Supplementary material for "Implicit Regularization in Nonconvex Statistical Estimation: Gradient Descent Converges Linearly for Phase Retrieval and Matrix Completion"

Contents

1 A general recipe for trajectory analysis

In this section, we sketch a general recipe for establishing performance guarantees of gradient descent, which conveys the key idea for proving the main results of this paper. The main challenge is to demonstrate that appropriate incoherence conditions are preserved throughout the trajectory of the algorithm. This requires exploiting statistical independence of the samples in a careful manner, in conjunction with generic optimization theory. Central to our approach is a leave-one-out perturbation argument, which allows to decouple the statistical dependency while controlling the component-wise incoherence measures.

General Recipe (a leave-one-out analysis)

Step 1: characterize restricted strong convexity and smoothness of f, and identify the region of incoherence and contraction (RIC).

Step 2: introduce leave-one-out sequences $\{\boldsymbol{X}^{t,(l)}\}$ and $\{\boldsymbol{H}^{t,(l)}\}$ for each l, where $\{\boldsymbol{X}^{t,(l)}\}$ (resp. $\{\boldsymbol{H}^{t,(l)}\}$) is independent of any sample involving $\boldsymbol{\phi}_l$ (resp. $\boldsymbol{\psi}_l$);

Step 3: establish the incoherence condition for $\{\boldsymbol{X}^t\}$ and $\{\boldsymbol{H}^t\}$ via induction. Suppose the iterates satisfy the claimed conditions in the tth iteration:

(a) show, via restricted strong convexity, that the true iterates $(\boldsymbol{X}^{t+1}, \boldsymbol{H}^{t+1})$ and the leave-one-out version $(\boldsymbol{X}^{t+1,(l)}, \boldsymbol{H}^{t+1,(l)})$ are exceedingly close;

(b) use statistical independence to show that $\boldsymbol{X}^{t+1,(l)} - \boldsymbol{X}^\natural$ (resp. $\boldsymbol{H}^{t+1,(l)} - \boldsymbol{H}^\natural$) is incoherent w.r.t. $\boldsymbol{\phi}_l$ (resp. $\boldsymbol{\psi}_l$), namely, $\|\boldsymbol{\phi}_l^*(\boldsymbol{X}^{t+1,(l)} - \boldsymbol{X}^\natural)\|_2$ and $\|\boldsymbol{\psi}_l^*(\boldsymbol{H}^{t+1,(l)} - \boldsymbol{H}^\natural)\|_2$ are both well-controlled;

(c) combine the bounds to establish the desired incoherence condition concerning $\max_l \|\boldsymbol{\phi}_l^*(\boldsymbol{X}^{t+1} - \boldsymbol{X}^\natural)\|_2$ and $\max_l \|\boldsymbol{\psi}_l^*(\boldsymbol{H}^{t+1} - \boldsymbol{H}^\natural)\|_2$.

1.1 General model

Consider the following problem where the samples are collected in a bilinear/quadratic form as

$$y_j = \boldsymbol{\psi}_j^* \boldsymbol{H}^\natural \boldsymbol{X}^{\natural *} \boldsymbol{\phi}_j, \qquad 1 \le j \le m, \tag{1}$$

where the objects of interest $\boldsymbol{H}^\natural, \boldsymbol{X}^\natural \in \mathbb{C}^{n\times r}$ or $\mathbb{R}^{n\times r}$ might be vectors or tall matrices taking either real or complex values. The design vectors $\{\boldsymbol{\psi}_j\}$ and $\{\boldsymbol{\phi}_j\}$ are in either $\mathbb{C}^n$ or $\mathbb{R}^n$, and can be either random or deterministic. This model is quite general and entails all three examples in this paper as special cases:

- *Phase retrieval*: $\boldsymbol{H}^\natural = \boldsymbol{X}^\natural = \boldsymbol{x}^\natural \in \mathbb{R}^n$, and $\boldsymbol{\psi}_j = \boldsymbol{\phi}_j = \boldsymbol{a}_j$;
- *Matrix completion*: $\boldsymbol{H}^\natural = \boldsymbol{X}^\natural \in \mathbb{R}^{n\times r}$ and $\boldsymbol{\psi}_j, \boldsymbol{\phi}_j \in \{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n\}$;
- *Blind deconvolution*: $\boldsymbol{H}^\natural = \boldsymbol{h}^\natural \in \mathbb{C}^K$, $\boldsymbol{X}^\natural = \boldsymbol{x}^\natural \in \mathbb{C}^K$, $\boldsymbol{\phi}_j = \boldsymbol{a}_j$, and $\boldsymbol{\psi}_j = \boldsymbol{b}_j$.

For this setting, the empirical loss function is given by

$$f(\boldsymbol{Z}) := f(\boldsymbol{H}, \boldsymbol{X}) = \frac{1}{m}\sum_{j=1}^{m} \left| \boldsymbol{\psi}_j^* \boldsymbol{H}\boldsymbol{X}^* \boldsymbol{\phi}_j - y_j \right|^2,$$

where we denote $\boldsymbol{Z} = (\boldsymbol{H}, \boldsymbol{X})$. To minimize $f(\boldsymbol{Z})$, we proceed with vanilla gradient descent

$$\boldsymbol{Z}^{t+1} = \boldsymbol{Z}^t - \eta \nabla f(\boldsymbol{Z}^t), \qquad \forall t \ge 0$$

following a standard spectral initialization, where η is the step size. As a remark, for complex-valued problems, the gradient (resp. Hessian) should be understood as the Wirtinger gradient (resp. Hessian).

It is clear from (1) that $\boldsymbol{Z}^\natural = (\boldsymbol{H}^\natural, \boldsymbol{X}^\natural)$ can only be recovered up to certain global ambiguity. For clarity of presentation, we assume in this section that such ambiguity has already been taken care of via proper global transformation.

1.2 Outline of the recipe

We are now positioned to outline the general recipe, which entails the following steps.

- **Step 1: characterizing local geometry in the RIC.** Our first step is to characterize a region $\mathcal{R}$ — which we term as the *region of incoherence and contraction* (RIC) — such that the Hessian matrix $\nabla^2 f(\boldsymbol{Z})$ obeys strong convexity and smoothness,

$$\mathbf{0} \prec \alpha \boldsymbol{I} \preceq \nabla^2 f(\boldsymbol{Z}) \preceq \beta \boldsymbol{I}, \qquad \forall \boldsymbol{Z} \in \mathcal{R}, \tag{2}$$

  or at least along certain directions (i.e. restricted strong convexity and smoothness), where β/α scales slowly (or even remains bounded) with the problem size. As revealed by optimization theory, this geometric property (2) immediately implies linear convergence with the contraction rate $1 - O(\alpha/\beta)$ for a properly chosen step size η, as long as all iterates stay within the RIC.

  A natural question then arises: what does the RIC $\mathcal{R}$ look like? As it turns out, the RIC typically contains all points such that the ℓ_2 error $\|\boldsymbol{Z} - \boldsymbol{Z}^\natural\|_\mathrm{F}$ is not too large and

$$\textbf{(incoherence)} \qquad \max_j \left\| \boldsymbol{\phi}_j^* (\boldsymbol{X} - \boldsymbol{X}^\natural) \right\|_2 \text{ and } \max_j \left\| \boldsymbol{\psi}_j^* (\boldsymbol{H} - \boldsymbol{H}^\natural) \right\|_2 \text{ are well-controlled.} \tag{3}$$

  In the three examples, the above incoherence condition translates to:

  – *Phase retrieval*: $\max_j \left| \boldsymbol{a}_j^\top (\boldsymbol{x} - \boldsymbol{x}^\natural) \right|$ is well-controlled;

  – *Matrix completion*: $\left\| \boldsymbol{X} - \boldsymbol{X}^\natural \right\|_{2,\infty}$ is well-controlled;

  – *Blind deconvolution*: $\max_j \left| \boldsymbol{a}_j^\top (\boldsymbol{x} - \boldsymbol{x}^\natural) \right|$ and $\max_j \left| \boldsymbol{b}_j^\top (\boldsymbol{h} - \boldsymbol{h}^\natural) \right|$ are well-controlled.

- **Step 2: introducing the leave-one-out sequences.** To justify that no iterates leave the RIC, we rely on the construction of auxiliary sequences. Specifically, for each l, produce an auxiliary sequence $\{\boldsymbol{Z}^{t,(l)} = (\boldsymbol{X}^{t,(l)}, \boldsymbol{H}^{t,(l)})\}$ such that $\boldsymbol{X}^{t,(l)}$ (resp. $\boldsymbol{H}^{t,(l)}$) is independent of any sample involving $\boldsymbol{\phi}_l$ (resp. $\boldsymbol{\psi}_l$). As an example, suppose that the $\boldsymbol{\phi}_l$'s and the $\boldsymbol{\psi}_l$'s are independently and randomly generated. Then for each l, one can consider a leave-one-out loss function

$$f^{(l)}(\boldsymbol{Z}) := \frac{1}{m} \sum_{j: j \neq l} \left| \boldsymbol{\psi}_j^* \boldsymbol{H} \boldsymbol{X}^* \boldsymbol{\phi}_j - y_j \right|^2$$

  that discards the lth sample. One further generates $\{\boldsymbol{Z}^{t,(l)}\}$ by running vanilla gradient descent w.r.t. this auxiliary loss function, with a spectral initialization that similarly discards the lth sample. Note that this procedure is only introduced to facilitate analysis and is never implemented in practice.

- **Step 3: establishing the incoherence condition.** We are now ready to establish the incoherence condition with the assistance of the auxiliary sequences. Usually the proof proceeds by induction, where our goal is to show that the next iterate remains within the RIC, given that the current one does.

  – **Step 3(a): proximity between the original and the leave-one-out iterates.** As one can anticipate, $\{\boldsymbol{Z}^t\}$ and $\{\boldsymbol{Z}^{t,(l)}\}$ remain "glued" to each other along the whole trajectory, since their constructions differ by only a single sample. In fact, as long as the initial estimates stay sufficiently close, their gaps will never explode. To intuitively see why, use the fact $\nabla f(\boldsymbol{Z}^t) \approx \nabla f^{(l)}(\boldsymbol{Z}^t)$ to discover that

$$\begin{aligned} \boldsymbol{Z}^{t+1} - \boldsymbol{Z}^{t+1,(l)} &= \boldsymbol{Z}^t - \eta \nabla f(\boldsymbol{Z}^t) - \big(\boldsymbol{Z}^{t,(l)} - \eta \nabla f^{(l)} \big(\boldsymbol{Z}^{t,(l)} \big) \big) \\ &\approx \boldsymbol{Z}^t - \boldsymbol{Z}^{t,(l)} - \eta \nabla^2 f(\boldsymbol{Z}^t) \big(\boldsymbol{Z}^t - \boldsymbol{Z}^{t,(l)} \big), \end{aligned}$$

  which together with the strong convexity condition implies ℓ_2 contraction

$$\left\| \boldsymbol{Z}^{t+1} - \boldsymbol{Z}^{t+1,(l)} \right\|_\mathrm{F} \approx \left\| \big(\boldsymbol{I} - \eta \nabla^2 f(\boldsymbol{Z}^t) \big) \big(\boldsymbol{Z}^t - \boldsymbol{Z}^{t,(l)} \big) \right\|_\mathrm{F} \leq \left\| \boldsymbol{Z}^t - \boldsymbol{Z}^{t,(l)} \right\|_2.$$

  Indeed, (restricted) strong convexity is crucial in controlling the size of leave-one-out perturbations.

– **Step 3(b): incoherence condition of the leave-one-out iterates.** The fact that $\boldsymbol{Z}^{t+1}$ and $\boldsymbol{Z}^{t+1,(l)}$ are exceedingly close motivates us to control the incoherence of $\boldsymbol{Z}^{t+1,(l)} - \boldsymbol{Z}^{\natural}$ instead, for $1 \leq l \leq m$. By construction, $\boldsymbol{X}^{t+1,(l)}$ (resp. $\boldsymbol{H}^{t+1,(l)}$) is statistically *independent* of any sample involving the design vector $\boldsymbol{\phi}_l$ (resp. $\boldsymbol{\psi}_l$), a fact that typically leads to a more friendly analysis for controlling $\left\|\boldsymbol{\phi}_l^*\left(\boldsymbol{X}^{t+1,(l)} - \boldsymbol{X}^{\natural}\right)\right\|_2$ and $\left\|\boldsymbol{\psi}_l^*\left(\boldsymbol{H}^{t+1,(l)} - \boldsymbol{H}^{\natural}\right)\right\|_2$.

– **Step 3(c): combining the bounds.** With these results in place, apply the triangle inequality to obtain

$$\left\|\boldsymbol{\phi}_l^*\left(\boldsymbol{X}^{t+1} - \boldsymbol{X}^{\natural}\right)\right\|_2 \leq \|\boldsymbol{\phi}_l\|_2 \left\|\boldsymbol{X}^{t+1} - \boldsymbol{X}^{t+1,(l)}\right\|_{\mathrm{F}} + \left\|\boldsymbol{\phi}_l^*\left(\boldsymbol{X}^{t+1,(l)} - \boldsymbol{X}^{\natural}\right)\right\|_2,$$

where the first term is controlled in Step 3(a) and the second term is controlled in Step 3(b). The term $\left\|\boldsymbol{\psi}_l^*\left(\boldsymbol{H}^{t+1} - \boldsymbol{H}^{\natural}\right)\right\|_2$ can be bounded similarly. By choosing the bounds properly, this establishes the incoherence condition for all $1 \leq l \leq m$ as desired.

2 Analysis for phase retrieval

In this section, we instantiate the general recipe presented in Section 1 to phase retrieval and prove Theorem 1. Similar to the Section 7.1 in [?], we are going to use $\eta_t = c_1/(\log n \cdot \|\boldsymbol{x}^{\natural}\|_2^2)$ instead of $c_1/(\log n \cdot \|\boldsymbol{x}_0\|_2^2)$ as the step size for analysis. This is because with high probability, $\|\boldsymbol{x}_0\|_2$ and $\|\boldsymbol{x}^{\natural}\|_2$ are rather close in the relative sense. Without loss of generality, we assume throughout this section that $\left\|\boldsymbol{x}^{\natural}\right\|_2 = 1$ and

$$\mathrm{dist}(\boldsymbol{x}^0, \boldsymbol{x}^{\natural}) = \|\boldsymbol{x}^0 - \boldsymbol{x}^{\natural}\|_2 \leq \|\boldsymbol{x}^0 + \boldsymbol{x}^{\natural}\|_2. \tag{4}$$

In addition, the gradient and the Hessian of $f(\cdot)$ for this problem (see (13)) are given respectively by

$$\nabla f(\boldsymbol{x}) = \frac{1}{m}\sum_{j=1}^{m}\left[\left(\boldsymbol{a}_j^{\top}\boldsymbol{x}\right)^2 - y_j\right]\left(\boldsymbol{a}_j^{\top}\boldsymbol{x}\right)\boldsymbol{a}_j, \tag{5}$$

$$\nabla^2 f(\boldsymbol{x}) = \frac{1}{m}\sum_{j=1}^{m}\left[3\left(\boldsymbol{a}_j^{\top}\boldsymbol{x}\right)^2 - y_j\right]\boldsymbol{a}_j\boldsymbol{a}_j^{\top}, \tag{6}$$

which are useful throughout the proof.

2.1 Step 1: characterizing local geometry in the RIC

2.1.1 Local geometry

We start by characterizing the region that enjoys both strong convexity and the desired level of smoothness. This is supplied in the following lemma, which plays a crucial role in the subsequent analysis.

Lemma 1 (Restricted strong convexity and smoothness for phase retrieval)**.** *Fix any sufficiently small constant $C_1 > 0$ and any sufficiently large constant $C_2 > 0$, and suppose the sample complexity obeys $m \geq c_0 n \log n$ for some sufficiently large constant $c_0 > 0$. With probability at least $1 - O(mn^{-10})$,*

$$\nabla^2 f(\boldsymbol{x}) \succeq (1/2) \cdot \boldsymbol{I}_n$$

holds simultaneously for all $\boldsymbol{x} \in \mathbb{R}^n$ satisfying $\left\|\boldsymbol{x} - \boldsymbol{x}^{\natural}\right\|_2 \leq 2C_1$; and

$$\nabla^2 f(\boldsymbol{x}) \preceq (5C_2(10 + C_2)\log n) \cdot \boldsymbol{I}_n$$

holds simultaneously for all $\boldsymbol{x} \in \mathbb{R}^n$ obeying

$$\left\|\boldsymbol{x} - \boldsymbol{x}^{\natural}\right\|_2 \leq 2C_1, \tag{7a}$$

$$\max_{1\leq j\leq m}\left|\boldsymbol{a}_j^{\top}\left(\boldsymbol{x} - \boldsymbol{x}^{\natural}\right)\right| \leq C_2\sqrt{\log n}. \tag{7b}$$

Proof. See Appendix 4.1. □

In words, Lemma 1 reveals that the Hessian matrix is positive definite and (almost) well-conditioned, if one restricts attention to the set of points that are (i) not far away from the truth (cf. (7a)) and (ii) incoherent with respect to the measurement vectors $\{\boldsymbol{a}_j\}_{1\leq j\leq m}$ (cf. (7b)).

2.1.2 Error contraction

As we point out before, the nice local geometry enables ℓ_2 contraction, which we formalize below.

Lemma 2. *With probability exceeding* $1 - O(mn^{-10})$, *one has*

$$\left\|\boldsymbol{x}^{t+1} - \boldsymbol{x}^{\natural}\right\|_2 \leq (1 - \eta/2) \left\|\boldsymbol{x}^{t} - \boldsymbol{x}^{\natural}\right\|_2 \tag{8}$$

for any $\boldsymbol{x}^t$ *obeying the conditions* (7), *provided that the step size satisfies* $0 < \eta \leq 1/\left[5C_2\left(10 + C_2\right)\log n\right]$.

Proof. This proof applies the standard argument when establishing the ℓ_2 error contraction of gradient descent for strongly convex and smooth functions. See Appendix 4.2. □

With the help of Lemma 2, we can turn the proof of Theorem 1 into ensuring that the trajectory $\{\boldsymbol{x}^t\}_{0\leq t\leq n}$ lies in the RIC specified by (9).[1] This is formally stated in the next lemma.

Lemma 3. *Suppose for all* $0 \leq t \leq T_0 := n$, *the trajectory* $\{\boldsymbol{x}^t\}$ *falls within the region of incoherence and contraction (termed the RIC), namely,*

$$\left\|\boldsymbol{x}^{t} - \boldsymbol{x}^{\natural}\right\|_2 \leq C_1, \tag{9a}$$

$$\max_{1\leq l\leq m} \left|\boldsymbol{a}_l^{\top}\left(\boldsymbol{x}^{t} - \boldsymbol{x}^{\natural}\right)\right| \leq C_2\sqrt{\log n}, \tag{9b}$$

then the claims in Theorem 1 hold true. Here and throughout this section, $C_1, C_2 > 0$ *are two absolute constants as specified in Lemma 1.*

Proof. See Appendix 4.3. □

2.2 Step 2: introducing the leave-one-out sequences

In comparison to the ℓ_2 error bound (9a) that captures the overall loss, the incoherence hypothesis (9b) — which concerns sample-wise control of the empirical risk — is more complicated to establish. This is partly due to the statistical dependence between $\boldsymbol{x}^t$ and the sampling vectors $\{\boldsymbol{a}_l\}$. As described in the general recipe, the key idea is the introduction of a *leave-one-out* version of the WF iterates, which removes a single measurement from consideration.

To be precise, for each $1 \leq l \leq m$, we define the leave-one-out empirical loss function as

$$f^{(l)}(\boldsymbol{x}) := \frac{1}{4m}\sum_{j:j\neq l}\left[\left(\boldsymbol{a}_j^{\top}\boldsymbol{x}\right)^2 - y_j\right]^2, \tag{10}$$

and the auxiliary trajectory $\left\{\boldsymbol{x}^{t,(l)}\right\}_{t\geq 0}$ is constructed by running WF w.r.t. $f^{(l)}(\boldsymbol{x})$. In addition, the spectral initialization $\boldsymbol{x}^{0,(l)}$ is computed based on the rescaled leading eigenvector of the leave-one-out data matrix

$$\boldsymbol{Y}^{(l)} := \frac{1}{m}\sum_{j:j\neq l} y_j \boldsymbol{a}_j \boldsymbol{a}_j^{\top}. \tag{11}$$

Clearly, the entire sequence $\left\{\boldsymbol{x}^{t,(l)}\right\}_{t\geq 0}$ is independent of the lth sampling vector $\boldsymbol{a}_l$. This auxiliary procedure is formally described in Algorithm 1.

2.3 Step 3: establishing the incoherence condition by induction

As revealed by Lemma 3, it suffices to prove that the iterates $\{\boldsymbol{x}^t\}_{0\leq t\leq T_0}$ satisfies (9) with high probability. Our proof will be inductive in nature. For the sake of clarity, we list all the induction hypotheses:

$$\left\|\boldsymbol{x}^{t} - \boldsymbol{x}^{\natural}\right\|_2 \leq C_1, \tag{13a}$$

[1]Here, we deliberately change $2C_1$ in (7a) to C_1 in the definition of the RIC (9a) to ensure the correctness of the analysis.

Algorithm 1 The lth leave-one-out sequence for phase retrieval

Input: $\{\boldsymbol{a}_j\}_{1\le j\le m, j\neq l}$ and $\{y_j\}_{1\le j\le m, j\neq l}$.

Spectral initialization: let $\lambda_1\big(\boldsymbol{Y}^{(l)}\big)$ and $\widetilde{\boldsymbol{x}}^{0,(l)}$ be the leading eigenvalue and eigenvector of

$$\boldsymbol{Y}^{(l)} = \frac{1}{m}\sum_{j:j\neq l} y_j \boldsymbol{a}_j \boldsymbol{a}_j^\top,$$

respectively, and set

$$\boldsymbol{x}^{0,(l)} = \begin{cases} \sqrt{\lambda_1\left(\boldsymbol{Y}^{(l)}\right)/3}\,\widetilde{\boldsymbol{x}}^{0,(l)}, & \text{if } \big\|\widetilde{\boldsymbol{x}}^{0,(l)} - \boldsymbol{x}^\natural\big\|_2 \le \big\|\widetilde{\boldsymbol{x}}^{0,(l)} + \boldsymbol{x}^\natural\big\|_2, \\ -\sqrt{\lambda_1\left(\boldsymbol{Y}^{(l)}\right)/3}\,\widetilde{\boldsymbol{x}}^{0,(l)}, & \text{else.} \end{cases}$$

Gradient updates: **for** $t = 0, 1, 2, \ldots, T-1$ **do**

$$\boldsymbol{x}^{t+1,(l)} = \boldsymbol{x}^{t,(l)} - \eta_t \nabla f^{(l)}\big(\boldsymbol{x}^{t,(l)}\big). \tag{12}$$

$$\max_{1\le l\le m}\big\|\boldsymbol{x}^t - \boldsymbol{x}^{t,(l)}\big\|_2 \le C_3\sqrt{\frac{\log n}{n}} \tag{13b}$$

$$\max_{1\le j\le m}\big|\boldsymbol{a}_j^\top\left(\boldsymbol{x}^t - \boldsymbol{x}^\natural\right)\big| \le C_2\sqrt{\log n}. \tag{13c}$$

Here $C_3 > 0$ is some universal constant. The induction on (13a), that is,

$$\big\|\boldsymbol{x}^{t+1} - \boldsymbol{x}^\natural\big\|_2 \le C_1, \tag{14}$$

has already been established in Lemma 2. This subsection is devoted to establishing (13b) and (13c) for the $(t+1)$th iteration, assuming that (13) holds true up to the tth iteration. We defer the justification of the base case (i.e. initialization at $t = 0$) to Section 2.4.

- **Step 3(a): proximity between the original and the leave-one-out iterates.** The leave-one-out sequence $\{\boldsymbol{x}^{t,(l)}\}$ behaves similarly to the true WF iterates $\{\boldsymbol{x}^t\}$ while maintaining statistical independence with $\boldsymbol{a}_l$, a key fact that allows us to control the incoherence of lth leave-one-out sequence w.r.t. $\boldsymbol{a}_l$. We will formally quantify the gap between $\boldsymbol{x}^{t+1}$ and $\boldsymbol{x}^{t+1,(l)}$ in the following lemma, which establishes the induction in (13b).

  Lemma 4. *Under the hypotheses (13), with probability at least* $1 - O(mn^{-10})$,

$$\max_{1\le l\le m}\big\|\boldsymbol{x}^{t+1} - \boldsymbol{x}^{t+1,(l)}\big\|_2 \le C_3\sqrt{\frac{\log n}{n}}, \tag{15}$$

  as long as the sample size obeys $m \gg n\log n$ *and the stepsize* $0 < \eta \le 1/\left[5C_2\left(10 + C_2\right)\log n\right]$.

  Proof. The proof relies heavily on the restricted strong convexity (see Lemma 1) and is deferred to Appendix 4.4. □

- **Step 3(b): incoherence of the leave-one-out iterates.** By construction, $\boldsymbol{x}^{t+1,(l)}$ is statistically independent of the sampling vector $\boldsymbol{a}_l$. One can thus invoke the standard Gaussian concentration results and the union bound to derive that with probability at least $1 - O\left(mn^{-10}\right)$,

$$\begin{aligned}\max_{1\le l\le m}\left|\boldsymbol{a}_l^\top\left(\boldsymbol{x}^{t+1,(l)} - \boldsymbol{x}^\natural\right)\right| &\le 5\sqrt{\log n}\big\|\boldsymbol{x}^{t+1,(l)} - \boldsymbol{x}^\natural\big\|_2 \\ &\overset{\text{(i)}}{\le} 5\sqrt{\log n}\left(\big\|\boldsymbol{x}^{t+1,(l)} - \boldsymbol{x}^{t+1}\big\|_2 + \big\|\boldsymbol{x}^{t+1} - \boldsymbol{x}^\natural\big\|_2\right)\end{aligned}$$

$$\overset{\text{(ii)}}{\leq} 5\sqrt{\log n}\left(C_3\sqrt{\frac{\log n}{n}}+C_1\right)$$
$$\leq C_4\sqrt{\log n} \tag{16}$$

holds for some constant $C_4 \geq 6C_1 > 0$ and n sufficiently large. Here, (i) comes from the triangle inequality, and (ii) arises from the proximity bound (15) and the condition (14).

- **Step 3(c): combining the bounds.** We are now prepared to establish (13c) for the $(t+1)$th iteration. Specifically,

$$\begin{aligned}\max_{1\leq l\leq m}\left|\boldsymbol{a}_l^\top\left(\boldsymbol{x}^{t+1}-\boldsymbol{x}^\natural\right)\right| &\leq \max_{1\leq l\leq m}\left|\boldsymbol{a}_l^\top\left(\boldsymbol{x}^{t+1}-\boldsymbol{x}^{t+1,(l)}\right)\right| + \max_{1\leq l\leq m}\left|\boldsymbol{a}_l^\top\left(\boldsymbol{x}^{t+1,(l)}-\boldsymbol{x}^\natural\right)\right| \\ &\overset{\text{(i)}}{\leq} \max_{1\leq l\leq m}\|\boldsymbol{a}_l\|_2\left\|\boldsymbol{x}^{t+1}-\boldsymbol{x}^{t+1,(l)}\right\|_2 + C_4\sqrt{\log n} \\ &\overset{\text{(ii)}}{\leq} \sqrt{6n}\cdot C_3\sqrt{\frac{\log n}{n}} + C_4\sqrt{\log n} \leq C_2\sqrt{\log n},\end{aligned} \tag{17}$$

where (i) follows from the Cauchy-Schwarz inequality and (16), the inequality (ii) is a consequence of (15) and (40), and the last inequality holds as long as $C_2/(C_3+C_4)$ is sufficiently large.

Using mathematical induction and the union bound, we establish (13) for all $t \leq T_0 = n$ with high probability. This in turn concludes the proof of Theorem 1, as long as the hypotheses are valid for the base case.

2.4 The base case: spectral initialization

In the end, we return to verify the induction hypotheses for the base case ($t = 0$), i.e. the spectral initialization obeys (13). The following lemma justifies (13a) by choosing δ sufficiently small.

Lemma 5. *Fix any small constant $\delta > 0$, and suppose $m > c_0 n\log n$ for some large constant $c_0 > 0$. Consider the two vectors $\boldsymbol{x}^0$ and $\widetilde{\boldsymbol{x}}^0$ as defined in Algorithm 1, and suppose without loss of generality that (4) holds. Then with probability exceeding $1 - O(n^{-10})$, one has*

$$\|\boldsymbol{Y} - \mathbb{E}[\boldsymbol{Y}]\| \leq \delta, \tag{18}$$

$$\|\boldsymbol{x}^0 - \boldsymbol{x}^\natural\|_2 \leq 2\delta \qquad \textit{and} \qquad \left\|\widetilde{\boldsymbol{x}}^0 - \boldsymbol{x}^\natural\right\|_2 \leq \sqrt{2}\delta. \tag{19}$$

Proof. This result follows directly from the Davis-Kahan sinΘ theorem. See Appendix 4.5. ☐

We then move on to justifying (13b), the proximity between the original and leave-one-out iterates for $t = 0$.

Lemma 6. *Suppose $m > c_0 n\log n$ for some large constant $c_0 > 0$. Then with probability at least $1 - O(mn^{-10})$, one has*

$$\max_{1\leq l\leq m}\left\|\boldsymbol{x}^0 - \boldsymbol{x}^{0,(l)}\right\|_2 \leq C_3\sqrt{\frac{\log n}{n}}. \tag{20}$$

Proof. This is also a consequence of the Davis-Kahan sinΘ theorem. See Appendix 4.6. ☐

The final claim (13c) can be proved using the same argument as in deriving (17), and hence is omitted.

3 Analysis for matrix completion

In this section, we instantiate the general recipe presented in Section 1 to matrix completion and prove Theorem 2. Before continuing, we first gather a few useful facts regarding the loss function for matrix completion. The gradient of it is given by

$$\nabla f(\boldsymbol{X}) = \frac{1}{p}\mathcal{P}_\Omega\left[\boldsymbol{X}\boldsymbol{X}^\top - \left(\boldsymbol{M}^\natural + \boldsymbol{E}\right)\right]\boldsymbol{X}. \tag{21}$$

We define the expected gradient (with respect to the sampling set Ω) to be

$$\nabla F\left(\boldsymbol{X}\right)=\left[\boldsymbol{X}\boldsymbol{X}^{\top}-\left(\boldsymbol{M}^{\natural}+\boldsymbol{E}\right)\right]\boldsymbol{X}$$

and also the (expected) gradient without noise to be

$$\nabla f_{\mathrm{clean}}\left(\boldsymbol{X}\right)=\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{X}\boldsymbol{X}^{\top}-\boldsymbol{M}^{\natural}\right)\boldsymbol{X}\qquad\text{and}\qquad\nabla F_{\mathrm{clean}}\left(\boldsymbol{X}\right)=\left(\boldsymbol{X}\boldsymbol{X}^{\top}-\boldsymbol{M}^{\natural}\right)\boldsymbol{X}.\tag{22}$$

In addition, we need the Hessian $\nabla^2 f_{\mathrm{clean}}(\boldsymbol{X})$, which is represented by an $nr\times nr$ matrix. Simple calculations reveal that for any $\boldsymbol{V}\in\mathbb{R}^{n\times r}$,

$$\mathrm{vec}\left(\boldsymbol{V}\right)^{\top}\nabla^{2}f_{\mathrm{clean}}\left(\boldsymbol{X}\right)\mathrm{vec}\left(\boldsymbol{V}\right)=\frac{1}{2p}\left\Vert \mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\top}+\boldsymbol{X}\boldsymbol{V}^{\top}\right)\right\Vert _{\mathrm{F}}^{2}+\frac{1}{p}\left\langle \mathcal{P}_{\Omega}\left(\boldsymbol{X}\boldsymbol{X}^{\top}-\boldsymbol{M}^{\natural}\right),\boldsymbol{V}\boldsymbol{V}^{\top}\right\rangle ,\tag{23}$$

where $\mathrm{vec}(\boldsymbol{V})\in\mathbb{R}^{nr}$ denotes the vectorization of $\boldsymbol{V}$.

And for reference issues, we re-list the theoretical guarantees on the vanilla GD iterates specified by Theorem 2, namely, with probability at least $1-O\left(n^{-3}\right)$, the iterates of Algorithm 2 satisfy

$$\left\Vert \boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\natural}\right\Vert _{\mathrm{F}}\leq\left(C_{4}\rho^{t}\mu r\frac{1}{\sqrt{np}}+C_{1}\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right)\left\Vert \boldsymbol{X}^{\natural}\right\Vert _{\mathrm{F}},\tag{24a}$$

$$\left\Vert \boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\natural}\right\Vert _{2,\infty}\leq\left(C_{5}\rho^{t}\mu r\sqrt{\frac{\log n}{np}}+C_{8}\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\right)\left\Vert \boldsymbol{X}^{\natural}\right\Vert _{2,\infty},\tag{24b}$$

$$\left\Vert \boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\natural}\right\Vert \leq\left(C_{9}\rho^{t}\mu r\frac{1}{\sqrt{np}}+C_{10}\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right)\left\Vert \boldsymbol{X}^{\natural}\right\Vert \tag{24c}$$

for all $0\leq t\leq T=O(n^5)$, where C_1, C_4, C_5, C_8, C_9 and C_{10} are some absolute positive constants and $1-(\sigma_{\min}/5)\cdot\eta\leq\rho<1$, provided that $0<\eta_t\equiv\eta\leq 2/\left(25\kappa\sigma_{\max}\right)$.

3.1 Step 1: characterizing local geometry in the RIC

3.1.1 Local geometry

The first step is to characterize the region where the empirical loss function enjoys restricted strong convexity and smoothness in an appropriate sense. This is formally stated in the following lemma.

Lemma 7 (Restricted strong convexity and smoothness for matrix completion)**.** *Suppose that the sample size obeys $n^2p\geq C\kappa^2\mu rn\log n$ for some sufficiently large constant $C>0$. Then with probability at least $1-O\left(n^{-10}\right)$, the Hessian $\nabla^2 f_{\mathrm{clean}}(\boldsymbol{X})$ as defined in* (23) *obeys*

$$\mathrm{vec}\left(\boldsymbol{V}\right)^{\top}\nabla^{2}f_{\mathrm{clean}}\left(\boldsymbol{X}\right)\mathrm{vec}\left(\boldsymbol{V}\right)\geq\frac{\sigma_{\min}}{2}\left\Vert \boldsymbol{V}\right\Vert _{\mathrm{F}}^{2}\qquad\text{and}\qquad\left\Vert \nabla^{2}f_{\mathrm{clean}}\left(\boldsymbol{X}\right)\right\Vert \leq\frac{5}{2}\sigma_{\max}\tag{25}$$

for all $\boldsymbol{X}$ and $\boldsymbol{V}=\boldsymbol{Y}\boldsymbol{H}_Y-\boldsymbol{Z}$, with $\boldsymbol{H}_Y:=\arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\left\Vert\boldsymbol{Y}\boldsymbol{R}-\boldsymbol{Z}\right\Vert_{\mathrm{F}}$, satisfying:

$$\left\Vert \boldsymbol{X}-\boldsymbol{X}^{\natural}\right\Vert _{2,\infty}\leq\epsilon\left\Vert \boldsymbol{X}^{\natural}\right\Vert _{2,\infty},\tag{26a}$$

$$\left\Vert \boldsymbol{Z}-\boldsymbol{X}^{\natural}\right\Vert \leq\delta\left\Vert \boldsymbol{X}^{\natural}\right\Vert ,\tag{26b}$$

where $\epsilon\ll 1/\sqrt{\kappa^3\mu r\log^2 n}$ and $\delta\ll 1/\kappa$.

Proof. See Appendix 5.1. ☐

Lemma 7 reveals that the Hessian matrix is well-conditioned in a neighborhood close to $\boldsymbol{X}^{\natural}$ that remains incoherent measured in the ℓ_2/ℓ_∞ norm (cf. (26a)), and along directions that point towards points which are not far away from the truth in the spectral norm (cf. (26b)).

Remark 1. The second condition (26b) is characterized using the spectral norm $\|\cdot\|$, while in previous works this is typically presented in the Frobenius norm $\|\cdot\|_{\mathrm{F}}$. It is also worth noting that the Hessian matrix — even in the infinite-sample and noiseless case — is rank-deficient and cannot be positive definite. As a result, we resort to the form of strong convexity by restricting attention to certain directions (see the conditions on $\boldsymbol{V}$).

3.1.2 Error contraction

Our goal is to demonstrate the error bounds (24) measured in three different norms. Notably, as long as the iterates satisfy (24) at the tth iteration, then $\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\|_{2,\infty}$ is sufficiently small. Under our sample complexity assumption, $\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t$ satisfies the ℓ_2/ℓ_∞ condition (26a) required in Lemma 7. Consequently, we can invoke Lemma 7 to arrive at the following error contraction result.

Lemma 8 (Contraction w.r.t. the Frobenius norm)**.** *Suppose $n^2p \geq C\kappa^3\mu^3r^3n\log^3 n$ and the noise satisfies* (24)*. If the iterates satisfy* (24a) *and* (24b) *at the tth iteration, then with probability at least $1 - O(n^{-10})$,*

$$\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1} - \boldsymbol{X}^\natural\right\|_{\mathrm{F}} \leq C_4\rho^{t+1}\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}} + C_1\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}}$$

holds as long as $0 < \eta \leq 2/(25\kappa\sigma_{\max})$, $1 - (\sigma_{\min}/4)\cdot\eta \leq \rho < 1$, and C_1 is sufficiently large.

Proof. The proof is built upon Lemma 7. See Appendix 5.2. ☐

Further, if the current iterate satisfies all three conditions in (24), then we can derive a stronger sense of error contraction, namely, contraction in terms of the spectral norm.

Lemma 9 (Contraction w.r.t. the spectral norm)**.** *Suppose $n^2p \geq C\kappa^3\mu^3r^3n\log^3 n$ and the noise satisfies* (24)*. If the iterates satisfy* (24) *at the tth iteration, then*

$$\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1} - \boldsymbol{X}^\natural\right\| \leq C_9\rho^{t+1}\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\natural\right\| + C_{10}\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\natural\right\| \tag{27}$$

holds with probability at least $1 - O(n^{-10})$, provided that $0 < \eta \leq 1/\left(2\sigma_{\max}\right)$ and $1 - (\sigma_{\min}/3)\cdot\eta \leq \rho < 1$.

Proof. The key observation is this: the iterate that proceeds according to the population-level gradient reduces the error w.r.t. $\|\cdot\|$, namely,

$$\left\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \eta\nabla F_{\mathrm{clean}}\big(\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t\big) - \boldsymbol{X}^\natural\right\| < \left\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\right\|,$$

as long as $\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t$ is sufficiently close to the truth. Notably, the orthonormal matrix $\widehat{\boldsymbol{H}}^t$ is still chosen to be the one that minimizes the $\|\cdot\|_{\mathrm{F}}$ distance (as opposed to $\|\cdot\|$), which yields a symmetry property $\boldsymbol{X}^{\natural\top}\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t = \big(\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t\big)^\top\boldsymbol{X}^\natural$, crucial for our analysis. See Appendix 5.3 for details. ☐

3.2 Step 2: introducing the leave-one-out sequences

In order to establish the incoherence properties (24b) for the entire trajectory, which is difficult to deal with directly due to the complicated statistical dependence, we introduce a collection of *leave-one-out* versions of $\{\boldsymbol{X}^t\}_{t\geq 0}$, denoted by $\{\boldsymbol{X}^{t,(l)}\}_{t\geq 0}$ for each $1 \leq l \leq n$. Specifically, $\{\boldsymbol{X}^{t,(l)}\}_{t\geq 0}$ is the iterates of gradient descent operating on the auxiliary loss function

$$f^{(l)}\left(\boldsymbol{X}\right) := \frac{1}{4p}\left\|\mathcal{P}_{\Omega^{-l}}\left[\boldsymbol{X}\boldsymbol{X}^\top - \left(\boldsymbol{M}^\natural + \boldsymbol{E}\right)\right]\right\|_{\mathrm{F}}^2 + \frac{1}{4}\left\|\mathcal{P}_l\left(\boldsymbol{X}\boldsymbol{X}^\top - \boldsymbol{M}^\natural\right)\right\|_{\mathrm{F}}^2. \tag{28}$$

Here, $\mathcal{P}_{\Omega_l}$ (resp. $\mathcal{P}_{\Omega^{-l}}$ and $\mathcal{P}_l$) represents the orthogonal projection onto the subspace of matrices which vanish outside of the index set $\Omega_l := \{(i,j) \in \Omega \mid i = l \text{ or } j = l\}$ (resp. $\Omega^{-l} := \{(i,j) \in \Omega \mid i \neq l, j \neq l\}$ and $\{(i,j) \mid i = l \text{ or } j = l\}$); that is, for any matrix $\boldsymbol{M}$,

$$\left[\mathcal{P}_{\Omega_l}\left(\boldsymbol{M}\right)\right]_{i,j} = \begin{cases} M_{i,j}, & \text{if } (i = l \text{ or } j = l) \text{ and } (i,j) \in \Omega, \\ 0, & \text{else}, \end{cases} \tag{29}$$

$$\left[\mathcal{P}_{\Omega^{-l}}\left(\boldsymbol{M}\right)\right]_{i,j} = \begin{cases} M_{i,j}, & \text{if } i \neq l \text{ and } j \neq l \text{ and } (i,j) \in \Omega \\ 0, & \text{else} \end{cases} \quad \text{and} \quad \left[\mathcal{P}_l\left(\boldsymbol{M}\right)\right]_{i,j} = \begin{cases} 0, & \text{if } i \neq l \text{ and } j \neq l, \\ M_{i,j}, & \text{if } i = l \text{ or } j = l. \end{cases} \tag{30}$$

The gradient of the leave-one-out loss function (28) is given by

$$\nabla f^{(l)}\left(\boldsymbol{X}\right)=\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left[\boldsymbol{X}\boldsymbol{X}^{\top}-\left(\boldsymbol{M}^{\natural}+\boldsymbol{E}\right)\right]\boldsymbol{X}+\mathcal{P}_{l}\left(\boldsymbol{X}\boldsymbol{X}^{\top}-\boldsymbol{M}^{\natural}\right)\boldsymbol{X}. \tag{31}$$

The full algorithm to obtain the leave-one-out sequence $\{\boldsymbol{X}^{t,(l)}\}_{t\geq 0}$ (including spectral initialization) is summarized in Algorithm 2.

Algorithm 2 The lth leave-one-out sequence for matrix completion

Input: $\boldsymbol{Y}=[Y_{i,j}]_{1\leq i,j\leq n}$, $\boldsymbol{M}^{\natural}_{\cdot,l}$, $\boldsymbol{M}^{\natural}_{l,\cdot}$, r, p.
Spectral initialization: Let $\boldsymbol{U}^{0,(l)}\boldsymbol{\Sigma}^{(l)}\boldsymbol{U}^{0,(l)\top}$ be the top-r eigendecomposition of

$$\boldsymbol{M}^{(l)}:=\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left(\boldsymbol{Y}\right)+\mathcal{P}_{l}\left(\boldsymbol{M}^{\natural}\right)=\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left(\boldsymbol{M}^{\natural}+\boldsymbol{E}\right)+\mathcal{P}_{l}\left(\boldsymbol{M}^{\natural}\right)$$

with $\mathcal{P}_{\Omega^{-l}}$ and $\mathcal{P}_{l}$ defined in (30), and set $\boldsymbol{X}^{0,(l)}=\boldsymbol{U}^{0,(l)}\big(\boldsymbol{\Sigma}^{(l)}\big)^{1/2}$.
Gradient updates: **for** $t=0,1,2,\dots,T-1$ **do**

$$\boldsymbol{X}^{t+1,(l)}=\boldsymbol{X}^{t,(l)}-\eta_{t}\nabla f^{(l)}\big(\boldsymbol{X}^{t,(l)}\big). \tag{32}$$

Remark 2. Rather than simply dropping all samples in the lth row/column, we replace the lth row/column with their respective population means. In other words, the leave-one-out gradient forms an unbiased surrogate for the true gradient, which is particularly important in ensuring high estimation accuracy.

3.3 Step 3: establishing the incoherence condition by induction

We will continue the proof of Theorem 2 in an inductive manner. As seen in Section 3.1.2, the induction hypotheses (24a) and (24c) hold for the $(t+1)$th iteration as long as (24) holds at the tth iteration. Therefore, we are left with proving the incoherence hypothesis (24b) for all $0\leq t\leq T=O(n^{5})$. For clarity of analysis, it is crucial to maintain a list of induction hypotheses, which includes a few more hypotheses that complement (24), and is given below.

$$\big\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\natural}\big\|_{\mathrm{F}}\leq\left(C_{4}\rho^{t}\mu r\frac{1}{\sqrt{np}}+C_{1}\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right)\big\|\boldsymbol{X}^{\natural}\big\|_{\mathrm{F}}, \tag{33a}$$

$$\big\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\natural}\big\|_{2,\infty}\leq\left(C_{5}\rho^{t}\mu r\sqrt{\frac{\log n}{np}}+C_{8}\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\right)\big\|\boldsymbol{X}^{\natural}\big\|_{2,\infty}, \tag{33b}$$

$$\big\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\natural}\big\|\leq\left(C_{9}\rho^{t}\mu r\frac{1}{\sqrt{np}}+C_{10}\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right)\big\|\boldsymbol{X}^{\natural}\big\|, \tag{33c}$$

$$\max_{1\leq l\leq n}\big\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\big\|_{\mathrm{F}}\leq\left(C_{3}\rho^{t}\mu r\sqrt{\frac{\log n}{np}}+C_{7}\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\right)\big\|\boldsymbol{X}^{\natural}\big\|_{2,\infty}, \tag{33d}$$

$$\max_{1\leq l\leq n}\big\|\big(\boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)}-\boldsymbol{X}^{\natural}\big)_{l,\cdot}\big\|_{2}\leq\left(C_{2}\rho^{t}\mu r\frac{1}{\sqrt{np}}+C_{6}\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\right)\big\|\boldsymbol{X}^{\natural}\big\|_{2,\infty} \tag{33e}$$

hold for some absolute constants $0<\rho<1$ and $C_{1},\cdots,C_{10}>0$. Here, $\widehat{\boldsymbol{H}}^{t,(l)}$ and $\boldsymbol{R}^{t,(l)}$ are orthonormal matrices defined by

$$\widehat{\boldsymbol{H}}^{t,(l)}:=\arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}-\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}}, \tag{34}$$

$$\boldsymbol{R}^{t,(l)}:=\arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\big\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}-\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}\big\|_{\mathrm{F}}. \tag{35}$$

Clearly, the first three hypotheses (33a)-(33c) constitute the conclusion of Theorem 2, i.e. (24). The last two hypotheses (33d) and (33e) are auxiliary properties connecting the true iterates and the auxiliary leave-one-out sequences. Moreover, we summarize below several immediate consequences of (33), which will be useful throughout.

Lemma 10. *Suppose* $n^2p \gg \kappa^3\mu^2r^2 n\log n$ *and the noise satisfies* (24)*. Under the hypotheses* (33)*, one has*

$$\left\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)}\right\|_{\mathrm{F}} \le 5\kappa\left\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\|_{\mathrm{F}}, \tag{36a}$$

$$\left\|\boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \boldsymbol{X}^\natural\right\|_{\mathrm{F}} \le \left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^\natural\right\|_{\mathrm{F}} \le \left\{2C_4\rho^t\mu r\frac{1}{\sqrt{np}} + 2C_1\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right\}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}}, \tag{36b}$$

$$\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^\natural\right\|_{2,\infty} \le \left\{(C_3+C_5)\rho^t\mu r\sqrt{\frac{\log n}{np}} + (C_8+C_7)\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\right\}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}, \tag{36c}$$

$$\left\|\boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \boldsymbol{X}^\natural\right\| \le \left\{2C_9\rho^t\mu r\frac{1}{\sqrt{np}} + 2C_{10}\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right\}\left\|\boldsymbol{X}^\natural\right\|. \tag{36d}$$

In particular, (36a) *follows from hypotheses* (33c) *and* (33d)*.*

Proof. See Appendix 5.4. □

In the sequel, we follow the general recipe outlined in Section 1 to establish the induction hypotheses. We only need to establish (33b), (33d) and (33e) for the $(t+1)$th iteration, since (33a) and (33c) have been established in Section 3.1.2. Specifically, we resort to the leave-one-out iterates by showing that: first, the true and the auxiliary iterates remain exceedingly close throughout; second, the lth leave-one-out sequence stays incoherent with $\boldsymbol{e}_l$ due to statistical independence.

- **Step 3(a): proximity between the original and the leave-one-out iterates.** We demonstrate that $\boldsymbol{X}^{t+1}$ is well approximated by $\boldsymbol{X}^{t+1,(l)}$, up to proper orthonormal transforms. This is precisely the induction hypothesis (33d) for the $(t+1)$th iteration.

  Lemma 11. *Suppose the sample complexity satisfies* $n^2p \gg \kappa^4\mu^3r^3n\log^3 n$ *and the noise satisfies* (24)*. Under the hypotheses* (33) *for the* t*th iteration, we have*

  $$\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1} - \boldsymbol{X}^{t+1,(l)}\boldsymbol{R}^{t+1,(l)}\right\|_{\mathrm{F}} \le C_3\rho^{t+1}\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} + C_7\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} \tag{37}$$

  with probability at least $1-O(n^{-10})$*, provided that* $0<\eta\le 2/(25\kappa\sigma_{\max})$*,* $1-(\sigma_{\min}/5)\cdot\eta\le\rho<1$ *and* $C_7>0$ *is sufficiently large.*

  Proof. The fact that this difference is well-controlled relies heavily on the benign geometric property of the Hessian revealed by Lemma 7. Two important remarks are in order: (1) both points $\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t$ and $\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}$ satisfy (26a); (2) the difference $\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}$ forms a valid direction for restricted strong convexity. These two properties together allow us to invoke Lemma 7. See Appendix 5.5. □

- **Step 3(b): incoherence of the leave-one-out iterates.** Given that $\boldsymbol{X}^{t+1,(l)}$ is sufficiently close to $\boldsymbol{X}^{t+1}$, we turn our attention to establishing the incoherence of this surrogate $\boldsymbol{X}^{t+1,(l)}$ w.r.t. $\boldsymbol{e}_l$. This amounts to proving the induction hypothesis (33e) for the $(t+1)$th iteration.

  Lemma 12. *Suppose the sample complexity meets* $n^2p \gg \kappa^3\mu^3r^3n\log^3 n$ *and the noise satisfies* (24)*. Under the hypotheses* (33) *for the* t*th iteration, one has*

  $$\left\|\left(\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t+1,(l)} - \boldsymbol{X}^\natural\right)_{l,\cdot}\right\|_2 \le C_2\rho^{t+1}\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} + C_6\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} \tag{38}$$

  with probability at least $1-O(n^{-10})$*, as long as* $0<\eta\le 1/\sigma_{\max}$*,* $1-(\sigma_{\min}/3)\cdot\eta\le\rho<1$*,* $C_2\gg\kappa C_9$ *and* $C_6\gg\kappa C_{10}/\sqrt{\log n}$*.*

Proof. The key observation is that $\boldsymbol{X}^{t+1,(l)}$ is statistically independent from any sample in the lth row/column of the matrix. Since there are an order of np samples in each row/column, we obtain enough information that helps establish the desired incoherence property. See Appendix 5.6. □

- **Step 3(c): combining the bounds.** The inequalities (33d) and (33e) taken collectively allow us to establish the induction hypothesis (33b). Specifically, for every $1 \leq l \leq n$, write

$$\left(\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1} - \boldsymbol{X}^{\natural}\right)_{l,\cdot} = \left(\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1} - \boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t+1,(l)}\right)_{l,\cdot} + \left(\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t+1,(l)} - \boldsymbol{X}^{\natural}\right)_{l,\cdot},$$

and the triangle inequality gives

$$\left\|\left(\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1} - \boldsymbol{X}^{\natural}\right)_{l,\cdot}\right\|_2 \leq \left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1} - \boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t+1,(l)}\right\|_{\mathrm{F}} + \left\|\left(\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t+1,(l)} - \boldsymbol{X}^{\natural}\right)_{l,\cdot}\right\|_2. \tag{39}$$

The second term has already been bounded by (38). Since we have established the induction hypotheses (33c) and (33d) for the $(t+1)$th iteration, the first term can be bounded by (36a) for the $(t+1)$th iteration, i.e.

$$\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1} - \boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t+1,(l)}\right\|_{\mathrm{F}} \leq 5\kappa \left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1} - \boldsymbol{X}^{t+1,(l)}\boldsymbol{R}^{t+1,(l)}\right\|_{\mathrm{F}}.$$

Plugging the above inequality, (37) and (38) into (39), we have

$$\begin{aligned}\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1} - \boldsymbol{X}^{\natural}\right\|_{2,\infty} &\leq 5\kappa\left(C_3\rho^{t+1}\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} + \frac{C_7}{\sigma_{\min}}\sigma\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\right) \\ &\quad + C_2\rho^{t+1}\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} + \frac{C_6}{\sigma_{\min}}\sigma\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} \\ &\leq C_5\rho^{t+1}\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} + \frac{C_8}{\sigma_{\min}}\sigma\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\end{aligned}$$

as long as $C_5/(\kappa C_3 + C_2)$ and $C_8/(\kappa C_7 + C_6)$ are sufficiently large. This establishes the induction hypothesis (33b) and finishes the proof.

3.4 The base case: spectral initialization

Finally, we return to check the base case, namely, we aim to show that the spectral initialization satisfies the induction hypotheses (33a)-(33e) for $t = 0$. This is accomplished via the following lemma.

Lemma 13. *Suppose the sample size obeys $n^2p \gg \mu^2 r^2 n \log n$, the noise satisfies (24), and $\kappa = \sigma_{\max}/\sigma_{\min} \asymp 1$. Then with probability at least $1 - O\left(n^{-10}\right)$, the claims in* (33a)-(33e) *hold simultaneously for $t = 0$.*

Proof. This follows by invoking the Davis-Kahan sinΘ theorem [?] as well as the entrywise eigenvector perturbation analysis in [?]. We defer the proof to Appendix 5.7. □

4 Proofs for phase retrieval

Before proceeding, we gather a few simple facts. The standard concentration inequality for χ^2 random variables together with the union bound reveals that the sampling vectors $\{\boldsymbol{a}_j\}$ obey

$$\max_{1\leq j\leq m}\|\boldsymbol{a}_j\|_2 \leq \sqrt{6n} \tag{40}$$

with probability at least $1 - O(me^{-1.5n})$. In addition, standard Gaussian concentration inequalities give

$$\max_{1\leq j\leq m}\left|\boldsymbol{a}_j^{\top}\boldsymbol{x}^{\natural}\right| \leq 5\sqrt{\log n} \tag{41}$$

with probability exceeding $1 - O(mn^{-10})$.

4.1 Proof of Lemma 1

We start with the smoothness bound, namely, $\nabla^2 f(\boldsymbol{x}) \preceq O(\log n)\cdot \boldsymbol{I}_n$. It suffices to prove the upper bound $\left\|\nabla^2 f\left(\boldsymbol{x}\right)\right\| \lesssim \log n$. To this end, we first decompose the Hessian (cf. (6)) into three components as follows:

$$\nabla^2 f\left(\boldsymbol{x}\right) = \underbrace{\frac{3}{m}\sum_{j=1}^{m}\left[\left(\boldsymbol{a}_j^\top \boldsymbol{x}\right)^2 - \left(\boldsymbol{a}_j^\top \boldsymbol{x}^\natural\right)^2\right]\boldsymbol{a}_j\boldsymbol{a}_j^\top}_{:=\boldsymbol{\Lambda}_1} + \underbrace{\frac{2}{m}\sum_{j=1}^{m}\left(\boldsymbol{a}_j^\top \boldsymbol{x}^\natural\right)^2\boldsymbol{a}_j\boldsymbol{a}_j^\top - 2\left(\boldsymbol{I}_n + 2\boldsymbol{x}^\natural\boldsymbol{x}^{\natural\top}\right)}_{:=\boldsymbol{\Lambda}_2} + \underbrace{2\left(\boldsymbol{I}_n + 2\boldsymbol{x}^\natural\boldsymbol{x}^{\natural\top}\right)}_{:=\boldsymbol{\Lambda}_3},$$

where we have used $y_j = (\boldsymbol{a}_j^\top \boldsymbol{x}^\natural)^2$. In the sequel, we control the three terms $\boldsymbol{\Lambda}_1$, $\boldsymbol{\Lambda}_2$ and $\boldsymbol{\Lambda}_3$ in reverse order.

- The third term $\boldsymbol{\Lambda}_3$ can be easily bounded by

$$\|\boldsymbol{\Lambda}_3\| \leq 2\left(\|\boldsymbol{I}_n\| + 2\left\|\boldsymbol{x}^\natural\boldsymbol{x}^{\natural\top}\right\|\right) = 6.$$

- The second term $\boldsymbol{\Lambda}_2$ can be controlled by means of Lemma 19:

$$\|\boldsymbol{\Lambda}_2\| \leq 2\delta$$

  for an arbitrarily small constant $\delta > 0$, as long as $m \geq c_0 n\log n$ for c_0 sufficiently large.

- It thus remains to control $\boldsymbol{\Lambda}_1$. Towards this we discover that

$$\|\boldsymbol{\Lambda}_1\| \leq \left\|\frac{3}{m}\sum_{j=1}^{m}\left|\boldsymbol{a}_j^\top\left(\boldsymbol{x} - \boldsymbol{x}^\natural\right)\right|\left|\boldsymbol{a}_j^\top\left(\boldsymbol{x} + \boldsymbol{x}^\natural\right)\right|\boldsymbol{a}_j\boldsymbol{a}_j^\top\right\|. \tag{42}$$

  Under the assumption $\max_{1\leq j\leq m}\left|\boldsymbol{a}_j^\top\left(\boldsymbol{x} - \boldsymbol{x}^\natural\right)\right| \leq C_2\sqrt{\log n}$ and the fact (41), we can also obtain

$$\max_{1\leq j\leq m}\left|\boldsymbol{a}_j^\top\left(\boldsymbol{x} + \boldsymbol{x}^\natural\right)\right| \leq 2\max_{1\leq j\leq m}\left|\boldsymbol{a}_j^\top\boldsymbol{x}^\natural\right| + \max_{1\leq j\leq m}\left|\boldsymbol{a}_j^\top\left(\boldsymbol{x} - \boldsymbol{x}^\natural\right)\right| \leq \left(10 + C_2\right)\sqrt{\log n}.$$

  Substitution into (42) leads to

$$\|\boldsymbol{\Lambda}_1\| \leq 3C_2\left(10 + C_2\right)\log n \cdot \left\|\frac{1}{m}\sum_{j=1}^{m}\boldsymbol{a}_j\boldsymbol{a}_j^\top\right\| \leq 4C_2\left(10 + C_2\right)\log n,$$

  where the last inequality is a direct consequence of Lemma 18.

Combining the above bounds on $\boldsymbol{\Lambda}_1$, $\boldsymbol{\Lambda}_2$ and $\boldsymbol{\Lambda}_3$ yields

$$\left\|\nabla^2 f\left(\boldsymbol{x}\right)\right\| \leq \|\boldsymbol{\Lambda}_1\| + \|\boldsymbol{\Lambda}_2\| + \|\boldsymbol{\Lambda}_3\| \leq 4C_2\left(10 + C_2\right)\log n + 2\delta + 6 \leq 5C_2\left(10 + C_2\right)\log n,$$

as long as n is sufficiently large. This establishes the claimed smoothness property.

Next we move on to the strong convexity lower bound. Picking a constant $C > 0$ and enforcing proper truncation, we get

$$\nabla^2 f\left(\boldsymbol{x}\right) = \frac{1}{m}\sum_{j=1}^{m}\left[3\left(\boldsymbol{a}_j^\top\boldsymbol{x}\right)^2 - y_j\right]\boldsymbol{a}_j\boldsymbol{a}_j^\top \succeq \underbrace{\frac{3}{m}\sum_{j=1}^{m}\left(\boldsymbol{a}_j^\top\boldsymbol{x}\right)^2\mathbb{1}_{\left\{\left|\boldsymbol{a}_j^\top\boldsymbol{x}\right|\leq C\right\}}\boldsymbol{a}_j\boldsymbol{a}_j^\top}_{:=\boldsymbol{\Lambda}_4} - \underbrace{\frac{1}{m}\sum_{j=1}^{m}\left(\boldsymbol{a}_j^\top\boldsymbol{x}^\natural\right)^2\boldsymbol{a}_j\boldsymbol{a}_j^\top}_{:=\boldsymbol{\Lambda}_5}.$$

We begin with the simpler term $\boldsymbol{\Lambda}_5$. Lemma 19 implies that with probability at least $1 - O(n^{-10})$,

$$\left\|\boldsymbol{\Lambda}_5 - \left(\boldsymbol{I}_n + 2\boldsymbol{x}^\natural\boldsymbol{x}^{\natural\top}\right)\right\| \leq \delta$$

holds for any small constant $\delta > 0$, as long as $m/(n\log n)$ is sufficiently large. This reveals that

$$\boldsymbol{\Lambda}_5 \preceq (1+\delta)\cdot\boldsymbol{I}_n + 2\boldsymbol{x}^\natural\boldsymbol{x}^{\natural\top}.$$

To bound $\mathbf{\Lambda}_4$, invoke Lemma 20 to conclude that with probability at least $1-c_3e^{-c_2m}$ (for some constants $c_2, c_3 > 0$),

$$\left\|\mathbf{\Lambda}_4 - 3\left(\beta_1 \boldsymbol{x}\boldsymbol{x}^\top + \beta_2\|\boldsymbol{x}\|_2^2\boldsymbol{I}_n\right)\right\| \leq \delta\|\boldsymbol{x}\|_2^2$$

for any small constant $\delta > 0$, provided that m/n is sufficiently large. Here,

$$\beta_1 := \mathbb{E}\left[\xi^4\,\mathbb{1}_{\{|\xi|\leq C\}}\right] - \mathbb{E}\left[\xi^2\,\mathbb{1}_{|\xi|\leq C}\right] \quad \text{and} \quad \beta_2 := \mathbb{E}\left[\xi^2\,\mathbb{1}_{|\xi|\leq C}\right],$$

where the expectation is taken with respect to $\xi \sim \mathcal{N}(0,1)$. By the assumption $\left\|\boldsymbol{x}-\boldsymbol{x}^\natural\right\|_2 \leq 2C_1$, one has

$$\|\boldsymbol{x}\|_2 \leq 1+2C_1, \quad \left|\|\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}^\natural\|_2^2\right| \leq 2C_1\left(4C_1+1\right), \quad \left\|\boldsymbol{x}^\natural\boldsymbol{x}^{\natural\top} - \boldsymbol{x}\boldsymbol{x}^\top\right\| \leq 6C_1\left(4C_1+1\right),$$

which leads to

$$\begin{aligned}\left\|\mathbf{\Lambda}_4 - 3\left(\beta_1\boldsymbol{x}^\natural\boldsymbol{x}^{\natural\top} + \beta_2\boldsymbol{I}_n\right)\right\| &\leq \left\|\mathbf{\Lambda}_4 - 3\left(\beta_1\boldsymbol{x}\boldsymbol{x}^\top + \beta_2\|\boldsymbol{x}\|_2^2\boldsymbol{I}_n\right)\right\| + 3\left\|\left(\beta_1\boldsymbol{x}^\natural\boldsymbol{x}^{\natural\top} + \beta_2\boldsymbol{I}_n\right) - \left(\beta_1\boldsymbol{x}\boldsymbol{x}^\top + \beta_2\|\boldsymbol{x}\|_2^2\boldsymbol{I}_n\right)\right\| \\ &\leq \delta\|\boldsymbol{x}\|_2^2 + 3\beta_1\left\|\boldsymbol{x}^\natural\boldsymbol{x}^{\natural\top} - \boldsymbol{x}\boldsymbol{x}^\top\right\| + 3\beta_2\left\|\boldsymbol{I}_n - \|\boldsymbol{x}\|_2^2\boldsymbol{I}_n\right\| \\ &\leq \delta\left(1+2C_1\right)^2 + 18\beta_1C_1\left(4C_1+1\right) + 6\beta_2C_1\left(4C_1+1\right).\end{aligned}$$

This further implies

$$\mathbf{\Lambda}_4 \succeq 3\left(\beta_1\boldsymbol{x}^\natural\boldsymbol{x}^{\natural\top} + \beta_2\boldsymbol{I}_n\right) - \left[\delta\left(1+2C_1\right)^2 + 18\beta_1C_1\left(4C_1+1\right) + 6\beta_2C_1\left(4C_1+1\right)\right]\boldsymbol{I}_n.$$

Recognizing that β_1 (resp. β_2) approaches 2 (resp. 1) as C grows, we can thus take C_1 small enough and C large enough to guarantee that

$$\mathbf{\Lambda}_4 \succeq 5\boldsymbol{x}^\natural\boldsymbol{x}^{\natural\top} + 2\boldsymbol{I}_n.$$

Putting the preceding two bounds on $\mathbf{\Lambda}_4$ and $\mathbf{\Lambda}_5$ together yields

$$\nabla^2 f\left(\boldsymbol{x}\right) \succeq 5\boldsymbol{x}^\natural\boldsymbol{x}^{\natural\top} + 2\boldsymbol{I}_n - \left[\left(1+\delta\right)\cdot\boldsymbol{I}_n + 2\boldsymbol{x}^\natural\boldsymbol{x}^{\natural\top}\right] \succeq (1/2)\cdot\boldsymbol{I}_n$$

as claimed.

4.2 Proof of Lemma 2

Using the update rule (cf. (15)) as well as the fundamental theorem of calculus [?, Chapter XIII, Theorem 4.2], we get

$$\boldsymbol{x}^{t+1} - \boldsymbol{x}^\natural = \boldsymbol{x}^t - \eta\nabla f\left(\boldsymbol{x}^t\right) - \left[\boldsymbol{x}^\natural - \eta\nabla f\left(\boldsymbol{x}^\natural\right)\right] = \left[\boldsymbol{I}_n - \eta\int_0^1 \nabla^2 f\left(\boldsymbol{x}\left(\tau\right)\right)\mathrm{d}\tau\right]\left(\boldsymbol{x}^t - \boldsymbol{x}^\natural\right),$$

where we denote $\boldsymbol{x}\left(\tau\right) = \boldsymbol{x}^\natural + \tau(\boldsymbol{x}^t - \boldsymbol{x}^\natural)$, $0 \leq \tau \leq 1$. Here, the first equality makes use of the fact that $\nabla f(\boldsymbol{x}^\natural) = \boldsymbol{0}$. Under the condition (7), it is self-evident that for all $0 \leq \tau \leq 1$,

$$\left\|\boldsymbol{x}\left(\tau\right) - \boldsymbol{x}^\natural\right\|_2 = \|\tau(\boldsymbol{x}^t - \boldsymbol{x}^\natural)\|_2 \leq 2C_1 \qquad \text{and}$$

$$\max_{1\leq l\leq m}\left|\boldsymbol{a}_l^\top\left(\boldsymbol{x}(\tau) - \boldsymbol{x}^\natural\right)\right| \leq \max_{1\leq l\leq m}\left|\boldsymbol{a}_l^\top\tau\left(\boldsymbol{x}^t - \boldsymbol{x}^\natural\right)\right| \leq C_2\sqrt{\log n}.$$

This means that for all $0 \leq \tau \leq 1$,

$$(1/2)\cdot\boldsymbol{I}_n \preceq \nabla^2 f\left(\boldsymbol{x}(\tau)\right) \preceq \left[5C_2\left(10+C_2\right)\log n\right]\cdot\boldsymbol{I}_n$$

in view of Lemma 1. Picking $\eta \leq 1/\left[5C_2\left(10+C_2\right)\log n\right]$ (and hence $\|\eta\nabla^2 f(\boldsymbol{x}(\tau))\| \leq 1$), one sees that

$$\boldsymbol{0} \preceq \boldsymbol{I}_n - \eta\int_0^1 \nabla^2 f\left(\boldsymbol{x}\left(\tau\right)\right)\mathrm{d}\tau \preceq \left(1-\eta/2\right)\cdot\boldsymbol{I}_n,$$

which immediately yields

$$\left\|\boldsymbol{x}^{t+1} - \boldsymbol{x}^\natural\right\|_2 \leq \left\|\boldsymbol{I}_n - \eta\int_0^1 \nabla^2 f\left(\boldsymbol{x}\left(\tau\right)\right)\mathrm{d}\tau\right\|\cdot\left\|\boldsymbol{x}^t - \boldsymbol{x}^\natural\right\|_2 \leq \left(1-\eta/2\right)\left\|\boldsymbol{x}^t - \boldsymbol{x}^\natural\right\|_2.$$

4.3 Proof of Lemma 3

We start with proving (17a). For all $0 \le t \le T_0$, invoke Lemma 2 recursively with the conditions (9) to reach

$$\left\|\boldsymbol{x}^t - \boldsymbol{x}^\natural\right\|_2 \le (1-\eta/2)^t \left\|\boldsymbol{x}^0 - \boldsymbol{x}^\natural\right\|_2 \le C_1 (1-\eta/2)^t \left\|\boldsymbol{x}^\natural\right\|_2. \tag{43}$$

This finishes the proof of (17a) for $0 \le t \le T_0$ and also reveals that

$$\left\|\boldsymbol{x}^{T_0} - \boldsymbol{x}^\natural\right\|_2 \le C_1 (1-\eta/2)^{T_0} \left\|\boldsymbol{x}^\natural\right\|_2 \ll \frac{1}{n}\left\|\boldsymbol{x}^\natural\right\|_2, \tag{44}$$

provided that $\eta \asymp 1/\log n$. Applying the Cauchy-Schwarz inequality and the fact (40) indicate that

$$\max_{1\le l\le m} \left|\boldsymbol{a}_l^\top \left(\boldsymbol{x}^{T_0} - \boldsymbol{x}^\natural\right)\right| \le \max_{1\le l\le m} \|\boldsymbol{a}_l\|_2 \|\boldsymbol{x}^{T_0} - \boldsymbol{x}^\natural\|_2 \le \sqrt{6n}\cdot \frac{1}{n}\|\boldsymbol{x}^\natural\|_2 \ll C_2\sqrt{\log n},$$

leading to the satisfaction of (7). Therefore, invoking Lemma 2 yields

$$\left\|\boldsymbol{x}^{T_0+1} - \boldsymbol{x}^\natural\right\|_2 \le (1-\eta/2)\left\|\boldsymbol{x}^{T_0} - \boldsymbol{x}^\natural\right\|_2 \ll \frac{1}{n}\|\boldsymbol{x}^\natural\|_2.$$

One can then repeat this argument to arrive at for all $t > T_0$

$$\left\|\boldsymbol{x}^t - \boldsymbol{x}^\natural\right\|_2 \le (1-\eta/2)^t \left\|\boldsymbol{x}^0 - \boldsymbol{x}^\natural\right\|_2 \le C_1 (1-\eta/2)^t \left\|\boldsymbol{x}^\natural\right\|_2 \ll \frac{1}{n}\|\boldsymbol{x}^\natural\|_2. \tag{45}$$

We are left with (17b). It is self-evident that the iterates from $0 \le t \le T_0$ satisfy (17b) by assumptions. For $t > T_0$, we can use the Cauchy-Schhwarz inequality to obtain

$$\max_{1\le j\le m} \left|\boldsymbol{a}_j^\top \left(\boldsymbol{x}^t - \boldsymbol{x}^\natural\right)\right| \le \max_{1\le j\le m} \|\boldsymbol{a}_j\|_2 \left\|\boldsymbol{x}^t - \boldsymbol{x}^\natural\right\|_2 \ll \sqrt{n}\cdot\frac{1}{n} \le C_2\sqrt{\log n},$$

where the penultimate relation uses the conditions (40) and (45).

4.4 Proof of Lemma 4

First, going through the same derivation as in (16) and (17) will result in

$$\max_{1\le l\le m} \left|\boldsymbol{a}_l^\top \left(\boldsymbol{x}^{t,(l)} - \boldsymbol{x}^\natural\right)\right| \le C_4\sqrt{\log n} \tag{46}$$

for some $C_4 < C_2$, which will be helpful for our analysis.

We use the gradient update rules once again to decompose

$$\begin{aligned}
\boldsymbol{x}^{t+1} - \boldsymbol{x}^{t+1,(l)} &= \boldsymbol{x}^t - \eta\nabla f\left(\boldsymbol{x}^t\right) - \left[\boldsymbol{x}^{t,(l)} - \eta\nabla f^{(l)}\left(\boldsymbol{x}^{t,(l)}\right)\right] \\
&= \boldsymbol{x}^t - \eta\nabla f\left(\boldsymbol{x}^t\right) - \left[\boldsymbol{x}^{t,(l)} - \eta\nabla f\left(\boldsymbol{x}^{t,(l)}\right)\right] - \eta\left[\nabla f\left(\boldsymbol{x}^{t,(l)}\right) - \nabla f^{(l)}\left(\boldsymbol{x}^{t,(l)}\right)\right] \\
&= \underbrace{\boldsymbol{x}^t - \boldsymbol{x}^{t,(l)} - \eta\left[\nabla f\left(\boldsymbol{x}^t\right) - \nabla f\left(\boldsymbol{x}^{t,(l)}\right)\right]}_{:=\boldsymbol{\nu}_1^{(l)}} \underbrace{- \eta\frac{1}{m}\left[\left(\boldsymbol{a}_l^\top\boldsymbol{x}^{t,(l)}\right)^2 - \left(\boldsymbol{a}_l^\top\boldsymbol{x}^\natural\right)^2\right]\left(\boldsymbol{a}_l^\top\boldsymbol{x}^{t,(l)}\right)\boldsymbol{a}_l}_{:=\boldsymbol{\nu}_2^{(l)}},
\end{aligned}$$

where the last line comes from the definition of $\nabla f(\cdot)$ and $\nabla f^{(l)}(\cdot)$.

1. We first control the term $\boldsymbol{\nu}_2^{(l)}$, which is easier to deal with. Specifically,

$$\begin{aligned}
\|\boldsymbol{\nu}_2^{(l)}\|_2 &\le \eta\frac{\|\boldsymbol{a}_l\|_2}{m}\left|\left(\boldsymbol{a}_l^\top\boldsymbol{x}^{t,(l)}\right)^2 - \left(\boldsymbol{a}_l^\top\boldsymbol{x}^\natural\right)^2\right|\left|\boldsymbol{a}_l^\top\boldsymbol{x}^{t,(l)}\right| \\
&\overset{\text{(i)}}{\lesssim} C_4(C_4+5)(C_4+10)\eta\frac{n\log n}{m}\sqrt{\frac{\log n}{n}} \overset{\text{(ii)}}{\le} c\eta\sqrt{\frac{\log n}{n}},
\end{aligned}$$

for any small constant $c > 0$. Here (i) follows since (40) and, in view of (41) and (46),

$$\left|\left(\boldsymbol{a}_l^\top \boldsymbol{x}^{t,(l)}\right)^2 - \left(\boldsymbol{a}_l^\top \boldsymbol{x}^\natural\right)^2\right| \le \left|\boldsymbol{a}_l^\top\left(\boldsymbol{x}^{t,(l)} - \boldsymbol{x}^\natural\right)\right| \left(\left|\boldsymbol{a}_l^\top\left(\boldsymbol{x}^{t,(l)} - \boldsymbol{x}^\natural\right)\right| + 2\left|\boldsymbol{a}_l^\top \boldsymbol{x}^\natural\right|\right) \le C_4(C_4+10)\log n,$$

$$\text{and} \qquad \left|\boldsymbol{a}_l^\top \boldsymbol{x}^{t,(l)}\right| \le \left|\boldsymbol{a}_l^\top\left(\boldsymbol{x}^{t,(l)} - \boldsymbol{x}^\natural\right)\right| + \left|\boldsymbol{a}_l^\top \boldsymbol{x}^\natural\right| \le (C_4+5)\sqrt{\log n}.$$

And (ii) holds as long as $m \gg n \log n$.

2. For the term $\boldsymbol{\nu}_1^{(l)}$, the fundamental theorem of calculus [**?**, Chapter XIII, Theorem 4.2] tells us that

$$\boldsymbol{\nu}_1^{(l)} = \left[\boldsymbol{I}_n - \eta \int_0^1 \nabla^2 f\left(\boldsymbol{x}\left(\tau\right)\right) \mathrm{d}\tau\right]\left(\boldsymbol{x}^t - \boldsymbol{x}^{t,(l)}\right),$$

where we abuse the notation and denote $\boldsymbol{x}\left(\tau\right) = \boldsymbol{x}^{t,(l)} + \tau(\boldsymbol{x}^t - \boldsymbol{x}^{t,(l)})$. By the induction hypotheses (13) and the condition (46), one can verify that

$$\left\|\boldsymbol{x}\left(\tau\right) - \boldsymbol{x}^\natural\right\|_2 \le \tau\left\|\boldsymbol{x}^t - \boldsymbol{x}^\natural\right\|_2 + (1-\tau)\left\|\boldsymbol{x}^{t,(l)} - \boldsymbol{x}^\natural\right\|_2 \le 2C_1 \qquad \text{and} \tag{47}$$

$$\max_{1\le l\le m}\left|\boldsymbol{a}_l^\top\left(\boldsymbol{x}\left(\tau\right) - \boldsymbol{x}^\natural\right)\right| \le \tau \max_{1\le l\le m}\left|\boldsymbol{a}_l^\top\left(\boldsymbol{x}^t - \boldsymbol{x}^\natural\right)\right| + (1-\tau)\max_{1\le l\le m}\left|\boldsymbol{a}_l^\top\left(\boldsymbol{x}^{t,(l)} - \boldsymbol{x}^\natural\right)\right| \le C_2\sqrt{\log n}$$

for all $0 \le \tau \le 1$, as long as $C_4 \le C_2$. The second line follows directly from (46). To see why (47) holds, we note that

$$\left\|\boldsymbol{x}^{t,(l)} - \boldsymbol{x}^\natural\right\|_2 \le \left\|\boldsymbol{x}^{t,(l)} - \boldsymbol{x}^t\right\|_2 + \left\|\boldsymbol{x}^t - \boldsymbol{x}^\natural\right\|_2 \le C_3\sqrt{\frac{\log n}{n}} + C_1,$$

where the second inequality follows from the induction hypotheses (13b) and (13a). This combined with (13a) gives

$$\left\|\boldsymbol{x}\left(\tau\right) - \boldsymbol{x}^\natural\right\|_2 \le \tau C_1 + (1-\tau)\left(C_3\sqrt{\frac{\log n}{n}} + C_1\right) \le 2C_1$$

as long as n is large enough, thus justifying (47). Hence by Lemma 1, $\nabla^2 f\left(\boldsymbol{x}\left(\tau\right)\right)$ is positive definite and almost well-conditioned. By choosing $0 < \eta \le 1/\left[5C_2\left(10 + C_2\right)\log n\right]$, we get

$$\left\|\boldsymbol{\nu}_1^{(l)}\right\|_2 \le (1 - \eta/2)\left\|\boldsymbol{x}^t - \boldsymbol{x}^{t,(l)}\right\|_2.$$

3. Combine the preceding bounds on $\boldsymbol{\nu}_1^{(l)}$ and $\boldsymbol{\nu}_2^{(l)}$ as well as the induction bound (13b) to arrive at

$$\left\|\boldsymbol{x}^{t+1} - \boldsymbol{x}^{t+1,(l)}\right\|_2 \le (1-\eta/2)\left\|\boldsymbol{x}^t - \boldsymbol{x}^{t,(l)}\right\|_2 + c\eta\sqrt{\frac{\log n}{n}} \le C_3\sqrt{\frac{\log n}{n}}. \tag{48}$$

This establishes (15) for the $(t+1)$th iteration.

4.5 Proof of Lemma 5

In view of the assumption (4) that $\left\|\boldsymbol{x}^0 - \boldsymbol{x}^\natural\right\|_2 \le \left\|\boldsymbol{x}^0 + \boldsymbol{x}^\natural\right\|_2$ and the fact that $\boldsymbol{x}^0 = \sqrt{\lambda_1\left(\boldsymbol{Y}\right)/3}\,\widetilde{\boldsymbol{x}}^0$ for some $\lambda_1\left(\boldsymbol{Y}\right) > 0$ (which we will verify below), it is straightforward to see that

$$\left\|\widetilde{\boldsymbol{x}}^0 - \boldsymbol{x}^\natural\right\|_2 \le \left\|\widetilde{\boldsymbol{x}}^0 + \boldsymbol{x}^\natural\right\|_2.$$

One can then invoke the Davis-Kahan sinΘ theorem [**?**, Corollary 1] to obtain

$$\left\|\widetilde{\boldsymbol{x}}^0 - \boldsymbol{x}^\natural\right\|_2 \le 2\sqrt{2}\frac{\left\|\boldsymbol{Y} - \mathbb{E}\left[\boldsymbol{Y}\right]\right\|}{\lambda_1\left(\mathbb{E}\left[\boldsymbol{Y}\right]\right) - \lambda_2\left(\mathbb{E}\left[\boldsymbol{Y}\right]\right)}.$$

Note that (18) — $\|\boldsymbol{Y}-\mathbb{E}[\boldsymbol{Y}]\| \le \delta$ — is a direct consequence of Lemma 19. Additionally, the fact that $\mathbb{E}[\boldsymbol{Y}] = \boldsymbol{I} + 2\boldsymbol{x}^{\natural}\boldsymbol{x}^{\natural\top}$ gives $\lambda_1(\mathbb{E}[\boldsymbol{Y}]) = 3$, $\lambda_2(\mathbb{E}[\boldsymbol{Y}]) = 1$, and $\lambda_1(\mathbb{E}[\boldsymbol{Y}]) - \lambda_2(\mathbb{E}[\boldsymbol{Y}]) = 2$. Combining this spectral gap and the inequality $\|\boldsymbol{Y}-\mathbb{E}[\boldsymbol{Y}]\| \le \delta$, we arrive at

$$\left\|\widetilde{\boldsymbol{x}}^0 - \boldsymbol{x}^{\natural}\right\|_2 \le \sqrt{2}\delta.$$

To connect this bound with $\boldsymbol{x}^0$, we need to take into account the scaling factor $\sqrt{\lambda_1(\boldsymbol{Y})/3}$. To this end, it follows from Weyl's inequality and (18) that

$$|\lambda_1(\boldsymbol{Y}) - 3| = |\lambda_1(\boldsymbol{Y}) - \lambda_1(\mathbb{E}[\boldsymbol{Y}])| \le \|\boldsymbol{Y}-\mathbb{E}[\boldsymbol{Y}]\| \le \delta$$

and, as a consequence, $\lambda_1(\boldsymbol{Y}) \ge 3-\delta > 0$ when $\delta \le 1$. This further implies that

$$\left|\sqrt{\frac{\lambda_1(\boldsymbol{Y})}{3}} - 1\right| = \left|\frac{\frac{\lambda_1(\boldsymbol{Y})}{3} - 1}{\sqrt{\frac{\lambda_1(\boldsymbol{Y})}{3}} + 1}\right| \le \left|\frac{\lambda_1(\boldsymbol{Y})}{3} - 1\right| \le \frac{1}{3}\delta, \tag{49}$$

where we have used the elementary identity $\sqrt{a}-\sqrt{b} = (a-b)/(\sqrt{a}+\sqrt{b})$. With these bounds in place, we can use the triangle inequality to get

$$\begin{aligned}\left\|\boldsymbol{x}^0 - \boldsymbol{x}^{\natural}\right\|_2 = \left\|\sqrt{\frac{\lambda_1(\boldsymbol{Y})}{3}}\widetilde{\boldsymbol{x}}^0 - \boldsymbol{x}^{\natural}\right\|_2 &= \left\|\sqrt{\frac{\lambda_1(\boldsymbol{Y})}{3}}\widetilde{\boldsymbol{x}}^0 - \widetilde{\boldsymbol{x}}^0 + \widetilde{\boldsymbol{x}}^0 - \boldsymbol{x}^{\natural}\right\|_2 \\ &\le \left|\sqrt{\frac{\lambda_1(\boldsymbol{Y})}{3}} - 1\right| + \left\|\widetilde{\boldsymbol{x}}^0 - \boldsymbol{x}^{\natural}\right\|_2 \\ &\le \frac{1}{3}\delta + \sqrt{2}\delta \le 2\delta.\end{aligned}$$

4.6 Proof of Lemma 6

To begin with, repeating the same argument as in Lemma 5 (which we omit here for conciseness), we see that for any fixed constant $\delta > 0$,

$$\left\|\boldsymbol{Y}^{(l)} - \mathbb{E}\left[\boldsymbol{Y}^{(l)}\right]\right\| \le \delta, \quad \|\boldsymbol{x}^{0,(l)} - \boldsymbol{x}^{\natural}\|_2 \le 2\delta, \quad \left\|\widetilde{\boldsymbol{x}}^{0,(l)} - \boldsymbol{x}^{\natural}\right\|_2 \le \sqrt{2}\delta, \qquad 1 \le l \le m \tag{50}$$

holds with probability at least $1 - O(mn^{-10})$ as long as $m \gg n\log n$. The ℓ_2 bound on $\|\boldsymbol{x}^0 - \boldsymbol{x}^{0,(l)}\|_2$ is derived as follows.

1. We start by controlling $\left\|\widetilde{\boldsymbol{x}}^0 - \widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2$. Combining (19) and (50) yields

$$\left\|\widetilde{\boldsymbol{x}}^0 - \widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2 \le \left\|\widetilde{\boldsymbol{x}}^0 - \boldsymbol{x}^{\natural}\right\|_2 + \left\|\widetilde{\boldsymbol{x}}^{0,(l)} - \boldsymbol{x}^{\natural}\right\|_2 \le 2\sqrt{2}\delta.$$

   For δ sufficiently small, this implies that $\left\|\widetilde{\boldsymbol{x}}^0 - \widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2 \le \left\|\widetilde{\boldsymbol{x}}^0 + \widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2$, and hence the Davis-Kahan sinΘ theorem [?] gives

$$\left\|\widetilde{\boldsymbol{x}}^0 - \widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2 \le \frac{\left\|\left(\boldsymbol{Y} - \boldsymbol{Y}^{(l)}\right)\widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2}{\lambda_1(\boldsymbol{Y}) - \lambda_2\left(\boldsymbol{Y}^{(l)}\right)} \le \left\|\left(\boldsymbol{Y} - \boldsymbol{Y}^{(l)}\right)\widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2. \tag{51}$$

   Here, the second inequality uses Weyl's inequality:

$$\begin{aligned}\lambda_1\left(\boldsymbol{Y}\right) - \lambda_2\left(\boldsymbol{Y}^{(l)}\right) &\ge \lambda_1(\mathbb{E}[\boldsymbol{Y}]) - \left\|\boldsymbol{Y} - \mathbb{E}[\boldsymbol{Y}]\right\| - \lambda_2(\mathbb{E}[\boldsymbol{Y}^{(l)}]) - \left\|\boldsymbol{Y}^{(l)} - \mathbb{E}[\boldsymbol{Y}^{(l)}]\right\| \\ &\ge 3 - \delta - 1 - \delta \ge 1,\end{aligned}$$

   with the proviso that $\delta \le 1/2$.

2. We now connect $\|\boldsymbol{x}^0 - \boldsymbol{x}^{0,(l)}\|_2$ with $\|\widetilde{\boldsymbol{x}}^0 - \widetilde{\boldsymbol{x}}^{0,(l)}\|_2$. Applying the Weyl's inequality and (18) yields

$$|\lambda_1(\boldsymbol{Y}) - 3| \le \|\boldsymbol{Y} - \mathbb{E}[\boldsymbol{Y}]\| \le \delta \qquad \Longrightarrow \qquad \lambda_1(\boldsymbol{Y}) \in [3-\delta, 3+\delta] \subseteq [2,4] \tag{52}$$

and, similarly, $\lambda_1(\boldsymbol{Y}^{(l)}), \|\boldsymbol{Y}\|, \|\boldsymbol{Y}^{(l)}\| \in [2,4]$. Invoke Lemma 21 to arrive at

$$\begin{aligned} \frac{1}{\sqrt{3}}\left\|\boldsymbol{x}^0 - \boldsymbol{x}^{0,(l)}\right\|_2 &\le \frac{\left\|\left(\boldsymbol{Y} - \boldsymbol{Y}^{(l)}\right)\widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2}{2\sqrt{2}} + \left(2 + \frac{4}{\sqrt{2}}\right)\left\|\widetilde{\boldsymbol{x}}^0 - \widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2 \\ &\le 6\left\|\left(\boldsymbol{Y} - \boldsymbol{Y}^{(l)}\right)\widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2, \end{aligned} \tag{53}$$

where the last inequality comes from (51).

3. Everything then boils down to controlling $\left\|\left(\boldsymbol{Y} - \boldsymbol{Y}^{(l)}\right)\widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2$. Towards this we observe that

$$\begin{aligned} \max_{1\le l\le m}\left\|\left(\boldsymbol{Y} - \boldsymbol{Y}^{(l)}\right)\widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2 &= \max_{1\le l\le m}\frac{1}{m}\left\|\left(\boldsymbol{a}_l^\top\boldsymbol{x}^\natural\right)^2\boldsymbol{a}_l\boldsymbol{a}_l^\top\widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2 \\ &\le \max_{1\le l\le m}\frac{\left(\boldsymbol{a}_l^\top\boldsymbol{x}^\natural\right)^2\left|\boldsymbol{a}_l^\top\widetilde{\boldsymbol{x}}^{0,(l)}\right|\left\|\boldsymbol{a}_l\right\|_2}{m} \\ &\overset{\text{(i)}}{\lesssim} \frac{\log n\cdot\sqrt{\log n}\cdot\sqrt{n}}{m} \\ &\asymp \sqrt{\frac{\log n}{n}}\cdot\frac{n\log n}{m}. \end{aligned} \tag{54}$$

The inequality (i) makes use of the fact $\max_l\left|\boldsymbol{a}_l^\top\boldsymbol{x}^\natural\right| \le 5\sqrt{\log n}$ (cf. (41)), the bound $\max_l\|\boldsymbol{a}_l\|_2 \le 6\sqrt{n}$ (cf. (40)), and $\max_l\left|\boldsymbol{a}_l^\top\widetilde{\boldsymbol{x}}^{0,(l)}\right| \le 5\sqrt{\log n}$ (due to statistical independence and standard Gaussian concentration). As long as $m/(n\log n)$ is sufficiently large, substituting the above bound (54) into (53) leads us to conclude that

$$\max_{1\le l\le m}\left\|\boldsymbol{x}^0 - \boldsymbol{x}^{0,(l)}\right\|_2 \le C_3\sqrt{\frac{\log n}{n}} \tag{55}$$

for any constant $C_3 > 0$.

5 Proofs for matrix completion

Before proceeding to the proofs, let us record an immediate consequence of the incoherence property (22):

$$\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} \le \sqrt{\frac{\kappa\mu}{n}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}} \le \sqrt{\frac{\kappa\mu r}{n}}\left\|\boldsymbol{X}^\natural\right\|. \tag{56}$$

where $\kappa = \sigma_{\max}/\sigma_{\min}$ is the condition number of $\boldsymbol{M}^\natural$. This follows since

$$\begin{aligned} \left\|\boldsymbol{X}^\natural\right\|_{2,\infty} &= \left\|\boldsymbol{U}^\natural\left(\boldsymbol{\Sigma}^\natural\right)^{1/2}\right\|_{2,\infty} \le \left\|\boldsymbol{U}^\natural\right\|_{2,\infty}\left\|\left(\boldsymbol{\Sigma}^\natural\right)^{1/2}\right\| \\ &\le \sqrt{\frac{\mu}{n}}\left\|\boldsymbol{U}^\natural\right\|_{\mathrm{F}}\left\|\left(\boldsymbol{\Sigma}^\natural\right)^{1/2}\right\| \le \sqrt{\frac{\mu}{n}}\left\|\boldsymbol{U}^\natural\right\|_{\mathrm{F}}\sqrt{\kappa\sigma_{\min}} \\ &\le \sqrt{\frac{\kappa\mu}{n}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}} \le \sqrt{\frac{\kappa\mu r}{n}}\left\|\boldsymbol{X}^\natural\right\|. \end{aligned}$$

Unless otherwise specified, we use the indicator variable $\delta_{j,k}$ to denote whether the entry in the location (j,k) is included in Ω. Under our model, $\delta_{j,k}$ is a Bernoulli random variable with mean p.

5.1 Proof of Lemma 7

By the expression of the Hessian in (23), one can decompose

$$\begin{aligned}\mathrm{vec}\left(\boldsymbol{V}\right)^{\top}\nabla^{2}f_{\mathrm{clean}}\left(\boldsymbol{X}\right)\mathrm{vec}\left(\boldsymbol{V}\right) &= \frac{1}{2p}\left\Vert \mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\top}+\boldsymbol{X}\boldsymbol{V}^{\top}\right)\right\Vert _{\mathrm{F}}^{2}+\frac{1}{p}\left\langle \mathcal{P}_{\Omega}\left(\boldsymbol{X}\boldsymbol{X}^{\top}-\boldsymbol{M}^{\natural}\right),\boldsymbol{V}\boldsymbol{V}^{\top}\right\rangle \\ =\underbrace{\frac{1}{2p}\left\Vert \mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\top}+\boldsymbol{X}\boldsymbol{V}^{\top}\right)\right\Vert _{\mathrm{F}}^{2}-\frac{1}{2p}\left\Vert \mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\natural\top}+\boldsymbol{X}^{\natural}\boldsymbol{V}^{\top}\right)\right\Vert _{\mathrm{F}}^{2}}_{:=\alpha_{1}} &+\underbrace{\frac{1}{p}\left\langle \mathcal{P}_{\Omega}\left(\boldsymbol{X}\boldsymbol{X}^{\top}-\boldsymbol{M}^{\natural}\right),\boldsymbol{V}\boldsymbol{V}^{\top}\right\rangle }_{:=\alpha_{2}}\\ +\underbrace{\frac{1}{2p}\left\Vert \mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\natural\top}+\boldsymbol{X}^{\natural}\boldsymbol{V}^{\top}\right)\right\Vert _{\mathrm{F}}^{2}-\frac{1}{2}\left\Vert \boldsymbol{V}\boldsymbol{X}^{\natural\top}+\boldsymbol{X}^{\natural}\boldsymbol{V}^{\top}\right\Vert _{\mathrm{F}}^{2}}_{:=\alpha_{3}} &+\underbrace{\frac{1}{2}\left\Vert \boldsymbol{V}\boldsymbol{X}^{\natural\top}+\boldsymbol{X}^{\natural}\boldsymbol{V}^{\top}\right\Vert _{\mathrm{F}}^{2}}_{:=\alpha_{4}}.\end{aligned}$$

The basic idea is to demonstrate that: (1) α_4 is bounded both from above and from below, and (2) the first three terms are sufficiently small in size compared to α_4.

1. We start by controlling α_4. It is immediate to derive the following upper bound

$$\alpha_{4}\leq\left\Vert \boldsymbol{V}\boldsymbol{X}^{\natural\top}\right\Vert _{\mathrm{F}}^{2}+\left\Vert \boldsymbol{X}^{\natural}\boldsymbol{V}^{\top}\right\Vert _{\mathrm{F}}^{2}\leq2\|\boldsymbol{X}^{\natural}\|^{2}\left\Vert \boldsymbol{V}\right\Vert _{\mathrm{F}}^{2}=2\sigma_{\max}\left\Vert \boldsymbol{V}\right\Vert _{\mathrm{F}}^{2}.$$

   When it comes to the lower bound, one discovers that

$$\begin{aligned}\alpha_{4} &= \frac{1}{2}\left\{ \left\Vert \boldsymbol{V}\boldsymbol{X}^{\natural\top}\right\Vert _{\mathrm{F}}^{2}+\left\Vert \boldsymbol{X}^{\natural}\boldsymbol{V}^{\top}\right\Vert _{\mathrm{F}}^{2}+2\mathrm{Tr}\left(\boldsymbol{X}^{\natural\top}\boldsymbol{V}\boldsymbol{X}^{\natural\top}\boldsymbol{V}\right)\right\} \\ &\geq\sigma_{\min}\left\Vert \boldsymbol{V}\right\Vert _{\mathrm{F}}^{2}+\mathrm{Tr}\left[\left(\boldsymbol{Z}+\boldsymbol{X}^{\natural}-\boldsymbol{Z}\right)^{\top}\boldsymbol{V}\left(\boldsymbol{Z}+\boldsymbol{X}^{\natural}-\boldsymbol{Z}\right)^{\top}\boldsymbol{V}\right]\\ &\geq\sigma_{\min}\left\Vert \boldsymbol{V}\right\Vert _{\mathrm{F}}^{2}+\mathrm{Tr}\left(\boldsymbol{Z}^{\top}\boldsymbol{V}\boldsymbol{Z}^{\top}\boldsymbol{V}\right)-2\left\Vert \boldsymbol{Z}-\boldsymbol{X}^{\natural}\right\Vert \left\Vert \boldsymbol{Z}\right\Vert \left\Vert \boldsymbol{V}\right\Vert _{\mathrm{F}}^{2}-\left\Vert \boldsymbol{Z}-\boldsymbol{X}^{\natural}\right\Vert ^{2}\left\Vert \boldsymbol{V}\right\Vert _{\mathrm{F}}^{2}\\ &\geq\left(\sigma_{\min}-5\delta\sigma_{\max}\right)\left\Vert \boldsymbol{V}\right\Vert _{\mathrm{F}}^{2}+\mathrm{Tr}\left(\boldsymbol{Z}^{\top}\boldsymbol{V}\boldsymbol{Z}^{\top}\boldsymbol{V}\right),\end{aligned}\tag{57}$$

   where the last line comes from the assumptions that

$$\left\Vert \boldsymbol{Z}-\boldsymbol{X}^{\natural}\right\Vert \leq\delta\left\Vert \boldsymbol{X}^{\natural}\right\Vert \leq\left\Vert \boldsymbol{X}^{\natural}\right\Vert \qquad\text{and}\qquad\left\Vert \boldsymbol{Z}\right\Vert \leq\left\Vert \boldsymbol{Z}-\boldsymbol{X}^{\natural}\right\Vert +\left\Vert \boldsymbol{X}^{\natural}\right\Vert \leq2\left\Vert \boldsymbol{X}^{\natural}\right\Vert .$$

   With our assumption $\boldsymbol{V}=\boldsymbol{Y}\boldsymbol{H}_{Y}-\boldsymbol{Z}$ in mind, it comes down to controlling

$$\mathrm{Tr}\left(\boldsymbol{Z}^{\top}\boldsymbol{V}\boldsymbol{Z}^{\top}\boldsymbol{V}\right)=\mathrm{Tr}\left[\boldsymbol{Z}^{\top}\left(\boldsymbol{Y}\boldsymbol{H}_{Y}-\boldsymbol{Z}\right)\boldsymbol{Z}^{\top}\left(\boldsymbol{Y}\boldsymbol{H}_{Y}-\boldsymbol{Z}\right)\right].$$

   From the definition of $\boldsymbol{H}_Y$, we see from Lemma 22 that $\boldsymbol{Z}^{\top}\boldsymbol{Y}\boldsymbol{H}_{Y}$ (and hence $\boldsymbol{Z}^{\top}\left(\boldsymbol{Y}\boldsymbol{H}_{Y}-\boldsymbol{Z}\right)$) is a symmetric matrix, which implies that

$$\mathrm{Tr}\left[\boldsymbol{Z}^{\top}\left(\boldsymbol{Y}\boldsymbol{H}_{Y}-\boldsymbol{Z}\right)\boldsymbol{Z}^{\top}\left(\boldsymbol{Y}\boldsymbol{H}_{Y}-\boldsymbol{Z}\right)\right]\geq0.$$

   Substitution into (57) gives

$$\alpha_{4}\geq\left(\sigma_{\min}-5\delta\sigma_{\max}\right)\left\Vert \boldsymbol{V}\right\Vert _{\mathrm{F}}^{2}\geq\frac{9}{10}\sigma_{\min}\left\Vert \boldsymbol{V}\right\Vert _{\mathrm{F}}^{2},$$

   provided that $\kappa\delta\leq1/50$.

2. For α_1, we consider the following quantity

$$\begin{aligned}\left\Vert \mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\top}+\boldsymbol{X}\boldsymbol{V}^{\top}\right)\right\Vert _{\mathrm{F}}^{2} &= \left\langle \mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\top}\right),\mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\top}\right)\right\rangle +\left\langle \mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\top}\right),\mathcal{P}_{\Omega}\left(\boldsymbol{X}\boldsymbol{V}^{\top}\right)\right\rangle \\ &\quad+\left\langle \mathcal{P}_{\Omega}\left(\boldsymbol{X}\boldsymbol{V}^{\top}\right),\mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\top}\right)\right\rangle +\left\langle \mathcal{P}_{\Omega}\left(\boldsymbol{X}\boldsymbol{V}^{\top}\right),\mathcal{P}_{\Omega}\left(\boldsymbol{X}\boldsymbol{V}^{\top}\right)\right\rangle \\ &= 2\left\langle \mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\top}\right),\mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\top}\right)\right\rangle +2\left\langle \mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\top}\right),\mathcal{P}_{\Omega}\left(\boldsymbol{X}\boldsymbol{V}^{\top}\right)\right\rangle .\end{aligned}$$

   Similar decomposition can be performed on $\left\Vert \mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\natural\top}+\boldsymbol{X}^{\natural}\boldsymbol{V}^{\top}\right)\right\Vert _{\mathrm{F}}^{2}$ as well. These identities yield

$$\alpha_{1}=\underbrace{\frac{1}{p}\left[\left\langle \mathcal{P}_{\Omega}(\boldsymbol{V}\boldsymbol{X}^{\top}),\mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\top}\right)\right\rangle -\left\langle \mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\natural\top}\right),\mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\natural\top}\right)\right\rangle \right]}_{:=\beta_{1}}$$

$$+\underbrace{\frac{1}{p}\left[\langle\mathcal{P}_\Omega(\boldsymbol{V}\boldsymbol{X}^\top),\mathcal{P}_\Omega\left(\boldsymbol{X}\boldsymbol{V}^\top\right)\rangle-\left\langle\mathcal{P}_\Omega\left(\boldsymbol{V}\boldsymbol{X}^{\natural\top}\right),\mathcal{P}_\Omega\left(\boldsymbol{X}^\natural\boldsymbol{V}^\top\right)\right\rangle\right]}_{:=\beta_2}.$$

For β_2, one has

$$\beta_2=\frac{1}{p}\left\langle\mathcal{P}_\Omega\left(\boldsymbol{V}\left(\boldsymbol{X}-\boldsymbol{X}^\natural\right)^\top\right),\mathcal{P}_\Omega\left(\left(\boldsymbol{X}-\boldsymbol{X}^\natural\right)\boldsymbol{V}^\top\right)\right\rangle$$
$$+\frac{1}{p}\left\langle\mathcal{P}_\Omega\left(\boldsymbol{V}\left(\boldsymbol{X}-\boldsymbol{X}^\natural\right)^\top\right),\mathcal{P}_\Omega\left(\boldsymbol{X}^\natural\boldsymbol{V}^\top\right)\right\rangle+\frac{1}{p}\left\langle\mathcal{P}_\Omega\left(\boldsymbol{V}\boldsymbol{X}^{\natural\top}\right),\mathcal{P}_\Omega\left(\left(\boldsymbol{X}-\boldsymbol{X}^\natural\right)\boldsymbol{V}^\top\right)\right\rangle$$

which together with the inequality $|\langle\boldsymbol{A},\boldsymbol{B}\rangle|\leq\|\boldsymbol{A}\|_{\mathrm{F}}\|\boldsymbol{B}\|_{\mathrm{F}}$ gives

$$|\beta_2|\leq\frac{1}{p}\left\|\mathcal{P}_\Omega\left(\boldsymbol{V}\left(\boldsymbol{X}-\boldsymbol{X}^\natural\right)^\top\right)\right\|_{\mathrm{F}}^2+\frac{2}{p}\left\|\mathcal{P}_\Omega\left(\boldsymbol{V}\left(\boldsymbol{X}-\boldsymbol{X}^\natural\right)^\top\right)\right\|_{\mathrm{F}}\left\|\mathcal{P}_\Omega\left(\boldsymbol{X}^\natural\boldsymbol{V}^\top\right)\right\|_{\mathrm{F}}. \tag{58}$$

This then calls for upper bounds on the following two terms

$$\frac{1}{\sqrt{p}}\left\|\mathcal{P}_\Omega\left(\boldsymbol{V}\left(\boldsymbol{X}-\boldsymbol{X}^\natural\right)^\top\right)\right\|_{\mathrm{F}}\qquad\text{and}\qquad\frac{1}{\sqrt{p}}\left\|\mathcal{P}_\Omega\left(\boldsymbol{X}^\natural\boldsymbol{V}^\top\right)\right\|_{\mathrm{F}}.$$

The injectivity of $\mathcal{P}_\Omega$ (cf. [**?**, Section 4.2] or Lemma 25)—when restricted to the tangent space of $\boldsymbol{M}^\natural$—gives: for any fixed constant $\gamma>0$,

$$\frac{1}{\sqrt{p}}\left\|\mathcal{P}_\Omega\left(\boldsymbol{X}^\natural\boldsymbol{V}^\top\right)\right\|_{\mathrm{F}}\leq(1+\gamma)\left\|\boldsymbol{X}^\natural\boldsymbol{V}^\top\right\|_{\mathrm{F}}\leq(1+\gamma)\left\|\boldsymbol{X}^\natural\right\|\|\boldsymbol{V}\|_{\mathrm{F}}$$

with probability at least $1-O\left(n^{-10}\right)$, provided that $n^2p/(\mu nr\log n)$ is sufficiently large. In addition,

$$\begin{aligned}\frac{1}{p}\left\|\mathcal{P}_\Omega\left(\boldsymbol{V}\left(\boldsymbol{X}-\boldsymbol{X}^\natural\right)^\top\right)\right\|_{\mathrm{F}}^2&=\frac{1}{p}\sum_{1\leq j,k\leq n}\delta_{j,k}\left[\boldsymbol{V}_{j,\cdot}\left(\boldsymbol{X}_{k,\cdot}-\boldsymbol{X}^\natural_{k,\cdot}\right)^\top\right]^2\\&=\sum_{1\leq j\leq n}\boldsymbol{V}_{j,\cdot}\left[\frac{1}{p}\sum_{1\leq k\leq n}\delta_{j,k}\left(\boldsymbol{X}_{k,\cdot}-\boldsymbol{X}^\natural_{k,\cdot}\right)^\top\left(\boldsymbol{X}_{k,\cdot}-\boldsymbol{X}^\natural_{k,\cdot}\right)\right]\boldsymbol{V}_{j,\cdot}^\top\\&\leq\max_{1\leq j\leq n}\left\|\frac{1}{p}\sum_{1\leq k\leq n}\delta_{j,k}\left(\boldsymbol{X}_{k,\cdot}-\boldsymbol{X}^\natural_{k,\cdot}\right)^\top\left(\boldsymbol{X}_{k,\cdot}-\boldsymbol{X}^\natural_{k,\cdot}\right)\right\|\|\boldsymbol{V}\|_{\mathrm{F}}^2\\&\leq\left\{\frac{1}{p}\max_{1\leq j\leq n}\sum_{1\leq k\leq n}\delta_{j,k}\right\}\left\{\max_{1\leq k\leq n}\left\|\boldsymbol{X}_{k,\cdot}-\boldsymbol{X}^\natural_{k,\cdot}\right\|_2^2\right\}\|\boldsymbol{V}\|_{\mathrm{F}}^2\\&\leq(1+\gamma)\,n\left\|\boldsymbol{X}-\boldsymbol{X}^\natural\right\|_{2,\infty}^2\|\boldsymbol{V}\|_{\mathrm{F}}^2,\end{aligned}$$

with probability exceeding $1-O\left(n^{-10}\right)$, which holds as long as $np/\log n$ is sufficiently large. Taken collectively, the above bounds yield that for any small constant $\gamma>0$,

$$\begin{aligned}|\beta_2|&\leq(1+\gamma)\,n\left\|\boldsymbol{X}-\boldsymbol{X}^\natural\right\|_{2,\infty}^2\|\boldsymbol{V}\|_{\mathrm{F}}^2+2\sqrt{(1+\gamma)\,n\left\|\boldsymbol{X}-\boldsymbol{X}^\natural\right\|_{2,\infty}^2\|\boldsymbol{V}\|_{\mathrm{F}}^2\cdot(1+\gamma)^2\left\|\boldsymbol{X}^\natural\right\|^2\|\boldsymbol{V}\|_{\mathrm{F}}^2}\\&\lesssim\left(\epsilon^2n\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}^2+\epsilon\sqrt{n}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}\left\|\boldsymbol{X}^\natural\right\|\right)\|\boldsymbol{V}\|_{\mathrm{F}}^2,\end{aligned}$$

where the last inequality makes use of the assumption $\|\boldsymbol{X}-\boldsymbol{X}^\natural\|_{2,\infty}\leq\epsilon\|\boldsymbol{X}^\natural\|_{2,\infty}$. The same analysis can be repeated to control β_1. Altogether, we obtain

$$\begin{aligned}|\alpha_1|\leq|\beta_1|+|\beta_2|&\lesssim\left(n\epsilon^2\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}^2+\sqrt{n}\epsilon\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}\left\|\boldsymbol{X}^\natural\right\|\right)\|\boldsymbol{V}\|_{\mathrm{F}}^2\\&\overset{\text{(i)}}{\leq}\left(n\epsilon^2\frac{\kappa\mu r}{n}+\sqrt{n}\epsilon\sqrt{\frac{\kappa\mu r}{n}}\right)\sigma_{\max}\|\boldsymbol{V}\|_{\mathrm{F}}^2\overset{\text{(ii)}}{\leq}\frac{1}{10}\sigma_{\min}\|\boldsymbol{V}\|_{\mathrm{F}}^2,\end{aligned}$$

where (i) utilizes the incoherence condition (56) and (ii) holds with the proviso that $\epsilon\sqrt{\kappa^3\mu r}\ll1$.

3. To bound α_2, apply the Cauchy-Schwarz inequality to get

$$|\alpha_2| = \left| \left\langle \boldsymbol{V},\ \frac{1}{p}\mathcal{P}_\Omega \left(\boldsymbol{X}\boldsymbol{X}^\top - \boldsymbol{M}^\natural\right) \boldsymbol{V} \right\rangle \right| \le \left\| \frac{1}{p}\mathcal{P}_\Omega \left(\boldsymbol{X}\boldsymbol{X}^\top - \boldsymbol{M}^\natural\right) \right\| \|\boldsymbol{V}\|_{\mathrm{F}}^2 .$$

   In view of Lemma 30, with probability at least $1 - O\left(n^{-10}\right)$,

$$\begin{aligned} \left\| \frac{1}{p}\mathcal{P}_\Omega \left(\boldsymbol{X}\boldsymbol{X}^\top - \boldsymbol{M}^\natural\right) \right\| &\le 2n\epsilon^2 \left\|\boldsymbol{X}^\natural\right\|_{2,\infty}^2 + 4\epsilon\sqrt{n}\log n \left\|\boldsymbol{X}^\natural\right\|_{2,\infty} \left\|\boldsymbol{X}^\natural\right\| \\ &\le \left(2n\epsilon^2 \frac{\kappa\mu r}{n} + 4\epsilon\sqrt{n}\log n \sqrt{\frac{\kappa\mu r}{n}} \right) \sigma_{\max} \le \frac{1}{10}\sigma_{\min} \end{aligned}$$

   as soon as $\epsilon\sqrt{\kappa^3\mu r}\log n \ll 1$, where we utilize the incoherence condition (56). This in turn implies that

$$|\alpha_2| \le \frac{1}{10}\sigma_{\min} \|\boldsymbol{V}\|_{\mathrm{F}}^2 .$$

   Notably, this bound holds uniformly over all $\boldsymbol{X}$ satisfying the condition in Lemma 7, regardless of the statistical dependence between $\boldsymbol{X}$ and the sampling set Ω.

4. The last term α_3 can also be controlled using the injectivity of $\mathcal{P}_\Omega$ when restricted to the tangent space of $\boldsymbol{M}^\natural$. Specifically, it follows from the bounds in [**?**, Section 4.2] or Lemma 25 that

$$|\alpha_3| \le \gamma \left\| \boldsymbol{V}\boldsymbol{X}^{\natural\top} + \boldsymbol{X}^\natural\boldsymbol{V}^\top \right\|_{\mathrm{F}}^2 \le 4\gamma\sigma_{\max} \|\boldsymbol{V}\|_{\mathrm{F}}^2 \le \frac{1}{10}\sigma_{\min} \|\boldsymbol{V}\|_{\mathrm{F}}^2$$

   for any $\gamma > 0$ such that $\kappa\gamma$ is a small constant, as soon as $n^2 p \gg \kappa^2\mu r n \log n$.

5. Taking all the preceding bounds collectively yields

$$\begin{aligned} \mathrm{vec}\left(\boldsymbol{V}\right)^\top \nabla^2 f_{\mathrm{clean}}\left(\boldsymbol{X}\right) \mathrm{vec}\left(\boldsymbol{V}\right) &\ge \alpha_4 - |\alpha_1| - |\alpha_2| - |\alpha_3| \\ &\ge \left(\frac{9}{10} - \frac{3}{10} \right) \sigma_{\min} \|\boldsymbol{V}\|_{\mathrm{F}}^2 \ge \frac{1}{2}\sigma_{\min} \|\boldsymbol{V}\|_{\mathrm{F}}^2 \end{aligned}$$

   for all $\boldsymbol{V}$ satisfying our assumptions, and

$$\begin{aligned} \left| \mathrm{vec}\left(\boldsymbol{V}\right)^\top \nabla^2 f_{\mathrm{clean}}\left(\boldsymbol{X}\right) \mathrm{vec}\left(\boldsymbol{V}\right) \right| &\le \alpha_4 + |\alpha_1| + |\alpha_2| + |\alpha_3| \\ &\le \left(2\sigma_{\max} + \frac{3}{10}\sigma_{\min} \right) \|\boldsymbol{V}\|_{\mathrm{F}}^2 \le \frac{5}{2}\sigma_{\max} \|\boldsymbol{V}\|_{\mathrm{F}}^2 \end{aligned}$$

   for all $\boldsymbol{V}$. Since this upper bound holds uniformly over all $\boldsymbol{V}$, we conclude that

$$\left\| \nabla^2 f_{\mathrm{clean}}\left(\boldsymbol{X}\right) \right\| \le \frac{5}{2}\sigma_{\max}$$

   as claimed.

5.2 Proof of Lemma 8

Given that $\widehat{\boldsymbol{H}}^{t+1}$ is chosen to minimize the error in terms of the Frobenius norm (cf. (23)), we have

$$\begin{aligned} \left\| \boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1} - \boldsymbol{X}^\natural \right\|_{\mathrm{F}} &\le \left\| \boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t} - \boldsymbol{X}^\natural \right\|_{\mathrm{F}} = \left\| \left[\boldsymbol{X}^t - \eta\nabla f\left(\boldsymbol{X}^t\right) \right] \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \right\|_{\mathrm{F}} \\ &\overset{\text{(i)}}{=} \left\| \boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \eta\nabla f\big(\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t\big) - \boldsymbol{X}^\natural \right\|_{\mathrm{F}} \\ &\overset{\text{(ii)}}{=} \left\| \boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \eta \left[\nabla f_{\mathrm{clean}}\big(\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t\big) - \frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{E}\right) \boldsymbol{X}^t\widehat{\boldsymbol{H}}^t \right] - \boldsymbol{X}^\natural \right\|_{\mathrm{F}} \end{aligned}$$

$$\leq \underbrace{\left\| \boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \eta \nabla f_{\text{clean}}\big(\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t\big) - \big(\boldsymbol{X}^\natural - \eta \nabla f_{\text{clean}}(\boldsymbol{X}^\natural)\big) \right\|_{\text{F}}}_{:=\alpha_1} + \underbrace{\eta \left\| \frac{1}{p} \mathcal{P}_\Omega(\boldsymbol{E})\, \boldsymbol{X}^t \widehat{\boldsymbol{H}}^t \right\|_{\text{F}}}_{:=\alpha_2}, \tag{59}$$

where (i) follows from the identity $\nabla f(\boldsymbol{X}^t \boldsymbol{R}) = \nabla f(\boldsymbol{X}^t)\, \boldsymbol{R}$ for any orthonormal matrix $\boldsymbol{R} \in \mathcal{O}^{r\times r}$, (ii) arises from the definitions of $\nabla f(\boldsymbol{X})$ and $\nabla f_{\text{clean}}(\boldsymbol{X})$ (see (21) and (22), respectively), and the last inequality (59) utilizes the triangle inequality and the fact that $\nabla f_{\text{clean}}(\boldsymbol{X}^\natural) = \boldsymbol{0}$. It thus suffices to control α_1 and α_2.

1. For the second term α_2 in (59), it is easy to see that

$$\alpha_2 \leq \eta \left\| \frac{1}{p} \mathcal{P}_\Omega(\boldsymbol{E}) \right\| \left\| \boldsymbol{X}^t \widehat{\boldsymbol{H}}^t \right\|_{\text{F}} \leq 2\eta \left\| \frac{1}{p} \mathcal{P}_\Omega(\boldsymbol{E}) \right\| \left\| \boldsymbol{X}^\natural \right\|_{\text{F}} \leq 2\eta C \sigma \sqrt{\frac{n}{p}} \|\boldsymbol{X}^\natural\|_{\text{F}}$$

for some absolute constant $C > 0$. Here, the second inequality holds because $\|\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t\|_{\text{F}} \leq \|\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\|_{\text{F}} + \|\boldsymbol{X}^\natural\|_{\text{F}} \leq 2\|\boldsymbol{X}^\natural\|_{\text{F}}$, following the hypothesis (24a) together with our assumptions on the noise and the sample complexity. The last inequality makes use of Lemma 27.

2. For the first term α_1 in (59), the fundamental theorem of calculus [?, Chapter XIII, Theorem 4.2] reveals

$$\begin{aligned}
&\text{vec}\left[\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \eta \nabla f_{\text{clean}}\big(\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t\big) - \big(\boldsymbol{X}^\natural - \eta \nabla f_{\text{clean}}(\boldsymbol{X}^\natural)\big)\right] \\
&\quad = \text{vec}\left[\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\right] - \eta \cdot \text{vec}\left[\nabla f_{\text{clean}}\big(\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t\big) - \nabla f_{\text{clean}}\left(\boldsymbol{X}^\natural\right)\right] \\
&\quad = \Bigg(\boldsymbol{I}_{nr} - \eta \underbrace{\int_0^1 \nabla^2 f_{\text{clean}}\left(\boldsymbol{X}(\tau)\right) \mathrm{d}\tau}_{:=\boldsymbol{A}}\Bigg) \text{vec}\left(\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\right),
\end{aligned} \tag{60}$$

where we denote $\boldsymbol{X}(\tau) := \boldsymbol{X}^\natural + \tau(\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural)$. Taking the squared Euclidean norm of both sides of the equality (60) leads to

$$\begin{aligned}
(\alpha_1)^2 &= \text{vec}\big(\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\big)^\top (\boldsymbol{I}_{nr} - \eta \boldsymbol{A})^2\, \text{vec}\big(\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\big) \\
&= \text{vec}\big(\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\big)^\top \big(\boldsymbol{I}_{nr} - 2\eta \boldsymbol{A} + \eta^2 \boldsymbol{A}^2\big)\, \text{vec}\big(\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\big) \\
&\leq \left\|\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\right\|_{\text{F}}^2 + \eta^2 \|\boldsymbol{A}\|^2 \left\|\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\right\|_{\text{F}}^2 - 2\eta\, \text{vec}\big(\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\big)^\top \boldsymbol{A}\, \text{vec}\big(\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\big),
\end{aligned} \tag{61}$$

where in (61) we have used the fact that

$$\text{vec}\big(\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\big)^\top \boldsymbol{A}^2 \text{vec}\big(\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\big) \leq \|\boldsymbol{A}\|^2 \left\|\text{vec}\big(\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\big)\right\|_2^2 = \|\boldsymbol{A}\|^2 \left\|\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\right\|_{\text{F}}^2.$$

Based on the condition (24b), it is easily seen that $\forall \tau \in [0,1]$,

$$\left\|\boldsymbol{X}(\tau) - \boldsymbol{X}^\natural\right\|_{2,\infty} \leq \left(C_5 \mu r \sqrt{\frac{\log n}{np}} + \frac{C_8}{\sigma_{\min}} \sigma \sqrt{\frac{n \log n}{p}}\right) \left\|\boldsymbol{X}^\natural\right\|_{2,\infty}.$$

Taking $\boldsymbol{X} = \boldsymbol{X}(\tau), \boldsymbol{Y} = \boldsymbol{X}^t$ and $\boldsymbol{Z} = \boldsymbol{X}^\natural$ in Lemma 7, one can easily verify the assumptions therein given our sample size condition $n^2 p \gg \kappa^3 \mu^3 r^3 n \log^3 n$ and the noise condition (24). As a result,

$$\text{vec}\big(\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\big)^\top \boldsymbol{A}\, \text{vec}\big(\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\big) \geq \frac{\sigma_{\min}}{2} \left\|\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\right\|_{\text{F}}^2 \qquad \text{and} \qquad \|\boldsymbol{A}\| \leq \frac{5}{2} \sigma_{\max}.$$

Substituting these two inequalities into (61) yields

$$(\alpha_1)^2 \leq \left(1 + \frac{25}{4} \eta^2 \sigma_{\max}^2 - \sigma_{\min} \eta\right) \left\|\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\right\|_{\text{F}}^2 \leq \left(1 - \frac{\sigma_{\min}}{2} \eta\right) \left\|\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\right\|_{\text{F}}^2$$

as long as $0 < \eta \leq (2\sigma_{\min})/(25\sigma_{\max}^2)$, which further implies that

$$\alpha_1 \leq \left(1 - \frac{\sigma_{\min}}{4} \eta\right) \left\|\boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\right\|_{\text{F}}.$$

3. Combining the preceding bounds on both α_1 and α_2 and making use of the hypothesis (24a), we have

$$\begin{aligned}
\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1}-\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}} &\le \left(1-\frac{\sigma_{\min}}{4}\eta\right)\left\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}}+2\eta C\sigma\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}} \\
&\le \left(1-\frac{\sigma_{\min}}{4}\eta\right)\left(C_4\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}}+C_1\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}}\right)+2\eta C\sigma\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}} \\
&\le \left(1-\frac{\sigma_{\min}}{4}\eta\right)C_4\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}}+\left[\left(1-\frac{\sigma_{\min}}{4}\eta\right)\frac{C_1}{\sigma_{\min}}+2\eta C\right]\sigma\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}} \\
&\le C_4\rho^{t+1}\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}}+C_1\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}}
\end{aligned}$$

as long as $0<\eta\le(2\sigma_{\min})/(25\sigma_{\max}^2)$, $1-(\sigma_{\min}/4)\cdot\eta\le\rho<1$ and C_1 is sufficiently large. This completes the proof of the contraction with respect to the Frobenius norm.

5.3 Proof of Lemma 9

To facilitate analysis, we construct an auxiliary matrix defined as follows

$$\widetilde{\boldsymbol{X}}^{t+1}:=\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\eta\frac{1}{p}\mathcal{P}_{\Omega}\left[\boldsymbol{X}^{t}\boldsymbol{X}^{t\top}-\left(\boldsymbol{M}^{\natural}+\boldsymbol{E}\right)\right]\boldsymbol{X}^{\natural}. \tag{62}$$

With this auxiliary matrix in place, we invoke the triangle inequality to bound

$$\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1}-\boldsymbol{X}^{\natural}\right\|\le\underbrace{\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1}-\widetilde{\boldsymbol{X}}^{t+1}\right\|}_{:=\alpha_1}+\underbrace{\left\|\widetilde{\boldsymbol{X}}^{t+1}-\boldsymbol{X}^{\natural}\right\|}_{:=\alpha_2}. \tag{63}$$

1. We start with the second term α_2 and show that the auxiliary matrix $\widetilde{\boldsymbol{X}}^{t+1}$ is also not far from the truth. The definition of $\widetilde{\boldsymbol{X}}^{t+1}$ allows one to express

$$\begin{aligned}
\alpha_2 &= \left\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\eta\frac{1}{p}\mathcal{P}_{\Omega}\left[\boldsymbol{X}^{t}\boldsymbol{X}^{t\top}-\left(\boldsymbol{M}^{\natural}+\boldsymbol{E}\right)\right]\boldsymbol{X}^{\natural}-\boldsymbol{X}^{\natural}\right\| \\
&\le \eta\left\|\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{E}\right)\right\|\left\|\boldsymbol{X}^{\natural}\right\|+\left\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\eta\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{X}^{t}\boldsymbol{X}^{t\top}-\boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\boldsymbol{X}^{\natural}-\boldsymbol{X}^{\natural}\right\| \qquad (64)\\
&\le \eta\left\|\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{E}\right)\right\|\left\|\boldsymbol{X}^{\natural}\right\|+\underbrace{\left\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\eta\left(\boldsymbol{X}^{t}\boldsymbol{X}^{t\top}-\boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\boldsymbol{X}^{\natural}-\boldsymbol{X}^{\natural}\right\|}_{:=\beta_1} \\
&\quad +\underbrace{\eta\left\|\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{X}^{t}\boldsymbol{X}^{t\top}-\boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\boldsymbol{X}^{\natural}-\left(\boldsymbol{X}^{t}\boldsymbol{X}^{t\top}-\boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\boldsymbol{X}^{\natural}\right\|}_{:=\beta_2}, \qquad (65)
\end{aligned}$$

where we have used the triangle inequality to separate the population-level component (i.e. β_1), the perturbation (i.e. β_2), and the noise component. In what follows, we will denote

$$\boldsymbol{\Delta}^{t}:=\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\natural}$$

which, by Lemma 22, satisfies the following symmetry property

$$\widehat{\boldsymbol{H}}^{t\top}\boldsymbol{X}^{t\top}\boldsymbol{X}^{\natural}=\boldsymbol{X}^{\natural\top}\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}\qquad\Longrightarrow\qquad\boldsymbol{\Delta}^{t\top}\boldsymbol{X}^{\natural}=\boldsymbol{X}^{\natural\top}\boldsymbol{\Delta}^{t}. \tag{66}$$

(a) The population-level component β_1 is easier to control. Specifically, we first simplify its expression as

$$\begin{aligned}
\beta_1 &= \left\|\boldsymbol{\Delta}^{t}-\eta\left(\boldsymbol{\Delta}^{t}\boldsymbol{\Delta}^{t\top}+\boldsymbol{\Delta}^{t}\boldsymbol{X}^{\natural\top}+\boldsymbol{X}^{\natural}\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^{\natural}\right\| \\
&\le \underbrace{\left\|\boldsymbol{\Delta}^{t}-\eta\left(\boldsymbol{\Delta}^{t}\boldsymbol{X}^{\natural\top}+\boldsymbol{X}^{\natural}\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^{\natural}\right\|}_{:=\gamma_1}+\underbrace{\eta\left\|\boldsymbol{\Delta}^{t}\boldsymbol{\Delta}^{t\top}\boldsymbol{X}^{\natural}\right\|}_{:=\gamma_2}.
\end{aligned}$$

The leading term γ_1 can be upper bounded by

$$\begin{aligned}\gamma_1 &= \left\|\boldsymbol{\Delta}^t - \eta\boldsymbol{\Delta}^t\boldsymbol{\Sigma}^\natural - \eta\boldsymbol{X}^\natural\boldsymbol{\Delta}^{t\top}\boldsymbol{X}^\natural\right\| = \left\|\boldsymbol{\Delta}^t - \eta\boldsymbol{\Delta}^t\boldsymbol{\Sigma}^\natural - \eta\boldsymbol{X}^\natural\boldsymbol{X}^{\natural\top}\boldsymbol{\Delta}^t\right\| \\ &= \left\|\frac{1}{2}\boldsymbol{\Delta}^t\left(\boldsymbol{I}_r - 2\eta\boldsymbol{\Sigma}^\natural\right) + \frac{1}{2}\left(\boldsymbol{I}_r - 2\eta\boldsymbol{M}^\natural\right)\boldsymbol{\Delta}^t\right\| \leq \frac{1}{2}\left(\left\|\boldsymbol{I}_r - 2\eta\boldsymbol{\Sigma}^\natural\right\| + \left\|\boldsymbol{I}_r - 2\eta\boldsymbol{M}^\natural\right\|\right)\left\|\boldsymbol{\Delta}^t\right\|\end{aligned}$$

where the second identity follows from the symmetry property (66). By choosing $\eta \leq 1/(2\sigma_{\max})$, one has $\boldsymbol{0} \preceq \boldsymbol{I}_r - 2\eta\boldsymbol{\Sigma}^\natural \preceq (1 - 2\eta\sigma_{\min})\,\boldsymbol{I}_r$ and $\boldsymbol{0} \preceq \boldsymbol{I}_r - 2\eta\boldsymbol{M}^\natural \preceq \boldsymbol{I}_r$, and further one can ensure

$$\gamma_1 \leq \frac{1}{2}\left[(1 - 2\eta\sigma_{\min}) + 1\right]\left\|\boldsymbol{\Delta}^t\right\| = (1 - \eta\sigma_{\min})\left\|\boldsymbol{\Delta}^t\right\|. \tag{67}$$

Next, regarding the higher order term γ_2, we can easily obtain

$$\gamma_2 \leq \eta\left\|\boldsymbol{\Delta}^t\right\|^2\left\|\boldsymbol{X}^\natural\right\|. \tag{68}$$

The bounds (67) and (68) taken collectively give

$$\beta_1 \leq (1 - \eta\sigma_{\min})\left\|\boldsymbol{\Delta}^t\right\| + \eta\left\|\boldsymbol{\Delta}^t\right\|^2\left\|\boldsymbol{X}^\natural\right\|. \tag{69}$$

(b) We now turn to the perturbation part β_2 by showing that

$$\begin{aligned}\frac{1}{\eta}\beta_2 &= \left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{\Delta}^t\boldsymbol{\Delta}^{t\top} + \boldsymbol{\Delta}^t\boldsymbol{X}^{\natural\top} + \boldsymbol{X}^\natural\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^\natural - \left[\boldsymbol{\Delta}^t\boldsymbol{\Delta}^{t\top} + \boldsymbol{\Delta}^t\boldsymbol{X}^{\natural\top} + \boldsymbol{X}^\natural\boldsymbol{\Delta}^{t\top}\right]\boldsymbol{X}^\natural\right\| \\ &\leq \underbrace{\left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{\Delta}^t\boldsymbol{X}^{\natural\top}\right)\boldsymbol{X}^\natural - \left(\boldsymbol{\Delta}^t\boldsymbol{X}^{\natural\top}\right)\boldsymbol{X}^\natural\right\|_{\mathrm{F}}}_{:=\theta_1} + \underbrace{\left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{X}^\natural\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^\natural - \left(\boldsymbol{X}^\natural\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^\natural\right\|_{\mathrm{F}}}_{:=\theta_2} \\ &\quad + \underbrace{\left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{\Delta}^t\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^\natural - \left(\boldsymbol{\Delta}^t\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^\natural\right\|_{\mathrm{F}}}_{:=\theta_3},\end{aligned} \tag{70}$$

where the last inequality holds due to the triangle inequality as well as the fact that $\|\boldsymbol{A}\| \leq \|\boldsymbol{A}\|_{\mathrm{F}}$. In the sequel, we shall bound the three terms separately.

- For the first term θ_1 in (70), the lth row of $\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{\Delta}^t\boldsymbol{X}^{\natural\top}\right)\boldsymbol{X}^\natural - \left(\boldsymbol{\Delta}^t\boldsymbol{X}^{\natural\top}\right)\boldsymbol{X}^\natural$ is given by

$$\frac{1}{p}\sum_{j=1}^{n}\left(\delta_{l,j} - p\right)\boldsymbol{\Delta}^t_{l,\cdot}\boldsymbol{X}^{\natural\top}_{j,\cdot}\boldsymbol{X}^\natural_{j,\cdot} = \boldsymbol{\Delta}^t_{l,\cdot}\left[\frac{1}{p}\sum_{j=1}^{n}\left(\delta_{l,j} - p\right)\boldsymbol{X}^{\natural\top}_{j,\cdot}\boldsymbol{X}^\natural_{j,\cdot}\right]$$

where, as usual, $\delta_{l,j} = \mathbb{1}_{\{(l,j)\in\Omega\}}$. Lemma 28 together with the union bound reveals that

$$\begin{aligned}\left\|\frac{1}{p}\sum_{j=1}^{n}\left(\delta_{l,j} - p\right)\boldsymbol{X}^{\natural\top}_{j,\cdot}\boldsymbol{X}^\natural_{j,\cdot}\right\| &\lesssim \frac{1}{p}\left(\sqrt{p\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}^2\left\|\boldsymbol{X}^\natural\right\|^2\log n} + \left\|\boldsymbol{X}^\natural\right\|_{2,\infty}^2\log n\right) \\ &\asymp \sqrt{\frac{\|\boldsymbol{X}^\natural\|_{2,\infty}^2\sigma_{\max}\log n}{p}} + \frac{\|\boldsymbol{X}^\natural\|_{2,\infty}^2\log n}{p}\end{aligned}$$

for all $1 \leq l \leq n$ with high probability. This gives

$$\begin{aligned}\left\|\boldsymbol{\Delta}^t_{l,\cdot}\left[\frac{1}{p}\sum_{j=1}^{n}\left(\delta_{l,j} - p\right)\boldsymbol{X}^{\natural\top}_{j,\cdot}\boldsymbol{X}^\natural_{j,\cdot}\right]\right\|_2 &\leq \left\|\boldsymbol{\Delta}^t_{l,\cdot}\right\|_2\left\|\frac{1}{p}\sum_{j}\left(\delta_{l,j} - p\right)\boldsymbol{X}^{\natural\top}_{j,\cdot}\boldsymbol{X}^\natural_{j,\cdot}\right\| \\ &\lesssim \left\|\boldsymbol{\Delta}^t_{l,\cdot}\right\|_2\left\{\sqrt{\frac{\|\boldsymbol{X}^\natural\|_{2,\infty}^2\sigma_{\max}\log n}{p}} + \frac{\|\boldsymbol{X}^\natural\|_{2,\infty}^2\log n}{p}\right\},\end{aligned}$$

which further reveals that

$$
\begin{aligned}
\theta_1 = \sqrt{\sum_{l=1}^{n}\left\|\frac{1}{p}\sum_j\left(\delta_{l,j}-p\right)\boldsymbol{\Delta}_{l,\cdot}^t\boldsymbol{X}_{j,\cdot}^{\natural\top}\boldsymbol{X}_{j,\cdot}^{\natural}\right\|_2^2} &\lesssim \left\|\boldsymbol{\Delta}^t\right\|_{\mathrm{F}}\left\{\sqrt{\frac{\|\boldsymbol{X}^\natural\|_{2,\infty}^2\sigma_{\max}\log n}{p}}+\frac{\|\boldsymbol{X}^\natural\|_{2,\infty}^2\log n}{p}\right\} \\
&\overset{\text{(i)}}{\lesssim} \left\|\boldsymbol{\Delta}^t\right\|\left\{\sqrt{\frac{\|\boldsymbol{X}^\natural\|_{2,\infty}^2 r\sigma_{\max}\log n}{p}}+\frac{\sqrt{r}\|\boldsymbol{X}^\natural\|_{2,\infty}^2\log n}{p}\right\} \\
&\overset{\text{(ii)}}{\lesssim} \left\|\boldsymbol{\Delta}^t\right\|\left\{\sqrt{\frac{\kappa\mu r^2\log n}{np}}+\frac{\kappa\mu r^{3/2}\log n}{np}\right\}\sigma_{\max} \\
&\overset{\text{(iii)}}{\le} \gamma\sigma_{\min}\left\|\boldsymbol{\Delta}^t\right\|,
\end{aligned}
$$

for arbitrarily small $\gamma > 0$. Here, (i) follows from $\|\boldsymbol{\Delta}^t\|_{\mathrm{F}} \le \sqrt{r}\,\|\boldsymbol{\Delta}^t\|$, (ii) holds owing to the incoherence condition (56), and (iii) follows as long as $n^2p \gg \kappa^3\mu r^2 n\log n$.

- For the second term θ_2 in (70), denote

$$
\boldsymbol{A} = \mathcal{P}_\Omega\left(\boldsymbol{X}^\natural\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^\natural - p\left(\boldsymbol{X}^\natural\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^\natural,
$$

whose lth row is given by

$$
\boldsymbol{A}_{l,\cdot} = \boldsymbol{X}_{l,\cdot}^\natural\sum_{j=1}^{n}\left(\delta_{l,j}-p\right)\boldsymbol{\Delta}_{j,\cdot}^{t\top}\boldsymbol{X}_{j,\cdot}^\natural. \tag{71}
$$

Recalling the induction hypotheses (24b) and (24c), we define

$$
\left\|\boldsymbol{\Delta}^t\right\|_{2,\infty} \le C_5\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} + C_8\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} := \xi \tag{72}
$$

$$
\left\|\boldsymbol{\Delta}^t\right\| \le C_9\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\natural\right\| + C_{10}\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\natural\right\| := \psi. \tag{73}
$$

With these two definitions in place, we now introduce a "truncation level"

$$
\omega := 2p\xi\sigma_{\max} \tag{74}
$$

that allows us to bound θ_2 in terms of the following two terms

$$
\theta_2 = \frac{1}{p}\sqrt{\sum_{l=1}^{n}\|\boldsymbol{A}_{l,\cdot}\|_2^2} \le \underbrace{\frac{1}{p}\sqrt{\sum_{l=1}^{n}\|\boldsymbol{A}_{l,\cdot}\|_2^2\,\mathbb{1}_{\{\|\boldsymbol{A}_{l,\cdot}\|_2\le\omega\}}}}_{:=\phi_1} + \underbrace{\frac{1}{p}\sqrt{\sum_{l=1}^{n}\|\boldsymbol{A}_{l,\cdot}\|_2^2\,\mathbb{1}_{\{\|\boldsymbol{A}_{l,\cdot}\|_2\ge\omega\}}}}_{:=\phi_2}.
$$

We will apply different strategies when upper bounding the terms ϕ_1 and ϕ_2, with their bounds given in the following two lemmas under the induction hypotheses (24b) and (24c).

Lemma 14. *Under the conditions in Lemma 9, there exist some constants $c, C > 0$ such that with probability exceeding $1 - c\exp(-Cnr\log n)$,*

$$
\phi_1 \lesssim \xi\sqrt{p\sigma_{\max}\|\boldsymbol{X}^\natural\|_{2,\infty}^2 nr\log^2 n} \tag{75}
$$

holds simultaneously for all $\boldsymbol{\Delta}^t$ obeying (72) and (73). Here, ξ is defined in (72).

Lemma 15. *Under the conditions in Lemma 9, with probability at least $1 - O\left(n^{-10}\right)$,*

$$
\phi_2 \lesssim \xi\sqrt{\kappa\mu r^2 p\log^2 n}\left\|\boldsymbol{X}^\natural\right\|^2 \tag{76}
$$

holds simultaneously for all $\boldsymbol{\Delta}^t$ obeying (72) and (73). Here, ξ is defined in (72).

The bounds (75) and (76) together with the incoherence condition (56) yield

$$\theta_2 \lesssim \frac{1}{p}\xi\sqrt{p\sigma_{\max}\|\boldsymbol{X}^\natural\|_{2,\infty}^2 nr\log^2 n} + \frac{1}{p}\xi\sqrt{\kappa\mu r^2 p\log^2 n}\left\|\boldsymbol{X}^\natural\right\|^2 \lesssim \sqrt{\frac{\kappa\mu r^2\log^2 n}{p}}\xi\sigma_{\max}.$$

- Next, we assert that the third term θ_3 in (70) has the same upper bound as θ_2. The proof follows by repeating the same argument used in bounding θ_2, and is hence omitted.

Take the previous three bounds on θ_1, θ_2 and θ_3 together to arrive at

$$\beta_2 \le \eta\left(|\theta_1|+|\theta_2|+|\theta_3|\right) \le \eta\gamma\sigma_{\min}\left\|\boldsymbol{\Delta}^t\right\| + \widetilde{C}\eta\sqrt{\frac{\kappa\mu r^2\log^2 n}{p}}\xi\sigma_{\max}$$

for some constant $\widetilde{C} > 0$.

(c) Substituting the preceding bounds on β_1 and β_2 into (65), we reach

$$\begin{aligned}
\alpha_2 &\overset{\text{(i)}}{\le} \left(1-\eta\sigma_{\min}+\eta\gamma\sigma_{\min}+\eta\left\|\boldsymbol{\Delta}^t\right\|\left\|\boldsymbol{X}^\natural\right\|\right)\left\|\boldsymbol{\Delta}^t\right\| + \eta\left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{E}\right)\right\|\left\|\boldsymbol{X}^\natural\right\| \\
&\quad + \widetilde{C}\eta\sqrt{\frac{\kappa\mu r^2\log^2 n}{p}}\sigma_{\max}\left(C_5\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} + C_8\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}\right) \\
&\overset{\text{(ii)}}{\le} \left(1-\frac{\sigma_{\min}}{2}\eta\right)\left\|\boldsymbol{\Delta}^t\right\| + \eta\left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{E}\right)\right\|\left\|\boldsymbol{X}^\natural\right\| \\
&\quad + \widetilde{C}\eta\sqrt{\frac{\kappa\mu r^2\log^2 n}{p}}\sigma_{\max}\left(C_5\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} + C_8\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}\right) \\
&\overset{\text{(iii)}}{\le} \left(1-\frac{\sigma_{\min}}{2}\eta\right)\left\|\boldsymbol{\Delta}^t\right\| + C\eta\sigma\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\natural\right\| \\
&\quad + \widetilde{C}\eta\sqrt{\frac{\kappa^2\mu^2 r^3\log^3 n}{np}}\sigma_{\max}\left(C_5\rho^t\mu r\sqrt{\frac{1}{np}} + C_8\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right)\left\|\boldsymbol{X}^\natural\right\|
\end{aligned} \tag{77}$$

for some constant $C > 0$. Here, (i) uses the definition of ξ (cf. (72)), (ii) holds if γ is small enough and $\|\boldsymbol{\Delta}^t\|\left\|\boldsymbol{X}^\natural\right\| \ll \sigma_{\min}$, and (iii) follows from Lemma 27 as well as the incoherence condition (56). An immediate consequence of (77) is that under the sample size condition and the noise condition of this lemma, one has

$$\left\|\widetilde{\boldsymbol{X}}^{t+1} - \boldsymbol{X}^\natural\right\|\left\|\boldsymbol{X}^\natural\right\| \le \sigma_{\min}/2 \tag{78}$$

if $0 < \eta \le 1/\sigma_{\max}$.

2. We then move on to the first term α_1 in (63), which can be rewritten as

$$\alpha_1 = \left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^t\boldsymbol{R}_1 - \widetilde{\boldsymbol{X}}^{t+1}\right\|,$$

with

$$\boldsymbol{R}_1 = \left(\widehat{\boldsymbol{H}}^t\right)^{-1}\widehat{\boldsymbol{H}}^{t+1} := \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^t\boldsymbol{R} - \boldsymbol{X}^\natural\right\|_{\mathrm{F}}. \tag{79}$$

(a) First, we claim that $\widetilde{\boldsymbol{X}}^{t+1}$ satisfies

$$\boldsymbol{I}_r = \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\left\|\widetilde{\boldsymbol{X}}^{t+1}\boldsymbol{R} - \boldsymbol{X}^\natural\right\|_{\mathrm{F}}, \tag{80}$$

meaning that $\widetilde{\boldsymbol{X}}^{t+1}$ is already rotated to the direction that is most "aligned" with $\boldsymbol{X}^\natural$. This important property eases the analysis. In fact, in view of Lemma 22, (80) follows if one can show that $\boldsymbol{X}^{\natural\top}\widetilde{\boldsymbol{X}}^{t+1}$ is symmetric and positive semidefinite. First of all, it follows from Lemma 22 that $\boldsymbol{X}^{\natural\top}\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t$ is symmetric and, hence, by definition,

$$\boldsymbol{X}^{\natural\top}\widetilde{\boldsymbol{X}}^{t+1} = \boldsymbol{X}^{\natural\top}\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \frac{\eta}{p}\boldsymbol{X}^{\natural\top}\mathcal{P}_\Omega\left[\boldsymbol{X}^t\boldsymbol{X}^{t\top} - \left(\boldsymbol{M}^\natural + \boldsymbol{E}\right)\right]\boldsymbol{X}^\natural$$

is also symmetric. Additionally,

$$\left\|\boldsymbol{X}^{\natural\top}\widetilde{\boldsymbol{X}}^{t+1} - \boldsymbol{M}^\natural\right\| \le \left\|\widetilde{\boldsymbol{X}}^{t+1} - \boldsymbol{X}^\natural\right\|\left\|\boldsymbol{X}^\natural\right\| \le \sigma_{\min}/2,$$

where the second inequality holds according to (78). Weyl's inequality guarantees that

$$\boldsymbol{X}^{\natural\top}\widetilde{\boldsymbol{X}}^{t+1} \succeq \frac{1}{2}\sigma_{\min}\boldsymbol{I}_r,$$

thus justifying (80) via Lemma 22.

(b) With (79) and (80) in place, we resort to Lemma 24 to establish the bound. Specifically, take $\boldsymbol{X}_1 = \widetilde{\boldsymbol{X}}^{t+1}$ and $\boldsymbol{X}_2 = \boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^t$, and it comes from (78) that

$$\left\|\boldsymbol{X}_1 - \boldsymbol{X}^\natural\right\|\left\|\boldsymbol{X}^\natural\right\| \le \sigma_{\min}/2.$$

Moreover, we have

$$\|\boldsymbol{X}_1 - \boldsymbol{X}_2\|\left\|\boldsymbol{X}^\natural\right\| = \left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^t - \widetilde{\boldsymbol{X}}^{t+1}\right\|\left\|\boldsymbol{X}^\natural\right\|,$$

in which

$$\begin{aligned}\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^t - \widetilde{\boldsymbol{X}}^{t+1} &= \left(\boldsymbol{X}^t - \eta\frac{1}{p}\mathcal{P}_\Omega\left[\boldsymbol{X}^t\boldsymbol{X}^{t\top} - \left(\boldsymbol{M}^\natural + \boldsymbol{E}\right)\right]\boldsymbol{X}^t\right)\widehat{\boldsymbol{H}}^t \\ &\quad - \left[\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \eta\frac{1}{p}\mathcal{P}_\Omega\left[\boldsymbol{X}^t\boldsymbol{X}^{t\top} - \left(\boldsymbol{M}^\natural + \boldsymbol{E}\right)\right]\boldsymbol{X}^\natural\right] \\ &= -\eta\frac{1}{p}\mathcal{P}_\Omega\left[\boldsymbol{X}^t\boldsymbol{X}^{t\top} - \left(\boldsymbol{M}^\natural + \boldsymbol{E}\right)\right]\left(\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\right).\end{aligned}$$

This allows one to derive

$$\begin{aligned}\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^t - \widetilde{\boldsymbol{X}}^{t+1}\right\| &\le \eta\left\|\frac{1}{p}\mathcal{P}_\Omega\left[\boldsymbol{X}^t\boldsymbol{X}^{t\top} - \boldsymbol{M}^\natural\right]\left(\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\right)\right\| + \eta\left\|\frac{1}{p}\mathcal{P}_\Omega(\boldsymbol{E})\left(\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\right)\right\| \\ &\le \eta\left(2n\left\|\boldsymbol{\Delta}^t\right\|_{2,\infty}^2 + 4\sqrt{n}\log n\left\|\boldsymbol{\Delta}^t\right\|_{2,\infty}\left\|\boldsymbol{X}^\natural\right\| + C\sigma\sqrt{\frac{n}{p}}\right)\left\|\boldsymbol{\Delta}^t\right\| \qquad (81)\end{aligned}$$

for some absolute constant $C > 0$. Here the last inequality follows from Lemma 27 and Lemma 30. As a consequence,

$$\|\boldsymbol{X}_1 - \boldsymbol{X}_2\|\left\|\boldsymbol{X}^\natural\right\| \le \eta\left(2n\left\|\boldsymbol{\Delta}^t\right\|_{2,\infty}^2 + 4\sqrt{n}\log n\left\|\boldsymbol{\Delta}^t\right\|_{2,\infty}\left\|\boldsymbol{X}^\natural\right\| + C\sigma\sqrt{\frac{n}{p}}\right)\left\|\boldsymbol{\Delta}^t\right\|\left\|\boldsymbol{X}^\natural\right\|.$$

Under our sample size condition and the noise condition (24) and the induction hypotheses (24), one can show

$$\|\boldsymbol{X}_1 - \boldsymbol{X}_2\|\left\|\boldsymbol{X}^\natural\right\| \le \sigma_{\min}/4.$$

Apply Lemma 24 and (81) to reach

$$\begin{aligned}\alpha_1 &\le 5\kappa\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^t - \widetilde{\boldsymbol{X}}^{t+1}\right\| \\ &\le 5\kappa\eta\left(2n\left\|\boldsymbol{\Delta}^t\right\|_{2,\infty}^2 + 2\sqrt{n}\log n\left\|\boldsymbol{\Delta}^t\right\|_{2,\infty}\left\|\boldsymbol{X}^\natural\right\| + C\sigma\sqrt{\frac{n}{p}}\right)\left\|\boldsymbol{\Delta}^t\right\|.\end{aligned}$$

3. Combining the above bounds on α_1 and α_2, we arrive at

$$\begin{aligned}\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1}-\boldsymbol{X}^{\natural}\right\| &\le \left(1-\frac{\sigma_{\min}}{2}\eta\right)\left\|\boldsymbol{\Delta}^{t}\right\|+\eta C\sigma\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^{\natural}\right\| \\ &\quad+\widetilde{C}\eta\sqrt{\frac{\kappa^{2}\mu^{2}r^{3}\log^{3}n}{np}}\sigma_{\max}\left(C_{5}\rho^{t}\mu r\sqrt{\frac{1}{np}}+\frac{C_{8}}{\sigma_{\min}}\sigma\sqrt{\frac{n}{p}}\right)\left\|\boldsymbol{X}^{\natural}\right\| \\ &\quad+5\eta\kappa\left(2n\left\|\boldsymbol{\Delta}^{t}\right\|_{2,\infty}^{2}+2\sqrt{n}\log n\left\|\boldsymbol{\Delta}^{t}\right\|_{2,\infty}\left\|\boldsymbol{X}^{\natural}\right\|+C\sigma\sqrt{\frac{n}{p}}\right)\left\|\boldsymbol{\Delta}^{t}\right\| \\ &\le C_{9}\rho^{t+1}\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^{\natural}\right\|+C_{10}\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|,\end{aligned}$$

with the proviso that $\rho\ge 1-(\sigma_{\min}/3)\cdot\eta$, κ is a constant, and $n^{2}p\gg\mu^{3}r^{3}n\log^{3}n$.

5.3.1 Proof of Lemma 14

In what follows, we first assume that the $\delta_{j,k}$'s are independent, and then use the standard decoupling trick to extend the result to symmetric sampling case (i.e. $\delta_{j,k}=\delta_{k,j}$).

To begin with, we justify the concentration bound for any $\boldsymbol{\Delta}^{t}$ independent of Ω, followed by the standard covering argument that extends the bound to all $\boldsymbol{\Delta}^{t}$. For any $\boldsymbol{\Delta}^{t}$ independent of Ω, one has

$$B:=\max_{1\le j\le n}\left\|\boldsymbol{X}_{l,\cdot}^{\natural}\left(\delta_{l,j}-p\right)\boldsymbol{\Delta}_{j,\cdot}^{t\top}\boldsymbol{X}_{j,\cdot}^{\natural}\right\|_{2}\le\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}^{2}\xi$$

$$\begin{aligned}\text{and}\qquad V&:=\left\|\mathbb{E}\left[\sum_{j=1}^{n}\left(\delta_{l,j}-p\right)^{2}\boldsymbol{X}_{l,\cdot}^{\natural}\boldsymbol{\Delta}_{j,\cdot}^{t\top}\boldsymbol{X}_{j,\cdot}^{\natural}\left(\boldsymbol{X}_{l,\cdot}^{\natural}\boldsymbol{\Delta}_{j,\cdot}^{t\top}\boldsymbol{X}_{j,\cdot}^{\natural}\right)^{\top}\right]\right\| \\ &\le p\left\|\boldsymbol{X}_{l,\cdot}^{\natural}\right\|_{2}^{2}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}^{2}\left\|\sum_{j=1}^{n}\boldsymbol{\Delta}_{j,\cdot}^{t\top}\boldsymbol{\Delta}_{j,\cdot}^{t}\right\| \\ &\le p\left\|\boldsymbol{X}_{l,\cdot}^{\natural}\right\|_{2}^{2}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}^{2}\psi^{2} \\ &\le 2p\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}^{2}\xi^{2}\sigma_{\max},\end{aligned}$$

where ξ and ψ are defined respectively in (72) and (73). Here, the last line makes use of the fact that

$$\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\psi\ll\xi\left\|\boldsymbol{X}^{\natural}\right\|=\xi\sqrt{\sigma_{\max}}, \tag{82}$$

as long as n is sufficiently large. Apply the matrix Bernstein inequality [**?**, Theorem 6.1.1] to get

$$\begin{aligned}\mathbb{P}\left\{\left\|\boldsymbol{A}_{l,\cdot}\right\|_{2}\ge t\right\}&\le 2r\exp\left(-\frac{ct^{2}}{2p\xi^{2}\sigma_{\max}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}^{2}+t\cdot\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}^{2}\xi}\right) \\ &\le 2r\exp\left(-\frac{ct^{2}}{4p\xi^{2}\sigma_{\max}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}^{2}}\right)\end{aligned}$$

for some constant $c>0$, provided that

$$t\le 2p\sigma_{\max}\xi.$$

This upper bound on t is exactly the truncation level ω we introduce in (74). With this in mind, we can easily verify that

$$\left\|\boldsymbol{A}_{l,\cdot}\right\|_{2}\mathbb{1}_{\left\{\left\|\boldsymbol{A}_{l,\cdot}\right\|_{2}\le\omega\right\}}$$

is a sub-Gaussian random variable with variance proxy not exceeding $O\left(p\xi^{2}\sigma_{\max}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}^{2}\log r\right)$. Therefore, invoking the concentration bounds for quadratic functions [**?**, Theorem 2.1] yields that for some constants

$C_0, C > 0$, with probability at least $1 - C_0 e^{-Cnr \log n}$,

$$\phi_1^2 = \sum_{l=1}^{n} \|\boldsymbol{A}_{l,\cdot}\|_2^2 \, \mathbb{1}_{\left\{\|\boldsymbol{A}_{l,\cdot}\|_2 \le \omega\right\}} \lesssim p\xi^2 \sigma_{\max} \|\boldsymbol{X}^\natural\|_{2,\infty}^2 nr \log^2 n.$$

Now that we have established an upper bound on any fixed matrix $\boldsymbol{\Delta}^t$ (which holds with exponentially high probability), we can proceed to invoke the standard epsilon-net argument to establish a uniform bound over all feasible $\boldsymbol{\Delta}^t$. This argument is fairly standard, and is thus omitted; see [**?**, Section 2.3.1] or the proof of Lemma 29. In conclusion, we have that with probability exceeding $1 - C_0 e^{-\frac{1}{2}Cnr \log n}$,

$$\phi_1 = \sqrt{\sum_{l=1}^{n} \|\boldsymbol{A}_{l,\cdot}\|_2^2 \, \mathbb{1}_{\left\{\|\boldsymbol{A}_{l,\cdot}\|_2 \le \omega\right\}}} \lesssim \sqrt{p\xi^2 \sigma_{\max} \|\boldsymbol{X}^\natural\|_{2,\infty}^2 nr \log^2 n} \tag{83}$$

holds simultaneously for all $\boldsymbol{\Delta}^t \in \mathbb{R}^{n \times r}$ obeying the conditions of the lemma.

In the end, we comment on how to extend the bound to the symmetric sampling pattern where $\delta_{j,k} = \delta_{k,j}$. Recall from (71) that the diagonal element $\delta_{l,l}$ cannot change the ℓ_2 norm of $\boldsymbol{A}_{l,\cdot}$ by more than $\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}^2 \xi$. As a result, changing all the diagonals $\{\delta_{l,l}\}$ cannot change the quantity of interest (i.e. ϕ_1) by more than $\sqrt{n} \left\|\boldsymbol{X}^\natural\right\|_{2,\infty}^2 \xi$. This is smaller than the right hand side of (83) under our incoherence and sample size conditions. Hence from now on we ignore the effect of $\{\delta_{l,l}\}$ and focus on off-diagonal terms. The proof then follows from the same argument as in [**?**, Theorem D.2]. More specifically, we can employ the construction of Bernoulli random variables introduced therein to demonstrate that the upper bound in (83) still holds if the indicator $\delta_{i,j}$ is replaced by $(\tau_{i,j} + \tau'_{i,j})/2$, where $\tau_{i,j}$ and $\tau'_{i,j}$ are independent copies of the symmetric Bernoulli random variables. Recognizing that $\sup_{\boldsymbol{\Delta}^t} \phi_1$ is a norm of the Bernoulli random variables $\tau_{i,j}$, one can repeat the decoupling argument in [**?**, Claim D.3] to finish the proof. We omit the details here for brevity.

5.3.2 Proof of Lemma 15

Observe from (71) that

$$\begin{aligned}
\|\boldsymbol{A}_{l,\cdot}\|_2 &\le \left\|\boldsymbol{X}^\natural\right\|_{2,\infty} \left\| \sum_{j=1}^{n} (\delta_{l,j} - p)\, \boldsymbol{\Delta}_{j,\cdot}^{t\top} \boldsymbol{X}_{j,\cdot}^\natural \right\| & (84)\\
&\le \left\|\boldsymbol{X}^\natural\right\|_{2,\infty} \left(\left\| \sum_{j=1}^{n} \delta_{l,j} \boldsymbol{\Delta}_{j,\cdot}^{t\top} \boldsymbol{X}_{j,\cdot}^\natural \right\| + p \left\|\boldsymbol{\Delta}^t\right\| \left\|\boldsymbol{X}^\natural\right\| \right) \\
&\le \left\|\boldsymbol{X}^\natural\right\|_{2,\infty} \left(\left\| \left[\delta_{l,1} \boldsymbol{\Delta}_{1,\cdot}^{t\top}, \cdots, \delta_{l,n} \boldsymbol{\Delta}_{n,\cdot}^{t\top}\right] \right\| \left\| \begin{bmatrix} \delta_{l,1} \boldsymbol{X}_{1,\cdot}^\natural \\ \vdots \\ \delta_{l,n} \boldsymbol{X}_{n,\cdot}^\natural \end{bmatrix} \right\| + p\psi \left\|\boldsymbol{X}^\natural\right\| \right) \\
&\le \left\|\boldsymbol{X}^\natural\right\|_{2,\infty} \left(\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}^t\right)\right\| \cdot 1.2\sqrt{p} \left\|\boldsymbol{X}^\natural\right\| + p\psi \left\|\boldsymbol{X}^\natural\right\| \right), & (85)
\end{aligned}$$

where ψ is as defined in (73) and $\boldsymbol{G}_l(\cdot)$ is as defined in Lemma 28. Here, the last inequality follows from Lemma 28, namely, for some constant $C > 0$, the following holds with probability at least $1 - O(n^{-10})$

$$\begin{aligned}
\left\| \begin{bmatrix} \delta_{l,1} \boldsymbol{X}_{1,\cdot}^\natural \\ \vdots \\ \delta_{l,n} \boldsymbol{X}_{n,\cdot}^\natural \end{bmatrix} \right\| &\le \left(p \left\|\boldsymbol{X}^\natural\right\|^2 + C\sqrt{p \|\boldsymbol{X}^\natural\|_{2,\infty}^2 \left\|\boldsymbol{X}^\natural\right\|^2 \log n} + C \|\boldsymbol{X}^\natural\|_{2,\infty}^2 \log n \right)^{\frac{1}{2}} \\
&\le \left(p + C\sqrt{p \frac{\kappa\mu r}{n} \log n} + C \frac{\kappa\mu r \log n}{n} \right)^{\frac{1}{2}} \left\|\boldsymbol{X}^\natural\right\| \le 1.2\sqrt{p} \left\|\boldsymbol{X}^\natural\right\|, & (86)
\end{aligned}$$

where we also use the incoherence condition (56) and the sample complexity condition $n^2p \gg \kappa\mu rn\log n$. Hence, the event

$$\left\|\boldsymbol{A}_{l,\cdot}\right\|_2 \geq \omega = 2p\sigma_{\max}\xi$$

together with (84) and (85) necessarily implies that

$$\left\|\sum_{j=1}^{n}\left(\delta_{l,j}-p\right)\boldsymbol{\Delta}_{j,\cdot}^{t\top}\boldsymbol{X}_{j,\cdot}^{\natural}\right\| \geq 2p\sigma_{\max}\frac{\xi}{\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}} \qquad \text{and}$$

$$\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}^t\right)\right\| \geq \frac{\frac{2p\sigma_{\max}\xi}{\left\|\boldsymbol{X}^{\natural}\right\|\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}}-p\psi}{1.2\sqrt{p}} \geq \frac{\frac{2\sqrt{p}\left\|\boldsymbol{X}^{\natural}\right\|\xi}{\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}}-\sqrt{p}\psi}{1.2} \geq 1.5\sqrt{p}\frac{\xi}{\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}}\left\|\boldsymbol{X}^{\natural}\right\|,$$

where the last inequality follows from the bound (82). As a result, with probability at least $1-O(n^{-10})$ (i.e. when (86) holds for all l's) we can upper bound ϕ_2 by

$$\phi_2 = \sqrt{\sum_{l=1}^{n}\left\|\boldsymbol{A}_{l,\cdot}\right\|_2^2 \mathbb{1}_{\left\{\left\|\boldsymbol{A}_{l,\cdot}\right\|_2\geq\omega\right\}}} \leq \sqrt{\sum_{l=1}^{n}\left\|\boldsymbol{A}_{l,\cdot}\right\|_2^2 \mathbb{1}_{\left\{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta}^t)\right\|\geq\frac{1.5\sqrt{p}\xi\sqrt{\sigma_{\max}}}{\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}}\right\}}},$$

where the indicator functions are now specified with respect to $\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}^t\right)\right\|$.

Next, we divide into multiple cases based on the size of $\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}^t\right)\right\|$. By Lemma 29, for some constants $c_1, c_2 > 0$, with probability at least $1-c_1\exp\left(-c_2nr\log n\right)$,

$$\sum_{l=1}^{n}\mathbb{1}_{\left\{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta}^t)\right\|\geq4\sqrt{p}\psi+\sqrt{2^k r}\xi\right\}} \leq \frac{\alpha n}{2^{k-3}} \tag{87}$$

for any $k \geq 0$ and any $\alpha \gtrsim \log n$. We claim that it suffices to consider the set of sufficiently large k obeying

$$\sqrt{2^k r}\xi \geq 4\sqrt{p}\psi \quad \text{or equivalently} \quad k \geq \log\frac{16p\psi^2}{r\xi^2}; \tag{88}$$

otherwise we can use (82) to obtain

$$4\sqrt{p}\psi+\sqrt{2^k r}\xi \leq 8\sqrt{p}\psi \ll 1.5\sqrt{p}\frac{\xi}{\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}}\left\|\boldsymbol{X}^{\natural}\right\|,$$

which contradicts the event $\left\|\boldsymbol{A}_{l,\cdot}\right\|_2 \geq \omega$. Consequently, we divide all indices into the following sets

$$S_k = \left\{1 \leq l \leq n : \left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}^t\right)\right\| \in \left(\sqrt{2^k r}\xi, \sqrt{2^{k+1} r}\xi\right]\right\} \tag{89}$$

defined for each integer k obeying (88). Under the condition (88), it follows from (87) that

$$\sum_{l=1}^{n}\mathbb{1}_{\left\{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta}^t)\right\|\geq\sqrt{2^{k+2}r}\xi\right\}} \leq \sum_{l=1}^{n}\mathbb{1}_{\left\{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta}^t)\right\|\geq4\sqrt{p}\psi+\sqrt{2^k r}\xi\right\}} \leq \frac{\alpha n}{2^{k-3}},$$

meaning that the cardinality of S_k satisfies

$$\left|S_{k+2}\right| \leq \frac{\alpha n}{2^{k-3}} \qquad \text{or} \qquad \left|S_k\right| \leq \frac{\alpha n}{2^{k-5}}$$

which decays exponentially fast as k increases. Therefore, when restricting attention to the set of indices within S_k, we can obtain

$$\sqrt{\sum_{l\in S_k}\left\|\boldsymbol{A}_{l,\cdot}\right\|_2^2} \overset{\text{(i)}}{\leq} \sqrt{\left|S_k\right|\cdot\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}^2\left(1.2\sqrt{2^{k+1}r}\xi\sqrt{p}\left\|\boldsymbol{X}^{\natural}\right\|+p\psi\left\|\boldsymbol{X}^{\natural}\right\|\right)^2}$$

$$
\begin{aligned}
&\leq \sqrt{\frac{\alpha n}{2^{k-5}}} \left\| \boldsymbol{X}^\natural \right\|_{2,\infty} \left(2\sqrt{2^{k+1} r}\xi\sqrt{p} \left\| \boldsymbol{X}^\natural \right\| + p\psi \left\| \boldsymbol{X}^\natural \right\| \right) \\
&\overset{\text{(ii)}}{\leq} 4\sqrt{\frac{\alpha n}{2^{k-5}}} \left\| \boldsymbol{X}^\natural \right\|_{2,\infty} \sqrt{2^{k+1} r}\xi\sqrt{p} \left\| \boldsymbol{X}^\natural \right\| \\
&\overset{\text{(iii)}}{\leq} 32\sqrt{\alpha\kappa\mu r^2 p}\xi \left\| \boldsymbol{X}^\natural \right\|^2,
\end{aligned}
$$

where (i) follows from the bound (85) and the constraint (89) in S_k, (ii) is a consequence of (88) and (iii) uses the incoherence condition (56).

Now that we have developed an upper bound with respect to each S_k, we can add them up to yield the final upper bound. Note that there are in total no more than $O(\log n)$ different sets, i.e. $S_k = \emptyset$ if $k \geq c_1 \log n$ for c_1 sufficiently large. This arises since

$$
\|\boldsymbol{G}_l(\boldsymbol{\Delta}^t)\| \leq \|\boldsymbol{\Delta}^t\|_{\mathrm{F}} \leq \sqrt{n}\|\boldsymbol{\Delta}^t\|_{2,\infty} \leq \sqrt{n}\xi \leq \sqrt{n}\sqrt{r}\xi
$$

and hence

$$
\mathbb{1}_{\left\{\|\boldsymbol{G}_l(\boldsymbol{\Delta}^t)\| \geq 4\sqrt{p}\psi + \sqrt{2^k r}\xi\right\}} = 0 \qquad \text{and} \qquad S_k = \emptyset
$$

if $k/\log n$ is sufficiently large. One can thus conclude that

$$
\phi_2^2 \leq \sum_{k=\log\frac{16p\psi^2}{r\xi^2}}^{c_1 \log n} \sum_{l \in S_k} \|\boldsymbol{A}_{l,\cdot}\|_2^2 \lesssim \left(\sqrt{\alpha\kappa\mu r^2 p}\xi \left\| \boldsymbol{X}^\natural \right\|^2 \right)^2 \cdot \log n,
$$

leading to $\phi_2 \lesssim \xi\sqrt{\alpha\kappa\mu r^2 p \log n} \left\| \boldsymbol{X}^\natural \right\|^2$. The proof is finished by taking $\alpha = c\log n$ for some sufficiently large constant $c > 0$.

5.4 Proof of Lemma 10

1. To obtain (36a), we invoke Lemma 24. Setting $\boldsymbol{X}_1 = \boldsymbol{X}^t \widehat{\boldsymbol{H}}^t$ and $\boldsymbol{X}_2 = \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)}$, we get

$$
\left\| \boldsymbol{X}_1 - \boldsymbol{X}^\natural \right\| \left\| \boldsymbol{X}^\natural \right\| \overset{\text{(i)}}{\leq} C_9 \rho^t \mu r \frac{1}{\sqrt{np}} \sigma_{\max} + \frac{C_{10}}{\sigma_{\min}} \sigma \sqrt{\frac{n \log n}{p}} \sigma_{\max} \overset{\text{(ii)}}{\leq} \frac{1}{2} \sigma_{\min},
$$

where (i) follows from (33c) and (ii) holds as long as $n^2 p \gg \kappa^2 \mu^2 r^2 n$ and the noise satisfies (24). In addition,

$$
\begin{aligned}
\|\boldsymbol{X}_1 - \boldsymbol{X}_2\| \left\| \boldsymbol{X}^\natural \right\| &\leq \|\boldsymbol{X}_1 - \boldsymbol{X}_2\|_{\mathrm{F}} \left\| \boldsymbol{X}^\natural \right\| \\
&\overset{\text{(i)}}{\leq} \left(C_3 \rho^t \mu r \sqrt{\frac{\log n}{np}} \left\| \boldsymbol{X}^\natural \right\|_{2,\infty} + \frac{C_7}{\sigma_{\min}} \sigma \sqrt{\frac{n \log n}{p}} \left\| \boldsymbol{X}^\natural \right\|_{2,\infty} \right) \left\| \boldsymbol{X}^\natural \right\| \\
&\overset{\text{(ii)}}{\leq} C_3 \rho^t \mu r \sqrt{\frac{\log n}{np}} \sigma_{\max} + \frac{C_7}{\sigma_{\min}} \sigma \sqrt{\frac{n \log n}{p}} \sigma_{\max} \\
&\overset{\text{(iii)}}{\leq} \frac{1}{2} \sigma_{\min},
\end{aligned}
$$

where (i) utilizes (33d), (ii) follows since $\left\| \boldsymbol{X}^\natural \right\|_{2,\infty} \leq \left\| \boldsymbol{X}^\natural \right\|$, and (iii) holds if $n^2 p \gg \kappa^2 \mu^2 r^2 n \log n$ and the noise satisfies (24). With these in place, Lemma 24 immediately yields (36a).

2. The first inequality in (36b) follows directly from the definition of $\widehat{\boldsymbol{H}}^{t,(l)}$. The second inequality is concerned with the estimation error of $\boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)}$ with respect to the Frobenius norm. Combining (33a), (33d) and the triangle inequality yields

$$
\left\| \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} - \boldsymbol{X}^\natural \right\|_{\mathrm{F}} \leq \left\| \boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \right\|_{\mathrm{F}} + \left\| \boldsymbol{X}^t \widehat{\boldsymbol{H}}^t - \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} \right\|_{\mathrm{F}}
$$

$$\leq C_4\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}}+\frac{C_1\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}}+C_3\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}+\frac{C_7\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}$$

$$\leq C_4\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}}+\frac{C_1\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}}+C_3\rho^t\mu r\sqrt{\frac{\log n}{np}}\sqrt{\frac{\kappa\mu}{n}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}}+\frac{C_7\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\sqrt{\frac{\kappa\mu}{n}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}}$$

$$\leq 2C_4\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}}+\frac{2C_1\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}}, \tag{90}$$

where the last step holds true as long as $n \gg \kappa\mu\log n$.

3. To obtain (36c), we use (33d) and (33b) to get

$$\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}-\boldsymbol{X}^\natural\right\|_{2,\infty}\leq\left\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t-\boldsymbol{X}^\natural\right\|_{2,\infty}+\left\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\|_{\mathrm{F}}$$

$$\leq C_5\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}+\frac{C_8\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}+C_3\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}+\frac{C_7\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}$$

$$\leq\left(C_3+C_5\right)\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}+\frac{C_8+C_7}{\sigma_{\min}}\sigma\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}.$$

4. Finally, to obtain (36d), one can take the triangle inequality

$$\begin{aligned}\left\|\boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)}-\boldsymbol{X}^\natural\right\|&\leq\left\|\boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)}-\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t\right\|_{\mathrm{F}}+\left\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t-\boldsymbol{X}^\natural\right\|\\&\leq 5\kappa\left\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\|_{\mathrm{F}}+\left\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t-\boldsymbol{X}^\natural\right\|,\end{aligned}$$

where the second line follows from (36a). Combine (33d) and (33c) to yield

$$\left\|\boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)}-\boldsymbol{X}^\natural\right\|$$

$$\leq 5\kappa\left(C_3\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}+\frac{C_7}{\sigma_{\min}}\sigma\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}\right)+C_9\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\natural\right\|+\frac{C_{10}}{\sigma_{\min}}\sigma\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\natural\right\|$$

$$\leq 5\kappa\sqrt{\frac{\kappa\mu r}{n}}\left\|\boldsymbol{X}^\natural\right\|\left(C_3\rho^t\mu r\sqrt{\frac{\log n}{np}}+\frac{C_7}{\sigma_{\min}}\sigma\sqrt{\frac{n\log n}{p}}\right)+C_9\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\natural\right\|+\frac{C_{10}\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\natural\right\|$$

$$\leq 2C_9\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\natural\right\|+\frac{2C_{10}\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\natural\right\|,$$

where the second inequality uses the incoherence of $\boldsymbol{X}^\natural$ (cf. (56)) and the last inequality holds as long as $n \gg \kappa^3\mu r\log n$.

5.5 Proof of Lemma 11

From the definition of $\boldsymbol{R}^{t+1,(l)}$ (see (35)), we must have

$$\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1}-\boldsymbol{X}^{t+1,(l)}\boldsymbol{R}^{t+1,(l)}\right\|_{\mathrm{F}}\leq\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{t+1,(l)}\boldsymbol{R}^{t,(l)}\right\|_{\mathrm{F}}.$$

The gradient update rules in (21) and (32) allow one to express

$$\begin{aligned}\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^t-\boldsymbol{X}^{t+1,(l)}\boldsymbol{R}^{t,(l)}&=\left[\boldsymbol{X}^t-\eta\nabla f\left(\boldsymbol{X}^t\right)\right]\widehat{\boldsymbol{H}}^t-\left[\boldsymbol{X}^{t,(l)}-\eta\nabla f^{(l)}\big(\boldsymbol{X}^{t,(l)}\big)\right]\boldsymbol{R}^{t,(l)}\\&=\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t-\eta\nabla f\big(\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t\big)-\left[\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}-\eta\nabla f^{(l)}\big(\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\big)\right]\\&=\big(\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\big)-\eta\left[\nabla f\big(\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t\big)-\nabla f\big(\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\big)\right]\end{aligned}$$

$$-\eta\left[\nabla f\big(\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\big)-\nabla f^{(l)}\big(\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\big)\right],$$

where we have again used the fact that $\nabla f\left(\boldsymbol{X}^{t}\right)\boldsymbol{R}=\nabla f(\boldsymbol{X}^{t}\boldsymbol{R})$ for any orthonormal matrix $\boldsymbol{R}\in\mathcal{O}^{r\times r}$ (similarly for $\nabla f^{(l)}\big(\boldsymbol{X}^{t,(l)}\big)$). Relate the right-hand side of the above equation with $\nabla f_{\text{clean}}\left(\boldsymbol{X}\right)$ to reach

$$\begin{aligned}\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{t+1,(l)}\boldsymbol{R}^{t,(l)}=&\underbrace{\big(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\big)-\eta\left[\nabla f_{\text{clean}}\big(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}\big)-\nabla f_{\text{clean}}\big(\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\big)\right]}_{:=\boldsymbol{B}_{1}^{(l)}}\\&\underbrace{-\eta\left[\frac{1}{p}\mathcal{P}_{\Omega_{l}}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top}-\boldsymbol{M}^{\natural}\right)-\mathcal{P}_{l}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top}-\boldsymbol{M}^{\natural}\right)\right]\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}}_{:=\boldsymbol{B}_{2}^{(l)}}\\&\underbrace{+\eta\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{E}\right)\left(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right)}_{:=\boldsymbol{B}_{3}^{(l)}}\underbrace{+\eta\frac{1}{p}\mathcal{P}_{\Omega_{l}}\left(\boldsymbol{E}\right)\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}}_{:=\boldsymbol{B}_{4}^{(l)}},\end{aligned}\tag{91}$$

where we have used the following relationship between $\nabla f^{(l)}\left(\boldsymbol{X}\right)$ and $\nabla f\left(\boldsymbol{X}\right)$:

$$\nabla f^{(l)}\left(\boldsymbol{X}\right)=\nabla f\left(\boldsymbol{X}\right)-\frac{1}{p}\mathcal{P}_{\Omega_{l}}\left[\boldsymbol{X}\boldsymbol{X}^{\top}-\left(\boldsymbol{M}^{\natural}+\boldsymbol{E}\right)\right]\boldsymbol{X}+\mathcal{P}_{l}\left(\boldsymbol{X}\boldsymbol{X}^{\top}-\boldsymbol{M}^{\natural}\right)\boldsymbol{X}\tag{92}$$

for all $\boldsymbol{X}\in\mathbb{R}^{n\times r}$ with $\mathcal{P}_{\Omega_{l}}$ and $\mathcal{P}_{l}$ defined respectively in (29) and (30). In the sequel, we control the four terms in reverse order.

1. The last term $\boldsymbol{B}_{4}^{(l)}$ is controlled via the following lemma.

   Lemma 16. *Suppose that the sample size obeys $n^{2}p>C\mu^{2}r^{2}n\log^{2}n$ for some sufficiently large constant $C>0$. Then with probability at least $1-O\left(n^{-10}\right)$, the matrix $\boldsymbol{B}_{4}^{(l)}$ as defined in (91) satisfies*

   $$\left\Vert \boldsymbol{B}_{4}^{(l)}\right\Vert _{\mathrm{F}}\lesssim\eta\sigma\sqrt{\frac{n\log n}{p}}\left\Vert \boldsymbol{X}^{\natural}\right\Vert _{2,\infty}.$$

2. The third term $\boldsymbol{B}_{3}^{(l)}$ can be bounded as follows

   $$\left\Vert \boldsymbol{B}_{3}^{(l)}\right\Vert _{\mathrm{F}}\leq\eta\left\Vert \frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{E}\right)\right\Vert \left\Vert \boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\Vert _{\mathrm{F}}\lesssim\eta\sigma\sqrt{\frac{n}{p}}\left\Vert \boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\Vert _{\mathrm{F}},$$

   where the second inequality comes from Lemma 27.

3. For the second term $\boldsymbol{B}_{2}^{(l)}$, we have the following lemma.

   Lemma 17. *Suppose that the sample size obeys $n^{2}p\gg\mu^{2}r^{2}n\log n$. Then with probability exceeding $1-O\left(n^{-10}\right)$, the matrix $\boldsymbol{B}_{2}^{(l)}$ as defined in (91) satisfies*

   $$\left\Vert \boldsymbol{B}_{2}^{(l)}\right\Vert _{\mathrm{F}}\lesssim\eta\sqrt{\frac{\kappa^{2}\mu^{2}r^{2}\log n}{np}}\left\Vert \boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}-\boldsymbol{X}^{\natural}\right\Vert _{2,\infty}\sigma_{\max}.\tag{93}$$

4. Regarding the first term $\boldsymbol{B}_{1}^{(l)}$, apply the fundamental theorem of calculus [?, Chapter XIII, Theorem 4.2] to get

   $$\mathrm{vec}\big(\boldsymbol{B}_{1}^{(l)}\big)=\left(\boldsymbol{I}_{nr}-\eta\int_{0}^{1}\nabla^{2}f_{\text{clean}}\left(\boldsymbol{X}(\tau)\right)\mathrm{d}\tau\right)\mathrm{vec}\left(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right),\tag{94}$$

   where we abuse the notation and denote $\boldsymbol{X}(\tau):=\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}+\tau\left(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right)$. Going through the same derivations as in the proof of Lemma 8 (see Appendix 5.2), we get

   $$\left\Vert \boldsymbol{B}_{1}^{(l)}\right\Vert _{\mathrm{F}}\leq\left(1-\frac{\sigma_{\min}}{4}\eta\right)\left\Vert \boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\Vert _{\mathrm{F}}\tag{95}$$

   with the proviso that $0<\eta\leq(2\sigma_{\min})/(25\sigma_{\max}^{2})$.

Applying the triangle inequality to (91) and invoking the preceding four bounds, we arrive at

$$
\begin{aligned}
&\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1}-\boldsymbol{X}^{t+1,(l)}\boldsymbol{R}^{t+1,(l)}\right\|_{\mathrm{F}} \\
&\quad\le\left(1-\frac{\sigma_{\min}}{4}\eta\right)\left\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\|_{\mathrm{F}}+\widetilde{C}\eta\sqrt{\frac{\kappa^2\mu^2r^2\log n}{np}}\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}-\boldsymbol{X}^{\natural}\right\|_{2,\infty}\sigma_{\max} \\
&\qquad+\widetilde{C}\eta\sigma\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\|_{\mathrm{F}}+\widetilde{C}\eta\sigma\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} \\
&\quad=\left(1-\frac{\sigma_{\min}}{4}\eta+\widetilde{C}\eta\sigma\sqrt{\frac{n}{p}}\right)\left\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\|_{\mathrm{F}}+\widetilde{C}\eta\sqrt{\frac{\kappa^2\mu^2r^2\log n}{np}}\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}-\boldsymbol{X}^{\natural}\right\|_{2,\infty}\sigma_{\max} \\
&\qquad+\widetilde{C}\eta\sigma\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} \\
&\quad\le\left(1-\frac{2\sigma_{\min}}{9}\eta\right)\left\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\|_{\mathrm{F}}+\widetilde{C}\eta\sqrt{\frac{\kappa^2\mu^2r^2\log n}{np}}\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}-\boldsymbol{X}^{\natural}\right\|_{2,\infty}\sigma_{\max} \\
&\qquad+\widetilde{C}\eta\sigma\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}
\end{aligned}
$$

for some absolute constant $\widetilde{C}>0$. Here the last inequality holds as long as $\sigma\sqrt{n/p}\ll\sigma_{\min}$, which is satisfied under our noise condition (24). This taken collectively with the hypotheses (33d) and (36c) leads to

$$
\begin{aligned}
&\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1}-\boldsymbol{X}^{t+1,(l)}\boldsymbol{R}^{t+1,(l)}\right\|_{\mathrm{F}} \\
&\quad\le\left(1-\frac{2\sigma_{\min}}{9}\eta\right)\left(C_3\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}+C_7\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\right) \\
&\qquad+\widetilde{C}\eta\sqrt{\frac{\kappa^2\mu^2r^2\log n}{np}}\left[\left(C_3+C_5\right)\rho^t\mu r\sqrt{\frac{\log n}{np}}+\left(C_8+C_7\right)\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\right]\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\sigma_{\max} \\
&\qquad+\widetilde{C}\eta\sigma\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} \\
&\quad\le\left(1-\frac{\sigma_{\min}}{5}\eta\right)C_3\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}+C_7\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}
\end{aligned}
$$

as long as $C_7>0$ is sufficiently large, where we have used the sample complexity assumption $n^2p\gg\kappa^4\mu^2r^2n\log n$ and the step size $0<\eta\le1/(2\sigma_{\max})\le1/(2\sigma_{\min})$. This finishes the proof.

5.5.1 Proof of Lemma 16

By the unitary invariance of the Frobenius norm, one has

$$
\left\|\boldsymbol{B}_4^{(l)}\right\|_{\mathrm{F}}=\frac{\eta}{p}\left\|\mathcal{P}_{\Omega_l}\left(\boldsymbol{E}\right)\boldsymbol{X}^{t,(l)}\right\|_{\mathrm{F}},
$$

where all nonzero entries of the matrix $\mathcal{P}_{\Omega_l}\left(\boldsymbol{E}\right)$ reside in the lth row/column. Decouple the effects of the lth row and the lth column of $\mathcal{P}_{\Omega_l}\left(\boldsymbol{E}\right)$ to reach

$$
\frac{p}{\eta}\left\|\boldsymbol{B}_4^{(l)}\right\|_{\mathrm{F}}\le\left\|\sum_{j=1}^{n}\underbrace{\delta_{l,j}E_{l,j}\boldsymbol{X}_{j,\cdot}^{t,(l)}}_{:=\boldsymbol{u}_j}\right\|_2+\underbrace{\left\|\sum_{j:j\neq l}\delta_{l,j}E_{l,j}\boldsymbol{X}_{l,\cdot}^{t,(l)}\right\|_2}_{:=\alpha},\tag{96}
$$

where $\delta_{l,j} := \mathbb{1}_{\{(l,j)\in\Omega\}}$ indicates whether the (l,j)-th entry is observed. Since $\boldsymbol{X}^{t,(l)}$ is independent of $\{\delta_{l,j}\}_{1\leq j\leq n}$ and $\{E_{l,j}\}_{1\leq j\leq n}$, we can treat the first term as a sum of independent vectors $\{\boldsymbol{u}_j\}$. It is easy to verify that

$$\left\|\|\boldsymbol{u}_j\|_2\right\|_{\psi_1} \leq \left\|\boldsymbol{X}^{t,(l)}\right\|_{2,\infty}\|\delta_{l,j}E_{l,j}\|_{\psi_1} \lesssim \sigma\left\|\boldsymbol{X}^{t,(l)}\right\|_{2,\infty},$$

where $\|\cdot\|_{\psi_1}$ denotes the sub-exponential norm [?, Section 6]. Further, one can calculate

$$V := \left\|\mathbb{E}\left[\sum_{j=1}^{n}(\delta_{l,j}E_{l,j})^2\boldsymbol{X}_{j,\cdot}^{t,(l)}\boldsymbol{X}_{j,\cdot}^{t,(l)\top}\right]\right\| \lesssim p\sigma^2\left\|\mathbb{E}\left[\sum_{j=1}^{n}\boldsymbol{X}_{j,\cdot}^{t,(l)}\boldsymbol{X}_{j,\cdot}^{t,(l)\top}\right]\right\| = p\sigma^2\left\|\boldsymbol{X}^{t,(l)}\right\|_{\mathrm{F}}^2.$$

Invoke the matrix Bernstein inequality [?, Proposition 2] to discover that with probability at least $1 - O\left(n^{-10}\right)$,

$$\begin{aligned}\left\|\sum_{j=1}^{n}\boldsymbol{u}_j\right\|_2 &\lesssim \sqrt{V\log n} + \left\|\|\boldsymbol{u}_j\|\right\|_{\psi_1}\log^2 n \\ &\lesssim \sqrt{p\sigma^2\left\|\boldsymbol{X}^{t,(l)}\right\|_{\mathrm{F}}^2\log n} + \sigma\left\|\boldsymbol{X}^{t,(l)}\right\|_{2,\infty}\log^2 n \\ &\lesssim \sigma\sqrt{np\log n}\left\|\boldsymbol{X}^{t,(l)}\right\|_{2,\infty} + \sigma\left\|\boldsymbol{X}^{t,(l)}\right\|_{2,\infty}\log^2 n \\ &\lesssim \sigma\sqrt{np\log n}\left\|\boldsymbol{X}^{t,(l)}\right\|_{2,\infty},\end{aligned}$$

where the third inequality follows from $\left\|\boldsymbol{X}^{t,(l)}\right\|_{\mathrm{F}}^2 \leq n\left\|\boldsymbol{X}^{t,(l)}\right\|_{2,\infty}^2$, and the last inequality holds as long as $np \gg \log^2 n$.

Additionally, the remaining term α in (96) can be controlled using the same argument, giving rise to

$$\alpha \lesssim \sigma\sqrt{np\log n}\left\|\boldsymbol{X}^{t,(l)}\right\|_{2,\infty}.$$

We then complete the proof by observing that

$$\left\|\boldsymbol{X}^{t,(l)}\right\|_{2,\infty} = \left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\|_{2,\infty} \leq \left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural}\right\|_{2,\infty} + \left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} \leq 2\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}, \tag{97}$$

where the last inequality follows by combining (36c), the sample complexity condition $n^2p \gg \mu^2r^2n\log n$, and the noise condition (24).

5.5.2 Proof of Lemma 17

For notational simplicity, we denote

$$\boldsymbol{C} := \boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{M}^{\natural} = \boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}. \tag{98}$$

Since the Frobenius norm is unitarily invariant, we have

$$\left\|\boldsymbol{B}_2^{(l)}\right\|_{\mathrm{F}} = \eta\left\|\underbrace{\left[\frac{1}{p}\mathcal{P}_{\Omega_l}(\boldsymbol{C}) - \mathcal{P}_l(\boldsymbol{C})\right]}_{:=\boldsymbol{W}}\boldsymbol{X}^{t,(l)}\right\|_{\mathrm{F}}.$$

Again, all nonzero entries of the matrix $\boldsymbol{W}$ reside in its lth row/column. We can deal with the lth row and the lth column of $\boldsymbol{W}$ separately as follows

$$\frac{p}{\eta}\left\|\boldsymbol{B}_2^{(l)}\right\|_{\mathrm{F}} \leq \left\|\sum_{j=1}^{n}(\delta_{l,j}-p)C_{l,j}\boldsymbol{X}_{j,\cdot}^{t,(l)}\right\|_2 + \sqrt{\sum_{j:j\neq l}(\delta_{l,j}-p)^2}\,\|\boldsymbol{C}\|_\infty\left\|\boldsymbol{X}_{l,\cdot}^{t,(l)}\right\|_2$$

$$\lesssim \left\| \sum_{j=1}^{n} \left(\delta_{l,j} - p\right) C_{l,j} \boldsymbol{X}_{j,\cdot}^{t,(l)} \right\|_{2} + \sqrt{np} \left\|\boldsymbol{C}\right\|_{\infty} \left\| \boldsymbol{X}_{l,\cdot}^{t,(l)} \right\|_{2},$$

where $\delta_{l,j} := \mathbb{1}_{\{(l,j)\in\Omega\}}$ and the second line relies on the fact that $\sum_{j:j\neq l}\left(\delta_{l,j}-p\right)^2 \asymp np$. It follows that

$$L := \max_{1\leq j\leq n} \left\| \left(\delta_{l,j} - p\right) C_{l,j} \boldsymbol{X}_{j,\cdot}^{t,(l)} \right\|_{2} \leq \left\|\boldsymbol{C}\right\|_{\infty} \left\| \boldsymbol{X}^{t,(l)} \right\|_{2,\infty} \overset{\text{(i)}}{\leq} 2 \left\|\boldsymbol{C}\right\|_{\infty} \left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty},$$

$$V := \left\| \sum_{j=1}^{n} \mathbb{E}\left[\left(\delta_{l,j} - p\right)^2\right] C_{l,j}^2 \boldsymbol{X}_{j,\cdot}^{t,(l)} \boldsymbol{X}_{j,\cdot}^{t,(l)\top} \right\| \leq p \|\boldsymbol{C}\|_{\infty}^2 \left\| \sum_{j=1}^{n} \boldsymbol{X}_{j,\cdot}^{t,(l)} \boldsymbol{X}_{j,\cdot}^{t,(l)\top} \right\|$$

$$= p \left\|\boldsymbol{C}\right\|_{\infty}^2 \left\| \boldsymbol{X}^{t,(l)} \right\|_{\mathrm{F}}^2 \overset{\text{(ii)}}{\leq} 4p \left\|\boldsymbol{C}\right\|_{\infty}^2 \left\|\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}}^2.$$

Here, (i) is a consequence of (97). In addition, (ii) follows from

$$\left\| \boldsymbol{X}^{t,(l)} \right\|_{\mathrm{F}} = \left\| \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} \right\|_{\mathrm{F}} \leq \left\| \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural} \right\|_{\mathrm{F}} + \left\|\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}} \leq 2 \left\|\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}},$$

where the last inequality comes from (36b), the sample complexity condition $n^2 p \gg \mu^2 r^2 n \log n$, and the noise condition (24). The matrix Bernstein inequality [**?**, Theorem 6.1.1] reveals that

$$\left\| \sum_{j=1}^{n} \left(\delta_{l,j} - p\right) C_{l,j} \boldsymbol{X}_{j,\cdot}^{t,(l)} \right\|_{2} \lesssim \sqrt{V \log n} + L \log n \lesssim \sqrt{p \left\|\boldsymbol{C}\right\|_{\infty}^2 \left\|\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}}^2 \log n} + \left\|\boldsymbol{C}\right\|_{\infty} \left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} \log n$$

with probability exceeding $1 - O\left(n^{-10}\right)$, and as a result,

$$\frac{p}{\eta} \left\| \boldsymbol{B}_2^{(l)} \right\|_{\mathrm{F}} \lesssim \sqrt{p \log n} \left\|\boldsymbol{C}\right\|_{\infty} \left\|\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}} + \sqrt{np} \left\|\boldsymbol{C}\right\|_{\infty} \left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} \tag{99}$$

as soon as $np \gg \log n$.

To finish up, we make the observation that

$$\begin{aligned}
\left\|\boldsymbol{C}\right\|_{\infty} &= \left\| \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} \left(\boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} \right)^{\top} - \boldsymbol{X}^{\natural} \boldsymbol{X}^{\natural\top} \right\|_{\infty} \\
&\leq \left\| \left(\boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural} \right) \left(\boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} \right)^{\top} \right\|_{\infty} + \left\| \boldsymbol{X}^{\natural} \left(\boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural} \right)^{\top} - \boldsymbol{X}^{\natural} \boldsymbol{X}^{\natural\top} \right\|_{\infty} \\
&\leq \left\| \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural} \right\|_{2,\infty} \left\| \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} \right\|_{2,\infty} + \left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} \left\| \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural} \right\|_{2,\infty} \\
&\leq 3 \left\| \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural} \right\|_{2,\infty} \left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty},
\end{aligned} \tag{100}$$

where the last line arises from (97). This combined with (99) gives

$$\begin{aligned}
\left\| \boldsymbol{B}_2^{(l)} \right\|_{\mathrm{F}} &\lesssim \eta \sqrt{\frac{\log n}{p}} \left\|\boldsymbol{C}\right\|_{\infty} \left\|\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}} + \eta \sqrt{\frac{n}{p}} \left\|\boldsymbol{C}\right\|_{\infty} \left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} \\
&\overset{\text{(i)}}{\lesssim} \eta \sqrt{\frac{\log n}{p}} \left\| \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural} \right\|_{2,\infty} \left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} \left\|\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}} + \eta \sqrt{\frac{n}{p}} \left\| \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural} \right\|_{2,\infty} \left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}^2 \\
&\overset{\text{(ii)}}{\lesssim} \eta \sqrt{\frac{\log n}{p}} \left\| \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural} \right\|_{2,\infty} \sqrt{\frac{\kappa \mu r^2}{n}} \sigma_{\max} + \eta \sqrt{\frac{n}{p}} \left\| \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural} \right\|_{2,\infty} \frac{\kappa \mu r}{n} \sigma_{\max} \\
&\lesssim \eta \sqrt{\frac{\kappa^2 \mu^2 r^2 \log n}{np}} \left\| \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural} \right\|_{2,\infty} \sigma_{\max},
\end{aligned}$$

where (i) comes from (100), and (ii) makes use of the incoherence condition (56).

5.6 Proof of Lemma 12

We first introduce an auxiliary matrix

$$\widetilde{\boldsymbol{X}}^{t+1,(l)} := \boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \eta\left[\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left[\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \left(\boldsymbol{M}^{\natural}+\boldsymbol{E}\right)\right] + \mathcal{P}_{l}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{M}^{\natural}\right)\right]\boldsymbol{X}^{\natural}. \tag{101}$$

With this in place, we can use the triangle inequality to obtain

$$\left\|\left(\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t+1,(l)} - \boldsymbol{X}^{\natural}\right)_{l,\cdot}\right\|_2 \le \underbrace{\left\|\left(\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t+1,(l)} - \widetilde{\boldsymbol{X}}^{t+1,(l)}\right)_{l,\cdot}\right\|_2}_{:=\alpha_1} + \underbrace{\left\|\left(\widetilde{\boldsymbol{X}}^{t+1,(l)} - \boldsymbol{X}^{\natural}\right)_{l,\cdot}\right\|_2}_{:=\alpha_2}. \tag{102}$$

In what follows, we bound the two terms α_1 and α_2 separately.

1. Regarding the second term α_2 of (102), we see from the definition of $\widetilde{\boldsymbol{X}}^{t+1,(l)}$ (see (101)) that

$$\left(\widetilde{\boldsymbol{X}}^{t+1,(l)} - \boldsymbol{X}^{\natural}\right)_{l,\cdot} = \left[\boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \eta\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\boldsymbol{X}^{\natural} - \boldsymbol{X}^{\natural}\right]_{l,\cdot}, \tag{103}$$

where we also utilize the definitions of $\mathcal{P}_{\Omega^{-l}}$ and $\mathcal{P}_l$ in (30). For notational convenience, we denote

$$\boldsymbol{\Delta}^{t,(l)} := \boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \boldsymbol{X}^{\natural}. \tag{104}$$

This allows us to rewrite (103) as

$$\begin{aligned}\left(\widetilde{\boldsymbol{X}}^{t+1,(l)} - \boldsymbol{X}^{\natural}\right)_{l,\cdot} &= \boldsymbol{\Delta}^{t,(l)}_{l,\cdot} - \eta\left[\left(\boldsymbol{\Delta}^{t,(l)}\boldsymbol{X}^{\natural\top} + \boldsymbol{X}^{\natural}\boldsymbol{\Delta}^{t,(l)\top}\right)\boldsymbol{X}^{\natural}\right]_{l,\cdot} - \eta\left[\boldsymbol{\Delta}^{t,(l)}\boldsymbol{\Delta}^{t,(l)\top}\boldsymbol{X}^{\natural}\right]_{l,\cdot}\\ &= \boldsymbol{\Delta}^{t,(l)}_{l,\cdot} - \eta\boldsymbol{\Delta}^{t,(l)}_{l,\cdot}\boldsymbol{\Sigma}^{\natural} - \eta\boldsymbol{X}^{\natural}_{l,\cdot}\boldsymbol{\Delta}^{t,(l)\top}\boldsymbol{X}^{\natural} - \eta\boldsymbol{\Delta}^{t,(l)}_{l,\cdot}\boldsymbol{\Delta}^{t,(l)\top}\boldsymbol{X}^{\natural},\end{aligned}$$

which further implies that

$$\begin{aligned}\alpha_2 &\le \left\|\boldsymbol{\Delta}^{t,(l)}_{l,\cdot} - \eta\boldsymbol{\Delta}^{t,(l)}_{l,\cdot}\boldsymbol{\Sigma}^{\natural}\right\|_2 + \eta\left\|\boldsymbol{X}^{\natural}_{l,\cdot}\boldsymbol{\Delta}^{t,(l)\top}\boldsymbol{X}^{\natural}\right\|_2 + \eta\left\|\boldsymbol{\Delta}^{t,(l)}_{l,\cdot}\boldsymbol{\Delta}^{t,(l)\top}\boldsymbol{X}^{\natural}\right\|_2\\ &\le \left\|\boldsymbol{\Delta}^{t,(l)}_{l,\cdot}\right\|_2\left\|\boldsymbol{I}_r - \eta\boldsymbol{\Sigma}^{\natural}\right\| + \eta\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\left\|\boldsymbol{X}^{\natural}\right\| + \eta\left\|\boldsymbol{\Delta}^{t,(l)}_{l,\cdot}\right\|_2\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\left\|\boldsymbol{X}^{\natural}\right\|\\ &\le \left\|\boldsymbol{\Delta}^{t,(l)}_{l,\cdot}\right\|_2\left\|\boldsymbol{I}_r - \eta\boldsymbol{\Sigma}^{\natural}\right\| + 2\eta\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\left\|\boldsymbol{X}^{\natural}\right\|.\end{aligned}$$

Here, the last line follows from the fact that $\left\|\boldsymbol{\Delta}^{t,(l)}_{l,\cdot}\right\|_2 \le \left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}$. To see this, one can use the induction hypothesis (33e) to get

$$\left\|\boldsymbol{\Delta}^{t,(l)}_{l,\cdot}\right\|_2 \le C_2\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} + C_6\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} \ll \left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} \tag{105}$$

as long as $np \gg \mu^2 r^2$ and $\sigma\sqrt{(n\log n)/p} \ll \sigma_{\min}$. By taking $0 < \eta \le 1/\sigma_{\max}$, we have $\boldsymbol{0} \preceq \boldsymbol{I}_r - \eta\boldsymbol{\Sigma}^{\natural} \preceq (1-\eta\sigma_{\min})\boldsymbol{I}_r$, and hence can obtain

$$\alpha_2 \le (1-\eta\sigma_{\min})\left\|\boldsymbol{\Delta}^{t,(l)}_{l,\cdot}\right\|_2 + 2\eta\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\left\|\boldsymbol{X}^{\natural}\right\|. \tag{106}$$

An immediate consequence of the above two inequalities and (36d) is

$$\alpha_2 \le \|\boldsymbol{X}^{\natural}\|_{2,\infty}. \tag{107}$$

2. The first term α_1 of (102) can be equivalently written as

$$\alpha_1 = \left\|\left(\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t,(l)}\boldsymbol{R}_1 - \widetilde{\boldsymbol{X}}^{t+1,(l)}\right)_{l,\cdot}\right\|_2,$$

where

$$\boldsymbol{R}_1 = \left(\widehat{\boldsymbol{H}}^{t,(l)}\right)^{-1}\widehat{\boldsymbol{H}}^{t+1,(l)} := \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \left\|\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t,(l)}\boldsymbol{R} - \boldsymbol{X}^{\natural}\right\|_{\mathrm{F}},$$

Simple algebra yields

$$\begin{aligned}
\alpha_1 &\le \left\|\left(\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \widetilde{\boldsymbol{X}}^{t+1,(l)}\right)_{l,\cdot}\boldsymbol{R}_1\right\|_2 + \left\|\widetilde{\boldsymbol{X}}^{t+1,(l)}_{l,\cdot}\right\|_2 \|\boldsymbol{R}_1 - \boldsymbol{I}_r\| \\
&\le \underbrace{\left\|\left(\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \widetilde{\boldsymbol{X}}^{t+1,(l)}\right)_{l,\cdot}\right\|_2}_{:=\beta_1} + 2\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\underbrace{\|\boldsymbol{R}_1 - \boldsymbol{I}_r\|}_{:=\beta_2}.
\end{aligned}$$

Here, to bound the the second term we have used

$$\left\|\widetilde{\boldsymbol{X}}^{t+1,(l)}_{l,\cdot}\right\|_2 \le \left\|\widetilde{\boldsymbol{X}}^{t+1,(l)}_{l,\cdot} - \boldsymbol{X}^{\natural}_{l,\cdot}\right\|_2 + \left\|\boldsymbol{X}^{\natural}_{l,\cdot}\right\|_2 = \alpha_2 + \left\|\boldsymbol{X}^{\natural}_{l,\cdot}\right\|_2 \le 2\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty},$$

where the last inequality follows from (107). It remains to upper bound β_1 and β_2. For both β_1 and β_2, a central quantity to control is $\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \widetilde{\boldsymbol{X}}^{t+1,(l)}$. By the definition of $\widetilde{\boldsymbol{X}}^{t+1,(l)}$ in (101) and the gradient update rule for $\boldsymbol{X}^{t+1,(l)}$ (see (32)), one has

$$\begin{aligned}
&\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \widetilde{\boldsymbol{X}}^{t+1,(l)} \\
&= \left\{\boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \eta\left[\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left[\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \left(\boldsymbol{M}^{\natural} + \boldsymbol{E}\right)\right] + \mathcal{P}_l\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{M}^{\natural}\right)\right]\boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)}\right\} \\
&\quad - \left\{\boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \eta\left[\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left[\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \left(\boldsymbol{M}^{\natural} + \boldsymbol{E}\right)\right] + \mathcal{P}_l\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{M}^{\natural}\right)\right]\boldsymbol{X}^{\natural}\right\} \\
&= -\eta\left[\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right) + \mathcal{P}_l\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\right]\boldsymbol{\Delta}^{t,(l)} + \frac{\eta}{p}\mathcal{P}_{\Omega^{-l}}\left(\boldsymbol{E}\right)\boldsymbol{\Delta}^{t,(l)}.
\end{aligned} \tag{108}$$

It is easy to verify that

$$\left\|\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left(\boldsymbol{E}\right)\right\| \overset{\text{(i)}}{\le} \left\|\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{E}\right)\right\| \overset{\text{(ii)}}{\lesssim} \sigma\sqrt{\frac{n}{p}} \overset{\text{(iii)}}{\le} \frac{\delta}{2}\sigma_{\min}$$

for $\delta > 0$ sufficiently small. Here, (i) uses the elementary fact that the spectral norm of a submatrix is no more than that of the matrix itself, (ii) arises from Lemma 27 and (iii) is a consequence of the noise condition (24). Therefore, in order to control (108), we need to upper bound the following quantity

$$\gamma := \left\|\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right) + \mathcal{P}_l\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\right\|. \tag{109}$$

To this end, we make the observation that

$$\begin{aligned}
\gamma &\le \underbrace{\left\|\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\right\|}_{:=\gamma_1} \\
&\quad + \underbrace{\left\|\frac{1}{p}\mathcal{P}_{\Omega_l}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right) - \mathcal{P}_l\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\right\|}_{:=\gamma_2},
\end{aligned} \tag{110}$$

where $\mathcal{P}_{\Omega_l}$ is defined in (29). An application of Lemma 30 reveals that

$$\gamma_1 \le 2n\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural}\right\|_{2,\infty}^2 + 4\sqrt{n}\log n\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural}\right\|_{2,\infty}\left\|\boldsymbol{X}^{\natural}\right\|,$$

where $\boldsymbol{R}^{t,(l)} \in \mathcal{O}^{r\times r}$ is defined in (35). Let $\boldsymbol{C} = \boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}$ as in (98), and one can bound the other term γ_2 by taking advantage of the triangle inequality and the symmetry property:

$$\gamma_2 \leq \frac{2}{p}\sqrt{\sum_{j=1}^{n}\left(\delta_{l,j}-p\right)^2 C_{l,j}^2} \overset{\text{(i)}}{\lesssim} \sqrt{\frac{n}{p}}\left\|\boldsymbol{C}\right\|_{\infty} \overset{\text{(ii)}}{\lesssim} \sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural}\right\|_{2,\infty}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty},$$

where (i) comes from the standard Chernoff bound $\sum_{j=1}^{n}\left(\delta_{l,j}-p\right)^2 \asymp np$, and in (ii) we utilize the bound established in (100). The previous two bounds taken collectively give

$$\begin{aligned}\gamma &\leq 2n\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural}\right\|_{2,\infty}^2 + 4\sqrt{n}\log n\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural}\right\|_{2,\infty}\left\|\boldsymbol{X}^{\natural}\right\| \\ &\quad + \widetilde{C}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural}\right\|_{2,\infty}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} \leq \frac{\delta}{2}\sigma_{\min} \end{aligned} \tag{111}$$

for some constant $\widetilde{C} > 0$ and $\delta > 0$ sufficiently small. The last inequality follows from (36c), the incoherence condition (56) and our sample size condition. In summary, we obtain

$$\left\|\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \widetilde{\boldsymbol{X}}^{t+1,(l)}\right\| \leq \eta\left(\gamma + \left\|\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left(\boldsymbol{E}\right)\right\|\right)\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \leq \eta\delta\sigma_{\min}\left\|\boldsymbol{\Delta}^{t,(l)}\right\|, \tag{112}$$

for $\delta > 0$ sufficiently small. With the estimate (112) in place, we can continue our derivation on β_1 and β_2.

(a) With regard to β_1, in view of (108) we can obtain

$$\begin{aligned}\beta_1 &\overset{\text{(i)}}{=} \eta\left\|\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)_{l,\cdot}\boldsymbol{\Delta}^{t,(l)}\right\|_2 \\ &\leq \eta\left\|\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)_{l,\cdot}\right\|_2\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \\ &\overset{\text{(ii)}}{=} \eta\left\|\left[\boldsymbol{\Delta}^{t,(l)}\left(\boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)}\right)^{\top} + \boldsymbol{X}^{\natural}\boldsymbol{\Delta}^{t,(l)\top}\right]_{l,\cdot}\right\|_2\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \\ &\leq \eta\left(\left\|\boldsymbol{\Delta}_{l,\cdot}^{t,(l)}\right\|_2\left\|\boldsymbol{X}^{t,(l)}\right\| + \left\|\boldsymbol{X}_{l,\cdot}^{\natural}\right\|_2\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\right)\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \\ &\leq \eta\left\|\boldsymbol{\Delta}_{l,\cdot}^{t,(l)}\right\|_2\left\|\boldsymbol{X}^{t,(l)}\right\|\left\|\boldsymbol{\Delta}^{t,(l)}\right\| + \eta\left\|\boldsymbol{X}_{l,\cdot}^{\natural}\right\|_2\left\|\boldsymbol{\Delta}^{t,(l)}\right\|^2, \end{aligned} \tag{113}$$

where (i) follows from the definitions of $\mathcal{P}_{\Omega^{-l}}$ and $\mathcal{P}_l$ (see (30) and note that all entries in the lth row of $\mathcal{P}_{\Omega^{-l}}(\cdot)$ are identically zero), and the identity (ii) is due to the definition of $\boldsymbol{\Delta}^{t,(l)}$ in (104).

(b) For β_2, we first claim that

$$\boldsymbol{I}_r := \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\left\|\widetilde{\boldsymbol{X}}^{t+1,(l)}\boldsymbol{R} - \boldsymbol{X}^{\natural}\right\|_{\mathrm{F}}, \tag{114}$$

whose justification follows similar reasonings as that of (80), and is therefore omitted. In particular, it gives rise to the facts that $\boldsymbol{X}^{\natural\top}\widetilde{\boldsymbol{X}}^{t+1,(l)}$ is symmetric and

$$\left(\widetilde{\boldsymbol{X}}^{t+1,(l)}\right)^{\top}\boldsymbol{X}^{\natural} \succeq \frac{1}{2}\sigma_{\min}\boldsymbol{I}_r. \tag{115}$$

We are now ready to invoke Lemma 23 to bound β_2. We abuse the notation and denote $\boldsymbol{C} := \left(\widetilde{\boldsymbol{X}}^{t+1,(l)}\right)^{\top}\boldsymbol{X}^{\natural}$ and $\boldsymbol{E} := \left(\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \widetilde{\boldsymbol{X}}^{t+1,(l)}\right)^{\top}\boldsymbol{X}^{\natural}$. We have

$$\left\|\boldsymbol{E}\right\| \leq \frac{1}{2}\sigma_{\min} \leq \sigma_r\left(\boldsymbol{C}\right).$$

The first inequality arises from (112), namely,

$$\left\|\boldsymbol{E}\right\| \leq \left\|\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \widetilde{\boldsymbol{X}}^{t+1,(l)}\right\|\left\|\boldsymbol{X}^{\natural}\right\| \leq \eta\delta\sigma_{\min}\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\left\|\boldsymbol{X}^{\natural}\right\|$$

$$\overset{\text{(i)}}{\leq} \eta\delta\sigma_{\min}\left\|\boldsymbol{X}^{\natural}\right\|^{2} \overset{\text{(ii)}}{\leq} \frac{1}{2}\sigma_{\min},$$

where (i) holds since $\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \leq \left\|\boldsymbol{X}^{\natural}\right\|$ and (ii) holds true for δ sufficiently small and $\eta \leq 1/\sigma_{\max}$. Invoke Lemma 23 to obtain

$$\begin{aligned}\beta_{2} = \left\|\boldsymbol{R}_{1}-\boldsymbol{I}_{r}\right\| &\leq \frac{2}{\sigma_{r-1}\left(\boldsymbol{C}\right)+\sigma_{r}\left(\boldsymbol{C}\right)}\left\|\boldsymbol{E}\right\| \\ &\leq \frac{2}{\sigma_{\min}}\left\|\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t,(l)}-\widetilde{\boldsymbol{X}}^{t+1,(l)}\right\|\left\|\boldsymbol{X}^{\natural}\right\| &(116)\\ &\leq 2\delta\eta\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\left\|\boldsymbol{X}^{\natural}\right\|, &(117)\end{aligned}$$

where (116) follows since $\sigma_{r-1}\left(\boldsymbol{C}\right) \geq \sigma_{r}\left(\boldsymbol{C}\right) \geq \sigma_{\min}/2$ from (115), and the last line comes from (112).

(c) Putting the previous bounds (113) and (117) together yields

$$\alpha_{1} \leq \eta\left\|\boldsymbol{\Delta}_{l,\cdot}^{t,(l)}\right\|_{2}\left\|\boldsymbol{X}^{t,(l)}\right\|\left\|\boldsymbol{\Delta}^{t,(l)}\right\|+\eta\left\|\boldsymbol{X}_{l,\cdot}^{\natural}\right\|_{2}\left\|\boldsymbol{\Delta}^{t,(l)}\right\|^{2}+4\delta\eta\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\left\|\boldsymbol{X}^{\natural}\right\|. \qquad (118)$$

3. Combine (102), (106) and (118) to reach

$$\begin{aligned}\left\|\left(\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t+1,(l)}-\boldsymbol{X}^{\natural}\right)_{l,\cdot}\right\|_{2} &\leq \left(1-\eta\sigma_{\min}\right)\left\|\boldsymbol{\Delta}_{l,\cdot}^{t,(l)}\right\|_{2}+2\eta\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\left\|\boldsymbol{X}^{\natural}\right\| \\ &\quad+\eta\left\|\boldsymbol{\Delta}_{l,\cdot}^{t,(l)}\right\|_{2}\left\|\boldsymbol{X}^{t,(l)}\right\|\left\|\boldsymbol{\Delta}^{t,(l)}\right\|+\eta\left\|\boldsymbol{X}_{l,\cdot}^{\natural}\right\|_{2}\left\|\boldsymbol{\Delta}^{t,(l)}\right\|^{2}+4\delta\eta\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\left\|\boldsymbol{X}^{\natural}\right\| \\ &\overset{\text{(i)}}{\leq} \left(1-\eta\sigma_{\min}+\eta\left\|\boldsymbol{X}^{t,(l)}\right\|\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\right)\left\|\boldsymbol{\Delta}_{l,\cdot}^{t,(l)}\right\|_{2}+4\eta\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\left\|\boldsymbol{X}^{\natural}\right\| \\ &\overset{\text{(ii)}}{\leq} \left(1-\frac{\sigma_{\min}}{2}\eta\right)\left(C_{2}\rho^{t}\mu r\frac{1}{\sqrt{np}}+\frac{C_{6}}{\sigma_{\min}}\sigma\sqrt{\frac{n\log n}{p}}\right)\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} \\ &\quad+4\eta\left\|\boldsymbol{X}^{\natural}\right\|\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\left(2C_{9}\rho^{t}\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^{\natural}\right\|+\frac{2C_{10}}{\sigma_{\min}}\sigma\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|\right) \\ &\overset{\text{(iii)}}{\leq} C_{2}\rho^{t+1}\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}+\frac{C_{6}}{\sigma_{\min}}\sigma\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}.\end{aligned}$$

Here, (i) follows since $\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \leq \left\|\boldsymbol{X}^{\natural}\right\|$ and δ is sufficiently small, (ii) invokes the hypotheses (33e) and (36d) and recognizes that

$$\left\|\boldsymbol{X}^{t,(l)}\right\|\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \leq 2\left\|\boldsymbol{X}^{\natural}\right\|\left(2C_{9}\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^{\natural}\right\|+\frac{2C_{10}}{\sigma_{\min}}\sigma\sqrt{\frac{n\log n}{np}}\left\|\boldsymbol{X}^{\natural}\right\|\right) \leq \frac{\sigma_{\min}}{2}$$

holds under the sample size and noise condition, while (iii) is valid as long as $1-(\sigma_{\min}/3)\cdot\eta \leq \rho < 1$, $C_{2} \gg \kappa C_{9}$ and $C_{6} \gg \kappa C_{10}/\sqrt{\log n}$.

5.7 Proof of Lemma 13

For notational convenience, we define the following two orthonormal matrices

$$\boldsymbol{Q} := \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\left\|\boldsymbol{U}^{0}\boldsymbol{R}-\boldsymbol{U}^{\natural}\right\|_{\mathrm{F}} \qquad \text{and} \qquad \boldsymbol{Q}^{(l)} := \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\left\|\boldsymbol{U}^{0,(l)}\boldsymbol{R}-\boldsymbol{U}^{\natural}\right\|_{\mathrm{F}}.$$

The problem of finding $\widehat{\boldsymbol{H}}^{t}$ (see (23)) is called the *orthogonal Procrustes problem* [?]. It is well-known that the minimizer $\widehat{\boldsymbol{H}}^{t}$ always exists and is given by

$$\widehat{\boldsymbol{H}}^{t} = \mathrm{sgn}\left(\boldsymbol{X}^{t\top}\boldsymbol{X}^{\natural}\right).$$

Here, the sign matrix $\mathrm{sgn}(\boldsymbol{B})$ is defined as

$$\mathrm{sgn}(\boldsymbol{B}) := \boldsymbol{U}\boldsymbol{V}^\top \tag{119}$$

for any matrix $\boldsymbol{B}$ with singular value decomposition $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$, where the columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ are left and right singular vectors, respectively.

Before proceeding, we make note of the following perturbation bounds on $\boldsymbol{M}^0$ and $\boldsymbol{M}^{(l)}$ (as defined in Algorithm 2 and Algorithm 2, respectively):

$$\begin{aligned}\left\|\boldsymbol{M}^0 - \boldsymbol{M}^\natural\right\| &\overset{\text{(i)}}{\leq} \left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{M}^\natural\right) - \boldsymbol{M}^\natural\right\| + \left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{E}\right)\right\| \\ &\overset{\text{(ii)}}{\leq} C\sqrt{\frac{n}{p}}\left\|\boldsymbol{M}^\natural\right\|_{2,\infty} + C\sigma\sqrt{\frac{n}{p}} = C\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}^2 + C\frac{\sigma}{\sqrt{\sigma_{\min}}}\sqrt{\frac{n}{p}}\sqrt{\sigma_{\min}} \\ &\overset{\text{(iii)}}{\leq} C\left\{\mu r\sqrt{\frac{1}{np}}\sqrt{\sigma_{\max}} + \frac{\sigma}{\sqrt{\sigma_{\min}}}\sqrt{\frac{n}{p}}\right\}\left\|\boldsymbol{X}^\natural\right\| \overset{\text{(iv)}}{\ll} \sigma_{\min},\end{aligned} \tag{120}$$

for some universal constant $C > 0$. Here, (i) arises from the triangle inequality, (ii) utilizes Lemma 26 and Lemma 27, (iii) follows from the incoherence condition (56) and (iv) holds under our sample complexity assumption that $n^2 p \gg \mu^2 r^2 n$ and the noise condition (24). Similarly, we have

$$\left\|\boldsymbol{M}^{(l)} - \boldsymbol{M}^\natural\right\| \lesssim \left\{\mu r\sqrt{\frac{1}{np}}\sqrt{\sigma_{\max}} + \frac{\sigma}{\sqrt{\sigma_{\min}}}\sqrt{\frac{n}{p}}\right\}\left\|\boldsymbol{X}^\natural\right\| \ll \sigma_{\min}. \tag{121}$$

Combine Weyl's inequality, (120) and (121) to obtain

$$\left\|\boldsymbol{\Sigma}^0 - \boldsymbol{\Sigma}^\natural\right\| \leq \left\|\boldsymbol{M}^0 - \boldsymbol{M}^\natural\right\| \ll \sigma_{\min} \quad \text{and} \quad \left\|\boldsymbol{\Sigma}^{(l)} - \boldsymbol{\Sigma}^\natural\right\| \leq \left\|\boldsymbol{M}^{(l)} - \boldsymbol{M}^\natural\right\| \ll \sigma_{\min}, \tag{122}$$

which further implies

$$\frac{1}{2}\sigma_{\min} \leq \sigma_r\left(\boldsymbol{\Sigma}^0\right) \leq \sigma_1\left(\boldsymbol{\Sigma}^0\right) \leq 2\sigma_{\max} \quad \text{and} \quad \frac{1}{2}\sigma_{\min} \leq \sigma_r\left(\boldsymbol{\Sigma}^{(l)}\right) \leq \sigma_1\left(\boldsymbol{\Sigma}^{(l)}\right) \leq 2\sigma_{\max}. \tag{123}$$

We start by proving (33a), (33b) and (33c). The key decomposition we need is the following

$$\boldsymbol{X}^0\widehat{\boldsymbol{H}}^0 - \boldsymbol{X}^\natural = \boldsymbol{U}^0\left(\boldsymbol{\Sigma}^0\right)^{1/2}\left(\widehat{\boldsymbol{H}}^0 - \boldsymbol{Q}\right) + \boldsymbol{U}^0\left[\left(\boldsymbol{\Sigma}^0\right)^{1/2}\boldsymbol{Q} - \boldsymbol{Q}\left(\boldsymbol{\Sigma}^\natural\right)^{1/2}\right] + \left(\boldsymbol{U}^0\boldsymbol{Q} - \boldsymbol{U}^\natural\right)\left(\boldsymbol{\Sigma}^\natural\right)^{1/2}. \tag{124}$$

1. For the spectral norm error bound in (33c), the triangle inequality together with (124) yields

$$\left\|\boldsymbol{X}^0\widehat{\boldsymbol{H}}^0 - \boldsymbol{X}^\natural\right\| \leq \left\|\left(\boldsymbol{\Sigma}^0\right)^{1/2}\right\|\left\|\widehat{\boldsymbol{H}}^0 - \boldsymbol{Q}\right\| + \left\|\left(\boldsymbol{\Sigma}^0\right)^{1/2}\boldsymbol{Q} - \boldsymbol{Q}\left(\boldsymbol{\Sigma}^\natural\right)^{1/2}\right\| + \sqrt{\sigma_{\max}}\left\|\boldsymbol{U}^0\boldsymbol{Q} - \boldsymbol{U}^\natural\right\|,$$

   where we have also used the fact that $\|\boldsymbol{U}^0\| = 1$. Recognizing that $\left\|\boldsymbol{M}^0 - \boldsymbol{M}^\natural\right\| \ll \sigma_{\min}$ (see (120)) and the assumption $\sigma_{\max}/\sigma_{\min} \lesssim 1$, we can apply Lemma 34, Lemma 33 and Lemma 32 to obtain

$$\left\|\widehat{\boldsymbol{H}}^0 - \boldsymbol{Q}\right\| \lesssim \frac{1}{\sigma_{\min}}\left\|\boldsymbol{M}^0 - \boldsymbol{M}^\natural\right\|, \tag{125a}$$

$$\left\|\left(\boldsymbol{\Sigma}^0\right)^{1/2}\boldsymbol{Q} - \boldsymbol{Q}\left(\boldsymbol{\Sigma}^\natural\right)^{1/2}\right\| \lesssim \frac{1}{\sqrt{\sigma_{\min}}}\left\|\boldsymbol{M}^0 - \boldsymbol{M}^\natural\right\|, \tag{125b}$$

$$\left\|\boldsymbol{U}^0\boldsymbol{Q} - \boldsymbol{U}^\natural\right\| \lesssim \frac{1}{\sigma_{\min}}\left\|\boldsymbol{M}^0 - \boldsymbol{M}^\natural\right\|. \tag{125c}$$

   These taken collectively imply the advertised upper bound

$$\begin{aligned}\left\|\boldsymbol{X}^0\widehat{\boldsymbol{H}}^0 - \boldsymbol{X}^\natural\right\| &\lesssim \sqrt{\sigma_{\max}}\frac{1}{\sigma_{\min}}\left\|\boldsymbol{M}^0 - \boldsymbol{M}^\natural\right\| + \frac{1}{\sqrt{\sigma_{\min}}}\left\|\boldsymbol{M}^0 - \boldsymbol{M}^\natural\right\| \lesssim \frac{1}{\sqrt{\sigma_{\min}}}\left\|\boldsymbol{M}^0 - \boldsymbol{M}^\natural\right\| \\ &\lesssim \left\{\mu r\sqrt{\frac{1}{np}}\sqrt{\frac{\sigma_{\max}}{\sigma_{\min}}} + \frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right\}\left\|\boldsymbol{X}^\natural\right\|,\end{aligned}$$

   where we also utilize the fact that $\left\|\left(\boldsymbol{\Sigma}^0\right)^{1/2}\right\| \leq \sqrt{2\sigma_{\max}}$ (see (123)) and the bounded condition number assumption, i.e. $\sigma_{\max}/\sigma_{\min} \lesssim 1$. This finishes the proof of (33c).

2. With regard to the Frobenius norm bound in (33a), one has

$$\begin{aligned}\left\|\boldsymbol{X}^0\widehat{\boldsymbol{H}}^0-\boldsymbol{X}^\natural\right\|_{\mathrm{F}} &\leq \sqrt{r}\|\boldsymbol{X}^0\widehat{\boldsymbol{H}}^0-\boldsymbol{X}^\natural\| \\ &\overset{\text{(i)}}{\lesssim} \left\{\mu r\sqrt{\frac{1}{np}}+\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right\}\sqrt{r}\left\|\boldsymbol{X}^\natural\right\| = \left\{\mu r\sqrt{\frac{1}{np}}+\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right\}\sqrt{r}\frac{\sqrt{\sigma_{\max}}}{\sqrt{\sigma_{\min}}}\sqrt{\sigma_{\min}} \\ &\overset{\text{(ii)}}{\lesssim} \left\{\mu r\sqrt{\frac{1}{np}}+\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right\}\sqrt{r}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}}.\end{aligned}$$

Here (i) arises from (33c) and (ii) holds true since $\sigma_{\max}/\sigma_{\min}\asymp 1$ and $\sqrt{r}\sqrt{\sigma_{\min}}\leq\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}}$, thus completing the proof of (33a).

3. The proof of (33b) follows from similar arguments as used in proving (33c). Combine (124) and the triangle inequality to reach

$$\begin{aligned}\left\|\boldsymbol{X}^0\widehat{\boldsymbol{H}}^0-\boldsymbol{X}^\natural\right\|_{2,\infty} \leq \|\boldsymbol{U}^0\|_{2,\infty}&\left\{\left\|\left(\boldsymbol{\Sigma}^0\right)^{1/2}\right\|\left\|\widehat{\boldsymbol{H}}^0-\boldsymbol{Q}\right\|+\left\|\left(\boldsymbol{\Sigma}^0\right)^{1/2}\boldsymbol{Q}-\boldsymbol{Q}\left(\boldsymbol{\Sigma}^\natural\right)^{1/2}\right\|\right\} \\ &+\sqrt{\sigma_{\max}}\left\|\boldsymbol{U}^0\boldsymbol{Q}-\boldsymbol{U}^\natural\right\|_{2,\infty}.\end{aligned}$$

Plugging in the estimates (120), (123), (125a) and (125b) results in

$$\left\|\boldsymbol{X}^0\widehat{\boldsymbol{H}}^0-\boldsymbol{X}^\natural\right\|_{2,\infty}\lesssim\left\{\mu r\sqrt{\frac{1}{np}}+\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right\}\left\|\boldsymbol{X}^\natural\right\|\left\|\boldsymbol{U}^0\right\|_{2,\infty}+\sqrt{\sigma_{\max}}\left\|\boldsymbol{U}^0\boldsymbol{Q}-\boldsymbol{U}^\natural\right\|_{2,\infty}.$$

It remains to study the component-wise error of $\boldsymbol{U}^0$. To this end, it has already been shown in [?, Lemma 14] that

$$\left\|\boldsymbol{U}^0\boldsymbol{Q}-\boldsymbol{U}^\natural\right\|_{2,\infty}\lesssim\left(\mu r\sqrt{\frac{1}{np}}+\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right)\left\|\boldsymbol{U}^\natural\right\|_{2,\infty}\quad\text{and}\quad\left\|\boldsymbol{U}^0\right\|_{2,\infty}\lesssim\left\|\boldsymbol{U}^\natural\right\|_{2,\infty}\tag{126}$$

under our assumptions. These combined with the previous inequality give

$$\left\|\boldsymbol{X}^0\widehat{\boldsymbol{H}}^0-\boldsymbol{X}^\natural\right\|_{2,\infty}\lesssim\left\{\mu r\sqrt{\frac{1}{np}}+\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right\}\sqrt{\sigma_{\max}}\left\|\boldsymbol{U}^\natural\right\|_{2,\infty}\lesssim\left\{\mu r\sqrt{\frac{1}{np}}+\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right\}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty},$$

where the last relation is due to the observation that

$$\sqrt{\sigma_{\max}}\left\|\boldsymbol{U}^\natural\right\|_{2,\infty}\lesssim\sqrt{\sigma_{\min}}\left\|\boldsymbol{U}^\natural\right\|_{2,\infty}\leq\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}.$$

4. We now move on to proving (33e). Recall that $\boldsymbol{Q}^{(l)}=\arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\left\|\boldsymbol{U}^{0,(l)}\boldsymbol{R}-\boldsymbol{U}^\natural\right\|_{\mathrm{F}}$. By the triangle inequality,

$$\begin{aligned}\left\|\left(\boldsymbol{X}^{0,(l)}\widehat{\boldsymbol{H}}^{0,(l)}-\boldsymbol{X}^\natural\right)_{l,\cdot}\right\|_2 &\leq\left\|\boldsymbol{X}^{0,(l)}_{l,\cdot}\left(\widehat{\boldsymbol{H}}^{0,(l)}-\boldsymbol{Q}^{(l)}\right)\right\|_2+\left\|\left(\boldsymbol{X}^{0,(l)}\boldsymbol{Q}^{(l)}-\boldsymbol{X}^\natural\right)_{l,\cdot}\right\|_2 \\ &\leq\left\|\boldsymbol{X}^{0,(l)}_{l,\cdot}\right\|_2\left\|\widehat{\boldsymbol{H}}^{0,(l)}-\boldsymbol{Q}^{(l)}\right\|+\left\|\left(\boldsymbol{X}^{0,(l)}\boldsymbol{Q}^{(l)}-\boldsymbol{X}^\natural\right)_{l,\cdot}\right\|_2.\end{aligned}\tag{127}$$

Note that $\boldsymbol{X}^\natural_{l,\cdot}=\boldsymbol{M}^\natural_{l,\cdot}\boldsymbol{U}^\natural\left(\boldsymbol{\Sigma}^\natural\right)^{-1/2}$ and, by construction of $\boldsymbol{M}^{(l)}$,

$$\boldsymbol{X}^{0,(l)}_{l,\cdot}=\boldsymbol{M}^{(l)}_{l,\cdot}\boldsymbol{U}^{0,(l)}\left(\boldsymbol{\Sigma}^{(l)}\right)^{-1/2}=\boldsymbol{M}^\natural_{l,\cdot}\boldsymbol{U}^{0,(l)}\left(\boldsymbol{\Sigma}^{(l)}\right)^{-1/2}.$$

We can thus decompose

$$\left(\boldsymbol{X}^{0,(l)}\boldsymbol{Q}^{(l)}-\boldsymbol{X}^\natural\right)_{l,\cdot}=\boldsymbol{M}^\natural_{l,\cdot}\left\{\boldsymbol{U}^{0,(l)}\left[\left(\boldsymbol{\Sigma}^{(l)}\right)^{-1/2}\boldsymbol{Q}^{(l)}-\boldsymbol{Q}^{(l)}\left(\boldsymbol{\Sigma}^\natural\right)^{-1/2}\right]+\left(\boldsymbol{U}^{0,(l)}\boldsymbol{Q}^{(l)}-\boldsymbol{U}^\natural\right)\left(\boldsymbol{\Sigma}^\natural\right)^{-1/2}\right\},$$

which further implies that

$$\left\|\left(\boldsymbol{X}^{0,(l)}\boldsymbol{Q}^{(l)}-\boldsymbol{X}^{\natural}\right)_{l,\cdot}\right\|_{2}\leq\left\|\boldsymbol{M}^{\natural}\right\|_{2,\infty}\left\{\left\|\left(\boldsymbol{\Sigma}^{(l)}\right)^{-1/2}\boldsymbol{Q}^{(l)}-\boldsymbol{Q}^{(l)}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-1/2}\right\|+\frac{1}{\sqrt{\sigma_{\min}}}\left\|\boldsymbol{U}^{0,(l)}\boldsymbol{Q}^{(l)}-\boldsymbol{U}^{\natural}\right\|\right\}.\tag{128}$$

In order to control this, we first see that

$$\begin{aligned}\left\|\left(\boldsymbol{\Sigma}^{(l)}\right)^{-1/2}\boldsymbol{Q}^{(l)}-\boldsymbol{Q}^{(l)}\left(\boldsymbol{\Sigma}^{\natural}\right)^{-1/2}\right\|&=\left\|\left(\boldsymbol{\Sigma}^{(l)}\right)^{-1/2}\left[\boldsymbol{Q}^{(l)}\left(\boldsymbol{\Sigma}^{\natural}\right)^{1/2}-\left(\boldsymbol{\Sigma}^{(l)}\right)^{1/2}\boldsymbol{Q}^{(l)}\right]\left(\boldsymbol{\Sigma}^{\natural}\right)^{-1/2}\right\|\\&\lesssim\frac{1}{\sigma_{\min}}\left\|\boldsymbol{Q}^{(l)}\left(\boldsymbol{\Sigma}^{\natural}\right)^{1/2}-\left(\boldsymbol{\Sigma}^{(l)}\right)^{-1/2}\boldsymbol{Q}^{(l)}\right\|\\&\lesssim\frac{1}{\sigma_{\min}^{3/2}}\left\|\boldsymbol{M}^{(l)}-\boldsymbol{M}^{\natural}\right\|,\end{aligned}$$

where the penultimate inequality uses (123) and the last inequality arises from Lemma 33. Additionally, Lemma 32 gives

$$\left\|\boldsymbol{U}^{0,(l)}\boldsymbol{Q}^{(l)}-\boldsymbol{U}^{\natural}\right\|\lesssim\frac{1}{\sigma_{\min}}\left\|\boldsymbol{M}^{(l)}-\boldsymbol{M}^{\natural}\right\|.$$

Plugging the previous two bounds into (128), we reach

$$\left\|\left(\boldsymbol{X}^{0,(l)}\boldsymbol{Q}^{(l)}-\boldsymbol{X}^{\natural}\right)_{l,\cdot}\right\|_{2}\lesssim\frac{1}{\sigma_{\min}^{3/2}}\left\|\boldsymbol{M}^{(l)}-\boldsymbol{M}^{\natural}\right\|\left\|\boldsymbol{M}^{\natural}\right\|_{2,\infty}\lesssim\left\{\mu r\sqrt{\frac{1}{np}}+\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right\}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}.$$

where the last relation follows from $\left\|\boldsymbol{M}^{\natural}\right\|_{2,\infty}=\left\|\boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right\|_{2,\infty}\leq\sqrt{\sigma_{\max}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}$ and the estimate (121). Note that this also implies that $\left\|\boldsymbol{X}^{0,(l)}_{l,\cdot}\right\|_{2}\leq 2\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}$. To see this, one has by the unitary invariance of $\left\|(\cdot)_{l,\cdot}\right\|_{2}$,

$$\left\|\boldsymbol{X}^{0,(l)}_{l,\cdot}\right\|_{2}=\left\|\boldsymbol{X}^{0,(l)}_{l,\cdot}\boldsymbol{Q}^{(l)}\right\|_{2}\leq\left\|\left(\boldsymbol{X}^{0,(l)}\boldsymbol{Q}^{(l)}-\boldsymbol{X}^{\natural}\right)_{l,\cdot}\right\|_{2}+\left\|\boldsymbol{X}^{\natural}_{l,\cdot}\right\|_{2}\leq 2\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}.$$

Substituting the above bounds back to (127) yields in

$$\begin{aligned}\left\|\left(\boldsymbol{X}^{0,(l)}\widehat{\boldsymbol{H}}^{0,(l)}-\boldsymbol{X}^{\natural}\right)_{l,\cdot}\right\|_{2}&\lesssim\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\left\|\widehat{\boldsymbol{H}}^{0,(l)}-\boldsymbol{Q}^{(l)}\right\|+\left\{\mu r\sqrt{\frac{1}{np}}+\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right\}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\\&\lesssim\left\{\mu r\sqrt{\frac{1}{np}}+\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right\}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty},\end{aligned}$$

where the second line relies on Lemma 34, the bound (121), and the condition $\sigma_{\max}/\sigma_{\min}\asymp 1$. This establishes (33e).

5. Our final step is to justify (33d). Define $\boldsymbol{B}:=\arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\left\|\boldsymbol{U}^{0,(l)}\boldsymbol{R}-\boldsymbol{U}^{0}\right\|_{\mathrm{F}}$. From the definition of $\boldsymbol{R}^{0,(l)}$ (cf. (35)), one has

$$\left\|\boldsymbol{X}^{0}\widehat{\boldsymbol{H}}^{0}-\boldsymbol{X}^{0,(l)}\boldsymbol{R}^{0,(l)}\right\|_{\mathrm{F}}\leq\left\|\boldsymbol{X}^{0,(l)}\boldsymbol{B}-\boldsymbol{X}^{0}\right\|_{\mathrm{F}}.$$

Recognizing that

$$\boldsymbol{X}^{0,(l)}\boldsymbol{B}-\boldsymbol{X}^{0}=\boldsymbol{U}^{0,(l)}\left[\left(\boldsymbol{\Sigma}^{(l)}\right)^{1/2}\boldsymbol{B}-\boldsymbol{B}\left(\boldsymbol{\Sigma}^{0}\right)^{1/2}\right]+\left(\boldsymbol{U}^{0,(l)}\boldsymbol{B}-\boldsymbol{U}^{0}\right)\left(\boldsymbol{\Sigma}^{0}\right)^{1/2},$$

we can use the triangle inequality to bound

$$\left\|\boldsymbol{X}^{0,(l)}\boldsymbol{B}-\boldsymbol{X}^{0}\right\|_{\mathrm{F}}\leq\left\|\left(\boldsymbol{\Sigma}^{(l)}\right)^{1/2}\boldsymbol{B}-\boldsymbol{B}\left(\boldsymbol{\Sigma}^{0}\right)^{1/2}\right\|_{\mathrm{F}}+\left\|\boldsymbol{U}^{0,(l)}\boldsymbol{B}-\boldsymbol{U}^{0}\right\|_{\mathrm{F}}\left\|\left(\boldsymbol{\Sigma}^{0}\right)^{1/2}\right\|.$$

In view of Lemma 33 and the bounds (120) and (121), one has

$$\left\|\left(\boldsymbol{\Sigma}^{(l)}\right)^{-1/2}\boldsymbol{B}-\boldsymbol{B}\boldsymbol{\Sigma}^{1/2}\right\|_{\mathrm{F}}\lesssim\frac{1}{\sqrt{\sigma_{\min}}}\left\|\left(\boldsymbol{M}^{0}-\boldsymbol{M}^{(l)}\right)\boldsymbol{U}^{0,(l)}\right\|_{\mathrm{F}}.$$

From Davis-Kahan's sinΘ theorem [?] we see that

$$\left\| \boldsymbol{U}^{0,(l)} \boldsymbol{B} - \boldsymbol{U}^{0} \right\|_{\mathrm{F}} \lesssim \frac{1}{\sigma_{\min}} \left\| (\boldsymbol{M}^{0} - \boldsymbol{M}^{(l)}) \boldsymbol{U}^{0,(l)} \right\|_{\mathrm{F}} .$$

These estimates taken together with (123) give

$$\left\| \boldsymbol{X}^{0,(l)} \boldsymbol{B} - \boldsymbol{X}^{0} \right\|_{\mathrm{F}} \lesssim \frac{1}{\sqrt{\sigma_{\min}}} \left\| (\boldsymbol{M}^{0} - \boldsymbol{M}^{(l)}) \boldsymbol{U}^{0,(l)} \right\|_{\mathrm{F}} .$$

It then boils down to controlling $\left\| \left(\boldsymbol{M}^{0} - \boldsymbol{M}^{(l)} \right) \boldsymbol{U}^{0,(l)} \right\|_{\mathrm{F}}$. Quantities of this type have showed up multiple times already, and hence we omit the proof details for conciseness (see Appendix 5.5). With probability at least $1 - O\left(n^{-10}\right)$,

$$\left\| (\boldsymbol{M}^{0} - \boldsymbol{M}^{(l)}) \boldsymbol{U}^{0,(l)} \right\|_{\mathrm{F}} \lesssim \left\{ \mu r \sqrt{\frac{\log n}{np}} \sigma_{\max} + \sigma \sqrt{\frac{n \log n}{p}} \right\} \left\| \boldsymbol{U}^{0,(l)} \right\|_{2,\infty} .$$

If one further has

$$\left\| \boldsymbol{U}^{0,(l)} \right\|_{2,\infty} \lesssim \| \boldsymbol{U}^{\natural} \|_{2,\infty} \lesssim \frac{1}{\sqrt{\sigma_{\min}}} \| \boldsymbol{X}^{\natural} \|_{2,\infty} , \tag{129}$$

then taking the previous bounds collectively establishes the desired bound

$$\left\| \boldsymbol{X}^{0} \widehat{\boldsymbol{H}}^{0} - \boldsymbol{X}^{0,(l)} \boldsymbol{R}^{0,(l)} \right\|_{\mathrm{F}} \lesssim \left\{ \mu r \sqrt{\frac{\log n}{np}} + \frac{\sigma}{\sigma_{\min}} \sqrt{\frac{n \log n}{p}} \right\} \| \boldsymbol{X}^{\natural} \|_{2,\infty} .$$

Proof of Claim (129). Denote by $\boldsymbol{M}^{(l),\mathrm{zero}}$ the matrix derived by zeroing out the lth row/column of $\boldsymbol{M}^{(l)}$, and $\boldsymbol{U}^{(l),\mathrm{zero}} \in \mathbb{R}^{n \times r}$ containing the leading r eigenvectors of $\boldsymbol{M}^{(l),\mathrm{zero}}$. On the one hand, [?, Lemma 4 and Lemma 14] demonstrate that

$$\max_{1 \leq l \leq n} \| \boldsymbol{U}^{(l),\mathrm{zero}} \|_{2,\infty} \lesssim \| \boldsymbol{U}^{\natural} \|_{2,\infty} .$$

On the other hand, by the Davis-Kahan $\sin \Theta$ theorem [?] we obtain

$$\left\| \boldsymbol{U}^{0,(l)} \mathrm{sgn} \left(\boldsymbol{U}^{0,(l)\top} \boldsymbol{U}^{(l),\mathrm{zero}} \right) - \boldsymbol{U}^{(l),\mathrm{zero}} \right\|_{\mathrm{F}} \lesssim \frac{1}{\sigma_{\min}} \left\| \left(\boldsymbol{M}^{(l)} - \boldsymbol{M}^{(l),\mathrm{zero}} \right) \boldsymbol{U}^{(l),\mathrm{zero}} \right\|_{\mathrm{F}} , \tag{130}$$

where $\mathrm{sgn}(\boldsymbol{A})$ denotes the sign matrix of $\boldsymbol{A}$. For any $j \neq l$, one has

$$\left(\boldsymbol{M}^{(l)} - \boldsymbol{M}^{(l),\mathrm{zero}} \right)_{j,\cdot} \boldsymbol{U}^{(l),\mathrm{zero}} = \left(\boldsymbol{M}^{(l)} - \boldsymbol{M}^{(l),\mathrm{zero}} \right)_{j,l} \boldsymbol{U}^{(l),\mathrm{zero}}_{l,\cdot} = \boldsymbol{0}_{1 \times r},$$

since the lth row of $\boldsymbol{U}^{(l),\mathrm{zero}}_{l,\cdot}$ is identically zero by construction. In addition,

$$\left\| \left(\boldsymbol{M}^{(l)} - \boldsymbol{M}^{(l),\mathrm{zero}} \right)_{l,\cdot} \boldsymbol{U}^{(l),\mathrm{zero}} \right\|_{2} = \left\| \boldsymbol{M}^{\natural}_{l,\cdot} \boldsymbol{U}^{(l),\mathrm{zero}} \right\|_{2} \leq \| \boldsymbol{M}^{\natural} \|_{2,\infty} \leq \sigma_{\max} \left\| \boldsymbol{U}^{\natural} \right\|_{2,\infty} .$$

As a consequence, one has

$$\left\| \left(\boldsymbol{M}^{(l)} - \boldsymbol{M}^{(l),\mathrm{zero}} \right) \boldsymbol{U}^{(l),\mathrm{zero}} \right\|_{\mathrm{F}} = \left\| \left(\boldsymbol{M}^{(l)} - \boldsymbol{M}^{(l),\mathrm{zero}} \right)_{l,\cdot} \boldsymbol{U}^{(l),\mathrm{zero}} \right\|_{2} \leq \sigma_{\max} \left\| \boldsymbol{U}^{\natural} \right\|_{2,\infty} ,$$

which combined with (130) and the assumption $\sigma_{\max}/\sigma_{\min} \asymp 1$ yields

$$\left\| \boldsymbol{U}^{0,(l)} \mathrm{sgn} \left(\boldsymbol{U}^{0,(l)\top} \boldsymbol{U}^{(l),\mathrm{zero}} \right) - \boldsymbol{U}^{(l),\mathrm{zero}} \right\|_{\mathrm{F}} \lesssim \left\| \boldsymbol{U}^{\natural} \right\|_{2,\infty}$$

The claim (129) then follows by combining the above estimates:

$$\begin{aligned} \left\| \boldsymbol{U}^{0,(l)} \right\|_{2,\infty} &= \left\| \boldsymbol{U}^{0,(l)} \mathrm{sgn} \left(\boldsymbol{U}^{0,(l)\top} \boldsymbol{U}^{(l),\mathrm{zero}} \right) \right\|_{2,\infty} \\ &\leq \| \boldsymbol{U}^{(l),\mathrm{zero}} \|_{2,\infty} + \left\| \boldsymbol{U}^{0,(l)} \mathrm{sgn} \left(\boldsymbol{U}^{0,(l)\top} \boldsymbol{U}^{(l),\mathrm{zero}} \right) - \boldsymbol{U}^{(l),\mathrm{zero}} \right\|_{\mathrm{F}} \lesssim \| \boldsymbol{U}^{\natural} \|_{2,\infty}, \end{aligned}$$

where we have utilized the unitary invariance of $\|\cdot\|_{2,\infty}$. □

6 Technical lemmas

6.1 Technical lemmas for phase retrieval

6.1.1 Matrix concentration inequalities

Lemma 18. *Suppose that* $\boldsymbol{a}_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$ *for every* $1 \le j \le m$*. Fix any small constant* $\delta > 0$*. With probability at least* $1 - C_2 e^{-c_2 m}$*, one has*

$$\left\| \frac{1}{m} \sum_{j=1}^{m} \boldsymbol{a}_j \boldsymbol{a}_j^\top - \boldsymbol{I}_n \right\| \le \delta,$$

as long as $m \ge c_0 n$ *for some sufficiently large constant* $c_0 > 0$*. Here,* $C_2, c_2 > 0$ *are some universal constants.*

Proof. This is an immediate consequence of [**?**, Corollary 5.35]. □

Lemma 19. *Suppose that* $\boldsymbol{a}_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$*, for every* $1 \le j \le m$*. Fix any small constant* $\delta > 0$*. With probability at least* $1 - O(n^{-10})$*, we have*

$$\left\| \frac{1}{m} \sum_{j=1}^{m} \left(\boldsymbol{a}_j^\top \boldsymbol{x}^\natural \right)^2 \boldsymbol{a}_j \boldsymbol{a}_j^\top - \left(\|\boldsymbol{x}^\natural\|_2^2 \boldsymbol{I}_n + 2\boldsymbol{x}^\natural \boldsymbol{x}^{\natural\top} \right) \right\| \le \delta \|\boldsymbol{x}^\natural\|_2^2,$$

provided that $m \ge c_0 n \log n$ *for some sufficiently large constant* $c_0 > 0$*.*

Proof. This is adapted from [**?**, Lemma 7.4]. □

Lemma 20. *Suppose that* $\boldsymbol{a}_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$*, for every* $1 \le j \le m$*. Fix any small constant* $\delta > 0$ *and any constant* $C > 0$*. Suppose* $m \ge c_0 n$ *for some sufficiently large constant* $c_0 > 0$*. Then with probability at least* $1 - C_2 e^{-c_2 m}$*,*

$$\left\| \frac{1}{m} \sum_{j=1}^{m} \left(\boldsymbol{a}_j^\top \boldsymbol{x} \right)^2 \mathbb{1}_{\{|\boldsymbol{a}_j^\top \boldsymbol{x}| \le C\}} \boldsymbol{a}_j \boldsymbol{a}_j^\top - \left(\beta_1 \boldsymbol{x}\boldsymbol{x}^\top + \beta_2 \|\boldsymbol{x}\|_2^2 \boldsymbol{I}_n \right) \right\| \le \delta \|\boldsymbol{x}\|_2^2, \quad \forall \boldsymbol{x} \in \mathbb{R}^n$$

holds for some absolute constants $c_2, C_2 > 0$*, where*

$$\beta_1 := \mathbb{E}\left[\xi^4 \mathbb{1}_{\{|\xi| \le C\}} \right] - \mathbb{E}\left[\xi^2 \mathbb{1}_{|\xi| \le C} \right] \quad \textit{and} \quad \beta_2 = \mathbb{E}\left[\xi^2 \mathbb{1}_{|\xi| \le C} \right]$$

with ξ *being a standard Gaussian random variable.*

Proof. This is supplied in [**?**, supplementary material]. □

6.1.2 Matrix perturbation bounds

Lemma 21. *Let* $\lambda_1(\boldsymbol{A})$*,* $\boldsymbol{u}$ *be the leading eigenvalue and eigenvector of a symmetric matrix* $\boldsymbol{A}$*, respectively, and* $\lambda_1(\widetilde{\boldsymbol{A}})$*,* $\widetilde{\boldsymbol{u}}$ *be the leading eigenvalue and eigenvector of a symmetric matrix* $\widetilde{\boldsymbol{A}}$*, respectively. Suppose that* $\lambda_1(\boldsymbol{A}), \lambda_1(\widetilde{\boldsymbol{A}}), \|\boldsymbol{A}\|, \|\widetilde{\boldsymbol{A}}\| \in [C_1, C_2]$ *for some* $C_1, C_2 > 0$*. Then,*

$$\left\| \sqrt{\lambda_1(\boldsymbol{A})}\, \boldsymbol{u} - \sqrt{\lambda_1(\widetilde{\boldsymbol{A}})}\, \widetilde{\boldsymbol{u}} \right\|_2 \le \frac{\left\| (\boldsymbol{A} - \widetilde{\boldsymbol{A}}) \boldsymbol{u} \right\|_2}{2\sqrt{C_1}} + \left(\sqrt{C_2} + \frac{C_2}{\sqrt{C_1}} \right) \|\boldsymbol{u} - \widetilde{\boldsymbol{u}}\|_2 .$$

Proof. Observe that

$$\left\| \sqrt{\lambda_1(\boldsymbol{A})}\, \boldsymbol{u} - \sqrt{\lambda_1(\widetilde{\boldsymbol{A}})}\, \widetilde{\boldsymbol{u}} \right\|_2 \le \left\| \sqrt{\lambda_1(\boldsymbol{A})}\, \boldsymbol{u} - \sqrt{\lambda_1(\widetilde{\boldsymbol{A}})}\, \boldsymbol{u} \right\|_2 + \left\| \sqrt{\lambda_1(\widetilde{\boldsymbol{A}})}\, \boldsymbol{u} - \sqrt{\lambda_1(\widetilde{\boldsymbol{A}})}\, \widetilde{\boldsymbol{u}} \right\|_2$$

$$\leq \left|\sqrt{\lambda_1(\boldsymbol{A})} - \sqrt{\lambda_1(\widetilde{\boldsymbol{A}})}\right| + \sqrt{\lambda_1(\widetilde{\boldsymbol{A}})}\,\|\boldsymbol{u} - \widetilde{\boldsymbol{u}}\|_2\,, \tag{131}$$

where the last inequality follows since $\|\boldsymbol{u}\|_2 = 1$. Using the identity $\sqrt{a} - \sqrt{b} = (a-b)/(\sqrt{a} + \sqrt{b})$, we have

$$\left|\sqrt{\lambda_1(\boldsymbol{A})} - \sqrt{\lambda_1(\widetilde{\boldsymbol{A}})}\right| = \frac{\left|\lambda_1(\boldsymbol{A}) - \lambda_1(\widetilde{\boldsymbol{A}})\right|}{\left|\sqrt{\lambda_1(\boldsymbol{A})} + \sqrt{\lambda_1(\widetilde{\boldsymbol{A}})}\right|} \leq \frac{\left|\lambda_1(\boldsymbol{A}) - \lambda_1(\widetilde{\boldsymbol{A}})\right|}{2\sqrt{C_1}},$$

where the last inequality comes from our assumptions on $\lambda_1(\boldsymbol{A})$ and $\lambda_1(\widetilde{\boldsymbol{A}})$. This combined with (131) yields

$$\left\|\sqrt{\lambda_1(\boldsymbol{A})}\,\boldsymbol{u} - \sqrt{\lambda_1(\widetilde{\boldsymbol{A}})}\,\widetilde{\boldsymbol{u}}\right\|_2 \leq \frac{\left|\lambda_1(\boldsymbol{A}) - \lambda_1(\widetilde{\boldsymbol{A}})\right|}{2\sqrt{C_1}} + \sqrt{C_2}\,\|\boldsymbol{u} - \widetilde{\boldsymbol{u}}\|_2\,. \tag{132}$$

To control $\left|\lambda_1(\boldsymbol{A}) - \lambda_1(\widetilde{\boldsymbol{A}})\right|$, use the relationship between the eigenvalue and the eigenvector to obtain

$$\begin{aligned}
\left|\lambda_1(\boldsymbol{A}) - \lambda_1(\widetilde{\boldsymbol{A}})\right| &= \left|\boldsymbol{u}^\top \boldsymbol{A}\boldsymbol{u} - \widetilde{\boldsymbol{u}}^\top \widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{u}}\right| \\
&\leq \left|\boldsymbol{u}^\top (\boldsymbol{A} - \widetilde{\boldsymbol{A}})\boldsymbol{u}\right| + \left|\boldsymbol{u}^\top \widetilde{\boldsymbol{A}}\boldsymbol{u} - \widetilde{\boldsymbol{u}}^\top \widetilde{\boldsymbol{A}}\boldsymbol{u}\right| + \left|\widetilde{\boldsymbol{u}}^\top \widetilde{\boldsymbol{A}}\boldsymbol{u} - \widetilde{\boldsymbol{u}}^\top \widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{u}}\right| \\
&\leq \left\|(\boldsymbol{A} - \widetilde{\boldsymbol{A}})\boldsymbol{u}\right\|_2 + 2\,\|\boldsymbol{u} - \widetilde{\boldsymbol{u}}\|_2\,\|\widetilde{\boldsymbol{A}}\|,
\end{aligned}$$

which together with (132) gives

$$\begin{aligned}
\left\|\sqrt{\lambda_1(\boldsymbol{A})}\,\boldsymbol{u} - \sqrt{\lambda_1(\widetilde{\boldsymbol{A}})}\,\widetilde{\boldsymbol{u}}\right\|_2 &\leq \frac{\left\|(\boldsymbol{A} - \widetilde{\boldsymbol{A}})\boldsymbol{u}\right\|_2 + 2\,\|\boldsymbol{u} - \widetilde{\boldsymbol{u}}\|_2\,\|\widetilde{\boldsymbol{A}}\|}{2\sqrt{C_1}} + \sqrt{C_2}\,\|\boldsymbol{u} - \widetilde{\boldsymbol{u}}\|_2 \\
&\leq \frac{\left\|(\boldsymbol{A} - \widetilde{\boldsymbol{A}})\boldsymbol{u}\right\|_2}{2\sqrt{C_1}} + \left(\frac{C_2}{\sqrt{C_1}} + \sqrt{C_2}\right)\|\boldsymbol{u} - \widetilde{\boldsymbol{u}}\|_2
\end{aligned}$$

as claimed. □

6.2 Technical lemmas for matrix completion

6.2.1 Orthogonal Procrustes problem

The orthogonal Procrustes problem is a matrix approximation problem which seeks an orthogonal matrix $\boldsymbol{R}$ to best "align" two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. Specifically, for $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times r}$, define $\widehat{\boldsymbol{R}}$ to be the minimizer of

$$\text{minimize}_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \quad \|\boldsymbol{A}\boldsymbol{R} - \boldsymbol{B}\|_{\mathrm{F}}\,. \tag{133}$$

The first lemma is concerned with the characterization of the minimizer $\widehat{\boldsymbol{R}}$ of (133).

Lemma 22. *For $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times r}$, $\widehat{\boldsymbol{R}}$ is the minimizer of (133) if and only if $\widehat{\boldsymbol{R}}^\top \boldsymbol{A}^\top \boldsymbol{B}$ is symmetric and positive semidefinite.*

Proof. This is an immediate consequence of [**?**, Theorem 2]. □

Let $\boldsymbol{A}^\top \boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ be the singular value decomposition of $\boldsymbol{A}^\top \boldsymbol{B} \in \mathbb{R}^{r\times r}$. It is easy to check that $\widehat{\boldsymbol{R}} := \boldsymbol{U}\boldsymbol{V}^\top$ satisfies the conditions that $\widehat{\boldsymbol{R}}^\top \boldsymbol{A}^\top \boldsymbol{B}$ is both symmetric and positive semidefinite. In view of Lemma 22, $\widehat{\boldsymbol{R}} = \boldsymbol{U}\boldsymbol{V}^\top$ is the minimizer of (133). In the special case when $\boldsymbol{C} := \boldsymbol{A}^\top \boldsymbol{B}$ is invertible, $\widehat{\boldsymbol{R}}$ enjoys the following equivalent form:

$$\widehat{\boldsymbol{R}} = \widehat{\boldsymbol{H}}(\boldsymbol{C}) := \boldsymbol{C}\left(\boldsymbol{C}^\top \boldsymbol{C}\right)^{-1/2}, \tag{134}$$

where $\widehat{\boldsymbol{H}}(\cdot)$ is an $\mathbb{R}^{r\times r}$-valued function on $\mathbb{R}^{r\times r}$. This motivates us to look at the perturbation bounds for the matrix-valued function $\widehat{\boldsymbol{H}}(\cdot)$, which is formulated in the following lemma.

Lemma 23. *Let $\boldsymbol{C} \in \mathbb{R}^{r\times r}$ be a nonsingular matrix. Then for any matrix $\boldsymbol{E} \in \mathbb{R}^{r\times r}$ with $\|\boldsymbol{E}\| \le \sigma_{\min}(\boldsymbol{C})$ and any unitarily invariant norm $\|\!|\cdot|\!\|$, one has*

$$\left\|\!\left|\widehat{\boldsymbol{H}}(\boldsymbol{C}+\boldsymbol{E}) - \widehat{\boldsymbol{H}}(\boldsymbol{C})\right|\!\right\| \le \frac{2}{\sigma_{r-1}(\boldsymbol{C}) + \sigma_r(\boldsymbol{C})} \|\!|\boldsymbol{E}|\!\|,$$

where $\widehat{\boldsymbol{H}}(\cdot)$ is defined above.

Proof. This is an immediate consequence of [**?**, Theorem 2.3]. □

With Lemma 23 in place, we are ready to present the following bounds on two matrices after "aligning" them with $\boldsymbol{X}^\natural$.

Lemma 24. *Instate the notation in Section 3.2. Suppose $\boldsymbol{X}_1, \boldsymbol{X}_2 \in \mathbb{R}^{n\times r}$ are two matrices such that*

$$\left\|\boldsymbol{X}_1 - \boldsymbol{X}^\natural\right\| \left\|\boldsymbol{X}^\natural\right\| \le \sigma_{\min}/2, \tag{135a}$$

$$\left\|\boldsymbol{X}_1 - \boldsymbol{X}_2\right\| \left\|\boldsymbol{X}^\natural\right\| \le \sigma_{\min}/4. \tag{135b}$$

Denote

$$\boldsymbol{R}_1 := \underset{\boldsymbol{R}\in\mathcal{O}^{r\times r}}{\operatorname{argmin}} \left\|\boldsymbol{X}_1\boldsymbol{R} - \boldsymbol{X}^\natural\right\|_{\mathrm{F}} \qquad \textit{and} \qquad \boldsymbol{R}_2 := \underset{\boldsymbol{R}\in\mathcal{O}^{r\times r}}{\operatorname{argmin}} \left\|\boldsymbol{X}_2\boldsymbol{R} - \boldsymbol{X}^\natural\right\|_{\mathrm{F}}.$$

Then the following two inequalities hold true:

$$\left\|\boldsymbol{X}_1\boldsymbol{R}_1 - \boldsymbol{X}_2\boldsymbol{R}_2\right\| \le 5\kappa \left\|\boldsymbol{X}_1 - \boldsymbol{X}_2\right\| \qquad \textit{and} \qquad \left\|\boldsymbol{X}_1\boldsymbol{R}_1 - \boldsymbol{X}_2\boldsymbol{R}_2\right\|_{\mathrm{F}} \le 5\kappa \left\|\boldsymbol{X}_1 - \boldsymbol{X}_2\right\|_{\mathrm{F}}.$$

Proof. Before proving the claims, we first gather some immediate consequences of the assumptions (135). Denote $\boldsymbol{C} = \boldsymbol{X}_1^\top \boldsymbol{X}^\natural$ and $\boldsymbol{E} = (\boldsymbol{X}_2 - \boldsymbol{X}_1)^\top \boldsymbol{X}^\natural$. It is easily seen that $\boldsymbol{C}$ is invertible since

$$\left\|\boldsymbol{C} - \boldsymbol{X}^{\natural\top}\boldsymbol{X}^\natural\right\| \le \left\|\boldsymbol{X}_1 - \boldsymbol{X}^\natural\right\| \left\|\boldsymbol{X}^\natural\right\| \overset{\text{(i)}}{\le} \sigma_{\min}/2 \qquad \overset{\text{(ii)}}{\Longrightarrow} \qquad \sigma_r(\boldsymbol{C}) \ge \sigma_{\min}/2, \tag{136}$$

where (i) follows from the assumption (135a) and (ii) is a direct application of Weyl's inequality. In addition, $\boldsymbol{C} + \boldsymbol{E} = \boldsymbol{X}_2^\top \boldsymbol{X}^\natural$ is also invertible since

$$\|\boldsymbol{E}\| \le \left\|\boldsymbol{X}_1 - \boldsymbol{X}_2\right\| \left\|\boldsymbol{X}^\natural\right\| \overset{\text{(i)}}{\le} \sigma_{\min}/4 \overset{\text{(ii)}}{<} \sigma_r(\boldsymbol{C}),$$

where (i) arises from the assumption (135b) and (ii) holds because of (136). When both $\boldsymbol{C}$ and $\boldsymbol{C}+\boldsymbol{E}$ are invertible, the orthonormal matrices $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ admit closed-form expressions as follows

$$\boldsymbol{R}_1 = \boldsymbol{C}\left(\boldsymbol{C}^\top\boldsymbol{C}\right)^{-1/2} \qquad \text{and} \qquad \boldsymbol{R}_2 = (\boldsymbol{C}+\boldsymbol{E})\left[(\boldsymbol{C}+\boldsymbol{E})^\top(\boldsymbol{C}+\boldsymbol{E})\right]^{-1/2}.$$

Moreover, we have the following bound on $\|\boldsymbol{X}_1\|$:

$$\|\boldsymbol{X}_1\| \overset{\text{(i)}}{\le} \left\|\boldsymbol{X}_1 - \boldsymbol{X}^\natural\right\| + \left\|\boldsymbol{X}^\natural\right\| \overset{\text{(ii)}}{\le} \frac{\sigma_{\min}}{2\left\|\boldsymbol{X}^\natural\right\|} + \left\|\boldsymbol{X}^\natural\right\| \le \frac{\sigma_{\max}}{2\left\|\boldsymbol{X}^\natural\right\|} + \left\|\boldsymbol{X}^\natural\right\| \overset{\text{(iii)}}{\le} 2\left\|\boldsymbol{X}^\natural\right\|, \tag{137}$$

where (i) is the triangle inequality, (ii) uses the assumption (135a) and (iii) arises from the fact that $\left\|\boldsymbol{X}^\natural\right\| = \sqrt{\sigma_{\max}}$.

With these in place, we turn to establishing the claimed bounds. We will focus on the upper bound on $\left\|\boldsymbol{X}_1\boldsymbol{R}_1 - \boldsymbol{X}_2\boldsymbol{R}_2\right\|_{\mathrm{F}}$, as the bound on $\left\|\boldsymbol{X}_1\boldsymbol{R}_1 - \boldsymbol{X}_2\boldsymbol{R}_2\right\|$ can be easily obtained using the same argument. Simple algebra reveals that

$$\begin{aligned}
\left\|\boldsymbol{X}_1\boldsymbol{R}_1 - \boldsymbol{X}_2\boldsymbol{R}_2\right\|_{\mathrm{F}} &= \left\|(\boldsymbol{X}_1 - \boldsymbol{X}_2)\boldsymbol{R}_2 + \boldsymbol{X}_1(\boldsymbol{R}_1 - \boldsymbol{R}_2)\right\|_{\mathrm{F}} \\
&\le \left\|\boldsymbol{X}_1 - \boldsymbol{X}_2\right\|_{\mathrm{F}} + \left\|\boldsymbol{X}_1\right\| \left\|\boldsymbol{R}_1 - \boldsymbol{R}_2\right\|_{\mathrm{F}} \\
&\le \left\|\boldsymbol{X}_1 - \boldsymbol{X}_2\right\|_{\mathrm{F}} + 2\left\|\boldsymbol{X}^\natural\right\| \left\|\boldsymbol{R}_1 - \boldsymbol{R}_2\right\|_{\mathrm{F}},
\end{aligned} \tag{138}$$

where the first inequality uses the fact that $\|\boldsymbol{R}_2\| = 1$ and the last inequality comes from (137). An application of Lemma 23 leads us to conclude that

$$\begin{aligned}
\|\boldsymbol{R}_1 - \boldsymbol{R}_2\|_{\mathrm{F}} &\leq \frac{2}{\sigma_r(\boldsymbol{C}) + \sigma_{r-1}(\boldsymbol{C})} \|\boldsymbol{E}\|_{\mathrm{F}} \\
&\leq \frac{2}{\sigma_{\min}} \left\| (\boldsymbol{X}_2 - \boldsymbol{X}_1)^\top \boldsymbol{X}^\natural \right\|_{\mathrm{F}} \qquad (139) \\
&\leq \frac{2}{\sigma_{\min}} \|\boldsymbol{X}_2 - \boldsymbol{X}_1\|_{\mathrm{F}} \left\|\boldsymbol{X}^\natural\right\|, \qquad (140)
\end{aligned}$$

where (139) utilizes (136). Combine (138) and (140) to reach

$$\begin{aligned}
\|\boldsymbol{X}_1\boldsymbol{R}_1 - \boldsymbol{X}_2\boldsymbol{R}_2\|_{\mathrm{F}} &\leq \|\boldsymbol{X}_1 - \boldsymbol{X}_2\|_{\mathrm{F}} + \frac{4}{\sigma_{\min}} \|\boldsymbol{X}_2 - \boldsymbol{X}_1\|_{\mathrm{F}} \left\|\boldsymbol{X}^\natural\right\|^2 \\
&\leq (1 + 4\kappa) \|\boldsymbol{X}_1 - \boldsymbol{X}_2\|_{\mathrm{F}},
\end{aligned}$$

which finishes the proof by noting that $\kappa \geq 1$. □

6.2.2 Matrix concentration inequalities

This section collects various measure concentration results regarding the Bernoulli random variables $\{\delta_{j,k}\}_{1\leq j,k\leq n}$, which is ubiquitous in the analysis for matrix completion.

Lemma 25. *Fix any small constant $\delta > 0$, and suppose that $m \gg \delta^{-2}\mu nr \log n$. Then with probability exceeding $1 - O\left(n^{-10}\right)$, one has*

$$(1-\delta)\|\boldsymbol{B}\|_{\mathrm{F}} \leq \frac{1}{\sqrt{p}}\|\mathcal{P}_\Omega(\boldsymbol{B})\|_{\mathrm{F}} \leq (1+\delta)\|\boldsymbol{B}\|_{\mathrm{F}}$$

holds simultaneously for all $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ lying within the tangent space of $\boldsymbol{M}^\natural$.

Proof. This result has been established in [?, Section 4.2] for asymmetric sampling patterns (where each (i,j), $i \neq j$ is included in Ω independently). It is straightforward to extend the proof and the result to symmetric sampling patterns (where each (i,j), $i \geq j$ is included in Ω independently). We omit the proof for conciseness. □

Lemma 26. *Fix a matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$. Suppose $n^2 p \geq c_0 n \log n$ for some sufficiently large constant $c_0 > 0$. With probability at least $1 - O\left(n^{-10}\right)$, one has*

$$\left\| \frac{1}{p}\mathcal{P}_\Omega(\boldsymbol{M}) - \boldsymbol{M} \right\| \leq C\sqrt{\frac{n}{p}}\, \|\boldsymbol{M}\|_\infty,$$

where $C > 0$ is some absolute constant.

Proof. See [?, Lemma 3.2]. Similar to Lemma 25, the result therein was provided for the asymmetric sampling patterns but can be easily extended to the symmetric case. □

Lemma 27. *Recall from Section 3.2 that $\boldsymbol{E} \in \mathbb{R}^{n\times n}$ is the symmetric noise matrix. Suppose the sample size obeys $n^2 p \geq c_0 n \log^2 n$ for some sufficiently large constant $c_0 > 0$. With probability at least $1 - O\left(n^{-10}\right)$, one has*

$$\left\| \frac{1}{p}\mathcal{P}_\Omega(\boldsymbol{E}) \right\| \leq C\sigma\sqrt{\frac{n}{p}},$$

where $C > 0$ is some universal constant.

Proof. See [?, Lemma 11]. □

Lemma 28. *Fix some matrix $\boldsymbol{A} \in \mathbb{R}^{n\times r}$ with $n \geq 2r$ and some $1 \leq l \leq n$. Suppose $\{\delta_{l,j}\}_{1\leq j\leq n}$ are independent Bernoulli random variables with means $\{p_j\}_{1\leq j\leq n}$ no more than p. Define*

$$\boldsymbol{G}_l(\boldsymbol{A}) := \left[\delta_{l,1}\boldsymbol{A}_{1,\cdot}^{\top}, \delta_{l,2}\boldsymbol{A}_{2,\cdot}^{\top}, \cdots, \delta_{l,n}\boldsymbol{A}_{n,\cdot}^{\top}\right] \in \mathbb{R}^{r\times n}.$$

Then one has

$$\mathsf{Median}\left[\|\boldsymbol{G}_l(\boldsymbol{A})\|\right] \leq \sqrt{p\|\boldsymbol{A}\|^2 + \sqrt{2p\|\boldsymbol{A}\|_{2,\infty}^2\|\boldsymbol{A}\|^2\log(4r)} + \frac{2\|\boldsymbol{A}\|_{2,\infty}^2}{3}\log(4r)}$$

and for any constant $C \geq 3$, with probability exceeding $1 - n^{-(1.5C-1)}$

$$\left\|\sum_{j=1}^{n}(\delta_{l,j} - p)\boldsymbol{A}_{j,\cdot}^{\top}\boldsymbol{A}_{j,\cdot}\right\| \leq C\left(\sqrt{p\|\boldsymbol{A}\|_{2,\infty}^2\|\boldsymbol{A}\|^2\log n} + \|\boldsymbol{A}\|_{2,\infty}^2\log n\right),$$

and

$$\|\boldsymbol{G}_l(\boldsymbol{A})\| \leq \sqrt{p\|\boldsymbol{A}\|^2 + C\left(\sqrt{p\|\boldsymbol{A}\|_{2,\infty}^2\|\boldsymbol{A}\|^2\log n} + \|\boldsymbol{A}\|_{2,\infty}^2\log n\right)}.$$

Proof. By the definition of $\boldsymbol{G}_l(\boldsymbol{A})$ and the triangle inequality, one has

$$\|\boldsymbol{G}_l(\boldsymbol{A})\|^2 = \left\|\boldsymbol{G}_l(\boldsymbol{A})\boldsymbol{G}_l(\boldsymbol{A})^{\top}\right\| = \left\|\sum_{j=1}^{n}\delta_{l,j}\boldsymbol{A}_{j,\cdot}^{\top}\boldsymbol{A}_{j,\cdot}\right\| \leq \left\|\sum_{j=1}^{n}(\delta_{l,j} - p_j)\boldsymbol{A}_{j,\cdot}^{\top}\boldsymbol{A}_{j,\cdot}\right\| + p\|\boldsymbol{A}\|^2.$$

Therefore, it suffices to control the first term. It can be seen that $\left\{(\delta_{l,j} - p_j)\boldsymbol{A}_{j,\cdot}^{\top}\boldsymbol{A}_{j,\cdot}\right\}_{1\leq j\leq n}$ are i.i.d. zero-mean random matrices. Letting

$$L := \max_{1\leq j\leq n}\left\|(\delta_{l,j} - p_j)\boldsymbol{A}_{j,\cdot}^{\top}\boldsymbol{A}_{j,\cdot}\right\| \leq \|\boldsymbol{A}\|_{2,\infty}^2$$

and $$V := \left\|\sum_{j=1}^{n}\mathbb{E}\left[(\delta_{l,j} - p_j)^2\boldsymbol{A}_{j,\cdot}^{\top}\boldsymbol{A}_{j,\cdot}\boldsymbol{A}_{j,\cdot}^{\top}\boldsymbol{A}_{j,\cdot}\right]\right\| \leq \mathbb{E}\left[(\delta_{l,j} - p_j)^2\right]\|\boldsymbol{A}\|_{2,\infty}^2\left\|\sum_{j=1}^{n}\boldsymbol{A}_{j,\cdot}^{\top}\boldsymbol{A}_{j,\cdot}\right\| \leq p\|\boldsymbol{A}\|_{2,\infty}^2\|\boldsymbol{A}\|^2$$

and invoking matrix Bernstein's inequality [**?**, Theorem 6.1.1], one has for all $t \geq 0$,

$$\mathbb{P}\left\{\left\|\sum_{j=1}^{n}(\delta_{l,j} - p_j)\boldsymbol{A}_{j,\cdot}^{\top}\boldsymbol{A}_{j,\cdot}\right\| \geq t\right\} \leq 2r\cdot\exp\left(\frac{-t^2/2}{p\|\boldsymbol{A}\|_{2,\infty}^2\|\boldsymbol{A}\|^2 + \|\boldsymbol{A}\|_{2,\infty}^2\cdot t/3}\right). \tag{141}$$

We can thus find an upper bound on $\mathsf{Median}\left[\left\|\sum_{j=1}^{n}(\delta_{l,j} - p_j)\boldsymbol{A}_{j,\cdot}^{\top}\boldsymbol{A}_{j,\cdot}\right\|\right]$ by finding a value t that ensures the right-hand side of (141) is smaller than $1/2$. Using this strategy and some simple calculations, we get

$$\mathsf{Median}\left[\left\|\sum_{j=1}^{n}(\delta_{l,j} - p_j)\boldsymbol{A}_{j,\cdot}^{\top}\boldsymbol{A}_{j,\cdot}\right\|\right] \leq \sqrt{2p\|\boldsymbol{A}\|_{2,\infty}^2\|\boldsymbol{A}\|^2\log(4r)} + \frac{2\|\boldsymbol{A}\|_{2,\infty}^2}{3}\log(4r)$$

and for any $C \geq 3$,

$$\left\|\sum_{j=1}^{n}(\delta_{l,j} - p_j)\boldsymbol{A}_{j,\cdot}^{\top}\boldsymbol{A}_{j,\cdot}\right\| \leq C\left(\sqrt{p\|\boldsymbol{A}\|_{2,\infty}^2\|\boldsymbol{A}\|^2\log n} + \|\boldsymbol{A}\|_{2,\infty}^2\log n\right)$$

holds with probability at least $1 - n^{-(1.5C-1)}$. As a consequence, we have

$$\mathsf{Median}\left[\|\boldsymbol{G}_l(\boldsymbol{A})\|\right] \leq \sqrt{p\|\boldsymbol{A}\|^2 + \sqrt{2p\|\boldsymbol{A}\|_{2,\infty}^2\|\boldsymbol{A}\|^2\log(4r)} + \frac{2\|\boldsymbol{A}\|_{2,\infty}^2}{3}\log(4r)},$$

and with probability exceeding $1-n^{-(1.5C-1)}$,

$$\|\boldsymbol{G}_l(\boldsymbol{A})\|^2 \le p\|\boldsymbol{A}\|^2 + C\left(\sqrt{p\|\boldsymbol{A}\|_{2,\infty}^2\|\boldsymbol{A}\|^2\log n} + \|\boldsymbol{A}\|_{2,\infty}^2\log n\right).$$

This completes the proof. □

Lemma 29. *Let $\{\delta_{l,j}\}_{1\le l\le j\le n}$ be i.i.d. Bernoulli random variables with mean p and $\delta_{l,j}=\delta_{j,l}$. For any $\boldsymbol{\Delta}\in\mathbb{R}^{n\times r}$, define*

$$\boldsymbol{G}_l(\boldsymbol{\Delta}) := \left[\delta_{l,1}\boldsymbol{\Delta}_{1,\cdot}^\top, \delta_{l,2}\boldsymbol{\Delta}_{2,\cdot}^\top, \cdots, \delta_{l,n}\boldsymbol{\Delta}_{n,\cdot}^\top\right] \in \mathbb{R}^{r\times n}.$$

Suppose the sample size obeys $n^2p \gg \kappa\mu rn\log^2 n$. Then for any $k>0$ and $\alpha>0$ large enough, with probability at least $1-c_1e^{-\alpha Cnr\log n/2}$,

$$\sum_{l=1}^{n}\mathbb{1}_{\left\{\|\boldsymbol{G}_l(\boldsymbol{\Delta})\|\ge 4\sqrt{p}\psi+2\sqrt{kr}\xi\right\}} \le \frac{2\alpha n\log n}{k}$$

holds simultaneously for all $\boldsymbol{\Delta}\in\mathbb{R}^{n\times r}$ obeying

$$\|\boldsymbol{\Delta}\|_{2,\infty} \le C_5\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} + C_8\sigma\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} := \xi$$

$$\text{and}\qquad \|\boldsymbol{\Delta}\| \le C_9\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\natural\right\| + C_{10}\sigma\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\natural\right\| := \psi,$$

where $c_1, C_5, C_8, C_9, C_{10}>0$ are some absolute constants.

Proof. For simplicity of presentation, we will prove the claim for the asymmetric case where $\{\delta_{l,j}\}_{1\le l,j\le n}$ are independent. The results immediately carry over to the symmetric case as claimed in this lemma. To see this, note that we can always divide $\boldsymbol{G}_l(\boldsymbol{\Delta})$ into

$$\boldsymbol{G}_l(\boldsymbol{\Delta}) = \boldsymbol{G}_l^{\text{upper}}(\boldsymbol{\Delta}) + \boldsymbol{G}_l^{\text{lower}}(\boldsymbol{\Delta}),$$

where all nonzero components of $\boldsymbol{G}_l^{\text{upper}}(\boldsymbol{\Delta})$ come from the upper triangular part (those blocks with $l\le j$), while all nonzero components of $\boldsymbol{G}_l^{\text{lower}}(\boldsymbol{\Delta})$ are from the lower triangular part (those blocks with $l>j$). We can then look at $\{\boldsymbol{G}_l^{\text{upper}}(\boldsymbol{\Delta}) \mid 1\le l\le n\}$ and $\{\boldsymbol{G}_l^{\text{upper}}(\boldsymbol{\Delta}) \mid 1\le l\le n\}$ separately using the argument we develop for the asymmetric case. From now on, we assume that $\{\delta_{l,j}\}_{1\le l,j\le n}$ are independent.

Suppose for the moment that $\boldsymbol{\Delta}$ is statistically independent of $\{\delta_{l,j}\}$. Clearly, for any $\boldsymbol{\Delta},\widetilde{\boldsymbol{\Delta}}\in\mathbb{R}^{n\times r}$,

$$\begin{aligned}
\left|\|\boldsymbol{G}_l(\boldsymbol{\Delta})\| - \|\boldsymbol{G}_l(\widetilde{\boldsymbol{\Delta}})\|\right| &\le \left\|\boldsymbol{G}_l(\boldsymbol{\Delta}) - \boldsymbol{G}_l(\widetilde{\boldsymbol{\Delta}})\right\| \le \left\|\boldsymbol{G}_l(\boldsymbol{\Delta}) - \boldsymbol{G}_l(\widetilde{\boldsymbol{\Delta}})\right\|_{\mathrm{F}} \\
&\le \sqrt{\sum_{j=1}^{n}\left\|\boldsymbol{\Delta}_{j,\cdot} - \widetilde{\boldsymbol{\Delta}}_{j,\cdot}\right\|_2^2} \\
&:= d(\boldsymbol{\Delta},\widetilde{\boldsymbol{\Delta}}),
\end{aligned}$$

which implies that $\|\boldsymbol{G}_l(\boldsymbol{\Delta})\|$ is 1-Lipschitz with respect to the metric $d(\cdot,\cdot)$. Moreover,

$$\max_{1\le j\le n}\|\delta_{l,j}\boldsymbol{\Delta}_{j,\cdot}\|_2 \le \|\boldsymbol{\Delta}\|_{2,\infty} \le \xi$$

according to our assumption. Hence, Talagrand's inequality [**?**, Proposition 1] reveals the existence of some absolute constants $C,c>0$ such that for all $\lambda>0$

$$\mathbb{P}\left\{\|\boldsymbol{G}_l(\boldsymbol{\Delta})\| - \mathsf{Median}\left[\|\boldsymbol{G}_l(\boldsymbol{\Delta})\|\right] \ge \lambda\xi\right\} \le C\exp\left(-c\lambda^2\right). \tag{142}$$

We then proceed to control Median $[\|\boldsymbol{G}_l(\boldsymbol{\Delta})\|]$. A direct application of Lemma 28 yields

$$\mathsf{Median}\left[\|\boldsymbol{G}_l(\boldsymbol{\Delta})\|\right] \le \sqrt{2p\psi^2 + \sqrt{p\log(4r)}\xi\psi + \frac{2\xi^2}{3}\log(4r)} \le 2\sqrt{p}\psi,$$

where the last relation holds since $p\psi^2 \gg \xi^2 \log r$, which follows by combining the definitions of ψ and ξ, the sample size condition $np \gg \kappa\mu r \log^2 n$, and the incoherence condition (56). Thus, substitution into (142) and taking $\lambda = \sqrt{kr}$ give

$$\mathbb{P}\left\{\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right)\right\| \geq 2\sqrt{p}\psi + \sqrt{kr}\xi\right\} \leq C\exp\left(-ckr\right) \tag{143}$$

for any $k \geq 0$. Furthermore, invoking [?, Corollary A.1.14] and using the bound (143), one has

$$\mathbb{P}\left(\sum_{l=1}^{n} \mathbb{1}_{\left\{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta})\right\| \geq 2\sqrt{p}\psi + \sqrt{kr}\xi\right\}} \geq tnC\exp\left(-ckr\right)\right) \leq 2\exp\left(-\frac{t\log t}{2} nC\exp\left(-ckr\right)\right)$$

for any $t \geq 6$. Choose $t = \alpha \log n / \left[kC\exp\left(-ckr\right)\right] \geq 6$ to obtain

$$\mathbb{P}\left(\sum_{l=1}^{n} \mathbb{1}_{\left\{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta})\right\| \geq 2\sqrt{p}\psi + \sqrt{kr}\xi\right\}} \geq \frac{\alpha n \log n}{k}\right) \leq 2\exp\left(-\frac{\alpha C}{2} nr\log n\right). \tag{144}$$

So far we have demonstrated that for any fixed $\boldsymbol{\Delta}$ obeying our assumptions, $\sum_{l=1}^{n} \mathbb{1}_{\left\{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta})\right\| \geq 2\sqrt{p}\psi + \sqrt{kr}\xi\right\}}$ is well controlled with exponentially high probability. In order to extend the results to all feasible $\boldsymbol{\Delta}$, we resort to the standard ϵ-net argument. Clearly, due to the homogeneity property of $\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right)\right\|$, it suffices to restrict attention to the following set:

$$\mathcal{S} = \left\{\boldsymbol{\Delta} \mid \min\left\{\xi, \psi\right\} \leq \left\|\boldsymbol{\Delta}\right\| \leq \psi\right\}, \tag{145}$$

where $\psi/\xi \lesssim \|\boldsymbol{X}^\natural\|/\|\boldsymbol{X}^\natural\|_{2,\infty} \lesssim \sqrt{n}$. We then proceed with the following steps.

1. Introduce the auxiliary function

$$\chi_l(\boldsymbol{\Delta}) = \begin{cases} 1, & \text{if } \left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right)\right\| \geq 4\sqrt{p}\psi + 2\sqrt{kr}\xi, \\ \frac{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta})\right\| - 2\sqrt{p}\psi - \sqrt{kr}\xi}{2\sqrt{p}\psi + \sqrt{kr}\xi}, & \text{if } \left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right)\right\| \in \left[2\sqrt{p}\psi + \sqrt{kr}\xi,\ 4\sqrt{p}\psi + 2\sqrt{kr}\xi\right], \\ 0, & \text{else.} \end{cases}$$

   Clearly, this function is sandwiched between two indicator functions

$$\mathbb{1}_{\left\{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta})\right\| \geq 4\sqrt{p}\psi + 2\sqrt{kr}\xi\right\}} \leq \chi_l(\boldsymbol{\Delta}) \leq \mathbb{1}_{\left\{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta})\right\| \geq 2\sqrt{p}\psi + \sqrt{kr}\xi\right\}}.$$

   Note that χ_l is more convenient to work with due to continuity.

2. Consider an ϵ-net $\mathcal{N}_\epsilon$ [?, Section 2.3.1] of the set $\mathcal{S}$ as defined in (145). For any $\epsilon = 1/n^{O(1)}$, one can find such a net with cardinality $\log|\mathcal{N}_\epsilon| \lesssim nr\log n$. Apply the union bound and (144) to yield

$$\begin{aligned} \mathbb{P}\left(\sum_{l=1}^{n} \chi_l(\boldsymbol{\Delta}) \geq \frac{\alpha n \log n}{k},\ \forall \boldsymbol{\Delta} \in \mathcal{N}_\epsilon\right) &\leq \mathbb{P}\left(\sum_{l=1}^{n} \mathbb{1}_{\left\{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta})\right\| \geq 2\sqrt{p}\psi + \sqrt{kr}\xi\right\}} \geq \frac{\alpha n \log n}{k},\ \forall \boldsymbol{\Delta} \in \mathcal{N}_\epsilon\right) \\ &\leq 2|\mathcal{N}_\epsilon|\exp\left(-\frac{\alpha C}{2} nr\log n\right) \leq 2\exp\left(-\frac{\alpha C}{4} nr\log n\right), \end{aligned}$$

   as long as α is chosen to be sufficiently large.

3. One can then use the continuity argument to extend the bound to all $\boldsymbol{\Delta}$ outside the ϵ-net, i.e. with exponentially high probability,

$$\begin{aligned} & \sum_{l=1}^{n} \chi_l(\boldsymbol{\Delta}) \leq \frac{2\alpha n \log n}{k}, \quad \forall \boldsymbol{\Delta} \in \mathcal{S} \\ \Longrightarrow \quad & \sum_{l=1}^{n} \mathbb{1}_{\left\{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta})\right\| \geq 4\sqrt{p}\psi + 2\sqrt{kr}\xi\right\}} \leq \sum_{l=1}^{n} \chi_l(\boldsymbol{\Delta}) \leq \frac{2\alpha n \log n}{k}, \quad \forall \boldsymbol{\Delta} \in \mathcal{S} \end{aligned}$$

   This is fairly standard (see, e.g. [?, Section 2.3.1]) and is thus omitted here.

We have thus concluded the proof. $\square$

Lemma 30. *Suppose the sample size obeys $n^2 p \geq C\kappa\mu r n \log n$ for some sufficiently large constant $C > 0$. Then with probability at least $1 - O\left(n^{-10}\right)$,*

$$\left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{X}\boldsymbol{X}^{\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right) \right\| \leq 2n\epsilon^2 \left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}^2 + 4\epsilon\sqrt{n}\log n \left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\left\|\boldsymbol{X}^{\natural}\right\|$$

holds simultaneously for all $\boldsymbol{X} \in \mathbb{R}^{n\times r}$ satisfying

$$\left\|\boldsymbol{X} - \boldsymbol{X}^{\natural}\right\|_{2,\infty} \leq \epsilon \left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}, \tag{146}$$

where $\epsilon > 0$ is any fixed constant.

Proof. To simplify the notations hereafter, we denote $\boldsymbol{\Delta} := \boldsymbol{X} - \boldsymbol{X}^{\natural}$. With this notation in place, one can decompose

$$\boldsymbol{X}\boldsymbol{X}^{\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top} = \boldsymbol{\Delta}\boldsymbol{X}^{\natural\top} + \boldsymbol{X}^{\natural}\boldsymbol{\Delta}^{\top} + \boldsymbol{\Delta}\boldsymbol{\Delta}^{\top},$$

which together with the triangle inequality implies that

$$\begin{aligned}\left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{X}\boldsymbol{X}^{\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right) \right\| &\leq \left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right)\right\| + \left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{X}^{\natural}\boldsymbol{\Delta}^{\top}\right)\right\| + \left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{\Delta}\boldsymbol{\Delta}^{\top}\right)\right\| \\ &= \underbrace{\left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{\Delta}\boldsymbol{\Delta}^{\top}\right)\right\|}_{:=\alpha_1} + 2\underbrace{\left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right)\right\|}_{:=\alpha_2}. \end{aligned} \tag{147}$$

In the sequel, we bound α_1 and α_2 separately.

1. Recall from [?, Theorem 2.5] the elementary inequality that

$$\|\boldsymbol{C}\| \leq \big\||\boldsymbol{C}|\big\|, \tag{148}$$

   where $|\boldsymbol{C}| := [|c_{i,j}|]_{1\leq i,j\leq n}$ for any matrix $\boldsymbol{C} = [c_{i,j}]_{1\leq i,j\leq n}$. In addition, for any matrix $\boldsymbol{D} := [d_{i,j}]_{1\leq i,j\leq n}$ such that $|d_{i,j}| \geq |c_{i,j}|$ for all i and j, one has $\big\||\boldsymbol{C}|\big\| \leq \big\||\boldsymbol{D}|\big\|$. Therefore

$$\alpha_1 \leq \left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\left|\boldsymbol{\Delta}\boldsymbol{\Delta}^{\top}\right|\right)\right\| \leq \|\boldsymbol{\Delta}\|_{2,\infty}^2 \left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\mathbf{1}\mathbf{1}^{\top}\right)\right\|.$$

   Lemma 26 then tells us that with probability at least $1 - O(n^{-10})$,

$$\left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\mathbf{1}\mathbf{1}^{\top}\right) - \mathbf{1}\mathbf{1}^{\top}\right\| \leq C\sqrt{\frac{n}{p}} \tag{149}$$

   for some universal constant $C > 0$, as long as $p \gg \log n/n$. This together with the triangle inequality yields

$$\left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\mathbf{1}\mathbf{1}^{\top}\right)\right\| \leq \left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\mathbf{1}\mathbf{1}^{\top}\right) - \mathbf{1}\mathbf{1}^{\top}\right\| + \left\|\mathbf{1}\mathbf{1}^{\top}\right\| \leq C\sqrt{\frac{n}{p}} + n \leq 2n, \tag{150}$$

   provided that $p \gg 1/n$. Putting together the previous bounds, we arrive at

$$\alpha_1 \leq 2n \|\boldsymbol{\Delta}\|_{2,\infty}^2. \tag{151}$$

2. Regarding the second term α_2, apply the elementary inequality (148) once again to get

$$\left\|\mathcal{P}_{\Omega}\left(\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right)\right\| \leq \left\|\mathcal{P}_{\Omega}\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right|\right)\right\|,$$

   which motivates us to look at $\left\|\mathcal{P}_{\Omega}\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right|\right)\right\|$ instead. A key step of this part is to take advantage of the $\ell_{2,\infty}$ norm constraint of $\mathcal{P}_{\Omega}\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right|\right)$. Specifically, we claim for the moment that with probability exceeding $1 - O(n^{-10})$,

$$\left\|\mathcal{P}_{\Omega}\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right|\right)\right\|_{2,\infty}^2 \leq 2p\sigma_{\max}\|\boldsymbol{\Delta}\|_{2,\infty}^2 := \theta \tag{152}$$

holds under our sample size condition. In addition, we also have the following trivial ℓ_∞ norm bound

$$\left\|\mathcal{P}_\Omega\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right|\right)\right\|_\infty \le \|\boldsymbol{\Delta}\|_{2,\infty}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} := \gamma. \tag{153}$$

In what follows, for simplicity of presentation, we will denote

$$\boldsymbol{A} := \mathcal{P}_\Omega\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right|\right). \tag{154}$$

(a) To facilitate the analysis of $\|\boldsymbol{A}\|$, we first introduce $k_0+1 = \frac{1}{2}\log(\kappa\mu r)$ auxiliary matrices[2] $\boldsymbol{B}_s \in \mathbb{R}^{n\times n}$ that satisfy

$$\|\boldsymbol{A}\| \le \|\boldsymbol{B}_{k_0}\| + \sum_{s=0}^{k_0-1}\|\boldsymbol{B}_s\|. \tag{155}$$

To be precise, each $\boldsymbol{B}_s$ is defined such that

$$[\boldsymbol{B}_s]_{j,k} = \begin{cases} \frac{1}{2^s}\gamma, & \text{if} \quad A_{j,k} \in (\frac{1}{2^{s+1}}\gamma, \frac{1}{2^s}\gamma], \\ 0, & \text{else,} \end{cases} \qquad \text{for } 0 \le s \le k_0-1 \qquad \text{and}$$

$$[\boldsymbol{B}_{k_0}]_{j,k} = \begin{cases} \frac{1}{2^{k_0}}\gamma, & \text{if} \quad A_{j,k} \le \frac{1}{2^{k_0}}\gamma, \\ 0, & \text{else,} \end{cases}$$

which clearly satisfy (155); in words, $\boldsymbol{B}_s$ is constructed by rounding up those entries of $\boldsymbol{A}$ within a prescribed magnitude interval. Thus, it suffices to bound $\|\boldsymbol{B}_s\|$ for every s. To this end, we start with $s = k_0$ and use the definition of $\boldsymbol{B}_{k_0}$ to get

$$\|\boldsymbol{B}_{k_0}\| \overset{\text{(i)}}{\le} \|\boldsymbol{B}_{k_0}\|_\infty\sqrt{(2np)^2} \overset{\text{(ii)}}{\le} 4np\frac{1}{\sqrt{\kappa\mu r}}\|\boldsymbol{\Delta}\|_{2,\infty}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} \overset{\text{(iii)}}{\le} 4\sqrt{np}\|\boldsymbol{\Delta}\|_{2,\infty}\left\|\boldsymbol{X}^\natural\right\|,$$

where (i) arises from Lemma 31, with $2np$ being a crude upper bound on the number of nonzero entries in each row and each column. This can be derived by applying the standard Chernoff bound on Ω. The second inequality (ii) relies on the definitions of γ and k_0. The last one (iii) follows from the incoherence condition (56). Besides, for any $0 \le s \le k_0-1$, by construction one has

$$\|\boldsymbol{B}_s\|_{2,\infty}^2 \le 4\theta = 8p\sigma_{\max}\|\boldsymbol{\Delta}\|_{2,\infty}^2 \qquad \text{and} \qquad \|\boldsymbol{B}_s\|_\infty = \frac{1}{2^s}\gamma,$$

where θ is as defined in (152). Here, we have used the fact that the magnitude of each entry of $\boldsymbol{B}_s$ is at most 2 times that of $\boldsymbol{A}$. An immediate implication is that there are at most

$$\frac{\|\boldsymbol{B}_s\|_{2,\infty}^2}{\|\boldsymbol{B}_s\|_\infty^2} \le \frac{8p\sigma_{\max}\|\boldsymbol{\Delta}\|_{2,\infty}^2}{\left(\frac{1}{2^s}\gamma\right)^2} := k_{\mathrm{r}}$$

nonzero entries in each row of $\boldsymbol{B}_s$ and at most

$$k_{\mathrm{c}} = 2np$$

nonzero entries in each column of $\boldsymbol{B}_s$, where k_{c} is derived from the standard Chernoff bound on Ω. Utilizing Lemma 31 once more, we discover that

$$\|\boldsymbol{B}_s\| \le \|\boldsymbol{B}_s\|_\infty\sqrt{k_{\mathrm{r}}k_{\mathrm{c}}} = \frac{1}{2^s}\gamma\sqrt{k_{\mathrm{r}}k_{\mathrm{c}}} = \sqrt{16np^2\sigma_{\max}\|\boldsymbol{\Delta}\|_{2,\infty}^2} = 4\sqrt{np}\|\boldsymbol{\Delta}\|_{2,\infty}\left\|\boldsymbol{X}^\natural\right\|$$

for each $0 \le s \le k_0-1$. Combining all, we arrive at

$$\|\boldsymbol{A}\| \le \sum_{s=0}^{k_0-1}\|\boldsymbol{B}_s\| + \|\boldsymbol{B}_{k_0}\| \le (k_0+1)\,4\sqrt{np}\|\boldsymbol{\Delta}\|_{2,\infty}\left\|\boldsymbol{X}^\natural\right\|$$

[2] For simplicity, we assume $\frac{1}{2}\log(\kappa\mu r)$ is an integer. The argument here can be easily adapted to the case when $\frac{1}{2}\log(\kappa\mu r)$ is not an integer.

$$\leq 2\sqrt{n}p \log(\kappa\mu r) \|\mathbf{\Delta}\|_{2,\infty} \left\|\mathbf{X}^\natural\right\|$$
$$\leq 2\sqrt{n}p \log n \|\mathbf{\Delta}\|_{2,\infty} \left\|\mathbf{X}^\natural\right\|,$$

where the last relation holds under the condition $n \geq \kappa\mu r$. This further gives

$$\alpha_2 \leq \frac{1}{p}\|\mathbf{A}\| \leq 2\sqrt{n}\log n \|\mathbf{\Delta}\|_{2,\infty} \left\|\mathbf{X}^\natural\right\|. \tag{156}$$

(b) In order to finish the proof of this part, we need to justify the claim (152). Observe that

$$\begin{aligned}\left\|\left[\mathcal{P}_\Omega\left(\left|\mathbf{\Delta}\mathbf{X}^{\natural\top}\right|\right)\right]_{l,\cdot}\right\|_2^2 &= \sum\nolimits_{j=1}^n \left(\mathbf{\Delta}_{l,\cdot}\mathbf{X}_{j,\cdot}^{\natural\top}\delta_{l,j}\right)^2 \\ &= \mathbf{\Delta}_{l,\cdot}\left(\sum\nolimits_{j=1}^n \delta_{l,j}\mathbf{X}_{j,\cdot}^{\natural\top}\mathbf{X}_{j,\cdot}^\natural\right)\mathbf{\Delta}_{l,\cdot}^\top \\ &\leq \|\mathbf{\Delta}\|_{2,\infty}^2 \left\|\sum\nolimits_{j=1}^n \delta_{l,j}\mathbf{X}_{j,\cdot}^{\natural\top}\mathbf{X}_{j,\cdot}^\natural\right\| \end{aligned} \tag{157}$$

for every $1 \leq l \leq n$, where $\delta_{l,j}$ indicates whether the entry with the index (l,j) is observed or not. Invoke Lemma 28 to yield

$$\begin{aligned}\left\|\sum\nolimits_{j=1}^n \delta_{l,j}\mathbf{X}_{j,\cdot}^{\natural\top}\mathbf{X}_{j,\cdot}^\natural\right\| &= \left\|\left[\delta_{l,1}\mathbf{X}_{1,\cdot}^{\natural\top}, \delta_{l,2}\mathbf{X}_{2,\cdot}^{\natural\top}, \cdots, \delta_{l,n}\mathbf{X}_{n,\cdot}^{\natural\top}\right]\right\|^2 \\ &\leq p\sigma_{\max} + C\left(\sqrt{p\|\mathbf{X}^\natural\|_{2,\infty}^2\|\mathbf{X}^\natural\|^2\log n} + \left\|\mathbf{X}^\natural\right\|_{2,\infty}^2\log n\right) \\ &\leq \left(p + C\sqrt{\frac{p\kappa\mu r\log n}{n}} + C\frac{\kappa\mu r\log n}{n}\right)\sigma_{\max} \\ &\leq 2p\sigma_{\max}, \end{aligned} \tag{158}$$

with high probability, as soon as $np \gg \kappa\mu r\log n$. Combining (157) and (158) yields

$$\left\|\left[\mathcal{P}_\Omega\left(\left|\mathbf{\Delta}\mathbf{X}^{\natural\top}\right|\right)\right]_{l,\cdot}\right\|_2^2 \leq 2p\sigma_{\max}\|\mathbf{\Delta}\|_{2,\infty}^2, \qquad 1 \leq l \leq n$$

as claimed in (152).

3. Taken together, the preceding bounds (147), (151) and (156) yield

$$\left\|\frac{1}{p}\mathcal{P}_\Omega\left(\mathbf{X}\mathbf{X}^\top - \mathbf{X}^\natural\mathbf{X}^{\natural\top}\right)\right\| \leq \alpha_1 + 2\alpha_2 \leq 2n\|\mathbf{\Delta}\|_{2,\infty}^2 + 4\sqrt{n}\log n\|\mathbf{\Delta}\|_{2,\infty}\left\|\mathbf{X}^\natural\right\|.$$

The proof is completed by substituting the assumption $\|\mathbf{\Delta}\|_{2,\infty} \leq \epsilon\left\|\mathbf{X}^\natural\right\|_{2,\infty}$. □

In the end of this subsection, we record a useful lemma to bound the spectral norm of a sparse Bernoulli matrix.

Lemma 31. *Let $\mathbf{A} \in \{0,1\}^{n_1\times n_2}$ be a binary matrix, and suppose that there are at most k_r and k_c nonzero entries in each row and column of $\mathbf{A}$, respectively. Then one has $\|\mathbf{A}\| \leq \sqrt{k_\mathrm{c}k_\mathrm{r}}$.*

Proof. This immediately follows from the elementary inequality $\|\mathbf{A}\|^2 \leq \|\mathbf{A}\|_{1\to1}\|\mathbf{A}\|_{\infty\to\infty}$ (see [**?**, equation (1.11)]), where $\|\mathbf{A}\|_{1\to1}$ and $\|\mathbf{A}\|_{\infty\to\infty}$ are the induced 1-norm (or maximum absolute column sum norm) and the induced ∞-norm (or maximum absolute row sum norm), respectively. □

6.2.3 Matrix perturbation bounds

Lemma 32. *Let $\mathbf{M} \in \mathbb{R}^{n\times n}$ be a symmetric matrix with the top-r eigendecomposition $\mathbf{U\Sigma U}^\top$. Assume $\left\|\mathbf{M} - \mathbf{M}^\natural\right\| \leq \sigma_{\min}/2$ and denote*

$$\widehat{\mathbf{Q}} := \underset{\mathbf{R}\in\mathcal{O}^{r\times r}}{\operatorname{argmin}}\left\|\mathbf{U}\mathbf{R} - \mathbf{U}^\natural\right\|_\mathrm{F}.$$

Then there is some numerical constant $c_3 > 0$ such that

$$\left\|\mathbf{U}\widehat{\mathbf{Q}} - \mathbf{U}^\natural\right\| \leq \frac{c_3}{\sigma_{\min}}\left\|\mathbf{M} - \mathbf{M}^\natural\right\|.$$

Proof. Define $\boldsymbol{Q} = \boldsymbol{U}^\top \boldsymbol{U}^\natural$. The triangle inequality gives

$$\left\|\boldsymbol{U}\widehat{\boldsymbol{Q}} - \boldsymbol{U}^\natural\right\| \le \left\|\boldsymbol{U}(\widehat{\boldsymbol{Q}} - \boldsymbol{Q})\right\| + \left\|\boldsymbol{U}\boldsymbol{Q} - \boldsymbol{U}^\natural\right\| \le \left\|\widehat{\boldsymbol{Q}} - \boldsymbol{Q}\right\| + \left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\natural - \boldsymbol{U}^\natural\right\|. \tag{159}$$

[?, Lemma 3] asserts that

$$\left\|\widehat{\boldsymbol{Q}} - \boldsymbol{Q}\right\| \le 4\left(\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\|/\sigma_{\min}\right)^2$$

as long as $\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\| \le \sigma_{\min}/2$. For the remaining term in (159), one can use $\boldsymbol{U}^{\natural\top}\boldsymbol{U}^\natural = \boldsymbol{I}_r$ to obtain

$$\left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\natural - \boldsymbol{U}^\natural\right\| = \left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\natural - \boldsymbol{U}^\natural\boldsymbol{U}^{\natural\top}\boldsymbol{U}^\natural\right\| \le \left\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\natural\boldsymbol{U}^{\natural\top}\right\|,$$

which together with the Davis-Kahan sinΘ theorem [?] reveals that

$$\left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\natural - \boldsymbol{U}^\natural\right\| \le \frac{c_2}{\sigma_{\min}}\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\|$$

for some constant $c_2 > 0$. Combine the estimates on $\left\|\widehat{\boldsymbol{Q}} - \boldsymbol{Q}\right\|$, $\left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\natural - \boldsymbol{U}^\natural\right\|$ and (159) to reach

$$\left\|\boldsymbol{U}\widehat{\boldsymbol{Q}} - \boldsymbol{U}^\natural\right\| \le \left(\frac{4}{\sigma_{\min}}\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\|\right)^2 + \frac{c_2}{\sigma_{\min}}\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\| \le \frac{c_3}{\sigma_{\min}}\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\|$$

for some numerical constant $c_3 > 0$, where we have utilized the fact that $\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\|/\sigma_{\min} \le 1/2$. □

Lemma 33. *Let $\boldsymbol{M}, \widetilde{\boldsymbol{M}} \in \mathbb{R}^{n\times n}$ be two symmetric matrices with top-r eigendecompositions $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^\top$ and $\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{U}}^\top$, respectively. Assume $\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\| \le \sigma_{\min}/4$ and $\left\|\widetilde{\boldsymbol{M}} - \boldsymbol{M}^\natural\right\| \le \sigma_{\min}/4$, and suppose $\sigma_{\max}/\sigma_{\min}$ is bounded by some constant $c_1 > 0$, with $\sigma_{\max}$ and $\sigma_{\min}$ the largest and the smallest singular values of $\boldsymbol{M}^\natural$, respectively. If we denote*

$$\boldsymbol{Q} := \operatorname*{argmin}_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \left\|\boldsymbol{U}\boldsymbol{R} - \widetilde{\boldsymbol{U}}\right\|_{\mathrm{F}},$$

then there exists some numerical constant $c_3 > 0$ such that

$$\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{Q} - \boldsymbol{Q}\widetilde{\boldsymbol{\Sigma}}^{1/2}\right\| \le \frac{c_3}{\sqrt{\sigma_{\min}}}\left\|\widetilde{\boldsymbol{M}} - \boldsymbol{M}\right\| \qquad \text{and} \qquad \left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{Q} - \boldsymbol{Q}\widetilde{\boldsymbol{\Sigma}}^{1/2}\right\|_{\mathrm{F}} \le \frac{c_3}{\sqrt{\sigma_{\min}}}\left\|(\widetilde{\boldsymbol{M}} - \boldsymbol{M})\boldsymbol{U}\right\|_{\mathrm{F}}.$$

Proof. Here, we focus on the Frobenius norm; the bound on the operator norm follows from the same argument, and hence we omit the proof. Since $\|\cdot\|_{\mathrm{F}}$ is unitarily invariant, we have

$$\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{Q} - \boldsymbol{Q}\widetilde{\boldsymbol{\Sigma}}^{1/2}\right\|_{\mathrm{F}} = \left\|\boldsymbol{Q}^\top\boldsymbol{\Sigma}^{1/2}\boldsymbol{Q} - \widetilde{\boldsymbol{\Sigma}}^{1/2}\right\|_{\mathrm{F}},$$

where $\boldsymbol{Q}^\top\boldsymbol{\Sigma}^{1/2}\boldsymbol{Q}$ and $\widetilde{\boldsymbol{\Sigma}}^{1/2}$ are the matrix square roots of $\boldsymbol{Q}^\top\boldsymbol{\Sigma}\boldsymbol{Q}$ and $\widetilde{\boldsymbol{\Sigma}}$, respectively. In view of the matrix square root perturbation bound [?, Lemma 2.1],

$$\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{Q} - \boldsymbol{Q}\widetilde{\boldsymbol{\Sigma}}^{1/2}\right\|_{\mathrm{F}} \le \frac{1}{\sigma_{\min}\left[(\boldsymbol{\Sigma})^{1/2}\right] + \sigma_{\min}\left[(\widetilde{\boldsymbol{\Sigma}})^{1/2}\right]}\left\|\boldsymbol{Q}^\top\boldsymbol{\Sigma}\boldsymbol{Q} - \widetilde{\boldsymbol{\Sigma}}\right\|_{\mathrm{F}} \le \frac{1}{\sqrt{\sigma_{\min}}}\left\|\boldsymbol{Q}^\top\boldsymbol{\Sigma}\boldsymbol{Q} - \widetilde{\boldsymbol{\Sigma}}\right\|_{\mathrm{F}}, \tag{160}$$

where the last inequality follows from the lower estimates

$$\sigma_{\min}(\boldsymbol{\Sigma}) \ge \sigma_{\min}\left(\boldsymbol{\Sigma}^\natural\right) - \left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\| \ge \sigma_{\min}/4$$

and, similarly, $\sigma_{\min}(\widetilde{\boldsymbol{\Sigma}}) \ge \sigma_{\min}/4$. Recognizing that $\boldsymbol{\Sigma} = \boldsymbol{U}^\top\boldsymbol{M}\boldsymbol{U}$ and $\widetilde{\boldsymbol{\Sigma}} = \widetilde{\boldsymbol{U}}^\top\widetilde{\boldsymbol{M}}\widetilde{\boldsymbol{U}}$, one gets

$$\begin{aligned}\left\|\boldsymbol{Q}^\top\boldsymbol{\Sigma}\boldsymbol{Q} - \widetilde{\boldsymbol{\Sigma}}\right\|_{\mathrm{F}} &= \left\|(\boldsymbol{U}\boldsymbol{Q})^\top\boldsymbol{M}(\boldsymbol{U}\boldsymbol{Q}) - \widetilde{\boldsymbol{U}}^\top\widetilde{\boldsymbol{M}}\widetilde{\boldsymbol{U}}\right\|_{\mathrm{F}} \\ &\le \left\|(\boldsymbol{U}\boldsymbol{Q})^\top\boldsymbol{M}(\boldsymbol{U}\boldsymbol{Q}) - (\boldsymbol{U}\boldsymbol{Q})^\top\widetilde{\boldsymbol{M}}(\boldsymbol{U}\boldsymbol{Q})\right\|_{\mathrm{F}} + \left\|(\boldsymbol{U}\boldsymbol{Q})^\top\widetilde{\boldsymbol{M}}(\boldsymbol{U}\boldsymbol{Q}) - \widetilde{\boldsymbol{U}}^\top\widetilde{\boldsymbol{M}}(\boldsymbol{U}\boldsymbol{Q})\right\|_{\mathrm{F}} \\ &\quad + \left\|\widetilde{\boldsymbol{U}}^\top\widetilde{\boldsymbol{M}}(\boldsymbol{U}\boldsymbol{Q}) - \widetilde{\boldsymbol{U}}^\top\widetilde{\boldsymbol{M}}\widetilde{\boldsymbol{U}}\right\|_{\mathrm{F}}\end{aligned}$$

$$\leq \left\|(\widetilde{\boldsymbol{M}}-\boldsymbol{M})\boldsymbol{U}\right\|_{\mathrm{F}} + 2\|\boldsymbol{U}\boldsymbol{Q}-\widetilde{\boldsymbol{U}}\|_{\mathrm{F}}\|\widetilde{\boldsymbol{M}}\| \leq \left\|(\widetilde{\boldsymbol{M}}-\boldsymbol{M})\boldsymbol{U}\right\|_{\mathrm{F}} + 4\sigma_{\max}\|\boldsymbol{U}\boldsymbol{Q}-\widetilde{\boldsymbol{U}}\|_{\mathrm{F}}, \tag{161}$$

where the last relation holds due to the upper estimate

$$\|\widetilde{\boldsymbol{M}}\| \leq \|\boldsymbol{M}^\natural\| + \|\widetilde{\boldsymbol{M}}-\boldsymbol{M}^\natural\| \leq \sigma_{\max}+\sigma_{\min}/4 \leq 2\sigma_{\max}.$$

Invoke the Davis-Kahan sinΘ theorem [?] to obtain

$$\|\boldsymbol{U}\boldsymbol{Q}-\widetilde{\boldsymbol{U}}\|_{\mathrm{F}} \leq \frac{c_2}{\sigma_r(\boldsymbol{M})-\sigma_{r+1}(\widetilde{\boldsymbol{M}})}\left\|(\widetilde{\boldsymbol{M}}-\boldsymbol{M})\boldsymbol{U}\right\|_{\mathrm{F}} \leq \frac{2c_2}{\sigma_{\min}}\left\|(\widetilde{\boldsymbol{M}}-\boldsymbol{M})\boldsymbol{U}\right\|_{\mathrm{F}}, \tag{162}$$

for some constant $c_2>0$, where the last inequality follows from the bounds

$$\sigma_r(\boldsymbol{M}) \geq \sigma_r\left(\boldsymbol{M}^\natural\right) - \|\boldsymbol{M}-\boldsymbol{M}^\natural\| \geq 3\sigma_{\min}/4,$$
$$\sigma_{r+1}(\widetilde{\boldsymbol{M}}) \leq \sigma_{r+1}\left(\boldsymbol{M}^\natural\right) + \|\widetilde{\boldsymbol{M}}-\boldsymbol{M}^\natural\| \leq \sigma_{\min}/4.$$

Combine (160), (161), (162) and the fact $\sigma_{\max}/\sigma_{\min}\leq c_1$ to reach

$$\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{Q}-\boldsymbol{Q}\widetilde{\boldsymbol{\Sigma}}^{1/2}\right\|_{\mathrm{F}} \leq \frac{c_3}{\sqrt{\sigma_{\min}}}\left\|(\widetilde{\boldsymbol{M}}-\boldsymbol{M})\boldsymbol{U}\right\|_{\mathrm{F}}$$

for some constant $c_3>0$. □

Lemma 34. *Let $\boldsymbol{M}\in\mathbb{R}^{n\times n}$ be a symmetric matrix with the top-r eigendecomposition $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^\top$. Denote $\boldsymbol{X}=\boldsymbol{U}\boldsymbol{\Sigma}^{1/2}$ and $\boldsymbol{X}^\natural=\boldsymbol{U}^\natural(\boldsymbol{\Sigma}^\natural)^{1/2}$, and define*

$$\widehat{\boldsymbol{Q}} := \operatorname*{argmin}_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\|\boldsymbol{U}\boldsymbol{R}-\boldsymbol{U}^\natural\|_{\mathrm{F}} \qquad \text{and} \qquad \widehat{\boldsymbol{H}} := \operatorname*{argmin}_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\|\boldsymbol{X}\boldsymbol{R}-\boldsymbol{X}^\natural\|_{\mathrm{F}}.$$

Assume $\|\boldsymbol{M}-\boldsymbol{M}^\natural\|\leq\sigma_{\min}/2$, and suppose $\sigma_{\max}/\sigma_{\min}$ is bounded by some constant $c_1>0$. Then there exists a numerical constant $c_3>0$ such that

$$\|\widehat{\boldsymbol{Q}}-\widehat{\boldsymbol{H}}\| \leq \frac{c_3}{\sigma_{\min}}\|\boldsymbol{M}-\boldsymbol{M}^\natural\|.$$

Proof. We first collect several useful facts about the spectrum of $\boldsymbol{\Sigma}$. Weyl's inequality tells us that $\|\boldsymbol{\Sigma}-\boldsymbol{\Sigma}^\natural\|\leq\|\boldsymbol{M}-\boldsymbol{M}^\natural\|\leq\sigma_{\min}/2$, which further implies that

$$\sigma_r(\boldsymbol{\Sigma}) \geq \sigma_r\left(\boldsymbol{\Sigma}^\natural\right) - \|\boldsymbol{\Sigma}-\boldsymbol{\Sigma}^\natural\| \geq \sigma_{\min}/2 \qquad \text{and} \qquad \|\boldsymbol{\Sigma}\| \leq \|\boldsymbol{\Sigma}^\natural\| + \|\boldsymbol{\Sigma}-\boldsymbol{\Sigma}^\natural\| \leq 2\sigma_{\max}.$$

Denote

$$\boldsymbol{Q}=\boldsymbol{U}^\top\boldsymbol{U}^\natural \qquad \text{and} \qquad \boldsymbol{H}=\boldsymbol{X}^\top\boldsymbol{X}^\natural.$$

Simple algebra yields

$$\boldsymbol{H}=\boldsymbol{\Sigma}^{1/2}\boldsymbol{Q}\left(\boldsymbol{\Sigma}^\natural\right)^{1/2} = \underbrace{\boldsymbol{\Sigma}^{1/2}(\boldsymbol{Q}-\widehat{\boldsymbol{Q}})\left(\boldsymbol{\Sigma}^\natural\right)^{1/2} + \left(\boldsymbol{\Sigma}^{1/2}\widehat{\boldsymbol{Q}}-\widehat{\boldsymbol{Q}}\boldsymbol{\Sigma}^{1/2}\right)\left(\boldsymbol{\Sigma}^\natural\right)^{1/2}}_{:=\boldsymbol{E}} + \underbrace{\widehat{\boldsymbol{Q}}\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\natural\right)^{1/2}}_{:=\boldsymbol{A}}.$$

It can be easily seen that $\sigma_{r-1}(\boldsymbol{A})\geq\sigma_r(\boldsymbol{A})\geq\sigma_{\min}/2$, and

$$\begin{aligned}\|\boldsymbol{E}\| &\leq \|\boldsymbol{\Sigma}^{1/2}\|\cdot\|\boldsymbol{Q}-\widehat{\boldsymbol{Q}}\|\cdot\|\left(\boldsymbol{\Sigma}^\natural\right)^{1/2}\| + \left\|\boldsymbol{\Sigma}^{1/2}\widehat{\boldsymbol{Q}}-\widehat{\boldsymbol{Q}}\boldsymbol{\Sigma}^{1/2}\right\|\cdot\|\left(\boldsymbol{\Sigma}^\natural\right)^{1/2}\| \\ &\leq 2\sigma_{\max}\underbrace{\|\boldsymbol{Q}-\widehat{\boldsymbol{Q}}\|}_{:=\alpha}+\sqrt{\sigma_{\max}}\underbrace{\left\|\boldsymbol{\Sigma}^{1/2}\widehat{\boldsymbol{Q}}-\widehat{\boldsymbol{Q}}\boldsymbol{\Sigma}^{1/2}\right\|}_{:=\beta},\end{aligned}$$

which can be controlled as follows.

- Regarding α, use [?, Lemma 3] to reach

$$\alpha = \left\|\boldsymbol{Q} - \widehat{\boldsymbol{Q}}\right\| \leq 4 \left\|\boldsymbol{M} - \boldsymbol{M}^{\natural}\right\|^2 / \sigma_{\min}^2.$$

- For β, one has

$$\beta \overset{\text{(i)}}{=} \left\|\widehat{\boldsymbol{Q}}^{\top} \boldsymbol{\Sigma}^{1/2} \widehat{\boldsymbol{Q}} - \boldsymbol{\Sigma}^{1/2}\right\| \overset{\text{(ii)}}{\leq} \frac{1}{2\sigma_r\left(\boldsymbol{\Sigma}^{1/2}\right)} \left\|\widehat{\boldsymbol{Q}}^{\top} \boldsymbol{\Sigma} \widehat{\boldsymbol{Q}} - \boldsymbol{\Sigma}\right\| \overset{\text{(iii)}}{=} \frac{1}{2\sigma_r\left(\boldsymbol{\Sigma}^{1/2}\right)} \left\|\boldsymbol{\Sigma}\widehat{\boldsymbol{Q}} - \widehat{\boldsymbol{Q}}\boldsymbol{\Sigma}\right\|,$$

where (i) and (iii) come from the unitary invariance of $\|\cdot\|$, and (ii) follows from the matrix square root perturbation bound [?, Lemma 2.1]. We can further take the triangle inequality to obtain

$$\begin{aligned} \left\|\boldsymbol{\Sigma}\widehat{\boldsymbol{Q}} - \widehat{\boldsymbol{Q}}\boldsymbol{\Sigma}\right\| &= \left\|\boldsymbol{\Sigma}\boldsymbol{Q} - \boldsymbol{Q}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}(\widehat{\boldsymbol{Q}} - \boldsymbol{Q}) - (\widehat{\boldsymbol{Q}} - \boldsymbol{Q})\boldsymbol{\Sigma}\right\| \\ &\leq \|\boldsymbol{\Sigma}\boldsymbol{Q} - \boldsymbol{Q}\boldsymbol{\Sigma}\| + 2\|\boldsymbol{\Sigma}\| \left\|\boldsymbol{Q} - \widehat{\boldsymbol{Q}}\right\| \\ &= \left\|\boldsymbol{U}\left(\boldsymbol{M} - \boldsymbol{M}^{\natural}\right)\boldsymbol{U}^{\natural\top} + \boldsymbol{Q}\left(\boldsymbol{\Sigma}^{\natural} - \boldsymbol{\Sigma}\right)\right\| + 2\|\boldsymbol{\Sigma}\| \left\|\boldsymbol{Q} - \widehat{\boldsymbol{Q}}\right\| \\ &\leq \left\|\boldsymbol{U}\left(\boldsymbol{M} - \boldsymbol{M}^{\natural}\right)\boldsymbol{U}^{\natural\top}\right\| + \left\|\boldsymbol{Q}\left(\boldsymbol{\Sigma}^{\natural} - \boldsymbol{\Sigma}\right)\right\| + 2\|\boldsymbol{\Sigma}\| \left\|\boldsymbol{Q} - \widehat{\boldsymbol{Q}}\right\| \\ &\leq 2\left\|\boldsymbol{M} - \boldsymbol{M}^{\natural}\right\| + 4\sigma_{\max}\alpha, \end{aligned}$$

where the last inequality uses the Weyl's inequality $\|\boldsymbol{\Sigma}^{\natural} - \boldsymbol{\Sigma}\| \leq \|\boldsymbol{M} - \boldsymbol{M}^{\natural}\|$ and the fact that $\|\boldsymbol{\Sigma}\| \leq 2\sigma_{\max}$.

- Rearrange the previous bounds to arrive at

$$\|\boldsymbol{E}\| \leq 2\sigma_{\max}\alpha + \sqrt{\sigma_{\max}} \frac{1}{\sqrt{\sigma_{\min}}} \left(2\left\|\boldsymbol{M} - \boldsymbol{M}^{\natural}\right\| + 4\sigma_{\max}\alpha\right) \leq c_2 \left\|\boldsymbol{M} - \boldsymbol{M}^{\natural}\right\|$$

for some numerical constant $c_2 > 0$, where we have used the assumption that $\sigma_{\max}/\sigma_{\min}$ is bounded.

Recognizing that $\widehat{\boldsymbol{Q}} = \operatorname{sgn}(\boldsymbol{A})$ (see definition in (119)), we are ready to invoke Lemma 23 to deduce that

$$\left\|\widehat{\boldsymbol{Q}} - \widehat{\boldsymbol{H}}\right\| \leq \frac{2}{\sigma_{r-1}(\boldsymbol{A}) + \sigma_r(\boldsymbol{A})} \|\boldsymbol{E}\| \leq \frac{c_3}{\sigma_{\min}} \left\|\boldsymbol{M} - \boldsymbol{M}^{\natural}\right\|$$

for some constant $c_3 > 0$. □