},$$

$$ndp^2 \gtrsim \kappa^8\mu^5 r^7\log^8(n+d), \qquad np \gtrsim \kappa^8\mu^4 r^6\log^6(n+d),$$

and

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \gtrsim \kappa^2\mu^{3/2}r^{5/2}\log^{5/2}(n+d)\left[\frac{\kappa\mu r\log(n+d)}{\sqrt{ndp}} + \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} + \frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}}\right]\sqrt{\frac{r}{d}}.$$

Then the confidence interval computed in Algorithm 6 obeys

$$\mathbb{P}\left(S_{i,j}^{\star}\in\mathsf{CI}_{i,j}^{1-\alpha}\right) = 1-\alpha+o(1).$$

C Analysis for matrix denoising (Theorem 9)

Before embarking on the proof of Theorem 9, let us define two matrices

$$\boldsymbol{H_U} \coloneqq \boldsymbol{U}^{\top}\boldsymbol{U}^{\natural} \qquad \text{and} \qquad \boldsymbol{H_V} \coloneqq \boldsymbol{V}^{\top}\boldsymbol{V}^{\natural},$$

which are closely related to the rotation matrices $\boldsymbol{R_U}$ and $\boldsymbol{R_V}$ but are often easier to work with.

C.1 A few key lemmas

We first collect a couple of facts that are useful for the proof. The first lemma below quantifies the spectral norm of the noise matrix, the perturbation bound for singular subspaces w.r.t. the spectral norm, as well as some basic facts about $\boldsymbol{H_U}$ and $\boldsymbol{H_V}$.

Lemma 1. *Assume that $\sigma\sqrt{n} \ll \sigma_r^{\natural}$. Then with probability exceeding $1-O(n^{-10})$, we have*

$$\|\boldsymbol{E}\| \lesssim \sigma\sqrt{n},$$

$$\max\left\{\left\|\boldsymbol{U}\boldsymbol{R_U}-\boldsymbol{U}^{\natural}\right\|,\left\|\boldsymbol{V}\boldsymbol{R_V}-\boldsymbol{V}^{\natural}\right\|\right\} \lesssim \frac{\sigma}{\sigma_r^{\natural}}\sqrt{n},$$

$$\max\left\{\left\|\boldsymbol{H_U}-\boldsymbol{R_U}\right\|,\left\|\boldsymbol{H_V}-\boldsymbol{R_V}\right\|\right\} \lesssim \frac{\sigma^2 n}{\sigma_r^{\natural 2}}, \tag{C.1}$$

and

$$\frac{1}{2}\leq\sigma_i\left(\boldsymbol{H_U}\right)\leq 2, \qquad \frac{1}{2}\leq\sigma_i\left(\boldsymbol{H_U}\right)\leq 2, \qquad 1\leq i\leq r.$$

Proof. This lemma follows from standard matrix tail bounds and the celebrated Wedin $\sin\mathbf{\Theta}$ theorem (Wedin, 1972). See Appendix C.3.1. $\square$

The next lemma shows that $\mathbf{\Sigma}$ and $\mathbf{\Sigma}^{\natural}$ remain close even after $\mathbf{\Sigma}$ is properly rotated (resp. modulated) by $\boldsymbol{R}_{\boldsymbol{U}}$ and $\boldsymbol{R}_{\boldsymbol{V}}$ (resp. $\boldsymbol{H}_{\boldsymbol{U}}$ and $\boldsymbol{H}_{\boldsymbol{V}}$).

Lemma 2. *Assume that $\sigma\sqrt{n\log n}\ll\sigma_r^{\natural}$ and $n_2\gtrsim\mu^{\natural}r$. Then with probability exceeding $1-O(n^{-10})$ we have*

$$\left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top}\mathbf{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}-\mathbf{\Sigma}^{\natural}\right\|\lesssim\kappa^{\natural}\frac{\sigma^2 n}{\sigma_r^{\natural}}+\sigma\sqrt{r\log n},$$

$$\left\|\boldsymbol{H}_{\boldsymbol{U}}^{\top}\mathbf{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\mathbf{\Sigma}^{\natural}\right\|\lesssim\frac{(\sigma\sqrt{n})^3}{\sigma_r^{\natural 2}}+\sigma\sqrt{r\log n}.$$

Proof. See Appendix C.3.2. $\square$

We now turn to the most crucial technical lemma, which establishes the intertwined connection between the $\ell_{2,\infty}$ norms of several matrices — all of these matrices reveal some sort of difference between the estimates and the ground truth. This lemma is the most technically involved in this subsection, which is established via the powerful leave-one-out analysis framework (see Chen et al. (2020c, Chapter 4) for an accessible introduction).

Lemma 3. *Assume that $\sigma\sqrt{n\log n}\ll\sigma_r^{\natural}$. Then with probability exceeding $1-O(n^{-10})$ we have*

$$\begin{aligned}\left\|\boldsymbol{U}\mathbf{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{M}\boldsymbol{V}^{\natural}\right\|_{2,\infty}&\lesssim\frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{r\log n}+\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}\\&\quad+(B\log n)\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right\|_{2,\infty}+\frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{n}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^{\natural}\right\|_{2,\infty};\\\left\|\boldsymbol{V}\mathbf{\Sigma}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{M}^{\top}\boldsymbol{U}^{\natural}\right\|_{2,\infty}&\lesssim\frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{r\log n}+\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural}}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty}\\&\quad+(B\log n)\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{n}\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right\|_{2,\infty}.\end{aligned}$$

Proof. See Appendix C.3.3. $\square$

With the above two lemmas in place, we can readily translate them to $\ell_{2,\infty}$ estimation guarantees (although the rotation matrices are replaced by $\boldsymbol{H}_{\boldsymbol{U}}$ and $\boldsymbol{H}_{\boldsymbol{V}}$), as stated below.

Lemma 4. *Assume that $\sigma\sqrt{n\log n}\ll\sigma_r^{\natural}/\kappa^{\natural}$ and $n_1\geq\mu^{\natural}r$. Then with probability exceeding $1-O(n^{-10})$ we have*

$$\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}\lesssim\frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n}+\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty},$$

$$\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right\|_{2,\infty}\lesssim\frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n}+\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty}.$$

Proof. See Appendix C.3.4. $\square$

C.2 Proof of Theorem 9

We are now well equipped to establish Theorem 9. It is worth noting that, while Lemma 4 delivers $\ell_{2,\infty}$ perturbation bounds for singular subspaces, it falls short of revealing the relation between the perturbation error and the desired first-order approximation $\boldsymbol{E}\boldsymbol{V}^{\natural}(\mathbf{\Sigma}^{\natural})^{-1}$ and $\boldsymbol{E}^{\top}\boldsymbol{V}^{\natural}(\mathbf{\Sigma}^{\natural})^{-1}$. In order to make explicit the role of the first-order approximation, we first make the observation that

$$\boldsymbol{U}^{\natural}+\boldsymbol{E}\boldsymbol{V}^{\natural}(\mathbf{\Sigma}^{\natural})^{-1}=\boldsymbol{U}^{\natural}-\boldsymbol{M}^{\natural}\boldsymbol{V}^{\natural}(\mathbf{\Sigma}^{\natural})^{-1}+\boldsymbol{M}\boldsymbol{V}^{\natural}(\mathbf{\Sigma}^{\natural})^{-1}=\boldsymbol{M}\boldsymbol{V}^{\natural}(\mathbf{\Sigma}^{\natural})^{-1},\tag{C.2}$$

which makes use of the decomposition $\boldsymbol{M}^{\natural}=\boldsymbol{U}^{\natural}\mathbf{\Sigma}^{\natural}\boldsymbol{V}^{\natural\top}$. This motivates us to pay attention to the matrix $\boldsymbol{M}\boldsymbol{V}^{\natural}(\mathbf{\Sigma}^{\natural})^{-1}$. In what follows, our proof consists of the following steps:

1. Establish the proximity of $\boldsymbol{M}\boldsymbol{V}^{\natural}$ and $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}$;
2. Establish the proximity of $\boldsymbol{M}\boldsymbol{V}^{\natural}$ and $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}$ by exploiting the closeness between $\boldsymbol{H}_{\boldsymbol{V}}$ and $\boldsymbol{R}_{\boldsymbol{V}}$;
3. Replacing $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}$ with $\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural}$, so that we can control the target difference between $\boldsymbol{M}\boldsymbol{V}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}$ and $\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}$.

C.2.1 Step 1: establishing the proximity of $\boldsymbol{M}\boldsymbol{V}^{\natural}$ and $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}$

Combine Lemma 4 with Lemma 3 to obtain

$$\begin{aligned}\left\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{M}\boldsymbol{V}^{\natural}\right\|_{2,\infty}&\lesssim\frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{r\log n}+\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}+(B\log n)\left\{\frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n}+\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty}\right\}\\&\quad+\frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{n}\left\{\frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n}+\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}\right\}\\&\asymp\frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{r\log n}+\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty},\end{aligned}\tag{C.3}$$

provided that $\sigma\sqrt{n\log n}\ll\sigma_r^{\natural}/\kappa^{\natural}$ and $B\lesssim\sigma\sqrt{n_1/(\mu^{\natural}\log n)}$, where we have used the fact that

$$(B\log n)\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty}\leq(B\log n)\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^{\natural}r}{n_2}}\lesssim\frac{\sigma\sqrt{n}}{\sigma_r^{\natural}}\sigma\sqrt{r\log n}.$$

C.2.2 Step 2: establishing the proximity of $\boldsymbol{M}\boldsymbol{V}^{\natural}$ and $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}$

Taking the result (C.3) together with Lemma 1 allows us to replace $\boldsymbol{H}_{\boldsymbol{V}}$ with $\boldsymbol{R}_{\boldsymbol{V}}$ as follows

$$\begin{aligned}\left\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}-\boldsymbol{M}\boldsymbol{V}^{\natural}\right\|_{2,\infty}&\leq\left\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{M}\boldsymbol{V}^{\natural}\right\|_{2,\infty}+\left\|\boldsymbol{U}\boldsymbol{\Sigma}\left(\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{R}_{\boldsymbol{V}}\right)\right\|_{2,\infty}\\&\leq\left\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{M}\boldsymbol{V}^{\natural}\right\|_{2,\infty}+\left\|\boldsymbol{U}\right\|_{2,\infty}\left\|\boldsymbol{\Sigma}\right\|\left\|\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{R}_{\boldsymbol{V}}\right\|\\&\overset{\text{(i)}}{\lesssim}\frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{r\log n}+\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\left(\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n}\right)\sigma_1^{\natural}\frac{\sigma^2 n}{\sigma_r^{\natural 2}}\\&\overset{\text{(ii)}}{\lesssim}\frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{r\log n}+\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}.\end{aligned}\tag{C.4}$$

Here, (i) invokes (C.3), Lemma 3, Lemma 1, Lemma 2, Lemma 4 as well as their direct consequences

$$\begin{aligned}\left\|\boldsymbol{U}\right\|_{2,\infty}\leq 2\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}\right\|_{2,\infty}&\lesssim\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}\lesssim\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n}+\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}\\&\lesssim\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n},\end{aligned}\tag{C.5a}$$

$$\left\|\boldsymbol{\Sigma}\right\|=\left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}\right\|\leq\left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}-\boldsymbol{\Sigma}^{\natural}\right\|+\left\|\boldsymbol{\Sigma}^{\natural}\right\|\leq\kappa^{\natural}\frac{\sigma^2 n}{\sigma_r^{\natural}}+\sigma\sqrt{r\log n}+\sigma_1^{\natural}\lesssim\sigma_1^{\natural},\tag{C.5b}$$

provided that $\sigma\sqrt{n}\lesssim\sigma_r^{\natural}/\sqrt{\kappa^{\natural}}$ and $\sigma\sqrt{n\log n}\ll\sigma_r^{\natural}$; (ii) is valid provided that $\sigma\sqrt{n}\ll\sigma_r^{\natural}$. In summary, from (C.4) and Lemma 2, we can write

$$\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}=\boldsymbol{M}\boldsymbol{V}^{\natural}+\boldsymbol{\Delta}_{\boldsymbol{U}}\qquad\text{and}\qquad\boldsymbol{\Sigma}^{\natural}=\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}+\boldsymbol{\Delta}_{\boldsymbol{\Sigma}}$$

where

$$\left\|\boldsymbol{\Delta}_{\boldsymbol{U}}\right\|_{2,\infty}\lesssim\frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{r\log n}+\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty},\tag{C.6a}$$

$$\left\|\boldsymbol{\Delta}_{\boldsymbol{\Sigma}}\right\|\lesssim\kappa^{\natural}\frac{\sigma^2 n}{\sigma_r^{\natural}}+\sigma\sqrt{r\log n}.\tag{C.6b}$$

C.2.3 Step 3: replacing $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}$ with $\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural}$

The preceding bounds in turn allow one to derive

$$\begin{aligned}\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural} &= \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}} + \boldsymbol{U}\left(\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural} - \boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}\right)\\ &= \boldsymbol{M}\boldsymbol{V}^{\natural} + \boldsymbol{\Delta}_{\boldsymbol{U}} + \boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\left(\boldsymbol{\Sigma}^{\natural} - \boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}\right)\\ &= \left(\boldsymbol{M}^{\natural} + \boldsymbol{E}\right)\boldsymbol{V}^{\natural} + \boldsymbol{\Delta}_{\boldsymbol{U}} + \boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{\Delta}_{\boldsymbol{\Sigma}}\\ &= \boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural} + \boldsymbol{E}\boldsymbol{V}^{\natural} + \boldsymbol{\Delta}_{\boldsymbol{U}} + \boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{\Delta}_{\boldsymbol{\Sigma}},\end{aligned}$$

or equivalently,

$$\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} = \boldsymbol{U}^{\natural} + \boldsymbol{E}\boldsymbol{V}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-1} + \underbrace{\boldsymbol{\Delta}_{\boldsymbol{U}}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-1} + \boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{\Delta}_{\boldsymbol{\Sigma}}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-1}}_{=:\boldsymbol{\Psi}_{\boldsymbol{U}}}. \tag{C.7}$$

The first two terms on the right-hand side of (C.7) coincide with the desired first-order approximation. Therefore, it suffices to control the size of $\boldsymbol{\Psi}_{\boldsymbol{U}}$. Towards this, invoke (C.6) to deduce that

$$\begin{aligned}\left\|\boldsymbol{\Psi}_{\boldsymbol{U}}\right\|_{2,\infty} &\leq \frac{1}{\sigma_r^{\natural}}\left\|\boldsymbol{\Delta}_{\boldsymbol{U}}\right\|_{2,\infty} + \frac{1}{\sigma_r^{\natural}}\left\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\right\|_{2,\infty}\left\|\boldsymbol{\Delta}_{\boldsymbol{\Sigma}}\right\|\\ &\leq \frac{1}{\sigma_r^{\natural}}\left\|\boldsymbol{\Delta}_{\boldsymbol{U}}\right\|_{2,\infty} + \frac{1}{\sigma_r^{\natural}}\left\|\boldsymbol{U}\right\|_{2,\infty}\left\|\boldsymbol{\Delta}_{\boldsymbol{\Sigma}}\right\|\\ &\lesssim \frac{\sigma\sqrt{n} + B\log n}{\sigma_r^{\natural 2}}\sigma\sqrt{r\log n} + \frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} + \frac{1}{\sigma_r^{\natural}}\left(\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} + \frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n}\right)\left(\kappa^{\natural}\frac{\sigma^2 n}{\sigma_r^{\natural}} + \sigma\sqrt{r\log n}\right)\\ &\lesssim \frac{\sigma^2\sqrt{nr}\log n}{\sigma_r^{\natural 2}} + \left(\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}} + \frac{\sigma\sqrt{r\log n}}{\sigma_r^{\natural}}\right)\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty},\end{aligned}$$

where the penultimate line relies on (C.6) and (C.5), and the last line is valid provided that $\sigma\sqrt{n} \ll \sigma_r^{\natural}/\kappa^{\natural}$ and $B \lesssim \sigma\sqrt{n_1/\left(\mu^{\natural}\log n\right)}$.

The advertised bound regarding $\boldsymbol{V}\boldsymbol{R}_{\boldsymbol{V}}$ follows from an analogous proof argument and is hence omitted.

C.3 Proof of auxiliary lemmas

Before we start proving the lemmas in Appendix C.1, we first introduce a useful fact that will be useful throughout; the proof is deferred to Appendix C.3.5.

Lemma 5. *For any fixed matrix $\boldsymbol{A}$, with probability exceeding $1 - O(n^{-10})$ we have*

$$\left\|\boldsymbol{E}\boldsymbol{A}\right\|_{2,\infty} \lesssim \sigma\left\|\boldsymbol{A}\right\|_{\mathrm{F}}\sqrt{\log n} + B\left\|\boldsymbol{A}\right\|_{2,\infty}\log n.$$

C.3.1 Proof of Lemma 1

From standard matrix tail bounds (e.g., Chen et al. (2020c, Theorem 3.1.4)), we know that

$$\|\boldsymbol{E}\| \lesssim \sigma\sqrt{n}$$

holds with probability exceeding $1 - O(n^{-10})$. Applying Weyl's inequality then yields

$$\sigma_r\left(\boldsymbol{M}\right) \geq \sigma_r^{\natural} - \|\boldsymbol{E}\| \geq \frac{1}{2}\sigma_r^{\natural}$$

as long as $\sigma\sqrt{n} \ll \sigma_r^{\natural}$. In view of Wedin's $\sin\boldsymbol{\Theta}$ Theorem (Chen et al., 2020c, Theorem 2.3.1), we obtain

$$\max\left\{\left\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^{\natural}\right\|, \left\|\boldsymbol{V}\boldsymbol{R}_{\boldsymbol{V}} - \boldsymbol{V}^{\natural}\right\|\right\} \leq \frac{\sqrt{2}\,\|\boldsymbol{E}\|}{\sigma_r\left(\boldsymbol{M}\right) - \sigma_{r+1}\left(\boldsymbol{M}^{\natural}\right)} \leq \frac{2\sqrt{2}\,\|\boldsymbol{E}\|}{\sigma_r^{\natural}} \lesssim \frac{\sigma\sqrt{n}}{\sigma_r^{\natural}}. \tag{C.8}$$

Next, we turn attention to $\left\|\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{R}_{\boldsymbol{U}}\right\|$. Given that both $\boldsymbol{U}$ and $\boldsymbol{U}^{\natural}$ are orthonormal, the SVD of $\boldsymbol{H}_{\boldsymbol{U}}=\boldsymbol{U}^{\top}\boldsymbol{U}^{\natural}$ can be written as

$$\boldsymbol{H}_{\boldsymbol{U}}=\boldsymbol{X}(\cos\boldsymbol{\Theta})\boldsymbol{Y}^{\top},$$

where $\boldsymbol{X},\boldsymbol{Y}\in\mathbb{R}^{r\times r}$ are orthonormal matrices and $\boldsymbol{\Theta}$ is a diagonal matrix composed of the principal angles between $\boldsymbol{U}$ and $\boldsymbol{U}^{\natural}$ (see Chen et al. (2020c, Section 2.1)). It is easily seen that one can write $\boldsymbol{R}_{\boldsymbol{U}}=\mathsf{sgn}(\boldsymbol{H})=\boldsymbol{X}\boldsymbol{Y}^{\top}$, and therefore,

$$\begin{aligned}\left\|\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{R}_{\boldsymbol{U}}\right\|&=\left\|\boldsymbol{H}-\mathsf{sgn}\left(\boldsymbol{H}\right)\right\|=\left\|\boldsymbol{X}\left(\cos\boldsymbol{\Theta}-\boldsymbol{I}_r\right)\boldsymbol{Y}^{\top}\right\|=\left\|\boldsymbol{I}_r-\cos\boldsymbol{\Theta}\right\|\\&=\left\|2\sin^2\left(\boldsymbol{\Theta}/2\right)\right\|\lesssim\left\|\sin\boldsymbol{\Theta}\right\|^2.\end{aligned}$$

Invoke Wedin's $\sin\boldsymbol{\Theta}$ Theorem and repeat the argument in (C.8) to yield

$$\left\|\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{R}_{\boldsymbol{U}}\right\|\lesssim\left\|\sin\boldsymbol{\Theta}\right\|^2\lesssim\frac{\sigma^2 n}{\sigma_r^{\natural 2}}.$$

Given that $\boldsymbol{R}_{\boldsymbol{U}}$ is a square orthonormal matrix, we immediately have

$$\begin{aligned}\sigma_{\max}(\boldsymbol{H}_{\boldsymbol{U}})&\le\sigma_{\max}(\boldsymbol{R}_{\boldsymbol{U}})+\left\|\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{R}_{\boldsymbol{U}}\right\|\le 1+O\Big(\frac{\sigma^2 n}{\sigma_r^{\natural 2}}\Big)\le 2,\\ \sigma_r(\boldsymbol{H}_{\boldsymbol{U}})&\ge\sigma_{\max}(\boldsymbol{R}_{\boldsymbol{U}})-\left\|\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{R}_{\boldsymbol{U}}\right\|\ge 1-O\Big(\frac{\sigma^2 n}{\sigma_r^{\natural 2}}\Big)\ge 1/2,\end{aligned}$$

provided that $\sigma^2 n\ll\sigma_r^{\natural 2}$. The claimed bounds on $\boldsymbol{H}_{\boldsymbol{V}}$ can be derived analogously.

C.3.2 Proof of Lemma 2

First, invoke the triangle inequality to decompose

$$\left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}-\boldsymbol{\Sigma}^{\natural}\right\|\le\underbrace{\left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}-\boldsymbol{H}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}\right\|}_{=:\,\alpha_1}+\underbrace{\left\|\boldsymbol{H}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{U}^{\natural\top}\boldsymbol{M}\boldsymbol{V}^{\natural}\right\|}_{=:\,\alpha_2}+\underbrace{\left\|\boldsymbol{U}^{\natural\top}\boldsymbol{M}\boldsymbol{V}^{\natural}-\boldsymbol{\Sigma}^{\natural}\right\|}_{=:\,\alpha_3},$$

leaving us with three terms to cope with.

- Regarding α_1, the triangle inequality gives

$$\begin{aligned}\alpha_1&\le\left\|\left(\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{R}_{\boldsymbol{U}}\right)^{\top}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}\right\|+\left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\left(\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{R}_{\boldsymbol{V}}\right)\right\|\\&\le\left\|\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{R}_{\boldsymbol{U}}\right\|\left\|\boldsymbol{\Sigma}\right\|\left\|\boldsymbol{H}_{\boldsymbol{V}}\right\|+\left\|\boldsymbol{R}_{\boldsymbol{U}}\right\|\left\|\boldsymbol{\Sigma}\right\|\left\|\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{R}_{\boldsymbol{V}}\right\|\\&\lesssim\frac{\sigma^2 n}{\sigma_r^{\natural 2}}\sigma_1^{\natural}\lesssim\kappa^{\natural}\frac{\sigma^2 n}{\sigma_r^{\natural}}.\end{aligned}$$

  Here, the last line makes use of (C.1) in Lemma 1 and its direct consequence

$$\max\left\{\left\|\boldsymbol{H}_{\boldsymbol{U}}\right\|,\left\|\boldsymbol{H}_{\boldsymbol{V}}\right\|\right\}\le 2,\tag{C.9}$$

  as well as an application of Weyl's inequality (assuming $\sigma\sqrt{n}\ll\sigma_r^{\natural}$):

$$\left\|\boldsymbol{\Sigma}\right\|\le\left\|\boldsymbol{\Sigma}^{\natural}\right\|+\left\|\boldsymbol{E}\right\|\lesssim\sigma_1^{\natural}+\sigma\sqrt{n}\lesssim\sigma_1^{\natural}.$$

- Regarding α_2, it is easily seen that

$$\boldsymbol{H}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{U}^{\natural\top}\boldsymbol{M}\boldsymbol{V}^{\natural}=\boldsymbol{U}^{\natural\top}\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\top}\boldsymbol{V}^{\natural}-\boldsymbol{U}^{\natural\top}\boldsymbol{M}\boldsymbol{V}^{\natural}=\boldsymbol{U}^{\natural\top}\boldsymbol{U}_{\perp}\boldsymbol{\Sigma}_{\perp}\boldsymbol{V}_{\perp}^{\top}\boldsymbol{V}^{\natural},$$

  where we denote the full SVD of $\boldsymbol{M}$ as

$$\boldsymbol{M}=\left[\begin{array}{cc}\boldsymbol{U}&\boldsymbol{U}_{\perp}\end{array}\right]\left[\begin{array}{cc}\boldsymbol{\Sigma}&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{\Sigma}_{\perp}\end{array}\right]\left[\begin{array}{c}\boldsymbol{V}^{\top}\\\boldsymbol{V}_{\perp}^{\top}\end{array}\right]=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\top}+\boldsymbol{U}_{\perp}\boldsymbol{\Sigma}_{\perp}\boldsymbol{V}_{\perp}^{\top}.$$

In view of Weyl's inequality and Lemma 1, we have

$$\|\boldsymbol{\Sigma}_{\perp}\| \leq \sigma_{r+1}\left(\boldsymbol{M}^{\natural}\right) + \|\boldsymbol{E}\| = \|\boldsymbol{E}\| \lesssim \sigma\sqrt{n}.$$

In addition, it follows from Chen et al. (2020c, Lemmas 2.1.2-2.1.3) and Lemma 1 that

$$\begin{aligned}
\left\|\boldsymbol{U}^{\natural\top}\boldsymbol{U}_{\perp}\right\| &\lesssim \left\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^{\natural}\right\| \lesssim \frac{\sigma\sqrt{n}}{\sigma_r^{\natural}}, \\
\left\|\boldsymbol{V}^{\natural\top}\boldsymbol{V}_{\perp}\right\| &\lesssim \left\|\boldsymbol{V}\boldsymbol{R}_{\boldsymbol{V}} - \boldsymbol{V}^{\natural}\right\| \lesssim \frac{\sigma\sqrt{n}}{\sigma_r^{\natural}}.
\end{aligned}$$

Putting the above bounds together, we arrive at

$$\alpha_2 = \left\|\boldsymbol{U}^{\natural\top}\boldsymbol{U}_{\perp}\boldsymbol{\Sigma}_{\perp}\boldsymbol{V}_{\perp}^{\top}\boldsymbol{V}^{\natural}\right\| \leq \left\|\boldsymbol{U}^{\natural\top}\boldsymbol{U}_{\perp}\right\|\left\|\boldsymbol{\Sigma}_{\perp}\right\|\left\|\boldsymbol{V}^{\natural\top}\boldsymbol{V}_{\perp}\right\| \lesssim \frac{\left(\sigma\sqrt{n}\right)^3}{\sigma_r^{\natural 2}}.$$

- It remains to control α_3. Towards this end, we make the observation that

$$\boldsymbol{U}^{\natural\top}\boldsymbol{M}\boldsymbol{V}^{\natural} - \boldsymbol{\Sigma}^{\natural} = \boldsymbol{U}^{\natural\top}\boldsymbol{M}^{\natural}\boldsymbol{V}^{\natural} + \boldsymbol{U}^{\natural\top}\boldsymbol{E}\boldsymbol{V}^{\natural} - \boldsymbol{\Sigma}^{\natural} = \boldsymbol{\Sigma}^{\natural} + \boldsymbol{U}^{\natural\top}\boldsymbol{E}\boldsymbol{V}^{\natural} - \boldsymbol{\Sigma}^{\natural} = \boldsymbol{U}^{\natural\top}\boldsymbol{E}\boldsymbol{V}^{\natural},$$

which motivates us to control the spectral norm of $\boldsymbol{U}^{\natural\top}\boldsymbol{E}\boldsymbol{V}^{\natural}$. This matrix admits the following decomposition

$$\boldsymbol{U}^{\natural\top}\boldsymbol{E}\boldsymbol{V}^{\natural} = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} E_{i,j}\boldsymbol{U}_{i,\cdot}^{\natural\top}\boldsymbol{V}_{j,\cdot}^{\natural}, \tag{C.10}$$

which can be controlled through the matrix Bernstein inequality. To do so, we are in need of calculating the following quantities:

$$\begin{aligned}
L &:= \max_{(i,j)\in[n_1]\times[n_2]}\left\|E_{i,j}\boldsymbol{U}_{i,\cdot}^{\natural\top}\boldsymbol{V}_{j,\cdot}^{\natural}\right\| \leq B\sqrt{\frac{\mu^{\natural}r}{n_1}}\sqrt{\frac{\mu^{\natural}r}{n_2}} = \frac{B\mu^{\natural}r}{\sqrt{n_1n_2}}, \\
V &:= \max\left\{\left\|\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\mathbb{E}\left[\left(E_{i,j}\left(\boldsymbol{U}_{i,\cdot}^{\natural}\right)^{\top}\boldsymbol{V}_{j,\cdot}^{\natural}\right)\left(E_{i,j}\left(\boldsymbol{U}_{i,\cdot}^{\natural}\right)^{\top}\boldsymbol{V}_{j,\cdot}^{\natural}\right)^{\top}\right]\right\|,\right. \\
&\qquad\qquad \left.\left\|\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\mathbb{E}\left[\left(E_{i,j}\left(\boldsymbol{U}_{i,\cdot}^{\natural}\right)^{\top}\boldsymbol{V}_{j,\cdot}^{\natural}\right)^{\top}\left(E_{i,j}\left(\boldsymbol{U}_{i,\cdot}^{\natural}\right)^{\top}\boldsymbol{V}_{j,\cdot}^{\natural}\right)\right]\right\|\right\} \\
&\leq \max\left\{\left\|\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sigma^2\left\|\boldsymbol{V}_{j,\cdot}^{\natural}\right\|_2^2\left(\boldsymbol{U}_{i,\cdot}^{\natural}\right)^{\top}\boldsymbol{U}_{i,\cdot}^{\natural}\right\|, \left\|\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sigma^2\left\|\boldsymbol{U}_{j,\cdot}^{\natural}\right\|_2^2\left(\boldsymbol{V}_{i,\cdot}^{\natural}\right)^{\top}\boldsymbol{V}_{i,\cdot}^{\natural}\right\|\right\} \\
&= \sigma^2\max\left\{\left\|\boldsymbol{V}^{\natural}\right\|_{\mathrm{F}}^2\left\|\boldsymbol{U}^{\natural\top}\boldsymbol{U}^{\natural}\right\|, \left\|\boldsymbol{U}^{\natural}\right\|_{\mathrm{F}}^2\left\|\boldsymbol{V}^{\natural\top}\boldsymbol{V}^{\natural}\right\|\right\} = \sigma^2 r,
\end{aligned}$$

where we have made use of Assumptions 2-3. Applying matrix Bernstein inequality (Tropp, 2015, Theorem 6.1.1) to the decomposition (C.10) then tells us that, with probability exceeding $1 - O(n^{-10})$,

$$\left\|\boldsymbol{U}^{\natural\top}\boldsymbol{E}\boldsymbol{V}^{\natural}\right\| \lesssim \sqrt{V\log n} + L\log n \lesssim \sigma\sqrt{r\log n} + \frac{B\mu^{\natural}r\log n}{\sqrt{n_1n_2}} \asymp \sigma\sqrt{r\log n}, \tag{C.11}$$

provided that $B \lesssim \sigma\sqrt{n_1/(\mu^{\natural}\log n)}$ as well as $n_2 \gtrsim \mu^{\natural}r$. This implies that

$$\alpha_3 \lesssim \sigma\sqrt{r\log n}.$$

Combine the preceding bounds on α_1, α_2 and α_3 to reach

$$\left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}} - \boldsymbol{\Sigma}^{\natural}\right\| \leq \alpha_1 + \alpha_2 + \alpha_3 \lesssim \kappa^{\natural}\frac{\sigma^2 n}{\sigma_r^{\natural}} + \frac{\left(\sigma\sqrt{n}\right)^3}{\sigma_r^{\natural 2}} + \sigma\sqrt{r\log n} \asymp \kappa^{\natural}\frac{\sigma^2 n}{\sigma_r^{\natural}} + \sigma\sqrt{r\log n},$$

where the last relation holds as long as $\sigma\sqrt{n} \ll \sigma_r^\natural$.

In addition, it is readily seen from the triangle inequality that

$$\begin{aligned}\left\|\boldsymbol{H}_{\boldsymbol{U}}^\top \boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{\Sigma}^\natural\right\| &\le \left\|\boldsymbol{H}_{\boldsymbol{U}}^\top \boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{U}^{\natural\top}\boldsymbol{M}\boldsymbol{V}^\natural\right\| + \left\|\boldsymbol{U}^{\natural\top}\boldsymbol{M}\boldsymbol{V}^\natural - \boldsymbol{\Sigma}^\natural\right\|, \\ &= \alpha_2 + \alpha_3 \lesssim \frac{\left(\sigma\sqrt{n}\right)^3}{\sigma_r^{\natural 2}} + \sigma\sqrt{r\log n}.\end{aligned}$$

C.3.3 Proof of Lemma 3

Recognizing the basic identity $\boldsymbol{U}\boldsymbol{\Sigma} = \boldsymbol{M}\boldsymbol{V}$, we can invoke the triangle inequality to obtain

$$\left\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{M}\boldsymbol{V}^\natural\right\|_{2,\infty} = \left\|\boldsymbol{M}\left(\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{V}^\natural\right)\right\|_{2,\infty} \le \left\|\boldsymbol{M}^\natural\left(\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{V}^\natural\right)\right\|_{2,\infty} + \left\|\boldsymbol{E}\left(\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{V}^\natural\right)\right\|_{2,\infty}.$$

This leaves us with two crucial terms to control, which forms the main content of this subsection. In particular, our proof relies heavily on the leave-one-out analysis framework that has proved effective in analyzing spectral methods (Abbe et al., 2020; Cai et al., 2021; Chen et al., 2020c).

Step 1: introduce leave-one-out auxiliary sequences. To begin with, let us introduce the leave-one-out sequences and estimates. For each $1 \le l \le n_1$, we define

$$\boldsymbol{M}^{(l)} \coloneqq \boldsymbol{M}^\natural + \mathcal{P}_{-l,\cdot}\left(\boldsymbol{E}\right),$$

where the operator $\mathcal{P}_{-l,\cdot}$ is defined in (A.1). In words, $\boldsymbol{M}^{(l)}$ is obtained from $\boldsymbol{M}$ by discarding all randomness in the l-th row. We shall define the leave-one-out estimates $\boldsymbol{U}^{(l)}$, $\boldsymbol{V}^{(l)}$, $\boldsymbol{H}_{\boldsymbol{U}}^{(l)}$ and $\boldsymbol{H}_{\boldsymbol{V}}^{(l)}$ w.r.t. $\boldsymbol{M}^{(l)}$, in the same way as how we define $\boldsymbol{U}$, $\boldsymbol{V}$, $\boldsymbol{H}_{\boldsymbol{U}}$ and $\boldsymbol{H}_{\boldsymbol{V}}$ w.r.t. $\boldsymbol{M}$. The key statistical benefits are: $\boldsymbol{U}^{(l)}$, $\boldsymbol{V}^{(l)}$, $\boldsymbol{H}_{\boldsymbol{U}}^{(l)}$ and $\boldsymbol{H}_{\boldsymbol{V}}^{(l)}$ are all statistically independent of the l-th row of $\boldsymbol{E}$, given that $\boldsymbol{M}^{(l)}$ does not contain any entry in the l-th row of $\boldsymbol{E}$.

Step 2: bounding $\|\boldsymbol{E}(\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{V}^\natural)\|_{2,\infty}$. For each $1 \le l \le n_1$, the triangle inequality gives

$$\left\|\boldsymbol{E}_{l,\cdot}\left(\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{V}^\natural\right)\right\|_2 \le \underbrace{\left\|\boldsymbol{E}_{l,\cdot}\left(\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)} - \boldsymbol{V}^\natural\right)\right\|_2}_{=:\,\alpha_1} + \underbrace{\left\|\boldsymbol{E}_{l,\cdot}\left(\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)} - \boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}\right)\right\|_2}_{=:\,\alpha_2}.$$

We shall bound α_1 and α_2 separately in what follows.

- Let us look at α_1 first. Note that $\boldsymbol{V}^{(l)}\boldsymbol{H}^{(l)} - \boldsymbol{V}^\natural$ is independent of $\boldsymbol{E}_{l,\cdot}$. Therefore, conditional on $\boldsymbol{V}^{(l)}\boldsymbol{H}^{(l)} - \boldsymbol{V}^\natural$, we can apply Lemma 5 to show that

$$\begin{aligned}\alpha_1 &\lesssim \sigma\sqrt{\log n}\left\|\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)} - \boldsymbol{V}^\natural\right\|_{\mathrm{F}} + (B\log n)\left\|\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)} - \boldsymbol{V}^\natural\right\|_{2,\infty} \\ &\lesssim \sigma\sqrt{\log n}\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{V}^\natural\right\|_{\mathrm{F}} + (B\log n)\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{V}^\natural\right\|_{2,\infty} + \left(\sigma\sqrt{\log n} + B\log n\right)\left\|\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)} - \boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}\right\|_{\mathrm{F}} \\ &\lesssim \sigma\sqrt{\log n}\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{V}^\natural\right\|_{\mathrm{F}} + (B\log n)\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{V}^\natural\right\|_{2,\infty} + (B\log n)\left\|\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)} - \boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}\right\|_{\mathrm{F}}\end{aligned}$$

  with probability exceeding $1 - O(n^{-10})$, where the last line is valid since $B \ge \sigma$. Moreover, the orthonormality of $\boldsymbol{V}^\natural$ allows one to obtain

$$\begin{aligned}\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{V}^\natural\right\|_{\mathrm{F}} &= \left\|\boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{V}^\natural - \boldsymbol{V}^\natural\right\|_{\mathrm{F}} = \left\|\left(\boldsymbol{V}\boldsymbol{V}^\top - \boldsymbol{V}^\natural\boldsymbol{V}^{\natural\top}\right)\boldsymbol{V}^\natural\right\|_{\mathrm{F}} \le \left\|\boldsymbol{V}\boldsymbol{V}^\top - \boldsymbol{V}^\natural\boldsymbol{V}^{\natural\top}\right\|_{\mathrm{F}} \\ &\lesssim \left\|\boldsymbol{V}\boldsymbol{R}_{\boldsymbol{V}} - \boldsymbol{V}^\natural\right\|_{\mathrm{F}} \lesssim \frac{\sigma}{\sigma_r^\natural}\sqrt{nr},\end{aligned}$$

  where the last line follows from Chen et al. (2020c, Lemma 2.1.3) and Lemma 1. As a consequence, the above results taken together lead us to

$$\alpha_1 \lesssim \frac{\sigma^2}{\sigma_r^\natural}\sqrt{nr\log n} + (B\log n)\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{V}^\natural\right\|_{2,\infty} + (B\log n)\left\|\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)} - \boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}\right\|_{\mathrm{F}}. \tag{C.12}$$

- Regarding α_2, we first observe that

$$\alpha_2 \leq \|\boldsymbol{E}_{l,\cdot}\|_2 \left\|\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)} - \boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}\right\| \lesssim \sigma\sqrt{n_2}\left\|\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)} - \boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}\right\|,$$

motivating us to pay attention to the proximity of $\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}$ and $\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)}$. We then proceed via the following several sub-steps.

  – First, let us upper bound $\|\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)} - \boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}\|_{\mathrm{F}}$ and $\|\boldsymbol{U}^{(l)}\boldsymbol{H}_{\boldsymbol{U}}^{(l)} - \boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}\|_{\mathrm{F}}$. As can be easily seen,

$$\left\|\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)} - \boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}\right\|_{\mathrm{F}} = \left\|\big(\boldsymbol{V}^{(l)}\boldsymbol{V}^{(l)\top} - \boldsymbol{V}\boldsymbol{V}^{\top}\big)\boldsymbol{V}^{\natural}\right\|_{\mathrm{F}} \leq \left\|\boldsymbol{V}^{(l)}\boldsymbol{V}^{(l)\top} - \boldsymbol{V}\boldsymbol{V}^{\top}\right\|_{\mathrm{F}}, \tag{C.13a}$$

$$\left\|\boldsymbol{U}^{(l)}\boldsymbol{H}_{\boldsymbol{U}}^{(l)} - \boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}\right\|_{\mathrm{F}} = \left\|\big(\boldsymbol{U}^{(l)}\boldsymbol{U}^{(l)\top} - \boldsymbol{U}\boldsymbol{U}^{\top}\big)\boldsymbol{U}^{\natural}\right\|_{\mathrm{F}} \leq \left\|\boldsymbol{U}^{(l)}\boldsymbol{U}^{(l)\top} - \boldsymbol{U}\boldsymbol{U}^{\top}\right\|_{\mathrm{F}}. \tag{C.13b}$$

  In view of Wedin's $\sin\boldsymbol{\Theta}$ Theorem (see Wedin (1972) or Chen et al. (2020c, Theorem 2.3.1)), we can deduce that

$$\begin{aligned}
&\max\left\{\left\|\boldsymbol{U}^{(l)}\boldsymbol{U}^{(l)\top} - \boldsymbol{U}\boldsymbol{U}^{\top}\right\|_{\mathrm{F}}, \left\|\boldsymbol{V}^{(l)}\boldsymbol{V}^{(l)\top} - \boldsymbol{V}\boldsymbol{V}^{\top}\right\|_{\mathrm{F}}\right\} \\
&\quad\lesssim \frac{\max\left\{\|(\boldsymbol{M}^{(l)} - \boldsymbol{M})\boldsymbol{V}^{(l)}\|_{\mathrm{F}}, \left\|\big(\boldsymbol{M}^{(l)} - \boldsymbol{M}\big)^{\top}\boldsymbol{U}^{(l)}\right\|_{\mathrm{F}}\right\}}{\sigma_r\left(\boldsymbol{M}^{(l)}\right) - \sigma_{r+1}\left(\boldsymbol{M}\right) - \|\boldsymbol{M}^{(l)} - \boldsymbol{M}\|} \\
&\quad\leq \frac{\max\left\{\|(\boldsymbol{M}^{(l)} - \boldsymbol{M})\boldsymbol{V}^{(l)}\|_{\mathrm{F}}, \left\|\big(\boldsymbol{M}^{(l)} - \boldsymbol{M}\big)^{\top}\boldsymbol{U}^{(l)}\right\|_{\mathrm{F}}\right\}}{\sigma_r\left(\boldsymbol{M}^{\star}\right) - \sigma_{r+1}\left(\boldsymbol{M}^{\star}\right) - \|\boldsymbol{E}\| - \|\boldsymbol{E}^{(l)}\| - \|\boldsymbol{M}^{(l)} - \boldsymbol{M}\|} \\
&\quad\lesssim \frac{\max\left\{\|(\boldsymbol{M}^{(l)} - \boldsymbol{M})\boldsymbol{V}^{(l)}\|_{\mathrm{F}}, \left\|\big(\boldsymbol{M}^{(l)} - \boldsymbol{M}\big)^{\top}\boldsymbol{U}^{(l)}\right\|_{\mathrm{F}}\right\}}{\sigma_r^{\natural}},
\end{aligned} \tag{C.14}$$

  where we have invoked Lemma 1 in the last inequality. We then need to bound the two terms in $\max\{\cdot,\cdot\}$ respectively. For the first term, it is seen that

$$\begin{aligned}
\left\|\big(\boldsymbol{M}^{(l)} - \boldsymbol{M}\big)\boldsymbol{V}^{(l)}\right\|_{\mathrm{F}} &= \left\|\boldsymbol{E}_{l,\cdot}\boldsymbol{V}^{(l)}\right\|_2 \overset{\text{(i)}}{\leq} 2\left\|\boldsymbol{E}_{l,\cdot}\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)}\right\|_2 \\
&\lesssim \|\boldsymbol{E}_{l,\cdot}\boldsymbol{V}^{\natural}\|_2 + \left\|\boldsymbol{E}_{l,\cdot}\left(\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)} - \boldsymbol{V}^{\natural}\right)\right\|_2 \\
&\overset{\text{(ii)}}{\lesssim} \sigma\sqrt{\log n}\,\|\boldsymbol{V}^{\natural}\|_{\mathrm{F}} + (B\log n)\,\|\boldsymbol{V}^{\natural}\|_{2,\infty} + \alpha_1 \\
&\lesssim \sigma\sqrt{r\log n} + B\sqrt{\frac{\mu^{\natural}r}{n_2}}\log n + \alpha_1 \\
&\overset{\text{(iii)}}{\lesssim} \sigma\sqrt{r\log n} + \alpha_1.
\end{aligned}$$

  Here (i) holds since $\sigma_{\min}(\boldsymbol{H}_{\boldsymbol{V}}^{(l)}) \geq 1/2$ (see Lemma 1), (ii) is a consequence of Lemma 5, and (iii) is valid under our noise condition $B \lesssim \sigma\sqrt{n_1/(\mu^{\natural}\log n)}$. When it comes to the other term $\|\big(\boldsymbol{M}^{(l)} - \boldsymbol{M}\big)^{\top}\boldsymbol{U}^{(l)}\|_{\mathrm{F}}$, we have

$$\begin{aligned}
\left\|\big(\boldsymbol{M}^{(l)} - \boldsymbol{M}\big)^{\top}\boldsymbol{U}^{(l)}\right\|_{\mathrm{F}} &= \left\|\boldsymbol{E}_{l,\cdot}^{\top}\boldsymbol{U}_{l,\cdot}^{(l)}\right\|_{\mathrm{F}} = \|\boldsymbol{E}_{l,\cdot}\|_2\left\|\boldsymbol{U}_{l,\cdot}^{(l)}\right\|_2 \\
&\lesssim \|\boldsymbol{E}\|\left\|\boldsymbol{U}_{l,\cdot}^{(l)}\boldsymbol{H}_{\boldsymbol{U}}^{(l)}\right\|_2 \lesssim \sigma\sqrt{n}\left(\|\boldsymbol{U}^{\natural}\|_{2,\infty} + \left\|\boldsymbol{U}^{(l)}\boldsymbol{H}_{\boldsymbol{U}}^{(l)} - \boldsymbol{U}^{\natural}\right\|_{2,\infty}\right) \\
&\lesssim \sigma\sqrt{n}\,\|\boldsymbol{U}^{\natural}\|_{2,\infty} + \sigma\sqrt{n}\left\|\boldsymbol{U}^{(l)}\boldsymbol{H}^{(l)} - \boldsymbol{U}^{\natural}\right\|_{2,\infty},
\end{aligned}$$

  where the first inequality holds true since $\sigma_{\min}(\boldsymbol{H}_{\boldsymbol{U}}^{(l)}) \geq 1/2$ (see Lemma 1), and the second inequality comes from Lemma 1 and the triangle inequality. Substitution of these two bounds into (C.14) and

(C.13) yields

$$\begin{aligned}
&\max\left\{\left\|\boldsymbol{U}^{(l)}\boldsymbol{H}_{\boldsymbol{U}}^{(l)}-\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}\right\|_{\mathrm{F}},\left\|\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)}-\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}\right\|_{\mathrm{F}}\right\}\\
&\quad\le\max\left\{\left\|\boldsymbol{U}^{(l)}\boldsymbol{U}^{(l)\top}-\boldsymbol{U}\boldsymbol{U}^{\top}\right\|_{\mathrm{F}},\left\|\boldsymbol{V}^{(l)}\boldsymbol{V}^{(l)\top}-\boldsymbol{V}\boldsymbol{V}^{\top}\right\|_{\mathrm{F}}\right\}\\
&\quad\lesssim\frac{1}{\sigma_r^\natural}\left[\sigma\sqrt{r\log n}+\alpha_1+\sigma\sqrt{n}\left\|\boldsymbol{U}^\natural\right\|_{2,\infty}+\sigma\sqrt{n}\left\|\boldsymbol{U}^{(l)}\boldsymbol{H}_{\boldsymbol{U}}^{(l)}-\boldsymbol{U}^\natural\right\|_{2,\infty}\right].
\end{aligned}\tag{C.15}$$

– Next, let us establish an upper bound on $\|\boldsymbol{U}^{(l)}\boldsymbol{H}_{\boldsymbol{U}}^{(l)}-\boldsymbol{U}^\natural\|_{2,\infty}$ using the above result. Applying the triangle inequality and (C.15) gives

$$\begin{aligned}
\left\|\boldsymbol{U}^{(l)}\boldsymbol{H}_{\boldsymbol{U}}^{(l)}-\boldsymbol{U}^\natural\right\|_{2,\infty}&\le\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^\natural\right\|_{2,\infty}+\left\|\boldsymbol{U}^{(l)}\boldsymbol{H}_{\boldsymbol{U}}^{(l)}-\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}\right\|_{\mathrm{F}}\\
&\lesssim\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^\natural\right\|_{2,\infty}+\frac{1}{\sigma_r^\natural}\left[\sigma\sqrt{r\log n}+\alpha_1+\sigma\sqrt{n}\left\|\boldsymbol{U}^\natural\right\|_{2,\infty}+\sigma\sqrt{n}\left\|\boldsymbol{U}^{(l)}\boldsymbol{H}_{\boldsymbol{U}}^{(l)}-\boldsymbol{U}^\natural\right\|_{2,\infty}\right],
\end{aligned}$$

where both sides of this inequality include the term $\|\boldsymbol{U}^{(l)}\boldsymbol{H}_{\boldsymbol{U}}^{(l)}-\boldsymbol{U}^\natural\|_{2,\infty}$. Under the assumption that $\sigma\sqrt{n}\ll\sigma_r^\natural$, we can rearrange terms to simplify it as

$$\left\|\boldsymbol{U}^{(l)}\boldsymbol{H}_{\boldsymbol{U}}^{(l)}-\boldsymbol{U}^\natural\right\|_{2,\infty}\lesssim\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^\natural\right\|_{2,\infty}+\frac{1}{\sigma_r^\natural}\left(\sigma\sqrt{r\log n}+\alpha_1+\sigma\sqrt{n}\left\|\boldsymbol{U}^\natural\right\|_{2,\infty}\right).\tag{C.16}$$

– Further, the inequality (C.16) taken together with (C.15) gives

$$\begin{aligned}
\left\|\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)}-\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}\right\|_{\mathrm{F}}&\lesssim\frac{1}{\sigma_r^\natural}\left[\sigma\sqrt{r\log n}+\alpha_1+\sigma\sqrt{n}\left\|\boldsymbol{U}^\natural\right\|_{2,\infty}\right]+\frac{\sigma}{\sigma_r^\natural}\sqrt{n}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^\natural\right\|_{2,\infty}\\
&\quad+\frac{\sigma}{\sigma_r^\natural}\sqrt{n}\cdot\frac{1}{\sigma_r^\natural}\left(\sigma\sqrt{r\log n}+\alpha_1+\sigma\sqrt{n}\left\|\boldsymbol{U}^\natural\right\|_{2,\infty}\right)\\
&\asymp\frac{1}{\sigma_r^\natural}\left[\sigma\sqrt{r\log n}+\alpha_1+\sigma\sqrt{n}\left\|\boldsymbol{U}^\natural\right\|_{2,\infty}\right]+\frac{\sigma}{\sigma_r^\natural}\sqrt{n}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^\natural\right\|_{2,\infty}
\end{aligned}\tag{C.17}$$

as long as $\sigma\sqrt{n}\ll\sigma_r^\natural$.

– Putting the above pieces together, we arrive at

$$\begin{aligned}
\alpha_2&\lesssim\sigma\sqrt{n_2}\left\|\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)}-\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}\right\|\\
&\lesssim\frac{\sigma^2}{\sigma_r^\natural}\sqrt{n_2 r\log n}+\frac{\sigma}{\sigma_r^\natural}\sqrt{n_2}\alpha_1+\frac{\sigma^2\sqrt{nn_2}}{\sigma_r^\natural}\left\|\boldsymbol{U}^\natural\right\|_{2,\infty}+\frac{\sigma^2}{\sigma_r^\natural}\sqrt{nn_2}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^\natural\right\|_{2,\infty}.
\end{aligned}\tag{C.18}$$

- Given that we have obtained an upper bound (C.17) on $\|\boldsymbol{V}^{(l)}\boldsymbol{H}_{\boldsymbol{V}}^{(l)}-\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}\|_{\mathrm{F}}$, we can revisit the bound (C.12) on α_1 to further derive

$$\begin{aligned}
\alpha_1&\lesssim\frac{\sigma^2}{\sigma_r^\natural}\sqrt{nr\log n}+(B\log n)\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^\natural\right\|_{2,\infty}\\
&\quad+\frac{B\log n}{\sigma_r^\natural}\left[\sigma\sqrt{r\log n}+\alpha_1+\sigma\sqrt{n}\left\|\boldsymbol{U}^\natural\right\|_{2,\infty}+\sigma\sqrt{n}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^\natural\right\|_{2,\infty}\right].
\end{aligned}$$

Importantly, both sides of this inequality include the term α_1. Under the assumption that $B\lesssim\sigma\sqrt{n_1/(\mu^\natural\log n)}$, we can rearrange terms to derived a simpler bound as follows:

$$\alpha_1\lesssim\frac{\sigma\sqrt{n}+B\log n}{\sigma_r^\natural}\sigma\sqrt{r\log n}+(B\log n)\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^\natural\right\|_{2,\infty}+\frac{B\log n}{\sigma_r^\natural}\sigma\sqrt{n}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^\natural\right\|_{2,\infty}.\tag{C.19}$$

Combine the bound (C.19) on α_1 and the bound (C.18) on α_2 to yield

$$\begin{aligned}
\left\|\boldsymbol{E}_{l,\cdot}\left(\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right)\right\|_2 &\leq \alpha_1+\alpha_2\\
&\leq \alpha_1+\frac{\sigma^2}{\sigma_r^{\natural}}\sqrt{n_2 r\log n}+\frac{\sigma^2}{\sigma_r^{\natural}}\sqrt{nn_2}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\frac{\sigma^2}{\sigma_r^{\natural}}\sqrt{nn_2}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}\\
&\lesssim \frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{r\log n}+\frac{\sigma^2}{\sigma_r^{\natural}}\sqrt{nn_2}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}\\
&\quad+(B\log n)\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right\|_{2,\infty}+\frac{\sigma\sqrt{n_2}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{n}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^{\natural}\right\|_{2,\infty},
\end{aligned}$$

where the second inequality holds as long as $\sigma\sqrt{n}\ll\sigma_r^{\natural}$. Given that this bound holds for all $l\in[n_1]$ with high probability, we can conclude that

$$\begin{aligned}
\left\|\boldsymbol{E}\left(\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right)\right\|_{2,\infty} &\lesssim \frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{r\log n}+\frac{\sigma^2 n}{\sigma_r^{\natural}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}\\
&\quad+(B\log n)\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right\|_{2,\infty}+\frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{n}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}.
\end{aligned}$$

Step 3: bounding $\|\boldsymbol{M}^{\natural}(\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural})\|_{2,\infty}$. Next, we observe that

$$\left\|\boldsymbol{M}^{\natural}\left(\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right)\right\|_{2,\infty}=\left\|\boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural}\boldsymbol{V}^{\natural\top}\left(\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right)\right\|_{2,\infty}\leq\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}\sigma_1^{\natural}\left\|\boldsymbol{V}^{\natural\top}\left(\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right)\right\|.$$

Let us denote by $\boldsymbol{X}(\cos\boldsymbol{\Theta})\boldsymbol{Y}^{\top}$ the SVD of $\boldsymbol{H}_{\boldsymbol{V}}=\boldsymbol{V}^{\top}\boldsymbol{V}^{\natural}$, where $\boldsymbol{X},\boldsymbol{Y}\in\mathbb{R}^{r\times r}$ are orthonormal matrices and $\boldsymbol{\Theta}$ is a diagonal matrix consisting of the principal angles between the subspaces spanned by the columns of $\boldsymbol{V}$ and $\boldsymbol{V}^{\natural}$. We can then deduce that

$$\begin{aligned}
\left\|\boldsymbol{V}^{\natural\top}\left(\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right)\right\| &= \left\|\boldsymbol{H}_{\boldsymbol{V}}^{\top}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{I}_r\right\|=\left\|\boldsymbol{Y}\left(\cos^2\boldsymbol{\Theta}\right)\boldsymbol{Y}^{\top}-\boldsymbol{I}_r\right\|\\
&=\left\|\boldsymbol{Y}\left(\cos^2\boldsymbol{\Theta}-\boldsymbol{I}_r\right)\boldsymbol{Y}^{\top}\right\|=\left\|\cos^2\boldsymbol{\Theta}-\boldsymbol{I}_r\right\|\\
&=\left\|\sin^2\boldsymbol{\Theta}\right\|=\left\|\sin\boldsymbol{\Theta}\right\|^2.
\end{aligned}$$

From Chen et al. (2020c, Lemma 2.1.2 and Lemma 2.1.3) and Lemma 1, we know that

$$\|\sin\boldsymbol{\Theta}\|\lesssim\left\|\boldsymbol{V}\boldsymbol{R}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right\|\lesssim\frac{\sigma}{\sigma_r^{\natural}}\sqrt{n},$$

thus indicating that

$$\left\|\boldsymbol{M}^{\natural}\left(\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right)\right\|_{2,\infty}\lesssim\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}\sigma_1^{\natural}\frac{\sigma^2 n}{\sigma_r^{\natural 2}}\lesssim\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}.$$

Step 4: putting all these results together. Combining the bounds in the preceding steps, we arrive at

$$\begin{aligned}
\left\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{M}\boldsymbol{V}^{\natural}\right\|_{2,\infty} &\leq \left\|\boldsymbol{M}^{\natural}\left(\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right)\right\|_{2,\infty}+\left\|\boldsymbol{E}\left(\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right)\right\|_{2,\infty}\\
&\lesssim \frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{r\log n}+\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}\\
&\quad+(B\log n)\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right\|_{2,\infty}+\frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{n}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}.
\end{aligned}$$

Similarly, we can also show that

$$\begin{aligned}
\left\|\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{M}^{\top}\boldsymbol{U}^{\natural}\right\|_{2,\infty} &\lesssim \frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{r\log n}+\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural}}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty}\\
&\quad+(B\log n)\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\sigma\sqrt{n}\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right\|_{2,\infty}.
\end{aligned}$$

This concludes the proof of this lemma.

C.3.4 Proof of Lemma 4

Let us start by decomposing

$$\begin{aligned}
\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^{\natural}\right\|_{2,\infty} &\leq \frac{1}{\sigma_r^{\natural}}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural}-\boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural}\right\|_{2,\infty} \\
&\leq \frac{1}{\sigma_r^{\natural}}\left\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural}\right\|_{2,\infty}+\frac{1}{\sigma_r^{\natural}}\left\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural}\right\|_{2,\infty} \\
&\leq \underbrace{\frac{1}{\sigma_r^{\natural}}\left\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{M}\boldsymbol{V}^{\natural}\right\|_{2,\infty}}_{=:\,\alpha_1}+\underbrace{\frac{1}{\sigma_r^{\natural}}\left\|\boldsymbol{E}\boldsymbol{V}^{\natural}\right\|_{2,\infty}}_{=:\,\alpha_2}+\underbrace{\frac{1}{\sigma_r^{\natural}}\left\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural}\right\|_{2,\infty}}_{=:\,\alpha_3},
\end{aligned}$$

where the last line relies on the identity $\boldsymbol{M}\boldsymbol{V}^{\natural}=\boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural}+\boldsymbol{E}\boldsymbol{V}^{\natural}$. Similarly, we can derive

$$\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right\|_{2,\infty} \leq \underbrace{\frac{1}{\sigma_r^{\natural}}\left\|\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{M}^{\top}\boldsymbol{U}^{\natural}\right\|_{2,\infty}}_{=:\,\beta_1}+\underbrace{\frac{1}{\sigma_r^{\natural}}\left\|\boldsymbol{E}^{\top}\boldsymbol{U}^{\natural}\right\|_{2,\infty}}_{=:\,\beta_2}+\underbrace{\frac{1}{\sigma_r^{\natural}}\left\|\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}\boldsymbol{\Sigma}^{\natural}\right\|_{2,\infty}}_{=:\,\beta_3}.$$

In the sequel, we shall bound α_1, α_2, α_3, β_1, β_2 and β_3 separately.

- Regarding α_1 and β_1, Lemma 3 tells us that

$$\begin{aligned}
\alpha_1 &\lesssim \frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n}+\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} \\
&\quad+\frac{B\log n}{\sigma_r^{\natural}}\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right\|_{2,\infty}+\frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\frac{\sigma}{\sigma_r^{\natural}}\sqrt{n}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^{\natural}\right\|_{2,\infty};\\
\beta_1 &\lesssim \frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n}+\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} \\
&\quad+\frac{B\log n}{\sigma_r^{\natural}}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\frac{\sigma\sqrt{n}+B\log n}{\sigma_r^{\natural}}\frac{\sigma}{\sigma_r^{\natural}}\sqrt{n}\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{V}^{\natural}\right\|_{2,\infty}.
\end{aligned}$$

- We now turn to α_2 and β_2. In view of Lemma 5, we know that with probability exceeding $1-O(n^{-10})$

$$\alpha_2 \lesssim \frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n}+\frac{B\log n}{\sigma_r^{\natural}}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty}\lesssim\frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n}+\frac{B\log n}{\sigma_r^{\natural}}\sqrt{\frac{\mu^{\natural}r}{n_2}}\lesssim\frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n}$$

  under Assumption 2 and our noise condition $B\lesssim\sigma\sqrt{n_1/(\mu^{\natural}\log n)}$. Similarly, we can deduce that

$$\beta_2 \lesssim \frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n}.$$

- With regards to α_3 and β_3, it is seen from Lemma 2 that

$$\begin{aligned}
\alpha_3 &\leq \frac{1}{\sigma_r^{\natural}}\left\|\boldsymbol{U}\right\|_{2,\infty}\left\|\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{H}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural}\right\| \lesssim \frac{1}{\sigma_r^{\natural}}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}\right\|_{2,\infty}\left\|\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{H}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural}\right\| \\
&\lesssim \frac{1}{\sigma_r^{\natural}}\left(\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}\right)\left\|\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{H}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural}\right\|,
\end{aligned}$$

  where the penultimate relation holds since $\sigma_r(\boldsymbol{H}_{\boldsymbol{U}})\geq 1/2$ (see Lemma 1). Furthermore,

$$\begin{aligned}
\left\|\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{H}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural}\right\| &= \left\|\left(\boldsymbol{H}_{\boldsymbol{U}}^{\top}\right)^{-1}\left(\boldsymbol{H}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{H}_{\boldsymbol{U}}^{\top}\boldsymbol{H}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural}\right)\right\| \\
&\overset{\text{(i)}}{\lesssim} \left\|\boldsymbol{H}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{H}_{\boldsymbol{U}}^{\top}\boldsymbol{H}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural}\right\| \\
&\leq \left\|\boldsymbol{H}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{H}_{\boldsymbol{V}}-\boldsymbol{\Sigma}^{\natural}\right\|+\left\|\left(\boldsymbol{H}_{\boldsymbol{U}}^{\top}\boldsymbol{H}_{\boldsymbol{U}}-\boldsymbol{I}_r\right)\boldsymbol{\Sigma}^{\natural}\right\|
\end{aligned}$$

$$\overset{\text{(ii)}}{\lesssim} \frac{\left(\sigma\sqrt{n}\right)^3}{\sigma_r^{\natural 2}} + \sigma\sqrt{r\log n} + \sigma_1^{\natural}\frac{\sigma^2 n}{\sigma_r^{\natural 2}}$$
$$\overset{\text{(iii)}}{\asymp} \sigma\sqrt{r\log n} + \kappa^{\natural}\frac{\sigma^2 n}{\sigma_r^{\natural}}.$$

Here (i) holds since $\|\boldsymbol{H}_{\boldsymbol{U}}^{-1}\| \leq 2$ (see Lemma 1); (ii) uses Lemma 2 as well as the following fact

$$\begin{aligned}\left\|\boldsymbol{H}_{\boldsymbol{U}}^{\top}\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{I}_r\right\| = \left\|\boldsymbol{H}_{\boldsymbol{U}}^{\top}\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{R}_{\boldsymbol{U}}\right\| &\leq \left\|\boldsymbol{H}_{\boldsymbol{U}}^{\top}\left(\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{R}_{\boldsymbol{U}}\right)\right\| + \left\|\left(\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{R}_{\boldsymbol{U}}\right)^{\top}\boldsymbol{R}_{\boldsymbol{U}}\right\| \\ &\lesssim \left(\frac{\sigma}{\sigma_r^{\natural}}\sqrt{n}\right)^2,\end{aligned}$$

where the last inequality is an immediate consequence of Lemma 1; and (iii) holds true provided that $\sigma\sqrt{n} \ll \sigma_r^{\natural}$. Consequently, these bounds allow us to derive

$$\alpha_3 \lesssim \frac{1}{\sigma_r^{\natural}}\left(\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} + \left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{U}^{\natural}\right\|_{2,\infty}\right)\left(\sigma\sqrt{r\log n} + \kappa^{\natural}\frac{\sigma^2 n}{\sigma_r^{\natural}}\right).$$

The bound on β_3 follows analogously.

Equipped with the above bounds on α_1, α_2 and α_3, we can readily show that

$$\begin{aligned}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{U}^{\natural}\right\|_{2,\infty} &\leq \alpha_1 + \alpha_2 + \alpha_3 \\ &\lesssim \frac{\sigma\sqrt{n} + B\log n}{\sigma_r^{\natural}}\frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n} + \frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} \\ &\quad + \frac{B\log n}{\sigma_r^{\natural}}\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{V}^{\natural}\right\|_{2,\infty} + \frac{\sigma\sqrt{n} + B\log n}{\sigma_r^{\natural}}\frac{\sigma}{\sigma_r^{\natural}}\sqrt{n}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{U}^{\natural}\right\|_{2,\infty} \\ &\quad + \frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n} + \left(\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} + \left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{U}^{\natural}\right\|_{2,\infty}\right)\left(\frac{\sigma\sqrt{r\log n}}{\sigma_r^{\natural}} + \kappa^{\natural}\frac{\sigma^2 n}{\sigma_r^{\natural 2}}\right).\end{aligned}$$

It is worth noting that both sides of the above inequality contain the term $\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{U}^{\natural}\|_{2,\infty}$. Under the circumstance where $\sigma\sqrt{n\log n} \ll \sigma_r^{\natural}/\kappa^{\natural}$, $B \lesssim \sigma\sqrt{n_1/(\mu^{\natural}\log n)}$ and $n_1 \geq \mu^{\natural} r$, we can rearrange terms and obtain

$$\begin{aligned}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{U}^{\natural}\right\|_{2,\infty} &\lesssim \frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n} + \frac{\sigma\sqrt{r\log n}}{\sigma_r^{\natural}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} + \frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} + \frac{B\log n}{\sigma_r^{\natural}}\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{V}^{\natural}\right\|_{2,\infty} \\ &\asymp \frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n} + \frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} + \frac{B\log n}{\sigma_r^{\natural}}\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{V}^{\natural}\right\|_{2,\infty}.\end{aligned}$$

Invoke an analogous argument to also show that

$$\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{V}^{\natural}\right\|_{2,\infty} \lesssim \frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n} + \frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} + \frac{B\log n}{\sigma_r^{\natural}}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{U}^{\natural}\right\|_{2,\infty}.$$

Taking the above two inequalities collectively, we can eliminate the appearance of $\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{V}^{\natural}\right\|_{2,\infty}$ and arrive at

$$\begin{aligned}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{U}^{\natural}\right\|_{2,\infty} &\lesssim \frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n} + \frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} \\ &\quad + \frac{B\log n}{\sigma_r^{\natural}}\left(\frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n} + \frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} + \frac{B\log n}{\sigma_r^{\natural}}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{U}^{\natural}\right\|_{2,\infty}\right) \\ &\lesssim \frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n} + \frac{B\log n}{\sigma_r^{\natural}}\frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} + \frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} + \frac{B^2\log^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{U}^{\natural}\right\|_{2,\infty} \\ &\asymp \frac{\sigma}{\sigma_r^{\natural}}\sqrt{r\log n} + \frac{\kappa^{\natural}\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} + \frac{B^2\log^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{U}^{\natural}\right\|_{2,\infty}.\end{aligned} \tag{C.20}$$

Here, the penultimate line holds true provided that $B \lesssim \sigma\sqrt{n_1/(\mu^\natural \log n)}$ and $\sigma\sqrt{n \log n} \ll \sigma_r^\natural/\kappa^\natural$, while the last line follows since

$$\frac{B\log n}{\sigma_r^\natural}\frac{\kappa^\natural\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{V}^\natural\right\|_{2,\infty} \le \frac{B\log n}{\sigma_r^\natural}\frac{\kappa^\natural\sigma^2 n}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^\natural r}{n_2}} \lesssim \frac{\sigma}{\sigma_r^\natural}\sqrt{r\log n},$$

which again makes use of the assumptions $B \lesssim \sigma\sqrt{n_1/(\mu^\natural \log n)}$, $\sigma\sqrt{n\log n} \ll \sigma_r^\natural/\kappa^\natural$ as well as Assumptions 2.

It is observed that $\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{U}^\natural\|_{2,\infty}$ appears on the both side of the above inequality (C.20). Given the assumption $B^2\log^2 n \lesssim (\sigma\sqrt{n\log n})^2 \ll \sigma_r^{\natural 2}$, one can rearrange terms to yield

$$\left\|\boldsymbol{U}\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{U}^\natural\right\|_{2,\infty} \lesssim \frac{\sigma}{\sigma_r^\natural}\sqrt{r\log n} + \frac{\kappa^\natural\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}^\natural\right\|_{2,\infty}.$$

Repeating a similar argument allows one to establish the other claim

$$\left\|\boldsymbol{V}\boldsymbol{H}_{\boldsymbol{V}} - \boldsymbol{V}^\natural\right\|_{2,\infty} \lesssim \frac{\sigma}{\sigma_r^\natural}\sqrt{r\log n} + \frac{\kappa^\natural\sigma^2 n}{\sigma_r^{\natural 2}}\left\|\boldsymbol{V}^\natural\right\|_{2,\infty}.$$

C.3.5 Proof of Lemma 5

For any $l \in [d]$, we can write

$$\boldsymbol{E}_{l,\cdot}\boldsymbol{A} = \sum_{j=1}^{n_2} E_{l,j}\boldsymbol{A}_{j,\cdot},$$

which can be viewed as a sum of independent random vectors. It is straightforward to compute that

$$L \coloneqq \max_{j\in[n_2]}\left\|E_{l,j}\boldsymbol{A}_{j,\cdot}\right\|_2 \le B\left\|\boldsymbol{A}\right\|_{2,\infty},$$
$$V \coloneqq \sum_{j=1}^{n_2}\mathbb{E}\left[\left\|E_{l,j}\boldsymbol{A}_{j,\cdot}\right\|_2^2\right] \le \sigma^2\sum_{j=1}^{n_2}\left\|\boldsymbol{A}_{j,\cdot}\right\|_2^2 = \sigma^2\left\|\boldsymbol{A}\right\|_{\mathrm{F}}^2.$$

Apply the matrix Bernstein inequality (Tropp, 2015, Theorem 6.1.1) to show that with probability exceeding $1 - O(n^{-11})$,

$$\left\|\boldsymbol{E}_{l,\cdot}\boldsymbol{A}\right\|_2 \lesssim \sqrt{V\log n} + L\log n \lesssim \sigma\left\|\boldsymbol{A}\right\|_{\mathrm{F}}\sqrt{\log n} + B\left\|\boldsymbol{A}\right\|_{2,\infty}\log n.$$

Since this holds for all $l \in [d]$, applying the union bound yields that with probability exceeding $1 - O(n^{-10})$,

$$\left\|\boldsymbol{E}\boldsymbol{A}\right\|_{2,\infty} = \max_{l\in[d]}\left\|\boldsymbol{E}_{l,\cdot}\boldsymbol{A}\right\|_2 \lesssim \sigma\left\|\boldsymbol{A}\right\|_{\mathrm{F}}\sqrt{\log n} + B\left\|\boldsymbol{A}\right\|_{2,\infty}\log n.$$

D Analysis for PCA: the SVD-based approach

In this section, we establish our theoretical guarantees for the SVD-based approach presented in Section 3.2. We shall establish our inference guarantees by connecting the PCA model with the matrix denoising problem studied in Section 6.1.

Before embarking on the analysis, we claim that: without loss of generality, we can work with the assumption that

$$\max_{j\in[n]}|\eta_{l,j}| \le C_{\mathsf{noise}}\omega_l^\star\sqrt{\log(n+d)} \qquad \text{for all } l \in [d], \tag{D.1}$$

for some absolute constant $C_{\mathsf{noise}} > 0$, in addition to the noise condition we have already imposed in Section 1.1. To see why this is valid, we note that by applying Lemma 51 (with $\delta = C_\delta(n+d)^{-100}$ for some sufficiently small constant $C_\delta > 0$) to $\{\eta_{l,j}\}$, we can produce a set of auxiliary random variables $\{\widetilde{\eta}_{l,j}\}$ such that

- the $\widetilde{\eta}_{l,j}$'s are independent sub-Gaussian random variables satisfying

$$\mathbb{E}[\widetilde{\eta}_{l,j}] = 0, \qquad \widetilde{\omega}_l^{\star 2} \coloneqq \mathbb{E}[\widetilde{\eta}_{l,j}^2] = \left[1 + O\left((n+d)^{-50}\right)\right]\omega_l^{\star 2}, \qquad \|\widetilde{\eta}_{l,j}\|_{\psi_2} \lesssim \omega_l^\star, \qquad |\widetilde{\eta}_{l,j}| \lesssim \omega_l^\star\sqrt{\log(n+d)};$$

- these auxiliary variables satisfy

$$\mathbb{P}\left(\eta_{l,j}=\widetilde{\eta}_{l,j} \text{ for all } l\in[d] \text{ and } j\in[n]\right)\geq 1-O\left((n+d)^{-98}\right).$$

The above properties suggest that: if we replace the noise matrix $\boldsymbol{N}=[\eta_{l,j}]_{l\in[d],j\in[n]}$ with $\widetilde{\boldsymbol{N}}=[\widetilde{\eta}_{l,j}]_{l\in[d],j\in[n]}$, then the observations remain unchanged (i.e. $\mathcal{P}_{\Omega}(\boldsymbol{X}+\boldsymbol{N})=\mathcal{P}_{\Omega}(\boldsymbol{X}+\widetilde{\boldsymbol{N}})$) with high probability, and consequently, the resulting estimates are also unchanged. In light of this, our analysis proceeds with the following steps.

- We shall start by proving Theorems 1-4 under the additional assumption (D.1); if this can be accomplished, then the results are clearly valid if we replace $\boldsymbol{N}$ (resp. $\{\omega_l^\star\}_{l\in[d]}$) with $\widetilde{\boldsymbol{N}}$ (resp. $\{\widetilde{\omega}_l^\star\}_{l=1}^d$).
- The next step then boils down to showing the proximity of $\widetilde{\omega}_l^\star$ and $\omega_l^\star$, which will then be invoked to establish these theorems without assuming (D.1).

D.1 Connection between PCA and matrix denoising

In order to invoke our theoretical guarantees for matrix denoising (i.e., Theorem 9) to assist in understanding PCA, we need to establish an explicit connection between these two models. This forms the main content of this subsection.

Specification of $\boldsymbol{M}^\natural$, $\boldsymbol{M}$ and $\boldsymbol{E}$. Recall that in our PCA model in Section 1.1, we assume the covariance matrix $\boldsymbol{S}^\star$ admits the eigen-decomposition $\boldsymbol{U}^\star\boldsymbol{\Sigma}^{\star 2}\boldsymbol{U}^{\star\top}$. As a result, the matrix $\boldsymbol{X}$ (cf. (2.1a)) can be equivalently expressed as

$$\boldsymbol{X}=\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\left[\boldsymbol{f}_1,\dots,\boldsymbol{f}_n\right]=\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\boldsymbol{F},\qquad\text{where}\qquad \boldsymbol{f}_i\overset{\text{i.i.d.}}{\sim}\mathcal{N}\left(\boldsymbol{0},\boldsymbol{I}_r\right).$$

Recalling that our observation matrix is $\boldsymbol{Y}=\mathcal{P}_{\Omega}(\boldsymbol{X}+\boldsymbol{N})$, we can specify the matrices $\boldsymbol{M}^\natural$, $\boldsymbol{M}$ and $\boldsymbol{E}$ as defined in Appendix 6.1 as follows:

$$\boldsymbol{M}:=\frac{1}{\sqrt{np}}\boldsymbol{Y},\qquad \boldsymbol{M}^\natural:=\mathbb{E}\left[\boldsymbol{M}\middle|\boldsymbol{F}\right]=\frac{1}{\sqrt{n}}\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\boldsymbol{F}\qquad\text{and}\qquad \boldsymbol{E}:=\boldsymbol{M}-\boldsymbol{M}^\natural. \tag{D.2}$$

As usual, we shall let the SVD of $\boldsymbol{M}^\natural$ be $\boldsymbol{M}^\natural=\boldsymbol{U}^\natural\boldsymbol{\Sigma}^\natural\boldsymbol{V}^{\natural\top}$, and use $\kappa^\natural,\mu^\natural,\sigma_r^\natural,\sigma_1^\natural$ to denote the conditional number, the incoherence parameter, the minimum and maximum singular value of $\boldsymbol{M}^\natural$, respectively. Here and below, we shall focus on the randomness of noise and missing data while treating $\boldsymbol{F}$ as given (even though it is generated randomly).

We shall also specify several useful relations between $(\boldsymbol{U}^\natural,\boldsymbol{V}^\natural,\boldsymbol{R}_{\boldsymbol{U}})$ and $(\boldsymbol{U}^\star,\boldsymbol{V}^\star,\boldsymbol{R})$, where the rotation matrix $\boldsymbol{R}$ is defined as $\boldsymbol{R}=\arg\min_{\boldsymbol{O}\in\mathcal{O}^{r\times r}}\|\boldsymbol{U}\boldsymbol{O}-\boldsymbol{U}^\star\|_{\mathrm{F}}$. First, observe that

$$\boldsymbol{V}^\natural=\boldsymbol{M}^{\natural\top}\boldsymbol{U}^\natural\left(\boldsymbol{\Sigma}^\natural\right)^{-1}=\frac{1}{\sqrt{n}}\boldsymbol{F}^\top\underbrace{\boldsymbol{\Sigma}^\star\boldsymbol{U}^{\star\top}\boldsymbol{U}^\natural\left(\boldsymbol{\Sigma}^\natural\right)^{-1}}_{=:\,\boldsymbol{J}}. \tag{D.3}$$

Given that $\boldsymbol{U}^\star$ and $\boldsymbol{U}^\natural$ represent the same column space, there exists $\boldsymbol{Q}\in\mathcal{O}^{r\times r}$ such that

$$\boldsymbol{U}^\natural=\boldsymbol{U}^\star\boldsymbol{Q}, \tag{D.4}$$

thus leading to the expression

$$\boldsymbol{J}=\boldsymbol{\Sigma}^\star\boldsymbol{Q}(\boldsymbol{\Sigma}^\natural)^{-1}. \tag{D.5}$$

The definition of $\boldsymbol{R}$ together with $\boldsymbol{U}^\natural=\boldsymbol{U}^\star\boldsymbol{Q}$ also allows one to derive

$$\boldsymbol{R}_{\boldsymbol{U}}=\arg\min_{\boldsymbol{O}\in\mathcal{O}^{r\times r}}\|\boldsymbol{U}\boldsymbol{O}-\boldsymbol{U}^\natural\|_{\mathrm{F}}=\arg\min_{\boldsymbol{O}\in\mathcal{O}^{r\times r}}\|\boldsymbol{U}\boldsymbol{O}\boldsymbol{Q}^\top-\boldsymbol{U}^\star\|_{\mathrm{F}}=\boldsymbol{R}\boldsymbol{Q}. \tag{D.6}$$

Statistical properties of $\boldsymbol{E} = [E_{i,j}]$. We now describe the statistical properties of the perturbation matrix $\boldsymbol{E}$. To begin with, it is readily seen from the definition of $\boldsymbol{E}$ that $\mathbb{E}[\boldsymbol{E} \mid \boldsymbol{F}] = \boldsymbol{0}$. Moreover, from the definition of $\boldsymbol{M}$ and $\boldsymbol{M}^{\natural}$, it is observed that

$$\boldsymbol{E} = \boldsymbol{M} - \boldsymbol{M}^{\natural\top} = \frac{1}{\sqrt{n}}\left[\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{F} + \boldsymbol{N}\right) - \boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{F}\right] = \frac{1}{\sqrt{n}}\left[\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{F}\right) - \boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{F}\right] + \frac{1}{\sqrt{n}}\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{N}\right), \tag{D.7}$$

which is clearly a zero-mean matrix conditional on $\boldsymbol{F}$. In addition, for location (i, j), we have

$$M_{i,j}^{\natural} = \frac{1}{\sqrt{n}}\left(\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{F}\right)_{i,j} = \frac{1}{\sqrt{n}}\left(\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{F}^{\star}\right)_{i,j} = \frac{1}{\sqrt{n}}\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_{j},$$

and hence the variance of $E_{i,j}$ can be calculated as

$$\sigma_{i,j}^{2} \coloneqq \mathsf{var}\left(E_{i,j}|\boldsymbol{F}\right) = \mathbb{E}\left[E_{i,j}^{2}|\boldsymbol{F}\right] = \frac{1-p}{np}\left(\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{F}\right)_{i,j}^{2} + \frac{\omega_{i}^{\star 2}}{np} = \frac{1-p}{np}\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_{j}\right)^{2} + \frac{\omega_{i}^{\star 2}}{np}. \tag{D.8}$$

A good event $\mathcal{E}_{\mathsf{good}}$. Finally, the lemma below defines a high-probability event $\mathcal{E}_{\mathsf{good}}$ under which the random quantities defined above enjoy appealing properties.

Lemma 6. *There exists an event $\mathcal{E}_{\mathsf{good}}$ with $\mathbb{P}(\mathcal{E}_{\mathsf{good}}) \geq 1 - O((n+d)^{-10})$, on which the following properties hold.*

- *$\mathcal{E}_{\mathsf{good}}$ is $\sigma(\boldsymbol{F})$-measurable, where $\sigma(\boldsymbol{F})$ is the σ-algebra generated by $\boldsymbol{F}$.*
- *If $n \gg \kappa^{2}(r + \log(n+d))$, then one has*

$$\left\|\boldsymbol{\Sigma}^{\natural} - \boldsymbol{\Sigma}^{\star}\right\| \lesssim \kappa\sqrt{\frac{r + \log\left(n+d\right)}{n}}\sigma_{r}^{\star}, \tag{D.9}$$

$$\sigma_{r}^{\natural} \asymp \sigma_{r}^{\star} \qquad \text{and} \qquad \sigma_{1}^{\natural} \asymp \sigma_{1}^{\star}. \tag{D.10}$$

- *The conditional number $\kappa^{\natural}$ and the incoherence parameter $\mu^{\natural}$ of $\boldsymbol{M}^{\natural}$ obey*

$$\kappa^{\natural} \asymp \sqrt{\kappa}, \tag{D.11}$$

$$\mu^{\natural} \lesssim \mu + \log\left(n+d\right). \tag{D.12}$$

In addition, we have the following $\ell_{2,\infty}$ norm bound for $\boldsymbol{U}^{\natural}$ and $\boldsymbol{V}^{\natural}$:

$$\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} \leq \sqrt{\frac{\mu r}{d}} \qquad \text{and} \qquad \left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} \lesssim \sqrt{\frac{r\log\left(n+d\right)}{n}}. \tag{D.13}$$

- *The noise levels $\{\sigma_{i,j}\}$ are upper bounded by*

$$\max_{i\in[d],j\in[n]}\sigma_{i,j}^{2} \lesssim \frac{\mu r\log\left(n+d\right)}{ndp}\sigma_{1}^{\star 2} + \frac{\omega_{\max}^{2}}{np} =: \sigma_{\mathsf{ub}}^{2}, \tag{D.14}$$

$$\begin{aligned}\max_{i\in[d],j\in[n]}\left|E_{i,j}\right| &\lesssim \max_{i\in[d]}\frac{1}{p}\sqrt{\frac{\log\left(n+d\right)}{n}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2} + \frac{\omega_{l}^{\star}}{p}\sqrt{\frac{\log\left(n+d\right)}{n}} \\ &\lesssim \frac{1}{p}\sqrt{\frac{\mu r\log\left(n+d\right)}{nd}}\sigma_{1}^{\star} + \frac{\omega_{\max}}{p}\sqrt{\frac{\log\left(n+d\right)}{n}} =: B. \end{aligned}\tag{D.15}$$

Here, Lemma 6 only lists part of the properties that hold when the event $\mathcal{E}_{\mathsf{good}}$ happens. See Appendix E.5 for a complete list of useful properties as well as the proofs.

D.2 Distributional characterization for principal subspace (Proof of Theorem 11)

Armed with the above connection between matrix denoising and PCA, we are ready to demonstrate how we can invoke the matrix denoising theorem (i.e., Theorem 9) to obtain a distributional characterization of $\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star}$ for the SVD-based approach (as stated in Theorem 11).

Step 1: first-order approximation and its tightness. To begin with, let us invoke Theorem 9 and the explicit connection between matrix denoising and PCA to derive a nearly tight first-order approximation of $\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}$ for the PCA problem. This is a crucial step that paves the way to our uncertainty quantification method.

Lemma 7. *Assume that*

$$(n \wedge d)\, p \gg \kappa^2 \mu r \log^2(n+d) \qquad \text{and} \qquad \frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{n+d}{np}} \ll \frac{1}{\kappa \log^{1/2}(n+d)}.$$

Then on the event $\mathcal{E}_{\mathsf{good}}$ (see Lemma 6) one has

$$\mathbb{P}\left(\left\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star} - \boldsymbol{Z}\right\|_{2,\infty} \lesssim \zeta_{\mathsf{2nd}} \,\middle|\, \boldsymbol{F}\right) \geq 1 - O\left((n+d)^{-10}\right),$$

where

$$\begin{aligned}
\boldsymbol{Z} &\coloneqq \boldsymbol{E}\boldsymbol{V}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top}, \\
\zeta_{\mathsf{2nd}} &\coloneqq \frac{\sigma_{\mathsf{ub}}^2\sqrt{(n+d)\,r}\log(n+d)}{\sigma_r^{\star 2}} + \left(\frac{\sqrt{\kappa}\sigma_{\mathsf{ub}}^2(n+d)}{\sigma_r^{\star 2}} + \frac{\sigma_{\mathsf{ub}}\sqrt{r\log(n+d)}}{\sigma_r^{\star}}\right)\left\|\boldsymbol{U}^{\star}\right\|_{2,\infty}.
\end{aligned}$$

Here, the rotation matrix $\boldsymbol{Q}$ has been introduced in (D.4), and σ_{ub} has been defined in (D.14).

Proof. See Appendix E.1.1. ☐

In the above lemma, the term $\boldsymbol{Z} = \boldsymbol{E}\boldsymbol{V}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top}$ is a linear mapping of the perturbation matrix $\boldsymbol{E}$, and can therefore be viewed as the first-order approximation. As we shall see, the higher-order error is well-controlled and dominated by the size of the first-order term.

Step 2: computing the covariance of the first-order approximation. We now proceed to characterize the covariance of the first-order approximation. Specifically, observe that the l-th row of $\boldsymbol{Z}$ ($1 \leq l \leq d$) is given by

$$\boldsymbol{Z}_{l,\cdot} = \boldsymbol{E}_{l,\cdot}\boldsymbol{V}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top} = \sum_{j=1}^{n} E_{l,j}\boldsymbol{V}_{j,\cdot}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top}. \tag{D.16}$$

The covariance matrix of this zero-mean random vector $\boldsymbol{Z}_{l,\cdot}$ (conditional on $\boldsymbol{F}$) can then be calculated as

$$\widetilde{\boldsymbol{\Sigma}}_l \coloneqq \boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{V}^{\natural\top}\mathsf{diag}\left\{\sigma_{l,1}^2,\dots,\sigma_{l,n}^2\right\}\boldsymbol{V}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top}.$$

It remains to see whether $\widetilde{\boldsymbol{\Sigma}}_l$ can be expressed directly in terms of the key quantities introduced in the PCA model. Towards this end, we find it convenient to introduce a deterministic matrix as follows

$$\boldsymbol{\Sigma}_{U,l}^{\star} = \left(\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{\omega_l^{\star 2}}{np}\right)(\boldsymbol{\Sigma}^{\star})^{-2} + \frac{2(1-p)}{np}\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^{\star}, \tag{D.17}$$

which, according to the next lemma, serves as a reasonably good proxy of $\widetilde{\boldsymbol{\Sigma}}_l$.

Lemma 8. *Suppose that $n \gg \kappa^5 r \log^3(n+d)$. On the event $\mathcal{E}_{\mathsf{good}}$ (cf. Lemma 6) , we have*

$$\left\| \widetilde{\mathbf{\Sigma}}_l - \mathbf{\Sigma}_{U,l}^\star \right\| \lesssim \sqrt{\frac{\kappa^5 r \log^3(n+d)}{n}} \lambda_{\min}\left(\mathbf{\Sigma}_{U,l}^\star\right),$$

$$\max\left\{\lambda_{\max}\big(\widetilde{\mathbf{\Sigma}}_l\big), \lambda_{\max}\left(\mathbf{\Sigma}_{U,l}^\star\right)\right\} \lesssim \frac{1-p}{np\sigma_r^{\star 2}} \left\|\boldsymbol{U}_{l,\cdot}^\star \mathbf{\Sigma}^\star\right\|_2^2 + \frac{\omega_l^{\star 2}}{np\sigma_r^{\star 2}},$$

$$\min\left\{\lambda_{\min}\big(\widetilde{\mathbf{\Sigma}}_l\big), \lambda_{\min}\left(\mathbf{\Sigma}_{U,l}^\star\right)\right\} \gtrsim \frac{1-p}{np\sigma_1^{\star 2}} \left\|\boldsymbol{U}_{l,\cdot}^\star \mathbf{\Sigma}^\star\right\|_2^2 + \frac{\omega_l^{\star 2}}{np\sigma_1^{\star 2}}.$$

Proof. See Appendix E.1.2. □

Step 3: distributional guarantees for $(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^\star)_{l,\cdot}$. In view of (D.16), the row $\boldsymbol{Z}_{l,\cdot}$ is a superposition (in fact, linear combination) of n independent random sources. From the insights derived from variants of central limit theorems (e.g., Chen et al. (2010)), it is natural to conjecture that $\boldsymbol{Z}_{l,\cdot}$ might admit a reasonably tight Gaussian approximation. This can indeed be rigorized. With the above covariance matrices in place, we now move on to formalize a Gaussian approximation for $\boldsymbol{Z}_{l,\cdot}$, which further results in distributional guarantees for each row of $\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^\star$.

Lemma 9. *Suppose that*

$$\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{n+d}{np}} \lesssim \frac{1}{\kappa^{1/2} r^{3/4} \log^{3/2}(n+d)} \wedge \frac{\sqrt{1\wedge (d/n)}}{\kappa \mu^{1/2} r^{3/4} \log^{1/2}(n+d)},$$

$$np \gtrsim \kappa^4 \mu r^{5/2} \log^3(n+d), \quad n \gtrsim \kappa^5 r^2 \log^4(n+d), \quad d \gtrsim \kappa \mu r^{5/2} \log^2(n+d),$$

and

$$\left\|\boldsymbol{U}_{l,\cdot}^\star \mathbf{\Sigma}^\star\right\|_2 \gtrsim \left[\frac{\kappa \mu r^{5/4} \log^{5/2}(n+d)}{\sqrt{(n\wedge d)\,p}} + \frac{\kappa^{3/2}\mu^{3/2} r^{5/4} \log^{3/2}(n+d)}{\sqrt{np}\wedge\sqrt{d^2 p/n}} + \frac{\kappa^{1/2}\mu r^{5/4}\log^{3/2}(n+d)}{\sqrt{d}}\right]\sqrt{\frac{r}{d}}\sigma_1^\star.$$

Then it follows that

$$\sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left((\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^\star)_{l,\cdot}\in\mathcal{C}\right) - \mathbb{P}\left(\mathcal{N}\left(\mathbf{0},\mathbf{\Sigma}_{U,l}^\star\right)\in\mathcal{C}\right)\right| \lesssim \frac{1}{\sqrt{\log(n+d)}} = o(1),$$

where $\mathscr{C}^r$ is the set of all convex sets in $\mathbb{R}^r$.

Proof. See Appendix E.1.3. □

This lemma makes clear that each row of the estimation error $\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^\star$ is well approximated by a zero-mean Gaussian vector with covariance matrix specified by $\mathbf{\Sigma}_{U,l}^\star$. We have thus established Theorem 11.

D.3 Validity of confidence regions (Proof of Theorem 12)

With the above distributional guarantees in place, the remaining challenge for constructing confidence regions for $\boldsymbol{U}^\star$ lies in how to accurately estimate the covariance matrix $\mathbf{\Sigma}_{U,l}^\star$, which we study in this subsection. Our focal point is the plug-in estimator

$$\mathbf{\Sigma}_{U,l} \coloneqq \left(\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}\mathbf{\Sigma}\right\|_2^2 + \frac{\omega_l^2}{np}\right)(\mathbf{\Sigma})^{-2} + \frac{2(1-p)}{np}\boldsymbol{U}_{l,\cdot}^\top \boldsymbol{U}_{l,\cdot} \quad \text{where} \quad \omega_l^2 \coloneqq \frac{\sum_{j=1}^n y_{l,j}^2 \mathbb{1}_{(l,j)\in\Omega}}{\sum_{j=1}^n \mathbb{1}_{(l,j)\in\Omega}} - S_{l,l}. \tag{D.18}$$

Step 1: fine-grained estimation guarantees for $\boldsymbol{U}$ and $\boldsymbol{\Sigma}$. In order to justify the faithfulness of the plug-in estimator, a starting point is to show that the components used in this plug-in estimator (e.g., $\boldsymbol{U}$, $\boldsymbol{\Sigma}$) are all reliable estimates of the corresponding ground truth. This is demonstrated in the following lemma.

Lemma 10. *Assume that*

$$(n \wedge d)\, p \gg \kappa^2 \mu r \log^2(n+d) \qquad \text{and} \qquad \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{(n+d)\log(n+d)}{np}} \ll \frac{1}{\kappa}.$$

Then with probability exceeding $1 - O((n+d)^{-10})$, we have

$$\left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{R} - \boldsymbol{U}_{l,\cdot}^\star\right\|_{2,\infty} \lesssim \frac{1}{\sigma_r^\star}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\sqrt{\frac{r\log^2(n+d)}{np}} + \frac{\omega_l^\star}{\sigma_r^\star}\sqrt{\frac{r\log(n+d)}{np}} + \zeta_{\mathsf{2nd}}, \tag{D.19}$$

$$\left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}\boldsymbol{Q}^\top - \boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_{2,\infty} \lesssim \left(\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \omega_l^\star\right)\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sigma_r^\star\sqrt{\frac{r+\log(n+d)}{n}}\left\|\boldsymbol{U}_{l,\cdot}^\star\right\|_2 + \sigma_r^\star\zeta_{\mathsf{2nd}}, \tag{D.20}$$

$$\left\|\boldsymbol{R}\left(\boldsymbol{\Sigma}^\star\right)^{-2}\boldsymbol{R}^\top - \boldsymbol{\Sigma}^{-2}\right\| \lesssim \frac{1}{\sigma_r^{\star 2}}\sqrt{\frac{\kappa^2 r\log(n+d)}{n}} + \frac{1}{\sigma_r^{\star 2}}\zeta_{\mathsf{2nd}}\sqrt{\frac{\kappa d}{\mu r}}. \tag{D.21}$$

Here, ζ_{2nd} is defined in Lemma 7.

Proof. See Appendix E.2.1. □

Step 2: faithfulness of the plug-in estimator. With the fine-grained estimation guarantees established in Lemma 10, we can convert them to demonstrate the desired accuracy of our plug-in estimators, including both ω_l^2 and $\boldsymbol{\Sigma}_{U,l}$.

Lemma 11. *Instate the conditions in Lemma 10. In addition, assume that $np \gtrsim \kappa^2 r^2 \log^2(n+d)$. Then with probability exceeding $1 - O((n+d)^{-10})$, for each $i, j \in [d]$,*

$$\begin{aligned}\left|S_{i,j} - S_{i,j}^\star\right| \le &\left(\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sqrt{\frac{r+\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \frac{r\log^2(n+d)}{np}\omega_{\max}^2 \\ &+ \left(\sqrt{\frac{r\log^2(n+d)}{np}}\omega_{\max} + \zeta_{\mathsf{2nd}}\sigma_r^\star\right)\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\right) + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}.\end{aligned}$$

In addition, with probability exceeding $1 - O((n+d)^{-10})$, for each $l \in [d]$,

$$\begin{aligned}\left|\omega_l^2 - \omega_l^{\star 2}\right| \lesssim &\left(\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sqrt{\frac{r+\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2 + \sqrt{\frac{\log^2(n+d)}{np}}\omega_{\max}^2 \\ &+ \left(\sqrt{\frac{r\log^2(n+d)}{np}}\omega_{\max} + \zeta_{\mathsf{2nd}}\sigma_r^\star\right)\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}.\end{aligned}$$

Proof. See Appendix E.2.2. □

Lemma 12. *Instate the conditions in Theorem 11, Lemma 10 and Lemma 11. For any $\delta \in (0,1)$, we further assume that $np \gtrsim \delta^{-2}\kappa^4 r\log^2(n+d)$ and $\zeta_{\mathsf{2nd}}\sqrt{d} \lesssim \delta\sqrt{\mu r/\kappa^3}$. Then with probability at least $1 - O((n+d)^{-10})$, one has*

$$\left\|\boldsymbol{\Sigma}_{U,l} - \boldsymbol{R}\boldsymbol{\Sigma}_{U,l}^\star\boldsymbol{R}^\top\right\| \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right).$$

Proof. See Appendix E.2.3. □

Step 3: validity of the constructed confidence regions. Given our Gaussian approximation in Theorem 11 and the accuracy of the covariance estimator $\mathbf{\Sigma}_{U,l}$, it is natural to construct confidence regions by "pretending" that the estimation error follows a Gaussian distribution with covariance matrix $\mathbf{\Sigma}_{U,l}$. This straightforward procedure turns out to be statistically valid, as asserted by the following lemma.

Lemma 13. *Instate the conditions in Theorem 11, Lemma 10 and Lemma 11. Further assume that* $n \wedge d \gtrsim \kappa^{3/2} r^{7/4} \log^{5/2}(n+d)$, $np \gtrsim \kappa^5 r^{5/2} \log^5(n+d)$

$$(n \wedge d)\, p \gtrsim \kappa^{7/2} \mu r^{7/4} \log^{5/2}(n+d), \qquad \text{and} \qquad \frac{\omega_{\max}}{\sigma_r^\star} \sqrt{\frac{n+d}{np}} \lesssim \frac{1}{\kappa^{5/4} r^{3/8} \log^{5/4}(n+d)}.$$

Then we have

$$\mathbb{P}\left(\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{R}^\top \in \mathsf{CR}_{U,l}^{1-\alpha}\right) = 1 - \alpha + O\left(\frac{1}{\sqrt{\log(n+d)}}\right) = 1 - \alpha + o(1).$$

Proof. See Appendix E.2.4. □

This concludes the proof of Theorem 12.

D.4 Entrywise distributional characterization for $\boldsymbol{S}^\star$ (Proof of Theorem 13)

The preceding distributional guarantees for principal subspace in turn allow one to perform inference on the covariance matrix $\boldsymbol{S}^\star$ of the noiseless data vectors $\{\boldsymbol{x}_j\}_{1 \le j \le n}$. In this subsection, we demonstrate how to establish the advertised distributional characterization for the matrix $\boldsymbol{S}$ returned by Algorithm 1. Towards this end, we begin with the following decomposition

$$\boldsymbol{S} - \boldsymbol{S}^\star = \underbrace{\boldsymbol{S} - \boldsymbol{M}^\natural \boldsymbol{M}^{\natural\top}}_{=: \boldsymbol{W}} + \underbrace{\boldsymbol{M}^\natural \boldsymbol{M}^{\natural\top} - \boldsymbol{S}^\star}_{=: \boldsymbol{A}},$$

and we shall write $\boldsymbol{W} = [W_{i,j}]_{1 \le i,j \le d}$ and $\boldsymbol{A} = [A_{i,j}]$ from now on. In what follows, our proof consists of the following main steps:

1. Show that conditional on $\boldsymbol{F}$, each entry $W_{i,j}$ is approximately a zero-mean Gaussian, whose variance concentrates around some deterministic quantity $\widetilde{v}_{i,j}$.
2. Show that each entry $A_{i,j}$ is approximately Gaussian with mean zero and variance $\overline{v}_{i,j}$.
3. Utilize the (near) independence of $W_{i,j}$ and $A_{i,j}$ to demonstrate that $S_{i,j} - S_{i,j}^\star$ is approximately Gaussian with mean zero and variance $\widetilde{v}_{i,j} + \overline{v}_{i,j}$.

Step 1: first- and second-order approximation of $\boldsymbol{W}$. To begin with, Lemma 7 allows one to derive a reasonably accurate first- and second-order approximation for $\boldsymbol{W} = \boldsymbol{S} - \boldsymbol{M}^\natural \boldsymbol{M}^{\natural\top}$. This is formalized in the following lemma, providing an explicit form of this approximation and the goodness of the approximation.

Lemma 14. *Instate the assumptions in Lemma 7. Then one can write*

$$\boldsymbol{W} = \boldsymbol{S} - \boldsymbol{M}^\natural \boldsymbol{M}^{\natural\top} = \boldsymbol{X} + \boldsymbol{\Phi},$$

where

$$\boldsymbol{X} \coloneqq \boldsymbol{E} \boldsymbol{M}^{\natural\top} + \boldsymbol{M}^\natural \boldsymbol{E}^\top \tag{D.22}$$

and the residual matrix $\boldsymbol{\Phi}$ *satisfies: conditional on* $\boldsymbol{F}$ *and on the* $\sigma(\boldsymbol{F})$*-measurable event* $\mathcal{E}_{\mathsf{good}}$ *(see Lemma 6),*

$$\begin{aligned} |\Phi_{i,j}| &\lesssim \zeta_{\mathsf{2nd}} \sigma_1^{\star 2} \left(\left\| \boldsymbol{U}_{i,\cdot}^\star \right\|_2 + \left\| \boldsymbol{U}_{j,\cdot}^\star \right\|_2 \right) + \frac{\kappa r \log^2(n+d)}{np} \omega_{\max} \left(\left\| \boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 + \left\| \boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \right) \\ &\quad + \frac{\kappa r \log^2(n+d)}{np} \left\| \boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \left\| \boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 + \frac{\kappa r \log^2(n+d)}{np} \omega_{\max}^2 + \zeta_{\mathsf{2nd}}^2 \sigma_1^{\star 2} \end{aligned} \tag{D.23}$$

holds for any $i, j \in [d]$ *with probability exceeding* $1 - O((n+d)^{-10})$. *Here,* ζ_{2nd} *is defined in Lemma 7.*

Proof. See Appendix E.3.1. □

Step 2: computing the entrywise variance of our approximation. Observe that the (i,j)-th entry of the matrix $\boldsymbol{X}$ (see (D.22)) is given by

$$X_{i,j}=\left[\boldsymbol{E}\boldsymbol{M}^{\natural\top}+\boldsymbol{M}^{\natural}\boldsymbol{E}^{\top}\right]_{i,j}=\sum_{l=1}^{n}\left\{M_{j,l}^{\natural}E_{i,l}+M_{i,l}^{\natural}E_{j,l}\right\}. \tag{D.24}$$

It is straightforward to calculate the variance of $X_{i,j}$ conditional on $\boldsymbol{F}$: when $i\neq j$,

$$\mathsf{var}\left(X_{i,j}|\boldsymbol{F}\right)=\sum_{l=1}^{n}M_{j,l}^{\natural 2}\sigma_{i,l}^{2}+\sum_{l=1}^{n}M_{i,l}^{\natural 2}\sigma_{j,l}^{2}; \tag{D.25}$$

and when $i=j$, we have

$$\mathsf{var}\left(X_{i,i}|\boldsymbol{F}\right)=4\sum_{l=1}^{n}M_{i,l}^{\natural 2}\sigma_{i,l}^{2}.$$

The next lemma shows that the conditional variance $\mathsf{var}(X_{i,j}|\boldsymbol{F})$ concentrates around some deterministic quantity $\widetilde{v}_{i,j}$ defined as follows:

$$\widetilde{v}_{i,j}\coloneqq\frac{2\left(1-p\right)}{np}\left(S_{i,i}^{\star}S_{j,j}^{\star}+2S_{i,j}^{\star 2}\right)+\frac{1}{np}\left(\omega_{i}^{\star 2}S_{j,j}^{\star}+\omega_{j}^{\star 2}S_{i,i}^{\star}\right) \tag{D.26a}$$

for any $i\neq j$, and

$$\widetilde{v}_{i,i}\coloneqq\frac{12\left(1-p\right)}{np}S_{i,i}^{\star 2}+\frac{4}{np}\omega_{i}^{\star 2}S_{i,i}^{\star} \tag{D.26b}$$

for any $i\in[d]$.

Lemma 15. *Suppose that $n\gg\log^{3}(n+d)$, and recall the definition of $\widetilde{v}_{i,j}$ in (D.26). On the event $\mathcal{E}_{\mathsf{good}}$ (cf. Lemma 6), we have*

$$\mathsf{var}\left(X_{i,j}|\boldsymbol{F}\right)=\widetilde{v}_{i,j}+O\left(\sqrt{\frac{\log^{3}\left(n+d\right)}{n}}\right)\widetilde{v}_{i,j} \tag{D.27}$$

for any $i,j\in[d]$. In addition, for any $i,j\in[d]$ it holds that

$$\widetilde{v}_{i,j}\gtrsim\frac{1-p}{np}\left\Vert\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert_{2}^{2}\left\Vert\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert_{2}^{2}+\frac{1}{np}\omega_{\min}^{2}\left(\left\Vert\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert_{2}^{2}+\left\Vert\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert_{2}^{2}\right). \tag{D.28}$$

Proof. See Appendix E.3.2. □

Step 3: establishing approximate Gaussianity of $W_{i,j}$. We are now ready to invoke the Berry-Esseen Theorem to show that $W_{i,j}$ is approximately Gaussian with mean zero and variance $\widetilde{v}_{i,j}$, as stated in the next lemma.

Lemma 16. *Suppose that $np\gtrsim\kappa^{2}r^{2}\log^{5}(n+d)$, $d\gtrsim\kappa^{2}\mu r^{2}\log^{2}(n+d)$,*

$$\frac{\omega_{\max}}{\sigma_{r}^{\star}}\sqrt{\frac{n+d}{np}}\lesssim\frac{1}{\sqrt{\kappa^{2}r\log^{2}\left(n+d\right)}}\wedge\sqrt{\frac{1\wedge\left(d/n\right)}{\kappa^{3}\mu r\log\left(n+d\right)}},$$

$$\min\left\{\left\Vert\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert_{2},\left\Vert\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert_{2}\right\}\gtrsim\left[\frac{\kappa^{3/2}\mu r\log^{5/2}\left(n+d\right)}{\sqrt{\left(n\wedge d\right)p}}+\frac{\kappa^{2}\mu^{3/2}r\log^{3/2}\left(n+d\right)}{\sqrt{np}\wedge\sqrt{d^{2}p/n}}+\frac{\kappa\mu r\log^{3/2}\left(n+d\right)}{\sqrt{d}}\right]\sqrt{\frac{r}{d}}\sigma_{1}^{\star}$$

and

$$\left\Vert\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert_{2}+\left\Vert\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert_{2}\gtrsim\sqrt{\kappa r}\log^{5/2}\left(n+d\right)\frac{\omega_{\max}}{\sigma_{r}^{\star}}\sqrt{\frac{d}{np}}\cdot\sqrt{\frac{r}{d}}\sigma_{1}^{\star}.$$

Then we have

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\left(\widetilde{v}_{i,j}\right)^{-1/2}W_{i,j}\leq t\,\middle|\,\boldsymbol{F}\right)-\Phi\left(t\right)\right|\lesssim\frac{1}{\sqrt{\log\left(n+d\right)}}=o\left(1\right).$$

Proof. See Appendix E.3.3. □

An important observation that one should bear in mind is that: conditional on $\boldsymbol{F}$, the distribution of $W_{i,j}$ is approximately $\mathcal{N}(0, \widetilde{v}_{i,j})$, where $\widetilde{v}_{i,j}$ does not depend on $\boldsymbol{F}$. This suggests that $W_{i,j}$ is nearly independent of the σ-algebra $\sigma(\boldsymbol{F})$.

Step 4: establishing approximate Gaussianity of $A_{i,j}$. We now move on to the matrix $\boldsymbol{A}$. It follows from (D.2) that

$$A_{i,j} = \left(\boldsymbol{M}^{\natural}\boldsymbol{M}^{\natural\top} - \boldsymbol{S}^{\star}\right)_{i,j} = \frac{1}{n}\sum_{l=1}^{n}\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)\left(\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right) - \boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star 2}\boldsymbol{U}_{j,\cdot}^{\star\top}. \tag{D.29}$$

In view of the independence of $\{\boldsymbol{f}_l\}_{1\leq l\leq n}$, the variance of $A_{i,j}$ can be calculated as follows

$$\begin{aligned}
\overline{v}_{i,j} &\coloneqq \mathsf{var}\left(A_{i,j}\right) = \frac{1}{n}\mathsf{var}\left[\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_1\right)\left(\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_1\right)\right] \\
&= \frac{1}{n}\left[\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + 2\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star 2}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2\right] - \frac{1}{n}\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star 2}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2 \\
&= \frac{1}{n}\left[\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star 2}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2\right] = \frac{1}{n}\left(S_{i,i}^{\star}S_{j,j}^{\star} + S_{i,j}^{\star 2}\right).
\end{aligned} \tag{D.30}$$

With the assistance of the Berry-Esseen Theorem, the lemma below demonstrates that $A_{i,j}$ is approximately Gaussian.

Lemma 17. *It holds that*

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}\left(\left(\overline{v}_{i,j}\right)^{-1/2}A_{i,j}\leq z\right) - \Phi\left(z\right)\right| \lesssim \frac{1}{\sqrt{n}}.$$

Proof. See Appendix E.3.4. □

Note that $A_{i,j}$ is $\sigma(\boldsymbol{F})$-measurable. This fact taken collectively with the near independence between $W_{i,j}$ and $\sigma(\boldsymbol{F})$ implies that $W_{i,j}$ and $A_{i,j}$ are nearly statistically independent.

Step 5: distributional characterization of $S_{i,j} - S_{i,j}^{\star}$. The approximate Gaussianity of $W_{i,j}$ and $A_{i,j}$, as well as the near independence between them, leads to the conjecture that $S_{i,j} - S_{i,j}^{\star} = W_{i,j} + A_{i,j}$ is approximately distributed as $\mathcal{N}(0, v_{i,j}^{\star})$ with

$$v_{ij}^{\star} \coloneqq \widetilde{v}_{i,j} + \overline{v}_{i,j}.$$

The following lemma rigorizes this conjecture, which in turn concludes the proof of Theorem 13.

Lemma 18. *Instate the assumptions of Lemma 16, and suppose that $n \gtrsim \log(n+d)$. Then we have*

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\left(S_{i,j} - S_{i,j}^{\star}\right)/\sqrt{v_{i,j}^{\star}}\leq t\right) - \Phi\left(t\right)\right| \lesssim \frac{1}{\sqrt{\log\left(n+d\right)}} = o\left(1\right).$$

Proof. See Appendix E.3.5. □

D.5 Validity of confidence intervals (Proof of Theorem 14)

In order to construct valid confidence interval for $S_{i,j}^{\star}$ based on the distributional theory in Lemma 18, it remains to estimate the variance $v_{i,j}^{\star}$ in a sufficiently accurate manner. We propose to estimate $v_{i,j}^{\star}$ by the following plug-in estimator $v_{i,j}$:

$$\begin{aligned}
v_{i,j} &= \frac{2-p}{np}S_{i,i}S_{j,j} + \frac{4-3p}{np}S_{i,j}^2 + \frac{1}{np}\left(\omega_i^2 S_{j,j} + \omega_j^2 S_{i,i}\right), \qquad \text{if } i\neq j, \\
v_{i,i} &= \frac{12-9p}{np}S_{i,i}^2 + \frac{4}{np}\omega_i^2 S_{i,i}.
\end{aligned}$$

Step 1: faithfulness of the plug-in estimator. Armed with the fine-grained estimation guarantees in Lemma 10 and Lemma 11, we are ready to justify that $v_{i,j}$ is a reliable estimate of $v_{i,j}^{\star}$.

Lemma 19. *Instate the conditions in Lemma 16. For any $\delta \in (0,1)$, we further assume that $n \gtrsim \delta^{-2}\kappa^2 r \log(n+d)$, $np \gtrsim \delta^{-2}\kappa r \log^2(n+d)$, and*

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \gtrsim \delta^{-1}\omega_{\max}\sqrt{\frac{r\log^2(n+d)}{np}}.$$

Then with probability exceeding $1 - O((n+d)^{-10})$, one has

$$\left|v_{i,j}^{\star} - v_{i,j}\right| \lesssim \delta v_{i,j}^{\star}.$$

Proof. See Appendix E.4.1. □

Step 2: validity of the constructed confidence intervals. Equipped with the Gaussian approximation in Lemma 18 as well as the faithfulness of the plug-in estimator $v_{i,j}$ in Lemma 19, we can readily construct a valid confidence interval for $S_{i,j}^{\star}$.

Lemma 20. *Instate the conditions in Lemma 16. Further assume that $n \gtrsim \kappa^3 r \log^3(n+d)$. Then the confidence region $\mathsf{CI}_{i,j}^{1-\alpha}$ returned from Algorithm 4 satisfies*

$$\mathbb{P}\left(S_{i,j}^{\star} \in \mathsf{CI}_{i,j}^{1-\alpha}\right) = 1 - \alpha + O\left(\frac{1}{\sqrt{\log(n+d)}}\right) = 1 - \alpha + o(1).$$

Proof. See Appendix E.4.2. □

E Auxiliary lemmas: the SVD-based approach

E.1 Auxiliary lemmas for Theorem 11

E.1.1 Proof of Lemma 7

Recall from that $\boldsymbol{R} \in \mathcal{O}^{r\times r}$ is the rotation matrix that best aligns $\boldsymbol{U}$ and $\boldsymbol{U}^{\star}$, and $\boldsymbol{U}^{\star}\boldsymbol{Q} = \boldsymbol{U}^{\natural}$ for some $\boldsymbol{Q} \in \mathcal{O}^{r\times r}$. It has been shown in (D.6) that $\boldsymbol{R}\boldsymbol{Q} \in \mathcal{O}^{r\times r}$ is the rotation matrix that best aligns $\boldsymbol{U}$ and $\boldsymbol{U}^{\natural}$. Conditional on $\boldsymbol{F}$, if the assumptions of Theorem 9 are satisfied (which we shall check momentarily), then we can invoke Theorem 9 to obtain

$$\boldsymbol{U}\boldsymbol{R}\boldsymbol{Q} - \boldsymbol{U}^{\natural} = \boldsymbol{E}\boldsymbol{V}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1} + \boldsymbol{\Psi},$$

where the residual matrix $\boldsymbol{\Psi}$ satisfies

$$\mathbb{P}\left(\|\boldsymbol{\Psi}\|_{2,\infty} \lesssim \zeta_{\mathsf{2nd}}(\boldsymbol{F}) \,\Big|\, \boldsymbol{F}\right) \geq 1 - O\left((n+d)^{-10}\right).$$

Here, the quantity $\zeta_{\mathsf{2nd}}(\boldsymbol{F})$ is given by

$$\zeta_{\mathsf{2nd}}(\boldsymbol{F}) := \frac{\sigma^2\sqrt{(n+d)\,r}\log(n+d)}{\sigma_r^{\natural 2}} + \left(\frac{\kappa^{\natural}\sigma^2(n+d)}{\sigma_r^{\natural 2}} + \frac{\sigma\sqrt{r\log(n+d)}}{\sigma_r^{\natural}}\right)\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}.$$

Since $\boldsymbol{U}^{\star} = \boldsymbol{U}^{\natural}\boldsymbol{Q}^{\top}$, we can rewrite the decomposition as

$$\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star} = \underbrace{\boldsymbol{E}\boldsymbol{V}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top}}_{=:\boldsymbol{Z}} + \boldsymbol{\Psi}\boldsymbol{Q}^{\top}.$$

When the event $\mathcal{E}_{\mathsf{good}}$ happens, it is readily seen from (D.10), (D.11) and (D.14) that

$$\zeta_{\mathsf{2nd}}(\boldsymbol{F}) \lesssim \underbrace{\frac{\sigma_{\mathsf{ub}}^2\sqrt{(n+d)\,r}\log(n+d)}{\sigma_r^{\star 2}} + \left(\frac{\sqrt{\kappa}\sigma_{\mathsf{ub}}^2(n+d)}{\sigma_r^{\star 2}} + \frac{\sigma_{\mathsf{ub}}\sqrt{r\log(n+d)}}{\sigma_r^\star}\right)\|\boldsymbol{U}^\star\|_{2,\infty}}_{=:\zeta_{\mathsf{2nd}}},$$

thus indicating that

$$\mathbb{P}\left(\|\boldsymbol{\Psi}\|_{2,\infty}\mathbb{1}_{\mathcal{E}_{\mathsf{good}}} \lesssim \zeta_{\mathsf{2nd}} \,\big|\, \boldsymbol{F}\right) \geq 1 - O\left((n+d)^{-10}\right)$$

Given that $\mathcal{E}_{\mathsf{good}}$ is $\sigma(\boldsymbol{F})$-measurable and that

$$\|\boldsymbol{\Psi}\|_{2,\infty} = \left\|\boldsymbol{\Psi}\boldsymbol{Q}^\top\right\|_{2,\infty} = \|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star - \boldsymbol{Z}\|_{2,\infty},$$

we conclude that on the event $\mathcal{E}_{\mathsf{good}}$,

$$\mathbb{P}\left(\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star - \boldsymbol{Z}\|_{2,\infty} \lesssim \zeta_{\mathsf{2nd}} \,\big|\, \boldsymbol{F}\right) \geq 1 - O\left((n+d)^{-10}\right).$$

We are now left with verifying the assumption of Theorem 9, which requires

$$\sigma_{\mathsf{ub}}\sqrt{(n+d)\log(n+d)}\kappa^\natural\,\mathbb{1}_{\mathcal{E}_{\mathsf{good}}} \ll \sigma_r^\natural \qquad \text{and} \qquad B \lesssim \sigma_{\mathsf{ub}}\sqrt{\frac{\min\{n,d\}}{\mu^\natural\log(n+d)}}$$

to hold whenever $\mathcal{E}_{\mathsf{good}}$ happens. In view of (D.14), (D.15) and (D.10), we know that these conditions can be guaranteed as long as

$$\left(\sqrt{\frac{\mu r\log(n+d)}{ndp}}\sigma_1^\star + \frac{\omega_{\max}}{\sqrt{np}}\right)\sqrt{(n+d)\,\kappa\log(n+d)} \ll \sigma_r^\star$$

and

$$\sqrt{\frac{\mu r\log(n+d)}{nd}}\sigma_1^\star + \frac{\omega_{\max}}{p}\sqrt{\frac{\log(n+d)}{n}} \lesssim \left(\frac{\sqrt{\mu r\log(n+d)}}{\sqrt{ndp}}\sigma_1^\star + \frac{\omega_{\max}}{\sqrt{np}}\right)\sqrt{\frac{\min\{n,d\}}{\mu\log^2(n+d)}},$$

which clearly hold true under our assumptions

$$\min\{n,d\}\,p \gg \kappa^2\mu r\log^2(n+d) \qquad \text{and} \qquad \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{(n+d)\log(n+d)}{np}} \ll \frac{1}{\kappa}.$$

E.1.2 Proof of Lemma 8

Throughout this section, it is assumed that the event $\mathcal{E}_{\mathsf{good}}$ happens. Recall that

$$\widetilde{\boldsymbol{\Sigma}}_l = \boldsymbol{Q}(\boldsymbol{\Sigma}^\natural)^{-1}\boldsymbol{V}^{\natural\top}\mathsf{diag}\left\{\sigma_{l,1}^2,\dots,\sigma_{l,n}^2\right\}\boldsymbol{V}^\natural(\boldsymbol{\Sigma}^\natural)^{-1}\boldsymbol{Q}^\top.$$

The matrices in the middle part satisfy

$$\boldsymbol{V}^{\natural\top}\mathsf{diag}\left\{\sigma_{l,1}^2,\dots,\sigma_{l,n}^2\right\}\boldsymbol{V}^\natural = \frac{1}{n}\boldsymbol{J}^\top\boldsymbol{F}\mathsf{diag}\left\{\sigma_{l,1}^2,\dots,\sigma_{l,n}^2\right\}\boldsymbol{F}^\top\boldsymbol{J} = \boldsymbol{J}^\top\left(\frac{1}{n}\sum_{j=1}^n\sigma_{l,j}^2\boldsymbol{f}_j\boldsymbol{f}_j^\top\right)\boldsymbol{J},$$

where we have made use of the identity (D.3). To control the term within the parentheses of the above identity, we can make use of the variance calculation in (D.8) to obtain

$$\frac{1}{n}\sum_{j=1}^n\sigma_{l,j}^2\boldsymbol{f}_j\boldsymbol{f}_j^\top = \frac{1-p}{n^2p}\sum_{j=1}^n\left(\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\boldsymbol{f}_j\right)^2\boldsymbol{f}_j\boldsymbol{f}_j^\top + \frac{\omega_l^{\star 2}}{n^2p}\sum_{j=1}^n\boldsymbol{f}_j\boldsymbol{f}_j^\top.$$

In addition, it is seen from (E.61) and (E.52) that

$$\left\|\frac{1}{n}\sum_{j=1}^{n}\left(\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_{j}\right)^{2}\boldsymbol{f}_{j}\boldsymbol{f}_{j}^{\top}-\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\boldsymbol{I}_{r}-2\boldsymbol{\Sigma}^{\star}\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|\lesssim\sqrt{\frac{r\log^{3}\left(n+d\right)}{n}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}$$

$$\text{and}\qquad\left\|\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{f}_{i}\boldsymbol{f}_{i}^{\top}-\boldsymbol{I}_{r}\right\|=\left\|\frac{1}{n}\boldsymbol{F}\boldsymbol{F}^{\top}-\boldsymbol{I}_{r}\right\|\lesssim\sqrt{\frac{r+\log\left(n+d\right)}{n}}.$$

These bounds taken together allow us to express

$$\frac{1}{n}\sum_{j=1}^{n}\sigma_{l,j}^{2}\boldsymbol{f}_{j}\boldsymbol{f}_{j}^{\top}=\frac{1-p}{np}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\boldsymbol{I}_{r}+2\boldsymbol{\Sigma}^{\star}\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right)+\frac{\omega_{l}^{\star2}}{np}\boldsymbol{I}_{r}+\boldsymbol{R}_{l,1}$$

for some residual matrix $\boldsymbol{R}_{l,1}$ satisfying

$$\left\|\boldsymbol{R}_{l,1}\right\|\lesssim\frac{1-p}{np}\sqrt{\frac{r\log^{3}\left(n+d\right)}{n}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}+\frac{\omega_{l}^{\star2}}{np}\sqrt{\frac{r+\log\left(n+d\right)}{n}}.$$

Putting the above pieces together, we arrive at

$$\begin{aligned}\widetilde{\boldsymbol{\Sigma}}_{l}&=\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{J}^{\top}\left(\frac{1}{n}\sum_{j=1}^{n}\sigma_{l,j}^{2}\boldsymbol{f}_{j}^{\star}\boldsymbol{f}_{j}^{\star\top}\right)\boldsymbol{J}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top}\\&=\underbrace{\left(\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}+\frac{\omega_{l}^{\star2}}{np}\right)\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{J}^{\top}\boldsymbol{J}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top}}_{=:\boldsymbol{\Sigma}_{l,1}}\\&\quad+\underbrace{\frac{2\left(1-p\right)}{np}\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{J}^{\top}\boldsymbol{\Sigma}^{\star}\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{J}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top}}_{=:\boldsymbol{\Sigma}_{l,2}}+\boldsymbol{R}_{l,2}\end{aligned}$$

for some residual matrix

$$\boldsymbol{R}_{l,2}=\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{J}^{\top}\boldsymbol{R}_{l,1}\boldsymbol{J}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top}.$$

This motivates us to look at $\boldsymbol{\Sigma}_{l,1}$ and $\boldsymbol{\Sigma}_{l,2}$ separately.

- Regarding the matrix $\boldsymbol{\Sigma}_{l,1}$, we make the observation that

$$\begin{aligned}\left\|(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{J}^{\top}\boldsymbol{J}(\boldsymbol{\Sigma}^{\natural})^{-1}-(\boldsymbol{\Sigma}^{\natural})^{-2}\right\|&\overset{\text{(i)}}{=}\left\|(\boldsymbol{\Sigma}^{\natural})^{-1}\left(\boldsymbol{J}^{\top}\boldsymbol{J}-\boldsymbol{Q}^{\top}\boldsymbol{Q}\right)(\boldsymbol{\Sigma}^{\natural})^{-1}\right\|\\&\leq\left\|(\boldsymbol{\Sigma}^{\natural})^{-1}\left(\boldsymbol{J}-\boldsymbol{Q}\right)^{\top}\boldsymbol{J}(\boldsymbol{\Sigma}^{\natural})^{-1}\right\|+\left\|(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top}\left(\boldsymbol{J}-\boldsymbol{Q}\right)(\boldsymbol{\Sigma}^{\natural})^{-1}\right\|\\&\overset{\text{(ii)}}{\lesssim}\frac{1}{\sigma_{r}^{\star2}}\left\|\boldsymbol{J}-\boldsymbol{Q}\right\|\left(\left\|\boldsymbol{Q}\right\|+\left\|\boldsymbol{J}\right\|\right)\overset{\text{(iii)}}{\lesssim}\frac{\kappa}{\sigma_{r}^{\star2}}\sqrt{\frac{r+\log\left(n+d\right)}{n}}.\end{aligned}\tag{E.1}$$

  Here, (i) comes from the fact that $\boldsymbol{Q}$ is a orthonormal matrix, (ii) follows from the property (D.10), whereas (iii) utilizes the properties (E.58) and (E.60) in Appendix E.5. In addition, from (E.58) and (D.10) we know that

$$\begin{aligned}\left\|\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-1}-\left(\boldsymbol{\Sigma}^{\star}\right)^{-1}\boldsymbol{Q}\right\|&\leq\left\|\left(\boldsymbol{\Sigma}^{\star}\right)^{-1}\left(\boldsymbol{\Sigma}^{\star}\boldsymbol{Q}-\boldsymbol{Q}\boldsymbol{\Sigma}^{\natural}\right)(\boldsymbol{\Sigma}^{\natural})^{-1}\right\|\lesssim\frac{1}{\sigma_{r}^{\star2}}\left\|\boldsymbol{\Sigma}^{\star}\boldsymbol{Q}-\boldsymbol{Q}\boldsymbol{\Sigma}^{\natural}\right\|\\&\lesssim\frac{\kappa}{\sigma_{r}^{\star}}\sqrt{\frac{r+\log\left(n+d\right)}{n}}.\end{aligned}\tag{E.2}$$

  This immediately leads to

$$\left\|\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-2}\boldsymbol{Q}^{\top}-\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\right\|=\left\|\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-1}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top}-\left(\boldsymbol{\Sigma}^{\star}\right)^{-1}\boldsymbol{Q}\boldsymbol{Q}^{\top}\left(\boldsymbol{\Sigma}^{\star}\right)^{-1}\right\|$$

$$
\begin{aligned}
&\leq \left\| \left[\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-1} - (\boldsymbol{\Sigma}^{\star})^{-1} \boldsymbol{Q} \right] (\boldsymbol{\Sigma}^{\natural})^{-1} \boldsymbol{Q}^{\top} \right\| + \left\| (\boldsymbol{\Sigma}^{\star})^{-1} \boldsymbol{Q} \left[\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-1} - (\boldsymbol{\Sigma}^{\star})^{-1} \boldsymbol{Q} \right]^{\top} \right\| \\
&\leq \left(\|(\boldsymbol{\Sigma}^{\natural})^{-1}\| + \|(\boldsymbol{\Sigma}^{\star})^{-1}\| \right) \|\boldsymbol{Q}\| \left\| \boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-1} - (\boldsymbol{\Sigma}^{\star})^{-1} \boldsymbol{Q} \right\| \\
&\lesssim \frac{\kappa}{\sigma_r^{\star 2}} \sqrt{\frac{r + \log(n+d)}{n}}, \qquad \text{(E.3)}
\end{aligned}
$$

where we have again used (D.10). Taking (E.1) and (E.3) collectively gives

$$
\begin{aligned}
&\left\| \boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-1} \boldsymbol{J}^{\top} \boldsymbol{J} (\boldsymbol{\Sigma}^{\natural})^{-1} \boldsymbol{Q}^{\top} - (\boldsymbol{\Sigma}^{\star})^{-2} \right\| \\
&\quad \leq \left\| \boldsymbol{Q} \left[(\boldsymbol{\Sigma}^{\natural})^{-1} \boldsymbol{J}^{\top} \boldsymbol{J} (\boldsymbol{\Sigma}^{\natural})^{-1} - (\boldsymbol{\Sigma}^{\natural})^{-2} \right] \boldsymbol{Q}^{\top} \right\| + \left\| \boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-2} \boldsymbol{Q}^{\top} - (\boldsymbol{\Sigma}^{\star})^{-2} \right\| \\
&\quad \leq \left\| (\boldsymbol{\Sigma}^{\natural})^{-1} \boldsymbol{J}^{\top} \boldsymbol{J} (\boldsymbol{\Sigma}^{\natural})^{-1} - (\boldsymbol{\Sigma}^{\natural})^{-2} \right\| + \left\| \boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-2} \boldsymbol{Q}^{\top} - (\boldsymbol{\Sigma}^{\star})^{-2} \right\| \\
&\quad \lesssim \frac{\kappa}{\sigma_r^{\star 2}} \sqrt{\frac{r + \log(n+d)}{n}}.
\end{aligned}
$$

Substitution into the definition of $\boldsymbol{\Sigma}_{l,1}$ allows us to conclude that

$$
\left\| \boldsymbol{\Sigma}_{l,1} - \left(\frac{1-p}{np} \|\boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star}\|_2^2 + \frac{\sigma_l^2}{np} \right) (\boldsymbol{\Sigma}^{\star})^{-2} \right\| \lesssim \left(\frac{1-p}{np} \|\boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star}\|_2^2 + \frac{\omega_l^{\star 2}}{np} \right) \frac{\kappa}{\sigma_r^{\star 2}} \sqrt{\frac{r + \log(n+d)}{n}}. \qquad \text{(E.4)}
$$

- When it comes to the remaining term $\boldsymbol{\Sigma}_{l,2}$, we first notice that

$$
\begin{aligned}
&\left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \boldsymbol{J} (\boldsymbol{\Sigma}^{\natural})^{-1} - \boldsymbol{U}_{l,\cdot}^{\natural} \right\|_2 \leq \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \boldsymbol{Q} (\boldsymbol{\Sigma}^{\natural})^{-1} - \boldsymbol{U}_{l,\cdot}^{\natural} \right\|_2 + \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} (\boldsymbol{J} - \boldsymbol{Q}) (\boldsymbol{\Sigma}^{\natural})^{-1} \right\|_2 \\
&\quad \leq \left\| \boldsymbol{U}_{l,\cdot}^{\star} \left(\boldsymbol{\Sigma}^{\star} \boldsymbol{Q} - \boldsymbol{Q} \boldsymbol{\Sigma}^{\natural} \right) (\boldsymbol{\Sigma}^{\natural})^{-1} \right\|_2 + \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{Q} - \boldsymbol{U}_{l,\cdot}^{\natural} \right\|_2 + \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} (\boldsymbol{J} - \boldsymbol{Q}) (\boldsymbol{\Sigma}^{\natural})^{-1} \right\|_2 \\
&\quad \lesssim \frac{1}{\sigma_r^{\star}} \|\boldsymbol{U}_{l,\cdot}^{\star}\|_2 \left\| \boldsymbol{\Sigma}^{\star} \boldsymbol{Q} - \boldsymbol{Q} \boldsymbol{\Sigma}^{\natural} \right\| + \frac{\sigma_1^{\star}}{\sigma_r^{\star}} \|\boldsymbol{U}_{l,\cdot}^{\star}\|_2 \|\boldsymbol{J} - \boldsymbol{Q}\| \\
&\quad \lesssim \kappa^{3/2} \sqrt{\frac{r + \log(n+d)}{n}} \|\boldsymbol{U}_{l,\cdot}^{\star}\|_2,
\end{aligned}
$$

where the penultimate line uses (D.10) and fact that $\boldsymbol{U}^{\star} \boldsymbol{Q} = \boldsymbol{U}^{\natural}$, and the last line relies on the property (E.58) in Appendix E.5. An immediate consequence is that

$$
\left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \boldsymbol{J} (\boldsymbol{\Sigma}^{\natural})^{-1} \right\|_2 \leq \left\| \boldsymbol{U}_{l,\cdot}^{\natural} \right\|_2 + \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \boldsymbol{J} (\boldsymbol{\Sigma}^{\natural})^{-1} - \boldsymbol{U}_{l,\cdot}^{\natural} \right\|_2 \lesssim \|\boldsymbol{U}_{l,\cdot}^{\star}\|_2,
$$

provided that $n \gg \kappa^3 r + \kappa^3 \log(n+d)$. This combined with the fact $\boldsymbol{U}^{\star} \boldsymbol{Q} = \boldsymbol{U}^{\natural}$ immediately yields

$$
\begin{aligned}
&\left\| \boldsymbol{Q} (\boldsymbol{\Sigma}^{\natural})^{-1} \boldsymbol{J}^{\top} \boldsymbol{\Sigma}^{\star} \boldsymbol{U}_{l,\cdot}^{\star\top} \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \boldsymbol{J} (\boldsymbol{\Sigma}^{\natural})^{-1} \boldsymbol{Q}^{\top} - \boldsymbol{U}_{l,\cdot}^{\star\top} \boldsymbol{U}_{l,\cdot}^{\star} \right\| \\
&\quad = \left\| (\boldsymbol{\Sigma}^{\natural})^{-1} \boldsymbol{J}^{\top} \boldsymbol{\Sigma}^{\star} \boldsymbol{U}_{l,\cdot}^{\star\top} \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \boldsymbol{J} (\boldsymbol{\Sigma}^{\natural})^{-1} - \boldsymbol{U}_{l,\cdot}^{\natural\top} \boldsymbol{U}_{l,\cdot}^{\natural} \right\| \\
&\quad \leq \left\| (\boldsymbol{\Sigma}^{\natural})^{-1} \boldsymbol{J}^{\top} \boldsymbol{\Sigma}^{\star} \boldsymbol{U}_{l,\cdot}^{\star\top} \left(\boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \boldsymbol{J} (\boldsymbol{\Sigma}^{\natural})^{-1} - \boldsymbol{U}_{l,\cdot}^{\natural} \right) \right\| + \left\| \left(\boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \boldsymbol{J} (\boldsymbol{\Sigma}^{\natural})^{-1} - \boldsymbol{U}_{l,\cdot}^{\natural} \right)^{\top} \boldsymbol{U}_{l,\cdot}^{\natural} \right\| \\
&\quad \leq \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \boldsymbol{J} (\boldsymbol{\Sigma}^{\natural})^{-1} - \boldsymbol{U}_{l,\cdot}^{\natural} \right\|_2 \left(\left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \boldsymbol{J} (\boldsymbol{\Sigma}^{\natural})^{-1} \right\|_2 + \left\| \boldsymbol{U}_{l,\cdot}^{\natural} \right\|_2 \right) \\
&\quad \lesssim \kappa^{3/2} \sqrt{\frac{r + \log(n+d)}{n}} \|\boldsymbol{U}_{l,\cdot}^{\star}\|_2^2,
\end{aligned}
$$

and as a result,

$$
\left\| \boldsymbol{\Sigma}_{l,2} - \frac{2(1-p)}{np} \boldsymbol{U}_{l,\cdot}^{\star\top} \boldsymbol{U}_{l,\cdot}^{\star} \right\| \lesssim \frac{1-p}{np} \kappa^{3/2} \sqrt{\frac{r + \log(n+d)}{n}} \|\boldsymbol{U}_{l,\cdot}^{\star}\|_2^2. \qquad \text{(E.5)}
$$

Combining the above bounds (E.4) and (E.5), we can demonstrate that

$$\widetilde{\boldsymbol{\Sigma}}_l = \boldsymbol{\Sigma}_{l,1} + \boldsymbol{\Sigma}_{l,2} + \boldsymbol{R}_{l,2} = \underbrace{\left(\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{\omega_l^{\star 2}}{np}\right)\left(\boldsymbol{\Sigma}^{\star}\right)^{-2} + \frac{2\left(1-p\right)}{np}\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^{\star}}_{=: \boldsymbol{\Sigma}_{U,l}^{\star}} + \boldsymbol{R}_l$$

holds for some residual matrix $\boldsymbol{R}_l$ satisfying

$$\begin{aligned}
\|\boldsymbol{R}_l\| &\leq \|\boldsymbol{R}_{l,2}\| + \left\|\boldsymbol{\Sigma}_{l,1} - \left(\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{\omega_l^{\star 2}}{np}\right)\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\right\| + \left\|\boldsymbol{\Sigma}_{l,2} - \frac{2\left(1-p\right)}{np}\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^{\star}\right\| \\
&\overset{\text{(i)}}{\lesssim} \frac{1}{\sigma_r^{\star 2}}\left(\frac{1-p}{np}\sqrt{\frac{r\log^3\left(n+d\right)}{n}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{\omega_l^{\star 2}}{np}\sqrt{\frac{r+\log\left(n+d\right)}{n}}\right) \\
&\quad + \left(\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{\omega_l^{\star 2}}{np}\right)\frac{\kappa}{\sigma_r^{\star 2}}\sqrt{\frac{r+\log\left(n+d\right)}{n}} + \frac{1-p}{np}\kappa^{3/2}\sqrt{\frac{r+\log\left(n+d\right)}{n}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2^2 \\
&\lesssim \frac{1-p}{np\sigma_r^{\star 2}}\left(\sqrt{\frac{r\log^3\left(n+d\right)}{n}} + \kappa^{3/2}\sqrt{\frac{r+\log\left(n+d\right)}{n}}\right)\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{\omega_l^{\star 2}}{np\sigma_r^{\star 2}}\kappa\sqrt{\frac{r+\log\left(n+d\right)}{n}}.
\end{aligned}$$

Here, (i) has made use of (D.10) and the property (E.60) in Appendix E.5.

It then boils down to proving that $\boldsymbol{\Sigma}_{U,l}^{\star}$ and $\widetilde{\boldsymbol{\Sigma}}_l$ are both well-conditioned. Towards this, we first observe that

$$\left(\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{\omega_l^{\star 2}}{np}\right)\left(\boldsymbol{\Sigma}^{\star}\right)^{-2} + \frac{2\left(1-p\right)}{np}\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^{\star} \succeq \boldsymbol{\Sigma}_{U,l}^{\star} \succeq \left(\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{\omega_l^{\star 2}}{np}\right)\left(\boldsymbol{\Sigma}^{\star}\right)^{-2},$$

and as a result,

$$\lambda_{\max}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) \leq \frac{3\left(1-p\right)}{np\sigma_r^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{\omega_l^{\star 2}}{np\sigma_r^{\star 2}}, \tag{E.6a}$$

$$\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) \geq \frac{1-p}{np\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{\omega_l^{\star 2}}{np\sigma_1^{\star 2}}. \tag{E.6b}$$

This immediately implies that the condition number of $\boldsymbol{\Sigma}_{U,l}^{\star}$ is at most 3κ. In addition, it follows from the preceding bounds that

$$\frac{\|\boldsymbol{R}_l\|}{\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)} \lesssim \sqrt{\frac{\kappa^2 r\log^3\left(n+d\right) + \kappa^5 r + \kappa^5\log\left(n+d\right)}{n}} \asymp \sqrt{\frac{\kappa^5 r\log^3\left(n+d\right)}{n}}.$$

This means that $\|\boldsymbol{R}_l\| \ll \lambda_{\min}(\boldsymbol{\Sigma}_{U,l}^{\star})$ holds as long as $n \gg \kappa^5 r\log^3(n+d)$, and as a consequence,

$$\lambda_{\min}\big(\widetilde{\boldsymbol{\Sigma}}_l\big) \in \left[\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) - \|\boldsymbol{R}_l\|, \lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) + \|\boldsymbol{R}_l\|\right] \quad \Longrightarrow \quad \lambda_{\min}\big(\widetilde{\boldsymbol{\Sigma}}_l\big) \asymp \lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right).$$

Therefore, one can conclude that

$$\left\|\widetilde{\boldsymbol{\Sigma}}_l - \boldsymbol{\Sigma}_{U,l}^{\star}\right\| = \|\boldsymbol{R}_l\| \lesssim \sqrt{\frac{\kappa^5 r\log^3\left(n+d\right)}{n}}\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right).$$

An immediate consequence is that the condition number of $\widetilde{\boldsymbol{\Sigma}}_l$ is at most 4κ provided that $n \gg \kappa^5 r\log^3(n+d)$. It then follows from Weyl's inequality and (E.6) that

$$\lambda_{\max}\big(\widetilde{\boldsymbol{\Sigma}}_l\big) \lesssim \frac{1-p}{np\sigma_r^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{\omega_l^{\star 2}}{np\sigma_r^{\star 2}},$$

$$\lambda_{\min}\big(\widetilde{\boldsymbol{\Sigma}}_l\big) \gtrsim \frac{1-p}{np\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{\omega_l^{\star 2}}{np\sigma_1^{\star 2}}.$$

E.1.3 Proof of Lemma 9

Step 1: Gaussian approximation of $\boldsymbol{Z}_{l,\cdot}$ using the Berry-Esseen Theorem. Let us write

$$\boldsymbol{Z}_{l,\cdot} = \sum_{j=1}^{n} \boldsymbol{Y}_j, \qquad \text{where} \quad \boldsymbol{Y}_j = E_{l,j}\boldsymbol{V}_{j,\cdot}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top}.$$

The Berry-Esseen theorem (see Theorem 20) tells us that

$$\sup_{\mathcal{C}\in\mathscr{C}^r} \left|\mathbb{P}\left(\boldsymbol{Z}_{l,\cdot} \in \mathcal{C} \middle| \boldsymbol{F}\right) - \mathbb{P}\left(\mathcal{N}(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}}_l) \in \mathcal{C}\middle|\boldsymbol{F}\right)\right| \lesssim r^{1/4}\gamma\left(\boldsymbol{F}\right), \tag{E.7}$$

where we define

$$\gamma\left(\boldsymbol{F}\right) \coloneqq \sum_{i=1}^{n} \mathbb{E}\left[\left\|\boldsymbol{Y}_j\widetilde{\boldsymbol{\Sigma}}_l^{-1/2}\right\|_2^3\middle|\boldsymbol{F}\right].$$

It is straightforward to bound

$$\max_{1\leq j\leq n}\left\|\boldsymbol{Y}_j\right\|_2 \leq \max_{l,j}\left|E_{l,j}\right| \cdot \max_j \left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} \cdot \left\|(\boldsymbol{\Sigma}^{\natural})^{-1}\right\|\left\|\boldsymbol{Q}\right\| \lesssim \frac{1}{\sigma_r^{\natural}}B\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty},$$

where B is defined in (D.15) in Appendix E.5. As a result, one can bound $\gamma\left(\boldsymbol{F}\right)$ as follows

$$\begin{aligned}\gamma\left(\boldsymbol{F}\right) &\leq \lambda_{\min}^{-3/2}(\widetilde{\boldsymbol{\Sigma}}_l)\sum_{j=1}^{n}\mathbb{E}\left[\left\|\boldsymbol{Y}_j\right\|_2^3\middle|\boldsymbol{F}\right] \leq \lambda_{\min}^{-3/2}(\widetilde{\boldsymbol{\Sigma}}_l)\max_{1\leq j\leq n}\left\|\boldsymbol{Y}_j\right\|_2\sum_{j=1}^{n}\mathbb{E}\left[\left\|\boldsymbol{Y}_j\right\|_2^2\middle|\boldsymbol{F}\right] \\ &= \lambda_{\min}^{-3/2}(\widetilde{\boldsymbol{\Sigma}}_l)\mathsf{tr}(\widetilde{\boldsymbol{\Sigma}}_l)\max_{j\in[n]}\left\|\boldsymbol{Y}_j\right\|_2,\end{aligned}$$

where the last identity uses the fact

$$\begin{aligned}\sum_{j=1}^{n}\mathbb{E}\left[\left\|\boldsymbol{Y}_j\right\|_2^2\middle|\boldsymbol{F}\right] &= \sum_{j=1}^{n}\mathbb{E}\left[\boldsymbol{Y}_j\boldsymbol{Y}_j^{\top}\middle|\boldsymbol{F}\right] = \sum_{j=1}^{n}\mathbb{E}\left[\mathsf{tr}\left(\boldsymbol{Y}_j^{\top}\boldsymbol{Y}_j\right)\middle|\boldsymbol{F}\right] = \mathsf{tr}\left[\sum_{j=1}^{n}\mathbb{E}\left(\boldsymbol{Y}_j^{\top}\boldsymbol{Y}_j\middle|\boldsymbol{F}\right)\right] \\ &= \mathsf{tr}\left(\mathbb{E}\left[\boldsymbol{Z}_{l,\cdot}^{\top}\boldsymbol{Z}_{l,\cdot}\middle|\boldsymbol{F}\right]\right) = \mathsf{tr}(\widetilde{\boldsymbol{\Sigma}}_l).\end{aligned} \tag{E.8}$$

Substitution into (E.7) then yields

$$\sup_{\mathcal{C}\in\mathscr{C}^r} \left|\mathbb{P}\left(\boldsymbol{Z}_{l,\cdot} \in \mathcal{C} \middle| \boldsymbol{F}\right) - \mathbb{P}\left(\mathcal{N}(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}}_l) \in \mathcal{C}\middle|\boldsymbol{F}\right)\right| \lesssim \frac{r^{1/4}\mathsf{tr}(\widetilde{\boldsymbol{\Sigma}}_l)}{\lambda_{\min}^{3/2}(\widetilde{\boldsymbol{\Sigma}}_l)}\frac{B}{\sigma_r^{\natural}}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty}$$

When the event $\mathcal{E}_{\mathsf{good}}$ occurs, we know from Lemma 8 that the condition number of $\widetilde{\boldsymbol{\Sigma}}_l$ is bounded above by $O(\kappa)$. This implies that

$$\frac{\mathsf{tr}(\widetilde{\boldsymbol{\Sigma}}_l)}{\lambda_{\min}^{3/2}(\widetilde{\boldsymbol{\Sigma}}_l)} \leq \frac{r\|\widetilde{\boldsymbol{\Sigma}}_l\|}{\lambda_{\min}^{3/2}(\widetilde{\boldsymbol{\Sigma}}_l)} \lesssim \frac{\kappa r}{\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l)},$$

and as a consequence,

$$\begin{aligned}\frac{r^{1/4}\mathsf{tr}(\widetilde{\boldsymbol{\Sigma}}_l)}{\lambda_{\min}^{3/2}(\widetilde{\boldsymbol{\Sigma}}_l)}\frac{B}{\sigma_r^{\natural}}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} &\lesssim \frac{\kappa r^{5/4}}{\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l)}\frac{B}{\sigma_r^{\natural}}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} \\ &\lesssim \frac{\kappa r^{5/4}}{\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l)}\frac{1}{\sigma_r^{\star}}\left(\frac{1}{p}\sqrt{\frac{\mu r\log(n+d)}{nd}}\sigma_1^{\star} + \frac{\omega_{\max}}{p}\sqrt{\frac{\log(n+d)}{n}}\right)\sqrt{\frac{\mu+\log(n+d)}{n}} \\ &\lesssim \underbrace{\frac{1}{\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l)}\frac{\kappa^{3/2}\mu r^{7/4}\log(n+d)}{\sqrt{n^2dp^2}}}_{=:\alpha_1} + \underbrace{\frac{1}{\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l)}\frac{\omega_{\max}}{\sigma_r^{\star}}\frac{\kappa\mu^{1/2}r^{5/4}\log(n+d)}{np}}_{=:\alpha_2}.\end{aligned}$$

Here, the penultimate relation uses (D.15), (D.10) and (D.12). In order to bound α_1 and α_2, we note that Lemma 8 tells us that

$$\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l) \gtrsim \frac{1}{\sqrt{np}\sigma_1^\star}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \frac{\omega_l^\star}{\sqrt{np}\sigma_1^\star}.$$

Recalling our assumption $\omega_l^\star \asymp \omega_{\max}$, we can obtain

$$\begin{aligned}
\alpha_1 &\lesssim \frac{\kappa^{3/2}\mu r^{7/4}\log(n+d)}{\sqrt{ndp}}\frac{\sigma_1^\star}{\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2} \lesssim \frac{1}{\sqrt{\log(n+d)}},\\
\alpha_2 &\lesssim \frac{\kappa^{3/2}\mu^{1/2}r^{5/4}\log(n+d)}{\sqrt{np}} \lesssim \frac{1}{\sqrt{\log(n+d)}},
\end{aligned}$$

provided that $np \gtrsim \kappa^3\mu r^{5/2}\log^3(n+d)$ and

$$\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \gtrsim \frac{\kappa^{3/2}\mu r^{7/4}\log^{3/2}(n+d)}{\sqrt{ndp}}\sigma_1^\star.$$

These results taken collectively guarantee that

$$\sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left(\boldsymbol{Z}_{l,\cdot}\in\mathcal{C}\mid\boldsymbol{F}\right) - \mathbb{P}\left(\mathcal{N}(\boldsymbol{0},\widetilde{\boldsymbol{\Sigma}}_l)\in\mathcal{C}\middle|\boldsymbol{F}\right)\right| \lesssim \alpha_1+\alpha_2 \lesssim \frac{1}{\sqrt{\log(n+d)}}. \tag{E.9}$$

Step 2: bounding the total-variation distance between Gaussian distributions. It has been shown in Lemma 8 that

$$\left\|\widetilde{\boldsymbol{\Sigma}}_l - \boldsymbol{\Sigma}_{U,l}^\star\right\| \lesssim \sqrt{\frac{\kappa^5 r\log^3(n+d)}{n}}\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right) \tag{E.10}$$

holds on the event $\mathcal{E}_{\mathsf{good}}$. It is also assumed that $\boldsymbol{\Sigma}_{U,l}^\star$ is non-singular. Therefore, conditional on $\boldsymbol{F}$ and the event $\mathcal{E}_{\mathsf{good}}$, we can apply Theorem 21 and (E.10) to show that

$$\begin{aligned}
&\sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left(\mathcal{N}(\boldsymbol{0},\widetilde{\boldsymbol{\Sigma}}_l)\in\mathcal{C}\mid\boldsymbol{F}\right) - \mathbb{P}\left(\mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}_{U,l}^\star\right)\in\mathcal{C}\right)\right| \le \mathsf{TV}\left(\mathcal{N}(\boldsymbol{0},\widetilde{\boldsymbol{\Sigma}}_l),\mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}_{U,l}^\star\right)\right)\\
&\asymp \left\|\left(\boldsymbol{\Sigma}_{U,l}^\star\right)^{-1/2}\widetilde{\boldsymbol{\Sigma}}_l\left(\boldsymbol{\Sigma}_{U,l}^\star\right)^{-1/2} - \boldsymbol{I}_d\right\|_{\mathrm{F}} = \left\|\left(\boldsymbol{\Sigma}_{U,l}^\star\right)^{-1/2}\left(\widetilde{\boldsymbol{\Sigma}}_l - \boldsymbol{\Sigma}_{U,l}^\star\right)\left(\boldsymbol{\Sigma}_{U,l}^\star\right)^{-1/2}\right\|_{\mathrm{F}}\\
&\lesssim \left\|\left(\boldsymbol{\Sigma}_{U,l}^\star\right)^{-1/2}\right\|\left\|\widetilde{\boldsymbol{\Sigma}}_l - \boldsymbol{\Sigma}_{U,l}^\star\right\|_{\mathrm{F}}\left\|\left(\boldsymbol{\Sigma}_{U,l}^\star\right)^{-1/2}\right\| \lesssim \sqrt{r}\left\|\left(\boldsymbol{\Sigma}_{U,l}^\star\right)^{-1}\right\|\left\|\widetilde{\boldsymbol{\Sigma}}_l - \boldsymbol{\Sigma}_{U,l}^\star\right\|\\
&\lesssim \frac{\sqrt{r}}{\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right)}\cdot\sqrt{\frac{\kappa^5 r\log^3(n+d)}{n}}\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right)\\
&\lesssim \sqrt{\frac{\kappa^5 r^2\log^3(n+d)}{n}},
\end{aligned}$$

where $\mathsf{TV}(\cdot,\cdot)$ represents the total-variation distance between two distributions (Tsybakov and Zaiats, 2009). Combine the above inequality with (E.9) to show that: on the event $\mathcal{E}_{\mathsf{good}}$, one has

$$\sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left(\boldsymbol{Z}_{l,\cdot}\in\mathcal{C}\mid\boldsymbol{F}\right) - \mathbb{P}\left(\mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}_{U,l}^\star\right)\in\mathcal{C}\right)\right| \lesssim \frac{1}{\sqrt{\log(n+d)}}, \tag{E.11}$$

with the proviso that $n \gtrsim \kappa^5 r^2\log^4(n+d)$.

Step 3: taking higher-order error into account. In this step, we shall always assume that the event $\mathcal{E}_{\mathsf{good}}$ happens. It has been shown in Lemma 7 that

$$\mathbb{P}\left(\left\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star - \boldsymbol{Z}\right\|_{2,\infty} \lesssim \zeta_{\mathsf{2nd}}\,\middle|\,\boldsymbol{F}\right) \ge 1 - O\left((n+d)^{-10}\right),$$

which immediately gives

$$\mathbb{P}\left(\left\|\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\left(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}-\boldsymbol{Z}\right)\right\|_{2,\infty}\leq\zeta\,\middle|\,\boldsymbol{F}\right)\geq1-O\left(\left(n+d\right)^{-10}\right). \tag{E.12}$$

Here, the quantity ζ is defined as

$$\zeta\coloneqq c_{\zeta}\zeta_{\mathsf{2nd}}\left(\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)\right)^{-\frac{1}{2}} \tag{E.13}$$

for some sufficiently large constant $c_{\zeta}>0$.

For any convex set $\mathcal{C}\in\mathscr{C}^{r}$ and any ε, recalling the definition of $\mathcal{C}^{\varepsilon}$ in (A.3) in Appendix A, we have

$$\begin{aligned}
\mathbb{P}\left(\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\boldsymbol{Z}_{l,\cdot}\in\mathcal{C}^{-\zeta}\,\middle|\,\boldsymbol{F}\right)&=\mathbb{P}\left(\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\boldsymbol{Z}_{l,\cdot}\in\mathcal{C}^{-\zeta},\left\|\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\left(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}-\boldsymbol{Z}\right)\right\|_{2,\infty}\leq\zeta\,\middle|\,\boldsymbol{F}\right)\\
&\quad+\mathbb{P}\left(\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\boldsymbol{Z}_{l,\cdot}\in\mathcal{C}^{-\zeta},\left\|\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\left(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}-\boldsymbol{Z}\right)\right\|_{2,\infty}>\zeta\,\middle|\,\boldsymbol{F}\right)\\
&\leq\mathbb{P}\left(\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\left(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}\right)_{l,\cdot}\in\mathcal{C}\,\middle|\,\boldsymbol{F}\right)+\mathbb{P}\left(\left\|\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\left(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}-\boldsymbol{Z}\right)\right\|_{2,\infty}>\zeta\,\middle|\,\boldsymbol{F}\right)\\
&\leq\mathbb{P}\left(\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\left(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}\right)_{l,\cdot}\in\mathcal{C}\,\middle|\,\boldsymbol{F}\right)+O\left(\left(n+d\right)^{-10}\right),
\end{aligned} \tag{E.14}$$

where the first inequality follows from the definition of $\mathcal{C}^{-\zeta}$, and the last inequality makes use of (E.12). Similarly,

$$\mathbb{P}\left(\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\left(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}\right)_{l,\cdot}\in\mathcal{C}\,\middle|\,\boldsymbol{F}\right)\leq\mathbb{P}\left(\boldsymbol{\Sigma}_{l}^{-1/2}\boldsymbol{Z}_{l,\cdot}\in\mathcal{C}^{\zeta}\,\middle|\,\boldsymbol{F}\right)+O\left(\left(n+d\right)^{-10}\right). \tag{E.15}$$

In addition, for any set $\mathcal{X}\subseteq\mathbb{R}^{r}$ and any matrix $\boldsymbol{A}\in\mathbb{R}^{r\times r}$, let us denote by $\boldsymbol{A}\mathcal{X}$ be the set $\{\boldsymbol{A}\boldsymbol{x}:\boldsymbol{x}\in\mathcal{X}\}$. It is easily seen that when $\boldsymbol{A}$ is non-singular, $\boldsymbol{A}\mathcal{C}\subseteq\mathscr{C}^{r}$ holds if and only if $\mathcal{C}\in\mathscr{C}^{r}$. We can then deduce that

$$\begin{aligned}
\mathbb{P}\left(\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\boldsymbol{Z}_{l,\cdot}\in\mathcal{C}^{\zeta}\,\middle|\,\boldsymbol{F}\right)&=\mathbb{P}\left(\boldsymbol{Z}_{l,\cdot}\in\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{1/2}\mathcal{C}^{\zeta}\,\middle|\,\boldsymbol{F}\right)\\
&\overset{\text{(i)}}{\leq}\mathbb{P}\left(\mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}_{U,l}^{\star}\right)\in\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{1/2}\mathcal{C}^{\zeta}\right)+O\left(\frac{1}{\sqrt{\log\left(n+d\right)}}\right)\\
&=\mathbb{P}\left(\mathcal{N}\left(\boldsymbol{0},\boldsymbol{I}_{r}\right)\in\mathcal{C}^{\zeta}\right)+O\left(\frac{1}{\sqrt{\log\left(n+d\right)}}\right)\\
&\overset{\text{(ii)}}{\leq}\mathbb{P}\left(\mathcal{N}\left(\boldsymbol{0},\boldsymbol{I}_{r}\right)\in\mathcal{C}\right)+\zeta\left(0.59r^{1/4}+0.21\right)+\left(\frac{1}{\sqrt{\log\left(n+d\right)}}\right)\\
&\overset{\text{(iii)}}{\leq}\mathbb{P}\left(\mathcal{N}\left(\boldsymbol{0},\boldsymbol{I}_{r}\right)\in\mathcal{C}\right)+O\left(\frac{1}{\sqrt{\log\left(n+d\right)}}\right),
\end{aligned}$$

where (i) uses (E.11), (ii) is a consequence of Theorem 22, and (iii) holds provided that $\zeta\lesssim1/(r^{1/4}\log^{1/2}(n+d))$. Similarly we can show that

$$\mathbb{P}\left(\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\boldsymbol{Z}_{l,\cdot}\in\mathcal{C}^{-\zeta}\,\middle|\,\boldsymbol{F}\right)\geq\mathbb{P}\left(\mathcal{N}\left(\boldsymbol{0},\boldsymbol{I}_{r}\right)\in\mathcal{C}\right)-O\left(\frac{1}{\sqrt{\log\left(n+d\right)}}\right).$$

Combine the above two inequalities with (E.14) and (E.15) to achieve

$$\left|\mathbb{P}\left(\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\left(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}\right)_{l,\cdot}\in\mathcal{C}\,\middle|\,\boldsymbol{F}\right)-\mathbb{P}\left(\mathcal{N}\left(\boldsymbol{0},\boldsymbol{I}_{r}\right)\in\mathcal{C}\right)\right|\lesssim\frac{1}{\sqrt{\log\left(n+d\right)}}.$$

It is worth noting that this inequality holds for all $\mathcal{C}\in\mathscr{C}^{r}$. As a result, on the event $\mathcal{E}_{\mathsf{good}}$ we can obtain

$$\sup_{\mathcal{C}\in\mathscr{C}^{r}}\left|\mathbb{P}\left(\left(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}\right)_{l,\cdot}\in\mathcal{C}\,\middle|\,\boldsymbol{F}\right)-\mathbb{P}\left(\mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}_{l}\right)\in\mathcal{C}\right)\right|$$

$$
\begin{aligned}
&= \sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left(\left(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^\star\right)_{l,\cdot}\in\boldsymbol{\Sigma}_l^{1/2}\mathcal{C}\right)-\mathbb{P}\left(\mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}_l\right)\in\boldsymbol{\Sigma}_l^{1/2}\mathcal{C}\right)\right| \\
&= \sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left(\boldsymbol{\Sigma}_l^{-1/2}\left(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^\star\right)_{l,\cdot}\in\mathcal{C}\right)-\mathbb{P}\left(\mathcal{N}\left(\boldsymbol{0},\boldsymbol{I}_r\right)\in\mathcal{C}\right)\right| \\
&\lesssim \frac{1}{\sqrt{\log\left(n+d\right)}},
\end{aligned}
\tag{E.16}
$$

where the first identity makes use of the fact that $\mathcal{C}\rightarrow(\boldsymbol{\Sigma}_{U,l}^\star)^{1/2}\mathcal{C}$ is a one-to one mapping from $\mathscr{C}^r$ to $\mathscr{C}^r$ (since $\boldsymbol{\Sigma}_{U,l}^\star$ has full rank).

We are left with checking the conditions required to guarantee $\zeta\lesssim 1/(r^{1/4}\log^{1/2}(n+d))$. More generally, we shall check the conditions required to guarantee

$$
\zeta_{\mathsf{2nd}}\lesssim\delta\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^\star\right)
$$

for some $\delta>0$. Recall that

$$
\begin{aligned}
\zeta_{\mathsf{2nd}} &= \frac{\sigma_{\mathsf{ub}}^2\sqrt{\left(n+d\right)r}\log\left(n+d\right)}{\sigma_r^{\star2}}+\left(\frac{\sqrt{\kappa}\sigma_{\mathsf{ub}}^2\left(n+d\right)}{\sigma_r^{\star2}}+\frac{\sigma_{\mathsf{ub}}\sqrt{r\log\left(n+d\right)}}{\sigma_r^\star}\right)\left\|\boldsymbol{U}^\star\right\|_{2,\infty} \\
&\asymp \underbrace{\frac{\sigma_{\mathsf{ub}}^2\sqrt{\left(n+d\right)r}\log\left(n+d\right)}{\sigma_r^{\star2}}}_{=:\alpha_1}+\underbrace{\frac{\sigma_{\mathsf{ub}}^2\left(n+d\right)}{\sigma_r^{\star2}}\sqrt{\frac{\kappa\mu r}{d}}}_{=:\alpha_2}+\underbrace{\frac{\sigma_{\mathsf{ub}}\sqrt{r\log\left(n+d\right)}}{\sigma_r^\star}\sqrt{\frac{\mu r}{d}}}_{=:\alpha_3},
\end{aligned}
$$

where

$$
\sigma_{\mathsf{ub}}\asymp\sqrt{\frac{\mu r\log\left(n+d\right)}{ndp}}\sigma_1^\star+\frac{\omega_{\max}}{\sqrt{np}}.
$$

In addition, we have learned from Lemma 8 that

$$
\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^\star\right)\gtrsim\frac{1}{\sqrt{np}\sigma_1^\star}\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\omega_l^\star\right).
$$

In what follows, we shall bound the terms α_1, α_2 and α_3 separately.

- Let us start with α_1, for which we have

$$
\begin{aligned}
\alpha_1 &\asymp \left(\frac{\mu r\log\left(n+d\right)}{ndp}\sigma_1^{\star2}+\frac{\omega_{\max}^2}{np}\right)\frac{\sqrt{\left(n+d\right)r}\log\left(n+d\right)}{\sigma_r^{\star2}} \\
&\asymp \frac{\kappa\mu r^{3/2}\sqrt{n+d}\log^2\left(n+d\right)}{ndp}+\frac{\omega_{\max}^2}{\sigma_r^{\star2}}\frac{\sqrt{\left(n+d\right)r}\log\left(n+d\right)}{np} \\
&\lesssim \delta\frac{1}{\sqrt{np}\sigma_1^\star}\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\omega_l^\star\right)\lesssim\delta\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^\star\right).
\end{aligned}
$$

  Here, the last line holds under Assumption 1 as well as the conditions that

$$
\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{n+d}{np}}\lesssim\delta\frac{1}{\sqrt{\kappa r\log^2\left(n+d\right)}}
$$

$$
\text{and}\quad\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\gtrsim\delta^{-1}\frac{\kappa\mu r\log^2\left(n+d\right)}{\sqrt{\left(n\wedge d\right)p}}\sqrt{\frac{r}{d}}\sigma_1^\star.
$$

- Regarding α_2, we make the observation that

$$
\alpha_2\asymp\left(\frac{\mu r\log\left(n+d\right)}{ndp}\sigma_1^{\star2}+\frac{\omega_{\max}^2}{np}\right)\frac{n+d}{\sigma_r^{\star2}}\sqrt{\frac{\kappa\mu r}{d}}
$$

$$
\begin{aligned}
&\asymp \frac{\kappa^{3/2}\mu^{3/2}r\log(n+d)}{(n\wedge d)\,p}\sqrt{\frac{r}{d}} + \frac{\omega_{\max}^2}{\sigma_r^{\star 2}}\frac{n+d}{np}\sqrt{\frac{\kappa\mu r}{d}} \\
&\lesssim \delta\frac{1}{\sqrt{np}\sigma_1^\star}\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \omega_l^\star\right) \lesssim \delta\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^\star\right).
\end{aligned}
$$

Here, the last line makes use of Assumption 1 as well as the assumptions

$$
\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{n+d}{np}} \lesssim \delta\frac{\sqrt{1\wedge(d/n)}}{\kappa\mu^{1/2}r^{1/2}}
$$

$$
\text{and}\quad \left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \gtrsim \delta^{-1}\frac{\kappa^{3/2}\mu^{3/2}r\log(n+d)}{\sqrt{np}\wedge\sqrt{d^2p/n}}\sqrt{\frac{r}{d}}\sigma_1^\star.
$$

- It remains to control α_3. Towards this end, we have

$$
\begin{aligned}
\alpha_3 &\asymp \left(\sqrt{\frac{\mu r\log(n+d)}{ndp}}\sigma_1^\star + \frac{\omega_{\max}}{\sqrt{np}}\right)\frac{\sqrt{r\log(n+d)}}{\sigma_r^\star}\sqrt{\frac{\mu r}{d}} \\
&\asymp \sqrt{\frac{\kappa\mu^2r^3\log^2(n+d)}{nd^2p}} + \frac{\omega_{\max}}{\sqrt{np}}\frac{\sqrt{r\log(n+d)}}{\sigma_r^\star}\sqrt{\frac{\mu r}{d}} \\
&\lesssim \delta\frac{1}{\sqrt{np}\sigma_1^\star}\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \omega_l^\star\right) \lesssim \delta\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^\star\right).
\end{aligned}
$$

Here, the last line holds under Assumption 1 and the conditions $d \gtrsim \delta^{-2}\kappa\mu r^2\log(n+d)$ and

$$
\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \gtrsim \delta^{-1}\frac{\kappa^{1/2}\mu r^{3/2}\log(n+d)}{d}\sigma_1^\star.
$$

Combining the above results, we see that $\zeta_{\mathsf{2nd}} \lesssim \delta\lambda_{\min}^{1/2}(\boldsymbol{\Sigma}_{U,l}^\star)$ occurs provided that $d \gtrsim \delta^{-2}\kappa\mu r^2\log(n+d)$,

$$
\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{n+d}{np}} \lesssim \delta\frac{1}{\sqrt{\kappa r\log^2(n+d)}} \wedge \delta\frac{\sqrt{1\wedge(d/n)}}{\kappa\mu^{1/2}r^{1/2}}
$$

and

$$
\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \gtrsim \delta^{-1}\left[\frac{\kappa\mu r\log^2(n+d)}{\sqrt{(n\wedge d)\,p}} + \frac{\kappa^{3/2}\mu^{3/2}r\log(n+d)}{\sqrt{np}\wedge\sqrt{d^2p/n}} + \frac{\kappa^{1/2}\mu r\log(n+d)}{\sqrt{d}}\right]\sqrt{\frac{r}{d}}\sigma_1^\star.
$$

By taking

$$
\delta = \frac{1}{r^{1/4}\log^{1/2}(n+d)},
$$

we can conclude that $\zeta \lesssim 1/(r^{1/4}\log^{1/2}(n+d))$ holds as long as $d \gtrsim \kappa\mu r^{5/2}\log^2(n+d)$,

$$
\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{n+d}{np}} \lesssim \frac{1}{\kappa^{1/2}r^{3/4}\log^{3/2}(n+d)} \wedge \frac{\sqrt{1\wedge(d/n)}}{\kappa\mu^{1/2}r^{3/4}\log^{1/2}(n+d)}
$$

and

$$
\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \gtrsim \left[\frac{\kappa\mu r^{5/4}\log^{5/2}(n+d)}{\sqrt{(n\wedge d)\,p}} + \frac{\kappa^{3/2}\mu^{3/2}r^{5/4}\log^{3/2}(n+d)}{\sqrt{np}\wedge\sqrt{d^2p/n}} + \frac{\kappa^{1/2}\mu r^{5/4}\log^{3/2}(n+d)}{\sqrt{d}}\right]\sqrt{\frac{r}{d}}\sigma_1^\star.
$$

Step 4: distributional characterization of $(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star})_{l,\cdot}$. For any convex set $\mathcal{C}\in\mathscr{C}^{r}$, it holds that

$$
\begin{aligned}
&\left|\mathbb{P}\left((\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star})_{l,\cdot}\in\mathcal{C}\right)-\mathbb{P}\left(\mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}_{U,l}^{\star}\right)\in\mathcal{C}\right)\right| \\
&\quad=\left|\mathbb{E}\left[\mathbb{P}\left((\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star})_{l,\cdot}\in\mathcal{C}\big|\boldsymbol{F}\right)-\mathbb{P}\left(\mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}_{U,l}^{\star}\right)\in\mathcal{C}\right)\right]\right| \\
&\quad\leq\left|\mathbb{E}\left[\left[\mathbb{P}\left((\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star})_{l,\cdot}\in\mathcal{C}\big|\boldsymbol{F}\right)-\mathbb{P}\left(\mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}_{U,l}^{\star}\right)\in\mathcal{C}\right)\right]\mathbb{1}_{\mathcal{E}_{\mathsf{good}}}\right]\right| \\
&\qquad+\left|\mathbb{E}\left[\left[\mathbb{P}\left((\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star})_{l,\cdot}\in\mathcal{C}\big|\boldsymbol{F}\right)-\mathbb{P}\left(\mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}_{U,l}^{\star}\right)\in\mathcal{C}\right)\right]\mathbb{1}_{\mathcal{E}_{\mathsf{good}}^{\mathsf{c}}}\right]\right| \\
&\quad\leq\frac{1}{\sqrt{\log\left(n+d\right)}}+2\mathbb{P}\left(\mathcal{E}_{\mathsf{good}}^{\mathsf{c}}\right) \\
&\quad\lesssim\frac{1}{\sqrt{\log\left(n+d\right)}}+\frac{1}{\left(n+d\right)^{100}}\lesssim\frac{1}{\sqrt{\log\left(n+d\right)}},
\end{aligned}
$$

where the penultimate line relies on (E.16). This allows one to conclude that

$$
\sup_{\mathcal{C}\in\mathscr{C}^{r}}\left|\mathbb{P}\left(\boldsymbol{Z}_{l,\cdot}\in\mathcal{C}\right)-\mathbb{P}\left(\mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}_{U,l}^{\star}\right)\in\mathcal{C}\right)\right|\lesssim\frac{1}{\sqrt{\log\left(n+d\right)}}=o\left(1\right).
$$

E.2 Auxiliary lemmas for Theorem 12

E.2.1 Proof of Lemma 10

1. We first bound $\|(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star})_{l,\cdot}\|_2$. Recall that

$$
\boldsymbol{Z}_{l,\cdot}=\sum_{j=1}^{n}E_{l,j}\boldsymbol{V}_{j,\cdot}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top},
$$

which is an independent sum of random vectors (where the randomness comes from $\{E_{l,j}\}_{1\leq j\leq n}$) conditional on $\boldsymbol{F}$. By carrying out the following calculation (see (D.7) and (D.8))

$$
\begin{aligned}
L&\coloneqq\max_{1\leq j\leq n}\left\|E_{l,j}\boldsymbol{V}_{j,\cdot}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top}\right\|_{2}\leq\left\{\max_{1\leq j\leq n}\left|E_{l,j}\right|\right\}\left\|\boldsymbol{V}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\right\|_{2,\infty} \\
&\lesssim\frac{1}{\sqrt{np}}\left(\max_{1\leq j\leq n}\left|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_{j}\right|+\omega_{l}^{\star}\sqrt{\log\left(n+d\right)}\right)\left\|\boldsymbol{V}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\right\|_{2,\infty}, \\
V&\coloneqq\sum_{j=1}^{n}\mathbb{E}\left[E_{l,j}^{2}\left\|\boldsymbol{V}_{j,\cdot}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top}\right\|_{2}^{2}\right]\leq\sum_{j=1}^{n}\mathbb{E}\left[E_{l,j}^{2}\right]\left\|\boldsymbol{V}_{j,\cdot}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\right\|_{2}^{2}\leq\max_{j}\mathbb{E}\left[E_{l,j}^{2}\right]\left\|\boldsymbol{V}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\right\|_{\mathrm{F}}^{2} \\
&\lesssim\frac{1}{np}\left[\max_{1\leq j\leq n}\left(\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_{j}\right)^{2}+\omega_{l}^{\star2}\right]\left\|\boldsymbol{V}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\right\|_{\mathrm{F}}^{2},
\end{aligned}
$$

we can invoke the Bernstein inequality (Chen et al., 2020c, Corollary 3.1.3) to demonstrate that

$$
\begin{aligned}
\left\|\boldsymbol{Z}_{l,\cdot}\right\|_{2}&\lesssim\sqrt{V\log\left(n+d\right)}+L\log\left(n+d\right) \\
&\lesssim\sqrt{\frac{\log\left(n+d\right)}{np}}\left[\max_{1\leq j\leq n}\left|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_{j}\right|+\omega_{l}^{\star}\right]\left\|\boldsymbol{V}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\right\|_{\mathrm{F}} \\
&\quad+\frac{\log\left(n+d\right)}{\sqrt{np}}\left(\max_{1\leq j\leq n}\left|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_{j}\right|+\omega_{l}^{\star}\sqrt{\log\left(n+d\right)}\right)\left\|\boldsymbol{V}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\right\|_{2,\infty}
\end{aligned}
$$

with probability at least $1-O((n+d)^{-10})$. On the event $\mathcal{E}_{\mathsf{good}}$, we can further derive

$$
\left\|\boldsymbol{Z}_{l,\cdot}\right\|_{2}\overset{\text{(i)}}{\lesssim}\frac{1}{\sigma_{r}^{\star}}\sqrt{\frac{r\log\left(n+d\right)}{np}}\left[\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\sqrt{\log\left(n+d\right)}+\omega_{l}^{\star}\right]
$$

$$
\begin{aligned}
&+\frac{1}{\sigma_r^\star}\frac{\log(n+d)}{\sqrt{np}}\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\sqrt{\log(n+d)}+\omega_l^\star\sqrt{\log(n+d)}\right)\sqrt{\frac{\log(n+d)}{n}}\\
&\asymp\frac{1}{\sigma_r^\star}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left(\sqrt{\frac{r\log^2(n+d)}{np}}+\frac{\log^{3/2}(n+d)}{np}\right)+\frac{\omega_l^\star}{\sigma_r^\star}\left(\sqrt{\frac{r\log(n+d)}{np}}+\frac{\log^2(n+d)}{np}\right)\\
&\overset{\text{(ii)}}{\lesssim}\frac{1}{\sigma_r^\star}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\sqrt{\frac{r\log^2(n+d)}{np}}+\frac{\omega_l^\star}{\sigma_r^\star}\sqrt{\frac{r\log(n+d)}{np}}
\end{aligned}
$$

where (i) uses (D.10), (E.57) as well as (D.13), and (ii) holds true as long as $np\gtrsim\log^3(n+d)$. By combining the above inequality with Lemma 7, we demonstrate that with probability exceeding $1-O((n+d)^{-10})$

$$
\left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{R}-\boldsymbol{U}_{l,\cdot}^\star\right\|_2\lesssim\frac{1}{\sigma_r^\star}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\sqrt{\frac{r\log^2(n+d)}{np}}+\frac{\omega_l^\star}{\sigma_r^\star}\sqrt{\frac{r\log(n+d)}{np}}+\zeta_{\mathsf{2nd}}.
$$

2. Next, we turn attention to the term $\|(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}\boldsymbol{Q}^\top-\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star)_{l,\cdot}\|$. Conditional on $\boldsymbol{F}$, we have learned from (C.4) that

$$
\left\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}-\boldsymbol{M}\boldsymbol{V}^\natural\right\|_{2,\infty}\lesssim\frac{\sigma\sqrt{n+d}+B\log(n+d)}{\sigma_r^\natural}\sigma\sqrt{r\log(n+d)}+\frac{\kappa^\natural\sigma^2(n+d)}{\sigma_r^\natural}\left\|\boldsymbol{U}^\natural\right\|_{2,\infty}
$$

with probability exceeding $1-O((n+d)^{-10})$. When $\mathcal{E}_{\mathsf{good}}$ happens, we have

$$
\begin{aligned}
\left\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}-\boldsymbol{M}\boldsymbol{V}^\natural\right\|_{2,\infty}&\overset{\text{(i)}}{\lesssim}\frac{\sigma_{\mathsf{ub}}\sqrt{n+d}+\sigma_{\mathsf{ub}}\sqrt{\log^3(n+d)/p}}{\sigma_r^\star}\sigma_{\mathsf{ub}}\sqrt{r\log(n+d)}+\frac{\sqrt{\kappa}\sigma_{\mathsf{ub}}^2(n+d)}{\sigma_r^\star}\left\|\boldsymbol{U}^\star\right\|_{2,\infty}\\
&\overset{\text{(ii)}}{\lesssim}\frac{\sigma_{\mathsf{ub}}^2\sqrt{(n+d)\,r\log(n+d)}}{\sigma_r^\star}+\frac{\sqrt{\kappa}\sigma_{\mathsf{ub}}^2(n+d)}{\sigma_r^\star}\left\|\boldsymbol{U}^\star\right\|_{2,\infty}\lesssim\sigma_r^\star\zeta_{\mathsf{2nd}}.
\end{aligned}
$$

where (i) uses (D.10), (D.14) and (D.11), and (ii) is valid provided that $(n\vee d)p\gtrsim\log^3(n+d)$. Note that

$$
\boldsymbol{M}\boldsymbol{V}^\natural=\boldsymbol{M}^\natural\boldsymbol{V}^\natural+\boldsymbol{E}\boldsymbol{V}^\natural=\boldsymbol{U}^\natural\boldsymbol{\Sigma}^\natural+\boldsymbol{E}\boldsymbol{V}^\natural.
$$

Similar to how we bound $\|\boldsymbol{Z}_{l,\cdot}\|_2$, we can show that with probability exceeding $1-O((n+d)^{-10})$

$$
\left\|\boldsymbol{E}_{l,\cdot}\boldsymbol{V}^\natural\right\|_2\lesssim\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\omega_l^\star\right)\sqrt{\frac{r\log^2(n+d)}{np}},
$$

thus indicating that

$$
\left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}-\boldsymbol{U}_{l,\cdot}^\natural\boldsymbol{\Sigma}^\natural\right\|_2\leq\left\|\boldsymbol{E}_{l,\cdot}\boldsymbol{V}^\natural\right\|_2+\left\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}-\boldsymbol{M}\boldsymbol{V}^\natural\right\|_{2,\infty}\leq\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\omega_l^\star\right)\sqrt{\frac{r\log^2(n+d)}{np}}+\sigma_r^\star\zeta_{\mathsf{2nd}}.
$$

In addition, note that $\boldsymbol{U}^\natural=\boldsymbol{U}^\star\boldsymbol{Q}$, and hence

$$
\begin{aligned}
\left\|\boldsymbol{U}_{l,\cdot}^\natural\boldsymbol{\Sigma}^\natural-\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\boldsymbol{Q}\right\|_2&\leq\left\|\boldsymbol{U}_{l,\cdot}^\star\left(\boldsymbol{Q}\boldsymbol{\Sigma}^\natural-\boldsymbol{\Sigma}^\star\boldsymbol{Q}\right)\right\|_2\leq\left\|\boldsymbol{U}_{l,\cdot}^\star\right\|_2\left\|\boldsymbol{Q}\boldsymbol{\Sigma}^\natural-\boldsymbol{\Sigma}^\star\boldsymbol{Q}\right\|\\
&\lesssim\kappa\sigma_r^\star\sqrt{\frac{r+\log(n+d)}{n}}\left\|\boldsymbol{U}_{l,\cdot}^\star\right\|_2,
\end{aligned}
$$

where the penultimate relation relies on (E.58). Therefore, we can arrive at

$$
\begin{aligned}
\left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}\boldsymbol{Q}^\top-\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2&\leq\left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}-\boldsymbol{U}_{l,\cdot}^\natural\boldsymbol{\Sigma}^\natural\right\|_2+\left\|\boldsymbol{U}_{l,\cdot}^\natural\boldsymbol{\Sigma}^\natural-\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\boldsymbol{Q}\right\|_2\\
&\lesssim\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\omega_l^\star\right)\sqrt{\frac{r\log^2(n+d)}{np}}+\kappa\sigma_r^\star\sqrt{\frac{r+\log(n+d)}{n}}\left\|\boldsymbol{U}_{l,\cdot}^\star\right\|_2+\sigma_r^\star\zeta_{\mathsf{2nd}}.
\end{aligned}
$$

3. Finally, let us upper bound $\|\boldsymbol{R}(\boldsymbol{\Sigma}^{\star})^{-2}\boldsymbol{R}^{\top} - (\boldsymbol{\Sigma})^{-2}\|$. From Lemma 2, we know that with probability exceeding $1 - O((n+d)^{-10})$

$$\left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}} - \boldsymbol{\Sigma}^{\natural}\right\| \lesssim \kappa^{\natural}\frac{\sigma^{2}(n+d)}{\sigma_{r}^{\natural}} + \sigma\sqrt{r\log(n+d)}.$$

When the event $\mathcal{E}_{\mathsf{good}}$ happens, it follows from (D.10), (D.11) and (D.14) that

$$\left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}} - \boldsymbol{\Sigma}^{\natural}\right\| \lesssim \sqrt{\kappa}\frac{\sigma_{\mathsf{ub}}^{2}(n+d)}{\sigma_{r}^{\star}} + \sigma_{\mathsf{ub}}\sqrt{r\log(n+d)} \lesssim \sigma_{r}^{\star}\zeta_{\mathsf{2nd}}\sqrt{\frac{d}{\mu r}}.$$

An immediate consequence is that when $\mathcal{E}_{\mathsf{good}}$ happens, by Weyl's inequality and (D.10) we have

$$\sigma_{1} \asymp \sigma_{1}^{\natural} \asymp \sigma_{1}^{\star} \qquad \text{and} \qquad \sigma_{r} \asymp \sigma_{r}^{\natural} \asymp \sigma_{r}^{\star}, \tag{E.17}$$

provided that $\zeta_{\mathsf{2nd}}\sqrt{d} \ll 1$. This further gives

$$\begin{aligned}
\left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}^{2}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{\Sigma}^{\natural 2}\right\| &\le \left\|\left(\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}} - \boldsymbol{\Sigma}^{\natural}\right)\left(\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}\right)^{\top}\right\| + \left\|\boldsymbol{\Sigma}^{\natural}\left(\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}} - \boldsymbol{\Sigma}^{\natural}\right)^{\top}\right\| \\
&\le \left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}} - \boldsymbol{\Sigma}^{\natural}\right\|\left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}\right\| + \left\|\boldsymbol{\Sigma}^{\natural}\right\|\left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}} - \boldsymbol{\Sigma}^{\natural}\right\| \\
&\lesssim \sigma_{r}^{\star}\sigma_{1}^{\star}\zeta_{\mathsf{2nd}}\sqrt{\frac{d}{\mu r}},
\end{aligned} \tag{E.18}$$

and as a result,

$$\begin{aligned}
\left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}^{-2}\boldsymbol{R}_{\boldsymbol{U}} - (\boldsymbol{\Sigma}^{\natural})^{-2}\right\| &= \left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}^{-2}\boldsymbol{R}_{\boldsymbol{U}}\left(\boldsymbol{\Sigma}^{\natural 2} - \boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}^{2}\boldsymbol{R}_{\boldsymbol{U}}\right)(\boldsymbol{\Sigma}^{\natural})^{-2}\right\| \\
&\le \left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}^{-2}\boldsymbol{R}_{\boldsymbol{U}}\right\|\left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}^{2}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{\Sigma}^{\natural 2}\right\|\left\|(\boldsymbol{\Sigma}^{\natural})^{-2}\right\| \\
&\lesssim \frac{1}{\sigma_{r}^{2}\sigma_{r}^{\natural 2}}\sigma_{r}^{\star}\sigma_{1}^{\star}\zeta_{\mathsf{2nd}} \asymp \frac{1}{\sigma_{r}^{\star 2}}\zeta_{\mathsf{2nd}}\sqrt{\frac{\kappa d}{\mu r}}.
\end{aligned}$$

In addition, from (E.58) and (D.10) we can derive

$$\begin{aligned}
\left\|\boldsymbol{\Sigma}^{\star 2} - \boldsymbol{Q}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{Q}^{\top}\right\| &= \left\|\boldsymbol{\Sigma}^{\star}\boldsymbol{Q}\boldsymbol{Q}^{\top}\boldsymbol{\Sigma}^{\star} - \boldsymbol{Q}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{Q}^{\top}\right\| \\
&\le \left\|\left(\boldsymbol{\Sigma}^{\star}\boldsymbol{Q} - \boldsymbol{Q}\boldsymbol{\Sigma}^{\natural}\right)\boldsymbol{Q}^{\top}\boldsymbol{\Sigma}^{\star}\right\| + \left\|\boldsymbol{Q}\boldsymbol{\Sigma}^{\natural}\left(\boldsymbol{\Sigma}^{\star}\boldsymbol{Q} - \boldsymbol{Q}\boldsymbol{\Sigma}^{\natural}\right)^{\top}\right\| \\
&\lesssim \sigma_{1}^{\star 2}\sqrt{\frac{r + \log(n+d)}{n}},
\end{aligned}$$

which in turn leads to

$$\begin{aligned}
\left\|\boldsymbol{Q}^{\top}(\boldsymbol{\Sigma}^{\star})^{-2}\boldsymbol{Q} - \boldsymbol{\Sigma}^{\natural -2}\right\| &= \left\|(\boldsymbol{\Sigma}^{\star})^{-2} - \boldsymbol{Q}\boldsymbol{\Sigma}^{\natural -2}\boldsymbol{Q}^{\top}\right\| \\
&\le \left\|(\boldsymbol{\Sigma}^{\star})^{-2}\right\|\left\|\boldsymbol{\Sigma}^{\star 2} - \boldsymbol{Q}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{Q}^{\top}\right\|\left\|\boldsymbol{Q}\boldsymbol{\Sigma}^{\natural -2}\boldsymbol{Q}^{\top}\right\| \\
&\lesssim \frac{1}{\sigma_{r}^{\star 2}}\sqrt{\frac{\kappa^{2} r\log(n+d)}{n}}.
\end{aligned}$$

Recalling that $\boldsymbol{R}_{\boldsymbol{U}} = \boldsymbol{R}\boldsymbol{Q}$ (see (D.6)), we reach

$$\begin{aligned}
\left\|\boldsymbol{R}(\boldsymbol{\Sigma}^{\star})^{-2}\boldsymbol{R}^{\top} - (\boldsymbol{\Sigma})^{-2}\right\| = \left\|\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{Q}^{\top}(\boldsymbol{\Sigma}^{\star})^{-2}\boldsymbol{Q}\boldsymbol{R}_{\boldsymbol{U}}^{\top} - (\boldsymbol{\Sigma})^{-2}\right\| &= \left\|\boldsymbol{Q}^{\top}(\boldsymbol{\Sigma}^{\star})^{-2}\boldsymbol{Q} - \boldsymbol{R}_{\boldsymbol{U}}^{\top}(\boldsymbol{\Sigma})^{-2}\boldsymbol{R}_{\boldsymbol{U}}\right\| \\
\le \left\|\boldsymbol{Q}^{\top}(\boldsymbol{\Sigma}^{\star})^{-2}\boldsymbol{Q} - (\boldsymbol{\Sigma}^{\natural})^{-2}\right\| + \left\|(\boldsymbol{\Sigma}^{\natural})^{-2} - \boldsymbol{R}_{\boldsymbol{U}}^{\top}(\boldsymbol{\Sigma})^{-2}\boldsymbol{R}_{\boldsymbol{U}}\right\| & \\
\lesssim \frac{1}{\sigma_{r}^{\star 2}}\sqrt{\frac{\kappa^{2} r\log(n+d)}{n}} + \frac{1}{\sigma_{r}^{\star 2}}\zeta_{\mathsf{2nd}}\sqrt{\frac{\kappa d}{\mu r}}. &
\end{aligned}$$

E.2.2 Proof of Lemma 11

Before proceeding, let us make the following useful observation that for each $l \in [d]$

$$
\begin{aligned}
\left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\right\|_2 &= \left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}\boldsymbol{Q}^\top\right\|_2 \leq \left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}} - \boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \\
&\overset{\text{(i)}}{\lesssim} \left(\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \omega_l^\star\right)\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sigma_r^\star\sqrt{\frac{r+\log(n+d)}{n}}\left\|\boldsymbol{U}_{l,\cdot}^\star\right\|_2 + \sigma_r^\star\zeta_{\mathsf{2nd}} + \left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \\
&\overset{\text{(ii)}}{\lesssim} \left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \omega_l^\star\sqrt{\frac{r\log^2(n+d)}{np}} + \sigma_r^\star\zeta_{\mathsf{2nd}},
\end{aligned}
\tag{E.19}
$$

where (i) uses (D.20), and (ii) holds as soon as $np \gtrsim r\log^2(n+d)$ and $n \gtrsim \kappa^2 r\log(n+d)$. We are now positioned to embark on the proof.

Step 1: bounding $|S_{i,j} - S_{i,j}^\star|$. Recall the definition of $\boldsymbol{S}$ and $\boldsymbol{S}^\star$ in Algorithm 1 and in (1.2), respectively.

$$
\begin{aligned}
\left|S_{i,j} - S_{i,j}^\star\right| &= \left|\left(\boldsymbol{U}_{i,\cdot}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}\boldsymbol{Q}^\top\right)\left(\boldsymbol{U}_{j,\cdot}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}\boldsymbol{Q}^\top\right)^\top - \boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\left(\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right)^\top\right| \\
&\leq \left\|\boldsymbol{U}_{i\cdot}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}\boldsymbol{Q}^\top - \boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}\boldsymbol{\Sigma}\right\|_2 + \left\|\boldsymbol{U}_{j\cdot}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}\boldsymbol{Q}^\top - \boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \\
&\overset{\text{(i)}}{\lesssim} \left\|\boldsymbol{U}_{i\cdot}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}\boldsymbol{Q}^\top - \boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left(\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \omega_{\max}\sqrt{\frac{r\log^2(n+d)}{np}} + \sigma_r^\star\zeta_{\mathsf{2nd}}\right) \\
&\quad + \left\|\boldsymbol{U}_{j\cdot}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}\boldsymbol{Q}^\top - \boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \\
&\overset{\text{(ii)}}{\lesssim} \left[\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \omega_l^\star\right)\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sigma_r^\star\sqrt{\frac{r+\log(n+d)}{n}}\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2 + \sigma_r^\star\zeta_{\mathsf{2nd}}\right] \\
&\qquad \cdot\left(\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \omega_{\max}\sqrt{\frac{r\log^2(n+d)}{np}} + \sigma_r^\star\zeta_{\mathsf{2nd}}\right) \\
&\quad + \left[\left(\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \omega_l^\star\right)\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sigma_r^\star\sqrt{\frac{r+\log(n+d)}{n}}\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2 + \sigma_r^\star\zeta_{\mathsf{2nd}}\right]\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \\
&\overset{\text{(iii)}}{\lesssim} \left(\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sqrt{\frac{r+\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \frac{r\log^2(n+d)}{np}\omega_{\max}^2 \\
&\quad + \left(\sqrt{\frac{r\log^2(n+d)}{np}}\omega_{\max} + \zeta_{\mathsf{2nd}}\sigma_r^\star\right)\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\right) + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}.
\end{aligned}
\tag{E.20}
$$

Here, (i) follows from (E.19), (ii) utilizes (D.20), while (iii) utilizes the AM-GM inequality and holds true as long as $np \gtrsim r\log^2(n+d)$ and $n \gtrsim \kappa^2 r\log(n+d)$. Specifically, taking $i = j = l$ in (E.20) yields

$$
\begin{aligned}
\left|S_{l,l} - S_{l,l}^\star\right| &\lesssim \left(\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sqrt{\frac{r+\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2 + \frac{r\log^2(n+d)}{np}\omega_{\max}^2 \\
&\quad + \left(\sqrt{\frac{r\log^2(n+d)}{np}}\omega_{\max} + \zeta_{\mathsf{2nd}}\sigma_r^\star\right)\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}.
\end{aligned}
\tag{E.21}
$$

Step 2: bounding $\left|\omega_l^2 - \omega_l^{\star 2}\right|$. With the above bound on $\left|S_{l,l} - S_{l,l}^{\star}\right|$ in place, we can move on to control $\left|\omega_l^2 - \omega_l^{\star 2}\right|$. First of all, invoke Chernoff's inequality (see (Vershynin, 2017, Exercise 2.3.5)) to show that

$$\mathbb{P}\left(\sum_{j=1}^{n} \mathbb{1}_{(l,j)\in\Omega} < \frac{1}{2}np\right) \leq \exp\left(-\frac{c}{4}np\right) \leq (n+d)^{-10}$$

as long as $np \gg \log(n+d)$. In what follows, we shall define the following event

$$\mathcal{E}_l \coloneqq \left\{\sum_{j=1}^{n} \mathbb{1}_{(l,j)\in\Omega} \geq np/2\right\},$$

which occurs with probability at least $1 - O((n+d)^{-10})$.

Recall the definition of $y_{l,j}$ in (1.3). Conditional on

$$\Omega_{l,\cdot} = \{(l,j) : (l,j) \in \Omega\},$$

one can verify that $\{y_{l,j} : (l,j) \in \Omega_{l,\cdot}\}$ are independent sub-Gaussian random variables with sub-Gaussian norm bounded above by

$$\|y_{l,j}\|_{\psi_2} = \|x_{l,j}\|_{\psi_2} + \|\eta_{l,j}\|_{\psi_2} \lesssim \sqrt{S_{l,l}^{\star}} + \omega_l^{\star}$$

for each $j \in [n]$. As a result, we know that: conditional on $\Omega_{l,\cdot}$, $\{y_{l,j}^2 : (l,j) \in \Omega_{l,\cdot}\}$ are independent sub-exponential random variables obeying

$$K \coloneqq \max_{1\leq j\leq n} \left\|y_{l,j}^2\right\|_{\psi_1} \leq \max_{1\leq j\leq n} \|y_{l,j}\|_{\psi_2}^2 \lesssim S_{l,l}^{\star} + \omega_l^{\star 2}.$$

In addition, it is easily observed that $\mathbb{E}[y_{l,j}^2] = S_{l,l}^{\star} + \omega_l^{\star 2}$. Then we can apply the Bernstein inequality (see Vershynin (2017, Theorem 2.8.1)) to demonstrate that: for any $t > 0$,

$$\mathbb{P}\left(\left|\frac{\sum_{j=1}^{n} y_{l,j}^2 \mathbb{1}_{(l,j)\in\Omega}}{\sum_{j=1}^{n} \mathbb{1}_{(l,j)\in\Omega}} - \omega_l^{\star 2} - S_{l,l}^{\star}\right| \geq t \,\middle|\, \Omega_{l,\cdot}\right) \leq 2\exp\left[-c\left(\sum_{j=1}^{n} \mathbb{1}_{(l,j)\in\Omega}\right)\min\left\{\frac{t^2}{K^2}, \frac{t}{K}\right\}\right]$$

for some universal constant $c > 0$. Therefore, when $\mathcal{E}_l$ happens, we have

$$\mathbb{P}\left(\left|\frac{\sum_{j=1}^{n} y_{l,j}^2 \mathbb{1}_{(l,j)\in\Omega}}{\sum_{j=1}^{n} \mathbb{1}_{(l,j)\in\Omega}} - \omega_l^{\star 2} - S_{l,l}^{\star}\right| \geq t \,\middle|\, \Omega_{l,\cdot}\right)\mathbb{1}_{\mathcal{E}_l} \leq 2\exp\left[-\frac{c}{2}np\min\left\{\frac{t^2}{K^2}, \frac{t}{K}\right\}\right].$$

Take expectation to further achieve that: for any $t > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{\sum_{j=1}^{n} y_{l,j}^2 \mathbb{1}_{(l,j)\in\Omega}}{\sum_{j=1}^{n} \mathbb{1}_{(l,j)\in\Omega}} - \omega_l^{\star 2} - S_{l,l}^{\star}\right| \geq t\right) = \mathbb{E}\left[\mathbb{P}\left(\left|\frac{\sum_{j=1}^{n} y_{l,j}^2 \mathbb{1}_{(l,j)\in\Omega}}{\sum_{j=1}^{n} \mathbb{1}_{(l,j)\in\Omega}} - \omega_l^{\star 2} - S_{l,l}^{\star}\right| \geq t \,\middle|\, \Omega_{l,\cdot}\right)\right] \\
&= \mathbb{E}\left[\mathbb{P}\left(\left|\frac{\sum_{j=1}^{n} y_{l,j}^2 \mathbb{1}_{(l,j)\in\Omega}}{\sum_{j=1}^{n} \mathbb{1}_{(l,j)\in\Omega}} - \omega_l^{\star 2} - S_{l,l}^{\star}\right| \geq t \,\middle|\, \Omega_{l,\cdot}\right)\mathbb{1}_{\mathcal{E}_l}\right] \\
&\quad + \mathbb{E}\left[\mathbb{P}\left(\left|\frac{\sum_{j=1}^{n} y_{l,j}^2 \mathbb{1}_{(l,j)\in\Omega}}{\sum_{j=1}^{n} \mathbb{1}_{(l,j)\in\Omega}} - \omega_l^{\star 2} - S_{l,l}^{\star}\right| \geq t \,\middle|\, \Omega_{l,\cdot}\right)\mathbb{1}_{\mathcal{E}_l^{\mathrm{c}}}\right] \\
&\leq 2\exp\left[-\frac{c}{2}np\min\left\{\frac{t^2}{K^2}, \frac{t}{K}\right\}\right] + \mathbb{P}\left(\mathcal{E}_l^{\mathrm{c}}\right) \lesssim 2\exp\left[-\frac{c}{2}np\min\left\{\frac{t^2}{K^2}, \frac{t}{K}\right\}\right] + (n+d)^{-10}.
\end{aligned}$$

By taking

$$t = \widetilde{C}K\sqrt{\frac{\log^2(n+d)}{np}} \asymp \widetilde{C}\left(\omega_l^{\star 2} + S_{l,l}^{\star}\right)\sqrt{\frac{\log^2(n+d)}{np}}$$

for some sufficiently large constant $\widetilde{C} > 0$, we see that with probability exceeding $1 - O((n+d)^{-10})$

$$\left| \frac{\sum_{j=1}^{n} y_{l,j}^2 \mathbb{1}_{(l,j)\in\Omega}}{\sum_{j=1}^{n} \mathbb{1}_{(l,j)\in\Omega}} - \omega_l^{\star 2} - S_{l,l}^{\star} \right| \lesssim \sqrt{\frac{\log^2(n+d)}{np}} \left(\omega_l^{\star 2} + S_{l,l}^{\star}\right). \tag{E.22}$$

This in turn allows one to conclude that

$$\begin{aligned}
\left|\omega_l^2 - \omega_l^{\star 2}\right| &\le \left| \frac{\sum_{j=1}^{n} y_{l,j}^2 \mathbb{1}_{(l,j)\in\Omega}}{\sum_{j=1}^{n} \mathbb{1}_{(l,j)\in\Omega}} - \omega_l^{\star 2} - S_{l,l}^{\star} \right| + \left|S_{l,l} - S_{l,l}^{\star}\right| \\
&\lesssim \sqrt{\frac{\log^2(n+d)}{np}} \left(\omega_l^{\star 2} + S_{l,l}^{\star}\right) + \left(\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sqrt{\frac{r+\log(n+d)}{n}}\right) \left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 \\
&\quad + \frac{r\log^2(n+d)}{np}\omega_{\max}^2 + \left(\sqrt{\frac{r\log^2(n+d)}{np}}\omega_{\max} + \zeta_{\mathsf{2nd}}\sigma_r^{\star}\right) \left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2} \\
&\lesssim \left(\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sqrt{\frac{r+\log(n+d)}{n}}\right) \left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \sqrt{\frac{\log^2(n+d)}{np}}\omega_{\max}^2 \\
&\quad + \left(\sqrt{\frac{r\log^2(n+d)}{np}}\omega_{\max} + \zeta_{\mathsf{2nd}}\sigma_r^{\star}\right) \left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2},
\end{aligned}$$

where the second inequality makes use of (E.21) and (E.22); the last relation holds as long as $np \gtrsim r\log^2(n+d)$.

E.2.3 Proof of Lemma 12

Recall from (D.17) and (D.18) that

$$\begin{aligned}
\boldsymbol{R}\boldsymbol{\Sigma}_{U,l}^{\star}\boldsymbol{R}^{\top} &= \underbrace{\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 \boldsymbol{R}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{R}^{\top}}_{=:\boldsymbol{A}_1} + \underbrace{\frac{\omega_l^{\star 2}}{np}\boldsymbol{R}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{R}^{\top}}_{=:\boldsymbol{A}_2} + \underbrace{\frac{2(1-p)}{np}\boldsymbol{R}\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{R}^{\top}}_{=:\boldsymbol{A}_3}; \\
\boldsymbol{\Sigma}_{U,l} &= \underbrace{\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\right\|_2^2 (\boldsymbol{\Sigma})^{-2}}_{=:\boldsymbol{B}_1} + \underbrace{\frac{\omega_l^2}{np}(\boldsymbol{\Sigma})^{-2}}_{=:\boldsymbol{B}_2} + \underbrace{\frac{2(1-p)}{np}\boldsymbol{U}_{l,\cdot}^{\top}\boldsymbol{U}_{l,\cdot}}_{=:\boldsymbol{B}_3}.
\end{aligned}$$

Before continuing, we single out several useful facts. It follows from Lemma 10 that for all $l \in [d]$,

$$\begin{aligned}
\left|\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 - \left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\right\|_2\right| &= \left|\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 - \left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}\boldsymbol{Q}^{\top}\right\|_2\right| \le \left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\boldsymbol{R}_{\boldsymbol{V}}\boldsymbol{Q}^{\top} - \boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \\
&\lesssim \left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right)\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sigma_r^{\star}\sqrt{\frac{r+\log(n+d)}{n}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2 + \sigma_r^{\star}\zeta_{\mathsf{2nd}},
\end{aligned}$$

which further reveals that

$$\begin{aligned}
&\left|\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 - \left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\right\|_2^2\right| = \left|\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 - \left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\right\|_2\right| \left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\right\|_2\right) \\
&\lesssim \left|\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 - \left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\right\|_2\right| \left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \left|\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 - \left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\right\|_2\right|\right) \\
&\lesssim \left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right)\sqrt{\frac{r\log^2(n+d)}{np}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \kappa\sqrt{\frac{r+\log(n+d)}{n}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \sigma_r^{\star}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \\
&\quad + \left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \omega_l^{\star 2}\right)\frac{r\log^2(n+d)}{np} + \kappa^2\sigma_r^{\star 2}\frac{r+\log(n+d)}{n}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2^2 + \sigma_r^{\star 2}\zeta_{\mathsf{2nd}}^2
\end{aligned}$$

$$\lesssim \omega_l^\star \left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2 \sqrt{\frac{r\log^2(n+d)}{np}} + \left[\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sqrt{\frac{r+\log(n+d)}{n}}\right] \left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2$$
$$+ \omega_l^{\star 2}\frac{r\log^2(n+d)}{np} + \sigma_r^\star \zeta_{\mathsf{2nd}} \left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2 + \sigma_r^{\star 2}\zeta_{\mathsf{2nd}}^2. \tag{E.23}$$

Also, we recall from Lemma 8 that

$$\lambda_{\min}\left(\boldsymbol{\Sigma}_l^\star\right) \gtrsim \frac{1-p}{np\sigma_1^{\star 2}} \left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2 + \frac{\omega_l^{\star 2}}{np\sigma_1^{\star 2}} \asymp \frac{1}{np\sigma_1^{\star 2}} \left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2 + \frac{1}{np\sigma_1^{\star 2}}\omega_l^{\star 2} + \frac{1}{np\sigma_1^{\star 2}}\omega_l^\star \left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2,$$

where the last relation uses the AM-GM inequality. Armed with these bounds, we are ready to bound the terms $\|\boldsymbol{A}_i - \boldsymbol{B}_i\|$ for $i = 1, 2, 3$, which we shall study separately.

- We start with $\|\boldsymbol{A}_1 - \boldsymbol{B}_1\|$, where we have

$$\begin{aligned}
\|\boldsymbol{A}_1 - \boldsymbol{B}_1\| &\le \frac{1-p}{np}\left|\left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2 - \left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\right\|_2^2\right| \left\|\boldsymbol{\Sigma}^{-2}\right\| + \left(\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2\right) \left\|\boldsymbol{R}\left(\boldsymbol{\Sigma}^\star\right)^{-2}\boldsymbol{R}^\top - \boldsymbol{\Sigma}^{-2}\right\| \\
&\lesssim \underbrace{\frac{1}{np}\frac{\omega_l^\star}{\sigma_r^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2 \sqrt{\frac{r\log^2(n+d)}{np}}}_{=:\alpha_{1,1}} + \underbrace{\frac{1}{np\sigma_r^\star}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2 + \frac{1}{np}\zeta_{\mathsf{2nd}}^2}_{=:\alpha_{1,2}} \\
&\quad + \underbrace{\frac{1}{np\sigma_r^{\star 2}}\left[\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sqrt{\frac{r+\log(n+d)}{n}}\right]\left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2}_{=:\alpha_{1,3}} + \underbrace{\frac{1}{np}\frac{\omega_l^{\star 2}}{\sigma_r^{\star 2}}\frac{r\log^2(n+d)}{np}}_{=:\alpha_{1,4}} \\
&\quad + \underbrace{\frac{1}{np\sigma_r^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2\sqrt{\frac{\kappa^2 r\log(n+d)}{n}}}_{=:\alpha_{1,5}} + \underbrace{\frac{1}{np}\left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2 \frac{1}{\sigma_r^{\star 2}}\zeta_{\mathsf{2nd}}\sqrt{\frac{\kappa d}{\mu r}}}_{=:\alpha_{1,6}} \\
&\lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right) + \frac{1}{np\sigma_r^\star}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2 + \frac{1}{np}\zeta_{\mathsf{2nd}}^2.
\end{aligned}$$

  Here, the penultimate relation uses (E.23) as well as (D.21), and the last line holds since

$$\begin{aligned}
\alpha_{1,1} &\lesssim \delta\frac{1}{np\sigma_1^{\star 2}}\omega_l^\star\left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2 \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right), \\
\alpha_{1,3} + \alpha_{1,5} + \alpha_{1,6} &\lesssim \delta\frac{1}{np\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2 \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right), \\
\alpha_{1,4} &\lesssim \delta\frac{1}{np\sigma_1^{\star 2}}\omega_l^{\star 2} \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right),
\end{aligned}$$

  provided that $np \gtrsim \delta^{-2}\kappa^2 r\log^2(n+d)$, $n \gtrsim \delta^{-2}\kappa^4(r+\log(n+d))$ and $\zeta_{\mathsf{2nd}}\sqrt{d} \lesssim \delta\sqrt{\mu r/\kappa^3}$.

- Regarding $\|\boldsymbol{A}_2 - \boldsymbol{B}_2\|$, it is observed that

$$\begin{aligned}
\|\boldsymbol{A}_2 - \boldsymbol{B}_2\| &= \left\|\frac{\omega_l^{\star 2}}{np}\boldsymbol{R}\left(\boldsymbol{\Sigma}^\star\right)^{-2}\boldsymbol{R}^\top - \frac{\omega_l^2}{np}\left(\boldsymbol{\Sigma}\right)^{-2}\right\| \\
&\le \underbrace{\frac{\omega_l^{\star 2}}{np}\left\|\boldsymbol{R}\left(\boldsymbol{\Sigma}^\star\right)^{-2}\boldsymbol{R}^\top - \left(\boldsymbol{\Sigma}\right)^{-2}\right\|}_{=:\alpha_{2,1}} + \underbrace{\frac{\left|\omega_l^{\star 2} - \omega_l^2\right|}{np}\left\|\left(\boldsymbol{\Sigma}\right)^{-2}\right\|}_{=:\alpha_{2,2}}.
\end{aligned}$$

  The first term $\alpha_{2,1}$ can be bounded by

$$\alpha_{2,1} \overset{\text{(i)}}{\lesssim} \frac{\omega_l^{\star 2}}{np}\frac{1}{\sigma_r^{\star 2}}\left(\sqrt{\frac{\kappa^2 r\log(n+d)}{n}} + \zeta_{\mathsf{2nd}}\sqrt{\frac{\kappa d}{\mu r}}\right) \overset{\text{(ii)}}{\lesssim} \frac{\delta}{np\sigma_1^{\star 2}}\omega_l^{\star 2} \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right).$$

Here, (i) comes from Lemma 11, whereas (ii) holds provided that $n \gtrsim \delta^{-2}\kappa^4 r\log(n+d)$ and $\zeta_{\mathsf{2nd}}\sqrt{d} \lesssim \delta\sqrt{\mu r/\kappa^3}$. For the second term $\alpha_{2,2}$, it holds that

$$\begin{aligned}
\alpha_{2,2} &\overset{\text{(i)}}{\lesssim} \frac{1}{np\sigma_r^{\star 2}}\left[\sqrt{\frac{\kappa r\log^2(n+d)}{np}}\left(\sqrt{\kappa}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2+\omega_l^{\star}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right)+\sqrt{\frac{\log^2(n+d)}{np}}\omega_l^{\star 2}+\sigma_1^{\star 2}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2+\sigma_1^{\star 2}\zeta_{\mathsf{2nd}}^2\right] \\
&\lesssim \frac{1}{np\sigma_r^{\star 2}}\sqrt{\frac{\kappa^2 r\log^2(n+d)}{np}}\left[\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2+\omega_l^{\star 2}+\omega_l^{\star}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right]+\frac{\kappa}{np}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2+\frac{\kappa}{np}\zeta_{\mathsf{2nd}}^2 \\
&\overset{\text{(ii)}}{\lesssim} \frac{\delta}{np\sigma_1^{\star 2}}\left[\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2+\omega_l^{\star 2}+\omega_l^{\star}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right]+\frac{\kappa}{np}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2+\frac{\kappa}{np}\zeta_{\mathsf{2nd}}^2 \\
&\lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)+\frac{\kappa}{np}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2+\frac{\kappa}{np}\zeta_{\mathsf{2nd}}^2.
\end{aligned}$$

Here, (i) utilizes (E.17) and Lemma 11, while (ii) is valid as long as $np \gtrsim \delta^{-2}\kappa^4 r\log^2(n+d)$. These results combined yield

$$\left\|\boldsymbol{A}_2-\boldsymbol{B}_2\right\| \le \alpha_{2,1}+\alpha_{2,2} \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)+\frac{\kappa}{np}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2+\frac{\kappa}{np}\zeta_{\mathsf{2nd}}^2,$$

provided that $np \gtrsim \delta^{-2}\kappa^4 r\log^2(n+d)$ and $\zeta_{\mathsf{2nd}} \lesssim \delta\sqrt{\mu r/\kappa^3}$.

- We are left with $\|\boldsymbol{A}_3-\boldsymbol{B}_3\|$, where we have

$$\begin{aligned}
\left\|\boldsymbol{A}_3-\boldsymbol{B}_3\right\| &= \frac{2(1-p)}{np}\left\|\boldsymbol{R}\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{R}^{\top}-\boldsymbol{U}_{l,\cdot}^{\top}\boldsymbol{U}_{l,\cdot}\right\| = \frac{2(1-p)}{np}\left\|\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^{\star}-\boldsymbol{R}^{\top}\boldsymbol{U}_{l,\cdot}^{\top}\boldsymbol{U}_{l,\cdot}\boldsymbol{R}\right\| \\
&\le \frac{2}{np}\left\|\left(\boldsymbol{U}_{l,\cdot}^{\star}-\boldsymbol{U}_l\boldsymbol{R}\right)^{\top}\boldsymbol{U}_{l,\cdot}^{\star}\right\|+\frac{2}{np}\left\|\left(\boldsymbol{U}_l\boldsymbol{R}\right)^{\top}\left(\boldsymbol{U}_{l,\cdot}^{\star}-\boldsymbol{U}_{l,\cdot}\boldsymbol{R}\right)\right\| \\
&= \frac{2}{np}\left\|\boldsymbol{U}_{l,\cdot}^{\star}-\boldsymbol{U}_l\boldsymbol{R}\right\|_2\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{l,\cdot}\right\|_2\right) \\
&\lesssim \underbrace{\frac{1}{np}\left\|\boldsymbol{U}_{l,\cdot}^{\star}-\boldsymbol{U}_l\boldsymbol{R}\right\|_2\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2}_{=:\alpha_{3,1}}+\underbrace{\frac{1}{np}\left\|\boldsymbol{U}_{l,\cdot}^{\star}-\boldsymbol{U}_l\boldsymbol{R}\right\|_2^2}_{=:\alpha_{3,2}}.
\end{aligned}$$

The first term $\alpha_{3,1}$ can be bounded by

$$\begin{aligned}
\alpha_{3,1} &\overset{\text{(i)}}{\lesssim} \frac{1}{np}\left[\frac{1}{\sigma_r^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_l^{\star}\right)\sqrt{\frac{r\log^2(n+d)}{np}}+\zeta_{\mathsf{2nd}}\right]\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2 \\
&\lesssim \frac{1}{np\sigma_r^{\star 2}}\left[\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2+\omega_l^{\star}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right]\sqrt{\frac{r\log^2(n+d)}{np}}+\frac{1}{np}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2 \\
&\overset{\text{(ii)}}{\lesssim} \frac{\delta}{np\sigma_1^{\star 2}}\left[\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2+\omega_l^{\star}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right]+\frac{1}{np}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2 \\
&\lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)+\frac{1}{np}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2.
\end{aligned}$$

Here, (i) invokes (D.19), and (ii) holds provided that $np \gtrsim \delta^{-2}\kappa^2 r\log^2(n+d)$. The second term $\alpha_{3,2}$ can be bounded by

$$\begin{aligned}
\alpha_{3,2} &\overset{\text{(i)}}{\lesssim} \frac{1}{np\sigma_r^{\star 2}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2+\omega_l^{\star 2}\right)\frac{r\log^2(n+d)}{np}+\frac{1}{np}\zeta_{\mathsf{2nd}}^2 \\
&\overset{\text{(ii)}}{\lesssim} \frac{\delta}{np\sigma_1^{\star 2}}\left[\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2+\omega_l^{\star 2}\right]+\frac{1}{np}\zeta_{\mathsf{2nd}}^2 \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)+\frac{1}{np}\zeta_{\mathsf{2nd}}^2,
\end{aligned}$$

where (i) relies on (D.19), and (ii) holds provided that $np \gtrsim \delta^{-1}\kappa r \log^2(n+d)$. As a result, we conclude that

$$\|\boldsymbol{A}_3 - \boldsymbol{B}_3\| \lesssim \alpha_{3,1} + \alpha_{3,2} \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) + \frac{1}{np}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2 + \frac{1}{np}\zeta_{\mathsf{2nd}}^2,$$

with the proviso that $np \gtrsim \delta^{-2}\kappa^2 r \log^2(n+d)$.

Finally, taking the above bounds together yields

$$\begin{aligned}\left\|\boldsymbol{R}\boldsymbol{\Sigma}_l^{\star}\boldsymbol{R}^{\top} - \boldsymbol{\Sigma}_l\right\| &\le \|\boldsymbol{A}_1 - \boldsymbol{B}_1\| + \|\boldsymbol{A}_2 - \boldsymbol{B}_2\| + \|\boldsymbol{A}_3 - \boldsymbol{B}_3\| \\ &\lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) + \frac{\kappa}{np\sigma_r^{\star}}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{\kappa}{np}\zeta_{\mathsf{2nd}}^2,\end{aligned}$$

with the proviso that $np \gtrsim \delta^{-2}\kappa^4 r \log^2(n+d)$ and $\zeta_{\mathsf{2nd}}\sqrt{d} \lesssim \delta\sqrt{\mu r/\kappa^3}$. Under the assumption of Lemma 9, it is seen that

$$\zeta_{\mathsf{2nd}}\lambda_{\min}^{-1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) \lesssim \frac{1}{r^{1/4}\log^{1/2}(n+d)} \ll 1.$$

This in turn guarantees that

$$\begin{aligned}\left\|\boldsymbol{R}\boldsymbol{\Sigma}_l^{\star}\boldsymbol{R}^{\top} - \boldsymbol{\Sigma}_l\right\| &\lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) + \frac{\kappa}{np\sigma_r^{\star}}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{\kappa}{np}\zeta_{\mathsf{2nd}}^2 \\ &\overset{\text{(i)}}{\lesssim} \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) + \lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_l^{\star}\right)\sqrt{\frac{\kappa^3}{np}}\zeta_{\mathsf{2nd}} + \frac{\kappa}{np}\zeta_{\mathsf{2nd}}^2 \\ &\lesssim \lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)\left[\delta + \sqrt{\frac{\kappa^3}{np}}\zeta_{\mathsf{2nd}}\lambda_{\min}^{-1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) + \frac{\kappa}{np}\zeta_{\mathsf{2nd}}^2\lambda_{\min}^{-1}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)\right] \qquad \text{(E.24)} \\ &\overset{\text{(ii)}}{\asymp} \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),\end{aligned}$$

where (i) makes use of

$$\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_l^{\star}\right) \gtrsim \frac{1}{\sqrt{np}\sigma_1^{\star}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2,$$

which is a direct consequence of Lemma 8; and (ii) holds true provided that $np \gtrsim \delta^{-2}\kappa^3$.

E.2.4 Proof of Lemma 13

Recall the definition of the Euclidean ball $\mathcal{B}_{1-\alpha}$ in Algorithm 3. It is easily seen that

$$\begin{aligned}\left(\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star}\right)_{l,\cdot}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2} \in \mathcal{B}_{1-\alpha} \quad &\iff \quad \left(\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star}\right)_{l,\cdot}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\boldsymbol{R} \in \mathcal{B}_{1-\alpha} \\ &\iff \quad \left(\boldsymbol{U} - \boldsymbol{U}^{\star}\boldsymbol{R}^{\top}\right)_{l,\cdot}\boldsymbol{R}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\boldsymbol{R}^{\top} \in \mathcal{B}_{1-\alpha},\end{aligned}$$

where the last line comes from the rotational invariance of $\mathcal{B}_{1-\alpha}$. From the definition of $\mathsf{CR}_{U,l}^{1-\alpha}$ in Algorithm 3, we also know that

$$\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{R}^{\top} \in \mathsf{CR}_{U,l}^{1-\alpha} \quad \iff \quad \left(\boldsymbol{U} - \boldsymbol{U}^{\star}\boldsymbol{R}^{\top}\right)_{l,\cdot}\boldsymbol{\Sigma}_{U,l}^{-1/2} \in \mathcal{B}_{1-\alpha}.$$

Let us define

$$\boldsymbol{\Delta} \coloneqq \left(\boldsymbol{U} - \boldsymbol{U}^{\star}\boldsymbol{R}^{\top}\right)_{l,\cdot}\boldsymbol{R}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\boldsymbol{R}^{\top} - \left(\boldsymbol{U} - \boldsymbol{U}^{\star}\boldsymbol{R}^{\top}\right)_{l,\cdot}\boldsymbol{\Sigma}_{U,l}^{-1/2},$$

then it is straightforward to check that with probability exceeding $1 - O((n+d)^{-10})$

$$\begin{aligned}\|\boldsymbol{\Delta}\|_2 &\le \left\|\left(\boldsymbol{U} - \boldsymbol{U}^{\star}\boldsymbol{R}^{\top}\right)_{l,\cdot}\right\|_2\left\|\boldsymbol{R}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\boldsymbol{R}^{\top} - \boldsymbol{\Sigma}_{U,l}^{-1/2}\right\| \\ &= \left\|\left(\boldsymbol{U} - \boldsymbol{U}^{\star}\boldsymbol{R}^{\top}\right)_{l,\cdot}\right\|_2\left\|\boldsymbol{R}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\boldsymbol{R}^{\top}\left(\boldsymbol{\Sigma}_{U,l}^{1/2} - \boldsymbol{R}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{1/2}\boldsymbol{R}^{\top}\right)\boldsymbol{\Sigma}_{U,l}^{-1/2}\right\| \\ &\le \left\|\left(\boldsymbol{U} - \boldsymbol{U}^{\star}\boldsymbol{R}^{\top}\right)_{l,\cdot}\right\|_2\left\|\boldsymbol{R}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\boldsymbol{R}^{\top}\right\|\left\|\boldsymbol{R}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{1/2}\boldsymbol{R}^{\top} - \boldsymbol{\Sigma}_{U,l}^{1/2}\right\|\left\|\boldsymbol{\Sigma}_{U,l}^{-1/2}\right\|\end{aligned}$$

$$\lesssim \left\| \left(\boldsymbol{U} - \boldsymbol{U}^\star \boldsymbol{R}^\top \right)_{l,\cdot} \right\|_2 \lambda_{\min}^{-1} \left(\boldsymbol{\Sigma}_{U,l}^\star \right) \left\| \boldsymbol{R} \left(\boldsymbol{\Sigma}_{U,l}^\star \right)^{1/2} \boldsymbol{R}^\top - \boldsymbol{\Sigma}_{U,l}^{1/2} \right\|. \tag{E.25}$$

Here the last line follows from an immediate result from Lemma 12 and Weyl's inequality:

$$\lambda_{\min} \left(\boldsymbol{\Sigma}_{U,l} \right) \asymp \lambda_{\min} \left(\boldsymbol{\Sigma}_{U,l}^\star \right), \tag{E.26}$$

which holds as long as $\delta \ll 1$. Notice that

$$\begin{aligned} \left\| \boldsymbol{R} \left(\boldsymbol{\Sigma}_l^\star \right)^{1/2} \boldsymbol{R}^\top - \boldsymbol{\Sigma}_l^{1/2} \right\| &\overset{\text{(i)}}{\lesssim} \frac{1}{\lambda_{\min}^{1/2} \left(\boldsymbol{\Sigma}_l^\star \right) + \lambda_{\min}^{1/2} \left(\boldsymbol{\Sigma}_l \right)} \left\| \boldsymbol{R} \boldsymbol{\Sigma}_l^\star \boldsymbol{R}^\top - \boldsymbol{\Sigma}_l \right\| \\ &\overset{\text{(ii)}}{\lesssim} \lambda_{\min}^{-1/2} \left(\boldsymbol{\Sigma}_l^\star \right) \left\| \boldsymbol{R} \boldsymbol{\Sigma}_l^\star \boldsymbol{R}^\top - \boldsymbol{\Sigma}_l \right\| \\ &\overset{\text{(iii)}}{\lesssim} \lambda_{\min}^{1/2} \left(\boldsymbol{\Sigma}_{U,l}^\star \right) \delta, \end{aligned} \tag{E.27}$$

where (i) follows from the perturbation bound of matrix square root (Schmitt, 1992, Lemma 2.1); (ii) arises from (E.26); and (iii) is a consequence of Lemma 12. We can combine (E.25) and (E.27) to achieve

$$\begin{aligned} \|\boldsymbol{\Delta}\|_2 &\lesssim \left\| \left(\boldsymbol{U} - \boldsymbol{U}^\star \boldsymbol{R}^\top \right)_{l,\cdot} \right\|_2 \lambda_{\min}^{-1/2} \left(\boldsymbol{\Sigma}_{U,l}^\star \right) \delta \\ &\lesssim \left[\frac{1}{\sigma_r^\star} \left(\left\| \boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 + \omega_l^\star \right) \sqrt{\frac{r \log^2 (n+d)}{np}} + \zeta_{\mathsf{2nd}} \right] \lambda_{\min}^{-1/2} \left(\boldsymbol{\Sigma}_l^\star \right) \delta \\ &\asymp \frac{1}{\sigma_r^\star} \left(\left\| \boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 + \omega_l^\star \right) \sqrt{\frac{r \log^2 (n+d)}{np}} \lambda_{\min}^{-1/2} \left(\boldsymbol{\Sigma}_{U,l}^\star \right) \delta + \zeta_{\mathsf{2nd}} \lambda_{\min}^{-1/2} \left(\boldsymbol{\Sigma}_{U,l}^\star \right) \delta \end{aligned}$$

with probability exceeding $1 - O((n+d)^{-10})$. Let

$$\zeta \coloneqq \widetilde{C} \left[\frac{1}{\sigma_r^\star} \left(\left\| \boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 + \omega_l^\star \right) \sqrt{\frac{r \log^2 (n+d)}{np}} \lambda_{\min}^{-1/2} \left(\boldsymbol{\Sigma}_{U,l}^\star \right) \delta + \zeta_{\mathsf{2nd}} \lambda_{\min}^{-1/2} \left(\boldsymbol{\Sigma}_{U,l}^\star \right) \delta \right]$$

for some sufficiently large constant $\widetilde{C} > 0$ such that $\mathbb{P}(\|\boldsymbol{\Delta}\|_2 \le \zeta) \ge 1 - O((n+d)^{-10})$. Recalling the definition (A.3) of $\mathcal{C}^\varepsilon$ for any convex set $\mathcal{C}$, we have

$$\begin{aligned} \mathbb{P} \left(\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{R}^\top \in \mathsf{CR}_{U,l}^{1-\alpha} \right) &= \mathbb{P} \left(\left(\boldsymbol{U} - \boldsymbol{U}^\star \boldsymbol{R}^\top \right)_{l,\cdot} \boldsymbol{\Sigma}_l^{-1/2} \in \mathcal{B}_{1-\alpha} \right) \\ &= \mathbb{P} \left(\left(\boldsymbol{U} - \boldsymbol{U}^\star \boldsymbol{R}^\top \right)_{l,\cdot} \boldsymbol{\Sigma}_l^{-1/2} \in \mathcal{B}_{1-\alpha}, \|\boldsymbol{\Delta}\|_2 \le \zeta \right) + \mathbb{P} \left(\left(\boldsymbol{U} - \boldsymbol{U}^\star \boldsymbol{R}^\top \right)_{l,\cdot} \boldsymbol{\Sigma}_l^{-1/2} \in \mathcal{B}_{1-\alpha}, \|\boldsymbol{\Delta}\|_2 > \zeta \right) \\ &\le \mathbb{P} \left(\left(\boldsymbol{U} - \boldsymbol{U}^\star \boldsymbol{R}^\top \right)_{l,\cdot} \boldsymbol{R} \left(\boldsymbol{\Sigma}_l^\star \right)^{-1/2} \boldsymbol{R}^\top \in \mathcal{B}_{1-\alpha}^\zeta \right) + \mathbb{P} \left(\|\boldsymbol{\Delta}\|_2 > \zeta \right) \\ &\overset{\text{(i)}}{=} \mathbb{P} \left(\left(\boldsymbol{U} \boldsymbol{R} - \boldsymbol{U}^\star \right)_{l,\cdot} \left(\boldsymbol{\Sigma}_l^\star \right)^{-1/2} \in \mathcal{B}_{1-\alpha}^\zeta \right) + O \left((n+d)^{-10} \right) \\ &\overset{\text{(ii)}}{\le} \mathcal{N} \left(\boldsymbol{0}, \boldsymbol{I}_r \right) \left\{ \mathcal{B}_{1-\alpha}^\zeta \right\} + O \left(\log^{-1/2} (n+d) \right) \\ &\overset{\text{(iii)}}{\le} \mathcal{N} \left(\boldsymbol{0}, \boldsymbol{I}_r \right) \left\{ \mathcal{B}_{1-\alpha} \right\} + \zeta \left(0.59 r^{1/4} + 0.21 \right) + O \left(\log^{-1/2} (n+d) \right) \\ &\overset{\text{(iv)}}{\le} 1 - \alpha + O \left(\log^{-1/2} (n+d) \right). \end{aligned} \tag{E.28}$$

Here (i) holds since $\mathcal{B}_{1-\alpha}^\zeta$ is rotational invariant; (ii) uses Lemma 9; (iii) invokes Theorem 22; and (iv) makes use of the definition of $\mathcal{B}_{1-\alpha}$ in Algorithm 3 and holds under the condition $\zeta r^{1/4} \lesssim 1/\sqrt{\log(n+d)}$, i.e.,

$$\frac{1}{\sigma_r^\star} \left(\left\| \boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 + \omega_l^\star \right) \sqrt{\frac{r \log^2 (n+d)}{np}} \lambda_{\min}^{-1/2} \left(\boldsymbol{\Sigma}_l^\star \right) \delta + \zeta_{\mathsf{2nd}} \lambda_{\min}^{-1/2} \left(\boldsymbol{\Sigma}_l^\star \right) \delta \lesssim \frac{1}{r^{1/4} \sqrt{\log (n+d)}}.$$

We have learned from the proof of Lemma 9 (Step 3 in Appendix E.1.3) that

$$\zeta_{\mathsf{2nd}}\lambda_{\min}^{-1/2}(\boldsymbol{\Sigma}_l^\star) \lesssim 1/(r^{1/4}\log^{1/2}(n+d))$$

holds under the assumptions of Lemma 9. In addition, it is seen that

$$\begin{aligned}
&\frac{1}{\sigma_r^\star}\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\omega_l^\star\right)\sqrt{\frac{r\log^2(n+d)}{np}}\lambda_{\min}^{-1/2}(\boldsymbol{\Sigma}_l^\star)\,\delta \\
&\lesssim \frac{1}{\sigma_r^\star}\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\omega_l^\star\right)\sqrt{\frac{r\log^2(n+d)}{np}}\left(\frac{1-p}{np\sigma_1^{\star2}}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2+\frac{\omega_l^{\star2}}{np\sigma_1^{\star2}}\right)^{-1/2}\delta \\
&\lesssim \sqrt{\kappa r\log^2(n+d)}\delta.
\end{aligned}$$

Therefore, the condition $\zeta r^{1/4}\lesssim 1/\sqrt{\log(n+d)}$ is guaranteed by taking

$$\delta=\frac{1}{\kappa^{1/2}r^{3/4}\log^{3/2}(n+d)}.$$

Similar to (E.28)we can show that

$$\mathbb{P}\left(\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{R}^\top\in\mathsf{CR}_{U,l}^{1-\alpha}\right)\geq 1-\alpha+O\left(\log^{-1/2}(n+d)\right) \tag{E.29}$$

Taking (E.28) and (E.29) collectively yields

$$\mathbb{P}\left(\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{R}^\top\in\mathsf{CR}_{U,l}^{1-\alpha}\right)= 1-\alpha+O\left(\log^{-1/2}(n+d)\right).$$

To finish up, we need to check that when δ is chosen as the above quantity, the conditions in Lemma 12 can be guaranteed by $np\gtrsim\kappa^5r^{5/2}\log^5(n+d)$ and

$$\zeta_{\mathsf{2nd}}\sqrt{d}\lesssim\frac{\mu^{1/2}}{\kappa^2r^{1/4}\log^{3/2}(n+d)}. \tag{E.30}$$

Towards this, we make note of the following decomposition

$$\begin{aligned}
\zeta_{\mathsf{2nd}}\sqrt{d}&=\frac{\sigma_{\mathsf{ub}}^2\sqrt{(n+d)\,dr}\log(n+d)}{\sigma_r^{\star2}}+\left(\frac{\sqrt{\kappa}\sigma_{\mathsf{ub}}^2(n+d)}{\sigma_r^{\star2}}+\frac{\sigma_{\mathsf{ub}}\sqrt{r\log(n+d)}}{\sigma_r^\star}\right)\|\boldsymbol{U}^\star\|_{2,\infty}\sqrt{d} \\
&\asymp\underbrace{\frac{\sigma_{\mathsf{ub}}^2\sqrt{(n+d)\,dr}\log(n+d)}{\sigma_r^{\star2}}}_{=:\alpha_1}+\underbrace{\frac{\sigma_{\mathsf{ub}}^2(n+d)}{\sigma_r^{\star2}}\sqrt{\kappa\mu r}}_{=:\alpha_2}+\underbrace{\frac{\sigma_{\mathsf{ub}}\sqrt{\mu r^2\log(n+d)}}{\sigma_r^\star}}_{=:\alpha_3},
\end{aligned}$$

where

$$\sigma_{\mathsf{ub}}\asymp\sqrt{\frac{\mu r\log(n+d)}{ndp}}\sigma_1^\star+\frac{\omega_{\max}}{\sqrt{np}}.$$

Then we proceed to bound α_1, α_2 and α_3, separately.

- We begin with α_1 by showing that

$$\begin{aligned}
\alpha_1&\asymp\left(\frac{\mu r\log(n+d)}{ndp}\sigma_1^{\star2}+\frac{\omega_{\max}^2}{np}\right)\frac{\sqrt{(n+d)\,dr}\log(n+d)}{\sigma_r^{\star2}} \\
&\asymp\frac{\kappa\mu r^{3/2}\sqrt{(n+d)\,d}\log^2(n+d)}{ndp}+\frac{\omega_{\max}^2}{\sigma_r^{\star2}}\frac{\sqrt{(n+d)\,dr}\log(n+d)}{np} \\
&\lesssim\frac{\mu^{1/2}}{\kappa^2r^{1/4}\log^{3/2}(n+d)},
\end{aligned}$$

provided that $(n \wedge d)p \gtrsim \kappa^3 \mu^{1/2} r^{7/4} \log^{7/2}(n+d)$ and

$$\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{n+d}{np}} \lesssim \frac{1}{\kappa r^{3/8}\log^{5/4}(n+d)}.$$

- When it comes to α_2, we observe that

$$\begin{aligned}
\alpha_2 &\asymp \left(\frac{\mu r \log(n+d)}{ndp}\sigma_1^{\star 2} + \frac{\omega_{\max}^2}{np}\right)\frac{n+d}{\sigma_r^{\star 2}}\sqrt{\kappa\mu r} \\
&\asymp \frac{\kappa^{3/2}\mu^{3/2}r^{3/2}\log(n+d)}{(n\wedge d)\,p} + \frac{\omega_{\max}^2}{\sigma_r^{\star 2}}\frac{n+d}{np}\sqrt{\kappa\mu r} \\
&\lesssim \frac{\mu^{1/2}}{\kappa^2 r^{1/4}\log^{3/2}(n+d)},
\end{aligned}$$

provided that $(n \wedge d)p \gtrsim \kappa^{7/2}\mu r^{7/4}\log^{5/2}(n+d)$ and

$$\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{n+d}{np}} \lesssim \frac{1}{\kappa^{5/4} r^{3/8}\log^{3/4}(n+d)}.$$

- We are left with the term α_3. Observe that

$$\begin{aligned}
\alpha_3 &\asymp \left(\sqrt{\frac{\mu r\log(n+d)}{ndp}}\sigma_1^\star + \frac{\omega_{\max}}{\sqrt{np}}\right)\frac{\sqrt{\mu r^2\log(n+d)}}{\sigma_r^\star} \\
&\asymp \sqrt{\frac{\kappa\mu^2 r^3\log^2(n+d)}{ndp}} + \frac{\omega_{\max}}{\sqrt{np}}\frac{\sqrt{\mu r^2\log(n+d)}}{\sigma_r^\star} \\
&\lesssim \frac{\mu^{1/2}}{\kappa^2 r^{1/4}\log^{3/2}(n+d)},
\end{aligned}$$

with the proviso that $ndp \gtrsim \kappa^5 \mu r^{7/2}\log^5(n+d)$ and

$$\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{1}{np}} \lesssim \frac{1}{\kappa^2 r^{5/4}\log^2(n+d)}.$$

With the above calculations in mind, we see that (E.30) can be guaranteed by the following assumptions

$$n\wedge d \gtrsim \kappa^{3/2} r^{7/4}\log^{5/2}(n+d), \quad (n\wedge d)\,p \gtrsim \kappa^{7/2}\mu r^{7/4}\log^{5/2}(n+d)$$

$$\text{and} \qquad \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{n+d}{np}} \lesssim \frac{1}{\kappa^{5/4} r^{3/8}\log^{5/4}(n+d)},$$

thus concluding the proof.

E.3 Auxiliary lemmas for Theorem 13

E.3.1 Proof of Lemma 14

Note that all the probabilistic arguments in this section are conditional on $\boldsymbol{F}$, and we shall always assume the occurrence of the $\sigma(\boldsymbol{F})$-measurable high-probability event $\mathcal{E}_{\mathsf{good}}$. We can follow the same analysis as in Appendix E.1.1 (Proof of Lemma 7) to obtain

$$\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^\natural = \boldsymbol{Z} + \boldsymbol{\Psi}_{\boldsymbol{U}}$$

with $\boldsymbol{R}_{\boldsymbol{U}} = \arg\min_{\boldsymbol{O}\in\mathcal{O}^{r\times r}} \|\boldsymbol{U}\boldsymbol{O} - \boldsymbol{U}^{\natural}\|_{\mathrm{F}}^2$, where

$$\boldsymbol{Z} = \boldsymbol{E}\boldsymbol{V}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}$$

and $\boldsymbol{\Psi}_{\boldsymbol{U}}$ is a residual matrix that satisfies

$$\mathbb{P}\left(\|\boldsymbol{\Psi}_{\boldsymbol{U}}\|_{2,\infty}\mathbb{1}_{\mathcal{E}_{\mathsf{good}}} \lesssim \zeta_{\mathsf{2nd}} \,\middle|\, \boldsymbol{F}\right) \geq 1 - O\left((n+d)^{-10}\right). \tag{E.31}$$

Here, ζ_{2nd} is a quantity defined in Lemma 7. From (E.18), we can see that

$$\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}^2\boldsymbol{R}_{\boldsymbol{U}} = \boldsymbol{\Sigma}^{\natural 2} + \boldsymbol{\Psi}_{\boldsymbol{\Sigma}}$$

holds for some matrix $\boldsymbol{\Psi}_{\boldsymbol{\Sigma}}$ satisfying

$$\mathbb{P}\left(\|\boldsymbol{\Psi}_{\boldsymbol{\Sigma}}\|\mathbb{1}_{\mathcal{E}_{\mathsf{good}}} \lesssim \sqrt{\frac{d}{\mu r}}\zeta_{\mathsf{2nd}}\sigma_r^{\star}\sigma_1^{\star} \,\middle|\, \boldsymbol{F}\right) \geq 1 - O\left((n+d)^{-10}\right). \tag{E.32}$$

This allows one to further derive

$$\begin{aligned}
\boldsymbol{S} - \boldsymbol{M}^{\natural}\boldsymbol{M}^{\natural\top} &= \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^{\top} - \boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{U}^{\natural\top} \\
&= \boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}^2\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{U}^{\top} - \boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{U}^{\natural\top} \\
&= \boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{U}^{\top} + \boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{\Psi}_{\boldsymbol{\Sigma}}\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{U}^{\top} - \boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{U}^{\natural\top} \\
&= \left(\boldsymbol{U}^{\natural} + \boldsymbol{Z} + \boldsymbol{\Psi}_{\boldsymbol{U}}\right)\boldsymbol{\Sigma}^{\natural 2}\left(\boldsymbol{U}^{\natural} + \boldsymbol{Z} + \boldsymbol{\Psi}_{\boldsymbol{U}}\right)^{\top} + \boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{\Psi}_{\boldsymbol{\Sigma}}\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{U}^{\top} - \boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{U}^{\natural\top} \\
&= \boldsymbol{X} + \boldsymbol{\Phi},
\end{aligned}$$

where the matrices $\boldsymbol{X}$ and $\boldsymbol{\Phi}$ are given by

$$\boldsymbol{X} = \boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{Z}^{\top} + \boldsymbol{Z}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{U}^{\natural} = \boldsymbol{E}\boldsymbol{M}^{\natural\top} + \boldsymbol{M}^{\natural}\boldsymbol{E}^{\top},$$

$$\boldsymbol{\Phi} = \boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{\Psi}_{\boldsymbol{U}}^{\top} + \boldsymbol{Z}\boldsymbol{\Sigma}^{\natural 2}\left(\boldsymbol{Z} + \boldsymbol{\Psi}_{\boldsymbol{U}}\right)^{\top} + \boldsymbol{\Psi}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural 2}\left(\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\right)^{\top} + \boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{\Psi}_{\boldsymbol{\Sigma}}\left(\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\right)^{\top}.$$

Recall from Lemma 40 that for each $l \in [d]$,

$$\|\boldsymbol{Z}_{l,\cdot}\|_2 \lesssim \frac{1}{\sigma_r^{\star}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\sqrt{\frac{r\log^2(n+d)}{np}} + \frac{\omega_l^{\star}}{\sigma_r^{\star}}\sqrt{\frac{r\log(n+d)}{np}} \tag{E.33}$$

and

$$\left\|(\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star})_{l,\cdot}\right\|_2 \lesssim \frac{1}{\sigma_r^{\star}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\sqrt{\frac{r\log^2(n+d)}{np}} + \frac{\omega_l^{\star}}{\sigma_r^{\star}}\sqrt{\frac{r\log(n+d)}{np}} + \zeta_{\mathsf{2nd}}$$

hold with probability at least $1 - O\big((n+d)^{-10}\big)$. As an immediate consequence, we obtain

$$\begin{aligned}
\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star}\|_{2,\infty} = \max_{l\in[d]}\left\|(\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star})_{l,\cdot}\right\|_2 &\lesssim \sqrt{\frac{\kappa\mu r^2\log^2(n+d)}{ndp}} + \frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{r\log(n+d)}{np}} + \zeta_{\mathsf{2nd}} \\
&\asymp \frac{\sigma_{\mathsf{ub}}}{\sigma_r^{\star}}\sqrt{r\log(n+d)} + \zeta_{\mathsf{2nd}},
\end{aligned}$$

which further implies that

$$\|\boldsymbol{U}\|_{2,\infty} \leq \|\boldsymbol{U}^{\star}\|_{2,\infty} + \|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star}\|_{2,\infty} \lesssim \sqrt{\frac{\mu r}{d}} + \frac{\sigma_{\mathsf{ub}}}{\sigma_r^{\star}}\sqrt{r\log(n+d)} + \zeta_{\mathsf{2nd}} \asymp \sqrt{\frac{\mu r}{d}}, \tag{E.34}$$

provided that $\sigma_{\mathsf{ub}}\sqrt{n+d} \lesssim \sigma_r^{\star}/\sqrt{\kappa\log(n+d)}$. Therefore, with probability exceeding $1 - O((n+d)^{-10})$, we have

$$|\Phi_{i,j}| \overset{\text{(i)}}{\lesssim} \sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2\|\boldsymbol{\Psi}_{\boldsymbol{U}}\|_{2,\infty} + \|\boldsymbol{Z}_{i,\cdot}\|_2\|\boldsymbol{Z}_{j,\cdot}\|_2 + \|\boldsymbol{Z}_{i,\cdot}\|_2\|\boldsymbol{\Psi}_{\boldsymbol{U}}\|_{2,\infty} + \|\boldsymbol{\Psi}_{\boldsymbol{U}}\|_{2,\infty}\|\boldsymbol{U}_{j,\cdot}\|_2\right)$$

$$
\begin{aligned}
&+ \|\boldsymbol{\Psi}_{\boldsymbol{\Sigma}}\| \|\boldsymbol{U}_{i,\cdot}\|_2 \|\boldsymbol{U}_{j,\cdot}\|_2 \\
&\overset{\text{(ii)}}{\lesssim} \zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \|\boldsymbol{Z}_{i,\cdot}\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 + \|\boldsymbol{Z}_{j,\cdot}\|_2 + \zeta_{\mathsf{2nd}}\right) + \sigma_1^{\star 2}\|\boldsymbol{Z}_{i,\cdot}\|_2\|\boldsymbol{Z}_{j,\cdot}\|_2 \\
&+ \sqrt{\frac{d}{\mu r}}\zeta_{\mathsf{2nd}}\sigma_r^{\star}\sigma_1^{\star}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \|\boldsymbol{Z}_{i,\cdot}\|_2 + \zeta_{\mathsf{2nd}}\right)\sqrt{\frac{\mu r}{d}} \\
&\lesssim \zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \|\boldsymbol{Z}_{i,\cdot}\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 + \|\boldsymbol{Z}_{j,\cdot}\|_2 + \zeta_{\mathsf{2nd}}\right) + \sigma_1^{\star 2}\|\boldsymbol{Z}_{i,\cdot}\|_2\|\boldsymbol{Z}_{j,\cdot}\|_2 \\
&\overset{\text{(iii)}}{\lesssim} \zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 + \frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{r\log(n+d)}{np}} + \zeta_{\mathsf{2nd}}\right) \\
&+ \frac{\kappa r\log^2(n+d)}{np}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right)\left(\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right) \\
&\overset{\text{(iv)}}{\lesssim} \zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right) + \frac{\kappa r\log^2(n+d)}{np}\omega_{\max}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right) \\
&+ \frac{\kappa r\log^2(n+d)}{np}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{\kappa r\log^2(n+d)}{np}\omega_{\max}^2 + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}
\end{aligned}
$$

for all $i, j \in [d]$. Here, (i) arises from (D.10) and the fact that $\boldsymbol{U}^{\natural} = \boldsymbol{U}^{\star}\boldsymbol{Q}$ for some orthonormal matrix $\boldsymbol{Q}$ (cf. (D.4)); (ii) invokes (E.31), (E.32) and (E.34); (iii) follows from (E.33) and holds as long as $np \gtrsim \kappa r\log^2(n+d)$; and (iv) invokes the AM-GM inequality.

E.3.2 Proof of Lemma 15

In this subsection, we shall focus on establishing the claimed result for the case when $i \neq j$; the case when $i = j$ can be proved in a similar (in fact, easier) manner. In view of the expression (D.25), we can write

$$
\mathsf{var}\left(X_{i,j}|\boldsymbol{F}\right) = \underbrace{\sum_{l=1}^{n} M_{j,l}^{\natural 2}\sigma_{i,l}^2}_{=:\alpha_1} + \underbrace{\sum_{l=1}^{n} M_{i,l}^{\natural 2}\sigma_{j,l}^2}_{=:\alpha_2},
$$

thus motivating us to study the behavior of α_1 and α_2, respectively.

- Let us begin with the term α_1. By virtue of (D.2) and (D.8), we have

$$
\begin{aligned}
\alpha_1 = \sum_{l=1}^{n} M_{j,l}^{\natural 2}\sigma_{i,l}^2 &= \sum_{l=1}^{n}\left(\frac{1}{\sqrt{n}}\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2\left[\frac{1-p}{np}\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2 + \frac{\omega_i^{\star 2}}{np}\right] \\
&= \frac{1-p}{n^2 p}\sum_{l=1}^{n}\left(\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2 + \frac{\omega_i^{\star 2}}{n^2 p}\sum_{l=1}^{n}\left(\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2.
\end{aligned}
$$

  On the event $\mathcal{E}_{\mathsf{good}}$ (see Lemma 6), we know from the basic facts in (E.62a) and (E.62b) that

$$
\begin{aligned}
\left|\frac{1}{n}\sum_{j=1}^{n}\left(\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2 - S_{i,i}^{\star}S_{j,j}^{\star} - 2S_{i,j}^{\star 2}\right| &\lesssim \sqrt{\frac{\log^3(n+d)}{n}}S_{i,i}^{\star}S_{j,j}^{\star}; \\
\left|\frac{1}{n}\sum_{l=1}^{n}\left(\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2 - S_{j,j}^{\star}\right| &\lesssim \sqrt{\frac{\log(n+d)}{n}}S_{j,j}^{\star}.
\end{aligned}
$$

  As a result, we can express α_1 as

$$
\alpha_1 = \underbrace{\frac{1-p}{np}\left(S_{i,i}^{\star}S_{j,j}^{\star} + 2S_{i,j}^{\star 2}\right) + \frac{\omega_i^{\star 2}}{np}S_{j,j}^{\star}}_{=:\alpha_1^{\star}} + r_1
$$

for some residual term r_1 satisfying

$$|r_1| \lesssim \sqrt{\frac{\log^3(n+d)}{n}} \alpha_1^\star.$$

- Akin to our analysis for α_1, we can also demonstrate that

$$\alpha_2 = \underbrace{\frac{1-p}{np}\left[S_{i,i}^\star S_{j,j}^\star + 2S_{i,j}^{\star 2}\right] + \frac{\omega_j^{\star 2}}{np} S_{i,i}^\star}_{=:\alpha_2^\star} + r_2$$

for some residual term r_2 obeying

$$|r_2| \lesssim \sqrt{\frac{\log^3(n+d)}{n}} \alpha_2^\star.$$

Therefore we conclude that

$$\mathsf{var}\left(X_{i,j}|\boldsymbol{F}\right) = \alpha_1 + \alpha_2 = \widetilde{v}_{i,j} + r_{i,j},$$

where

$$\widetilde{v}_{i,j} \coloneqq \alpha_1^\star + \alpha_2^\star = \frac{2(1-p)}{np}\left(S_{i,i}^\star S_{j,j}^\star + 2S_{i,j}^{\star 2}\right) + \frac{1}{np}\left(\omega_i^{\star 2} S_{j,j}^\star + \omega_j^{\star 2} S_{i,i}^\star\right)$$

and the residual term $r_{i,j}$ is bounded in magnitude by

$$|r_{i,j}| \leq |r_1| + |r_2| \lesssim \sqrt{\frac{\log^3(n+d)}{n}} \left(\alpha_1^\star + \alpha_2^\star\right) \lesssim \sqrt{\frac{\log^3(n+d)}{n}} \widetilde{v}_{i,j}.$$

It is also straightforward to obtain a sharp lower bound on $\widetilde{v}_{i,j}$ as follows:

$$\widetilde{v}_{i,j} \gtrsim \frac{1-p}{np}\left\|\boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2 \left\|\boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2 + \frac{1}{np}\omega_{\min}^2 \left(\left\|\boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2 + \left\|\boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2\right). \tag{E.35}$$

E.3.3 Proof of Lemma 16

Note that all the probabilistic arguments in this subsection are conditional on $\boldsymbol{F}$, and we shall always assume the occurrence of the $\sigma(\boldsymbol{F})$-measurable high-probability event $\mathcal{E}_{\mathsf{good}}$.

Step 1: Gaussian approximation of $X_{i,j}$ using the Berry-Esseen Theorem. We can write $X_{i,j}$ (cf. (D.24)) as a sum of independent mean-zero random variables

$$X_{i,j} = \sum_{l=1}^{n} \underbrace{\left[M_{j,l}^\natural E_{i,l} + M_{i,l}^\natural E_{j,l}\right]}_{=:\,Y_l}.$$

Then we can apply the Berry-Esseen Theorem (cf. Theorem 19) to arrive at

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(X_{i,j}/\sqrt{\mathsf{var}\left(X_{i,j}|\boldsymbol{F}\right)} \leq t \,\middle|\, \boldsymbol{F}\right) - \Phi(t)\right| \lesssim \gamma(\boldsymbol{F}),$$

where

$$\gamma(\boldsymbol{F}) \coloneqq \frac{1}{\mathsf{var}^{3/2}\left(X_{i,j}|\boldsymbol{F}\right)} \sum_{l=1}^{n} \mathbb{E}\left[\left|Y_l\right|^3 \middle| \boldsymbol{F}\right].$$

It thus boils down to controlling $\gamma(\boldsymbol{F})$. For any $l \in [n]$, we observe that

$$|Y_l| \leq B_i \left\|\boldsymbol{M}_{j,\cdot}^{\natural}\right\|_{\infty} + B_j \left\|\boldsymbol{M}_{i,\cdot}^{\natural}\right\|_{\infty} \leq B_i \sigma_1^{\natural} \left\|\boldsymbol{U}_{j,\cdot}^{\natural}\right\|_2 \left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} + B_j \sigma_1^{\natural} \left\|\boldsymbol{U}_{i,\cdot}^{\natural}\right\|_2 \left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty},$$

where B_i and B_j are two quantities obeying

$$\max_{l\in[n]} |E_{i,l}| \leq B_i \qquad \text{and} \qquad \max_{l\in[n]} |E_{j,l}| \leq B_j.$$

In view of (D.10), (D.15), (D.13) and the fact that $\boldsymbol{U}^{\natural} = \boldsymbol{U}^{\star}\boldsymbol{Q}$, we can derive

$$\begin{aligned}
\max_{l\in[d]} |Y_l| &\leq B_i \sigma_1^{\natural} \left\|\boldsymbol{U}_{j,\cdot}^{\natural}\right\|_2 \left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} + B_j \sigma_1^{\natural} \left\|\boldsymbol{U}_{i,\cdot}^{\natural}\right\|_2 \left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} \\
&\lesssim \frac{1}{p}\sqrt{\frac{\log(n+d)}{n}} \left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_i^{\star}\right) \sigma_1^{\star} \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \sqrt{\frac{\log(n+d)}{n}} \\
&\quad + \frac{1}{p}\sqrt{\frac{\log(n+d)}{n}} \left(\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_j^{\star}\right) \sigma_1^{\star} \left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 \sqrt{\frac{\log(n+d)}{n}} \\
&\lesssim \frac{\sqrt{\kappa}\log(n+d)}{np} \left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{\log(n+d)}{np}\omega_{\max}\sigma_1^{\star} \left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right).
\end{aligned} \tag{E.36}$$

The preceding bounds allow us to reach

$$\begin{aligned}
\gamma(\boldsymbol{F}) &\leq \frac{\max_{l\in[d]}|Y_l|}{\mathsf{var}^{3/2}(X_{i,j}|\boldsymbol{F})} \sum_{l=1}^{n} \mathbb{E}\left[Y_l^2 \middle| \boldsymbol{F}\right] \overset{\text{(i)}}{=} \frac{\max_{l\in[d]}|Y_l|}{\mathsf{var}^{1/2}(X_{i,j}|\boldsymbol{F})} \\
&\overset{\text{(ii)}}{\lesssim} \frac{1}{\widetilde{v}_{i,j}^{1/2}} \frac{\sqrt{\kappa}\log(n+d)}{np} \left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{1}{\widetilde{v}_{i,j}^{1/2}} \frac{\log(n+d)}{np}\omega_{\max}\sigma_1^{\star}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right) \\
&\overset{\text{(iii)}}{\lesssim} \frac{\sqrt{\kappa}\log(n+d)}{\sqrt{np}} \lesssim \frac{1}{\sqrt{\log(n+d)}}.
\end{aligned}$$

Here, (i) follows from the fact that $\mathsf{var}(X_{i,j}|\boldsymbol{F}) = \sum_{l=1}^{n} \mathbb{E}[Y_l^2|\boldsymbol{F}]$; (ii) follows from Lemma 15 and (E.36), provided that $n \gg \log^3(n+d)$; (iii) follows from

$$\widetilde{v}_{i,j}^{1/2} \gtrsim \frac{1}{\sqrt{np}} \left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{1}{\sqrt{np}}\omega_{\min}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right), \tag{E.37}$$

which is a direct consequence of (D.28) in Lemma 15; and (iv) is valid with the proviso that $np \gtrsim \kappa\log^3(n+d)$. As a consequence, we arrive at

$$\sup_{t\in\mathbb{R}} \left|\mathbb{P}\left(X_{i,j}/\sqrt{\mathsf{var}(X_{i,j}|\boldsymbol{F})} \leq t \,\middle|\, \boldsymbol{F}\right) - \Phi(t)\right| \lesssim \frac{1}{\sqrt{\log(n+d)}}, \tag{E.38}$$

provide that $np \gtrsim \kappa\log^3(n+d)$ and $n \gg \log^3(n+d)$.

Step 2: derandomizing the conditional variance. In this step, we intend to replace $\mathsf{var}(X_{i,j}|\boldsymbol{F})$ in (E.38) with $\widetilde{v}_{i,j}$. Towards this, it is first observed that

$$\begin{aligned}
\mathbb{P}\left(X_{i,j}/\sqrt{\widetilde{v}_{i,j}} \leq t \,\middle|\, \boldsymbol{F}\right) - \Phi(t) &= \underbrace{\mathbb{P}\left(X_{i,j}/\sqrt{\widetilde{v}_{i,j}} \leq t \,\middle|\, \boldsymbol{F}\right) - \mathbb{P}\left(X_{i,j}/\sqrt{\mathsf{var}(X_{i,j}|\boldsymbol{F})} \leq t \,\middle|\, \boldsymbol{F}\right)}_{=:\alpha_1} \\
&\quad + \underbrace{\mathbb{P}\left(X_{i,j}/\sqrt{\mathsf{var}(X_{i,j}|\boldsymbol{F})} \leq t \,\middle|\, \boldsymbol{F}\right) - \Phi(t)}_{=:\alpha_2}.
\end{aligned}$$

Regarding the second term α_2, it follows from (E.38) that: when $\mathcal{E}_{\mathsf{good}}$ occurs, one has

$$|\alpha_2| \lesssim \frac{1}{\sqrt{\log(n+d)}}.$$

We then turn attention to bounding α_1. In view of Lemma 15, we know that: when $\mathcal{E}_{\mathsf{good}}$ happens, one has

$$\mathsf{var}(X_{i,j}|\boldsymbol{F}) = \widetilde{v}_{i,j} + O\left(\sqrt{\frac{\log^3(n+d)}{n}}\right)\widetilde{v}_{i,j}. \tag{E.39}$$

An immediate consequence is that

$$\frac{1}{2}\widetilde{v}_{i,j} \le \mathsf{var}(X_{i,j}|\boldsymbol{F}) \le 2\widetilde{v}_{i,j}, \tag{E.40}$$

with the proviso that $n \gg \log^3(n+d)$. Taking the above two relations collectively yields

$$\left|\sqrt{\mathsf{var}(X_{i,j}|\boldsymbol{F})} - \sqrt{\widetilde{v}_{i,j}}\right| = \frac{|\mathsf{var}(X_{i,j}|\boldsymbol{F}) - \widetilde{v}_{i,j}|}{\sqrt{\mathsf{var}(X_{i,j}|\boldsymbol{F})} + \sqrt{\widetilde{v}_{i,j}}} \le \underbrace{\widetilde{C}\sqrt{\frac{\log(n+d)^3}{n}}}_{=:\delta}\sqrt{\widetilde{v}_{i,j}} \tag{E.41}$$

for some sufficiently large constant $\widetilde{C} > 0$. Consequently, we arrive at

$$\begin{aligned}
\mathbb{P}\left(X_{i,j}/\sqrt{\widetilde{v}_{i,j}} \le t \,\middle|\, \boldsymbol{F}\right) &= \mathbb{P}\left(X_{i,j} \le t\sqrt{\widetilde{v}_{i,j}} \,\middle|\, \boldsymbol{F}\right) \overset{\text{(i)}}{\le} \mathbb{P}\left(X_{i,j} \le t\sqrt{\mathsf{var}(X_{i,j}|\boldsymbol{F})} + t\delta\sqrt{\widetilde{v}_{i,j}} \,\middle|\, \boldsymbol{F}\right) \\
&\overset{\text{(ii)}}{\le} \mathbb{P}\left(X_{i,j}/\sqrt{\mathsf{var}(X_{i,j}|\boldsymbol{F})} \le t + \sqrt{2}t\delta \,\middle|\, \boldsymbol{F}\right) \\
&\overset{\text{(iii)}}{\le} \Phi\left(t + \sqrt{2}t\delta\right) + O\left(\frac{1}{\sqrt{\log(n+d)}}\right) \overset{\text{(iv)}}{\le} \Phi(t) + O\left(\frac{1}{\sqrt{\log(n+d)}}\right) \\
&\overset{\text{(v)}}{\le} \mathbb{P}\left(X_{i,j}/\sqrt{\mathsf{var}(X_{i,j}|\boldsymbol{F})} \le t \,\middle|\, \boldsymbol{F}\right) + O\left(\frac{1}{\sqrt{\log(n+d)}}\right).
\end{aligned}$$

Here, (i) follows from (E.41); (ii) is a consequence of (E.40); (iii) and (v) come from (E.38); and (iv) arises from

$$\Phi\left(t + \sqrt{2}t\delta\right) - \Phi(t) = \int_t^{t+\sqrt{2}t\delta} \phi(s)\,\mathrm{d}s \le \sqrt{2}t\phi(t)\,\delta \le \sqrt{2}\cdot\sqrt{\frac{1}{2\pi e}}\delta \lesssim \frac{1}{\sqrt{\log(n+d)}}$$

provided that $n \gtrsim \log^4(n+d)$, where we use the fact that $\sup_{t\in\mathbb{R}} t\phi(t) = \phi(1) = \sqrt{1/(2\pi e)}$. Similarly, we can develop the following lower bound

$$\mathbb{P}\left(X_{i,j}/\sqrt{\widetilde{v}_{i,j}} \le t \,\middle|\, \boldsymbol{F}\right) \ge \left[\mathbb{P}\left(X_{i,j}/\sqrt{\mathsf{var}(X_{i,j}|\boldsymbol{F})} \le t \,\middle|\, \boldsymbol{F}\right) - O\left(\frac{1}{\sqrt{\log(n+d)}}\right)\right].$$

As a consequence, we arrive at

$$|\alpha_1| \lesssim \frac{1}{\sqrt{\log(n+d)}}.$$

Combine the preceding bounds on α_1 and α_2 to reach

$$\left|\mathbb{P}\left(X_{i,j}/\sqrt{\widetilde{v}_{i,j}} \le t \,\middle|\, \boldsymbol{F}\right) - \Phi(t)\right| \le (|\alpha_1| + |\alpha_2|) \lesssim \frac{1}{\sqrt{\log(n+d)}}, \qquad \forall t \in \mathbb{R}.$$

Step 3: taking higher-order errors into account. Similar to Step 3 in the proof of Lemma 9 (cf. Appendix E.1.3), we can see that

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\widetilde{v}_{i,j}^{-1/2}\left(\boldsymbol{S}-\boldsymbol{M}^{\natural}\boldsymbol{M}^{\natural\top}\right)_{i,j}\leq t\,\middle|\,\boldsymbol{F}\right)-\Phi\left(t\right)\right|\lesssim\frac{1}{\sqrt{\log\left(n+d\right)}}$$

holds as long as we can show that

$$\mathbb{P}\left(\widetilde{v}_{i,j}^{-1/2}\left|\Phi_{i,j}\right|\lesssim\log^{-1/2}\left(n+d\right)\,\middle|\,\boldsymbol{F}\right)\geq1-O\left(\left(n+d\right)^{-10}\right). \tag{E.42}$$

Therefore, it suffices to verify (E.42), or more generally, to derive sufficient conditions to guarantee $|\Phi_{i,j}|\lesssim\delta\widetilde{v}_{i,j}$ with high probability for any $\delta>0$. It comes from Lemma 14 that with probability exceeding $1-O((n+d)^{-10})$,

$$\begin{aligned}|\Phi_{i,j}|\lesssim\zeta_{i,j}\asymp&\underbrace{\zeta_{\mathsf{2nd}}\sigma_1^{\star2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)}_{=:\beta_1}+\underbrace{\frac{\kappa r\log^2\left(n+d\right)}{np}\omega_{\max}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right)}_{=:\beta_2}\\&+\underbrace{\frac{\kappa r\log^2\left(n+d\right)}{np}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2}_{=:\beta_3}+\underbrace{\frac{\kappa r\log^2\left(n+d\right)}{np}\omega_{\max}^2}_{=:\beta_4}+\underbrace{\zeta_{\mathsf{2nd}}^2\sigma_1^{\star2}}_{=:\beta_5}.\end{aligned}$$

Recalling the lower bound on $\widetilde{v}_{i,j}^{1/2}$ in (E.37), we can proceed to control these terms separately.

- Regarding β_1, we recall from the proof of Lemma 9 that

$$\zeta_{\mathsf{2nd}}\lesssim\frac{\delta}{\sqrt{\kappa}}\frac{1}{\sqrt{np}\sigma_1^{\star}}\left(\min\left\{\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2,\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right\}+\omega_{\min}\right), \tag{E.43}$$

  provided that $d\gtrsim\delta^{-2}\kappa^2\mu r^2\log(n+d)$,

$$\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{n+d}{np}}\lesssim\frac{1}{\sqrt{\kappa^2r\log^2\left(n+d\right)}}\wedge\delta\frac{\sqrt{1\wedge(d/n)}}{\kappa^{3/2}\mu^{1/2}r^{1/2}}$$

  and

$$\min\left\{\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2,\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right\}\gtrsim\delta^{-1}\left[\frac{\kappa^{3/2}\mu r\log^2\left(n+d\right)}{\sqrt{\left(n\wedge d\right)p}}+\frac{\kappa^2\mu^{3/2}r\log\left(n+d\right)}{\sqrt{np}\wedge\sqrt{d^2p/n}}+\frac{\kappa\mu r\log\left(n+d\right)}{\sqrt{d}}\right]\sqrt{\frac{r}{d}}\sigma_1^{\star}.$$

  This immediately leads to

$$\begin{aligned}\beta_1&\lesssim\frac{\delta}{\sqrt{\kappa}}\frac{1}{\sqrt{np}}\left(\min\left\{\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2,\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right\}+\omega_{\min}\right)\sigma_1^{\star}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)\\&\lesssim\delta\sqrt{\frac{1}{np}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\delta\sqrt{\frac{1}{np}}\omega_{\min}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right)\\&\lesssim\delta\widetilde{v}_{i,j}^{1/2}.\end{aligned}$$

- Regarding β_2, we can derive

$$\beta_2\lesssim\frac{\delta}{\sqrt{np}}\omega_{\min}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right)\lesssim\delta\widetilde{v}_{i,j}^{1/2},$$

  provided that $np\gtrsim\delta^{-2}\kappa^2r^2\log^4(n+d)$.

- When it comes to β_3, it is seen that

$$\beta_3 \lesssim \frac{\delta}{\sqrt{np}} \left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 \left\| \boldsymbol{U}_{j,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 \lesssim \delta \widetilde{v}_{i,j}^{1/2},$$

provided that $np \gtrsim \delta^{-2} \kappa^2 r^2 \log^4(n+d)$.

- Regarding β_4, we can calculate

$$\beta_4 \lesssim \frac{\delta}{\sqrt{np}} \omega_{\min} \left(\left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 + \left\| \boldsymbol{U}_{j,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 \right) \lesssim \delta \widetilde{v}_{i,j}^{1/2},$$

provided that

$$\left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 + \left\| \boldsymbol{U}_{j,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 \gtrsim \delta^{-1} \sqrt{\kappa r} \log^2 (n+d) \frac{\omega_{\max}}{\sigma_r^{\star}} \sqrt{\frac{d}{np}} \cdot \sqrt{\frac{r}{d}} \sigma_1^{\star}.$$

- Regarding β_5, we invoke (E.43) to obtain

$$\begin{aligned}
\beta_5 &\lesssim \frac{\delta^2}{\kappa} \frac{1}{np} \left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 \left\| \boldsymbol{U}_{j,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 + \frac{\delta^2}{\kappa} \frac{1}{np} \omega_{\min}^2 \\
&\lesssim \frac{\delta}{\sqrt{np}} \left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 \left\| \boldsymbol{U}_{j,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 + \frac{\delta}{\sqrt{np}} \omega_{\min} \left(\left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 + \left\| \boldsymbol{U}_{j,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 \right) \\
&\lesssim \delta \widetilde{v}_{i,j}^{1/2},
\end{aligned}$$

provided that $np \gtrsim 1$ and

$$\left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 + \left\| \boldsymbol{U}_{j,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 \gtrsim \frac{1}{\kappa^{3/2}} \frac{\omega_{\min}}{\sigma_r^{\star}} \sqrt{\frac{d}{np}} \cdot \sqrt{\frac{1}{d}} \sigma_1^{\star}.$$

With the above inequalities in place, we can conclude that

$$\widetilde{v}_{i,j}^{-1/2} \zeta_{i,j} \lesssim \widetilde{v}_{i,j}^{-1/2} \left(\beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 \right) \lesssim \delta,$$

with the proviso that $np \gtrsim \delta^{-2} \kappa^2 r^2 \log^4(n+d)$, $d \gtrsim \delta^{-2} \kappa^2 \mu r^2 \log(n+d)$,

$$\frac{\omega_{\max}}{\sigma_r^{\star}} \sqrt{\frac{n+d}{np}} \lesssim \frac{1}{\sqrt{\kappa^2 r \log^2 (n+d)}} \wedge \delta \frac{\sqrt{1 \wedge (d/n)}}{\kappa^{3/2} \mu^{1/2} r^{1/2}},$$

$$\min \left\{ \left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2, \left\| \boldsymbol{U}_{j,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 \right\} \gtrsim \delta^{-1} \left[\frac{\kappa^{3/2} \mu r \log^2 (n+d)}{\sqrt{(n \wedge d) p}} + \frac{\kappa^2 \mu^{3/2} r \log (n+d)}{\sqrt{np} \wedge \sqrt{d^2 p/n}} + \frac{\kappa \mu r \log (n+d)}{\sqrt{d}} \right] \sqrt{\frac{r}{d}} \sigma_1^{\star},$$

and

$$\left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 + \left\| \boldsymbol{U}_{j,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 \gtrsim \delta^{-1} \sqrt{\kappa r} \log^2 (n+d) \frac{\omega_{\max}}{\sigma_r^{\star}} \sqrt{\frac{d}{np}} \cdot \sqrt{\frac{r}{d}} \sigma_1^{\star}.$$

Taking $\delta \asymp \log^{-1/2}(n+d)$ then establishes the advertised result.

E.3.4 Proof of Lemma 17

Define

$$a_l = n^{-1} (\boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \boldsymbol{f}_l)(\boldsymbol{U}_{j,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \boldsymbol{f}_l)$$

for each $l = 1, \dots, n$. In view of the expression (D.29), we can write

$$A_{i,j} = \sum_{l=1}^{n} \left(a_l - \mathbb{E}\left[a_l \right] \right).$$

Apply Theorem 19 to show that

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}\left(\overline{v}_{i,j}^{-1/2}A_{i,j}\leq z\right)-\Phi\left(z\right)\right|\lesssim\gamma,$$

where $\overline{v}_{i,j}$ is defined in (D.30) and

$$\gamma\coloneqq\overline{v}_{i,j}^{-3/2}\sum_{l=1}^{n}\mathbb{E}\left[\left|a_l^3\right|\right].$$

It remains to control the term γ. Given that $\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\sim\mathcal{N}(0,\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\|_2^2)$ and $\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\sim\mathcal{N}(0,\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\|_2^2)$, it is straightforward to check that for each $l\in[n]$,

$$\mathbb{E}\left[\left|a_l^3\right|\right]\leq\frac{1}{n^3}\mathbb{E}\left[\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^6\right]^{1/2}\mathbb{E}\left[\left(\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^6\right]^{1/2}\lesssim\frac{1}{n^3}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^3\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^3.$$

Recognizing that

$$\overline{v}_{i,j}=\frac{1}{n}\left(S_{i,i}^{\star}S_{j,j}^{\star}+S_{i,j}^{\star 2}\right)\geq\frac{1}{n}S_{i,i}^{\star}S_{j,j}^{\star}=\frac{1}{n}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2,$$

we can combine the above bounds to arrive at

$$\gamma=\overline{v}_{i,j}^{-3/2}\sum_{l=1}^{n}\mathbb{E}\left[\left|a_l^3\right|\right]\lesssim\frac{n^{-2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^3\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^3}{n^{-3/2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^3\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^3}\lesssim\frac{1}{\sqrt{n}}$$

as claimed.

E.3.5 Proof of Lemma 18

For any $z\in\mathbb{R}$, we can decompose

$$\begin{aligned}\mathbb{P}\left(\left(S_{i,j}-S_{i,j}^{\star}\right)/\sqrt{v_{i,j}^{\star}}\leq z\right)&=\mathbb{P}\left(A_{i,j}+W_{i,j}\leq\sqrt{v_{i,j}^{\star}}z\right)=\mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{A_{i,j}+W_{i,j}\leq\sqrt{v_{i,j}^{\star}}z}\,\middle|\,\boldsymbol{F}\right]\right]\\&=\underbrace{\mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{A_{i,j}+W_{i,j}\leq\sqrt{v_{i,j}^{\star}}z}\mathbb{1}_{\mathcal{E}_{\mathsf{good}}}\,\middle|\,\boldsymbol{F}\right]\right]}_{=:\alpha_1}+\underbrace{\mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{A_{i,j}+W_{i,j}\leq\sqrt{v_{i,j}^{\star}}z}\mathbb{1}_{\mathcal{E}_{\mathsf{good}}^{\mathsf{c}}}\,\middle|\,\boldsymbol{F}\right]\right]}_{=:\alpha_2},\end{aligned}$$

leaving us with two terms to control.

- Regarding the first term α_1, we note that $\mathcal{E}_{\mathsf{good}}$ is $\sigma(\boldsymbol{F})$-measurable, and consequently,

$$\begin{aligned}\mathbb{E}\left[\mathbb{1}_{A_{i,j}+W_{i,j}\leq\sqrt{v_{i,j}^{\star}}z}\mathbb{1}_{\mathcal{E}_{\mathsf{good}}}\,\middle|\,\boldsymbol{F}\right]&=\mathbb{1}_{\mathcal{E}_{\mathsf{good}}}\mathbb{E}\left[\mathbb{1}_{A_{i,j}+W_{i,j}\leq\sqrt{v_{i,j}^{\star}}z}\,\middle|\,\boldsymbol{F}\right]\\&=\mathbb{1}_{\mathcal{E}_{\mathsf{good}}}\mathbb{P}\left(W_{i,j}\leq\sqrt{v_{i,j}^{\star}}z-A_{i,j}\,\middle|\,\boldsymbol{F}\right).\end{aligned}$$

  In view of Lemma 16, we can see that on the event $\mathcal{E}_{\mathsf{good}}$,

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\left(\widetilde{v}_{i,j}\right)^{-1/2}W_{i,j}\leq t\,\middle|\,\boldsymbol{F}\right)-\Phi\left(t\right)\right|\lesssim\frac{1}{\sqrt{\log\left(n+d\right)}}.$$

  Since $A_{i,j}$ is $\sigma(\boldsymbol{F})$-measurable, by choosing $t=\left(\widetilde{v}_{i,j}\right)^{-1/2}(\sqrt{v_{i,j}^{\star}}z-A_{i,j})$ we have

$$\left|\mathbb{P}\left(W_{i,j}\leq\sqrt{v_{i,j}^{\star}}z-A_{i,j}\,\middle|\,\boldsymbol{F}\right)-\Phi\left(\frac{\sqrt{v_{i,j}^{\star}}z-A_{i,j}}{\widetilde{v}_{i,j}^{1/2}}\right)\right|\mathbb{1}_{\mathcal{E}_{\mathsf{good}}}\lesssim\frac{1}{\sqrt{\log\left(n+d\right)}}.$$

  This in turn leads to

$$\alpha_1=\mathbb{E}\left[\Phi\left(\frac{\sqrt{v_{i,j}^{\star}}z-A_{i,j}}{\widetilde{v}_{i,j}^{1/2}}\right)\mathbb{1}_{\mathcal{E}_{\mathsf{good}}}\right]+O\left(\frac{1}{\sqrt{\log\left(n+d\right)}}\right)$$

$$
\begin{aligned}
&= \mathbb{E}\left[\Phi\left(\frac{\sqrt{v_{i,j}^{\star}}z - A_{i,j}}{\widetilde{v}_{i,j}^{1/2}}\right)\right] - \mathbb{E}\left[\Phi\left(\frac{\sqrt{v_{i,j}^{\star}}z - A_{i,j}}{\widetilde{v}_{i,j}^{1/2}}\right)\mathbb{1}_{\mathcal{E}_{\mathsf{good}}^{\mathsf{c}}}\right] + O\left(\frac{1}{\sqrt{\log(n+d)}}\right) \\
&= \mathbb{E}\left[\Phi\left(\frac{\sqrt{v_{i,j}^{\star}}z - A_{i,j}}{\widetilde{v}_{i,j}^{1/2}}\right)\right] + O\left(\frac{1}{\sqrt{\log(n+d)}}\right),
\end{aligned}
$$

where the last identity is valid since

$$
\mathbb{E}\left[\Phi\left(\frac{\sqrt{v_{i,j}^{\star}}z - A_{i,j}}{\widetilde{v}_{i,j}^{1/2}}\right)\mathbb{1}_{\mathcal{E}_{\mathsf{good}}^{\mathsf{c}}}\right] \le \mathbb{P}\left(\mathcal{E}_{\mathsf{good}}^{\mathsf{c}}\right) \lesssim (n+d)^{-10} \lesssim \frac{1}{\sqrt{\log(n+d)}}.
$$

Let $\phi(\cdot)$ denote the probability density of $\mathcal{N}(0,1)$, then it is readily seen that

$$
\begin{aligned}
\mathbb{E}\left[\Phi\left(\frac{\sqrt{v_{i,j}^{\star}}z - A_{i,j}}{\widetilde{v}_{i,j}^{1/2}}\right)\right] &= \mathbb{E}\left[\int_{-\infty}^{+\infty}\phi(t)\,\mathbb{1}_{t\le(\widetilde{v}_{i,j})^{-1/2}\left(\sqrt{v_{i,j}^{\star}}z - A_{i,j}\right)}\,\mathrm{d}t\right] \\
&\overset{\text{(i)}}{=} \int_{-\infty}^{+\infty}\mathbb{E}\left[\phi(t)\,\mathbb{1}_{t\le(\widetilde{v}_{i,j})^{-1/2}\left(\sqrt{v_{i,j}^{\star}}z - A_{i,j}\right)}\right]\mathrm{d}t \\
&= \int_{-\infty}^{+\infty}\phi(t)\,\mathbb{P}\left(t \le \left(\widetilde{v}_{i,j}\right)^{-1/2}\left(\sqrt{v_{i,j}^{\star}}z - A_{i,j}\right)\right)\mathrm{d}t \\
&= \int_{-\infty}^{+\infty}\phi(t)\,\mathbb{P}\left(A_{i,j} \le \sqrt{v_{i,j}^{\star}}z - t\widetilde{v}_{i,j}^{1/2}\right)\mathrm{d}t \\
&\overset{\text{(ii)}}{=} \int_{-\infty}^{+\infty}\phi(t)\,\Phi\left[\left(\overline{v}_{i,j}\right)^{-1/2}\left(\sqrt{v_{i,j}^{\star}}z - t\widetilde{v}_{i,j}^{1/2}\right)\right]\mathrm{d}t + O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

Here, (i) invokes Fubini's Theorem for nonnegative measurable functions, whereas (ii) follows from Lemma 17. Finally, letting U and V be two independent $\mathcal{N}(0,1)$ random variables, we can readily calculate

$$
\begin{aligned}
\int_{-\infty}^{+\infty}\phi(t)\,\Phi\left(\frac{z\sqrt{v_{i,j}^{\star}} - t\sqrt{\widetilde{v}_{i,j}}}{\sqrt{\overline{v}_{i,j}}}\right)\mathrm{d}t &= \mathbb{P}\left(U \le \frac{z\sqrt{v_{i,j}^{\star}} - V\sqrt{\widetilde{v}_{i,j}}}{\sqrt{\overline{v}_{i,j}}}\right) \\
&= \mathbb{P}\left(U\sqrt{\overline{v}_{i,j}} + V\sqrt{\widetilde{v}_{i,j}} \le z\sqrt{v_{i,j}^{\star}}\right) \\
&= \Phi(z),
\end{aligned}
$$

where the last relation follows from the fact that $U\sqrt{\overline{v}_{i,j}} + V\sqrt{\widetilde{v}_{i,j}} \sim \mathcal{N}(0, v_{i,j}^{\star})$. This allows us to conclude that

$$
\alpha_1 = \Phi(z) + O\left(\frac{1}{\sqrt{\log(n+d)}} + \frac{1}{\sqrt{n}}\right).
$$

- We then move on to the other term α_2. Towards this, it is straightforward to derive that

$$
\alpha_2 \le \mathbb{P}\left(\mathcal{E}_{\mathsf{good}}^{\mathsf{c}}\right) \lesssim (n+d)^{-10}.
$$

Note that the above analysis holds for all $z \in \mathbb{R}$. Therefore, for any $z \in \mathbb{R}$, taking the above calculation together yields

$$
\left|\mathbb{P}\left(\left(S_{i,j} - S_{i,j}^{\star}\right)/\sqrt{v_{i,j}^{\star}} \le z\right) - \Phi(z)\right| \le |\alpha_1 - \Phi(z)| + \alpha_2 \lesssim \frac{1}{\sqrt{\log(n+d)}} + \frac{1}{\sqrt{n}} \asymp \frac{1}{\sqrt{\log(n+d)}}
$$

as claimed, provided that $n \gtrsim \log(n+d)$.

E.4 Auxiliary lemmas for Theorem 14

E.4.1 Proof of Lemma 19

We focus on the case when $i \neq j$. The analysis for $i = j$ is similar and even more simple. Denote

$$v_{i,j}^{\star} = \underbrace{\frac{2-p}{np} S_{i,i}^{\star} S_{j,j}^{\star}}_{=:\alpha_1} + \underbrace{\frac{4-3p}{np} S_{i,j}^{\star 2}}_{=:\alpha_2} + \underbrace{\frac{1}{np}\left(\omega_i^{\star 2} S_{j,j}^{\star} + \omega_j^{\star 2} S_{i,i}^{\star}\right)}_{=:\alpha_3}$$

and

$$v_{i,j} = \underbrace{\frac{2-p}{np} S_{i,i} S_{j,j}}_{=:\beta_1} + \underbrace{\frac{4-3p}{np} S_{i,j}^{2}}_{=:\beta_2} + \underbrace{\frac{1}{np}\left(\omega_i^{2} S_{j,j} + \omega_j^{2} S_{i,i}\right)}_{=:\beta_3}.$$

From Lemma 15 we know that

$$v_{i,j}^{\star} \gtrsim \frac{1}{np} \left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{1}{np}\omega_{\min}^2 \left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\right).$$

Recall from the proof of Lemma 9 that

$$\zeta_{\mathsf{2nd}} \lesssim \frac{\varepsilon}{\sqrt{np\kappa}\sigma_1^{\star}} \left(\min\left\{\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2, \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right\} + \omega_{\min}\right) \tag{E.44}$$

and

$$\begin{aligned}\varepsilon v_{i,j}^{\star 1/2} &\gtrsim \zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right) + \frac{\kappa r \log^2(n+d)}{np}\omega_{\max}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right)\\ &\quad + \frac{\kappa r \log^2(n+d)}{np}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{\kappa r \log^2(n+d)}{np}\omega_{\max}^2 + \varepsilon\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}\end{aligned} \tag{E.45}$$

hold for any given $\varepsilon > 0$ provided that $np \gtrsim \varepsilon^{-2}\kappa^2 r^2 \log^4(n+d)$, $d \gtrsim \varepsilon^{-2}\kappa^2 \mu r^2 \log(n+d)$,

$$\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{n+d}{np}} \lesssim \frac{1}{\sqrt{\kappa^2 r \log^2(n+d)}} \wedge \varepsilon\frac{\sqrt{1\wedge(d/n)}}{\kappa^{3/2}\mu^{1/2}r^{1/2}},$$

$$\min\left\{\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2, \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right\} \gtrsim \varepsilon^{-1}\left[\frac{\kappa^{3/2}\mu r \log^2(n+d)}{\sqrt{(n\wedge d)p}} + \frac{\kappa^2\mu^{3/2} r \log(n+d)}{\sqrt{np}\wedge\sqrt{d^2p/n}} + \frac{\kappa\mu r\log(n+d)}{\sqrt{d}}\right]\sqrt{\frac{r}{d}}\sigma_1^{\star},$$

and

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \gtrsim \varepsilon^{-1}\sqrt{\kappa r}\log^2(n+d)\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}}\cdot\sqrt{\frac{r}{d}}\sigma_1^{\star}.$$

Step 1: bounding $|\alpha_1 - \beta_1|$. Recall from Lemma 11 that

$$\begin{aligned}\left|S_{i,i} - S_{i,i}^{\star}\right| &\lesssim \left(\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sqrt{\frac{r+\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{r\log^2(n+d)}{np}\omega_{\max}^2\\ &\quad + \left(\sqrt{\frac{r\log^2(n+d)}{np}}\omega_{\max} + \zeta_{\mathsf{2nd}}\sigma_r^{\star}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2};\end{aligned} \tag{E.46a}$$

$$\left|S_{j,j} - S_{j,j}^{\star}\right| \lesssim \left(\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sqrt{\frac{r+\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{r\log^2(n+d)}{np}\omega_{\max}^2$$

$$+\left(\sqrt{\frac{r\log^2(n+d)}{np}}\omega_{\max}+\zeta_{\mathsf{2nd}}\sigma_r^\star\right)\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}; \tag{E.46b}$$

Then we have

$$\begin{aligned}|\alpha_1-\beta_1| &\lesssim \frac{1}{np}\left|S_{i,i}S_{j,j}-S_{i,i}^\star S_{j,j}^\star\right| \lesssim \frac{1}{np}S_{i.i}\left|S_{j,j}-S_{j,j}^\star\right|+\frac{1}{np}S_{j,j}^\star\left|S_{i,i}-S_{i,i}^\star\right| \\ &\lesssim \underbrace{\frac{1}{np}S_{i.i}^\star\left|S_{j,j}-S_{j,j}^\star\right|+\frac{1}{np}S_{j,j}^\star\left|S_{i,i}-S_{i,i}^\star\right|}_{=:\gamma_{1,1}}+\underbrace{\frac{1}{np}\left|S_{i,i}-S_{i,i}^\star\right|\left|S_{j,j}-S_{j,j}^\star\right|}_{=:\gamma_{1,2}}.\end{aligned}$$

- Regarding $\gamma_{1,1}$, we can utilize (E.46) to obtain

$$\begin{aligned}\gamma_{1,1} &\lesssim \underbrace{\frac{1}{np}\left(\sqrt{\frac{r\log^2(n+d)}{np}}+\kappa\sqrt{\frac{r+\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2}_{=:\gamma_{1,1,1}} \\ &\quad+\underbrace{\frac{r\log^2(n+d)}{n^2p^2}\omega_{\max}^2\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\right)}_{=:\gamma_{1,1,2}} \\ &\quad+\underbrace{\frac{1}{np}\sqrt{\frac{r\log^2(n+d)}{np}}\omega_{\max}\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\right)}_{=:\gamma_{1,1,3}} \\ &\quad+\underbrace{\frac{1}{np}\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\right)}_{=:\gamma_{1,1,4}} \\ &\quad+\underbrace{\frac{1}{np}\zeta_{\mathsf{2nd}}\sigma_r^\star\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\right)}_{=:\gamma_{1,1,5}}\lesssim\delta v_{i,j}^\star,\end{aligned}$$

  where the last relation holds since

$$\begin{aligned}\gamma_{1,1,1} &\lesssim \frac{\delta}{np}\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\lesssim\delta v_{i,j}^\star, \\ \gamma_{1,1,2} &\lesssim \frac{\delta}{np}\omega_{\min}^2\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\right)\lesssim\delta v_{i,j}^\star, \\ \gamma_{1,1,3} &\lesssim \delta\frac{\omega_{\min}}{np}\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\right) \\ &\overset{\text{(i)}}{\lesssim} \frac{\delta}{np}\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2+\frac{\delta}{np}\omega_{\min}^2\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\right)\lesssim\delta v_{i,j}^\star, \\ \gamma_{1,1,4} &\lesssim \frac{1}{np}\left[\zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right)\right]^2\overset{\text{(ii)}}{\lesssim}\frac{\varepsilon^2}{np}v_{i,j}^\star\lesssim\delta v_{i,j}^\star \\ \gamma_{1,1,5} &\overset{\text{(iii)}}{\lesssim} \frac{1}{np}\frac{\varepsilon}{\sqrt{np}\kappa}\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2+\frac{\varepsilon\omega_{\min}}{n^{3/2}p^{3/2}\kappa}\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\right) \\ &\overset{\text{(iv)}}{\lesssim} \frac{\delta}{np}\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2+\frac{\delta}{np}\omega_{\min}^2\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\right)\lesssim\delta v_{i,j}^\star,\end{aligned}$$

  provided that $np \gtrsim \delta^{-2}r\log^2(n+d)$, $n\gtrsim\delta^{-2}\kappa^2 r\log(n+d)$, $\varepsilon\lesssim 1$. Here (i) and (iv) invoke the AM-GM inequality; (ii) utilizes (E.45); (iii) utilizes (E.44).

- Regarding $\gamma_{1,2}$, we also use (E.46) to achieve

$$
\begin{aligned}
\gamma_{1,2} &\lesssim \left(\sqrt{\frac{r\log^{2}(n+d)}{np}}+\kappa\sqrt{\frac{r+\log(n+d)}{n}}\right)\gamma_{1,1} \\
&\quad+\frac{1}{np}\left[\frac{r\log^{2}(n+d)}{np}\omega_{\max}^{2}+\left(\sqrt{\frac{r\log^{2}(n+d)}{np}}\omega_{\max}+\zeta_{\mathsf{2nd}}\sigma_{r}^{\star}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\zeta_{\mathsf{2nd}}^{2}\sigma_{1}^{\star2}\right] \\
&\quad\cdot\left[\frac{r\log^{2}(n+d)}{np}\omega_{\max}^{2}+\left(\sqrt{\frac{r\log^{2}(n+d)}{np}}\omega_{\max}+\zeta_{\mathsf{2nd}}\sigma_{r}^{\star}\right)\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\zeta_{\mathsf{2nd}}^{2}\sigma_{1}^{\star2}\right] \\
&\lesssim \delta v_{i,j}^{\star}+\underbrace{\frac{r^{2}\log^{4}(n+d)}{n^{3}p^{3}}\omega_{\max}^{4}}_{=:\gamma_{1,2,1}}+\underbrace{\frac{1}{np}\left(\frac{r\log^{2}(n+d)}{np}\omega_{\max}^{2}+\zeta_{\mathsf{2nd}}^{2}\sigma_{r}^{\star2}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}}_{=:\gamma_{1,2,2}} \\
&\quad+\underbrace{\frac{1}{np}\zeta_{\mathsf{2nd}}^{4}\sigma_{1}^{\star4}}_{=:\gamma_{1,2,3}}+\underbrace{\frac{r^{3/2}\log^{3}(n+d)}{n^{5/2}p^{5/2}}\omega_{\max}^{3}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\right)}_{=:\gamma_{1,2,4}} \\
&\quad+\underbrace{\frac{1}{np}\zeta_{\mathsf{2nd}}^{3}\sigma_{1}^{\star2}\sigma_{r}^{\star}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\right)}_{=:\gamma_{1,2,5}} \\
&\lesssim \delta v_{i,j}^{\star}.
\end{aligned}
$$

Here the penultimate relation follows from the bound on $\gamma_{1,1}$ as well as the assumption that $\theta \lesssim 1$ and $n \gtrsim \kappa^{3} r\log(n+d)$, and the last line holds since

$$
\begin{aligned}
\gamma_{1,2,1} &\lesssim \frac{1}{np}\left[\frac{\kappa r\log^{2}(n+d)}{np}\omega_{\max}^{2}\right]^{2}\overset{\text{(i)}}{\lesssim}\frac{\varepsilon^{2}}{np}v_{i,j}^{\star}\le\delta v_{i,j}^{\star}, \\
\gamma_{1,2,2} &\lesssim \frac{1}{\sqrt{np}}\left(\frac{r\log^{2}(n+d)}{np}\omega_{\max}^{2}+\zeta_{\mathsf{2nd}}^{2}\sigma_{r}^{\star2}\right)\cdot\frac{1}{\sqrt{np}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2} \\
&\overset{\text{(ii)}}{\lesssim}\frac{1}{\sqrt{np}}\varepsilon(v_{i,j}^{\star})^{1/2}\cdot\frac{1}{\sqrt{np}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\lesssim\frac{\varepsilon}{\sqrt{np}}v_{i,j}^{\star}\lesssim\delta v_{i,j}^{\star}, \\
\gamma_{1,2,3} &\lesssim \frac{1}{np}\left(\zeta_{\mathsf{2nd}}^{2}\sigma_{1}^{\star2}\right)^{2}\overset{\text{(ii)}}{\lesssim}\frac{\varepsilon^{2}}{np}v_{i,j}^{\star}\lesssim\delta v_{i,j}^{\star}, \\
\gamma_{1,2,4} &\lesssim \frac{r^{1/2}\log(n+d)}{np}\cdot\frac{\kappa r\log^{2}(n+d)}{np}\omega_{\max}^{2}\cdot\frac{1}{\sqrt{np}}\omega_{\min}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\right) \\
&\overset{\text{(iii)}}{\lesssim}\frac{r^{1/2}\log(n+d)}{np}\cdot\varepsilon(v_{i,j}^{\star})^{1/2}\cdot\frac{1}{\sqrt{np}}\omega_{\min}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\right) \\
&\lesssim \frac{r^{1/2}\log(n+d)}{np}\varepsilon v_{i,j}^{\star}\lesssim\delta v_{i,j}^{\star}, \\
\gamma_{1,2,5} &\lesssim \frac{1}{np}\cdot\zeta_{\mathsf{2nd}}^{2}\sigma_{1}^{\star2}\cdot\zeta_{\mathsf{2nd}}\sigma_{1}^{\star2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\right)\overset{\text{(iv)}}{\lesssim}\frac{\varepsilon^{2}}{np}v_{i,j}^{\star}\lesssim\delta v_{i,j}^{\star},
\end{aligned}
$$

provided that $np \gtrsim \delta^{-2} r^{1/2}\log(n+d)$, $\varepsilon \lesssim 1$. Here (i)-(iv) utilize (E.45).

Taking the bounds on $\gamma_{1,1}$ and $\gamma_{1,2}$ collectively yields

$$
|\alpha_{1}-\beta_{1}|\lesssim|\gamma_{1,1}|+|\gamma_{1,2}|\lesssim\delta v_{i,j}^{\star}
$$

provided that $np \gtrsim \delta^{-2} r\log^{2}(n+d)$, $n \gtrsim \delta^{-2}\kappa^{2} r\log(n+d)$ and $\varepsilon \lesssim 1$.

Step 2: bounding $|\alpha_2 - \beta_2|$. From Lemma 11 we know that

$$\left|S_{i,j} - S_{i,j}^{\star}\right| \lesssim \left(\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sqrt{\frac{r+\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{r\log^2(n+d)}{np}\omega_{\max}^2$$
$$+\left(\sqrt{\frac{r\log^2(n+d)}{np}}\omega_{\max} + \zeta_{\mathsf{2nd}}\sigma_r^{\star}\right)\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right) + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}.$$

Then we have

$$|\alpha_2 - \beta_2| \lesssim \frac{1}{np}\left|S_{i,j}^{\star 2} - S_{i,j}^2\right| \lesssim \underbrace{\frac{1}{np}S_{i,j}^{\star}\left|S_{i,j} - S_{i,j}^{\star}\right|}_{=:\gamma_{2,1}} + \underbrace{\frac{1}{np}\left|S_{i,j} - S_{i,j}^{\star}\right|^2}_{=:\gamma_{2,2}}.$$

- Regarding $\gamma_{2,1}$, we have

$$\begin{aligned}
\gamma_{2,1} &\lesssim \underbrace{\frac{1}{np}\left(\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sqrt{\frac{r+\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2}_{=:\gamma_{2,1,1}} \\
&\quad + \underbrace{\frac{r\log^2(n+d)}{n^2p^2}\omega_{\max}^2\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2}_{=:\gamma_{2,1,2}} \\
&\quad + \underbrace{\frac{1}{np}\sqrt{\frac{r\log^2(n+d)}{np}}\omega_{\max}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right)}_{=:\gamma_{2,1,3}} \\
&\quad + \underbrace{\frac{1}{np}\zeta_{\mathsf{2nd}}\sigma_1^{\star}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right)}_{=:\gamma_{2,1,4}} \\
&\quad + \underbrace{\frac{1}{np}\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2}_{=:\gamma_{2,1,5}} \lesssim \delta v_{i,j}^{\star}.
\end{aligned}$$

  Here the last relation holds since

$$\begin{aligned}
\gamma_{2,1,1} &\lesssim \frac{\delta}{np}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 \lesssim \delta v_{i,j}^{\star}, \\
\gamma_{2,1,2} &\lesssim \frac{1}{\sqrt{np}}\cdot\frac{\kappa r\log^2(n+d)}{np}\omega_{\max}^2\cdot\frac{1}{\sqrt{np}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \\
&\overset{\text{(i)}}{\lesssim} \frac{\varepsilon}{\sqrt{np}}(v_{i,j}^{\star})^{1/2}\cdot\frac{1}{\sqrt{np}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \lesssim \delta v_{i,j}^{\star}, \\
\gamma_{2,1,3} &\overset{\text{(ii)}}{\lesssim} \frac{\delta}{np}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{\delta}{np}\omega_{\min}^2\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\right) \lesssim \delta v_{i,j}^{\star}, \\
\gamma_{2,1,4} &\lesssim \frac{1}{\sqrt{np}}\cdot\frac{1}{\sqrt{np}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\cdot\zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right) \\
&\overset{\text{(iii)}}{\lesssim} \frac{\varepsilon}{\sqrt{np}}(v_{i,j}^{\star})^{1/2}\cdot\frac{1}{\sqrt{np}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \lesssim \delta v_{i,j}^{\star}, \\
\gamma_{2,1,5} &\lesssim \frac{1}{\sqrt{np}}\cdot\frac{1}{\sqrt{np}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\cdot\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}
\end{aligned}$$

$$\overset{\text{(iv)}}{\lesssim} \frac{\varepsilon}{\sqrt{np}}(v_{i,j}^\star)^{1/2} \cdot \frac{1}{\sqrt{np}} \left\|\boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2 \left\|\boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2 \lesssim \frac{\varepsilon}{\sqrt{np}} v_{i,j}^\star \lesssim \delta v_{i,j}^\star,$$

provided that $np \gtrsim \delta^{-2} r \log^2(n+d)$, $n \gtrsim \delta^{-2}\kappa^2 r \log(n+d)$ and $\varepsilon \lesssim 1$. Here (i), (iii) and (iv) follow from (E.45); (ii) follows from the AM-GM inequality.

- Regarding $\gamma_{2,2}$, we have

$$\begin{aligned}\gamma_{2,2} &\lesssim \underbrace{\frac{1}{np}\left(\frac{r\log^2(n+d)}{np} + \frac{\kappa^2 r\log(n+d)}{n}\right)\left\|\boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2 \left\|\boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2}_{=:\gamma_{2,2,1}} + \underbrace{\frac{r^2\log^4(n+d)}{n^3p^3}\omega_{\max}^4}_{=:\gamma_{2,2,2}} + \underbrace{\frac{1}{np}\zeta_{\mathsf{2nd}}^4\sigma_1^{\star 4}}_{=:\gamma_{2,2,3}} \\ &\quad + \underbrace{\frac{r\log^2(n+d)}{n^2p^2}\omega_{\max}^2\left(\left\|\boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2 + \left\|\boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2\right)}_{=:\gamma_{2,2,4}} + \underbrace{\frac{1}{np}\zeta_{\mathsf{2nd}}^2\sigma_r^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2 + \left\|\boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2\right)}_{=:\gamma_{2,2,5}} \\ &\lesssim \delta v_{i,j}^\star.\end{aligned}$$

Here the last line holds since

$$\begin{aligned}\gamma_{2,2,1} &\lesssim \frac{\delta}{np}\left\|\boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2 \left\|\boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2 \lesssim \delta v_{i,j}^\star, \\ \gamma_{2,2,2} &\overset{\text{(i)}}{\lesssim} \frac{1}{np}\left(\frac{\kappa r\log^2(n+d)}{np}\omega_{\max}^2\right)^2 \lesssim \frac{\varepsilon^2}{np} v_{i,j}^\star \lesssim \delta v_{i,j}^\star, \\ \gamma_{2,2,3} &\overset{\text{(ii)}}{\lesssim} \frac{1}{np}\left(\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}\right)^2 \lesssim \frac{\varepsilon^2}{np} v_{i,j}^\star \lesssim \delta v_{i,j}^\star, \\ \gamma_{2,2,4} &\lesssim \frac{\delta}{np}\omega_{\min}^2\left(\left\|\boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2 + \left\|\boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2\right) \lesssim \delta v_{i,j}^\star, \\ \gamma_{2,2,5} &\lesssim \frac{1}{np}\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 4}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2^2 + \left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2^2\right) \overset{\text{(iii)}}{\lesssim} \frac{\varepsilon^2}{np} v_{i,j}^\star \lesssim \delta v_{i,j}^\star.\end{aligned}$$

provided that $np \gtrsim \delta^{-1} r\log^2(n+d)$, $n \gtrsim \delta^{-1}\kappa^2 r\log(n+d)$ and $\varepsilon \lesssim 1$. Here (i)-(iii) follows from (E.45).

Taking the bounds on $\gamma_{2,1}$ and $\gamma_{2,2}$ collectively yields

$$|\alpha_2 - \beta_2| \lesssim |\gamma_{2,1}| + |\gamma_{2,2}| \lesssim \delta v_{i,j}^\star$$

provided that $np \gtrsim \delta^{-2} r\log^2(n+d)$, $n \gtrsim \delta^{-2}\kappa^2 r\log(n+d)$ and $\varepsilon \lesssim 1$.

Step 3: bounding $|\alpha_3 - \beta_3|$. For each $l \in [d]$, we first define UB_l to be

$$\begin{aligned}\mathsf{UB}_l \coloneqq &\left(\sqrt{\frac{r\log^2(n+d)}{np}} + \kappa\sqrt{\frac{r+\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2 + \frac{r\log^2(n+d)}{np}\omega_{\max}^2 \\ &+ \left(\sqrt{\frac{r\log^2(n+d)}{np}}\omega_{\max} + \zeta_{\mathsf{2nd}}\sigma_r^\star\right)\left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2 + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}.\end{aligned}$$

In view of Lemma 41, we know that

$$\left|\omega_i^2 - \omega_i^{\star 2}\right| \lesssim \sqrt{\frac{\log^2(n+d)}{np}}\omega_i^{\star 2} + \mathsf{UB}_i.$$

We also learn from (E.46) that UB_i (resp. UB_j) is the upper bound of $|S_{i,i} - S_{i,i}^\star|$ (resp. $|S_{j,j} - S_{j,j}^\star|$) we used in Step 1. We have

$$\frac{1}{np}\left|\omega_i^2 S_{j,j} - \omega_i^{\star 2} S_{j,j}^\star\right| \lesssim \frac{1}{np} S_{j,j}\left|\omega_i^2 - \omega_i^{\star 2}\right| + \frac{1}{np}\omega_i^{\star 2}\left|S_{j,j} - S_{j,j}^\star\right|$$

$$\lesssim \underbrace{\frac{1}{np}S_{j,j}^{\star}\left|\omega_{i}^{2}-\omega_{i}^{\star 2}\right|}_{=:\gamma_{3,1}}+\underbrace{\frac{1}{np}\omega_{i}^{\star 2}\left|S_{j,j}-S_{j,j}^{\star}\right|}_{=:\gamma_{3,2}}+\underbrace{\frac{1}{np}\left|S_{j,j}-S_{j,j}^{\star}\right|\left|\omega_{i}^{2}-\omega_{i}^{\star 2}\right|}_{=:\gamma_{3,3}}.$$

- Regarding $\gamma_{3,1}$, we have

$$\gamma_{3,1}\lesssim\underbrace{\frac{1}{np}\left\Vert \boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}^{2}\sqrt{\frac{\log^{2}\left(n+d\right)}{np}}\omega_{i}^{\star 2}}_{=:\gamma_{3,1,1}}+\underbrace{\frac{1}{np}S_{j,j}^{\star}\mathsf{UB}_{i}}_{=:\gamma_{3,1,2}}\lesssim\delta v_{i,j}^{\star}.$$

Here the last relation holds since (i) the first term

$$\gamma_{3,1,1}\lesssim\sqrt{\frac{\kappa\log^{2}\left(n+d\right)}{n^{3}p^{3}}}\omega_{i}^{\star 2}\sigma_{r}^{\star 2}\left\Vert \boldsymbol{U}_{j,\cdot}^{\star}\right\Vert _{2}^{2}\lesssim\sqrt{\frac{\kappa\log^{2}\left(n+d\right)}{np}}v_{i,j}^{\star}\lesssim\delta v_{i,j}^{\star}$$

provided that $np\gtrsim\delta^{-2}\kappa\log^{2}(n+d)$; and (ii) $\gamma_{3,1,2}\lesssim\delta v_{i,j}^{\star}$ because it is straightforward to check that the second term $\gamma_{3,1,2}$ admits the same upper bound as $\gamma_{1,1}$.

- Regarding $\gamma_{3,2}$, we have

$$\begin{aligned}\gamma_{3,2} & \lesssim\underbrace{\frac{1}{np}\omega_{i}^{\star 2}\left(\sqrt{\frac{r\log^{2}\left(n+d\right)}{np}}+\kappa\sqrt{\frac{r+\log\left(n+d\right)}{n}}\right)\left\Vert \boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}^{2}}_{=:\gamma_{3,2,1}}+\underbrace{\frac{r\log^{2}\left(n+d\right)}{n^{2}p^{2}}\omega_{\max}^{4}}_{=:\gamma_{3,2,2}}\\
 & \quad+\underbrace{\frac{1}{np}\sqrt{\frac{r\log^{2}\left(n+d\right)}{np}}\omega_{\max}^{3}\left\Vert \boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}}_{=:\gamma_{3,2,3}}+\underbrace{\frac{1}{np}\omega_{i}^{\star 2}\zeta_{\mathsf{2nd}}\sigma_{r}^{\star}\left\Vert \boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}}_{=:\gamma_{3,2,4}}+\underbrace{\frac{1}{np}\omega_{i}^{\star 2}\zeta_{\mathsf{2nd}}^{2}\sigma_{1}^{\star 2}}_{=:\gamma_{3,2,5}}\\
 & \lesssim\delta v_{i,j}^{\star}.\end{aligned}$$

Here the last line holds since

$$\begin{aligned}\gamma_{3,2,1}+\gamma_{3,2,2}+\gamma_{3,2,3} & \lesssim\frac{\delta}{np}\omega_{\min}^{2}\left(\left\Vert \boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}^{2}+\left\Vert \boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}^{2}\right)\lesssim\delta v_{i,j}^{\star},\\
\gamma_{3,2,4} & \overset{\text{(i)}}{\lesssim}\frac{1}{np}\omega_{i}^{\star 2}\frac{\varepsilon}{\sqrt{np\kappa}\sigma_{1}^{\star}}\left(\min\left\{ \left\Vert \boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2},\left\Vert \boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}\right\} +\omega_{\min}\right)\sigma_{r}^{\star}\left\Vert \boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}\\
 & \lesssim\frac{1}{np}\omega_{i}^{\star 2}\frac{\varepsilon}{\sqrt{np\kappa}}\left\Vert \boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}^{2}+\frac{1}{np}\frac{\varepsilon}{\sqrt{np\kappa}}\omega_{\min}^{3}\left\Vert \boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}\\
 & \lesssim\frac{\delta}{np}\omega_{\min}^{2}\left(\left\Vert \boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}^{2}+\left\Vert \boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}^{2}\right)\lesssim\delta v_{i,j}^{\star},\\
\gamma_{3,2,5} & \overset{\text{(ii)}}{\lesssim}\frac{1}{np}\omega_{i}^{\star 2}\frac{\varepsilon^{2}}{np\kappa}\left(\left\Vert \boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}^{2}+\omega_{\min}^{2}\right)\lesssim\frac{\delta}{np}\omega_{\min}^{2}\left(\left\Vert \boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}^{2}+\left\Vert \boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}^{2}\right)\lesssim\delta v_{i,j}^{\star},\end{aligned}$$

provided that $np\gtrsim\delta^{-2}r\log^{2}(n+d)$, $n\gtrsim\delta^{-2}\kappa^{2}r\log(n+d)$, $\varepsilon\lesssim1$ and

$$\left\Vert \boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}+\left\Vert \boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}\gtrsim\delta^{-1}\omega_{\max}\sqrt{\frac{r\log^{2}\left(n+d\right)}{np}}.$$

Here (i) and (ii) follows from (E.44).

- Regarding $\gamma_{3,3}$, we have

$$\gamma_{3,3}\lesssim\underbrace{\frac{1}{np}\left|S_{j,j}-S_{j,j}^{\star}\right|\sqrt{\frac{\log^{2}\left(n+d\right)}{np}}\omega_{i}^{\star 2}}_{=:\gamma_{3,3,1}}+\underbrace{\frac{1}{np}\left|S_{j,j}-S_{j,j}^{\star}\right|\mathsf{UB}_{i}}_{=:\gamma_{3,3,2}}\lesssim\delta v_{i,j}^{\star}.$$

The last relation holds since (i) the first term

$$\gamma_{3,3,1} \lesssim \sqrt{\frac{\log^2(n+d)}{np}}\gamma_{3,2} \lesssim \delta v_{i,j}^{\star}$$

provided that $np \gtrsim \log^2(n+d)$; and (ii) $\gamma_{3,3,2} \lesssim \delta v_{i,j}^{\star}$ because it is easy to check that the second term $\gamma_{3,3,2}$ admits the same upper bound as $\gamma_{1,2}$.

Therefore we conclude that

$$\frac{1}{np}\left|\omega_i^2 S_{j,j} - \omega_i^{\star 2} S_{j,j}^{\star}\right| \lesssim \gamma_{3,1} + \gamma_{3,2} + \gamma_{3,3} \lesssim \delta v_{i,j}^{\star},$$

and similarly we can show that

$$\frac{1}{np}\left|\omega_j^2 S_{i,i} - \omega_j^{\star 2} S_{i,i}^{\star}\right| \lesssim \delta v_{i,j}^{\star}.$$

This allows us to achieve

$$|\alpha_3 - \beta_3| \lesssim \frac{1}{np}\left|\omega_i^2 S_{j,j} - \omega_i^{\star 2} S_{j,j}^{\star}\right| + \frac{1}{np}\left|\omega_j^2 S_{i,i} - \omega_j^{\star 2} S_{i,i}^{\star}\right| \lesssim \delta v_{i,j}^{\star}$$

provided that $np \gtrsim \delta^{-2}\kappa r \log^2(n+d)$, $n \gtrsim \delta^{-2}\kappa^2 r \log(n+d)$, $\varepsilon \lesssim 1$ and

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \gtrsim \delta^{-1}\omega_{\max}\sqrt{\frac{r\log^2(n+d)}{np}}.$$

Step 4: putting everything together. Taking the bounds on $|\alpha_1 - \beta_1|$, $|\alpha_2 - \beta_2|$, $|\alpha_3 - \beta_3|$ collectively yields

$$\left|v_{i,j} - v_{i,j}^{\star}\right| \le |\alpha_1 - \beta_1| + |\alpha_2 - \beta_2| + |\alpha_3 - \beta_3| \lesssim \delta v_{i,j}^{\star}$$

provided that $n \gtrsim \delta^{-2}\kappa^2 r \log(n+d)$, $np \gtrsim \delta^{-2}\kappa r \log^2(n+d)$, $\varepsilon \lesssim 1$ and

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \gtrsim \delta^{-1}\omega_{\max}\sqrt{\frac{r\log^2(n+d)}{np}}.$$

Therefore we can take $\varepsilon = 1$, therefore (E.44) and (E.45) are all guaranteed by the conditions of Lemma 16.

E.4.2 Proof of Lemma 20

We know that

$$\begin{aligned} S_{i,j}^{\star} \in \mathsf{CI}_{i,j}^{1-\alpha} \quad &\Longleftrightarrow \quad S_{i,j}^{\star} \in \left[S_{i,j} \pm \Phi^{-1}(1-\alpha/2)\sqrt{v_{i,j}}\right] \\ &\Longleftrightarrow \quad \frac{S_{i,j} - S_{i,j}^{\star}}{\sqrt{v_{i,j}}} \in \left[-\Phi^{-1}(1-\alpha/2), \Phi^{-1}(1-\alpha/2)\right]. \end{aligned} \tag{E.47}$$

From Lemma 18 we know that for any $t \in \mathbb{R}$,

$$\mathbb{P}\left(\frac{S_{i,j} - S_{i,j}^{\star}}{\sqrt{v_{i,j}^{\star}}} \le t\right) = \Phi(t) + O\left(\frac{1}{\sqrt{\log(n+d)}}\right). \tag{E.48}$$

From Lemma 19 we know that with probability exceeding $1 - O((n+d)^{-10})$,

$$\left|v_{i,j}^{\star} - v_{i,j}\right| \lesssim \delta v_{i,j}^{\star},$$

where δ is the unspecified constant in Lemma 19. An immediate result is that $v_{i,j} \asymp v_{i,j}^{\star}$ as long as $\delta \ll 1$. Therefore

$$\Delta := \left|\frac{S_{i,j} - S_{i,j}^{\star}}{\sqrt{v_{i,j}^{\star}}} - \frac{S_{i,j} - S_{i,j}^{\star}}{\sqrt{v_{i,j}}}\right| = \left|S_{i,j} - S_{i,j}^{\star}\right|\left|\frac{v_{i,j}^{\star} - v_{i,j}}{\sqrt{v_{i,j}^{\star} v_{i,j}}\left(\sqrt{v_{i,j}^{\star}} + \sqrt{v_{i,j}}\right)}\right| \lesssim \delta\left|S_{i,j} - S_{i,j}^{\star}\right| / \sqrt{v_{i,j}^{\star}}.$$

To bound Δ, we need the following claim.

Claim 1. Instate the conditions in Lemma 14. Suppose that $np \gtrsim \log^4(n+d)$. Then with probability exceeding $1 - O((n+d)^{-10})$, we have

$$\left|S_{i,j} - S_{i,j}^\star\right| \lesssim \zeta_{i,j} + \sqrt{\frac{\kappa \log(n+d)}{np}} \left\|\boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2 \left\|\boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2 + \frac{\omega_{\max}}{\sqrt{np}} \left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right) \sigma_1^\star.$$

From the proof of Lemma 16 (more specifically, Step 3 in Appendix E.3.3) that

$$(v_{i,j}^\star)^{-1/2} \zeta_{i,j} \le \widetilde{v}_{i,j}^{-1/2} \zeta_{i,j} \lesssim \frac{1}{\sqrt{\log(n+d)}}.$$

holds under the conditions of Lemma 16. In addition, we also learn from Lemma 15 that

$$\sqrt{v_{i,j}^\star} \gtrsim \frac{1}{\sqrt{np}} \left\|\boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2 \left\|\boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2 + \frac{1}{\sqrt{np}} \omega_{\min} \left(\left\|\boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2\right).$$

Therefore we know that with probability exceeding $1 - O((n+d)^{-10})$,

$$\Delta \lesssim \delta \sqrt{\kappa \log(n+d)} + \frac{1}{\sqrt{\log(n+d)}} \lesssim \frac{1}{\sqrt{\log(n+d)}}$$

if we choose

$$\delta \asymp \frac{1}{\kappa^{1/2} \log(n+d)}.$$

Then we know that for any $t \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{P}\left(\frac{S_{i,j} - S_{i,j}^\star}{\sqrt{v_{i,j}}} \le t\right) &\le \mathbb{P}\left(\frac{S_{i,j} - S_{i,j}^\star}{\sqrt{v_{i,j}}} \le t, \Delta \lesssim \frac{1}{\sqrt{\log(n+d)}}\right) + O\left((n+d)^{-10}\right) \\
&\le \mathbb{P}\left(\frac{S_{i,j} - S_{i,j}^\star}{\sqrt{v_{i,j}^\star}} \le t + \frac{1}{\sqrt{\log(n+d)}}\right) + O\left((n+d)^{-10}\right) \\
&\overset{\text{(i)}}{\le} \Phi\left(t + \frac{1}{\sqrt{\log(n+d)}}\right) + O\left(\frac{1}{\sqrt{\log(n+d)}}\right) \\
&\overset{\text{(ii)}}{\le} \Phi(t) + O\left(\frac{1}{\sqrt{\log(n+d)}}\right).
\end{aligned}$$

Here (i) follows from (E.48); (ii) holds since $\Phi(\cdot)$ is a $1/\sqrt{2\pi}$-Lipschitz continuous function. Similarly we can show that

$$\mathbb{P}\left(\frac{S_{i,j} - S_{i,j}^\star}{\sqrt{v_{i,j}}} \le t\right) \ge \Phi(t) + O\left(\frac{1}{\sqrt{\log(n+d)}}\right).$$

Therefore we have

$$\mathbb{P}\left(\frac{S_{i,j} - S_{i,j}^\star}{\sqrt{v_{i,j}}} \le t\right) = \Phi(t) + O\left(\frac{1}{\sqrt{\log(n+d)}}\right). \tag{E.49}$$

By taking $t = \Phi^{-1}(1-\alpha/2)$ in (E.49), we have

$$\mathbb{P}\left(\frac{S_{i,j} - S_{i,j}^\star}{\sqrt{v_{i,j}}} \le \Phi^{-1}(1-\alpha/2)\right) = 1 - \frac{\alpha}{2} + O\left(\frac{1}{\sqrt{\log(n+d)}}\right). \tag{E.50}$$

In addition, for all $\varepsilon \in (0, 1/\sqrt{\log(n+d)}]$, by taking $t = -\Phi^{-1}(1-\alpha/2) - \varepsilon$ in (E.49), we have

$$\mathbb{P}\left(\frac{S_{i,j} - S_{i,j}^\star}{\sqrt{v_{i,j}}} \le -\Phi^{-1}(1-\alpha/2) - \varepsilon\right) = \Phi\left(-\Phi^{-1}(1-\alpha/2) - \varepsilon\right) + O\left(\frac{1}{\sqrt{\log(n+d)}}\right)$$

$$= \frac{\alpha}{2} + O\left(\frac{1}{\sqrt{\log(n+d)}}\right),$$

where the constant hide in $O(\cdot)$ is independent of ε. Here the last relation holds since $\Phi(\cdot)$ is a $1/\sqrt{2\pi}$-Lipschitz continuous function and $\varepsilon \lesssim 1/\sqrt{\log(n+d)}$. As a result,

$$\mathbb{P}\left(\frac{S_{i,j} - S_{i,j}^{\star}}{\sqrt{v_{i,j}}} < -\Phi^{-1}(1-\alpha/2)\right) = \lim_{\varepsilon \searrow 0} \mathbb{P}\left(\frac{S_{i,j} - S_{i,j}^{\star}}{\sqrt{v_{i,j}}} \leq -\Phi^{-1}(1-\alpha/2) - \varepsilon\right)$$
$$= \frac{\alpha}{2} + O\left(\frac{1}{\sqrt{\log(n+d)}}\right). \tag{E.51}$$

Taking (E.47), (E.49) and (E.51) collectively yields

$$\begin{aligned}\mathbb{P}\left(S_{i,j}^{\star} \in \mathsf{CI}_{i,j}^{1-\alpha}\right) &= \mathbb{P}\left(\frac{S_{i,j} - S_{i,j}^{\star}}{\sqrt{v_{i,j}}} \in \left[-\Phi^{-1}(1-\alpha/2), \Phi^{-1}(1-\alpha/2)\right]\right) \\ &= \mathbb{P}\left(\frac{S_{i,j} - S_{i,j}^{\star}}{\sqrt{v_{i,j}}} \leq \Phi^{-1}(1-\alpha/2)\right) - \mathbb{P}\left(\frac{S_{i,j} - S_{i,j}^{\star}}{\sqrt{v_{i,j}}} < -\Phi^{-1}(1-\alpha/2)\right) \\ &= 1 - \frac{\alpha}{2} + O\left(\frac{1}{\sqrt{\log(n+d)}}\right).\end{aligned}$$

It remains to present the conditions of Lemma 19 when we take $\delta \asymp 1/\sqrt{\kappa \log^2(n+d)}$, which reads $n \gtrsim \kappa^3 r \log^3(n+d)$, $np \gtrsim \kappa^2 r \log^4(n+d)$ and

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \gtrsim \omega_{\max}\sqrt{\frac{\kappa r \log^4(n+d)}{np}}.$$

Proof of Claim 1. Denote

$$X_{i,j} = \sum_{l=1}^{n}\big[\underbrace{M_{j,l}^{\natural}E_{i,l}}_{=:a_l} + \underbrace{M_{i,l}^{\natural}E_{j,l}}_{=:b_l}\big].$$

Then we bound $\sum_{l=1}^{n} a_l$ and $\sum_{l=1}^{n} b_l$ respectively.

We first bound $\sum_{l=1}^{n} a_l$. It is straightforward to calculate

$$L_a \coloneqq \max_{1\leq l\leq n}|a_l| \leq B_i\left\|\boldsymbol{M}_{j,\cdot}^{\natural}\right\|_{\infty} \leq B_i\left\|\boldsymbol{U}_{j,\cdot}^{\natural}\right\|_2\sigma_1^{\natural}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty},$$
$$V_a \coloneqq \sum_{l=1}^{n}\mathsf{var}\left(M_{j,l}^{\natural}E_{i,l}|\boldsymbol{F}\right) = \sum_{l=1}^{n}M_{j,l}^{\natural 2}\sigma_{i,l}^2 = \sum_{l=1}^{n}M_{j,l}^{\natural 2}\sigma_{i,l}^2 = \sum_{l=1}^{n}M_{j,l}^{\natural 2}\left[\frac{1-p}{np}\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2 + \frac{\omega_i^{\star 2}}{np}\right],$$

whereB_i and B_j are two (random) quantities such that

$$\max_{l\in[n]}|E_{i,l}| \leq B_i \qquad \text{and} \qquad \max_{l\in[n]}|E_{j,l}| \leq B_j.$$

On the event $\mathcal{E}_{\mathsf{good}}$, we know that

$$\begin{aligned}L_a &\overset{\text{(i)}}{\lesssim} \frac{1}{p}\sqrt{\frac{\log(n+d)}{n}}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_i^{\star}\right)\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\sigma_1^{\star}\sqrt{\frac{\log(n+d)}{n}}, \\ &\lesssim \frac{\log(n+d)}{np}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_i^{\star}\right)\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\sigma_1^{\star},\end{aligned}$$

and

$$V_a \overset{\text{(ii)}}{\lesssim} \sum_{l=1}^{n}M_{j,l}^{\natural 2}\left[\frac{1-p}{np}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\log(n+d) + \frac{\omega_i^{\star 2}}{np}\right]$$

$$
\begin{aligned}
&= \left\|\boldsymbol{M}_{j,\cdot}^{\natural}\right\|_2^2 \left[\frac{1-p}{np}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 \log(n+d) + \frac{\omega_i^{\star 2}}{np}\right] \\
&\lesssim \left\|\boldsymbol{U}_{j,\cdot}^{\natural}\right\|_2^2 \sigma_1^{\natural 2}\left[\frac{1-p}{np}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 \log(n+d) + \frac{\omega_i^{\star 2}}{np}\right] \\
&\overset{\text{(iii)}}{\lesssim} \left[\frac{1-p}{np}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 \log(n+d) + \frac{\omega_i^{\star 2}}{np}\right]\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2 \sigma_1^{\star 2}.
\end{aligned}
$$

Here (i) follows from (D.10), (D.13) and (D.15); (ii) follows from (E.57); (iii) follows from (D.10). Therefore in view of the Bernstein inequality (Vershynin, 2017, Theorem 2.8.4), conditional on $\boldsymbol{F}$, with probability exceeding $1-O((n+d)^{-10})$,

$$
\begin{aligned}
\sum_{l=1}^{n} a_l &\lesssim \sqrt{V_a \log(n+d)} + L_a \log(n+d) \\
&\lesssim \left[\sqrt{\frac{\log(n+d)}{np}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{\omega_i^{\star}}{\sqrt{np}}\right]\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \sigma_1^{\star} \frac{\log^2(n+d)}{np}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_j^{\star}\right)\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \sigma_1^{\star} \\
&\lesssim \left[\sqrt{\frac{\log(n+d)}{np}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{\omega_i^{\star}}{\sqrt{np}}\right]\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \sigma_1^{\star},
\end{aligned}
$$

provided that $np \gtrsim \log^4(n+d)$. Similarly we can show that with probability exceeding $1-O((n+d)^{-10})$,

$$
\sum_{l=1}^{n} b_l \lesssim \left[\sqrt{\frac{\log(n+d)}{np}}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{\omega_j^{\star}}{\sqrt{np}}\right]\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 \sigma_1^{\star}.
$$

Therefore

$$
\sum_{l=1}^{n} (a_l + b_l) \lesssim \sqrt{\frac{\kappa \log(n+d)}{np}}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{\omega_{\max}}{\sqrt{np}}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)\sigma_1^{\star}.
$$

Combining the above result with Lemma 14 yields

$$
\left|S_{i,j} - S_{i,j}^{\star}\right| \lesssim \zeta_{i,j} + \sqrt{\frac{\kappa \log(n+d)}{np}}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{\omega_{\max}}{\sqrt{np}}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)\sigma_1^{\star}.
$$

□

E.5 Other basic facts

In this section we will establish a couple of useful properties that occur with high probability.

1. In view of standard results on Gaussian random matrices, we have

$$
\left\|\frac{1}{n}\boldsymbol{F}\boldsymbol{F}^{\top} - \boldsymbol{I}_r\right\| \lesssim \sqrt{\frac{r}{n}} + \sqrt{\frac{\log(n+d)}{n}} + \frac{r}{n} + \frac{\log(n+d)}{n} \lesssim \sqrt{\frac{r+\log(n+d)}{n}} \tag{E.52}
$$

   with probability exceeding $1-O((n+d)^{-100})$, provided that $n \gtrsim r + \log(n+d)$. As an immediate consequence, it is seen from the definition (D.2) that

$$
\begin{aligned}
\left\|\boldsymbol{M}^{\natural}\boldsymbol{M}^{\natural\top} - \boldsymbol{S}^{\star}\right\| &= \left\|\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\left(\frac{1}{n}\boldsymbol{F}\boldsymbol{F}^{\top} - \boldsymbol{I}_r\right)\boldsymbol{\Sigma}^{\star}\boldsymbol{U}^{\star\top}\right\| \le \left\|\boldsymbol{U}^{\star}\right\|^2 \left\|\boldsymbol{\Sigma}^{\star}\right\|^2 \left\|\frac{1}{n}\boldsymbol{F}\boldsymbol{F}^{\top} - \boldsymbol{I}_r\right\| \\
&\lesssim \sqrt{\frac{r+\log(n+d)}{n}}\sigma_1^{\star 2}.
\end{aligned} \tag{E.53}
$$

Weyl's inequality then tells us that

$$\left\|\boldsymbol{\Sigma}^{\natural 2}-\boldsymbol{\Sigma}^{\star 2}\right\| \leq \left\|\boldsymbol{M}^{\natural}\boldsymbol{M}^{\natural\top}-\boldsymbol{S}^{\star}\right\| \lesssim \sqrt{\frac{r+\log\left(n+d\right)}{n}}\sigma_{1}^{\star 2}, \tag{E.54}$$

thus indicating that

$$\sigma_{r}^{\natural} \asymp \sigma_{r}^{\star} \qquad \text{and} \qquad \sigma_{1}^{\natural} \asymp \sigma_{1}^{\star} \tag{E.55}$$

as long as $n \gg \frac{\sigma_1^{\star 4}}{\sigma_r^{\star 4}}\big(r+\log(n+d)\big) = \kappa^2\big(r+\log(n+d)\big)$ (see the definition of κ in (3.1)). Consequently,

$$\left\|\boldsymbol{\Sigma}^{\natural}-\boldsymbol{\Sigma}^{\star}\right\| \leq \frac{\left\|\boldsymbol{\Sigma}^{\natural 2}-\boldsymbol{\Sigma}^{\star 2}\right\|}{\sigma_{r}^{\star}} \lesssim \sqrt{\frac{r+\log\left(n+d\right)}{n}}\frac{\sigma_{1}^{\star 2}}{\sigma_{r}^{\star}} \asymp \kappa\sqrt{\frac{\left(r+\log\left(n+d\right)\right)}{n}}\sigma_{r}^{\star}. \tag{E.56}$$

2. Given the calculation (D.8), let us define

$$\sigma^{2} \coloneqq \max_{i\in[d],j\in[n]}\sigma_{i,j}^{2} \asymp \frac{1-p}{np}\max_{i\in[d],j\in[n]}\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_{j}\right)^{2}+\frac{\omega_{i}^{\star 2}}{np}.$$

Note that for each $i\in[d]$ and $j\in[n]$, one has $\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_{j}\sim\mathcal{N}(0,\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\|_2^2)$, thus revealing that

$$\max_{j\in[n]}\left|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_{j}\right| \lesssim \left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\sqrt{\log\left(n+d\right)} \tag{E.57}$$

with probability exceeding $1-O((n+d)^{-100})$. As a result, taking the union bound gives

$$\sigma^{2} \lesssim \frac{\|\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\|_{2,\infty}^{2}\log\left(n+d\right)+\omega_{\max}^{2}}{np} \lesssim \frac{\mu r\log\left(n+d\right)}{ndp}\sigma_{1}^{\star 2}+\frac{\omega_{\max}^{2}}{np} \eqqcolon \sigma_{\mathsf{ub}}^{2},$$

where

$$\sigma_{\mathsf{ub}} \asymp \sqrt{\frac{\mu r\log\left(n+d\right)}{ndp}}\sigma_{1}^{\star}+\frac{\omega_{\max}}{\sqrt{np}}.$$

3. In addition, it follows from the expression (D.7) and the property (E.57) that

$$\begin{aligned}
\max_{i\in[d],j\in[n]}\left|E_{i,j}\right| &\leq \frac{1}{\sqrt{np}}\max_{i\in[d],j\in[n]}\left|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_{j}\right|+\frac{1}{\sqrt{np}}\max_{i\in[d],j\in[n]}\left|N_{i,j}\right| \\
&\lesssim \frac{1}{\sqrt{np}}\left\|\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2,\infty}\sqrt{\log\left(n+d\right)}+\frac{\omega_{\max}\sqrt{\log\left(n+d\right)}}{\sqrt{np}} \\
&\lesssim \frac{1}{p}\sqrt{\frac{\mu r\log\left(n+d\right)}{nd}}\sigma_{1}^{\star}+\frac{\omega_{\max}}{p}\sqrt{\frac{\log\left(n+d\right)}{n}}
\end{aligned}$$

occurs with probability exceeding $1-O((n+d)^{-100})$. Therefore, we shall take

$$B \coloneqq \frac{1}{p}\sqrt{\frac{\mu r\log\left(n+d\right)}{nd}}\sigma_{1}^{\star}+\frac{\omega_{\max}}{p}\sqrt{\frac{\log\left(n+d\right)}{n}} \asymp \sigma_{\mathsf{ub}}\sqrt{\frac{\log\left(n+d\right)}{p}}.$$

4. Recall that the top-r eigen-decomposition of $\boldsymbol{M}^{\natural}\boldsymbol{M}^{\natural\top}$ and $\boldsymbol{S}^{\star}$ are denoted by $\boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{U}^{\natural\top}$ and $\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{U}^{\star\top}$, respectively, and that $\boldsymbol{Q}$ is a rotation matrix such that $\boldsymbol{U}^{\natural}=\boldsymbol{U}^{\star}\boldsymbol{Q}$. Therefore, the matrix $\boldsymbol{J}$ defined in (D.5) obeys

$$\begin{aligned}
\left\|\boldsymbol{Q}-\boldsymbol{J}\right\| &= \left\|\boldsymbol{Q}-\boldsymbol{\Sigma}^{\star}\boldsymbol{Q}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-1}\right\| \leq \frac{1}{\sigma_{r}^{\natural}}\left\|\boldsymbol{Q}\boldsymbol{\Sigma}^{\natural}-\boldsymbol{\Sigma}^{\star}\boldsymbol{Q}\right\| = \frac{1}{\sigma_{r}^{\natural}}\left\|\boldsymbol{Q}\boldsymbol{\Sigma}^{\natural}\boldsymbol{Q}^{\top}-\boldsymbol{\Sigma}^{\star}\right\| \\
&= \frac{1}{\sigma_{r}^{\natural}}\left\|\boldsymbol{U}^{\star}\left(\boldsymbol{Q}\boldsymbol{\Sigma}^{\natural}\boldsymbol{Q}^{\top}-\boldsymbol{\Sigma}^{\star}\right)\boldsymbol{U}^{\star\top}\right\| = \frac{1}{\sigma_{r}^{\natural}}\left\|\boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural}\boldsymbol{U}^{\natural\top}-\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{U}^{\star\top}\right\|.
\end{aligned}$$

Invoke the perturbation bound for matrix square roots (Schmitt, 1992, Lemma 2.1) to derive

$$\begin{aligned}\left\|\boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural}\boldsymbol{U}^{\natural\top}-\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{U}^{\star\top}\right\| &\lesssim \frac{1}{\sigma_r^{\natural}+\sigma_r^{\star}}\left\|\boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{U}^{\natural\top}-\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star 2}\boldsymbol{U}^{\star\top}\right\| \\ &\asymp \frac{1}{\sigma_r^{\star}}\left\|\boldsymbol{M}^{\natural}\boldsymbol{M}^{\natural\top}-\boldsymbol{S}^{\star}\right\|,\end{aligned}$$

where the last step follows from (D.10). To summarize, with probability exceeding $1-O((n+d)^{-100})$ one has

$$\|\boldsymbol{Q}-\boldsymbol{J}\| \lesssim \frac{1}{\sigma_r^{\star}}\left\|\boldsymbol{Q}\boldsymbol{\Sigma}^{\natural}-\boldsymbol{\Sigma}^{\star}\boldsymbol{Q}\right\| \lesssim \frac{1}{\sigma_r^{\star 2}}\left\|\boldsymbol{M}^{\natural}\boldsymbol{M}^{\natural\top}-\boldsymbol{S}^{\star}\right\| \lesssim \kappa\sqrt{\frac{r+\log(n+d)}{n}}, \tag{E.58}$$

where we have made used of the properties (E.55) and (E.53).

5. In view of (E.55), we know that with probability exceeding $1-O((n+d)^{-100})$, the conditional number of $\boldsymbol{M}^{\natural}$ satisfies

$$\kappa^{\natural} \asymp \sqrt{\kappa}.$$

Recalling that $\boldsymbol{U}^{\natural}=\boldsymbol{U}^{\star}\boldsymbol{Q}$, we can see from the incoherence assumption that

$$\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}=\|\boldsymbol{U}^{\star}\boldsymbol{Q}\|_{2,\infty}=\|\boldsymbol{U}^{\star}\|_{2,\infty}\leq\sqrt{\frac{\mu r}{d}}. \tag{E.59}$$

In addition, it is readily seen from (D.3) that

$$\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty}=\left\|\frac{1}{\sqrt{n}}\boldsymbol{F}^{\top}\boldsymbol{J}\right\|_{2,\infty}\overset{\text{(i)}}{\lesssim}\sqrt{\frac{1}{n}}\left\|\boldsymbol{F}^{\top}\right\|_{2,\infty}\overset{\text{(ii)}}{\lesssim}\sqrt{\frac{r\log(n+d)}{n}}$$

with probability exceeding $1-O((n+d)^{-100})$. Here, (i) holds since, according to (E.58),

$$\|\boldsymbol{J}\|\leq\|\boldsymbol{Q}\|+\|\boldsymbol{J}-\boldsymbol{Q}\|\leq 1+\kappa\sqrt{\frac{r+\log(n+d)}{n}}\leq 2 \tag{E.60}$$

holds as long as $n\gg\kappa^2(r+\log(n+d))$; and (ii) follows from the standard Gaussian concentration inequality. Combine (E.59) and (D.13) to reach

$$\mu^{\natural}\lesssim\mu+\log(n+d)$$

with probability exceeding $1-O((n+d)^{-100})$.

6. Invoke Chen et al. (2019a, Lemma 14) to show that, for all $l\in[d]$,

$$\left\|\frac{1}{n}\sum_{j=1}^{n}\left(\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_j\right)^2\boldsymbol{f}_j\boldsymbol{f}_j^{\top}-\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\boldsymbol{I}_r-2\boldsymbol{\Sigma}^{\star}\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|\lesssim\sqrt{\frac{r\log^3(n+d)}{n}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 \tag{E.61}$$

holds with probability exceeding $1-O((n+d)^{-100})$, provided that $n\gg r\log^3(n+d)$. In addition, we make note of the following lemmas.

Lemma 21. *Assume that $n\gg\log(n+d)$. For any fixed vector $\boldsymbol{u}\in\mathbb{R}^r$, with probability exceeding $1-O((n+d)^{-100})$ we have*

$$\left|\frac{1}{n}\sum_{j=1}^{n}\left(\boldsymbol{u}^{\top}\boldsymbol{f}_j\right)^2-\|\boldsymbol{u}\|_2^2\right|\lesssim\sqrt{\frac{\log(n+d)}{n}}\|\boldsymbol{u}\|_2^2.$$

Lemma 22. *Assume that $n\gg\log(n+d)$. For any fixed unit vectors $\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^r$, with probability exceeding $1-O((n+d)^{-100})$ we have*

$$\left|\frac{1}{n}\sum_{j=1}^{n}\left(\boldsymbol{u}^{\top}\boldsymbol{f}_j\right)^2\left(\boldsymbol{v}^{\top}\boldsymbol{f}_j\right)^2-\|\boldsymbol{u}\|_2^2\|\boldsymbol{v}\|_2^2-2\left(\boldsymbol{u}^{\top}\boldsymbol{v}\right)^2\right|\lesssim\sqrt{\frac{\log(n+d)^3}{n}}\|\boldsymbol{u}\|_2^2\|\boldsymbol{v}\|_2^2.$$

In view of Lemma 21 and Lemma 22, we know that for each $i, l \in [d]$

$$\left| \frac{1}{n} \sum_{j=1}^{n} \left(\boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \boldsymbol{f}_j \right)^2 \left(\boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \boldsymbol{f}_j \right)^2 - \left[\left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2 \left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2 + 2 S_{i,l}^{\star 2} \right] \right| \lesssim \sqrt{\frac{\log (n+d)^3}{n}} \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2 \left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2 \tag{E.62a}$$

$$\text{and} \qquad \left| \frac{1}{n} \sum_{j=1}^{n} \left(\boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \boldsymbol{f}_j \right)^2 - \left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2 \right| \lesssim \sqrt{\frac{\log (n+d)}{n}} \left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2 \tag{E.62b}$$

hold with probability exceeding $1 - O((n+d)^{-100})$.

7. With the above properties in place, we can now formally define the "good" event $\mathcal{E}_{\mathsf{good}}$ as follows

$$\mathcal{E}_{\mathsf{good}} \coloneqq \{\text{All equations from (E.52) to (E.62b) hold}\}.$$

It is immediately seen from the above analysis that

$$\mathbb{P}\left(\mathcal{E}_{\mathsf{good}}\right) \geq 1 - O\left((n+d)^{-100}\right).$$

By construction, it is self-evident that the event $\mathcal{E}_{\mathsf{good}}$ is $\sigma(\boldsymbol{F})$-measurable.

Proof of Lemma 21. Recognizing that $\boldsymbol{u}^\top \boldsymbol{f}_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \|\boldsymbol{u}\|_2^2)$, we have

$$\frac{1}{n} \sum_{j=1}^{n} \left(\boldsymbol{u}^\top \boldsymbol{f}_j \right)^2 / \|\boldsymbol{u}\|_2^2 \sim \chi_n^2,$$

where χ_n^2 denotes the chi-square distribution with n degrees of freedom. One can then apply the tail bound for chi-square random variables (see, e.g., Wainwright (2019, Example 2.5)) to establish the desired result. ☐

Proof of Lemma 22. Given that $\boldsymbol{u}^\top \boldsymbol{f}_j \sim \mathcal{N}(0, \|\boldsymbol{u}\|_2^2)$ for all $j \in [n]$, we have, with probability exceeding $1 - O((n+d)^{-100})$, that

$$\max_{j \in [n]} \left| \boldsymbol{u}^\top \boldsymbol{f}_j \right| \leq C_1 \left\| \boldsymbol{u} \right\|_2 \sqrt{\log (n+d)}$$

for some sufficiently large constant $C_1 > 0$. For each $j \in [n]$, let $X_j = (\boldsymbol{u}^\top \boldsymbol{f}_j)^2$ and $Y_j = (\boldsymbol{v}^\top \boldsymbol{f}_j)^2$ and define the event

$$\mathcal{A}_j \coloneqq \left\{ \left| \boldsymbol{u}^\top \boldsymbol{f}_j \right| \leq C_1 \left\| \boldsymbol{u} \right\|_2 \sqrt{\log (n+d)} \right\}.$$

Then with probability exceeding $1 - O((n+d)^{-100})$, it holds that

$$\frac{1}{n} \sum_{j=1}^{n} X_j Y_j = \frac{1}{n} \sum_{j=1}^{n} X_j Y_j \mathbb{1}_{\mathcal{A}_j},$$

which motivates us to decompose

$$\left| \frac{1}{n} \sum_{j=1}^{n} \left(X_j Y_j - \mathbb{E}\left[X_j Y_j \right] \right) \right| \leq \underbrace{\left| \frac{1}{n} \sum_{j=1}^{n} \left\{ X_j Y_j \mathbb{1}_{\mathcal{A}_j} - \mathbb{E}\left[X_j Y_j \mathbb{1}_{\mathcal{A}_j} \right] \right\} \right|}_{=:\alpha_1} + \underbrace{\left| \mathbb{E}\left[X_1 Y_1 \mathbb{1}_{\mathcal{A}_1^{\mathsf{c}}} \right] \right|}_{=:\alpha_2}.$$

- Let us first bound α_2. It is straightforward to derive that

$$\begin{aligned}
\alpha_2 &= \mathbb{E}\left[\left(\boldsymbol{u}^\top \boldsymbol{f}_1 \right)^2 \left(\boldsymbol{v}^\top \boldsymbol{f}_1 \right)^2 \mathbb{1}_{|\boldsymbol{u}^\top \boldsymbol{f}_1| > C_1 \|\boldsymbol{u}\|_2 \sqrt{\log(n+d)}} \right] \\
&\overset{\text{(i)}}{\leq} \left(\mathbb{E}\left[\left(\boldsymbol{u}^\top \boldsymbol{f}_1 \right)^6 \right] \right)^{\frac{1}{3}} \left(\mathbb{E}\left[\left(\boldsymbol{v}^\top \boldsymbol{f}_1 \right)^6 \right] \right)^{\frac{1}{3}} \left[\mathbb{P}\left(\left| \boldsymbol{u}^\top \boldsymbol{f}_1 \right| > C_1 \left\| \boldsymbol{u} \right\|_2 \sqrt{\log (n+d)} \right) \right]^{1/3} \\
&\lesssim \left\| \boldsymbol{u} \right\|_2^2 \left\| \boldsymbol{v} \right\|_2^2 \exp\left(-\frac{C_1^2 \log (n+d)}{6} \right) \overset{\text{(ii)}}{\lesssim} \left\| \boldsymbol{u} \right\|_2^2 \left\| \boldsymbol{v} \right\|_2^2 (n+d)^{-100},
\end{aligned}$$

  where (i) comes from Hölder's inequality, and (ii) holds for C_1 large enough.

- Next, we shall apply the Bernstein inequality (Vershynin, 2017, Theorem 2.8.2) to bound α_1. Note that for each $j \in [n]$

$$\left\| X_j Y_j \,\mathbb{1}_{\mathcal{A}_j} - \mathbb{E}\left[X_j Y_j \,\mathbb{1}_{\mathcal{A}_j} \right] \right\|_{\psi_1} \leq \left\| X_j Y_j \,\mathbb{1}_{\mathcal{A}_j} \right\|_{\psi_1} + \left\| \mathbb{E}\left[X_j Y_j \,\mathbb{1}_{\mathcal{A}_j} \right] \right\|_{\psi_1}, \tag{E.63}$$

where $\|\cdot\|_{\psi_1}$ denotes the sub-exponential norm (Vershynin, 2012). The first term of (E.63) obeys

$$\begin{aligned} \left\| X_j Y_j \,\mathbb{1}_{\mathcal{A}_j} \right\|_{\psi_1} &\leq C_1^2 \left\| \boldsymbol{u} \right\|_2^2 \log(n+d) \left\| Y_j \right\|_{\psi_1} \lesssim C_1^2 \left\| \boldsymbol{u} \right\|_2^2 \log(n+d) \left\| \boldsymbol{v}^\top \boldsymbol{f}_j \right\|_{\psi_2}^2 \\ &\lesssim C_1^2 \log(n+d) \left\| \boldsymbol{u} \right\|_2^2 \left\| \boldsymbol{v} \right\|_2^2, \end{aligned}$$

where $\|\cdot\|_{\psi_2}$ denotes the sub-Gaussian norm (Vershynin, 2012). Turning to the second term of (E.63), we have

$$\begin{aligned} \left\| \mathbb{E}\left[X_j Y_j \,\mathbb{1}_{\mathcal{A}_j} \right] \right\|_{\psi_1} \lesssim \mathbb{E}\left[X_j Y_j \,\mathbb{1}_{\mathcal{A}_j} \right] &\leq \mathbb{E}\left[\left(\boldsymbol{u}^\top \boldsymbol{f}_j \right)^2 \left(\boldsymbol{v}^\top \boldsymbol{f}_j \right)^2 \right] \\ &= \left\| \boldsymbol{u} \right\|_2^2 \left\| \boldsymbol{v} \right\|_2^2 + 2 \left(\boldsymbol{u}^\top \boldsymbol{v} \right)^2 \leq 3 \left\| \boldsymbol{u} \right\|_2^2 \left\| \boldsymbol{v} \right\|_2^2, \end{aligned}$$

where the last step arises from Cauchy-Schwarz. The above results taken collectively give

$$\left\| X_j Y_j \,\mathbb{1}_{\mathcal{A}_j} - \mathbb{E}\left[X_j Y_j \,\mathbb{1}_{\mathcal{A}_j} \right] \right\|_{\psi_1} \lesssim C_1^2 \log(n+d) \left\| \boldsymbol{u} \right\|_2^2 \left\| \boldsymbol{v} \right\|_2^2.$$

Applying the Bernstein inequality (Vershynin, 2017, Theorem 2.8.2) then yields

$$\alpha_1 \lesssim \sqrt{\frac{\log(n+d)^3}{n}} \left\| \boldsymbol{u} \right\|_2^2 \left\| \boldsymbol{v} \right\|_2^2 + \frac{\log^2(n+d)}{n} \left\| \boldsymbol{u} \right\|_2^2 \left\| \boldsymbol{v} \right\|_2^2 \lesssim \sqrt{\frac{\log(n+d)^3}{n}} \left\| \boldsymbol{u} \right\|_2^2 \left\| \boldsymbol{v} \right\|_2^2$$

with probability exceeding $1 - O((n+d)^{-100})$, provided that $n \gg \log(n+d)$.

Combine the preceding bounds on α_1 and α_2 to achieve

$$\begin{aligned} \left| \frac{1}{n} \sum_{j=1}^n \left[X_j Y_j - \mathbb{E}\left[X_j Y_j \right] \right] \right| \leq \alpha_1 + \alpha_2 &\lesssim \sqrt{\frac{\log(n+d)^3}{n}} \left\| \boldsymbol{u} \right\|_2^2 \left\| \boldsymbol{v} \right\|_2^2 + (n+d)^{-10} \left\| \boldsymbol{u} \right\|_2^2 \left\| \boldsymbol{v} \right\|_2^2 \\ &\lesssim \sqrt{\frac{\log(n+d)^3}{n}} \left\| \boldsymbol{u} \right\|_2^2 \left\| \boldsymbol{v} \right\|_2^2 \end{aligned}$$

with probability exceeding $1 - O((n+d)^{-100})$. It is straightforward to verify that

$$\mathbb{E}\left[X_j Y_j \right] = \left\| \boldsymbol{u} \right\|_2^2 \left\| \boldsymbol{v} \right\|_2^2 + 2 \left(\boldsymbol{u}^\top \boldsymbol{v} \right)^2$$

for each $j \in [n]$, thus concluding the proof. □

F Analysis for HeteroPCA applied to subspace estimation (Theorem 10)

This section outlines the proof that establishes our statistical guarantees stated in Theorem 10. We shall begin by isolating several useful lemmas, and then combine these lemmas to complete the proof.

F.1 A few key lemmas

We now state below a couple of key lemmas that lead to improved statistical guarantees of Algorithm 7. From now on, we will denote

$$\boldsymbol{G} = \boldsymbol{G}^{t_0}, \qquad \boldsymbol{U} = \boldsymbol{U}^{t_0} \qquad \text{and} \qquad \boldsymbol{\Sigma} = (\boldsymbol{\Lambda}^{t_0})^{1/2}$$

and let

$$\boldsymbol{H} \coloneqq \boldsymbol{U}^\top \boldsymbol{U}^\natural.$$

To begin with, the first lemma controls the discrepancy between the sample gram matrix $\boldsymbol{G}$ and the ground truth $\boldsymbol{G}^\natural$, which improves upon prior theory developed in Zhang et al. (2018).

Lemma 23. *Instate the assumptions in Theorem 10. Suppose that the number of iterations exceeds* $t_0 \geq \log\left(\frac{\sigma_1^{\star 2}}{\zeta_{\mathsf{op}}}\right)$. *Then with probability exceeding* $1 - O(n^{-10})$, *the iterate* $\boldsymbol{G} := \boldsymbol{G}^{t_0}$ *computed in Algorithm 7 satisfies*

$$\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}-\boldsymbol{G}^{\natural}\right)\right\| \lesssim \kappa^{\natural 2}\sqrt{\frac{\mu^{\natural} r}{n_1}}\zeta_{\mathsf{op}}, \tag{F.1}$$

$$\left\|\boldsymbol{G}-\boldsymbol{G}^{\natural}\right\| \lesssim \zeta_{\mathsf{op}}, \tag{F.2}$$

where ζ_{op} *is defined in (6.16).*

Proof. See Appendix F.3.2. □

Lemma (23) makes clear that $\boldsymbol{G}$ converges to $\boldsymbol{G}^{\natural}$ as the signal-to-noise ratio increases. We pause to compare this lemma with its counterpart in Zhang et al. (2018). Specifically, Zhang et al. (2018, Theorem 7) and its proof demonstrated that $\|\mathcal{P}_{\mathsf{diag}}(\boldsymbol{G}-\boldsymbol{G}^{\natural})\| \lesssim \zeta_{\mathsf{op}}$. In comparison our result (F.1) strengthens the prior estimation bound by a factor of $1/\sqrt{n_1}$ for the scenario with $\kappa^{\natural}, \mu^{\natural}, r = O(1)$. This improvement serves as one of the key analysis ingredients that allows us to sharpen the statistical guarantees for $\mathsf{HeteroPCA}$.

The above bound on the difference between $\boldsymbol{G}$ and $\boldsymbol{G}^{\natural}$ in turn allows one to develop perturbation bounds for the eigenspace measured by the spectral norm, as stated in the following lemma. In the meantime, this lemma also contains some basic facts regarding $\boldsymbol{H}$ and $\boldsymbol{R}_{\boldsymbol{U}}$.

Lemma 24. *Instate the assumptions in Theorem 10, and recall the definition of* ζ_{op} *in (6.16). Then with probability exceeding* $1 - O(n^{-10})$, *we have*

$$\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\star}\| \lesssim \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}} \qquad \textit{and} \qquad \|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^{\star}\| \lesssim \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}, \tag{F.3}$$

$$\|\boldsymbol{H} - \boldsymbol{R}_{\boldsymbol{U}}\| \lesssim \frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}} \qquad \textit{and} \qquad \left\|\boldsymbol{H}^{\top}\boldsymbol{H} - \boldsymbol{I}_r\right\| \lesssim \frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}},$$

and

$$\frac{1}{2} \leq \sigma_i\left(\boldsymbol{H}\right) \leq 2, \qquad \forall 1 \leq i \leq r.$$

Proof. See Appendix F.3.3. □

While the above two lemmas focus on spectral norm metrics, the following lemma takes one substantial step further by characterizing the difference between $\boldsymbol{G}$ and $\boldsymbol{G}^{\natural}$ in each row, when projected onto the subspace spanned by $\boldsymbol{U}^{\natural}$.

Lemma 25. *Suppose that* $n_1 \gtrsim \kappa^{\natural 4}\mu^{\natural} r$. *Then with probability exceeding* $1 - O(n^{-10})$, *we have*

$$\left\|\left(\boldsymbol{G}-\boldsymbol{G}^{\natural}\right)_{m,\cdot}\boldsymbol{U}^{\natural}\right\|_2 \lesssim \zeta_{\mathsf{op}}\sqrt{\frac{\mu^{\natural} r}{n_1}}$$

simultaneously for each $m \in [n_1]$, *with* ζ_{op} *defined in (6.16).*

Proof. See Appendix F.3.4. □

Next, we present a crucial technical lemma that uncovers the intertwined relation between $\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{H}$ and $\boldsymbol{G}\boldsymbol{U}^{\natural}$ in an $\ell_{2,\infty}$ sense. To establish the $\ell_{2,\infty}$ bound in this lemma, we invoke the powerful leave-one-out analysis framework to decouple complicated statistical dependency.

Lemma 26. *Suppose that* $n_1 \gtrsim \kappa^{\natural 4}\mu^{\natural} r$ *and* $\zeta_{\mathsf{op}} \ll \sigma_r^{\natural 2}/\kappa^{\natural 2}$. *Then with probability exceeding* $1 - O(n^{-10})$, *we have*

$$\left\|\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{H} - \boldsymbol{G}\boldsymbol{U}^{\natural}\right\|_{2,\infty} = \left\|\boldsymbol{G}\left(\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\natural}\right)\right\|_{2,\infty} \lesssim \zeta_{\mathsf{op}}\left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\natural}\right\|_{2,\infty} + \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^{\natural} r}{n_1}}.$$

Here, ζ_{op} *is a quantity defined in (6.16).*

Proof. See Appendix F.3.5. ☐

Furthermore, the lemma below reveals that $\boldsymbol{\Sigma}^2$ and $\boldsymbol{\Sigma}^{\natural 2}$ remain close even after $\boldsymbol{\Sigma}^2$ is rotated by the rotation matrix $\boldsymbol{R}_{\boldsymbol{U}}$.

Lemma 27. *Suppose that $n_1 \gtrsim \mu^{\natural 2} r \log^2 n$ and $n_2 \gtrsim r \log^4 n$. Then with probability exceeding $1 - O(n^{-10})$ we have*

$$\left\|\boldsymbol{R}_{\boldsymbol{U}}^{\top} \boldsymbol{\Sigma}^2 \boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{\Sigma}^{\natural 2}\right\| \lesssim \kappa^{\natural 2} \sqrt{\frac{\mu^{\natural} r}{n_1}} \zeta_{\mathsf{op}} + \kappa^{\natural 2} \frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}},$$

where ζ_{op} is defined in (6.16).

Proof. See Appendix F.3.6. ☐

With the above auxiliary results in place, we can put them together to yield the following $\ell_{2,\infty}$ statistical guarantees.

Lemma 28. *Suppose that $\zeta_{\mathsf{op}} \ll \sigma_r^{\natural 2} / \kappa^{\natural 2}$ and that*

$$n_1 \gtrsim \kappa^{\natural 4} \mu^{\natural} r + \mu^{\natural 2} r \log^2 n, \qquad n_2 \gtrsim r \log^4 n.$$

Then with probability at least $1 - O(n^{-10})$, one has

$$\left\|\boldsymbol{U} \boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^{\natural}\right\|_{2,\infty} \lesssim \kappa^{\natural 2} \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}} \sqrt{\frac{\mu^{\natural} r}{n_1}},$$

where ζ_{op} is defined in (6.16).

Proof. See Appendix F.3.7. ☐

F.2 Proof of Theorem 10

Armed with the above lemmas, we are in a position to establish Theorem 10. It is worth noting that, while Lemma 28 delivers $\ell_{2,\infty}$ perturbation bounds for the eigenspace, it falls short of revealing the relation between the estimation error and the desired approximation

$$\boldsymbol{Z} = \boldsymbol{E} \boldsymbol{V}^{\natural} (\boldsymbol{\Sigma}^{\natural})^{-1} + \mathcal{P}_{\mathsf{off\text{-}diag}}(\boldsymbol{E}\boldsymbol{E}^{\top}) \boldsymbol{U}^{\natural} (\boldsymbol{\Sigma}^{\natural})^{-2}. \tag{F.4}$$

In order to justify the tightness of this approximation $\boldsymbol{Z}$, we intend to establish each of the following steps:

$$\boldsymbol{U} \boldsymbol{R}_{\boldsymbol{U}} \boldsymbol{\Sigma}^{\natural 2} \overset{\text{Step 1}}{\approx} \boldsymbol{U} \boldsymbol{\Sigma}^2 \boldsymbol{R}_{\boldsymbol{U}} \overset{\text{Step 2}}{\approx} \boldsymbol{U} \boldsymbol{\Sigma}^2 \boldsymbol{H} \overset{\text{Step 3}}{\approx} \boldsymbol{G} \boldsymbol{U}^{\natural} \overset{\text{Step 4}}{\approx} \boldsymbol{U}^{\natural} \boldsymbol{\Sigma}^{\natural 2} + \underbrace{\boldsymbol{E} \boldsymbol{V}^{\natural} \boldsymbol{\Sigma}^{\natural} + \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^{\top}\right) \boldsymbol{U}^{\natural}}_{= \boldsymbol{Z} (\boldsymbol{\Sigma}^{\natural})^2}, \tag{F.5}$$

which would in turn ensure that

$$\boldsymbol{U} \boldsymbol{R}_{\boldsymbol{U}} \approx \boldsymbol{U}^{\natural} + \boldsymbol{Z}$$

as advertised.

F.2.1 Step 1: establishing the proximity of $\boldsymbol{U} \boldsymbol{R}_{\boldsymbol{U}} \boldsymbol{\Sigma}^{\natural 2}$ and $\boldsymbol{U} \boldsymbol{\Sigma}^2 \boldsymbol{R}_{\boldsymbol{U}}$.

Lemma 27 tells us that

$$\begin{aligned}
\left\|\boldsymbol{U} \boldsymbol{R}_{\boldsymbol{U}} \boldsymbol{\Sigma}^{\natural 2} - \boldsymbol{U} \boldsymbol{\Sigma}^2 \boldsymbol{R}_{\boldsymbol{U}}\right\|_{2,\infty} &= \left\|\boldsymbol{U} \boldsymbol{R}_{\boldsymbol{U}} \left(\boldsymbol{\Sigma}^{\natural 2} - \boldsymbol{R}_{\boldsymbol{U}}^{\top} \boldsymbol{\Sigma}^2 \boldsymbol{R}_{\boldsymbol{U}}\right)\right\|_{2,\infty} \leq \left\|\boldsymbol{U} \boldsymbol{R}_{\boldsymbol{U}}\right\|_{2,\infty} \left\|\boldsymbol{\Sigma}^{\natural 2} - \boldsymbol{R}_{\boldsymbol{U}}^{\top} \boldsymbol{\Sigma}^2 \boldsymbol{R}_{\boldsymbol{U}}\right\| \\
&\leq \sqrt{\frac{\mu^{\natural} r}{n_1}} \left\|\boldsymbol{\Sigma}^{\natural 2} - \boldsymbol{R}_{\boldsymbol{U}}^{\top} \boldsymbol{\Sigma}^2 \boldsymbol{R}_{\boldsymbol{U}}\right\| \\
&\asymp \sqrt{\frac{\mu^{\natural} r}{n_1}} \left(\kappa^{\natural 2} \sqrt{\frac{\mu^{\natural} r}{n_1}} \zeta_{\mathsf{op}} + \kappa^{\natural 2} \frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\right).
\end{aligned} \tag{F.6}$$

Here, the penultimate inequality follows since — according to Lemma 28 — we have

$$\|\boldsymbol{U}\|_{2,\infty} \leq \left\|\boldsymbol{U}^\natural\right\|_{2,\infty} + \left\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^\natural\right\|_{2,\infty} \lesssim \sqrt{\frac{\mu^\natural r}{n_1}} + \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}} \lesssim \sqrt{\frac{\mu^\natural r}{n_1}} \tag{F.7}$$

under the condition $\zeta_{\mathsf{op}}/\sigma_r^{\natural 2} \lesssim 1/\kappa^{\natural 2}$, while the last relation arises from Lemma 27.

F.2.2 Step 2: replacing $\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{R}_{\boldsymbol{U}}$ with $\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{H}$.

Given that $\boldsymbol{R}_{\boldsymbol{U}}$ and $\boldsymbol{H}$ are fairly close, one can expect that replacing $\boldsymbol{R}_{\boldsymbol{U}}$ with $\boldsymbol{H}$ in $\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{R}_{\boldsymbol{U}}$ does not change the matrix by much. To formalize this, we invoke Lemma 24 to reach

$$\left\|\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{H} - \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{R}_{\boldsymbol{U}}\right\|_{2,\infty} \leq \|\boldsymbol{U}\|_{2,\infty}\left\|\boldsymbol{\Sigma}^2\right\|\|\boldsymbol{H} - \boldsymbol{R}_{\boldsymbol{U}}\| \lesssim \sqrt{\frac{\mu^\natural r}{n_1}}\sigma_1^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}} \asymp \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^\natural r}{n_1}}. \tag{F.8}$$

Here, the penultimate relation follows from Lemma 24, (F.7), as well as a direct consequence of Lemma 27:

$$\left\|\boldsymbol{\Sigma}^2\right\| = \left\|\boldsymbol{R}_{\boldsymbol{U}}^\top\boldsymbol{\Sigma}^2\boldsymbol{R}_{\boldsymbol{U}}\right\| \leq \left\|\boldsymbol{\Sigma}^{\natural 2}\right\| + \left\|\boldsymbol{R}_{\boldsymbol{U}}^\top\boldsymbol{\Sigma}^2\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{\Sigma}^{\natural 2}\right\| \leq \sigma_1^{\natural 2} + \kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\zeta_{\mathsf{op}} + \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}} \asymp \sigma_1^{\natural 2}, \tag{F.9}$$

provided that $\zeta_{\mathsf{op}}/\sigma_r^{\natural 2} \lesssim 1$ and $n_1 \gtrsim \mu^\natural r$.

F.2.3 Step 3: establishing the proximity of $\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{H}$ and $\boldsymbol{G}\boldsymbol{U}^\natural$.

It is readily seen from Lemma 26 that

$$\left\|\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{H} - \boldsymbol{G}\boldsymbol{U}^\natural\right\|_{2,\infty} \lesssim \zeta_{\mathsf{op}}\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right\|_{2,\infty} + \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^\natural r}{n_1}} \asymp \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^\natural r}{n_1}}, \tag{F.10}$$

where the last inequality results from Lemma 28.

F.2.4 Step 4: investigating the statistical properties of $\boldsymbol{G}\boldsymbol{U}^\natural$.

It then remains to decompose $\boldsymbol{G}\boldsymbol{U}^\natural$ as claimed in (F.5). To begin with, we make the observation that

$$\begin{aligned}
\boldsymbol{G} &= \mathcal{P}_{\mathsf{off\text{-}diag}}\left[\left(\boldsymbol{M}^\natural + \boldsymbol{E}\right)\left(\boldsymbol{M}^\natural + \boldsymbol{E}\right)^\top\right] + \mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^\natural\right) + \mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G} - \boldsymbol{G}^\natural\right) \\
&= \boldsymbol{G}^\natural + \mathcal{P}_{\mathsf{off\text{-}diag}}\left[\boldsymbol{E}\boldsymbol{M}^{\natural\top} + \boldsymbol{M}^\natural\boldsymbol{E}^\top + \boldsymbol{E}\boldsymbol{E}^\top\right] + \mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G} - \boldsymbol{G}^\natural\right),
\end{aligned}$$

which together with the eigen-decomposition $\boldsymbol{G}^\natural = \boldsymbol{U}^\natural\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{U}^{\natural\top}$ and the definition (F.4) of $\boldsymbol{Z}$ allows one to derive

$$\begin{aligned}
\boldsymbol{G}\boldsymbol{U}^\natural - \left(\boldsymbol{U}^\natural + \boldsymbol{Z}\right)\boldsymbol{\Sigma}^{\natural 2} &= \boldsymbol{U}^\natural\boldsymbol{\Sigma}^{\natural 2} + \mathcal{P}_{\mathsf{off\text{-}diag}}\left[\boldsymbol{E}\boldsymbol{M}^{\natural\top} + \boldsymbol{M}^\natural\boldsymbol{E}^\top + \boldsymbol{E}\boldsymbol{E}^\top\right]\boldsymbol{U}^\natural + \mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G} - \boldsymbol{G}^\natural\right)\boldsymbol{U}^\natural \\
&\quad - \boldsymbol{U}^\natural\boldsymbol{\Sigma}^{\natural 2} - \left[\boldsymbol{E}\boldsymbol{M}^{\natural\top} + \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]\boldsymbol{U}^\natural \\
&= \underbrace{\boldsymbol{M}^\natural\boldsymbol{E}^\top\boldsymbol{U}^\natural}_{=:\boldsymbol{R}_1} - \underbrace{\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{E}\boldsymbol{M}^{\natural\top} + \boldsymbol{M}^\natural\boldsymbol{E}^\top\right)\boldsymbol{U}^\natural}_{=:\boldsymbol{R}_2} + \underbrace{\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G} - \boldsymbol{G}^\natural\right)\boldsymbol{U}^\natural}_{=:\boldsymbol{R}_3}.
\end{aligned}$$

This motivates us to control the terms $\boldsymbol{R}_1$, $\boldsymbol{R}_2$ and $\boldsymbol{R}_3$ separately.

- Regarding $\boldsymbol{R}_1$, note that we have shown in (C.11) that

$$\left\|\boldsymbol{U}^{\natural\top}\boldsymbol{E}\boldsymbol{V}^\natural\right\| \lesssim \sigma\sqrt{r\log n} + \frac{B\mu^\natural r\log n}{\sqrt{n_1 n_2}} \lesssim \sigma\sqrt{r\log n} + \frac{\sigma\mu^\natural r}{\sqrt[4]{n_1 n_2}}$$

with probability exceeding $1 - O(n^{-10})$, where the last inequality holds under Assumption 5. Therefore, we can derive

$$\begin{aligned}\|\boldsymbol{R}_1\|_{2,\infty} &= \left\|\boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural}\left(\boldsymbol{U}^{\natural\top}\boldsymbol{E}\boldsymbol{V}^{\natural}\right)^{\top}\right\|_{2,\infty} \le \left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}\left\|\boldsymbol{\Sigma}^{\natural}\right\|\left\|\boldsymbol{U}^{\natural\top}\boldsymbol{E}\boldsymbol{V}^{\natural}\right\| \\ &\lesssim \sqrt{\frac{\mu^{\natural}r}{n_1}}\sigma_1^{\natural}\left(\sigma\sqrt{r\log n} + \frac{\sigma\mu^{\natural}r}{\sqrt[4]{n_1 n_2}}\right) \lesssim \frac{\mu^{\natural}r}{n_1}\zeta_{\mathsf{op}},\end{aligned}$$

where the last relation holds provided that $n_1 n_2 \gtrsim \mu^{\natural 2} r^2$.

- With regards to $\boldsymbol{R}_2$, we observe from the incoherence condition that

$$\|\boldsymbol{R}_2\|_{2,\infty} \le \max_{1\le i\le n_1}\left\{\left|\sum_{j=1}^{n_2} E_{i,j}M_{i,j}^{\natural}\right|\left\|\boldsymbol{U}_{i,\cdot}^{\natural}\right\|_2\right\} \le \max_{1\le i\le n_1}\left|\sum_{j=1}^{n_2} E_{i,j}M_{i,j}^{\natural}\right|\sqrt{\frac{\mu^{\natural}r}{n_1}}.$$

  Recalling from (F.44) that

$$\max_{1\le i\le n_1}\left|\sum_{j=1}^{n_2} E_{i,j}M_{i,j}^{\natural}\right| \lesssim \frac{\sqrt{\mu^{\natural}r}}{n_1}\zeta_{\mathsf{op}}$$

  holds with probability exceeding $1 - O(n^{-10})$, we can upper bound

$$\|\boldsymbol{R}_2\|_{2,\infty} \lesssim \zeta_{\mathsf{op}}\frac{\mu^{\natural}r}{n_1^{3/2}}.$$

- It remains to control $\|\boldsymbol{R}_3\|_{2,\infty}$, towards which we can apply Lemma 23 to reach

$$\|\boldsymbol{R}_3\|_{2,\infty} = \max_{1\le i\le n_1}\left\{\left|G_{i,i} - G_{i,i}^{\natural}\right|\left\|\boldsymbol{U}_{i,\cdot}^{\natural}\right\|_2\right\} \le \left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G} - \boldsymbol{G}^{\natural}\right)\right\|\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} \lesssim \kappa^{\natural 2}\frac{\mu^{\natural}r}{n_1}\zeta_{\mathsf{op}}.$$

Taking the preceding bounds on $\|\boldsymbol{R}_1\|_{2,\infty}$, $\|\boldsymbol{R}_2\|_{2,\infty}$ and $\|\boldsymbol{R}_3\|_{2,\infty}$ collectively, we arrive at

$$\left\|\boldsymbol{G}\boldsymbol{U}^{\natural} - \left(\boldsymbol{U}^{\natural} + \boldsymbol{Z}\right)\boldsymbol{\Sigma}^{\natural 2}\right\|_{2,\infty} \le \|\boldsymbol{R}_1\|_{2,\infty} + \|\boldsymbol{R}_2\|_{2,\infty} + \|\boldsymbol{R}_3\|_{2,\infty} \lesssim \kappa^{\natural 2}\frac{\mu^{\natural}r}{n_1}\zeta_{\mathsf{op}}. \tag{F.11}$$

F.2.5 Step 5: putting everything together.

To finish up, combine (F.6), (F.8), (F.10) and (F.11) to conclude that

$$\begin{aligned}\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^{\star} - \boldsymbol{Z}\|_{2,\infty} &\le \frac{1}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural 2} - \left(\boldsymbol{U}^{\natural} + \boldsymbol{Z}\right)\boldsymbol{\Sigma}^{\natural 2}\right\|_{2,\infty} \\ &\le \frac{1}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural 2} - \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{R}_{\boldsymbol{U}}\right\|_{2,\infty} + \frac{1}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{H} - \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{R}_{\boldsymbol{U}}\right\|_{2,\infty} \\ &\quad + \frac{1}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{H} - \boldsymbol{G}\boldsymbol{U}^{\natural}\right\|_{2,\infty} + \frac{1}{\sigma_r^{\natural 2}}\left\|\boldsymbol{G}\boldsymbol{U}^{\natural} - \left(\boldsymbol{U}^{\natural} + \boldsymbol{Z}\right)\boldsymbol{\Sigma}^{\natural 2}\right\|_{2,\infty} \\ &\lesssim \kappa^{\natural 2}\frac{\mu^{\natural}r}{n_1}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}} + \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}}\sqrt{\frac{\mu^{\natural}r}{n_1}}\end{aligned}$$

as claimed.

F.3 Proof of auxiliary lemmas

In this section, we establish the auxiliary lemmas needed in the proof of Theorem 10. At the core of our analysis lies a leave-one-out (and leave-two-out) analysis framework that has been previously adopted to analyze spectral methods (Abbe et al., 2020; Cai et al., 2021; Chen et al., 2020c, 2019b), which we shall introduce below.

F.3.1 Preparation: leave-one-out and leave-two-out auxiliary estimates

To begin with, let us introduce the leave-one-out and leave-two-out auxiliary sequences and the associated estimates, which play a pivotal role in enabling fine-grained statistical analysis.

Leave-one-out auxiliary estimates. For each $m \in [n_1]$, define

$$\boldsymbol{M}^{(m)} := \mathcal{P}_{-m,\cdot}(\boldsymbol{M}) + \mathcal{P}_{m,\cdot}(\boldsymbol{M}^{\natural}) \qquad \text{and} \qquad \boldsymbol{G}^{(m)} := \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{M}^{(m)}\boldsymbol{M}^{(m)\top}\right) + \mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{M}^{\natural}\boldsymbol{M}^{\natural\top}\right), \tag{F.12}$$

where we recall that $\mathcal{P}_{-m,\cdot}(\boldsymbol{M}) \in \mathbb{R}^{n_1\times n_2}$ is obtained by setting to zero all entries in the m-th row of $\boldsymbol{M}$, and $\mathcal{P}_{m,\cdot}(\boldsymbol{M}) = \boldsymbol{M} - \mathcal{P}_{-m,\cdot}(\boldsymbol{M})$. Throughout this section, we let $\boldsymbol{U}^{(m)}\boldsymbol{\Lambda}^{(m)}\boldsymbol{U}^{(m)\top}$ be the top-r eigen-decomposition of $\boldsymbol{G}^{(m)}$, and we define

$$\boldsymbol{H}^{(m)} := \boldsymbol{U}^{(m)\top}\boldsymbol{U}^{\natural}. \tag{F.13}$$

Two features are particularly worth emphasizing:

- The matrices $\boldsymbol{U}^{(m)}$, $\boldsymbol{G}^{(m)}$ and $\boldsymbol{H}^{(m)}$ are all statistically independent of the m-th row of the noise matrix $\boldsymbol{E}$, given that $\boldsymbol{M}^{(m)}$ does not contain any randomness arising in the m-th row of $\boldsymbol{E}$.
- The estimate $\boldsymbol{U}^{(m)}$ (resp. $\boldsymbol{H}^{(m)}$ and $\boldsymbol{G}^{(m)}$) is expected to be extremely close to the original estimate $\boldsymbol{U}$ (resp. $\boldsymbol{H} = \boldsymbol{U}^{\top}\boldsymbol{U}^{\natural}$ and $\boldsymbol{G}$), given that we have only dropped a small fraction of data when generating the leave-one-out estimate.

Informally, the above two features taken together allow one to show the weak statistical dependency between $(\boldsymbol{U}, \boldsymbol{G}, \boldsymbol{H})$ and the m-th row of $\boldsymbol{E}$.

Leave-two-out auxiliary estimates. As it turns out, we are also in need of a collection of slightly more complicated leave-two-out estimates to assist in our analysis. Specifically, for each $m \in [n_1]$ and $l \in [n_2]$, we let

$$\boldsymbol{M}^{(m,l)} = \mathcal{P}_{-m,-l}(\boldsymbol{M}) + \mathcal{P}_{m,l}(\boldsymbol{M}^{\natural}) \qquad \text{and} \qquad \boldsymbol{G}^{(m,l)} = \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{M}^{(m,l)}\boldsymbol{M}^{(m,l)\top}\right) + \mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{M}^{\natural}\boldsymbol{M}^{\natural\top}\right). \tag{F.14}$$

Here, $\mathcal{P}_{-m,-l}(\boldsymbol{M}) \in \mathbb{R}^{n_1\times n_2}$ is obtained by zeroing out all entries in the m-th row and the l-th column of $\boldsymbol{M}$, and $\mathcal{P}_{m,l}(\boldsymbol{M}) = \boldsymbol{M} - \mathcal{P}_{-m,-l}(\boldsymbol{M})$. We let $\boldsymbol{U}^{(m,l)}\boldsymbol{\Lambda}^{(m,l)}\boldsymbol{U}^{(m,l)\top}$ represent the top-r eigen-decomposition of $\boldsymbol{G}^{(m,l)}$, and define

$$\boldsymbol{H}^{(m,l)} := \boldsymbol{U}^{(m,l)\top}\boldsymbol{U}^{\natural}. \tag{F.15}$$

Akin to the leave-one-out counterpart, the matrices $\boldsymbol{U}^{(m,l)}$, $\boldsymbol{G}^{(m,l)}$ and $\boldsymbol{H}^{(m,l)}$ we construct are all statistically independent of the m-th row and the l-th column of $\boldsymbol{E}$, and we shall also exploit the proximity of $\boldsymbol{U}^{(m,l)}$, $\boldsymbol{U}^{(m)}$ and $\boldsymbol{U}$ in the subsequent analysis.

Useful properties concerning these auxiliary estimates. In the sequel, we collect a couple of useful lemmas that are concerned with the leave-one-out and leave-two-out estimates, several of which are adapted from Cai et al. (2021). The first lemma controls the difference between the leave-one-out gram matrix $\boldsymbol{G}^{(m)}$ and the original gram matrix $\boldsymbol{G}$, as well as the difference between $\boldsymbol{G}^{(m)}$ and the ground-truth gram matrix.

Lemma 29. *Suppose that $n_1 \gtrsim \kappa^{\natural 4}\mu^{\natural}r$. Then with probability exceeding $1 - O(n^{-10})$,*

$$\begin{aligned}
\left\|\boldsymbol{G}^{(m)} - \boldsymbol{G}\right\| &\lesssim \sigma^2\sqrt{n_1 n_2}\log n + \kappa^{\natural 2}\sigma\sigma_1^{\natural}\sqrt{\mu^{\natural}r\log n} \\
\left\|\boldsymbol{G}^{(m)} - \boldsymbol{G}^{\natural}\right\| &\lesssim \zeta_{\mathsf{op}} \asymp \sigma^2\sqrt{n_1 n_2}\log n + \sigma\sigma_1^{\natural}\sqrt{n_1\log n}
\end{aligned}$$

hold simultaneously for all $1 \le m \le n_1$.

Proof. See Appendix F.3.8. □

The next lemma confirms that the leave-one-out estimate $\boldsymbol{U}^{(m)}$ and the leave-two-out estimate $\boldsymbol{U}^{(m,l)}$ are exceedingly close.

Lemma 30. *Instate the assumptions in Theorem 10. Then with probability exceeding $1 - O(n^{-10})$, we have*

$$\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top} - \boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\| \lesssim \frac{1}{\sigma_r^{\natural 2}}\left(B\log n + \sigma\sqrt{n_1 \log n}\right)^2 \left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} + \frac{\sigma^2}{\sigma_r^{\natural 2}} + \frac{1}{\sigma_r^{\natural 2}}\left(B\log n + \sigma\sqrt{n_1 \log n}\right)\left\|\boldsymbol{M}^{\natural\top}\right\|_{2,\infty}$$

simultaneously for all $1 \le m \le n_1$ and $1 \le l \le n_2$.

Proof. See Appendix F.3.9. □

The following two lemmas study the size of $\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^{\natural}$ and $\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}$ when projected towards several important directions.

Lemma 31. *Instate the assumptions in Theorem 10. Then with probability exceeding $1 - O(n^{-10})$, we have, for all $l \in [n_2]$ and all $m \in [n_1]$,*

$$\left\|\boldsymbol{e}_l^\top \left[\mathcal{P}_{-m,\cdot}(\boldsymbol{M})\right]^\top \left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^{\natural}\right)\right\|_2 \lesssim \frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}}\left\|\boldsymbol{M}^{\natural\top}\right\|_{2,\infty} + \left(B\log n + \sigma\sqrt{n_1\log n}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^{\natural}\right\|_{2,\infty} + \left(\left\|\boldsymbol{M}^{\natural\top}\right\|_{2,\infty} + B\log n + \sigma\sqrt{n_1\log n}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top} - \boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\|; \tag{F.16}$$

$$\left\|\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{M})\right]^\top \boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} \lesssim \left(B\log n + \sigma\sqrt{n_1\log n}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} + \sigma_1^{\natural}\sqrt{\frac{\mu^{\natural} r}{n_2}}. \tag{F.17}$$

Proof. See Appendix F.3.10. □

Lemma 32. *Instate the assumptions in Theorem 10. Then with probability exceeding $1 - O(n^{-10})$, we have*

$$\left\|\boldsymbol{E}_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{M})\right]^\top \left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^{\natural}\right)\right\|_2 \lesssim \zeta_{\mathsf{op}}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^{\natural}\right\|_{2,\infty} + \frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\left(\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} + \sqrt{\frac{\mu^{\natural} r}{n_1}}\right) \tag{F.18}$$

$$\left\|\boldsymbol{E}_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{M})\right]^\top \boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_2 \lesssim \zeta_{\mathsf{op}}\left(\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} + \sqrt{\frac{\mu^{\natural} r}{n_1}}\right) \tag{F.19}$$

simultaneously for all $1 \le m \le n_1$.

Proof. See Appendix F.3.11. □

Finally, we justify the proximity of the leave-one-out estimate $\boldsymbol{U}^{(m)}$ and the original estimate $\boldsymbol{U}$, which turns out to be a consequence of the preceding results.

Lemma 33. *Instate the assumptions in Theorem 10. Then with probability exceeding $1 - O(n^{-10})$,*

$$\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top} - \boldsymbol{U}\boldsymbol{U}^\top\right\| \lesssim \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\left(\left\|\boldsymbol{U}\boldsymbol{H}\right\|_{2,\infty} + \sqrt{\frac{\mu^{\natural} r}{n_1}}\right)$$

holds simultaneously for all $1 \le m \le n_1$.

Proof. See Appendix F.3.12. □

F.3.2 Proof of Lemma 23

Recall that the initialization $\boldsymbol{G}^0$ is obtained by dropping all diagonal entries of $\boldsymbol{M}\boldsymbol{M}^\top$, namely,

$$\boldsymbol{G}^0 = \mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{M}\boldsymbol{M}^\top\right),$$

and $\boldsymbol{U}^0$ consists of the top-r eigenvectors of $\boldsymbol{G}^0$. These are precisely the subjects studied in Cai et al. (2021). In light of this, we state below the following two lemmas borrowed from Cai et al. (2021), which assist in proving Lemma 23.

Lemma 34. *Instate the assumptions in Theorem 10. With probability exceeding* $1 - O(n^{-10})$,

$$\left\|\boldsymbol{G}^0 - \mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{G}^\natural\right)\right\| = \left\|\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{G}^0 - \boldsymbol{G}^\natural\right)\right\| \le C^{\textsf{dd}}\zeta_{\textsf{op}}$$
$$\left\|\boldsymbol{G}^0 - \boldsymbol{G}^\natural\right\| \le C^{\textsf{dd}}\zeta_{\textsf{op}} + \left\|\boldsymbol{M}^\natural\right\|_{2,\infty}^2$$

hold for some universal constant $C^{\textsf{dd}} > 0$.

Proof. See Cai et al. (2021, Lemma 1). □

Lemma 35. *Instate the assumptions in Theorem 10. With probability exceeding* $1 - O(n^{-10})$,

$$\left\|\boldsymbol{U}^0\boldsymbol{R}^0 - \boldsymbol{U}^\natural\right\| \le C_{\textsf{op}}^{\textsf{dd}}\frac{\zeta_{\textsf{op}} + \left\|\boldsymbol{M}^\natural\right\|_{2,\infty}^2}{\sigma_r^{\natural 2}}$$
$$\left\|\boldsymbol{U}^0\boldsymbol{R}^0 - \boldsymbol{U}^\natural\right\|_{2,\infty} \le C_{\infty}^{\textsf{dd}}\frac{\kappa^{\natural 2}\left(\zeta_{\textsf{op}} + \left\|\boldsymbol{M}^\natural\right\|_{2,\infty}^2\right)}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^\natural r}{n_1}}$$

hold for some universal constants $C_{\textsf{op}}^{\textsf{dd}}, C_{\infty}^{\textsf{dd}} > 0$. *Here,* $\boldsymbol{R}_0$ *is defined to be the following rotation matrix*

$$\boldsymbol{R}^0 \coloneqq \arg\min_{\boldsymbol{O}\in\mathcal{O}^{r\times r}}\left\|\boldsymbol{U}^0\boldsymbol{O} - \boldsymbol{U}^\natural\right\|_{\mathrm{F}}^2. \tag{F.20}$$

Proof. See Cai et al. (2021, Theorem 1). □

Armed with the above lemmas, we are ready to show by induction that: for $1 \le s \le t_0$,

$$\left\|\mathcal{P}_{\textsf{diag}}\left(\boldsymbol{G}^s - \boldsymbol{G}^\natural\right)\right\| \le C_0\kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\left\|\boldsymbol{G}^{s-1} - \boldsymbol{G}^\natural\right\| \tag{F.21}$$

for some sufficiently large constant $C_0 > 0$.

Base case: $s = 1$**.** Let us start with the base case by making the observation that

$$\begin{aligned}\left\|\mathcal{P}_{\textsf{diag}}\left(\boldsymbol{G}^1 - \boldsymbol{G}^\natural\right)\right\| &= \left\|\mathcal{P}_{\textsf{diag}}\left(\boldsymbol{U}^0\boldsymbol{\Lambda}^0\boldsymbol{U}^{0\top} - \boldsymbol{G}^\natural\right)\right\| = \left\|\mathcal{P}_{\textsf{diag}}\left(\mathcal{P}_{\boldsymbol{U}^0}(\boldsymbol{G}^0) - \boldsymbol{G}^\natural\right)\right\| \\ &\le \underbrace{\left\|\mathcal{P}_{\textsf{diag}}\left[\mathcal{P}_{\boldsymbol{U}^0}\left(\boldsymbol{G}^0 - \boldsymbol{G}^\natural\right)\right]\right\|}_{=:\alpha_1} + \underbrace{\left\|\mathcal{P}_{\textsf{diag}}\left(\mathcal{P}_{\boldsymbol{U}_\perp^0}\boldsymbol{G}^\natural\right)\right\|}_{=:\alpha_2},\end{aligned}$$

where $\boldsymbol{U}_\perp^0 \in \mathbb{R}^{n_1\times(n_1-r)}$ is a matrix whose orthonormal columns span the orthogonal complement of the column space of $\boldsymbol{U}^0$.

- Regarding the first term α_1, we have seen from Lemma 35 that

$$\left\|\boldsymbol{U}^0\right\|_{2,\infty} = \left\|\boldsymbol{U}^0\boldsymbol{R}^0\right\|_{2,\infty} \le \left\|\boldsymbol{U}^0\boldsymbol{R}^0 - \boldsymbol{U}^\natural\right\|_{2,\infty} + \left\|\boldsymbol{U}^\natural\right\|_{2,\infty} \le C_{\infty}^{\textsf{dd}}\frac{\kappa^{\natural 2}\left(\zeta_{\textsf{op}} + \left\|\boldsymbol{M}^\natural\right\|_{2,\infty}^2\right)}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^\natural r}{n_1}} + \sqrt{\frac{\mu^\natural r}{n_1}}$$

$$\overset{\text{(i)}}{\leq} C_{\infty}^{\mathsf{dd}} \kappa^{\natural 2} \left(\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}} + \frac{\mu^{\natural} r}{d} \right) \sqrt{\frac{\mu^{\natural} r}{n_1}} + \sqrt{\frac{\mu^{\natural} r}{n_1}} \overset{\text{(ii)}}{\leq} \sqrt{\frac{4\mu^{\natural} r}{n_1}}, \tag{F.22}$$

where (i) holds since

$$\left\| \boldsymbol{M}^{\natural} \right\|_{2,\infty} = \left\| \boldsymbol{U}^{\natural} \boldsymbol{\Sigma}^{\natural} \boldsymbol{V}^{\natural \top} \right\|_{2,\infty} \leq \left\| \boldsymbol{U}^{\natural} \right\|_{2,\infty} \left\| \boldsymbol{\Sigma}^{\natural} \right\| \left\| \boldsymbol{V}^{\natural} \right\| \leq \sqrt{\frac{\mu^{\natural} r}{n_1}} \sigma_1^{\natural}, \tag{F.23}$$

and (ii) holds provided that $\zeta_{\mathsf{op}}/\sigma_r^{\natural 2} \ll 1/\kappa^{\natural 2}$ and $d \gg \kappa^{\natural 2} \mu^{\natural} r$. This in turn allows us to use Zhang et al. (2018, Lemma 1) to reach

$$\alpha_1 \leq 2\sqrt{\frac{\mu^{\natural} r}{n_1}} \left\| \boldsymbol{G}^0 - \boldsymbol{G}^{\natural} \right\|. \tag{F.24}$$

- Regarding the second term α_2, invoke Zhang et al. (2018, Lemma 1) again to arrive at

$$\alpha_2 = \left\| \mathcal{P}_{\mathsf{diag}} \left(\mathcal{P}_{\boldsymbol{U}_{\perp}^0} \boldsymbol{G}^{\natural} \mathcal{P}_{\boldsymbol{U}^{\natural}} \right) \right\| \leq \sqrt{\frac{\mu^{\natural} r}{n_1}} \left\| \mathcal{P}_{\boldsymbol{U}_{\perp}^0} \boldsymbol{G}^{\natural} \right\|, \tag{F.25}$$

which has made use of the fact that $\boldsymbol{G}^{\natural} \mathcal{P}_{\boldsymbol{U}^{\natural}} = \boldsymbol{G}^{\natural}$. Moreover, it is seen that

$$\left\| \mathcal{P}_{\boldsymbol{U}_{\perp}^0} \boldsymbol{G}^{\natural} \right\| = \left\| \boldsymbol{U}_{\perp}^0 \boldsymbol{U}_{\perp}^{0\top} \boldsymbol{U}^{\natural} \boldsymbol{\Sigma}^{\natural 2} \boldsymbol{U}^{\natural \top} \right\| \leq \sigma_1^{\natural 2} \left\| \boldsymbol{U}_{\perp}^{0\top} \boldsymbol{U}^{\natural} \right\| = \sigma_1^{\natural 2} \left\| \sin \boldsymbol{\Theta} \left(\boldsymbol{U}^0, \boldsymbol{U}^{\natural} \right) \right\|, \tag{F.26}$$

where $\boldsymbol{\Theta}\left(\boldsymbol{U}^0, \boldsymbol{U}^{\natural}\right)$ is a diagonal matrix whose diagonal entries correspond to the principal angles between $\boldsymbol{U}^0$ and $\boldsymbol{U}^{\natural}$, and the last identity follows from Chen et al. (2020c, Lemma 2.1.2). In view of the Davis-Kahan $\sin \boldsymbol{\Theta}$ Theorem (Chen et al., 2020c, Theorem 2.2.1), we can demonstrate that

$$\left\| \sin \boldsymbol{\Theta} \left(\boldsymbol{U}^0, \boldsymbol{U}^{\natural} \right) \right\| \leq \frac{\sqrt{2} \left\| \boldsymbol{G}^0 - \boldsymbol{G}^{\natural} \right\|}{\lambda_r \left(\boldsymbol{G}^0 \right) - \lambda_{r+1} \left(\boldsymbol{G}^{\natural} \right)} = \frac{\sqrt{2} \left\| \boldsymbol{G}^0 - \boldsymbol{G}^{\natural} \right\|}{\lambda_r \left(\boldsymbol{G}^0 \right)} \leq \frac{2\sqrt{2} \left\| \boldsymbol{G}^0 - \boldsymbol{G}^{\natural} \right\|}{\sigma_r^{\natural 2}}. \tag{F.27}$$

Here, the identity comes from the fact $\lambda_{r+1}(\boldsymbol{G}^{\natural}) = 0$; the last inequality follows from a direct application of Weyl's inequality:

$$\lambda_r \left(\boldsymbol{G}^0 \right) \geq \lambda_r \left(\boldsymbol{G}^{\natural} \right) - \left\| \boldsymbol{G}^0 - \boldsymbol{G}^{\natural} \right\| \overset{\text{(i)}}{\geq} \sigma_r^{\natural 2} - \zeta_{\mathsf{op}} - \left\| \boldsymbol{M}^{\natural} \right\|_{2,\infty}^2 \overset{\text{(ii)}}{\geq} \sigma_r^{\natural 2} - \zeta_{\mathsf{op}} - \frac{\mu^{\natural} r}{n_1} \sigma_1^{\natural 2} \overset{\text{(iii)}}{\geq} \frac{1}{2} \sigma_r^{\natural 2},$$

where (i) is a consequence of Lemma 34, (ii) follows from (F.23), and (iii) holds as long as $\zeta_{\mathsf{op}} \ll \sigma_r^{\natural 2}$ and $n_1 \gg \kappa^{\natural 2} \mu^{\natural} r$. Combine (F.25), (F.26) and (F.27) to reach

$$\alpha_2 \leq 2\sqrt{2} \kappa^{\natural 2} \sqrt{\frac{\mu^{\natural} r}{n_1}} \left\| \boldsymbol{G}^0 - \boldsymbol{G}^{\natural} \right\|.$$

Combine the above bounds on α_1 and α_2 to yield

$$\left\| \mathcal{P}_{\mathsf{diag}} \left(\boldsymbol{G}^1 - \boldsymbol{G}^{\natural} \right) \right\| \leq \alpha_1 + \alpha_2 \leq \left(2 + 2\sqrt{2} \kappa^{\natural 2} \right) \sqrt{\frac{\mu^{\natural} r}{n_1}} \left\| \boldsymbol{G}^0 - \boldsymbol{G}^{\natural} \right\| \leq C_0 \kappa^{\natural 2} \sqrt{\frac{\mu^{\natural} r}{n_1}} \left\| \boldsymbol{G}^0 - \boldsymbol{G}^{\natural} \right\|,$$

with the proviso that the constant C_0 is sufficiently large.

Induction step. For any given $t > 1$, suppose that (F.21) holds for all $s = 1, 2 \ldots, t$, and we'd like to show that it continues to hold for $s = t+1$. From the induction hypothesis, we know that for any $1 \leq \tau \leq t$, one has

$$\left\| \mathcal{P}_{\mathsf{diag}} \left(\boldsymbol{G}^{\tau} - \boldsymbol{G}^{\natural} \right) \right\| \leq C_0 \kappa^{\natural 2} \sqrt{\frac{\mu^{\natural} r}{n_1}} \left\| \boldsymbol{G}^{\tau - 1} - \boldsymbol{G}^{\natural} \right\| \leq C_0 \kappa^{\natural 2} \sqrt{\frac{\mu^{\natural} r}{n_1}} \left(\left\| \mathcal{P}_{\mathsf{diag}} \left(\boldsymbol{G}^{\tau - 1} - \boldsymbol{G}^{\natural} \right) \right\| + \left\| \mathcal{P}_{\mathsf{off\text{-}diag}} \left(\boldsymbol{G}^{\tau - 1} - \boldsymbol{G}^{\natural} \right) \right\| \right)$$

$$= C_0\kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\left(\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^{\tau-1}-\boldsymbol{G}^\natural\right)\right\|+\left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}^0-\boldsymbol{G}^\natural\right)\right\|\right), \tag{F.28}$$

where the last line holds since, by construction, $\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}^\tau\right)=\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}^0\right)$ for all τ. Applying the above inequality recursively gives

$$\begin{aligned}
\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^t-\boldsymbol{G}^\natural\right)\right\| &\le C_0\kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\left(\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^{t-1}-\boldsymbol{G}^\natural\right)\right\|+\left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}^0-\boldsymbol{G}^\natural\right)\right\|\right)\le\cdots\\
&\le\left(C_0\kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\right)^t\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^0-\boldsymbol{G}^\natural\right)\right\|+\sum_{i=1}^t\left(C_0\kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\right)^i\left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}^0-\boldsymbol{G}^\natural\right)\right\|\\
&\le\left(C_0\kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\right)^t\left\|\boldsymbol{M}^\natural\right\|_{2,\infty}^2+2C_0\kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}^0-\boldsymbol{G}^\natural\right)\right\|\\
&\le\left(C_0\kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\right)^t\frac{\mu^\natural r}{n_1}\sigma_1^{\natural 2}+2C_0C^{\mathsf{dd}}\kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\zeta_{\mathsf{op}}.
\end{aligned}\tag{F.29}$$

Here, the penultimate line holds as long as $n_1\gg\kappa^{\natural 4}\mu^\natural r$, and the last line follows from (F.23) and Lemma 34. An immediate consequence is that

$$\begin{aligned}
\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^t\right)\right\| &\le\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^\natural\right)\right\|+\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^t-\boldsymbol{G}^\natural\right)\right\|=\left\|\boldsymbol{M}^\natural\right\|_{2,\infty}^2+\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^t-\boldsymbol{G}^\natural\right)\right\|\\
&\le\frac{\mu^\natural r}{n_1}\sigma_1^{\natural 2}+\left(C_0\kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\right)^t\frac{\mu^\natural r}{n_1}\sigma_1^{\natural 2}+2C_0C^{\mathsf{dd}}\kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\zeta_{\mathsf{op}}\\
&\le\frac{2\mu^\natural r}{n_1}\sigma_1^{\natural 2}+2C_0C^{\mathsf{dd}}\kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\zeta_{\mathsf{op}},
\end{aligned}$$

where the second line follows from (F.23) and (F.29), and the last relation is valid as long as $n_1\gg\kappa^{\natural 4}\mu^\natural r$. This together with the fact $\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}^t\right)=\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}^0\right)$ (by construction) allows one to obtain

$$\left\|\boldsymbol{G}^t-\boldsymbol{G}^0\right\|=\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^t\right)\right\|\le\frac{2\mu^\natural r}{n_1}\sigma_1^{\natural 2}+2C_0C^{\mathsf{dd}}\kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\zeta_{\mathsf{op}}. \tag{F.30}$$

In view of Weyl's inequality, we have

$$\begin{aligned}
\lambda_r\left(\boldsymbol{G}^t\right) &\ge\lambda_r\left(\boldsymbol{G}^\natural\right)-\left\|\boldsymbol{G}^t-\boldsymbol{G}^\natural\right\|\ge\lambda_r\left(\boldsymbol{G}^\natural\right)-\left\|\boldsymbol{G}^t-\boldsymbol{G}^0\right\|-\left\|\boldsymbol{G}^\natural-\boldsymbol{G}^0\right\|\\
&\overset{\text{(i)}}{\ge}\sigma_r^{\natural 2}-\frac{2\mu^\natural r}{n_1}\sigma_1^{\natural 2}-2C_0C^{\mathsf{dd}}\kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\zeta_{\mathsf{op}}-C^{\mathsf{dd}}\zeta_{\mathsf{op}}-\left\|\boldsymbol{M}^\natural\right\|_{2,\infty}^2\\
&\overset{\text{(ii)}}{\ge}\sigma_r^{\natural 2}-\frac{3\mu^\natural r}{n_1}\sigma_1^{\natural 2}-2C_0C^{\mathsf{dd}}\kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\zeta_{\mathsf{op}}-C^{\mathsf{dd}}\zeta_{\mathsf{op}}\overset{\text{(iii)}}{\ge}\frac{\sigma_r^{\natural 2}}{2}.
\end{aligned}\tag{F.31}$$

Here, (i) follows from (F.30) and Lemma 34, (ii) comes from (F.23), and (iii) is guaranteed to hold as long as $n_1\gg\kappa^{\natural 4}\mu^\natural r$ and $\zeta_{\mathsf{op}}\ll\sigma_r^{\natural 2}$. Then by virtue of Davis Kahan's $\sin\boldsymbol{\Theta}$ Theorem (Chen et al., 2020c, Theorem 2.2.1),

$$\left\|\boldsymbol{U}^t\boldsymbol{R}^t-\boldsymbol{U}^0\right\|\le\frac{\left\|\boldsymbol{G}^t-\boldsymbol{G}^0\right\|}{\lambda_r\left(\boldsymbol{G}^t\right)-\lambda_{r+1}\left(\boldsymbol{G}^\natural\right)}\le\frac{4\kappa^{\natural 2}\mu^\natural r}{n_1}+4C_0C^{\mathsf{dd}}\kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}},$$

where the last relation arises from $\lambda_{r+1}(\boldsymbol{G}^{\natural}) = 0$, (F.30) and (F.31). This indicates that

$$\begin{aligned}
\left\|\boldsymbol{U}^{t}\right\|_{2,\infty} &\leq \left\|\boldsymbol{U}^{0}\right\|_{2,\infty} + \left\|\boldsymbol{U}^{t}\boldsymbol{R}^{t} - \boldsymbol{U}^{0}\right\|_{2,\infty} \\
&\leq \sqrt{\frac{4\mu^{\natural}r}{n_1} + \frac{4\kappa^{\natural 2}\mu^{\natural}r}{n_1}} + 4C_0C^{\mathsf{dd}}\kappa^{\natural 2}\sqrt{\frac{\mu^{\natural}r}{n_1}}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}} \\
&\leq 3\sqrt{\frac{\mu^{\natural}r}{n_1}},
\end{aligned} \tag{F.32}$$

where the penultimate relation follows from (F.22), and the last relation holds provided that $n_1 \gg \kappa^{\natural 4}\mu^{\natural}r$ and $\zeta_{\mathsf{op}}/\sigma_r^{\natural 2} \ll 1/\kappa^{\natural 2}$.

To proceed, we recall that the diagonal entries of $\boldsymbol{G}^{t+1}$ are set to be the diagonal entries of $\boldsymbol{U}^{t}\boldsymbol{\Lambda}^{t}\boldsymbol{U}^{t\top}$, thus revealing that

$$\begin{aligned}
\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^{t+1} - \boldsymbol{G}^{\natural}\right)\right\| &= \left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{U}^{t}\boldsymbol{\Lambda}^{t}\boldsymbol{U}^{t\top} - \boldsymbol{G}^{\natural}\right)\right\| = \left\|\mathcal{P}_{\mathsf{diag}}\left(\mathcal{P}_{\boldsymbol{U}^{t}}\left(\boldsymbol{G}^{t}\right) - \boldsymbol{G}^{\natural}\right)\right\| \\
&\leq \underbrace{\left\|\mathcal{P}_{\mathsf{diag}}\left[\mathcal{P}_{\boldsymbol{U}^{t}}\left(\boldsymbol{G}^{t} - \boldsymbol{G}^{\natural}\right)\right]\right\|}_{=:\beta_1} + \underbrace{\left\|\mathcal{P}_{\mathsf{diag}}\left(\mathcal{P}_{\boldsymbol{U}_{\perp}^{t}}\boldsymbol{G}^{\natural}\right)\right\|}_{=:\beta_2}.
\end{aligned}$$

Here, $\boldsymbol{U}_{\perp}^{t}$ represents the orthogonal complement of the subspace $\boldsymbol{U}^{t}$. Similar to how we bound α_1 and α_2 for the base case, we can invoke Zhang et al. (2018, Lemma 1) and (F.32) to reach

$$\beta_1 \leq 3\sqrt{\frac{\mu^{\natural}r}{n_1}}\left\|\boldsymbol{G}^{t} - \boldsymbol{G}^{\natural}\right\|, \tag{F.33}$$

$$\beta_2 \leq \sqrt{\frac{\mu^{\natural}r}{n_1}}\left\|\mathcal{P}_{\boldsymbol{U}_{\perp}^{t}}\boldsymbol{G}^{\natural}\right\| \leq \sigma_1^{\natural 2}\left\|\left(\boldsymbol{U}_{\perp}^{t}\right)^{\top}\boldsymbol{U}^{\natural}\right\| = \sigma_1^{\natural 2}\left\|\sin\boldsymbol{\Theta}\left(\boldsymbol{U}^{t}, \boldsymbol{U}^{\natural}\right)\right\|, \tag{F.34}$$

where $\boldsymbol{\Theta}\left(\boldsymbol{U}^{t}, \boldsymbol{U}^{\natural}\right)$ is a diagonal matrix whose diagonal entries are the principal angles between $\boldsymbol{U}^{t}$ and $\boldsymbol{U}^{\natural}$. Apply the Davis Kahan $\sin\boldsymbol{\Theta}$ Theorem (Chen et al., 2020c, Theorem 2.2.1) to obtain

$$\left\|\sin\boldsymbol{\Theta}\left(\boldsymbol{U}^{t}, \boldsymbol{U}^{\natural}\right)\right\| \leq \frac{\left\|\boldsymbol{G}^{t} - \boldsymbol{G}^{\natural}\right\|}{\lambda_r\left(\boldsymbol{G}^{t}\right) - \lambda_{r+1}\left(\boldsymbol{G}^{\natural}\right)} \leq \frac{2\left\|\boldsymbol{G}^{t} - \boldsymbol{G}^{\natural}\right\|}{\sigma_r^{\natural 2}}. \tag{F.35}$$

Here, the last inequality comes from $\lambda_{r+1}(\boldsymbol{G}^{\natural}) = 0$ and (F.31). Combine (F.34) and (F.35) to reach

$$\beta_2 \leq \sqrt{\frac{\mu^{\natural}r}{n_1}}\sigma_1^{\natural 2}\frac{2\left\|\boldsymbol{G}^{t} - \boldsymbol{G}^{\natural}\right\|}{\sigma_r^{\natural 2}} \leq 2\kappa^{\natural 2}\sqrt{\frac{\mu^{\natural}r}{n_1}}\left\|\boldsymbol{G}^{t} - \boldsymbol{G}^{\natural}\right\|.$$

Taking together the preceding bounds on β_1 and β_2, we arrive at

$$\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^{t+1} - \boldsymbol{G}^{\natural}\right)\right\| \leq \beta_1 + \beta_2 \leq \left(3 + 2\kappa^{\natural 2}\right)\sqrt{\frac{\mu^{\natural}r}{n_1}}\left\|\boldsymbol{G}^{t} - \boldsymbol{G}^{\natural}\right\| \leq C_0\kappa^{\natural 2}\sqrt{\frac{\mu^{\natural}r}{n_1}}\left\|\boldsymbol{G}^{t} - \boldsymbol{G}^{\natural}\right\|,$$

provided that the constant C_0 is large enough.

Invoking the inequality (F.21) to establish the lemma. The above induction steps taken together establish the hypothesis (F.21), namely, for all $t \geq 1$,

$$\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^{t} - \boldsymbol{G}^{\natural}\right)\right\| \leq C_0\kappa^{\natural 2}\sqrt{\frac{\mu^{\natural}r}{n_1}}\left\|\boldsymbol{G}^{t-1} - \boldsymbol{G}^{\natural}\right\|.$$

Follow the same procedure used to derive (F.29), we can show that

$$\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^{t} - \boldsymbol{G}^{\natural}\right)\right\| \leq \left(C_0\kappa^{\natural 2}\sqrt{\frac{\mu^{\natural}r}{n_1}}\right)^{t}\frac{\mu^{\natural}r}{n_1}\sigma_1^{\natural 2} + 2C_0C^{\mathsf{dd}}\kappa^{\natural 2}\sqrt{\frac{\mu^{\natural}r}{n_1}}\zeta_{\mathsf{op}}.$$

If the number of iterations t_0 satisfies

$$t_0 \geq \frac{\log\left(\left(C^{\mathsf{dd}}\right)^{-1}(\kappa^\natural)^{-2}\sqrt{\mu^\natural r/n_1}\sigma_1^{\natural 2}/\zeta_{\mathsf{op}}\right)}{\log\left((\kappa^\natural)^{-2}\sqrt{n_1/\left(\mu^\natural r\right)}\right)},$$

we can guarantee that

$$\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^{t_0}-\boldsymbol{G}^\natural\right)\right\| \leq 3C_0C^{\mathsf{dd}}\kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\zeta_{\mathsf{op}}$$

as long as $n_1 \gg \kappa^{\natural 4}\mu^\natural r$. In addition, note that when $n_1 \gg \kappa^{\natural 4}\mu^\natural r$, one has

$$\frac{\log\left(\left(C^{\mathsf{dd}}\right)^{-1}(\kappa^\natural)^{-2}\sqrt{\mu^\natural r/n_1}\sigma_1^{\natural 2}/\zeta_{\mathsf{op}}\right)}{\log\left((\kappa^\natural)^{-2}\sqrt{n_1/\left(\mu^\natural r\right)}\right)} \leq \log\left(\frac{\sigma_1^{\natural 2}}{\zeta_{\mathsf{op}}}\right).$$

Therefore, it suffices to take $t_0 \geq \log\left(\frac{\sigma_1^{\natural 2}}{\zeta_{\mathsf{op}}}\right)$ as claimed.

To finish up, we observe that

$$\begin{aligned}
\left\|\boldsymbol{G}^{t_0}-\boldsymbol{G}^\natural\right\| &\leq \left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^{t_0}-\boldsymbol{G}^\natural\right)\right\| + \left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}^{t_0}-\boldsymbol{G}^\natural\right)\right\| \\
&\overset{\text{(i)}}{=} \left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^{t_0}-\boldsymbol{G}^\natural\right)\right\| + \left\|\boldsymbol{G}^{0}-\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}^\natural\right)\right\| \\
&\overset{\text{(ii)}}{\lesssim} \kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\zeta_{\mathsf{op}} + \zeta_{\mathsf{op}} \overset{\text{(iii)}}{\lesssim} \zeta_{\mathsf{op}}.
\end{aligned}$$

Here, (i) follows from the construction $\boldsymbol{G}^0 = \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}^0\right) = \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}^{t_0}\right)$; (ii) follows from Lemma 34; and (iii) holds provided that $n_1 \gtrsim \kappa^{\natural 4}\mu^\natural r$.

F.3.3 Proof of Lemma 24

We first apply Weyl's inequality to demonstrate that

$$\lambda_r\left(\boldsymbol{G}\right) \geq \sigma_r^{\natural 2} - \left\|\boldsymbol{G}-\boldsymbol{G}^\natural\right\| \geq \sigma_r^{\natural 2} - \widetilde{C}\zeta_{\mathsf{op}} \geq \frac{1}{2}\sigma_r^{\natural 2}$$

for some constant $\widetilde{C} > 0$, where the penultimate step comes from Lemma 23, and the last step holds true provided that $\zeta_{\mathsf{op}} \ll \sigma_r^{\natural 2}$. In view of the Davis-Kahan $\sin\boldsymbol{\Theta}$ Theorem (Chen et al., 2020c, Theorem 2.2.1), we obtain

$$\left\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\natural\boldsymbol{U}^{\natural\top}\right\| \leq \frac{\sqrt{2}\left\|\boldsymbol{G}-\boldsymbol{G}^\natural\right\|}{\lambda_r\left(\boldsymbol{G}\right)-\lambda_{r+1}\left(\boldsymbol{G}^\natural\right)} \leq \frac{2\sqrt{2}\left\|\boldsymbol{G}-\boldsymbol{G}^\natural\right\|}{\sigma_r^{\natural 2}} \lesssim \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}, \tag{F.36}$$

which immediately leads to the advertised bound on $\left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^\natural\right\|$ as follows

$$\left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^\natural\right\| = \left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\natural - \boldsymbol{U}^\natural\boldsymbol{U}^{\natural\top}\boldsymbol{U}^\natural\right\| \leq \left\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\natural\boldsymbol{U}^{\natural\top}\right\| \lesssim \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}.$$

Next, we turn attention to $\left\|\boldsymbol{H}-\boldsymbol{R}_{\boldsymbol{U}}\right\|$. Given that both $\boldsymbol{U}$ and $\boldsymbol{U}^\natural$ have orthonormal columns, the SVD of $\boldsymbol{H} = \boldsymbol{U}^\top\boldsymbol{U}^\natural$ can be written as

$$\boldsymbol{H} = \boldsymbol{X}(\cos\boldsymbol{\Theta})\boldsymbol{Y}^\top,$$

where $\boldsymbol{X},\boldsymbol{Y} \in \mathbb{R}^{r\times r}$ are orthonormal matrices and $\boldsymbol{\Theta}$ is a diagonal matrix composed of the principal angles between $\boldsymbol{U}$ and $\boldsymbol{U}^\natural$ (see Chen et al. (2020c, Section 2.1)). It is well known that one can write $\boldsymbol{R}_{\boldsymbol{U}} = \mathsf{sgn}(\boldsymbol{H}) = \boldsymbol{X}\boldsymbol{Y}^\top$, and therefore,

$$\left\|\boldsymbol{H}-\boldsymbol{R}_{\boldsymbol{U}}\right\| = \left\|\boldsymbol{X}\left(\cos\boldsymbol{\Theta}-\boldsymbol{I}_r\right)\boldsymbol{Y}^\top\right\| = \left\|\boldsymbol{I}_r - \cos\boldsymbol{\Theta}\right\|$$

$$= \left\|2\sin^2\left(\boldsymbol{\Theta}/2\right)\right\| \lesssim \|\sin\boldsymbol{\Theta}\|^2 \asymp \left\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\natural\boldsymbol{U}^{\natural\top}\right\|^2 \lesssim \frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}}.$$

Here the penultimate relation comes from Chen et al. (2020c, Lemma 2.1.2) and the last relation invokes (F.36). Given that $\boldsymbol{R}_{\boldsymbol{U}}$ is a square orthonormal matrix, we immediately have

$$\sigma_{\max}(\boldsymbol{H}_{\boldsymbol{U}}) \le \sigma_{\max}(\boldsymbol{R}_{\boldsymbol{U}}) + \|\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{R}_{\boldsymbol{U}}\| \le 1 + O\Big(\frac{\sigma^2 n}{\sigma_r^{\natural 2}}\Big) \le 2,$$
$$\sigma_r(\boldsymbol{H}_{\boldsymbol{U}}) \ge \sigma_{\max}(\boldsymbol{R}_{\boldsymbol{U}}) - \|\boldsymbol{H}_{\boldsymbol{U}} - \boldsymbol{R}_{\boldsymbol{U}}\| \ge 1 - O\Big(\frac{\sigma^2 n}{\sigma_r^{\natural 2}}\Big) \ge 1/2,$$

provided that $\zeta_{\mathsf{op}} \ll \sigma_r^{\natural 2}$. In addition, we can similarly derive

$$\left\|\boldsymbol{H}^\top\boldsymbol{H} - \boldsymbol{I}_r\right\| = \left\|\boldsymbol{Y}\left[\cos^2\boldsymbol{\Theta} - \boldsymbol{I}_r\right]\boldsymbol{Y}^\top\right\| = \left\|\cos^2\boldsymbol{\Theta} - \boldsymbol{I}_r\right\| = \left\|\sin^2\boldsymbol{\Theta}\right\| = \|\sin\boldsymbol{\Theta}\|^2 \lesssim \frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}}.$$

An immediate consequence of the proximity between $\boldsymbol{H}$ and $\boldsymbol{R}_{\boldsymbol{U}}$ is that

$$\begin{aligned}\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right\| &\le \left\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^\natural\right\| + \|\boldsymbol{U}(\boldsymbol{H} - \boldsymbol{R}_{\boldsymbol{U}})\| \le \left\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^\natural\right\| + \|\boldsymbol{H} - \boldsymbol{R}_{\boldsymbol{U}}\| \\ &\lesssim \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}} + \frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}} \lesssim \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}},\end{aligned}$$

provided that $\zeta_{\mathsf{op}} \ll \sigma_r^{\natural 2}$.

F.3.4 Proof of Lemma 25

We begin with the triangle inequality:

$$\begin{aligned}\left\|\left(\boldsymbol{G} - \boldsymbol{G}^\natural\right)_{m,\cdot}\boldsymbol{U}^\natural\right\|_2 &\le \left\|\left[\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G} - \boldsymbol{G}^\natural\right)\right]_{m,\cdot}\boldsymbol{U}^\natural\right\|_2 + \left\|\left[\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G} - \boldsymbol{G}^\natural\right)\right]_{m,\cdot}\boldsymbol{U}^\natural\right\|_2 \\ &= \underbrace{\left\|\left[\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}^0 - \boldsymbol{G}^\natural\right)\right]_{m,\cdot}\boldsymbol{U}^\natural\right\|_2}_{=:\,\alpha_1} + \underbrace{\left\|\left[\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G} - \boldsymbol{G}^\natural\right)\right]_{m,\cdot}\boldsymbol{U}^\natural\right\|_2}_{=:\,\alpha_2},\end{aligned}$$

where the last inequality makes use of our construction $\mathcal{P}_{\mathsf{off\text{-}diag}}(\boldsymbol{G}) = \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}^0\right)$. Given that the properties of $\boldsymbol{G}^0$ (which is the diagonal-deleted version of the sample gram matrix) have been studied in Cai et al. (2021), we can readily borrow Cai et al. (2021, Lemma 2) to bound

$$\alpha_1 \lesssim \zeta_{\mathsf{op}}\sqrt{\frac{\mu^\natural r}{n_1}}$$

with probability exceeding $1 - O(d^{-10})$. The careful reader might remark that the above bound is slightly different from Cai et al. (2021, Lemma 2) in the sense that the bound therein contains an additional term $\|\boldsymbol{M}^\natural\|_{2,\infty}^2\sqrt{\mu^\natural r/n_1}$; note, however, that this extra term is caused by the effect of the diagonal part $\mathcal{P}_{\mathsf{diag}}(\boldsymbol{G}^\natural)$ in their analysis, which has been removed in the above term α_1. The interested reader is referred to Cai et al. (2021, Appendix B.3) for details. In addition, it is seen that

$$\alpha_2 = \left|G_{m,m} - G_{m,m}^\natural\right|\left\|\boldsymbol{U}_{m,\cdot}^\natural\right\|_2 \le \left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G} - \boldsymbol{G}^\natural\right)\right\|\left\|\boldsymbol{U}^\natural\right\|_{2,\infty} \lesssim \kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\zeta_{\mathsf{op}}\sqrt{\frac{\mu^\natural r}{n_1}} \lesssim \zeta_{\mathsf{op}}\sqrt{\frac{\mu^\natural r}{n_1}},$$

where the penultimate relation invokes Lemma 23, and the last inequality holds true provided that $n_1 \gtrsim \kappa^{\natural 4}\mu^\natural r$. We can thus conclude that

$$\left\|\left(\boldsymbol{G} - \boldsymbol{G}^\natural\right)_{m,\cdot}\boldsymbol{U}^\natural\right\|_2 \le \alpha_1 + \alpha_2 \lesssim \zeta_{\mathsf{op}}\sqrt{\frac{\mu^\natural r}{n_1}}.$$

F.3.5 Proof of Lemma 26

Recall that for each $1 \le m \le n_1$, we employ the notation $\mathcal{P}_{-m,\cdot}(\boldsymbol{M})$ to represent an $n_1 \times n_2$ matrix such that

$$\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{M})\right]_{i,\cdot} = \begin{cases} \boldsymbol{0}, & \text{if } i = m, \\ \boldsymbol{M}_{i,\cdot}, & \text{if } i \ne m. \end{cases}$$

In other words, it is obtained by zeroing out the m-th row of $\boldsymbol{M}$. Simple algebra then reveals that the m-th row of $\boldsymbol{G}$ can be decomposed into

$$\begin{aligned} \boldsymbol{G}_{m,\cdot} &= \boldsymbol{M}_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{M})\right]^\top + G_{m,m}\boldsymbol{e}_m^\top \\ &= \boldsymbol{M}^\natural_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}\left(\boldsymbol{M}^\natural\right)\right]^\top + \boldsymbol{M}^\natural_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{E})\right]^\top + \boldsymbol{E}_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{M})\right]^\top + G_{m,m}\boldsymbol{e}_m^\top \\ &= \boldsymbol{M}^\natural_{m,\cdot}\boldsymbol{M}^{\natural\top} + \boldsymbol{M}^\natural_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{E})\right]^\top + \boldsymbol{E}_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{M})\right]^\top + \left(G_{m,m} - G^\natural_{m,m}\right)\boldsymbol{e}_m^\top, \end{aligned}$$

where we have used $\boldsymbol{G}^\natural = \boldsymbol{M}^\natural\boldsymbol{M}^{\natural\top}$. Apply the triangle inequality once again to yield

$$\begin{aligned} \left\|\boldsymbol{G}_{m,\cdot}\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right)\right\|_2 \le{}& \underbrace{\left\|\boldsymbol{M}^\natural_{m,\cdot}\boldsymbol{M}^{\natural\top}\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right)\right\|_2}_{=:\alpha_1} + \underbrace{\left\|\boldsymbol{M}^\natural_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{E})\right]^\top\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right)\right\|_2}_{=:\alpha_2} \\ &+ \underbrace{\left\|\boldsymbol{E}_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{M})\right]^\top\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right)\right\|_2}_{=:\alpha_3} + \underbrace{\left\|\left(G_{m,m} - G^\natural_{m,m}\right)\boldsymbol{e}_m^\top\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right)\right\|_2}_{=:\alpha_4} \end{aligned}$$

for each $1 \le m \le n_1$. In what follows, we shall bound the terms $\alpha_1, \cdots, \alpha_4$ separately.

- Let us begin with α_1. Write

$$\left\|\boldsymbol{U}^{\natural\top}\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right)\right\| = \left\|\boldsymbol{U}^{\natural\top}\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\natural - \boldsymbol{I}_r\right\| = \left\|\boldsymbol{H}^\top\boldsymbol{H} - \boldsymbol{I}_r\right\| \lesssim \frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}}, \tag{F.37}$$

  where the last relation follows from Lemma 24. This allows us to upper bound α_1 as follows:

$$\begin{aligned} \alpha_1 &= \left\|\boldsymbol{e}_m^\top\boldsymbol{M}^\natural\boldsymbol{M}^{\natural\top}\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right)\right\|_2 \le \left\|\boldsymbol{e}_m^\top\boldsymbol{U}^\natural\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{U}^{\natural\top}\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right)\right\|_2 \\ &\le \left\|\boldsymbol{U}^\natural\right\|_{2,\infty}\left\|\boldsymbol{\Sigma}^\natural\right\|^2\left\|\boldsymbol{U}^{\natural\top}\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right)\right\| \\ &\lesssim \sigma_1^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}}\left\|\boldsymbol{U}^\natural\right\|_{2,\infty} \lesssim \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^\natural r}{n_1}}. \end{aligned}$$

- Regarding α_2, it is readily seen that

$$\alpha_2 \le \left\|\boldsymbol{M}^\natural_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{E})\right]^\top\right\|_2\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right\|.$$

  Regarding the first term $\left\|\boldsymbol{M}^\natural_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{E})\right]^\top\right\|_2$, we notice that

$$\begin{aligned} \left\|\boldsymbol{M}^\natural_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{E})\right]^\top\right\|_2^2 &\le \left\|\boldsymbol{M}^\natural_{m,\cdot}\boldsymbol{E}^\top\right\|_2^2 = \sum_{i=1}^{n_1}\left(\sum_{j=1}^{n_2} M^\natural_{m,j}E_{i,j}\right)^2 \\ &\lesssim \left\|\boldsymbol{M}^\natural_{m,\cdot}\right\|_2^2\left(\sigma^2 n_1 + \sigma^2\log^2 n\right) + \left\|\boldsymbol{M}^\natural_{m,\cdot}\right\|_\infty^2 B^2\log^3 n \\ &\lesssim \frac{\mu^\natural r}{n_1}\sigma_1^{\natural 2}\left(\sigma^2 n_1 + \sigma^2\log^2 n\right) + \frac{\mu^\natural r}{n_1 n_2}\sigma_1^{\natural 2}B^2\log^3 n \\ &\lesssim \mu^\natural r\sigma_1^{\natural 2}\sigma^2, \end{aligned}$$

where the second line follows from Cai et al. (2021, Lemma 14), the third line comes from (F.23) and the following bound

$$\left\|\boldsymbol{M}^{\natural}\right\|_{\infty}=\left\|\boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural}\boldsymbol{V}^{\natural\top}\right\|_{\infty}\leq\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}\left\|\boldsymbol{\Sigma}^{\natural}\right\|\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty}\leq\frac{\mu^{\natural}r}{\sqrt{n_1n_2}}\sigma_1^{\natural},\tag{F.38}$$

and the last line holds provided that $n_1\gtrsim\log^2 n$, $n_2\gtrsim\log^2 n$ and $B\lesssim\sigma\sqrt[4]{n_1n_2}/\sqrt{\log n}$. Therefore, we can demonstrate that

$$\begin{aligned}\alpha_2&\lesssim\left\|\boldsymbol{M}_{m,\cdot}^{\natural}\left[\mathcal{P}_{-m,\cdot}\left(\boldsymbol{E}\right)\right]^{\top}\right\|_2\left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\natural}\right\|\overset{\text{(i)}}{\lesssim}\sqrt{\mu^{\natural}r}\sigma_1^{\natural}\sigma\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\\&\asymp\kappa^{\natural}\sqrt{\mu^{\natural}r}\frac{\sigma\zeta_{\mathsf{op}}}{\sigma_r^{\natural}}\overset{\text{(ii)}}{\lesssim}\sqrt{\frac{\mu^{\natural}r}{n_1}}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}.\end{aligned}$$

Here, (i) utilizes Lemma 24 and (ii) comes from the definition of $\zeta_{\mathsf{op}}\geq\sigma\sigma_1^{\natural}\sqrt{n_1\log n}$.

- When it comes to α_3, our starting point is

$$\alpha_3\leq\underbrace{\left\|\boldsymbol{E}_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}\left(\boldsymbol{M}\right)\right]^{\top}\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^{\natural}\right)\right\|_2}_{=:\,\beta_1}+\underbrace{\left\|\boldsymbol{E}_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}\left(\boldsymbol{M}\right)\right]^{\top}\left(\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right)\right\|_2}_{=:\,\beta_2}.$$

  - The first term β_1 can be controlled using Lemma 32 as follows

$$\beta_1\leq\zeta_{\mathsf{op}}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\left(\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}+\sqrt{\frac{\mu^{\natural}r}{n_1}}\right).\tag{F.39}$$

  Additionally, it follows from Lemma 33 that

$$\begin{aligned}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}\boldsymbol{H}\right\|\leq\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}\boldsymbol{U}^{\top}\right\|&\leq\kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\left(\left\|\boldsymbol{U}\boldsymbol{H}\right\|_{2,\infty}+\sqrt{\frac{\mu^{\natural}r}{n_1}}\right)\\&\leq\kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\left(\left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\sqrt{\frac{\mu^{\natural}r}{n_1}}\right)\\&\asymp\kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\left(\left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\sqrt{\frac{\mu^{\natural}r}{n_1}}\right),\end{aligned}$$

  which together with the triangle inequality gives

$$\begin{aligned}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}&\leq\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}\boldsymbol{H}\right\|+\left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}\\&\lesssim\kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\left(\left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\sqrt{\frac{\mu^{\natural}r}{n_1}}\right)+\left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}\\&\asymp\kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^{\natural}r}{n_1}}+\left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}\end{aligned}\tag{F.40}$$

  as long as $\zeta_{\mathsf{op}}\ll\sigma_r^{\natural 2}/\kappa^{\natural 2}$. An immediate consequence is that

$$\begin{aligned}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}&\leq\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}\\&\lesssim\sqrt{\frac{\mu^{\natural}r}{n_1}}+\kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^{\natural}r}{n_1}}+\left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}\end{aligned}$$

$$\asymp \sqrt{\frac{\mu^\natural r}{n_1}} + \left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right\|_{2,\infty}, \tag{F.41}$$

with the proviso that $\zeta_{\mathsf{op}} \ll \sigma_r^{\natural 2}/\kappa^{\natural 2}$. Therefore we can invoke (F.40) and (F.41) to refine (F.39) as

$$\begin{aligned}\beta_1 &\lesssim \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^\natural r}{n_1}} + \zeta_{\mathsf{op}}\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right\|_{2,\infty} + \frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\left(\sqrt{\frac{\mu^\natural r}{n_1}} + \left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right\|_{2,\infty}\right) \\ &\asymp \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^\natural r}{n_1}} + \zeta_{\mathsf{op}}\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right\|_{2,\infty},\end{aligned}$$

provided that $\zeta_{\mathsf{op}} \ll \sigma_r^{\natural 2}$.

– With regards to the second term β_2, we see that

$$\begin{aligned}\beta_2 &= \left\|\boldsymbol{E}_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{M})\right]^\top\left(\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}\right)\boldsymbol{U}^\natural\right\|_2 \\ &\leq \left\|\boldsymbol{E}_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{M})\right]^\top\right\|_2\left\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}\right\|.\end{aligned}$$

It has already been proved in Cai et al. (2021, Appendix C.2) that with probability exceeding $1 - O(n^{-10})$,

$$\begin{aligned}\left\|\boldsymbol{E}_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{M})\right]^\top\right\|_2 &\lesssim \sigma^2\sqrt{n_1 n_2\log n} + \sigma\sqrt{n_2\log n}\left\|\boldsymbol{M}^{\natural\top}\right\|_{2,\infty} \\ &\lesssim \sigma^2\sqrt{n_1 n_2\log n} + \sigma\sqrt{\mu^\natural r\log n}\,\sigma_1^\natural \\ &\lesssim \zeta_{\mathsf{op}},\end{aligned}$$

where the second relation holds since

$$\left\|\boldsymbol{M}^{\natural\top}\right\|_{2,\infty} = \left\|\boldsymbol{V}^\natural\boldsymbol{\Sigma}^\natural\boldsymbol{U}^{\natural\top}\right\|_{2,\infty} \leq \left\|\boldsymbol{V}^\natural\right\|_{2,\infty}\left\|\boldsymbol{\Sigma}^\natural\right\|\left\|\boldsymbol{U}^\natural\right\| \leq \sigma_1^\natural\sqrt{\frac{\mu^\natural r}{n_2}}, \tag{F.42}$$

and the last relation holds as long as $n_1 \gtrsim \mu^\natural r$. This together with Lemma 33 provides an upper bound on β_2 as follows:

$$\begin{aligned}\beta_2 &\leq \left\|\boldsymbol{E}_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{M})\right]^\top\right\|_2\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\| \lesssim \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\left(\left\|\boldsymbol{U}\boldsymbol{H}\right\|_{2,\infty} + \sqrt{\frac{\mu^\natural r}{n_1}}\right) \\ &\lesssim \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\left(\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right\|_{2,\infty} + \sqrt{\frac{\mu^\natural r}{n_1}}\right).\end{aligned}$$

– Combine the preceding bounds on β_1 and β_2 to arrive at

$$\alpha_3 \leq \beta_1 + \beta_2 \lesssim \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^\natural r}{n_1}} + \zeta_{\mathsf{op}}\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right\|_{2,\infty},$$

provided that $\zeta_{\mathsf{op}}/\sigma_r^{\natural 2} \ll 1/\kappa^{\natural 2}$.

- For α_4, Lemma 23 tells us that

$$\alpha_4 \leq \left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G} - \boldsymbol{G}^\natural\right)\right\|\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right\|_{2,\infty} \lesssim \kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\zeta_{\mathsf{op}}\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^\natural\right\|_{2,\infty}.$$

Thus far, we have developed upper bounds on $\alpha_1, \cdots, \alpha_4$, which taken collectively lead to

$$\begin{aligned}
\left\|\boldsymbol{G}_{m,\cdot}\left(\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\natural}\right)\right\|_2 &\leq \alpha_1+\alpha_2+\alpha_3+\alpha_4 \\
&\lesssim \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^{\natural} r}{n_1}}+\zeta_{\mathsf{op}}\left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\kappa^{\natural 2}\sqrt{\frac{\mu^{\natural} r}{n_1}}\zeta_{\mathsf{op}}\left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\natural}\right\|_{2,\infty} \\
&\asymp \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^{\natural} r}{n_1}}+\zeta_{\mathsf{op}}\left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\natural}\right\|_{2,\infty},
\end{aligned}$$

provided that $n_1 \gtrsim \kappa^{\natural 4}\mu^{\natural} r$.

F.3.6 Proof of Lemma 27

To control the target quantity, we make note of the following decomposition

$$\left\|\boldsymbol{R}^{\top}\boldsymbol{\Sigma}^2\boldsymbol{R}-\boldsymbol{\Sigma}^{\natural 2}\right\| \leq \underbrace{\left\|\boldsymbol{R}^{\top}\boldsymbol{\Sigma}^2\boldsymbol{R}-\boldsymbol{H}^{\top}\boldsymbol{\Sigma}^2\boldsymbol{H}\right\|}_{=:\,\alpha_1}+\underbrace{\left\|\boldsymbol{H}^{\top}\boldsymbol{\Sigma}^2\boldsymbol{H}-\boldsymbol{U}^{\natural\top}\boldsymbol{G}\boldsymbol{U}^{\natural}\right\|}_{=:\,\alpha_2}+\underbrace{\left\|\boldsymbol{U}^{\natural\top}\boldsymbol{G}\boldsymbol{U}^{\natural}-\boldsymbol{\Sigma}^{\natural 2}\right\|}_{=:\,\alpha_3}.$$

In the sequel, we shall upper bound each of these terms separately.

Step 1: bounding α_1. Lemma 24 tells us that

$$\begin{aligned}
\alpha_1 &\leq \left\|(\boldsymbol{H}-\boldsymbol{R})^{\top}\boldsymbol{\Sigma}^2\boldsymbol{H}\right\|+\left\|\boldsymbol{R}^{\top}\boldsymbol{\Sigma}^2(\boldsymbol{H}-\boldsymbol{R})\right\| \\
&\leq \|(\boldsymbol{H}-\boldsymbol{R})\|\,\|\boldsymbol{\Lambda}\|\,(\|\boldsymbol{H}\|+\|\boldsymbol{R}\|) \\
&\lesssim \frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}}\sigma_1^{\natural 2} \asymp \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}.
\end{aligned}$$

Here, the penultimate inequality results from an application of Weyl's inequality:

$$\|\boldsymbol{\Lambda}\| \leq \left\|\boldsymbol{\Lambda}^{\natural}\right\|+\left\|\boldsymbol{G}-\boldsymbol{G}^{\natural}\right\| \lesssim \sigma_1^{\natural 2}+\zeta_{\mathsf{op}} \asymp \sigma_1^{\natural 2},$$

where we have used Lemma 23 and the assumption $\zeta_{\mathsf{op}} \lesssim \sigma_1^{\natural 2}$.

Step 2: bounding α_2. Regarding α_2, it is easily seen that

$$\begin{aligned}
\boldsymbol{H}^{\top}\boldsymbol{\Sigma}^2\boldsymbol{H}-\boldsymbol{U}^{\natural\top}\boldsymbol{G}\boldsymbol{U}^{\natural} &= \boldsymbol{U}^{\natural\top}\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^{\top}\boldsymbol{U}^{\natural}-\boldsymbol{U}^{\natural\top}\boldsymbol{G}\boldsymbol{U}^{\natural} = \boldsymbol{U}^{\natural\top}\left(\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^{\top}-\boldsymbol{G}\right)\boldsymbol{U}^{\natural} \\
&= -\boldsymbol{U}^{\natural\top}\boldsymbol{U}_{\perp}\boldsymbol{\Lambda}_{\perp}\boldsymbol{U}_{\perp}^{\top}\boldsymbol{U}^{\natural},
\end{aligned}$$

where we denote the full SVD of $\boldsymbol{G}$ as

$$\boldsymbol{G} = \left[\begin{array}{cc}\boldsymbol{U} & \boldsymbol{U}_{\perp}\end{array}\right]\left[\begin{array}{cc}\boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_{\perp}\end{array}\right]\left[\begin{array}{c}\boldsymbol{V}^{\top} \\ \boldsymbol{V}_{\perp}^{\top}\end{array}\right] = \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{V}^{\top}+\boldsymbol{U}_{\perp}\boldsymbol{\Sigma}_{\perp}^2\boldsymbol{V}_{\perp}^{\top}.$$

From Chen et al. (2020c, Lemma 2.1.2) and (F.36), we have learned that

$$\left\|\boldsymbol{U}^{\natural\top}\boldsymbol{U}_{\perp}\right\| = \left\|\boldsymbol{U}\boldsymbol{U}^{\top}-\boldsymbol{U}^{\natural}\boldsymbol{U}^{\natural\top}\right\| \lesssim \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}.$$

In view of Weyl's inequality and Lemma 23, we obtain

$$\|\boldsymbol{\Lambda}_{\perp}\| \leq \lambda_{r+1}\left(\boldsymbol{G}^{\natural}\right)+\left\|\boldsymbol{G}-\boldsymbol{G}^{\natural}\right\| \lesssim \zeta_{\mathsf{op}},$$

thus indicating that

$$\alpha_2 = \left\|\boldsymbol{U}^{\natural\top}\boldsymbol{U}_{\perp}\boldsymbol{\Lambda}_{\perp}\boldsymbol{U}_{\perp}^{\top}\boldsymbol{U}^{\natural}\right\| \leq \left\|\boldsymbol{U}^{\natural\top}\boldsymbol{U}_{\perp}\right\|^2\|\boldsymbol{\Lambda}_{\perp}\| \lesssim \frac{\zeta_{\mathsf{op}}^3}{\sigma_r^{\natural 4}}.$$

Step 3: bounding α_3. Let us decompose

$$\begin{aligned}
\boldsymbol{G} &= \mathcal{P}_{\mathsf{off\text{-}diag}}\left[\left(\boldsymbol{M}^{\natural}+\boldsymbol{E}\right)\left(\boldsymbol{M}^{\natural}+\boldsymbol{E}\right)^{\top}\right]+\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}^{\natural}\right)+\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}-\boldsymbol{G}^{\natural}\right)\\
&= \boldsymbol{G}^{\natural}+\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{M}^{\natural\top}+\boldsymbol{M}^{\natural}\boldsymbol{E}^{\top}+\boldsymbol{E}\boldsymbol{E}^{\top}\right)+\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}-\boldsymbol{G}^{\natural}\right)\\
&= \boldsymbol{G}^{\natural}+\boldsymbol{E}\boldsymbol{M}^{\natural\top}+\boldsymbol{M}^{\natural}\boldsymbol{E}^{\top}+\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^{\top}\right)+\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}-\boldsymbol{G}^{\natural}\right)-\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{E}\boldsymbol{M}^{\natural\top}+\boldsymbol{M}^{\natural}\boldsymbol{E}^{\top}\right),
\end{aligned}$$

which in turn implies that

$$\begin{aligned}
\boldsymbol{U}^{\natural\top}\boldsymbol{G}\boldsymbol{U}^{\natural}-\boldsymbol{\Sigma}^{\natural2} &= \underbrace{\boldsymbol{U}^{\natural\top}\left(\boldsymbol{E}\boldsymbol{M}^{\natural\top}+\boldsymbol{M}^{\natural}\boldsymbol{E}^{\top}\right)\boldsymbol{U}^{\natural}}_{=:\boldsymbol{J}_1}+\underbrace{\boldsymbol{U}^{\natural\top}\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^{\top}\right)\boldsymbol{U}^{\natural}}_{=:\boldsymbol{J}_2}\\
&\quad+\underbrace{\boldsymbol{U}^{\natural\top}\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}-\boldsymbol{G}^{\natural}\right)\boldsymbol{U}^{\natural}}_{=:\boldsymbol{J}_3}-\underbrace{\boldsymbol{U}^{\natural\top}\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{E}\boldsymbol{M}^{\natural\top}+\boldsymbol{M}^{\natural}\boldsymbol{E}^{\top}\right)\boldsymbol{U}^{\natural}}_{=:\boldsymbol{J}_4}.
\end{aligned}$$

We shall then bound the spectral norm of $\boldsymbol{J}_1$, $\boldsymbol{J}_2$, $\boldsymbol{J}_3$ and $\boldsymbol{J}_4$ separately.

Step 3.1: bounding $\|\boldsymbol{J}_1\|$. Note that

$$\left\|\boldsymbol{U}^{\natural\top}\boldsymbol{E}\boldsymbol{M}^{\natural\top}\boldsymbol{U}^{\natural}\right\|=\left\|\boldsymbol{U}^{\natural\top}\boldsymbol{E}\boldsymbol{V}^{\natural}\boldsymbol{\Sigma}^{\natural}\right\|\leq\sigma_1^{\natural}\left\|\boldsymbol{U}^{\natural\top}\boldsymbol{E}\boldsymbol{V}^{\natural}\right\|.$$

Then we can utilize (C.11) to show that with probability exceeding $1-O(d^{-10})$

$$\left\|\boldsymbol{U}^{\natural\top}\boldsymbol{E}\boldsymbol{V}^{\natural}\right\|\overset{\text{(i)}}{\lesssim}\sigma\sqrt{r\log n}+\frac{B\mu^{\natural}r\log n}{\sqrt{n_1n_2}}\overset{\text{(i)}}{\lesssim}\sigma\sqrt{r\log n}+\frac{\sigma\mu^{\natural}r\sqrt{\log n}}{\sqrt[4]{n_1n_2}}$$

Note that (i) uses an intermediate result in (C.11) since here the condition on B (Assumption 5) is different from what we have assumed for the matrix denoising model (Assumption 3), whereas (ii) holds as long as $B\lesssim\sigma\sqrt[4]{n_1n_2}/\sqrt{\log n}$. Therefore, we obtain

$$\begin{aligned}
\|\boldsymbol{J}_1\| &\leq 2\left\|\boldsymbol{U}^{\natural\top}\boldsymbol{E}\boldsymbol{M}^{\natural\top}\boldsymbol{U}^{\natural}\right\|=2\left\|\boldsymbol{U}^{\natural\top}\boldsymbol{E}\boldsymbol{V}^{\natural\top}\boldsymbol{\Sigma}^{\natural}\right\|\leq2\sigma_1^{\natural}\left\|\boldsymbol{U}^{\natural\top}\boldsymbol{E}\boldsymbol{V}^{\natural}\right\|\\
&\lesssim\sigma\sigma_1^{\natural}\sqrt{r\log n}+\frac{\sigma\sigma_1^{\natural}\mu^{\natural}r\sqrt{\log n}}{\sqrt[4]{n_1n_2}}\lesssim\sqrt{\frac{\mu^{\natural}r}{n_1}}\zeta_{\mathsf{op}}
\end{aligned}$$

as long as $n_1n_2\gtrsim\mu^{\natural2}r^2$.

Step 3.2: bounding $\|\boldsymbol{J}_2\|$. This matrix $\boldsymbol{J}_2$ can be expressed as a sum of independent and zero-mean random matrices:

$$\boldsymbol{J}_2=\sum_{l=1}^{n_2}\underbrace{\boldsymbol{U}^{\natural\top}\left(\boldsymbol{E}_{\cdot,l}\boldsymbol{E}_{\cdot,l}^{\top}-\boldsymbol{D}_l\right)\boldsymbol{U}^{\natural}}_{=:\boldsymbol{X}_l};$$

here and below, we denote

$$\boldsymbol{D}_l\coloneqq\mathsf{diag}\left\{E_{1,l}^2,\ldots,E_{n_1,l}^2\right\}. \tag{F.43}$$

We will invoke the truncated matrix Bernstein inequality (Chen et al., 2020c, Theorem 3.1.1) to bound its spectral norm. To do so, we need to calculate the following quantities:

- We first study

$$\begin{aligned}
v\coloneqq\left\|\sum_{l=1}^{n_2}\mathbb{E}\left[\boldsymbol{X}_l\boldsymbol{X}_l^{\top}\right]\right\| &\leq\sup_{\|\boldsymbol{v}\|_2=1}\sum_{l=1}^{n_2}\mathbb{E}\left[\boldsymbol{v}^{\top}\boldsymbol{U}^{\natural\top}\left(\boldsymbol{E}_{\cdot,l}\boldsymbol{E}_{\cdot,l}^{\top}-\boldsymbol{D}_l\right)\boldsymbol{U}^{\natural}\boldsymbol{U}^{\natural\top}\left(\boldsymbol{E}_{\cdot,l}\boldsymbol{E}_{\cdot,l}^{\top}-\boldsymbol{D}_l\right)\boldsymbol{U}^{\natural}\boldsymbol{v}\right]\\
&=\sup_{\|\boldsymbol{v}\|_2=1}\sum_{l=1}^{n_2}\sum_{k=1}^{r}\mathbb{E}\left[\boldsymbol{v}^{\top}\boldsymbol{U}^{\natural\top}\left(\boldsymbol{E}_{\cdot,l}\boldsymbol{E}_{\cdot,l}^{\top}-\boldsymbol{D}_l\right)\boldsymbol{U}_{\cdot,k}^{\natural}\boldsymbol{U}_{\cdot,k}^{\natural\top}\left(\boldsymbol{E}_{\cdot,l}\boldsymbol{E}_{\cdot,l}^{\top}-\boldsymbol{D}_l\right)\boldsymbol{U}^{\natural}\boldsymbol{v}\right]
\end{aligned}$$

$$= \sup_{\|\boldsymbol{v}\|_2=1} \sum_{l=1}^{n_2}\sum_{k=1}^{r} \mathbb{E}\left[\left(\boldsymbol{v}^\top \boldsymbol{U}^{\natural\top}\left(\boldsymbol{E}_{\cdot,l}\boldsymbol{E}_{\cdot,l}^\top - \boldsymbol{D}_l\right)\boldsymbol{U}^{\natural}_{\cdot,k}\right)^2\right].$$

For any $\boldsymbol{v} \in \mathbb{R}^r$ with unit norm, let $\boldsymbol{x} = \boldsymbol{U}^{\natural}\boldsymbol{v}$ and derive

$$\begin{aligned}
\sum_{l=1}^{n_2}\sum_{k=1}^{r} \mathbb{E}\left[\left(\boldsymbol{v}^\top \boldsymbol{U}^{\natural\top}\left(\boldsymbol{E}_{\cdot,l}\boldsymbol{E}_{\cdot,l}^\top - \boldsymbol{D}_l\right)\boldsymbol{U}^{\natural}_{\cdot,k}\right)^2\right] &= \sum_{l=1}^{n_2}\sum_{k=1}^{r} \mathbb{E}\left[\left(\boldsymbol{x}^\top \left(\boldsymbol{E}_{\cdot,l}\boldsymbol{E}_{\cdot,l}^\top - \boldsymbol{D}_l\right)\boldsymbol{U}^{\natural}_{\cdot,k}\right)^2\right] \\
&= \sum_{l=1}^{n_2}\sum_{k=1}^{r} \mathsf{var}\left(\boldsymbol{x}^\top \boldsymbol{E}_{\cdot,l}\boldsymbol{E}_{\cdot,l}^\top \boldsymbol{U}^{\natural}_{\cdot,k}\right) = \sum_{l=1}^{n_2}\sum_{k=1}^{r} \mathsf{var}\left(\sum_{i=1}^{n_1}\sum_{j=1}^{n_1} x_i U^{\natural}_{j,k} E_{i,l}E_{j,l}\right) \\
&= \sum_{l=1}^{n_2}\sum_{k=1}^{r}\sum_{i=1}^{n_1}\sum_{j=1}^{n_1} x_i^2 U^{\natural 2}_{j,k}\sigma_{i,l}^2\sigma_{j,l}^2 \le \sigma^4 n_2 \|\boldsymbol{x}\|_2^2 \left\|\boldsymbol{U}^{\natural}\right\|_{\mathrm{F}}^2 = \sigma^4 n_2 r.
\end{aligned}$$

To sum up, we have obtained

$$v \le \sigma^4 n_2 r.$$

- Note that for each $l \in [n_2]$, one can decompose

$$\|\boldsymbol{X}_l\| \le \left\|\boldsymbol{U}^{\natural\top}\boldsymbol{E}_{\cdot,l}\right\|_2^2 + \left\|\boldsymbol{U}^{\natural\top}\boldsymbol{D}_l\boldsymbol{U}^{\natural}\right\| = \underbrace{\left\|\sum_{i=1}^{n_1} E_{i,l}\boldsymbol{U}^{\natural}_{i,\cdot}\right\|_2^2}_{=:\beta_1} + \underbrace{\left\|\sum_{i=1}^{n_1} E_{i,l}^2\boldsymbol{U}^{\natural\top}_{i,\cdot}\boldsymbol{U}^{\natural}_{i,\cdot}\right\|}_{=:\beta_2}.$$

  - For β_1, it is straightforward to calculate

$$\begin{aligned}
L_{\beta_1} &\coloneqq \max_{i\in[n_1]} \left\|E_{i,l}\boldsymbol{U}^{\natural}_{i,\cdot}\right\|_2 \le B\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} \le B\sqrt{\frac{\mu^{\natural} r}{n_1}}, \\
V_{\beta_1} &\coloneqq \sum_{i=1}^{n_1} \mathbb{E}\left[E_{i,l}^2\right]\left\|\boldsymbol{U}^{\natural}_{i,\cdot}\right\|_2^2 \le \sigma^2\left\|\boldsymbol{U}^{\natural}\right\|_{\mathrm{F}}^2 = \sigma^2 r.
\end{aligned}$$

  The matrix Bernstein inequality (Tropp, 2015, Theorem 6.1.1) tells us that

$$\mathbb{P}(\beta_1 \ge t) \le r\exp\left(\frac{-t/2}{V_{\beta_1} + L_{\beta_1}\sqrt{t}/3}\right).$$

  - Regarding β_2, we can also calculate

$$\begin{aligned}
L_{\beta_2} &\coloneqq \max_{i\in[n_1]} \left\|E_{i,l}^2\boldsymbol{U}^{\natural\top}_{i,\cdot}\boldsymbol{U}^{\natural}_{i,\cdot}\right\|_2 \le B^2\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}^2 \le B^2\frac{\mu^{\natural} r}{n_1}, \\
V_{\beta_2} &\coloneqq \left\|\sum_{i=1}^{n_1} \mathbb{E}\left[E_{i,l}^4\right]\left\|\boldsymbol{U}^{\natural}_{i,\cdot}\right\|_2^2\boldsymbol{U}^{\natural\top}_{i,\cdot}\boldsymbol{U}^{\natural}_{i,\cdot}\right\| \le \sigma^2B^2\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}^2 \le \sigma^2B^2\frac{\mu^{\natural} r}{n_1}.
\end{aligned}$$

  By virtue of the matrix Bernstein inequality (Tropp, 2015, Theorem 6.1.1), we can obtain

$$\mathbb{P}(\beta_2 \ge t) \le r\exp\left(\frac{-t^2/2}{V_{\beta_2} + L_{\beta_2}t/3}\right).$$

  - Combine these tail probability bounds for β_1 and β_2 to achieve

$$\begin{aligned}
\mathbb{P}(\|\boldsymbol{X}_l\| \ge t) &\le \mathbb{P}\left(\beta_1 \ge \frac{t}{2}\right) + \mathbb{P}\left(\beta_2 \ge \frac{t}{2}\right) \\
&\le r\exp\left(\frac{-t/4}{V_{\beta_1} + L_{\beta_1}\sqrt{t}/5}\right) + r\exp\left(\frac{-t^2/8}{V_{\beta_2} + L_{\beta_2}t/6}\right)
\end{aligned}$$

$$\leq r\exp\left(-\min\left\{\frac{t}{8V_{\beta_1}},\frac{5\sqrt{t}}{8L_{\beta_1}}\right\}\right)+r\exp\left(-\min\left\{\frac{t^2}{16V_{\beta_2}},\frac{3t}{8L_{\beta_2}}\right\}\right).$$

Therefore, we can set

$$L \coloneqq \widetilde{C}\left(V_{\beta_1}\log^2 n+L_{\beta_1}^2\log^2 n+\sqrt{V_{\beta_2}}\log n+L_{\beta_2}\log n\right)\asymp\sigma^2 r\log^2 n+\frac{B^2\mu^\natural r}{n_1}\log^2 n$$

for some sufficiently large constant $\widetilde{C}>0$, so as to guarantee that

$$\mathbb{P}\left(\|\boldsymbol{X}_l\|\geq L\right)\leq q_0\coloneqq\frac{1}{n^{10}},\qquad\text{for all } l\in[n_1].$$

- Based on the above choice of L, we can see that

$$\begin{aligned} q_1 &\coloneqq \left\|\mathbb{E}\left[\boldsymbol{X}_l\,\mathbb{1}\left\{\|\boldsymbol{X}_l\|\geq L\right\}\right]\right\|\leq\mathbb{E}\left[\|\boldsymbol{X}_l\|\,\mathbb{1}\left\{\|\boldsymbol{X}_l\|\geq L\right\}\right]\\ &=\int_0^\infty\mathbb{P}\left(\|\boldsymbol{X}_l\|\,\mathbb{1}\left\{\|\boldsymbol{X}_l\|\geq L\right\}>t\right)\mathrm{d}t\\ &=\int_0^L\mathbb{P}\left(\|\boldsymbol{X}_l\|\geq L\right)\mathrm{d}t+\int_L^\infty\mathbb{P}\left(\|\boldsymbol{X}_l\|>t\right)\mathrm{d}t\\ &\leq\frac{L}{n^{10}}+\int_L^\infty\mathbb{P}\left(\|\boldsymbol{X}_l\|>t\right)\mathrm{d}t, \end{aligned}$$

where the second inequality uses Jensen's inequality. Notice that for $t\geq L$, we have $t\gg V_{\beta_1}$ and $t\gg\sqrt{V_{\beta_2}}$ as long as the constant $\widetilde{C}$ is sufficiently large. As a consequence,

$$\begin{aligned} \mathbb{P}\left(\|\boldsymbol{X}_l\|\geq t\right) &\leq r\exp\left(-\min\left\{\frac{t}{8V_{\beta_1}},\frac{5\sqrt{t}}{8L_{\beta_1}}\right\}\right)+r\exp\left(-\min\left\{\frac{t^2}{16V_{\beta_2}},\frac{3t}{8L_{\beta_2}}\right\}\right)\\ &\leq r\exp\left(-\min\left\{\frac{\sqrt{t}}{\sqrt{V_{\beta_1}}},\frac{5\sqrt{t}}{8L_{\beta_1}}\right\}\right)+r\exp\left(-\min\left\{\frac{t}{\sqrt{V_{\beta_2}}},\frac{3t}{8L_{\beta_2}}\right\}\right)\\ &\leq r\exp\left(-\frac{\sqrt{t}}{\max\left\{\sqrt{V_{\beta_1}},2L_{\beta_1}\right\}}\right)+r\exp\left(-\frac{t}{\max\left\{\sqrt{V_{\beta_2}},3L_{\beta_2}\right\}}\right), \end{aligned}$$

provided that $\widetilde{C}$ is sufficiently large. Consequently, we can deduce that

$$\int_L^\infty\mathbb{P}\left(\|\boldsymbol{X}_l\|\geq t\right)\mathrm{d}t\leq r\underbrace{\int_L^\infty\exp\left(-\frac{\sqrt{t}}{\max\left\{\sqrt{V_{\beta_1}},2L_{\beta_1}\right\}}\right)\mathrm{d}t}_{=:I_1}+r\underbrace{\int_L^\infty\exp\left(-\frac{t}{\max\left\{\sqrt{V_{\beta_2}},3L_{\beta_2}\right\}}\right)\mathrm{d}t}_{=:I_2}.$$

  - The first integral obeys

$$\begin{aligned} I_1 &\overset{\text{(i)}}{=} 2\max\left\{\sqrt{V_{\beta_1}},2L_{\beta_1}\right\}\sqrt{L}\exp\left(-\frac{\sqrt{L}}{\max\left\{\sqrt{V_{\beta_1}},2L_{\beta_1}\right\}}\right)\\ &\quad+2\max\left\{\sqrt{V_{\beta_1}},2L_{\beta_1}\right\}^2\exp\left(-\frac{\sqrt{L}}{\max\left\{\sqrt{V_{\beta_1}},2L_{\beta_1}\right\}}\right)\\ &\leq\left(\frac{4}{\sqrt{\widetilde{C}}}+\frac{8}{\widetilde{C}}\right)L\exp\left(-\frac{\sqrt{\widetilde{C}}}{2}\log n\right)\overset{\text{(iii)}}{\leq}\frac{L}{n^{20}}. \end{aligned}$$

    Here, (i) follows from the following formula that holds for any constants $\alpha,\beta>0$:

$$\int_\beta^\infty\exp\left(-\alpha\sqrt{x}\right)\mathrm{d}x\overset{y=\sqrt{x}}{=}2\int_{\sqrt{\beta}}^\infty y\exp\left(-\alpha y\right)\mathrm{d}y=\left[-\frac{2}{\alpha}y\exp\left(-\alpha y\right)\right]\Bigg|_{\sqrt{\beta}}^\infty+\frac{2}{\alpha}\int_{\sqrt{\beta}}^\infty\exp\left(-\alpha y\right)\mathrm{d}y$$

$$= \frac{2\sqrt{\beta}}{\alpha} \exp\left(-\alpha\sqrt{\beta}\right) + \left[-\frac{2}{\alpha^2}\exp\left(-\alpha y\right)\right]\Bigg|_{\sqrt{\beta}}^{\infty}$$
$$= \frac{2\sqrt{\beta}}{\alpha} \exp\left(-\alpha\sqrt{\beta}\right) + \frac{2}{\alpha^2}\exp\left(-\alpha\sqrt{\beta}\right);$$

(ii) comes from the definition of L:

$$\sqrt{L} \geq \sqrt{\widetilde{C}\left(V_{\beta_1}\log^2 n + L_{\beta_1}^2 \log^2 n\right)} \geq \sqrt{\widetilde{C}} \max\left\{\sqrt{V_{\beta_1}}, L_{\beta_1}\right\} \log n;$$

and (iii) holds provided that $\widetilde{C}$ is sufficiently large.

– The second integral satisfies

$$I_2 = \max\left\{\sqrt{V_{\beta_2}}, 3L_{\beta_2}\right\} \exp\left(-\frac{L}{\max\left\{\sqrt{V_{\beta_2}}, 3L_{\beta_2}\right\}}\right) \leq \frac{3L}{\widetilde{C}} \exp\left(-\frac{\widetilde{C}}{3}\log n\right) \leq \frac{L}{n^{20}}.$$

Here the penultimate inequality follows from

$$L \geq \widetilde{C}\left(\sqrt{V_{\beta_2}}\log n + L_{\beta_2}\log n\right) \geq \widetilde{C}\max\left\{\sqrt{V_{\beta_2}}, L_{\beta_2}\right\}\log n,$$

and the last inequality holds provided that $\widetilde{C}$ is sufficiently large.

Therefore we conclude that

$$q_1 \leq \frac{L}{n^{10}} + rI_1 + rI_2 \leq \frac{L}{n^{10}} + 2r\frac{L}{n^{20}} \leq \frac{L}{n^9}.$$

- With the above quantities in mind, we are ready to use the truncated matrix Bernstein inequality (Chen et al., 2020c, Theorem 3.1.1) to show that with probability exceeding $1 - O(n^{-10})$,

$$\begin{aligned}
\|\boldsymbol{J}_2\| &\lesssim \sqrt{v\log n} + L\log n + nq_1 \lesssim \sigma^2\sqrt{n_2 r\log n} + \sigma^2 r\log^3 n + \frac{B^2\mu^\natural r}{n_1}\log^3 n \\
&\overset{\text{(i)}}{\lesssim} \sigma^2\sqrt{n_2 r\log n} + \sigma^2 r\log^3 n + \frac{\sigma^2\sqrt{n_1 n_2}\mu^\natural r}{n_1}\log^2 n \\
&\asymp \left(\sqrt{\frac{r}{n_1}} + \frac{r\log^2 n}{\sqrt{n_1 n_2}} + \frac{\mu^\natural r\log n}{n_1}\right)\sigma^2\sqrt{n_1 n_2}\log n \overset{\text{(ii)}}{\lesssim} \sqrt{\frac{r}{n_1}}\zeta_{\mathsf{op}}.
\end{aligned}$$

Here, (i) makes use of the noise condition $B \lesssim \sigma\sqrt[4]{n_1 n_2}/\sqrt{\log n}$, whereas (ii) holds provided that $n_1 \gtrsim \mu^{\natural 2} r\log^2 dn$ and $n_2 \gtrsim r\log^4 n$.

Step 3.3: bounding $\|\boldsymbol{J}_3\|$. Regarding $\boldsymbol{J}_3$, it is easy to show that

$$\|\boldsymbol{J}_3\| \leq \left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G} - \boldsymbol{G}^\natural\right)\right\| \lesssim \kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{n_1}}\zeta_{\mathsf{op}},$$

which results from Lemma 23.

Step 3.4: bounding $\|\boldsymbol{J}_4\|$. We are now left with bounding $\boldsymbol{J}_4$. Note that

$$\|\boldsymbol{J}_4\| \leq \left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{E}\boldsymbol{M}^{\natural\top} + \boldsymbol{M}^\natural\boldsymbol{E}^\top\right)\right\| \leq 2\max_{i\in[n_1]}\left|\sum_{j=1}^{n_2} E_{i,j}M_{i,j}^\natural\right|.$$

For any $i \in [n_1]$, it is straightforward to calculate that

$$L_i \coloneqq \max_{j\in[n_2]} \left|E_{i,j} M^{\natural}_{i,j}\right| \le B \left\|\boldsymbol{M}^{\natural}\right\|_{\infty},$$

$$V_i \coloneqq \mathsf{var}\left(\sum_{j=1}^{n_2} E_{i,j} M^{\natural}_{i,j}\right) = \sum_{j=1}^{n_2} \sigma_{i,j}^2 \left|M^{\natural}_{i,j}\right|^2 \le \sigma^2 \left\|\boldsymbol{M}^{\natural}\right\|_{2,\infty}^2 .$$

In view of the Bernstein inequality (Vershynin, 2017, Theorem 2.8.4),

$$\begin{aligned}
\left|\sum_{j=1}^{n_2} E_{i,j} M^{\natural}_{i,j}\right| &\lesssim \sqrt{V_i \log n} + L_i \log n \lesssim \sigma \left\|\boldsymbol{M}^{\natural}\right\|_{2,\infty} \sqrt{\log n} + B \left\|\boldsymbol{M}^{\natural}\right\|_{\infty} \log n \\
&\lesssim \sigma\sigma_1^{\natural} \left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} \sqrt{\log n} + \sigma\sqrt{n_2}\sqrt{\frac{\mu^{\natural} r}{n_1 n_2}}\sigma_1^{\natural} \\
&\lesssim \sigma\sigma_1^{\natural}\sqrt{\frac{\mu^{\natural} r \log n}{n_1}} \lesssim \frac{\sqrt{\mu^{\natural} r}}{n_1}\zeta_{\mathsf{op}} \qquad \text{(F.44)}
\end{aligned}$$

with probability exceeding $1 - O(n^{-11})$. Combining the above bounds and applying the union bound show that with probability exceeding $1 - O(n^{-10})$,

$$\|\boldsymbol{J}_4\| \lesssim \frac{\sqrt{\mu^{\natural} r}}{n_1}\zeta_{\mathsf{op}}.$$

Step 3.5: putting all this together. Taking the previous bounds on the spectral norm of $\boldsymbol{J}_1$, $\boldsymbol{J}_2$, $\boldsymbol{J}_3$, $\boldsymbol{J}_4$ collectively yields

$$\alpha_3 \le \|\boldsymbol{J}_1\| + \|\boldsymbol{J}_2\| + \|\boldsymbol{J}_3\| + \|\boldsymbol{J}_4\| \lesssim \kappa^{\natural 2}\sqrt{\frac{\mu^{\natural} r}{n_1}}\zeta_{\mathsf{op}} + \frac{\mu^{\natural} r}{n_1}\zeta_{\mathsf{op}} \asymp \kappa^{\natural 2}\sqrt{\frac{\mu^{\natural} r}{n_1}}\zeta_{\mathsf{op}},$$

where the last relation holds as long as $n_1 \gtrsim \mu^{\natural} r$.

Step 4: combining the bounds on α_1, α_2 and α_3. Taking the bounds on α_1, α_2, α_3 together leads to

$$\left\|\boldsymbol{R}^{\top}\boldsymbol{\Sigma}^2\boldsymbol{R} - \boldsymbol{\Sigma}^{\natural 2}\right\| \le \alpha_1 + \alpha_2 + \alpha_3 \lesssim \kappa^{\natural 2}\sqrt{\frac{\mu^{\natural} r}{n_1}}\zeta_{\mathsf{op}} + \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}} + \frac{\zeta_{\mathsf{op}}^3}{\sigma_r^{\natural 4}} \asymp \kappa^{\natural 2}\sqrt{\frac{\mu^{\natural} r}{n_1}}\zeta_{\mathsf{op}} + \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}$$

with probability exceeding $1 - O(n^{-10})$, where the last relation is valid under the condition that $\zeta_{\mathsf{op}} \ll \sigma_r^{\natural 2}$.

F.3.7 Proof of Lemma 28

To begin with, observe that $\boldsymbol{G}^{\natural} = \boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{U}^{\natural\top}$ (and hence $\boldsymbol{G}^{\natural}\boldsymbol{U}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-2} = \boldsymbol{U}^{\natural}$), which allows us to decompose

$$\begin{aligned}
\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\natural}\right\|_{2,\infty} &= \left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{G}\boldsymbol{U}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-2} + \boldsymbol{G}\boldsymbol{U}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-2} - \boldsymbol{G}^{\natural}\boldsymbol{U}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-2}\right\|_{2,\infty} \\
&\le \underbrace{\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{G}\boldsymbol{U}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-2}\right\|_{2,\infty}}_{=:\,\alpha_1} + \underbrace{\left\|\left(\boldsymbol{G} - \boldsymbol{G}^{\natural}\right)\boldsymbol{U}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-2}\right\|_{2,\infty}}_{=:\,\alpha_2}.
\end{aligned}$$

We then proceed to bound the terms α_1 the α_2.

- Regarding α_1, we learn from Abbe et al. (2020, Lemma 1) that for each $m \in [n_1]$,

$$\left\|\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{G}\boldsymbol{U}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-2}\right)_{m,\cdot}\right\|_2 \lesssim \underbrace{\frac{1}{\sigma_r^{\natural 4}}\left\|\boldsymbol{G} - \boldsymbol{G}^{\natural}\right\|\left\|\boldsymbol{G}_{m,\cdot}\boldsymbol{U}^{\natural}\right\|_2}_{=:\beta_1} + \underbrace{\frac{1}{\sigma_r^{\natural 2}}\left\|\boldsymbol{G}_{m,\cdot}\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\natural}\right)\right\|_2}_{=:\beta_2}.$$

We first bound β_1. From Lemma 25 we know that

$$\begin{aligned}\left\|\boldsymbol{G}_{m,\cdot}\boldsymbol{U}^{\natural}\right\|_2 &\le \left\|\left(\boldsymbol{G} - \boldsymbol{G}^{\natural}\right)_{m,\cdot}\boldsymbol{U}^{\natural}\right\|_2 + \left\|\boldsymbol{G}^{\natural}_{m,\cdot}\boldsymbol{U}^{\natural}\right\|_2 \le \left\|\left(\boldsymbol{G} - \boldsymbol{G}^{\natural}\right)_{m,\cdot}\boldsymbol{U}^{\natural}\right\|_2 + \left\|\boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{U}^{\natural\top}\boldsymbol{U}^{\natural}\right\|_{2,\infty}\\ &= \left\|\left(\boldsymbol{G} - \boldsymbol{G}^{\natural}\right)_{m,\cdot}\boldsymbol{U}^{\natural}\right\|_2 + \left\|\boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural 2}\right\|_{2,\infty}\\ &\lesssim \zeta_{\mathsf{op}}\sqrt{\frac{\mu^{\natural}r}{n_1}} + \sigma_1^{\natural 2}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} \lesssim \sigma_1^{\natural 2}\sqrt{\frac{\mu^{\natural}r}{n_1}},\end{aligned}$$

where the last line holds provided that $\zeta_{\mathsf{op}} \lesssim \sigma_1^{\natural 2}$. This combined with Lemma 23 yields

$$\beta_1 \lesssim \frac{1}{\sigma_r^{\natural 4}}\zeta_{\mathsf{op}}\sigma_1^{\natural 2}\sqrt{\frac{\mu^{\natural}r}{n_1}} \lesssim \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^{\natural}r}{n_1}}.$$

Additionally, Lemma 26 tells us that

$$\beta_2 \le \frac{1}{\sigma_r^{\natural 2}}\left\|\boldsymbol{G}\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\natural}\right)\right\|_{2,\infty} \lesssim \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\natural}\right\|_{2,\infty} + \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}}\sqrt{\frac{\mu^{\natural}r}{n_1}}.$$

Therefore, for all $m \in [n_1]$ we have

$$\left\|\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{G}\boldsymbol{U}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-2}\right)_{m,\cdot}\right\|_2 \lesssim \beta_1 + \beta_2 \lesssim \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^{\natural}r}{n_1}} + \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\natural}\right\|_{2,\infty},$$

where the last inequality holds provided that $\zeta_{\mathsf{op}} \ll \sigma_r^{\natural 2}$. As a result, we arrive at

$$\alpha_1 = \max_{1\le m\le n_1}\left\|\left(\boldsymbol{U}\boldsymbol{H} - \boldsymbol{G}\boldsymbol{U}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-2}\right)_{m,\cdot}\right\|_2 \lesssim \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^{\natural}r}{n_1}} + \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\natural}\right\|_{2,\infty}.$$

- In view of Lemma 25, the second term α_2 can be bounded by

$$\alpha_2 \le \frac{1}{\sigma_r^{\natural 2}}\left\|\left(\boldsymbol{G} - \boldsymbol{G}^{\natural}\right)\boldsymbol{U}^{\natural}\right\|_{2,\infty} \lesssim \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^{\natural}r}{n_1}}.$$

The preceding bounds taken together allow us to conclude that

$$\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\natural}\right\|_{2,\infty} \le \alpha_1 + \alpha_2 \lesssim \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^{\natural}r}{n_1}} + \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\natural}\right\|_{2,\infty}.$$

As long as $\zeta_{\mathsf{op}} \ll \sigma_r^{\natural 2}$, we can rearrange terms to arrive at

$$\left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\natural}\right\|_{2,\infty} \lesssim \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^{\natural}r}{n_1}}, \tag{F.45}$$

which combined with Lemma 24 gives

$$\left\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^{\natural}\right\|_{2,\infty} \le \left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\natural}\right\|_{2,\infty} + \left\|\boldsymbol{U}\left(\boldsymbol{H} - \boldsymbol{R}_{\boldsymbol{U}}\right)\right\|_{2,\infty} \lesssim \left\|\boldsymbol{U}\boldsymbol{H} - \boldsymbol{U}^{\natural}\right\|_{2,\infty} + \left\|\boldsymbol{U}\right\|_{2,\infty}\left\|\boldsymbol{H} - \boldsymbol{R}_{\boldsymbol{U}}\right\|$$

$$\lesssim \left\|\boldsymbol{U}\boldsymbol{H}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\left(\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\left\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}\right)\left\|\boldsymbol{H}-\boldsymbol{R}_{\boldsymbol{U}}\right\|$$

$$\lesssim \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^{\natural}r}{n_1}}+\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}}\sqrt{\frac{\mu^{\natural}r}{n_1}}+\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}}\left\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}.$$

Once again, when $\zeta_{\mathsf{op}} \ll \sigma_r^{\natural 2}$, one can rearrange terms to yield

$$\left\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}\lesssim \kappa^{\natural 2}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\sqrt{\frac{\mu^{\natural}r}{n_1}}.$$

F.3.8 Proof of Lemma 29

We start with $\|\boldsymbol{G}^{(m)}-\boldsymbol{G}\|$, for which the triangle inequality yields

$$\left\|\boldsymbol{G}^{(m)}-\boldsymbol{G}\right\| \le \left\|\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{G}^{(m)}-\boldsymbol{G}\right)\right\|+\left\|\mathcal{P}_{\textsf{diag}}\left(\boldsymbol{G}^{(m)}-\boldsymbol{G}\right)\right\|. \tag{F.46}$$

- Regarding the first term on the right-hand side of (F.46), it is observed that $\mathcal{P}_{\textsf{off-diag}}(\boldsymbol{G}^{(m)}-\boldsymbol{G})$ in the current paper is the same as the matrix $\boldsymbol{G}^{(m)}-\boldsymbol{G}$ in Cai et al. (2021) (due to the diagonal deletion strategy employed therein). One can then apply Cai et al. (2021, Lemma 6) to show that

$$\begin{aligned}\left\|\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{G}^{(m)}-\boldsymbol{G}\right)\right\| &\lesssim \sigma\sqrt{n_2}\left(\sigma\sqrt{n_1}+\left\|\boldsymbol{M}^{\natural\top}\right\|_{2,\infty}\right)\sqrt{\log n}\\ &\lesssim \sigma^2\sqrt{n_1n_2\log n}+\sigma\sqrt{n_2}\sqrt{\frac{\mu^{\natural}r}{n_2}}\sigma_1^{\natural}\sqrt{\log n}\\ &\asymp \sigma^2\sqrt{n_1n_2\log n}+\sigma\sigma_1^{\natural}\sqrt{\mu^{\natural}r\log n}\end{aligned}$$

  with probability exceeding $1-O(n^{-11})$, where the second inequality relies on (F.42). Here, we have replaced σ_{col} (resp. σ_{row}) in Cai et al. (2021, Lemma 6) with $\sigma\sqrt{n_1}$ (resp. $\sigma\sqrt{n_2}$) under our setting.

- Recalling that the diagonal of $\boldsymbol{G}^{(m)}$ coincides with the true diagonal of $\boldsymbol{G}^{\natural}$, we can invoke Lemma 23 to bound the second term on the right-hand side of (F.46) as follows

$$\left\|\mathcal{P}_{\textsf{diag}}\left(\boldsymbol{G}^{(m)}-\boldsymbol{G}\right)\right\| = \left\|\mathcal{P}_{\textsf{diag}}\left(\boldsymbol{G}^{\natural}-\boldsymbol{G}\right)\right\| \lesssim \kappa^{\natural 2}\sqrt{\frac{\mu^{\natural}r}{n_1}}\zeta_{\mathsf{op}}.$$

Combining the above two bounds and invoke the union bound lead to

$$\begin{aligned}\left\|\boldsymbol{G}^{(m)}-\boldsymbol{G}\right\| &\le \left\|\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{G}^{(m)}-\boldsymbol{G}\right)\right\|+\left\|\mathcal{P}_{\textsf{diag}}\left(\boldsymbol{G}^{(m)}-\boldsymbol{G}\right)\right\|\\ &\lesssim \left\{\sigma^2\sqrt{n_1n_2\log n}+\sigma\sigma_1^{\natural}\sqrt{\mu^{\natural}r\log n}\right\}+\kappa^{\natural 2}\sqrt{\frac{\mu^{\natural}r}{n_1}}\zeta_{\mathsf{op}}\\ &\lesssim \sigma^2\sqrt{n_1n_2\log n}+\sigma\sigma_1^{\natural}\sqrt{\mu^{\natural}r\log n}+\kappa^{\natural 2}\sqrt{\frac{\mu^{\natural}r}{n_1}}\left(\sigma^2\sqrt{n_1n_2}\log n+\sigma\sigma_1^{\natural}\sqrt{n_1\log n}\right)\\ &\lesssim \sigma^2\sqrt{n_1n_2}\log n+\kappa^{\natural 2}\sigma\sigma_1^{\natural}\sqrt{\mu^{\natural}r\log n}\end{aligned}$$

simultaneously for all $1 \le m \le n_1$, as long as $n_1 \gtrsim \kappa^{\natural 4}\mu^{\natural}r$. Here, the penultimate inequality relies on the definition of ζ_{op}.

To finish up, taking the above inequality and Lemma 23 collectively yields

$$\left\|\boldsymbol{G}^{(m)}-\boldsymbol{G}^{\natural}\right\| \le \left\|\boldsymbol{G}^{(m)}-\boldsymbol{G}\right\|+\left\|\boldsymbol{G}-\boldsymbol{G}^{\natural}\right\| \lesssim \sigma^2\sqrt{n_1n_2\log n}+\kappa^{\natural 2}\sigma\sigma_1^{\natural}\sqrt{\mu^{\natural}r\log n}+\zeta_{\mathsf{op}} \asymp \zeta_{\mathsf{op}},$$

with the proviso that $n_1 \gtrsim \kappa^{\natural 4}\mu^{\natural}r$.

F.3.9 Proof of Lemma 30

The proof of Lemma 30 can be directly adapted from the proof of Cai et al. (2021, Lemma 9) (see Cai et al. (2021, Appendix C.5)). Two observations are crucial:

- The off-diagonal part of $\boldsymbol{G}^{(m)}$ (resp. $\boldsymbol{G}^{(m,l)}$) in this paper is the same as that of $\boldsymbol{G}^{(m)}$ (resp. $\boldsymbol{G}^{(m,l)}$) defined in Cai et al. (2021).
- Regarding the diagonal, this paper imputes the diagonals of both $\boldsymbol{G}^{(m)}$ and $\boldsymbol{G}^{(m,l)}$ with the diagonal of the ground truth $\boldsymbol{G}^{\natural}$, while the diagonal entries of both $\boldsymbol{G}^{(m)}$ and $\boldsymbol{G}^{(m,l)}$ in Cai et al. (2021) are all zeros.

Therefore, in both the current paper and Cai et al. (2021), we end up dealing with the same matrix $\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}$, and consequently, the bound on $\|\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)}\|$ established in Cai et al. (2021, Appendix C.5.1) remains valid when it comes to our setting.

In addition, we can easily check that the bound on $\|(\boldsymbol{G}^{(m)} - \boldsymbol{G}^{(m,l)})\boldsymbol{U}^{(m,l)}\|$ derived in Cai et al. (2021, Appendix C.5.2) also holds under our setting; this is simply because the analysis in Cai et al. (2021, Appendix C.5.2) remains valid if $\boldsymbol{G}$ and $\boldsymbol{G}^{(l)}$ have the same deterministic diagonal. By replacing σ_{col} with $\sigma\sqrt{n_1}$ in Cai et al. (2021, Lemma 9), we arrive at the result claimed in Lemma 30.

F.3.10 Proof of Lemma 31

For the sake of brevity, we shall only focus on proving (F.16). The proof of (F.17) is similar to — and in fact, simpler than — the proof of (F.16), and can also be directly adapted from the proof of Cai et al. (2021, Lemma 7).

For notational simplicity, we denote $\boldsymbol{B} := \mathcal{P}_{-m,\cdot}(\boldsymbol{M})$. For any $l \in [n_2]$, we can write

$$\begin{aligned}
\left\|\boldsymbol{e}_l^{\top}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{M})\right]^{\top}\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^{\natural}\right)\right\|_2 &= \left\|\boldsymbol{B}_{\cdot,l}^{\top}\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^{\natural}\right)\right\|_2 \\
&\leq \underbrace{\left\|\mathbb{E}\left(\boldsymbol{B}_{\cdot,l}\right)^{\top}\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^{\natural}\right)\right\|_2}_{=:\alpha_1} + \underbrace{\left\|\left[\boldsymbol{B}_{\cdot,l} - \mathbb{E}\left(\boldsymbol{B}_{\cdot,l}\right)\right]^{\top}\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right)\right\|_2}_{=:\alpha_2} \\
&\quad + \underbrace{\left\|\left[\boldsymbol{B}_{\cdot,l} - \mathbb{E}\left(\boldsymbol{B}_{\cdot,l}\right)\right]^{\top}\left(\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)} - \boldsymbol{U}^{\natural}\right)\right\|_2}_{=:\alpha_3}.
\end{aligned}$$

Therefore, we seek to bound α_1, α_2 and α_3 separately.

- Let us begin with the quantity α_1. It is straightforward to see that

$$\begin{aligned}
\alpha_1 &= \left\|\boldsymbol{M}_{\cdot,l}^{\natural\top}\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^{\natural}\right)\right\|_2 = \left\|\boldsymbol{M}_{\cdot,l}^{\natural\top}\boldsymbol{U}^{\natural}\boldsymbol{U}^{\natural\top}\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^{\natural}\right)\right\|_2 \\
&\leq \left\|\boldsymbol{M}^{\natural\top}\right\|_{2,\infty}\left\|\boldsymbol{U}^{\natural\top}\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^{\natural}\right)\right\| \\
&= \left\|\boldsymbol{M}^{\natural\top}\right\|_{2,\infty}\left\|\boldsymbol{H}^{(m)\top}\boldsymbol{H}^{(m)} - \boldsymbol{I}_r\right\|,
\end{aligned} \tag{F.47}$$

  where the last identity arises from the following relation

$$\left\|\boldsymbol{U}^{\natural\top}\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^{\natural}\right)\right\| = \left\|\boldsymbol{U}^{\natural\top}\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}\boldsymbol{U}^{\natural} - \boldsymbol{I}_r\right\| = \left\|\boldsymbol{H}^{(m)\top}\boldsymbol{H}^{(m)} - \boldsymbol{I}_r\right\|.$$

  Let us write the SVD of $\boldsymbol{H}^{(m)} = \boldsymbol{U}^{(m)\top}\boldsymbol{U}^{\natural}$ as $\boldsymbol{X}(\cos\boldsymbol{\Theta})\boldsymbol{Y}^{\top}$, where $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{r\times r}$ are square orthonormal matrices and $\boldsymbol{\Theta}$ is a diagonal matrix composed of the principal angles between $\boldsymbol{U}^{(m)}$ and $\boldsymbol{U}^{\natural}$. This allows us to deduce that

$$\left\|\boldsymbol{H}^{(m)\top}\boldsymbol{H}^{(m)} - \boldsymbol{I}_r\right\| = \left\|\boldsymbol{Y}\left(\boldsymbol{I}_r - \cos^2\boldsymbol{\Theta}\right)\boldsymbol{Y}^{\top}\right\| = \left\|\boldsymbol{I}_r - \cos^2\boldsymbol{\Theta}\right\| = \left\|\sin^2\boldsymbol{\Theta}\right\| = \left\|\sin\boldsymbol{\Theta}\right\|^2.$$

  In view of Davis-Kahan's $\sin\boldsymbol{\Theta}$ Theorem (Chen et al., 2020c, Theorem 2.2.1), we have

$$\left\|\sin\boldsymbol{\Theta}\right\| \leq \frac{\left\|\boldsymbol{G}^{(m)} - \boldsymbol{G}^{\natural}\right\|}{\lambda_r\left(\boldsymbol{G}^{(m)}\right) - \lambda_{r+1}\left(\boldsymbol{G}^{\natural}\right)} \lesssim \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}},$$

where the last inequality follows from Lemma 29 and an application of Weyl's inequality:

$$\lambda_r\left(\boldsymbol{G}^{(m)}\right) \geq \lambda_r\left(\boldsymbol{G}^{\natural}\right) - \left\|\boldsymbol{G}^{(m)} - \boldsymbol{G}^{\natural}\right\| \overset{\text{(i)}}{\geq} \sigma_r^{\natural 2} - \widetilde{C}\zeta_{\mathsf{op}} \overset{\text{(ii)}}{\geq} \frac{1}{2}\sigma_r^{\natural 2},$$

with $\widetilde{C} > 0$ representing some absolute constant. Here, (i) results from Lemma 29, while (ii) holds provided that $\zeta_{\mathsf{op}} \ll \sigma_r^{\natural 2}$.. Therefore, we arrive at

$$\left\|\boldsymbol{H}^{(m)\top}\boldsymbol{H}^{(m)} - \boldsymbol{I}_r\right\| \leq \left\|\sin\boldsymbol{\Theta}\right\|^2 \lesssim \frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}}.$$

Substitution into (F.47) yields

$$\alpha_1 \lesssim \frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}}\left\|\boldsymbol{M}^{\natural\top}\right\|_{2,\infty}.$$

- Regarding α_2, it is observed that

$$\begin{aligned}
\alpha_2 &\leq \left(\left\|\boldsymbol{B}_{\cdot,l}\right\|_2 + \left\|\mathbb{E}\left(\boldsymbol{B}_{\cdot,l}\right)\right\|_2\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)}\right\| \\
&\leq \left(\left\|\boldsymbol{M}_{\cdot,l}\right\|_2 + \left\|\boldsymbol{M}_{\cdot,l}^{\natural}\right\|_2\right)\left\|\left(\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top} - \boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right)\boldsymbol{U}^{\natural}\right\| \\
&\leq \left(\left\|\boldsymbol{M}^{\natural\top}\right\|_{2,\infty} + B\sqrt{\log n} + \sigma\sqrt{n_1}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top} - \boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\|.
\end{aligned}$$

  Here, the last inequality arises from Cai et al. (2021, Lemma 12).

- We are now left with the quantity α_3, which can be expressed as

$$\alpha_3 = \left\|\sum_{i\in[n_1]\setminus\{m\}} E_{i,l}\left(\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)} - \boldsymbol{U}^{\natural}\right)_{i,\cdot}\right\|_2.$$

  Conditional on $\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)} - \boldsymbol{U}^{\natural}$, this term can be viewed as the spectral norm of a sum of independent mean-zero random vectors (where the randomness comes from $\{E_{i,l}\}_{i\in[n_1]\setminus\{m\}}$). To control this term, we first calculate

$$\begin{aligned}
L &\coloneqq \max_{i\in[n_1]\setminus\{m\}}\left\|E_{i,l}\left(\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)} - \boldsymbol{U}^{\natural}\right)_{i,\cdot}\right\| \leq B\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)} - \boldsymbol{U}^{\natural}\right\|_{2,\infty}, \\
V &\coloneqq \sum_{i\in[n_1]\setminus\{m\}}\mathbb{E}\left[E_{i,l}^2\right]\left\|\left(\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)} - \boldsymbol{U}^{\natural}\right)_{i,\cdot}\right\|_2^2 \leq \sigma^2 n_1\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)} - \boldsymbol{U}^{\natural}\right\|_{2,\infty}^2.
\end{aligned}$$

  In view of the matrix Bernstein inequality (Tropp, 2015, Theorem 6.1.1),

$$\begin{aligned}
\alpha_3 &= \left\|\sum_{i\in[n_1]\setminus\{m\}} E_{i,l}\left(\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)} - \boldsymbol{U}^{\natural}\right)_{i,\cdot}\right\| \lesssim \sqrt{V\log n} + L\log n \\
&\lesssim \left(\sigma\sqrt{n_1\log n} + B\log n\right)\left\|\boldsymbol{U}^{(m,l)}\boldsymbol{H}^{(m,l)} - \boldsymbol{U}^{\natural}\right\|_{2,\infty}
\end{aligned}$$

  with probability exceeding $1 - O(n^{-10})$. In addition, the triangle inequality gives

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

  and therefore,

$$\alpha_3 \lesssim \left(\sigma\sqrt{n_1\log n} + B\log n\right)\left(\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^{\natural}\right\|_{2,\infty} + \left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top} - \boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\|\right).$$

Combine the preceding bounds on α_1, α_2 and α_3 to arrive at

$$\begin{aligned}
\left\|\boldsymbol{e}_l^\top \left[\mathcal{P}_{-m,\cdot}\left(\boldsymbol{M}\right)\right]^\top \left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right)\right\|_2 &\le \alpha_1 + \alpha_2 + \alpha_3 \\
&\lesssim \frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}} \left\|\boldsymbol{M}^{\natural\top}\right\|_{2,\infty} + \left(\sigma\sqrt{n_1 \log n} + B\log n\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right\|_{2,\infty} \\
&\quad + \left(\left\|\boldsymbol{M}^{\natural\top}\right\|_{2,\infty} + B\log n + \sigma\sqrt{n_1\log n}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top} - \boldsymbol{U}^{(m,l)}\boldsymbol{U}^{(m,l)\top}\right\|
\end{aligned}$$

as claimed.

F.3.11 Proof of Lemma 32

In this subsection, for the sake of brevity, we shall only focus on proving (F.18). The proof of (F.19) is similar to that of (F.18), and can also be easily adapted from the proof of Cai et al. (2021, Lemma 7).

For notational simplicity, we denote $\boldsymbol{B} \coloneqq \mathcal{P}_{-m,\cdot}(\boldsymbol{M})$, allowing us to express

$$\boldsymbol{E}_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}\left(\boldsymbol{M}\right)\right]^\top \left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right) = \sum_{j=1}^{n_2} E_{m,j}\left[\boldsymbol{B}^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right)\right]_{j,\cdot}.$$

Conditional on $\boldsymbol{B}$ and $\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural$, the above term can be viewed as a sum of independent zero-mean random vectors, where the randomness comes from $\{E_{m,j}\}_{j\in[n_2]}$. We can calculate

$$\begin{aligned}
L &\coloneqq \max_{j\in[n_2]}\left\|E_{m,j}\left[\boldsymbol{B}^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right)\right]_{j,\cdot}\right\| \le B\left\|\boldsymbol{B}^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right)\right\|_{2,\infty}, \\
V &\coloneqq \sum_{j\in[n_2]}\mathbb{E}\left(E_{m,j}^2\right)\left\|\boldsymbol{B}^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right)_{j,\cdot}\right\|_2^2 \le \sigma^2\left\|\boldsymbol{B}^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right)\right\|_{\mathrm{F}}^2.
\end{aligned}$$

In view of the matrix Bernstein inequality (Tropp, 2015, Theorem 6.1.1), with probability exceeding $1 - O(n^{-11})$

$$\begin{aligned}
\left\|\sum_{j=1}^{n_2} E_{m,j}\left[\boldsymbol{B}^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right)\right]_{j,\cdot}\right\|_2 &\lesssim \sqrt{V\log n} + L\log n \\
&\lesssim \underbrace{\sigma\left\|\boldsymbol{B}^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right)\right\|_{\mathrm{F}}\sqrt{\log n}}_{=:\alpha_1} + \underbrace{B\left\|\boldsymbol{B}^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right)\right\|_{2,\infty}\log n}_{=:\alpha_2}.
\end{aligned}$$

For the first term α_1, with probability exceeding $1 - O(n^{-11})$ we have

$$\begin{aligned}
\left\|\boldsymbol{B}^\top\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right)\right\|_{\mathrm{F}} &\le \|\boldsymbol{B}\|\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right\|_{\mathrm{F}} \le \|\boldsymbol{M}\|\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right\|_{\mathrm{F}} \\
&\le \left(\|\boldsymbol{M}^\natural\| + \|\boldsymbol{E}\|\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right\|_{\mathrm{F}} \\
&\lesssim \left(\sigma_1^\natural + \sigma\sqrt{n}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right\|_{\mathrm{F}} \\
&\asymp \left(\sigma_1^\natural + \sigma\sqrt{n_2}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right\|_{\mathrm{F}}.
\end{aligned}$$

Here, the penultimate inequality uses $\|\boldsymbol{E}\| \lesssim \sigma\sqrt{n}$ with probability exceeding $1 - O(n^{-11})$, which follows from standard matrix tail bounds (e.g., Chen et al. (2020c, Theorem 3.1.4)); and the last relation holds since $n = \max\{n_1, n_2\}$ and $\sigma\sqrt{n_1} \lesssim \zeta_{\mathsf{op}}/\sigma_1^\natural \ll \sigma_1^\natural$. As a result, we reach

$$\begin{aligned}
\alpha_1 &\lesssim \sigma\left(\sigma_1^\natural + \sigma\sqrt{n_2}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right\|_{\mathrm{F}}\sqrt{\log n} \\
&\lesssim \left(\sigma\sigma_1^\natural\sqrt{n_1\log n} + \sigma^2\sqrt{n_1 n_2\log n}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)} - \boldsymbol{U}^\natural\right\|_{2,\infty}
\end{aligned}$$

$$\lesssim \zeta_{\mathsf{op}} \left\| \boldsymbol{U}^{(m)} \boldsymbol{H}^{(m)} - \boldsymbol{U}^{\natural} \right\|_{2,\infty}.$$

Regarding the second term α_2, we know from (F.16) in Lemma 31 that with probability exceeding $1 - O(n^{-10})$,

$$\alpha_2 \leq \underbrace{(B \log n) \frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 4}} \left\| \boldsymbol{M}^{\natural\top} \right\|_{2,\infty}}_{=: \beta_1} + \underbrace{(B \log n) \left(\sigma \sqrt{n_1 \log n} + B \log n \right) \left\| \boldsymbol{U}^{(m)} \boldsymbol{H}^{(m)} - \boldsymbol{U}^{\natural} \right\|_{2,\infty}}_{=: \beta_2}$$
$$+ \underbrace{(B \log n) \left(\left\| \boldsymbol{M}^{\natural\top} \right\|_{2,\infty} + B \log n + \sigma \sqrt{n_1 \log n} \right) \left\| \boldsymbol{U}^{(m)} \boldsymbol{U}^{(m)\top} - \boldsymbol{U}^{(m,l)} \boldsymbol{U}^{(m,l)\top} \right\|}_{=: \beta_3}$$

holds for all $m \in [n_1]$. In what follows, we shall bound β_1, β_2 and β_3 respectively.

- Regarding β_1, we first observe that

$$(B \log n) \left\| \boldsymbol{M}^{\natural\top} \right\|_{2,\infty} \leq (B \log n) \sqrt{\frac{\mu^{\natural} r}{n_2}} \sigma_1^{\natural} \lesssim \sigma \sqrt{n_2 \log n} \sqrt{\frac{\mu^{\natural} r}{n_2}} \sigma_1^{\natural} \lesssim \sigma \sigma_1^{\natural} \sqrt{\mu^{\natural} r \log n} \lesssim \sqrt{\frac{\mu^{\natural} r}{n_1}} \zeta_{\mathsf{op}}, \quad \text{(F.48)}$$

where we have used the noise condition $B \lesssim \sigma \sqrt{n_2 / \log n}$. Therefore

$$\beta_1 \lesssim \sqrt{\frac{\mu^{\natural} r}{n_1}} \frac{\zeta_{\mathsf{op}}^3}{\sigma_r^{\natural 4}}.$$

- When it comes to β_2, we notice that

$$\left(\sigma \sqrt{n_1 \log n} + B \log n \right)^2 \asymp \sigma^2 n_1 \log n + B^2 \log^2 d \lesssim \sigma^2 n_1 \log n + \sigma^2 \sqrt{n_1 n_2} \log n \lesssim \zeta_{\mathsf{op}}, \quad \text{(F.49)}$$

as long as $\sigma \sqrt{n_1 \log n} \ll \sigma_1^{\natural}$. Here, we have used the noise assumption $B \lesssim \sigma \sqrt[4]{n_1 n_2} / \sqrt{\log n}$. This in turn leads us to

$$\beta_2 \leq \left(\sigma \sqrt{n_1 \log n} + B \log n \right)^2 \left\| \boldsymbol{U}^{(m)} \boldsymbol{H}^{(m)} - \boldsymbol{U}^{\natural} \right\|_{2,\infty} \lesssim \zeta_{\mathsf{op}} \left\| \boldsymbol{U}^{(m)} \boldsymbol{H}^{(m)} - \boldsymbol{U}^{\natural} \right\|_{2,\infty}.$$

- We are left with the term β_3. From Lemma 30, we see that with probability exceeding $1 - O(n^{-10})$,

$$\begin{aligned}
&(B \log n) \left\| \boldsymbol{U}^{(m)} \boldsymbol{U}^{(m)\top} - \boldsymbol{U}^{(m,l)} \boldsymbol{U}^{(m,l)\top} \right\| \\
&\lesssim \frac{B \log n}{\sigma_r^{\natural 2}} \left[\left(B \log n + \sigma \sqrt{n_1 \log n} \right)^2 \left\| \boldsymbol{U}^{(m)} \boldsymbol{H}^{(m)} \right\|_{2,\infty} + \sigma^2 \right] + \frac{B \log n}{\sigma_r^{\natural 2}} \left(B \log n + \sigma \sqrt{n_1 \log n} \right) \left\| \boldsymbol{M}^{\natural\top} \right\|_{2,\infty} \\
&\lesssim \frac{B \log n}{\sigma_r^{\natural 2}} \left[\zeta_{\mathsf{op}} \left\| \boldsymbol{U}^{(m)} \boldsymbol{H}^{(m)} \right\|_{2,\infty} + \sigma^2 \right] + \left(B \log n + \sigma \sqrt{n_1 \log n} \right) \sqrt{\frac{\mu^{\natural} r}{n_1}} \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}} \\
&\lesssim \frac{B \log n}{\sigma_r^{\natural 2}} \zeta_{\mathsf{op}} \left\| \boldsymbol{U}^{(m)} \boldsymbol{H}^{(m)} \right\|_{2,\infty} + \left(B \log n + \sigma \sqrt{n_1 \log n} \right) \sqrt{\frac{\mu^{\natural} r}{n_1}} \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}
\end{aligned}$$

simultaneously for all m and l, where the penultimate inequality follows from (F.48) and (F.49), and the last inequality holds since

$$\frac{B \log n}{\sigma_r^{\natural 2}} \sigma^2 \lesssim \frac{B \log n}{\sigma_r^{\natural 2}} \sqrt{\frac{\mu^{\natural} r}{n_1}} \sigma^2 \sqrt{n_1 n_2} \lesssim B \log n \sqrt{\frac{\mu^{\natural} r}{n_1}} \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}.$$

Therefore, with probability exceeding $1 - O(n^{-10})$, we have

$$\beta_3 \lesssim \left(\left\| \boldsymbol{M}^{\natural\top} \right\|_{2,\infty} + B \log n + \sigma \sqrt{n_1 \log n} \right) \frac{B \log n}{\sigma_r^{\natural 2}} \zeta_{\mathsf{op}} \left\| \boldsymbol{U}^{(m)} \boldsymbol{H}^{(m)} \right\|_{2,\infty}$$

$$
\begin{aligned}
&+\left(\left\|\boldsymbol{M}^{\natural\top}\right\|_{2,\infty}+B\log n+\sigma\sqrt{n_1\log n}\right)\left(B\log n+\sigma\sqrt{n_1\log n}\right)\sqrt{\frac{\mu^{\natural}r}{n_1}}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\\
&\lesssim \left\|\boldsymbol{M}^{\natural\top}\right\|_{2,\infty}\frac{B\log n}{\sigma_r^{\natural 2}}\zeta_{\mathsf{op}}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}+\frac{\left(B\log n+\sigma\sqrt{n_1\log n}\right)^2}{\sigma_r^{\natural 2}}\zeta_{\mathsf{op}}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}\\
&\quad+\left\|\boldsymbol{M}^{\natural\top}\right\|_{2,\infty}\left(B\log n+\sigma\sqrt{n_1\log n}\right)\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}+\left(B\log n+\sigma\sqrt{n_1\log n}\right)^2\sqrt{\frac{\mu^{\natural}r}{n_1}}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\\
&\overset{\text{(i)}}{\lesssim}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}+\frac{\mu^{\natural}r}{n_1}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}+\left\|\boldsymbol{M}^{\natural\top}\right\|_{2,\infty}\sigma\sqrt{\mu^{\natural}r\log n}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}+\sqrt{\frac{\mu^{\natural}r}{n_1}}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\\
&\overset{\text{(ii)}}{\lesssim}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}+\frac{\mu^{\natural}r}{n_1}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}+\frac{\mu^{\natural}r}{\sqrt{n_1n_2}}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}+\sqrt{\frac{\mu^{\natural}r}{n_1}}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\\
&\overset{\text{(iii)}}{\lesssim}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}+\sqrt{\frac{\mu^{\natural}r}{n_1}}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}
\end{aligned}
$$

simultaneously for all $m\in[n_1]$, where (i) uses (F.48) and (F.49); (ii) comes from

$$
\left\|\boldsymbol{M}^{\natural\top}\right\|_{2,\infty}\sigma\sqrt{\mu^{\natural}r\log n}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\lesssim\sigma\sigma_1^{\natural}\sqrt{n_1\log n}\frac{\mu^{\natural}r}{\sqrt{n_1n_2}}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\lesssim\frac{\mu^{\natural}r}{\sqrt{n_1n_2}}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}};
$$

and (iii) holds provided that $n_1\gtrsim\mu^{\natural}r$ and $n_2\gtrsim\mu^{\natural}r$.

Putting the above pieces together, we conclude that

$$
\begin{aligned}
\left\|\boldsymbol{E}_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}(\boldsymbol{M})\right]^{\top}\left(\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^{\natural}\right)\right\|_2&\le\alpha_1+\alpha_2\le\alpha_1+\beta_1+\beta_2+\beta_3\\
&\lesssim\zeta_{\mathsf{op}}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}-\boldsymbol{U}^{\natural}\right\|_{2,\infty}+\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}+\sqrt{\frac{\mu^{\natural}r}{n_1}}\frac{\zeta_{\mathsf{op}}^2}{\sigma_r^{\natural 2}}
\end{aligned}
$$

simultaneously for all $m\in[n_1]$, provided that $\zeta_{\mathsf{op}}\lesssim\sigma_r^{\natural 2}$.

F.3.12 Proof of Lemma 33

In view of the Davis-Kahan $\sin\boldsymbol{\Theta}$ Theorem (Chen et al., 2020c, Theorem 2.2.1), we have

$$
\begin{aligned}
\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}\boldsymbol{U}^{\top}\right\|&\le\frac{\left\|\left(\boldsymbol{G}-\boldsymbol{G}^{(m)}\right)\boldsymbol{U}^{(m)}\right\|}{\lambda_r\left(\boldsymbol{G}^{(m)}\right)-\lambda_{r+1}\left(\boldsymbol{G}\right)}\le\frac{2\left\|\left(\boldsymbol{G}-\boldsymbol{G}^{(m)}\right)\boldsymbol{U}^{(m)}\right\|}{\sigma_r^{\natural 2}}\\
&\le\underbrace{\frac{2\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}-\boldsymbol{G}^{(m)}\right)\boldsymbol{U}^{(m)}\right\|}{\sigma_r^{\natural 2}}}_{=:\alpha_1}+\underbrace{\frac{2\left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}-\boldsymbol{G}^{(m)}\right)\boldsymbol{U}^{(m)}\right\|}{\sigma_r^{\natural 2}}}_{=:\alpha_2},
\end{aligned}
$$

where the penultimate inequality follows from Weyl's inequality:

$$
\begin{aligned}
\lambda_r\left(\boldsymbol{G}^{(m)}\right)&\ge\lambda_r\left(\boldsymbol{G}^{\natural}\right)-\left\|\boldsymbol{G}^{(m)}-\boldsymbol{G}^{\natural}\right\|\overset{\text{(i)}}{\ge}\sigma_r^{\natural 2}-\widetilde{C}\zeta_{\mathsf{op}}\overset{\text{(ii)}}{\ge}\frac{3}{4}\sigma_r^{\natural 2},\\
\lambda_{r+1}\left(\boldsymbol{G}\right)&\le\lambda_{r+1}\left(\boldsymbol{G}^{\natural}\right)+\left\|\boldsymbol{G}-\boldsymbol{G}^{\natural}\right\|\overset{\text{(iii)}}{\le}\widetilde{C}\zeta_{\mathsf{op}}\overset{\text{(iv)}}{\le}\frac{1}{4}\sigma_r^{\natural 2},
\end{aligned}
$$

with $\widetilde{C}>0$ some absolute constant. Here, (i) comes from Lemma 29; (iii) comes from Lemma 23 and the fact that $\lambda_{r+1}(\boldsymbol{G}^{\natural})=0$; (ii) and (iv) are valid provided that $\zeta_{\mathsf{op}}\ll\sigma_r^{\natural 2}$.

Bounding α_1. Recalling that $\mathcal{P}_{\mathsf{diag}}(\boldsymbol{G}^{(m)}) = \mathcal{P}_{\mathsf{diag}}(\boldsymbol{G}^{\natural})$, we obtain

$$\alpha_1 = \frac{2\left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}-\boldsymbol{G}^{\natural}\right)\right\|\left\|\boldsymbol{U}^{(m)}\right\|}{\sigma_r^{\natural 2}} \lesssim \kappa^{\natural 2}\sqrt{\frac{\mu^{\natural} r}{n_1}}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}},$$

where the last inequality follows from Lemma 23.

Bounding α_2. Observe that the symmetric matrix $\mathcal{P}_{\mathsf{off\text{-}diag}}(\boldsymbol{G}-\boldsymbol{G}^{(m)})$ is supported on the m-th row and the m-th column, and

$$\left(\boldsymbol{G}-\boldsymbol{G}^{(m)}\right)_{m,i} = \left(\boldsymbol{G}-\boldsymbol{G}^{(m)}\right)_{i,m} = \boldsymbol{E}_{m,\cdot}\boldsymbol{M}_{i,\cdot}^{\top}, \qquad \forall i \in [n_1]. \tag{F.50}$$

Therefore, we can derive

$$\begin{aligned}
\left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}-\boldsymbol{G}^{(m)}\right)\boldsymbol{U}^{(m)}\right\| &\leq \left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}-\boldsymbol{G}^{(m)}\right)\boldsymbol{U}^{(m)}\right\|_{\mathrm{F}} \overset{\text{(i)}}{\leq} 2\left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}-\boldsymbol{G}^{(m)}\right)\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{\mathrm{F}} \\
&\leq 2\left\|\mathcal{P}_{m,\cdot}\left(\boldsymbol{G}-\boldsymbol{G}^{(m)}\right)\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{\mathrm{F}} + 2\left\|\mathcal{P}_{\cdot,m}\left(\boldsymbol{G}-\boldsymbol{G}^{(m)}\right)\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{\mathrm{F}} \\
&\overset{\text{(ii)}}{=} 2\underbrace{\left\|\boldsymbol{E}_{m,\cdot}\left[\mathcal{P}_{-m,\cdot}\left(\boldsymbol{M}\right)\right]^{\top}\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{\mathrm{F}}}_{=:\alpha_{2,1}} + 2\underbrace{\left\|\left(\boldsymbol{G}-\boldsymbol{G}^{(m)}\right)_{m,\cdot}\right\|_2\left\|\boldsymbol{U}_{m,\cdot}^{(m)}\boldsymbol{H}^{(m)}\right\|_2}_{=:\alpha_{2,2}}.
\end{aligned}$$

Here (i) follows from Lemma 24, and (ii) follows from (F.50).

- Regarding α_1, we can invoke (F.19) in Lemma 32 to achieve that with probability exceeding $1-O(n^{-11})$,

$$\alpha_{2,1} \lesssim \zeta_{\mathsf{op}}\left(\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} + \sqrt{\frac{\mu^{\natural} r}{n_1}}\right).$$

- With regards to α_2, we can invoke Lemma 29 to show that with probability exceeding $1-O(n^{-11})$,

$$\alpha_{2,2} \leq \left\|\boldsymbol{G}-\boldsymbol{G}^{(m)}\right\|_2\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} \leq \left(\sigma^2\sqrt{n_1 n_2}\log n + \kappa^{\natural 2}\sigma\sigma_1^{\natural}\sqrt{\mu^{\natural} r\log n}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}.$$

Taking the above bounds on $\alpha_{2,1}$ and $\alpha_{2,2}$ collectively yields

$$\begin{aligned}
\left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}-\boldsymbol{G}^{(m)}\right)\boldsymbol{U}^{(m)}\right\| &\lesssim \alpha_{2,1}+\alpha_{2,2} \\
&\lesssim \zeta_{\mathsf{op}}\left(\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} + \sqrt{\frac{\mu^{\natural} r}{n_1}}\right) + \left(\sigma^2\sqrt{n_1 n_2}\log n + \kappa^{\natural 2}\sigma\sigma_1^{\natural}\sqrt{\mu^{\natural} r\log n}\right)\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} \\
&\lesssim \zeta_{\mathsf{op}}\left(\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} + \sqrt{\frac{\mu^{\natural} r}{n_1}}\right),
\end{aligned}$$

provided that $n_1 \gtrsim \kappa^{\natural 4}\mu^{\natural} r$. Therefore, we arrive at

$$\alpha_2 \lesssim \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\left(\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty} + \sqrt{\frac{\mu^{\natural} r}{n_1}}\right).$$

Combining bounds on α_1 and α_2. The preceding bounds on α_1 and α_2 combined allow one to derive

$$\begin{aligned}\left\|\boldsymbol{U}^{(m)}\boldsymbol{U}^{(m)\top}-\boldsymbol{U}\boldsymbol{U}^{\top}\right\| &\le \alpha_1+\alpha_2 \lesssim \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\left(\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}+\sqrt{\frac{\mu^{\natural}r}{n_1}}\right)+\kappa^{\natural 2}\sqrt{\frac{\mu^{\natural}r}{n_1}}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}} \\ &\lesssim \frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\left\|\boldsymbol{U}^{(m)}\boldsymbol{H}^{(m)}\right\|_{2,\infty}+\kappa^{\natural 2}\sqrt{\frac{\mu^{\natural}r}{n_1}}\frac{\zeta_{\mathsf{op}}}{\sigma_r^{\natural 2}}\end{aligned}$$

with probability at least $1-O(n^{-11})$. This together with the union bound concludes the proof.

G Analysis for PCA: the approach based on HeteroPCA

This section is dedicated to proving our theoretical guarantees for the approach based on HeteroPCA — the ones presented in Section 3.3. Akin to Appendix D, we shall begin by studying a closely related subspace estimation model, and exploit the connection between this model and PCA in order to establish our main results. Similar to our analysis for the SVD-based approach (see Appendix D), it suffices to prove the theorems under the additional conditions

$$\max_{j\in[n]}|\eta_{l,j}| \le C_{\mathsf{noise}}\omega_l^{\star}\sqrt{\log(n+d)} \qquad \text{for all } l\in[d], \tag{G.1}$$

for some absolute constant $C_{\mathsf{noise}}>0$. See Appendix D for detailed reasoning.

G.1 Connection between PCA and subspace estimation

Our theoretical guarantees for subspace estimation (i.e., Theorem 10) provide a plausible path for us to analyze the performance of HeteroPCA for the PCA model. In order to do so, one needs to first make explicit the connection between the two models under consideration. By virtue of the close similarity between subspace estimation and matrix denoising (cf. Section 6.1), many of the useful connections have already been established previously in Appendix D.1. For instance:

- The definition of $\boldsymbol{M}$, $\boldsymbol{M}^{\natural}$ and $\boldsymbol{E}$ can be found in (D.2);
- The expressions of $\boldsymbol{V}^{\natural}$, $\boldsymbol{Q}$, $\boldsymbol{J}$ and $\boldsymbol{R}_{\boldsymbol{U}}$ can be found in (D.3), (D.4), (D.5) and (D.6), respectively;
- The entrywise variance of the zero-mean noise matrix $\boldsymbol{E}$ has been quantified in (D.8).

Given the slight difference in the incoherence condition (note that Assumption 4 is different from Assumption 2 for matrix denoising), we need to make slight modification of the bound on $\mu^{\natural}$ (cf. (D.12)) in Lemma 6 therein. For ease of reference, the following lemma summarizes the useful properties that assist in analyzing HeteroPCA.

Lemma 36. *There is an event $\mathcal{E}_{\mathsf{good}}$ with $\mathbb{P}(\mathcal{E}_{\mathsf{good}})\ge 1-O\big((n+d)^{-10}\big)$, on which the following properties hold.*

- *$\mathcal{E}_{\mathsf{good}}$ is $\sigma(\boldsymbol{F})$-measurable, where $\sigma(\boldsymbol{F})$ is the σ-algebra generated by $\boldsymbol{F}$.*
- *If $n\gg\kappa^2(r+\log(n+d))$, then one has*

$$\left\|\boldsymbol{\Sigma}^{\natural}-\boldsymbol{\Sigma}^{\star}\right\| \lesssim \kappa\sqrt{\frac{r+\log(n+d)}{n}}\sigma_r^{\star}, \tag{G.2}$$

$$\left\|\boldsymbol{\Sigma}^{\natural 2}-\boldsymbol{\Sigma}^{\star 2}\right\| \lesssim \sqrt{\frac{r+\log(n+d)}{n}}\sigma_1^{\star 2}, \tag{G.3}$$

$$\sigma_r^{\natural}\asymp\sigma_r^{\star} \qquad \text{and} \qquad \sigma_1^{\natural}\asymp\sigma_1^{\star}. \tag{G.4}$$

- *The conditional number $\kappa^{\natural}$ and the incoherence parameter $\mu^{\natural}$ of $\boldsymbol{M}^{\natural}$ obey*

$$\kappa^{\natural} \asymp \sqrt{\kappa}, \tag{G.5}$$

$$\mu^{\natural} \lesssim \kappa\mu \log(n+d). \tag{G.6}$$

In addition, we have the following $\ell_{2,\infty}$ norm bound for $\boldsymbol{U}^{\natural}$ and $\boldsymbol{V}^{\natural}$:

$$\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty} \leq \sqrt{\frac{\mu r}{d}} \qquad \text{and} \qquad \left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} \lesssim \sqrt{\frac{r\log(n+d)}{n}}. \tag{G.7}$$

- *The noise levels $\{\sigma_{i,j}\}$ are upper bounded by*

$$\max_{i\in[d],j\in[n]} \sigma_{i,j}^2 \lesssim \frac{\mu r \log(n+d)}{ndp}\sigma_1^{\star 2} + \frac{\omega_{\max}^2}{np} =: \sigma_{\mathsf{ub}}^2, \tag{G.8}$$

$$\begin{aligned}\max_{i\in[d],j\in[n]} |E_{i,j}| &\lesssim \max_{i\in[d]} \frac{1}{p}\sqrt{\frac{\log(n+d)}{n}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{\omega_l^{\star}}{p}\sqrt{\frac{\log(n+d)}{n}} \\ &\lesssim \frac{1}{p}\sqrt{\frac{\mu r\log(n+d)}{nd}}\sigma_1^{\star} + \frac{\omega_{\max}}{p}\sqrt{\frac{\log(n+d)}{n}} =: B.\end{aligned} \tag{G.9}$$

- *The matrices $\boldsymbol{J}$ and $\boldsymbol{Q}$ are close in the sense that*

$$\left\|\boldsymbol{Q}-\boldsymbol{J}\right\| \lesssim \frac{1}{\sigma_r^{\star}}\left\|\boldsymbol{Q}\boldsymbol{\Sigma}^{\natural}-\boldsymbol{\Sigma}^{\star}\boldsymbol{Q}\right\| \lesssim \kappa\sqrt{\frac{r+\log(n+d)}{n}}. \tag{G.10}$$

- *For each $i\in[d]$, one has*

$$\max_{j\in[n]}\left|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_j\right| \lesssim \left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\sqrt{\log(n+d)}; \tag{G.11}$$

for each $i,l\in[d]$, we have

$$\left|\frac{1}{n}\sum_{j=1}^{n}\left(\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_j\right)^2\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_j\right)^2 - \left(S_{l,l}^{\star}S_{i,i}^{\star}+2S_{i,l}^{\star 2}\right)\right| \lesssim \sqrt{\frac{\log(n+d)^3}{n}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 \tag{G.12a}$$

$$\text{and} \qquad \left|\frac{1}{n}\sum_{j=1}^{n}\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_j\right)^2 - S_{i,i}^{\star}\right| \lesssim \sqrt{\frac{\log(n+d)}{n}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2. \tag{G.12b}$$

Proof. The proofs of these properties are identical to the ones provided in Appendix E.5, except for that of (G.6). The relation (G.6) can be immediately justified by taking together (G.7) and the following inequality

$$\begin{aligned}\left\|\boldsymbol{M}^{\natural}\right\|_{\infty} &= \max_{i,j}\left|\boldsymbol{U}_{i,\cdot}^{\natural}\boldsymbol{\Sigma}^{\natural}\boldsymbol{V}_{j,\cdot}^{\natural\top}\right| \leq \sigma_1^{\natural}\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} \overset{\text{(i)}}{\lesssim} \sqrt{\frac{\mu\log(n+d)}{nd}}\sigma_1^{\natural}\sqrt{r} \\ &\lesssim \sqrt{\frac{\mu\log(n+d)}{nd}}\kappa^{\natural}\left\|\boldsymbol{M}^{\natural}\right\|_{\mathrm{F}} \overset{\text{(ii)}}{\asymp} \sqrt{\frac{\kappa\mu\log(n+d)}{nd}}\left\|\boldsymbol{M}^{\natural}\right\|_{\mathrm{F}}.\end{aligned}$$

Here, (i) follows from (G.7), whereas (ii) arises from (G.5). □

G.2 Distributional characterization for principal subspace (Proof of Theorem 15)

With the above connection between the subspace estimation model and PCA in place, we can readily move on to invoke Theorem 10 to establish our distributional characterization of $\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}$ for $\mathsf{HeteroPCA}$ (as stated in Theorem 15).

Step 1: first- and second-order approximation and the tightness. As a starting point, Theorem 9 taken together with the explicit connection between the subspace estimation model and PCA allows one to approximate $\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}$ in a concise form, which is a crucial first step that enables our subsequent development of the distributional theory. The proof can be found in Appendix H.1.1.

Lemma 37. *Assume that $d \gtrsim \kappa^3\mu^2 r\log^4(n+d)$ $n \gtrsim r\log^4(n+d)$, $ndp^2 \gg \kappa^4\mu^2r^2\log^4(n+d)$, $np \gg \kappa^4\mu r\log^2(n+d)$,*

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \ll \frac{1}{\kappa\log(n+d)} \qquad \text{and} \qquad \frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}} \ll \frac{1}{\sqrt{\kappa^3\log(n+d)}}.$$

Then we have

$$\mathbb{P}\left(\left\|\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}-\boldsymbol{Z}\right\|_{2,\infty}\mathbb{1}_{\mathcal{E}_{\mathsf{good}}} \lesssim \zeta_{\mathsf{2nd}}\,\middle|\,\boldsymbol{F}\right) \geq 1-O\left((n+d)^{-10}\right)$$

almost surely, where

$$\boldsymbol{Z} \coloneqq \left[\boldsymbol{E}\boldsymbol{M}^{\natural\top}+\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^{\top}\right)\right]\boldsymbol{U}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-2}\boldsymbol{Q}^{\top} \tag{G.13}$$

and

$$\zeta_{\mathsf{2nd}} \coloneqq \frac{\kappa^2\mu r\log(n+d)}{d}\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}+\frac{\zeta_{\mathsf{1st}}^2}{\sigma_r^{\star 4}}\sqrt{\frac{\kappa^3\mu r\log(n+d)}{d}}, \tag{G.14a}$$

$$\zeta_{\mathsf{1st}} \coloneqq \frac{\mu r\log^2(n+d)}{\sqrt{ndp}}\sigma_1^{\star 2}+\frac{\omega_{\max}^2}{p}\sqrt{\frac{d}{n}}\log(n+d)+\sigma_1^{\star 2}\sqrt{\frac{\mu r}{np}}\log(n+d)+\sigma_1^{\star}\omega_{\max}\sqrt{\frac{d\log(n+d)}{np}}. \tag{G.14b}$$

In particular, we expect ζ_{2nd} to be negligible, so that the approximation $\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star} \approx +\boldsymbol{Z}$ is nearly tight. It is worth noting that the approximation $\boldsymbol{Z}$ consists of both linear and second-order effects of the perturbation matrix $\boldsymbol{E}$.

Step 2: computing the covariance of the first- and second-order approximation. In order to pin down the distribution of $\boldsymbol{Z}$, an important step lies in characterizing its covariance. To be precise, observe that the l-th row of $\boldsymbol{Z}$ $(1 \leq l \leq d)$ defined in (G.13) satisfies

$$\begin{aligned}
\boldsymbol{Z}_{l,\cdot} &= \boldsymbol{e}_l^{\top}\left[\boldsymbol{E}\boldsymbol{M}^{\natural\top}+\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^{\top}\right)\right]\boldsymbol{U}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-2}\boldsymbol{Q}^{\top} \\
&= \boldsymbol{E}_{l,\cdot}\boldsymbol{V}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}+\boldsymbol{E}_{l,\cdot}\left[\mathcal{P}_{-l,\cdot}\left(\boldsymbol{E}\right)\right]^{\top}\boldsymbol{U}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-2}\boldsymbol{Q}^{\top} \\
&= \sum_{j=1}^{n}E_{l,j}\left[\boldsymbol{V}_{j,\cdot}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}+\left[\mathcal{P}_{-l,\cdot}\left(\boldsymbol{E}_{\cdot,j}\right)\right]^{\top}\boldsymbol{U}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-2}\right]\boldsymbol{Q}^{\top},
\end{aligned} \tag{G.15}$$

where we recall that $\mathcal{P}_{-l,\cdot}(\boldsymbol{E})$ is obtained by zeroing out the l-th row of $\boldsymbol{E}$. Conditional on $\boldsymbol{F}$, the covariance matrix of this zero-mean random vector $\boldsymbol{Z}_{l,\cdot}$ can thus be calculated as

$$\begin{aligned}
\widetilde{\boldsymbol{\Sigma}}_l \coloneqq\ & \boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{V}^{\natural\top}\mathsf{diag}\left\{\sigma_{l,1}^2,\dots,\sigma_{l,n}^2\right\}\boldsymbol{V}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top} \\
&+\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-2}\boldsymbol{U}^{\natural\top}\mathsf{diag}\left\{\sum_{j:j\neq l}\sigma_{l,j}^2\sigma_{1,j}^2,\dots,\sum_{j:j\neq l}\sigma_{l,j}^2\sigma_{d,j}^2\right\}\boldsymbol{U}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-2}\boldsymbol{Q}^{\top}.
\end{aligned} \tag{G.16}$$

Given that the above expression of $\widetilde{\boldsymbol{\Sigma}}_l$ contains components like $\boldsymbol{\Sigma}^{\natural}$ and $\boldsymbol{V}^{\natural}$ (which are introduced in order to use the subspace estimation model), it is natural to see whether one can express $\widetilde{\boldsymbol{\Sigma}}_l$ directly in terms of the corresponding quantities introduced for the PCA model. To do so, we first single out a deterministic matrix as follows

$$\boldsymbol{\Sigma}_{U,l}^{\star} \coloneqq \left(\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2+\frac{\omega_l^{\star 2}}{np}\right)\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}+\frac{2(1-p)}{np}\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^{\star}+\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{U}^{\star\top}\mathsf{diag}\left\{\left[d_{l,i}^{\star}\right]_{i=1}^{d}\right\}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}, \tag{G.17}$$

where we define

$$d_{l,i}^{\star} \coloneqq \frac{1}{np^{2}}\left[\omega_{l}^{\star2}+(1-p)\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\right]\left[\omega_{i}^{\star2}+(1-p)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\right]+\frac{2\left(1-p\right)^{2}}{np^{2}}S_{i,l}^{\star2}. \tag{G.18}$$

In view of the following lemma, $\boldsymbol{\Sigma}_{U,l}^{\star}$ approximates $\widetilde{\boldsymbol{\Sigma}}_{l}$ in a reasonably well fashion.

Lemma 38. *Suppose that $n \gg \kappa^{8}\mu^{2}r^{3}\log^{3}(n+d)$. On the event $\mathcal{E}_{\mathsf{good}}$ (cf. Lemma 36) , we have*

$$\left\|\widetilde{\boldsymbol{\Sigma}}_{l}-\boldsymbol{\Sigma}_{U,l}^{\star}\right\| \lesssim \sqrt{\frac{\kappa^{8}\mu^{2}r^{3}\log^{3}\left(n+d\right)}{n}}\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),$$

$$\max\left\{\lambda_{\max}\left(\widetilde{\boldsymbol{\Sigma}}_{l}\right),\lambda_{\max}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)\right\} \lesssim \frac{1-p}{np\sigma_{r}^{\star2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}+\frac{\omega_{l}^{\star2}}{np\sigma_{r}^{\star2}}+\frac{\kappa\mu r\left(1-p\right)^{2}}{ndp^{2}\sigma_{r}^{\star2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}+\frac{\kappa\mu r\left(1-p\right)}{ndp^{2}\sigma_{r}^{\star2}}\omega_{\max}^{2}+\frac{\omega_{l}^{\star2}\omega_{\max}^{2}}{np^{2}\sigma_{r}^{\star4}},$$

$$\min\left\{\lambda_{\min}\left(\widetilde{\boldsymbol{\Sigma}}_{l}\right),\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)\right\} \gtrsim \frac{1-p}{np\sigma_{1}^{\star2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}+\frac{\omega_{l}^{\star2}}{np\sigma_{1}^{\star2}}+\frac{\left(1-p\right)^{2}}{ndp^{2}\kappa\sigma_{1}^{\star2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}+\frac{1-p}{ndp^{2}\kappa\sigma_{1}^{\star2}}\omega_{l}^{\star2}+\frac{\omega_{l}^{\star2}\omega_{\min}^{2}}{np^{2}\sigma_{1}^{\star4}}.$$

In addition, the condition number of $\widetilde{\boldsymbol{\Sigma}}_{l}$ is bounded above by $O(\kappa^{3}\mu r)$.

Proof. See Appendix H.1.2. □

Step 3: establishing distributional guarantees for $(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star})_{l,\cdot}$. By virtue of the decomposition (G.15), each row $\boldsymbol{Z}_{l,\cdot}$ can be viewed as the sum of a collection of independent zero-mean random variables/vectors. This suggests that $\boldsymbol{Z}_{l,\cdot}$ might be well approximated by certain multivariate Gaussian distributions. If so, then the zero-mean nature of $\boldsymbol{Z}_{l,\cdot}$ in conjunction with the above-mentioned covariance formula allows us to pin down the distribution of each row of $\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}$ approximately.

Lemma 39 (Gaussian approximation of $\boldsymbol{Z}_{l,\cdot}$). *Suppose that the following conditions hold:*

$$n \gtrsim \kappa^{8}\mu^{2}r^{4}\log^{4}(n+d), \qquad d \gtrsim \kappa^{6}\mu^{2}r^{5/2}\log^{5}(n+d),$$

$$np \gtrsim \kappa^{8}\mu^{3}r^{11/2}\log^{5}\left(n+d\right), \qquad ndp^{2} \gtrsim \kappa^{8}\mu^{4}r^{13/2}\log^{6}\left(n+d\right),$$

$$\frac{\omega_{\max}}{\sigma_{r}^{\star}}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\kappa^{3}\mu^{1/2}r^{3/4}\log^{2}\left(n+d\right)}, \qquad \frac{\omega_{\max}^{2}}{p\sigma_{r}^{\star2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\kappa^{4}\mu^{3/2}r^{11/4}\log^{5/2}\left(n+d\right)},$$

and

$$\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2} \gtrsim \left[\frac{\kappa^{4}\mu^{5/2}r^{9/4}\log^{5}\left(n+d\right)}{\sqrt{ndp^{2}}}+\frac{\kappa^{7/2}\mu^{3/2}r^{5/4}\log^{3}\left(n+d\right)}{\sqrt{np}}\sigma_{1}^{\star}+\frac{\kappa^{7/2}\mu^{2}r^{7/4}\log^{7/2}\left(n+d\right)}{\sqrt{d}}\right]\sqrt{\frac{r}{d}}\sigma_{1}^{\star}.$$

Then it is guaranteed that

$$\sup_{\mathcal{C}\in\mathscr{C}^{r}}\left|\mathbb{P}\left(\left(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}\right)_{l,\cdot}\in\mathcal{C}\right)-\mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}_{U,l}^{\star}\right)\{\mathcal{C}\}\right| \lesssim \frac{1}{\sqrt{\log\left(n+d\right)}}=o\left(1\right),$$

where $\mathscr{C}^{r}$ denotes the set of all convex sets in $\mathbb{R}^{r}$.

Proof. See Appendix H.1.3. □

Once Lemma 39 is established, we have solidified the advertised Gaussian approximation of each row of $\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}$, thus concluding the proof of Theorem 15.

G.3 Validity of confidence regions (Proof of Theorem 16)

Our distributional characterization of $(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star})_{l,\cdot}$ hints at the possibility of constructing valid confidence region for $\boldsymbol{U}^{\star}$, provided that the covariance matrix $\boldsymbol{\Sigma}_{U,l}^{\star}$ (cf. (G.17)) can be reliably estimated. Similar to the SVD-based approach, we attempt to estimate $\boldsymbol{\Sigma}_{U,l}^{\star}$ by means of the following plug-in estimator:

$$\boldsymbol{\Sigma}_{U,l} \coloneqq \left(\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\right\|_2^2+\frac{\omega_l^2}{np}\right)\boldsymbol{\Sigma}^{-2}+\frac{2\left(1-p\right)}{np}\boldsymbol{U}_{l,\cdot}^{\top}\boldsymbol{U}_{l,\cdot}+\left(\boldsymbol{\Sigma}\right)^{-2}\boldsymbol{U}^{\top}\mathsf{diag}\left\{\left[d_{l,i}\right]_{1\le i\le d}\right\}\boldsymbol{U}\left(\boldsymbol{\Sigma}\right)^{-2}, \tag{G.19}$$

where for each $i\in[d]$, we define

$$d_{l,i} \coloneqq \frac{1}{np^2}\left[\omega_l^2+\left(1-p\right)\left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\right\|_2^2\right]\left[\omega_i^2+\left(1-p\right)\left\|\boldsymbol{U}_{i,\cdot}\boldsymbol{\Sigma}\right\|_2^2\right]+\frac{2\left(1-p\right)^2}{np^2}S_{l,i}^2, \tag{G.20}$$

$$\omega_i^2 \coloneqq \frac{\sum_{j=1}^n y_{i,j}^2\,\mathbb{1}_{(i,j)\in\Omega}}{\sum_{j=1}^n \mathbb{1}_{(i,j)\in\Omega}}-S_{i,i}. \tag{G.21}$$

Step 1: fine-grained estimation guarantees for $\boldsymbol{U}$ and $\boldsymbol{\Sigma}$. To begin with, we need to show that the components in the plug-in estimator are all reliable estimates of their deterministic counterpart (after proper rotation).

Lemma 40. *Recall the definition of ζ_{1st} and ζ_{2nd} in (G.14). Assume that $\zeta_{\mathsf{1st}}/\sigma_r^{\star2}\lesssim 1/\sqrt{\kappa^3\mu}$, $n\gg r+\log(n+d)$ and $d\gtrsim\kappa^4\mu^2 r\log(n+d)$. Let*

$$\theta \coloneqq \sqrt{\frac{\kappa r\log^2\left(n+d\right)}{np}}\left(1+\frac{\sigma_{\mathsf{ub}}}{\sigma_r^{\star}}\sqrt{n}\right) \tag{G.22}$$

Then with probability exceeding $1-O((n+d)^{-10})$, we have

$$\left\|\left(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}\right)_{l,\cdot}\right\|_2 \lesssim \frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_l^{\star}\right)+\zeta_{\mathsf{2nd}}, \tag{G.23a}$$

$$\left\|\left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{R}-\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\right)_{l,\cdot}\right\|_2 \lesssim \theta\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_l^{\star}\right)+\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2\sqrt{\frac{\kappa^2\left(r+\log\left(n+d\right)\right)}{n}}\sigma_1^{\star}+\zeta_{\mathsf{2nd}}\sigma_1^{\star}, \tag{G.23b}$$

$$\left\|\boldsymbol{R}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{R}^{\top}-\boldsymbol{\Sigma}^{-2}\right\| \lesssim \sqrt{\frac{d}{\kappa\mu r\log\left(n+d\right)}}\frac{\zeta_{\mathsf{2nd}}}{\sigma_r^{\star2}}+\sqrt{\frac{\kappa^3\left(r+\log\left(n+d\right)\right)}{n}}\frac{1}{\sigma_r^{\star2}}, \tag{G.23c}$$

$$\left\|\boldsymbol{U}\boldsymbol{\Sigma}^{-2}\boldsymbol{R}-\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\right\| \lesssim \frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star4}}+\sqrt{\frac{\kappa^3\left(r+\log\left(n+d\right)\right)}{n}}\frac{1}{\sigma_r^{\star2}}, \tag{G.23d}$$

$$\left\|\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}\right\|_{2,\infty} \lesssim \frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star2}}\sqrt{\frac{r\log\left(n+d\right)}{d}}. \tag{G.23e}$$

Proof. See Appendix H.2.1. ☐

Step 2: faithfulness of the plug-in estimator. With Lemma 40 in hand, we move forward to show that $\boldsymbol{S}$ and $\{\omega_i^2\}_{i=1}^d$ are reliable estimators of the covariance matrix $\boldsymbol{S}^{\star}$ and the noise levels $\{\omega_i^{\star2}\}_{i=1}^d$, respectively.

Lemma 41. *Instate the conditions in Lemma 40. In addition, assume that $n\gtrsim\kappa^3 r\log(n+d)$, $\zeta_{\mathsf{1st}}/\sigma_r^{\star2}\lesssim 1/\sqrt{\log(n+d)}$. Then with probability exceeding $1-O((n+d)^{-10})$, we have*

$$\left\|\boldsymbol{S}-\boldsymbol{S}^{\star}\right\|_{\infty} \lesssim \left(\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star}}\sqrt{\frac{\kappa r\log\left(n+d\right)}{d}}+\sqrt{\frac{\kappa^2\mu r^2\log\left(n+d\right)}{nd}}\sigma_1^{\star}\right)\sqrt{\frac{\mu r}{d}}\sigma_1^{\star} \tag{G.24}$$

and, for each $i,j\in[d]$,

$$\left|S_{i,j}-S_{i,j}^{\star}\right| \lesssim \left(\theta+\sqrt{\frac{\kappa^3 r\log\left(n+d\right)}{n}}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\theta^2\omega_{\max}^2$$

$$+ \left(\theta\omega_{\max} + \zeta_{\mathsf{2nd}}\sigma_1^\star\right)\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\right) + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}. \tag{G.25}$$

Here, the quantity θ is defined in (G.22). In addition, with probability exceeding $1 - O((n+d)^{-10})$, we have

$$\left|\omega_i^2 - \omega_i^{\star 2}\right| \lesssim \sqrt{\frac{\log^2(n+d)}{np}}\omega_i^{\star 2} + \zeta_{\mathsf{1st}}\frac{\sqrt{\kappa^2\mu r^2\log(n+d)}}{d} + \sqrt{\frac{\kappa^2\mu^2 r^3\log(n+d)}{nd^2}}\sigma_1^{\star 2} \tag{G.26}$$

for all $i \in [d]$, and

$$\left|\omega_l^2 - \omega_l^{\star 2}\right| \lesssim \left(\sqrt{\frac{\log^2(n+d)}{np}} + \theta^2\right)\omega_l^{\star 2} + \left(\theta + \sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2$$
$$+ \left(\theta\omega_{\max} + \zeta_{\mathsf{2nd}}\sigma_1^\star\right)\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}. \tag{G.27}$$

Proof. See Appendix H.2.2. □

The above two lemmas taken together allow us to demonstrate that our plug-in estimator $\boldsymbol{\Sigma}_{U,l}$ is a faithful estimate of $\boldsymbol{\Sigma}_{U,l}^\star$, as stated below.

Lemma 42. *Instate the conditions in Lemma 39, Lemma 40 and Lemma 42. Consider any $\delta \in (0,1)$, and we further suppose that $n \gtrsim \delta^{-2}\kappa^9\mu^2 r^3\log(n+d)$, $d \gtrsim \kappa^4\mu r\log(n+d)$,*

$$ndp^2 \gtrsim \delta^{-2}\kappa^8\mu^4 r^4\log^5(n+d), \qquad np \gtrsim \delta^{-2}\kappa^8\mu^3 r^3\log^3(n+d),$$

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{\delta}{\kappa^3\mu r\log^{3/2}(n+d)} \qquad \text{and} \qquad \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d}{np}} \lesssim \frac{\delta}{\kappa^{7/2}\mu r\log(n+d)}.$$

Then with probability exceeding $1 - O((n+d)^{-10})$, we have

$$\left\|\boldsymbol{\Sigma}_{U,l} - \boldsymbol{R}\boldsymbol{\Sigma}_{U,l}^\star\boldsymbol{R}^\top\right\| \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right).$$

Proof. See Appendix H.2.3. □

Step 3: validity of the constructed confidence regions. Finally, we can combine the Gaussian approximation of $\boldsymbol{U}_{l,\cdot}\boldsymbol{R} - \boldsymbol{U}^\star$ established in Theorem 15 and the estimation guarantee of the covariance matrix established in Lemma 42 to justify the validity of the constructed confidence region $\mathsf{CR}_{U,l}^{1-\alpha}$ (cf. Algorithm 5).

Lemma 43. *Instate the conditions in Theorem 15. Suppose that $n \gtrsim \kappa^{12}\mu^3 r^{11/2}\log^5(n+d)$, $d \gtrsim \kappa^4\mu r\log(n+d)$,*

$$ndp^2 \gtrsim \kappa^{11}\mu^5 r^{13/2}\log^9(n+d), \qquad np \gtrsim \kappa^{11}\mu^4 r^{11/2}\log^7(n+d),$$

and

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\kappa^{9/2}\mu^{3/2}r^{9/4}\log^{7/2}(n+d)}, \qquad \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\kappa^5\mu^{3/2}r^{9/4}\log^3(n+d)}.$$

Then it holds that

$$\mathbb{P}\left(\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{R}^\top \in \mathsf{CR}_{U,l}^{1-\alpha}\right) = 1 - \alpha + O\left(\frac{1}{\sqrt{\log(n+d)}}\right) = 1 - \alpha + o(1).$$

Proof. See Appendix H.2.4. □

With this lemma in place, we have concluded the proof of Theorem 16.

G.4 Entrywise distributional characterization for $\boldsymbol{S}^\star$ (Proof of Theorem 17)

Based on our distributional theory for principal subspace, we now turn attention to statistical inference for the spiked covariance matrix $\boldsymbol{S}^\star$ based on the estimate $\boldsymbol{S}$ returned by Algorithm 6. Similar to the SVD-based approach, we first decompose

$$\boldsymbol{S}-\boldsymbol{S}^\star=\underbrace{\boldsymbol{S}-\boldsymbol{M}^\natural\boldsymbol{M}^{\natural\top}}_{=:\boldsymbol{W}}+\underbrace{\boldsymbol{M}^\natural\boldsymbol{M}^{\natural\top}-\boldsymbol{S}^\star}_{=:\boldsymbol{A}}.$$

Then we prove Theorem 17 by the following steps:

1. Show that conditional on $\boldsymbol{F}$, on the $\sigma(\boldsymbol{F})$-measurable high probability event $\mathcal{E}_{\mathsf{good}}$, each entry $W_{i,j}$ is approximately Gaussian with mean zero and variance concentrating around some deterministic quantity $\widetilde{v}_{i,j}$.
2. Show that each $A_{i,j}$ is approximately a zero-mean Gaussian with variance $\overline{v}_{i,j}$.
3. Conclude that $S_{i,j}-S^\star_{i,j}$ is approximately a Gaussian with mean zero and variance $\widetilde{v}_{i,j}+\overline{v}_{i,j}$, which is established using the (near) independence of $W_{i,j}$ (whose distribution is nearly independent of $\boldsymbol{F}$) and $A_{i,j}$ (which is $\sigma(\boldsymbol{F})$-measurable).

Step 1: first- and second-order approximation of $\boldsymbol{W}$. We first utilize Lemma 37 to obtain an explicit entrywise first- and second-order approximation of $\boldsymbol{W}$. This approximation is crucial for our entrywise distributional characterization.

Lemma 44. *Instate the assumptions in Lemma 37. Then one can write*

$$\boldsymbol{W}=\boldsymbol{S}-\boldsymbol{M}^\natural\boldsymbol{M}^{\natural\top}=\boldsymbol{X}+\boldsymbol{\Phi},$$

where

$$\boldsymbol{X}\coloneqq\boldsymbol{E}\boldsymbol{M}^{\natural\top}+\boldsymbol{M}^\natural\boldsymbol{E}^\top+\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\boldsymbol{U}^\natural\boldsymbol{U}^{\natural\top}+\boldsymbol{U}^\natural\boldsymbol{U}^{\natural\top}\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\tag{G.28}$$

and the residual matrix $\boldsymbol{\Phi}$ satisfies: conditional on $\boldsymbol{F}$ and on the $\sigma(\boldsymbol{F})$-measurable event $\mathcal{E}_{\mathsf{good}}$ (see Lemma 36),

$$|\Phi_{i,j}|\lesssim\underbrace{\left(\zeta_{\mathsf{2nd}}\sigma_1^{\star2}+\theta^2\omega_{\max}\sigma_1^\star\right)\left(\left\|\boldsymbol{U}^\star_{i,\cdot}\right\|_2+\left\|\boldsymbol{U}^\star_{j,\cdot}\right\|_2\right)+\theta^2\left\|\boldsymbol{U}^\star_{i,\cdot}\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}^\star_{j,\cdot}\boldsymbol{\Sigma}^\star\right\|_2+\theta^2\omega_{\max}^2+\zeta_{\mathsf{2nd}}^2\sigma_1^{\star2}}_{=:\zeta_{i,j}}\tag{G.29}$$

holds for any $i,j\in[d]$ with probability exceeding $1-O((n+d)^{-10})$. Here, ζ_{2nd} is defined in Lemma 37, and θ is defined in (G.22).

Proof. See Appendix H.3.1. ∎

Step 2: computing the entrywise variance of our approximation. We can check that the (i,j)-th entry of the matrix $\boldsymbol{X}$ (cf. (G.28)) is given by

$$\begin{aligned}X_{i,j}&=\left[\boldsymbol{E}\boldsymbol{M}^{\natural\top}+\boldsymbol{M}^\natural\boldsymbol{E}^\top+\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\boldsymbol{U}^\natural\boldsymbol{U}^{\natural\top}+\boldsymbol{U}^\natural\boldsymbol{U}^{\natural\top}\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]_{i,j}\\&=\sum_{l=1}^n\left\{M^\natural_{j,l}E_{i,l}+M^\natural_{i,l}E_{j,l}+E_{i,l}\left[\mathcal{P}_{-i,\cdot}\left(\boldsymbol{E}_{\cdot,l}\right)\right]^\top\boldsymbol{U}^\star\left(\boldsymbol{U}^\star_{j,\cdot}\right)^\top+E_{j,l}\left[\mathcal{P}_{-j,\cdot}\left(\boldsymbol{E}_{\cdot,l}\right)\right]^\top\boldsymbol{U}^\star\left(\boldsymbol{U}^\star_{i,\cdot}\right)^\top\right\}\\&=\sum_{l=1}^n\left[M^\natural_{j,l}E_{i,l}+M^\natural_{i,l}E_{j,l}+\sum_{k:k\neq i}E_{i,l}E_{k,l}\left(\boldsymbol{U}^\star_{k,\cdot}\left(\boldsymbol{U}^\star_{j,\cdot}\right)^\top\right)+\sum_{k:k\neq j}E_{j,l}E_{k,l}\left(\boldsymbol{U}^\star_{k,\cdot}\left(\boldsymbol{U}^\star_{i,\cdot}\right)^\top\right)\right].\end{aligned}\tag{G.30}$$

Here, the penultimate step uses the fact that $\boldsymbol{U}^\natural=\boldsymbol{U}^\star\boldsymbol{Q}$ for some orthonormal matrix $\boldsymbol{Q}$ (see (D.4)). This allows one to calculate the variance of $X_{i,j}$ conditional on $\boldsymbol{F}$: when $i\neq j$,

$$\mathsf{var}\left(X_{i,j}|\boldsymbol{F}\right)=\sum_{l=1}^nM_{j,l}^{\natural2}\sigma_{i,l}^2+\sum_{l=1}^nM_{i,l}^{\natural2}\sigma_{j,l}^2+\sum_{l=1}^n\sum_{k:k\neq i}\sigma_{i,l}^2\sigma_{k,l}^2\left(\boldsymbol{U}^\star_{k,\cdot}\boldsymbol{U}^{\star\top}_{j,\cdot}\right)^2+\sum_{l=1}^n\sum_{k:k\neq j}\sigma_{j,l}^2\sigma_{k,l}^2\left(\boldsymbol{U}^\star_{k,\cdot}\boldsymbol{U}^{\star\top}_{i,\cdot}\right)^2;\tag{G.31}$$

when $i=j$, we have

$$\begin{aligned}
\mathsf{var}\left(X_{i,i}|\boldsymbol{F}\right) &= 4\sum_{l=1}^{n} M_{i,l}^{\natural 2}\sigma_{i,l}^{2} + 4\sum_{l=1}^{n}\sum_{k:k\neq i}\sigma_{i,l}^{2}\sigma_{k,l}^{2}\left(\boldsymbol{U}_{k,\cdot}^{\natural}\boldsymbol{U}_{i,\cdot}^{\natural\top}\right)^{2} \\
&= 4\sum_{l=1}^{n} M_{i,l}^{\natural 2}\sigma_{i,l}^{2} + 4\sum_{l=1}^{n}\sum_{k:k\neq i}\sigma_{i,l}^{2}\sigma_{k,l}^{2}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{i,\cdot}^{\star\top}\right)^{2}.
\end{aligned}$$

The next lemma states that $\mathsf{var}(X_{i,j}|\boldsymbol{F})$ concentrates around some deterministic quantity $\widetilde{v}_{i,j}$ defined as follows:

$$\begin{aligned}
\widetilde{v}_{i,j} &\coloneqq \frac{2\left(1-p\right)}{np}\left(S_{i,i}^{\star}S_{j,j}^{\star}+2S_{i,j}^{\star 2}\right)+\frac{1}{np}\left(\omega_{i}^{\star 2}S_{j,j}^{\star}+\omega_{j}^{\star 2}S_{i,i}^{\star}\right) \\
&\quad+\frac{1}{np^{2}}\sum_{k=1}^{d}\left\{\left[\omega_{i}^{\star 2}+\left(1-p\right)S_{i,i}^{\star}\right]\left[\omega_{k}^{\star 2}+\left(1-p\right)S_{k,k}^{\star}\right]+2\left(1-p\right)^{2}S_{i,k}^{\star 2}\right\}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^{2} \\
&\quad+\frac{1}{np^{2}}\sum_{k=1}^{d}\left\{\left[\omega_{j}^{\star 2}+\left(1-p\right)S_{j,j}^{\star}\right]\left[\omega_{k}^{\star 2}+\left(1-p\right)S_{k,k}^{\star}\right]+2\left(1-p\right)^{2}S_{j,k}^{\star 2}\right\}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{i,\cdot}^{\star\top}\right)^{2}
\end{aligned}\tag{G.32a}$$

for any $i\neq j$, and

$$\begin{aligned}
\widetilde{v}_{i,i} &\coloneqq \frac{12\left(1-p\right)}{np}S_{i,i}^{\star 2}+\frac{4}{np}\omega_{i}^{\star 2}S_{i,i}^{\star} \\
&\quad+\frac{4}{np^{2}}\sum_{k=1}^{d}\left\{\left[\omega_{i}^{\star 2}+\left(1-p\right)S_{i,i}^{\star}\right]\left[\omega_{k}^{\star 2}+\left(1-p\right)S_{k,k}^{\star}\right]+2\left(1-p\right)^{2}S_{i,k}^{\star 2}\right\}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{i,\cdot}^{\star\top}\right)^{2}
\end{aligned}\tag{G.32b}$$

for any $i\in[d]$.

Lemma 45. *Suppose that $n\gg\log^{3}(n+d)$, and recall the definition of $\widetilde{v}_{i,j}$ in (G.32). On the event $\mathcal{E}_{\mathsf{good}}$ (cf. Lemma 36), we have*

$$\mathsf{var}\left(X_{i,j}|\boldsymbol{F}\right)=\widetilde{v}_{i,j}+O\left(\sqrt{\frac{\log^{3}\left(n+d\right)}{n}}+\frac{\kappa\mu^{2}r^{2}}{d}\right)\widetilde{v}_{i,j}\tag{G.33}$$

for any $i,j\in[d]$. In addition, for any $i,j\in[d]$ it holds that

$$\widetilde{v}_{i,j}\gtrsim\frac{1}{ndp^{2}\kappa\wedge np}\left\Vert\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert_{2}^{2}\left\Vert\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert_{2}^{2}+\left(\frac{\omega_{\min}^{2}\sigma_{r}^{\star 2}}{ndp^{2}\wedge np}+\frac{\omega_{\min}^{4}}{np^{2}}\right)\left(\left\Vert\boldsymbol{U}_{i,\cdot}^{\star}\right\Vert_{2}^{2}+\left\Vert\boldsymbol{U}_{j,\cdot}^{\star}\right\Vert_{2}^{2}\right).\tag{G.34}$$

Proof. See Appendix H.3.2. □

Step 3: establishing approximate Gaussianity of $W_{i,j}$, $A_{i,j}$ and $S_{i,j}-S_{i,j}^{\star}$. Notice that $X_{i,j}$ can be viewed as a sum of independent random variables. In the next lemma, we invoke the Berry-Esseen Theorem to show that $X_{i,j}$ is approximately Gaussian with mean zero and variance $\widetilde{v}_{i,j}$, which further leads to distributional characterizations of $W_{i,j}$.

Lemma 46. *Suppose that $d\gtrsim\kappa^{7}\mu^{2}r^{2}\log^{5}(n+d)$, $np\gtrsim\kappa^{3}r^{2}\log^{5}(n+d)$, $ndp^{2}\gtrsim\kappa^{5}\mu^{2}r^{4}\log^{7}(n+d)$,*

$$\frac{\omega_{\max}}{\sigma_{r}^{\star}}\sqrt{\frac{d}{np}}\lesssim\frac{1}{\kappa^{7/2}\mu^{1/2}r^{1/2}\log^{2}\left(n+d\right)},\qquad\frac{\omega_{\max}^{2}}{p\sigma_{r}^{\star 2}}\sqrt{\frac{d}{n}}\lesssim\frac{1}{\kappa^{3}\mu^{1/2}r\log^{3}\left(n+d\right)},$$

$$\min\left\{\left\Vert\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert_{2},\left\Vert\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert_{2}\right\}\gtrsim\left[\frac{\kappa^{9/2}\mu^{5/2}r^{2}\log^{5}\left(n+d\right)}{\sqrt{ndp^{2}}}+\frac{\kappa^{4}\mu^{3/2}r\log^{3}\left(n+d\right)}{\sqrt{np}}+\frac{\kappa^{4}\mu^{2}r^{3/2}\log^{7/2}\left(n+d\right)}{\sqrt{d}}\right]\sqrt{\frac{r}{d}}\sigma_{1}^{\star},$$

and

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \gtrsim \kappa r^{1/2}\log^{5/2}(n+d)\left(\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}}+\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}\right)\sqrt{\frac{r}{d}}.$$

Then it holds that

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\left(\widetilde{v}_{i,j}\right)^{-1/2}W_{i,j}\le t \,\middle|\, \boldsymbol{F}\right)-\Phi(t)\right|\lesssim\frac{1}{\sqrt{\log(n+d)}}=o(1).$$

Proof. See Appendix H.3.3. □

Recall that we have already shown in Lemma 17 that $A_{i,j}$ is approximately $\mathcal{N}(0,\overline{v}_{i,j})$ with $\overline{v}_{i,j}$ defined as

$$\overline{v}_{i,j}\coloneqq\frac{1}{n}\left[\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2+\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star 2}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2\right]=\frac{1}{n}\left(S_{i,i}^{\star}S_{j,j}^{\star}+S_{i,j}^{\star 2}\right). \tag{G.35}$$

For completeness we repeat the result in the following lemma.

Lemma 47. *It holds that*

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}\left(\left(\overline{v}_{i,j}\right)^{-1/2}A_{i,j}\le z\right)-\Phi(z)\right|\lesssim\frac{1}{\sqrt{n}}.$$

Lemma 46 states that conditional distribution of $W_{i,j}$ (conditional on $\boldsymbol{F}$) is nearly invariant, suggesting that $W_{i,j}$ is almost independent of $\sigma(\boldsymbol{F})$. In addition, $A_{i,j}$ is $\sigma(\boldsymbol{F})$-measurable, thus indicating that $W_{i,j}$ and $A_{i,j}$ are nearly statistically independent. Akin to Lemma 18, we can rigorously show that $S_{i,j}-S_{i,j}^{\star}=W_{i,j}+A_{i,j}$ is approximately Gaussian with mean zero and variance

$$v_{ij}^{\star}\coloneqq\widetilde{v}_{i,j}+\overline{v}_{i,j}.$$

This is formally stated and proved in the next lemma, which in turn concludes the proof of Theorem 17.

Lemma 48. *Instate the assumptions of Lemma 46, and suppose that* $n\gtrsim\log(n+d)$. *Then we have*

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\left(S_{i,j}-S_{i,j}^{\star}\right)/\sqrt{v_{i,j}^{\star}}\le t\right)-\Phi(t)\right|\lesssim\frac{1}{\sqrt{\log(n+d)}}=o(1).$$

The proof of Lemma 48 is exactly the same as that of Lemma 18 in Appendix E.3.5, and is hence omitted here for brevity.

G.5 Validity of confidence intervals (Proof of Theorem 18)

Armed with the above entrywise distributional theory for $\boldsymbol{S}-\boldsymbol{S}^{\star}$, we hope to construct valid confidence interval for each $S_{i,j}^{\star}$ based on the estimate $\boldsymbol{S}$ returned by HeteroPCA. This requires a faithful estimate of the variance $v_{i,j}^{\star}$. Towards this, we define the following plug-in estimator: if $i\neq j$, let

$$\begin{aligned}v_{i,j}&=\frac{2-p}{np}S_{i,i}S_{j,j}+\frac{4-3p}{np}S_{i,j}^2+\frac{1}{np}\left(\omega_i^2S_{j,j}+\omega_j^2S_{i,i}\right)\\&\quad+\frac{1}{np^2}\sum_{k=1}^{d}\left\{\left[\omega_i^2+(1-p)S_{i,i}\right]\left[\omega_k^2+(1-p)S_{k,k}\right]+2(1-p)^2S_{i,k}^2\right\}\left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\top}\right)^2\\&\quad+\frac{1}{np^2}\sum_{k=1}^{d}\left\{\left[\omega_j^2+(1-p)S_{j,j}\right]\left[\omega_k^2+(1-p)S_{k,k}\right]+2(1-p)^2S_{j,k}^2\right\}\left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{i,\cdot}^{\top}\right)^2;\end{aligned} \tag{G.36a}$$

otherwise, let

$$\begin{aligned}v_{i,i}&=\frac{12-9p}{np}S_{i,i}^2+\frac{4}{np}\omega_i^2S_{i,i}\\&\quad+\frac{4}{np^2}\sum_{k=1}^{d}\left\{\left[\omega_i^2+(1-p)S_{i,i}\right]\left[\omega_k^2+(1-p)S_{k,k}\right]+2(1-p)^2S_{i,k}^2\right\}\left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{i,\cdot}^{\top}\right)^2.\end{aligned} \tag{G.36b}$$

Step 1: faithfulness of the plug-in estimator. With the fine-grained estimation guarantees in Lemma 40 and Lemma 41 in place, we can demonstrate that $v_{i,j}$ is a reliable estimate of $v_{i,j}^{\star}$, as formally stated below.

Lemma 49. *Instate the conditions in Lemma 46. For any $\delta \in (0,1)$, we further assume that $n \gtrsim \delta^{-2}\kappa^7\mu^2 r^3 \log(n+d)$, $d \gtrsim \kappa^2\mu\log(n+d)$,*

$$ndp^2 \gtrsim \delta^{-2}\kappa^6\mu^4 r^6 \log^5(n+d), \qquad np \gtrsim \delta^{-2}\kappa^6\mu^3 r^5 \log^3(n+d),$$

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{\delta}{\kappa^2\mu r^2\log^{3/2}(n+d)}, \qquad \frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}} \lesssim \frac{\delta}{\kappa^{5/2}\mu r^2\log(n+d)},$$

and

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \gtrsim \delta^{-1}\kappa\mu r^2\log(n+d)\left[\frac{\kappa\mu r\log(n+d)}{\sqrt{ndp}} + \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} + \frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}}\right]\sqrt{\frac{r}{d}}.$$

Then with probability exceeding $1-O((n+d)^{-10})$, the quantity $v_{i,j}$ defined in (G.36) obeys

$$\left|v_{i,j}^{\star} - v_{i,j}\right| \lesssim \delta v_{i,j}^{\star}.$$

Proof. See Appendix H.4.1. □

Step 2: validity of the constructed confidence intervals. Armed with the Gaussian approximation in Lemma 48 as well as the faithfulness of $v_{i,j}$ as an estimate of $v_{i,j}^{\star}$ in the previous lemma, we show in the next lemma that the confidence interval constructed in Algorithm 6 is valid and nearly accurate.

Lemma 50. *Instate the conditions in Lemma 46. Further suppose that $n \gtrsim \kappa^9\mu^3 r^4\log^4(n+d)$, $d \gtrsim \kappa^2\mu\log(n+d)$,*

$$ndp^2 \gtrsim \kappa^8\mu^5 r^7\log^8(n+d), \qquad np \gtrsim \kappa^8\mu^4 r^6\log^6(n+d),$$

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\kappa^3\mu^{3/2}r^{5/2}\log^3(n+d)}, \qquad \frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\kappa^{7/2}\mu^{3/2}r^{5/2}\log^{5/2}(n+d)},$$

and

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \gtrsim \kappa^2\mu^{3/2}r^{5/2}\log^{5/2}(n+d)\left[\frac{\kappa\mu r\log(n+d)}{\sqrt{ndp}} + \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} + \frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}}\right]\sqrt{\frac{r}{d}}.$$

Then the confidence region $\mathsf{CI}_{i,j}^{1-\alpha}$ returned from Algorithm 6 satisfies

$$\mathbb{P}\left(S_{i,j}^{\star} \in \mathsf{CI}_{i,j}^{1-\alpha}\right) = 1-\alpha+O\left(\frac{1}{\sqrt{\log(n+d)}}\right) = 1-\alpha+o(1).$$

Proof. See Appendix H.4.2. □

H Auxiliary lemmas: the approach based on HeteroPCA

H.1 Auxiliary lemmas for Theorem 15

H.1.1 Proof of Lemma 37

As before, we remind the reader that $\boldsymbol{R} \in \mathcal{O}^{r\times r}$ represents the rotation matrix that best aligns $\boldsymbol{U}$ and $\boldsymbol{U}^{\star}$, and the rotation matrix $\boldsymbol{Q}$ is chosen to satisfy $\boldsymbol{U}^{\star}\boldsymbol{Q} = \boldsymbol{U}^{\natural}$. In addition, we have also shown in (D.6) that $\boldsymbol{R}\boldsymbol{Q} \in \mathcal{O}^{r\times r}$ is the rotation matrix that best aligns $\boldsymbol{U}$ and $\boldsymbol{U}^{\natural}$. Suppose for the moment that the assumptions of Theorem 10 are satisfied (which we shall verify shortly). Then conditional on $\boldsymbol{F}$, invoking Theorem 10 leads to

$$\boldsymbol{URQ} - \boldsymbol{U}^{\natural} = \left[\boldsymbol{E}\boldsymbol{M}^{\natural\top} + \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^{\top}\right)\right]\boldsymbol{U}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-2} + \boldsymbol{\Psi}, \tag{H.1}$$

where the residual matrix $\boldsymbol{\Psi}$ can be controlled with high probability as follows

$$\mathbb{P}\left(\left\|\boldsymbol{\Psi}\right\|_{2,\infty} \lesssim \zeta_{\mathsf{2nd}}\left(\boldsymbol{F}\right) \,\middle|\, \boldsymbol{F}\right) \geq 1 - O\left(\left(n+d\right)^{-10}\right).$$

Here, the quantity $\zeta_{\mathsf{2nd}}\left(\boldsymbol{F}\right)$ is defined as

$$\zeta_{\mathsf{2nd}}\left(\boldsymbol{F}\right) \coloneqq \kappa^{\natural 2}\frac{\mu^{\natural}r}{d}\frac{\zeta_{\mathsf{1st}}\left(\boldsymbol{F}\right)}{\sigma_r^{\natural 2}} + \kappa^{\natural 2}\frac{\zeta_{\mathsf{1st}}^2\left(\boldsymbol{F}\right)}{\sigma_r^{\natural 4}}\sqrt{\frac{\mu^{\natural}r}{d}}$$

with the quantity $\zeta_{\mathsf{1st}}\left(\boldsymbol{F}\right)$ given by

$$\zeta_{\mathsf{1st}}\left(\boldsymbol{F}\right) \coloneqq \sigma^2\sqrt{nd}\log\left(n+d\right) + \sigma\sigma_1^{\natural}\sqrt{d\log\left(n+d\right)}.$$

Given that $\boldsymbol{U}^{\star} = \boldsymbol{U}^{\natural}\boldsymbol{Q}^{\top}$, the above decomposition (H.1) can alternatively be written as

$$\boldsymbol{UR} - \boldsymbol{U}^{\star} = \underbrace{\left[\boldsymbol{E}\boldsymbol{M}^{\natural\top} + \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^{\top}\right)\right]\boldsymbol{U}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-2}\boldsymbol{Q}^{\top}}_{=:\boldsymbol{Z}} + \boldsymbol{\Psi}\boldsymbol{Q}^{\top}.$$

When the event $\mathcal{E}_{\mathsf{good}}$ occurs, we can see from (G.4), (G.5), (G.6) and (G.8) that

$$\begin{aligned}\zeta_{\mathsf{1st}}\left(\boldsymbol{F}\right) &\asymp \sigma_{\mathsf{ub}}^2\sqrt{nd}\log\left(n+d\right) + \sigma_{\mathsf{ub}}\sigma_1^{\natural}\sqrt{d\log\left(n+d\right)}\\ &\asymp \underbrace{\frac{\mu r\log^2\left(n+d\right)}{\sqrt{ndp}}\sigma_1^{\star 2} + \frac{\omega_{\max}^2}{p}\sqrt{\frac{d}{n}}\log\left(n+d\right) + \sigma_1^{\star 2}\sqrt{\frac{\mu r}{np}}\log\left(n+d\right) + \sigma_1^{\star}\omega_{\max}\sqrt{\frac{d\log\left(n+d\right)}{np}}}_{=:\zeta_{\mathsf{1st}}}\end{aligned}$$

and

$$\zeta_{\mathsf{2nd}}\left(\boldsymbol{F}\right) \lesssim \underbrace{\frac{\kappa^2\mu r\log\left(n+d\right)}{d}\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}} + \frac{\zeta_{\mathsf{1st}}^2}{\sigma_r^{\star 4}}\sqrt{\frac{\kappa^3\mu r\log\left(n+d\right)}{d}}}_{=:\zeta_{\mathsf{2nd}}}.$$

These bounds taken together imply that

$$\mathbb{P}\left(\left\|\boldsymbol{\Psi}\right\|_{2,\infty}\mathbb{1}_{\mathcal{E}_{\mathsf{good}}} \lesssim \zeta_{\mathsf{2nd}} \,\middle|\, \boldsymbol{F}\right) \geq 1 - O\left(\left(n+d\right)^{-10}\right).$$

Additionally, in view of the facts that $\mathcal{E}_{\mathsf{good}}$ is $\sigma(\boldsymbol{F})$-measurable and

$$\left\|\boldsymbol{\Psi}\right\|_{2,\infty} = \left\|\boldsymbol{\Psi}\boldsymbol{Q}^{\top}\right\|_{2,\infty} = \left\|\boldsymbol{UR} - \boldsymbol{U}^{\star} - \boldsymbol{Z}\right\|_{2,\infty},$$

one can readily demonstrate that

$$\mathbb{P}\left(\left\|\boldsymbol{UR} - \boldsymbol{U}^{\star} - \boldsymbol{Z}\right\|_{2,\infty} \lesssim \zeta_{\mathsf{2nd}} \,\middle|\, \boldsymbol{F}\right) \geq 1 - O\left(\left(n+d\right)^{-10}\right)$$

on the high-probability event $\mathcal{E}_{\mathsf{good}}$.

It remains to verify the assumptions of Theorem 10, which requires

$$d \gtrsim \kappa^{\natural 4}\mu^{\natural}r + \mu^{\natural 2}r\log^2\left(n+d\right), \quad n \gtrsim r\log^4\left(n+d\right), \quad B \lesssim \frac{\sigma_{\mathsf{ub}}\min\left\{\sqrt{n_2}, \sqrt[4]{n_1 n_2}\right\}}{\sqrt{\log n}}, \quad \zeta_{\mathsf{1st}} \ll \frac{\sigma_r^{\natural 2}}{\kappa^{\natural 2}} \tag{H.2}$$

whenever $\mathcal{E}_{\mathsf{good}}$ occurs. According to (G.4), (G.5), (G.6), (G.9) and the definition of ζ_{2nd}, the conditions in (H.2) are guaranteed to hold as long as

$$d \gtrsim \kappa^3\mu^2 r\log^4\left(n+d\right), \qquad n \gtrsim r\log^4\left(n+d\right), \qquad \text{and} \qquad \zeta_{\mathsf{1st}} \ll \frac{\sigma_r^{\star 2}}{\kappa}.$$

Finally, we would like to take a closer inspection on the condition $\zeta_{\mathsf{1st}} \ll \sigma_r^{\star 2}/\kappa$; in fact, we seek to derive sufficient conditions which guarantee that $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2} \ll \delta$ for any given $\delta > 0$, which will also be useful in subsequent analysis. It is straightforward to check that $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2} \lesssim \delta$ is equivalent to

$$ndp^2 \gg \delta^{-2}\kappa^2\mu^2 r^2 \log^4(n+d), \qquad np \gg \delta^{-2}\kappa^2\mu r \log^2(n+d), \tag{H.3a}$$

$$\text{and} \qquad \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \ll \frac{\delta}{\log(n+d)}, \qquad \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d}{np}} \ll \frac{\delta}{\sqrt{\kappa \log(n+d)}}. \tag{H.3b}$$

By taking $\delta \coloneqq 1/\kappa$, we see that $\zeta_{\mathsf{1st}} \ll \sigma_r^{\star 2}/\kappa$ is guaranteed as long as the following conditions hold:

$$ndp^2 \gg \kappa^4\mu^2 r^2 \log^4(n+d), \qquad np \gg \kappa^4\mu r \log^2(n+d)$$

$$\text{and} \qquad \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \ll \frac{1}{\kappa\log(n+d)}, \qquad \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d}{np}} \ll \frac{1}{\sqrt{\kappa^3 \log(n+d)}}.$$

This concludes the proof.

H.1.2 Proof of Lemma 38

Let us write $\widetilde{\boldsymbol{\Sigma}}_l$ as the superposition of two components:

$$\widetilde{\boldsymbol{\Sigma}}_l = \underbrace{\boldsymbol{Q}(\boldsymbol{\Sigma}^\natural)^{-1}\boldsymbol{V}^{\natural\top}\mathsf{diag}\left\{\sigma_{l,1}^2,\dots,\sigma_{l,n}^2\right\}\boldsymbol{V}^\natural(\boldsymbol{\Sigma}^\natural)^{-1}\boldsymbol{Q}^\top}_{=:\widetilde{\boldsymbol{\Sigma}}_{l,1}}$$

$$+\underbrace{\boldsymbol{Q}(\boldsymbol{\Sigma}^\natural)^{-2}\boldsymbol{U}^{\natural\top}\mathsf{diag}\left\{\sum_{j:j\neq l}\sigma_{l,j}^2\sigma_{1,j}^2,\dots,\sum_{j:j\neq l}\sigma_{l,j}^2\sigma_{d,j}^2\right\}\boldsymbol{U}^\natural(\boldsymbol{\Sigma}^\natural)^{-2}\boldsymbol{Q}^\top}_{=:\widetilde{\boldsymbol{\Sigma}}_{l,2}}.$$

We shall control $\widetilde{\boldsymbol{\Sigma}}_{l,1}$ and $\widetilde{\boldsymbol{\Sigma}}_{l,2}$ separately. Throughout this subsection we assume that $\mathcal{E}_{\mathsf{good}}$ happens.

Step 1: identifying a good approximation of $\widetilde{\boldsymbol{\Sigma}}_{l,1}$. To begin with, the concentration of the first component $\widetilde{\boldsymbol{\Sigma}}_{l,1}$ has already been studied in Lemma 8, which tells us that

$$\widetilde{\boldsymbol{\Sigma}}_{l,1} = \boldsymbol{\Sigma}_{l,1}^\star + \boldsymbol{R}_1.$$

Here, the matrix $\boldsymbol{\Sigma}_{l,1}^\star$ is defined as

$$\boldsymbol{\Sigma}_{l,1}^\star = \left(\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2 + \frac{\omega_l^{\star 2}}{np}\right)(\boldsymbol{\Sigma}^\star)^{-2} + \frac{2(1-p)}{np}\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^\star,$$

and $\boldsymbol{R}_1$ is some residual matrix satisfying

$$\|\boldsymbol{R}_1\| \lesssim \sqrt{\frac{\kappa^5 r\log^3(n+d)}{n}}\lambda_{\min}\left(\boldsymbol{\Sigma}_{l,1}^\star\right).$$

Step 2: identifying a good approximation of $\widetilde{\boldsymbol{\Sigma}}_{l,2}$. We now turn attention to approximating $\widetilde{\boldsymbol{\Sigma}}_{l,2}$, and it suffices to study $\sum_{j\neq l}^n \sigma_{l,j}^2\sigma_{i,j}^2$ for each $i \in [d]$. In view of the expression (D.8), we have

$$\sum_{j:j\neq l}\sigma_{l,j}^2\sigma_{i,j}^2 = \sum_{j:j\neq l}\left[\frac{1-p}{np}\left(\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\boldsymbol{f}_j\right)^2 + \frac{\omega_l^{\star 2}}{np}\right]\left[\frac{1-p}{np}\left(\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\boldsymbol{f}_j\right)^2 + \frac{\omega_i^{\star 2}}{np}\right]$$

$$= \left(\frac{1-p}{np}\right)^2\underbrace{\sum_{j:j\neq l}\left(\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\boldsymbol{f}_j\right)^2\left(\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\boldsymbol{f}_j\right)^2}_{=:\alpha_1} + \frac{(1-p)\,\omega_i^{\star 2}}{n^2p^2}\underbrace{\sum_{j:j\neq l}\left(\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\boldsymbol{f}_j\right)^2}_{=:\alpha_2}$$

$$+\frac{(1-p)\,\omega_l^{\star 2}}{n^2p^2}\underbrace{\sum_{j:j\neq l}\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_j\right)^2}_{=:\alpha_3}+\frac{(n-1)\,\omega_l^{\star 2}\omega_i^{\star 2}}{n^2p^2}. \tag{H.4}$$

To proceed, we can see from (G.12a), (G.12b) and (E.57) that

$$\left|\frac{1}{n}\alpha_1-\left[\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2+2S_{l,i}^{\star 2}\right]\right|\lesssim\sqrt{\frac{\log^3(n+d)}{n}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2,$$

$$\left|\frac{1}{n}\alpha_2-\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\right|\lesssim\sqrt{\frac{\log(n+d)}{n}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2,$$

$$\left|\frac{1}{n}\alpha_3-\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\right|\lesssim\sqrt{\frac{\log(n+d)}{n}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2.$$

Substitution into (H.4) yields

$$\sum_{j:j\neq l}\sigma_{l,j}^2\sigma_{i,j}^2=\underbrace{\frac{(1-p)^2}{np^2}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2+2S_{l,i}^{\star 2}\right)+\frac{1-p}{np^2}\left(\omega_i^{\star 2}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2+\omega_l^{\star 2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\right)+\frac{\omega_l^{\star 2}\omega_i^{\star 2}}{np^2}}_{=:\,d_{l,i}}+r_i \tag{H.5}$$

for some residual term r_i satisfying

$$|r_i|\lesssim\sqrt{\frac{\log^3(n+d)}{n}}d_{l,i}. \tag{H.6}$$

As a consequence, the above calculations together with the definition of $\widetilde{\boldsymbol{\Sigma}}_{l,2}$ allow one to write

$$\widetilde{\boldsymbol{\Sigma}}_{l,2}=\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-2}\boldsymbol{U}^{\natural\top}\mathsf{diag}\left\{d_{l,i}\right\}_{i=1}^{d}\boldsymbol{U}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-2}\boldsymbol{Q}^{\top}+\boldsymbol{R}_2 \tag{H.7}$$

for some residual matrix $\boldsymbol{R}_2$ satisfying

$$\begin{aligned}\|\boldsymbol{R}_2\|&\leq\left\|\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-2}\boldsymbol{U}^{\natural\top}\mathsf{diag}\left\{|r_1|,\ldots,|r_d|\right\}\boldsymbol{U}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-2}\boldsymbol{Q}^{\top}\right\|\\&\lesssim\frac{1}{\left(\sigma_r(\boldsymbol{\Sigma}^{\natural})\right)^4}\max_{1\leq i\leq d}|r_i|\lesssim\sqrt{\frac{\log^3(n+d)}{n}}\max_{1\leq i\leq d}\frac{1}{\sigma_r^{\star 4}}d_{l,i},\end{aligned}$$

where the last inequality makes use of (G.4) and (H.6).

Further, it turns out that the term $\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-2}\boldsymbol{U}^{\natural\top}$ used in (H.7) can be well approximated by $(\boldsymbol{\Sigma}^{\star})^{-2}\boldsymbol{Q}$. To see this, we note that

$$\begin{aligned}\left\|\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-2}-(\boldsymbol{\Sigma}^{\star})^{-2}\boldsymbol{Q}\right\|&\leq\left\|\left[\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-1}-(\boldsymbol{\Sigma}^{\star})^{-1}\boldsymbol{Q}\right](\boldsymbol{\Sigma}^{\natural})^{-1}\right\|+\left\|(\boldsymbol{\Sigma}^{\star})^{-1}\left[\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-1}-(\boldsymbol{\Sigma}^{\star})^{-1}\boldsymbol{Q}\right]\right\|\\&\overset{\text{(i)}}{\asymp}\frac{1}{\sigma_r^{\star}}\left\|\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-1}-(\boldsymbol{\Sigma}^{\star})^{-1}\boldsymbol{Q}\right\|\overset{\text{(ii)}}{\lesssim}\frac{\kappa}{\sigma_r^{\star 2}}\sqrt{\frac{r+\log(n+d)}{n}},\end{aligned}$$

where (i) comes from (G.4), and (ii) results from (E.2). This combined with the identity $\boldsymbol{U}^{\star}=\boldsymbol{U}^{\natural}\boldsymbol{Q}^{\top}$ gives

$$\begin{aligned}\left\|\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-2}\boldsymbol{U}^{\natural\top}-(\boldsymbol{\Sigma}^{\star})^{-2}\boldsymbol{U}^{\star\top}\right\|&=\left\|\left[\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-2}-(\boldsymbol{\Sigma}^{\star})^{-2}\boldsymbol{Q}\right]\boldsymbol{U}^{\natural\top}\right\|\\&\leq\left\|\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-2}-(\boldsymbol{\Sigma}^{\star})^{-2}\boldsymbol{Q}\right\|\\&\lesssim\frac{\kappa}{\sigma_r^{\star 2}}\sqrt{\frac{r+\log(n+d)}{n}}.\end{aligned}$$

Therefore, substituting this into (H.7) reveals that

$$
\begin{aligned}
&\left\|\widetilde{\boldsymbol{\Sigma}}_{l,2}-\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{U}^{\star\top}\mathsf{diag}\left\{d_{l,i}\right\}_{i=1}^{d}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\right\| \\
&\quad\leq\left\|\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-2}\boldsymbol{U}^{\natural\top}\mathsf{diag}\left\{d_{l,i}\right\}_{i=1}^{d}\boldsymbol{U}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-2}\boldsymbol{Q}^{\top}-\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{U}^{\star\top}\mathsf{diag}\left\{d_{l,i}\right\}_{i=1}^{d}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\right\|+\|\boldsymbol{R}_{2}\| \\
&\quad\leq\left\|\left(\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-2}\boldsymbol{U}^{\natural\top}-\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{U}^{\star\top}\right)\mathsf{diag}\left\{d_{l,i}\right\}_{i=1}^{d}\boldsymbol{U}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-2}\boldsymbol{Q}^{\top}\right\| \\
&\qquad+\left\|\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{U}^{\star\top}\mathsf{diag}\left\{d_{l,i}\right\}_{i=1}^{d}\left(\boldsymbol{Q}(\boldsymbol{\Sigma}^{\natural})^{-2}\boldsymbol{U}^{\natural\top}-\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{U}^{\star\top}\right)^{\top}\right\|+\|\boldsymbol{R}_{2}\| \\
&\quad\lesssim\frac{\kappa}{\sigma_{r}^{\star4}}\sqrt{\frac{r+\log\left(n+d\right)}{n}}\max_{1\leq i\leq d}d_{l,i}+\|\boldsymbol{R}_{2}\|,
\end{aligned}
$$

where the last relation relies on (G.4). These bounds taken collectively allow one to express

$$
\widetilde{\boldsymbol{\Sigma}}_{l,2}=\boldsymbol{\Sigma}_{l,2}^{\star}+\boldsymbol{R}_{3}, \tag{H.8}
$$

where

$$
\boldsymbol{\Sigma}_{l,2}^{\star}\coloneqq\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{U}^{\star\top}\mathsf{diag}\left\{d_{l,i}\right\}_{i=1}^{d}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2} \tag{H.9}
$$

and $\boldsymbol{R}_3$ is some residual matrix satisfying

$$
\|\boldsymbol{R}_{3}\|\lesssim\|\boldsymbol{R}_{2}\|+\frac{\kappa}{\sigma_{r}^{\star4}}\sqrt{\frac{r+\log\left(n+d\right)}{n}}\max_{1\leq i\leq d}d_{l,i}\lesssim\sqrt{\frac{\kappa^{2}r\log^{3}\left(n+d\right)}{n}}\frac{1}{\sigma_{r}^{\star4}}\max_{1\leq i\leq d}d_{l,i}.
$$

Step 3: controlling the spectrum of $\widetilde{\boldsymbol{\Sigma}}_{l,2}$. Next, we will investigate the spectrum of $\widetilde{\boldsymbol{\Sigma}}_{l,2}$. It is straightforward to show that the matrix $\boldsymbol{\Sigma}_{l,2}^{\star}$ defined in (H.9) obeys

$$
\left\|\boldsymbol{\Sigma}_{l,2}^{\star}\right\|\leq\frac{1}{\sigma_{r}^{\star4}}\max_{1\leq i\leq d}|d_{l,i}|\leq\frac{3\kappa\mu r\left(1-p\right)^{2}}{ndp^{2}\sigma_{r}^{\star2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}+\frac{2\kappa\mu r\left(1-p\right)}{ndp^{2}\sigma_{r}^{\star2}}\omega_{\max}^{\star2}+\frac{\omega_{l}^{\star2}\omega_{\max}^{\star2}}{np^{2}\sigma_{r}^{\star4}}. \tag{H.10}
$$

In addition, we claim that the following inequality is valid.

***Claim* 2.** It holds that

$$
\boldsymbol{U}^{\star\top}\mathsf{diag}\left\{\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\right\}_{i=1}^{d}\boldsymbol{U}^{\star}\succeq\frac{\sigma_{r}^{\star2}}{4d}\boldsymbol{I}_{r}. \tag{H.11}
$$

With this claim in place, we can combine it with the definition (H.5) of $d_{l,i}$ to bound

$$
\begin{aligned}
\boldsymbol{\Sigma}_{l,2}^{\star}&=\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{U}^{\star\top}\mathsf{diag}\left\{d_{l,i}\right\}_{i=1}^{d}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2} \\
&\succeq\frac{\left(1-p\right)^{2}}{np^{2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{U}^{\star\top}\mathsf{diag}\left\{\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\right\}_{i=1}^{d}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2} \\
&\quad+\frac{1-p}{np^{2}}\omega_{l}^{\star2}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{U}^{\star\top}\mathsf{diag}\left\{\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\right\}_{i=1}^{d}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2} \\
&\quad+\frac{1}{np^{2}}\omega_{l}^{\star2}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{U}^{\star\top}\mathsf{diag}\left\{\omega_{i}^{\star2}\right\}_{i=1}^{d}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2} \\
&\succeq\left[\frac{\left(1-p\right)^{2}}{4ndp^{2}}\sigma_{r}^{\star2}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}+\frac{1-p}{4ndp^{2}}\sigma_{r}^{\star2}\omega_{l}^{\star2}+\frac{\omega_{l}^{\star2}\omega_{\min}^{2}}{np^{2}}\right]\left(\boldsymbol{\Sigma}^{\star}\right)^{-4},
\end{aligned}
$$

which in turn leads to

$$
\begin{aligned}
\lambda_{\min}\left(\boldsymbol{\Sigma}_{l,2}^{\star}\right)&\geq\frac{1}{\sigma_{1}^{\star4}}\left[\frac{\left(1-p\right)^{2}}{4ndp^{2}}\sigma_{r}^{\star2}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}+\frac{1-p}{4ndp^{2}}\sigma_{r}^{\star2}\omega_{l}^{\star2}+\frac{\omega_{l}^{\star2}\omega_{\min}^{2}}{np^{2}}\right] \\
&=\frac{\left(1-p\right)^{2}}{4ndp^{2}\kappa\sigma_{1}^{\star2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}+\frac{1-p}{4ndp^{2}\kappa\sigma_{1}^{\star2}}\omega_{l}^{\star2}+\frac{\omega_{l}^{\star2}\omega_{\min}^{2}}{np^{2}\sigma_{1}^{\star4}}.
\end{aligned} \tag{H.12}
$$

The above bounds (H.10) and (H.12) imply that the condition number of $\mathbf{\Sigma}_{l,2}^{\star}$ — denoted by $\kappa(\mathbf{\Sigma}_{l,2}^{\star})$ — is upper bounded by

$$\kappa(\mathbf{\Sigma}_{l,2}^{\star}) = \frac{\left\|\mathbf{\Sigma}_{l,2}^{\star}\right\|}{\lambda_{\min}\left(\mathbf{\Sigma}_{l,2}^{\star}\right)} \leq 12\kappa^{3}\mu r + 8\kappa^{3}\mu r\frac{\omega_{\max}^{\star 2}}{\omega_{l}^{\star 2}} + \kappa^{2}\frac{\omega_{\max}^{\star 2}}{\omega_{\min}^{2}} \lesssim \kappa^{3}\mu r.$$

Moreover, we can also obtain

$$\begin{aligned}\|\boldsymbol{R}_{3}\| &\lesssim \sqrt{\frac{\kappa^{2}r\log^{3}(n+d)}{n}}\left(\frac{3\kappa\mu r(1-p)^{2}}{ndp^{2}\sigma_{r}^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\mathbf{\Sigma}^{\star}\right\|_{2}^{2} + \frac{2\kappa\mu r(1-p)}{ndp^{2}\sigma_{r}^{\star 2}}\omega_{\max}^{\star 2} + \frac{\omega_{l}^{\star 2}\omega_{\max}^{\star 2}}{np^{2}\sigma_{r}^{\star 4}}\right)\\ &\lesssim \sqrt{\frac{\kappa^{2}r\log^{3}(n+d)}{n}}\kappa^{3}\mu r\lambda_{\min}\left(\mathbf{\Sigma}_{l,2}^{\star}\right) \asymp \sqrt{\frac{\kappa^{8}\mu^{2}r^{3}\log^{3}(n+d)}{n}}\lambda_{\min}\left(\mathbf{\Sigma}_{l,2}^{\star}\right).\end{aligned}$$

Combining the above results, we are allowed to express

$$\widetilde{\mathbf{\Sigma}}_{l} = \mathbf{\Sigma}_{U,l}^{\star} + \boldsymbol{R}_{l}$$

where $\mathbf{\Sigma}_{U,l}^{\star} = \mathbf{\Sigma}_{l,1}^{\star} + \mathbf{\Sigma}_{l,2}^{\star}$ and the residual matrix $\boldsymbol{R}_{l}$ satisfies

$$\begin{aligned}\|\boldsymbol{R}_{l}\| \leq \|\boldsymbol{R}_{1}\| + \|\boldsymbol{R}_{3}\| &\lesssim \sqrt{\frac{\kappa^{5}r\log^{3}(n+d)}{n}}\lambda_{\min}\left(\mathbf{\Sigma}_{l,1}^{\star}\right) + \sqrt{\frac{\kappa^{8}\mu^{2}r^{3}\log^{3}(n+d)}{n}}\lambda_{\min}\left(\mathbf{\Sigma}_{l,2}^{\star}\right)\\ &\lesssim \sqrt{\frac{\kappa^{8}\mu^{2}r^{3}\log^{3}(n+d)}{n}}\lambda_{\min}\left(\mathbf{\Sigma}_{U,l}^{\star}\right).\end{aligned}$$

An immediate consequence is that the conditional number of $\mathbf{\Sigma}_{U,l}^{\star}$ is upper bounded by the maximum of the conditional numbers of $\mathbf{\Sigma}_{l,1}^{\star}$ and $\mathbf{\Sigma}_{l,2}^{\star}$, which is at most $O(\kappa^{3}\mu r)$. Consequently, when $n \gg \kappa^{8}\mu^{2}r^{3}\log^{3}(n+d)$, the conditional number of $\widetilde{\mathbf{\Sigma}}_{l}$ is also upper bounded by $O(\kappa^{3}\mu r)$. This finishes the proof, as long as Claim 2 can be justified.

Proof of Claim 2. For any $c \in (0,1)$, we define an index set

$$\mathcal{I}_{c} := \{i \in [d] : \|\boldsymbol{U}_{i,\cdot}^{\star}\|_{2} \geq \sqrt{c/d}\}.$$

Then for any $\boldsymbol{v} \in \mathbb{R}^{d}$, one has

$$\begin{aligned}\boldsymbol{v}^{\top}\boldsymbol{U}^{\star\top}\mathsf{diag}\left\{\left\|\boldsymbol{U}_{i,\cdot}^{\star}\mathbf{\Sigma}^{\star}\right\|_{2}^{2}\right\}_{i=1}^{d}\boldsymbol{U}^{\star}\boldsymbol{v} &= \sum_{i=1}^{d}\left(\boldsymbol{v}^{\top}\boldsymbol{U}_{i,\cdot}^{\star}\right)^{2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\mathbf{\Sigma}^{\star}\right\|_{2}^{2} \geq \sigma_{r}^{\star 2}\sum_{i=1}^{d}\left(\boldsymbol{v}^{\top}\boldsymbol{U}_{i,\cdot}^{\star}\right)^{2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}^{2}\\ &\geq \sigma_{r}^{\star 2}\sum_{i\in\mathcal{I}_{c}}\left(\boldsymbol{v}^{\top}\boldsymbol{U}_{i,\cdot}^{\star}\right)^{2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}^{2} \geq \frac{c\sigma_{r}^{\star 2}}{d}\sum_{i\in\mathcal{I}_{c}}\left(\boldsymbol{v}^{\top}\boldsymbol{U}_{i,\cdot}^{\star}\right)^{2}\\ &= \frac{c\sigma_{r}^{\star 2}}{d}\left[\|\boldsymbol{v}\|_{2}^{2} - \sum_{i\in[d]\setminus\mathcal{I}_{c}}\left(\boldsymbol{v}^{\top}\boldsymbol{U}_{i,\cdot}^{\star}\right)^{2}\right]\\ &\geq \frac{c\sigma_{r}^{\star 2}}{d}\left[\|\boldsymbol{v}\|_{2}^{2} - \sum_{i\in[d]\setminus\mathcal{I}_{c}}\|\boldsymbol{v}\|_{2}^{2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}^{2}\right]\\ &\geq \frac{c\sigma_{r}^{\star 2}}{d}\left[\|\boldsymbol{v}\|_{2}^{2} - \sum_{i\in[d]\setminus\mathcal{I}_{c}}\|\boldsymbol{v}\|_{2}^{2}\frac{c}{d}\right]\\ &\geq \frac{c\sigma_{r}^{\star 2}}{d}\left[\|\boldsymbol{v}\|_{2}^{2} - c\|\boldsymbol{v}\|_{2}^{2}\right] \geq \frac{c(1-c)\sigma_{r}^{\star 2}}{d}\|\boldsymbol{v}\|_{2}^{2}.\end{aligned}$$

Since the above inequality holds for arbitrary $c \in (0,1)$ and any $\boldsymbol{v} \in \mathbb{R}^d$, taking $c = 1/2$ leads to

$$\inf_{\boldsymbol{v}\in\mathbb{R}^d} \boldsymbol{v}^\top \boldsymbol{U}^{\star\top} \mathsf{diag}\left\{\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\right\}_{i=1}^d \boldsymbol{U}^{\star}\boldsymbol{v} \geq \frac{\sigma_r^{\star 2}}{4d}\left\|\boldsymbol{v}\right\|_2^2.$$

Therefore, we can conclude that

$$\boldsymbol{U}^{\star\top} \mathsf{diag}\left\{\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\right\}_{i=1}^d \boldsymbol{U}^{\star} \succeq \frac{\sigma_r^{\star 2}}{4d}\boldsymbol{I}_r.$$

□

H.1.3 Proof of Lemma 39

Step 1: Gaussian approximation of $\boldsymbol{Z}_{l,\cdot}$ using the Berry-Esseen Theorem. It is first seen that

$$\boldsymbol{Z}_{l,\cdot} = \sum_{j=1}^n \boldsymbol{Y}_j, \qquad \text{where} \quad \boldsymbol{Y}_j = E_{l,j}\left[\boldsymbol{V}_{j,\cdot}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1} + \left[\mathcal{P}_{-l,\cdot}\left(\boldsymbol{E}_{\cdot,j}\right)\right]^\top \boldsymbol{U}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-2}\right]\boldsymbol{Q}^\top.$$

Invoking the Berry-Esseen theorem (see Theorem 20) then yields

$$\sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left(\boldsymbol{Z}_{l,\cdot}\in\mathcal{C}\middle|\boldsymbol{F}\right) - \mathbb{P}\left(\mathcal{N}\big(\boldsymbol{0},\widetilde{\boldsymbol{\Sigma}}_l\big)\in\mathcal{C}\middle|\boldsymbol{F}\right)\right| \lesssim r^{1/4}\gamma\left(\boldsymbol{F}\right), \tag{H.13}$$

where $\widetilde{\boldsymbol{\Sigma}}_l$ is the covariance matrix of $\boldsymbol{Z}_{l,\cdot}$ that has been calculated in (G.16), and $\gamma(\boldsymbol{F})$ is defined to be

$$\gamma\left(\boldsymbol{F}\right) \coloneqq \sum_{i=1}^n \mathbb{E}\left[\left\|\boldsymbol{Y}_j\widetilde{\boldsymbol{\Sigma}}_l^{-1/2}\right\|_2^3\middle|\boldsymbol{F}\right].$$

In the sequel, let us develop an upper bound on $\gamma(\boldsymbol{F})$. Towards this, we find it helpful to define the quantity B_l such that

$$\max_{j\in[n]}\left|E_{l,j}\right| \leq B_l.$$

- Towards this, we first provide a (conditional) high probability bound for each $\|\boldsymbol{Y}_j\|$. In view of Cai et al. (2021, Lemma 12), with probability exceeding $1-O((n+d)^{-101})$ we have

$$\begin{aligned}\left\|\left[\mathcal{P}_{-l,\cdot}\left(\boldsymbol{E}_{\cdot,j}\right)\right]^\top\boldsymbol{U}^{\natural}\right\|_2 = \left\|\sum_{i:i\neq l}E_{i,j}\boldsymbol{U}_{i,\cdot}^{\natural}\right\|_2 &\lesssim \left(B\log\left(n+d\right)+\sigma\sqrt{d\log\left(n+d\right)}\right)\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}\\ &\asymp \left(B\log\left(n+d\right)+\sigma\sqrt{d\log\left(n+d\right)}\right)\left\|\boldsymbol{U}^{\star}\right\|_{2,\infty}\end{aligned}$$

  holds for all $j\in[n]$, where the last relation makes use of the fact that $\boldsymbol{U}^{\natural} = \boldsymbol{U}^{\star}\boldsymbol{Q}$ for some orthonormal matrix $\boldsymbol{Q}$ (see (D.4)). Consequently, with probability exceeding $1-O((n+d)^{-101})$,

$$\left\|\boldsymbol{Y}_j\right\|_2 \leq \frac{1}{\sigma_r^{\natural}}B_l\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} + \frac{1}{\sigma_r^{\natural 2}}B_l\left(B\log\left(n+d\right)+\sigma\sqrt{d\log\left(n+d\right)}\right)\left\|\boldsymbol{U}^{\star}\right\|_{2,\infty}$$

  holds for each $j\in[n]$. If we define

$$C_{\mathsf{prob}} \coloneqq \widetilde{C}_1 B_l\left[\frac{1}{\sigma_r^{\natural}}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} + \frac{1}{\sigma_r^{\natural 2}}\left(B\log\left(n+d\right)+\sigma\sqrt{d\log\left(n+d\right)}\right)\left\|\boldsymbol{U}^{\star}\right\|_{2,\infty}\right]$$

  for some sufficiently large constant $\widetilde{C}_1 > 0$, then the union bound over $1\leq j\leq n$ guarantees that

$$\mathbb{P}\left(\max_{1\leq j\leq n}\left\|\boldsymbol{Y}_j\right\|_2 \leq C_{\mathsf{prob}}\middle|\boldsymbol{F}\right) \geq 1-O\left(\left(n+d\right)^{-100}\right). \tag{H.14}$$

- In addition, we also know that $\|\boldsymbol{Y}_j\|_2$ admits a trivial deterministic upper bound as follows

$$\|\boldsymbol{Y}_j\|_2 \lesssim \frac{1}{\sigma_r^{\natural}} B_l \left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} + \frac{1}{\sigma_r^{\natural 2}} B_l \left\|\boldsymbol{E}_{\cdot,j}\right\|_2 \lesssim \frac{1}{\sigma_r^{\natural}} B_l \left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} + \frac{1}{\sigma_r^{\natural 2}} B_l B\sqrt{d}.$$

  This means that by defining the quantity

$$C_{\mathsf{det}} \coloneqq \widetilde{C}_2 B_l \left(\frac{1}{\sigma_r^{\natural}} \left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} + \frac{1}{\sigma_r^{\natural 2}} B\sqrt{d}\right)$$

  for some sufficiently large constant $\widetilde{C}_2 > 0$, we can ensure that

$$\max_{1\le j\le n} \|\boldsymbol{Y}_j\|_2 \le C_{\mathsf{det}}. \tag{H.15}$$

The above arguments taken together allow one to bound $\gamma(\boldsymbol{F})$ as follows

$$\begin{aligned}
\gamma(\boldsymbol{F}) &\le \lambda_{\min}^{-3/2}(\widetilde{\boldsymbol{\Sigma}}_l) \sum_{j=1}^{n} \mathbb{E}\left[\|\boldsymbol{Y}_j\|_2^3 \,\middle|\, \boldsymbol{F}\right] \\
&= \lambda_{\min}^{-3/2}(\widetilde{\boldsymbol{\Sigma}}_l) \sum_{j=1}^{n} \mathbb{E}\left[\|\boldsymbol{Y}_j\|_2^3 \mathbb{1}_{\|\boldsymbol{Y}_j\|_2 \le C_{\mathsf{prob}}} \,\middle|\, \boldsymbol{F}\right] + \lambda_{\min}^{-3/2}(\widetilde{\boldsymbol{\Sigma}}_l) \sum_{j=1}^{n} \mathbb{E}\left[\|\boldsymbol{Y}_j\|_2^3 \mathbb{1}_{\|\boldsymbol{Y}_j\|_2 > C_{\mathsf{prob}}} \,\middle|\, \boldsymbol{F}\right] \\
&\overset{\text{(i)}}{\lesssim} \lambda_{\min}^{-3/2}(\widetilde{\boldsymbol{\Sigma}}_l) C_{\mathsf{prob}} \sum_{j=1}^{n} \mathbb{E}\left[\|\boldsymbol{Y}_j\|_2^2 \,\middle|\, \boldsymbol{F}\right] + \lambda_{\min}^{-3/2}(\widetilde{\boldsymbol{\Sigma}}_l) \sum_{j=1}^{n} C_{\mathsf{det}}^3 \mathbb{P}\left(\max_{1\le j\le n} \|\boldsymbol{Y}_j\|_2 \le C_{\mathsf{prob}} \middle| \boldsymbol{F}\right) \\
&\overset{\text{(ii)}}{\lesssim} \lambda_{\min}^{-3/2}(\widetilde{\boldsymbol{\Sigma}}_l) \mathsf{tr}(\widetilde{\boldsymbol{\Sigma}}_l) C_{\mathsf{prob}} + \lambda_{\min}^{-3/2}(\widetilde{\boldsymbol{\Sigma}}_l) C_{\mathsf{det}}^3 (n+d)^{-99}.
\end{aligned}$$

Here (i) relies on (H.15); (ii) follows from (H.14) as well as the fact that $\sum_{j=1}^{n} \mathbb{E}\left[\|\boldsymbol{Y}_j\|_2^2 \,\middle|\, \boldsymbol{F}\right] = \mathsf{tr}(\widetilde{\boldsymbol{\Sigma}}_l)$, which we have already shown in (E.8). Substitution into (H.13) then yields

$$\begin{aligned}
&\sup_{\mathcal{C}\in\mathscr{C}^r} \left|\mathbb{P}\left(\boldsymbol{Z}_{l,\cdot} \in \mathcal{C} \middle| \boldsymbol{F}\right) - \mathbb{P}\left(\mathcal{N}(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}}_l) \in \mathcal{C} \middle| \boldsymbol{F}\right)\right| \\
&\qquad \lesssim \underbrace{r^{1/4} \lambda_{\min}^{-3/2}(\widetilde{\boldsymbol{\Sigma}}_l) \mathsf{tr}(\widetilde{\boldsymbol{\Sigma}}_l) C_{\mathsf{prob}}}_{=:\alpha} + \underbrace{r^{1/4} \lambda_{\min}^{-3/2}(\widetilde{\boldsymbol{\Sigma}}_l) C_{\mathsf{det}}^3 (n+d)^{-99}}_{=:\beta}.
\end{aligned} \tag{H.16}$$

When the high-probability event $\mathcal{E}_{\mathsf{good}}$ occurs, we know from Lemma 38 that the condition number of $\widetilde{\boldsymbol{\Sigma}}_l$ is bounded by $O(\kappa^3 \mu r)$. This implies that

$$\frac{\mathsf{tr}(\widetilde{\boldsymbol{\Sigma}}_l)}{\lambda_{\min}^{3/2}(\widetilde{\boldsymbol{\Sigma}}_l)} \le \frac{r\|\widetilde{\boldsymbol{\Sigma}}_l\|}{\lambda_{\min}^{3/2}(\widetilde{\boldsymbol{\Sigma}}_l)} \lesssim \frac{\kappa^3 \mu r^2}{\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l)}.$$

and as a consequence,

$$\begin{aligned}
\alpha &\lesssim \frac{\kappa^3 \mu r^{9/4}}{\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l)} C_{\mathsf{prob}} \\
&\asymp \underbrace{\frac{\kappa^3 \mu r^{9/4}}{\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l)} \frac{B_l}{\sigma_r^{\natural}} \left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty}}_{=:\alpha_1} + \underbrace{\frac{\kappa^3 \mu r^{9/4} \log(n+d)}{\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l)} \frac{B_l B}{\sigma_r^{\natural 2}} \|\boldsymbol{U}^\star\|_{2,\infty}}_{=:\alpha_2} + \underbrace{\frac{\kappa^3 \mu r^{9/4} \log^{3/2}(n+d)}{\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l)} \frac{\sigma B_l}{\sigma_r^{\natural 2}} \sqrt{d} \|\boldsymbol{U}^\star\|_{2,\infty}}_{=:\alpha_3}.
\end{aligned}$$

All of these terms involve $\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l)$, which can be lower bounded using Lemma 38 as follows

$$\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l) \gtrsim \frac{1}{\sqrt{np}\sigma_1^\star} \left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2 + \frac{\omega_l^\star}{\sqrt{np}\sigma_1^\star} + \frac{1}{\sqrt{ndp^2\kappa}\sigma_1^\star} \left\|\boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2 + \frac{1}{\sqrt{ndp^2\kappa}\sigma_1^\star} \omega_l^\star + \frac{\omega_{\min}^2}{\sqrt{np^2}\sigma_1^{\star 2}}. \tag{H.17}$$

Moreover, it is seen from (G.9) that

$$B_l \lesssim \frac{1}{p}\sqrt{\frac{\log(n+d)}{n}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{\omega_l^{\star}}{p}\sqrt{\frac{\log(n+d)}{n}} \asymp \frac{1}{p}\sqrt{\frac{\log(n+d)}{n}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right). \tag{H.18}$$

In the sequel, we shall bound α_1, α_2, and α_3 separately.

- We start with bounding α_1, where we have

$$\begin{aligned}
\alpha_1 &\overset{\text{(i)}}{\lesssim} \frac{\kappa^3\mu r^{9/4}}{\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l)}\frac{1}{\sigma_r^{\star}}\frac{1}{p}\sqrt{\frac{\log(n+d)}{n}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right)\sqrt{\frac{\log(n+d)}{n}} \\
&\lesssim \frac{1}{\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l)}\frac{\kappa^3\mu r^{9/4}\log^{1/2}(n+d)}{np\sigma_r^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right) \\
&\overset{\text{(ii)}}{\lesssim} \frac{\kappa^3\mu r^{9/4}\log^{1/2}(n+d)}{np\sigma_r^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right)\left[\frac{1}{\sqrt{np}\sigma_1^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right)\right]^{-1} \\
&\asymp \frac{\kappa^{7/2}\mu r^{9/4}\log^{1/2}(n+d)}{\sqrt{np}} \overset{\text{(iii)}}{\lesssim} \frac{1}{\sqrt{\log(n+d)}}.
\end{aligned}$$

  Here (i) follows from (G.4), (G.7), (G.9) and (H.18); (ii) makes use of (H.17); and (iii) is valid with the proviso that $np \gtrsim \kappa^7\mu^2 r^{9/2}\log^2(n+d)$.

- Regarding α_2, we know that

$$\begin{aligned}
\alpha_2 &\overset{\text{(i)}}{\lesssim} \frac{\kappa^3\mu r^{9/4}\log(n+d)}{\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l)}\frac{1}{\sigma_r^{\star 2}}\left(\frac{1}{p^2}\sqrt{\frac{\mu r\log^2(n+d)}{n^2 d}}\sigma_1^{\star} + \frac{\omega_{\max}}{np^2}\log(n+d)\right)\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right)\left\|\boldsymbol{U}^{\star}\right\|_{2,\infty} \\
&\lesssim \frac{1}{\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l)}\frac{1}{\sigma_r^{\star 2}}\left(\frac{\kappa^3\mu^2 r^{13/4}\log^2(n+d)}{ndp^2}\sigma_1^{\star} + \frac{\omega_{\max}}{n\sqrt{d}p^2}\kappa^3\mu^{3/2}r^{11/4}\log^2(n+d)\right)\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right) \\
&\overset{\text{(ii)}}{\lesssim} \frac{1}{\sigma_r^{\star 2}}\left(\frac{\kappa^3\mu^2 r^{13/4}\log^2(n+d)}{ndp^2}\sigma_1^{\star} + \frac{\omega_{\max}}{n\sqrt{d}p^2}\kappa^3\mu^{3/2}r^{11/4}\log^2(n+d)\right) \\
&\qquad \cdot\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right)\left[\frac{1}{\sqrt{ndp^2\kappa}\sigma_1^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right)\right]^{-1} \\
&\lesssim \frac{1}{\sigma_r^{\star}}\left(\frac{\kappa^4\mu^2 r^{13/4}\log^2(n+d)}{\sqrt{ndp^2}}\sigma_1^{\star} + \frac{\omega_{\max}}{\sqrt{np^2}}\kappa^4\mu^{3/2}r^{11/4}\log^2(n+d)\right) \overset{\text{(iii)}}{\lesssim} \frac{1}{\sqrt{\log(n+d)}}.
\end{aligned}$$

  Here, (i) follows from (G.4), (G.9) and (H.18); (ii) makes use of (H.17); and (iii) holds provided that $ndp^2 \gtrsim \kappa^8\mu^4 r^{13/2}\log^5(n+d)$ and

$$\frac{\omega_{\max}}{\sigma_r^{\star}\sqrt{np^2}} \lesssim \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} + \frac{1}{\sqrt{ndp}} \lesssim \frac{1}{\kappa^4\mu^{3/2}r^{11/4}\log^{5/2}(n+d)},$$

  which can be guaranteed by $ndp^2 \gtrsim \kappa^8\mu^3 r^{11/2}\log^5(n+d)$ and

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\kappa^4\mu^{3/2}r^{11/4}\log^{5/2}(n+d)}.$$

- When it comes to α_3, we have

$$\alpha_3 \overset{\text{(i)}}{\lesssim} \frac{\kappa^3\mu r^{9/4}\log^{3/2}(n+d)}{\lambda_{\min}^{1/2}(\widetilde{\boldsymbol{\Sigma}}_l)}\frac{1}{\sigma_r^{\star 2}}\left(\sqrt{\frac{\mu r\log(n+d)}{ndp}}\sigma_1^{\star} + \frac{\omega_{\max}}{\sqrt{np}}\right)\frac{1}{p}\sqrt{\frac{\log(n+d)}{n}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right)\sqrt{d}\left\|\boldsymbol{U}^{\star}\right\|_{2,\infty}$$

$$\lesssim \frac{1}{\lambda_{\min}^{1/2}(\widetilde{\mathbf{\Sigma}}_l)}\frac{1}{\sigma_r^{\star 2}}\left(\frac{\kappa^3\mu^2 r^{13/4}\log^{5/2}(n+d)}{\sqrt{n^2dp^3}}\sigma_1^\star+\omega_{\max}\frac{\kappa^3\mu^{3/2}r^{11/4}\log^2(n+d)}{\sqrt{n^2p^3}}\right)\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2+\omega_l^\star\right)$$

$$\overset{\text{(ii)}}{\lesssim} \frac{1}{\sigma_r^{\star 2}}\left(\frac{\kappa^3\mu^2 r^{13/4}\log^{5/2}(n+d)}{\sqrt{n^2dp^3}}\sigma_1^\star+\omega_{\max}\frac{\kappa^3\mu^{3/2}r^{11/4}\log^2(n+d)}{\sqrt{n^2p^3}}\right)$$

$$\cdot\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2+\omega_l^\star\right)\left[\frac{1}{\sqrt{np}\sigma_1^\star}\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2+\omega_l^\star\right)\right]^{-1}$$

$$\lesssim \frac{\kappa^4\mu^2 r^{13/4}\log^{5/2}(n+d)}{\sqrt{ndp^2}}\sigma_1^\star+\frac{\omega_{\max}}{\sigma_r^\star}\frac{\kappa^{7/2}\mu^{3/2}r^{11/4}\log^2(n+d)}{\sqrt{np^2}}\overset{\text{(iii)}}{\lesssim}\frac{1}{\sqrt{\log(n+d)}}.$$

Here, (i) follows from (G.8), (G.9) and (H.18); (ii) makes use of (H.17); and (iii) holds provided that $ndp^2 \gtrsim \kappa^8\mu^4 r^{13/2}\log^6(n+d)$ and

$$\frac{\omega_{\max}}{\sigma_r^\star\sqrt{np^2}}\lesssim\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}+\frac{1}{\sqrt{ndp}}\lesssim\frac{1}{\kappa^{7/2}\mu^{3/2}r^{11/4}\log^{5/2}(n+d)},$$

which can be guaranteed by $ndp^2 \gtrsim \kappa^7\mu^3 r^{11/2}\log^5(n+d)$ and

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}\lesssim\frac{1}{\kappa^{7/2}\mu^{3/2}r^{11/4}\log^{5/2}(n+d)}.$$

In addition, we have learned from (H.17) and (H.18) that

$$B_l\lesssim\frac{1}{p}\sqrt{\frac{\log(n+d)}{n}}\sqrt{np}\sigma_1^\star\lambda_{\min}^{1/2}(\widetilde{\mathbf{\Sigma}}_l)\asymp\sqrt{\frac{\log(n+d)}{p}}\sigma_1^\star\lambda_{\min}^{1/2}(\widetilde{\mathbf{\Sigma}}_l). \tag{H.19}$$

Therefore the term β defined in (H.16) can be bounded by

$$\beta\overset{\text{(i)}}{\lesssim} r^{1/4}\lambda_{\min}^{-3/2}(\widetilde{\mathbf{\Sigma}}_l)B_l^3\left(\frac{1}{\sigma_r^\star}\sqrt{\frac{r\log(n+d)}{n}}+\frac{1}{\sigma_r^{\star 2}}B\sqrt{d}\right)^3(n+d)^{-99}$$

$$\overset{\text{(ii)}}{\lesssim}\kappa^{3/2}r^{1/4}\frac{\log^{3/2}(n+d)}{p^{3/2}}\left(1+\frac{1}{\sigma_r^\star}B\sqrt{d}\right)^3(n+d)^{-99}$$

$$\overset{\text{(iii)}}{\lesssim}\kappa^{3/2}r^{1/4}\frac{\log^{3/2}(n+d)}{p^{3/2}}d^{3/2}(n+d)^{-99}\overset{\text{(iv)}}{\lesssim}(n+d)^{-50}.$$

Here (i) utilizes (G.4) and (G.7); (ii) follows from (H.19); (iii) holds as long as $B\lesssim\sigma_r^\star$, which can be guaranteed by $ndp^2\gtrsim\kappa\mu r\log(n+d)$ and

$$\frac{\omega_{\max}}{p\sigma_r^\star}\sqrt{\frac{1}{n}}\lesssim\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}+\frac{1}{\sqrt{ndp}}\lesssim\frac{1}{\sqrt{\log(n+d)}};$$

and (iv) is valid provided that $np\gtrsim\kappa r\log(n+d)$.

Therefore, the preceding bounds allow one to conclude that

$$\sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left(\boldsymbol{Z}_{l,\cdot}\in\mathcal{C}\middle|\boldsymbol{F}\right)-\mathbb{P}\left(\mathcal{N}\left(\mathbf{0},\widetilde{\mathbf{\Sigma}}_l\right)\in\mathcal{C}\middle|\boldsymbol{F}\right)\right|\lesssim\alpha_1+\alpha_2+\alpha_3+\beta\lesssim\frac{1}{\sqrt{\log(n+d)}}, \tag{H.20}$$

provided that $np\gtrsim\kappa^7\mu^2 r^{9/2}\log^2(n+d)$, $ndp^2\gtrsim\kappa^8\mu^4 r^{13/2}\log^6(n+d)$ and

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}\lesssim\frac{1}{\kappa^4\mu^{3/2}r^{11/4}\log^{5/2}(n+d)}.$$

Step 2: bounding TV distance between Gaussian distributions. In view of Lemma 38, we know that

$$\left\|\widetilde{\boldsymbol{\Sigma}}_l - \boldsymbol{\Sigma}_{U,l}^{\star}\right\| \lesssim \sqrt{\frac{\kappa^8\mu^2 r^3 \log^3(n+d)}{n}}\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) \tag{H.21}$$

holds on the event $\mathcal{E}_{\mathsf{good}}$. In addition, $\boldsymbol{\Sigma}_{U,l}^{\star}$ is assumed to be non-singular. As a result, conditional on $\boldsymbol{F}$ and the event $\mathcal{E}_{\mathsf{good}}$, one can invoke Theorem 21 and (H.21) to arrive at

$$\begin{aligned}
&\sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left(\mathcal{N}(\mathbf{0},\widetilde{\boldsymbol{\Sigma}}_l)\in\mathcal{C}\mid\boldsymbol{F}\right)-\mathbb{P}\left(\mathcal{N}\left(\mathbf{0},\boldsymbol{\Sigma}_{U,l}^{\star}\right)\in\mathcal{C}\right)\right| \leq \mathsf{TV}\left(\mathcal{N}(\mathbf{0},\widetilde{\boldsymbol{\Sigma}}_l),\mathcal{N}\left(\mathbf{0},\boldsymbol{\Sigma}_{U,l}^{\star}\right)\right)\\
&\asymp \left\|\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\widetilde{\boldsymbol{\Sigma}}_l\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}-\boldsymbol{I}_d\right\|_{\mathrm{F}} = \left\|\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\left(\widetilde{\boldsymbol{\Sigma}}_l-\boldsymbol{\Sigma}_{U,l}^{\star}\right)\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\right\|_{\mathrm{F}}\\
&\lesssim \left\|\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\right\|\left\|\widetilde{\boldsymbol{\Sigma}}_l-\boldsymbol{\Sigma}_{U,l}^{\star}\right\|_{\mathrm{F}}\left\|\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\right\| \lesssim \sqrt{r}\left\|\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1}\right\|\left\|\widetilde{\boldsymbol{\Sigma}}_l-\boldsymbol{\Sigma}_{U,l}^{\star}\right\|\\
&\lesssim \frac{\sqrt{r}}{\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)}\cdot\sqrt{\frac{\kappa^8\mu^2 r^4\log^3(n+d)}{n}}\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)\\
&\lesssim \sqrt{\frac{\kappa^8\mu^2 r^4\log^3(n+d)}{n}}.
\end{aligned}$$

where $\mathsf{TV}(\cdot,\cdot)$ represents the total-variation distance between two distributions (Tsybakov and Zaiats, 2009). The above inequality taken together with (H.20) yields

$$\sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left(\boldsymbol{Z}_{l,\cdot}\in\mathcal{C}\middle|\boldsymbol{F}\right)-\mathbb{P}\left(\mathcal{N}\left(\mathbf{0},\boldsymbol{\Sigma}_l^{\star}\right)\in\mathcal{C}\right)\right| \lesssim \frac{1}{\sqrt{\log(n+d)}} \tag{H.22}$$

on the event $\mathcal{E}_{\mathsf{good}}$, provided that $n \gtrsim \kappa^8\mu^2 r^4\log^4(n+d)$.

Step 3: accounting for higher-order errors. Let us assume that the high-probability event $\mathcal{E}_{\mathsf{good}}$ happens. Lemma 37 tells us that

$$\mathbb{P}\left(\|\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}-\boldsymbol{Z}\|_{2,\infty}\lesssim\zeta_{\mathsf{2nd}}\mid\boldsymbol{F}\right)\geq 1-O\left((n+d)^{-10}\right),$$

thus indicating that

$$\mathbb{P}\left(\left\|\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}-\boldsymbol{Z})\right\|_{2,\infty}\leq\zeta\mid\boldsymbol{F}\right)\geq 1-O\left((n+d)^{-10}\right).$$

Here, we define the quantity ζ as

$$\zeta \coloneqq c_{\zeta}\zeta_{\mathsf{2nd}}\big(\lambda_{\min}(\boldsymbol{\Sigma}_{U,l}^{\star})\big)^{-\frac{1}{2}}$$

for some large enough constant $c_{\zeta}>0$. Follow the same analysis as Step 3 in Appendix E.1.3 (the proof of Lemma 9), we can demonstrate that

$$\sup_{\mathcal{C}\in\mathscr{C}^r}\left|\mathbb{P}\left((\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star})_{l,\cdot}\in\mathcal{C}\mid\boldsymbol{F}\right)-\mathbb{P}\left(\mathcal{N}\left(\mathbf{0},\boldsymbol{\Sigma}_l^{\star}\right)\in\mathcal{C}\right)\right|\lesssim\frac{1}{\sqrt{\log(n+d)}}. \tag{H.23}$$

provided that $\zeta \lesssim 1/(r^{1/4}\log^{1/2}(n+d))$.

It remains to verify the conditions required to guarantee $\zeta \lesssim 1/(r^{1/4}\log^{1/2}(n+d))$. More generally, we shall check that under what conditions we can guarantee

$$\zeta_{\mathsf{2nd}} \lesssim \delta\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)$$

for some $\delta>0$. Recall from Lemma 37 that

$$\zeta_{\mathsf{2nd}} \lesssim \underbrace{\frac{\kappa^2\mu r\log(n+d)}{d}\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}}_{=:\gamma_1}+\underbrace{\frac{\zeta_{\mathsf{1st}}^2}{\sigma_r^{\star 4}}\sqrt{\frac{\kappa^3\mu r\log(n+d)}{d}}}_{=:\gamma_2},$$

and from Lemma 38 that

$$\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) \gtrsim \frac{1}{\sqrt{np}\sigma_1^{\star}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{\omega_l^{\star}}{\sqrt{np}\sigma_1^{\star}} + \frac{1}{\sqrt{ndp^2\kappa}\sigma_1^{\star}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{1}{\sqrt{ndp^2\kappa}\sigma_1^{\star}}\omega_l^{\star} + \frac{\omega_{\min}^2}{\sqrt{np^2}\sigma_1^{\star 2}}.$$

- Regarding γ_1, we can derive

$$\begin{aligned}\gamma_1 &\asymp \underbrace{\frac{\kappa^3\mu^2r^2\log^3(n+d)}{\sqrt{nd^3p^2}}}_{=:\gamma_{1,1}} + \underbrace{\kappa^2\mu r\log^2(n+d)\frac{\omega_{\max}^2}{\sigma_r^{\star 2}}\sqrt{\frac{1}{ndp^2}}}_{=:\gamma_{1,2}} \\ &\quad + \underbrace{\frac{\kappa^3\mu^{3/2}r^{3/2}\log^2(n+d)}{\sqrt{np}d}}_{=:\gamma_{1,3}} + \underbrace{\kappa^{5/2}\mu r\log^{3/2}(n+d)\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{1}{ndp}}}_{=:\gamma_{1,4}} \\ &\lesssim \delta\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),\end{aligned}$$

  where the last line holds since

$$\begin{aligned}\gamma_{1,1} &\lesssim \delta\frac{1}{\sqrt{ndp^2\kappa}\sigma_1^{\star}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2, \\ \gamma_{1,2} &\lesssim \delta\frac{\omega_{\min}^2}{\sqrt{np^2}\sigma_1^{\star 2}}, \\ \gamma_{1,3} &\lesssim \delta\frac{1}{\sqrt{np}\sigma_1^{\star}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2, \\ \gamma_{1,4} &\lesssim \delta\frac{\omega_l^{\star}}{\sqrt{np}\sigma_1^{\star}},\end{aligned}$$

  provided that $d \gtrsim \delta^{-2}\kappa^6\mu^2r^2\log^4(n+d)$ and

$$\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \gtrsim \delta^{-1}\frac{\kappa^{7/2}\mu^2r^2\log^3(n+d)}{d}\sigma_1^{\star}.$$

- When it comes to γ_2, we observe that

$$\begin{aligned}\gamma_2 &\asymp \underbrace{\frac{\kappa^{7/2}\mu^{5/2}r^{5/2}\log^{9/2}(n+d)}{nd^{3/2}p^2}}_{=:\gamma_{2,1}} + \underbrace{\kappa^{3/2}\mu^{1/2}r^{1/2}\log^{5/2}(n+d)\frac{\omega_{\max}^4}{p^2\sigma_r^{\star 4}}\frac{d^{1/2}}{n}}_{=:\gamma_{2,2}} \\ &\quad + \underbrace{\frac{\kappa^{7/2}\mu^{3/2}r^{3/2}\log^{5/2}(n+d)}{nd^{1/2}p}}_{=:\gamma_{2,3}} + \underbrace{\frac{\omega_{\max}^2}{\sigma_r^{\star 2}}\frac{\sqrt{d}\kappa^{5/2}\mu^{1/2}r^{1/2}\log^{3/2}(n+d)}{np}}_{=:\gamma_{2,4}} \\ &\lesssim \delta\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),\end{aligned}$$

  where the last line holds since

$$\begin{aligned}\gamma_{2,1} &\lesssim \delta\frac{1}{\sqrt{ndp^2\kappa}\sigma_1^{\star}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2, \\ \gamma_{2,2}+\gamma_{2,4} &\lesssim \delta\frac{\omega_{\min}^2}{\sqrt{np^2}\sigma_1^{\star 2}}, \\ \gamma_{2,3} &\lesssim \delta\frac{1}{\sqrt{np}\sigma_1^{\star}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2,\end{aligned}$$

  provided that

$$\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}} \lesssim \frac{\delta}{\kappa^3\mu^{1/2}r^{1/2}\log^{3/2}(n+d)}, \qquad \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{\delta}{\kappa^{5/2}\mu^{1/2}r^{1/2}\log^{5/2}(n+d)},$$

$$\text{and} \qquad \left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2} \gtrsim \delta^{-1}\frac{\kappa^{4}\mu^{5/2}r^{5/2}\log^{9/2}(n+d)}{\sqrt{nd^{2}p^{2}}}\sigma_{1}^{\star} + \delta^{-1}\frac{\kappa^{7/2}\mu^{3/2}r^{3/2}\log^{5/2}(n+d)}{\sqrt{ndp}}\sigma_{1}^{\star}.$$

Therefore, $\zeta_{\mathsf{2nd}} \lesssim \delta\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)$ is guaranteed to hold as long as $d \gtrsim \delta^{-2}\kappa^{6}\mu^{2}r^{2}\log^{4}(n+d)$,

$$\frac{\omega_{\max}}{\sigma_{r}^{\star}}\sqrt{\frac{d}{np}} \lesssim \frac{\delta}{\kappa^{3}\mu^{1/2}r^{1/2}\log^{3/2}(n+d)}, \qquad \frac{\omega_{\max}^{2}}{p\sigma_{r}^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{\delta}{\kappa^{5/2}\mu^{1/2}r^{1/2}\log^{5/2}(n+d)},$$

and

$$\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2} \gtrsim \delta^{-1}\left[\frac{\kappa^{4}\mu^{2}r^{2}\log^{9/2}(n+d)}{\sqrt{ndp^{2}}} + \frac{\kappa^{7/2}\mu r\log^{5/2}(n+d)}{\sqrt{np}}\sigma_{1}^{\star} + \frac{\kappa^{7/2}\mu^{3/2}r^{3/2}\log^{3}(n+d)}{\sqrt{d}}\right]\sqrt{\frac{\mu r}{d}}\sigma_{1}^{\star}.$$

By taking $\delta = 1/(r^{1/4}\log^{1/2}(n+d))$, we see that $\zeta \lesssim 1/(r^{1/4}\log^{1/2}(n+d))$ holds provided that $d \gtrsim \kappa^{6}\mu^{2}r^{5/2}\log^{5}(n+d)$,

$$\frac{\omega_{\max}}{\sigma_{r}^{\star}}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\kappa^{3}\mu^{1/2}r^{3/4}\log^{2}(n+d)}, \qquad \frac{\omega_{\max}^{2}}{p\sigma_{r}^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\kappa^{5/2}\mu^{1/2}r^{3/4}\log^{3}(n+d)},$$

and

$$\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2} \gtrsim \left[\frac{\kappa^{4}\mu^{2}r^{9/4}\log^{5}(n+d)}{\sqrt{ndp^{2}}} + \frac{\kappa^{7/2}\mu r^{5/4}\log^{3}(n+d)}{\sqrt{np}}\sigma_{1}^{\star} + \frac{\kappa^{7/2}\mu^{3/2}r^{7/4}\log^{7/2}(n+d)}{\sqrt{d}}\right]\sqrt{\frac{\mu r}{d}}\sigma_{1}^{\star}.$$

Step 4: distributional characterization of $\boldsymbol{Z}_{l,\cdot}$. Following the same analysis as Step 4 in Appendix E.1.3 (the proof of Lemma 9), we can derive from (H.23) that

$$\sup_{\mathcal{C}\in\mathscr{C}^{r}}\left|\mathbb{P}\left((\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star})_{l,\cdot}\in\mathcal{C}\right) - \mathbb{P}\left(\mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}_{l}\right)\in\mathcal{C}\right)\right| \lesssim \frac{1}{\sqrt{\log(n+d)}}$$

as claimed.

H.2 Auxiliary lemmas for Theorem 16

H.2.1 Proof of Lemma 40

This lemma is concerned with several different quantities, which we seek to control separately.

Bounding $\|(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star})_{l,\cdot}\|_{2}$ and $\|\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}\|_{2,\infty}$. In view of Lemma 37, we can begin with the decomposition

$$\begin{aligned}\left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{R}-\boldsymbol{U}_{l,\cdot}^{\star}\right\|_{2} &\leq \|\boldsymbol{Z}_{l,\cdot}\|_{2} + \zeta_{\mathsf{2nd}} = \left\|\left[\boldsymbol{E}\boldsymbol{M}^{\natural\top} + \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^{\top}\right)\right]_{l,\cdot}\boldsymbol{U}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-2}\boldsymbol{Q}^{\top}\right\|_{2} + \zeta_{\mathsf{2nd}} \\ &\leq \underbrace{\left\|\boldsymbol{E}_{l,\cdot}\boldsymbol{V}^{\natural}(\boldsymbol{\Sigma}^{\natural})^{-1}\boldsymbol{Q}^{\top}\right\|_{2}}_{=:\alpha_{1}} + \underbrace{\left\|\boldsymbol{E}_{l,\cdot}\left[\mathcal{P}_{-l,\cdot}\left(\boldsymbol{E}\right)\right]^{\top}\boldsymbol{U}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-2}\boldsymbol{Q}^{\top}\right\|_{2}}_{=:\alpha_{2}} + \zeta_{\mathsf{2nd}},\end{aligned}$$

where $\boldsymbol{Z}$ and ζ_{2nd} are defined in (G.13). Note that we have shown in Appendix E.2.1 that with probability exceeding $1-O((n+d)^{-10})$

$$\alpha_{1} \lesssim \frac{1}{\sigma_{r}^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2} + \omega_{l}^{\star}\right)\sqrt{\frac{r\log^{2}(n+d)}{np}}.$$

It thus boils down to bounding α_{2}.

In order to do so, we know from (G.4) that on the event $\mathcal{E}_{\mathsf{good}}$,

$$\alpha_2 \, \mathbb{1}_{\mathcal{E}_{\mathsf{good}}} \lesssim \frac{1}{\sigma_r^{\star 2}} \left\| \boldsymbol{E}_{l,\cdot} \left[\mathcal{P}_{-l,\cdot} \left(\boldsymbol{E} \right) \right]^\top \boldsymbol{U}^\natural \right\|_2 \asymp \frac{1}{\sigma_r^{\star 2}} \left\| \boldsymbol{E}_{l,\cdot} \left[\mathcal{P}_{-l,\cdot} \left(\boldsymbol{E} \right) \right]^\top \boldsymbol{U}^\star \right\|_2, \tag{H.24}$$

where the last relation arises from the fact $\boldsymbol{U}^\natural = \boldsymbol{U}^\star \boldsymbol{Q}$ for some orthonormal matrix $\boldsymbol{Q}$ (see (D.4)). Therefore, it suffices to bound

$$\left\| \boldsymbol{E}_{l,\cdot} \left[\mathcal{P}_{-l,\cdot} \left(\boldsymbol{E} \right) \right]^\top \boldsymbol{U}^\star \right\|_2 = \left\| \sum_{j=1}^n E_{l,j} \boldsymbol{C}_{j,\cdot} \boldsymbol{U}^\star \right\|_2 = \left\| \sum_{j=1}^n \boldsymbol{X}_j \right\|_2,$$

where $\boldsymbol{C} = [\mathcal{P}_{-l,\cdot}(\boldsymbol{E})]^\top$ and $\boldsymbol{X}_j = E_{l,j} \boldsymbol{C}_{j,\cdot} \boldsymbol{U}^\star$ for $j \in [n]$. Recognizing that $\sum_{j=1}^n \boldsymbol{X}_j$ is a sum of independent random vectors (conditional on $\boldsymbol{F}$), we can employ the truncated Bernstein inequality (see, e.g., Chen et al. (2020c, Theorem 3.1.1)) to bound the above quantity. From now on, we will always assume occurrence of $\mathcal{E}_{\mathsf{good}}$ when bounding $\|(\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|_2$ (recall that $\mathcal{E}_{\mathsf{good}}$ is $\sigma(\boldsymbol{F})$-measurable).

- It is first observed that

$$\max_{1 \le j \le n} \|\boldsymbol{X}_j\|_2 \le \max_{1 \le j \le n} |E_{l,j}| \, \|\boldsymbol{C}_{j,\cdot} \boldsymbol{U}^\star\|_2 .$$

  Recall that for each $j \in [n]$,

$$E_{l,j} = \frac{1}{\sqrt{np}} \left[\left(\delta_{l,j} - 1 \right) \boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star \boldsymbol{f}_j + N_{l,j} \right]$$

  with $\delta_{l,j} = \mathbb{1}_{(l,j) \in \Omega}$. It is seen from (G.11) that

$$|E_{l,j}| \le \widetilde{C}_1 \sqrt{\frac{1}{np^2}} \left(\left\| \boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 + \omega_l^\star \right) \sqrt{\log\left(n+d\right)} =: L_E$$

  for some sufficiently large constant $\widetilde{C}_1 > 0$. In addition, for each $j \in [n]$, we can write $\boldsymbol{C}_{j,\cdot} \boldsymbol{U}^\star = \sum_{i: i \ne l} E_{i,j} \boldsymbol{U}_{i,\cdot}^\star$. It is then straightforward to compute

$$\begin{aligned} L_1 &\coloneqq \max_{i: i \ne l} \left\| E_{i,j} \boldsymbol{U}_{i,\cdot}^\star \right\|_2 \lesssim \sqrt{\frac{\log\left(n+d\right)}{np^2}} \left(\|\boldsymbol{U}^\star \boldsymbol{\Sigma}^\star\|_{2,\infty} + \omega_{\max} \right) \sqrt{\frac{\mu r}{d}} \\ &\asymp \sqrt{\frac{\mu r \log\left(n+d\right)}{ndp^2}} \left(\|\boldsymbol{U}^\star \boldsymbol{\Sigma}^\star\|_{2,\infty} + \omega_{\max} \right) \lesssim \sqrt{\frac{\mu r \log\left(n+d\right)}{dp}} \sigma_{\mathsf{ub}}, \\ V_1 &\coloneqq \sum_{i: i \ne l} \mathbb{E}\left[E_{i,j}^2 \left\| \boldsymbol{U}_{i,\cdot}^\star \right\|_2^2 \right] \le \sum_{i: i \ne l} \sigma_{i,j}^2 \left\| \boldsymbol{U}_{i,\cdot}^\star \right\|_2^2 \le \sigma_{\mathsf{ub}}^2 \|\boldsymbol{U}^\star\|_{\mathrm{F}}^2 = r \sigma_{\mathsf{ub}}^2, \end{aligned}$$

  where σ_{ub} is defined in (D.14). Apply the matrix Bernstein inequality (Tropp, 2015, Theorem 6.1.1) to achieve

$$\mathbb{P}\left(\|\boldsymbol{C}_{j,\cdot} \boldsymbol{U}^\star\|_2 \ge t \,\middle|\, \boldsymbol{F} \right) \le (d+1) \exp\left(\frac{-t^2/2}{V_1 + L_1 t/3} \right).$$

  If we take

$$L_C \coloneqq \widetilde{C}_2 \left[\sqrt{V_1 \log\left(n+d\right)} + L_1 \log\left(n+d\right) \right].$$

  for some sufficiently large constant $\widetilde{C}_2 > 0$, then the above inequality tells us that for each $j \in [n]$

$$\mathbb{P}\left(\|\boldsymbol{C}_{j,\cdot} \boldsymbol{U}^\star\|_2 \ge L_C \,\middle|\, \boldsymbol{F} \right) \le \left(n+d\right)^{-20}.$$

  Consequently, by setting

$$L \coloneqq L_E L_C \asymp \sqrt{\frac{\log\left(n+d\right)}{np^2}} \left(\left\| \boldsymbol{U}_{l,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 + \omega_l^\star \right) \left[\sqrt{V_1 \log\left(n+d\right)} + L_1 \log\left(n+d\right) \right]$$

$$\lesssim \sqrt{\frac{r\log^2(n+d)}{np^2}}\sigma_{\mathsf{ub}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_l^{\star}\right)\left[1+\sqrt{\frac{\mu}{dp}}\right],$$

we can further derive

$$\mathbb{P}\left(\|\boldsymbol{X}_j\|\geq L\,\middle|\,\boldsymbol{F}\right)\leq \mathbb{P}\left(\|\boldsymbol{C}_{j,\cdot}\boldsymbol{U}^{\star}\|_2\geq L_C\,\middle|\,\boldsymbol{F}\right)\leq (n+d)^{-20}\coloneqq q_0.$$

- Next, we can develop the following upper bound

$$\begin{aligned}
q_1 &\coloneqq \left\|\mathbb{E}\left[\boldsymbol{X}_j\,\mathbb{1}_{\|\boldsymbol{X}_j\|\geq L}\,\middle|\,\boldsymbol{F}\right]\right\| \overset{\text{(i)}}{\leq} \mathbb{E}\left[\|\boldsymbol{X}_j\|\,\mathbb{1}_{\|\boldsymbol{X}_j\|\geq L}\,\middle|\,\boldsymbol{F}\right]\\
&\overset{\text{(ii)}}{\leq} L_E\mathbb{E}\left[\|\boldsymbol{C}_{j,\cdot}\boldsymbol{U}^{\star}\|_2\,\mathbb{1}_{\|\boldsymbol{C}_{j,\cdot}\boldsymbol{U}^{\star}\|_2\geq L_C}\,\middle|\,\boldsymbol{F}\right]\\
&= L_E\int_0^{\infty}\mathbb{P}\left(\|\boldsymbol{C}_{j,\cdot}\boldsymbol{U}^{\star}\|_2\,\mathbb{1}_{\|\boldsymbol{C}_{j,\cdot}\boldsymbol{U}^{\star}\|_2\geq L_C}\geq t\,\middle|\,\boldsymbol{F}\right)\mathrm{d}t\\
&= L_E\int_0^{L_C}\mathbb{P}\left(\|\boldsymbol{C}_{j,\cdot}\boldsymbol{U}^{\star}\|_2\geq L_C\,\middle|\,\boldsymbol{F}\right)\mathrm{d}t + L_E\int_{L_C}^{\infty}\mathbb{P}\left(\|\boldsymbol{C}_{j,\cdot}\boldsymbol{U}^{\star}\|_2\geq t\,\middle|\,\boldsymbol{F}\right)\mathrm{d}t\\
&\leq L_EL_C(n+d)^{-20}+L_E\int_{L_C}^{\infty}\mathbb{P}\left(\|\boldsymbol{C}_{j,\cdot}\boldsymbol{U}^{\star}\|_2\geq t\,\middle|\,\boldsymbol{F}\right)\mathrm{d}t.
\end{aligned}$$

Here, (i) makes use of Jensen's inequality, while (ii) holds since the conditions $\|\boldsymbol{X}_j\|\geq L$ and $|E_{l,j}|\leq L_E$ taken together imply that $\|\boldsymbol{C}_{j,\cdot}\boldsymbol{U}^{\star}\|_2\geq L_C$. Note that for any $t\geq L_C$, we have $t\gg\sqrt{V_1\log(n+d)}$ and $t\gg L_1\log(n+d)$ as long as $\widetilde{C}_2$ is sufficiently large. Therefore, we arrive at

$$\mathbb{P}\left(\|\boldsymbol{C}_{j,\cdot}\boldsymbol{U}^{\star}\|_2\geq t\,\middle|\,\boldsymbol{F}\right)\leq (d+1)\exp\left(-t/\max\left\{4\sqrt{V_1/\log(n+d)},4L_1/3\right\}\right),$$

which immediately gives

$$\begin{aligned}
&\int_{L_C}^{\infty}\mathbb{P}\left(\|\boldsymbol{C}_{j,\cdot}\boldsymbol{U}^{\star}\|_2\geq t\,\middle|\,\boldsymbol{F}\right)\mathrm{d}t\leq (d+1)\int_{L_C}^{\infty}\exp\left(-t/\max\left\{4\sqrt{V_1/\log(n+d)},4L_1/3\right\}\right)\mathrm{d}t\\
&\quad\leq (d+1)\max\left\{4\sqrt{V_1/\log(n+d)},4L_1/3\right\}\exp\left(-L_C/\max\left\{4\sqrt{V_1/\log(n+d)},4L_1/3\right\}\right)\\
&\quad\leq 4(d+1)\widetilde{C}_2^{-1}L_C\exp\left(-4\widetilde{C}_2\log(n+d)\right)\leq L_C(n+d)^{-20},
\end{aligned}$$

provided that $\widetilde{C}_2$ is sufficiently large. As a consequence, we reach

$$q_1\leq 2L_EL_C(n+d)^{-20}\leq L(n+d)^{-19}.$$

- Finally, let us calculate the variance statistics as follows

$$\begin{aligned}
v &\coloneqq \sum_{j=1}^n\mathbb{E}\left[\|\boldsymbol{X}_j\|_2^2\,\middle|\,\boldsymbol{F}\right]=\sum_{j=1}^n\mathbb{E}\left[\|E_{l,j}\boldsymbol{C}_{j,\cdot}\boldsymbol{U}^{\star}\|_2^2\,\middle|\,\boldsymbol{F}\right]=\sum_{j=1}^n\sigma_{l,j}^2\mathbb{E}\left[\|\boldsymbol{C}_{j,\cdot}\boldsymbol{U}^{\star}\|_2^2\,\middle|\,\boldsymbol{F}\right]\\
&\leq \max_{1\leq j\leq n}\sigma_{l,j}^2\mathbb{E}\left[\|\boldsymbol{C}\boldsymbol{U}^{\star}\|_{\mathrm{F}}^2\,\middle|\,\boldsymbol{F}\right]=\max_{1\leq j\leq n}\sigma_{l,j}^2\mathbb{E}\left[\left\|\left[\mathcal{P}_{-l,\cdot}(\boldsymbol{E})\right]^{\top}\boldsymbol{U}^{\star}\right\|_{\mathrm{F}}^2\,\middle|\,\boldsymbol{F}\right]\\
&\leq \frac{\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\log(n+d)+\omega_l^{\star 2}}{np}\mathbb{E}\left[\left\|\left[\mathcal{P}_{-l,\cdot}(\boldsymbol{E})\right]^{\top}\boldsymbol{U}^{\star}\right\|_{\mathrm{F}}^2\,\middle|\,\boldsymbol{F}\right],
\end{aligned}$$

where the last inequality results from the following relation

$$\max_{1\leq j\leq n}\sigma_{l,j}^2\asymp\frac{1-p}{np}\max_{1\leq j\leq n}\left(\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_j\right)^2+\frac{\omega_l^{\star 2}}{np}\lesssim\frac{\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\log(n+d)+\omega_l^{\star 2}}{np}.$$

Notice that

$$\begin{aligned}\mathbb{E}\left[\left\|\left[\mathcal{P}_{-l,\cdot}\left(\boldsymbol{E}\right)\right]^{\top}\boldsymbol{U}^{\star}\right\|_{\mathrm{F}}^{2}\mid\boldsymbol{F}\right] &= \mathbb{E}\left[\mathsf{tr}\left(\boldsymbol{U}^{\star\top}\mathcal{P}_{-l,\cdot}\left(\boldsymbol{E}\right)\left[\mathcal{P}_{-l,\cdot}\left(\boldsymbol{E}\right)\right]^{\top}\boldsymbol{U}^{\star}\right)\mid\boldsymbol{F}\right]\\ &= \mathbb{E}\left[\mathsf{tr}\left(\mathcal{P}_{-l,\cdot}\left(\boldsymbol{E}\right)\left[\mathcal{P}_{-l,\cdot}\left(\boldsymbol{E}\right)\right]^{\top}\boldsymbol{U}^{\star}\boldsymbol{U}^{\star\top}\right)\mid\boldsymbol{F}\right]\\ &= \mathsf{tr}\left\{\boldsymbol{D}\boldsymbol{U}^{\star}\boldsymbol{U}^{\star\top}\right\},\end{aligned}$$

where $\boldsymbol{D} = \mathbb{E}\big[\mathcal{P}_{-l,\cdot}(\boldsymbol{E})[\mathcal{P}_{-l,\cdot}(\boldsymbol{E})]^{\top}|\boldsymbol{F}\big] \in \mathbb{R}^{d\times d}$ is a diagonal matrix with the i-th diagonal entry given by

$$D_{i,i} = \begin{cases}\sum_{j=1}^{n}\sigma_{i,j}^{2} & \text{if } i\neq l,\\ 0 & \text{if } i=l.\end{cases}$$

As a result, we have demonstrated that

$$\begin{aligned}\mathbb{E}\left[\left\|\left[\mathcal{P}_{-l,\cdot}\left(\boldsymbol{E}\right)\right]^{\top}\boldsymbol{U}^{\star}\right\|_{\mathrm{F}}^{2}\mid\boldsymbol{F}\right] &= \mathsf{tr}\left\{\boldsymbol{D}\boldsymbol{U}^{\star}\boldsymbol{U}^{\star\top}\right\} = \sum_{i=1}^{d}D_{i,i}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}^{2}\\ &\leq \sum_{i=1}^{d}\left(\sum_{j=1}^{n}\sigma_{i,j}^{2}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}^{2}\leq n\sigma_{\mathsf{ub}}^{2}\left\|\boldsymbol{U}^{\star}\right\|_{\mathrm{F}}^{2} = nr\sigma_{\mathsf{ub}}^{2}.\end{aligned}$$

Taking the above inequalities collectively yields

$$v \lesssim \frac{r\sigma_{\mathsf{ub}}^{2}}{p}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\log\left(n+d\right)+\omega_{l}^{\star2}\right).$$

Equipped with the above quantities, we are ready to invoke the truncated matrix Bernstein inequality (Chen et al., 2020c, Theorem 3.1.1) to show that

$$\begin{aligned}\left\|\sum_{j=1}^{n}\boldsymbol{X}_{j}\right\|_{2} &\lesssim \sqrt{v\log\left(n+d\right)}+L\log\left(n+d\right)+nq_{1}\\ &\lesssim \sigma_{\mathsf{ub}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\sqrt{\log\left(n+d\right)}+\omega_{l}^{\star}\right)\sqrt{\frac{r\log\left(n+d\right)}{p}}+\sqrt{\frac{r\log^{4}\left(n+d\right)}{np^{2}}}\sigma_{\mathsf{ub}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\omega_{l}^{\star}\right)\left[1+\sqrt{\frac{\mu}{dp}}\right]\\ &\lesssim \sigma_{\mathsf{ub}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\omega_{l}^{\star}\right)\sqrt{\frac{r\log^{2}\left(n+d\right)}{p}}\end{aligned}$$

with probability exceeding $1-O((n+d)^{-10})$, where the last line holds as long as $np \gtrsim \log^{2}(n+d)$ and $ndp^{2} \gtrsim \mu\log^{2}(n+d)$. Substitution into (H.24) yields

$$\alpha_{2}\,\mathbb{1}_{\mathcal{E}_{\mathsf{good}}} \lesssim \frac{1}{\sigma_{r}^{\star2}}\sigma_{\mathsf{ub}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\omega_{l}^{\star}\right)\sqrt{\frac{r\log^{2}\left(n+d\right)}{p}}.$$

Putting the above bounds together allows one to conclude that

$$\begin{aligned}\left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{R}-\boldsymbol{U}_{l,\cdot}^{\star}\right\|_{2} &\leq \alpha_{1}+\alpha_{2}+\zeta_{\mathsf{2nd}}\\ &\lesssim \frac{1}{\sigma_{r}^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\omega_{l}^{\star}\right)\sqrt{\frac{r\log^{2}\left(n+d\right)}{np}}+\frac{\sigma_{\mathsf{ub}}}{\sigma_{r}^{\star2}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\omega_{l}^{\star}\right)\sqrt{\frac{r\log^{2}\left(n+d\right)}{p}}+\zeta_{\mathsf{2nd}}\\ &\lesssim \frac{1}{\sigma_{r}^{\star}}\sqrt{\frac{r\log^{2}\left(n+d\right)}{np}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\omega_{l}^{\star}\right)\left(1+\frac{\sigma_{\mathsf{ub}}}{\sigma_{r}^{\star}}\sqrt{n}\right)+\zeta_{\mathsf{2nd}}\\ &\asymp \frac{\theta}{\sqrt{\kappa}\sigma_{r}^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\omega_{l}^{\star}\right)+\zeta_{\mathsf{2nd}},\end{aligned}\tag{H.25}$$

where we define

$$\theta := \sqrt{\frac{\kappa r \log^2(n+d)}{np}}\left(1+\frac{\sigma_{\mathsf{ub}}}{\sigma_r^\star}\sqrt{n}\right)$$

for notational simplicity. By taking the supremum over $l \in [d]$, we further arrive at

$$\begin{aligned}
\|\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^\star\|_{2,\infty} &= \max_{l\in[d]}\left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{R}-\boldsymbol{U}_{l,\cdot}^\star\right\|_2 \\
&\lesssim \frac{1}{\sigma_r^\star}\sqrt{\frac{r\log^2(n+d)}{np}}\left(\|\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\|_{2,\infty}+\omega_{\max}\right)\left[1+\frac{\sigma_{\mathsf{ub}}}{\sigma_r^\star}\sqrt{n}\right]+\zeta_{\mathsf{2nd}} \\
&\lesssim \frac{\sigma_{\mathsf{ub}}}{\sigma_r^\star}\left(1+\frac{\sigma_{\mathsf{ub}}}{\sigma_r^\star}\sqrt{n}\right)\sqrt{r\log^2(n+d)}+\zeta_{\mathsf{2nd}} \\
&\lesssim \frac{1}{\sigma_r^{\star 2}}\left[\sigma_{\mathsf{ub}}\sqrt{d\log(n+d)}+\sigma_{\mathsf{ub}}^2\sqrt{nd}\log(n+d)\right]\sqrt{\frac{r\log(n+d)}{d}}+\zeta_{\mathsf{2nd}} \\
&\asymp \frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}\sqrt{\frac{r\log(n+d)}{d}}+\zeta_{\mathsf{2nd}} \asymp \frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}\sqrt{\frac{r\log(n+d)}{d}},
\end{aligned} \tag{H.26}$$

where the last relation holds provided that $d \gtrsim \kappa^4\mu^2 r\log(n+d)$ and $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2} \lesssim 1/\sqrt{\kappa^3\mu}$.

Bounding $\|\boldsymbol{R}^\top\boldsymbol{\Sigma}^{-2}\boldsymbol{R}-(\boldsymbol{\Sigma}^\star)^{-2}\|$. Conditional on $\boldsymbol{F}$, we learn from Lemma 27 that with probability exceeding $1-O((n+d)^{-10})$

$$\begin{aligned}
\left\|\boldsymbol{Q}^\top\boldsymbol{R}^\top\boldsymbol{\Sigma}^2\boldsymbol{R}\boldsymbol{Q}-\boldsymbol{\Sigma}^{\natural 2}\right\| &\overset{\text{(i)}}{=} \left\|\boldsymbol{R}_{\boldsymbol{U}}^\top\boldsymbol{\Sigma}^2\boldsymbol{R}_{\boldsymbol{U}}-\boldsymbol{\Sigma}^{\natural 2}\right\| \overset{\text{(ii)}}{\lesssim} \kappa^{\natural 2}\sqrt{\frac{\mu^\natural r}{d}}\zeta_{\mathsf{1st}}+\kappa^{\natural 2}\frac{\zeta_{\mathsf{1st}}^2}{\sigma_r^{\natural 2}} \\
&\overset{\text{(iii)}}{\lesssim} \sqrt{\frac{\kappa^3\mu r\log(n+d)}{d}}\zeta_{\mathsf{1st}}+\kappa\frac{\zeta_{\mathsf{1st}}^2}{\sigma_r^{\star 2}} \asymp \sqrt{\frac{d}{\kappa\mu r\log(n+d)}}\zeta_{\mathsf{2nd}}\sigma_r^{\star 2}.
\end{aligned} \tag{H.27}$$

Here, (i) follows from the fact that $\boldsymbol{R}_{\boldsymbol{U}} = \boldsymbol{R}\boldsymbol{Q}$; (ii) follows from Lemma 27; and (iii) utilizes (G.4), (G.6) and (G.5). We can then readily derive

$$\sigma_1^2 \overset{\text{(i)}}{\le} \sigma_1^{\natural 2}+\left\|\boldsymbol{Q}^\top\boldsymbol{R}^\top\boldsymbol{\Sigma}^2\boldsymbol{R}\boldsymbol{Q}-\boldsymbol{\Sigma}^{\natural 2}\right\| \overset{\text{(ii)}}{\lesssim} \sigma_1^{\natural 2}+\sqrt{\frac{\kappa^3\mu r\log(n+d)}{d}}\zeta_{\mathsf{1st}}+\kappa\frac{\zeta_{\mathsf{1st}}^2}{\sigma_r^{\star 2}} \overset{\text{(iii)}}{\le} 2\sigma_1^{\natural 2} \overset{\text{(iv)}}{\le} 4\sigma_1^{\star 2}, \tag{H.28}$$

$$\sigma_r^2 \overset{\text{(v)}}{\ge} \sigma_r^{\natural 2}-\left\|\boldsymbol{Q}^\top\boldsymbol{R}^\top\boldsymbol{\Sigma}^2\boldsymbol{R}\boldsymbol{Q}-\boldsymbol{\Sigma}^{\natural 2}\right\| \overset{\text{(vi)}}{\gtrsim} \sigma_r^{\natural 2}-\sqrt{\frac{\kappa^3\mu r\log(n+d)}{d}}\zeta_{\mathsf{1st}}-\kappa\frac{\zeta_{\mathsf{1st}}^2}{\sigma_r^{\star 2}} \overset{\text{(vii)}}{\ge} \frac{1}{2}\sigma_r^{\natural 2} \overset{\text{(viii)}}{\ge} \frac{1}{4}\sigma_r^{\star 2}. \tag{H.29}$$

Here, (i) and (v) follow from Weyl's inequality; (ii) and (vi) rely on (H.27); (iii) and (vii) make use of (G.4) and hold true provided that $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2} \ll 1/\sqrt{\kappa}$ and $d \gtrsim \kappa^2\mu r\log(n+d)$; (iv) and (viii) utilize (E.54) and are valid as long as $n \gg r+\log(n+d)$. On the event $\mathcal{E}_{\mathsf{good}}$, it is seen from (G.10) that

$$\left\|\boldsymbol{Q}\boldsymbol{\Sigma}^\natural\boldsymbol{Q}^\top-\boldsymbol{\Sigma}^\star\right\| = \left\|\boldsymbol{Q}\boldsymbol{\Sigma}^\natural-\boldsymbol{\Sigma}^\star\boldsymbol{Q}\right\| \lesssim \sqrt{\frac{\kappa(r+\log(n+d))}{n}}\sigma_1^\star,$$

and as a result,

$$\begin{aligned}
\left\|\boldsymbol{Q}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{Q}^\top-\boldsymbol{\Sigma}^{\star 2}\right\| &= \left\|\left(\boldsymbol{Q}\boldsymbol{\Sigma}^\natural\boldsymbol{Q}^\top-\boldsymbol{\Sigma}^\star\right)\boldsymbol{Q}\boldsymbol{\Sigma}^\natural\boldsymbol{Q}^\top+\boldsymbol{\Sigma}^\star\left(\boldsymbol{Q}\boldsymbol{\Sigma}^\natural\boldsymbol{Q}^\top-\boldsymbol{\Sigma}^\star\right)\right\| \\
&\le \left\|\boldsymbol{Q}\boldsymbol{\Sigma}^\natural\boldsymbol{Q}^\top-\boldsymbol{\Sigma}^\star\right\|\left(\sigma_1^\natural+\sigma_1^\star\right) \\
&\lesssim \sqrt{\frac{\kappa(r+\log(n+d))}{n}}\sigma_1^{\star 2},
\end{aligned} \tag{H.30}$$

where the last line relies on (G.4). Taking (H.27) and (H.30) together yields

$$\left\|\boldsymbol{R}^\top\boldsymbol{\Sigma}^2\boldsymbol{R}-\boldsymbol{\Sigma}^{\star 2}\right\| \le \left\|\boldsymbol{R}^\top\boldsymbol{\Sigma}^2\boldsymbol{R}-\boldsymbol{Q}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{Q}^\top\right\|+\left\|\boldsymbol{Q}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{Q}^\top-\boldsymbol{\Sigma}^{\star 2}\right\|$$

$$
\begin{aligned}
&= \left\|\boldsymbol{Q}^\top\boldsymbol{R}^\top\boldsymbol{\Sigma}^2\boldsymbol{R}\boldsymbol{Q}-\boldsymbol{\Sigma}^{\natural 2}\right\| + \left\|\boldsymbol{Q}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{Q}^\top-\boldsymbol{\Sigma}^{\star 2}\right\| \\
&\lesssim \sqrt{\frac{d}{\kappa\mu r\log(n+d)}}\zeta_{\mathsf{2nd}}\sigma_r^{\star 2} + \sqrt{\frac{\kappa\left(r+\log(n+d)\right)}{n}}\sigma_1^{\star 2}.
\end{aligned}
$$

In view of the perturbation bound for matrix square roots (Schmitt, 1992, Lemma 2.2), we obtain

$$
\begin{aligned}
\left\|\boldsymbol{R}^\top\boldsymbol{\Sigma}\boldsymbol{R}-\boldsymbol{\Sigma}^\star\right\| &\le \frac{1}{\sigma_r^\star+\sigma_r}\left\|\boldsymbol{R}^\top\boldsymbol{\Sigma}^2\boldsymbol{R}-\boldsymbol{\Sigma}^{\star 2}\right\| \le \frac{1}{\sigma_r^\star}\left\|\boldsymbol{R}^\top\boldsymbol{\Sigma}^2\boldsymbol{R}-\boldsymbol{\Sigma}^{\star 2}\right\| \\
&\lesssim \sqrt{\frac{d}{\kappa\mu r\log(n+d)}}\zeta_{\mathsf{2nd}}\sigma_r^\star + \sqrt{\frac{\kappa^2\left(r+\log(n+d)\right)}{n}}\sigma_1^\star, \qquad \text{(H.31)}
\end{aligned}
$$

which in turn leads to the upper bound

$$
\begin{aligned}
\left\|\boldsymbol{R}^\top\boldsymbol{\Sigma}^{-2}\boldsymbol{R}-\left(\boldsymbol{\Sigma}^\star\right)^{-2}\right\| &= \left\|\boldsymbol{R}^\top\boldsymbol{\Sigma}^{-2}\boldsymbol{R}\left(\boldsymbol{R}^\top\boldsymbol{\Sigma}^2\boldsymbol{R}-\boldsymbol{\Sigma}^{\star 2}\right)\left(\boldsymbol{\Sigma}^\star\right)^{-2}\right\| \\
&\le \left\|\boldsymbol{R}^\top\boldsymbol{\Sigma}^{-2}\boldsymbol{R}\right\|\left\|\boldsymbol{R}^\top\boldsymbol{\Sigma}^2\boldsymbol{R}-\boldsymbol{\Sigma}^{\star 2}\right\|\left\|\left(\boldsymbol{\Sigma}^\star\right)^{-2}\right\| \\
&\lesssim \frac{1}{\sigma_r^{\star 4}}\left\|\boldsymbol{R}^\top\boldsymbol{\Sigma}^2\boldsymbol{R}-\boldsymbol{\Sigma}^{\star 2}\right\| \\
&\lesssim \sqrt{\frac{d}{\kappa\mu r\log(n+d)}}\frac{\zeta_{\mathsf{2nd}}}{\sigma_r^{\star 2}} + \sqrt{\frac{\kappa^3\left(r+\log(n+d)\right)}{n}}\frac{1}{\sigma_r^{\star 2}}. \qquad \text{(H.32)}
\end{aligned}
$$

Here, the penultimate line comes from (H.28).

Bounding $\|(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{R}-\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star)_{l,\cdot}\|_2$. We start by bounding

$$
\begin{aligned}
\left\|\boldsymbol{U}_{l\cdot}\boldsymbol{\Sigma}\boldsymbol{R}-\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 &\le \left\|\left(\boldsymbol{U}_{l\cdot}\boldsymbol{R}-\boldsymbol{U}_{l,\cdot}^\star\right)\boldsymbol{R}^\top\boldsymbol{\Sigma}\boldsymbol{R}+\boldsymbol{U}_{l,\cdot}^\star\left(\boldsymbol{R}^\top\boldsymbol{\Sigma}\boldsymbol{R}-\boldsymbol{\Sigma}^\star\right)\right\|_2 \\
&\overset{\text{(i)}}{\le} 2\left\|\boldsymbol{U}_{l\cdot}\boldsymbol{R}-\boldsymbol{U}_{l,\cdot}^\star\right\|_2\sigma_1^\star + \left\|\boldsymbol{U}_{l,\cdot}^\star\right\|_2\left\|\boldsymbol{R}^\top\boldsymbol{\Sigma}\boldsymbol{R}-\boldsymbol{\Sigma}^\star\right\| \\
&\overset{\text{(ii)}}{\lesssim} \left(\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\omega_l^\star\right)\left[\sqrt{\frac{\kappa r\log^2(n+d)}{np}}+\frac{1}{\sigma_r^\star}\sigma_{\mathsf{ub}}\sqrt{\frac{\kappa r\log^2(n+d)}{p}}\right] \\
&\quad + \left\|\boldsymbol{U}_{l,\cdot}^\star\right\|_2\left(\sqrt{\frac{d}{\kappa\mu r\log(n+d)}}\zeta_{\mathsf{2nd}}\sigma_r^\star+\sqrt{\frac{\kappa^2\left(r+\log(n+d)\right)}{n}}\sigma_1^\star\right)+\zeta_{\mathsf{2nd}}\sigma_1^\star \\
&\lesssim \sqrt{\frac{\kappa r\log^2(n+d)}{np}}\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\omega_l^\star\right)\left[1+\frac{\sigma_{\mathsf{ub}}}{\sigma_r^\star}\sqrt{n}\right] \\
&\quad + \left\|\boldsymbol{U}_{l,\cdot}^\star\right\|_2\sqrt{\frac{\kappa^2\left(r+\log(n+d)\right)}{n}}\sigma_1^\star+\zeta_{\mathsf{2nd}}\sigma_1^\star \\
&\asymp \theta\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\omega_l^\star\right)+\left\|\boldsymbol{U}_{l,\cdot}^\star\right\|_2\sqrt{\frac{\kappa^2\left(r+\log(n+d)\right)}{n}}\sigma_1^\star+\zeta_{\mathsf{2nd}}\sigma_1^\star. \qquad \text{(H.33)}
\end{aligned}
$$

Here, (i) arises from (H.29), whereas (ii) follows from (H.25) and (H.31). In addition, we observe that

$$
\begin{aligned}
\left\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{R}-\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\right\|_{2,\infty} &\le \left\|\left(\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^\star\right)\boldsymbol{R}^\top\boldsymbol{\Sigma}\boldsymbol{R}+\boldsymbol{U}^\star\left(\boldsymbol{R}^\top\boldsymbol{\Sigma}\boldsymbol{R}-\boldsymbol{\Sigma}^\star\right)\right\|_{2,\infty} \\
&\overset{\text{(i)}}{\le} \left\|\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^\star\right\|_{2,\infty}\sigma_1^\star+\left\|\boldsymbol{U}^\star\right\|_{2,\infty}\left\|\boldsymbol{R}^\top\boldsymbol{\Sigma}\boldsymbol{R}-\boldsymbol{\Sigma}^\star\right\| \\
&\overset{\text{(ii)}}{\lesssim} \frac{\zeta_{\mathsf{1st}}}{\sigma_r^\star}\sqrt{\frac{\kappa r\log(n+d)}{d}}+\sqrt{\frac{\mu r}{d}}\left(\sqrt{\frac{d}{\kappa\mu r\log(n+d)}}\zeta_{\mathsf{2nd}}\sigma_r^\star+\sqrt{\frac{\kappa^2\left(r+\log(n+d)\right)}{n}}\sigma_1^\star\right) \\
&\asymp \frac{\zeta_{\mathsf{1st}}}{\sigma_r^\star}\sqrt{\frac{\kappa r\log(n+d)}{d}}+\sqrt{\frac{1}{\kappa\log(n+d)}}\zeta_{\mathsf{2nd}}\sigma_r^\star+\sqrt{\frac{\kappa^2\mu r^2\log(n+d)}{nd}}\sigma_1^\star.
\end{aligned}
$$

$$\overset{\text{(iii)}}{\lesssim} \frac{\zeta_{\mathsf{1st}}}{\sigma_r^\star}\sqrt{\frac{\kappa r \log(n+d)}{d}} + \sqrt{\frac{\kappa^2 \mu r^2 \log(n+d)}{nd}}\sigma_1^\star. \tag{H.34}$$

Here, (i) results from (H.29); (ii) follows from (H.26) and (H.31); (iii) holds true provided that $d \gtrsim \kappa^2\mu^2 r$ and $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2} \lesssim 1/\sqrt{\kappa\mu}$.

Bounding $\|\boldsymbol{U}\boldsymbol{\Sigma}^{-2}\boldsymbol{R} - \boldsymbol{U}^\star(\boldsymbol{\Sigma}^\star)^{-2}\|$. It is first observed that, on the event $\mathcal{E}_{\mathsf{good}}$,

$$\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\| \overset{\text{(i)}}{=} \left\|\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^\natural\right\| \overset{\text{(ii)}}{\lesssim} \frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\natural 2}} \overset{\text{(iii)}}{\lesssim} \frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}. \tag{H.35}$$

Here, (i) comes from the facts that $\boldsymbol{R}_{\boldsymbol{U}} = \boldsymbol{R}\boldsymbol{Q}$ and $\boldsymbol{U}^\natural = \boldsymbol{U}^\star\boldsymbol{Q}$ for some orthonormal matrix $\boldsymbol{Q}$ (see (D.4)); (ii) follows from Lemma 24; and (iii) is a consequence of (G.4). This immediately gives

$$\begin{aligned}
\left\|\boldsymbol{U}\boldsymbol{\Sigma}^{-2}\boldsymbol{R} - \boldsymbol{U}^\star(\boldsymbol{\Sigma}^\star)^{-2}\right\| &= \left\|(\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star)\boldsymbol{R}^\top\boldsymbol{\Sigma}^{-2}\boldsymbol{R} + \boldsymbol{U}^\star\left[\boldsymbol{R}^\top\boldsymbol{\Sigma}^{-2}\boldsymbol{R} - (\boldsymbol{\Sigma}^\star)^{-2}\right]\right\| \\
&\le \frac{1}{\sigma_r^2}\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\| + \|\boldsymbol{U}^\star\|\left\|\boldsymbol{R}^\top\boldsymbol{\Sigma}^{-2}\boldsymbol{R} - (\boldsymbol{\Sigma}^\star)^{-2}\right\| \\
&\overset{\text{(i)}}{\lesssim} \frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 4}} + \sqrt{\frac{d}{\kappa\mu r\log(n+d)}}\frac{\zeta_{\mathsf{2nd}}}{\sigma_r^{\star 2}} + \sqrt{\frac{\kappa^3(r+\log(n+d))}{n}}\frac{1}{\sigma_r^{\star 2}} \\
&\asymp \frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 4}} + \sqrt{\frac{\kappa^3\mu r\log(n+d)}{d}}\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 4}} + \kappa\frac{\zeta_{\mathsf{1st}}^2}{\sigma_r^{\star 6}} + \sqrt{\frac{\kappa^3(r+\log(n+d))}{n}}\frac{1}{\sigma_r^{\star 2}} \\
&\overset{\text{(ii)}}{\asymp} \frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 4}} + \sqrt{\frac{\kappa^3(r+\log(n+d))}{n}}\frac{1}{\sigma_r^{\star 2}},
\end{aligned}$$

where (i) follows from (H.32) and (H.35), and (ii) holds true as long as $d \gtrsim \kappa^3\mu r\log(n+d)$ and $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2} \lesssim 1/\sqrt{\kappa}$.

H.2.2 Proof of Lemma 41

Before proceeding, let us make some useful observations: for all $l \in [d]$,

$$\begin{aligned}
\|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\|_2 = \|\boldsymbol{U}_{l,\cdot}\boldsymbol{\Sigma}\boldsymbol{R}\|_2 &\le \left\|\boldsymbol{U}_{l\cdot}\boldsymbol{\Sigma}\boldsymbol{R} - \boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \\
&\overset{\text{(i)}}{\lesssim} \theta\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \omega_l^\star\right) + \left\|\boldsymbol{U}_{l,\cdot}^\star\right\|_2\sqrt{\frac{\kappa^2(r+\log(n+d))}{n}}\sigma_1^\star + \zeta_{\mathsf{2nd}}\sigma_1^\star + \left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \\
&\lesssim \theta\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \omega_l^\star\right) + \frac{1}{\sigma_r^\star}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\sqrt{\frac{\kappa^2(r+\log(n+d))}{n}}\sigma_1^\star + \zeta_{\mathsf{2nd}}\sigma_1^\star + \left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \\
&\overset{\text{(ii)}}{\asymp} \left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \theta\omega_l^\star + \zeta_{\mathsf{2nd}}\sigma_1^\star,
\end{aligned} \tag{H.36}$$

where θ is defined in (G.22). Here, (i) relies on (G.23b), while (ii) holds true as long as $n \gtrsim \kappa^3 r\log(n+d)$ and $\theta \ll 1$. Additionally, note that

$$\begin{aligned}
\theta &\asymp \sqrt{\frac{\kappa r\log^2(n+d)}{np}}\left(1 + \sqrt{\frac{\kappa\mu r\log(n+d)}{dp}} + \frac{\omega_{\max}}{\sigma_r^\star\sqrt{p}}\right) \\
&\asymp \sqrt{\frac{\kappa r\log^2(n+d)}{np}} + \sqrt{\frac{\kappa^2\mu r^2\log^3(n+d)}{ndp^2}} + \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{\kappa r\log^2(n+d)}{np^2}}.
\end{aligned} \tag{H.37}$$

In view of the following relation (which makes use of the AM-GM inequality)

$$\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}} \ge \frac{\kappa\mu r\log^2(n+d)}{\sqrt{ndp}} + \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}\log(n+d) \ge \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{\kappa\mu r\log^3(n+d)}{np^2}}, \tag{H.38}$$

we can see that

$$\theta \lesssim \frac{1}{\sqrt{\mu}} \cdot \frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}. \tag{H.39}$$

This means that $\theta \ll 1$ can be guaranteed by the condition $\zeta_{\mathsf{1st}} \ll \sigma_r^{\star 2}$. We are now positioned to embark on the proof.

Step 1: bounding $|S_{i,j} - S_{i,j}^{\star}|$. We first develop an entrywise upper bound on $\boldsymbol{S} - \boldsymbol{S}^{\star}$ as follows

$$\begin{aligned}
\|\boldsymbol{S} - \boldsymbol{S}^{\star}\|_{\infty} &= \left\|(\boldsymbol{U\Sigma R})(\boldsymbol{U\Sigma R})^{\top} - \boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}(\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star})^{\top}\right\|_{\infty} \\
&= \left\|(\boldsymbol{U\Sigma R} - \boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star})(\boldsymbol{U\Sigma R})^{\top} + \boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}(\boldsymbol{U\Sigma R} - \boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star})^{\top}\right\|_{\infty} \\
&\leq \|\boldsymbol{U\Sigma R} - \boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\|_{2,\infty}\|\boldsymbol{U\Sigma}\|_{2,\infty} + \|\boldsymbol{U\Sigma R} - \boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\|_{2,\infty}\|\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\|_{2,\infty} \\
&\lesssim \|\boldsymbol{U\Sigma R} - \boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\|_{2,\infty}\|\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\|_{2,\infty} + \|\boldsymbol{U\Sigma R} - \boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\|_{2,\infty}^{2} \\
&\lesssim \left(\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star}}\sqrt{\frac{\kappa r\log(n+d)}{d}} + \sqrt{\frac{\kappa^2\mu r^2\log(n+d)}{nd}}\sigma_1^{\star}\right)\sqrt{\frac{\mu r}{d}}\sigma_1^{\star} \\
&\quad + \left(\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star}}\sqrt{\frac{\kappa r\log(n+d)}{d}} + \sqrt{\frac{\kappa^2\mu r^2\log(n+d)}{nd}}\sigma_1^{\star}\right)^{2} \\
&\asymp \left(\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star}}\sqrt{\frac{\kappa r\log(n+d)}{d}} + \sqrt{\frac{\kappa^2\mu r^2\log(n+d)}{nd}}\sigma_1^{\star}\right)\sqrt{\frac{\mu r}{d}}\sigma_1^{\star}.
\end{aligned} \tag{H.40}$$

Here the penultimate relation follows from (H.34), and the last relation holds provided that

$$\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star}}\sqrt{\frac{\kappa r\log(n+d)}{d}} + \sqrt{\frac{\kappa^2\mu r^2\log(n+d)}{nd}}\sigma_1^{\star} \lesssim \sqrt{\frac{\mu r}{d}}\sigma_1^{\star},$$

which can be guaranteed by $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2} \lesssim 1/\sqrt{\log(n+d)}$ and $n \gtrsim \kappa^2 r\log(n+d)$. Focusing on the (i,j)-th entry, we can obtain

$$\begin{aligned}
\left|S_{i,j} - S_{i,j}^{\star}\right| &= \left|(\boldsymbol{U}_{i,\cdot}\boldsymbol{\Sigma R})(\boldsymbol{U}_{j,\cdot}\boldsymbol{\Sigma R})^{\top} - \boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}(\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star})^{\top}\right| \\
&\leq \left\|\boldsymbol{U}_{i\cdot}\boldsymbol{\Sigma R} - \boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}\boldsymbol{\Sigma}\right\|_2 + \left\|\boldsymbol{U}_{j\cdot}\boldsymbol{\Sigma R} - \boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \\
&\overset{\text{(i)}}{\lesssim} \left\|\boldsymbol{U}_{i\cdot}\boldsymbol{\Sigma R} - \boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{i\cdot}\boldsymbol{\Sigma R} - \boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left(\theta\omega_j^{\star} + \zeta_{\mathsf{2nd}}\sigma_1^{\star}\right) \\
&\quad + \left\|\boldsymbol{U}_{j\cdot}\boldsymbol{\Sigma R} - \boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \\
&\overset{\text{(ii)}}{\lesssim} \left[\theta\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_i^{\star}\right) + \left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2\sqrt{\frac{\kappa^2(r+\log(n+d))}{n}}\sigma_1^{\star} + \zeta_{\mathsf{2nd}}\sigma_1^{\star}\right]\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \\
&\quad + \left[\theta\left(\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_j^{\star}\right) + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\sqrt{\frac{\kappa^2(r+\log(n+d))}{n}}\sigma_1^{\star} + \zeta_{\mathsf{2nd}}\sigma_1^{\star}\right]\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \\
&\quad + \left[\theta\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_i^{\star}\right) + \left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2\sqrt{\frac{\kappa^2(r+\log(n+d))}{n}}\sigma_1^{\star} + \zeta_{\mathsf{2nd}}\sigma_1^{\star}\right]\left(\theta\omega_j^{\star} + \zeta_{\mathsf{2nd}}\sigma_1^{\star}\right) \\
&\overset{\text{(iii)}}{\lesssim} \left(\theta + \sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \\
&\quad + \left(\theta\omega_{\max} + \zeta_{\mathsf{2nd}}\sigma_1^{\star}\right)\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right) + \theta^2\omega_{\max}^2 + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}.
\end{aligned} \tag{H.41}$$

Here, (i) follows from (H.36), (ii) arises from (G.23b) and the fact that $\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \leq \frac{1}{\sigma_r^{\star}}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2$, while (iii) holds true as long as $\theta \ll 1$ (which is guaranteed when $\zeta_{\mathsf{1st}} \ll \sigma_r^{\star 2}$) and $n \gtrsim \kappa^3 r\log(n+d)$.

Step 2: bounding $|\omega_i^2 - \omega_i^{\star 2}|$. We can follow the same analysis as in Step 2 in Appendix E.2.2 to show that, for all $i \in [d]$,

$$\left|\omega_i^2 - \omega_i^{\star 2}\right| \lesssim \sqrt{\frac{\log^2(n+d)}{np}}\left(\omega_i^{\star 2} + S_{i,i}^\star\right) + \left|S_{i,i} - S_{i,i}^\star\right|.$$

In view of (H.40) and the fact that $\|\boldsymbol{S}^\star\|_\infty = \|\boldsymbol{U}^\star\|_{2,\infty}^2 \|\boldsymbol{\Sigma}^{\star 2}\| \le \frac{\mu r}{d}\sigma_1^{\star 2}$, the above inequality implies that, for all $i \in [d]$,

$$\begin{aligned}
\left|\omega_i^2 - \omega_i^{\star 2}\right| &\lesssim \sqrt{\frac{\log^2(n+d)}{np}}\left(\omega_i^{\star 2} + \frac{\mu r}{d}\sigma_1^{\star 2}\right) + \left(\frac{\zeta_{\mathsf{1st}}}{\sigma_r^\star}\sqrt{\frac{\kappa r\log(n+d)}{d}} + \sqrt{\frac{\kappa^2\mu r^2\log(n+d)}{nd}}\sigma_1^\star\right)\sqrt{\frac{\mu r}{d}}\sigma_1^\star \\
&\asymp \sqrt{\frac{\log^2(n+d)}{np}}\omega_i^{\star 2} + \sqrt{\frac{\log^2(n+d)}{np}}\frac{\mu r}{d}\sigma_1^{\star 2} + \zeta_{\mathsf{1st}}\frac{\sqrt{\kappa^2\mu r^2\log(n+d)}}{d} + \sqrt{\frac{\kappa^2\mu^2 r^3\log(n+d)}{nd^2}}\sigma_1^{\star 2} \\
&\asymp \sqrt{\frac{\log^2(n+d)}{np}}\omega_i^{\star 2} + \zeta_{\mathsf{1st}}\frac{\sqrt{\kappa^2\mu r^2\log(n+d)}}{d} + \sqrt{\frac{\kappa^2\mu^2 r^3\log(n+d)}{nd^2}}\sigma_1^{\star 2},
\end{aligned}$$

where the last line makes use of the definition of ζ_{1st} in (G.14b). Additionally, in view of (H.41), we can derive for $i = l$ that

$$\begin{aligned}
\left|\omega_l^2 - \omega_l^{\star 2}\right| &\lesssim \sqrt{\frac{\log^2(n+d)}{np}}\left(\omega_l^{\star 2} + S_{l,l}^\star\right) + \left(\theta + \sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2 \\
&\quad + \left(\theta\omega_{\max} + \zeta_{\mathsf{2nd}}\sigma_1^\star\right)\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \theta^2\omega_{\max}^2 + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2} \\
&\asymp \left(\sqrt{\frac{\log^2(n+d)}{np}} + \theta^2\right)\omega_{\max}^2 + \left(\theta + \sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2 \\
&\quad + \left(\theta\omega_{\max} + \zeta_{\mathsf{2nd}}\sigma_1^\star\right)\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}.
\end{aligned}$$

Here, the last relation invokes the following condition

$$\sqrt{\frac{\log^2(n+d)}{np}}S_{l,l}^\star \lesssim \theta\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2,$$

which is a direct consequence of (H.37).

H.2.3 Proof of Lemma 42

Before proceeding, let us recall a few facts as follows. To begin with, from the definition (G.17) of $\boldsymbol{\Sigma}_{U,l}^\star$, we can decompose

$$\begin{aligned}
\boldsymbol{R}\boldsymbol{\Sigma}_{U,l}^\star\boldsymbol{R}^\top &= \underbrace{\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\boldsymbol{R}\left(\boldsymbol{\Sigma}^\star\right)^{-2}\boldsymbol{R}^\top}_{=:\boldsymbol{A}_1} + \underbrace{\frac{\omega_l^{\star 2}}{np}\boldsymbol{R}\left(\boldsymbol{\Sigma}^\star\right)^{-2}\boldsymbol{R}^\top}_{=:\boldsymbol{A}_2} + \underbrace{\frac{2(1-p)}{np}\boldsymbol{R}\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{R}^\top}_{=:\boldsymbol{A}_3} \\
&\quad + \underbrace{\boldsymbol{R}\left(\boldsymbol{\Sigma}^\star\right)^{-2}\boldsymbol{U}^{\star\top}\mathsf{diag}\left\{\left[d_{l,i}^\star\right]_{i=1}^d\right\}\boldsymbol{U}^\star\left(\boldsymbol{\Sigma}^\star\right)^{-2}\boldsymbol{R}^\top}_{=:\boldsymbol{A}_4},
\end{aligned}$$

where we recall that

$$d_{l,i}^\star := \frac{1}{np^2}\left[\omega_l^{\star 2} + (1-p)\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\right]\left[\omega_i^{\star 2} + (1-p)\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\right] + \frac{2(1-p)^2}{np^2}S_{i,l}^{\star 2}.$$

Regarding $\mathbf{\Sigma}_{U,l}$, we remind the reader of its definition in (G.19) as follows

$$\mathbf{\Sigma}_{U,l}=\underbrace{\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}\mathbf{\Sigma}\right\|_2^2\mathbf{\Sigma}^{-2}}_{=:\boldsymbol{B}_1}+\underbrace{\frac{\omega_l^2}{np}\mathbf{\Sigma}^{-2}}_{=:\boldsymbol{B}_2}+\underbrace{\frac{2(1-p)}{np}\boldsymbol{U}_{l,\cdot}^\top\boldsymbol{U}_{l,\cdot}}_{=:\boldsymbol{B}_3}+\underbrace{\mathbf{\Sigma}^{-2}\boldsymbol{U}^\top\mathsf{diag}\left\{[d_{l,i}]_{1\le i\le d}\right\}\boldsymbol{U}\mathbf{\Sigma}^{-2}}_{=:\boldsymbol{B}_4},$$

where we define

$$d_{l,i}\coloneqq\frac{1}{np^2}\left[\omega_l^2+(1-p)\left\|\boldsymbol{U}_{l,\cdot}\mathbf{\Sigma}\right\|_2^2\right]\left[\omega_i^2+(1-p)\left\|\boldsymbol{U}_{i,\cdot}\mathbf{\Sigma}\right\|_2^2\right]+\frac{2(1-p)^2}{np^2}S_{l,i}^2.$$

In addition, we also recall from Lemma 38 that

$$\lambda_{\min}\left(\mathbf{\Sigma}_{U,l}^\star\right)\gtrsim\frac{1-p}{np\sigma_1^{\star2}}\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2^2+\frac{\omega_l^{\star2}}{np\sigma_1^{\star2}}+\frac{(1-p)^2}{ndp^2\kappa\sigma_1^{\star2}}\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2^2+\frac{1-p}{4ndp^2\kappa\sigma_1^{\star2}}\omega_l^{\star2}+\frac{\omega_l^{\star2}\omega_{\min}^2}{np^2\sigma_1^{\star4}}.$$

We are now ready to present the proof, with the focus of bounding $\|\boldsymbol{A}_i-\boldsymbol{B}_i\|$ for each $1\le i\le 4$ as well as $\left|d_{l,i}^\star-d_{l,i}\right|$.

Step 1: controlling $\|\boldsymbol{A}_1-\boldsymbol{B}_1\|$. It follows from the triangle inequality and Lemma 40 that

$$\begin{aligned}\left|\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2-\left\|\boldsymbol{U}_{l,\cdot}\mathbf{\Sigma}\right\|_2\right|&=\left|\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2-\left\|\boldsymbol{U}_{l,\cdot}\mathbf{\Sigma}\boldsymbol{R}\right\|_2\right|\le\left\|\boldsymbol{U}_{l\cdot}\mathbf{\Sigma}\boldsymbol{R}-\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2\\&\lesssim\theta\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2+\omega_l^\star\right)+\left\|\boldsymbol{U}_{l,\cdot}^\star\right\|_2\sqrt{\frac{\kappa^2(r+\log(n+d))}{n}}\sigma_1^\star+\zeta_{\mathsf{2nd}}\sigma_1^\star.\end{aligned}$$

Here, the last line follows from (G.23b). This immediately gives

$$\begin{aligned}&\left|\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2^2-\left\|\boldsymbol{U}_{l,\cdot}\mathbf{\Sigma}\right\|_2^2\right|\lesssim\left|\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2-\left\|\boldsymbol{U}_{l,\cdot}\mathbf{\Sigma}\right\|_2\right|\left|\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2+\left\|\boldsymbol{U}_{l,\cdot}\mathbf{\Sigma}\right\|_2\right|\\&\quad\asymp\left|\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2-\left\|\boldsymbol{U}_{l,\cdot}\mathbf{\Sigma}\right\|_2\right|\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2+\left|\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2-\left\|\boldsymbol{U}_{l,\cdot}\mathbf{\Sigma}\right\|_2\right|^2\\&\quad\lesssim\theta\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2+\omega_l^\star\right)\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2+\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{l,\cdot}^\star\right\|_2\sqrt{\frac{\kappa^2(r+\log(n+d))}{n}}\sigma_1^\star\\&\quad\quad+\zeta_{\mathsf{2nd}}\sigma_1^\star\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2+\theta^2\left(\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2^2+\omega_l^{\star2}\right)+\left\|\boldsymbol{U}_{l,\cdot}^\star\right\|_2^2\frac{\kappa^2(r+\log(n+d))}{n}\sigma_1^{\star2}+\zeta_{\mathsf{2nd}}^2\sigma_1^{\star2}\\&\quad\lesssim\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2^2\left(\theta+\sqrt{\frac{\kappa^3(r+\log(n+d))}{n}}\right)+\theta\omega_l^\star\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2\\&\quad\quad+\zeta_{\mathsf{2nd}}\sigma_1^\star\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2+\theta^2\omega_l^{\star2}+\zeta_{\mathsf{2nd}}^2\sigma_1^{\star2}.\end{aligned}\tag{H.42}$$

Here the last relation holds provided that $\theta\ll1$ and $n\gtrsim\kappa^3r\log(n+d)$. In addition, from (G.23c) in Lemma 40, we know that

$$\left\|\boldsymbol{R}(\mathbf{\Sigma}^\star)^{-2}\boldsymbol{R}^\top-\mathbf{\Sigma}^{-2}\right\|\lesssim\sqrt{\frac{d}{\kappa\mu r\log(n+d)}}\frac{\zeta_{\mathsf{2nd}}}{\sigma_r^{\star2}}+\sqrt{\frac{\kappa^3(r+\log(n+d))}{n}}\frac{1}{\sigma_r^{\star2}}.\tag{H.43}$$

These two inequalities help us derive

$$\begin{aligned}\|\boldsymbol{A}_1-\boldsymbol{B}_1\|&\le\frac{1-p}{np}\left|\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2^2-\left\|\boldsymbol{U}_{l,\cdot}\mathbf{\Sigma}\right\|_2^2\right|\left\|\boldsymbol{R}\mathbf{\Sigma}^{-2}\boldsymbol{R}^\top\right\|+\frac{1-p}{np}\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2^2\left\|\boldsymbol{R}(\mathbf{\Sigma}^\star)^{-2}\boldsymbol{R}^\top-\mathbf{\Sigma}^{-2}\right\|\\&\overset{\text{(i)}}{\lesssim}\underbrace{\frac{1}{np\sigma_r^{\star2}}\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2^2\left(\theta+\sqrt{\frac{\kappa^3(r+\log(n+d))}{n}}\right)}_{=:\alpha_{1,1}}+\underbrace{\frac{1}{np\sigma_r^{\star2}}\theta\omega_l^\star\left\|\boldsymbol{U}_{l,\cdot}^\star\mathbf{\Sigma}^\star\right\|_2}_{=:\alpha_{1,2}}\end{aligned}$$

$$+\frac{1}{np\sigma_r^{\star 2}}\zeta_{\mathsf{2nd}}\sigma_1^\star\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\underbrace{\frac{1}{np\sigma_r^{\star 2}}\theta^2\omega_l^{\star 2}}_{=:\alpha_{1,3}}+\frac{1}{np\sigma_r^{\star 2}}\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}+\underbrace{\frac{1}{np}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\sqrt{\frac{d}{\kappa\mu r\log(n+d)}}\frac{\zeta_{\mathsf{2nd}}}{\sigma_r^{\star 2}}}_{=:\alpha_{1,4}}$$

$$\overset{\text{(ii)}}{\lesssim}\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right)+\frac{\sqrt{\kappa}}{np\sigma_r^\star}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\frac{\kappa}{np}\zeta_{\mathsf{2nd}}^2.$$

Here, (i) results from (H.28), (H.42) and (H.43); (ii) holds true due to the following inequalities:

$$\alpha_{1,1}+\alpha_{1,4}\lesssim\delta\frac{1-p}{np\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\lesssim\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right),$$

$$\alpha_{1,2}\lesssim\delta\frac{1-p}{np\sigma_1^{\star 2}}\omega_l^\star\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\lesssim\delta\frac{1-p}{np\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2+\delta\frac{\omega_l^{\star 2}}{np\sigma_1^{\star 2}}\lesssim\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right),$$

$$\alpha_{1,3}\lesssim\delta\frac{\omega_l^{\star 2}}{np\sigma_1^{\star 2}}\lesssim\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right),$$

provided that $\theta\lesssim\delta/\kappa$, $n\gtrsim\delta^{-2}\kappa^5 r\log(n+d)$ and $\zeta_{\mathsf{2nd}}\sqrt{d}\lesssim\delta\sqrt{\mu r\log(n+d)}$.

Step 2: controlling $\|\boldsymbol{A}_2-\boldsymbol{B}_2\|$. Recall from (G.27) that

$$\left|\omega_l^2-\omega_l^{\star 2}\right|\lesssim\left(\sqrt{\frac{\log^2(n+d)}{np}}+\theta^2\right)\omega_l^{\star 2}+\left(\theta+\sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2$$
$$+\theta\omega_{\max}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\zeta_{\mathsf{2nd}}\sigma_1^\star\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}. \tag{H.44}$$

Armed with this inequality, we can derive

$$\left\|\boldsymbol{A}_2-\boldsymbol{B}_2\right\|=\left\|\frac{\omega_l^{\star 2}}{np}\left[\boldsymbol{R}\left(\boldsymbol{\Sigma}^\star\right)^{-2}\boldsymbol{R}^\top-\boldsymbol{\Sigma}^{-2}\right]+\frac{\omega_l^{\star 2}-\omega_l^2}{np}\boldsymbol{\Sigma}^{-2}\right\|$$
$$\overset{\text{(i)}}{\lesssim}\frac{\omega_l^{\star 2}}{np}\left\|\boldsymbol{R}\left(\boldsymbol{\Sigma}^\star\right)^{-2}\boldsymbol{R}^\top-\boldsymbol{\Sigma}^{-2}\right\|+\frac{\left|\omega_l^{\star 2}-\omega_l^2\right|}{np\sigma_r^{\star 2}}$$
$$\overset{\text{(ii)}}{\lesssim}\underbrace{\frac{\omega_l^{\star 2}}{np}\sqrt{\frac{d}{\kappa\mu r\log(n+d)}}\frac{\zeta_{\mathsf{2nd}}}{\sigma_r^{\star 2}}}_{=:\alpha_{2,1}}+\underbrace{\frac{\omega_l^{\star 2}}{np\sigma_r^{\star 2}}\sqrt{\frac{\kappa^3(r+\log(n+d))}{n}}}_{=:\alpha_{2,2}}$$
$$+\underbrace{\frac{1}{np\sigma_r^{\star 2}}\left(\sqrt{\frac{\log^2(n+d)}{np}}+\theta^2\right)\omega_l^{\star 2}}_{=:\alpha_{2,3}}+\underbrace{\frac{1}{np\sigma_r^{\star 2}}\left(\theta+\sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2}_{=:\alpha_{2,4}}$$
$$+\underbrace{\frac{1}{np\sigma_r^{\star 2}}\theta\omega_{\max}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2}_{=:\alpha_{2,5}}+\frac{1}{np\sigma_r^{\star 2}}\zeta_{\mathsf{2nd}}\sigma_1^\star\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\frac{1}{np\sigma_r^{\star 2}}\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}$$
$$\overset{\text{(iii)}}{\lesssim}\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right)+\frac{\sqrt{\kappa}}{np\sigma_r^\star}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\frac{\kappa}{np}\zeta_{\mathsf{2nd}}^2.$$

Here, (i) arises from (H.28); (ii) utilizes (H.43) and (H.44); (iii) follows from the inequalities below:

$$\alpha_{2,1}+\alpha_{2,2}+\alpha_{2,3}\lesssim\delta\frac{\omega_l^{\star 2}}{np\sigma_1^{\star 2}}\lesssim\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right),$$

$$\alpha_{2,4}\lesssim\delta\frac{1-p}{np\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\lesssim\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right),$$

$$\alpha_{2,5}\lesssim\delta\frac{1-p}{np\sigma_1^{\star 2}}\omega_l^\star\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\lesssim\delta\frac{1-p}{np\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2+\delta\frac{\omega_l^{\star 2}}{np\sigma_1^{\star 2}}\lesssim\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right),$$

provided that $\zeta_{\mathsf{2nd}}\sqrt{d}\lesssim\delta\sqrt{\mu r\log(n+d)/\kappa}$, $n\gtrsim\delta^{-2}\kappa^5 r\log(n+d)$, $np\gtrsim\delta^{-2}\log^2(n+d)$ and $\theta\lesssim\delta/\kappa$.

Step 3: controlling $\|\boldsymbol{A}_3 - \boldsymbol{B}_3\|$. We have learned from (G.23a) in Lemma 40 we know that

$$\left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{R} - \boldsymbol{U}_{l,\cdot}^{\star}\right\|_2 \lesssim \frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right) + \zeta_{\mathsf{2nd}}.$$

This immediately gives

$$\begin{aligned}
\left\|\boldsymbol{U}_{l,\cdot}\right\|_2 = \left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{R}\right\|_2 &\leq \left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{l,\cdot}\boldsymbol{R} - \boldsymbol{U}_{l,\cdot}^{\star}\right\|_2 \lesssim \left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2 + \frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right) + \zeta_{\mathsf{2nd}} \\
&\leq \left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2 + \frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2\sigma_1^{\star} + \omega_l^{\star}\right) + \zeta_{\mathsf{2nd}} \\
&\lesssim \left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2 + \frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\omega_l^{\star} + \zeta_{\mathsf{2nd}},
\end{aligned} \tag{H.45}$$

where the last line holds provided that $\theta \ll 1$. As a consequence, we arrive at

$$\begin{aligned}
\left\|\boldsymbol{A}_3 - \boldsymbol{B}_3\right\| &= \frac{2(1-p)}{np}\left\|\boldsymbol{R}\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{R}^{\top} - \boldsymbol{U}_{l,\cdot}^{\top}\boldsymbol{U}_{l,\cdot}\right\| = \frac{2(1-p)}{np}\left\|\boldsymbol{U}_{l,\cdot}^{\star\top}\boldsymbol{U}_{l,\cdot}^{\star} - \boldsymbol{R}^{\top}\boldsymbol{U}_{l,\cdot}^{\top}\boldsymbol{U}_{l,\cdot}\boldsymbol{R}\right\| \\
&\leq \frac{2(1-p)}{np}\left\|\left(\boldsymbol{U}^{\star} - \boldsymbol{U}\boldsymbol{R}\right)_{l,\cdot}^{\top}\boldsymbol{U}_{l,\cdot}^{\star}\right\| + \frac{2(1-p)}{np}\left\|\left(\boldsymbol{U}_{l}\boldsymbol{R}\right)^{\top}\left(\boldsymbol{U}_{l,\cdot}^{\star} - \boldsymbol{U}_{l,\cdot}\boldsymbol{R}\right)\right\| \\
&\lesssim \frac{1}{np}\left\|\boldsymbol{U}_{l,\cdot}^{\star} - \boldsymbol{U}_{l,\cdot}\boldsymbol{R}\right\|_2\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{l,\cdot}\right\|_2\right) \\
&\overset{\text{(i)}}{\lesssim} \frac{1}{np}\left[\frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right) + \zeta_{\mathsf{2nd}}\right]\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2 + \frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\omega_l^{\star} + \zeta_{\mathsf{2nd}}\right) \\
&\overset{\text{(ii)}}{\lesssim} \underbrace{\frac{1}{np}\frac{\theta}{\sqrt{\kappa}\sigma_r^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2}_{=:\alpha_{3,1}} + \underbrace{\frac{1}{np}\frac{\theta}{\sqrt{\kappa}\sigma_r^{\star 2}}\omega_l^{\star}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2}_{=:\alpha_{3,2}} + \frac{1}{np}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2 + \underbrace{\frac{1}{np}\frac{\theta^2}{\kappa\sigma_r^{\star 2}}\omega_l^{\star 2}}_{=:\alpha_{3,3}} \\
&\quad + \frac{1}{np}\zeta_{\mathsf{2nd}}\frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\omega_l^{\star} + \frac{1}{np}\zeta_{\mathsf{2nd}}^2 \\
&\overset{\text{(iii)}}{\lesssim} \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) + \frac{1}{np}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\right\|_2 + \frac{1}{np}\zeta_{\mathsf{2nd}}\frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\omega_l^{\star} + \frac{1}{np}\zeta_{\mathsf{2nd}}^2.
\end{aligned}$$

Here, (i) follows from (G.23a) and (H.45); (ii) holds provided that $\theta \ll 1$; and (iii) is valid due to the following facts

$$\begin{aligned}
\alpha_{3,1} &\lesssim \delta\frac{1-p}{np\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right), \\
\alpha_{3,2} &\lesssim \delta\frac{1-p}{np\sigma_1^{\star 2}}\omega_l^{\star}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \lesssim \delta\frac{1-p}{np\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \delta\frac{\omega_l^{\star 2}}{np\sigma_1^{\star 2}} \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right), \\
\alpha_{3,3} &\lesssim \delta\frac{\omega_l^{\star 2}}{np\sigma_1^{\star 2}} \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),
\end{aligned}$$

provided that $\theta \lesssim \delta/\sqrt{\kappa}$.

Step 4: bounding $\|\boldsymbol{A}_4 - \boldsymbol{B}_4\|$. To begin with, we recall from (G.23d) in Lemma 40 that

$$\left\|\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{R}^{\top} - \boldsymbol{U}\boldsymbol{\Sigma}^{-2}\right\| \lesssim \frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 4}} + \sqrt{\frac{\kappa^3\left(r + \log\left(n+d\right)\right)}{n}}\frac{1}{\sigma_r^{\star 2}}. \tag{H.46}$$

This allows one to upper bound

$$\left\|\boldsymbol{A}_4 - \boldsymbol{B}_4\right\| \leq \left\|\left(\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{R}^{\top} - \boldsymbol{U}\boldsymbol{\Sigma}^{-2}\right)^{\top}\mathsf{diag}\left\{\left[d_{l,i}^{\star}\right]_{i=1}^{d}\right\}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{R}^{\top}\right\|$$

$$
\begin{aligned}
&+\left\|\boldsymbol{U}\boldsymbol{\Sigma}^{-2}\mathsf{diag}\left\{\left[d_{l,i}^{\star}-d_{l,i}\right]_{i=1}^{d}\right\}\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{R}^{\top}\right\| \\
&+\left\|\boldsymbol{U}\boldsymbol{\Sigma}^{-2}\mathsf{diag}\left\{\left[d_{l,i}\right]_{i=1}^{d}\right\}\left(\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{R}^{\top}-\boldsymbol{U}\boldsymbol{\Sigma}^{-2}\right)\right\| \\
&\overset{\text{(i)}}{\lesssim}\frac{1}{\sigma_r^{\star 2}}\left\|\boldsymbol{U}^{\star}\left(\boldsymbol{\Sigma}^{\star}\right)^{-2}\boldsymbol{R}^{\top}-\boldsymbol{U}\boldsymbol{\Sigma}^{-2}\right\|\max_{1\le i\le d}\left(d_{l,i}^{\star}+d_{l,i}\right)+\frac{1}{\sigma_r^{\star 4}}\max_{1\le i\le d}\left|d_{l,i}^{\star}-d_{l,i}\right| \\
&\overset{\text{(ii)}}{\lesssim}\left(\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 6}}+\frac{1}{\sigma_r^{\star 4}}\sqrt{\frac{\kappa^3\left(r+\log\left(n+d\right)\right)}{n}}\right)\max_{1\le i\le d}\left\{d_{l,i}^{\star}+\left|d_{l,i}^{\star}-d_{l,i}\right|\right\}+\frac{1}{\sigma_r^{\star 4}}\max_{1\le i\le d}\left|d_{l,i}^{\star}-d_{l,i}\right| \\
&\overset{\text{(iii)}}{\lesssim}\underbrace{\left(\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 6}}+\frac{1}{\sigma_r^{\star 4}}\sqrt{\frac{\kappa^3\left(r+\log\left(n+d\right)\right)}{n}}\right)\max_{1\le i\le d}d_{l,i}^{\star}}_{=:\alpha_4}+\underbrace{\frac{1}{\sigma_r^{\star 4}}\max_{1\le i\le d}\left|d_{l,i}^{\star}-d_{l,i}\right|}_{=:\beta}. \qquad \text{(H.47)}
\end{aligned}
$$

Here, (i) relies on (H.28); (ii) comes from (H.46); (iii) holds true provided that $\zeta_{\mathsf{1st}} \lesssim \sigma_r^{\star 2}$ and $n \gtrsim \kappa^3 r\log(n+d)$. Note that, for each $i\in[d]$,

$$
\begin{aligned}
d_{l,i}^{\star}&=\frac{1}{np^2}\left[\omega_l^{\star 2}+(1-p)\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\right]\left[\omega_i^{\star 2}+(1-p)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\right]+\frac{2\left(1-p\right)^2}{np^2}\left(\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star 2}\boldsymbol{U}_{i,\cdot}^{\star\top}\right)^2 \\
&\lesssim\frac{1}{np^2}\left[\omega_l^{\star 2}+(1-p)\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\right]\left[\omega_{\max}^2+\frac{\mu r}{d}\sigma_1^{\star 2}\right]+\frac{\mu r}{ndp^2}\sigma_1^{\star 2}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 \\
&\lesssim\frac{1}{np^2}\omega_{\max}^4+\frac{\mu r}{ndp^2}\sigma_1^{\star 2}\omega_{\max}^2+\frac{\mu r}{ndp^2}\sigma_1^{\star 2}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2, \qquad \text{(H.48)}
\end{aligned}
$$

which in turn results in

$$
\begin{aligned}
\alpha_4&\overset{\text{(i)}}{\lesssim}\underbrace{\left(\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 6}}+\frac{1}{\sigma_r^{\star 4}}\sqrt{\frac{\kappa^3 r\log\left(n+d\right)}{n}}\right)\frac{\omega_{\max}^4}{np^2}}_{=:\alpha_{4,1}}+\underbrace{\frac{\kappa\mu r}{ndp^2}\left(\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 4}}+\frac{1}{\sigma_r^{\star 2}}\sqrt{\frac{\kappa^3 r\log\left(n+d\right)}{n}}\right)\omega_{\max}^2}_{=:\alpha_{4,2}} \\
&\quad+\underbrace{\frac{\kappa\mu r}{ndp^2}\left(\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 4}}+\frac{1}{\sigma_r^{\star 2}}\sqrt{\frac{\kappa^3 r\log\left(n+d\right)}{n}}\right)\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2}_{=:\alpha_{4,4}} \\
&\overset{\text{(ii)}}{\lesssim}\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right).
\end{aligned}
$$

Here, (i) follows from (H.48), while (ii) holds since

$$
\begin{aligned}
\alpha_{4,1}&\lesssim\delta\frac{\omega_l^{\star 2}\omega_{\min}^2}{np^2\sigma_1^{\star 4}}\lesssim\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right), \\
\alpha_{4,2}&\lesssim\delta\frac{1-p}{ndp^2\kappa\sigma_1^{\star 2}}\omega_l^{\star 2}\lesssim\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right), \\
\alpha_{4,3}&\lesssim\delta\frac{\left(1-p\right)^2}{ndp^2\kappa\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\lesssim\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),
\end{aligned}
$$

provided that $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2}\lesssim\delta/(\kappa^3\mu r)$ and $n\gtrsim\delta^{-2}\kappa^9 r\log(n+d)$. Note that we still need to bound the term β in (H.47), which we leave to the next step.

Step 5: bounding $|d_{l,i}^{\star}-d_{l,i}|$. For each $i\in[d]$, we can decompose

$$
\begin{aligned}
\frac{1}{\sigma_r^{\star 4}}\left|d_{l,i}^{\star}-d_{l,i}\right|&\le\frac{1}{np^2\sigma_r^{\star 4}}\left|\left[\omega_l^{\star 2}+(1-p)\,S_{l,l}^{\star}\right]\left[\omega_i^{\star 2}+(1-p)\,S_{i,i}^{\star}\right]-\left[\omega_l^2+(1-p)\,S_{l,l}\right]\left[\omega_i^2+(1-p)\,S_{i,i}\right]\right| \\
&\quad+\frac{2\left(1-p\right)^2}{np^2\sigma_r^{\star 4}}\left|S_{i,l}^{\star 2}-S_{i,l}^2\right|
\end{aligned}
$$

$$
\begin{aligned}
&\le \underbrace{\frac{1}{np^2\sigma_r^{\star 4}}\left|\omega_l^{\star 2}+(1-p)\,S_{l,l}^{\star}\right|\left|\omega_i^{\star 2}+(1-p)\,S_{i,i}^{\star}-\omega_i^2-(1-p)\,S_{i,i}\right|}_{=:\beta_1} \\
&\quad+\underbrace{\frac{1}{np^2\sigma_r^{\star 4}}\left|\omega_l^{\star 2}+(1-p)\,S_{l,l}^{\star}-\omega_l^2-(1-p)\,S_{l,l}\right|\left|\omega_i^{\star 2}+(1-p)\,S_{i,i}^{\star}\right|}_{=:\beta_2} \\
&\quad+\underbrace{\frac{1}{np^2\sigma_r^{\star 4}}\left|\omega_l^{\star 2}+(1-p)\,S_{l,l}^{\star}-\omega_l^2-(1-p)\,S_{l,l}\right|\left|\omega_i^{2}+(1-p)\,S_{i,i}-\omega_i^{\star 2}-(1-p)\,S_{i,i}^{\star}\right|}_{=:\beta_3} \\
&\quad+\underbrace{\frac{2\,(1-p)^2}{np^2\sigma_r^{\star 4}}\left|S_{i,l}^{\star 2}-S_{i,l}^2\right|}_{=:\beta_4}.
\end{aligned}
$$

Denote $\Delta_i \coloneqq \omega_i^{\star 2}+(1-p)S_{i,i}^{\star}-\omega_i^2-(1-p)S_{i,i}$. We know from (G.26) and (G.24) in Lemma 41 that for each $i\in[d]$,

$$
\begin{aligned}
|\Delta_i| &\le \left|\omega_i^{\star 2}-\omega_i^2\right|+\|\boldsymbol{S}-\boldsymbol{S}^{\star}\|_{\infty} \\
&\lesssim \sqrt{\frac{\log^2(n+d)}{np}}\omega_{\max}^{\star 2}+\zeta_{\mathsf{1st}}\frac{\sqrt{\kappa^2\mu r^2\log(n+d)}}{d}+\sqrt{\frac{\kappa^2\mu^2 r^3\log(n+d)}{nd^2}}\sigma_1^{\star 2}; \qquad \text{(H.49)}
\end{aligned}
$$

and we have also learned from (G.27) and (G.25) that

$$
\begin{aligned}
|\Delta_l| &\le \left|\omega_l^{\star 2}-\omega_l^2\right|+\left|S_{l,l}-S_{l,l}^{\star}\right| \\
&\lesssim \left(\sqrt{\frac{\log^2(n+d)}{np}}+\theta^2\right)\omega_{\max}^2+\left(\theta+\sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 \\
&\quad+\theta\omega_{\max}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\zeta_{\mathsf{2nd}}\sigma_1^{\star}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}. \qquad \text{(H.50)}
\end{aligned}
$$

In addition, it is straightforward to verify that for each $i\in[d]$,

$$
\left|\omega_i^{\star 2}+(1-p)\,S_{i,i}^{\star}\right|\le\omega_{\max}^2+\frac{\mu r}{d}\sigma_1^{\star 2} \qquad \text{(H.51)}
$$

and

$$
\left|\omega_l^{\star 2}+(1-p)\,S_{l,l}^{\star}\right|\le\omega_l^{\star 2}+\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2. \qquad \text{(H.52)}
$$

- Regarding β_1, we can derive

$$
\begin{aligned}
\beta_1 &= \frac{1}{np^2\sigma_r^{\star 4}}\left|\omega_l^{\star 2}+(1-p)\,S_{l,l}^{\star}\right||\Delta_i| \\
&\overset{\text{(i)}}{\lesssim} \frac{1}{np^2\sigma_r^{\star 4}}\left(\omega_l^{\star 2}+\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\right)\left(\sqrt{\frac{\log^2(n+d)}{np}}\omega_{\max}^{\star 2}+\zeta_{\mathsf{1st}}\frac{\sqrt{\kappa^2\mu r^2\log(n+d)}}{d}+\sqrt{\frac{\kappa^2\mu^2 r^3\log(n+d)}{nd^2}}\sigma_1^{\star 2}\right) \\
&\asymp \underbrace{\frac{\omega_{\max}^{\star 4}}{np^2\sigma_r^{\star 4}}\sqrt{\frac{\log^2(n+d)}{np}}}_{=:\beta_{1,1}}+\underbrace{\frac{\omega_l^{\star 2}}{np^2\sigma_r^{\star 4}}\zeta_{\mathsf{1st}}\frac{\sqrt{\kappa^2\mu r^2\log(n+d)}}{d}}_{=:\beta_{1,2}}+\underbrace{\frac{\omega_l^{\star 2}}{np^2\sigma_r^{\star 2}}\sqrt{\frac{\kappa^4\mu^2 r^3\log(n+d)}{nd^2}}}_{=:\beta_{1,3}} \\
&\quad+\underbrace{\frac{\omega_{\max}^{\star 2}}{np^2\sigma_r^{\star 4}}\sqrt{\frac{\log^2(n+d)}{np}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2}_{=:\beta_{1,4}}+\underbrace{\frac{1}{np^2\sigma_r^{\star 4}}\zeta_{\mathsf{1st}}\frac{\sqrt{\kappa^2\mu r^2\log(n+d)}}{d}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2}_{=:\beta_{1,5}} \\
&\quad+\underbrace{\frac{1}{np^2\sigma_r^{\star 2}}\sqrt{\frac{\kappa^4\mu^2 r^3\log(n+d)}{nd^2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2}_{=:\beta_{1,6}}\overset{\text{(ii)}}{\lesssim}\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right).
\end{aligned}
$$

Here, (i) follows from (H.49) and (H.52); (ii) holds since

$$\beta_{1,1} \lesssim \delta \frac{\omega_l^{\star 2}\omega_{\min}^2}{np^2\sigma_1^{\star 4}} \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),$$

$$\beta_{1,2}+\beta_{1,3}+\beta_{1,4} \lesssim \delta \frac{1-p}{ndp^2\kappa\sigma_1^{\star 2}}\omega_l^{\star 2} \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),$$

$$\beta_{1,5}+\beta_{1,6} \lesssim \delta \frac{(1-p)^2}{ndp^2\kappa\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),$$

provided that $np \gtrsim \delta^{-2}\kappa^6\mu^2r^2\log^2(n+d)$, $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2} \lesssim \delta/\sqrt{\kappa^6\mu r^2\log(n+d)}$, $n \gtrsim \delta^{-2}\kappa^8\mu^2r^3\log(n+d)$.

- When it comes to β_2, we can see that

$$\begin{aligned}
\beta_2 &= \frac{1}{np^2\sigma_r^{\star 4}}\left|\omega_i^{\star 2}+(1-p)S_{i,i}^{\star}\right|\left|\Delta_l\right| \\
&\overset{\text{(i)}}{\lesssim} \underbrace{\frac{\omega_{\max}^4}{np^2\sigma_r^{\star 4}}\left(\sqrt{\frac{\log^2(n+d)}{np}}+\theta^2\right)}_{=:\beta_{2,1}} + \underbrace{\frac{\kappa\mu r}{ndp^2\sigma_r^{\star 2}}\omega_{\max}^2\left(\sqrt{\frac{\log^2(n+d)}{np}}+\theta^2\right)}_{=:\beta_{2,2}} \\
&\quad + \underbrace{\frac{1}{np^2\sigma_r^{\star 4}}\omega_{\max}^2\left(\theta+\sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2}_{=:\beta_{2,3}} + \underbrace{\frac{\kappa\mu r}{ndp^2\sigma_r^{\star 2}}\left(\theta+\sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2}_{=:\beta_{2,4}} \\
&\quad + \underbrace{\frac{1}{np^2\sigma_r^{\star 4}}\omega_{\max}^3\theta\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2}_{\beta_{2,5}} + \underbrace{\frac{\sqrt{\kappa}}{np^2\sigma_r^{\star 3}}\omega_{\max}^2\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2}_{=:\beta_{2,6}} + \underbrace{\frac{\kappa}{np^2\sigma_r^{\star 2}}\omega_{\max}^2\zeta_{\mathsf{2nd}}^2}_{=:\beta_{2,7}} \\
&\quad + \underbrace{\frac{\kappa\mu r}{ndp^2\sigma_r^{\star 2}}\theta\omega_{\max}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2}_{=:\beta_{2,8}} + \frac{\kappa^{3/2}\mu r}{ndp^2\sigma_r^{\star}}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{\kappa^2\mu r}{ndp^2}\zeta_{\mathsf{2nd}}^2 \\
&\overset{\text{(ii)}}{\lesssim} \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) + \frac{\kappa\mu r}{ndp^2\sigma_r^{\star 2}}\zeta_{\mathsf{2nd}}\sigma_1^{\star}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{\kappa\mu r}{ndp^2\sigma_r^{\star 2}}\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}.
\end{aligned}$$

Here, (i) follows from (H.50) and (H.51); (ii) holds since

$$\beta_{2,1} \lesssim \delta\frac{\omega_l^{\star 2}\omega_{\min}^2}{np^2\sigma_1^{\star 4}} \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),$$

$$\beta_{2,2}+\beta_{2,3}+\beta_{2,6}+\beta_{2,7} \lesssim \delta\frac{1-p}{4ndp^2\kappa\sigma_1^{\star 2}}\omega_l^{\star 2} \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),$$

$$\beta_{2,4} \lesssim \delta\frac{(1-p)^2}{ndp^2\kappa\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2,$$

$$\beta_{2,5} \lesssim \delta\frac{\omega_{\max}^3}{np^2\sigma_1^{\star 3}}\sqrt{\frac{1}{\kappa d}} \lesssim \delta\frac{1-p}{ndp^2\kappa\sigma_1^{\star 2}}\omega_l^{\star 2} + \delta\frac{\omega_l^{\star 2}\omega_{\min}^2}{np^2\sigma_1^{\star 4}} \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),$$

$$\beta_{2,8} \lesssim \delta\frac{1}{ndp^2\kappa\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\omega_l^{\star} \lesssim \delta\frac{(1-p)^2}{ndp^2\kappa\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \delta\frac{1-p}{4ndp^2\kappa\sigma_1^{\star 2}}\omega_l^{\star 2} \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),$$

provided that $np \gtrsim \delta^{-2}\kappa^6\mu^2r^2\log^2(n+d)$, $\theta \lesssim \delta/(\kappa^3\mu r)$, $n \gtrsim \delta^{-2}\kappa^9\mu^2r^3\log(n+d)$ and $\zeta_{\mathsf{2nd}}\sqrt{d} \lesssim \delta/(\kappa^3\sqrt{\mu r})$.

- With regards to β_3, we notice that

$$|\Delta_i| \lesssim \sqrt{\frac{\log^2(n+d)}{np}}\omega_{\max}^{\star 2} + \zeta_{\mathsf{1st}}\frac{\sqrt{\kappa^2\mu r^2\log(n+d)}}{d} + \sqrt{\frac{\kappa^2\mu^2 r^3\log(n+d)}{nd^2}}\sigma_1^{\star 2} \lesssim \omega_{\max}^2 + \frac{\mu r}{d}\sigma_1^{\star 2}$$

provided that $np \gtrsim \log^2(n+d)$, $n \gtrsim \kappa^2 r\log(n+d)$ and $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2} \lesssim 1/\sqrt{\log(n+d)}$. As a consequence, we can upper bound

$$\beta_3 = \frac{1}{np^2\sigma_r^{\star 4}}|\Delta_i||\Delta_l| \lesssim \frac{1}{np^2\sigma_r^{\star 4}}\left|\omega_i^{\star 2} + (1-p)S_{i,i}^{\star}\right||\Delta_l|.$$

This immediately suggests that β_3 satisfies the same upper bound we derive for β_2.

- Regarding β_4, it is seen that

$$\begin{aligned}\beta_4 &\lesssim \frac{1}{np^2\sigma_r^{\star 4}}\left|S_{i,l}^{\star 2} - S_{i,l}^2\right| \lesssim \frac{1}{np^2\sigma_r^{\star 4}}\left|S_{i,l}^{\star} - S_{i,l}\right|\left|S_{i,l}^{\star} + S_{i,l}\right| \\ &\lesssim \underbrace{\frac{1}{np^2\sigma_r^{\star 4}}\left|S_{i,l}^{\star} - S_{i,l}\right|S_{i,l}^{\star}}_{=:\beta_{4,1}} + \underbrace{\frac{1}{np^2\sigma_r^{\star 4}}\left|S_{i,l}^{\star} - S_{i,l}\right|^2}_{=:\beta_{4,2}}.\end{aligned}$$

Recall from (G.25) in Lemma 41 that

$$\begin{aligned}\left|S_{i,l} - S_{i,l}^{\star}\right| &\lesssim \left(\theta + \sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \theta^2\omega_{\max}^2 \\ &\quad + \left(\theta\omega_{\max} + \zeta_{\mathsf{2nd}}\sigma_1^{\star}\right)\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right) + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2} \\ &\lesssim \left(\theta + \sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\sqrt{\frac{\mu r}{d}}\sigma_1^{\star}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \theta^2\omega_{\max}^2 \\ &\quad + \theta\omega_{\max}\sqrt{\frac{\mu r}{d}}\sigma_1^{\star} + \zeta_{\mathsf{2nd}}\sigma_1^{\star}\sqrt{\frac{\mu r}{d}}\sigma_1^{\star} + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}.\end{aligned} \tag{H.53}$$

The first term $\beta_{4,1}$ can be upper bounded by

$$\begin{aligned}\beta_{4,1} &\lesssim \frac{1}{np^2\sigma_r^{\star 4}}\left|S_{i,l}^{\star} - S_{i,l}\right|\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \lesssim \frac{1}{np^2\sigma_r^{\star 4}}\left|S_{i,l}^{\star} - S_{i,l}\right|\sqrt{\frac{\mu r}{d}}\sigma_1^{\star}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \\ &\overset{\text{(i)}}{\lesssim} \underbrace{\frac{\kappa\mu r}{ndp^2\sigma_r^{\star 2}}\left(\theta + \sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2}_{=:\beta_{4,1,1}} + \underbrace{\frac{1}{np^2\sigma_r^{\star 4}}\theta^2\omega_{\max}^2\sqrt{\frac{\mu r}{d}}\sigma_1^{\star}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2}_{=:\beta_{4,1,2}} \\ &\quad + \underbrace{\frac{\kappa\mu r}{ndp^2\sigma_r^{\star 2}}\theta\omega_{\max}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2}_{=:\beta_{4,1,3}} + \frac{\kappa^{3/2}\mu r}{ndp^2\sigma_r^{\star}}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{1}{np^2\sigma_r^{\star 4}}\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}\sqrt{\frac{\mu r}{d}}\sigma_1^{\star}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \\ &\overset{\text{(ii)}}{\lesssim} \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) + \frac{\kappa^{3/2}\mu r}{ndp^2\sigma_r^{\star}}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \frac{\kappa^{3/2}}{np^2\sigma_r^{\star}}\zeta_{\mathsf{2nd}}^2\sqrt{\frac{\mu r}{d}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2.\end{aligned}$$

$$\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) \gtrsim \frac{1-p}{np\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{\omega_l^{\star 2}}{np\sigma_1^{\star 2}} + \frac{(1-p)^2}{ndp^2\kappa\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \frac{1-p}{ndp^2\kappa\sigma_1^{\star 2}}\omega_l^{\star 2} + \frac{\omega_l^{\star 2}\omega_{\min}^2}{np^2\sigma_1^{\star 4}}.$$

Here, (i) follows from (H.53), and (ii) holds since

$$\begin{aligned}\beta_{4,1,1} &\lesssim \delta\frac{(1-p)^2}{ndp^2\kappa\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right), \\ \beta_{4,1,2} &\lesssim \delta\frac{1-p}{ndp^2\kappa\sigma_1^{\star 2}}\omega_l^{\star 2} \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),\end{aligned}$$

$$\beta_{4,1,3} \lesssim \delta \frac{1}{ndp^2\kappa\sigma_1^{\star 2}} \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 \omega_l^{\star} \lesssim \delta \frac{(1-p)^2}{ndp^2\kappa\sigma_1^{\star 2}} \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2 + \delta \frac{1-p}{4ndp^2\kappa\sigma_1^{\star 2}} \omega_l^{\star 2} \lesssim \delta \lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),$$

provided that $\theta \lesssim \delta/(\kappa^3\mu r)$ and $n \gtrsim \delta^{-2}\kappa^9\mu^2 r^3 \log(n+d)$. The second term $\beta_{4,2}$ can be controlled as follows

$$\begin{aligned}
\beta_{4,2} &\overset{\text{(i)}}{\lesssim} \underbrace{\frac{\kappa\mu r}{ndp^2\sigma_r^{\star 2}} \left(\theta + \sqrt{\frac{\kappa^3 r \log(n+d)}{n}}\right)^2 \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2}_{=:\beta_{4,2,1}} + \underbrace{\frac{1}{np^2\sigma_r^{\star 4}} \theta^4 \omega_{\max}^4}_{=:\beta_{4,2,2}} \\
&\quad + \underbrace{\frac{\kappa\mu r}{ndp^2\sigma_r^{\star 2}} \theta^2 \omega_{\max}^2}_{=:\beta_{4,2,3}} + \frac{\kappa^2\mu r}{ndp^2} \zeta_{\mathsf{2nd}}^2 + \frac{\kappa^2}{np^2} \zeta_{\mathsf{2nd}}^4 \\
&\overset{\text{(ii)}}{\lesssim} \delta \lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) + \frac{\kappa^2\mu r}{ndp^2} \zeta_{\mathsf{2nd}}^2 + \frac{\kappa^2}{np^2} \zeta_{\mathsf{2nd}}^4.
\end{aligned}$$

Here, (i) follows from (H.53), and (ii) holds since

$$\begin{aligned}
\beta_{4,2,1} &\lesssim \delta \frac{(1-p)^2}{ndp^2\kappa\sigma_1^{\star 2}} \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2 \lesssim \delta \lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right), \\
\beta_{4,2,2} &\lesssim \delta \frac{\omega_l^{\star 2} \omega_{\min}^2}{np^2\sigma_1^{\star 4}} \lesssim \delta \lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right), \\
\beta_{4,2,3} &\lesssim \delta \frac{1-p}{ndp^2\kappa\sigma_1^{\star 2}} \omega_l^{\star 2} \lesssim \delta \frac{(1-p)^2}{ndp^2\kappa\sigma_1^{\star 2}} \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2 + \delta \frac{1-p}{4ndp^2\kappa\sigma_1^{\star 2}} \omega_l^{\star 2} \lesssim \delta \lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),
\end{aligned}$$

provided that $\theta \lesssim \sqrt{\delta/(\kappa^3\mu r)}$ and $n \gtrsim \delta^{-1}\kappa^6\mu r^2 \log(n+d)$. Combine the bounds on $\beta_{4,1}$ and $\beta_{4,2}$ to reach

$$\beta_4 \leq \beta_{4,1} + \beta_{4,2} \lesssim \delta \lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) + \frac{\kappa^{3/2}\mu r}{ndp^2\sigma_r^{\star}} \zeta_{\mathsf{2nd}} \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 + \frac{\kappa^2\mu r}{ndp^2} \zeta_{\mathsf{2nd}}^2,$$

provided that $\zeta_{\mathsf{2nd}}\sqrt{d} \lesssim \sqrt{\mu r}$.

Taking together the bounds on β_1, β_2, β_3 and β_4 yields

$$\begin{aligned}
\frac{1}{\sigma_r^{\star 4}} \left| d_{l,i}^{\star} - d_{l,i} \right| &\leq \beta_1 + \beta_2 + \beta_3 + \beta_4 \\
&\lesssim \delta \lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) + \frac{\kappa^{3/2}\mu r}{ndp^2\sigma_r^{\star}} \zeta_{\mathsf{2nd}} \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 + \frac{\kappa^2\mu r}{ndp^2} \zeta_{\mathsf{2nd}}^2,
\end{aligned}$$

with the proviso that $np \gtrsim \delta^{-2}\kappa^6\mu^2 r^2 \log^2(n+d)$, $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2} \lesssim \delta/\sqrt{\kappa^6\mu r^2 \log(n+d)}$, $\theta \lesssim \delta/(\kappa^3\mu r)$, $\zeta_{\mathsf{2nd}}\sqrt{d} \lesssim \delta/(\kappa^3\sqrt{\mu r})$ and $n \gtrsim \delta^{-2}\kappa^9\mu^2 r^3 \log(n+d)$. Given that this holds for all $i \in [d]$, it follows that

$$\beta = \frac{1}{\sigma_r^{\star 4}} \max_{1 \leq i \leq d} \left| d_{l,i}^{\star} - d_{l,i} \right| \lesssim \delta \lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) + \frac{\kappa^{3/2}\mu r}{ndp^2\sigma_r^{\star}} \zeta_{\mathsf{2nd}} \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 + \frac{\kappa^2\mu r}{ndp^2} \zeta_{\mathsf{2nd}}^2.$$

Step 6: putting all pieces together. From the above steps, we can demonstrate that

$$\begin{aligned}
\left\| \boldsymbol{R} \boldsymbol{\Sigma}_{U,l}^{\star} \boldsymbol{R}^{\top} - \boldsymbol{\Sigma}_{U,l} \right\| &\leq \sum_{i=1}^{4} \left\| \boldsymbol{A}_i - \boldsymbol{B}_i \right\| \leq \sum_{i=1}^{3} \left\| \boldsymbol{A}_i - \boldsymbol{B}_i \right\| + \alpha_4 + \beta \\
&\lesssim \delta \lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right) + \underbrace{\frac{\sqrt{\kappa}}{np\sigma_r^{\star}} \zeta_{\mathsf{2nd}} \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2}_{=:\gamma_1} + \underbrace{\frac{\kappa}{np} \zeta_{\mathsf{2nd}}^2}_{=:\gamma_2} + \underbrace{\frac{1}{np} \zeta_{\mathsf{2nd}} \frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}} \omega_l^{\star}}_{=:\gamma_3}
\end{aligned}$$

$$+\underbrace{\frac{\kappa^{3/2}\mu r}{ndp^2\sigma_r^\star}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2}_{=:\gamma_4}+\underbrace{\frac{\kappa^2\mu r}{ndp^2}\zeta_{\mathsf{2nd}}^2}_{=:\gamma_5}.$$

Note that under the assumption of Lemma 39, we have shown in Appendix H.1.3 that

$$\zeta_{\mathsf{2nd}}\lambda_{\min}^{-1/2}\left(\boldsymbol{\Sigma}_{U,l}^\star\right)\lesssim\frac{1}{r^{1/4}\log^{1/2}(n+d)}\ll 1. \tag{H.54}$$

In addition, it follows from Lemma 38 that

$$\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^\star\right)\gtrsim\frac{1}{\sqrt{np}\sigma_1^\star}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\frac{\omega_l^\star}{\sqrt{np}\sigma_1^\star}+\frac{1}{\sqrt{ndp^2\kappa}\sigma_1^\star}\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2. \tag{H.55}$$

As a consequence, we can derive the following upper bounds

$$\begin{aligned}
\gamma_1&\lesssim\frac{\sqrt{\kappa}}{np\sigma_r^\star}\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^\star\right)\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\lesssim\frac{\kappa}{\sqrt{np}}\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right)\lesssim\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right),\\
\gamma_2&\lesssim\frac{\kappa}{np}\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right)\lesssim\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right),\\
\gamma_3&\lesssim\frac{1}{np}\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^\star\right)\frac{\theta}{\sqrt{\kappa}\sigma_r^\star}\omega_l^\star\lesssim\frac{\theta}{\sqrt{np}}\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right)\lesssim\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right),\\
\gamma_4&\lesssim\frac{\kappa^{3/2}\mu r}{ndp^2\sigma_r^\star}\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^\star\right)\left\|\boldsymbol{U}_{l,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\lesssim\frac{\kappa^{5/2}\mu r}{\sqrt{ndp^2}}\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right)\lesssim\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right),\\
\gamma_5&\lesssim\frac{\kappa^2\mu r}{ndp^2}\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right)\lesssim\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right),
\end{aligned}$$

provided that $np\gtrsim\delta^{-2}\kappa^2$ and $ndp^2\gtrsim\delta^{-2}\kappa^5\mu^2r^2$. These allow us to conclude that

$$\left\|\boldsymbol{R}\boldsymbol{\Sigma}_{U,l}^\star\boldsymbol{R}^\top-\boldsymbol{\Sigma}_{U,l}\right\|\lesssim\delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^\star\right),$$

as long as the following assumptions hold: $np\gtrsim\delta^{-2}\kappa^6\mu^2r^2\log^2(n+d)$, $n\gtrsim\delta^{-2}\kappa^9\mu^2r^3\log(n+d)$, $\zeta_{\mathsf{1st}}/\sigma_r^{\star2}\lesssim\delta/(\kappa^3\mu r\sqrt{\log(n+d)})$, $\theta\lesssim\delta/(\kappa^3\mu r)$, and $\zeta_{\mathsf{2nd}}\sqrt{d}\lesssim\delta/(\kappa^3\sqrt{\mu r})$. In what follows, we take a closer look at the last three assumptions.

- In view of (H.3), it is readily seen that $\zeta_{\mathsf{1st}}/\sigma_r^{\star2}\lesssim\delta/(\kappa^3\mu r\sqrt{\log(n+d)})$ can be guaranteed by

  $$ndp^2\gtrsim\delta^{-2}\kappa^8\mu^4r^4\log^5(n+d),\qquad np\gtrsim\delta^{-2}\kappa^8\mu^3r^3\log^3(n+d),$$

  and

  $$\frac{\omega_{\max}^2}{p\sigma_r^{\star2}}\sqrt{\frac{d}{n}}\lesssim\frac{\delta}{\kappa^3\mu r\log^{3/2}(n+d)},\qquad\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d}{np}}\lesssim\frac{\delta}{\kappa^{7/2}\mu r\log(n+d)}.$$

- By virtue of (H.39), it is straightforward to see that $\theta\lesssim\delta/(\kappa^3\mu r)$ can be guaranteed by $\zeta_{\mathsf{1st}}/\sigma_r^{\star2}\lesssim\delta/(\kappa^3\mu r)$ — the latter is already guaranteed by $\zeta_{\mathsf{1st}}/\sigma_r^{\star2}\lesssim\delta/(\kappa^3\mu r\sqrt{\log(n+d)})$.

- In addition, we know that $\zeta_{\mathsf{2nd}}\sqrt{d}\lesssim\delta/(\kappa^3\sqrt{\mu r})$ is equivalent to

  $$\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star2}}\lesssim\frac{\delta\sqrt{d}}{\kappa^5\mu^{3/2}r^{3/2}\log(n+d)}\wedge\frac{\delta^{1/2}}{\kappa^{9/4}\mu^{1/2}r^{1/2}\log^{1/4}(n+d)},$$

  which can be guaranteed by $\zeta_{\mathsf{1st}}/\sigma_r^{\star2}\lesssim\delta/(\kappa^3\mu r\sqrt{\log(n+d)})$ as long as $d\gtrsim\kappa^4\mu r\log(n+d)$.

To summarize, we can conclude that the required assumptions for the above results to hold are: $n\gtrsim\delta^{-2}\kappa^9\mu^2r^3\log(n+d)$, $d\gtrsim\kappa^4\mu r\log(n+d)$,

$$ndp^2\gtrsim\delta^{-2}\kappa^8\mu^4r^4\log^5(n+d),\qquad np\gtrsim\delta^{-2}\kappa^8\mu^3r^3\log^3(n+d),$$

and

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star2}}\sqrt{\frac{d}{n}}\lesssim\frac{\delta}{\kappa^3\mu r\log^{3/2}(n+d)},\qquad\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d}{np}}\lesssim\frac{\delta}{\kappa^{7/2}\mu r\log(n+d)}.$$

H.2.4 Proof of Lemma 43

The proof of Lemma 43 is similar to the proof of Lemma 13 in Appendix E.2.4. Define

$$\boldsymbol{\Delta} \coloneqq \left(\boldsymbol{U}-\boldsymbol{U}^{\star}\boldsymbol{R}^{\top}\right)_{l,\cdot}\boldsymbol{R}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)^{-1/2}\boldsymbol{R}^{\top}-\left(\boldsymbol{U}-\boldsymbol{U}^{\star}\boldsymbol{R}^{\top}\right)_{l,\cdot}\boldsymbol{\Sigma}_{U,l}^{-1/2}.$$

Following the same analysis as in Appendix E.2.4, we can show that with probability exceeding $1-O((n+d)^{-10})$

$$\begin{aligned}\|\boldsymbol{\Delta}\|_2 &\lesssim \left\|\left(\boldsymbol{U}-\boldsymbol{U}^{\star}\boldsymbol{R}^{\top}\right)_{l,\cdot}\right\|_2 \lambda_{\min}^{-1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)\delta \\ &\lesssim \left[\frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_l^{\star}\right)+\zeta_{\mathsf{2nd}}\right]\lambda_{\min}^{-1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)\delta.\end{aligned}$$

Here, $\delta\in(0,1)$ is the (unspecified) quantity appearing in Lemma 42 such that

$$\left\|\boldsymbol{R}\boldsymbol{\Sigma}_{U,l}^{\star}\boldsymbol{R}^{\top}-\boldsymbol{\Sigma}_{U,l}\right\| \lesssim \delta\lambda_{\min}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),$$

and the last line follows from (G.23a) in Lemma 40. Let

$$\zeta \coloneqq \widetilde{C}\left[\frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_l^{\star}\right)\lambda_{\min}^{-1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)\delta+\zeta_{\mathsf{2nd}}\lambda_{\min}^{-1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)\delta\right]$$

for some sufficiently large constant $\tilde{C}>0$ such that $\mathbb{P}(\|\boldsymbol{\Delta}\|_2\le\zeta)\ge 1-O((n+d)^{-10})$. Following the same analysis in Appendix E.2.4 allows us to show that

$$\mathbb{P}\left(\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{R}^{\top}\in\mathsf{CR}_{U,l}^{1-\alpha}\right)=1-\alpha+O\left(\log^{-1/2}(n+d)\right),$$

provided that $\zeta r^{1/4}\lesssim 1/\sqrt{\log(n+d)}$, or equivalently,

$$\frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_l^{\star}\right)\lambda_{\min}^{-1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)\delta+\zeta_{\mathsf{2nd}}\lambda_{\min}^{-1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)\delta\lesssim\frac{1}{r^{1/4}\sqrt{\log(n+d)}}. \tag{H.56}$$

It remains to choose $\delta\in(0,1)$ to satisfy the above condition. First, we have learned from the proof of Lemma 39 (more specifically, Step 3 in Appendix H.1.3) that

$$\zeta_{\mathsf{2nd}}\lambda_{\min}^{-1/2}(\boldsymbol{\Sigma}_l^{\star})\lesssim 1/(r^{1/4}\log^{1/2}(n+d))$$

holds under the conditions of Lemma 39. As a result, it is sufficient to verify that

$$\frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_l^{\star}\right)\lambda_{\min}^{-1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)\delta\lesssim\frac{1}{r^{1/4}\sqrt{\log(n+d)}}. \tag{H.57}$$

To this end, recall from Lemma 38 that

$$\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)\gtrsim\left(\frac{1}{\sqrt{np}\sigma_1^{\star}}+\frac{1}{\sqrt{ndp^2\kappa}\sigma_1^{\star}}\right)\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_l^{\star}\right)+\frac{\omega_{\min}^2}{\sqrt{np^2}\sigma_1^{\star 2}}.$$

In view of the definition of θ (cf. (G.22)), we can show that

$$\begin{aligned}\frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_l^{\star}\right) &\asymp \underbrace{\frac{1}{\sigma_r^{\star}}\sqrt{\frac{r\log^2(n+d)}{np}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_l^{\star}\right)}_{=:\alpha_1}+\underbrace{\frac{1}{\sigma_r^{\star}}\sqrt{\frac{\kappa\mu r^2\log^3(n+d)}{ndp^2}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_l^{\star}\right)}_{=:\alpha_2} \\ &\quad+\underbrace{\frac{\omega_{\max}}{\sigma_r^{\star 2}}\sqrt{\frac{r\log^2(n+d)}{np^2}}\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2}_{=:\alpha_3}+\underbrace{\frac{\omega_{\max}^2}{\sigma_r^{\star 2}}\sqrt{\frac{r\log^2(n+d)}{np^2}}}_{=:\alpha_4}\end{aligned}$$

$$\lesssim \sqrt{\kappa^3\mu r^2\log^3(n+d)}\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),$$

where the last line holds since

$$\alpha_1 \lesssim \sqrt{\kappa r\log^2(n+d)}\left[\frac{1}{\sqrt{np}\sigma_1^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_l^{\star}\right)\right] \lesssim \sqrt{\kappa r\log^2(n+d)}\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),$$

$$\alpha_2 \lesssim \sqrt{\kappa^3\mu r^2\log^3(n+d)}\left[\frac{1}{\sqrt{ndp^2\kappa}\sigma_1^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_l^{\star}\right)\right] \lesssim \sqrt{\kappa^3\mu r^2\log^3(n+d)}\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),$$

$$\alpha_3 \lesssim \sqrt{\kappa^3\mu r^2\log^2(n+d)}\left(\frac{1}{\sqrt{ndp^2\kappa}\sigma_1^{\star}}\omega_l^{\star}\right) \lesssim \sqrt{\kappa^3\mu r^2\log^2(n+d)}\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right),$$

$$\alpha_4 \lesssim \sqrt{\kappa^2 r\log^2(n+d)}\left(\frac{\omega_{\min}^2}{\sqrt{np^2}\sigma_1^{\star 2}}\right) \lesssim \sqrt{\kappa^2 r\log^2(n+d)}\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right).$$

The above bound immediately gives

$$\frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_l^{\star}\right)\lambda_{\min}^{-1/2}\left(\boldsymbol{\Sigma}_{U,l}^{\star}\right)\delta \lesssim \sqrt{\kappa^3\mu r^2\log^3(n+d)}\delta.$$

Therefore, the desired condition (H.57) — and hence $\zeta r^{1/4} \lesssim 1/\sqrt{\log(n+d)}$ — can be guaranteed by taking

$$\delta = \frac{1}{\kappa^{3/2}\mu^{1/2}r^{5/4}\log^2(n+d)}.$$

After δ is specified, we can easily check that under our choice of δ, the assumptions in Lemma 42 read $n \gtrsim \kappa^{12}\mu^3 r^{11/2}\log^5(n+d)$, $d \gtrsim \kappa^4\mu r\log(n+d)$,

$$ndp^2 \gtrsim \kappa^{11}\mu^5 r^{13/2}\log^9(n+d), \qquad np \gtrsim \kappa^{11}\mu^4 r^{11/2}\log^7(n+d),$$

and

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\kappa^{9/2}\mu^{3/2}r^{9/4}\log^{7/2}(n+d)}, \qquad \frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\kappa^5\mu^{3/2}r^{9/4}\log^3(n+d)}.$$

This concludes the proof.

H.3 Auxiliary lemmas for Theorem 17

H.3.1 Proof of Lemma 44

Throughout this section, all the probabilistic arguments are conditional on $\boldsymbol{F}$, and we shall always assume that the $\sigma(\boldsymbol{F})$-measurable high-probability event $\mathcal{E}_{\mathsf{good}}$ occurs. Following the same analysis as in Appendix H.1.1 (Proof of Lemma 37), we can obtain

$$\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}} - \boldsymbol{U}^{\natural} = \boldsymbol{Z} + \boldsymbol{\Psi}_{\boldsymbol{U}}$$

with $\boldsymbol{R}_{\boldsymbol{U}} = \arg\min_{\boldsymbol{O}\in\mathcal{O}^{r\times r}}\|\boldsymbol{U}\boldsymbol{O}-\boldsymbol{U}^{\natural}\|_{\mathrm{F}}^2$, where

$$\boldsymbol{Z} = \left[\boldsymbol{E}\boldsymbol{M}^{\natural\top} + \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^{\top}\right)\right]\boldsymbol{U}^{\natural}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-2}$$

and $\boldsymbol{\Psi}_{\boldsymbol{U}}$ is a residual matrix obeying

$$\mathbb{P}\left(\left\|\boldsymbol{\Psi}_{\boldsymbol{U}}\right\|_{2,\infty}\mathbb{1}_{\mathcal{E}_{\mathsf{good}}} \lesssim \zeta_{\mathsf{2nd}} \,\middle|\, \boldsymbol{F}\right) \geq 1 - O\left((n+d)^{-10}\right). \tag{H.58}$$

Here, ζ_{2nd} is a quantity defined in Lemma 37. From the proof of Lemma 40, we know that

$$\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}^2\boldsymbol{R}_{\boldsymbol{U}} = \boldsymbol{\Sigma}^{\natural 2} + \boldsymbol{\Psi}_{\boldsymbol{\Sigma}}$$

holds for some matrix $\boldsymbol{\Psi}_{\boldsymbol{\Sigma}}$ satisfying

$$\mathbb{P}\left(\|\boldsymbol{\Psi}_{\boldsymbol{\Sigma}}\| \mathbb{1}_{\mathcal{E}_{\mathsf{good}}} \lesssim \sqrt{\frac{d}{\kappa\mu r\log(n+d)}}\zeta_{\mathsf{2nd}}\sigma_r^{\star 2} \,\Big|\, \boldsymbol{F}\right) \geq 1 - O\left((n+d)^{-10}\right). \tag{H.59}$$

Armed with the above facts, we can write

$$\begin{aligned}
\boldsymbol{S} - \boldsymbol{M}^{\natural}\boldsymbol{M}^{\natural\top} &= \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^{\top} - \boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{U}^{\natural\top} \\
&= \boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{\Sigma}^2\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{U}^{\top} - \boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{U}^{\natural\top} \\
&= \boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{U}^{\top} + \boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{\Psi}_{\boldsymbol{\Sigma}}\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{U}^{\top} - \boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{U}^{\natural\top} \\
&= \left(\boldsymbol{U}^{\natural} + \boldsymbol{Z} + \boldsymbol{\Psi}_{\boldsymbol{U}}\right)\boldsymbol{\Sigma}^{\natural 2}\left(\boldsymbol{U}^{\natural} + \boldsymbol{Z} + \boldsymbol{\Psi}_{\boldsymbol{U}}\right)^{\top} + \boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{\Psi}_{\boldsymbol{\Sigma}}\boldsymbol{R}_{\boldsymbol{U}}^{\top}\boldsymbol{U}^{\top} - \boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{U}^{\natural\top} \\
&= \boldsymbol{X} + \boldsymbol{\Phi},
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{X} &= \boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{Z}^{\top} + \boldsymbol{Z}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{U}^{\natural} \\
&= \boldsymbol{E}\boldsymbol{M}^{\natural\top} + \boldsymbol{M}^{\natural}\boldsymbol{E}^{\top} + \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^{\top}\right)\boldsymbol{U}^{\natural}\boldsymbol{U}^{\natural\top} + \boldsymbol{U}^{\natural}\boldsymbol{U}^{\natural\top}\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^{\top}\right)
\end{aligned}$$

and

$$\boldsymbol{\Phi} = \boldsymbol{U}^{\natural}\boldsymbol{\Sigma}^{\natural 2}\boldsymbol{\Psi}_{\boldsymbol{U}}^{\top} + \boldsymbol{Z}\boldsymbol{\Sigma}^{\natural 2}\left(\boldsymbol{Z} + \boldsymbol{\Psi}_{\boldsymbol{U}}\right)^{\top} + \boldsymbol{\Psi}_{\boldsymbol{U}}\boldsymbol{\Sigma}^{\natural 2}\left(\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\right)^{\top} + \boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\boldsymbol{\Psi}_{\boldsymbol{\Sigma}}\left(\boldsymbol{U}\boldsymbol{R}_{\boldsymbol{U}}\right)^{\top}.$$

It has already been shown in Lemma 40 that for each $l \in [d]$,

$$\|\boldsymbol{Z}_{l,\cdot}\|_2 \lesssim \frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\left(\left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right) \tag{H.60}$$

and

$$\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star}\|_{2,\infty} \lesssim \frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}\sqrt{\frac{r\log(n+d)}{d}}$$

hold with probability at least $1 - O\big((n+d)^{-10}\big)$. As an immediate consequence,

$$\|\boldsymbol{U}\|_{2,\infty} \leq \|\boldsymbol{U}^{\star}\|_{2,\infty} + \|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star}\|_{2,\infty} \lesssim \sqrt{\frac{\mu r}{d}} + \frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}\sqrt{\frac{r\log(n+d)}{d}} \lesssim \sqrt{\frac{\mu r}{d}}, \tag{H.61}$$

provided that $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2} \lesssim 1/\sqrt{\log(n+d)}$. Then with probability exceeding $1 - O((n+d)^{-10})$, we have

$$\begin{aligned}
|\Phi_{i,j}| &\overset{\text{(i)}}{\lesssim} \sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2\|\boldsymbol{\Psi}_{\boldsymbol{U}}\|_{2,\infty} + \|\boldsymbol{Z}_{i,\cdot}\|_2\|\boldsymbol{Z}_{j,\cdot}\|_2 + \|\boldsymbol{Z}_{i,\cdot}\|_2\|\boldsymbol{\Psi}_{\boldsymbol{U}}\|_{2,\infty} + \|\boldsymbol{\Psi}_{\boldsymbol{U}}\|_{2,\infty}\|\boldsymbol{U}_{j,\cdot}\|_2\right) \\
&\quad + \|\boldsymbol{\Psi}_{\boldsymbol{\Sigma}}\|\|\boldsymbol{U}_{i,\cdot}\|_2\|\boldsymbol{U}_{j,\cdot}\|_2 \\
&\overset{\text{(ii)}}{\lesssim} \zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \|\boldsymbol{Z}_{i,\cdot}\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 + \|\boldsymbol{Z}_{j,\cdot}\|_2 + \zeta_{\mathsf{2nd}}\right) + \sigma_1^{\star 2}\|\boldsymbol{Z}_{i,\cdot}\|_2\|\boldsymbol{Z}_{j,\cdot}\|_2 \\
&\quad + \sqrt{\frac{d}{\kappa\mu r\log(n+d)}}\zeta_{\mathsf{2nd}}\sigma_r^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \|\boldsymbol{Z}_{i,\cdot}\|_2 + \zeta_{\mathsf{2nd}}\right)\sqrt{\frac{\mu r}{d}} \\
&\lesssim \zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \|\boldsymbol{Z}_{i,\cdot}\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 + \|\boldsymbol{Z}_{j,\cdot}\|_2 + \zeta_{\mathsf{2nd}}\right) + \sigma_1^{\star 2}\|\boldsymbol{Z}_{i,\cdot}\|_2\|\boldsymbol{Z}_{j,\cdot}\|_2 \\
&\overset{\text{(iii)}}{\lesssim} \zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 + \frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\omega_{\max} + \zeta_{\mathsf{2nd}}\right) + \theta^2\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right)\left(\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_l^{\star}\right) \\
&\overset{\text{(iv)}}{\lesssim} \left(\zeta_{\mathsf{2nd}}\sigma_1^{\star 2} + \theta^2\omega_{\max}\sigma_1^{\star}\right)\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right) + \theta^2\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \theta^2\omega_{\max}^2 + \zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}
\end{aligned}$$

for each $i, j \in [d]$. Here, (i) makes use of (G.4) and the fact that $\boldsymbol{U}^{\natural} = \boldsymbol{U}^{\star}\boldsymbol{Q}$ for some orthonormal matrix $\boldsymbol{Q}$ (cf. (D.4)); (ii) follows from (H.58), (H.59) and (H.61); (iii) comes from (H.60) and holds provided that $\theta \lesssim 1$; and (iv) invokes the AM-GM inequality.

H.3.2 Proof of Lemma 45

In this subsection, we shall focus on establishing the claimed result for the case when $i \neq j$; the case when $i = j$ can be proved in a similar (in fact, easier) manner. In view of the expression (G.31), we can write

$$\mathsf{var}\left(X_{i,j}|\boldsymbol{F}\right) = \underbrace{\sum_{l=1}^{n} M_{j,l}^{\natural 2}\sigma_{i,l}^{2}}_{=:\alpha_1} + \underbrace{\sum_{l=1}^{n} M_{i,l}^{\natural 2}\sigma_{j,l}^{2}}_{=:\alpha_2} + \underbrace{\sum_{l=1}^{n}\sum_{k:k\neq i}\sigma_{i,l}^{2}\sigma_{k,l}^{2}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^{2}}_{=:\alpha_3} + \underbrace{\sum_{l=1}^{n}\sum_{k:k\neq j}\sigma_{j,l}^{2}\sigma_{k,l}^{2}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{i,\cdot}^{\star\top}\right)^{2}}_{=:\alpha_4},$$

thus motivating us to study the behavior of α_1, α_2, α_3 and α_4 respectively.

- To begin with, the concentration of α_1 and α_2 have already been studied in Lemma 15, which gives

$$\alpha_1 = \underbrace{\frac{1-p}{np}\left(S_{i,i}^{\star}S_{j,j}^{\star} + 2S_{i,j}^{\star 2}\right) + \frac{\omega_i^{\star 2}}{np}S_{j,j}^{\star}}_{=:\alpha_1^{\star}} + r_1$$

  and

$$\alpha_2 = \underbrace{\frac{1-p}{np}\left[S_{i,i}^{\star}S_{j,j}^{\star} + 2S_{i,j}^{\star 2}\right] + \frac{\omega_j^{\star 2}}{np}S_{i,i}^{\star}}_{=:\alpha_2^{\star}} + r_2,$$

  where

$$|r_1| \lesssim \sqrt{\frac{\log^3(n+d)}{n}}\alpha_1^{\star}, \qquad |r_2| \lesssim \sqrt{\frac{\log^3(n+d)}{n}}\alpha_2^{\star}.$$

- When it comes to α_3, we make the observation that

$$\begin{aligned}
\alpha_3 &= \sum_{l=1}^{n}\sum_{k:k\neq i}\left[\frac{1-p}{np}\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2 + \frac{\omega_i^{\star 2}}{np}\right]\left[\frac{1-p}{np}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2 + \frac{\omega_k^{\star 2}}{np}\right]\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2 \\
&= \left(\frac{1-p}{np}\right)^2\sum_{k:k\neq i}\left[\sum_{l=1}^{n}\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2\right]\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2 \\
&\quad + \frac{1-p}{n^2p^2}\sum_{k:k\neq i}\sum_{l=1}^{n}\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2\omega_k^{\star 2}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2 \\
&\quad + \frac{(1-p)\,\omega_i^{\star 2}}{n^2p^2}\sum_{k:k\neq i}\sum_{l=1}^{n}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2 + \frac{\omega_i^{\star 2}}{np^2}\sum_{k:k\neq i}\omega_k^{\star 2}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2.
\end{aligned}$$

  On the event $\mathcal{E}_{\mathsf{good}}$, it is seen from (G.12a) and (G.12b) that

$$\begin{aligned}
\left|\frac{1}{n}\sum_{l=1}^{n}\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2 - S_{k,k}^{\star}S_{i,i}^{\star} - 2S_{i,k}^{\star 2}\right| &\lesssim \sqrt{\frac{\log^3(n+d)}{n}}S_{k,k}^{\star}S_{i,i}^{\star} \\
\left|\frac{1}{n}\sum_{l=1}^{n}\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2 - S_{i,i}^{\star}\right| &\lesssim \sqrt{\frac{\log(n+d)}{n}}S_{i,i}^{\star} \\
\left|\frac{1}{n}\sum_{l=1}^{n}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2 - S_{k,k}^{\star}\right| &\lesssim \sqrt{\frac{\log(n+d)}{n}}S_{k,k}^{\star}
\end{aligned}$$

  for each $k \in [d]\backslash\{i\}$. As a consequence, one can express

$$\alpha_3 = \frac{\omega_i^{\star 2}}{np^2}\sum_{k:k\neq i}\omega_k^{\star 2}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2 + \frac{1-p}{np^2}S_{i,i}^{\star}\sum_{k:k\neq i}\omega_k^{\star 2}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2$$

$$+\frac{(1-p)\,\omega_i^{\star 2}}{np^2}\sum_{k:k\neq i} S_{k,k}^{\star}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2+\frac{(1-p)^2}{np^2}\sum_{k:k\neq i}\left(S_{k,k}^{\star}S_{i,i}^{\star}+2S_{i,k}^{\star 2}\right)\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2+\widetilde{r}_3$$

for some residual term $\widetilde{r}_3$ satisfying

$$|\widetilde{r}_3|\lesssim\sqrt{\frac{\log^3(n+d)}{n}}\,|\alpha_3-\widetilde{r}_3|\asymp\sqrt{\frac{\log^3(n+d)}{n}}\,|\alpha_3|\,,$$

where the last relation holds provided that $n\gg\log^3(n+d)$. This allows one to decompose α_3 as follows

$$\begin{aligned}\alpha_3&=\frac{1}{np^2}\sum_{k:k\neq i}\left\{\left[\omega_i^{\star 2}+(1-p)\,S_{i,i}^{\star}\right]\left[\omega_k^{\star 2}+(1-p)\,S_{k,k}^{\star}\right]+2\,(1-p)^2\,S_{i,k}^{\star 2}\right\}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2+\widetilde{r}_3\\&=\underbrace{\frac{1}{np^2}\sum_{k=1}^{d}\left\{\left[\omega_i^{\star 2}+(1-p)\,S_{i,i}^{\star}\right]\left[\omega_k^{\star 2}+(1-p)\,S_{k,k}^{\star}\right]+2\,(1-p)^2\,S_{i,k}^{\star 2}\right\}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2}_{=:\alpha_3^{\star}}+r_3,\end{aligned}$$

where we define

$$r_3=\widetilde{r}_3-\underbrace{\frac{1}{np^2}\left\{\left[\omega_i^{\star 2}+(1-p)\,S_{i,i}^{\star}\right]^2+2\,(1-p)^2\,S_{i,i}^{\star 2}\right\}\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2}_{=:\delta}.$$

Recalling from Claim 2 that

$$\sum_{k=1}^{d}S_{k,k}^{\star}\boldsymbol{U}_{k,\cdot}^{\star\top}\boldsymbol{U}_{k,\cdot}^{\star}=\sum_{k=1}^{d}\left\|\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\boldsymbol{U}_{k,\cdot}^{\star\top}\boldsymbol{U}_{k,\cdot}^{\star}\succeq\frac{\sigma_r^{\star 2}}{4d}\boldsymbol{I}_r, \tag{H.62}$$

we can reach

$$\begin{aligned}\alpha_3^{\star}&\geq\frac{1}{np^2}\sum_{k=1}^{d}\left\{\left[\omega_i^{\star 2}+(1-p)\,S_{i,i}^{\star}\right]\left[\omega_k^{\star 2}+(1-p)\,S_{k,k}^{\star}\right]\right\}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2\\&\gtrsim\frac{1}{np^2}\left[\omega_i^{\star 2}+(1-p)\,S_{i,i}^{\star}\right]\boldsymbol{U}_{j,\cdot}^{\star}\left[\sum_{k=1}^{d}\omega_k^{\star 2}\boldsymbol{U}_{k,\cdot}^{\star\top}\boldsymbol{U}_{k,\cdot}^{\star}+\sum_{k=1}^{d}S_{k,k}^{\star}\boldsymbol{U}_{k,\cdot}^{\star\top}\boldsymbol{U}_{k,\cdot}^{\star}\right]\boldsymbol{U}_{j,\cdot}^{\star\top}\\&\gtrsim\frac{1}{np^2}\left[\omega_i^{\star 2}+(1-p)\,S_{i,i}^{\star}\right]\left[\omega_{\min}^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2+\frac{\sigma_r^{\star 2}}{d}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right] &\text{(H.63)}\\&\gtrsim\frac{1}{np^2}\left[\omega_i^{\star 2}+(1-p)\,S_{i,i}^{\star}\right]\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\left(\omega_{\min}^2+\frac{\sigma_r^{\star 2}}{d}\right), &\text{(H.64)}\end{aligned}$$

where the second line uses the assumption that p is strictly bounded away from 1 (so that $1-p\asymp 1$), and the penultimate line makes use of (H.62). This immediately leads to

$$\begin{aligned}\delta&\lesssim\frac{1}{np^2}\left[\omega_i^{\star 2}+(1-p)\,S_{i,i}^{\star}\right]^2\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\\&\lesssim\frac{\mu r}{ndp^2}\left[\omega_i^{\star 2}+(1-p)\,S_{i,i}^{\star}\right]\left(\omega_{\min}^2+\frac{\mu r}{d}\sigma_1^{\star 2}\right)\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\\&\lesssim\frac{\kappa\mu^2r^2}{d}\alpha_3^{\star},\end{aligned}$$

where the penultimate line comes from the assumption $\omega_{\min}\asymp\omega_{\max}$, $1-p\asymp 1$, and the fact that $S_{i,i}^{\star}=\|\boldsymbol{U}_{i,\cdot}^{\star}\|_2^2\|\boldsymbol{\Sigma}^{\star}\|^2\leq\frac{\mu r}{d}\sigma_1^{\star 2}$, and the last inequality results from (H.64). As a consequence, we arrive at

$$|r_3|\leq|\widetilde{r}_3|+|\delta|\lesssim\sqrt{\frac{\log^3(n+d)}{n}}\,|\alpha_3|+\frac{\kappa\mu^2r^2}{d}\alpha_3^{\star}\lesssim\sqrt{\frac{\log^3(n+d)}{n}}\,(\alpha_3^{\star}+|r_3|)+\frac{\kappa\mu^2r^2}{d}\alpha_3^{\star}$$

$$\lesssim \left(\sqrt{\frac{\log^3(n+d)}{n}} + \frac{\kappa\mu^2 r^2}{d}\right)\alpha_3^\star,$$

where the last relation holds true as long as $n \gg \log^3(n+d)$.

- Repeating our analysis for α_3 allows one to show that

$$\alpha_4 = \underbrace{\frac{1}{np^2}\sum_{k=1}^{d}\left\{\left[\omega_j^{\star 2} + (1-p)\,S_{j,j}^\star\right]\left[\omega_k^{\star 2} + (1-p)\,S_{k,k}^\star\right] + 2\,(1-p)^2\,S_{j,k}^{\star 2}\right\}\left(\boldsymbol{U}_{k,\cdot}^\star \boldsymbol{U}_{i,\cdot}^{\star\top}\right)^2}_{=:\alpha_4^\star} + r_4,$$

where the residual term r_4 obeys

$$|r_4| \lesssim \left(\sqrt{\frac{\log^3(n+d)}{n}} + \frac{\kappa\mu^2 r^2}{d}\right)\alpha_4^\star.$$

Putting the above bounds together, we can conclude that

$$\mathsf{var}\left(X_{i,j}|\boldsymbol{F}\right) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = \widetilde{v}_{i,j} + r_{i,j},$$

where

$$\begin{aligned}
\widetilde{v}_{i,j} &\coloneqq \alpha_1^\star + \alpha_2^\star + \alpha_3^\star + \alpha_4^\star \\
&= \frac{2\,(1-p)}{np}\left(S_{i,i}^\star S_{j,j}^\star + 2S_{i,j}^{\star 2}\right) + \frac{1}{np}\left(\omega_i^{\star 2} S_{j,j}^\star + \omega_j^{\star 2} S_{i,i}^\star\right) \\
&\quad + \frac{1}{np^2}\sum_{k=1}^{d}\left\{\left[\omega_i^{\star 2} + (1-p)\,S_{i,i}^\star\right]\left[\omega_k^{\star 2} + (1-p)\,S_{k,k}^\star\right] + 2\,(1-p)^2\,S_{i,k}^{\star 2}\right\}\left(\boldsymbol{U}_{k,\cdot}^\star \boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2 \\
&\quad + \frac{1}{np^2}\sum_{k=1}^{d}\left\{\left[\omega_j^{\star 2} + (1-p)\,S_{j,j}^\star\right]\left[\omega_k^{\star 2} + (1-p)\,S_{k,k}^\star\right] + 2\,(1-p)^2\,S_{j,k}^{\star 2}\right\}\left(\boldsymbol{U}_{k,\cdot}^\star \boldsymbol{U}_{i,\cdot}^{\star\top}\right)^2,
\end{aligned}$$

and the residual term $r_{i,j}$ is bounded in magnitude by

$$\begin{aligned}
|r_{i,j}| &\le |r_1| + |r_2| + |r_3| + |r_4| \\
&\lesssim \sqrt{\frac{\log^3(n+d)}{n}}\left(\alpha_1^\star + \alpha_2^\star\right) + \left(\sqrt{\frac{\log^3(n+d)}{n}} + \frac{\kappa\mu^2 r^2}{d}\right)\left(\alpha_3^\star + \alpha_4^\star\right) \\
&\lesssim \left(\sqrt{\frac{\log^3(n+d)}{n}} + \frac{\kappa\mu^2 r^2}{d}\right)\widetilde{v}_{i,j}.
\end{aligned}$$

To finish up, it remains to develop a lower bound on $\widetilde{v}_{i,j}$. Towards this end, we first observe that

$$\begin{aligned}
\alpha_1^\star + \alpha_2^\star &= \frac{2\,(1-p)}{np}\left(S_{i,i}^\star S_{j,j}^\star + 2S_{i,j}^{\star 2}\right) + \frac{1}{np}\left(\omega_i^{\star 2} S_{j,j}^\star + \omega_j^{\star 2} S_{i,i}^\star\right) \\
&\gtrsim \frac{1}{np} S_{i,i}^\star S_{j,j}^\star + \frac{\omega_{\min}^2}{np}\left(S_{j,j}^\star + S_{i,i}^\star\right) \\
&\gtrsim \frac{1}{np}\left\|\boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2 \left\|\boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star\right\|_2^2 + \frac{\omega_{\min}^2 \sigma_r^{\star 2}}{np}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2^2 + \left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2^2\right),
\end{aligned}$$

where we have made use of the assumption $1-p \asymp 1$ and the elementary inequality $S^\star_{i,i} = \left\|\boldsymbol{U}^\star_{i,\cdot}\boldsymbol{\Sigma}^\star\right\|_2^2 \geq \sigma_r^{\star 2}\left\|\boldsymbol{U}^\star_{i,\cdot}\right\|_2^2$. In view of (H.63), the assumption $1-p \asymp 1$ as well as the bound $\left\|\boldsymbol{U}^\star_{j,\cdot}\boldsymbol{\Sigma}^\star\right\|_2^2 \leq \sigma_1^{\star 2}\left\|\boldsymbol{U}^\star_{j,\cdot}\right\|_2^2$, we can further lower bound

$$\begin{aligned}
\alpha_3^\star &\gtrsim \frac{1}{np^2}\left[\omega_i^{\star 2} + (1-p)\,S^\star_{i,i}\right]\left\|\boldsymbol{U}^\star_{j,\cdot}\right\|_2^2\left(\omega_{\min}^2 + \frac{\sigma_r^{\star 2}}{d}\right) \\
&\geq \frac{\omega_i^{\star 2}\omega_{\min}^2}{np^2}\left\|\boldsymbol{U}^\star_{j,\cdot}\right\|_2^2 + \frac{\omega_i^{\star 2}}{np^2}\frac{\sigma_r^{\star 2}}{d}\left\|\boldsymbol{U}^\star_{j,\cdot}\right\|_2^2 + \frac{(1-p)\,S^\star_{i,i}}{np^2}\left\|\boldsymbol{U}^\star_{j,\cdot}\right\|_2^2\frac{\sigma_r^{\star 2}}{d} \\
&\gtrsim \frac{\omega_{\min}^4}{np^2}\left\|\boldsymbol{U}^\star_{j,\cdot}\right\|_2^2 + \frac{\omega_{\min}^2\sigma_r^{\star 2}}{ndp^2}\left\|\boldsymbol{U}^\star_{j,\cdot}\right\|_2^2 + \frac{1}{ndp^2\kappa}\left\|\boldsymbol{U}^\star_{i,\cdot}\boldsymbol{\Sigma}^\star\right\|_2^2\left\|\boldsymbol{U}^\star_{j,\cdot}\boldsymbol{\Sigma}^\star\right\|_2^2,
\end{aligned}$$

and similarly,

$$\alpha_4^\star \gtrsim \frac{\omega_{\min}^4}{np^2}\left\|\boldsymbol{U}^\star_{i,\cdot}\right\|_2^2 + \frac{\omega_{\min}^2\sigma_r^{\star 2}}{ndp^2}\left\|\boldsymbol{U}^\star_{i,\cdot}\right\|_2^2 + \frac{1}{ndp^2\kappa}\left\|\boldsymbol{U}^\star_{i,\cdot}\boldsymbol{\Sigma}^\star\right\|_2^2\left\|\boldsymbol{U}^\star_{j,\cdot}\boldsymbol{\Sigma}^\star\right\|_2^2.$$

The above bounds taken collectively yield

$$\begin{aligned}
\widetilde{v}_{i,j} &= \alpha_1^\star + \alpha_2^\star + \alpha_3^\star + \alpha_4^\star \\
&\gtrsim \frac{1}{\min\left\{ndp^2\kappa, np\right\}}\left\|\boldsymbol{U}^\star_{i,\cdot}\boldsymbol{\Sigma}^\star\right\|_2^2\left\|\boldsymbol{U}^\star_{j,\cdot}\boldsymbol{\Sigma}^\star\right\|_2^2 + \left(\frac{\omega_{\min}^2\sigma_r^{\star 2}}{\min\left\{ndp^2, np\right\}} + \frac{\omega_{\min}^4}{np^2}\right)\left(\left\|\boldsymbol{U}^\star_{i,\cdot}\right\|_2^2 + \left\|\boldsymbol{U}^\star_{j,\cdot}\right\|_2^2\right)
\end{aligned}$$

as claimed.

H.3.3 Proof of Lemma 46

As before, all the probabilistic arguments in this subsection are conditional on $\boldsymbol{F}$, and it is assumed, unless otherwise noted, that the $\sigma(\boldsymbol{F})$-measurable high-probability event $\mathcal{E}_{\mathsf{good}}$ occurs.

Step 1: Gaussian approximation of $X_{i,j}$ using the Berry-Esseen Theorem. Recalling the definition (G.30) of $X_{i,j}$, let us denote

$$X_{i,j} = \sum_{l=1}^n \underbrace{\left\{M^\natural_{j,l}E_{i,l} + M^\natural_{i,l}E_{j,l} + E_{i,l}\left[\mathcal{P}_{-i,\cdot}\left(\boldsymbol{E}_{\cdot,l}\right)\right]^\top \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}_{j,\cdot} + E_{j,l}\left[\mathcal{P}_{-j,\cdot}\left(\boldsymbol{E}_{\cdot,l}\right)\right]^\top \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}_{i,\cdot}\right\}}_{=: Y_l},$$

where the Y_l's are statistically independent. Apply the Berry-Esseen Theorem (cf. Theorem 19) to reach

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(X_{i,j}/\sqrt{\mathsf{var}\left(X_{i,j}|\boldsymbol{F}\right)} \leq t \,\middle|\, \boldsymbol{F}\right) - \Phi\left(t\right)\right| \lesssim \gamma\left(\boldsymbol{F}\right),$$

where

$$\gamma\left(\boldsymbol{F}\right) \coloneqq \frac{1}{\mathsf{var}^{3/2}\left(X_{i,j}|\boldsymbol{F}\right)}\sum_{l=1}^n \mathbb{E}\left[\left|Y_l\right|^3 \middle| \boldsymbol{F}\right].$$

It thus boils down to controlling the quantity $\gamma(\boldsymbol{F})$, which forms the main content of the remaining proof.

- We first develop a high-probability bound on each $|Y_l|$. For any $l \in [n]$, observe that

$$\left[\mathcal{P}_{-i,\cdot}\left(\boldsymbol{E}_{\cdot,l}\right)\right]^\top \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}_{j,\cdot} = \sum_{k:k\neq i} E_{k,l}\left(\boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right)_{k,j}.$$

  It is straightforward to calculate that

$$L \coloneqq \max_{k:k\neq i}\left|E_{k,l}\left(\boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right)_{k,j}\right| \leq B\left\|\boldsymbol{U}^\star_{j,\cdot}\right\|_2\left\|\boldsymbol{U}^\star\right\|_{2,\infty} \lesssim B\sqrt{\frac{\mu r}{d}}\left\|\boldsymbol{U}^\star_{j,\cdot}\right\|_2,$$

$$V := \sum_{k:k\neq i} \mathsf{var}\left(E_{k,l}\left(\boldsymbol{U}^{\star}\boldsymbol{U}^{\star\top}\right)_{k,j}\right) \leq \sum_{k=1}^{d} \sigma_{k,l}^{2}\left(\boldsymbol{U}^{\star}\boldsymbol{U}^{\star\top}\right)_{k,j}^{2} \leq \sigma_{\mathsf{ub}}^{2}\left\|\boldsymbol{U}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right\|_{2}^{2}.$$

In view of the Bernstein inequality (Vershynin, 2017, Theorem 2.8.4), with probability exceeding $1 - O((n+d)^{-101})$ we have

$$\begin{aligned}
\left|\left[\mathcal{P}_{-i,\cdot}\left(\boldsymbol{E}_{\cdot,l}\right)\right]^{\top}\boldsymbol{U}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right| &\lesssim \sqrt{V\log\left(n+d\right)} + L\log\left(n+d\right) \\
&\lesssim \sigma_{\mathsf{ub}}\left\|\boldsymbol{U}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right\|_{2}\sqrt{\log\left(n+d\right)} + B\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\sqrt{\frac{\mu r}{d}}\log\left(n+d\right) \\
&\lesssim \sigma_{\mathsf{ub}}\sqrt{\log\left(n+d\right)}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2} + B\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\sqrt{\frac{\mu r}{d}}\log\left(n+d\right).
\end{aligned}$$

By defining two quantities B_i and B_j such that

$$\max_{l\in[n]}\left|E_{i,l}\right| \leq B_{i} \qquad \text{and} \qquad \max_{l\in[n]}\left|E_{j,l}\right| \leq B_{j},$$

we can immediately obtain from the preceding inequality that

$$\left|E_{i,l}\left[\mathcal{P}_{-i,\cdot}\left(\boldsymbol{E}_{\cdot,l}\right)\right]^{\top}\boldsymbol{U}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right| \lesssim \sigma_{\mathsf{ub}}B_{i}\sqrt{\log\left(n+d\right)}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2} + BB_{i}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\sqrt{\frac{\mu r}{d}}\log\left(n+d\right). \tag{H.65}$$

Similarly, we can also show that with probability exceeding $1 - O((n+d)^{-101})$,

$$\left|E_{j,l}\left[\mathcal{P}_{-j,\cdot}\left(\boldsymbol{E}_{\cdot,l}\right)\right]^{\top}\boldsymbol{U}^{\star}\boldsymbol{U}_{i,\cdot}^{\star\top}\right| \lesssim \sigma_{\mathsf{ub}}B_{j}\sqrt{\log\left(n+d\right)}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2} + BB_{j}\left\|\boldsymbol{U}_{i,\cdot}\right\|_{2}\sqrt{\frac{\mu r}{d_{1}}}\log\left(n+d\right).$$

In addition, it is easily seen from the inequality $|M_{i,l}^{\natural}| \leq \|\boldsymbol{U}_{i,\cdot}^{\natural}\|_{2}\|\boldsymbol{\Sigma}^{\natural}\|\|\boldsymbol{V}^{\natural}\|_{2,\infty}$ that

$$\begin{aligned}
\left|M_{j,l}^{\natural}E_{i,l} + M_{i,l}^{\natural}E_{j,l}\right| &\leq \sigma_{1}^{\natural}\left\|\boldsymbol{U}_{j,\cdot}^{\natural}\right\|_{2}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty}B_{i} + \sigma_{1}^{\natural}\left\|\boldsymbol{U}_{i,\cdot}^{\natural}\right\|_{2}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty}B_{j} \\
&= \sigma_{1}^{\natural}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty}\left(B_{i}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2} + B_{j}\left\|\boldsymbol{U}_{i,\cdot}\right\|_{2}\right).
\end{aligned}$$

Therefore we know that with probability exceeding $1 - O((n+d)^{-101})$

$$\left|Y_{l}\right| \lesssim \left[\sigma_{\mathsf{ub}}\sqrt{\log\left(n+d\right)} + B\sqrt{\frac{\mu r}{d}}\log\left(n+d\right) + \sigma_{1}^{\natural}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty}\right]\left(B_{i}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2} + B_{j}\left\|\boldsymbol{U}_{i,\cdot}\right\|_{2}\right).$$

Let

$$C_{\mathsf{prob}} := \widetilde{C}_{1}\left[\sigma_{\mathsf{ub}}\sqrt{\log\left(n+d\right)} + B\sqrt{\frac{\mu r}{d}}\log\left(n+d\right) + \sigma_{1}^{\natural}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty}\right]\left(B_{i}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2} + B_{j}\left\|\boldsymbol{U}_{i,\cdot}\right\|_{2}\right)$$

for some sufficiently large constant $\tilde{C} > 0$ such that with probability exceeding $1 - O((n+d)^{-101})$,

$$\max_{l\in[n]}\left|Y_{l}\right| \leq C_{\mathsf{prob}}. \tag{H.66}$$

- In addition, we are also in need of a deterministic upper bound on each $|Y_l|$. Observe that

$$\left|\left[\mathcal{P}_{-i,\cdot}\left(\boldsymbol{E}_{\cdot,l}\right)\right]^{\top}\boldsymbol{U}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right| \leq \left\|\mathcal{P}_{-i,\cdot}\left(\boldsymbol{E}_{\cdot,l}\right)\right\|_{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2} \leq \left\|\boldsymbol{E}_{\cdot,l}\right\|_{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2} \leq \sqrt{d}B\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2},$$

and similarly,

$$\left|\left[\mathcal{P}_{-j,\cdot}\left(\boldsymbol{E}_{\cdot,l}\right)\right]^{\top}\boldsymbol{U}^{\star}\boldsymbol{U}_{i,\cdot}^{\star\top}\right| \leq \sqrt{d}B\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}. \tag{H.67}$$

As a result, we can derive

$$\begin{aligned}
\left|Y_{l}\right| &\leq \sigma_{1}^{\natural 2}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty}\left(B_{i}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2} + B_{j}\left\|\boldsymbol{U}_{i,\cdot}\right\|_{2}\right) + B_{i}\sqrt{d}B\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2} + B_{j}\sqrt{d}B\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2} \\
&\leq \left(\sigma_{1}^{\natural 2}\left\|\boldsymbol{V}^{\natural}\right\|_{2,\infty} + B\sqrt{d}\right)\left(B_{i}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2} + B_{j}\left\|\boldsymbol{U}_{i,\cdot}\right\|_{2}\right) =: C_{\mathsf{det}}
\end{aligned} \tag{H.68}$$

With the above probabilistic and deterministic bounds in place (see (H.66) and (H.68)), we can decompose $\mathbb{E}\left[|Y_l|^3 \,|\boldsymbol{F}\right]$ for each $l \in [n]$ as follows

$$\begin{aligned}
\mathbb{E}\left[|Y_l|^3 \,|\boldsymbol{F}\right] &= \mathbb{E}\left[|Y_l|^3 \,\mathbb{1}_{|Y_l|\le C_{\mathsf{prob}}} \,|\boldsymbol{F}\right] + \mathbb{E}\left[|Y_l|^3 \,\mathbb{1}_{|Y_l|> C_{\mathsf{prob}}} \,|\boldsymbol{F}\right] \\
&\le C_{\mathsf{prob}}\mathbb{E}\left[Y_l^2|\boldsymbol{F}\right] + C_{\mathsf{det}}^3\mathbb{P}\left(|Y_l| > C_{\mathsf{prob}}\right) \\
&\lesssim C_{\mathsf{prob}}\mathbb{E}\left[Y_l^2|\boldsymbol{F}\right] + C_{\mathsf{det}}^3\left(n+d\right)^{-101}.
\end{aligned}$$

Recognizing that $\sum_{l=1}^n \mathbb{E}[Y_l^2|\boldsymbol{F}] = \mathsf{var}(X_{i,j}|\boldsymbol{F})$, we obtain

$$\gamma\left(\boldsymbol{F}\right) \le \frac{1}{\mathsf{var}^{3/2}\left(X_{i,j}|\boldsymbol{F}\right)}\sum_{l=1}^{n}\mathbb{E}\left[|Y_l|^3 \,|\boldsymbol{F}\right] \lesssim \underbrace{\frac{C_{\mathsf{prob}}}{\mathsf{var}^{1/2}\left(X_{i,j}|\boldsymbol{F}\right)}}_{=:\alpha} + \underbrace{\frac{C_{\mathsf{det}}^3\left(n+d\right)^{-100}}{\mathsf{var}^{3/2}\left(X_{i,j}|\boldsymbol{F}\right)}}_{\beta}. \tag{H.69}$$

It remains to control the terms $\mathsf{var}\left(X_{i,j}|\boldsymbol{F}\right)$, C_{prob} and C_{det}. On the event $\mathcal{E}_{\mathsf{good}}$, we have learned from Lemma 45 that

$$\begin{aligned}
\mathsf{var}^{1/2}\left(X_{i,j}|\boldsymbol{F}\right) \asymp \widetilde{v}_{i,j}^{1/2} &\gtrsim \frac{1}{\sqrt{ndp^2\kappa\wedge np}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 \\
&\quad + \frac{\sigma_r^{\star}}{\sqrt{ndp^2\kappa\wedge np}}\omega_{\min}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right) + \frac{\omega_{\min}^2}{\sqrt{np}}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right),
\end{aligned} \tag{H.70}$$

provided that $n \gg \log^3(n+d)$ and $d \gg \kappa\mu^2 r^2$. Moreover, it is seen from (G.9) that

$$B_i \lesssim \frac{1}{p}\sqrt{\frac{\log\left(n+d\right)}{n}}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_i^{\star}\right) \qquad \text{and} \qquad B_j \lesssim \frac{1}{p}\sqrt{\frac{\log\left(n+d\right)}{n}}\left(\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_j^{\star}\right), \tag{H.71}$$

which immediately lead to

$$B_i\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 + B_j\left\|\boldsymbol{U}_{i,\cdot}\right\|_2 \lesssim \frac{1}{p}\sqrt{\frac{\log\left(n+d\right)}{n}}\left[\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\sigma_1^{\star} + \omega_{\max}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)\right].$$

As a result, we can bound

$$\begin{aligned}
C_{\mathsf{prob}} &\lesssim \left[\frac{\sigma_{\mathsf{ub}}}{p}\sqrt{\frac{\log^2\left(n+d\right)}{n}} + B\sqrt{\frac{\mu r\log\left(n+d\right)}{ndp^2}}\log\left(n+d\right) + \sigma_1^{\star}\frac{\log\left(n+d\right)}{np}\right] \\
&\quad\cdot\left[\frac{1}{\sigma_r^{\star}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_{\max}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)\right]
\end{aligned}$$

as well as

$$\begin{aligned}
C_{\mathsf{det}} &\lesssim \left(\sigma_1^{\star}\frac{\log\left(n+d\right)}{np} + \frac{B}{p}\sqrt{\frac{d\log\left(n+d\right)}{n}}\right) \\
&\quad\cdot\left[\frac{1}{\sigma_r^{\star}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_{\max}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)\right],
\end{aligned}$$

where we have utilized (G.4) and (G.7). In what follows, we bound the terms α and β in (H.69) separately.

- Regarding α, one first observes that

$$\alpha\widetilde{v}_{i,j}^{1/2} \lesssim \underbrace{\frac{\sigma_{\mathsf{ub}}}{p\sigma_r^{\star}}\sqrt{\frac{\log^2\left(n+d\right)}{n}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2}_{=:\alpha_1} + \underbrace{\frac{\sigma_{\mathsf{ub}}}{p}\sqrt{\frac{\log^2\left(n+d\right)}{n}}\omega_{\max}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)}_{=:\alpha_2}$$

$$
\begin{aligned}
&+ \underbrace{B\sqrt{\frac{\mu r \log^3(n+d)}{ndp^2}}\frac{1}{\sigma_r^\star}\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2}_{=:\alpha_3} + \underbrace{B\sqrt{\frac{\mu r \log^3(n+d)}{ndp^2}}\omega_{\max}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right)}_{=:\alpha_4} \\
&+ \underbrace{\frac{\sqrt{\kappa}\log(n+d)}{np}\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2}_{=:\alpha_5} + \underbrace{\sigma_1^\star\frac{\log(n+d)}{np}\omega_{\max}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right)}_{=:\alpha_6} \\
\lesssim\ & \frac{1}{\sqrt{\log(n+d)}}\widetilde{v}_{i,j}^{1/2}.
\end{aligned}
$$

To see why the last step holds, we note that due to (G.8), (G.9) and the assumption $\omega_{\max} \asymp \omega_{\min}$, the following inequalities hold:

$$
\begin{aligned}
\alpha_1 &\asymp \frac{1}{p\sigma_r^\star}\left(\sqrt{\frac{\mu r\log(n+d)}{ndp}}\sigma_1^\star + \frac{\omega_{\max}}{\sqrt{np}}\right)\sqrt{\frac{\log^2(n+d)}{n}}\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \\
&\lesssim \frac{1}{\sqrt{\log(n+d)}}\left(\frac{1}{\sqrt{ndp^2\kappa}}+\frac{1}{\sqrt{np}}\right)\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \lesssim \frac{1}{\sqrt{\log(n+d)}}\widetilde{v}_{i,j}^{1/2}, \\
\alpha_2 &\asymp \frac{1}{p}\left(\sqrt{\frac{\mu r\log(n+d)}{ndp}}\sigma_1^\star + \frac{\omega_{\max}}{\sqrt{np}}\right)\sqrt{\frac{\log^2(n+d)}{n}}\omega_{\max}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right) \\
&\lesssim \frac{1}{\sqrt{\log(n+d)}}\left(\frac{1}{\sqrt{ndp^2}}+\frac{1}{\sqrt{np}}\right)\omega_{\min}\sigma_r^\star\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right) \lesssim \frac{1}{\sqrt{\log(n+d)}}\widetilde{v}_{i,j}^{1/2}, \\
\alpha_3 &\asymp \left(\frac{1}{p}\sqrt{\frac{\kappa\mu^2r^2\log^4(n+d)}{n^2d^2p^2}} + \frac{\omega_{\max}}{p\sigma_r^\star}\sqrt{\frac{\mu r\log^4(n+d)}{n^2dp^2}}\right)\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \\
&\lesssim \frac{1}{\sqrt{\log(n+d)}}\frac{1}{\sqrt{ndp^2\kappa}}\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \lesssim \frac{1}{\sqrt{\log(n+d)}}\widetilde{v}_{i,j}^{1/2}, \\
\alpha_4 &\asymp \left(\frac{\mu r\log^2(n+d)}{ndp^2}\sigma_1^\star + \frac{\omega_{\max}}{p}\sqrt{\frac{\mu r\log^4(n+d)}{n^2dp^2}}\right)\omega_{\max}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right) \\
&\lesssim \frac{1}{\sqrt{\log(n+d)}}\frac{\sigma_r^\star}{\sqrt{ndp^2}}\omega_{\min}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right) \lesssim \frac{1}{\sqrt{\log(n+d)}}\widetilde{v}_{i,j}^{1/2}, \\
\alpha_5 &\lesssim \frac{1}{\sqrt{\log(n+d)}}\frac{1}{\sqrt{np}}\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 \lesssim \frac{1}{\sqrt{\log(n+d)}}\widetilde{v}_{i,j}^{1/2}, \\
\alpha_6 &\lesssim \frac{1}{\sqrt{\log(n+d)}}\frac{\sigma_r^\star}{\sqrt{np}}\omega_{\min}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right) \lesssim \frac{1}{\sqrt{\log(n+d)}}\widetilde{v}_{i,j}^{1/2},
\end{aligned}
$$

provided that $np \gtrsim \kappa^2\mu r\log^4(n+d)$, $ndp^2 \gtrsim \kappa^2\mu^2r^2\log^5(n+d)$,

$$
\omega_{\max}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\sqrt{\kappa\log^3(n+d)}}, \qquad \text{and} \qquad \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{1}{np^2}} \lesssim \frac{1}{\sqrt{\kappa\mu r\log^5(n+d)}}. \tag{H.72}
$$

In view of (H.38), we know that the second condition in (H.72) can be guaranteed by $ndp^2 \gtrsim \kappa\mu r\log^5(n+d)$ and

$$
\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\sqrt{\kappa\mu r\log^5(n+d)}}.
$$

As a result, we conclude that

$$
\alpha \lesssim \frac{1}{\sqrt{\log(n+d)}}
$$

holds as long as $np \gtrsim \kappa^2 \mu r \log^4(n+d)$, $ndp^2 \gtrsim \kappa^2\mu^2 r^2 \log^5(n+d)$,

$$\omega_{\max}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\sqrt{\kappa \log^3(n+d)}}, \qquad \text{and} \qquad \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\sqrt{\kappa\mu r \log^5(n+d)}}.$$

- Turning to the term β, we make the observation that

$$\begin{aligned}
C_{\mathsf{det}}\widetilde{v}_{i,j}^{-1/2} &\overset{\text{(i)}}{\lesssim} \left(\sigma_1^\star \frac{\log(n+d)}{np} + \frac{B}{p}\sqrt{\frac{d\log(n+d)}{n}}\right)\left[\frac{1}{\sigma_r^\star}\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \omega_{\max}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right)\right] \\
&\quad \cdot \left\{\frac{1}{\sqrt{np}}\left[\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \omega_{\min}\sigma_r^\star\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right)\right]\right\}^{-1} \\
&\lesssim \left(\sigma_1^\star \frac{\log(n+d)}{np} + \frac{B}{p}\sqrt{\frac{d\log(n+d)}{n}}\right)\frac{\sqrt{np}}{\sigma_r^\star} \overset{\text{(ii)}}{\lesssim} \left(\frac{\log(n+d)}{np} + \frac{1}{p}\sqrt{\frac{d\log(n+d)}{n}}\right)\sqrt{\kappa np} \\
&\lesssim \frac{\sqrt{\kappa}\log(n+d)}{\sqrt{np}} + \sqrt{\frac{\kappa d\log(n+d)}{p}} \overset{\text{(iii)}}{\lesssim} \sqrt{nd}.
\end{aligned}$$

Here (i) follows from (H.70); (ii) holds as long as $B \lesssim \sigma_r^\star$, which can be guaranteed by $ndp^2 \gtrsim \kappa\mu r \log(n+d)$ and

$$\frac{\omega_{\max}}{p\sigma_r^\star}\sqrt{\frac{1}{n}} \lesssim \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} + \frac{1}{\sqrt{ndp}} \lesssim \frac{1}{\sqrt{\log(n+d)}};$$

and (iii) holds as long as $np \gtrsim \kappa\log(n+d)$. This immediately results in

$$\beta \lesssim \frac{C_{\mathsf{det}}^3 (n+d)^{-100}}{\widetilde{v}_{i,j}^{3/2}} \lesssim (nd)^{3/2}(n+d)^{-100} \lesssim (n+d)^{-50}.$$

Putting the above pieces together, we have shown that

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(X_{i,j}/\sqrt{\mathsf{var}(X_{i,j}|\boldsymbol{F})} \leq t \,\middle|\, \boldsymbol{F}\right) - \Phi(t)\right| \lesssim \gamma(\boldsymbol{F}) \lesssim \alpha + \beta \lesssim \frac{1}{\sqrt{\log(n+d)}}, \tag{H.73}$$

provided that $np \gtrsim \kappa^2\mu r\log^4(n+d)$, $ndp^2 \gtrsim \kappa^2\mu^2 r^2\log^5(n+d)$,

$$\omega_{\max}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\sqrt{\kappa \log^3(n+d)}}, \qquad \text{and} \qquad \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\sqrt{\kappa\mu r \log^5(n+d)}}.$$

Step 2: derandomizing the conditional variance. Given Lemma 45, we can follow the same analysis as Step 2 in Appendix E.3.3 (the proof of Lemma 16) to demonstrate that

$$\left|\mathbb{P}\left(X_{i,j}/\sqrt{\widetilde{v}_{i,j}} \leq t \,\middle|\, \boldsymbol{F}\right) - \Phi(t)\right| \leq (|\beta_1| + |\beta_2|) \lesssim \frac{1}{\sqrt{\log(n+d)}}$$

for all $t\in\mathbb{R}$, provided that $n \gtrsim \log^4(n+d)$ and $d \gtrsim \kappa\mu^2 r^2\sqrt{\log(n+d)}$.

Step 3: taking higher-order errors into account. By following the same analysis as Step 3 in Appendix E.1.3 (proof of Lemma 9), we know that if one can show

$$\mathbb{P}\left(\widetilde{v}_{i,j}^{-1/2}\left|\Phi_{i,j}\right| \lesssim \log^{-1/2}(n+d) \,\middle|\, \boldsymbol{F}\right) \geq 1 - O\left((n+d)^{-10}\right), \tag{H.74}$$

then it holds that

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\widetilde{v}_{i,j}^{-1/2}\left(\boldsymbol{S}-\boldsymbol{M}^{\natural}\boldsymbol{M}^{\natural\top}\right)_{i,j}\le t\,\middle|\,\boldsymbol{F}\right)-\Phi\left(t\right)\right|\lesssim\frac{1}{\sqrt{\log\left(n+d\right)}}.$$

As a result, we shall focus on proving (H.74) from now on. Recall from Lemma 44 that with probability exceeding $1-O((n+d)^{-10})$,

$$\begin{aligned}\left|\Phi_{i,j}\right|\lesssim\zeta_{i,j}\asymp&\underbrace{\theta^{2}\left[\sigma_{1}^{\star}\omega_{\max}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\right)+\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\right]}_{=:\gamma_{1}}\\&+\underbrace{\zeta_{\mathsf{2nd}}\sigma_{1}^{\star2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\right)}_{=:\gamma_{2}}+\underbrace{\theta^{2}\omega_{\max}^{2}}_{=:\gamma_{3}}+\underbrace{\zeta_{\mathsf{2nd}}^{2}\sigma_{1}^{\star2}}_{=:\gamma_{4}},\end{aligned}\tag{H.75}$$

and from Lemma 45 that

$$\begin{aligned}\widetilde{v}_{i,j}^{1/2}\gtrsim&\frac{1}{\sqrt{\min\left\{ndp^{2}\kappa,np\right\}}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\frac{\sigma_{r}^{\star}}{\sqrt{\min\left\{ndp^{2},np\right\}}}\omega_{\min}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\right)\\&+\frac{\omega_{\min}^{2}}{\sqrt{np}}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\right).\end{aligned}$$

Armed with these bounds, we seek to derive sufficient conditions that guarantee $\zeta_{i,j}\lesssim\delta\widetilde{v}_{i,j}^{1/2}$.

- Regarding the quantity γ_1 defined in (H.75), we note that

$$\begin{aligned}\gamma_{1}\lesssim&\left(\frac{\kappa r\log^{2}\left(n+d\right)}{np}+\frac{\kappa^{2}\mu r^{2}\log^{3}\left(n+d\right)}{ndp^{2}}+\frac{\omega_{\max}^{2}}{\sigma_{r}^{\star2}}\frac{\kappa r\log^{2}\left(n+d\right)}{np^{2}}\right)\\&\cdot\left[\sigma_{1}^{\star}\omega_{\max}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\right)+\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\right]\\\lesssim&\,\delta\frac{1}{\sqrt{\min\left\{ndp^{2}\kappa,np\right\}}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\delta\frac{\sigma_{r}^{\star}}{\sqrt{\min\left\{ndp^{2},np\right\}}}\omega_{\min}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\right)\\\lesssim&\,\delta\widetilde{v}_{i,j}^{1/2},\end{aligned}$$

  provided that $np\gtrsim\delta^{-2}\kappa^{3}r^{2}\log^{4}(n+d)$, $ndp^{2}\gtrsim\delta^{-2}\kappa^{5}\mu^{2}r^{4}\log^{6}(n+d)$, and

$$\frac{\omega_{\max}^{2}}{p\sigma_{r}^{\star2}}\sqrt{\frac{d}{n}}\lesssim\frac{\delta}{\kappa^{3/2}r\log^{2}\left(n+d\right)}.$$

- Regarding the term γ_2 defined in (H.75), we recall from the proof of Lemma 39 (Step 3 in Appendix H.1.3) that with probability exceeding $1-O((n+d)^{-10})$,

$$\begin{aligned}\zeta_{\mathsf{2nd}}\lesssim&\frac{\delta}{\sqrt{\kappa}}\min\left\{\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,i}^{\star}\right),\lambda_{\min}^{1/2}\left(\boldsymbol{\Sigma}_{U,j}^{\star}\right)\right\}\\\lesssim&\frac{\delta}{\sqrt{\kappa}}\frac{1}{\sqrt{\min\left\{ndp^{2}\kappa,np\right\}}\sigma_{1}^{\star}}\left(\min\left\{\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2},\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\right\}+\omega_{\min}\right)+\frac{\delta}{\sqrt{\kappa}}\frac{\omega_{\min}^{2}}{\sqrt{np^{2}}\sigma_{1}^{\star2}},\end{aligned}\tag{H.76}$$

  provided that $d\gtrsim\delta^{-2}\kappa^{7}\mu^{2}r^{2}\log^{4}(n+d)$,

$$\frac{\omega_{\max}}{\sigma_{r}^{\star}}\sqrt{\frac{d}{np}}\lesssim\frac{\delta}{\kappa^{7/2}\mu^{1/2}r^{1/2}\log^{3/2}\left(n+d\right)},\qquad\frac{\omega_{\max}^{2}}{p\sigma_{r}^{\star2}}\sqrt{\frac{d}{n}}\lesssim\frac{\delta}{\kappa^{3}\mu^{1/2}r^{1/2}\log^{5/2}\left(n+d\right)},$$

  and

$$\min\left\{\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2},\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\right\}$$

$$\gtrsim \delta^{-1}\left[\frac{\kappa^{9/2}\mu^{5/2}r^{2}\log^{9/2}(n+d)}{\sqrt{ndp^{2}}}+\frac{\kappa^{4}\mu^{3/2}r\log^{5/2}(n+d)}{\sqrt{np}}+\frac{\kappa^{4}\mu^{2}r^{3/2}\log^{3}(n+d)}{\sqrt{d}}\right]\sqrt{\frac{r}{d}}\sigma_1^{\star}.$$

Consequently, we obtain

$$\begin{aligned}
\gamma_2 &\lesssim \frac{\delta}{\sqrt{\kappa}}\frac{1}{\sqrt{\min\{ndp^2\kappa, np\}}\sigma_1^{\star}}\left(\min\left\{\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2,\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right\}+\omega_{\min}\right)\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)\\
&\quad+\frac{\delta}{\sqrt{\kappa}}\frac{\omega_{\min}^2}{\sqrt{np^2}\sigma_1^{\star 2}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)\\
&\lesssim \frac{\delta}{\sqrt{\kappa}}\frac{1}{\sqrt{\min\{ndp^2\kappa, np\}}}\left[\sqrt{\kappa}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_{\min}\sigma_1^{\star}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)\right]\\
&\quad+\frac{\delta}{\sqrt{\kappa}}\frac{\omega_{\min}^2}{\sqrt{np^2}}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)\\
&\lesssim \delta\widetilde{v}_{i,j}^{1/2}.
\end{aligned}$$

- We now move on to the term γ_3 defined in (H.75), which obeys

$$\gamma_3 \asymp \underbrace{\frac{\kappa r\log^2(n+d)}{np}\omega_{\max}^2}_{\gamma_{3,1}}+\underbrace{\frac{\kappa^2\mu r^2\log^3(n+d)}{ndp^2}\omega_{\max}^2}_{=:\gamma_{3,2}}+\underbrace{\frac{\omega_{\max}^4}{\sigma_r^{\star 2}}\frac{\kappa r\log^2(n+d)}{np^2}}_{=:\gamma_{3,3}}\lesssim \delta\widetilde{v}_{i,j}^{1/2}.$$

Here, the last relation holds since

$$\gamma_{3,1}\lesssim \delta\frac{\sigma_r^{\star}}{\sqrt{np}}\omega_{\min}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)\lesssim\delta\widetilde{v}_{i,j}^{1/2},$$

$$\gamma_{3,2}+\gamma_{3,3}\lesssim \delta\frac{\omega_{\min}^2}{\sqrt{np}}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)\lesssim\delta\widetilde{v}_{i,j}^{1/2},$$

with the proviso that

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\gtrsim \delta^{-1}\kappa r\log^2(n+d)\left[\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}}+\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}+\frac{\kappa\mu r\log(n+d)}{\sqrt{ndp^2}}\right]\sqrt{\frac{1}{d}}. \tag{H.77}$$

- With regards to the term γ_4 defined in (H.75), we can see from (H.76) that

$$\begin{aligned}
\zeta_{\mathsf{2nd}} &\lesssim \frac{\delta}{\sqrt{\kappa}}\frac{1}{\sqrt{\min\{ndp^2\kappa, np\}}\sigma_1^{\star}}\left(\min\left\{\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2,\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\right\}+\omega_{\min}\right)+\frac{\delta}{\sqrt{\kappa}}\frac{\omega_{\min}^2}{\sqrt{np^2}\sigma_1^{\star 2}}\\
&\lesssim \left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2,
\end{aligned}$$

provided that $np\gtrsim\delta^{-2}/\kappa$, $ndp^2\gtrsim\delta^{-2}/\kappa^2$,

$$\begin{aligned}
\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 &\gtrsim \frac{\delta}{\sqrt{\kappa}}\frac{1}{\sqrt{\min\{ndp^2\kappa, np\}}\sigma_1^{\star}}\omega_{\min}+\frac{\delta}{\sqrt{\kappa}}\frac{\omega_{\min}^2}{\sqrt{np^2}\sigma_1^{\star 2}}\\
&\asymp \delta\left(\frac{1}{\kappa^2}\frac{\omega_{\min}}{\sigma_r^{\star}}\frac{1}{\sqrt{np^2}}+\frac{1}{\kappa}\frac{\omega_{\min}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}}+\frac{1}{\kappa^{3/2}}\frac{\omega_{\min}^2}{\sigma_r^{\star 2}}\sqrt{\frac{d}{np^2}}\right)\sqrt{\frac{1}{d}}.
\end{aligned} \tag{H.78}$$

In view of the fact that

$$\frac{\omega_{\max}}{p\sigma_r^{\star}}\sqrt{\frac{1}{n}}\lesssim\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}+\frac{1}{\sqrt{ndp}},$$

the above condition (H.78) can already be guaranteed by (H.77). We then conclude that

$$\gamma_4\asymp\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}\lesssim\zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)\asymp\gamma_3,$$

and we have already bounded γ_3 above.

As a consequence, we have demonstrated that with probability exceeding $1-O((n+d)^{-10})$,

$$\left(\widetilde{v}_{i,j}\right)^{-1/2}\zeta_{i,j}\lesssim\left(\widetilde{v}_{i,j}\right)^{-1/2}\left(\gamma_{1}+\gamma_{2}+\gamma_{3}+\gamma_{4}\right)\lesssim\delta$$

holds as long as $d\gtrsim\delta^{-2}\kappa^{7}\mu^{2}r^{2}\log^{4}(n+d)$, $np\gtrsim\delta^{-2}\kappa^{3}r^{2}\log^{4}(n+d)$, $ndp^{2}\gtrsim\delta^{-2}\kappa^{5}\mu^{2}r^{4}\log^{6}(n+d)$,

$$\frac{\omega_{\max}}{\sigma_{r}^{\star}}\sqrt{\frac{d}{np}}\lesssim\frac{\delta}{\kappa^{7/2}\mu^{1/2}r^{1/2}\log^{3/2}\left(n+d\right)},\qquad\frac{\omega_{\max}^{2}}{p\sigma_{r}^{\star2}}\sqrt{\frac{d}{n}}\lesssim\frac{\delta}{\kappa^{3}\mu^{1/2}r\log^{5/2}\left(n+d\right)},$$

$$\begin{aligned}&\min\left\{\left\Vert \boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2},\left\Vert \boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\Vert _{2}\right\}\\&\quad\gtrsim\delta^{-1}\left[\frac{\kappa^{9/2}\mu^{5/2}r^{2}\log^{9/2}\left(n+d\right)}{\sqrt{ndp^{2}}}+\frac{\kappa^{4}\mu^{3/2}r\log^{5/2}\left(n+d\right)}{\sqrt{np}}+\frac{\kappa^{4}\mu^{2}r^{3/2}\log^{3}\left(n+d\right)}{\sqrt{d}}\right]\sqrt{\frac{r}{d}}\sigma_{1}^{\star},\end{aligned}$$

and

$$\left\Vert \boldsymbol{U}_{i,\cdot}^{\star}\right\Vert _{2}+\left\Vert \boldsymbol{U}_{j,\cdot}^{\star}\right\Vert _{2}\gtrsim\delta^{-1}\kappa r\log^{2}\left(n+d\right)\left[\frac{\omega_{\max}}{\sigma_{r}^{\star}}\sqrt{\frac{d}{np}}+\frac{\omega_{\max}^{2}}{p\sigma_{r}^{\star2}}\sqrt{\frac{d}{n}}+\frac{\kappa\mu r\log\left(n+d\right)}{\sqrt{ndp^{2}}}\right]\sqrt{\frac{1}{d}}.$$

Taking $\delta\asymp\log^{-1/2}(n+d)$ in the above bounds directly establishes the advertised result.

H.4 Auxiliary lemmas for Theorem 18

H.4.1 Proof of Lemma 49

In the sequel, we shall only consider the case when $i\neq j$; the analysis for $i=j$ is similar and in fact simpler, and hence we omit it here for brevity. Let us denote

$$\begin{aligned}v_{i,j}^{\star}=&\underbrace{\frac{2-p}{np}S_{i,i}^{\star}S_{j,j}^{\star}}_{=:\alpha_{1}}+\underbrace{\frac{4-3p}{np}S_{i,j}^{\star2}}_{=:\alpha_{2}}+\underbrace{\frac{1}{np}\left(\omega_{i}^{\star2}S_{j,j}^{\star}+\omega_{j}^{\star2}S_{i,i}^{\star}\right)}_{=:\alpha_{3}}\\&+\underbrace{\frac{1}{np^{2}}\sum_{k=1}^{d}\left[\omega_{i}^{\star2}+\left(1-p\right)S_{i,i}^{\star}\right]\left[\omega_{k}^{\star2}+\left(1-p\right)S_{k,k}^{\star}\right]\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^{2}}_{=:\alpha_{4}}+\underbrace{\frac{2\left(1-p\right)^{2}}{np^{2}}\sum_{k=1}^{d}S_{i,k}^{\star2}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^{2}}_{=:\alpha_{5}}\\&+\underbrace{\frac{1}{np^{2}}\sum_{k=1}^{d}\left[\omega_{j}^{\star2}+\left(1-p\right)S_{j,j}^{\star}\right]\left[\omega_{k}^{\star2}+\left(1-p\right)S_{k,k}^{\star}\right]\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{i,\cdot}^{\star\top}\right)^{2}}_{=:\alpha_{6}}+\underbrace{\frac{2\left(1-p\right)^{2}}{np^{2}}\sum_{k=1}^{d}S_{j,k}^{\star2}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{i,\cdot}^{\star\top}\right)^{2}}_{=:\alpha_{7}}\end{aligned}$$

and

$$\begin{aligned}v_{i,j}=&\underbrace{\frac{2-p}{np}S_{i,i}S_{j,j}}_{=:\beta_{1}}+\underbrace{\frac{4-3p}{np}S_{i,j}^{2}}_{=:\beta_{2}}+\underbrace{\frac{1}{np}\left(\omega_{i}^{2}S_{j,j}+\omega_{j}^{2}S_{i,i}\right)}_{=:\beta_{3}}\\&+\underbrace{\frac{1}{np^{2}}\sum_{k=1}^{d}\left[\omega_{i}^{2}+\left(1-p\right)S_{i,i}\right]\left[\omega_{k}^{2}+\left(1-p\right)S_{k,k}\right]\left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\top}\right)^{2}}_{=:\beta_{4}}+\underbrace{\frac{2\left(1-p\right)^{2}}{np^{2}}\sum_{k=1}^{d}S_{i,k}^{2}\left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\top}\right)^{2}}_{=:\beta_{5}}\\&+\underbrace{\frac{1}{np^{2}}\sum_{k=1}^{d}\left[\omega_{j}^{2}+\left(1-p\right)S_{j,j}\right]\left[\omega_{k}^{2}+\left(1-p\right)S_{k,k}\right]\left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{i,\cdot}^{\top}\right)^{2}}_{=:\beta_{6}}+\underbrace{\frac{2\left(1-p\right)^{2}}{np^{2}}\sum_{k=1}^{d}S_{j,k}^{2}\left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{i,\cdot}^{\top}\right)^{2}}_{=:\beta_{7}}.\end{aligned}$$

It follows from Lemma 45 that

$$v_{i,j}^{\star} \gtrsim \frac{1}{\min\left\{ndp^{2}\kappa, np\right\}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}+\left(\frac{\omega_{\min}^{2}\sigma_{r}^{\star2}}{\min\left\{ndp^{2}, np\right\}}+\frac{\omega_{\min}^{4}}{np^{2}}\right)\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}^{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}^{2}\right).$$

In view of the bound on γ_2 in Step 3 of Appendix H.3.3 as well as (H.76), we know that for any $\varepsilon \in (0,1)$, one has

$$\zeta_{\mathsf{2nd}}\sigma_{1}^{\star2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\right)\lesssim\varepsilon\widetilde{v}_{i,j}^{1/2}\lesssim\varepsilon\left(v_{i,j}^{\star}\right)^{1/2} \tag{H.79}$$

and

$$\zeta_{\mathsf{2nd}}\lesssim\frac{\varepsilon}{\sqrt{\kappa}}\frac{1}{\sqrt{\min\left\{ndp^{2}\kappa, np\right\}}\sigma_{1}^{\star}}\left(\min\left\{\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2},\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\right\}+\omega_{\min}\right)+\frac{\varepsilon}{\sqrt{\kappa}}\frac{\omega_{\min}^{2}}{\sqrt{np^{2}}\sigma_{1}^{\star2}}, \tag{H.80}$$

provided that the following conditions hold: $d \gtrsim \varepsilon^{-2}\kappa^{7}\mu^{2}r^{2}\log^{4}(n+d)$,

$$\frac{\omega_{\max}}{\sigma_{r}^{\star}}\sqrt{\frac{d}{np}}\lesssim\frac{\varepsilon}{\kappa^{7/2}\mu^{1/2}r^{1/2}\log^{3/2}\left(n+d\right)},\qquad\frac{\omega_{\max}^{2}}{p\sigma_{r}^{\star2}}\sqrt{\frac{d}{n}}\lesssim\frac{\varepsilon}{\kappa^{3}\mu^{1/2}r^{1/2}\log^{5/2}\left(n+d\right)},\qquad\text{and}$$

$$\min\left\{\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2},\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\right\}\gtrsim\varepsilon^{-1}\left[\frac{\kappa^{9/2}\mu^{5/2}r^{2}\log^{9/2}\left(n+d\right)}{\sqrt{ndp^{2}}}+\frac{\kappa^{4}\mu^{3/2}r\log^{5/2}\left(n+d\right)}{\sqrt{np}}+\frac{\kappa^{4}\mu^{2}r^{3/2}\log^{3}\left(n+d\right)}{\sqrt{d}}\right]\sqrt{\frac{r}{d}}\sigma_{1}^{\star}.$$

In view of the bounds on γ_3 and γ_4 in Step 3 of Appendix H.3.3, we can see that

$$\theta^{2}\omega_{\max}^{2}+\zeta_{\mathsf{2nd}}^{2}\sigma_{1}^{\star2}\lesssim\varepsilon\widetilde{v}_{i,j}^{1/2}\lesssim\varepsilon\left(v_{i,j}^{\star}\right)^{1/2}, \tag{H.81}$$

provided that

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\gtrsim\varepsilon^{-1}\kappa r\log^{2}\left(n+d\right)\left[\frac{\omega_{\max}}{\sigma_{r}^{\star}}\sqrt{\frac{d}{np}}+\frac{\omega_{\max}^{2}}{p\sigma_{r}^{\star2}}\sqrt{\frac{d}{n}}+\frac{\kappa\mu r\log\left(n+d\right)}{\sqrt{ndp^{2}}}\right]\sqrt{\frac{1}{d}}.$$

These basic facts will be very useful for us to control the difference between $v_{i,j}^{\star}$ and $v_{i,j}$, towards which we shall bound $\alpha_i-\beta_i$, $1 \le i \le 7$, separately.

Step 1: bounding $|\alpha_1-\beta_1|$. Recall from Lemma 41 that

$$\left|S_{i,i}-S_{i,i}^{\star}\right|\lesssim\left(\theta+\sqrt{\frac{\kappa^{3}r\log\left(n+d\right)}{n}}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}+\theta^{2}\omega_{\max}^{2}+\left(\theta\omega_{\max}+\zeta_{\mathsf{2nd}}\sigma_{1}^{\star}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\zeta_{\mathsf{2nd}}^{2}\sigma_{1}^{\star2}, \tag{H.82a}$$

$$\left|S_{j,j}-S_{j,j}^{\star}\right|\lesssim\left(\theta+\sqrt{\frac{\kappa^{3}r\log\left(n+d\right)}{n}}\right)\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}+\theta^{2}\omega_{\max}^{2}+\left(\theta\omega_{\max}+\zeta_{\mathsf{2nd}}\sigma_{1}^{\star}\right)\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\zeta_{\mathsf{2nd}}^{2}\sigma_{1}^{\star2}. \tag{H.82b}$$

We proceed with the following elementary inequality:

$$\begin{aligned}\left|\alpha_{1}-\beta_{1}\right| &\lesssim\frac{1}{np}\left|S_{i,i}S_{j,j}-S_{i,i}^{\star}S_{j,j}^{\star}\right|\lesssim\frac{1}{np}S_{i.i}\left|S_{j,j}-S_{j,j}^{\star}\right|+\frac{1}{np}S_{j,j}^{\star}\left|S_{i,i}-S_{i,i}^{\star}\right|\\ &\lesssim\underbrace{\frac{1}{np}S_{i.i}^{\star}\left|S_{j,j}-S_{j,j}^{\star}\right|+\frac{1}{np}S_{j,j}^{\star}\left|S_{i,i}-S_{i,i}^{\star}\right|}_{=:\gamma_{1,1}}+\underbrace{\frac{1}{np}\left|S_{i,i}-S_{i,i}^{\star}\right|\left|S_{j,j}-S_{j,j}^{\star}\right|}_{=:\gamma_{1,2}}.\end{aligned}$$

- Regarding $\gamma_{1,1}$, it is seen that

$$\gamma_{1,1}\lesssim\underbrace{\frac{1}{np}\left(\theta+\sqrt{\frac{\kappa^{3}r\log\left(n+d\right)}{n}}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}}_{=:\gamma_{1,1,1}}+\underbrace{\frac{1}{np}\theta^{2}\omega_{\max}^{2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\right)}_{=:\gamma_{1,1,2}}$$

$$
+\underbrace{\frac{1}{np}\theta\omega_{\max}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\right)}_{=:\gamma_{1,1,3}}+\underbrace{\frac{1}{np}\zeta_{\mathsf{2nd}}^{2}\sigma_{1}^{\star2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\right)}_{=:\gamma_{1,1,4}}
$$

$$
+\underbrace{\frac{1}{np}\zeta_{\mathsf{2nd}}\sigma_{1}^{\star}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\right)}_{=:\gamma_{1,1,5}}\lesssim\delta v_{i,j}^{\star},
$$

where the last relation holds true since

$$
\begin{aligned}
\gamma_{1,1,1} &\lesssim \frac{\delta}{np}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\lesssim\delta v_{i,j}^{\star},\\
\gamma_{1,1,2} &\lesssim \delta\frac{\omega_{\min}^{2}\sigma_{r}^{\star2}}{np}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}^{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}^{2}\right)\lesssim\delta v_{i,j}^{\star},\\
\gamma_{1,1,3} &\lesssim \delta\frac{\omega_{\min}\sigma_{r}^{\star}}{np}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\right)\\
&\overset{\text{(i)}}{\lesssim} \delta\frac{1}{np}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}+\delta\frac{\omega_{\min}^{2}\sigma_{r}^{\star2}}{np}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}^{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}^{2}\right)\lesssim\delta v_{i,j}^{\star},\\
\gamma_{1,1,4} &\lesssim \frac{1}{np}\zeta_{\mathsf{2nd}}^{2}\sigma_{1}^{\star4}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}^{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}^{2}\right)\overset{\text{(ii)}}{\lesssim}\frac{\varepsilon^{2}}{np}v_{i,j}^{\star}\lesssim\delta v_{i,j}^{\star},\\
\gamma_{1,1,5} &\lesssim \frac{1}{np}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\cdot\zeta_{\mathsf{2nd}}\sigma_{1}^{\star2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\right)\overset{\text{(iii)}}{\lesssim}\frac{\varepsilon}{\sqrt{np}}v_{i,j}^{\star}\lesssim\delta v_{i,j}^{\star},
\end{aligned}
$$

provided that $\theta\lesssim\delta/\sqrt{\kappa}$, $n\gtrsim\delta^{-2}\kappa^{3}r\log(n+d)$, $np\gtrsim\delta^{-2}$ and $\varepsilon\lesssim1$. Here, the relation (i) invokes the AM-GM inequality, while (ii) and (iii) make use of (H.79).

- Regarding $\gamma_{1,2}$, we make the observation that

$$
\begin{aligned}
\gamma_{1,2} &\lesssim \left(\theta+\sqrt{\frac{\kappa^{3}r\log\left(n+d\right)}{n}}\right)\gamma_{1,1}+\frac{1}{np}\left[\theta^{2}\omega_{\max}^{2}+\left(\theta\omega_{\max}+\zeta_{\mathsf{2nd}}\sigma_{1}^{\star}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\zeta_{\mathsf{2nd}}^{2}\sigma_{1}^{\star2}\right]\\
&\qquad\cdot\left[\theta^{2}\omega_{\max}^{2}+\left(\theta\omega_{\max}+\zeta_{\mathsf{2nd}}\sigma_{1}^{\star}\right)\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\zeta_{\mathsf{2nd}}^{2}\sigma_{1}^{\star2}\right]\\
&\lesssim \delta v_{i,j}^{\star}+\underbrace{\frac{1}{np}\theta^{4}\omega_{\max}^{4}}_{=:\gamma_{1,2,1}}+\underbrace{\frac{1}{np}\left(\theta^{2}\omega_{\max}^{2}+\zeta_{\mathsf{2nd}}^{2}\sigma_{1}^{\star2}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}}_{=:\gamma_{1,2,2}}\\
&\quad+\underbrace{\frac{1}{np}\zeta_{\mathsf{2nd}}^{4}\sigma_{1}^{\star4}}_{=:\gamma_{1,2,3}}+\underbrace{\frac{1}{np}\theta^{3}\omega_{\max}^{3}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\right)}_{=:\gamma_{1,2,4}}+\underbrace{\frac{1}{np}\zeta_{\mathsf{2nd}}^{3}\sigma_{1}^{\star3}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\right)}_{=:\gamma_{1,2,5}}\\
&\lesssim \delta v_{i,j}^{\star}.
\end{aligned}
$$

Here, the penultimate relation follows from the previous bound on $\gamma_{1,1}$ as well as the assumptions that $\theta\lesssim1$ and $n\gtrsim\kappa^{3}r\log(n+d)$; the last line is valid since

$$
\begin{aligned}
\gamma_{1,2,1} &\overset{\text{(i)}}{\lesssim} \frac{\varepsilon^{2}}{np}v_{i,j}^{\star}\leq\delta v_{i,j}^{\star},\\
\gamma_{1,2,2} &\overset{\text{(ii)}}{\lesssim} \frac{1}{np}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}\varepsilon\left(v_{i,j}^{\star}\right)^{1/2}\lesssim\frac{\varepsilon}{\sqrt{np}}v_{i,j}^{\star}\lesssim\delta v_{i,j}^{\star},\\
\gamma_{1,2,3} &\overset{\text{(ii)}}{\lesssim} \frac{\varepsilon^{2}}{np}v_{i,j}^{\star}\lesssim\delta v_{i,j}^{\star},\\
\gamma_{1,2,4} &\lesssim \sqrt{\frac{\kappa}{np}}\theta\cdot\theta^{2}\omega_{\max}^{2}\cdot\frac{1}{\sqrt{np}}\omega_{\max}\sigma_{r}^{\star}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\right)\overset{\text{(iii)}}{\lesssim}\sqrt{\frac{\kappa}{np}}\theta\varepsilon v_{i,j}^{\star}\lesssim\delta v_{i,j}^{\star},
\end{aligned}
$$

$$\gamma_{1,2,5} \lesssim \frac{1}{np} \cdot \zeta_{\mathsf{2nd}}^2 \sigma_1^{\star 2} \cdot \zeta_{\mathsf{2nd}} \sigma_1^{\star 2} \left(\left\| \boldsymbol{U}_{i,\cdot}^\star \right\|_2 + \left\| \boldsymbol{U}_{j,\cdot}^\star \right\|_2 \right) \overset{\text{(iv)}}{\lesssim} \frac{\varepsilon^2}{np} v_{i,j}^\star \lesssim \delta v_{i,j}^\star,$$

provided that $np \gtrsim \delta^{-1}$, $\varepsilon \lesssim 1$ and $\theta \lesssim \delta/\sqrt{\kappa}$. Here, the inequalities (i)-(iv) rely on (H.79) and (H.81).

Taking the above bounds on $\gamma_{1,1}$ and $\gamma_{1,2}$ collectively, we arrive at

$$|\alpha_1 - \beta_1| \lesssim |\gamma_{1,1}| + |\gamma_{1,2}| \lesssim \delta v_{i,j}^\star,$$

provided that $\theta \lesssim \delta/\sqrt{\kappa}$, $n \gtrsim \delta^{-2} \kappa^3 r \log(n+d)$, $\varepsilon \lesssim 1$, and $np \gtrsim \delta^{-2}$.

Step 2: bounding $|\alpha_2 - \beta_2|$. It comes from Lemma 41 that

$$\left| S_{i,j} - S_{i,j}^\star \right| \lesssim \left(\theta + \sqrt{\frac{\kappa^3 r \log(n+d)}{n}} \right) \left\| \boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \left\| \boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 + \theta^2 \omega_{\max}^2 + (\theta \omega_{\max} + \zeta_{\mathsf{2nd}} \sigma_1^\star) \left(\left\| \boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 + \left\| \boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \right) + \zeta_{\mathsf{2nd}}^2 \sigma_1^{\star 2}.$$

With this in place, we shall proceed to decompose $|\alpha_2 - \beta_2|$ as follows:

$$|\alpha_2 - \beta_2| \lesssim \frac{1}{np} \left| S_{i,j}^{\star 2} - S_{i,j}^2 \right| \lesssim \underbrace{\frac{1}{np} S_{i,j}^\star \left| S_{i,j} - S_{i,j}^\star \right|}_{=:\gamma_{2,1}} + \underbrace{\frac{1}{np} \left| S_{i,j} - S_{i,j}^\star \right|^2}_{=:\gamma_{2,2}}.$$

- With regards to $\gamma_{2,1}$, it is observed that

$$\begin{aligned}
\gamma_{2,1} \lesssim\; & \underbrace{\frac{1}{np} \left(\theta + \sqrt{\frac{\kappa^3 r \log(n+d)}{n}} \right) \left\| \boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2^2 \left\| \boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2^2}_{=:\gamma_{2,1,1}} + \underbrace{\frac{1}{np} \theta^2 \omega_{\max}^2 \left\| \boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \left\| \boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2}_{=:\gamma_{2,1,2}} \\
& + \underbrace{\frac{1}{np} \theta \omega_{\max} \left\| \boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \left\| \boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \left(\left\| \boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 + \left\| \boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \right)}_{=:\gamma_{2,1,3}} \\
& + \underbrace{\frac{1}{np} \zeta_{\mathsf{2nd}} \sigma_1^\star \left\| \boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \left\| \boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \left(\left\| \boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 + \left\| \boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \right)}_{=:\gamma_{2,1,4}} \\
& + \underbrace{\frac{1}{np} \zeta_{\mathsf{2nd}}^2 \sigma_1^{\star 2} \left\| \boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \left\| \boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2}_{=:\gamma_{2,1,5}} \lesssim \delta v_{i,j}^\star.
\end{aligned}$$

Here, the last relation holds due to the following bounds

$$\begin{aligned}
\gamma_{2,1,1} &\lesssim \frac{\delta}{np} \left\| \boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2^2 \left\| \boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2^2 \lesssim \delta v_{i,j}^\star, \\
\gamma_{2,1,2} &\lesssim \frac{1}{\sqrt{np}} \cdot \theta^2 \omega_{\max}^2 \cdot \frac{1}{\sqrt{np}} \left\| \boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \left\| \boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \overset{\text{(i)}}{\lesssim} \frac{\varepsilon}{\sqrt{np}} v_{i,j}^\star \lesssim \delta v_{i,j}^\star, \\
\gamma_{2,1,3} &\lesssim \sqrt{\kappa} \theta \cdot \frac{1}{\sqrt{np}} \left\| \boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \left\| \boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \cdot \frac{1}{\sqrt{np}} \omega_{\max} \sigma_r^\star \left(\left\| \boldsymbol{U}_{i,\cdot}^\star \right\|_2 + \left\| \boldsymbol{U}_{j,\cdot}^\star \right\|_2 \right) \\
&\lesssim \sqrt{\kappa} \theta v_{i,j}^\star \lesssim \delta v_{i,j}^\star, \\
\gamma_{2,1,4} &\lesssim \frac{1}{\sqrt{np}} \cdot \frac{1}{\sqrt{np}} \left\| \boldsymbol{U}_{i,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \left\| \boldsymbol{U}_{j,\cdot}^\star \boldsymbol{\Sigma}^\star \right\|_2 \cdot \zeta_{\mathsf{2nd}} \sigma_1^{\star 2} \left(\left\| \boldsymbol{U}_{i,\cdot}^\star \right\|_2 + \left\| \boldsymbol{U}_{j,\cdot}^\star \right\|_2 \right) \\
&\overset{\text{(ii)}}{\lesssim} \frac{\varepsilon}{\sqrt{np}} v_{i,j}^\star \lesssim \delta v_{i,j}^\star,
\end{aligned}$$

$$\gamma_{2,1,5} \lesssim \frac{1}{np} \left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 \left\| \boldsymbol{U}_{j,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 \cdot \zeta_{\mathsf{2nd}}^2 \sigma_1^{\star 2} \overset{\text{(iii)}}{\lesssim} \frac{\varepsilon}{\sqrt{np}} v_{i,j}^{\star} \lesssim \delta v_{i,j}^{\star},$$

with the proviso that $\theta \lesssim \delta/\sqrt{\kappa}$, $n \gtrsim \delta^{-2}\kappa^3 r \log(n+d)$, $\varepsilon \lesssim 1$ and $np \gtrsim \delta^{-2}$. Note that (i)-(iii) follow from (H.79) and (H.81).

- Regarding $\gamma_{2,2}$, we obtain

$$\begin{aligned}
\gamma_{2,2} &\lesssim \underbrace{\frac{1}{np}\left(\theta^2 + \frac{\kappa^3 r \log(n+d)}{n}\right) \left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2 \left\| \boldsymbol{U}_{j,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2}_{=:\gamma_{2,2,1}} + \underbrace{\frac{1}{np}\theta^4 \omega_{\max}^4}_{=:\gamma_{2,2,2}} + \underbrace{\frac{1}{np}\zeta_{\mathsf{2nd}}^4 \sigma_1^{\star 4}}_{=:\gamma_{2,2,3}} \\
&\quad + \underbrace{\frac{1}{np}\theta^2 \omega_{\max}^2 \left(\left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2 + \left\| \boldsymbol{U}_{j,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2\right)}_{=:\gamma_{2,2,4}} + \underbrace{\frac{1}{np}\zeta_{\mathsf{2nd}}^2 \sigma_1^{\star 2} \left(\left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2 + \left\| \boldsymbol{U}_{j,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2\right)}_{=:\gamma_{2,2,5}} \\
&\lesssim \delta v_{i,j}^{\star},
\end{aligned}$$

where the last inequality holds true since

$$\begin{aligned}
\gamma_{2,2,1} &\lesssim \left(\theta + \sqrt{\frac{\kappa^3 r \log(n+d)}{n}}\right) \gamma_{2,1,1} \lesssim \delta v_{i,j}^{\star}, \\
\gamma_{2,2,2} + \gamma_{2,2,3} &\overset{\text{(i)}}{\lesssim} \frac{\varepsilon^2}{np} v_{i,j}^{\star} \lesssim \delta v_{i,j}^{\star}, \\
\gamma_{2,2,4} &\lesssim \theta^2 \kappa \cdot \frac{1}{np}\omega_{\max}^2 \sigma_r^{\star 2} \left(\left\| \boldsymbol{U}_{i,\cdot}^{\star} \right\|_2^2 + \left\| \boldsymbol{U}_{j,\cdot}^{\star} \right\|_2^2\right) \lesssim \theta^2 \kappa v_{i,j}^{\star} \lesssim \delta v_{i,j}^{\star}, \\
\gamma_{2,2,5} &\lesssim \frac{1}{np}\zeta_{\mathsf{2nd}}^2 \sigma_1^{\star 4} \left(\left\| \boldsymbol{U}_{i,\cdot}^{\star} \right\|_2^2 + \left\| \boldsymbol{U}_{j,\cdot}^{\star} \right\|_2^2\right) \overset{\text{(ii)}}{\lesssim} \frac{\varepsilon^2}{np} v_{i,j}^{\star} \lesssim \delta v_{i,j}^{\star},
\end{aligned}$$

provided that $\theta \lesssim \delta/\sqrt{\kappa}$, $n \gtrsim \kappa^3 r \log(n+d)$, $\varepsilon \lesssim 1$ and $np \geq \delta^{-1}$. Here, (i) follows from (H.81), whereas (ii) follows from (H.79).

Taking the above bounds on $\gamma_{2,1}$ and $\gamma_{2,2}$ together yields

$$|\alpha_2 - \beta_2| \lesssim |\gamma_{2,1}| + |\gamma_{2,2}| \lesssim \delta v_{i,j}^{\star},$$

as long as $\theta \lesssim \delta/\sqrt{\kappa}$, $n \gtrsim \delta^{-2}\kappa^3 r \log(n+d)$, $\varepsilon \lesssim 1$ and $np \gtrsim \delta^{-2}$.

Step 3: bounding $|\alpha_3 - \beta_3|$. For each $l \in [d]$, let us first define UB_l as follows

$$\mathsf{UB}_l := \left(\theta + \sqrt{\frac{\kappa^3 r \log(n+d)}{n}}\right) \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2 + \theta^2 \omega_{\max}^2 + \left(\theta \omega_{\max} + \zeta_{\mathsf{2nd}} \sigma_1^{\star}\right) \left\| \boldsymbol{U}_{l,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2 + \zeta_{\mathsf{2nd}}^2 \sigma_1^{\star 2}.$$

According to Lemma 41, we can obtain

$$\left| S_{l,l} - S_{l,l}^{\star} \right| \lesssim \mathsf{UB}_l \qquad \forall\, l \in [d]\,.$$

Lemma 41 also tells us that

$$\left| \omega_i^2 - \omega_i^{\star 2} \right| \lesssim \sqrt{\frac{\log^2(n+d)}{np}} \omega_i^{\star 2} + \mathsf{UB}_i.$$

With these basic bounds in mind, we proceed with the following decomposition:

$$\begin{aligned}
\frac{1}{np}\left| \omega_i^2 S_{j,j} - \omega_i^{\star 2} S_{j,j}^{\star} \right| &\lesssim \frac{1}{np} S_{j,j} \left| \omega_i^2 - \omega_i^{\star 2} \right| + \frac{1}{np}\omega_i^{\star 2} \left| S_{j,j} - S_{j,j}^{\star} \right| \\
&\lesssim \underbrace{\frac{1}{np} S_{j,j}^{\star} \left| \omega_i^2 - \omega_i^{\star 2} \right|}_{=:\gamma_{3,1}} + \underbrace{\frac{1}{np}\omega_i^{\star 2} \left| S_{j,j} - S_{j,j}^{\star} \right|}_{=:\gamma_{3,2}} + \underbrace{\frac{1}{np}\left| S_{j,j} - S_{j,j}^{\star} \right| \left| \omega_i^2 - \omega_i^{\star 2} \right|}_{=:\gamma_{3,3}},
\end{aligned}$$

leaving us with three terms to cope with.

- Regarding $\gamma_{3,1}$, we can upper bound

$$\gamma_{3,1} \lesssim \underbrace{\frac{1}{np}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\sqrt{\frac{\log^{2}(n+d)}{np}}\omega_{i}^{\star2}}_{=:\gamma_{3,1,1}}+\underbrace{\frac{1}{np}S_{j,j}^{\star}\mathsf{UB}_{i}}_{=:\gamma_{3,1,2}} \lesssim \delta v_{i,j}^{\star}.$$

Here, the last relation holds since (i) the first term

$$\gamma_{3,1,1} \lesssim \sqrt{\frac{\kappa\log^{2}(n+d)}{n^{3}p^{3}}}\omega_{i}^{\star2}\sigma_{r}^{\star2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}^{2} \lesssim \sqrt{\frac{\kappa\log^{2}(n+d)}{np}}v_{i,j}^{\star} \lesssim \delta v_{i,j}^{\star}$$

provided that $np \gtrsim \delta^{-2}\kappa\log^{2}(n+d)$; and we can also demonstrate that the second term obeys $\gamma_{3,1,2} \lesssim \delta v_{i,j}^{\star}$, since it is easy seen that $\gamma_{3,1,2}$ admits the same upper bound as for $\gamma_{1,1}$.

- Regarding $\gamma_{3,2}$, it can be seen that

$$\begin{aligned}\gamma_{3,2} &\lesssim \underbrace{\frac{1}{np}\omega_{i}^{\star2}\left(\theta+\sqrt{\frac{\kappa^{3}r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}}_{=:\gamma_{3,2,1}}+\underbrace{\frac{1}{np}\omega_{i}^{\star2}\theta^{2}\omega_{\max}^{2}}_{=:\gamma_{3,2,2}}\\ &\quad+\underbrace{\frac{1}{np}\omega_{i}^{\star2}\theta\omega_{\max}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}}_{=:\gamma_{3,2,3}}+\underbrace{\frac{1}{np}\omega_{i}^{\star2}\zeta_{\mathsf{2nd}}\sigma_{1}^{\star}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}}_{=:\gamma_{3,2,4}}+\underbrace{\frac{1}{np}\omega_{i}^{\star2}\zeta_{\mathsf{2nd}}^{2}\sigma_{1}^{\star2}}_{=:\gamma_{3,2,5}}\\ &\lesssim \delta v_{i,j}^{\star},\end{aligned}$$

where the last line holds true since

$$\begin{aligned}\gamma_{3,2,1} &\lesssim \frac{\delta}{np}\omega_{i}^{\star2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2}^{2} \lesssim \delta v_{i,j}^{\star},\\ \gamma_{3,2,2}+\gamma_{3,2,3} &\lesssim \delta\frac{\omega_{\min}^{2}\sigma_{r}^{\star2}}{np}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}^{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}^{2}\right) \lesssim \delta v_{i,j}^{\star},\\ \gamma_{3,2,4} &\lesssim \frac{1}{np}\omega_{i}^{\star2}\cdot\zeta_{\mathsf{2nd}}\sigma_{1}^{\star2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\right) \overset{\text{(i)}}{\lesssim} \frac{\varepsilon}{np}\omega_{i}^{\star2}v_{i,j}^{\star1/2} \lesssim \delta\frac{\omega_{\min}\sigma_{r}^{\star}}{\sqrt{np}}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\right)v_{i,j}^{\star1/2} \lesssim \delta v_{i,j}^{\star},\\ \gamma_{3,2,5} &\lesssim \frac{1}{np}\omega_{i}^{\star2}\cdot\zeta_{\mathsf{2nd}}^{2}\sigma_{1}^{\star2} \overset{\text{(ii)}}{\lesssim} \frac{\varepsilon}{np}\omega_{i}^{\star2}v_{i,j}^{\star1/2} \lesssim \delta\frac{\omega_{\min}\sigma_{r}^{\star}}{\sqrt{np}}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}\right)v_{i,j}^{\star1/2} \lesssim \delta v_{i,j}^{\star},\end{aligned}$$

provided that $\theta \lesssim \delta$, $n \gtrsim \delta^{-2}\kappa^{3}r\log(n+d)$, $\varepsilon \lesssim 1$, $np \gtrsim \delta^{-2}\kappa$ and

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2} \gtrsim \delta^{-1}\kappa^{1/2}\theta\frac{\omega_{\max}}{\sigma_{r}^{\star}}.$$

Here, (i) follows from (H.79), while (ii) arises from (H.81).

- When it comes to $\gamma_{3,3}$, we obtain

$$\gamma_{3,3} \lesssim \underbrace{\frac{1}{np}\left|S_{j,j}-S_{j,j}^{\star}\right|\sqrt{\frac{\log^{2}(n+d)}{np}}\omega_{i}^{\star2}}_{=:\gamma_{3,3,1}}+\underbrace{\frac{1}{np}\left|S_{j,j}-S_{j,j}^{\star}\right|\mathsf{UB}_{i}}_{=:\gamma_{3,3,2}} \lesssim \delta v_{i,j}^{\star}.$$

The last relation holds since (i) the first term

$$\gamma_{3,3,1} \lesssim \sqrt{\frac{\log^{2}(n+d)}{np}}\gamma_{3,2} \lesssim \delta v_{i,j}^{\star}$$

provided that $np \gtrsim \log^{2}(n+d)$; and (ii) one can easily check that the second term $\gamma_{3,3,2}$ admits the same upper bound as for $\gamma_{1,2}$.

With the above results in hand, one can conclude that

$$\frac{1}{np}\left|\omega_i^2 S_{j,j} - \omega_i^{\star 2} S_{j,j}^{\star}\right| \lesssim \gamma_{3,1} + \gamma_{3,2} + \gamma_{3,3} \lesssim \delta v_{i,j}^{\star},$$

and similarly,

$$\frac{1}{np}\left|\omega_j^2 S_{i,i} - \omega_j^{\star 2} S_{i,i}^{\star}\right| \lesssim \delta v_{i,j}^{\star}.$$

These allow us to reach

$$|\alpha_3 - \beta_3| \lesssim \frac{1}{np}\left|\omega_i^2 S_{j,j} - \omega_i^{\star 2} S_{j,j}^{\star}\right| + \frac{1}{np}\left|\omega_j^2 S_{i,i} - \omega_j^{\star 2} S_{i,i}^{\star}\right| \lesssim \delta v_{i,j}^{\star},$$

provided that $np \gtrsim \delta^{-2}\kappa \log^2(n+d)$, $\theta \lesssim \delta$, $n \gtrsim \delta^{-2}\kappa^3 r \log(n+d)$, $\varepsilon \lesssim 1$, and

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \gtrsim \delta^{-1}\kappa^{1/2}\theta\frac{\omega_{\max}}{\sigma_r^{\star}}. \tag{H.83}$$

In view of (H.37), we know that

$$\begin{aligned}
\theta\frac{\omega_{\max}}{\sigma_r^{\star}} &\asymp \left(\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d\kappa\log^2(n+d)}{np}} + \frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{\kappa^2\mu r\log^3(n+d)}{np^2}} + \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d\kappa\log^2(n+d)}{n}}\right)\cdot\sqrt{\frac{r}{d}} \\
&\asymp \sqrt{\kappa\log^2(n+d)}\left(\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}} + \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} + \frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{\kappa\mu r\log(n+d)}{np^2}}\right)\cdot\sqrt{\frac{r}{d}} \\
&\asymp \sqrt{\kappa\log^2(n+d)}\left[\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}} + \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} + \frac{\kappa\mu r\log(n+d)}{\sqrt{ndp^2}}\right]\cdot\sqrt{\frac{r}{d}},
\end{aligned} \tag{H.84}$$

where we have used the AM-GM inequality in the last line. As a result, (H.83) is guaranteed by

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \gtrsim \delta^{-1}\kappa\log(n+d)\left[\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}} + \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} + \frac{\kappa\mu r\log(n+d)}{\sqrt{ndp^2}}\right]\cdot\sqrt{\frac{r}{d}}.$$

Step 4: bounding $|\alpha_4 - \beta_4|$ and $|\alpha_6 - \beta_6|$. For each $l \in [d]$, let us denote

$$\Delta_l \coloneqq \left|\omega_l^{\star 2} + (1-p)S_{l,l}^{\star 2} - \omega_l^2 - (1-p)S_{l,l}^2\right|.$$

Lemma 41 tells us that, for each $l \in [d]$,

$$\Delta_l \lesssim \sqrt{\frac{\log^2(n+d)}{np}}\omega_{\max}^2 + \zeta_{\mathsf{1st}}\frac{\sqrt{\kappa^2\mu r^2\log(n+d)}}{d} + \sqrt{\frac{\kappa^2\mu^2 r^3\log(n+d)}{nd^2}}\sigma_1^{\star 2}. \tag{H.85}$$

We also know that

$$\Delta_i \lesssim \sqrt{\frac{\log^2(n+d)}{np}}\omega_{\max}^2 + \mathsf{UB}_i \qquad \text{and} \qquad \Delta_j \lesssim \sqrt{\frac{\log^2(n+d)}{np}}\omega_{\max}^2 + \mathsf{UB}_j. \tag{H.86}$$

In addition, for each $l \in [d]$, it holds that

$$\omega_l^{\star 2} + (1-p)\,S_{l,l}^{\star} \le \omega_{\max}^2 + \left\|\boldsymbol{U}_{l,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 \lesssim \omega_{\max}^2 + \frac{\mu r}{d}\sigma_1^{\star 2}. \tag{H.87}$$

We also have the following bound

$$\left\|\boldsymbol{U}_{j,\cdot}\right\|_2 \le \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 + \left\|\left(\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star}\right)_{j,\cdot}\right\|_2 \overset{\text{(i)}}{\lesssim} \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 + \frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\left(\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_j^{\star}\right) + \zeta_{\mathsf{2nd}}$$

$$\overset{\text{(ii)}}{\lesssim} \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}+\frac{\theta}{\sqrt{\kappa}\sigma_{r}^{\star}}\omega_{j}^{\star}, \tag{H.88}$$

where (i) follows from Lemma 40, and (ii) holds provided that $\theta \lesssim 1$ and $\zeta_{\mathsf{2nd}} \lesssim \|\boldsymbol{U}_{j,\cdot}^{\star}\|_{2}$, which can be guaranteed by (H.59) as long as $\varepsilon \lesssim 1$, $ndp^2 \gtrsim 1$ and $np \gtrsim 1$. We can thus decompose

$$\begin{aligned}
\left|\alpha_{4}-\beta_{4}\right| \lesssim & \underbrace{\frac{1}{np^{2}}\left[\omega_{i}^{\star 2}+(1-p)\, S_{i,i}^{\star}\right] \sum_{k=1}^{d} \Delta_{k}\left(\boldsymbol{U}_{k,\cdot} \boldsymbol{U}_{j,\cdot}^{\top}\right)^{2}}_{=:\gamma_{4,1}}+\underbrace{\frac{\Delta_{i}}{np^{2}} \sum_{k=1}^{d}\left[\omega_{k}^{2}+(1-p)\, S_{k,k}\right]\left(\boldsymbol{U}_{k,\cdot} \boldsymbol{U}_{j,\cdot}^{\top}\right)^{2}}_{=:\gamma_{4,2}} \\
& +\underbrace{\frac{1}{np^{2}}\left[\omega_{i}^{\star 2}+(1-p)\, S_{i,i}^{\star}\right] \sum_{k=1}^{d}\left[\omega_{k}^{\star 2}+(1-p)\, S_{k,k}^{\star}\right]\left|\left(\boldsymbol{U}_{k,\cdot}^{\star} \boldsymbol{U}_{j,\cdot}^{\star \top}\right)^{2}-\left(\boldsymbol{U}_{k,\cdot} \boldsymbol{U}_{j,\cdot}^{\top}\right)^{2}\right|}_{=:\gamma_{4,3}}.
\end{aligned}$$

Step 4.1: bounding $\gamma_{4,1}$. We have learned from (H.88) that

$$\begin{aligned}
\gamma_{4,1} & \lesssim \frac{1}{np^{2}}\left[\omega_{i}^{\star 2}+(1-p)\, S_{i,i}^{\star}\right]\left\|\boldsymbol{U}_{j,\cdot}\right\|_{2}^{2} \max_{1 \leq k \leq d} \Delta_{k} \\
& \lesssim \underbrace{\frac{1}{np^{2}}\left[\omega_{i}^{\star 2}+(1-p)\, S_{i,i}^{\star}\right]\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}^{2} \max_{1 \leq k \leq d} \Delta_{k}}_{=:\gamma_{4,1,1}}+\underbrace{\frac{\theta \omega_{i}^{\star}}{np^{2} \sigma_{1}^{\star}}\left[\omega_{i}^{\star 2}+(1-p)\, S_{i,i}^{\star}\right] \max_{1 \leq k \leq d} \Delta_{k}}_{=:\gamma_{4,1,2}}.
\end{aligned}$$

Regarding $\gamma_{4,1,1}$, we can derive

$$\begin{aligned}
\gamma_{4,1,1} \lesssim & \underbrace{\frac{1}{np^{2}} \omega_{\max}^{4}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}^{2} \sqrt{\frac{\log^{2}(n+d)}{np}}}_{=:\gamma_{4,1,1,1}}+\underbrace{\frac{\sqrt{\kappa^{2} \mu r^{2} \log (n+d)}}{ndp^{2}} \omega_{\max}^{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}^{2} \zeta_{\mathsf{1st}}}_{=:\gamma_{4,1,1,2}} \\
& +\underbrace{\frac{1}{np^{2}} \omega_{\max}^{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}^{2} \sqrt{\frac{\kappa^{2} \mu^{2} r^{3} \log (n+d)}{nd^{2}}} \sigma_{1}^{\star 2}}_{=:\gamma_{4,1,1,3}}+\underbrace{\frac{1}{np^{2}}\left\|\boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}^{2} \sqrt{\frac{\log^{2}(n+d)}{np}} \omega_{\max}^{2}}_{=:\gamma_{4,1,1,4}} \\
& +\underbrace{\frac{\sqrt{\kappa^{2} \mu r^{2} \log (n+d)}}{ndp^{2}}\left\|\boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}^{2} \zeta_{\mathsf{1st}}}_{=:\gamma_{4,1,1,5}}+\underbrace{\frac{1}{np^{2}}\left\|\boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}^{2} \sqrt{\frac{\kappa^{2} \mu^{2} r^{3} \log (n+d)}{nd^{2}}} \sigma_{1}^{\star 2}}_{=:\gamma_{4,1,1,6}} \\
\lesssim & \delta v_{i,j}^{\star}.
\end{aligned}$$

Here, the last line holds since

$$\begin{aligned}
\gamma_{4,1,1,1} & \lesssim \delta \frac{\omega_{\min}^{4}}{np^{2}}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}^{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}^{2}\right) \lesssim \delta v_{i,j}^{\star}, \\
\gamma_{4,1,1,2}+\gamma_{4,1,1,3}+\gamma_{4,1,1,4} & \lesssim \delta \frac{\omega_{\min}^{2} \sigma_{r}^{\star 2}}{ndp^{2}}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_{2}^{2}+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_{2}^{2}\right) \lesssim \delta v_{i,j}^{\star}, \\
\gamma_{4,1,1,5}+\gamma_{4,1,1,6} & \lesssim \delta \frac{1}{ndp^{2} \kappa}\left\|\boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star}\right\|_{2}^{2}\left\|\boldsymbol{U}_{j,\cdot}^{\star} \boldsymbol{\Sigma}^{\star}\right\|_{2}^{2} \lesssim \delta v_{i,j}^{\star},
\end{aligned}$$

provided that $np \gtrsim \delta^{-2}\kappa^{2}\mu^{2}r^{2}\log^{2}(n+d)$, $n \gtrsim \delta^{-2}\kappa^{4}\mu^{2}r^{3}\log(n+d)$ and

$$\frac{\zeta_{\mathsf{1st}}}{\sigma_{r}^{\star 2}} \lesssim \frac{\delta}{\sqrt{\kappa^{2} \mu r^{2} \log (n+d)}}.$$

Regarding $\gamma_{4,1,2}$, we can upper bound

$$\gamma_{4,1,2} \lesssim \underbrace{\frac{1}{np^{2}} \frac{\omega_{\max}^{6}}{\sigma_{r}^{\star 2}} \frac{\theta^{2}}{\kappa} \sqrt{\frac{\log^{2}(n+d)}{np}}}_{=:\gamma_{4,1,2,1}}+\underbrace{\frac{\sqrt{\kappa^{2} \mu r^{2} \log (n+d)}}{ndp^{2}} \frac{\omega_{\max}^{4}}{\sigma_{r}^{\star 2}} \frac{\theta^{2}}{\kappa} \zeta_{\mathsf{1st}}}_{=:\gamma_{4,1,2,2}}$$

$$+\underbrace{\frac{1}{np^2}\omega_{\max}^4\theta^2\sqrt{\frac{\kappa^2\mu^2r^3\log(n+d)}{nd^2}}}_{=:\gamma_{4,1,2,3}}+\underbrace{\frac{1}{np^2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\frac{\theta^2}{\kappa}\frac{\omega_{\max}^4}{\sigma_r^{\star2}}\sqrt{\frac{\log^2(n+d)}{np}}}_{=:\gamma_{4,1,2,4}}$$

$$+\underbrace{\frac{\sqrt{\kappa^2\mu r^2\log(n+d)}}{ndp^2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\frac{\theta^2}{\kappa}\frac{\omega_{\max}^2}{\sigma_r^{\star2}}\zeta_{\mathsf{1st}}}_{=:\gamma_{4,1,2,5}}+\underbrace{\frac{1}{np^2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\theta^2\omega_{\max}^2\sqrt{\frac{\kappa^2\mu^2r^3\log(n+d)}{nd^2}}}_{=:\gamma_{4,1,2,6}}.$$

There are six terms on the right-hand side of the above inequality, which we shall control separately.

- With regards to $\gamma_{4,1,2,1}$, we observe that

$$\gamma_{4,1,2,1}\asymp\underbrace{\frac{\omega_{\max}^6}{\sigma_r^{\star2}}\frac{r\log^3(n+d)}{n^{2.5}p^{3.5}}}_{=:\gamma_{4,1,2,1,1}}+\underbrace{\frac{\omega_{\max}^6}{\sigma_r^{\star2}}\frac{\kappa\mu r^2\log^4(n+d)}{n^{2.5}dp^{4.5}}}_{=:\gamma_{4,1,2,1,2}}+\underbrace{\frac{\omega_{\max}^8}{\sigma_r^{\star4}}\frac{r\log^3(n+d)}{n^{2.5}p^{4.5}}}_{=:\gamma_{4,1,2,1,3}}\lesssim\delta v_{i,j}^{\star}.$$

  Here, the last relation holds since

$$\gamma_{4,1,2,1,1}+\gamma_{4,1,2,1,3}\lesssim\delta\frac{\omega_{\min}^4}{np^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right)\lesssim\delta v_{i,j}^{\star},$$
$$\gamma_{4,1,2,1,2}\lesssim\delta\frac{\omega_{\min}^2\sigma_r^{\star2}}{ndp^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right)\lesssim\delta v_{i,j}^{\star},$$

  provided that

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\gtrsim\delta^{-1/2}\frac{\sqrt{\log^3(n+d)}}{n^{1/4}p^{1/4}}\left[\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}}+\frac{\omega_{\max}^2}{p\sigma_r^{\star2}}\sqrt{\frac{d}{n}}\right]\sqrt{\frac{r}{d}}.$$

- When it comes to $\gamma_{4,1,2,2}$, we obtain

$$\gamma_{4,1,2,2}\asymp\underbrace{\frac{\omega_{\max}^4}{\sigma_r^{\star2}}\frac{\kappa\mu^{1/2}r^2\log^{5/2}(n+d)}{n^2dp^3}\zeta_{\mathsf{1st}}}_{=:\gamma_{4,1,2,2,1}}+\underbrace{\frac{\omega_{\max}^4}{\sigma_r^{\star2}}\frac{\kappa^2\mu^{3/2}r^3\log^{7/2}(n+d)}{n^2d^2p^4}\zeta_{\mathsf{1st}}}_{=:\gamma_{4,1,2,2,2}}$$
$$+\underbrace{\frac{\omega_{\max}^6}{\sigma_r^{\star4}}\frac{\kappa\mu^{1/2}r^2\log^{5/2}(n+d)}{n^2dp^4}\zeta_{\mathsf{1st}}}_{=:\gamma_{4,1,2,2,3}}\lesssim\delta v_{i,j}^{\star}.$$

  Here, the last relation holds since

$$\gamma_{4,1,2,2,1}+\gamma_{4,1,2,2,3}\lesssim\delta\frac{\omega_{\min}^2\sigma_r^{\star2}}{ndp^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right)\lesssim\delta v_{i,j}^{\star},$$
$$\gamma_{4,1,2,2,2}\lesssim\delta\frac{\omega_{\min}^4}{np^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right)\lesssim\delta v_{i,j}^{\star},$$

  provided that

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\gtrsim\delta^{-1/2}\sqrt{\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star2}}}\left[\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d\kappa\mu^{1/2}r\log^{5/2}(n+d)}{np}}+\sqrt{\frac{\kappa^2\mu^{3/2}r^2\log^{7/2}(n+d)}{ndp^2}}\right.$$
$$\left.+\frac{\omega_{\max}^2}{p\sigma_r^{\star2}}\sqrt{\frac{d\kappa\mu^{1/2}r\log^{5/2}(n+d)}{n}}\right]\sqrt{\frac{r}{d}}.$$

- Regarding $\gamma_{4,1,2,3}$, we can obtain

$$\gamma_{4,1,2,3} \asymp \underbrace{\frac{\omega_{\max}^4}{n^{2.5}dp^3}\kappa^2\mu r^{5/2}\log^{5/2}(n+d)}_{=:\gamma_{4,1,2,3,1}} + \underbrace{\frac{\omega_{\max}^4}{n^{2.5}d^2p^4}\kappa^3\mu^2 r^{7/2}\log^{7/2}(n+d)}_{=:\gamma_{4,1,2,3,2}}$$
$$+ \underbrace{\frac{\omega_{\max}^6}{\sigma_r^{\star 2}}\frac{\kappa^2\mu r^{5/2}\log^{5/2}(n+d)}{n^{5/2}dp^4}}_{=:\gamma_{4,1,2,3,3}} \lesssim \delta v_{i,j}^{\star}.$$

Here, the last relation holds since

$$\gamma_{4,1,2,3,1} + \gamma_{4,1,2,3,3} \lesssim \delta\frac{\omega_{\min}^2\sigma_r^{\star 2}}{ndp^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right) \lesssim \delta v_{i,j}^{\star},$$
$$\gamma_{4,1,2,3,2} \lesssim \delta\frac{\omega_{\min}^4}{np^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right) \lesssim \delta v_{i,j}^{\star},$$

provided that

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \gtrsim \delta^{-1/2}\frac{\kappa^{1/2}\mu^{1/4}r^{1/4}}{n^{1/4}}\left[\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d\kappa\mu^{1/2}r\log^{5/2}(n+d)}{np}} + \sqrt{\frac{\kappa^2\mu^{3/2}r^2\log^{7/2}(n+d)}{ndp^2}}\right.$$
$$\left. + \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d\kappa\mu^{1/2}r\log^{5/2}(n+d)}{n}}\right]\sqrt{\frac{r}{d}}.$$

- For $\gamma_{4,1,2,4}$, $\gamma_{4,1,2,5}$ and $\gamma_{4,1,2,6}$, it is straightforward to check that

$$\gamma_{4,1,2,4} \lesssim \delta\frac{\omega_{\min}^4}{np^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right) \lesssim \delta v_{i,j}^{\star},$$
$$\gamma_{4,1,2,5} + \gamma_{4,1,2,6} \lesssim \delta\frac{\omega_{\min}^2\sigma_r^{\star 2}}{ndp^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right) \lesssim \delta v_{i,j}^{\star},$$

provided that $\theta \lesssim 1$, $np \gtrsim \delta^{-2}\log^2(n+d)$, $n \gtrsim \delta^{-2}\kappa^4\mu^2 r^3\log(n+d)$ and

$$\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}} \lesssim \frac{\delta}{\sqrt{\kappa^2\mu r^2\log(n+d)}}.$$

Taking the bounds on $\gamma_{4,1,2,1}$ to $\gamma_{4,1,2,6}$ collectively yields

$$\gamma_{4,1,2} \lesssim \gamma_{4,1,2,1} + \gamma_{4,1,2,2} + \gamma_{4,1,2,3} + \gamma_{4,1,2,4} + \gamma_{4,1,2,5} + \gamma_{4,1,2,6} \lesssim \delta v_{i,j}^{\star}.$$

Then we can combine the bounds on $\gamma_{4,1,1}$ and $\gamma_{4,1,2}$ to arrive at

$$\gamma_{4,1} \lesssim \gamma_{4,1,1} + \gamma_{4,1,2} \lesssim \delta v_{i,j}^{\star},$$

provided that $\theta \lesssim 1$, $np \gtrsim \delta^{-2}\kappa^2\mu^2 r^2\log^2(n+d)$, $n \gtrsim \delta^{-2}\kappa^4\mu^2 r^3\log(n+d)$ and

$$\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}} \lesssim \frac{\delta}{\sqrt{\kappa^2\mu r^2\log(n+d)}},$$
$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \gtrsim \delta^{-1/2}\frac{\sqrt{\log^3(n+d)}}{n^{1/4}p^{1/4}}\left[\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}} + \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}\right]\sqrt{\frac{r}{d}},$$

and

$$\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \gtrsim \delta^{-1/2}\left(\sqrt{\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}} + \frac{\kappa^{1/2}\mu^{1/4}r^{1/4}}{n^{1/4}}\right)\left[\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d\kappa\mu^{1/2}r\log^{5/2}(n+d)}{np}}\right.$$
$$\left. + \sqrt{\frac{\kappa^2\mu^{3/2}r^2\log^{7/2}(n+d)}{ndp^2}} + \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d\kappa\mu^{1/2}r\log^{5/2}(n+d)}{n}}\right]\sqrt{\frac{r}{d}}.$$

Step 4.2: bounding $\gamma_{4,2}$. In view of (H.86), (H.87) and (H.88), we can develop the following upper bound:

$$
\begin{aligned}
\gamma_{4,2} &\lesssim \frac{\Delta_i}{np^2}\max_{1\le k\le d}\left[\omega_k^2+(1-p)\,S_{k,k}\right]\|\boldsymbol{U}_{j,\cdot}\|_2^2 \\
&\lesssim \frac{1}{np^2}\left(\sqrt{\frac{\log^2(n+d)}{np}}\omega_{\max}^2+\mathsf{UB}_i\right)\left(\omega_{\max}^2+\frac{\mu r}{d}\sigma_1^{\star 2}\right)\left(\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2+\frac{\theta^2}{\kappa\sigma_r^{\star 2}}\omega_j^{\star 2}\right) \\
&\lesssim \delta v_{i,j}^{\star}+\frac{1}{np^2}\mathsf{UB}_i\left(\omega_{\max}^2+\frac{\mu r}{d}\sigma_1^{\star 2}\right)\left(\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2+\frac{\theta^2}{\kappa\sigma_r^{\star 2}}\omega_j^{\star 2}\right) \\
&\quad+\frac{1}{np^2}\sqrt{\frac{\log^2(n+d)}{np}}\omega_{\max}^2\frac{\mu r}{d}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2+\frac{\theta^2}{\kappa\sigma_r^{\star 2}}\omega_j^{\star 2}\right) \\
&\lesssim \delta v_{i,j}^{\star}+\underbrace{\frac{1}{np^2}\mathsf{UB}_i\omega_{\max}^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2}_{=:\gamma_{4,2,1}}+\underbrace{\frac{1}{np^2}\mathsf{UB}_i\omega_{\max}^4\frac{\theta^2}{\kappa\sigma_r^{\star 2}}}_{=:\gamma_{4,2,2}}+\underbrace{\frac{\mu r}{ndp^2}\mathsf{UB}_i\sigma_1^{\star 2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2}_{=:\gamma_{4,2,3}}+\underbrace{\frac{\mu r}{ndp^2}\mathsf{UB}_i\theta^2\omega_j^{\star 2}}_{=:\gamma_{4,2,4}} \\
&\quad+\underbrace{\frac{\mu r}{ndp^2}\sqrt{\frac{\log^2(n+d)}{np}}\omega_{\max}^2\sigma_1^{\star 2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2}_{=:\gamma_{4,2,5}}+\underbrace{\frac{\mu r}{ndp^2}\sqrt{\frac{\log^2(n+d)}{np}}\omega_{\max}^4\theta^2}_{=:\gamma_{4,2,6}},
\end{aligned}
$$

where the penultimate relation holds since

$$
\frac{1}{np^2}\sqrt{\frac{\log^2(n+d)}{np}}\omega_{\max}^4\left(\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2+\frac{\theta^2}{\kappa\sigma_r^{\star 2}}\omega_j^{\star 2}\right)\lesssim\gamma_{4,1}\lesssim\delta v_{i,j}^{\star}.
$$

In what follows, we shall bound the six terms from $\gamma_{4,2,1}$ to $\gamma_{4,2,6}$ separately.

- Regarding $\gamma_{4,2,1}$, we can derive

$$
\begin{aligned}
\gamma_{4,2,1} &\asymp \underbrace{\frac{1}{np^2}\left(\theta+\sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\omega_{\max}^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2}_{=:\gamma_{4,2,1,1}}+\underbrace{\frac{1}{np^2}\theta^2\omega_{\max}^4\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2}_{=:\gamma_{4,2,1,2}} \\
&\quad+\underbrace{\frac{1}{np^2}\theta\omega_{\max}^3\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2}_{=:\gamma_{4,2,1,3}}+\underbrace{\frac{1}{np^2}\zeta_{\mathsf{2nd}}\sigma_1^{\star}\omega_{\max}^2\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2}_{=:\gamma_{4,2,1,4}} \\
&\quad+\underbrace{\frac{1}{np^2}\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}\omega_{\max}^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2}_{=:\gamma_{4,2,1,5}}\lesssim\delta v_{i,j}^{\star}.
\end{aligned}
$$

  Here, the last relation holds since

$$
\begin{aligned}
\gamma_{4,2,1,1}+\gamma_{4,2,1,3}+\gamma_{4,2,1,4}+\gamma_{4,2,1,5} &\lesssim \delta\frac{\omega_{\min}^2\sigma_r^{\star 2}}{ndp^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right)\lesssim\delta v_{i,j}^{\star}, \\
\gamma_{4,2,1,2} &\lesssim \delta\frac{\omega_{\min}^4}{np^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right)\lesssim\delta v_{i,j}^{\star},
\end{aligned}
$$

  provided that $\theta\lesssim\delta/(\kappa\mu r)$, $n\gtrsim\delta^{-2}\kappa^5\mu^2r^3\log(n+d)$, $\zeta_{\mathsf{2nd}}\sqrt{d}\lesssim\delta/\sqrt{\kappa^2\mu r}$,

$$
\theta\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{d}\lesssim\frac{\delta}{\sqrt{\kappa\mu r}}. \tag{H.89}
$$

  Note that in view of (H.37) and (H.38), (H.89) is guaranteed by $ndp^2\gtrsim\delta^{-2}\kappa^4\mu^3r^4\log^4(n+d)$,

$$
\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}}\lesssim\frac{\delta}{\sqrt{\kappa^2\mu r^2\log^2(n+d)}},\qquad\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}\lesssim\frac{\delta}{\sqrt{\kappa^2\mu r^2\log^2(n+d)}}.
$$

- Regarding $\gamma_{4,2,2}$, we have

$$\gamma_{4,2,2} \asymp \underbrace{\frac{1}{np^2}\left(\theta+\sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\omega_{\max}^4\frac{\theta^2}{\kappa\sigma_r^{\star 2}}}_{=:\gamma_{4,2,2,1}}+\underbrace{\frac{1}{np^2}\theta^2\omega_{\max}^6\frac{\theta^2}{\kappa\sigma_r^{\star 2}}}_{=:\gamma_{4,2,2,2}}$$
$$+\underbrace{\frac{1}{np^2}\theta\omega_{\max}^5\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\frac{\theta^2}{\kappa\sigma_r^{\star 2}}}_{=:\gamma_{4,2,2,3}}+\underbrace{\frac{1}{np^2}\zeta_{\mathsf{2nd}}\omega_{\max}^4\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\frac{\theta^2}{\sqrt{\kappa}\sigma_r^{\star}}}_{=:\gamma_{4,2,2,4}}$$
$$+\underbrace{\frac{1}{np^2}\zeta_{\mathsf{2nd}}^2\omega_{\max}^4\theta^2}_{=:\gamma_{4,2,2,5}}\lesssim\delta v_{i,j}^{\star}.$$

Here, the last relation holds since

$$\gamma_{4,2,2,1}\lesssim\delta\frac{\omega_{\min}^4}{np^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right)\lesssim\delta v_{i,j}^{\star},$$
$$\gamma_{4,2,2,2}\lesssim\frac{1}{np^2\kappa}\frac{\omega_{\max}^2}{\sigma_r^{\star 2}}\cdot\theta^4\omega_{\max}^4\overset{\text{(i)}}{\lesssim}\frac{1}{np^2\kappa}\frac{\omega_{\max}^2}{\sigma_r^{\star 2}}\varepsilon^2 v_{i,j}^{\star}\lesssim\delta v_{i,j}^{\star},$$
$$\gamma_{4,2,2,3}\lesssim\theta\frac{\omega_{\max}}{\sqrt{\kappa}\sigma_r^{\star}}\frac{1}{\sqrt{np^2}}\cdot\frac{\omega_{\min}^2}{\sqrt{np^2}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2\cdot\theta^2\omega_{\max}^2\overset{\text{(ii)}}{\lesssim}\varepsilon\theta\frac{\omega_{\max}}{\sqrt{\kappa}\sigma_r^{\star}}\frac{1}{\sqrt{np^2}}v_{i,j}^{\star}\lesssim\delta v_{i,j}^{\star},$$
$$\gamma_{4,2,2,4}\lesssim\frac{1}{\sqrt{ndp^2}}\zeta_{\mathsf{2nd}}\sqrt{d}\cdot\omega_{\max}^2\theta^2\cdot\frac{\omega_{\min}^2}{\sqrt{np^2}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2\overset{\text{(iii)}}{\lesssim}\frac{\varepsilon}{\sqrt{ndp^2}}\zeta_{\mathsf{2nd}}\sqrt{d}v_{i,j}^{\star}\lesssim\delta v_{i,j}^{\star},$$
$$\gamma_{4,2,2,5}\lesssim\frac{\omega_{\max}^2}{\sigma_1^{\star 2}}\frac{1}{np^2}\cdot\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}\cdot\theta^2\omega_{\max}^2\lesssim\frac{\omega_{\max}^2}{\sigma_1^{\star 2}}\frac{1}{np^2}\varepsilon^2 v_{i,j}^{\star}\lesssim\delta v_{i,j}^{\star},$$

provided that $\theta\lesssim\delta$, $n\gtrsim\kappa^3 r\log(n+d)$, $\varepsilon\lesssim 1$, $\zeta_{\mathsf{2nd}}\sqrt{d}\lesssim\delta$, $ndp^2\gtrsim 1$,

$$\frac{\omega_{\max}^2}{\sigma_r^{\star 2}}\frac{1}{np^2\kappa}\lesssim\delta.$$

Here, (i)-(iii) rely on (H.81). In view of the fact that

$$\frac{1}{\sqrt{ndp^2}}\cdot\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}=\frac{\omega_{\max}^2}{\sigma_r^{\star 2}}\frac{1}{np^2},$$

the last condition above is guaranteed by $ndp^2\gtrsim 1$ and

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}\lesssim\delta.$$

- Regarding $\gamma_{4,2,3}$, we have

$$\gamma_{4,2,3}\asymp\underbrace{\frac{\mu r}{ndp^2}\sigma_1^{\star 2}\left(\theta+\sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2}_{=:\gamma_{4,2,3,1}}+\underbrace{\frac{\mu r}{ndp^2}\sigma_1^{\star 2}\theta^2\omega_{\max}^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2}_{=:\gamma_{4,2,3,2}}$$
$$+\underbrace{\frac{\mu r}{ndp^2}\sigma_1^{\star 2}\theta\omega_{\max}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2}_{=:\gamma_{4,2,3,3}}+\underbrace{\frac{\mu r}{ndp^2}\sigma_1^{\star 3}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2}_{=:\gamma_{4,2,3,4}}$$
$$+\underbrace{\frac{\mu r}{ndp^2}\sigma_1^{\star 2}\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2}_{=:\gamma_{4,2,3,5}}\lesssim\delta v_{i,j}^{\star}.$$

Here, the last relation holds since

$$\begin{aligned}
\gamma_{4,2,3,1} &\lesssim \frac{\delta}{ndp^2\kappa}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 \lesssim \delta v_{i,j}^{\star},\\
\gamma_{4,2,3,2} &\lesssim \delta\frac{\omega_{\min}^2\sigma_r^{\star 2}}{ndp^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right) \lesssim \delta v_{i,j}^{\star},\\
\gamma_{4,2,3,3} &\overset{\text{(i)}}{\lesssim} \delta\frac{1}{ndp^2\kappa}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2+\delta\frac{\omega_{\min}^2\sigma_r^{\star 2}}{ndp^2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2 \lesssim \delta v_{i,j}^{\star},\\
\gamma_{4,2,3,4} &\lesssim \frac{\kappa\mu r}{\sqrt{ndp^2}}\cdot\frac{1}{\sqrt{ndp^2\kappa}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\cdot\zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)\\
&\overset{\text{(ii)}}{\lesssim} \frac{\kappa\mu r}{\sqrt{ndp^2}}\cdot\frac{1}{\sqrt{ndp^2\kappa}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\cdot\varepsilon(v_{i,j}^{\star})^{1/2} \lesssim \frac{\kappa\mu r}{\sqrt{ndp^2}}\varepsilon v_{i,j}^{\star} \lesssim \delta v_{i,j}^{\star},\\
\gamma_{4,2,3,5} &\lesssim \frac{\mu r}{ndp^2}\left[\zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)\right]^2 \overset{\text{(iii)}}{\lesssim} \frac{\mu r}{ndp^2}\varepsilon v_{i,j}^{\star} \lesssim \delta v_{i,j}^{\star},
\end{aligned}$$

provided that $\theta \lesssim \delta/(\kappa^2\mu r)$, $n \gtrsim \delta^{-2}\kappa^7\mu^2 r^3\log(n+d)$, $\varepsilon \lesssim 1$ and $ndp^2 \gtrsim \delta^{-2}\kappa^2\mu^2 r^2$. Here, (i) utilizes the AM-GM inequality, whereas (ii) and (iii) utilize (H.80).

- Regarding $\gamma_{4,2,4}$, it follows that

$$\begin{aligned}
\gamma_{4,2,4} \asymp{}& \underbrace{\frac{\mu r}{ndp^2}\theta^2\omega_j^{\star 2}\left(\theta+\sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2}_{=:\gamma_{4,2,4,1}}+\underbrace{\frac{\mu r}{ndp^2}\theta^4\omega_{\max}^4}_{=:\gamma_{4,2,4,2}}\\
&+\underbrace{\frac{\mu r}{ndp^2}\theta^3\omega_{\max}^3\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2}_{=:\gamma_{4,2,4,3}}+\underbrace{\frac{\mu r}{ndp^2}\theta^2\omega_j^{\star 2}\sigma_1^{\star}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2}_{=:\gamma_{4,2,4,4}}\\
&+\underbrace{\frac{\mu r}{ndp^2}\theta^2\omega_j^{\star 2}\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2}}_{=:\gamma_{4,2,4,5}} \lesssim \delta v_{i,j}^{\star},
\end{aligned}$$

where the last relation holds since

$$\begin{aligned}
\gamma_{4,2,4,1} &\lesssim \delta\frac{\omega_{\min}^2\sigma_r^{\star 2}}{ndp^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right) \lesssim \delta v_{i,j}^{\star},\\
\gamma_{4,2,4,2} &\overset{\text{(i)}}{\lesssim} \frac{\mu r}{ndp^2}\varepsilon^2 v_{i,j}^{\star} \lesssim \delta v_{i,j}^{\star},\\
\gamma_{4,2,4,3} &\overset{\text{(ii)}}{\lesssim} \frac{\sqrt{\kappa}\mu r}{\sqrt{ndp^2}}\theta\cdot\theta^2\omega_{\max}^2\cdot\frac{\omega_{\max}\sigma_r^{\star}}{\sqrt{ndp^2}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 \lesssim \frac{\sqrt{\kappa}\mu r}{\sqrt{ndp^2}}\varepsilon\theta v_{i,j}^{\star} \lesssim \delta v_{i,j}^{\star},\\
\gamma_{4,2,4,4} &\lesssim \frac{\mu r}{ndp^2}\cdot\theta^2\omega_j^{\star 2}\cdot\zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right) \overset{\text{(iii)}}{\lesssim} \frac{\mu r\varepsilon^2}{ndp^2}v_{i,j}^{\star} \lesssim \delta v_{i,j}^{\star},\\
\gamma_{4,2,4,5} &\lesssim \frac{\mu r}{ndp^2}\cdot\theta^2\omega_j^{\star 2}\cdot\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2} \overset{\text{(iv)}}{\lesssim} \frac{\mu r}{ndp^2}v_{i,j}^{\star} \lesssim \delta v_{i,j}^{\star},
\end{aligned}$$

as long as $\theta \lesssim \delta/(\kappa\mu r)$, $n \gtrsim \kappa^3 r\log(n+d)$, $ndp^2 \gtrsim \delta^{-2}\mu r$, $\varepsilon \lesssim 1$. Here, (i), (ii) and (iv) utilize (H.81); (iii) utilizes (H.79).

- Regarding $\gamma_{4,2,5}$, we can derive

$$\gamma_{4,2,5} \lesssim \delta\frac{\omega_{\min}^2\sigma_r^{\star 2}}{ndp^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right) \lesssim \delta v_{i,j}^{\star}$$

provided that $np \gtrsim \delta^{-2}\kappa^2\mu^2 r^2\log^2(n+d)$.

- Finally, when it comes to $\gamma_{4,2,6}$, we can bound

$$\gamma_{4,2,6} \overset{\text{(i)}}{\lesssim} \theta^2 \sqrt{\frac{1}{np}} \omega_{\max}^4 \theta^2 \overset{\text{(ii)}}{\lesssim} \sqrt{\frac{1}{np}} v_{i,j}^{\star} \overset{\text{(iii)}}{\lesssim} \delta v_{i,j}^{\star}.$$

  Here (i) follows from (H.37), (ii) comes from (H.81), whereas (iii) holds provided that $np \gtrsim \delta^{-2}$.

Take the bounds on the terms (from $\gamma_{4,2,1}$ to $\gamma_{4,2,6}$) together to arrive at

$$\gamma_{4,2} \lesssim \gamma_{4,2,1} + \gamma_{4,2,2} + \gamma_{4,2,3} + \gamma_{4,2,4} + \gamma_{4,2,5} + \gamma_{4,2,6} \lesssim \delta v_{i,j}^{\star},$$

provided that $\theta \lesssim \delta/(\kappa^2 \mu r)$, $\varepsilon \lesssim 1$, $\zeta_{\mathsf{2nd}}\sqrt{d} \lesssim \delta/\sqrt{\kappa^2 \mu r}$, $n \gtrsim \delta^{-2}\kappa^7\mu^2 r^3 \log(n+d)$, $ndp^2 \gtrsim \delta^{-2}\kappa^4\mu^3 r^4 \log^4(n+d)$, $np \gtrsim \delta^{-2}\kappa^2\mu^2 r^2 \log^2(n+d)$

$$\frac{\omega_{\max}}{\sigma_r^{\star}} \sqrt{\frac{d}{np}} \lesssim \frac{\delta}{\sqrt{\kappa^2 \mu r^2 \log^2(n+d)}} \qquad \text{and} \qquad \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}} \sqrt{\frac{d}{n}} \lesssim \frac{\delta}{\sqrt{\kappa^2 \mu r^2 \log^2(n+d)}}.$$

Step 4.3: bounding $\gamma_{4,3}$. In view of (H.87), we can upper bound

$$\begin{aligned}
\gamma_{4,3} &\lesssim \frac{1}{np^2}\left(\omega_{\max}^2 + \left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\right)\left(\omega_{\max}^2 + \frac{\mu r}{d}\sigma_1^{\star 2}\right)\sum_{k=1}^{d}\left|\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2 - \left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\top}\right)^2\right| \\
&\lesssim \underbrace{\frac{\omega_{\max}^4}{np^2}\sum_{k=1}^{d}\left|\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2 - \left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\top}\right)^2\right|}_{=:\gamma_{4,3,1}} + \underbrace{\frac{\mu r}{ndp^2}\sigma_1^{\star 2}\omega_{\max}^2\sum_{k=1}^{d}\left|\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2 - \left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\top}\right)^2\right|}_{=:\gamma_{4,3,2}} \\
&\quad + \underbrace{\frac{\mu r}{ndp^2}\sigma_1^{\star 2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\sum_{k=1}^{d}\left|\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2 - \left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\top}\right)^2\right|}_{=:\gamma_{4,3,3}}.
\end{aligned}$$

Note that for each $k \in [d]$,

$$\begin{aligned}
\left|\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top} - \boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\top}\right| &= \left|\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top} - \boldsymbol{U}_{k,\cdot}\boldsymbol{R}\left(\boldsymbol{U}_{j,\cdot}\boldsymbol{R}\right)^{\top}\right| \\
&\le \left|\left(\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star}\right)_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\star}\right| + \left|\left(\boldsymbol{U}\boldsymbol{R}\right)_{k,\cdot}\left(\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star}\right)_{j,\cdot}^{\top}\right| \\
&\le \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \left\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star}\right\|_{2,\infty} + \left\|\boldsymbol{U}_{k,\cdot}\right\|_2 \left\|\left(\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star}\right)_{j,\cdot}\right\|_2 \\
&\overset{\text{(i)}}{\lesssim} \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}\sqrt{\frac{r\log(n+d)}{d}} + \left\|\boldsymbol{U}_{k,\cdot}\right\|_2 \left[\frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\left(\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2 + \omega_j^{\star}\right) + \zeta_{\mathsf{2nd}}\right] \\
&\overset{\text{(ii)}}{\lesssim} \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}\sqrt{\frac{r\log(n+d)}{d}} + \left\|\boldsymbol{U}_{k,\cdot}\right\|_2 \frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\omega_j^{\star} + \left\|\boldsymbol{U}_{k,\cdot}\right\|_2 \zeta_{\mathsf{2nd}}.
\end{aligned}$$

Here, (i) follows from Lemma 40 as well as a direct consequence of (G.23e):

$$\left\|\boldsymbol{U}\right\|_{2,\infty} \le \left\|\boldsymbol{U}^{\star}\right\|_{2,\infty} + \left\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star}\right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r}{d}} + \frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}\sqrt{\frac{r\log(n+d)}{d}} \lesssim \sqrt{\frac{\mu r}{d}},$$

provided that $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2} \lesssim 1/\sqrt{\log(n+d)}$; and (ii) utilizes (H.39). Therefore, one can derive

$$\sum_{k=1}^{d}\left|\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2 - \left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\top}\right)^2\right| \lesssim \sum_{k=1}^{d}\left|\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top} - \boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\top}\right|\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\left\|\boldsymbol{U}_{k,\cdot}^{\star}\right\|_2 + \sum_{k=1}^{d}\left|\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top} - \boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\top}\right|^2$$

$$
\begin{aligned}
&\lesssim \sum_{k=1}^{d}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2 \frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}\sqrt{\frac{r\log(n+d)}{d}}\left\|\boldsymbol{U}_{k,\cdot}^{\star}\right\|_2+\sum_{k=1}^{d}\left\|\boldsymbol{U}_{k,\cdot}\right\|_2\left\|\boldsymbol{U}_{k,\cdot}^{\star}\right\|_2\frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\omega_j^{\star}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \\
&\quad+\sum_{k=1}^{d}\left\|\boldsymbol{U}_{k,\cdot}\right\|_2\left\|\boldsymbol{U}_{k,\cdot}^{\star}\right\|_2\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2+\sum_{k=1}^{d}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\frac{\zeta_{\mathsf{1st}}^2}{\sigma_r^{\star 4}}\frac{r\log(n+d)}{d} \\
&\quad+\sum_{k=1}^{d}\left\|\boldsymbol{U}_{k,\cdot}\right\|_2^2\frac{\theta^2}{\kappa\sigma_r^{\star 2}}\omega_j^{\star 2}+\sum_{k=1}^{d}\left\|\boldsymbol{U}_{k,\cdot}\right\|_2^2\zeta_{\mathsf{2nd}}^2 \\
&\overset{\text{(i)}}{\lesssim}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}\sqrt{r^2\log(n+d)}+\frac{\theta r}{\sqrt{\kappa}\sigma_r^{\star}}\omega_j^{\star}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2+\zeta_{\mathsf{2nd}}r\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \\
&\quad+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\frac{\zeta_{\mathsf{1st}}^2}{\sigma_r^{\star 4}}r^2\log(n+d)+\frac{\theta^2 r}{\kappa\sigma_r^{\star 2}}\omega_j^{\star 2}+\zeta_{\mathsf{2nd}}^2 r \\
&\overset{\text{(ii)}}{\lesssim}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}\sqrt{r^2\log(n+d)}+\frac{\theta r}{\sqrt{\kappa}\sigma_r^{\star}}\omega_j^{\star}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2+\zeta_{\mathsf{2nd}}r\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2+\frac{\theta^2 r}{\kappa\sigma_r^{\star 2}}\omega_j^{\star 2}.
\end{aligned} \tag{H.90}
$$

Here, (i) follows from the Cauchy-Schwarz inequality

$$
\sum_{k=1}^{d}\left\|\boldsymbol{U}_{k,\cdot}^{\star}\right\|_2\leq\sqrt{d\sum_{k=1}^{d}\left\|\boldsymbol{U}_{k,\cdot}^{\star}\right\|_2^2}\leq\sqrt{d\left\|\boldsymbol{U}^{\star}\right\|_{\mathrm{F}}^2}\leq\sqrt{dr},
$$

as well as the following bound

$$
\begin{aligned}
\sum_{k=1}^{d}\left\|\boldsymbol{U}_{k,\cdot}\right\|_2\left\|\boldsymbol{U}_{k,\cdot}^{\star}\right\|_2&\leq\sum_{k=1}^{d}\left\|\boldsymbol{U}_{k,\cdot}^{\star}\right\|_2^2+\left\|\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^{\star}\right\|_{2,\infty}\sum_{k=1}^{d}\left\|\boldsymbol{U}_{k,\cdot}^{\star}\right\|_2 \\
&\lesssim\left\|\boldsymbol{U}^{\star}\right\|_{\mathrm{F}}^2+\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}\sqrt{\frac{r\log(n+d)}{d}}\cdot\sqrt{dr}\lesssim r,
\end{aligned}
$$

which utilizes Lemma 40 and holds provided that $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2}\lesssim 1/\sqrt{\log(n+d)}$; (ii) holds provided that $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2}\lesssim 1/\sqrt{\log(n+d)}$. Then we shall bound the terms $\gamma_{4,3,1}$, $\gamma_{4,3,2}$ and $\gamma_{4,3,3}$ respectively.

- Regarding $\gamma_{4,3,1}$, we have

$$
\begin{aligned}
\gamma_{4,3,1}&\lesssim\underbrace{\frac{\sqrt{r^2\log(n+d)}}{np^2}\omega_{\max}^4\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}}_{=:\gamma_{4,3,1,1}}+\underbrace{\frac{r}{np^2}\omega_{\max}^5\frac{\theta}{\sqrt{\kappa}\sigma_r^{\star}}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2}_{=:\gamma_{4,3,1,2}} \\
&\quad+\underbrace{\frac{r}{np^2}\omega_{\max}^4\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2}_{=:\gamma_{4,3,1,3}}+\underbrace{\frac{r}{np^2}\omega_{\max}^6\frac{\theta^2}{\kappa\sigma_r^{\star 2}}}_{=:\gamma_{4,3,1,4}}\lesssim\delta v_{i,j}^{\star}.
\end{aligned}
$$

  Here, the last relation holds since

$$
\begin{aligned}
\gamma_{4,3,1,1}+\gamma_{4,3,1,2}&\lesssim\delta\frac{\omega_{\min}^4}{np^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right)\lesssim\delta v_{i,j}^{\star}, \\
\gamma_{4,3,1,3}&\overset{\text{(i)}}{\lesssim}\frac{r}{np^2}\omega_{\max}^4\frac{\varepsilon}{\sqrt{\kappa}}\left[\frac{1}{\sqrt{\min\{ndp^2\kappa,np\}}\sigma_1^{\star}}\left(\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\sigma_1^{\star}+\omega_{\min}\right)+\frac{\omega_{\min}^2}{\sqrt{np^2}\sigma_1^{\star 2}}\right]\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \\
&\lesssim\frac{\varepsilon r}{\sqrt{\kappa}}\frac{1}{\sqrt{\min\{ndp^2\kappa,np\}}}\frac{\omega_{\max}^4}{np^2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\left(\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2+\frac{\omega_{\min}}{\sigma_1^{\star}}\right)+\frac{\varepsilon r}{\sqrt{\kappa}}\frac{\omega_{\min}^6}{n^{3/2}p^3\sigma_1^{\star 2}}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 \\
&\lesssim\delta\frac{\omega_{\min}^4}{np^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right)\lesssim\delta v_{i,j}^{\star},
\end{aligned}
$$

$$\gamma_{4,3,1,4} \lesssim \theta^2\omega_{\max}^2 \cdot \frac{r}{np^2}\frac{\omega_{\max}^4}{\kappa\sigma_r^{\star 2}} \overset{\text{(ii)}}{\lesssim} \varepsilon(v_{i,j}^\star)^{1/2} \cdot \frac{r}{np^2}\frac{\omega_{\max}^4}{\kappa\sigma_r^{\star 2}} \lesssim \varepsilon(v_{i,j}^\star)^{1/2} \cdot \delta\frac{\omega_{\min}^2}{\sqrt{np^2}}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right)$$
$$\lesssim \delta v_{i,j}^\star,$$

provided that $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2} \lesssim \delta/\sqrt{r^2\log(n+d)}$, $\varepsilon \lesssim 1$, $ndp^2 \gtrsim \delta^{-2}r^2$, $np \gtrsim \delta^{-2}r^2$,

$$\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2 \gtrsim \delta^{-1}r\frac{\omega_{\max}}{\sigma_1^\star}\theta + \delta^{-1}\frac{\omega_{\min}}{\sigma_1^\star}\frac{r}{\sqrt{\min\{ndp^2\kappa, np\}}} + \delta^{-1}r\frac{\omega_{\max}^2}{\sigma_1^{\star 2}}\sqrt{\frac{1}{np^2}}.$$

Here (i) utilizes (H.80), while (ii) utilizes (H.81).

- Regarding $\gamma_{4,3,2}$, we have

$$\gamma_{4,3,2} \lesssim \underbrace{\frac{\mu r^2\sqrt{\log(n+d)}}{ndp^2}\sigma_1^{\star 2}\omega_{\max}^2\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2^2\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}}_{=:\gamma_{4,3,2,1}} + \underbrace{\frac{\mu r^2}{ndp^2}\sigma_1^\star\omega_{\max}^3\theta\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2}_{=:\gamma_{4,3,2,2}}$$
$$+ \underbrace{\frac{\mu r^2}{ndp^2}\sigma_1^{\star 2}\omega_{\max}^2\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2}_{=:\gamma_{4,3,2,3}} + \underbrace{\frac{\mu r^2}{ndp^2}\omega_{\max}^4\theta^2}_{=:\gamma_{4,3,2,4}} \lesssim \delta v_{i,j}^\star.$$

Here, the last relation holds since

$$\gamma_{4,3,2,1} + \gamma_{4,3,2,2} \lesssim \delta\frac{\omega_{\min}^2\sigma_r^{\star 2}}{ndp^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2^2 + \left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2^2\right) \lesssim \delta v_{i,j}^\star,$$
$$\gamma_{4,3,2,3} \lesssim \frac{\mu r^2}{ndp^2}\omega_{\max}^2 \cdot \zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2 \overset{\text{(i)}}{\lesssim} \frac{\mu r^2}{ndp^2}\omega_{\max}^2 \cdot \varepsilon(v_{i,j}^\star)^{1/2},$$
$$\lesssim \delta\frac{\omega_{\min}^2}{\sqrt{np^2}}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right) \cdot \varepsilon(v_{i,j}^\star)^{1/2} \lesssim \delta v_{i,j}^\star,$$
$$\gamma_{4,3,2,4} \lesssim \frac{\mu r^2}{ndp^2}\omega_{\max}^2 \cdot \omega_{\max}^2\theta^2 \overset{\text{(ii)}}{\lesssim} \frac{\mu r^2}{ndp^2}\omega_{\max}^2 \cdot \varepsilon(v_{i,j}^\star)^{1/2},$$
$$\lesssim \delta\frac{\omega_{\min}^2}{\sqrt{np^2}}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right) \cdot \varepsilon(v_{i,j}^\star)^{1/2} \lesssim \delta v_{i,j}^\star,$$

provided that $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2} \lesssim \delta/\sqrt{\kappa^2\mu^2r^4\log(n+d)}$ and

$$\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2 \gtrsim \delta^{-1}\sqrt{\kappa}\mu r^2\frac{\omega_{\max}}{\sigma_r^\star}\theta + \delta^{-1}\frac{\mu r^2}{\sqrt{ndp^2}} \cdot \sqrt{\frac{1}{d}}.$$

Here, (i) utilizes (H.79) and (ii) follows from (H.81).

- Regarding $\gamma_{4,3,3}$, we have

$$\gamma_{4,3,3} \lesssim \underbrace{\frac{\mu r^2\sqrt{\log(n+d)}}{ndp^2}\sigma_1^{\star 2}\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2^2\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}}_{=:\gamma_{4,3,3,1}} + \underbrace{\frac{\mu r^2}{ndp^2}\sigma_1^\star\omega_{\max}\theta\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2}_{=:\gamma_{4,3,3,2}}$$
$$+ \underbrace{\frac{\mu r^2}{ndp^2}\sigma_1^{\star 2}\zeta_{\mathsf{2nd}}\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2}_{=:\gamma_{4,3,3,3}} + \underbrace{\frac{\mu r^2}{ndp^2}\omega_{\max}^2\theta^2\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2}_{=:\gamma_{4,3,3,4}} \lesssim \delta v_{i,j}^\star.$$

Here, the last relation holds since

$$\gamma_{4,3,3,1} \lesssim \frac{\delta}{ndp^2\kappa}\left\|\boldsymbol{U}_{i,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2^2 \lesssim \delta v_{i,j}^\star,$$

$$
\begin{aligned}
\gamma_{4,3,3,2} &\overset{\text{(i)}}{\lesssim} \frac{\delta}{ndp^2\kappa}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2+\delta\frac{\omega_{\min}^2\sigma_r^{\star 2}}{ndp^2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2\lesssim\delta v_{i,j}^{\star},\\
\gamma_{4,3,3,3} &\lesssim \frac{\kappa\mu r^2}{\sqrt{ndp^2}}\cdot\zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)\cdot\frac{1}{\sqrt{ndp^2\kappa}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\\
&\overset{\text{(ii)}}{\lesssim} \frac{\kappa\mu r^2}{\sqrt{ndp^2}}\varepsilon(v_{i,j}^{\star})^{1/2}\cdot\frac{1}{\sqrt{ndp^2\kappa}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\lesssim\delta v_{i,j}^{\star}\\
\gamma_{4,3,3,4} &\lesssim \delta\frac{\omega_{\min}^2\sigma_r^{\star 2}}{ndp^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right)\lesssim\delta v_{i,j}^{\star},
\end{aligned}
$$

provided that $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2}\lesssim\delta/(\kappa^2\mu r^2\sqrt{\log(n+d)})$, $\theta\lesssim\delta/(\kappa\mu r^2)$, $\varepsilon\lesssim 1$, $ndp^2\gtrsim\delta^{-2}\kappa^2\mu^2r^4$. Here (i) invokes the AM-GM inequality, whereas (ii) follows from (H.79).

Taking the bounds on $\gamma_{4,3,1}$, $\gamma_{4,3,2}$ and $\gamma_{4,3,3}$ collectively yields

$$
\gamma_{4,3}\lesssim\gamma_{4,3,1}+\gamma_{4,3,2}+\gamma_{4,3,3}\lesssim\delta v_{i,j}^{\star},
$$

provided that $\zeta_{\mathsf{1st}}/\sigma_r^{\star 2}\lesssim\delta/(\kappa^2\mu r^2\sqrt{\log(n+d)})$, $\theta\lesssim\delta/(\kappa\mu r^2)$, $\varepsilon\lesssim 1$, $ndp^2\gtrsim\delta^{-2}\kappa^2\mu^2r^4$, $np\gtrsim\delta^{-2}r^2$, and

$$
\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\gtrsim\delta^{-1}\sqrt{\kappa}\mu r^2\frac{\omega_{\max}}{\sigma_r^{\star}}\theta+\delta^{-1}\frac{\omega_{\min}}{\sigma_1^{\star}}\frac{r}{\sqrt{\min\{ndp^2\kappa,np\}}}+\delta^{-1}r\frac{\omega_{\max}^2}{\sigma_1^{\star 2}}\sqrt{\frac{1}{np^2}}+\delta^{-1}\frac{\mu r^2}{\sqrt{ndp^2}}\cdot\sqrt{\frac{1}{d}}. \tag{H.91}
$$

In view of (H.84), we know that the above condition (H.91) is guaranteed by

$$
\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\gtrsim\delta^{-1}\kappa\mu r^2\log(n+d)\left[\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}}+\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}+\frac{\kappa\mu r\log(n+d)}{\sqrt{ndp^2}}\right]\cdot\sqrt{\frac{r}{d}}.
$$

Step 4.4: putting the bounds on $\gamma_{4,1}$, $\gamma_{4,2}$ and $\gamma_{4,3}$ together. In view of (H.39), we can take the bounds on $\gamma_{4,1}$, $\gamma_{4,2}$ and $\gamma_{4,3}$ together to reach

$$
|\alpha_4-\beta_4|\lesssim\gamma_{4,1}+\gamma_{4,2}+\gamma_{4,3}\lesssim\delta v_{i,j}^{\star},
$$

provided that $\varepsilon\lesssim 1$, $n\gtrsim\delta^{-2}\kappa^7\mu^2r^3\log(n+d)$, $ndp^2\gtrsim\delta^{-2}\kappa^4\mu^3r^4\log^4(n+d)$, $np\gtrsim\delta^{-2}\kappa^2\mu^2r^2\log^2(n+d)$,

$$
\frac{\zeta_{\mathsf{1st}}}{\sigma_r^{\star 2}}\lesssim\frac{\delta}{\kappa^2\mu r^2\sqrt{\log(n+d)}},\qquad\zeta_{\mathsf{2nd}}\sqrt{d}\lesssim\frac{\delta}{\sqrt{\kappa^2\mu r}}, \tag{H.92}
$$

$$
\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}}\lesssim\frac{\delta}{\sqrt{\kappa^2\mu r^2\log^2(n+d)}},\qquad\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}\lesssim\frac{\delta}{\sqrt{\kappa^2\mu r^2\log^2(n+d)}},
$$

and

$$
\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\gtrsim\delta^{-1}\kappa\mu r^2\log(n+d)\left[\frac{\kappa\mu r\log(n+d)}{\sqrt{ndp}}+\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}+\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}}\right]\sqrt{\frac{r}{d}}.
$$

Let us now take a closer look at (H.92). In view of (H.38) and the definition of ζ_{2nd}, we know that (H.92) is equivalent to

$$
\begin{aligned}
&ndp^2\gtrsim\delta^{-2}\kappa^6\mu^4r^6\log^5(n+d),\qquad np\gtrsim\delta^{-2}\kappa^6\mu^3r^5\log^3(n+d),\\
&\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}\lesssim\frac{\delta}{\kappa^2\mu r^2\log^{3/2}(n+d)},\qquad\frac{\omega_{\max}}{\sigma_r^{\star}}\sqrt{\frac{d}{np}}\lesssim\frac{\delta}{\kappa^{5/2}\mu r^2\log(n+d)},
\end{aligned}
$$

provided that $d\gtrsim\kappa^2\mu\log(n+d)$. This concludes our bound on $|\alpha_4-\beta_4|$ and the required conditions.

Similarly, we can also prove that (which we omit here for the sake of brevity)

$$
|\alpha_6-\beta_6|\lesssim\delta v_{i,j}^{\star}
$$

under the above conditions.

Step 5: bounding $|\alpha_5-\beta_5|$ and $|\alpha_7-\beta_7|$. Regarding $|\alpha_5-\beta_5|$, we first make the observation that

$$\begin{aligned}
|\alpha_5-\beta_5| &\lesssim \frac{1}{np^2}\left|\sum_{k=1}^{d} S_{i,k}^{\star 2}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2-\sum_{k=1}^{d} S_{i,k}^{2}\left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\top}\right)^2\right| \\
&\lesssim \frac{1}{np^2}\left|\sum_{k=1}^{d}\left(S_{i,k}^{2}-S_{i,k}^{\star 2}\right)\left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\top}\right)^2\right|+\frac{1}{np^2}\left|\sum_{k=1}^{d} S_{i,k}^{\star 2}\left[\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2-\left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\top}\right)^2\right]\right| \\
&\lesssim \frac{1}{np^2}\left\|\boldsymbol{U}_{j,\cdot}\right\|_2^2\max_{k\in[d]}\left|S_{i,k}^{2}-S_{i,k}^{\star 2}\right|+\frac{1}{np^2}\left\|\boldsymbol{U}_{i,\cdot}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2,\infty}^2\sum_{k=1}^{d}\left|\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2-\left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\top}\right)^2\right| \\
&\overset{\text{(i)}}{\lesssim} \frac{1}{np^2}\left(\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2+\frac{\theta^2}{\kappa\sigma_r^{\star 2}}\omega_j^{\star 2}\right)\max_{k\in[d]}\left|S_{i,k}^{2}-S_{i,k}^{\star 2}\right|+\frac{\mu r}{ndp^2}\sigma_1^{\star 2}\left\|\boldsymbol{U}_{i,\cdot}\boldsymbol{\Sigma}^{\star}\right\|_2^2\sum_{k=1}^{d}\left|\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2-\left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\top}\right)^2\right| \\
&\overset{\text{(ii)}}{\lesssim} \frac{1}{np^2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\max_{k\in[d]}\left|S_{i,k}^{2}-S_{i,k}^{\star 2}\right|+\frac{1}{np^2}\frac{\theta^2}{\kappa\sigma_r^{\star 2}}\omega_j^{\star 2}\max_{k\in[d]}\left|S_{i,k}^{2}-S_{i,k}^{\star 2}\right|+\delta v_{i,j}^{\star} \\
&\overset{\text{(iii)}}{\lesssim} \underbrace{\frac{1}{np^2}\sqrt{\frac{\mu r}{d}}\sigma_1^{\star}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\max_{k\in[d]}\left|S_{i,k}-S_{i,k}^{\star}\right|}_{=:\gamma_{5,1}}+\underbrace{\frac{1}{np^2}\sqrt{\frac{\mu r}{d}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\frac{\theta^2}{\sqrt{\kappa}\sigma_r^{\star}}\omega_j^{\star 2}\max_{k\in[d]}\left|S_{i,k}-S_{i,k}^{\star}\right|}_{=:\gamma_{5,2}} \\
&\quad+\underbrace{\frac{1}{np^2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\max_{k\in[d]}\left|S_{i,k}-S_{i,k}^{\star}\right|^2}_{=:\gamma_{5,3}}+\underbrace{\frac{1}{np^2}\frac{\theta^2}{\kappa\sigma_r^{\star 2}}\omega_j^{\star 2}\max_{k\in[d]}\left|S_{i,k}-S_{i,k}^{\star}\right|^2}_{=:\gamma_{5,4}}+\delta v_{i,j}^{\star}.
\end{aligned}$$

Here, (i) follows from (H.88); (ii) holds since

$$\frac{\mu r}{ndp^2}\sigma_1^{\star 2}\left\|\boldsymbol{U}_{i,\cdot}\boldsymbol{\Sigma}^{\star}\right\|_2^2\sum_{k=1}^{d}\left|\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2-\left(\boldsymbol{U}_{k,\cdot}\boldsymbol{U}_{j,\cdot}^{\top}\right)^2\right|=\gamma_{4,3,3}\lesssim\delta v_{i,j}^{\star};$$

and (iii) follows from the fact that

$$\begin{aligned}
\max_{k\in[d]}\left|S_{i,k}^{2}-S_{i,k}^{\star 2}\right| &\lesssim \max_{k\in[d]}\left|S_{i,k}-S_{i,k}^{\star}\right|S_{i,k}^{\star}+\max_{k\in[d]}\left|S_{i,k}-S_{i,k}^{\star}\right|^2 \\
&\lesssim \left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_{2,\infty}\max_{k\in[d]}\left|S_{i,k}-S_{i,k}^{\star}\right|+\max_{k\in[d]}\left|S_{i,k}-S_{i,k}^{\star}\right|^2 \\
&\lesssim \sqrt{\frac{\mu r}{d}}\sigma_1^{\star}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\max_{k\in[d]}\left|S_{i,k}-S_{i,k}^{\star}\right|+\max_{k\in[d]}\left|S_{i,k}-S_{i,k}^{\star}\right|^2.
\end{aligned}$$

In view of Lemma 41, we know that for each $k\in[d]$,

$$\begin{aligned}
\left|S_{i,k}-S_{i,k}^{\star}\right| &\lesssim \left(\theta+\sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\sqrt{\frac{\mu r}{d}}\sigma_1^{\star}+\theta^2\omega_{\max}^2 \\
&\quad+\left(\theta\omega_{\max}+\zeta_{\mathsf{2nd}}\sigma_1^{\star}\right)\sqrt{\frac{\mu r}{d}}\sigma_1^{\star}+\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 2} \\
&\lesssim \left(\theta+\sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\sqrt{\frac{\mu r}{d}}\sigma_1^{\star}+\theta\omega_{\max}\sqrt{\frac{\mu r}{d}}\sigma_1^{\star}+\zeta_{\mathsf{2nd}}\sigma_1^{\star}\sqrt{\frac{\mu r}{d}}\sigma_1^{\star},
\end{aligned}$$

where the last relation holds due to (H.89) and (H.80), provided that $\varepsilon\lesssim 1$, $np\gtrsim 1$ and $ndp^2\gtrsim 1$. With these preparations in place, we now proceed to bound the terms $\gamma_{5,1}$, $\gamma_{5,2}$, $\gamma_{5,3}$ and $\gamma_{5,4}$ separately.

- Regarding $\gamma_{5,1}$, we have

$$\gamma_{5,1}\lesssim\underbrace{\frac{\mu r}{ndp^2}\sigma_1^{\star 2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\left(\theta+\sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)}_{=:\gamma_{5,1,1}}$$

$$
+ \underbrace{\frac{\mu r}{ndp^2}\sigma_1^{\star 2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\theta\omega_{\max}}_{=:\gamma_{5,1,2}} + \underbrace{\frac{\mu r}{ndp^2}\sigma_1^{\star 2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\zeta_{\mathsf{2nd}}\sigma_1^{\star}}_{=:\gamma_{5,1,3}}
$$

$$
\lesssim \delta v_{i,j}^{\star},
$$

where the last relation holds since

$$
\begin{aligned}
\gamma_{5,1,1} &\lesssim \frac{\delta}{ndp^2\kappa}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 \lesssim \delta v_{i,j}^{\star}, \\
\gamma_{5,1,2} &\overset{\text{(i)}}{\lesssim} \frac{\delta}{ndp^2\kappa}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2 + \delta\frac{\omega_{\min}^2\sigma_r^{\star 2}}{ndp^2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2 \lesssim \delta v_{i,j}^{\star}, \\
\gamma_{5,1,3} &\lesssim \frac{\kappa\mu r}{\sqrt{ndp^2}}\cdot\sqrt{\frac{1}{ndp^2\kappa}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\cdot\zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right) \\
&\overset{\text{(ii)}}{\lesssim} \frac{\kappa\mu r}{\sqrt{ndp^2}}\cdot\sqrt{\frac{1}{ndp^2\kappa}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\cdot\varepsilon(v_{i,j}^{\star})^{1/2} \lesssim \frac{\kappa\mu r}{\sqrt{ndp^2}}\varepsilon v_{i,j}^{\star} \lesssim \delta v_{i,j}^{\star},
\end{aligned}
$$

provided that $\theta \lesssim \delta/(\kappa^2\mu r)$, $n \gtrsim \delta^{-2}\kappa^7\mu^2 r^3\log(n+d)$, $\varepsilon \lesssim 1$ and $ndp^2 \gtrsim \delta^{-2}\kappa^2\mu^2 r^2$. In the above relations, (i) uses the AM-GM inequality, where (ii) utilizes (H.79).

- Regarding $\gamma_{5,2}$, we can derive

$$
\begin{aligned}
\gamma_{5,2} &\lesssim \underbrace{\frac{\mu r}{ndp^2}\sigma_1^{\star}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\frac{\theta^2}{\sqrt{\kappa}\sigma_r^{\star}}\omega_j^{\star 2}\left(\theta + \sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)}_{=:\gamma_{5,2,1}} \\
&\quad + \underbrace{\frac{\mu r}{ndp^2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\theta^3\omega_{\max}^3}_{=:\gamma_{5,2,2}} + \underbrace{\frac{\mu r}{ndp^2}\sigma_1^{\star}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\theta^2\omega_j^{\star 2}\zeta_{\mathsf{2nd}}}_{=:\gamma_{5,2,3}} \lesssim \delta v_{i,j}^{\star},
\end{aligned}
$$

where the last relation holds since

$$
\begin{aligned}
\gamma_{5,2,1} &\lesssim \delta\frac{\omega_{\min}^2\sigma_r^{\star 2}}{ndp^2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2^2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\right) \lesssim \delta v_{i,j}^{\star}, \\
\gamma_{5,2,2} &\lesssim \frac{\mu r}{ndp^2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\theta^3\omega_{\max}^3 \lesssim \theta^2\gamma_{4,3,2,2} \lesssim \delta v_{i,j}^{\star}, \\
\gamma_{5,2,3} &\lesssim \frac{\mu r}{ndp^2}\cdot\zeta_{\mathsf{2nd}}\sigma_1^{\star 2}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)\cdot\theta^2\omega_j^{\star 2} \overset{\text{(i)}}{\lesssim} \frac{\mu r}{ndp^2}\varepsilon^2 v_{i,j}^{\star} \lesssim \delta v_{i,j}^{\star},
\end{aligned}
$$

provided that $\theta \lesssim \delta/(\kappa\mu r)$, $n \gtrsim \kappa^3 r\log(n+d)$, $\varepsilon \lesssim 1$ and $ndp^2 \gtrsim \delta^{-1}\mu r$. Here, the relation (i) in the above inequality arises from (H.79) and (H.81).

- With regards to $\gamma_{5,3}$, we make the observation that

$$
\begin{aligned}
\gamma_{5,3} &\lesssim \underbrace{\frac{\mu r}{ndp^2}\sigma_1^{\star 2}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\left(\theta^2 + \frac{\kappa^3 r\log(n+d)}{n}\right)}_{=:\gamma_{5,3,1}} \\
&\quad + \underbrace{\frac{\mu r}{ndp^2}\sigma_1^{\star 2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\theta^2\omega_{\max}^2}_{=:\gamma_{5,3,2}} + \underbrace{\frac{\mu r}{ndp^2}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2\zeta_{\mathsf{2nd}}^2\sigma_1^{\star 4}}_{=:\gamma_{5,3,3}} \lesssim \delta v_{i,j}^{\star},
\end{aligned}
$$

where the last relation holds since

$$
\gamma_{5,3,1} \lesssim \left(\theta + \sqrt{\frac{\kappa^3 r\log(n+d)}{n}}\right)\gamma_{5,1,1} \lesssim \delta v_{i,j}^{\star},
$$

$$\gamma_{5,3,2} \lesssim \delta \frac{\omega_{\min}^2 \sigma_r^{\star 2}}{ndp^2} \left(\left\| \boldsymbol{U}_{i,\cdot}^{\star} \right\|_2^2 + \left\| \boldsymbol{U}_{j,\cdot}^{\star} \right\|_2^2 \right) \lesssim \delta v_{i,j}^{\star},$$

$$\gamma_{5,3,3} \lesssim \frac{\mu r}{ndp^2} \left[\zeta_{\mathsf{2nd}} \sigma_1^{\star 2} \left(\left\| \boldsymbol{U}_{i,\cdot}^{\star} \right\|_2 + \left\| \boldsymbol{U}_{j,\cdot}^{\star} \right\|_2 \right) \right]^2 \overset{\text{(i)}}{\lesssim} \frac{\mu r}{ndp^2} \varepsilon^2 v_{i,j}^{\star} \lesssim \delta v_{i,j}^{\star},$$

provided that $\theta \lesssim \sqrt{\delta/(\kappa\mu r)}$, $n \gtrsim \kappa^3 r \log(n+d)$, $\varepsilon \lesssim 1$ and $ndp^2 \gtrsim \delta^{-1}\mu r$. Note that the inequality (i) in the above relation results from (H.79).

- When it comes to $\gamma_{5,4}$, we have the following upper bound

$$\gamma_{5,4} \lesssim \underbrace{\frac{\mu r}{ndp^2} \left\| \boldsymbol{U}_{i,\cdot}^{\star} \boldsymbol{\Sigma}^{\star} \right\|_2^2 \theta^2 \omega_j^{\star 2} \left(\theta^2 + \frac{\kappa^3 r \log(n+d)}{n} \right)}_{=:\gamma_{5,4,1}} + \underbrace{\frac{\mu r}{ndp^2} \theta^4 \omega_{\max}^4}_{=:\gamma_{5,4,2}} + \underbrace{\frac{\mu r}{ndp^2} \frac{\theta^2}{\kappa \sigma_r^{\star 2}} \omega_j^{\star 2} \zeta_{\mathsf{2nd}}^2 \sigma_1^{\star 4}}_{=:\gamma_{5,4,3}} \lesssim \delta v_{i,j}^{\star},$$

where the last inequality holds true since

$$\gamma_{5,4,1} \lesssim \delta \frac{\omega_{\min}^2 \sigma_r^{\star 2}}{ndp^2} \left(\left\| \boldsymbol{U}_{i,\cdot}^{\star} \right\|_2^2 + \left\| \boldsymbol{U}_{j,\cdot}^{\star} \right\|_2^2 \right) \lesssim \delta v_{i,j}^{\star},$$

$$\gamma_{5,4,2} \overset{\text{(i)}}{\lesssim} \frac{\mu r}{ndp^2} \varepsilon^2 v_{i,j}^{\star} \lesssim \delta v_{i,j}^{\star},$$

$$\gamma_{5,4,3} \overset{\text{(ii)}}{\lesssim} \frac{\mu r}{ndp^2 \kappa} \varepsilon^2 v_{i,j}^{\star} \lesssim \delta v_{i,j}^{\star},$$

provided that $\theta \lesssim \delta/(\kappa\mu r)$, $n \gtrsim \kappa^3 r \log(n+d)$ and $ndp^2 \gtrsim \delta^{-1}\mu r$. Here, the inequalities (i) and (ii) follow from (H.81).

Combining the above bounds on $\gamma_{5,1}$, $\gamma_{5,2}$, $\gamma_{5,3}$ and $\gamma_{5,4}$ yields

$$|\alpha_5 - \beta_5| \lesssim \gamma_{5,1} + \gamma_{5,2} + \gamma_{5,3} + \gamma_{5,4} \lesssim \delta v_{i,j}^{\star},$$

provided that $\theta \lesssim \delta/(\kappa^2\mu r)$, $n \gtrsim \delta^{-2}\kappa^7\mu^2 r^3 \log(n+d)$, $\varepsilon \lesssim 1$ and $ndp^2 \gtrsim \delta^{-2}\kappa^2\mu^2 r^2$.

Similarly we can also show that (which we omit here for brevity)

$$|\alpha_7 - \beta_7| \lesssim \delta v_{i,j}^{\star}$$

holds true under these conditions.

Step 6: putting everything together. We are now ready to combine the above bounds on $|\alpha_k - \beta_k|$, $k = 1, \dots, 7$, to conclude that

$$\left| v_{i,j} - v_{i,j}^{\star} \right| \le \sum_{k=1}^{7} |\alpha_k - \beta_k| \lesssim \delta v_{i,j}^{\star},$$

provided that $\varepsilon \lesssim 1$, $n \gtrsim \delta^{-2}\kappa^7\mu^2 r^3 \log(n+d)$, $d \gtrsim \kappa^2\mu \log(n+d)$,

$$ndp^2 \gtrsim \delta^{-2}\kappa^6\mu^4 r^6 \log^5(n+d), \qquad np \gtrsim \delta^{-2}\kappa^6\mu^3 r^5 \log^3(n+d),$$

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}} \sqrt{\frac{d}{n}} \lesssim \frac{\delta}{\kappa^2 \mu r^2 \log^{3/2}(n+d)}, \qquad \frac{\omega_{\max}}{\sigma_r^{\star}} \sqrt{\frac{d}{np}} \lesssim \frac{\delta}{\kappa^{5/2} \mu r^2 \log(n+d)},$$

and

$$\left\| \boldsymbol{U}_{i,\cdot}^{\star} \right\|_2 + \left\| \boldsymbol{U}_{j,\cdot}^{\star} \right\|_2 \gtrsim \delta^{-1}\kappa\mu r^2 \log(n+d) \left[\frac{\kappa\mu r \log(n+d)}{\sqrt{ndp}} + \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}} \sqrt{\frac{d}{n}} + \frac{\omega_{\max}}{\sigma_r^{\star}} \sqrt{\frac{d}{np}} \right] \sqrt{\frac{r}{d}}.$$

To finish up, we shall take $\varepsilon \asymp 1$, and note that (H.79), (H.80) as well as H.81 are guaranteed by the conditions of Lemma 46.

H.4.2 Proof of Lemma 50

To begin with, note that we have learned from Lemma 49 that with probability exceeding $1-O((n+d)^{-10})$,

$$\left|v_{i,j}^{\star}-v_{i,j}\right|\lesssim\delta v_{i,j}^{\star},$$

where δ is the (unspecified) quantity that has appeared in Lemma 49. When $\delta\ll 1$, an immediate result is that $v_{i,j}\asymp v_{i,j}^{\star}$, thus indicating that

$$\Delta\coloneqq\left|\frac{S_{i,j}-S_{i,j}^{\star}}{\sqrt{v_{i,j}^{\star}}}-\frac{S_{i,j}-S_{i,j}^{\star}}{\sqrt{v_{i,j}}}\right|=\left|S_{i,j}-S_{i,j}^{\star}\right|\left|\frac{v_{i,j}^{\star}-v_{i,j}}{\sqrt{v_{i,j}^{\star}v_{i,j}}\left(\sqrt{v_{i,j}^{\star}}+\sqrt{v_{i,j}}\right)}\right|\lesssim\delta\left|S_{i,j}-S_{i,j}^{\star}\right|/\sqrt{v_{i,j}^{\star}}.$$

We can then follow the same analysis as in the proof of Lemma 20 (in Appendix E.4.2) to show that

$$\mathbb{P}\left(S_{i,j}^{\star}\in\mathsf{CI}_{i,j}^{1-\alpha}\right)=1-\alpha+O\left(\frac{1}{\sqrt{\log\left(n+d\right)}}\right)$$

as long as

$$\Delta\lesssim\frac{1}{\sqrt{\log\left(n+d\right)}}. \tag{H.93}$$

holds with probability exceeding $1-O((n+d)^{-10})$. With the above calculations in mind, everything boils down to bounding the quantity Δ to the desired level. Towards this, we are in need of the following lemma.

Claim 3. Instate the conditions in Lemma 44. Suppose $np\gtrsim\log^4(n+d)$ and $ndp^2\gtrsim\mu r\log^5(n+d)$. Then with probability exceeding $1-O((n+d)^{-10})$, we have

$$\begin{aligned}\left|S_{i,j}-S_{i,j}^{\star}\right|&\lesssim\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left(\sqrt{\frac{\kappa\log\left(n+d\right)}{np}}+\sqrt{\frac{\kappa\mu r\log^2\left(n+d\right)}{ndp^2}}\right)\\&\quad+\left[\omega_{\max}\sigma_r^{\star}\left(\sqrt{\frac{\kappa}{np}}+\sqrt{\frac{\kappa\mu r\log^2\left(n+d\right)}{ndp^2}}\right)+\omega_{\max}^2\sqrt{\frac{\kappa\log^2\left(n+d\right)}{np^2}}\right]\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right)+\zeta_{i,j}.\end{aligned}$$

Recall from the proof of Lemma 46 (more specifically, Step 3 in Appendix H.3.3) that

$$(v_{i,j}^{\star})^{-1/2}\zeta_{i,j}\leq\widetilde{v}_{i,j}^{-1/2}\zeta_{i,j}\lesssim\frac{1}{\sqrt{\log\left(n+d\right)}} \tag{H.94}$$

holds under the assumptions of Lemma 46. In addition, recall from Lemma 45 that

$$\sqrt{v_{i,j}^{\star}}\gtrsim\frac{1}{\sqrt{\min\left\{ndp^2\kappa,np\right\}}}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\left(\frac{\omega_{\min}\sigma_r^{\star}}{\sqrt{\min\left\{ndp^2,np\right\}}}+\frac{\omega_{\min}^2}{\sqrt{np^2}}\right)\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\right). \tag{H.95}$$

Therefore we can take (H.94), (H.95) and Claim 3 collectively to show that

$$\Delta\lesssim\delta\sqrt{\kappa^2\mu r\log^2\left(n+d\right)}+\frac{1}{\sqrt{\log\left(n+d\right)}}\lesssim\frac{1}{\sqrt{\log\left(n+d\right)}}$$

with probability exceeding $1-O((n+d)^{-10})$, as long as we take

$$\delta\asymp\frac{1}{\kappa\mu^{1/2}r^{1/2}\log^{3/2}\left(n+d\right)}.$$

This in turn confirms (H.93).

It remains to specify the conditions of Lemma 49 when we choose $\delta \asymp \kappa^{-1}\mu^{-1/2}r^{-1/2}\log^{-3/2}(n+d)$. In this case, these conditions read $n \gtrsim \kappa^9\mu^3 r^4\log^4(n+d)$, $d \gtrsim \kappa^2\mu\log(n+d)$,

$$ndp^2 \gtrsim \kappa^8\mu^5 r^7 \log^8(n+d), \qquad np \gtrsim \kappa^8\mu^4 r^6\log^6(n+d),$$
$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\kappa^3\mu^{3/2}r^{5/2}\log^3(n+d)}, \qquad \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\kappa^{7/2}\mu^{3/2}r^{5/2}\log^{5/2}(n+d)},$$

and

$$\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2 \gtrsim \kappa^2\mu^{3/2}r^{5/2}\log^{5/2}(n+d)\left[\frac{\kappa\mu r\log(n+d)}{\sqrt{ndp}} + \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} + \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d}{np}}\right]\sqrt{\frac{r}{d}},$$

in addition to all the assumptions of Lemma 46. This concludes the proof of this lemma, as long as Claim 3 can be established.

Proof of Claim 3. Let us start by decomposing

$$X_{i,j} = \sum_{l=1}^n \Bigg[\underbrace{M_{j,l}^\natural E_{i,l}}_{=:a_l} + \underbrace{M_{i,l}^\natural E_{j,l}}_{=:b_l} + \underbrace{\sum_{k:k\neq i} E_{i,l}E_{k,l}\left(\boldsymbol{U}_{k,\cdot}^\star(\boldsymbol{U}_{j,\cdot}^\star)^\top\right)}_{=:c_l} + \underbrace{\sum_{k:k\neq j} E_{j,l}E_{k,l}\left(\boldsymbol{U}_{k,\cdot}^\star(\boldsymbol{U}_{i,\cdot}^\star)^\top\right)}_{=:d_l}\Bigg].$$

We shall then bound $\sum_{l=1}^n a_l$, $\sum_{l=1}^n b_l$, $\sum_{l=1}^n c_l$, and $\sum_{l=1}^n d_l$ separately.

- To begin with, the concentration of $\sum_{l=1}^n a_l$ and $\sum_{l=1}^n b_l$ have already been studied in Lemma 1, which reveals that with probability exceeding $1-O((n+d)^{-10})$,

$$\sum_{l=1}^n (a_l+b_l) \lesssim \sqrt{\frac{\kappa\log(n+d)}{np}}\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2 + \frac{\omega_{\max}}{\sqrt{np}}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2 + \left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right)\sigma_1^\star,$$

  provided that $np \gtrsim \log^4(n+d)$.

- We then move on to bound $\sum_{l=1}^n c_l$. It has been shown in (H.65) that with probability exceeding $1-O((n+d)^{-101})$,

$$\max_{1\le l\le n}|c_l| \le \underbrace{\widetilde{C}\sigma_{\mathsf{ub}}B_i\sqrt{\log(n+d)}\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2 + \widetilde{C}BB_i\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\sqrt{\frac{\mu r}{d}}\log(n+d)}_{=:C_{\mathsf{prob}}},$$

  where $\widetilde{C}>0$ is some sufficiently large constant. We also know from (H.67) that $\max_{1\le l\le n}|c_l|$ satisfies the following deterministic bound:

$$\max_{1\le l\le n}|c_l| \le \underbrace{B_i\sqrt{d}B\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2}_{=:C_{\mathsf{det}}}.$$

  Then with probability exceeding $1-O((n+d)^{-101})$, one has

$$\sum_{l=1}^n c_l = \sum_{l=1}^n c_l\,\mathbb{1}_{|c_l|\le C_{\mathsf{prob}}}.$$

  It is then straightforward to calculate that

$$L_c \coloneqq \max_{1\le l\le n}\left|c_l\,\mathbb{1}_{|c_l|\le C_{\mathsf{prob}}}\right| \le C_{\mathsf{prob}},$$
$$V_c \coloneqq \sum_{l=1}^n \mathsf{var}\left(c_l\,\mathbb{1}_{|c_l|\le C_{\mathsf{prob}}}\right) \le \sum_{l=1}^n \mathbb{E}\left[c_l^2\,\mathbb{1}_{|c_l|\le C_{\mathsf{prob}}}\right] \le \sum_{l=1}^n \mathbb{E}\left[c_l^2\right] = \sum_{l=1}^n \mathsf{var}\left(c_l\right)$$

$$= \sum_{l=1}^{n} \mathsf{var}\left[\sum_{k:k\neq i} E_{i,l}E_{k,l}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)\right] \leq \sum_{l=1}^{n}\sum_{k=1}^{d}\sigma_{i,l}^{2}\sigma_{k,l}^{2}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^{2},$$

$$M_c \coloneqq \sum_{l=1}^{n}\mathbb{E}\left[c_l\,\mathbb{1}_{|c_l|>C_{\mathsf{prob}}}\right] \leq C_{\mathsf{det}}\sum_{l=1}^{n}\mathbb{P}\left(|c_l|>C_{\mathsf{prob}}\right) \lesssim C_{\mathsf{det}}\left(n+d\right)^{-100}.$$

On the event $\mathcal{E}_{\mathsf{good}}$, it is seen that

$$\begin{aligned}
L_c &\lesssim \sigma_{\mathsf{ub}}B_i\sqrt{\log\left(n+d\right)}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 + BB_i\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\sqrt{\frac{\mu r}{d}}\log\left(n+d\right),\\
&\overset{\text{(i)}}{\lesssim} \sigma_{\mathsf{ub}}B_i\sqrt{\log\left(n+d\right)}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2 + \sigma_{\mathsf{ub}}\sqrt{\frac{\log\left(n+d\right)}{p}}B_i\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\sqrt{\frac{\mu r}{d}}\log\left(n+d\right),\\
&\lesssim \sigma_{\mathsf{ub}}B_i\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\sqrt{\log\left(n+d\right)}\left(1+\sqrt{\frac{\mu r}{dp}}\log\left(n+d\right)\right),
\end{aligned}$$

$$\begin{aligned}
V_c &\leq \sigma_{\mathsf{ub}}^2\sum_{l=1}^{n}\sigma_{i,l}^2\sum_{k=1}^{d}\left(\boldsymbol{U}_{k,\cdot}^{\star}\boldsymbol{U}_{j,\cdot}^{\star\top}\right)^2 = n\sigma_{\mathsf{ub}}^2\left(\frac{1-p}{np}\left(\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{f}_l\right)^2+\frac{\omega_i^{\star2}}{np}\right)\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2,\\
&\overset{\text{(ii)}}{\lesssim} n\sigma_{\mathsf{ub}}^2\left(\frac{1-p}{np}\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2^2\log\left(n+d\right)+\frac{\omega_i^{\star2}}{np}\right)\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2^2,
\end{aligned}$$

and

$$\begin{aligned}
M_c &\lesssim B_i\sqrt{d}B\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\left(n+d\right)^{-100} \overset{\text{(iii)}}{\lesssim} \sigma_{\mathsf{ub}}B_i\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\sqrt{\frac{d\log\left(n+d\right)}{p}}\left(n+d\right)^{-100}\\
&\overset{\text{(iv)}}{\lesssim} \sigma_{\mathsf{ub}}B_i\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\sqrt{nd}\left(n+d\right)^{-100} \lesssim \sigma_{\mathsf{ub}}B_i\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\left(n+d\right)^{-98}.
\end{aligned}$$

Here, (i) and (iii) follow from (G.9) and (G.8); (ii) arises from (G.11); and (iv) holds true provided that $np \gtrsim \log(n+d)$. Therefore, by virtue of the Bernstein inequality (Vershynin, 2017, Theorem 2.8.4), we see that conditional on $\boldsymbol{F}$,

$$\begin{aligned}
\sum_{l=1}^{n}c_l &= \sum_{l=1}^{n}c_l\,\mathbb{1}_{|c_l|\leq C_{\mathsf{prob}}} \lesssim M_c+\sqrt{V_c\log\left(n+d\right)}+L_c\log\left(n+d\right)\\
&\lesssim \sigma_{\mathsf{ub}}B_i\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\left(n+d\right)^{-98}+\sigma_{\mathsf{ub}}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\sqrt{\frac{\log\left(n+d\right)}{p}}\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2\sqrt{\log\left(n+d\right)}+\omega_i^{\star}\right)\\
&\quad+\sigma_{\mathsf{ub}}B_i\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\log^{3/2}\left(n+d\right)\left(1+\sqrt{\frac{\mu r}{dp}}\log\left(n+d\right)\right)\\
&\overset{\text{(i)}}{\lesssim} \sigma_{\mathsf{ub}}\left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_i^{\star}\right)\sqrt{\frac{\log\left(n+d\right)}{p}}\\
&\lesssim \left\|\boldsymbol{U}_{j,\cdot}^{\star}\right\|_2\left(\left\|\boldsymbol{U}_{i,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_i^{\star}\right)\left(\sqrt{\frac{\mu r\log^2\left(n+d\right)}{ndp^2}}\sigma_1^{\star}+\omega_{\max}\sqrt{\frac{\log\left(n+d\right)}{np^2}}\right)
\end{aligned}$$

holds with probability exceeding $1-O((n+d)^{-10})$. Here, (i) holds true provided that $np \gtrsim \log^3(n+d)$ and $ndp^2 \gtrsim \mu r\log^5(n+d)$. Similarly we can demonstrate that with probability exceeding $1-O((n+d)^{-10})$,

$$\sum_{l=1}^{n}d_l \lesssim \left\|\boldsymbol{U}_{i,\cdot}^{\star}\right\|_2\left(\left\|\boldsymbol{U}_{j,\cdot}^{\star}\boldsymbol{\Sigma}^{\star}\right\|_2+\omega_j^{\star}\right)\left(\sqrt{\frac{\mu r\log^2\left(n+d\right)}{ndp^2}}\sigma_1^{\star}+\omega_{\max}\sqrt{\frac{\log\left(n+d\right)}{np^2}}\right).$$

Therefore, the above two bounds taken together lead to

$$\sum_{l=1}^{n}(c_l+d_l) \lesssim \left[\frac{1}{\sigma_r^\star}\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2+\omega_{\max}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right)\right] \cdot\left(\sqrt{\frac{\mu r\log^2(n+d)}{ndp^2}}\sigma_1^\star+\omega_{\max}\sqrt{\frac{\log(n+d)}{np^2}}\right).$$

Therefore, putting the above bounds together, one can conclude that

$$\begin{aligned}
|X_{i,j}| &\le \sum_{l=1}^{n}(a_l+b_l+c_l+d_l)\\
&\lesssim \left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left(\sqrt{\frac{\kappa\log(n+d)}{np}}+\sqrt{\frac{\kappa\mu r\log^2(n+d)}{ndp^2}}\right)\\
&\quad+\omega_{\max}\sigma_r^\star\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right)\left(\sqrt{\frac{\kappa}{np}}+\sqrt{\frac{\kappa\mu r\log^2(n+d)}{ndp^2}}\right)\\
&\quad+\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{\log(n+d)}{np^2}}+\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right)\omega_{\max}^2\sqrt{\frac{\log(n+d)}{np^2}}\\
&\lesssim \left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left(\sqrt{\frac{\kappa\log(n+d)}{np}}+\sqrt{\frac{\kappa\mu r\log^2(n+d)}{ndp^2}}\right)\\
&\quad+\left[\omega_{\max}\sigma_r^\star\left(\sqrt{\frac{\kappa}{np}}+\sqrt{\frac{\kappa\mu r\log^2(n+d)}{ndp^2}}\right)+\omega_{\max}^2\sqrt{\frac{\kappa\log^2(n+d)}{np^2}}\right]\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right),
\end{aligned}$$

where the last relation holds since (due to the AM-GM inequality)

$$\begin{aligned}
&\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\sqrt{\frac{\kappa\mu r\log^2(n+d)}{ndp^2}}+\omega_{\max}^2\sqrt{\frac{\kappa\log^2(n+d)}{np^2}}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right)\\
&\ge 2\left[\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\sqrt{\frac{\kappa\mu r\log^2(n+d)}{ndp^2}}\omega_{\max}^2\sqrt{\frac{\kappa\log^2(n+d)}{np^2}}\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right)\right]^{1/2}\\
&\ge 2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{\log(n+d)}{np^2}}.
\end{aligned}$$

Combine the above result with Lemma 44 to arrive at

$$\begin{aligned}
\left|S_{i,j}-S_{i,j}^\star\right| &\lesssim \zeta_{i,j}+\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left\|\boldsymbol{U}_{j,\cdot}^\star\boldsymbol{\Sigma}^\star\right\|_2\left(\sqrt{\frac{\kappa\log(n+d)}{np}}+\sqrt{\frac{\kappa\mu r\log^2(n+d)}{ndp^2}}\right).\\
&\quad+\left[\omega_{\max}\sigma_r^\star\left(\sqrt{\frac{\kappa}{np}}+\sqrt{\frac{\kappa\mu r\log^2(n+d)}{ndp^2}}\right)+\omega_{\max}^2\sqrt{\frac{\kappa\log^2(n+d)}{np^2}}\right]\left(\left\|\boldsymbol{U}_{i,\cdot}^\star\right\|_2+\left\|\boldsymbol{U}_{j,\cdot}^\star\right\|_2\right).
\end{aligned}$$

□

I Other useful lemmas

In this section, we gather a few useful results from prior literature that prove useful for our analysis. The first theorem is a non-asymptotic version of the Berry-Esseen Theorem, which has been established using

Stein's method; see Chen et al. (2010, Theorem 3.7).

Theorem 19. *Let $\xi_1, \ldots, \xi_n$ be independent random variables with zero means, satisfying $\sum_{i=1}^n \mathsf{var}(\xi_i) = 1$. Then $W = \sum_{i=1}^n \xi_i$ satisfies*

$$\sup_{z\in\mathbb{R}} \left| \mathbb{P}(W \le z) - \Phi(z) \right| \le 10\gamma, \qquad \text{where } \gamma = \sum_{i=1}^n \mathbb{E}\left[|\xi_i|^3\right].$$

The next theorem, which we borrow from Raič (2019, Theorem 1.1), generalizes Theorem 19 to the vector case.

Theorem 20. *Let $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n$ be independent, $\mathbb{R}^d$-valued random vectors with zero means. Let $\boldsymbol{W} = \sum_{i=1}^n \boldsymbol{\xi}_i$ and assume $\boldsymbol{\Sigma} = \mathsf{cov}(\boldsymbol{W})$ is invertible. Let $\boldsymbol{Z}$ be a d-dimensional Gaussian random vector with zero mean and covariance matrix $\boldsymbol{\Sigma}$. Then we have*

$$\sup_{\mathcal{C}\in\mathscr{C}^d} \left|\mathbb{P}(\boldsymbol{W}\in\mathcal{C}) - \mathbb{P}(\boldsymbol{Z}\in\mathcal{C})\right| \le \left(42d^{1/4}+16\right)\gamma, \qquad \text{where } \gamma = \sum_{i=1}^n \mathbb{E}\left[\left\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\xi}_i\right\|_2^3\right].$$

Here, $\mathscr{C}^d$ represents the set of all convex sets in $\mathbb{R}^d$.

Moreover, deriving our distributional theory involves some basic results about the total-variation distance between two Gaussian distributions, as stated below. This result can be found in Devroye et al. (2018, Theorem 1.1).

Theorem 21. *Let $\boldsymbol{\mu}\in\mathbb{R}^d$, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ be positive definite $d\times d$ matrices. Then the total-variation distance between $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}_1)$ and $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}_2)$ — denoted by $\mathsf{TV}(\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}_1), \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}_2))$ — satisfies*

$$\frac{1}{100} \le \frac{\mathsf{TV}(\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}_1), \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}_2))}{\min\left\{1, \left\|\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1/2} - \boldsymbol{I}_d\right\|_{\mathrm{F}}\right\}} \le \frac{3}{2}.$$

Recall the definition of $\mathcal{C}^\varepsilon$ in (A.3). Then for any convex set $\mathcal{C}\in\mathscr{C}^d$, let us define the following quantity related to Gaussian distributions

$$\gamma(\mathcal{C}) \coloneqq \sup_{\varepsilon>0} \max\left\{\frac{1}{\varepsilon}\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_d)\{\mathcal{C}^\varepsilon\setminus\mathcal{C}\}, \frac{1}{\varepsilon}\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_d)\left\{\mathcal{C}\setminus\mathcal{C}^{-\varepsilon}\right\}\right\}, \tag{I.1}$$

and, in addition,

$$\gamma_d \coloneqq \sup_{\mathcal{C}\in\mathscr{C}^d} \gamma(\mathcal{C}). \tag{I.2}$$

The following theorem from Raič (2019, Theorem 1.2) delivers an upper bound on the quantity γ_d.

Theorem 22. *For all $d\in\mathbb{N}$, we have*

$$\gamma_d < 0.59d^{1/4} + 0.21.$$

Finally, we are in need of the following basic lemma in order to translate results derived for bounded random variables to the ones concerned with sub-Gaussian random variables.

Lemma 51. *There exist two universal constants $C_\delta, C_\sigma > 0$ such that: for any sub-Gaussian random variable X with $\mathbb{E}[X]=0$, $\mathsf{var}(X)=\sigma^2$, $\|X\|_{\psi_2}\lesssim\sigma$, and any $\delta\in(0, C_\delta\sigma)$, one can construct a random variable $\widetilde{X}$ satisfying the following properties:*

1. *$\widetilde{X}$ is equal to X with probability at least $1-\delta$;*
2. *$\mathbb{E}[\widetilde{X}]=0$;*
3. *$\widetilde{X}$ is a bounded random variable: $|\widetilde{X}| \le C_\sigma\sigma\sqrt{\log(\delta^{-1})}$;*
4. *The variance of $\widetilde{X}$ obeys: $\mathsf{var}(\widetilde{X}) = (1+O(\sqrt{\delta}))\sigma^2$;*
5. *$\widetilde{X}$ is a sub-Gaussian random variable obeying $\|\widetilde{X}\|_{\psi_2}\lesssim\sigma$.*

I.1 Proof of Lemma 51

Step 1: lower bounding $\mathbb{E}|X|$. For any $t>0$, it is easily seen that

$$\mathbb{E}\left[X^2\,\mathbb{1}_{|X|>t}\right] \overset{\text{(i)}}{\le} \left(\mathbb{E}\left[X^4\right]\right)^{\frac{1}{2}}\left(\mathbb{P}\left(|X|>t\right)\right)^{\frac{1}{2}} \overset{\text{(ii)}}{\lesssim} \sigma^2\exp\left(-\frac{t^2}{C\sigma^2}\right),$$

where $C>0$ is some absolute constant. Here, (i) results from the Cauchy-Schwarz inequality, whereas (ii) follows from standard properties of sub-Gaussian random variables (Vershynin, 2017, Proposition 2.5.2). By taking $t=c_{\mathsf{lb}}^{-1}\sigma/4$ for some sufficiently small constant $c_{\mathsf{lb}}>0$, we can guarantee that $\mathbb{E}[X^2\,\mathbb{1}_{|X|>t}]\le\sigma^2/2$, which in turn results in

$$\mathbb{E}\left[X^2\,\mathbb{1}_{|X|\le t}\right]=\mathbb{E}\left[X^2\right]-\mathbb{E}\left[X^2\,\mathbb{1}_{|X|>t}\right]\ge\sigma^2-\frac{1}{2}\sigma^2\ge\frac{1}{2}\sigma^2.$$

As a consequence, we obtain (with the above choice $t=c_{\mathsf{lb}}^{-1}\sigma/4$)

$$\mathbb{E}\left[|X|\right]\ge\mathbb{E}\left[|X|\,\mathbb{1}_{|X|\le t}\right]\ge\frac{\mathbb{E}\left[X^2\,\mathbb{1}_{|X|\le t}\right]}{t}\ge 2c_{\mathsf{lb}}\sigma. \tag{I.3}$$

Step 2: constructing $\widetilde{X}$ by truncating X randomly. For notational simplicity, let us define $X^+=X\vee 0$ and $X^-=(-X)\vee 0$. Given that $\mathbb{E}[X]=0$, we have

$$\mathbb{E}\left[X^+\right]=\mathbb{E}\left[X^-\right]=\frac{1}{2}\mathbb{E}\left[|X|\right]\ge c_{\mathsf{lb}}\sigma, \tag{I.4}$$

where the last inequality comes from (I.3). Define a function $f:\mathbb{R}^+\mapsto\mathbb{R}^+$ as follows

$$f(x)\coloneqq\mathbb{E}\left[X\,\mathbb{1}_{X\ge x}\right].$$

It is straightforward to check that $f(x)$ is monotonically non-increasing within the domain $x\in[0,\infty)$. In view of the monotone convergence theorem, we know that $f(x)$ is a left continuous function with $\lim_{x\searrow 0}f(x)=\mathbb{E}[X^+]$ and $\lim_{x\to+\infty}f(x)=0$. In addition, for any $x\in\mathbb{R}^+$, one has

$$f(x)=\mathbb{E}\left[X\,\mathbb{1}_{X\ge x}\right]\overset{\text{(i)}}{\le}\left(\mathbb{E}\left[X^2\right]\right)^{\frac{1}{2}}\left(\mathbb{P}\left(X\ge x\right)\right)^{\frac{1}{2}}\overset{\text{(ii)}}{\le}\sigma\exp\left(-\frac{x^2}{C\sigma^2}\right), \tag{I.5}$$

where $C>0$ is some absolute constant. Here, (i) comes from the Cauchy-Schwarz inequality, while (ii) follows from standard properties of sub-Gaussian random variables (Vershynin, 2017, Proposition 2.5.2).

For any given $\varepsilon\in(0,c_{\mathsf{lb}}\sigma/2)$, we take

$$x_\varepsilon\coloneqq\sup\left\{x\in\mathbb{R}^+:f(x)\ge\varepsilon\right\}.$$

Since f is known to be left continuous, we know that

$$\lim_{x\nearrow x_\varepsilon}f(x)=f(x_\varepsilon)\ge\varepsilon\ge\lim_{x\searrow x_\varepsilon}f(x).$$

Taking the definition of x_ε and (I.5) collectively yields

$$\varepsilon\le f(x_\varepsilon)\le\sigma\exp\left(-\frac{x_\varepsilon^2}{C\sigma^2}\right),$$

which further gives an upper bound on x_ε as follows

$$x_\varepsilon\le\sigma\sqrt{C\log\left(\sigma/\varepsilon\right)}. \tag{I.6}$$

In addition, we can also lower bound x_ε by observing that

$$\mathbb{E}\left[X^+\right]=\mathbb{E}\left[X^+\,\mathbb{1}_{X\le x_\varepsilon}\right]+\mathbb{E}\left[X\,\mathbb{1}_{X>x_\varepsilon}\right]=\mathbb{E}\left[X^+\,\mathbb{1}_{X\le x_\varepsilon}\right]+\lim_{x\to x_\varepsilon^+}f(x)$$

$$\le x_\varepsilon + \varepsilon \le x_\varepsilon + \frac{1}{2}c_{\mathsf{lb}}\sigma,$$

which taken collectively with (I.4) yields

$$x_\varepsilon \ge \frac{1}{2}c_{\mathsf{lb}}\sigma. \tag{I.7}$$

With these calculations in place, we define $\widetilde{X}^+$ as follows:

- If $\lim_{x\searrow x_\varepsilon} f(x) = f(x_\varepsilon)$, we immediately know that $f(x_\varepsilon) = \varepsilon$. Then we can set

$$\widetilde{X}^+ \coloneqq X^+ \mathbb{1}_{X^+<x_\varepsilon}.$$

This construction gives $\widetilde{X}^+ \le x_\varepsilon \le \sigma\sqrt{C\log(\sigma/\varepsilon)}$ and

$$\mathbb{E}\left[\widetilde{X}^+\right] = \mathbb{E}\left[X^+\right] - f(x_\varepsilon) = \mathbb{E}\left[X^+\right] - \varepsilon.$$

We can also derive from (I.7) that

$$\mathbb{P}\left(\widetilde{X}^+ \ne X^+\right) = \mathbb{P}\left(X^+ \ge x_\varepsilon\right) \le \frac{\mathbb{E}\left[X^+ \mathbb{1}_{X^+\ge x_\varepsilon}\right]}{x_\varepsilon} = \frac{f(x_\varepsilon)}{x_\varepsilon} = \frac{\varepsilon}{x_\varepsilon} \le \frac{2\varepsilon}{c_{\mathsf{lb}}\sigma}.$$

- If $\lim_{x\searrow x_\varepsilon} f(x) < f(x_\varepsilon)$, we know that

$$\mathbb{E}\left[X^+ \mathbb{1}_{X^+=x_\varepsilon}\right] = f(x_\varepsilon) - \lim_{x\searrow x_\varepsilon} > 0. \tag{I.8}$$

Then one can set

$$\widetilde{X}^+ \coloneqq X^+ \mathbb{1}_{X^+<x_\varepsilon} + X \mathbb{1}_{X^+=x_\varepsilon} Q,$$

where Q is a Bernoulli random variable (independent of X) with parameter

$$q = \frac{f(x_\varepsilon) - \varepsilon}{f(x_\varepsilon) - \lim_{x\searrow x_\varepsilon}},$$

i.e., $\mathbb{P}(Q=1) = 1 - \mathbb{P}(Q=0) = q$. This construction gives $\widetilde{X}^+ \le x_\varepsilon \le \sigma\sqrt{C\log(\sigma/\varepsilon)}$ and

$$\mathbb{E}\left[\widetilde{X}^+\right] = \mathbb{E}\left[X^+ \mathbb{1}_{X^+<x_\varepsilon}\right] + q\mathbb{E}\left[X \mathbb{1}_{X^+=x_\varepsilon}\right] \tag{I.9}$$

$$= \mathbb{E}\left[X^+\right] - f(x_\varepsilon) + q\left[f(x_\varepsilon) - \lim_{x\to x_\varepsilon^+} f(x)\right]$$

$$= \mathbb{E}\left[X^+\right] - \varepsilon. \tag{I.10}$$

In addition, we have

$$\begin{aligned}\mathbb{P}\left(\widetilde{X}^+ \ne X^+\right) &= \mathbb{P}\left(X^+ > x_\varepsilon\right) + \mathbb{P}\left(X^+ = x_\varepsilon, Q = 0\right) = \mathbb{P}\left(X^+ > x_\varepsilon\right) + (1-q)\,\mathbb{P}\left(X^+ = x_\varepsilon\right) \\ &\overset{\text{(i)}}{\le} \frac{\mathbb{E}\left[X^+ \mathbb{1}_{X^+>x_\varepsilon}\right]}{x_\varepsilon} + (1-q)\frac{\mathbb{E}\left[X^+ \mathbb{1}_{X^+=x_\varepsilon}\right]}{x_\varepsilon} \\ &= \frac{\mathbb{E}\left[X^+\right] - \mathbb{E}\left[X^+ \mathbb{1}_{X^+<x_\varepsilon}\right] - q\mathbb{E}\left[X^+ \mathbb{1}_{X^+=x_\varepsilon}\right]}{x_\varepsilon} \overset{\text{(ii)}}{=} \frac{\mathbb{E}\left[X^+\right] - \mathbb{E}\left[\widetilde{X}^+\right]}{x_\varepsilon} \\ &\overset{\text{(iii)}}{=} \frac{\varepsilon}{x_\varepsilon} \overset{\text{(iv)}}{\le} \frac{2\varepsilon}{c_{\mathsf{lb}}\sigma}.\end{aligned}$$

Here, (i) holds since $\mathbb{E}[X^+ \mathbb{1}_{X^+>x_\varepsilon}] \ge x_\varepsilon\mathbb{P}(X^+ > x_\varepsilon)$ and $\mathbb{E}[X^+ \mathbb{1}_{X^+=x_\varepsilon}] = x_\varepsilon \cdot \mathbb{P}(X^+ = x_\varepsilon)$; (ii) follows from (I.9); (iii) is a consequence of (I.10); and (iv) follows from (I.7).

We have thus constructed a random variable $\widetilde{X}^+$ that satisfies: (i) $\widetilde{X}^+$ equals either X^+ or 0; (ii) $\mathbb{P}(\widetilde{X}^+ \ne X^+) \le 2\varepsilon/(c_{\mathsf{lb}}\sigma)$; (iii) $\mathbb{E}[\widetilde{X}^+] = \mathbb{E}[X^+] - \varepsilon$; and (iv) $0 \le \widetilde{X}^+ \le \sigma\sqrt{C\log(\sigma/\varepsilon)}$. Similarly, we can also construct another random variable $\widetilde{X}^-$ satisfying: (i) $\widetilde{X}^-$ equals either X^- or 0; (ii) $\mathbb{P}(\widetilde{X}^- \ne X^-) \le 2\varepsilon/(c_{\mathsf{lb}}\sigma)$; (iii) $\mathbb{E}[\widetilde{X}^-] = \mathbb{E}[X^-] - \varepsilon$; and (iv) $0 \le \widetilde{X}^- \le \sigma\sqrt{C\log(\sigma/\varepsilon)}$. We shall then construct $\widetilde{X}$ as follows

$$\widetilde{X} \coloneqq \widetilde{X}^+ - \widetilde{X}^-.$$

Step 3: verifying the advertised properties of $\widetilde{X}$. To finish up, we can check that the following properties are satisfied:

1. $\widetilde{X}$ has mean zero, namely,

$$\mathbb{E}\left[\widetilde{X}\right]=\mathbb{E}\left[\widetilde{X}^{+}\right]-\mathbb{E}\left[\widetilde{X}^{-}\right]=\mathbb{E}\left[X^{+}\right]-\varepsilon-\mathbb{E}\left[X^{-}\right]+\varepsilon=\mathbb{E}\left[X\right]=0.$$

2. $\widetilde{X}$ is identical to X with high probability, namely,

$$\mathbb{P}\left(X\neq\widetilde{X}\right)=\mathbb{P}\left(X^{+}\neq\widetilde{X}^{+}\right)+\mathbb{P}\left(X^{-}\neq\widetilde{X}^{-}\right)\leq\frac{4\varepsilon}{c_{\mathsf{lb}}\sigma}.$$

3. $\widetilde{X}$ is a bounded random variable in the sense that $|\widetilde{X}|\leq\sigma\sqrt{C\log(\sigma/\varepsilon)}\lesssim\sigma\sqrt{\log(\sigma/\varepsilon)}$.

4. The variance of $\widetilde{X}$ is close to σ^2 in the sense that

$$\mathsf{var}(\widetilde{X})=\mathbb{E}\left[\widetilde{X}^{2}\right]=\mathbb{E}\left[X^{2}\right]-\mathbb{E}\left[X^{2}\,\mathbb{1}_{X\neq\widetilde{X}}\right]=\left(1+O\left(\sqrt{\varepsilon/\sigma}\right)\right)\sigma^{2},$$

   where the last relation holds due to the following observation

$$\mathbb{E}\left[X^{2}\,\mathbb{1}_{X\neq\widetilde{X}}\right]\overset{\text{(i)}}{\leq}\left(\mathbb{E}\left[X^{4}\right]\right)^{\frac{1}{2}}\left(\mathbb{P}\left(X\neq\widetilde{X}\right)\right)^{\frac{1}{2}}\overset{\text{(ii)}}{\lesssim}\sigma^{2}\sqrt{\frac{\varepsilon}{\sigma}}.$$

   Here, (i) invokes Cauchy-Schwarz, whereas (ii) is valid due to standard properties of sub-Gaussian random variables (Vershynin, 2017, Proposition 2.5.2).

5. By construction, we can see that for any $t\geq0$,

$$\mathbb{P}\left(\left|\widetilde{X}\right|\geq t\right)\leq\mathbb{P}\left(|X|\geq t\right)\leq2\exp\left(-\frac{t^{2}}{\widetilde{c}\sigma^{2}}\right)$$

   holds for some absolute constant $\widetilde{c}>0$, where the last relation follows from Vershynin (2017, Proposition 2.5.2). By invoking the definition of sub-Gaussian random variables (Vershynin, 2017, Definition 2.5.6) as well as standard properties of sub-Gaussian random variables (Vershynin, 2017, Proposition 2.5.2), we can conclude that $\widetilde{X}$ is sub-Gaussian obeying $\|\widetilde{X}\|_{\psi_2}\lesssim\sigma$.

By taking $\varepsilon=\delta c_{\mathsf{lb}}\sigma/4$ for any $\delta\in(0,2)$, we establish the desired result.

References

Abbe, E., Fan, J., Wang, K., Zhong, Y., et al. (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *Annals of Statistics*, 48(3):1452–1474.

Agterberg, J., Lubberts, Z., and Priebe, C. (2021). Entrywise estimation of singular vectors of low-rank matrices with heteroskedasticity and dependence. *arXiv preprint arXiv:2105.13346*.

Bai, J. and Wang, P. (2016). Econometric analysis of large factor models. *Annual Review of Economics*, 8:53–80.

Balzano, L., Chi, Y., and Lu, Y. M. (2018). Streaming PCA and subspace tracking: The missing data case. *Proceedings of the IEEE*, 106(8):1293–1310.

Bao, Z., Ding, X., Wang, J., and Wang, K. (2020). Statistical inference for principal components of spiked covariance matrix. *arXiv preprint arXiv:2008.11903*.

Bao, Z., Ding, X., and Wang, K. (2018). Singular vector and singular subspace distribution for the matrix denoising model. *accepted to Annals of Statistics*.

Bloemendal, A., Knowles, A., Yau, H.-T., and Yin, J. (2016). On the principal components of sample covariance matrices. *Probability theory and related fields*, 164(1-2):459–552.

Cai, C., Li, G., Chi, Y., Poor, H. V., and Chen, Y. (2021). Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees. *The Annals of Statistics*, 49(2):944–967.

Cai, C., Li, G., Poor, H. V., and Chen, Y. (2020a). Nonconvex low-rank tensor completion from noisy data. *Accepted to Operations Research*.

Cai, C., Poor, H. V., and Chen, Y. (2020b). Uncertainty quantification for nonconvex tensor completion: Confidence intervals, heteroscedasticity and optimality. In *International Conference on Machine Learning*, pages 1271–1282.

Cai, T. T. and Guo, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *The Annals of statistics*, 45(2):615–646.

Cai, T. T. and Zhang, A. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *The Annals of Statistics*, 46(1):60–89.

Candès, E. (2014). Mathematics of sparsity (and a few other things). In *Proceedings of the International Congress of Mathematicians, Seoul, South Korea*, volume 123.

Candès, E. and Plan, Y. (2010). Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925 –936.

Candès, E. and Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772.

Cape, J., Tang, M., and Priebe, C. E. (2019). The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *The Annals of Statistics*, 47(5):2405–2439.

Celentano, M., Montanari, A., and Wei, Y. (2020). The Lasso with general Gaussian designs with applications to hypothesis testing. *arXiv preprint arXiv:2007.13716*.

Chen, J., Liu, D., and Li, X. (2020a). Nonconvex rectangular matrix completion via gradient descent without $\ell_{2,\infty}$ regularization. *IEEE Transactions on Information Theory*.

Chen, L. H., Goldstein, L., and Shao, Q.-M. (2010). *Normal approximation by Stein's method*. Springer Science & Business Media.

Chen, P., Gao, C., and Zhang, A. Y. (2020b). Partial recovery for top-k ranking: Optimality of MLE and sub-optimality of spectral method. *arXiv preprint arXiv:2006.16485*.

Chen, Y., Cheng, C., and Fan, J. (2021). Asymmetry helps: Eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices. *Annals of Statistics*, 49(1):435–458.

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

Chen, Y., Chi, Y., Fan, J., and Ma, C. (2020c). Spectral methods for data science: A statistical perspective. *arXiv preprint arXiv:2012.08496*.

Chen, Y., Chi, Y., Fan, J., Ma, C., and Yan, Y. (2020d). Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *SIAM Journal on Optimization*, 30(4):3098–3121.

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

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

Chen, Y., Fan, J., Ma, C., and Yan, Y. (2020e). Bridging convex and nonconvex optimization in robust pca: Noise, outliers, and missing data. *arXiv preprint arXiv:2001.05484, accepted to The Annals of Statistics*.

Chen, Y. and Wainwright, M. J. (2015). Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees. *arXiv:1509.03025*.

Cheng, C., Wei, Y., and Chen, Y. (2020). Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise. *arXiv preprint arXiv:2001.04620*.

Chernozhukov, V., Hansen, C., Liao, Y., and Zhu, Y. (2021). Inference for low-rank models. *arXiv preprint arXiv:2107.02602*.

Chi, Y., Lu, Y. M., and Chen, Y. (2019). Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Transactions on Signal Processing*, 67(20):5239 – 5269.

Cho, J., Kim, D., and Rohe, K. (2017). Asymptotic theory for estimating the singular vectors and values of a partially-observed low rank matrix with noise. *Statistica Sinica*, pages 1921–1948.

Devroye, L., Mehrabian, A., and Reddad, T. (2018). The total variation distance between high-dimensional gaussians. *arXiv preprint arXiv:1810.08693*.

Ding, X. (2020). High dimensional deformed rectangular matrices with applications in matrix denoising. *Bernoulli*, 26(1):387–417.

El Karoui, N. (2015). 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.

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

Eldridge, J., Belkin, M., and Wang, Y. (2018). Unperturbed: spectral analysis beyond Davis-Kahan. In *Algorithmic Learning Theory*, pages 321–358.

Fan, J., Fan, Y., Han, X., and Lv, J. (2020). Asymptotic theory of eigenvectors for random matrices with diverging spikes. *Journal of the American Statistical Association*, pages 1–14.

Fan, J., Li, K., and Liao, Y. (2021a). Recent developments on factor models and its applications in econometric learning. *Annual Review of Financial Economics*, 13:to appear.

Fan, J., Wang, K., Zhong, Y., and Zhu, Z. (2021b). Robust high-dimensional factor models with applications to statistical machine learning. *Statistical Science*, 36(2):303–327.

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

Fan, J. and Yao, Q. (2017). *The elements of financial econometrics*. Cambridge University Press.

Florescu, L. and Perkins, W. (2016). Spectral thresholds in the bipartite stochastic block model. In *Conference on Learning Theory*, pages 943–959.

Gagliardini, P., Ossola, E., and Scaillet, O. (2019). Estimation of large dimensional conditional factor models in finance. *Swiss Finance Institute Research Paper*, (19-46).

Javanmard, A. and Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research*, 15(1):2869–2909.

Johnstone, I. M. and Paul, D. (2018). PCA in high dimensions: An orientation. *Proceedings of the IEEE*, 106(8):1277–1292.

Jolliffe, I. T. (1986). Principal components in regression analysis. In *Principal component analysis*, pages 129–155. Springer.

Keshavan, R. H., Montanari, A., and Oh, S. (2010a). Matrix completion from a few entries. *IEEE Transactions on Information Theory*, 56(6):2980 –2998.

Keshavan, R. H., Montanari, A., and Oh, S. (2010b). Matrix completion from noisy entries. *J. Mach. Learn. Res.*, 11:2057–2078.

Koltchinskii, V., Löffler, M., and Nickl, R. (2020). Efficient estimation of linear functionals of principal components. *The Annals of Statistics*, 48(1):464–490.

Lei, L. (2019). Unified $\ell_{2,\infty}$ eigenspace perturbation theory for symmetric random matrices. *arXiv preprint arXiv:1909.04798*.

Li, G., Cai, C., Gu, Y., Poor, H. V., and Chen, Y. (2021). Minimax estimation of linear functions of eigenvectors in the face of small eigen-gaps. *arXiv preprint arXiv:2104.03298*.

Ling, S. (2020). Near-optimal performance bounds for orthogonal and permutation group synchronization via spectral methods. *arXiv preprint arXiv:2008.05341*.

Löffler, M., Zhang, A. Y., and Zhou, H. H. (2019). Optimality of spectral clustering in the Gaussian mixture model. *arXiv preprint arXiv:1911.00538*.

Loh, P.-L. and Wainwright, M. J. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *The Annals of Statistics*, 40(3):1637–1664.

Lounici, K. (2014). High-dimensional covariance matrix estimation with missing observations. *Bernoulli*, 20(3):1029–1058.

Ma, C., Wang, K., Chi, Y., and Chen, Y. (2020). Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion, and blind deconvolution. *Foundations of Computational Mathematics*, 20(3):451–632.

Montanari, A., Ruan, F., and Yan, J. (2018). Adapting to unknown noise distribution in matrix denoising. *arXiv preprint arXiv:1810.02954*.

Montanari, A. and Sun, N. (2018). Spectral algorithms for tensor completion. *Communications on Pure and Applied Mathematics*, 71(11):2381–2425.

Negahban, S. and Wainwright, M. (2012). Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *Journal of Machine Learning Research*, pages 1665–1697.

Ning, Y. and Liu, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *The Annals of Statistics*, 45(1):158–195.

Pavez, E. and Ortega, A. (2020). Covariance matrix estimation with non uniform and data dependent missing observations. *IEEE Transactions on Information Theory*.

Raič, M. (2019). A multivariate Berry-Esseen theorem with explicit constants. *Bernoulli*, 25(4A):2824–2853.

Ren, Z., Sun, T., Zhang, C., and Zhou, H. (2015). Asymptotic normality and optimalities in estimation of large gaussian graphical models. *The Annals of Statistics*, 43(3):991–1026.

Rohe, K., Chatterjee, S., and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915.

Schmitt, B. A. (1992). Perturbation bounds for matrix square roots and Pythagorean sums. *Linear Algebra Appl.*, 174:215–227.

Sun, R. and Luo, Z.-Q. (2016). Guaranteed matrix completion via non-convex factorization. *IEEE Transactions on Information Theory*, 62(11):6535–6579.

Tropp, J. A. (2015). An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.*, 8(1-2):1–230.

Tsybakov, A. B. and Zaiats, V. (2009). *Introduction to nonparametric estimation*, volume 11. Springer.

van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202.

Vaswani, N., Chi, Y., and Bouwmans, T. (2018). Rethinking PCA for modern data sets: Theory, algorithms, and applications. *Proceedings of the IEEE*, 106(8):1274–1276.

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

Vershynin, R. (2017). High dimensional probability.

Wainwright, M. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wedin, P.-Å. (1972). Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, 12(1):99–111.

Xia, D. (2019a). Confidence region of singular subspaces for low-rank matrix regression. *IEEE Transactions on Information Theory*, 65(11):7437–7459.

Xia, D. (2019b). Normal approximation and confidence region of singular subspaces. *arXiv preprint arXiv:1901.00304*.

Xia, D. and Yuan, M. (2021). Statistical inferences of linear forms for noisy matrix completion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 83(1):58–77.

Xie, F. (2021). Entrywise limit theorems of eigenvectors and their one-step refinement for sparse random graphs. *arXiv preprint arXiv:2106.09840*.

Zhang, A., Cai, T. T., and Wu, Y. (2018). Heteroskedastic PCA: Algorithm, optimality, and applications. *accepted to Annals of Statistics*.

Zhang, C.-H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B*, 76(1):217–242.

Zheng, Q. and Lafferty, J. (2016). Convergence analysis for rectangular matrix completion using Burer-Monteiro factorization and gradient descent. *arXiv:1605.07051*.

Zhong, Y. and Boumal, N. (2018). Near-optimal bound for phase synchronization. *SIAM Journal on Optimization*.

Zhu, Z., Wang, T., and Samworth, R. J. (2019). High-dimensional principal component analysis with heterogeneous missingness. *arXiv preprint arXiv:1906.12125*.