A 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

$$\max_{1\le j\le m} \|\boldsymbol{a}_j\|_2 \le \sqrt{6n} \tag{98}$$

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

$$\max_{1\le j\le m} \left|\boldsymbol{a}_j^\top \boldsymbol{x}^\star\right| \le 5\sqrt{\log n} \tag{99}$$

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

A.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 $\left\|\nabla^2 f(\boldsymbol{x})\right\| \lesssim \log n$. To this end, we first decompose the Hessian (cf. (44)) into three components as follows:

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

where we have used $y_j = (\boldsymbol{a}_j^\top \boldsymbol{x}^\star)^2$. 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

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

- The second term $\boldsymbol{\Lambda}_2$ can be controlled by means of Lemma 32:

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

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

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

$$\|\boldsymbol{\Lambda}_1\| \le \left\|\frac{3}{m}\sum_{j=1}^m \left|\boldsymbol{a}_j^\top(\boldsymbol{x}-\boldsymbol{x}^\star)\right|\left|\boldsymbol{a}_j^\top(\boldsymbol{x}+\boldsymbol{x}^\star)\right|\boldsymbol{a}_j\boldsymbol{a}_j^\top\right\|. \tag{100}$$

  Under the assumption $\max_{1\le j\le m}\left|\boldsymbol{a}_j^\top(\boldsymbol{x}-\boldsymbol{x}^\star)\right| \le C_2\sqrt{\log n}$ and the fact (99), we can also obtain

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

  Substitution into (100) leads to

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

  where the last inequality is a direct consequence of Lemma 31.

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

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

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

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

We begin with the simpler term $\boldsymbol{\Lambda}_5$. Lemma 32 implies that with probability at least $1 - O(n^{-10})$,

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

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

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

To bound $\boldsymbol{\Lambda}_4$, invoke Lemma 33 to conclude that with probability at least $1 - c_3 e^{-c_2 m}$ (for some constants $c_2, c_3 > 0$),

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

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

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

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

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

which leads to

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

This further implies

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

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

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

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

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

as claimed.

A.2 Proof of Lemma 2

Using the update rule (cf. (17)) as well as the fundamental theorem of calculus [Lan93, Chapter XIII, Theorem 4.2], we get

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

where we denote $\boldsymbol{x}(\tau) = \boldsymbol{x}^\star + \tau(\boldsymbol{x}^t - \boldsymbol{x}^\star)$, $0 \leq \tau \leq 1$. Here, the first equality makes use of the fact that $\nabla f(\boldsymbol{x}^\star) = \boldsymbol{0}$. Under the condition (45), it is self-evident that for all $0 \leq \tau \leq 1$,

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

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

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

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

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

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

which immediately yields

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

A.3 Proof of Lemma 3

We start with proving (19a). For all $0 \leq t \leq T_0$, invoke Lemma 2 recursively with the conditions (47) to reach

$$\left\|\boldsymbol{x}^t - \boldsymbol{x}^\star\right\|_2 \leq (1-\eta/2)^t\left\|\boldsymbol{x}^0 - \boldsymbol{x}^\star\right\|_2 \leq C_1(1-\eta/2)^t\left\|\boldsymbol{x}^\star\right\|_2. \tag{101}$$

This finishes the proof of (19a) for $0 \leq t \leq T_0$ and also reveals that

$$\left\|\boldsymbol{x}^{T_0} - \boldsymbol{x}^\star\right\|_2 \leq C_1(1-\eta/2)^{T_0}\left\|\boldsymbol{x}^\star\right\|_2 \ll \frac{1}{n}\left\|\boldsymbol{x}^\star\right\|_2, \tag{102}$$

provided that $\eta \asymp 1/\log n$. Applying the Cauchy-Schwarz inequality and the fact (98) indicate that

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

leading to the satisfaction of (45). Therefore, invoking Lemma 2 yields

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

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

$$\left\|\boldsymbol{x}^t - \boldsymbol{x}^\star\right\|_2 \leq (1-\eta/2)^t\left\|\boldsymbol{x}^0 - \boldsymbol{x}^\star\right\|_2 \leq C_1(1-\eta/2)^t\left\|\boldsymbol{x}^\star\right\|_2 \ll \frac{1}{n}\|\boldsymbol{x}^\star\|_2. \tag{103}$$

We are left with (19b). It is self-evident that the iterates from $0 \leq t \leq T_0$ satisfy (19b) by assumptions. For $t > T_0$, we can use the Cauchy-Schhwarz inequality to obtain

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

where the penultimate relation uses the conditions (98) and (103).

A.4 Proof of Lemma 4

First, going through the same derivation as in (54) and (55) will result in

$$\max_{1\le l\le m}\left|\boldsymbol{a}_l^\top\left(\boldsymbol{x}^{t,(l)}-\boldsymbol{x}^\star\right)\right|\le C_4\sqrt{\log n} \tag{104}$$

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

We use the gradient update rules once again to decompose

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

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,

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

for any small constant $c>0$. Here (i) follows since (98) and, in view of (99) and (104),

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

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

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

2. For the term $\boldsymbol{\nu}_1^{(l)}$, the fundamental theorem of calculus [Lan93, 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 $\boldsymbol{x}(\tau)=\boldsymbol{x}^{t,(l)}+\tau(\boldsymbol{x}^t-\boldsymbol{x}^{t,(l)})$. By the induction hypotheses (51) and the condition (104), one can verify that

$$\left\|\boldsymbol{x}(\tau)-\boldsymbol{x}^\star\right\|_2\le\tau\left\|\boldsymbol{x}^t-\boldsymbol{x}^\star\right\|_2+(1-\tau)\left\|\boldsymbol{x}^{t,(l)}-\boldsymbol{x}^\star\right\|_2\le 2C_1\qquad\text{and} \tag{105}$$

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

for all $0\le\tau\le 1$, as long as $C_4\le C_2$. The second line follows directly from (104). To see why (105) holds, we note that

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

where the second inequality follows from the induction hypotheses (51b) and (51a). This combined with (51a) gives

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

as long as n is large enough, thus justifying (105). Hence by Lemma 1, $\nabla^2 f(\boldsymbol{x}(\tau))$ 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 (51b) to arrive at

$$\left\|\boldsymbol{x}^{t+1}-\boldsymbol{x}^{t+1,(l)}\right\|_2 \leq (1-\eta/2)\left\|\boldsymbol{x}^{t}-\boldsymbol{x}^{t,(l)}\right\|_2 + c\eta\sqrt{\frac{\log n}{n}} \leq C_3\sqrt{\frac{\log n}{n}}. \tag{106}$$

This establishes (53) for the $(t+1)$th iteration.

A.5 Proof of Lemma 5

In view of the assumption (42) that $\left\|\boldsymbol{x}^0-\boldsymbol{x}^\star\right\|_2 \leq \left\|\boldsymbol{x}^0+\boldsymbol{x}^\star\right\|_2$ 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

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

One can then invoke the Davis-Kahan sinΘ theorem [YWS15, Corollary 1] to obtain

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

Note that (56) — $\|\boldsymbol{Y}-\mathbb{E}[\boldsymbol{Y}]\| \leq \delta$ — is a direct consequence of Lemma 32. Additionally, the fact that $\mathbb{E}\left[\boldsymbol{Y}\right] = \boldsymbol{I}+2\boldsymbol{x}^\star\boldsymbol{x}^{\star\top}$ gives $\lambda_1\left(\mathbb{E}\left[\boldsymbol{Y}\right]\right)=3$, $\lambda_2\left(\mathbb{E}\left[\boldsymbol{Y}\right]\right)=1$, and $\lambda_1\left(\mathbb{E}\left[\boldsymbol{Y}\right]\right)-\lambda_2\left(\mathbb{E}\left[\boldsymbol{Y}\right]\right)=2$. Combining this spectral gap and the inequality $\|\boldsymbol{Y}-\mathbb{E}[\boldsymbol{Y}]\| \leq \delta$, we arrive at

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

To connect this bound with $\boldsymbol{x}^0$, we need to take into account the scaling factor $\sqrt{\lambda_1\left(\boldsymbol{Y}\right)/3}$. To this end, it follows from Weyl's inequality and (56) that

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

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

$$\left|\sqrt{\frac{\lambda_1\left(\boldsymbol{Y}\right)}{3}}-1\right| = \left|\frac{\frac{\lambda_1\left(\boldsymbol{Y}\right)}{3}-1}{\sqrt{\frac{\lambda_1\left(\boldsymbol{Y}\right)}{3}}+1}\right| \leq \left|\frac{\lambda_1\left(\boldsymbol{Y}\right)}{3}-1\right| \leq \frac{1}{3}\delta, \tag{107}$$

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

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

A.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 $\delta>0$,

$$\left\|\boldsymbol{Y}^{(l)}-\mathbb{E}\left[\boldsymbol{Y}^{(l)}\right]\right\| \leq \delta, \quad \|\boldsymbol{x}^{0,(l)}-\boldsymbol{x}^\star\|_2 \leq 2\delta, \quad \left\|\widetilde{\boldsymbol{x}}^{0,(l)}-\boldsymbol{x}^\star\right\|_2 \leq \sqrt{2}\delta, \qquad 1\leq l\leq m \tag{108}$$

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 (57) and (108) yields

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

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 $\sin\Theta$ theorem [DK70] gives

$$\left\|\widetilde{\boldsymbol{x}}^0-\widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2 \le \frac{\left\|\left(\boldsymbol{Y}-\boldsymbol{Y}^{(l)}\right)\widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2}{\lambda_1\left(\boldsymbol{Y}\right)-\lambda_2\left(\boldsymbol{Y}^{(l)}\right)} \le \left\|\left(\boldsymbol{Y}-\boldsymbol{Y}^{(l)}\right)\widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2. \tag{109}$$

Here, the second inequality uses Weyl's inequality:

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

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 (56) yields

$$\left|\lambda_1\left(\boldsymbol{Y}\right)-3\right| \le \|\boldsymbol{Y}-\mathbb{E}[\boldsymbol{Y}]\| \le \delta \qquad \Longrightarrow \qquad \lambda_1(\boldsymbol{Y})\in[3-\delta,3+\delta]\subseteq[2,4] \tag{110}$$

and, similarly, $\lambda_1(\boldsymbol{Y}^{(l)}), \|\boldsymbol{Y}\|, \|\boldsymbol{Y}^{(l)}\|\in[2,4]$. Invoke Lemma 34 to arrive at

$$\begin{aligned}\frac{1}{\sqrt{3}}\left\|\boldsymbol{x}^0-\boldsymbol{x}^{0,(l)}\right\|_2 &\le \frac{\left\|\left(\boldsymbol{Y}-\boldsymbol{Y}^{(l)}\right)\widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2}{2\sqrt{2}}+\left(2+\frac{4}{\sqrt{2}}\right)\left\|\widetilde{\boldsymbol{x}}^0-\widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2\\ &\le 6\left\|\left(\boldsymbol{Y}-\boldsymbol{Y}^{(l)}\right)\widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2,\end{aligned} \tag{111}$$

where the last inequality comes from (109).

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

$$\begin{aligned}\max_{1\le l\le m}\left\|\left(\boldsymbol{Y}-\boldsymbol{Y}^{(l)}\right)\widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2 &= \max_{1\le l\le m}\frac{1}{m}\left\|\left(\boldsymbol{a}_l^\top\boldsymbol{x}^\star\right)^2\boldsymbol{a}_l\boldsymbol{a}_l^\top\widetilde{\boldsymbol{x}}^{0,(l)}\right\|_2\\ &\le \max_{1\le l\le m}\frac{\left(\boldsymbol{a}_l^\top\boldsymbol{x}^\star\right)^2\left|\boldsymbol{a}_l^\top\widetilde{\boldsymbol{x}}^{0,(l)}\right|\left\|\boldsymbol{a}_l\right\|_2}{m}\\ &\overset{\text{(i)}}{\lesssim} \frac{\log n\cdot\sqrt{\log n}\cdot\sqrt{n}}{m}\\ &\asymp \sqrt{\frac{\log n}{n}}\cdot\frac{n\log n}{m}.\end{aligned} \tag{112}$$

The inequality (i) makes use of the fact $\max_l\left|\boldsymbol{a}_l^\top\boldsymbol{x}^\star\right| \le 5\sqrt{\log n}$ (cf. (99)), the bound $\max_l\|\boldsymbol{a}_l\|_2 \le 6\sqrt{n}$ (cf. (98)), 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 (112) into (111) leads us to conclude that

$$\max_{1\le l\le m}\left\|\boldsymbol{x}^0-\boldsymbol{x}^{0,(l)}\right\|_2 \le C_3\sqrt{\frac{\log n}{n}} \tag{113}$$

for any constant $C_3>0$.

B Proofs for matrix completion

Before proceeding to the proofs, let us record an immediate consequence of the incoherence property (25):

$$\|\boldsymbol{X}^\star\|_{2,\infty} \le \sqrt{\frac{\kappa\mu}{n}}\,\|\boldsymbol{X}^\star\|_{\mathrm{F}} \le \sqrt{\frac{\kappa\mu r}{n}}\,\|\boldsymbol{X}^\star\|. \tag{114}$$

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

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

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.

B.1 Proof of Lemma 7

By the expression of the Hessian in (61), one can decompose

$$\begin{aligned}&\mathrm{vec}\left(\boldsymbol{V}\right)^\top\nabla^2 f_{\mathrm{clean}}\left(\boldsymbol{X}\right)\mathrm{vec}\left(\boldsymbol{V}\right) = \frac{1}{2p}\left\|\mathcal{P}_\Omega\left(\boldsymbol{V}\boldsymbol{X}^\top+\boldsymbol{X}\boldsymbol{V}^\top\right)\right\|_{\mathrm{F}}^2 + \frac{1}{p}\left\langle\mathcal{P}_\Omega\left(\boldsymbol{X}\boldsymbol{X}^\top-\boldsymbol{M}^\star\right),\boldsymbol{V}\boldsymbol{V}^\top\right\rangle \\ &= \underbrace{\frac{1}{2p}\left\|\mathcal{P}_\Omega\left(\boldsymbol{V}\boldsymbol{X}^\top+\boldsymbol{X}\boldsymbol{V}^\top\right)\right\|_{\mathrm{F}}^2 - \frac{1}{2p}\left\|\mathcal{P}_\Omega\left(\boldsymbol{V}\boldsymbol{X}^{\star\top}+\boldsymbol{X}^\star\boldsymbol{V}^\top\right)\right\|_{\mathrm{F}}^2}_{:=\alpha_1} + \underbrace{\frac{1}{p}\left\langle\mathcal{P}_\Omega\left(\boldsymbol{X}\boldsymbol{X}^\top-\boldsymbol{M}^\star\right),\boldsymbol{V}\boldsymbol{V}^\top\right\rangle}_{:=\alpha_2} \\ &\quad + \underbrace{\frac{1}{2p}\left\|\mathcal{P}_\Omega\left(\boldsymbol{V}\boldsymbol{X}^{\star\top}+\boldsymbol{X}^\star\boldsymbol{V}^\top\right)\right\|_{\mathrm{F}}^2 - \frac{1}{2}\left\|\boldsymbol{V}\boldsymbol{X}^{\star\top}+\boldsymbol{X}^\star\boldsymbol{V}^\top\right\|_{\mathrm{F}}^2}_{:=\alpha_3} + \underbrace{\frac{1}{2}\left\|\boldsymbol{V}\boldsymbol{X}^{\star\top}+\boldsymbol{X}^\star\boldsymbol{V}^\top\right\|_{\mathrm{F}}^2}_{:=\alpha_4}.\end{aligned}$$

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

$$\alpha_4 \leq \left\|\boldsymbol{V}\boldsymbol{X}^{\star\top}\right\|_{\mathrm{F}}^2 + \left\|\boldsymbol{X}^\star\boldsymbol{V}^\top\right\|_{\mathrm{F}}^2 \leq 2\|\boldsymbol{X}^\star\|^2\|\boldsymbol{V}\|_{\mathrm{F}}^2 = 2\sigma_{\max}\|\boldsymbol{V}\|_{\mathrm{F}}^2.$$

When it comes to the lower bound, one discovers that

$$\begin{aligned}\alpha_4 &= \frac{1}{2}\left\{\left\|\boldsymbol{V}\boldsymbol{X}^{\star\top}\right\|_{\mathrm{F}}^2 + \left\|\boldsymbol{X}^\star\boldsymbol{V}^\top\right\|_{\mathrm{F}}^2 + 2\mathrm{Tr}\left(\boldsymbol{X}^{\star\top}\boldsymbol{V}\boldsymbol{X}^{\star\top}\boldsymbol{V}\right)\right\} \\ &\geq \sigma_{\min}\|\boldsymbol{V}\|_{\mathrm{F}}^2 + \mathrm{Tr}\left[\left(\boldsymbol{Z}+\boldsymbol{X}^\star-\boldsymbol{Z}\right)^\top\boldsymbol{V}\left(\boldsymbol{Z}+\boldsymbol{X}^\star-\boldsymbol{Z}\right)^\top\boldsymbol{V}\right] \\ &\geq \sigma_{\min}\|\boldsymbol{V}\|_{\mathrm{F}}^2 + \mathrm{Tr}\left(\boldsymbol{Z}^\top\boldsymbol{V}\boldsymbol{Z}^\top\boldsymbol{V}\right) - 2\|\boldsymbol{Z}-\boldsymbol{X}^\star\|\|\boldsymbol{Z}\|\|\boldsymbol{V}\|_{\mathrm{F}}^2 - \|\boldsymbol{Z}-\boldsymbol{X}^\star\|^2\|\boldsymbol{V}\|_{\mathrm{F}}^2 \\ &\geq \left(\sigma_{\min}-5\delta\sigma_{\max}\right)\|\boldsymbol{V}\|_{\mathrm{F}}^2 + \mathrm{Tr}\left(\boldsymbol{Z}^\top\boldsymbol{V}\boldsymbol{Z}^\top\boldsymbol{V}\right),\end{aligned}\tag{115}$$

where the last line comes from the assumptions that

$$\|\boldsymbol{Z}-\boldsymbol{X}^\star\| \leq \delta\|\boldsymbol{X}^\star\| \leq \|\boldsymbol{X}^\star\| \qquad \text{and} \qquad \|\boldsymbol{Z}\| \leq \|\boldsymbol{Z}-\boldsymbol{X}^\star\| + \|\boldsymbol{X}^\star\| \leq 2\|\boldsymbol{X}^\star\|.$$

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 35 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

$$\mathrm{Tr}\left[\boldsymbol{Z}^\top\left(\boldsymbol{Y}\boldsymbol{H}_Y-\boldsymbol{Z}\right)\boldsymbol{Z}^\top\left(\boldsymbol{Y}\boldsymbol{H}_Y-\boldsymbol{Z}\right)\right] \geq 0.$$

Substitution into (115) gives

$$\alpha_4 \geq \left(\sigma_{\min}-5\delta\sigma_{\max}\right)\|\boldsymbol{V}\|_{\mathrm{F}}^2 \geq \frac{9}{10}\sigma_{\min}\|\boldsymbol{V}\|_{\mathrm{F}}^2,$$

provided that $\kappa\delta \leq 1/50$.

2. For α_1, we consider the following quantity

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

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

$$\begin{aligned}\alpha_1 &= \underbrace{\frac{1}{p}\left[\left\langle\mathcal{P}_\Omega(\boldsymbol{V}\boldsymbol{X}^\top),\mathcal{P}_\Omega\left(\boldsymbol{V}\boldsymbol{X}^\top\right)\right\rangle-\left\langle\mathcal{P}_\Omega\left(\boldsymbol{V}\boldsymbol{X}^{\star\top}\right),\mathcal{P}_\Omega\left(\boldsymbol{V}\boldsymbol{X}^{\star\top}\right)\right\rangle\right]}_{:=\beta_1} \\ &\quad+\underbrace{\frac{1}{p}\left[\left\langle\mathcal{P}_\Omega(\boldsymbol{V}\boldsymbol{X}^\top),\mathcal{P}_\Omega\left(\boldsymbol{X}\boldsymbol{V}^\top\right)\right\rangle-\left\langle\mathcal{P}_\Omega\left(\boldsymbol{V}\boldsymbol{X}^{\star\top}\right),\mathcal{P}_\Omega\left(\boldsymbol{X}^\star\boldsymbol{V}^\top\right)\right\rangle\right]}_{:=\beta_2}.\end{aligned}$$

For β_2, one has

$$\begin{aligned}\beta_2 &= \frac{1}{p}\left\langle\mathcal{P}_\Omega\left(\boldsymbol{V}\left(\boldsymbol{X}-\boldsymbol{X}^\star\right)^\top\right),\mathcal{P}_\Omega\left(\left(\boldsymbol{X}-\boldsymbol{X}^\star\right)\boldsymbol{V}^\top\right)\right\rangle \\ &\quad+\frac{1}{p}\left\langle\mathcal{P}_\Omega\left(\boldsymbol{V}\left(\boldsymbol{X}-\boldsymbol{X}^\star\right)^\top\right),\mathcal{P}_\Omega\left(\boldsymbol{X}^\star\boldsymbol{V}^\top\right)\right\rangle+\frac{1}{p}\left\langle\mathcal{P}_\Omega\left(\boldsymbol{V}\boldsymbol{X}^{\star\top}\right),\mathcal{P}_\Omega\left(\left(\boldsymbol{X}-\boldsymbol{X}^\star\right)\boldsymbol{V}^\top\right)\right\rangle\end{aligned}$$

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

$$|\beta_2|\leq\frac{1}{p}\left\|\mathcal{P}_\Omega\left(\boldsymbol{V}\left(\boldsymbol{X}-\boldsymbol{X}^\star\right)^\top\right)\right\|_{\mathrm{F}}^2+\frac{2}{p}\left\|\mathcal{P}_\Omega\left(\boldsymbol{V}\left(\boldsymbol{X}-\boldsymbol{X}^\star\right)^\top\right)\right\|_{\mathrm{F}}\left\|\mathcal{P}_\Omega\left(\boldsymbol{X}^\star\boldsymbol{V}^\top\right)\right\|_{\mathrm{F}}. \tag{116}$$

This then calls for upper bounds on the following two terms

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

The injectivity of $\mathcal{P}_\Omega$ (cf. [CR09, Section 4.2] or Lemma 38)—when restricted to the tangent space of $\boldsymbol{M}^\star$—gives: for any fixed constant $\gamma>0$,

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

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

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

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 $\gamma>0$,

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

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

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

where (i) utilizes the incoherence condition (114) and (ii) holds with the proviso that $\epsilon\sqrt{\kappa^3\mu r}\ll 1$.

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

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

In view of Lemma 43, with probability at least $1-O\left(n^{-10}\right)$,

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

as soon as $\epsilon\sqrt{\kappa^3\mu r}\log n\ll 1$, where we utilize the incoherence condition (114). 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 $\boldsymbol{M}^\star$. Specifically, it follows from the bounds in [CR09, Section 4.2] or Lemma 38 that

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

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

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.

B.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. (26)), we have

$$\begin{aligned}
\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1}-\boldsymbol{X}^{\star}\right\|_{\mathrm{F}} &\leq \left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\star}\right\|_{\mathrm{F}} = \left\|\left[\boldsymbol{X}^{t}-\eta\nabla f\left(\boldsymbol{X}^{t}\right)\right]\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\star}\right\|_{\mathrm{F}} \\
&\overset{\text{(i)}}{=} \left\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\eta\nabla f\big(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}\big)-\boldsymbol{X}^{\star}\right\|_{\mathrm{F}} \\
&\overset{\text{(ii)}}{=} \left\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\eta\left[\nabla f_{\mathrm{clean}}\big(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}\big)-\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{E}\right)\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}\right]-\boldsymbol{X}^{\star}\right\|_{\mathrm{F}} \\
&\leq \underbrace{\left\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\eta\nabla f_{\mathrm{clean}}\big(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}\big)-\big(\boldsymbol{X}^{\star}-\eta\nabla f_{\mathrm{clean}}(\boldsymbol{X}^{\star})\big)\right\|_{\mathrm{F}}}_{:=\alpha_1} + \underbrace{\eta\left\|\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{E}\right)\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}\right\|_{\mathrm{F}}}_{:=\alpha_2}, \qquad (117)
\end{aligned}$$

where (i) follows from the identity $\nabla f(\boldsymbol{X}^{t}\boldsymbol{R}) = \nabla f\left(\boldsymbol{X}^{t}\right)\boldsymbol{R}$ for any orthonormal matrix $\boldsymbol{R}\in\mathcal{O}^{r\times r}$, (ii) arises from the definitions of $\nabla f\left(\boldsymbol{X}\right)$ and $\nabla f_{\mathrm{clean}}\left(\boldsymbol{X}\right)$ (see (59) and (60), respectively), and the last inequality (117) utilizes the triangle inequality and the fact that $\nabla f_{\mathrm{clean}}(\boldsymbol{X}^{\star}) = \boldsymbol{0}$. It thus suffices to control α_1 and α_2.

1. For the second term α_2 in (117), it is easy to see that with probability at least $1-O\left(n^{-10}\right)$,

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

for some absolute constant $C>0$. Here, the second inequality holds because $\left\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}\right\|_{\mathrm{F}} \leq \left\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\star}\right\|_{\mathrm{F}}+\left\|\boldsymbol{X}^{\star}\right\|_{\mathrm{F}} \leq 2\left\|\boldsymbol{X}^{\star}\right\|_{\mathrm{F}}$, following the hypothesis (28a) together with our assumptions on the noise and the sample complexity. The last inequality makes use of Lemma 40.

2. For the first term α_1 in (117), the fundamental theorem of calculus [Lan93, Chapter XIII, Theorem 4.2] reveals

$$\begin{aligned}
&\mathrm{vec}\left[\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\eta\nabla f_{\mathrm{clean}}\big(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}\big)-\big(\boldsymbol{X}^{\star}-\eta\nabla f_{\mathrm{clean}}(\boldsymbol{X}^{\star})\big)\right] \\
&\quad= \mathrm{vec}\left[\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\star}\right]-\eta\cdot\mathrm{vec}\left[\nabla f_{\mathrm{clean}}\big(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}\big)-\nabla f_{\mathrm{clean}}\left(\boldsymbol{X}^{\star}\right)\right] \\
&\quad= \Bigg(\boldsymbol{I}_{nr}-\eta\underbrace{\int_0^1\nabla^2 f_{\mathrm{clean}}\left(\boldsymbol{X}(\tau)\right)\mathrm{d}\tau}_{:=\boldsymbol{A}}\Bigg)\mathrm{vec}\left(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\star}\right), \qquad (118)
\end{aligned}$$

where we denote $\boldsymbol{X}(\tau) := \boldsymbol{X}^{\star}+\tau(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\star})$. Taking the squared Euclidean norm of both sides of the equality (118) leads to

$$\begin{aligned}
\left(\alpha_1\right)^2 &= \mathrm{vec}\big(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\star}\big)^{\top}\left(\boldsymbol{I}_{nr}-\eta\boldsymbol{A}\right)^2\mathrm{vec}\big(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\star}\big) \\
&= \mathrm{vec}\big(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\star}\big)^{\top}\left(\boldsymbol{I}_{nr}-2\eta\boldsymbol{A}+\eta^2\boldsymbol{A}^2\right)\mathrm{vec}\big(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\star}\big) \\
&\leq \left\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\star}\right\|_{\mathrm{F}}^2+\eta^2\left\|\boldsymbol{A}\right\|^2\left\|\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\star}\right\|_{\mathrm{F}}^2-2\eta\,\mathrm{vec}\big(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\star}\big)^{\top}\boldsymbol{A}\,\mathrm{vec}\big(\boldsymbol{X}^{t}\widehat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\star}\big), \qquad (119)
\end{aligned}$$

where in (119) we have used the fact that

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

Based on the condition (28b), it is easily seen that $\forall\tau\in[0,1]$,

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

Taking $\boldsymbol{X} = \boldsymbol{X}(\tau), \boldsymbol{Y} = \boldsymbol{X}^t$ and $\boldsymbol{Z} = \boldsymbol{X}^\star$ in Lemma 7, one can easily verify the assumptions therein given our sample size condition $n^2p \gg \kappa^3\mu^3r^3n\log^3 n$ and the noise condition (27). As a result,

$$\mathrm{vec}\big(\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\star\big)^\top \boldsymbol{A}\ \mathrm{vec}\big(\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\star\big) \ge \frac{\sigma_{\min}}{2}\big\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\star\big\|_{\mathrm{F}}^2 \qquad \text{and} \qquad \|\boldsymbol{A}\| \le \frac{5}{2}\sigma_{\max}.$$

Substituting these two inequalities into (119) yields

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

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

$$\alpha_1 \le \left(1 - \frac{\sigma_{\min}}{4}\eta\right)\big\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\star\big\|_{\mathrm{F}}.$$

3. Combining the preceding bounds on both α_1 and α_2 and making use of the hypothesis (28a), we have

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

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.

B.3 Proof of Lemma 9

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

$$\widetilde{\boldsymbol{X}}^{t+1} := \boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \eta\frac{1}{p}\mathcal{P}_\Omega\left[\boldsymbol{X}^t\boldsymbol{X}^{t\top} - (\boldsymbol{M}^\star + \boldsymbol{E})\right]\boldsymbol{X}^\star. \tag{120}$$

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

$$\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1} - \boldsymbol{X}^\star\right\| \le \underbrace{\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1} - \widetilde{\boldsymbol{X}}^{t+1}\right\|}_{:=\alpha_1} + \underbrace{\left\|\widetilde{\boldsymbol{X}}^{t+1} - \boldsymbol{X}^\star\right\|}_{:=\alpha_2}. \tag{121}$$

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

$$\begin{aligned}
\alpha_2 &= \left\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \eta\frac{1}{p}\mathcal{P}_\Omega\left[\boldsymbol{X}^t\boldsymbol{X}^{t\top} - (\boldsymbol{M}^\star + \boldsymbol{E})\right]\boldsymbol{X}^\star - \boldsymbol{X}^\star\right\| \\
&\le \eta\left\|\frac{1}{p}\mathcal{P}_\Omega(\boldsymbol{E})\right\|\|\boldsymbol{X}^\star\| + \left\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \eta\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{X}^t\boldsymbol{X}^{t\top} - \boldsymbol{X}^\star\boldsymbol{X}^{\star\top}\right)\boldsymbol{X}^\star - \boldsymbol{X}^\star\right\| \qquad (122) \\
&\le \eta\left\|\frac{1}{p}\mathcal{P}_\Omega(\boldsymbol{E})\right\|\|\boldsymbol{X}^\star\| + \underbrace{\left\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \eta\left(\boldsymbol{X}^t\boldsymbol{X}^{t\top} - \boldsymbol{X}^\star\boldsymbol{X}^{\star\top}\right)\boldsymbol{X}^\star - \boldsymbol{X}^\star\right\|}_{:=\beta_1} \\
&\quad + \underbrace{\eta\left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{X}^t\boldsymbol{X}^{t\top} - \boldsymbol{X}^\star\boldsymbol{X}^{\star\top}\right)\boldsymbol{X}^\star - \left(\boldsymbol{X}^t\boldsymbol{X}^{t\top} - \boldsymbol{X}^\star\boldsymbol{X}^{\star\top}\right)\boldsymbol{X}^\star\right\|}_{:=\beta_2}, \qquad (123)
\end{aligned}$$

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

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

which, by Lemma 35, satisfies the following symmetry property

$$\widehat{\boldsymbol{H}}^{t\top}\boldsymbol{X}^{t\top}\boldsymbol{X}^\star = \boldsymbol{X}^{\star\top}\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t \qquad \Longrightarrow \qquad \boldsymbol{\Delta}^{t\top}\boldsymbol{X}^\star = \boldsymbol{X}^{\star\top}\boldsymbol{\Delta}^t. \tag{124}$$

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

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

The leading term γ_1 can be upper bounded by

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

where the second identity follows from the symmetry property (124). By choosing $\eta \le 1/(2\sigma_{\max})$, one has $\boldsymbol{0} \preceq \boldsymbol{I}_r - 2\eta\boldsymbol{\Sigma}^\star \preceq (1-2\eta\sigma_{\min})\,\boldsymbol{I}_r$ and $\boldsymbol{0} \preceq \boldsymbol{I}_n - 2\eta\boldsymbol{M}^\star \preceq \boldsymbol{I}_n$, and further one can ensure

$$\gamma_1 \le \frac{1}{2}\left[(1-2\eta\sigma_{\min}) + 1\right]\left\|\boldsymbol{\Delta}^t\right\| = (1-\eta\sigma_{\min})\left\|\boldsymbol{\Delta}^t\right\|. \tag{125}$$

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

$$\gamma_2 \le \eta\left\|\boldsymbol{\Delta}^t\right\|^2\left\|\boldsymbol{X}^\star\right\|. \tag{126}$$

The bounds (125) and (126) taken collectively give

$$\beta_1 \le (1-\eta\sigma_{\min})\left\|\boldsymbol{\Delta}^t\right\| + \eta\left\|\boldsymbol{\Delta}^t\right\|^2\left\|\boldsymbol{X}^\star\right\|. \tag{127}$$

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

$$\begin{aligned}\frac{1}{\eta}\beta_2 &= \left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{\Delta}^t\boldsymbol{\Delta}^{t\top} + \boldsymbol{\Delta}^t\boldsymbol{X}^{\star\top} + \boldsymbol{X}^\star\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^\star - \left[\boldsymbol{\Delta}^t\boldsymbol{\Delta}^{t\top} + \boldsymbol{\Delta}^t\boldsymbol{X}^{\star\top} + \boldsymbol{X}^\star\boldsymbol{\Delta}^{t\top}\right]\boldsymbol{X}^\star\right\| \\ &\le \underbrace{\left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{\Delta}^t\boldsymbol{X}^{\star\top}\right)\boldsymbol{X}^\star - \left(\boldsymbol{\Delta}^t\boldsymbol{X}^{\star\top}\right)\boldsymbol{X}^\star\right\|_{\mathrm{F}}}_{:=\theta_1} + \underbrace{\left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{X}^\star\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^\star - \left(\boldsymbol{X}^\star\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^\star\right\|_{\mathrm{F}}}_{:=\theta_2} \\ &\quad + \underbrace{\left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{\Delta}^t\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^\star - \left(\boldsymbol{\Delta}^t\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^\star\right\|_{\mathrm{F}}}_{:=\theta_3},\end{aligned} \tag{128}$$

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

- For the first term θ_1 in (128), the lth row of $\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{\Delta}^t\boldsymbol{X}^{\star\top}\right)\boldsymbol{X}^\star - \left(\boldsymbol{\Delta}^t\boldsymbol{X}^{\star\top}\right)\boldsymbol{X}^\star$ is given by

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

where, as usual, $\delta_{l,j} = \mathbb{1}_{\{(l,j)\in\Omega\}}$. Lemma 41 together with the union bound reveals that

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

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

$$\left\| \boldsymbol{\Delta}_{l,\cdot}^{t} \left[\frac{1}{p} \sum_{j=1}^{n} (\delta_{l,j} - p) \boldsymbol{X}_{j,\cdot}^{\star\top} \boldsymbol{X}_{j,\cdot}^{\star} \right] \right\|_{2} \leq \left\| \boldsymbol{\Delta}_{l,\cdot}^{t} \right\|_{2} \left\| \frac{1}{p} \sum_{j} (\delta_{l,j} - p) \boldsymbol{X}_{j,\cdot}^{\star\top} \boldsymbol{X}_{j,\cdot}^{\star} \right\|$$
$$\lesssim \left\| \boldsymbol{\Delta}_{l,\cdot}^{t} \right\|_{2} \left\{ \sqrt{\frac{\left\| \boldsymbol{X}^{\star} \right\|_{2,\infty}^{2} \sigma_{\max} \log n}{p}} + \frac{\left\| \boldsymbol{X}^{\star} \right\|_{2,\infty}^{2} \log n}{p} \right\},$$

which further reveals that

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

for arbitrarily small $\gamma > 0$. Here, (i) follows from $\left\| \boldsymbol{\Delta}^{t} \right\|_{\mathrm{F}} \leq \sqrt{r} \left\| \boldsymbol{\Delta}^{t} \right\|$, (ii) holds owing to the incoherence condition (114), and (iii) follows as long as $n^{2}p \gg \kappa^{3} \mu r^{2} n \log n$.

- For the second term θ_2 in (128), denote

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

whose lth row is given by

$$\boldsymbol{A}_{l,\cdot} = \boldsymbol{X}_{l,\cdot}^{\star} \sum_{j=1}^{n} (\delta_{l,j} - p) \boldsymbol{\Delta}_{j,\cdot}^{t\top} \boldsymbol{X}_{j,\cdot}^{\star}. \tag{129}$$

Recalling the induction hypotheses (28b) and (28c), we define

$$\left\| \boldsymbol{\Delta}^{t} \right\|_{2,\infty} \leq C_{5} \rho^{t} \mu r \sqrt{\frac{\log n}{np}} \left\| \boldsymbol{X}^{\star} \right\|_{2,\infty} + C_{8} \frac{\sigma}{\sigma_{\min}} \sqrt{\frac{n \log n}{p}} \left\| \boldsymbol{X}^{\star} \right\|_{2,\infty} := \xi \tag{130}$$

$$\left\| \boldsymbol{\Delta}^{t} \right\| \leq C_{9} \rho^{t} \mu r \frac{1}{\sqrt{np}} \left\| \boldsymbol{X}^{\star} \right\| + C_{10} \frac{\sigma}{\sigma_{\min}} \sqrt{\frac{n}{p}} \left\| \boldsymbol{X}^{\star} \right\| := \psi. \tag{131}$$

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

$$\omega := 2p\xi\sigma_{\max} \tag{132}$$

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

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

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 (28b) and (28c).

Lemma 22. *Under the conditions in Lemma 9, there exist some constants $c, C > 0$ such that with probability exceeding $1 - c\exp(-Cnr\log n)$,*

$$\phi_1 \lesssim \xi\sqrt{p\sigma_{\max}\|\boldsymbol{X}^\star\|_{2,\infty}^2 nr\log^2 n} \tag{133}$$

holds simultaneously for all $\boldsymbol{\Delta}^t$ obeying (130) and (131). Here, ξ is defined in (130).

Lemma 23. *Under the conditions in Lemma 9, with probability at least $1 - O\left(n^{-10}\right)$,*

$$\phi_2 \lesssim \xi\sqrt{\kappa\mu r^2 p\log^2 n}\,\|\boldsymbol{X}^\star\|^2 \tag{134}$$

holds simultaneously for all $\boldsymbol{\Delta}^t$ obeying (130) and (131). Here, ξ is defined in (130).

The bounds (133) and (134) together with the incoherence condition (114) yield

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

- Next, we assert that the third term θ_3 in (128) 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

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

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

(c) Substituting the preceding bounds on β_1 and β_2 into (123), we reach

$$\begin{aligned}
\alpha_2 &\overset{\text{(i)}}{\leq} \left(1 - \eta\sigma_{\min} + \eta\gamma\sigma_{\min} + \eta\left\|\boldsymbol{\Delta}^t\right\|\|\boldsymbol{X}^\star\|\right)\left\|\boldsymbol{\Delta}^t\right\| + \eta\left\|\frac{1}{p}\mathcal{P}_\Omega(\boldsymbol{E})\right\|\|\boldsymbol{X}^\star\| \\
&\quad + \widetilde{C}\eta\sqrt{\frac{\kappa\mu r^2\log^2 n}{p}}\sigma_{\max}\left(C_5\rho^t\mu r\sqrt{\frac{\log n}{np}}\|\boldsymbol{X}^\star\|_{2,\infty} + C_8\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\|\boldsymbol{X}^\star\|_{2,\infty}\right) \\
&\overset{\text{(ii)}}{\leq} \left(1 - \frac{\sigma_{\min}}{2}\eta\right)\left\|\boldsymbol{\Delta}^t\right\| + \eta\left\|\frac{1}{p}\mathcal{P}_\Omega(\boldsymbol{E})\right\|\|\boldsymbol{X}^\star\| \\
&\quad + \widetilde{C}\eta\sqrt{\frac{\kappa\mu r^2\log^2 n}{p}}\sigma_{\max}\left(C_5\rho^t\mu r\sqrt{\frac{\log n}{np}}\|\boldsymbol{X}^\star\|_{2,\infty} + C_8\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\|\boldsymbol{X}^\star\|_{2,\infty}\right) \\
&\overset{\text{(iii)}}{\leq} \left(1 - \frac{\sigma_{\min}}{2}\eta\right)\left\|\boldsymbol{\Delta}^t\right\| + C\eta\sigma\sqrt{\frac{n}{p}}\|\boldsymbol{X}^\star\| \\
&\quad + \widetilde{C}\eta\sqrt{\frac{\kappa^2\mu^2 r^3\log^3 n}{np}}\sigma_{\max}\left(C_5\rho^t\mu r\sqrt{\frac{1}{np}} + C_8\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right)\|\boldsymbol{X}^\star\|
\end{aligned} \tag{135}$$

for some constant $C > 0$. Here, (i) uses the definition of ξ (cf. (130)), (ii) holds if γ is small enough and $\|\boldsymbol{\Delta}^t\|\,\|\boldsymbol{X}^\star\| \ll \sigma_{\min}$, and (iii) follows from Lemma 40 as well as the incoherence condition (114). An immediate consequence of (135) is that under the sample size condition and the noise condition of this lemma, one has

$$\left\|\widetilde{\boldsymbol{X}}^{t+1}-\boldsymbol{X}^{\star}\right\|\left\|\boldsymbol{X}^{\star}\right\| \leq \sigma_{\min}/2 \tag{136}$$

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

2. We then move on to the first term α_1 in (121), 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

$$\boldsymbol{R}_1=\left(\widehat{\boldsymbol{H}}^t\right)^{-1}\widehat{\boldsymbol{H}}^{t+1}:=\arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^t\boldsymbol{R}-\boldsymbol{X}^{\star}\right\|_{\mathrm{F}}. \tag{137}$$

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

$$\boldsymbol{I}_r=\arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\left\|\widetilde{\boldsymbol{X}}^{t+1}\boldsymbol{R}-\boldsymbol{X}^{\star}\right\|_{\mathrm{F}}, \tag{138}$$

meaning that $\widetilde{\boldsymbol{X}}^{t+1}$ is already rotated to the direction that is most "aligned" with $\boldsymbol{X}^{\star}$. This important property eases the analysis. In fact, in view of Lemma 35, (138) follows if one can show that $\boldsymbol{X}^{\star\top}\widetilde{\boldsymbol{X}}^{t+1}$ is symmetric and positive semidefinite. First of all, it follows from Lemma 35 that $\boldsymbol{X}^{\star\top}\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t$ is symmetric and, hence, by definition,

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

is also symmetric. Additionally,

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

where the second inequality holds according to (136). Weyl's inequality guarantees that

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

thus justifying (138) via Lemma 35.

(b) With (137) and (138) in place, we resort to Lemma 37 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 (136) that

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

Moreover, we have

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

in which

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

This allows one to derive

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

$$\leq \eta \left(2n \left\|\boldsymbol{\Delta}^t\right\|_{2,\infty}^2 + 4\sqrt{n}\log n \left\|\boldsymbol{\Delta}^t\right\|_{2,\infty} \|\boldsymbol{X}^\star\| + C\sigma\sqrt{\frac{n}{p}}\right) \left\|\boldsymbol{\Delta}^t\right\| \tag{139}$$

for some absolute constant $C > 0$. Here the last inequality follows from Lemma 40 and Lemma 43. As a consequence,

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

Under our sample size condition and the noise condition (27) and the induction hypotheses (28), one can show

$$\|\boldsymbol{X}_1 - \boldsymbol{X}_2\| \|\boldsymbol{X}^\star\| \leq \sigma_{\min}/4.$$

Apply Lemma 37 and (139) to reach

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

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

$$\begin{aligned}
\left\|\boldsymbol{X}^{t+1}\widehat{\boldsymbol{H}}^{t+1} - \boldsymbol{X}^\star\right\| &\leq \left(1 - \frac{\sigma_{\min}}{2}\eta\right) \left\|\boldsymbol{\Delta}^t\right\| + \eta C\sigma\sqrt{\frac{n}{p}} \|\boldsymbol{X}^\star\| \\
&\quad + \widetilde{C}\eta\sqrt{\frac{\kappa^2\mu^2 r^3 \log^3 n}{np}}\sigma_{\max}\left(C_5\rho^t\mu r\sqrt{\frac{1}{np}} + \frac{C_8}{\sigma_{\min}}\sigma\sqrt{\frac{n}{p}}\right)\|\boldsymbol{X}^\star\| \\
&\quad + 5\eta\kappa\left(2n \left\|\boldsymbol{\Delta}^t\right\|_{2,\infty}^2 + 2\sqrt{n}\log n \left\|\boldsymbol{\Delta}^t\right\|_{2,\infty} \|\boldsymbol{X}^\star\| + C\sigma\sqrt{\frac{n}{p}}\right) \left\|\boldsymbol{\Delta}^t\right\| \\
&\leq C_9\rho^{t+1}\mu r\frac{1}{\sqrt{np}}\|\boldsymbol{X}^\star\| + C_{10}\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\|\boldsymbol{X}^\star\|,
\end{aligned}$$

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

B.3.1 Proof of Lemma 22

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

$$B := \max_{1\leq j\leq n} \left\|\boldsymbol{X}_{l,\cdot}^\star \left(\delta_{l,j} - p\right) \boldsymbol{\Delta}_{j,\cdot}^{t\top} \boldsymbol{X}_{j,\cdot}^\star\right\|_2 \leq \|\boldsymbol{X}^\star\|_{2,\infty}^2 \xi$$

$$\begin{aligned}
\text{and} \qquad V &:= \left\|\mathbb{E}\left[\sum_{j=1}^n \left(\delta_{l,j} - p\right)^2 \boldsymbol{X}_{l,\cdot}^\star \boldsymbol{\Delta}_{j,\cdot}^{t\top} \boldsymbol{X}_{j,\cdot}^\star \left(\boldsymbol{X}_{l,\cdot}^\star \boldsymbol{\Delta}_{j,\cdot}^{t\top} \boldsymbol{X}_{j,\cdot}^\star\right)^\top\right]\right\| \\
&\leq p \left\|\boldsymbol{X}_{l,\cdot}^\star\right\|_2^2 \|\boldsymbol{X}^\star\|_{2,\infty}^2 \left\|\sum_{j=1}^n \boldsymbol{\Delta}_{j,\cdot}^{t\top}\boldsymbol{\Delta}_{j,\cdot}^t\right\| \\
&\leq p \left\|\boldsymbol{X}_{l,\cdot}^\star\right\|_2^2 \|\boldsymbol{X}^\star\|_{2,\infty}^2 \psi^2 \\
&\leq 2p \|\boldsymbol{X}^\star\|_{2,\infty}^2 \xi^2 \sigma_{\max},
\end{aligned}$$

where ξ and ψ are defined respectively in (130) and (131). Here, the last line makes use of the fact that

$$\|\boldsymbol{X}^\star\|_{2,\infty} \psi \ll \xi \|\boldsymbol{X}^\star\| = \xi\sqrt{\sigma_{\max}}, \tag{140}$$

as long as n is sufficiently large. Apply the matrix Bernstein inequality [Tro15b, Theorem 6.1.1] to get

$$\mathbb{P}\left\{\|\boldsymbol{A}_{l,\cdot}\|_2 \geq t\right\} \leq 2r \exp\left(-\frac{ct^2}{2p\xi^2\sigma_{\max}\|\boldsymbol{X}^\star\|_{2,\infty}^2 + t\cdot\|\boldsymbol{X}^\star\|_{2,\infty}^2\xi}\right)$$
$$\leq 2r\exp\left(-\frac{ct^2}{4p\xi^2\sigma_{\max}\|\boldsymbol{X}^\star\|_{2,\infty}^2}\right)$$

for some constant $c > 0$, provided that

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

This upper bound on t is exactly the truncation level ω we introduce in (132). With this in mind, we can easily verify that

$$\|\boldsymbol{A}_{l,\cdot}\|_2 \mathbb{1}_{\left\{\|\boldsymbol{A}_{l,\cdot}\|_2\leq\omega\right\}}$$

is a sub-Gaussian random variable with variance proxy not exceeding $O\left(p\xi^2\sigma_{\max}\|\boldsymbol{X}^\star\|_{2,\infty}^2\log r\right)$. Therefore, invoking the concentration bounds for quadratic functions [HKZ12, Theorem 2.1] yields that for some constants $C_0, C > 0$, with probability at least $1 - C_0e^{-Cnr\log n}$,

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

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 [Tao12, Section 2.3.1] or the proof of Lemma 42. In conclusion, we have that with probability exceeding $1 - C_0e^{-\frac{1}{2}Cnr\log n}$,

$$\phi_1 = \sqrt{\sum_{l=1}^{n}\|\boldsymbol{A}_{l,\cdot}\|_2^2\mathbb{1}_{\left\{\|\boldsymbol{A}_{l,\cdot}\|_2\leq\omega\right\}}} \lesssim \sqrt{p\xi^2\sigma_{\max}\|\boldsymbol{X}^\star\|_{2,\infty}^2 nr\log^2 n} \tag{141}$$

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 (129) that the diagonal element $\delta_{l,l}$ cannot change the ℓ_2 norm of $\boldsymbol{A}_{l,\cdot}$ by more than $\|\boldsymbol{X}^\star\|_{2,\infty}^2\xi$. As a result, changing all the diagonals $\{\delta_{l,l}\}$ cannot change the quantity of interest (i.e. ϕ_1) by more than $\sqrt{n}\|\boldsymbol{X}^\star\|_{2,\infty}^2\xi$. This is smaller than the right hand side of (141) 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 [GLM16, Theorem D.2]. More specifically, we can employ the construction of Bernoulli random variables introduced therein to demonstrate that the upper bound in (141) 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 [GLM16, Claim D.3] to finish the proof. We omit the details here for brevity.

B.3.2 Proof of Lemma 23

Observe from (129) that

$$\|\boldsymbol{A}_{l,\cdot}\|_2 \leq \|\boldsymbol{X}^\star\|_{2,\infty}\left\|\sum_{j=1}^{n}(\delta_{l,j}-p)\,\boldsymbol{\Delta}_{j,\cdot}^{t\top}\boldsymbol{X}_{j,\cdot}^\star\right\| \tag{142}$$
$$\leq \|\boldsymbol{X}^\star\|_{2,\infty}\left(\left\|\sum_{j=1}^{n}\delta_{l,j}\boldsymbol{\Delta}_{j,\cdot}^{t\top}\boldsymbol{X}_{j,\cdot}^\star\right\| + p\left\|\boldsymbol{\Delta}^t\right\|\|\boldsymbol{X}^\star\|\right)$$

$$
\begin{aligned}
&\le \|\boldsymbol{X}^\star\|_{2,\infty}\left(\left\|\left[\delta_{l,1}\boldsymbol{\Delta}_{1,\cdot}^{t\top},\cdots,\delta_{l,n}\boldsymbol{\Delta}_{n,\cdot}^{t\top}\right]\right\|\left\|\begin{bmatrix}\delta_{l,1}\boldsymbol{X}_{1,\cdot}^\star\\ \vdots\\ \delta_{l,n}\boldsymbol{X}_{n,\cdot}^\star\end{bmatrix}\right\|+p\psi\left\|\boldsymbol{X}^\star\right\|\right)\\
&\le \|\boldsymbol{X}^\star\|_{2,\infty}\left(\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}^t\right)\right\|\cdot 1.2\sqrt{p}\left\|\boldsymbol{X}^\star\right\|+p\psi\left\|\boldsymbol{X}^\star\right\|\right),
\end{aligned}
\tag{143}
$$

where ψ is as defined in (131) and $\boldsymbol{G}_l(\cdot)$ is as defined in Lemma 41. Here, the last inequality follows from Lemma 41, namely, for some constant $C>0$, the following holds with probability at least $1-O(n^{-10})$

$$
\begin{aligned}
\left\|\begin{bmatrix}\delta_{l,1}\boldsymbol{X}_{1,\cdot}^\star\\ \vdots\\ \delta_{l,n}\boldsymbol{X}_{n,\cdot}^\star\end{bmatrix}\right\| &\le \left(p\left\|\boldsymbol{X}^\star\right\|^2+C\sqrt{p\|\boldsymbol{X}^\star\|_{2,\infty}^2\left\|\boldsymbol{X}^\star\right\|^2\log n}+C\|\boldsymbol{X}^\star\|_{2,\infty}^2\log n\right)^{\frac{1}{2}}\\
&\le \left(p+C\sqrt{p\frac{\kappa\mu r}{n}\log n}+C\frac{\kappa\mu r\log n}{n}\right)^{\frac{1}{2}}\left\|\boldsymbol{X}^\star\right\|\le 1.2\sqrt{p}\left\|\boldsymbol{X}^\star\right\|,
\end{aligned}
\tag{144}
$$

where we also use the incoherence condition (114) and the sample complexity condition $n^2p\gg\kappa\mu rn\log n$. Hence, the event

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

together with (142) and (143) necessarily implies that

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

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

where the last inequality follows from the bound (140). As a result, with probability at least $1-O(n^{-10})$ (i.e. when (144) holds for all l's) we can upper bound ϕ_2 by

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

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

Next, we divide into multiple cases based on the size of $\|\boldsymbol{G}_l(\boldsymbol{\Delta}^t)\|$. By Lemma 42, for some constants $c_1,c_2>0$, with probability at least $1-c_1\exp(-c_2nr\log n)$,

$$
\sum_{l=1}^n\mathbb{1}_{\left\{\|\boldsymbol{G}_l(\boldsymbol{\Delta}^t)\|\ge 4\sqrt{p}\psi+\sqrt{2^kr}\xi\right\}}\le\frac{\alpha n}{2^{k-3}}
\tag{145}
$$

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

$$
\sqrt{2^kr}\xi\ge 4\sqrt{p}\psi\quad\text{or equivalently}\quad k\ge\log\frac{16p\psi^2}{r\xi^2};
\tag{146}
$$

otherwise we can use (140) to obtain

$$
4\sqrt{p}\psi+\sqrt{2^kr}\xi\le 8\sqrt{p}\psi\ll 1.5\sqrt{p}\frac{\xi}{\|\boldsymbol{X}^\star\|_{2,\infty}}\left\|\boldsymbol{X}^\star\right\|,
$$

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

$$
S_k=\left\{1\le l\le n:\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}^t\right)\right\|\in\left(\sqrt{2^kr}\xi,\sqrt{2^{k+1}r}\xi\right]\right\}
\tag{147}
$$

defined for each integer k obeying (146). Under the condition (146), it follows from (145) that

$$\sum_{l=1}^{n} \mathbb{1}_{\left\{\|\boldsymbol{G}_l(\boldsymbol{\Delta}^t)\| \geq \sqrt{2^{k+2} r}\xi\right\}} \leq \sum_{l=1}^{n} \mathbb{1}_{\left\{\|\boldsymbol{G}_l(\boldsymbol{\Delta}^t)\| \geq 4\sqrt{p}\psi + \sqrt{2^k r}\xi\right\}} \leq \frac{\alpha n}{2^{k-3}},$$

meaning that the cardinality of S_k satisfies

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

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

$$\begin{aligned}
\sqrt{\sum_{l \in S_k} \|\boldsymbol{A}_{l,\cdot}\|_2^2} &\overset{\text{(i)}}{\leq} \sqrt{|S_k| \cdot \|\boldsymbol{X}^\star\|_{2,\infty}^2 \left(1.2\sqrt{2^{k+1} r}\xi\sqrt{p}\,\|\boldsymbol{X}^\star\| + p\psi\,\|\boldsymbol{X}^\star\|\right)^2} \\
&\leq \sqrt{\frac{\alpha n}{2^{k-5}}}\,\|\boldsymbol{X}^\star\|_{2,\infty} \left(2\sqrt{2^{k+1} r}\xi\sqrt{p}\,\|\boldsymbol{X}^\star\| + p\psi\,\|\boldsymbol{X}^\star\|\right) \\
&\overset{\text{(ii)}}{\leq} 4\sqrt{\frac{\alpha n}{2^{k-5}}}\,\|\boldsymbol{X}^\star\|_{2,\infty} \sqrt{2^{k+1} r}\xi\sqrt{p}\,\|\boldsymbol{X}^\star\| \\
&\overset{\text{(iii)}}{\leq} 32\sqrt{\alpha\kappa\mu r^2 p}\xi\,\|\boldsymbol{X}^\star\|^2,
\end{aligned}$$

where (i) follows from the bound (143) and the constraint (147) in S_k, (ii) is a consequence of (146) and (iii) uses the incoherence condition (114).

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

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

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

B.4 Proof of Lemma 10

1. To obtain (73a), we invoke Lemma 37. 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

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

where (i) follows from (70c) and (ii) holds as long as $n^2 p \gg \kappa^2\mu^2 r^2 n$ and the noise satisfies (27). In addition,

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

$$\overset{\text{(ii)}}{\leq} C_3\rho^t\mu r\sqrt{\frac{\log n}{np}}\sigma_{\max} + \frac{C_7}{\sigma_{\min}}\sigma\sqrt{\frac{n\log n}{p}}\sigma_{\max}$$
$$\overset{\text{(iii)}}{\leq} \frac{1}{2}\sigma_{\min},$$

where (i) utilizes (70d), (ii) follows since $\|\boldsymbol{X}^\star\|_{2,\infty} \leq \|\boldsymbol{X}^\star\|$, and (iii) holds if $n^2p \gg \kappa^2\mu^2r^2n\log n$ and the noise satisfies (27). With these in place, Lemma 37 immediately yields (73a).

2. The first inequality in (73b) 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 (70a), (70d) and the triangle inequality yields

$$\begin{aligned}
&\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^\star\right\|_{\mathrm{F}} \leq \left\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\star\right\|_{\mathrm{F}} + \left\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\|_{\mathrm{F}} \\
&\quad\leq C_4\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\star\right\|_{\mathrm{F}} + \frac{C_1\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\star\right\|_{\mathrm{F}} + C_3\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^\star\right\|_{2,\infty} + \frac{C_7\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\star\right\|_{2,\infty} \\
&\quad\leq C_4\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\star\right\|_{\mathrm{F}} + \frac{C_1\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\star\right\|_{\mathrm{F}} + C_3\rho^t\mu r\sqrt{\frac{\log n}{np}}\sqrt{\frac{\kappa\mu}{n}}\left\|\boldsymbol{X}^\star\right\|_{\mathrm{F}} + \frac{C_7\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\sqrt{\frac{\kappa\mu}{n}}\left\|\boldsymbol{X}^\star\right\|_{\mathrm{F}} \\
&\quad\leq 2C_4\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\star\right\|_{\mathrm{F}} + \frac{2C_1\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\star\right\|_{\mathrm{F}},
\end{aligned}\tag{148}$$

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

3. To obtain (73c), we use (70d) and (70b) to get

$$\begin{aligned}
&\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^\star\right\|_{2,\infty} \leq \left\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^\star\right\|_{2,\infty} + \left\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\|_{\mathrm{F}} \\
&\quad\leq C_5\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^\star\right\|_{2,\infty} + \frac{C_8\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\star\right\|_{2,\infty} + C_3\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^\star\right\|_{2,\infty} + \frac{C_7\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\star\right\|_{2,\infty} \\
&\quad\leq \left(C_3 + C_5\right)\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^\star\right\|_{2,\infty} + \frac{C_8 + C_7}{\sigma_{\min}}\sigma\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\star\right\|_{2,\infty}.
\end{aligned}$$

4. Finally, to obtain (73d), one can take the triangle inequality

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

where the second line follows from (73a). Combine (70d) and (70c) to yield

$$\begin{aligned}
&\left\|\boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \boldsymbol{X}^\star\right\| \\
&\quad\leq 5\kappa\left(C_3\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^\star\right\|_{2,\infty} + \frac{C_7}{\sigma_{\min}}\sigma\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\star\right\|_{2,\infty}\right) + C_9\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\star\right\| + \frac{C_{10}}{\sigma_{\min}}\sigma\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\star\right\| \\
&\quad\leq 5\kappa\sqrt{\frac{\kappa\mu r}{n}}\left\|\boldsymbol{X}^\star\right\|\left(C_3\rho^t\mu r\sqrt{\frac{\log n}{np}} + \frac{C_7}{\sigma_{\min}}\sigma\sqrt{\frac{n\log n}{p}}\right) + C_9\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\star\right\| + \frac{C_{10}\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\star\right\| \\
&\quad\leq 2C_9\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\star\right\| + \frac{2C_{10}\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\star\right\|,
\end{aligned}$$

where the second inequality uses the incoherence of $\boldsymbol{X}^\star$ (cf. (114)) and the last inequality holds as long as $n \gg \kappa^3\mu r\log n$.

B.5 Proof of Lemma 11

From the definition of $\boldsymbol{R}^{t+1,(l)}$ (see (72)), 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 (24) and (69) allow one to express

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

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 $\nabla f_{\mathrm{clean}}\left(\boldsymbol{X}\right)$ to reach

$$\begin{aligned}
\boldsymbol{X}^{t+1} \widehat{\boldsymbol{H}}^{t} - \boldsymbol{X}^{t+1,(l)} \boldsymbol{R}^{t,(l)} &= \underbrace{\big(\boldsymbol{X}^{t} \widehat{\boldsymbol{H}}^{t} - \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)}\big) - \eta \left[\nabla f_{\mathrm{clean}}\big(\boldsymbol{X}^{t} \widehat{\boldsymbol{H}}^{t}\big) - \nabla f_{\mathrm{clean}}\big(\boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)}\big)\right]}_{:=\boldsymbol{B}_1^{(l)}} \\
&\quad \underbrace{- \eta \left[\frac{1}{p} \mathcal{P}_{\Omega_l}\left(\boldsymbol{X}^{t,(l)} \boldsymbol{X}^{t,(l)\top} - \boldsymbol{M}^{\star}\right) - \mathcal{P}_l\left(\boldsymbol{X}^{t,(l)} \boldsymbol{X}^{t,(l)\top} - \boldsymbol{M}^{\star}\right)\right] \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)}}_{:=\boldsymbol{B}_2^{(l)}} \\
&\quad \underbrace{+ \eta \frac{1}{p} \mathcal{P}_{\Omega}\left(\boldsymbol{E}\right) \left(\boldsymbol{X}^{t} \widehat{\boldsymbol{H}}^{t} - \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)}\right)}_{:=\boldsymbol{B}_3^{(l)}} + \underbrace{\eta \frac{1}{p} \mathcal{P}_{\Omega_l}\left(\boldsymbol{E}\right) \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)}}_{:=\boldsymbol{B}_4^{(l)}}, \qquad (149)
\end{aligned}$$

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

$$\nabla f^{(l)}\left(\boldsymbol{X}\right) = \nabla f\left(\boldsymbol{X}\right) - \frac{1}{p} \mathcal{P}_{\Omega_l}\left[\boldsymbol{X}\boldsymbol{X}^{\top} - \left(\boldsymbol{M}^{\star} + \boldsymbol{E}\right)\right] \boldsymbol{X} + \mathcal{P}_l\left(\boldsymbol{X}\boldsymbol{X}^{\top} - \boldsymbol{M}^{\star}\right) \boldsymbol{X} \qquad (150)$$

for all $\boldsymbol{X} \in \mathbb{R}^{n \times r}$ with $\mathcal{P}_{\Omega_l}$ and $\mathcal{P}_l$ defined respectively in (66) and (67). 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 24. *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 (149) satisfies*

   $$\left\| \boldsymbol{B}_4^{(l)} \right\|_{\mathrm{F}} \lesssim \eta \sigma \sqrt{\frac{n \log n}{p}} \left\| \boldsymbol{X}^{\star} \right\|_{2,\infty}.$$

2. The third term $\boldsymbol{B}_3^{(l)}$ can be bounded as follows

   $$\left\| \boldsymbol{B}_3^{(l)} \right\|_{\mathrm{F}} \leq \eta \left\| \frac{1}{p} \mathcal{P}_{\Omega}\left(\boldsymbol{E}\right) \right\| \left\| \boldsymbol{X}^{t} \widehat{\boldsymbol{H}}^{t} - \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} \right\|_{\mathrm{F}} \lesssim \eta \sigma \sqrt{\frac{n}{p}} \left\| \boldsymbol{X}^{t} \widehat{\boldsymbol{H}}^{t} - \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} \right\|_{\mathrm{F}},$$

   where the second inequality comes from Lemma 40.

3. For the second term $\boldsymbol{B}_2^{(l)}$, we have the following lemma.

Lemma 25. *Suppose that the sample size obeys $n^2p \gg \mu^2 r^2 n \log n$. Then with probability exceeding $1 - O\left(n^{-10}\right)$, the matrix $\boldsymbol{B}_2^{(l)}$ as defined in (149) satisfies*

$$\left\|\boldsymbol{B}_2^{(l)}\right\|_{\mathrm{F}} \lesssim \eta\sqrt{\frac{\kappa^2\mu^2r^2\log n}{np}}\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^\star\right\|_{2,\infty}\sigma_{\max}. \tag{151}$$

4. Regarding the first term $\boldsymbol{B}_1^{(l)}$, apply the fundamental theorem of calculus [Lan93, Chapter XIII, Theorem 4.2] to get
$$\mathrm{vec}\big(\boldsymbol{B}_1^{(l)}\big) = \left(\boldsymbol{I}_{nr} - \eta\int_0^1 \nabla^2 f_{\mathrm{clean}}\left(\boldsymbol{X}(\tau)\right)\mathrm{d}\tau\right)\mathrm{vec}\left(\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right), \tag{152}$$
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 B.2), we get
$$\|\boldsymbol{B}_1^{(l)}\|_{\mathrm{F}} \le \left(1 - \frac{\sigma_{\min}}{4}\eta\right)\left\|\boldsymbol{X}^t\widehat{\boldsymbol{H}}^t - \boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\|_{\mathrm{F}} \tag{153}$$
with the proviso that $0 < \eta \le (2\sigma_{\min})/(25\sigma_{\max}^2)$.

Applying the triangle inequality to (149) and invoking the preceding four bounds, we arrive at

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

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 (27). This taken collectively with the hypotheses (70d) and (73c) leads to

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

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

as long as $C_7 > 0$ is sufficiently large, where we have used the sample complexity assumption $n^2 p \gg \kappa^4 \mu^2 r^2 n \log n$ and the step size $0 < \eta \leq 1/(2\sigma_{\max}) \leq 1/(2\sigma_{\min})$. This finishes the proof.

B.5.1 Proof of Lemma 24

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

$$\frac{p}{\eta} \left\| \boldsymbol{B}_4^{(l)} \right\|_{\mathrm{F}} \leq \left\| \sum_{j=1}^{n} \underbrace{\delta_{l,j} E_{l,j} \boldsymbol{X}_{j,\cdot}^{t,(l)}}_{:= \boldsymbol{u}_j} \right\|_2 + \underbrace{\left\| \sum_{j: j \neq l} \delta_{l,j} E_{l,j} \boldsymbol{X}_{l,\cdot}^{t,(l)} \right\|_2}_{:= \alpha}, \tag{154}$$

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

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

where $\| \cdot \|_{\psi_1}$ denotes the sub-exponential norm [Kol11, Section A.1]. Further, one can calculate

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

Invoke the matrix Bernstein inequality [Kol11, Theorem 2.7] to discover that with probability at least $1 - O\left(n^{-10}\right)$,

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

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 (154) 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

$$\left\| \boldsymbol{X}^{t,(l)} \right\|_{2,\infty} = \left\| \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} \right\|_{2,\infty} \leq \left\| \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} - \boldsymbol{X}^\star \right\|_{2,\infty} + \left\| \boldsymbol{X}^\star \right\|_{2,\infty} \leq 2 \left\| \boldsymbol{X}^\star \right\|_{2,\infty}, \tag{155}$$

where the last inequality follows by combining (73c), the sample complexity condition $n^2 p \gg \mu^2 r^2 n \log n$, and the noise condition (27).

B.5.2 Proof of Lemma 25

For notational simplicity, we denote

$$\boldsymbol{C} := \boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{M}^{\star} = \boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\star}\boldsymbol{X}^{\star\top}. \tag{156}$$

Since the Frobenius norm is unitarily invariant, we have

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

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

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

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

$$\begin{aligned}
L &:= \max_{1\leq j\leq n}\left\|\left(\delta_{l,j}-p\right)C_{l,j}\boldsymbol{X}_{j,\cdot}^{t,(l)}\right\|_2 \leq \left\|\boldsymbol{C}\right\|_{\infty}\left\|\boldsymbol{X}^{t,(l)}\right\|_{2,\infty} \overset{\text{(i)}}{\leq} 2\left\|\boldsymbol{C}\right\|_{\infty}\left\|\boldsymbol{X}^{\star}\right\|_{2,\infty}, \\
V &:= \left\|\sum_{j=1}^{n}\mathbb{E}\left[\left(\delta_{l,j}-p\right)^2\right]C_{l,j}^2\boldsymbol{X}_{j,\cdot}^{t,(l)}\boldsymbol{X}_{j,\cdot}^{t,(l)\top}\right\| \leq p\|\boldsymbol{C}\|_{\infty}^2\left\|\sum_{j=1}^{n}\boldsymbol{X}_{j,\cdot}^{t,(l)}\boldsymbol{X}_{j,\cdot}^{t,(l)\top}\right\| \\
&= p\left\|\boldsymbol{C}\right\|_{\infty}^2\left\|\boldsymbol{X}^{t,(l)}\right\|_{\mathrm{F}}^2 \overset{\text{(ii)}}{\leq} 4p\left\|\boldsymbol{C}\right\|_{\infty}^2\left\|\boldsymbol{X}^{\star}\right\|_{\mathrm{F}}^2.
\end{aligned}$$

Here, (i) is a consequence of (155). In addition, (ii) follows from

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

where the last inequality comes from (73b), the sample complexity condition $n^2p \gg \mu^2 r^2 n\log n$, and the noise condition (27). The matrix Bernstein inequality [Tro15b, Theorem 6.1.1] reveals that

$$\left\|\sum_{j=1}^{n}\left(\delta_{l,j}-p\right)C_{l,j}\boldsymbol{X}_{j,\cdot}^{t,(l)}\right\|_2 \lesssim \sqrt{V\log n} + L\log n \lesssim \sqrt{p\left\|\boldsymbol{C}\right\|_{\infty}^2\left\|\boldsymbol{X}^{\star}\right\|_{\mathrm{F}}^2\log n} + \left\|\boldsymbol{C}\right\|_{\infty}\left\|\boldsymbol{X}^{\star}\right\|_{2,\infty}\log n$$

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

$$\frac{p}{\eta}\left\|\boldsymbol{B}_2^{(l)}\right\|_{\mathrm{F}} \lesssim \sqrt{p\log n}\left\|\boldsymbol{C}\right\|_{\infty}\left\|\boldsymbol{X}^{\star}\right\|_{\mathrm{F}} + \sqrt{np}\left\|\boldsymbol{C}\right\|_{\infty}\left\|\boldsymbol{X}^{\star}\right\|_{2,\infty} \tag{157}$$

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

To finish up, we make the observation that

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

$$\leq 3 \left\| \boldsymbol{X}^{t,(l)} \boldsymbol{R}^{t,(l)} - \boldsymbol{X}^\star \right\|_{2,\infty} \left\| \boldsymbol{X}^\star \right\|_{2,\infty}, \tag{158}$$

where the last line arises from (155). This combined with (157) gives

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

where (i) comes from (158), and (ii) makes use of the incoherence condition (114).

B.6 Proof of Lemma 12

We first introduce an auxiliary matrix

$$\widetilde{\boldsymbol{X}}^{t+1,(l)} := \boldsymbol{X}^{t,(l)} \widehat{\boldsymbol{H}}^{t,(l)} - \eta \left[\frac{1}{p} \mathcal{P}_{\Omega^{-l}} \left[\boldsymbol{X}^{t,(l)} \boldsymbol{X}^{t,(l)\top} - (\boldsymbol{M}^\star + \boldsymbol{E}) \right] + \mathcal{P}_l \left(\boldsymbol{X}^{t,(l)} \boldsymbol{X}^{t,(l)\top} - \boldsymbol{M}^\star \right) \right] \boldsymbol{X}^\star. \tag{159}$$

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

$$\left\| \left(\boldsymbol{X}^{t+1,(l)} \widehat{\boldsymbol{H}}^{t+1,(l)} - \boldsymbol{X}^\star \right)_{l,\cdot} \right\|_2 \leq \underbrace{\left\| \left(\boldsymbol{X}^{t+1,(l)} \widehat{\boldsymbol{H}}^{t+1,(l)} - \widetilde{\boldsymbol{X}}^{t+1,(l)} \right)_{l,\cdot} \right\|_2}_{:=\alpha_1} + \underbrace{\left\| \left(\widetilde{\boldsymbol{X}}^{t+1,(l)} - \boldsymbol{X}^\star \right)_{l,\cdot} \right\|_2}_{:=\alpha_2}. \tag{160}$$

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

1. Regarding the second term α_2 of (160), we see from the definition of $\widetilde{\boldsymbol{X}}^{t+1,(l)}$ (see (159)) that

$$\left(\widetilde{\boldsymbol{X}}^{t+1,(l)} - \boldsymbol{X}^\star \right)_{l,\cdot} = \left[\boldsymbol{X}^{t,(l)} \widehat{\boldsymbol{H}}^{t,(l)} - \eta \left(\boldsymbol{X}^{t,(l)} \boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^\star \boldsymbol{X}^{\star\top} \right) \boldsymbol{X}^\star - \boldsymbol{X}^\star \right]_{l,\cdot}, \tag{161}$$

where we also utilize the definitions of $\mathcal{P}_{\Omega^{-l}}$ and $\mathcal{P}_l$ in (67). For notational convenience, we denote

$$\boldsymbol{\Delta}^{t,(l)} := \boldsymbol{X}^{t,(l)} \widehat{\boldsymbol{H}}^{t,(l)} - \boldsymbol{X}^\star. \tag{162}$$

This allows us to rewrite (161) as

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

which further implies that

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

Here, the last line follows from the fact that $\left\|\boldsymbol{\Delta}_{l,\cdot}^{t,(l)}\right\|_2 \leq \|\boldsymbol{X}^\star\|_{2,\infty}$. To see this, one can use the induction hypothesis (70e) to get

$$\left\|\boldsymbol{\Delta}_{l,\cdot}^{t,(l)}\right\|_2 \leq C_2\rho^t\mu r\frac{1}{\sqrt{np}}\|\boldsymbol{X}^\star\|_{2,\infty} + C_6\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\|\boldsymbol{X}^\star\|_{2,\infty} \ll \|\boldsymbol{X}^\star\|_{2,\infty} \tag{163}$$

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

$$\alpha_2 \leq (1-\eta\sigma_{\min})\left\|\boldsymbol{\Delta}_{l,\cdot}^{t,(l)}\right\|_2 + 2\eta\|\boldsymbol{X}^\star\|_{2,\infty}\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\|\boldsymbol{X}^\star\|. \tag{164}$$

An immediate consequence of the above two inequalities and (73d) is

$$\alpha_2 \leq \|\boldsymbol{X}^\star\|_{2,\infty}. \tag{165}$$

2. The first term α_1 of (160) 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

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

Simple algebra yields

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

Here, to bound the the second term we have used

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

where the last inequality follows from (165). 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 (159) and the gradient update rule for $\boldsymbol{X}^{t+1,(l)}$ (see (69)), one has

$$\begin{aligned}&\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \widetilde{\boldsymbol{X}}^{t+1,(l)} \\ &= \left\{\boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \eta\left[\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left[\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - (\boldsymbol{M}^\star + \boldsymbol{E})\right] + \mathcal{P}_l\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{M}^\star\right)\right]\boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)}\right\} \\ &\quad - \left\{\boldsymbol{X}^{t,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \eta\left[\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left[\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - (\boldsymbol{M}^\star + \boldsymbol{E})\right] + \mathcal{P}_l\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{M}^\star\right)\right]\boldsymbol{X}^\star\right\} \\ &= -\eta\left[\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^\star\boldsymbol{X}^{\star\top}\right) + \mathcal{P}_l\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^\star\boldsymbol{X}^{\star\top}\right)\right]\boldsymbol{\Delta}^{t,(l)} + \frac{\eta}{p}\mathcal{P}_{\Omega^{-l}}(\boldsymbol{E})\boldsymbol{\Delta}^{t,(l)}.\end{aligned} \tag{166}$$

It is easy to verify that

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

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 40 and (iii) is a consequence of the noise condition (27). Therefore, in order to control (166), we need to upper bound the following quantity

$$\gamma := \left\| \frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\star}\boldsymbol{X}^{\star\top}\right) + \mathcal{P}_{l}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\star}\boldsymbol{X}^{\star\top}\right) \right\|. \tag{167}$$

To this end, we make the observation that

$$\begin{aligned}\gamma &\leq \underbrace{\left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\star}\boldsymbol{X}^{\star\top}\right) \right\|}_{:=\gamma_1} \\ &\quad + \underbrace{\left\| \frac{1}{p}\mathcal{P}_{\Omega_l}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\star}\boldsymbol{X}^{\star\top}\right) - \mathcal{P}_{l}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\star}\boldsymbol{X}^{\star\top}\right) \right\|}_{:=\gamma_2},\end{aligned} \tag{168}$$

where $\mathcal{P}_{\Omega_l}$ is defined in (66). An application of Lemma 43 reveals that

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

where $\boldsymbol{R}^{t,(l)} \in \mathcal{O}^{r\times r}$ is defined in (72). Let $\boldsymbol{C} = \boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\star}\boldsymbol{X}^{\star\top}$ as in (156), and one can bound the other term γ_2 by taking advantage of the triangle inequality and the symmetry property:

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

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 (158). The previous two bounds taken collectively give

$$\begin{aligned}\gamma &\leq 2n\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\star}\right\|_{2,\infty}^2 + 4\sqrt{n}\log n\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\star}\right\|_{2,\infty}\left\|\boldsymbol{X}^{\star}\right\| \\ &\quad + \widetilde{C}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\star}\right\|_{2,\infty}\left\|\boldsymbol{X}^{\star}\right\|_{2,\infty} \leq \frac{\delta}{2}\sigma_{\min}\end{aligned} \tag{169}$$

for some constant $\widetilde{C} > 0$ and $\delta > 0$ sufficiently small. The last inequality follows from (73c), the incoherence condition (114) and our sample size condition. In summary, we obtain

$$\left\|\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \widetilde{\boldsymbol{X}}^{t+1,(l)}\right\| \leq \eta\left(\gamma + \left\|\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left(\boldsymbol{E}\right)\right\|\right)\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \leq \eta\delta\sigma_{\min}\left\|\boldsymbol{\Delta}^{t,(l)}\right\|, \tag{170}$$

for $\delta > 0$ sufficiently small. With the estimate (170) in place, we can continue our derivation on β_1 and β_2.

(a) With regard to β_1, in view of (166) we can obtain

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

$$\leq \eta \left\|\boldsymbol{\Delta}_{l,\cdot}^{t,(l)}\right\|_2 \left\|\boldsymbol{X}^{t,(l)}\right\| \left\|\boldsymbol{\Delta}^{t,(l)}\right\| + \eta \left\|\boldsymbol{X}_{l,\cdot}^{\star}\right\|_2 \left\|\boldsymbol{\Delta}^{t,(l)}\right\|^2, \tag{171}$$

where (i) follows from the definitions of $\mathcal{P}_{\Omega^{-l}}$ and $\mathcal{P}_l$ (see (67) 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 (162).

(b) For β_2, we first claim that

$$\boldsymbol{I}_r := \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \left\|\widetilde{\boldsymbol{X}}^{t+1,(l)}\boldsymbol{R} - \boldsymbol{X}^{\star}\right\|_{\mathrm{F}}, \tag{172}$$

whose justification follows similar reasonings as that of (138), and is therefore omitted. In particular, it gives rise to the facts that $\boldsymbol{X}^{\star\top}\widetilde{\boldsymbol{X}}^{t+1,(l)}$ is symmetric and

$$\left(\widetilde{\boldsymbol{X}}^{t+1,(l)}\right)^{\top}\boldsymbol{X}^{\star} \succeq \frac{1}{2}\sigma_{\min}\boldsymbol{I}_r. \tag{173}$$

We are now ready to invoke Lemma 36 to bound β_2. We abuse the notation and denote $\boldsymbol{C} := \left(\widetilde{\boldsymbol{X}}^{t+1,(l)}\right)^{\top}\boldsymbol{X}^{\star}$ and $\boldsymbol{E} := \left(\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \widetilde{\boldsymbol{X}}^{t+1,(l)}\right)^{\top}\boldsymbol{X}^{\star}$. We have

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

The first inequality arises from (170), namely,

$$\begin{aligned}\|\boldsymbol{E}\| &\leq \left\|\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \widetilde{\boldsymbol{X}}^{t+1,(l)}\right\| \|\boldsymbol{X}^{\star}\| \leq \eta\delta\sigma_{\min}\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \|\boldsymbol{X}^{\star}\| \\ &\overset{\text{(i)}}{\leq} \eta\delta\sigma_{\min}\|\boldsymbol{X}^{\star}\|^2 \overset{\text{(ii)}}{\leq} \frac{1}{2}\sigma_{\min},\end{aligned}$$

where (i) holds since $\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \leq \|\boldsymbol{X}^{\star}\|$ and (ii) holds true for δ sufficiently small and $\eta \leq 1/\sigma_{\max}$. Invoke Lemma 36 to obtain

$$\begin{aligned}\beta_2 = \|\boldsymbol{R}_1 - \boldsymbol{I}_r\| &\leq \frac{2}{\sigma_{r-1}(\boldsymbol{C}) + \sigma_r(\boldsymbol{C})}\|\boldsymbol{E}\| \\ &\leq \frac{2}{\sigma_{\min}}\left\|\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t,(l)} - \widetilde{\boldsymbol{X}}^{t+1,(l)}\right\| \|\boldsymbol{X}^{\star}\| \qquad (174) \\ &\leq 2\delta\eta\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \|\boldsymbol{X}^{\star}\|, \qquad (175)\end{aligned}$$

where (174) follows since $\sigma_{r-1}(\boldsymbol{C}) \geq \sigma_r(\boldsymbol{C}) \geq \sigma_{\min}/2$ from (173), and the last line comes from (170).

(c) Putting the previous bounds (171) and (175) together yields

$$\alpha_1 \leq \eta\left\|\boldsymbol{\Delta}_{l,\cdot}^{t,(l)}\right\|_2 \left\|\boldsymbol{X}^{t,(l)}\right\| \left\|\boldsymbol{\Delta}^{t,(l)}\right\| + \eta\left\|\boldsymbol{X}_{l,\cdot}^{\star}\right\|_2\left\|\boldsymbol{\Delta}^{t,(l)}\right\|^2 + 4\delta\eta\|\boldsymbol{X}^{\star}\|_{2,\infty}\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \|\boldsymbol{X}^{\star}\|. \tag{176}$$

3. Combine (160), (164) and (176) to reach

$$\begin{aligned}\left\|\left(\boldsymbol{X}^{t+1,(l)}\widehat{\boldsymbol{H}}^{t+1,(l)} - \boldsymbol{X}^{\star}\right)_{l,\cdot}\right\|_2 &\leq (1-\eta\sigma_{\min})\left\|\boldsymbol{\Delta}_{l,\cdot}^{t,(l)}\right\|_2 + 2\eta\|\boldsymbol{X}^{\star}\|_{2,\infty}\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \|\boldsymbol{X}^{\star}\| \\ &\quad + \eta\left\|\boldsymbol{\Delta}_{l,\cdot}^{t,(l)}\right\|_2\left\|\boldsymbol{X}^{t,(l)}\right\|\left\|\boldsymbol{\Delta}^{t,(l)}\right\| + \eta\left\|\boldsymbol{X}_{l,\cdot}^{\star}\right\|_2\left\|\boldsymbol{\Delta}^{t,(l)}\right\|^2 + 4\delta\eta\|\boldsymbol{X}^{\star}\|_{2,\infty}\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\|\boldsymbol{X}^{\star}\| \\ &\overset{\text{(i)}}{\leq} \left(1 - \eta\sigma_{\min} + \eta\left\|\boldsymbol{X}^{t,(l)}\right\|\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\right)\left\|\boldsymbol{\Delta}_{l,\cdot}^{t,(l)}\right\|_2 + 4\eta\|\boldsymbol{X}^{\star}\|_{2,\infty}\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\|\boldsymbol{X}^{\star}\| \\ &\overset{\text{(ii)}}{\leq} \left(1 - \frac{\sigma_{\min}}{2}\eta\right)\left(C_2\rho^t\mu r\frac{1}{\sqrt{np}} + \frac{C_6}{\sigma_{\min}}\sigma\sqrt{\frac{n\log n}{p}}\right)\|\boldsymbol{X}^{\star}\|_{2,\infty} \\ &\quad + 4\eta\|\boldsymbol{X}^{\star}\|\|\boldsymbol{X}^{\star}\|_{2,\infty}\left(2C_9\rho^t\mu r\frac{1}{\sqrt{np}}\|\boldsymbol{X}^{\star}\| + \frac{2C_{10}}{\sigma_{\min}}\sigma\sqrt{\frac{n}{p}}\|\boldsymbol{X}^{\star}\|\right)\end{aligned}$$

$$\stackrel{\text{(iii)}}{\leq} C_2\rho^{t+1}\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\star\right\|_{2,\infty}+\frac{C_6}{\sigma_{\min}}\sigma\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\star\right\|_{2,\infty}.$$

Here, (i) follows since $\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \leq \left\|\boldsymbol{X}^\star\right\|$ and δ is sufficiently small, (ii) invokes the hypotheses (70e) and (73d) and recognizes that

$$\left\|\boldsymbol{X}^{t,(l)}\right\|\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \leq 2\left\|\boldsymbol{X}^\star\right\|\left(2C_9\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\star\right\|+\frac{2C_{10}}{\sigma_{\min}}\sigma\sqrt{\frac{n\log n}{np}}\left\|\boldsymbol{X}^\star\right\|\right) \leq \frac{\sigma_{\min}}{2}$$

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 $C_6 \gg \kappa C_{10}/\sqrt{\log n}$.

B.7 Proof of Lemma 13

For notational convenience, we define the following two orthonormal matrices

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

The problem of finding $\widehat{\boldsymbol{H}}^t$ (see (26)) is called the *orthogonal Procrustes problem* [TB77]. It is well-known that the minimizer $\widehat{\boldsymbol{H}}^t$ always exists and is given by

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

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

$$\operatorname{sgn}(\boldsymbol{B}) := \boldsymbol{U}\boldsymbol{V}^\top \tag{177}$$

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 5, respectively):

$$\begin{aligned}\left\|\boldsymbol{M}^0-\boldsymbol{M}^\star\right\| &\stackrel{\text{(i)}}{\leq} \left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{M}^\star\right)-\boldsymbol{M}^\star\right\|+\left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{E}\right)\right\| \\ &\stackrel{\text{(ii)}}{\leq} C\sqrt{\frac{n}{p}}\left\|\boldsymbol{M}^\star\right\|_\infty + C\sigma\sqrt{\frac{n}{p}} = C\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\star\right\|_{2,\infty}^2 + C\frac{\sigma}{\sqrt{\sigma_{\min}}}\sqrt{\frac{n}{p}}\sqrt{\sigma_{\min}} \\ &\stackrel{\text{(iii)}}{\leq} C\left\{\mu r\sqrt{\frac{1}{np}}\sqrt{\sigma_{\max}}+\frac{\sigma}{\sqrt{\sigma_{\min}}}\sqrt{\frac{n}{p}}\right\}\left\|\boldsymbol{X}^\star\right\| \stackrel{\text{(iv)}}{\ll} \sigma_{\min},\end{aligned} \tag{178}$$

for some universal constant $C>0$. Here, (i) arises from the triangle inequality, (ii) utilizes Lemma 39 and Lemma 40, (iii) follows from the incoherence condition (114) and (iv) holds under our sample complexity assumption that $n^2p \gg \mu^2 r^2 n$ and the noise condition (27). Similarly, we have

$$\left\|\boldsymbol{M}^{(l)}-\boldsymbol{M}^\star\right\| \lesssim \left\{\mu r\sqrt{\frac{1}{np}}\sqrt{\sigma_{\max}}+\frac{\sigma}{\sqrt{\sigma_{\min}}}\sqrt{\frac{n}{p}}\right\}\left\|\boldsymbol{X}^\star\right\| \ll \sigma_{\min}. \tag{179}$$

Combine Weyl's inequality, (178) and (179) to obtain

$$\left\|\boldsymbol{\Sigma}^0-\boldsymbol{\Sigma}^\star\right\| \leq \left\|\boldsymbol{M}^0-\boldsymbol{M}^\star\right\| \ll \sigma_{\min} \quad \text{and} \quad \left\|\boldsymbol{\Sigma}^{(l)}-\boldsymbol{\Sigma}^\star\right\| \leq \left\|\boldsymbol{M}^{(l)}-\boldsymbol{M}^\star\right\| \ll \sigma_{\min}, \tag{180}$$

which further implies

$$\frac{1}{2}\sigma_{\min} \leq \sigma_r\left(\boldsymbol{\Sigma}^0\right) \leq \sigma_1\left(\boldsymbol{\Sigma}^0\right) \leq 2\sigma_{\max} \quad \text{and} \quad \frac{1}{2}\sigma_{\min} \leq \sigma_r\left(\boldsymbol{\Sigma}^{(l)}\right) \leq \sigma_1\left(\boldsymbol{\Sigma}^{(l)}\right) \leq 2\sigma_{\max}. \tag{181}$$

We start by proving (70a), (70b) and (70c). The key decomposition we need is the following

$$\boldsymbol{X}^0\widehat{\boldsymbol{H}}^0-\boldsymbol{X}^\star = \boldsymbol{U}^0\left(\boldsymbol{\Sigma}^0\right)^{1/2}\left(\widehat{\boldsymbol{H}}^0-\boldsymbol{Q}\right)+\boldsymbol{U}^0\left[\left(\boldsymbol{\Sigma}^0\right)^{1/2}\boldsymbol{Q}-\boldsymbol{Q}\left(\boldsymbol{\Sigma}^\star\right)^{1/2}\right]+\left(\boldsymbol{U}^0\boldsymbol{Q}-\boldsymbol{U}^\star\right)\left(\boldsymbol{\Sigma}^\star\right)^{1/2}. \tag{182}$$

1. For the spectral norm error bound in (70c), the triangle inequality together with (182) yields

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

where we have also used the fact that $\|\boldsymbol{U}^0\|=1$. Recognizing that $\left\|\boldsymbol{M}^0-\boldsymbol{M}^\star\right\| \ll \sigma_{\min}$ (see (178)) and the assumption $\sigma_{\max}/\sigma_{\min} \lesssim 1$, we can apply Lemma 47, Lemma 46 and Lemma 45 to obtain

$$\left\|\widehat{\boldsymbol{H}}^0-\boldsymbol{Q}\right\| \lesssim \frac{1}{\sigma_{\min}}\left\|\boldsymbol{M}^0-\boldsymbol{M}^\star\right\|, \tag{183a}$$

$$\left\|\left(\boldsymbol{\Sigma}^0\right)^{1/2}\boldsymbol{Q}-\boldsymbol{Q}\left(\boldsymbol{\Sigma}^\star\right)^{1/2}\right\| \lesssim \frac{1}{\sqrt{\sigma_{\min}}}\left\|\boldsymbol{M}^0-\boldsymbol{M}^\star\right\|, \tag{183b}$$

$$\left\|\boldsymbol{U}^0\boldsymbol{Q}-\boldsymbol{U}^\star\right\| \lesssim \frac{1}{\sigma_{\min}}\left\|\boldsymbol{M}^0-\boldsymbol{M}^\star\right\|. \tag{183c}$$

These taken collectively imply the advertised upper bound

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

where we also utilize the fact that $\left\|\left(\boldsymbol{\Sigma}^0\right)^{1/2}\right\| \le \sqrt{2\sigma_{\max}}$ (see (181)) and the bounded condition number assumption, i.e. $\sigma_{\max}/\sigma_{\min} \lesssim 1$. This finishes the proof of (70c).

2. With regard to the Frobenius norm bound in (70a), one has

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

Here (i) arises from (70c) and (ii) holds true since $\sigma_{\max}/\sigma_{\min} \asymp 1$ and $\sqrt{r}\sqrt{\sigma_{\min}} \le \|\boldsymbol{X}^\star\|_{\mathrm{F}}$, thus completing the proof of (70a).

3. The proof of (70b) follows from similar arguments as used in proving (70c). Combine (182) and the triangle inequality to reach

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

Plugging in the estimates (178), (181), (183a) and (183b) results in

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

It remains to study the component-wise error of $\boldsymbol{U}^0$. To this end, it has already been shown in [AFWZ17, Lemma 14] that

$$\left\|\boldsymbol{U}^0\boldsymbol{Q}-\boldsymbol{U}^\star\right\|_{2,\infty} \lesssim \left(\mu r\sqrt{\frac{1}{np}} + \frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right)\left\|\boldsymbol{U}^\star\right\|_{2,\infty} \quad \text{and} \quad \left\|\boldsymbol{U}^0\right\|_{2,\infty} \lesssim \left\|\boldsymbol{U}^\star\right\|_{2,\infty} \tag{184}$$

under our assumptions. These combined with the previous inequality give

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

where the last relation is due to the observation that

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

4. We now move on to proving (70e). Recall that $\boldsymbol{Q}^{(l)}=\arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\left\|\boldsymbol{U}^{0,(l)}\boldsymbol{R}-\boldsymbol{U}^\star\right\|_{\mathrm{F}}$. By the triangle inequality,

$$\begin{aligned}\left\|\left(\boldsymbol{X}^{0,(l)}\widehat{\boldsymbol{H}}^{0,(l)}-\boldsymbol{X}^\star\right)_{l,\cdot}\right\|_2&\leq\left\|\boldsymbol{X}_{l,\cdot}^{0,(l)}\left(\widehat{\boldsymbol{H}}^{0,(l)}-\boldsymbol{Q}^{(l)}\right)\right\|_2+\left\|\left(\boldsymbol{X}^{0,(l)}\boldsymbol{Q}^{(l)}-\boldsymbol{X}^\star\right)_{l,\cdot}\right\|_2\\&\leq\left\|\boldsymbol{X}_{l,\cdot}^{0,(l)}\right\|_2\left\|\widehat{\boldsymbol{H}}^{0,(l)}-\boldsymbol{Q}^{(l)}\right\|+\left\|\left(\boldsymbol{X}^{0,(l)}\boldsymbol{Q}^{(l)}-\boldsymbol{X}^\star\right)_{l,\cdot}\right\|_2.\end{aligned}\tag{185}$$

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

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

We can thus decompose

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

which further implies that

$$\left\|\left(\boldsymbol{X}^{0,(l)}\boldsymbol{Q}^{(l)}-\boldsymbol{X}^\star\right)_{l,\cdot}\right\|_2\leq\left\|\boldsymbol{M}^\star\right\|_{2,\infty}\left\{\left\|\left(\boldsymbol{\Sigma}^{(l)}\right)^{-1/2}\boldsymbol{Q}^{(l)}-\boldsymbol{Q}^{(l)}\left(\boldsymbol{\Sigma}^\star\right)^{-1/2}\right\|+\frac{1}{\sqrt{\sigma_{\min}}}\left\|\boldsymbol{U}^{0,(l)}\boldsymbol{Q}^{(l)}-\boldsymbol{U}^\star\right\|\right\}.\tag{186}$$

In order to control this, we first see that

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

where the penultimate inequality uses (181) and the last inequality arises from Lemma 46. Additionally, Lemma 45 gives

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

Plugging the previous two bounds into (186), we reach

$$\left\|\left(\boldsymbol{X}^{0,(l)}\boldsymbol{Q}^{(l)}-\boldsymbol{X}^\star\right)_{l,\cdot}\right\|_2\lesssim\frac{1}{\sigma_{\min}^{3/2}}\left\|\boldsymbol{M}^{(l)}-\boldsymbol{M}^\star\right\|\left\|\boldsymbol{M}^\star\right\|_{2,\infty}\lesssim\left\{\mu r\sqrt{\frac{1}{np}}+\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right\}\left\|\boldsymbol{X}^\star\right\|_{2,\infty}.$$

where the last relation follows from $\left\|\boldsymbol{M}^\star\right\|_{2,\infty}=\left\|\boldsymbol{X}^\star\boldsymbol{X}^{\star\top}\right\|_{2,\infty}\leq\sqrt{\sigma_{\max}}\left\|\boldsymbol{X}^\star\right\|_{2,\infty}$ and the estimate (179). Note that this also implies that $\left\|\boldsymbol{X}_{l,\cdot}^{0,(l)}\right\|_2\leq 2\left\|\boldsymbol{X}^\star\right\|_{2,\infty}$. To see this, one has by the unitary invariance of $\left\|(\cdot)_{l,\cdot}\right\|_2$,

$$\left\|\boldsymbol{X}_{l,\cdot}^{0,(l)}\right\|_2=\left\|\boldsymbol{X}_{l,\cdot}^{0,(l)}\boldsymbol{Q}^{(l)}\right\|_2\leq\left\|\left(\boldsymbol{X}^{0,(l)}\boldsymbol{Q}^{(l)}-\boldsymbol{X}^\star\right)_{l,\cdot}\right\|_2+\left\|\boldsymbol{X}^\star_{l,\cdot}\right\|_2\leq 2\left\|\boldsymbol{X}^\star\right\|_{2,\infty}.$$

Substituting the above bounds back to (185) yields in

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

where the second line relies on Lemma 47, the bound (179), and the condition $\sigma_{\max}/\sigma_{\min} \asymp 1$. This establishes (70e).

5. Our final step is to justify (70d). 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. (72)), 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 46 and the bounds (178) and (179), 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 [DK70] we see that

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

These estimates taken together with (181) give

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

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 B.5). With probability at least $1-O\left(n^{-10}\right)$,

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

If one further has

$$\left\|\boldsymbol{U}^{0,(l)}\right\|_{2,\infty} \lesssim \left\|\boldsymbol{U}^{\star}\right\|_{2,\infty} \lesssim \frac{1}{\sqrt{\sigma_{\min}}}\left\|\boldsymbol{X}^{\star}\right\|_{2,\infty}, \tag{187}$$

then taking the previous bounds collectively establishes the desired bound

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

Proof of Claim (187). 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, [AFWZ17, Lemma 4 and Lemma 14] demonstrate that

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

On the other hand, by the Davis-Kahan $\sin\Theta$ theorem [DK70] we obtain

$$\left\| \boldsymbol{U}^{0,(l)} \mathrm{sgn}\left(\boldsymbol{U}^{0,(l)\top} \boldsymbol{U}^{(l),\mathrm{zero}} \right) - \boldsymbol{U}^{(l),\mathrm{zero}} \right\|_{\mathrm{F}} \lesssim \frac{1}{\sigma_{\min}} \left\| \left(\boldsymbol{M}^{(l)} - \boldsymbol{M}^{(l),\mathrm{zero}} \right) \boldsymbol{U}^{(l),\mathrm{zero}} \right\|_{\mathrm{F}}, \tag{188}$$

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,

$$\left\| \left(\boldsymbol{M}^{(l)} - \boldsymbol{M}^{(l),\mathrm{zero}} \right)_{l,\cdot} \boldsymbol{U}^{(l),\mathrm{zero}} \right\|_2 = \left\| \boldsymbol{M}^{\star}_{l,\cdot} \boldsymbol{U}^{(l),\mathrm{zero}} \right\|_2 \leq \|\boldsymbol{M}^{\star}\|_{2,\infty} \leq \sigma_{\max} \|\boldsymbol{U}^{\star}\|_{2,\infty}.$$

As a consequence, one has

$$\left\| \left(\boldsymbol{M}^{(l)} - \boldsymbol{M}^{(l),\mathrm{zero}} \right) \boldsymbol{U}^{(l),\mathrm{zero}} \right\|_{\mathrm{F}} = \left\| \left(\boldsymbol{M}^{(l)} - \boldsymbol{M}^{(l),\mathrm{zero}} \right)_{l,\cdot} \boldsymbol{U}^{(l),\mathrm{zero}} \right\|_2 \leq \sigma_{\max} \|\boldsymbol{U}^{\star}\|_{2,\infty},$$

which combined with (188) and the assumption $\sigma_{\max}/\sigma_{\min} \asymp 1$ yields

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

The claim (187) then follows by combining the above estimates:

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

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

C Proofs for blind deconvolution

Before proceeding to the proofs, we make note of the following concentration results. The standard Gaussian concentration inequality and the union bound give

$$\max_{1\leq l\leq m} \left| \boldsymbol{a}_l^{\mathsf{H}} \boldsymbol{x}^{\star} \right| \leq 5\sqrt{\log m} \tag{189}$$

with probability at least $1 - O(m^{-10})$. In addition, with probability exceeding $1 - Cm\exp(-cK)$ for some constants $c, C > 0$,

$$\max_{1\leq l\leq m} \|\boldsymbol{a}_l\|_2 \leq 3\sqrt{K}. \tag{190}$$

In addition, the population/expected Wirtinger Hessian at the truth $\boldsymbol{z}^{\star}$ is given by

$$\nabla^2 F\left(\boldsymbol{z}^{\star}\right) = \begin{bmatrix} \boldsymbol{I}_K & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{h}^{\star}\boldsymbol{x}^{\star\top} \\ \boldsymbol{0} & \boldsymbol{I}_K & \boldsymbol{x}^{\star}\boldsymbol{h}^{\star\top} & \boldsymbol{0} \\ \boldsymbol{0} & \left(\boldsymbol{x}^{\star}\boldsymbol{h}^{\star\top}\right)^{\mathsf{H}} & \boldsymbol{I}_K & \boldsymbol{0} \\ \left(\boldsymbol{h}^{\star}\boldsymbol{x}^{\star\top}\right)^{\mathsf{H}} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_K \end{bmatrix}. \tag{191}$$

C.1 Proof of Lemma 14

First, we find it convenient to decompose the Wirtinger Hessian (cf. (80)) into the expected Wirtinger Hessian at the truth (cf. (191)) and the perturbation part as follows:

$$\nabla^2 f\left(\boldsymbol{z}\right) = \nabla^2 F\left(\boldsymbol{z}^{\star}\right) + \left(\nabla^2 f\left(\boldsymbol{z}\right) - \nabla^2 F\left(\boldsymbol{z}^{\star}\right) \right). \tag{192}$$

The proof then proceeds by showing that (i) the population Hessian $\nabla^2 F\left(\boldsymbol{z}^{\star}\right)$ satisfies the restricted strong convexity and smoothness properties as advertised, and (ii) the perturbation $\nabla^2 f\left(\boldsymbol{z}\right) - \nabla^2 F\left(\boldsymbol{z}^{\star}\right)$ is well-controlled under our assumptions. We start by controlling the population Hessian in the following lemma.

Lemma 26. *Instate the notation and the conditions of Lemma 14. We have*

$$\left\|\nabla^2 F\big(\boldsymbol{z}^\star\big)\right\| = 2 \qquad \textit{and} \qquad \boldsymbol{u}^{\mathsf{H}}\left[\boldsymbol{D}\nabla^2 F\big(\boldsymbol{z}^\star\big) + \nabla^2 F\big(\boldsymbol{z}^\star\big)\boldsymbol{D}\right]\boldsymbol{u} \geq \|\boldsymbol{u}\|_2^2 .$$

The next step is to bound the perturbation. To this end, we define the set

$$\mathcal{S} := \{\boldsymbol{z} :\ \boldsymbol{z} \text{ satisfies } (82)\},$$

and derive the following lemma.

Lemma 27. *Suppose the sample complexity satisfies $m \gg \mu^2 K \log^9 m$, $c > 0$ is a sufficiently small constant, and $\delta = c/\log^2 m$. Then with probability at least $1 - O\left(m^{-10} + e^{-K}\log m\right)$, one has*

$$\sup_{\boldsymbol{z}\in\mathcal{S}} \left\|\nabla^2 f(\boldsymbol{z}) - \nabla^2 F(\boldsymbol{z}^\star)\right\| \leq 1/4.$$

Combining the two lemmas, we can easily see that for $\boldsymbol{z} \in \mathcal{S}$,

$$\left\|\nabla^2 f(\boldsymbol{z})\right\| \leq \left\|\nabla^2 F(\boldsymbol{z}^\star)\right\| + \left\|\nabla^2 f(\boldsymbol{z}) - \nabla^2 F(\boldsymbol{z}^\star)\right\| \leq 2 + 1/4 \leq 3,$$

which verifies the smoothness upper bound. In addition,

$$\begin{aligned}
&\boldsymbol{u}^{\mathsf{H}}\left[\boldsymbol{D}\nabla^2 f(\boldsymbol{z}) + \nabla^2 f(\boldsymbol{z})\boldsymbol{D}\right]\boldsymbol{u} \\
&\quad = \boldsymbol{u}^{\mathsf{H}}\left[\boldsymbol{D}\nabla^2 F(\boldsymbol{z}^\star) + \nabla^2 F(\boldsymbol{z}^\star)\boldsymbol{D}\right]\boldsymbol{u} + \boldsymbol{u}^{\mathsf{H}}\boldsymbol{D}\left[\nabla^2 f(\boldsymbol{z}) - \nabla^2 F(\boldsymbol{z}^\star)\right]\boldsymbol{u} + \boldsymbol{u}^{\mathsf{H}}\left[\nabla^2 f(\boldsymbol{z}) - \nabla^2 F(\boldsymbol{z}^\star)\right]\boldsymbol{D}\boldsymbol{u} \\
&\quad \overset{\text{(i)}}{\geq} \boldsymbol{u}^{\mathsf{H}}\left[\boldsymbol{D}\nabla^2 F(\boldsymbol{z}^\star) + \nabla^2 F(\boldsymbol{z}^\star)\boldsymbol{D}\right]\boldsymbol{u} - 2\|\boldsymbol{D}\|\left\|\nabla^2 f(\boldsymbol{z}) - \nabla^2 F(\boldsymbol{z}^\star)\right\|\|\boldsymbol{u}\|_2^2 \\
&\quad \overset{\text{(ii)}}{\geq} \|\boldsymbol{u}\|_2^2 - 2(1+\delta)\cdot\frac{1}{4}\|\boldsymbol{u}\|_2^2 \\
&\quad \overset{\text{(iii)}}{\geq} \frac{1}{4}\|\boldsymbol{u}\|_2^2,
\end{aligned}$$

where (i) uses the triangle inequality, (ii) holds because of Lemma 27 and the fact that $\|\boldsymbol{D}\| \leq 1+\delta$, and (iii) follows if $\delta \leq 1/2$. This establishes the claim on the restricted strong convexity.

C.1.1 Proof of Lemma 26

We start by proving the identity $\left\|\nabla^2 F(\boldsymbol{z}^\star)\right\| = 2$. Let

$$\boldsymbol{u}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}\boldsymbol{h}^\star\\ \boldsymbol{0}\\ \boldsymbol{0}\\ \overline{\boldsymbol{x}^\star}\end{bmatrix}, \qquad \boldsymbol{u}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{x}^\star\\ \overline{\boldsymbol{h}^\star}\\ \boldsymbol{0}\end{bmatrix}, \qquad \boldsymbol{u}_3 = \frac{1}{\sqrt{2}}\begin{bmatrix}\boldsymbol{h}^\star\\ \boldsymbol{0}\\ \boldsymbol{0}\\ -\overline{\boldsymbol{x}^\star}\end{bmatrix}, \qquad \boldsymbol{u}_4 = \frac{1}{\sqrt{2}}\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{x}^\star\\ -\overline{\boldsymbol{h}^\star}\\ \boldsymbol{0}\end{bmatrix}.$$

Recalling that $\|\boldsymbol{h}^\star\|_2 = \|\boldsymbol{x}^\star\|_2 = 1$, we can easily check that these four vectors form an orthonormal set. A little algebra reveals that

$$\nabla^2 F(\boldsymbol{z}^\star) = \boldsymbol{I}_{4K} + \boldsymbol{u}_1\boldsymbol{u}_1^{\mathsf{H}} + \boldsymbol{u}_2\boldsymbol{u}_2^{\mathsf{H}} - \boldsymbol{u}_3\boldsymbol{u}_3^{\mathsf{H}} - \boldsymbol{u}_4\boldsymbol{u}_4^{\mathsf{H}},$$

which immediately implies

$$\left\|\nabla^2 F(\boldsymbol{z}^\star)\right\| = 2.$$

We now turn attention to the restricted strong convexity. Since $\boldsymbol{u}^{\mathsf{H}}\boldsymbol{D}\nabla^2 F(\boldsymbol{z}^\star)\boldsymbol{u}$ is the complex conjugate of $\boldsymbol{u}^{\mathsf{H}}\nabla^2 F(\boldsymbol{z}^\star)\boldsymbol{D}\boldsymbol{u}$ as both $\nabla^2 F(\boldsymbol{z}^\star)$ and $\boldsymbol{D}$ are Hermitian, we will focus on the first term $\boldsymbol{u}^{\mathsf{H}}\boldsymbol{D}\nabla^2 F(\boldsymbol{z}^\star)\boldsymbol{u}$. This term can be rewritten as

$$\boldsymbol{u}^{\mathsf{H}}\boldsymbol{D}\nabla^2 F(\boldsymbol{z}^\star)\boldsymbol{u}$$

$$\overset{\text{(i)}}{=}\left[(\boldsymbol{h}_1-\boldsymbol{h}_2)^{\mathsf{H}},(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}},\left(\overline{\boldsymbol{h}_1-\boldsymbol{h}_2}\right)^{\mathsf{H}},\left(\overline{\boldsymbol{x}_1-\boldsymbol{x}_2}\right)^{\mathsf{H}}\right]\boldsymbol{D}\begin{bmatrix}\boldsymbol{I}_K & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{h}^\star\boldsymbol{x}^{\star\top}\\ \boldsymbol{0} & \boldsymbol{I}_K & \boldsymbol{x}^\star\boldsymbol{h}^{\star\top} & \boldsymbol{0}\\ \boldsymbol{0} & \left(\boldsymbol{x}^\star\boldsymbol{h}^{\star\top}\right)^{\mathsf{H}} & \boldsymbol{I}_K & \boldsymbol{0}\\ \left(\boldsymbol{h}^\star\boldsymbol{x}^{\star\top}\right)^{\mathsf{H}} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_K\end{bmatrix}\begin{bmatrix}\boldsymbol{h}_1-\boldsymbol{h}_2\\ \boldsymbol{x}_1-\boldsymbol{x}_2\\ \overline{\boldsymbol{h}_1-\boldsymbol{h}_2}\\ \overline{\boldsymbol{x}_1-\boldsymbol{x}_2}\end{bmatrix}$$

$$\overset{\text{(ii)}}{=}\left[\gamma_1(\boldsymbol{h}_1-\boldsymbol{h}_2)^{\mathsf{H}},\gamma_2(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}},\gamma_1\left(\overline{\boldsymbol{h}_1-\boldsymbol{h}_2}\right)^{\mathsf{H}},\gamma_2\left(\overline{\boldsymbol{x}_1-\boldsymbol{x}_2}\right)^{\mathsf{H}}\right]\begin{bmatrix}\boldsymbol{h}_1-\boldsymbol{h}_2+\boldsymbol{h}^\star\boldsymbol{x}^{\star\top}\overline{(\boldsymbol{x}_1-\boldsymbol{x}_2)}\\ \boldsymbol{x}_1-\boldsymbol{x}_2+\boldsymbol{x}^\star\boldsymbol{h}^{\star\top}\overline{(\boldsymbol{h}_1-\boldsymbol{h}_2)}\\ \left(\boldsymbol{x}^\star\boldsymbol{h}^{\star\top}\right)^{\mathsf{H}}(\boldsymbol{x}_1-\boldsymbol{x}_2)+\overline{(\boldsymbol{h}_1-\boldsymbol{h}_2)}\\ \left(\boldsymbol{h}^\star\boldsymbol{x}^{\star\top}\right)^{\mathsf{H}}(\boldsymbol{h}_1-\boldsymbol{h}_2)+\overline{(\boldsymbol{x}_1-\boldsymbol{x}_2)}\end{bmatrix}$$

$$=\left[\gamma_1(\boldsymbol{h}_1-\boldsymbol{h}_2)^{\mathsf{H}},\gamma_2(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}},\gamma_1\left(\overline{\boldsymbol{h}_1-\boldsymbol{h}_2}\right)^{\mathsf{H}},\gamma_2\left(\overline{\boldsymbol{x}_1-\boldsymbol{x}_2}\right)^{\mathsf{H}}\right]\begin{bmatrix}\boldsymbol{h}_1-\boldsymbol{h}_2+\boldsymbol{h}^\star(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}}\boldsymbol{x}^\star\\ \boldsymbol{x}_1-\boldsymbol{x}_2+\boldsymbol{x}^\star(\boldsymbol{h}_1-\boldsymbol{h}_2)^{\mathsf{H}}\boldsymbol{h}^\star\\ \overline{\boldsymbol{h}_1-\boldsymbol{h}_2}+\overline{\boldsymbol{h}^\star(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}}\boldsymbol{x}^\star}\\ \overline{\boldsymbol{x}_1-\boldsymbol{x}_2}+\overline{\boldsymbol{x}^\star(\boldsymbol{h}_1-\boldsymbol{h}_2)^{\mathsf{H}}\boldsymbol{h}^\star}\end{bmatrix}$$

$$\begin{aligned}&=2\gamma_1\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2^2+2\gamma_2\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2\\&\quad+(\gamma_1+\gamma_2)\underbrace{(\boldsymbol{h}_1-\boldsymbol{h}_2)^{\mathsf{H}}\boldsymbol{h}^\star(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}}\boldsymbol{x}^\star}_{:=\beta}+(\gamma_1+\gamma_2)\underbrace{\overline{(\boldsymbol{h}_1-\boldsymbol{h}_2)^{\mathsf{H}}\boldsymbol{h}^\star(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}}\boldsymbol{x}^\star}}_{=\overline{\beta}},\end{aligned}\tag{193}$$

where (i) uses the definitions of $\boldsymbol{u}$ and $\nabla^2F(\boldsymbol{z}^\star)$, and (ii) follows from the definition of $\boldsymbol{D}$. In view of the assumption (84), we can obtain

$$2\gamma_1\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2^2+2\gamma_2\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2\geq 2\min\{\gamma_1,\gamma_2\}\left(\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2^2+\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2\right)\geq(1-\delta)\|\boldsymbol{u}\|_2^2,$$

where the last inequality utilizes the identity

$$2\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2^2+2\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2=\|\boldsymbol{u}\|_2^2.\tag{194}$$

It then boils down to controlling β. Toward this goal, we decompose β into the following four terms

$$\begin{aligned}\beta=&\underbrace{(\boldsymbol{h}_1-\boldsymbol{h}_2)^{\mathsf{H}}\boldsymbol{h}_2(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}}\boldsymbol{x}_2}_{:=\beta_1}+\underbrace{(\boldsymbol{h}_1-\boldsymbol{h}_2)^{\mathsf{H}}(\boldsymbol{h}^\star-\boldsymbol{h}_2)(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}}(\boldsymbol{x}^\star-\boldsymbol{x}_2)}_{:=\beta_2}\\&+\underbrace{(\boldsymbol{h}_1-\boldsymbol{h}_2)^{\mathsf{H}}(\boldsymbol{h}^\star-\boldsymbol{h}_2)(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}}\boldsymbol{x}_2}_{:=\beta_3}+\underbrace{(\boldsymbol{h}_1-\boldsymbol{h}_2)^{\mathsf{H}}\boldsymbol{h}_2(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}}(\boldsymbol{x}^\star-\boldsymbol{x}_2)}_{:=\beta_4}.\end{aligned}$$

Since $\|\boldsymbol{h}_2-\boldsymbol{h}^\star\|_2$ and $\|\boldsymbol{x}_2-\boldsymbol{x}^\star\|_2$ are both small by (83), β_2,β_3 and β_4 are well-bounded. Specifically, regarding β_2, we discover that

$$|\beta_2|\leq\|\boldsymbol{h}^\star-\boldsymbol{h}_2\|_2\|\boldsymbol{x}^\star-\boldsymbol{x}_2\|_2\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2\leq\delta^2\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2\leq\delta\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2,$$

where the second inequality is due to (83) and the last one holds since $\delta<1$. Similarly, we can obtain

$$\begin{aligned}&|\beta_3|\leq\delta\|\boldsymbol{x}_2\|_2\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2\leq2\delta\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2,\\ \text{and}\qquad&|\beta_4|\leq\delta\|\boldsymbol{h}_2\|_2\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2\leq2\delta\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2,\end{aligned}$$

where both lines make use of the facts that

$$\|\boldsymbol{x}_2\|_2\leq\|\boldsymbol{x}_2-\boldsymbol{x}^\star\|_2+\|\boldsymbol{x}^\star\|_2\leq1+\delta\leq2\qquad\text{and}\qquad\|\boldsymbol{h}_2\|_2\leq\|\boldsymbol{h}_2-\boldsymbol{h}^\star\|_2+\|\boldsymbol{h}^\star\|_2\leq1+\delta\leq2.\tag{195}$$

Combine the previous three bounds to reach

$$|\beta_2|+|\beta_3|+|\beta_4|\leq5\delta\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2\leq5\delta\frac{\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2^2+\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2}{2}=\frac{5}{4}\delta\|\boldsymbol{u}\|_2^2,$$

where we utilize the elementary inequality $ab\leq(a^2+b^2)/2$ and the identity (194).

The only remaining term is thus β_1. Recalling that $(\boldsymbol{h}_1,\boldsymbol{x}_1)$ and $(\boldsymbol{h}_2,\boldsymbol{x}_2)$ are aligned by our assumption, we can invoke Lemma 56 to obtain

$$(\boldsymbol{h}_1-\boldsymbol{h}_2)^{\mathsf{H}}\boldsymbol{h}_2=\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2+\boldsymbol{x}_2^{\mathsf{H}}(\boldsymbol{x}_1-\boldsymbol{x}_2)-\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2^2,$$

which allows one to rewrite β_1 as

$$\begin{aligned}\beta_1&=\left\{\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2+\boldsymbol{x}_2^{\mathsf{H}}(\boldsymbol{x}_1-\boldsymbol{x}_2)-\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2^2\right\}\cdot(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}}\boldsymbol{x}_2\\&=(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}}\boldsymbol{x}_2\left(\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2-\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2^2\right)+\left|(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}}\boldsymbol{x}_2\right|^2.\end{aligned}$$

Consequently,

$$\begin{aligned}\beta_1+\overline{\beta_1}&=2\left|(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}}\boldsymbol{x}_2\right|_2^2+2\mathrm{Re}\left[(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}}\boldsymbol{x}_2\right]\left(\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2-\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2^2\right)\\&\geq 2\mathrm{Re}\left[(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}}\boldsymbol{x}_2\right]\left(\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2-\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2^2\right)\\&\overset{\text{(i)}}{\geq}-\left|(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathsf{H}}\boldsymbol{x}_2\right|\|\boldsymbol{u}\|_2^2\\&\overset{\text{(ii)}}{\geq}-4\delta\|\boldsymbol{u}\|_2^2.\end{aligned}$$

Here, (i) arises from the triangle inequality that

$$\left|\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2-\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2^2\right|\leq\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2+\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2^2=\frac{1}{2}\|\boldsymbol{u}\|_2^2,$$

and (ii) occurs since $\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2\leq\|\boldsymbol{x}_1-\boldsymbol{x}^\star\|_2+\|\boldsymbol{x}_2-\boldsymbol{x}^\star\|_2\leq 2\delta$ and $\|\boldsymbol{x}_2\|_2\leq 2$ (see (195)).

To finish up, note that $\gamma_1+\gamma_2\leq 2(1+\delta)\leq 3$ for $\delta<1/2$. Substitute these bounds into (193) to obtain

$$\begin{aligned}\boldsymbol{u}^{\mathsf{H}}\boldsymbol{D}\nabla^2F(\boldsymbol{z}^\star)\boldsymbol{u}&\geq(1-\delta)\|\boldsymbol{u}\|_2^2+(\gamma_1+\gamma_2)\left(\beta+\overline{\beta}\right)\\&\geq(1-\delta)\|\boldsymbol{u}\|_2^2+(\gamma_1+\gamma_2)\left(\beta_1+\overline{\beta_1}\right)-2(\gamma_1+\gamma_2)(|\beta_2|+|\beta_3|+|\beta_4|)\\&\geq(1-\delta)\|\boldsymbol{u}\|_2^2-12\delta\|\boldsymbol{u}\|_2^2-6\cdot\frac{5}{4}\delta\|\boldsymbol{u}\|_2^2\\&\geq(1-20.5\delta)\|\boldsymbol{u}\|_2^2\\&\geq\frac{1}{2}\|\boldsymbol{u}\|_2^2\end{aligned}$$

as long as δ is small enough.

C.1.2 Proof of Lemma 27

In view of the expressions of $\nabla^2 f(\boldsymbol{z})$ and $\nabla^2F(\boldsymbol{z}^\star)$ (cf. (80) and (191)) and the triangle inequality, we get

$$\left\|\nabla^2 f(\boldsymbol{z})-\nabla^2F(\boldsymbol{z}^\star)\right\|\leq 2\alpha_1+2\alpha_2+4\alpha_3+4\alpha_4,\tag{196}$$

where the four terms on the right-hand side are defined as follows

$$\begin{aligned}\alpha_1&=\left\|\sum_{j=1}^m\left|\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}\right|^2\boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}-\boldsymbol{I}_K\right\|,&\alpha_2&=\left\|\sum_{j=1}^m\left|\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}\right|^2\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}-\boldsymbol{I}_K\right\|,\\\alpha_3&=\left\|\sum_{j=1}^m\left(\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}\boldsymbol{x}^{\mathsf{H}}\boldsymbol{a}_j-y_j\right)\boldsymbol{b}_j\boldsymbol{a}_j^{\mathsf{H}}\right\|,&\alpha_4&=\left\|\sum_{j=1}^m\boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}\left(\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}\right)^\top-\boldsymbol{h}^\star\boldsymbol{x}^{\star\top}\right\|.\end{aligned}$$

In what follows, we shall control $\sup_{\boldsymbol{z}\in\mathcal{S}}\alpha_j$ for $j=1,2,3,4$ separately.

1. Regarding the first term α_1, the triangle inequality gives

$$\alpha_1 \leq \underbrace{\left\| \sum_{j=1}^{m} \left| \boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x} \right|^2 \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} - \sum_{j=1}^{m} \left| \boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^\star \right|^2 \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} \right\|}_{:=\beta_1} + \underbrace{\left\| \sum_{j=1}^{m} \left| \boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^\star \right|^2 \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} - \boldsymbol{I}_K \right\|}_{:=\beta_2}.$$

   - To control β_1, the key observation is that $\boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}$ and $\boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^\star$ are extremely close. We can rewrite β_1 as

$$\beta_1 = \left\| \sum_{j=1}^{m} \left(\left| \boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x} \right|^2 - \left| \boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^\star \right|^2 \right) \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} \right\| \leq \left\| \sum_{j=1}^{m} \left| \left| \boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x} \right|^2 - \left| \boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^\star \right|^2 \right| \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} \right\|, \tag{197}$$

     where

$$\begin{aligned} \left| \left| \boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x} \right|^2 - \left| \boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^\star \right|^2 \right| &\overset{\text{(i)}}{=} \left| \left[\boldsymbol{a}_j^{\mathsf{H}} \left(\boldsymbol{x} - \boldsymbol{x}^\star \right) \right]^{\mathsf{H}} \boldsymbol{a}_j^{\mathsf{H}} \left(\boldsymbol{x} - \boldsymbol{x}^\star \right) + \left[\boldsymbol{a}_j^{\mathsf{H}} \left(\boldsymbol{x} - \boldsymbol{x}^\star \right) \right]^{\mathsf{H}} \boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^\star + \left(\boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^\star \right)^{\mathsf{H}} \boldsymbol{a}_j^{\mathsf{H}} \left(\boldsymbol{x} - \boldsymbol{x}^\star \right) \right| \\ &\overset{\text{(ii)}}{\leq} \left| \boldsymbol{a}_j^{\mathsf{H}} \left(\boldsymbol{x} - \boldsymbol{x}^\star \right) \right|^2 + 2 \left| \boldsymbol{a}_j^{\mathsf{H}} \left(\boldsymbol{x} - \boldsymbol{x}^\star \right) \right| \left| \boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^\star \right| \\ &\overset{\text{(iii)}}{\leq} 4 C_3^2 \frac{1}{\log^3 m} + 4 C_3 \frac{1}{\log^{3/2} m} \cdot 5 \sqrt{\log m} \\ &\lesssim C_3 \frac{1}{\log m}. \end{aligned}$$

     Here, the first line (i) uses the identity for $u, v \in \mathbb{C}$,

$$|u|^2 - |v|^2 = u^{\mathsf{H}} u - v^{\mathsf{H}} v = (u - v)^{\mathsf{H}} (u - v) + (u - v)^{\mathsf{H}} v + v^{\mathsf{H}} (u - v),$$

     the second relation (ii) comes from the triangle inequality, and the third line (iii) follows from (189) and the assumption (82b). Substitution into (197) gives

$$\beta_1 \leq \max_{1 \leq j \leq m} \left| \left| \boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x} \right|^2 - \left| \boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^\star \right|^2 \right| \left\| \sum_{j=1}^{m} \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} \right\| \lesssim C_3 \frac{1}{\log m},$$

     where the last inequality comes from the fact that $\sum_{j=1}^{m} \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} = \boldsymbol{I}_K$.

   - The other term β_2 can be bounded through Lemma 59, which reveals that with probability $1 - O\left(m^{-10}\right)$,

$$\beta_2 \lesssim \sqrt{\frac{K}{m} \log m}.$$

   Taken collectively, the preceding two bounds give

$$\sup_{\boldsymbol{z} \in \mathcal{S}} \alpha_1 \lesssim \sqrt{\frac{K}{m} \log m} + C_3 \frac{1}{\log m}.$$

   Hence $\mathbb{P}(\sup_{\boldsymbol{z} \in \mathcal{S}} \alpha_1 \leq 1/32) = 1 - O(m^{-10})$.

2. We are going to prove that $\mathbb{P}(\sup_{\boldsymbol{z} \in \mathcal{S}} \alpha_2 \leq 1/32) = 1 - O(m^{-10})$. The triangle inequality allows us to bound α_2 as

$$\alpha_2 \leq \underbrace{\left\| \sum_{j=1}^{m} \left| \boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{h} \right|^2 \boldsymbol{a}_j \boldsymbol{a}_j^{\mathsf{H}} - \|\boldsymbol{h}\|_2^2 \boldsymbol{I}_K \right\|}_{:=\theta_1(\boldsymbol{h})} + \underbrace{\left\| \|\boldsymbol{h}\|_2^2 \boldsymbol{I}_K - \boldsymbol{I}_K \right\|}_{:=\theta_2(\boldsymbol{h})}.$$

The second term $\theta_2(\boldsymbol{h})$ is easy to control. To see this, we have

$$\theta_2(\boldsymbol{h}) = \left|\|\boldsymbol{h}\|_2^2 - 1\right| = \left|\,\|\boldsymbol{h}\|_2 - 1\right| (\|\boldsymbol{h}\|_2 + 1) \le 3\delta < 1/64,$$

where the penultimate relation uses the assumption that $\|\boldsymbol{h} - \boldsymbol{h}^\star\|_2 \le \delta$ and hence

$$\left|\,\|\boldsymbol{h}\|_2 - 1\right| \le \delta, \qquad \|\boldsymbol{h}\|_2 \le 1 + \delta \le 2.$$

For the first term $\theta_1(\boldsymbol{h})$, we define a new set

$$\mathcal{H} := \left\{ \boldsymbol{h} \in \mathbb{C}^K : \ \|\boldsymbol{h} - \boldsymbol{h}^\star\|_2 \le \delta \quad \text{and} \quad \max_{1\le j\le m} \left|\boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{h}\right| \le \frac{2C_4 \mu \log^2 m}{\sqrt{m}} \right\}.$$

It is easily seen that $\sup_{\boldsymbol{z}\in\mathcal{S}} \theta_1 \le \sup_{\boldsymbol{h}\in\mathcal{H}} \theta_1$. We plan to use the standard covering argument to show that

$$\mathbb{P}\left(\sup_{\boldsymbol{h}\in\mathcal{H}} \theta_1(\boldsymbol{h}) \le 1/64 \right) = 1 - O(m^{-10}). \tag{198}$$

To this end, we define $c_j(\boldsymbol{h}) = |\boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{h}|^2$ for every $1 \le j \le m$. It is straightforward to check that

$$\theta_1(\boldsymbol{h}) = \left\| \sum_{j=1}^m c_j(\boldsymbol{h}) \left(\boldsymbol{a}_j \boldsymbol{a}_j^{\mathsf{H}} - \boldsymbol{I}_K \right) \right\|, \qquad \max_{1\le j\le m} |c_j| \le \left(\frac{2C_4 \mu \log^2 m}{\sqrt{m}} \right)^2, \tag{199}$$

$$\sum_{j=1}^m c_j^2 = \sum_{j=1}^m |\boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{h}|^4 \le \left\{ \max_{1\le j\le m} |\boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{h}|^2 \right\} \sum_{j=1}^m |\boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{h}|^2 = \left\{ \max_{1\le j\le m} |\boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{h}|^2 \right\} \|\boldsymbol{h}\|_2^2 \le 4 \left(\frac{2C_4 \mu \log^2 m}{\sqrt{m}} \right)^2 \tag{200}$$

for $\boldsymbol{h} \in \mathcal{H}$. In the above argument, we have used the facts that $\sum_{j=1}^m \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} = \boldsymbol{I}_K$ and

$$\sum_{j=1}^m |\boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{h}|^2 = \boldsymbol{h}^{\mathsf{H}} \left(\sum_{j=1}^m \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} \right) \boldsymbol{h} = \|\boldsymbol{h}\|_2^2 \le (1+\delta)^2 \le 4,$$

together with the definition of $\mathcal{H}$. Lemma 57 combined with (199) and (200) readily yields that for any fixed $\boldsymbol{h} \in \mathcal{H}$ and any $t \ge 0$,

$$\begin{aligned} \mathbb{P}(\theta_1(\boldsymbol{h}) \ge t) &\le 2 \exp\left(\widetilde{C}_1 K - \widetilde{C}_2 \min\left\{ \frac{t}{\max_{1\le j\le m} |c_j|}, \frac{t^2}{\sum_{j=1}^m c_j^2} \right\} \right) \\ &\le 2 \exp\left(\widetilde{C}_1 K - \widetilde{C}_2 \frac{mt \min\{1, t/4\}}{4C_4^2 \mu^2 \log^4 m} \right), \end{aligned} \tag{201}$$

where $\widetilde{C}_1, \widetilde{C}_2 > 0$ are some universal constants.

Now we are in a position to strengthen this bound to obtain uniform control of θ_1 over $\mathcal{H}$. Note that for any $\boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathcal{H}$,

$$\begin{aligned} |\theta_1(\boldsymbol{h}_1) - \theta_1(\boldsymbol{h}_2)| &\le \left\| \sum_{j=1}^m \left(|\boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{h}_1|^2 - |\boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{h}_2|^2 \right) \boldsymbol{a}_j \boldsymbol{a}_j^{\mathsf{H}} \right\| + \left| \|\boldsymbol{h}_1\|_2^2 - \|\boldsymbol{h}_2\|_2^2 \right| \\ &= \max_{1\le j\le m} \left| |\boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{h}_1|^2 - |\boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{h}_2|^2 \right| \left\| \sum_{j=1}^m \boldsymbol{a}_j \boldsymbol{a}_j^{\mathsf{H}} \right\| + \left| \|\boldsymbol{h}_1\|_2^2 - \|\boldsymbol{h}_2\|_2^2 \right|, \end{aligned}$$

where

$$\left| |\boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{h}_2|^2 - |\boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{h}_1|^2 \right| = \left| (\boldsymbol{h}_2 - \boldsymbol{h}_1)^{\mathsf{H}} \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{h}_2 + \boldsymbol{h}_1^{\mathsf{H}} \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} (\boldsymbol{h}_2 - \boldsymbol{h}_1) \right|$$

$$
\begin{aligned}
&\le 2\max\{\|\boldsymbol{h}_1\|_2, \|\boldsymbol{h}_2\|_2\}\|\boldsymbol{h}_2-\boldsymbol{h}_1\|_2\|\boldsymbol{b}_j\|_2^2 \\
&\le 4\|\boldsymbol{h}_2-\boldsymbol{h}_1\|_2\|\boldsymbol{b}_j\|_2^2 \le \frac{4K}{m}\|\boldsymbol{h}_2-\boldsymbol{h}_1\|_2
\end{aligned}
$$

and

$$
\left|\|\boldsymbol{h}_1\|_2^2-\|\boldsymbol{h}_2\|_2^2\right| = \left|\boldsymbol{h}_1^{\mathsf{H}}(\boldsymbol{h}_1-\boldsymbol{h}_2)-(\boldsymbol{h}_1-\boldsymbol{h}_2)^{\mathsf{H}}\boldsymbol{h}_2\right| \le 2\max\{\|\boldsymbol{h}_1\|_2, \|\boldsymbol{h}_2\|_2\}\|\boldsymbol{h}_2-\boldsymbol{h}_1\|_2 \le 4\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2.
$$

Define an event $\mathcal{E}_0 = \left\{\left\|\sum_{j=1}^m \boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\right\| \le 2m\right\}$. When $\mathcal{E}_0$ happens, the previous estimates give

$$
|\theta_1(\boldsymbol{h}_1)-\theta_1(\boldsymbol{h}_2)| \le (8K+4)\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2 \le 10K\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2, \qquad \forall \boldsymbol{h}_1,\boldsymbol{h}_2\in\mathcal{H}.
$$

Let $\varepsilon = 1/(1280K)$, and $\widetilde{\mathcal{H}}$ be an ε-net covering $\mathcal{H}$ (see [Ver12, Definition 5.1]). We have

$$
\left(\left\{\sup_{\boldsymbol{h}\in\widetilde{\mathcal{H}}}\theta_1(\boldsymbol{h}) \le \frac{1}{128}\right\}\cap\mathcal{E}_0\right) \subseteq \left\{\sup_{\boldsymbol{h}\in\mathcal{H}}\theta_1 \le \frac{1}{64}\right\}
$$

and as a result,

$$
\mathbb{P}\left(\sup_{\boldsymbol{h}\in\mathcal{H}}\theta_1(\boldsymbol{h}) \ge \frac{1}{64}\right) \le \mathbb{P}\left(\sup_{\boldsymbol{h}\in\widetilde{\mathcal{H}}}\theta_1(\boldsymbol{h}) \ge \frac{1}{128}\right) + \mathbb{P}(\mathcal{E}_0^c) \le |\widetilde{\mathcal{H}}|\cdot\max_{\boldsymbol{h}\in\widetilde{\mathcal{H}}}\mathbb{P}\left(\theta_1(\boldsymbol{h}) \ge \frac{1}{128}\right) + \mathbb{P}(\mathcal{E}_0^c).
$$

Lemma 57 forces that $\mathbb{P}(\mathcal{E}_0^c) = O(m^{-10})$. Additionally, we have $\log|\widetilde{\mathcal{H}}| \le \widetilde{C}_3 K\log K$ for some absolute constant $\widetilde{C}_3 > 0$ according to [Ver12, Lemma 5.2]. Hence (201) leads to

$$
\begin{aligned}
|\widetilde{\mathcal{H}}|\cdot\max_{\boldsymbol{h}\in\widetilde{\mathcal{H}}}\mathbb{P}\left(\theta_1(\boldsymbol{h}) \ge \frac{1}{128}\right) &\le 2\exp\left(\widetilde{C}_3K\log K + \widetilde{C}_1K - \widetilde{C}_2\frac{m(1/128)\min\{1,(1/128)/4\}}{4C_4^2\mu^2\log^4 m}\right) \\
&\le 2\exp\left(2\widetilde{C}_3K\log m - \frac{\widetilde{C}_4 m}{\mu^2\log^4 m}\right)
\end{aligned}
$$

for some constant $\widetilde{C}_4 > 0$. Under the sample complexity $m \gg \mu^2 K\log^5 m$, the right-hand side of the above display is at most $O\left(m^{-10}\right)$. Combine the estimates above to establish the desired high-probability bound for $\sup_{\boldsymbol{z}\in\mathcal{S}}\alpha_2$.

3. Next, we will demonstrate that

$$
\mathbb{P}(\sup_{\boldsymbol{z}\,\in\,\mathcal{S}}\alpha_3 \le 1/96) = 1 - O\left(m^{-10} + e^{-K}\log m\right).
$$

To this end, we let

$$
\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1^{\mathsf{H}} \\ \vdots \\ \boldsymbol{a}_m^{\mathsf{H}} \end{bmatrix} \in \mathbb{C}^{m\times K}, \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{b}_1^{\mathsf{H}} \\ \vdots \\ \boldsymbol{b}_m^{\mathsf{H}} \end{bmatrix} \in \mathbb{C}^{m\times K}, \quad \boldsymbol{C} = \begin{bmatrix} c_1(\boldsymbol{z}) & & & \\ & c_2(\boldsymbol{z}) & & \\ & & \cdots & \\ & & & c_m(\boldsymbol{z}) \end{bmatrix} \in \mathbb{C}^{m\times m},
$$

where for each $1 \le j \le m$,

$$
c_j(\boldsymbol{z}) := \boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}\boldsymbol{x}^{\mathsf{H}}\boldsymbol{a}_j - y_j = \boldsymbol{b}_j^{\mathsf{H}}(\boldsymbol{h}\boldsymbol{x}^{\mathsf{H}} - \boldsymbol{h}^\star\boldsymbol{x}^{\star\mathsf{H}})\boldsymbol{a}_j.
$$

As a consequence, we can write $\alpha_3 = \|\boldsymbol{B}^{\mathsf{H}}\boldsymbol{C}\boldsymbol{A}\|$.

The key observation is that both the ℓ_∞ norm and the Frobenius norm of $\boldsymbol{C}$ are well-controlled. Specifically, we claim for the moment that with probability at least $1 - O\left(m^{-10}\right)$,

$$
\|\boldsymbol{C}\|_\infty = \max_{1\le j\le m}|c_j| \le C\frac{\mu\log^{5/2} m}{\sqrt{m}}; \tag{202a}
$$

$$
\|\boldsymbol{C}\|_{\mathrm{F}}^{2}=\sum_{j=1}^{m}|c_{j}|^{2}\leq 12\delta^{2}, \tag{202b}
$$

where $C>0$ is some absolute constant. This motivates us to divide the entries in $\boldsymbol{C}$ into multiple groups based on their magnitudes.

To be precise, introduce $R:=1+\lceil\log_{2}(C\mu\log^{7/2}m)\rceil$ sets $\{\mathcal{I}_{r}\}_{1\leq r\leq R}$, where

$$
\mathcal{I}_{r}=\left\{j\in[m]:\ \frac{C\mu\log^{5/2}m}{2^{r}\sqrt{m}}<|c_{j}|\leq\frac{C\mu\log^{5/2}m}{2^{r-1}\sqrt{m}}\right\},\qquad 1\leq r\leq R-1
$$

and $\mathcal{I}_{R}=\{1,\cdots,m\}\setminus\left(\bigcup_{r=1}^{R-1}\mathcal{I}_{r}\right)$. An immediate consequence of the definition of $\mathcal{I}_{r}$ and the norm constraints in (202) is the following cardinality bound

$$
|\mathcal{I}_{r}|\leq\frac{\|\boldsymbol{C}\|_{\mathrm{F}}^{2}}{\min_{j\in\mathcal{I}_{r}}|c_{j}|^{2}}\leq\frac{12\delta^{2}}{\left(\frac{C\mu\log^{5/2}m}{2^{r}\sqrt{m}}\right)^{2}}=\underbrace{\frac{12\delta^{2}4^{r}}{C^{2}\mu^{2}\log^{5}m}}_{\delta_{r}}m \tag{203}
$$

for $1\leq r\leq R-1$. Since $\{\mathcal{I}_{r}\}_{1\leq r\leq R}$ form a partition of the index set $\{1,\cdots,m\}$, it is easy to see that

$$
\boldsymbol{B}^{\mathsf{H}}\boldsymbol{C}\boldsymbol{A}=\sum_{r=1}^{R}(\boldsymbol{B}_{\mathcal{I}_{r},\cdot})^{\mathsf{H}}\boldsymbol{C}_{\mathcal{I}_{r},\mathcal{I}_{r}}\boldsymbol{A}_{\mathcal{I}_{r},\cdot},
$$

where $\boldsymbol{D}_{\mathcal{I},\mathcal{J}}$ denotes the submatrix of $\boldsymbol{D}$ induced by the rows and columns of $\boldsymbol{D}$ having indices from $\mathcal{I}$ and $\mathcal{J}$, respectively, and $\boldsymbol{D}_{\mathcal{I},\cdot}$ refers to the submatrix formed by the rows from the index set $\mathcal{I}$. As a result, one can invoke the triangle inequality to derive

$$
\alpha_{3}\leq\sum_{r=1}^{R-1}\|\boldsymbol{B}_{\mathcal{I}_{r},\cdot}\|\cdot\|\boldsymbol{C}_{\mathcal{I}_{r},\mathcal{I}_{r}}\|\cdot\|\boldsymbol{A}_{\mathcal{I}_{r},\cdot}\|+\|\boldsymbol{B}_{\mathcal{I}_{R},\cdot}\|\cdot\|\boldsymbol{C}_{\mathcal{I}_{R},\mathcal{I}_{R}}\|\cdot\|\boldsymbol{A}_{\mathcal{I}_{R},\cdot}\|. \tag{204}
$$

Recognizing that $\boldsymbol{B}^{\mathsf{H}}\boldsymbol{B}=\boldsymbol{I}_{K}$, we obtain

$$
\|\boldsymbol{B}_{\mathcal{I}_{r},\cdot}\|\leq\|\boldsymbol{B}\|=1
$$

for every $1\leq r\leq R$. In addition, by construction of $\mathcal{I}_{r}$, we have

$$
\|\boldsymbol{C}_{\mathcal{I}_{r},\mathcal{I}_{r}}\|=\max_{j\in\mathcal{I}_{r}}|c_{j}|\leq\frac{C\mu\log^{5/2}m}{2^{r-1}\sqrt{m}}
$$

for $1\leq r\leq R$, and specifically for R, one has

$$
\|\boldsymbol{C}_{\mathcal{I}_{R},\mathcal{I}_{R}}\|=\max_{j\in\mathcal{I}_{R}}|c_{j}|\leq\frac{C\mu\log^{5/2}m}{2^{R-1}\sqrt{m}}\leq\frac{1}{\sqrt{m}\log m},
$$

which follows from the definition of R, i.e. $R=1+\lceil\log_{2}(C\mu\log^{7/2}m)\rceil$. Regarding $\|\boldsymbol{A}_{\mathcal{I}_{r},\cdot}\|$, we discover that $\|\boldsymbol{A}_{I_{R},\cdot}\|\leq\|\boldsymbol{A}\|$ and in view of (203),

$$
\|\boldsymbol{A}_{\mathcal{I}_{r},\cdot}\|\leq\sup_{\mathcal{I}:|\mathcal{I}|\leq\delta_{r}m}\|\boldsymbol{A}_{\mathcal{I},\cdot}\|,\qquad 1\leq r\leq R-1.
$$

Substitute the above estimates into (204) to get

$$
\alpha_{3}\leq\sum_{r=1}^{R-1}\frac{C\mu\log^{5/2}m}{2^{r-1}\sqrt{m}}\sup_{\mathcal{I}:|\mathcal{I}|\leq\delta_{r}m}\|\boldsymbol{A}_{\mathcal{I},\cdot}\|+\frac{\|\boldsymbol{A}\|}{\sqrt{m}\log m}. \tag{205}
$$

It remains to upper bound $\|\boldsymbol{A}\|$ and $\sup_{\mathcal{I}:|\mathcal{I}|\leq\delta_r m}\|\boldsymbol{A}_{\mathcal{I},\cdot}\|$. Lemma 57 tells us that $\|\boldsymbol{A}\|\leq 2\sqrt{m}$ with probability at least $1-O\left(m^{-10}\right)$. Furthermore, we can invoke Lemma 58 to bound $\sup_{\mathcal{I}:|\mathcal{I}|\leq\delta_r m}\|\boldsymbol{A}_{\mathcal{I},\cdot}\|$ for each $1\leq r\leq R-1$. It is easily seen from our assumptions $m\gg\mu^2 K\log^9 m$ and $\delta=c/\log^2 m$ that $\delta_r\gg K/m$. In addition,

$$\delta_r\leq\frac{12\delta^2 4^{R-1}}{C^2\mu^2\log^5 m}\leq\frac{12\delta^2 4^{1+\log_2(C\mu\log^{7/2}m)}}{C^2\mu^2\log^5 m}=48\delta^2\log^2 m=\frac{48c}{\log^2 m}\ll 1.$$

By Lemma 58 we obtain that for some constants $\widetilde{C}_2,\widetilde{C}_3>0$

$$\begin{aligned}\mathbb{P}\left(\sup_{\mathcal{I}:|\mathcal{I}|\leq\delta_r m}\|\boldsymbol{A}_{\mathcal{I},\cdot}\|\geq\sqrt{4\widetilde{C}_3\delta_r m\log(e/\delta_r)}\right)&\leq 2\exp\left(-\frac{\widetilde{C}_2\widetilde{C}_3}{3}\delta_r m\log(e/\delta_r)\right)\\&\leq 2\exp\left(-\frac{\widetilde{C}_2\widetilde{C}_3}{3}\delta_r m\right)\leq 2e^{-K}.\end{aligned}$$

Taking the union bound and substituting the estimates above into (205), we see that with probability at least $1-O\left(m^{-10}\right)-O\left((R-1)e^{-K}\right)$,

$$\begin{aligned}\alpha_3&\leq\sum_{r=1}^{R-1}\frac{C\mu\log^{5/2}m}{2^{r-1}\sqrt{m}}\cdot\sqrt{4\widetilde{C}_3\delta_r m\log(e/\delta_r)}+\frac{2\sqrt{m}}{\sqrt{m}\log m}\\&\leq\sum_{r=1}^{R-1}4\delta\sqrt{12\widetilde{C}_3\log(e/\delta_r)}+\frac{2}{\log m}\\&\lesssim(R-1)\delta\sqrt{\log(e/\delta_1)}+\frac{1}{\log m}.\end{aligned}$$

Note that $\mu\leq\sqrt{m}$, $R-1=\lceil\log_2(C\mu\log^{7/2}m)\rceil\lesssim\log m$, and

$$\sqrt{\log\frac{e}{\delta_1}}=\sqrt{\log\left(\frac{eC^2\mu^2\log^5 m}{48\delta^2}\right)}\lesssim\log m.$$

Therefore, with probability exceeding $1-O\left(m^{-10}\right)-O\left(e^{-K}\log m\right)$,

$$\sup_{\boldsymbol{z}\in\mathcal{S}}\alpha_3\lesssim\delta\log^2 m+\frac{1}{\log m}.$$

By taking c to be small enough in $\delta=c/\log^2 m$, we get

$$\mathbb{P}\left(\sup_{\boldsymbol{z}\in\mathcal{S}}\alpha_3\geq 1/96\right)\leq O\left(m^{-10}\right)+O\left(e^{-K}\log m\right)$$

as claimed.

Finally, it remains to justify (202). For all $\boldsymbol{z}\in\mathcal{S}$, the triangle inequality tells us that

$$\begin{aligned}|c_j|&\leq\left|\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}(\boldsymbol{x}-\boldsymbol{x}^\star)^{\mathsf{H}}\boldsymbol{a}_j\right|+\left|\boldsymbol{b}_j^{\mathsf{H}}(\boldsymbol{h}-\boldsymbol{h}^\star)\boldsymbol{x}^{\star\mathsf{H}}\boldsymbol{a}_j\right|\\&\leq\left|\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}\right|\cdot\left|\boldsymbol{a}_j^{\mathsf{H}}(\boldsymbol{x}-\boldsymbol{x}^\star)\right|+\left(\left|\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}\right|+\left|\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}^\star\right|\right)\cdot\left|\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^\star\right|\\&\leq\frac{2C_4\mu\log^2 m}{\sqrt{m}}\cdot\frac{2C_3}{\log^{3/2}m}+\left(\frac{2C_4\mu\log^2 m}{\sqrt{m}}+\frac{\mu}{\sqrt{m}}\right)5\sqrt{\log m}\\&\leq C\frac{\mu\log^{5/2}m}{\sqrt{m}},\end{aligned}$$

for some large constant $C > 0$, where we have used the definition of $\mathcal{S}$ and the fact (189). The claim (202b) follows directly from [LLSW18, Lemma 5.14]. To avoid confusion, we use μ_1 to refer to the parameter μ therein. Let $L = m$, $N = K$, $d_0 = 1$, $\mu_1 = C_4 \mu \log^2 m/2$, and $\varepsilon = 1/15$. Then

$$\mathcal{S} \subseteq \mathcal{N}_{d_0} \cap \mathcal{N}_{\mu_1} \cap \mathcal{N}_{\varepsilon},$$

and the sample complexity condition $L \gg \mu_1^2 (K + N) \log^2 L$ is satisfied because we have assumed $m \gg \mu^2 K \log^6 m$. Therefore with probability exceeding $1 - O\left(m^{-10} + e^{-K}\right)$, we obtain that for all $\boldsymbol{z} \in \mathcal{S}$,

$$\|\boldsymbol{C}\|_{\mathrm{F}}^2 \le \frac{5}{4} \left\|\boldsymbol{h}\boldsymbol{x}^{\mathsf{H}} - \boldsymbol{h}^\star \boldsymbol{x}^{\star\mathsf{H}}\right\|_{\mathrm{F}}^2.$$

The claim (202b) can then be justified by observing that

$$\left\|\boldsymbol{h}\boldsymbol{x}^{\mathsf{H}} - \boldsymbol{h}^\star \boldsymbol{x}^{\star\mathsf{H}}\right\|_{\mathrm{F}} = \left\|\boldsymbol{h}\left(\boldsymbol{x} - \boldsymbol{x}^\star\right)^{\mathsf{H}} + \left(\boldsymbol{h} - \boldsymbol{h}^\star\right)\boldsymbol{x}^{\star\mathsf{H}}\right\|_{\mathrm{F}} \le \|\boldsymbol{h}\|_2 \|\boldsymbol{x} - \boldsymbol{x}^\star\|_2 + \|\boldsymbol{h} - \boldsymbol{h}^\star\|_2 \|\boldsymbol{x}^\star\|_2 \le 3\delta.$$

4. It remains to control α_4, for which we make note of the following inequality

$$\alpha_4 \le \underbrace{\left\|\sum_{j=1}^m \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} (\boldsymbol{h}\boldsymbol{x}^\top - \boldsymbol{h}^\star \boldsymbol{x}^{\star\top}) \overline{\boldsymbol{a}_j}\, \overline{\boldsymbol{a}_j}^{\mathsf{H}}\right\|}_{\theta_3} + \underbrace{\left\|\sum_{j=1}^m \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{h}^\star \boldsymbol{x}^{\star\top} (\overline{\boldsymbol{a}_j}\, \overline{\boldsymbol{a}_j}^{\mathsf{H}} - \boldsymbol{I}_K)\right\|}_{\theta_4}$$

with $\overline{\boldsymbol{a}_j}$ denoting the entrywise conjugate of $\boldsymbol{a}_j$. Since $\{\overline{\boldsymbol{a}_j}\}$ has the same joint distribution as $\{\boldsymbol{a}_j\}$, by the same argument used for bounding α_3 we obtain control of the first term, namely,

$$\mathbb{P}\left(\sup_{\boldsymbol{z} \in \mathcal{S}} \theta_3 \ge 1/96\right) = O(m^{-10} + e^{-K} \log m).$$

Note that $m \gg \mu^2 K \log m / \delta^2$ and $\delta \ll 1$. According to [LLSW18, Lemma 5.20],

$$\mathbb{P}\left(\sup_{\boldsymbol{z} \in \mathcal{S}} \theta_4 \ge 1/96\right) \le \mathbb{P}\left(\sup_{\boldsymbol{z} \in \mathcal{S}} \theta_4 \ge \delta\right) = O(m^{-10}).$$

Putting together the above bounds, we reach $\mathbb{P}(\sup_{\boldsymbol{z} \in \mathcal{S}} \alpha_4 \le 1/48) = 1 - O(m^{-10} + e^{-K} \log m)$.

5. Combining all the previous bounds for $\sup_{\boldsymbol{z} \in \mathcal{S}} \alpha_j$ and (196), we deduce that with probability $1 - O(m^{-10} + e^{-K} \log m)$,

$$\left\|\nabla^2 f\left(\boldsymbol{z}\right) - \nabla^2 F\left(\boldsymbol{z}^\star\right)\right\| \le 2 \cdot \frac{1}{32} + 2 \cdot \frac{1}{32} + 4 \cdot \frac{1}{96} + 4 \cdot \frac{1}{48} = \frac{1}{4}.$$

C.2 Proofs of Lemma 15 and Lemma 16

Proof of Lemma 15. In view of the definition of α^{t+1} (see (38)), one has

$$\operatorname{dist}\left(\boldsymbol{z}^{t+1}, \boldsymbol{z}^\star\right)^2 = \left\|\frac{1}{\overline{\alpha^{t+1}}} \boldsymbol{h}^{t+1} - \boldsymbol{h}^\star\right\|_2^2 + \left\|\alpha^{t+1} \boldsymbol{x}^{t+1} - \boldsymbol{x}^\star\right\|_2^2 \le \left\|\frac{1}{\overline{\alpha^t}} \boldsymbol{h}^{t+1} - \boldsymbol{h}^\star\right\|_2^2 + \left\|\alpha^t \boldsymbol{x}^{t+1} - \boldsymbol{x}^\star\right\|_2^2.$$

The gradient update rules (79) imply that

$$\begin{aligned}
\frac{1}{\overline{\alpha^t}} \boldsymbol{h}^{t+1} &= \frac{1}{\overline{\alpha^t}} \left(\boldsymbol{h}^t - \frac{\eta}{\|\boldsymbol{x}^t\|_2^2} \nabla_{\boldsymbol{h}} f\left(\boldsymbol{z}^t\right)\right) = \widetilde{\boldsymbol{h}}^t - \frac{\eta}{\|\widetilde{\boldsymbol{x}}^t\|_2^2} \nabla_{\boldsymbol{h}} f\left(\widetilde{\boldsymbol{z}}^t\right), \\
\alpha^t \boldsymbol{x}^{t+1} &= \alpha^t \left(\boldsymbol{x}^t - \frac{\eta}{\|\boldsymbol{h}^t\|_2^2} \nabla_{\boldsymbol{x}} f\left(\boldsymbol{z}^t\right)\right) = \widetilde{\boldsymbol{x}}^t - \frac{\eta}{\|\widetilde{\boldsymbol{h}}^t\|_2^2} \nabla_{\boldsymbol{x}} f\left(\widetilde{\boldsymbol{z}}^t\right),
\end{aligned}$$

where we denote $\widetilde{\boldsymbol{h}}^t = \frac{1}{\overline{\alpha^t}}\boldsymbol{h}^t$ and $\widetilde{\boldsymbol{x}}^t = \alpha^t \boldsymbol{x}^t$ as in (81). Let $\widehat{\boldsymbol{h}}^{t+1} = \frac{1}{\overline{\alpha^t}}\boldsymbol{h}^{t+1}$ and $\widehat{\boldsymbol{x}}^{t+1} = \alpha^t \boldsymbol{x}^{t+1}$. We further get

$$\begin{bmatrix} \widehat{\boldsymbol{h}}^{t+1} - \boldsymbol{h}^\star \\ \widehat{\boldsymbol{x}}^{t+1} - \boldsymbol{x}^\star \\ \overline{\widehat{\boldsymbol{h}}^{t+1} - \boldsymbol{h}^\star} \\ \overline{\widehat{\boldsymbol{x}}^{t+1} - \boldsymbol{x}^\star} \end{bmatrix} = \begin{bmatrix} \widetilde{\boldsymbol{h}}^{t} - \boldsymbol{h}^\star \\ \widetilde{\boldsymbol{x}}^{t} - \boldsymbol{x}^\star \\ \overline{\widetilde{\boldsymbol{h}}^{t} - \boldsymbol{h}^\star} \\ \overline{\widetilde{\boldsymbol{x}}^{t} - \boldsymbol{x}^\star} \end{bmatrix} - \eta \underbrace{\begin{bmatrix} \|\widetilde{\boldsymbol{x}}^t\|_2^{-2}\boldsymbol{I}_K & & & \\ & \|\widetilde{\boldsymbol{h}}^t\|_2^{-2}\boldsymbol{I}_K & & \\ & & \|\widetilde{\boldsymbol{x}}^t\|_2^{-2}\boldsymbol{I}_K & \\ & & & \|\widetilde{\boldsymbol{h}}^t\|_2^{-2}\boldsymbol{I}_K \end{bmatrix}}_{:=\boldsymbol{D}} \begin{bmatrix} \nabla_{\boldsymbol{h}} f(\widetilde{\boldsymbol{z}}^t) \\ \nabla_{\boldsymbol{x}} f(\widetilde{\boldsymbol{z}}^t) \\ \overline{\nabla_{\boldsymbol{h}} f(\widetilde{\boldsymbol{z}}^t)} \\ \overline{\nabla_{\boldsymbol{x}} f(\widetilde{\boldsymbol{z}}^t)} \end{bmatrix}. \tag{206}$$

The fundamental theorem of calculus (see Appendix D.3.1) together with the fact that $\nabla f(\boldsymbol{z}^\star) = \boldsymbol{0}$ tells us

$$\begin{bmatrix} \nabla_{\boldsymbol{h}} f(\widetilde{\boldsymbol{z}}^t) \\ \nabla_{\boldsymbol{x}} f(\widetilde{\boldsymbol{z}}^t) \\ \overline{\nabla_{\boldsymbol{h}} f(\widetilde{\boldsymbol{z}}^t)} \\ \overline{\nabla_{\boldsymbol{x}} f(\widetilde{\boldsymbol{z}}^t)} \end{bmatrix} = \begin{bmatrix} \nabla_{\boldsymbol{h}} f(\widetilde{\boldsymbol{z}}^t) - \nabla_{\boldsymbol{h}} f(\boldsymbol{z}^\star) \\ \nabla_{\boldsymbol{x}} f(\widetilde{\boldsymbol{z}}^t) - \nabla_{\boldsymbol{x}} f(\boldsymbol{z}^\star) \\ \overline{\nabla_{\boldsymbol{h}} f(\widetilde{\boldsymbol{z}}^t) - \nabla_{\boldsymbol{h}} f(\boldsymbol{z}^\star)} \\ \overline{\nabla_{\boldsymbol{x}} f(\widetilde{\boldsymbol{z}}^t) - \nabla_{\boldsymbol{x}} f(\boldsymbol{z}^\star)} \end{bmatrix} = \underbrace{\int_0^1 \nabla^2 f(\boldsymbol{z}(\tau))\,\mathrm{d}\tau}_{:=\boldsymbol{A}} \begin{bmatrix} \widetilde{\boldsymbol{h}}^{t} - \boldsymbol{h}^\star \\ \widetilde{\boldsymbol{x}}^{t} - \boldsymbol{x}^\star \\ \overline{\widetilde{\boldsymbol{h}}^{t} - \boldsymbol{h}^\star} \\ \overline{\widetilde{\boldsymbol{x}}^{t} - \boldsymbol{x}^\star} \end{bmatrix}, \tag{207}$$

where we denote $\boldsymbol{z}(\tau) := \boldsymbol{z}^\star + \tau(\widetilde{\boldsymbol{z}}^t - \boldsymbol{z}^\star)$ and $\nabla^2 f$ is the Wirtinger Hessian. To further simplify notation, denote $\widehat{\boldsymbol{z}}^{t+1} = \begin{bmatrix} \widehat{\boldsymbol{h}}^{t+1} \\ \widehat{\boldsymbol{x}}^{t+1} \end{bmatrix}$. The identity (207) allows us to rewrite (206) as

$$\begin{bmatrix} \widehat{\boldsymbol{z}}^{t+1} - \boldsymbol{z}^\star \\ \overline{\widehat{\boldsymbol{z}}^{t+1} - \boldsymbol{z}^\star} \end{bmatrix} = (\boldsymbol{I} - \eta \boldsymbol{D}\boldsymbol{A}) \begin{bmatrix} \widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star \\ \overline{\widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star} \end{bmatrix}. \tag{208}$$

Take the squared Euclidean norm of both sides of (208) to reach

$$\begin{aligned} \left\|\widehat{\boldsymbol{z}}^{t+1} - \boldsymbol{z}^\star\right\|_2^2 &= \frac{1}{2} \begin{bmatrix} \widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star \\ \overline{\widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star} \end{bmatrix}^{\mathsf{H}} (\boldsymbol{I} - \eta \boldsymbol{D}\boldsymbol{A})^{\mathsf{H}} (\boldsymbol{I} - \eta \boldsymbol{D}\boldsymbol{A}) \begin{bmatrix} \widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star \\ \overline{\widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star} \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} \widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star \\ \overline{\widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star} \end{bmatrix}^{\mathsf{H}} \left(\boldsymbol{I} + \eta^2 \boldsymbol{A}\boldsymbol{D}^2\boldsymbol{A} - \eta(\boldsymbol{D}\boldsymbol{A} + \boldsymbol{A}\boldsymbol{D})\right) \begin{bmatrix} \widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star \\ \overline{\widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star} \end{bmatrix} \\ &\leq (1 + \eta^2\|\boldsymbol{A}\|^2\|\boldsymbol{D}\|^2)\left\|\widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star\right\|_2^2 - \frac{\eta}{2} \begin{bmatrix} \widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star \\ \overline{\widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star} \end{bmatrix}^{\mathsf{H}} (\boldsymbol{D}\boldsymbol{A} + \boldsymbol{A}\boldsymbol{D}) \begin{bmatrix} \widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star \\ \overline{\widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star} \end{bmatrix}. \end{aligned} \tag{209}$$

Since $\boldsymbol{z}(\tau)$ lies between $\widetilde{\boldsymbol{z}}^t$ and $\boldsymbol{z}^\star$, we conclude from the assumptions (85) that for all $0 \leq \tau \leq 1$,

$$\max\left\{\|\boldsymbol{h}(\tau) - \boldsymbol{h}^\star\|_2, \|\boldsymbol{x}(\tau) - \boldsymbol{x}^\star\|_2\right\} \leq \mathrm{dist}\left(\boldsymbol{z}^t, \boldsymbol{z}^\star\right) \leq \xi \leq \delta;$$
$$\max_{1\leq j\leq m}\left|\boldsymbol{a}_j^{\mathsf{H}}(\boldsymbol{x}(\tau) - \boldsymbol{x}^\star)\right| \leq C_3 \frac{1}{\log^{3/2} m};$$
$$\max_{1\leq j\leq m}\left|\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}(\tau)\right| \leq C_4 \frac{\mu}{\sqrt{m}}\log^2 m$$

for $\xi > 0$ sufficiently small. Moreover, it is straightforward to see that

$$\gamma_1 := \left\|\widetilde{\boldsymbol{x}}^t\right\|_2^{-2} \qquad \text{and} \qquad \gamma_2 := \left\|\widetilde{\boldsymbol{h}}^t\right\|_2^{-2}$$

satisfy

$$\max\left\{|\gamma_1 - 1|, |\gamma_2 - 1|\right\} \lesssim \max\left\{\left\|\widetilde{\boldsymbol{h}}^t - \boldsymbol{h}^\star\right\|_2, \left\|\widetilde{\boldsymbol{x}}^t - \boldsymbol{x}^\star\right\|_2\right\} \leq \delta$$

as long as $\xi > 0$ is sufficiently small. We can now readily invoke Lemma 14 to arrive at

$$\|\boldsymbol{A}\|\,\|\boldsymbol{D}\| \leq 3(1+\delta) \leq 4 \qquad \text{and}$$

$$\begin{bmatrix} \widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star \\ \overline{\widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star} \end{bmatrix}^{\mathsf{H}} (\boldsymbol{D}\boldsymbol{A} + \boldsymbol{A}\boldsymbol{D}) \begin{bmatrix} \widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star \\ \overline{\widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star} \end{bmatrix} \geq \frac{1}{4}\left\|\begin{bmatrix} \widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star \\ \overline{\widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star} \end{bmatrix}\right\|_2^2 = \frac{1}{2}\left\|\widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^\star\right\|_2^2.$$

Substitution into (209) indicates that

$$\left\|\widehat{\boldsymbol{z}}^{t+1}-\boldsymbol{z}^{\star}\right\|_2^2 \leq \left(1+16\eta^2-\eta/4\right)\left\|\widetilde{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\star}\right\|_2^2.$$

When $0<\eta\leq 1/128$, this implies that

$$\left\|\widehat{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\star}\right\|_2^2 \leq (1-\eta/8)\left\|\widetilde{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\star}\right\|_2^2,$$

and hence

$$\left\|\widetilde{\boldsymbol{z}}^{t+1}-\boldsymbol{z}^{\star}\right\|_2 \leq \left\|\widehat{\boldsymbol{z}}^{t+1}-\boldsymbol{z}^{\star}\right\|_2 \leq (1-\eta/8)^{1/2}\left\|\widetilde{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\star}\right\|_2 \leq (1-\eta/16)\,\mathrm{dist}(\boldsymbol{z}^{t},\boldsymbol{z}^{\star}). \tag{210}$$

This completes the proof of Lemma 15. □

Proof of Lemma 16. Reuse the notation in this subsection, namely, $\widehat{\boldsymbol{z}}^{t+1}=\begin{bmatrix}\widehat{\boldsymbol{h}}^{t+1}\\ \widehat{\boldsymbol{x}}^{t+1}\end{bmatrix}$ with $\widehat{\boldsymbol{h}}^{t+1}=\frac{1}{\overline{\alpha^{t}}}\boldsymbol{h}^{t+1}$ and $\widehat{\boldsymbol{x}}^{t+1}=\alpha^{t}\boldsymbol{x}^{t+1}$. From (210), one can tell that

$$\left\|\widetilde{\boldsymbol{z}}^{t+1}-\boldsymbol{z}^{\star}\right\|_2 \leq \left\|\widehat{\boldsymbol{z}}^{t+1}-\boldsymbol{z}^{\star}\right\|_2 \leq \mathrm{dist}(\boldsymbol{z}^{t},\boldsymbol{z}^{\star}).$$

Invoke Lemma 52 with $\beta=\alpha^{t}$ to get

$$\left|\alpha^{t+1}-\alpha^{t}\right| \lesssim \left\|\widehat{\boldsymbol{z}}^{t+1}-\boldsymbol{z}^{\star}\right\|_2 \leq \mathrm{dist}(\boldsymbol{z}^{t},\boldsymbol{z}^{\star}).$$

This combined with the assumption $||\alpha^{t}|-1|\leq 1/2$ implies that

$$\left|\alpha^{t}\right| \geq \frac{1}{2} \qquad \text{and} \qquad \left|\frac{\alpha^{t+1}}{\alpha^{t}}-1\right| = \left|\frac{\alpha^{t+1}-\alpha^{t}}{\alpha^{t}}\right| \lesssim \mathrm{dist}(\boldsymbol{z}^{t},\boldsymbol{z}^{\star}) \lesssim C_1\frac{1}{\log^2 m}.$$

This finishes the proof of the first claim.

The second claim can be proved by induction. Suppose that $\left||\alpha^{s}|-1\right|\leq 1/2$ and $\mathrm{dist}(\boldsymbol{z}^{s},\boldsymbol{z}^{\star})\leq C_1(1-\eta/16)^{s}/\log^2 m$ hold for all $0\leq s\leq\tau\leq t$, then using our result in the first part gives

$$\begin{aligned}\left||\alpha^{\tau+1}|-1\right| &\leq \left||\alpha^{0}|-1\right| + \sum_{s=0}^{\tau}\left|\alpha^{s+1}-\alpha^{s}\right| \leq \frac{1}{4}+c\sum_{s=0}^{\tau}\mathrm{dist}(\boldsymbol{z}^{s},\boldsymbol{z}^{\star})\\ &\leq \frac{1}{4}+\frac{cC_1}{\frac{\eta}{16}\log^2 m} \leq \frac{1}{2}\end{aligned}$$

for m sufficiently large. The proof is then complete by induction. □

C.3 Proof of Lemma 17

Define the alignment parameter between $\boldsymbol{z}^{t,(l)}$ and $\widetilde{\boldsymbol{z}}^{t}$ as

$$\alpha_{\mathrm{mutual}}^{t,(l)} := \underset{\alpha\in\mathbb{C}}{\mathrm{argmin}} \left\|\frac{1}{\overline{\alpha}}\boldsymbol{h}^{t,(l)}-\frac{1}{\overline{\alpha^{t}}}\boldsymbol{h}^{t}\right\|_2^2 + \left\|\alpha\boldsymbol{x}^{t,(l)}-\alpha^{t}\boldsymbol{x}^{t}\right\|_2^2.$$

Further denote, for simplicity of presentation, $\widehat{\boldsymbol{z}}^{t,(l)}=\begin{bmatrix}\widehat{\boldsymbol{h}}^{t,(l)}\\ \widehat{\boldsymbol{x}}^{t,(l)}\end{bmatrix}$ with

$$\widehat{\boldsymbol{h}}^{t,(l)} := \frac{1}{\overline{\alpha_{\mathrm{mutual}}^{t,(l)}}}\boldsymbol{h}^{t,(l)} \qquad \text{and} \qquad \widehat{\boldsymbol{x}}^{t,(l)} := \alpha_{\mathrm{mutual}}^{t,(l)}\boldsymbol{x}^{t,(l)}.$$

Clearly, $\widehat{\boldsymbol{z}}^{t,(l)}$ is aligned with $\widetilde{\boldsymbol{z}}^{t}$.

Armed with the above notation, we have

$$\mathrm{dist}\big(\boldsymbol{z}^{t+1,(l)},\widetilde{\boldsymbol{z}}^{t+1}\big) = \min_{\alpha}\sqrt{\left\|\frac{1}{\overline{\alpha}}\boldsymbol{h}^{t+1,(l)}-\frac{1}{\overline{\alpha^{t+1}}}\boldsymbol{h}^{t+1}\right\|_2^2 + \left\|\alpha\boldsymbol{x}^{t+1,(l)}-\alpha^{t+1}\boldsymbol{x}^{t+1}\right\|_2^2}$$

$$= \min_{\alpha} \sqrt{\left\| \left(\frac{\overline{\alpha^{t}}}{\overline{\alpha^{t+1}}} \right) \left(\frac{1}{\overline{\alpha}} \frac{\overline{\alpha^{t+1}}}{\overline{\alpha^{t}}} \boldsymbol{h}^{t+1,(l)} - \frac{1}{\overline{\alpha^{t}}} \boldsymbol{h}^{t+1} \right) \right\|_2^2 + \left\| \left(\frac{\alpha^{t+1}}{\alpha^{t}} \right) \left(\alpha \frac{\alpha^{t}}{\alpha^{t+1}} \boldsymbol{x}^{t+1,(l)} - \alpha^{t} \boldsymbol{x}^{t+1} \right) \right\|_2^2}$$

$$\leq \sqrt{\left\| \left(\frac{\overline{\alpha^{t}}}{\overline{\alpha^{t+1}}} \right) \left(\frac{1}{\overline{\alpha_{\text{mutual}}^{t,(l)}}} \boldsymbol{h}^{t+1,(l)} - \frac{1}{\overline{\alpha^{t}}} \boldsymbol{h}^{t+1} \right) \right\|_2^2 + \left\| \left(\frac{\alpha^{t+1}}{\alpha^{t}} \right) \left(\alpha_{\text{mutual}}^{t,(l)} \boldsymbol{x}^{t+1,(l)} - \alpha^{t} \boldsymbol{x}^{t+1} \right) \right\|_2^2} \tag{211}$$

$$\leq \max\left\{ \left| \frac{\alpha^{t+1}}{\alpha^{t}} \right|, \left| \frac{\alpha^{t}}{\alpha^{t+1}} \right| \right\} \left\| \begin{bmatrix} \frac{1}{\overline{\alpha_{\text{mutual}}^{t,(l)}}} \boldsymbol{h}^{t+1,(l)} - \frac{1}{\overline{\alpha^{t}}} \boldsymbol{h}^{t+1} \\ \alpha_{\text{mutual}}^{t,(l)} \boldsymbol{x}^{t+1,(l)} - \alpha^{t} \boldsymbol{x}^{t+1} \end{bmatrix} \right\|_2, \tag{212}$$

where (211) follows by taking $\alpha = \frac{\alpha^{t+1}}{\alpha^{t}} \alpha_{\text{mutual}}^{t,(l)}$. The latter bound is more convenient to work with when controlling the gap between $\boldsymbol{z}^{t,(l)}$ and $\boldsymbol{z}^{t}$.

We can then apply the gradient update rules (79) and (89) to get

$$\begin{bmatrix} \frac{1}{\overline{\alpha_{\text{mutual}}^{t,(l)}}} \boldsymbol{h}^{t+1,(l)} - \frac{1}{\overline{\alpha^{t}}} \boldsymbol{h}^{t+1} \\ \alpha_{\text{mutual}}^{t,(l)} \boldsymbol{x}^{t+1,(l)} - \alpha^{t} \boldsymbol{x}^{t+1} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{\overline{\alpha_{\text{mutual}}^{t,(l)}}} \left(\boldsymbol{h}^{t,(l)} - \frac{\eta}{\|\boldsymbol{x}^{t,(l)}\|_2^2} \nabla_{\boldsymbol{h}} f^{(l)}\left(\boldsymbol{h}^{t,(l)}, \boldsymbol{x}^{t,(l)}\right) \right) - \frac{1}{\overline{\alpha^{t}}} \left(\boldsymbol{h}^{t} - \frac{\eta}{\|\boldsymbol{x}^{t}\|_2^2} \nabla_{\boldsymbol{h}} f\left(\boldsymbol{h}^{t}, \boldsymbol{x}^{t}\right) \right) \\ \alpha_{\text{mutual}}^{t,(l)} \left(\boldsymbol{x}^{t,(l)} - \frac{\eta}{\|\boldsymbol{h}^{t,(l)}\|_2^2} \nabla_{\boldsymbol{x}} f^{(l)}\left(\boldsymbol{h}^{t,(l)}, \boldsymbol{x}^{t,(l)}\right) \right) - \alpha^{t} \left(\boldsymbol{x}^{t} - \frac{\eta}{\|\boldsymbol{h}^{t}\|_2^2} \nabla_{\boldsymbol{x}} f\left(\boldsymbol{h}^{t}, \boldsymbol{x}^{t}\right) \right) \end{bmatrix}$$
$$= \begin{bmatrix} \widehat{\boldsymbol{h}}^{t,(l)} - \frac{\eta}{\|\widehat{\boldsymbol{x}}^{t,(l)}\|_2^2} \nabla_{\boldsymbol{h}} f^{(l)}\big(\widehat{\boldsymbol{h}}^{t,(l)}, \widehat{\boldsymbol{x}}^{t,(l)}\big) - \left(\widetilde{\boldsymbol{h}}^{t} - \frac{\eta}{\|\widetilde{\boldsymbol{x}}^{t}\|_2^2} \nabla_{\boldsymbol{h}} f\big(\widetilde{\boldsymbol{h}}^{t}, \widetilde{\boldsymbol{x}}^{t}\big) \right) \\ \widehat{\boldsymbol{x}}^{t,(l)} - \frac{\eta}{\|\widehat{\boldsymbol{h}}^{t,(l)}\|_2^2} \nabla_{\boldsymbol{x}} f^{(l)}\big(\widehat{\boldsymbol{h}}^{t,(l)}, \widehat{\boldsymbol{x}}^{t,(l)}\big) - \left(\widetilde{\boldsymbol{x}}^{t} - \frac{\eta}{\|\widetilde{\boldsymbol{h}}^{t}\|_2^2} \nabla_{\boldsymbol{x}} f\big(\widetilde{\boldsymbol{h}}^{t}, \widetilde{\boldsymbol{x}}^{t}\big) \right) \end{bmatrix}.$$

By construction, we can write the leave-one-out gradients as

$$\nabla_{\boldsymbol{h}} f^{(l)}(\boldsymbol{h}, \boldsymbol{x}) = \nabla_{\boldsymbol{h}} f(\boldsymbol{h}, \boldsymbol{x}) - \left(\boldsymbol{b}_l^{\mathsf{H}} \boldsymbol{h} \boldsymbol{x}^{\mathsf{H}} \boldsymbol{a}_l - y_l \right) \boldsymbol{b}_l \boldsymbol{a}_l^{\mathsf{H}} \boldsymbol{x} \qquad \text{and}$$
$$\nabla_{\boldsymbol{x}} f^{(l)}(\boldsymbol{h}, \boldsymbol{x}) = \nabla_{\boldsymbol{h}} f(\boldsymbol{h}, \boldsymbol{x}) - \overline{(\boldsymbol{b}_l^{\mathsf{H}} \boldsymbol{h} \boldsymbol{x}^{\mathsf{H}} \boldsymbol{a}_l - y_l)} \boldsymbol{a}_l \boldsymbol{b}_l^{\mathsf{H}} \boldsymbol{h},$$

which allow us to continue the derivation and obtain

$$\begin{bmatrix} \frac{1}{\overline{\alpha_{\text{mutual}}^{t,(l)}}} \boldsymbol{h}^{t+1,(l)} - \frac{1}{\overline{\alpha^{t}}} \boldsymbol{h}^{t+1} \\ \alpha_{\text{mutual}}^{t,(l)} \boldsymbol{x}^{t+1,(l)} - \alpha^{t} \boldsymbol{x}^{t+1} \end{bmatrix} = \begin{bmatrix} \widehat{\boldsymbol{h}}^{t,(l)} - \frac{\eta}{\|\widehat{\boldsymbol{x}}^{t,(l)}\|_2^2} \nabla_{\boldsymbol{h}} f\big(\widehat{\boldsymbol{h}}^{t,(l)}, \widehat{\boldsymbol{x}}^{t,(l)}\big) - \left(\widetilde{\boldsymbol{h}}^{t} - \frac{\eta}{\|\widetilde{\boldsymbol{x}}^{t}\|_2^2} \nabla_{\boldsymbol{h}} f\big(\widetilde{\boldsymbol{h}}^{t}, \widetilde{\boldsymbol{x}}^{t}\big) \right) \\ \widehat{\boldsymbol{x}}^{t,(l)} - \frac{\eta}{\|\widehat{\boldsymbol{h}}^{t,(l)}\|_2^2} \nabla_{\boldsymbol{x}} f\big(\widehat{\boldsymbol{h}}^{t,(l)}, \widehat{\boldsymbol{x}}^{t,(l)}\big) - \left(\widetilde{\boldsymbol{x}}^{t} - \frac{\eta}{\|\widetilde{\boldsymbol{h}}^{t}\|_2^2} \nabla_{\boldsymbol{x}} f\big(\widetilde{\boldsymbol{h}}^{t}, \widetilde{\boldsymbol{x}}^{t}\big) \right) \end{bmatrix}$$
$$- \eta \underbrace{\begin{bmatrix} \frac{1}{\|\widehat{\boldsymbol{x}}^{t,(l)}\|_2^2} \left(\boldsymbol{b}_l^{\mathsf{H}} \widehat{\boldsymbol{h}}^{t,(l)} \widehat{\boldsymbol{x}}^{t,(l)\mathsf{H}} \boldsymbol{a}_l - y_l \right) \boldsymbol{b}_l \boldsymbol{a}_l^{\mathsf{H}} \widehat{\boldsymbol{x}}^{t,(l)} \\ \frac{1}{\|\widehat{\boldsymbol{h}}^{t,(l)}\|_2^2} \overline{\left(\boldsymbol{b}_l^{\mathsf{H}} \widehat{\boldsymbol{h}}^{t,(l)} \widehat{\boldsymbol{x}}^{t,(l)\mathsf{H}} \boldsymbol{a}_l - y_l \right)} \boldsymbol{a}_l \boldsymbol{b}_l^{\mathsf{H}} \widehat{\boldsymbol{h}}^{t,(l)} \end{bmatrix}}_{:= \boldsymbol{J}_3}.$$

This further gives

$$\begin{bmatrix} \frac{1}{\overline{\alpha_{\text{mutual}}^{t,(l)}}} \boldsymbol{h}^{t+1,(l)} - \frac{1}{\overline{\alpha^{t}}} \boldsymbol{h}^{t+1} \\ \alpha_{\text{mutual}}^{t,(l)} \boldsymbol{x}^{t+1,(l)} - \alpha^{t} \boldsymbol{x}^{t+1} \end{bmatrix} = \underbrace{\begin{bmatrix} \widehat{\boldsymbol{h}}^{t,(l)} - \frac{\eta}{\|\widehat{\boldsymbol{x}}^{t,(l)}\|_2^2} \nabla_{\boldsymbol{h}} f\left(\widehat{\boldsymbol{h}}^{t,(l)}, \widehat{\boldsymbol{x}}^{t,(l)}\right) - \left(\widetilde{\boldsymbol{h}}^{t} - \frac{\eta}{\|\widehat{\boldsymbol{x}}^{t,(l)}\|_2^2} \nabla_{\boldsymbol{h}} f\left(\widetilde{\boldsymbol{h}}^{t}, \widetilde{\boldsymbol{x}}^{t}\right) \right) \\ \widehat{\boldsymbol{x}}^{t,(l)} - \frac{\eta}{\|\widehat{\boldsymbol{h}}^{t,(l)}\|_2^2} \nabla_{\boldsymbol{x}} f\left(\widehat{\boldsymbol{h}}^{t,(l)}, \widehat{\boldsymbol{x}}^{t,(l)}\right) - \left(\widetilde{\boldsymbol{x}}^{t} - \frac{\eta}{\|\widehat{\boldsymbol{h}}^{t,(l)}\|_2^2} \nabla_{\boldsymbol{x}} f\left(\widetilde{\boldsymbol{h}}^{t}, \widetilde{\boldsymbol{x}}^{t}\right) \right) \end{bmatrix}}_{:= \boldsymbol{\nu}_1}$$
$$+ \eta \underbrace{\begin{bmatrix} \left(\frac{1}{\|\widetilde{\boldsymbol{x}}^{t}\|_2^2} - \frac{1}{\|\widehat{\boldsymbol{x}}^{t,(l)}\|_2^2} \right) \nabla_{\boldsymbol{h}} f\left(\widetilde{\boldsymbol{h}}^{t}, \widetilde{\boldsymbol{x}}^{t}\right) \\ \left(\frac{1}{\|\widetilde{\boldsymbol{h}}^{t}\|_2^2} - \frac{1}{\|\widehat{\boldsymbol{h}}^{t,(l)}\|_2^2} \right) \nabla_{\boldsymbol{x}} f\left(\widetilde{\boldsymbol{h}}^{t}, \widetilde{\boldsymbol{x}}^{t}\right) \end{bmatrix}}_{:= \boldsymbol{\nu}_2} - \eta \boldsymbol{\nu}_3. \tag{213}$$

In what follows, we bound the three terms $\boldsymbol{\nu}_1$, $\boldsymbol{\nu}_2$, and $\boldsymbol{\nu}_3$ separately.

1. Regarding the first term $\boldsymbol{\nu}_1$, one can adopt the same strategy as in Appendix C.2. Specifically, write

$$\left[\begin{array}{c} \widehat{\boldsymbol{h}}^{t,(l)} - \frac{\eta}{\left\|\widehat{\boldsymbol{x}}^{t,(l)}\right\|_2^2} \nabla_{\boldsymbol{h}} f\left(\widehat{\boldsymbol{z}}^{t,(l)}\right) - \left(\widetilde{\boldsymbol{h}}^{t} - \frac{\eta}{\left\|\widehat{\boldsymbol{x}}^{t,(l)}\right\|_2^2} \nabla_{\boldsymbol{h}} f\left(\widetilde{\boldsymbol{z}}^{t}\right)\right) \\ \widehat{\boldsymbol{x}}^{t,(l)} - \frac{\eta}{\left\|\widehat{\boldsymbol{h}}^{t,(l)}\right\|_2^2} \nabla_{\boldsymbol{x}} f\left(\widehat{\boldsymbol{z}}^{t,(l)}\right) - \left(\widetilde{\boldsymbol{x}}^{t} - \frac{\eta}{\left\|\widehat{\boldsymbol{h}}^{t,(l)}\right\|_2^2} \nabla_{\boldsymbol{x}} f\left(\widetilde{\boldsymbol{z}}^{t}\right)\right) \\ \hline \overline{\widehat{\boldsymbol{h}}^{t,(l)} - \frac{\eta}{\|\widehat{\boldsymbol{x}}^{t,(l)}\|_2^2} \nabla_{\boldsymbol{h}} f\left(\widehat{\boldsymbol{z}}^{t,(l)}\right) - \left(\widetilde{\boldsymbol{h}}^{t} - \frac{\eta}{\|\widehat{\boldsymbol{x}}^{t,(l)}\|_2^2} \nabla_{\boldsymbol{h}} f\left(\widetilde{\boldsymbol{z}}^{t}\right)\right)} \\ \hline \overline{\widehat{\boldsymbol{x}}^{t,(l)} - \frac{\eta}{\|\widehat{\boldsymbol{h}}^{t,(l)}\|_2^2} \nabla_{\boldsymbol{x}} f\left(\widehat{\boldsymbol{z}}^{t,(l)}\right) - \left(\widetilde{\boldsymbol{x}}^{t} - \frac{\eta}{\|\widehat{\boldsymbol{h}}^{t,(l)}\|_2^2} \nabla_{\boldsymbol{x}} f\left(\widetilde{\boldsymbol{z}}^{t}\right)\right)} \end{array}\right] = \left[\begin{array}{c} \widehat{\boldsymbol{h}}^{t,(l)} - \widetilde{\boldsymbol{h}}^{t} \\ \widehat{\boldsymbol{x}}^{t,(l)} - \widetilde{\boldsymbol{x}}^{t} \\ \hline \overline{\widehat{\boldsymbol{h}}^{t,(l)} - \widetilde{\boldsymbol{h}}^{t}} \\ \hline \overline{\widehat{\boldsymbol{x}}^{t,(l)} - \widetilde{\boldsymbol{x}}^{t}} \end{array}\right]$$

$$- \eta \underbrace{\left[\begin{array}{cccc} \left\|\widehat{\boldsymbol{x}}^{t,(l)}\right\|_2^{-2} \boldsymbol{I}_K & & & \\ & \left\|\widehat{\boldsymbol{h}}^{t,(l)}\right\|_2^{-2} \boldsymbol{I}_K & & \\ & & \left\|\widehat{\boldsymbol{x}}^{t,(l)}\right\|_2^{-2} \boldsymbol{I}_K & \\ & & & \left\|\widehat{\boldsymbol{h}}^{t,(l)}\right\|_2^{-2} \boldsymbol{I}_K \end{array}\right]}_{:=\boldsymbol{D}} \left[\begin{array}{c} \nabla_{\boldsymbol{h}} f\left(\widehat{\boldsymbol{z}}^{t,(l)}\right) - \nabla_{\boldsymbol{h}} f\left(\widetilde{\boldsymbol{z}}^{t}\right) \\ \nabla_{\boldsymbol{x}} f\left(\widehat{\boldsymbol{z}}^{t,(l)}\right) - \nabla_{\boldsymbol{x}} f\left(\widetilde{\boldsymbol{z}}^{t}\right) \\ \hline \overline{\nabla_{\boldsymbol{h}} f\left(\widehat{\boldsymbol{z}}^{t,(l)}\right) - \nabla_{\boldsymbol{h}} f\left(\widetilde{\boldsymbol{z}}^{t}\right)} \\ \hline \overline{\nabla_{\boldsymbol{x}} f\left(\widehat{\boldsymbol{z}}^{t,(l)}\right) - \nabla_{\boldsymbol{x}} f\left(\widetilde{\boldsymbol{z}}^{t}\right)} \end{array}\right].$$

The fundamental theorem of calculus (see Appendix D.3.1) reveals that

$$\left[\begin{array}{c} \nabla_{\boldsymbol{h}} f\left(\widehat{\boldsymbol{z}}^{t,(l)}\right) - \nabla_{\boldsymbol{h}} f\left(\widetilde{\boldsymbol{z}}^{t}\right) \\ \nabla_{\boldsymbol{x}} f\left(\widehat{\boldsymbol{z}}^{t,(l)}\right) - \nabla_{\boldsymbol{x}} f\left(\widetilde{\boldsymbol{z}}^{t}\right) \\ \hline \overline{\nabla_{\boldsymbol{h}} f\left(\widehat{\boldsymbol{z}}^{t,(l)}\right) - \nabla_{\boldsymbol{h}} f\left(\widetilde{\boldsymbol{z}}^{t}\right)} \\ \hline \overline{\nabla_{\boldsymbol{x}} f\left(\widehat{\boldsymbol{z}}^{t,(l)}\right) - \nabla_{\boldsymbol{x}} f\left(\widetilde{\boldsymbol{z}}^{t}\right)} \end{array}\right] = \underbrace{\int_0^1 \nabla^2 f\left(\boldsymbol{z}\left(\tau\right)\right) \mathrm{d}\tau}_{:=\boldsymbol{A}} \left[\begin{array}{c} \widehat{\boldsymbol{h}}^{t,(l)} - \widetilde{\boldsymbol{h}}^{t} \\ \widehat{\boldsymbol{x}}^{t,(l)} - \widetilde{\boldsymbol{x}}^{t} \\ \hline \overline{\widehat{\boldsymbol{h}}^{t,(l)} - \widetilde{\boldsymbol{h}}^{t}} \\ \hline \overline{\widehat{\boldsymbol{x}}^{t,(l)} - \widetilde{\boldsymbol{x}}^{t}} \end{array}\right],$$

where we abuse the notation and denote $\boldsymbol{z}\left(\tau\right) = \widetilde{\boldsymbol{z}}^{t} + \tau\left(\widehat{\boldsymbol{z}}^{t,(l)} - \widetilde{\boldsymbol{z}}^{t}\right)$. In order to invoke Lemma 14, we need to verify the conditions required therein. Recall the induction hypothesis (90b) that

$$\operatorname{dist}\left(\boldsymbol{z}^{t,(l)}, \widetilde{\boldsymbol{z}}^{t}\right) = \left\|\widehat{\boldsymbol{z}}^{t,(l)} - \widetilde{\boldsymbol{z}}^{t}\right\|_2 \le C_2 \frac{\mu}{\sqrt{m}} \sqrt{\frac{\mu^2 K \log^9 m}{m}},$$

and the fact that $\boldsymbol{z}\left(\tau\right)$ lies between $\widehat{\boldsymbol{z}}^{t,(l)}$ and $\widetilde{\boldsymbol{z}}^{t}$. For all $0 \le \tau \le 1$:

(a) If $m \gg \mu^2 \sqrt{K} \log^{13/2} m$, then

$$\begin{aligned} \left\|\boldsymbol{z}\left(\tau\right) - \boldsymbol{z}^{\star}\right\|_2 &\le \max\left\{\left\|\widehat{\boldsymbol{z}}^{t,(l)} - \boldsymbol{z}^{\star}\right\|_2, \left\|\widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^{\star}\right\|_2\right\} \le \left\|\widetilde{\boldsymbol{z}}^{t} - \boldsymbol{z}^{\star}\right\|_2 + \left\|\widehat{\boldsymbol{z}}^{t,(l)} - \widetilde{\boldsymbol{z}}^{t}\right\|_2 \\ &\le C_1 \frac{1}{\log^2 m} + C_2 \frac{\mu}{\sqrt{m}} \sqrt{\frac{\mu^2 K \log^9 m}{m}} \le 2C_1 \frac{1}{\log^2 m}, \end{aligned}$$

where we have used the induction hypotheses (90a) and (90b);

(b) If $m \gg \mu^2 K \log^6 m$, then

$$\begin{aligned} \max_{1\le j\le m} \left|\boldsymbol{a}_j^{\mathsf{H}}\left(\boldsymbol{x}\left(\tau\right) - \boldsymbol{x}^{\star}\right)\right| &= \max_{1\le j\le m} \left|\tau \boldsymbol{a}_j^{\mathsf{H}}\left(\widehat{\boldsymbol{x}}^{t,(l)} - \widetilde{\boldsymbol{x}}^{t}\right) + \boldsymbol{a}_j^{\mathsf{H}}\left(\widetilde{\boldsymbol{x}}^{t} - \boldsymbol{x}^{\star}\right)\right| \\ &\le \max_{1\le j\le m} \left|\boldsymbol{a}_j^{\mathsf{H}}\left(\widehat{\boldsymbol{x}}^{t,(l)} - \widetilde{\boldsymbol{x}}^{t}\right)\right| + \max_{1\le j\le m} \left|\boldsymbol{a}_j^{\mathsf{H}}\left(\widetilde{\boldsymbol{x}}^{t} - \boldsymbol{x}^{\star}\right)\right| \\ &\le \max_{1\le j\le m} \left\|\boldsymbol{a}_j\right\|_2 \left\|\widehat{\boldsymbol{z}}^{t,(l)} - \widetilde{\boldsymbol{z}}^{t}\right\|_2 + C_3 \frac{1}{\log^{3/2} m} \\ &\le 3\sqrt{K} \cdot C_2 \frac{\mu}{\sqrt{m}} \sqrt{\frac{\mu^2 K \log^9 m}{m}} + C_3 \frac{1}{\log^{3/2} m} \le 2C_3 \frac{1}{\log^{3/2} m}, \end{aligned} \tag{214}$$

which follows from the bound (190) and the induction hypotheses (90b) and (90c);

(c) If $m \gg \mu K \log^{5/2} m$, then

$$\begin{aligned}
\max_{1\le j\le m}\left|\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}(\tau)\right| &= \max_{1\le j\le m}\left|\tau\boldsymbol{b}_j^{\mathsf{H}}\big(\widehat{\boldsymbol{h}}^{t,(l)}-\widetilde{\boldsymbol{h}}^t\big)+\boldsymbol{b}_j^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\right| \\
&\le \max_{1\le j\le m}\left|\boldsymbol{b}_j^{\mathsf{H}}\big(\widehat{\boldsymbol{h}}^{t,(l)}-\widetilde{\boldsymbol{h}}^t\big)\right| + \max_{1\le j\le m}\left|\boldsymbol{b}_j^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\right| \\
&\le \max_{1\le j\le m}\|\boldsymbol{b}_j\|_2\left\|\widehat{\boldsymbol{h}}^{t,(l)}-\widetilde{\boldsymbol{h}}^t\right\|_2 + \max_{1\le j\le m}\left|\boldsymbol{b}_j^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\right| \\
&\le \sqrt{\frac{K}{m}}\cdot C_2\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^2K\log^9 m}{m}} + C_4\frac{\mu}{\sqrt{m}}\log^2 m \le 2C_4\frac{\mu}{\sqrt{m}}\log^2 m, \qquad (215)
\end{aligned}$$

which makes use of the fact $\|\boldsymbol{b}_j\|_2=\sqrt{K/m}$ as well as the induction hypotheses (90b) and (90d).

These properties satisfy the condition (82) required in Lemma 14. The other two conditions (83) and (84) are also straightforward to check and hence we omit it. Thus, we can repeat the argument used in Appendix C.2 to obtain

$$\|\boldsymbol{\nu}_1\|_2 \le (1-\eta/16)\cdot\left\|\widehat{\boldsymbol{z}}^{t,(l)}-\widetilde{\boldsymbol{z}}^t\right\|_2.$$

2. In terms of the second term $\boldsymbol{\nu}_2$, it is easily seen that

$$\|\boldsymbol{\nu}_2\|_2 \le \max\left\{\left|\frac{1}{\left\|\widetilde{\boldsymbol{x}}^t\right\|_2^2}-\frac{1}{\left\|\widehat{\boldsymbol{x}}^{t,(l)}\right\|_2^2}\right|,\left|\frac{1}{\left\|\widetilde{\boldsymbol{h}}^t\right\|_2^2}-\frac{1}{\left\|\widehat{\boldsymbol{h}}^{t,(l)}\right\|_2^2}\right|\right\}\left\|\begin{bmatrix}\nabla_{\boldsymbol{h}}f(\widetilde{\boldsymbol{z}}^t)\\ \nabla_{\boldsymbol{x}}f(\widetilde{\boldsymbol{z}}^t)\end{bmatrix}\right\|_2.$$

We first note that the upper bound on $\|\nabla^2 f(\cdot)\|$ (which essentially provides a Lipschitz constant on the gradient) in Lemma 14 forces

$$\left\|\begin{bmatrix}\nabla_{\boldsymbol{h}}f(\widetilde{\boldsymbol{z}}^t)\\ \nabla_{\boldsymbol{x}}f(\widetilde{\boldsymbol{z}}^t)\end{bmatrix}\right\|_2 = \left\|\begin{bmatrix}\nabla_{\boldsymbol{h}}f(\widetilde{\boldsymbol{z}}^t)-\nabla_{\boldsymbol{h}}f(\boldsymbol{z}^\star)\\ \nabla_{\boldsymbol{x}}f(\widetilde{\boldsymbol{z}}^t)-\nabla_{\boldsymbol{x}}f(\boldsymbol{z}^\star)\end{bmatrix}\right\|_2 \lesssim \left\|\widetilde{\boldsymbol{z}}^t-\boldsymbol{z}^\star\right\|_2 \lesssim C_1\frac{1}{\log^2 m},$$

where the first identity follows since $\nabla_{\boldsymbol{h}}f(\boldsymbol{z}^\star)=\boldsymbol{0}$, and the last inequality comes from the induction hypothesis (90a). Additionally, recognizing that $\|\widetilde{\boldsymbol{x}}^t\|_2 \asymp \left\|\widehat{\boldsymbol{x}}^{t,(l)}\right\|_2 \asymp 1$, one can easily verify that

$$\left|\frac{1}{\left\|\widetilde{\boldsymbol{x}}^t\right\|_2^2}-\frac{1}{\left\|\widehat{\boldsymbol{x}}^{t,(l)}\right\|_2^2}\right| = \left|\frac{\left\|\widehat{\boldsymbol{x}}^{t,(l)}\right\|_2^2-\left\|\widetilde{\boldsymbol{x}}^t\right\|_2^2}{\left\|\widetilde{\boldsymbol{x}}^t\right\|_2^2\cdot\left\|\widehat{\boldsymbol{x}}^{t,(l)}\right\|_2^2}\right| \lesssim \left|\left\|\widehat{\boldsymbol{x}}^{t,(l)}\right\|_2-\left\|\widetilde{\boldsymbol{x}}^t\right\|_2\right| \lesssim \left\|\widehat{\boldsymbol{x}}^{t,(l)}-\widetilde{\boldsymbol{x}}^t\right\|_2.$$

A similar bound holds for the other term involving $\boldsymbol{h}$. Combining the estimates above thus yields

$$\|\boldsymbol{\nu}_2\|_2 \lesssim C_1\frac{1}{\log^2 m}\left\|\widehat{\boldsymbol{z}}^{t,(l)}-\widetilde{\boldsymbol{z}}^t\right\|_2.$$

3. When it comes to the last term $\boldsymbol{\nu}_3$, one first sees that

$$\left\|\left(\boldsymbol{b}_l^{\mathsf{H}}\widehat{\boldsymbol{h}}^{t,(l)}\widehat{\boldsymbol{x}}^{t,(l)\mathsf{H}}\boldsymbol{a}_l-y_l\right)\boldsymbol{b}_l\boldsymbol{a}_l^{\mathsf{H}}\widehat{\boldsymbol{x}}^{t,(l)}\right\|_2 \le \left|\boldsymbol{b}_l^{\mathsf{H}}\widehat{\boldsymbol{h}}^{t,(l)}\widehat{\boldsymbol{x}}^{t,(l)\mathsf{H}}\boldsymbol{a}_l-y_l\right|\|\boldsymbol{b}_l\|_2\left|\boldsymbol{a}_l^{\mathsf{H}}\widehat{\boldsymbol{x}}^{t,(l)}\right|. \qquad (216)$$

The bounds (189) and (214) taken collectively yield

$$\left|\boldsymbol{a}_l^{\mathsf{H}}\widehat{\boldsymbol{x}}^{t,(l)}\right| \le \left|\boldsymbol{a}_l^{\mathsf{H}}\boldsymbol{x}^\star\right| + \left|\boldsymbol{a}_l^{\mathsf{H}}\big(\widehat{\boldsymbol{x}}^{t,(l)}-\boldsymbol{x}^\star\big)\right| \lesssim \sqrt{\log m}+C_3\frac{1}{\log^{3/2} m} \asymp \sqrt{\log m}.$$

In addition, the same argument as in obtaining (215) tells us that

$$\left|\boldsymbol{b}_l^{\mathsf{H}}(\widehat{\boldsymbol{h}}^{t,(l)}-\boldsymbol{h}^\star)\right| \lesssim C_4\frac{\mu}{\sqrt{m}}\log^2 m.$$

Combine the previous two bounds to obtain

$$\left|\boldsymbol{b}_l^{\mathsf{H}}\widehat{\boldsymbol{h}}^{t,(l)}\widehat{\boldsymbol{x}}^{t,(l)\mathsf{H}}\boldsymbol{a}_l-y_l\right| \le \left|\boldsymbol{b}_l^{\mathsf{H}}\widehat{\boldsymbol{h}}^{t,(l)}(\widehat{\boldsymbol{x}}^{t,(l)}-\boldsymbol{x}^\star)^{\mathsf{H}}\boldsymbol{a}_l\right| + \left|\boldsymbol{b}_l^{\mathsf{H}}(\widehat{\boldsymbol{h}}^{t,(l)}-\boldsymbol{h}^\star)\boldsymbol{x}^{\star\mathsf{H}}\boldsymbol{a}_l\right|$$

$$
\begin{aligned}
&\le \left|\boldsymbol{b}_l^{\mathsf{H}}\widehat{\boldsymbol{h}}^{t,(l)}\right| \cdot \left|\boldsymbol{a}_l^{\mathsf{H}}(\widehat{\boldsymbol{x}}^{t,(l)} - \boldsymbol{x}^\star)\right| + \left|\boldsymbol{b}_l^{\mathsf{H}}(\widehat{\boldsymbol{h}}^{t,(l)} - \boldsymbol{h}^\star)\right| \cdot \left|\boldsymbol{a}_l^{\mathsf{H}}\boldsymbol{x}^\star\right| \\
&\le \left(\left|\boldsymbol{b}_l^{\mathsf{H}}(\widehat{\boldsymbol{h}}^{t,(l)} - \boldsymbol{h}^\star)\right| + \left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{h}^\star\right|\right) \cdot \left|\boldsymbol{a}_l^{\mathsf{H}}(\widehat{\boldsymbol{x}}^{t,(l)} - \boldsymbol{x}^\star)\right| + \left|\boldsymbol{b}_l^{\mathsf{H}}(\widehat{\boldsymbol{h}}^{t,(l)} - \boldsymbol{h}^\star)\right| \cdot \left|\boldsymbol{a}_l^{\mathsf{H}}\boldsymbol{x}^\star\right| \\
&\lesssim \left(C_4\mu\frac{\log^2 m}{\sqrt{m}} + \frac{\mu}{\sqrt{m}}\right) \cdot C_3\frac{1}{\log^{3/2} m} + C_4\mu\frac{\log^2 m}{\sqrt{m}} \cdot \sqrt{\log m} \lesssim C_4\mu\frac{\log^{5/2} m}{\sqrt{m}}.
\end{aligned}
$$

Substitution into (216) gives

$$
\left\|\left(\boldsymbol{b}_l^{\mathsf{H}}\widehat{\boldsymbol{h}}^{t,(l)}\widehat{\boldsymbol{x}}^{t,(l)\mathsf{H}}\boldsymbol{a}_l - y_l\right)\boldsymbol{b}_l\boldsymbol{a}_l^{\mathsf{H}}\widehat{\boldsymbol{x}}^{t,(l)}\right\|_2 \lesssim C_4\mu\frac{\log^{5/2} m}{\sqrt{m}} \cdot \sqrt{\frac{K}{m}} \cdot \sqrt{\log m}. \tag{217}
$$

Similarly, we can also derive

$$
\begin{aligned}
\left\|\overline{\left(\boldsymbol{b}_l^{\mathsf{H}}\widehat{\boldsymbol{h}}^{t,(l)}\widehat{\boldsymbol{x}}^{t,(l)\mathsf{H}}\boldsymbol{a}_l - y_l\right)}\boldsymbol{a}_l\boldsymbol{b}_l^{\mathsf{H}}\widehat{\boldsymbol{h}}^{t,(l)}\right\| &\le \left|\boldsymbol{b}_l^{\mathsf{H}}\widehat{\boldsymbol{h}}^{t,(l)}\widehat{\boldsymbol{x}}^{t,(l)\mathsf{H}}\boldsymbol{a}_l - y_l\right| \left\|\boldsymbol{a}_l\right\|_2 \left|\boldsymbol{b}_l^{\mathsf{H}}\widehat{\boldsymbol{h}}^{t,(l)}\right| \\
&\lesssim C_4\mu\frac{\log^{5/2} m}{\sqrt{m}} \cdot \sqrt{K} \cdot C_4\frac{\mu}{\sqrt{m}}\log^2 m
\end{aligned}
$$

Putting these bounds together indicates that

$$
\left\|\boldsymbol{\nu}_3\right\|_2 \lesssim (C_4)^2\,\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^2 K\log^9 m}{m}}.
$$

The above bounds taken together with (212) and (213) ensure the existence of a constant $C > 0$ such that

$$
\begin{aligned}
\operatorname{dist}\left(\boldsymbol{z}^{t+1,(l)}, \widetilde{\boldsymbol{z}}^{t+1}\right) &\le \max\left\{\left|\frac{\alpha^{t+1}}{\alpha^t}\right|, \left|\frac{\alpha^t}{\alpha^{t+1}}\right|\right\}\left\{\left(1 - \frac{\eta}{16} + CC_1\eta\frac{1}{\log^2 m}\right)\left\|\widehat{\boldsymbol{z}}^{t,(l)} - \widetilde{\boldsymbol{z}}^t\right\|_2 + C\left(C_4\right)^2\eta\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^2 K\log^9 m}{m}}\right\} \\
&\overset{\text{(i)}}{\le} \frac{1 - \eta/21}{1 - \eta/20}\left\{\left(1 - \frac{\eta}{20}\right)\left\|\widehat{\boldsymbol{z}}^{t,(l)} - \widetilde{\boldsymbol{z}}^t\right\|_2 + C\left(C_4\right)^2\eta\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^2 K\log^9 m}{m}}\right\} \\
&\le \left(1 - \frac{\eta}{21}\right)\left\|\widehat{\boldsymbol{z}}^{t,(l)} - \widetilde{\boldsymbol{z}}^t\right\|_2 + 2C\left(C_4\right)^2\eta\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^2 K\log^9 m}{m}} \\
&= \left(1 - \frac{\eta}{21}\right)\operatorname{dist}\left(\boldsymbol{z}^{t,(l)}, \widetilde{\boldsymbol{z}}^t\right) + 2C\left(C_4\right)^2\eta\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^2 K\log^9 m}{m}} \\
&\overset{\text{(ii)}}{\le} C_2\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^2 K\log^9 m}{m}}.
\end{aligned}
$$

Here, (i) holds as long as m is sufficiently large such that $CC_1 1/\log^2 m \ll 1$ and

$$
\max\left\{\left|\frac{\alpha^{t+1}}{\alpha^t}\right|, \left|\frac{\alpha^t}{\alpha^{t+1}}\right|\right\} < \frac{1 - \eta/21}{1 - \eta/20}, \tag{218}
$$

which is guaranteed by Lemma 16. The inequality (ii) arises from the induction hypothesis (90b) and taking $C_2 > 0$ is sufficiently large.

Finally we establish the second inequality claimed in the lemma. Take $(\boldsymbol{h}_1, \boldsymbol{x}_1) = (\widetilde{\boldsymbol{h}}^{t+1}, \widetilde{\boldsymbol{x}}^{t+1})$ and $(\boldsymbol{h}_2, \boldsymbol{x}_2) = (\widehat{\boldsymbol{h}}^{t+1,(l)}, \widehat{\boldsymbol{x}}^{t+1,(l)})$ in Lemma 55. Since both $(\boldsymbol{h}_1, \boldsymbol{x}_1)$ and $(\boldsymbol{h}_2, \boldsymbol{x}_2)$ are close enough to $(\boldsymbol{h}^\star, \boldsymbol{x}^\star)$, we deduce that

$$
\left\|\widetilde{\boldsymbol{z}}^{t+1,(l)} - \widetilde{\boldsymbol{z}}^{t+1}\right\|_2 \lesssim \left\|\widehat{\boldsymbol{z}}^{t+1,(l)} - \widetilde{\boldsymbol{z}}^{t+1}\right\|_2 \lesssim C_2\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^2 K\log^9 m}{m}}
$$

as claimed.

C.4 Proof of Lemma 18

Before going forward, we make note of the following inequality

$$\max_{1\le l\le m}\left|\boldsymbol{b}_l^{\mathsf{H}}\frac{1}{\overline{\alpha^{t+1}}}\boldsymbol{h}^{t+1}\right| \le \left|\frac{\alpha^t}{\alpha^{t+1}}\right|\max_{1\le l\le m}\left|\boldsymbol{b}_l^{\mathsf{H}}\frac{1}{\overline{\alpha^t}}\boldsymbol{h}^{t+1}\right| \le (1+\delta)\max_{1\le l\le m}\left|\boldsymbol{b}_l^{\mathsf{H}}\frac{1}{\overline{\alpha^t}}\boldsymbol{h}^{t+1}\right|$$

for some small $\delta \asymp \log^{-2} m$, where the last relation follows from Lemma 16 that

$$\left|\frac{\alpha^{t+1}}{\alpha^t}-1\right| \lesssim \frac{1}{\log^2 m} \le \delta$$

for m sufficiently large. In view of the above inequality, the focus of our subsequent analysis will be to control $\max_l \left|\boldsymbol{b}_l^{\mathsf{H}}\frac{1}{\overline{\alpha^t}}\boldsymbol{h}^{t+1}\right|$.

The gradient update rule for $\boldsymbol{h}^{t+1}$ (cf. (79a)) gives

$$\frac{1}{\overline{\alpha^t}}\boldsymbol{h}^{t+1} = \widetilde{\boldsymbol{h}}^t - \eta\xi\sum_{j=1}^m \boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}\left(\widetilde{\boldsymbol{h}}^t\widetilde{\boldsymbol{x}}^{t\mathsf{H}} - \boldsymbol{h}^\star\boldsymbol{x}^{\star\mathsf{H}}\right)\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\widetilde{\boldsymbol{x}}^t,$$

where $\widetilde{\boldsymbol{h}}^t = \frac{1}{\overline{\alpha^t}}\boldsymbol{h}^t$ and $\widetilde{\boldsymbol{x}}^t = \alpha^t\boldsymbol{x}^t$. Here and below, we denote $\xi = 1/\|\widetilde{\boldsymbol{x}}^t\|_2^2$ for notational convenience. The above formula can be further decomposed into the following terms

$$\begin{aligned}
\frac{1}{\overline{\alpha^t}}\boldsymbol{h}^{t+1} &= \widetilde{\boldsymbol{h}}^t - \eta\xi\sum_{j=1}^m \boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\left|\boldsymbol{a}_j^{\mathsf{H}}\widetilde{\boldsymbol{x}}^t\right|^2 + \eta\xi\sum_{j=1}^m \boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}^\star\boldsymbol{x}^{\star\mathsf{H}}\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\widetilde{\boldsymbol{x}}^t \\
&= \left(1-\eta\xi\left\|\boldsymbol{x}^\star\right\|_2^2\right)\widetilde{\boldsymbol{h}}^t - \underbrace{\eta\xi\sum_{j=1}^m \boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\left(\left|\boldsymbol{a}_j^{\mathsf{H}}\widetilde{\boldsymbol{x}}^t\right|^2 - \left|\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^\star\right|^2\right)}_{:=\boldsymbol{v}_1} \\
&\quad - \underbrace{\eta\xi\sum_{j=1}^m \boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\left(\left|\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^\star\right|^2 - \left\|\boldsymbol{x}^\star\right\|_2^2\right)}_{:=\boldsymbol{v}_2} + \underbrace{\eta\xi\sum_{j=1}^m \boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}^\star\boldsymbol{x}^{\star\mathsf{H}}\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\widetilde{\boldsymbol{x}}^t}_{:=\boldsymbol{v}_3},
\end{aligned}$$

where we use the fact that $\sum_{j=1}^m \boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}} = \boldsymbol{I}_K$. In the sequel, we shall control each term separately.

1. We start with $|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{v}_1|$ by making the observation that

$$\begin{aligned}
\frac{1}{\eta\xi}\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{v}_1\right| &= \left|\sum_{j=1}^m \boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\left[\boldsymbol{a}_j^{\mathsf{H}}\left(\widetilde{\boldsymbol{x}}^t-\boldsymbol{x}^\star\right)\left(\boldsymbol{a}_j^{\mathsf{H}}\widetilde{\boldsymbol{x}}^t\right)^{\mathsf{H}} + \boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^\star\left(\boldsymbol{a}_j^{\mathsf{H}}\left(\widetilde{\boldsymbol{x}}^t-\boldsymbol{x}^\star\right)\right)^{\mathsf{H}}\right]\right| \\
&\le \sum_{j=1}^m\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{b}_j\right|\left\{\max_{1\le j\le m}\left|\boldsymbol{b}_j^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\right|\right\}\left\{\max_{1\le j\le m}\left|\boldsymbol{a}_j^{\mathsf{H}}\left(\widetilde{\boldsymbol{x}}^t-\boldsymbol{x}^\star\right)\right|\left(\left|\boldsymbol{a}_j^{\mathsf{H}}\widetilde{\boldsymbol{x}}^t\right| + \left|\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^\star\right|\right)\right\}. \qquad (219)
\end{aligned}$$

Combining the induction hypothesis (90c) and the condition (189) yields

$$\max_{1\le j\le m}\left|\boldsymbol{a}_j^{\mathsf{H}}\widetilde{\boldsymbol{x}}^t\right| \le \max_{1\le j\le m}\left|\boldsymbol{a}_j^{\mathsf{H}}\left(\widetilde{\boldsymbol{x}}^t-\boldsymbol{x}^\star\right)\right| + \max_{1\le j\le m}\left|\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^\star\right| \le C_3\frac{1}{\log^{3/2} m} + 5\sqrt{\log m} \le 6\sqrt{\log m}$$

as long as m is sufficiently large. This further implies

$$\max_{1\le j\le m}\left|\boldsymbol{a}_j^{\mathsf{H}}\left(\widetilde{\boldsymbol{x}}^t-\boldsymbol{x}^\star\right)\right|\left(\left|\boldsymbol{a}_j^{\mathsf{H}}\widetilde{\boldsymbol{x}}^t\right| + \left|\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^\star\right|\right) \le C_3\frac{1}{\log^{3/2} m}\cdot 11\sqrt{\log m} \le 11C_3\frac{1}{\log m}.$$

Substituting it into (219) and taking Lemma 48, we arrive at

$$\frac{1}{\eta\xi}\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{v}_1\right| \lesssim \log m\cdot\left\{\max_{1\le j\le m}\left|\boldsymbol{b}_j^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\right|\right\}\cdot C_3\frac{1}{\log m} \lesssim C_3\max_{1\le j\le m}\left|\boldsymbol{b}_j^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\right| \le 0.1\max_{1\le j\le m}\left|\boldsymbol{b}_j^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\right|,$$

with the proviso that C_3 is sufficiently small.

2. We then move on to $|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{v}_3|$, which obeys

$$\frac{1}{\eta\xi}\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{v}_3\right| \leq \left|\sum_{j=1}^{m}\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}^{\star}\boldsymbol{x}^{\star\mathsf{H}}\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^{\star}\right| + \left|\sum_{j=1}^{m}\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}^{\star}\boldsymbol{x}^{\star\mathsf{H}}\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\left(\widetilde{\boldsymbol{x}}^t-\boldsymbol{x}^{\star}\right)\right|. \tag{220}$$

Regarding the first term, we have the following lemma, whose proof is given in Appendix C.4.1.

Lemma 28. *Suppose $m \geq CK\log^2 m$ for some sufficiently large constant $C>0$. Then with probability at least $1-O\left(m^{-10}\right)$, one has*

$$\left|\sum_{j=1}^{m}\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}^{\star}\boldsymbol{x}^{\star\mathsf{H}}\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^{\star} - \boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{h}^{\star}\right| \lesssim \frac{\mu}{\sqrt{m}}.$$

For the remaining term, we apply the same strategy as in bounding $|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{v}_1|$ to get

$$\begin{aligned}
\left|\sum_{j=1}^{m}\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}^{\star}\boldsymbol{x}^{\star\mathsf{H}}\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\left(\widetilde{\boldsymbol{x}}^t-\boldsymbol{x}^{\star}\right)\right| &\leq \sum_{j=1}^{m}\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{b}_j\right|\left\{\max_{1\leq j\leq m}\left|\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}^{\star}\right|\right\}\left\{\max_{1\leq j\leq m}\left|\boldsymbol{a}_j^{\mathsf{H}}\left(\widetilde{\boldsymbol{x}}^t-\boldsymbol{x}^{\star}\right)\right|\right\}\left\{\max_{1\leq j\leq m}\left|\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^{\star}\right|\right\} \\
&\leq 4\log m\cdot\frac{\mu}{\sqrt{m}}\cdot C_3\frac{1}{\log^{3/2}m}\cdot 5\sqrt{\log m} \\
&\lesssim C_3\frac{\mu}{\sqrt{m}},
\end{aligned}$$

where the second line follows from the incoherence (36), the induction hypothesis (90c), the condition (189) and Lemma 48. Combining the above three inequalities and the incoherence (36) yields

$$\frac{1}{\eta\xi}\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{v}_3\right| \lesssim \left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{h}^{\star}\right| + \frac{\mu}{\sqrt{m}} + C_3\frac{\mu}{\sqrt{m}} \lesssim \left(1+C_3\right)\frac{\mu}{\sqrt{m}}.$$

3. Finally, we need to control $\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{v}_2\right|$. For convenience of presentation, we will only bound $\left|\boldsymbol{b}_1^{\mathsf{H}}\boldsymbol{v}_2\right|$ in the sequel, but the argument easily extends to all other $\boldsymbol{b}_l$'s. The idea is to group $\{\boldsymbol{b}_j\}_{1\leq j\leq m}$ into bins each containing τ adjacent vectors, and to look at each bin separately. Here, $\tau \asymp \operatorname{poly}\log(m)$ is some integer to be specified later. For notational simplicity, we assume m/τ to be an integer, although all arguments continue to hold when m/τ is not an integer. For each $0\leq l\leq m-\tau$, the following summation over τ adjacent data obeys

$$\begin{aligned}
&\boldsymbol{b}_1^{\mathsf{H}}\sum_{j=1}^{\tau}\boldsymbol{b}_{l+j}\boldsymbol{b}_{l+j}^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\left(\left|\boldsymbol{a}_{l+j}^{\mathsf{H}}\boldsymbol{x}^{\star}\right|^2-\|\boldsymbol{x}^{\star}\|_2^2\right) \\
&= \boldsymbol{b}_1^{\mathsf{H}}\sum_{j=1}^{\tau}\boldsymbol{b}_{l+1}\boldsymbol{b}_{l+1}^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\left(\left|\boldsymbol{a}_{l+j}^{\mathsf{H}}\boldsymbol{x}^{\star}\right|^2-\|\boldsymbol{x}^{\star}\|_2^2\right) + \boldsymbol{b}_1^{\mathsf{H}}\sum_{j=1}^{\tau}\left(\boldsymbol{b}_{l+j}\boldsymbol{b}_{l+j}^{\mathsf{H}}-\boldsymbol{b}_{l+1}\boldsymbol{b}_{l+1}^{\mathsf{H}}\right)\widetilde{\boldsymbol{h}}^t\left(\left|\boldsymbol{a}_{l+j}^{\mathsf{H}}\boldsymbol{x}^{\star}\right|^2-\|\boldsymbol{x}^{\star}\|_2^2\right) \\
&= \left\{\sum_{j=1}^{\tau}\left(\left|\boldsymbol{a}_{l+j}^{\mathsf{H}}\boldsymbol{x}^{\star}\right|^2-\|\boldsymbol{x}^{\star}\|_2^2\right)\right\}\boldsymbol{b}_1^{\mathsf{H}}\boldsymbol{b}_{l+1}\boldsymbol{b}_{l+1}^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t + \boldsymbol{b}_1^{\mathsf{H}}\sum_{j=1}^{\tau}\left(\boldsymbol{b}_{l+j}-\boldsymbol{b}_{l+1}\right)\boldsymbol{b}_{l+j}^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\left(\left|\boldsymbol{a}_{l+j}^{\mathsf{H}}\boldsymbol{x}^{\star}\right|^2-\|\boldsymbol{x}^{\star}\|_2^2\right) \\
&\quad + \boldsymbol{b}_1^{\mathsf{H}}\sum_{j=1}^{\tau}\boldsymbol{b}_{l+1}\left(\boldsymbol{b}_{l+j}-\boldsymbol{b}_{l+1}\right)^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\left(\left|\boldsymbol{a}_{l+j}^{\mathsf{H}}\boldsymbol{x}^{\star}\right|^2-\|\boldsymbol{x}^{\star}\|_2^2\right).
\end{aligned} \tag{221}$$

We will now bound each term in (221) separately.

- Before bounding the first term in (221), we first bound the pre-factor $\left|\sum_{j=1}^{\tau}\left(|\boldsymbol{a}_{l+j}^{\mathsf{H}}\boldsymbol{x}^{\star}|^2-\|\boldsymbol{x}^{\star}\|_2^2\right)\right|$. Notably, the fluctuation of this quantity does not grow fast as it is the sum of i.i.d. random variables over a group of relatively large size, i.e. τ. Since $2\left|\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^\star\right|^2$ follows the χ_2^2 distribution, by standard concentration results (e.g. [RV13, Theorem 1.1]), with probability exceeding $1-O\left(m^{-10}\right)$,

$$\left|\sum_{j=1}^{\tau}\left(\left|\boldsymbol{a}_{l+j}^{\mathsf{H}}\boldsymbol{x}^\star\right|^2-\|\boldsymbol{x}^\star\|_2^2\right)\right|\lesssim\sqrt{\tau\log m}.$$

With this result in place, we can bound the first term in (221) as

$$\left|\left\{\sum_{j=1}^{\tau}\left(\left|\boldsymbol{a}_{l+j}^{\mathsf{H}}\boldsymbol{x}^\star\right|^2-\|\boldsymbol{x}^\star\|_2^2\right)\right\}\boldsymbol{b}_1^{\mathsf{H}}\boldsymbol{b}_{l+1}\boldsymbol{b}_{l+1}^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\right|\lesssim\sqrt{\tau\log m}\left|\boldsymbol{b}_1^{\mathsf{H}}\boldsymbol{b}_{l+1}\right|\max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\right|.$$

Taking the summation over all bins gives

$$\sum_{k=0}^{\frac{m}{\tau}-1}\left|\left\{\sum_{j=1}^{\tau}\left(\left|\boldsymbol{a}_{k\tau+j}^{\mathsf{H}}\boldsymbol{x}^\star\right|^2-\|\boldsymbol{x}^\star\|_2^2\right)\right\}\boldsymbol{b}_1^{\mathsf{H}}\boldsymbol{b}_{k\tau+1}\boldsymbol{b}_{k\tau+1}^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\right|\lesssim\sqrt{\tau\log m}\sum_{k=0}^{\frac{m}{\tau}-1}\left|\boldsymbol{b}_1^{\mathsf{H}}\boldsymbol{b}_{k\tau+1}\right|\max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\right|. \tag{222}$$

It is straightforward to see from the proof of Lemma 48 that

$$\sum_{k=0}^{\frac{m}{\tau}-1}\left|\boldsymbol{b}_1^{\mathsf{H}}\boldsymbol{b}_{k\tau+1}\right|=\|\boldsymbol{b}_1\|_2^2+\sum_{k=1}^{\frac{m}{\tau}-1}\left|\boldsymbol{b}_1^{\mathsf{H}}\boldsymbol{b}_{k\tau+1}\right|\leq\frac{K}{m}+O\left(\frac{\log m}{\tau}\right). \tag{223}$$

Substituting (223) into the previous inequality (222) gives

$$\begin{aligned}\sum_{k=0}^{\frac{m}{\tau}-1}\left|\left\{\sum_{j=1}^{\tau}\left(\left|\boldsymbol{a}_{k\tau+j}^{\mathsf{H}}\boldsymbol{x}^\star\right|^2-\|\boldsymbol{x}^\star\|_2^2\right)\right\}\boldsymbol{b}_1^{\mathsf{H}}\boldsymbol{b}_{k\tau+1}\boldsymbol{b}_{k\tau+1}^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\right|&\lesssim\left(\frac{K\sqrt{\tau\log m}}{m}+\sqrt{\frac{\log^3 m}{\tau}}\right)\max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\right|\\&\leq 0.1\max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\right|,\end{aligned}$$

as long as $m\gg K\sqrt{\tau\log m}$ and $\tau\gg\log^3 m$.

- The second term of (221) obeys

$$\begin{aligned}&\left|\boldsymbol{b}_1^{\mathsf{H}}\sum_{j=1}^{\tau}\left(\boldsymbol{b}_{l+j}-\boldsymbol{b}_{l+1}\right)\boldsymbol{b}_{l+j}^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\left(\left|\boldsymbol{a}_{l+j}^{\mathsf{H}}\boldsymbol{x}^\star\right|^2-\|\boldsymbol{x}^\star\|_2^2\right)\right|\\&\quad\leq\max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\right|\sqrt{\sum_{j=1}^{\tau}\left|\boldsymbol{b}_1^{\mathsf{H}}\left(\boldsymbol{b}_{l+j}-\boldsymbol{b}_{l+1}\right)\right|^2}\sqrt{\sum_{j=1}^{\tau}\left(\left|\boldsymbol{a}_{l+j}^{\mathsf{H}}\boldsymbol{x}^\star\right|^2-\|\boldsymbol{x}^\star\|_2^2\right)^2}\\&\quad\lesssim\sqrt{\tau}\max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\right|\sqrt{\sum_{j=1}^{\tau}\left|\boldsymbol{b}_1^{\mathsf{H}}\left(\boldsymbol{b}_{l+j}-\boldsymbol{b}_{l+1}\right)\right|^2},\end{aligned}$$

where the first inequality is due to Cauchy-Schwarz, and the second one holds because of the following lemma, whose proof can be found in Appendix C.4.2.

Lemma 29. *Suppose $\tau\geq C\log^4 m$ for some sufficiently large constant $C>0$. Then with probability exceeding $1-O\left(m^{-10}\right)$,*

$$\sum_{j=1}^{\tau}\left(\left|\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^\star\right|^2-\|\boldsymbol{x}^\star\|_2^2\right)^2\lesssim\tau.$$

With the above bound in mind, we can sum over all bins of size τ to obtain

$$\left|\boldsymbol{b}_1^{\mathsf{H}}\sum_{k=0}^{\frac{m}{\tau}-1}\sum_{j=1}^{\tau}\left(\boldsymbol{b}_{k\tau+j}-\boldsymbol{b}_{k\tau+1}\right)\boldsymbol{b}_{k\tau+j}^{\mathsf{H}}\widetilde{\boldsymbol{h}}^t\left\{\left|\boldsymbol{a}_{l+j}^{\mathsf{H}}\boldsymbol{x}^\star\right|^2-\|\boldsymbol{x}^\star\|_2^2\right\}\right|$$

$$\lesssim \left\{ \sqrt{\tau} \sum_{k=0}^{\frac{m}{\tau}-1} \sqrt{\sum_{j=1}^{\tau} \left| \boldsymbol{b}_1^{\mathsf{H}} \left(\boldsymbol{b}_{k\tau+j} - \boldsymbol{b}_{k\tau+1} \right) \right|^2} \right\} \max_{1 \le l \le m} \left| \boldsymbol{b}_l^{\mathsf{H}} \widetilde{\boldsymbol{h}}^t \right|$$
$$\le 0.1 \max_{1 \le l \le m} \left| \boldsymbol{b}_l^{\mathsf{H}} \widetilde{\boldsymbol{h}}^t \right| .$$

Here, the last line arises from Lemma 51, which says that for any small constant $c > 0$, as long as $m \gg \tau K \log m$

$$\sum_{k=0}^{\frac{m}{\tau}-1} \sqrt{\sum_{j=1}^{\tau} \left| \boldsymbol{b}_1^{\mathsf{H}} \left(\boldsymbol{b}_{k\tau+j} - \boldsymbol{b}_{k\tau+1} \right) \right|^2} \le c \frac{1}{\sqrt{\tau}}.$$

- The third term of (221) obeys

$$\left| \boldsymbol{b}_1^{\mathsf{H}} \sum_{j=1}^{\tau} \boldsymbol{b}_{l+1} \left(\boldsymbol{b}_{l+j} - \boldsymbol{b}_{l+1} \right)^{\mathsf{H}} \widetilde{\boldsymbol{h}}^t \left\{ \left| \boldsymbol{a}_{l+j}^{\mathsf{H}} \boldsymbol{x}^\star \right|^2 - \|\boldsymbol{x}^\star\|_2^2 \right\} \right|$$
$$\le \left| \boldsymbol{b}_1^{\mathsf{H}} \boldsymbol{b}_{l+1} \right| \left\{ \sum_{j=1}^{\tau} \left| \left| \boldsymbol{a}_{l+j}^{\mathsf{H}} \boldsymbol{x}^\star \right|^2 - \|\boldsymbol{x}^\star\|_2^2 \right| \right\} \max_{0 \le l \le m-\tau,\, 1 \le j \le \tau} \left| \left(\boldsymbol{b}_{l+j} - \boldsymbol{b}_{l+1} \right)^{\mathsf{H}} \widetilde{\boldsymbol{h}}^t \right|$$
$$\lesssim \tau \left| \boldsymbol{b}_1^{\mathsf{H}} \boldsymbol{b}_{l+1} \right| \max_{0 \le l \le m-\tau,\, 1 \le j \le \tau} \left| \left(\boldsymbol{b}_{l+j} - \boldsymbol{b}_{l+1} \right)^{\mathsf{H}} \widetilde{\boldsymbol{h}}^t \right| ,$$

where the last line relies on the inequality

$$\sum_{j=1}^{\tau} \left| \left| \boldsymbol{a}_{l+j}^{\mathsf{H}} \boldsymbol{x}^\star \right|^2 - \|\boldsymbol{x}^\star\|_2^2 \right| \le \sqrt{\tau} \sqrt{\sum_{j=1}^{\tau} \left(\left| \boldsymbol{a}_{l+j}^{\mathsf{H}} \boldsymbol{x}^\star \right|^2 - \|\boldsymbol{x}^\star\|_2^2 \right)^2} \lesssim \tau$$

owing to Lemma 29 and the Cauchy-Schwarz inequality. Summing over all bins gives

$$\sum_{k=0}^{\frac{m}{\tau}-1} \left| \boldsymbol{b}_1^{\mathsf{H}} \sum_{j=1}^{\tau} \boldsymbol{b}_{k\tau+1} \left(\boldsymbol{b}_{k\tau+j} - \boldsymbol{b}_{k\tau+1} \right)^{\mathsf{H}} \widetilde{\boldsymbol{h}}^t \left\{ \left| \boldsymbol{a}_{k\tau+j}^{\mathsf{H}} \boldsymbol{x}^\star \right|^2 - \|\boldsymbol{x}^\star\|_2^2 \right\} \right|$$
$$\lesssim \tau \sum_{k=0}^{\frac{m}{\tau}-1} \left| \boldsymbol{b}_1^{\mathsf{H}} \boldsymbol{b}_{k\tau+1} \right| \max_{0 \le l \le m-\tau,\, 1 \le j \le \tau} \left| \left(\boldsymbol{b}_{l+j} - \boldsymbol{b}_{l+1} \right)^{\mathsf{H}} \widetilde{\boldsymbol{h}}^t \right|$$
$$\lesssim \log m \max_{0 \le l \le m-\tau,\, 1 \le j \le \tau} \left| \left(\boldsymbol{b}_{l+j} - \boldsymbol{b}_{l+1} \right)^{\mathsf{H}} \widetilde{\boldsymbol{h}}^t \right| ,$$

where the last relation makes use of (223) with the proviso that $m \gg K\tau$. It then boils down to bounding $\max_{0 \le l \le m-\tau,\, 1 \le j \le \tau} \left| \left(\boldsymbol{b}_{l+j} - \boldsymbol{b}_{l+1} \right)^{\mathsf{H}} \widetilde{\boldsymbol{h}}^t \right|$. Without loss of generality, it suffices to look at $\left| (\boldsymbol{b}_j - \boldsymbol{b}_1)^{\mathsf{H}} \widetilde{\boldsymbol{h}}^t \right|$ for all $1 \le j \le \tau$. Specifically, we claim for the moment that

$$\max_{1 \le j \le \tau} \left| \left(\boldsymbol{b}_j - \boldsymbol{b}_1 \right)^{\mathsf{H}} \widetilde{\boldsymbol{h}}^t \right| \le c C_4 \frac{\mu}{\sqrt{m}} \log m \tag{224}$$

for some sufficiently small constant $c > 0$, provided that $m \gg \tau K \log^4 m$. As a result,

$$\sum_{k=0}^{\frac{m}{\tau}-1} \left| \boldsymbol{b}_1^{\mathsf{H}} \sum_{j=1}^{\tau} \boldsymbol{b}_{k\tau+1} \left(\boldsymbol{b}_{k\tau+j} - \boldsymbol{b}_{k\tau+1} \right)^{\mathsf{H}} \widetilde{\boldsymbol{h}}^t \left\{ \left| \boldsymbol{a}_{k\tau+j}^{\mathsf{H}} \boldsymbol{x}^\star \right|^2 - \|\boldsymbol{x}^\star\|_2^2 \right\} \right| \lesssim c C_4 \frac{\mu}{\sqrt{m}} \log^2 m.$$

- Putting the above results together, we get

$$\frac{1}{\eta \xi} \left| \boldsymbol{b}_1^{\mathsf{H}} \boldsymbol{v}_2 \right| \le \sum_{k=0}^{\frac{m}{\tau}-1} \left| \boldsymbol{b}_1^{\mathsf{H}} \sum_{j=1}^{\tau} \boldsymbol{b}_{k\tau+j} \boldsymbol{b}_{k\tau+j}^{\mathsf{H}} \widetilde{\boldsymbol{h}}^t \left\{ \left| \boldsymbol{a}_{k\tau+j}^{\mathsf{H}} \boldsymbol{x}^\star \right|^2 - \|\boldsymbol{x}^\star\|_2^2 \right\} \right| \le 0.2 \max_{1 \le l \le m} \left| \boldsymbol{b}_l^{\mathsf{H}} \widetilde{\boldsymbol{h}}^t \right| + O\left(c C_4 \frac{\mu}{\sqrt{m}} \log^2 m \right).$$

4. Combining the preceding bounds guarantees the existence of some constant $C_8 > 0$ such that

$$\begin{aligned}\left|\boldsymbol{b}_l^{\mathsf{H}}\widetilde{\boldsymbol{h}}^{t+1}\right| &\leq (1+\delta)\left\{(1-\eta\xi)\left|\boldsymbol{b}_l^{\mathsf{H}}\widetilde{\boldsymbol{h}}^{t}\right| + 0.3\eta\xi\max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\widetilde{\boldsymbol{h}}^{t}\right| + C_8(1+C_3)\eta\xi\frac{\mu}{\sqrt{m}} + C_8\eta\xi c C_4\frac{\mu}{\sqrt{m}}\log^2 m\right\}\\ &\overset{\text{(i)}}{\leq}\left(1+O\left(\frac{1}{\log^2 m}\right)\right)\left\{(1-0.7\eta\xi)\,C_4\frac{\mu}{\sqrt{m}}\log^2 m + C_8(1+C_3)\eta\xi\frac{\mu}{\sqrt{m}} + C_8\eta\xi c C_4\frac{\mu}{\sqrt{m}}\log^2 m\right\}\\ &\overset{\text{(ii)}}{\leq} C_4\frac{\mu}{\sqrt{m}}\log^2 m.\end{aligned}$$

Here, (i) uses the induction hypothesis (90d), and (ii) holds as long as $c > 0$ is sufficiently small (so that $(1+\delta)C_8\eta\xi c \ll 1$) and $\eta > 0$ is some sufficiently small constant. In order for the proof to go through, it suffices to pick

$$\tau = c_{10}\log^4 m$$

for some sufficiently large constant $c_{10} > 0$. Accordingly, we need the sample size to exceed

$$m \gg \mu^2\tau K\log^4 m \asymp \mu^2 K\log^8 m.$$

Finally, it remains to verify the claim (224), which we accomplish in Appendix C.4.3.

C.4.1 Proof of Lemma 28

Denote

$$w_j = \boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}^\star\boldsymbol{x}^{\star\mathsf{H}}\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^\star.$$

Recognizing that $\mathbb{E}[\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}] = \boldsymbol{I}_K$ and $\sum_{j=1}^m \boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}} = \boldsymbol{I}_K$, we can write the quantity of interest as the sum of independent random variables, namely,

$$\sum_{j=1}^m \boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}^\star\boldsymbol{x}^{\star\mathsf{H}}\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^\star - \boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{h}^\star = \sum_{j=1}^m \left(w_j - \mathbb{E}\left[w_j\right]\right).$$

Further, the sub-exponential norm (see definition in [Ver12]) of $w_j - \mathbb{E}[w_j]$ obeys

$$\left\|w_j - \mathbb{E}\left[w_j\right]\right\|_{\psi_1} \overset{\text{(i)}}{\leq} 2\left\|w_j\right\|_{\psi_1} \overset{\text{(ii)}}{\leq} 4\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{b}_j\right|\left|\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}^\star\right|\left\|\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^\star\right\|_{\psi_2}^2 \overset{\text{(iii)}}{\lesssim} \left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{b}_j\right|\frac{\mu}{\sqrt{m}} \overset{\text{(iv)}}{\leq} \frac{\mu\sqrt{K}}{m},$$

where (i) arises from the centering property of the sub-exponential norm (see [Ver12, Remark 5.18]), (ii) utilizes the relationship between the sub-exponential norm and the sub-Gaussian norm [Ver12, Lemma 5.14] and (iii) is a consequence of the incoherence condition (36) and the fact that $\left\|\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^\star\right\|_{\psi_2} \lesssim 1$, and (iv) follows from $\|\boldsymbol{b}_j\|_2 = \sqrt{K/m}$. Let $M = \max_{j\in[m]}\left\|w_j - \mathbb{E}\left[w_j\right]\right\|_{\psi_1}$ and

$$V^2 = \sum_{j=1}^m \left\|w_j - \mathbb{E}\left[w_j\right]\right\|_{\psi_1}^2 \lesssim \sum_{j=1}^m\left(\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{b}_j\right|\frac{\mu}{\sqrt{m}}\right)^2 = \frac{\mu^2}{m}\left\|\boldsymbol{b}_l\right\|_2^2 = \frac{\mu^2 K}{m^2},$$

which follows since $\sum_{j=1}^m\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{b}_j\right|^2 = \boldsymbol{b}_l^{\mathsf{H}}\left(\sum_{j=1}^m\boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}\right)\boldsymbol{b}_l = \|\boldsymbol{b}_l\|_2^2 = K/m$. Let $a_j = \left\|w_j - \mathbb{E}\left[w_j\right]\right\|_{\psi_1}$ and $X_j = (w_j - \mathbb{E}[w_j])/a_j$. Since $\|X_j\|_{\psi_1} = 1$, $\sum_{j=1}^m a_j^2 = V^2$ and $\max_{j\in[m]}|a_j| = M$, we can invoke [Ver12, Proposition 5.16] to obtain that

$$\mathbb{P}\left(\left|\sum_{j=1}^m a_j X_j\right| \geq t\right) \leq 2\exp\left(-c\min\left\{\frac{t}{M}, \frac{t^2}{V^2}\right\}\right),$$

where $c > 0$ is some universal constant. By taking $t = \mu/\sqrt{m}$, we see there exists some constant c' such that

$$\mathbb{P}\left(\left|\sum_{j=1}^m \boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{b}_j\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}^\star\boldsymbol{x}^{\star\mathsf{H}}\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^\star - \boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{h}^\star\right| \geq \frac{\mu}{\sqrt{m}}\right) \leq 2\exp\left(-c\min\left\{\frac{\mu/\sqrt{m}}{M}, \frac{\mu^2/m}{V^2}\right\}\right)$$

$$
\begin{aligned}
&\le 2\exp\left(-c'\min\left\{\frac{\mu/\sqrt{m}}{\mu\sqrt{K}/m},\frac{\mu^2/m}{\mu^2K/m^2}\right\}\right)\\
&=2\exp\left(-c'\min\left\{\sqrt{m/K},m/K\right\}\right).
\end{aligned}
$$

We conclude the proof by observing that $m\gg K\log^2 m$ as stated in the assumption.

C.4.2 Proof of Lemma 29

From the elementary inequality $(a-b)^2\le 2\left(a^2+b^2\right)$, we see that

$$
\sum_{j=1}^{\tau}\left(\left|\boldsymbol{a}_j^{\mathsf H}\boldsymbol{x}^\star\right|^2-\|\boldsymbol{x}^\star\|_2^2\right)^2\le 2\sum_{j=1}^{\tau}\left(\left|\boldsymbol{a}_j^{\mathsf H}\boldsymbol{x}^\star\right|^4+\|\boldsymbol{x}^\star\|_2^4\right)=2\sum_{j=1}^{\tau}\left|\boldsymbol{a}_j^{\mathsf H}\boldsymbol{x}^\star\right|^4+2\tau, \tag{225}
$$

where the last identity holds true since $\|\boldsymbol{x}^\star\|_2=1$. It thus suffices to control $\sum_{j=1}^{\tau}\left|\boldsymbol{a}_j^{\mathsf H}\boldsymbol{x}^\star\right|^4$. Let $\xi_i=\boldsymbol{a}_j^{\mathsf H}\boldsymbol{x}^\star$, which is a standard complex Gaussian random variable. Since the ξ_i's are statistically independent, one has

$$
\mathsf{Var}\left(\sum_{i=1}^{\tau}|\xi_i|^4\right)\le C_4\tau
$$

for some constant $C_4>0$. It then follows from the hypercontractivity concentration result for Gaussian polynomials that [SS12, Theorem 1.9]

$$
\begin{aligned}
\mathbb{P}\left\{\sum_{i=1}^{\tau}\left(|\xi_i|^4-\mathbb{E}\left[|\xi_i|^4\right]\right)\ge c\tau\right\}&\le C\exp\left(-c_2\left(\frac{c^2\tau^2}{\mathsf{Var}\left(\sum_{i=1}^{\tau}|\xi_i|^4\right)}\right)^{1/4}\right)\\
&\le C\exp\left(-c_2\left(\frac{c^2\tau^2}{C_4\tau}\right)^{1/4}\right)=C\exp\left(-c_2\left(\frac{c^2}{C_4}\right)^{1/4}\tau^{1/4}\right)\\
&\le O(m^{-10}),
\end{aligned}
$$

for some constants $c,c_2,C>0$, with the proviso that $\tau\gg\log^4 m$. As a consequence, with probability at least $1-O(m^{-10})$,

$$
\sum_{j=1}^{\tau}\left|\boldsymbol{a}_j^{\mathsf H}\boldsymbol{x}^\star\right|^4\lesssim\tau+\sum_{j=1}^{\tau}\mathbb{E}\left[\left|\boldsymbol{a}_j^{\mathsf H}\boldsymbol{x}^\star\right|^4\right]\asymp\tau,
$$

which together with (225) concludes the proof.

C.4.3 Proof of Claim (224)

We will prove the claim by induction. Again, observe that

$$
\left|(\boldsymbol{b}_j-\boldsymbol{b}_1)^{\mathsf H}\widetilde{\boldsymbol{h}}^t\right|=\left|(\boldsymbol{b}_j-\boldsymbol{b}_1)^{\mathsf H}\frac{1}{\overline{\alpha^t}}\boldsymbol{h}^t\right|=\left|\frac{\overline{\alpha^{t-1}}}{\overline{\alpha^t}}\right|\left|(\boldsymbol{b}_j-\boldsymbol{b}_1)^{\mathsf H}\frac{1}{\overline{\alpha^{t-1}}}\boldsymbol{h}^t\right|\le(1+\delta)\left|(\boldsymbol{b}_j-\boldsymbol{b}_1)^{\mathsf H}\frac{1}{\overline{\alpha^{t-1}}}\boldsymbol{h}^t\right|
$$

for some $\delta\asymp\log^{-2}m$, which allows us to look at $(\boldsymbol{b}_j-\boldsymbol{b}_1)^{\mathsf H}\frac{1}{\overline{\alpha^{t-1}}}\boldsymbol{h}^t$ instead.

Use the gradient update rule for $\boldsymbol{h}^t$ (cf. (79a)) once again to get

$$
\begin{aligned}
\frac{1}{\overline{\alpha^{t-1}}}\boldsymbol{h}^t&=\frac{1}{\overline{\alpha^{t-1}}}\left(\boldsymbol{h}^{t-1}-\frac{\eta}{\|\boldsymbol{x}^{t-1}\|_2^2}\sum_{l=1}^{m}\boldsymbol{b}_l\boldsymbol{b}_l^{\mathsf H}\left(\boldsymbol{h}^{t-1}\boldsymbol{x}^{t-1\mathsf H}-\boldsymbol{h}^\star\boldsymbol{x}^{\star\mathsf H}\right)\boldsymbol{a}_l\boldsymbol{a}_l^{\mathsf H}\boldsymbol{x}^{t-1}\right)\\
&=\widetilde{\boldsymbol{h}}^{t-1}-\eta\theta\sum_{l=1}^{m}\boldsymbol{b}_l\boldsymbol{b}_l^{\mathsf H}\left(\widetilde{\boldsymbol{h}}^{t-1}\widetilde{\boldsymbol{x}}^{t-1\mathsf H}-\boldsymbol{h}^\star\boldsymbol{x}^{\star\mathsf H}\right)\boldsymbol{a}_l\boldsymbol{a}_l^{\mathsf H}\widetilde{\boldsymbol{x}}^{t-1},
\end{aligned}
$$

where we denote $\theta := 1/\left\|\widetilde{\boldsymbol{x}}^{t-1}\right\|_2^2$. This further gives rise to

$$
\begin{aligned}
\left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^{\mathsf{H}} \frac{1}{\overline{\alpha^{t-1}}}\boldsymbol{h}^t &= \left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^{\mathsf{H}}\widetilde{\boldsymbol{h}}^{t-1} - \eta\theta\left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^{\mathsf{H}}\sum_{l=1}^m \boldsymbol{b}_l\boldsymbol{b}_l^{\mathsf{H}}\left(\widetilde{\boldsymbol{h}}^{t-1}\widetilde{\boldsymbol{x}}^{t-1\mathsf{H}} - \boldsymbol{h}^{\star}\boldsymbol{x}^{\star\mathsf{H}}\right)\boldsymbol{a}_l\boldsymbol{a}_l^{\mathsf{H}}\widetilde{\boldsymbol{x}}^{t-1} \\
&= \left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^{\mathsf{H}}\widetilde{\boldsymbol{h}}^{t-1} - \eta\theta\left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^{\mathsf{H}}\sum_{l=1}^m \boldsymbol{b}_l\boldsymbol{b}_l^{\mathsf{H}}\left(\widetilde{\boldsymbol{h}}^{t-1}\widetilde{\boldsymbol{x}}^{t-1\mathsf{H}} - \boldsymbol{h}^{\star}\boldsymbol{x}^{\star\mathsf{H}}\right)\widetilde{\boldsymbol{x}}^{t-1} \\
&\quad - \eta\theta\left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^{\mathsf{H}}\sum_{l=1}^m \boldsymbol{b}_l\boldsymbol{b}_l^{\mathsf{H}}\left(\widetilde{\boldsymbol{h}}^{t-1}\widetilde{\boldsymbol{x}}^{t-1\mathsf{H}} - \boldsymbol{h}^{\star}\boldsymbol{x}^{\star\mathsf{H}}\right)\left(\boldsymbol{a}_l\boldsymbol{a}_l^{\mathsf{H}} - \boldsymbol{I}_K\right)\widetilde{\boldsymbol{x}}^{t-1} \\
&= \left(1-\eta\theta\|\widetilde{\boldsymbol{x}}^{t-1}\|_2^2\right)\left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^{\mathsf{H}}\widetilde{\boldsymbol{h}}^{t-1} + \underbrace{\eta\theta\left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^{\mathsf{H}}\boldsymbol{h}^{\star}\left(\boldsymbol{x}^{\star\mathsf{H}}\widetilde{\boldsymbol{x}}^{t-1}\right)}_{:=\beta_1} \\
&\quad \underbrace{- \eta\theta\left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^{\mathsf{H}}\sum_{l=1}^m \boldsymbol{b}_l\boldsymbol{b}_l^{\mathsf{H}}\left(\widetilde{\boldsymbol{h}}^{t-1}\widetilde{\boldsymbol{x}}^{t-1\mathsf{H}} - \boldsymbol{h}^{\star}\boldsymbol{x}^{\star\mathsf{H}}\right)\left(\boldsymbol{a}_l\boldsymbol{a}_l^{\mathsf{H}} - \boldsymbol{I}_K\right)\widetilde{\boldsymbol{x}}^{t-1},}_{:=\beta_2}
\end{aligned}
$$

where the last identity makes use of the fact that $\sum_{l=1}^m \boldsymbol{b}_l\boldsymbol{b}_l^{\mathsf{H}} = \boldsymbol{I}_K$. For β_1, one can get

$$
\frac{1}{\eta\theta}\left|\beta_1\right| \leq \left|\left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^{\mathsf{H}}\boldsymbol{h}^{\star}\right| \|\boldsymbol{x}^{\star}\|_2\|\widetilde{\boldsymbol{x}}^{t-1}\|_2 \leq 4\frac{\mu}{\sqrt{m}},
$$

where we utilize the incoherence condition (36) and the fact that $\widetilde{\boldsymbol{x}}^{t-1}$ and $\boldsymbol{x}^{\star}$ are extremely close, i.e.

$$
\left\|\widetilde{\boldsymbol{x}}^{t-1} - \boldsymbol{x}^{\star}\right\|_2 \leq \operatorname{dist}\left(\boldsymbol{z}^{t-1},\boldsymbol{z}^{\star}\right) \ll 1 \qquad \Longrightarrow \qquad \|\widetilde{\boldsymbol{x}}^{t-1}\|_2 \leq 2.
$$

Regarding the second term β_2, we have

$$
\frac{1}{\eta\theta}\left|\beta_2\right| \leq \left\{\sum_{l=1}^m \left|\left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^{\mathsf{H}}\boldsymbol{b}_l\right|\right\} \underbrace{\max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\left(\widetilde{\boldsymbol{h}}^{t-1}\widetilde{\boldsymbol{x}}^{t-1\mathsf{H}} - \boldsymbol{h}^{\star}\boldsymbol{x}^{\star\mathsf{H}}\right)\left(\boldsymbol{a}_l\boldsymbol{a}_l^{\mathsf{H}} - \boldsymbol{I}_K\right)\widetilde{\boldsymbol{x}}^{t-1}\right|}_{:=\psi}.
$$

The term ψ can be bounded as follows

$$
\begin{aligned}
\psi &\leq \max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\widetilde{\boldsymbol{h}}^{t-1}\widetilde{\boldsymbol{x}}^{t-1\mathsf{H}}\left(\boldsymbol{a}_l\boldsymbol{a}_l^{\mathsf{H}} - \boldsymbol{I}\right)\widetilde{\boldsymbol{x}}^{t-1}\right| + \max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{h}^{\star}\boldsymbol{x}^{\star\mathsf{H}}\left(\boldsymbol{a}_l\boldsymbol{a}_l^{\mathsf{H}} - \boldsymbol{I}_K\right)\widetilde{\boldsymbol{x}}^{t-1}\right| \\
&\leq \max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\widetilde{\boldsymbol{h}}^{t-1}\right| \max_{1\leq l\leq m}\left|\widetilde{\boldsymbol{x}}^{t-1\mathsf{H}}\left(\boldsymbol{a}_l\boldsymbol{a}_l^{\mathsf{H}} - \boldsymbol{I}_K\right)\widetilde{\boldsymbol{x}}^{t-1}\right| + \max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{h}^{\star}\right| \max_{1\leq l\leq m}\left|\boldsymbol{x}^{\star\mathsf{H}}\left(\boldsymbol{a}_l\boldsymbol{a}_l^{\mathsf{H}} - \boldsymbol{I}_K\right)\widetilde{\boldsymbol{x}}^{t-1}\right| \\
&\lesssim \log m\left\{\max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\widetilde{\boldsymbol{h}}^{t-1}\right| + \frac{\mu}{\sqrt{m}}\right\}.
\end{aligned}
$$

Here, we have used the incoherence condition (36) and the facts that

$$
\begin{aligned}
\left|(\widetilde{\boldsymbol{x}}^{t-1})^{\mathsf{H}}\left(\boldsymbol{a}_l\boldsymbol{a}_l^{\mathsf{H}} - \boldsymbol{I}\right)\widetilde{\boldsymbol{x}}^{t-1}\right| &\leq \left\|\boldsymbol{a}_l^{\mathsf{H}}\widetilde{\boldsymbol{x}}^{t-1}\right\|_2^2 + \left\|\widetilde{\boldsymbol{x}}^{t-1}\right\|_2^2 \lesssim \log m, \\
\left|\boldsymbol{x}^{\star\mathsf{H}}\left(\boldsymbol{a}_l\boldsymbol{a}_l^{\mathsf{H}} - \boldsymbol{I}\right)\widetilde{\boldsymbol{x}}^{t-1}\right| &\leq \left\|\boldsymbol{a}_l^{\mathsf{H}}\widetilde{\boldsymbol{x}}^{t-1}\right\|_2\left\|\boldsymbol{a}_l^{\mathsf{H}}\boldsymbol{x}^{\star}\right\|_2 + \left\|\widetilde{\boldsymbol{x}}^{t-1}\right\|_2\|\boldsymbol{x}^{\star}\|_2 \lesssim \log m,
\end{aligned}
$$

which are immediate consequences of (90c) and (189). Combining this with Lemma 50, we see that for any small constant $c > 0$

$$
\frac{1}{\eta\theta}\left|\beta_2\right| \leq c\frac{1}{\log m}\left\{\max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\widetilde{\boldsymbol{h}}^{t-1}\right| + \frac{\mu}{\sqrt{m}}\right\}
$$

holds as long as $m \gg \tau K \log^4 m$.

To summarize, we arrive at

$$
\left|\left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^{\mathsf{H}}\widetilde{\boldsymbol{h}}^{t}\right| \leq (1+\delta)\left\{\left(1-\eta\theta\left\|\widetilde{\boldsymbol{x}}^{t-1}\right\|_2^2\right)\left|\left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^{\mathsf{H}}\widetilde{\boldsymbol{h}}^{t-1}\right| + 4\eta\theta\frac{\mu}{\sqrt{m}} + c\eta\theta\frac{1}{\log m}\left[\max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\widetilde{\boldsymbol{h}}^{t-1}\right| + \frac{\mu}{\sqrt{m}}\right]\right\}.
$$

Making use of the induction hypothesis (85c) and the fact that $\left\|\widetilde{\boldsymbol{x}}^{t-1}\right\|_2^2 \geq 0.9$, we reach

$$\left|(\boldsymbol{b}_j - \boldsymbol{b}_1)^{\mathsf{H}} \widetilde{\boldsymbol{h}}^t\right| \leq (1+\delta)\left\{(1-0.9\eta\theta)\left|(\boldsymbol{b}_j - \boldsymbol{b}_1)^{\mathsf{H}} \widetilde{\boldsymbol{h}}^{t-1}\right| + cC_4\eta\theta\frac{\mu}{\sqrt{m}}\log m + \frac{c\mu\eta\theta}{\sqrt{m}\log m}\right\}.$$

Recall that $\delta \asymp 1/\log^2 m$. As a result, if $\eta > 0$ is some sufficiently small constant and if

$$\left|(\boldsymbol{b}_j - \boldsymbol{b}_1)^{\mathsf{H}} \widetilde{\boldsymbol{h}}^{t-1}\right| \leq 10c\left(C_4\frac{\mu}{\sqrt{m}}\log m + \frac{\mu}{\eta\theta\sqrt{m}\log m}\right) \leq 20cC_4\frac{\mu}{\sqrt{m}}\log m$$

holds, then one has

$$\left|(\boldsymbol{b}_j - \boldsymbol{b}_1)^{\mathsf{H}} \widetilde{\boldsymbol{h}}^t\right| \leq 20cC_4\frac{\mu}{\sqrt{m}}\log m.$$

Therefore, this concludes the proof of the claim (224) by induction, provided that the base case is true, i.e. for some $c > 0$ sufficiently small

$$\left|(\boldsymbol{b}_j - \boldsymbol{b}_1)^{\mathsf{H}} \widetilde{\boldsymbol{h}}^0\right| \leq 20cC_4\frac{\mu}{\sqrt{m}}\log m. \tag{226}$$

The claim (226) is proved in Appendix C.6 (see Lemma 30).

C.5 Proof of Lemma 19

Recall that $\check{\boldsymbol{h}}^0$ and $\check{\boldsymbol{x}}^0$ are the leading left and right singular vectors of $\boldsymbol{M}$, respectively. Applying a variant of Wedin's sinΘ theorem [Dop00, Theorem 2.1], we derive that

$$\min_{\alpha\in\mathbb{C},|\alpha|=1}\left\{\left\|\alpha\check{\boldsymbol{h}}^0 - \boldsymbol{h}^\star\right\|_2 + \left\|\alpha\check{\boldsymbol{x}}^0 - \boldsymbol{x}^\star\right\|_2\right\} \leq \frac{c_1\left\|\boldsymbol{M} - \mathbb{E}\left[\boldsymbol{M}\right]\right\|}{\sigma_1\left(\mathbb{E}\left[\boldsymbol{M}\right]\right) - \sigma_2\left(\boldsymbol{M}\right)}, \tag{227}$$

for some universal constant $c_1 > 0$. Regarding the numerator of (227), it has been shown in [LLSW18, Lemma 5.20] that for any $\xi > 0$,

$$\left\|\boldsymbol{M} - \mathbb{E}\left[\boldsymbol{M}\right]\right\| \leq \xi \tag{228}$$

with probability exceeding $1 - O(m^{-10})$, provided that

$$m \geq \frac{c_2\mu^2 K\log^2 m}{\xi^2}$$

for some universal constant $c_2 > 0$. For the denominator of (227), we can take (228) together with Weyl's inequality to demonstrate that

$$\sigma_1\left(\mathbb{E}\left[\boldsymbol{M}\right]\right) - \sigma_2\left(\boldsymbol{M}\right) \geq \sigma_1\left(\mathbb{E}\left[\boldsymbol{M}\right]\right) - \sigma_2\left(\mathbb{E}\left[\boldsymbol{M}\right]\right) - \left\|\boldsymbol{M} - \mathbb{E}\left[\boldsymbol{M}\right]\right\| \geq 1 - \xi,$$

where the last inequality utilizes the facts that $\sigma_1\left(\mathbb{E}\left[\boldsymbol{M}\right]\right) = 1$ and $\sigma_2\left(\mathbb{E}[\boldsymbol{M}]\right) = 0$. These together with (227) reveal that

$$\min_{\alpha\in\mathbb{C},|\alpha|=1}\left\{\left\|\alpha\check{\boldsymbol{h}}^0 - \boldsymbol{h}^\star\right\|_2 + \left\|\alpha\check{\boldsymbol{x}}^0 - \boldsymbol{x}^\star\right\|_2\right\} \leq \frac{c_1\xi}{1-\xi} \leq 2c_1\xi \tag{229}$$

as long as $\xi \leq 1/2$.

Now we connect the preceding bound (229) with the scaled singular vectors $\boldsymbol{h}^0 = \sqrt{\sigma_1\left(\boldsymbol{M}\right)}\,\check{\boldsymbol{h}}^0$ and $\boldsymbol{x}^0 = \sqrt{\sigma_1\left(\boldsymbol{M}\right)}\,\check{\boldsymbol{x}}^0$. For any $\alpha \in \mathbb{C}$ with $|\alpha| = 1$, from the definition of $\boldsymbol{h}^0$ and $\boldsymbol{x}^0$ we have

$$\left\|\alpha\boldsymbol{h}^0 - \boldsymbol{h}^\star\right\|_2 + \left\|\alpha\boldsymbol{x}^0 - \boldsymbol{x}^\star\right\|_2 = \left\|\sqrt{\sigma_1\left(\boldsymbol{M}\right)}\left(\alpha\check{\boldsymbol{h}}^0\right) - \boldsymbol{h}^\star\right\|_2 + \left\|\sqrt{\sigma_1\left(\boldsymbol{M}\right)}\left(\alpha\check{\boldsymbol{x}}^0\right) - \boldsymbol{x}^\star\right\|_2.$$

Since $\alpha\check{\boldsymbol{h}}^0, \alpha\check{\boldsymbol{x}}^0$ are also the leading left and right singular vectors of $\boldsymbol{M}$, we can invoke Lemma 60 to get

$$\left\|\alpha\boldsymbol{h}^0 - \boldsymbol{h}^\star\right\|_2 + \left\|\alpha\boldsymbol{x}^0 - \boldsymbol{x}^\star\right\|_2 \leq \sqrt{\sigma_1(\mathbb{E}[\boldsymbol{M}])}\left(\left\|\alpha\check{\boldsymbol{h}}^0 - \boldsymbol{h}^\star\right\|_2 + \left\|\alpha\check{\boldsymbol{x}}^0 - \boldsymbol{x}^\star\right\|_2\right) + \frac{2\left|\sigma_1(\boldsymbol{M}) - \sigma_1\left(\mathbb{E}\left[\boldsymbol{M}\right]\right)\right|}{\sqrt{\sigma_1(\boldsymbol{M})} + \sqrt{\sigma_1\left(\mathbb{E}\left[\boldsymbol{M}\right]\right)}}$$

$$= \left\|\alpha\check{\boldsymbol{h}}^0 - \boldsymbol{h}^\star\right\|_2 + \left\|\alpha\check{\boldsymbol{x}}^0 - \boldsymbol{x}^\star\right\|_2 + \frac{2\left|\sigma_1(\boldsymbol{M}) - \sigma_1\left(\mathbb{E}\left[\boldsymbol{M}\right]\right)\right|}{\sqrt{\sigma_1(\boldsymbol{M})} + 1}. \tag{230}$$

In addition, we can apply Weyl's inequality once again to deduce that

$$|\sigma_1(\boldsymbol{M}) - \sigma_1(\mathbb{E}[\boldsymbol{M}])| \leq \|\boldsymbol{M} - \mathbb{E}[\boldsymbol{M}]\| \leq \xi, \tag{231}$$

where the last inequality comes from (228). Substitute (231) into (230) to obtain

$$\left\|\alpha\boldsymbol{h}^0 - \boldsymbol{h}^\star\right\|_2 + \left\|\alpha\boldsymbol{x}^0 - \boldsymbol{x}^\star\right\|_2 \leq \left\|\alpha\check{\boldsymbol{h}}^0 - \boldsymbol{h}^\star\right\|_2 + \left\|\alpha\check{\boldsymbol{x}}^0 - \boldsymbol{x}^\star\right\|_2 + 2\xi. \tag{232}$$

Taking the minimum over α, one can thus conclude that

$$\min_{\alpha\in\mathbb{C},|\alpha|=1}\left\{\left\|\alpha\boldsymbol{h}^0 - \boldsymbol{h}^\star\right\|_2 + \left\|\alpha\boldsymbol{x}^0 - \boldsymbol{x}^\star\right\|_2\right\} \leq \min_{\alpha\in\mathbb{C},|\alpha|=1}\left\{\left\|\alpha\check{\boldsymbol{h}}^0 - \boldsymbol{h}^\star\right\|_2 + \left\|\alpha\check{\boldsymbol{x}}^0 - \boldsymbol{x}^\star\right\|_2\right\} + 2\xi \leq 2c_1\xi + 2\xi,$$

where the last inequality comes from (229). Since ξ is arbitrary, by taking $m/(\mu^2 K \log^2 m)$ to be large enough, we finish the proof for (92). Carrying out similar arguments (which we omit here), we can also establish (93).

The last claim in Lemma 19 that $||\alpha_0| - 1| \leq 1/4$ is a direct corollary of (92) and Lemma 52.

C.6 Proof of Lemma 20

The proof is composed of three steps:

- In the first step, we show that the normalized singular vectors of $\boldsymbol{M}$ and $\boldsymbol{M}^{(l)}$ are close enough; see (240).
- We then proceed by passing this proximity result to the scaled singular vectors; see (243).
- Finally, we translate the usual ℓ_2 distance metric to the distance function we defined in (34); see (245). Along the way, we also prove the incoherence of $\boldsymbol{h}^0$ with respect to $\{\boldsymbol{b}_l\}$.

Here comes the formal proof. Recall that $\check{\boldsymbol{h}}^0$ and $\check{\boldsymbol{x}}^0$ are respectively the leading left and right singular vectors of $\boldsymbol{M}$, and $\check{\boldsymbol{h}}^{0,(l)}$ and $\check{\boldsymbol{x}}^{0,(l)}$ are respectively the leading left and right singular vectors of $\boldsymbol{M}^{(l)}$. Invoke Wedin's sinΘ theorem [Dop00, Theorem 2.1] to obtain

$$\min_{\alpha\in\mathbb{C},|\alpha|=1}\left\{\left\|\alpha\check{\boldsymbol{h}}^0 - \check{\boldsymbol{h}}^{0,(l)}\right\|_2 + \left\|\alpha\check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(l)}\right\|_2\right\} \leq c_1\frac{\left\|\left(\boldsymbol{M} - \boldsymbol{M}^{(l)}\right)\check{\boldsymbol{x}}^{0,(l)}\right\|_2 + \left\|\check{\boldsymbol{h}}^{0,(l)\mathsf{H}}\left(\boldsymbol{M} - \boldsymbol{M}^{(l)}\right)\right\|_2}{\sigma_1\left(\boldsymbol{M}^{(l)}\right) - \sigma_2\left(\boldsymbol{M}\right)}$$

for some universal constant $c_1 > 0$. Using the Weyl's inequality we get

$$\begin{aligned}\sigma_1\left(\boldsymbol{M}^{(l)}\right) - \sigma_2\left(\boldsymbol{M}\right) &\geq \sigma_1\left(\mathbb{E}[\boldsymbol{M}^{(l)}]\right) - \|\boldsymbol{M}^{(l)} - \mathbb{E}[\boldsymbol{M}^{(l)}]\| - \sigma_2\left(\mathbb{E}[\boldsymbol{M}]\right) - \|\boldsymbol{M} - \mathbb{E}[\boldsymbol{M}]\| \\ &\geq 3/4 - \|\boldsymbol{M}^{(l)} - \mathbb{E}[\boldsymbol{M}^{(l)}]\| - \|\boldsymbol{M} - \mathbb{E}[\boldsymbol{M}]\| \geq 1/2,\end{aligned}$$

where the penultimate inequality follows from

$$\sigma_1\left(\mathbb{E}[\boldsymbol{M}^{(l)}]\right) \geq 3/4$$

for m sufficiently large, and the last inequality comes from [LLSW18, Lemma 5.20], provided that $m \geq c_2\mu^2 K \log^2 m$ for some sufficiently large constant $c_2 > 0$. As a result, denoting

$$\beta^{0,(l)} := \operatorname*{argmin}_{\alpha\in\mathbb{C},|\alpha|=1}\left\{\left\|\alpha\check{\boldsymbol{h}}^0 - \check{\boldsymbol{h}}^{0,(l)}\right\|_2 + \left\|\alpha\check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(l)}\right\|_2\right\} \tag{233}$$

allows us to obtain

$$\left\|\beta^{0,(l)}\check{\boldsymbol{h}}^0 - \check{\boldsymbol{h}}^{0,(l)}\right\|_2 + \left\|\beta^{0,(l)}\check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(l)}\right\|_2 \leq 2c_1\left\{\left\|\left(\boldsymbol{M} - \boldsymbol{M}^{(l)}\right)\check{\boldsymbol{x}}^{0,(l)}\right\|_2 + \left\|\check{\boldsymbol{h}}^{0,(l)\mathsf{H}}\left(\boldsymbol{M} - \boldsymbol{M}^{(l)}\right)\right\|_2\right\}. \tag{234}$$

It then boils down to controlling the two terms on the right-hand side of (234). By construction,

$$\boldsymbol{M} - \boldsymbol{M}^{(l)} = \boldsymbol{b}_l\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{h}^\star\boldsymbol{x}^{\star\mathsf{H}}\boldsymbol{a}_l\boldsymbol{a}_l^{\mathsf{H}}.$$

- To bound the first term, observe that

$$\begin{aligned}\left\|\left(\boldsymbol{M}-\boldsymbol{M}^{(l)}\right)\check{\boldsymbol{x}}^{0,(l)}\right\|_2 = \left\|\boldsymbol{b}_l\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{h}^{\star}\boldsymbol{x}^{\star\mathsf{H}}\boldsymbol{a}_l\boldsymbol{a}_l^{\mathsf{H}}\check{\boldsymbol{x}}^{0,(l)}\right\|_2 &= \|\boldsymbol{b}_l\|_2\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{h}^{\star}\right|\left|\boldsymbol{a}_l^{\mathsf{H}}\boldsymbol{x}^{\star}\right|\cdot\left|\boldsymbol{a}_l^{\mathsf{H}}\check{\boldsymbol{x}}^{0,(l)}\right| \\ &\le 30\frac{\mu}{\sqrt{m}}\cdot\sqrt{\frac{K\log^2 m}{m}}, \end{aligned} \tag{235}$$

  where we use the fact that $\|\boldsymbol{b}_l\|_2=\sqrt{K/m}$, the incoherence condition (36), the bound (189) and the fact that with probability exceeding $1-O\left(m^{-10}\right)$,

$$\max_{1\le l\le m}\left|\boldsymbol{a}_l^{\mathsf{H}}\check{\boldsymbol{x}}^{0,(l)}\right|\le 5\sqrt{\log m},$$

  due to the independence between $\check{\boldsymbol{x}}^{0,(l)}$ and $\boldsymbol{a}_l$.

- To bound the second term, for any $\widetilde{\alpha}$ obeying $|\widetilde{\alpha}|=1$ one has

$$\begin{aligned}\left\|\check{\boldsymbol{h}}^{0,(l)\mathsf{H}}\left(\boldsymbol{M}-\boldsymbol{M}^{(l)}\right)\right\|_2 = \left\|\check{\boldsymbol{h}}^{0,(l)\mathsf{H}}\boldsymbol{b}_l\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{h}^{\star}\boldsymbol{x}^{\star\mathsf{H}}\boldsymbol{a}_l\boldsymbol{a}_l^{\mathsf{H}}\right\|_2 &= \|\boldsymbol{a}_l\|_2\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{h}^{\star}\right|\left|\boldsymbol{a}_l^{\mathsf{H}}\boldsymbol{x}^{\star}\right|\cdot\left|\boldsymbol{b}_l^{\mathsf{H}}\check{\boldsymbol{h}}^{0,(l)}\right| \\ &\overset{\text{(i)}}{\le} 3\sqrt{K}\cdot\frac{\mu}{\sqrt{m}}\cdot 5\sqrt{\log m}\cdot\left|\boldsymbol{b}_l^{\mathsf{H}}\check{\boldsymbol{h}}^{0,(l)}\right| \\ &\overset{\text{(ii)}}{\le} 15\sqrt{\frac{\mu^2K\log m}{m}}\left|\widetilde{\alpha}\boldsymbol{b}_l^{\mathsf{H}}\check{\boldsymbol{h}}^0\right| + 15\sqrt{\frac{\mu^2K\log m}{m}}\left|\boldsymbol{b}_l^{\mathsf{H}}\left(\widetilde{\alpha}\check{\boldsymbol{h}}^0-\check{\boldsymbol{h}}^{0,(l)}\right)\right| \\ &\overset{\text{(iii)}}{\le} 15\sqrt{\frac{\mu^2K\log m}{m}}\left|\boldsymbol{b}_l^{\mathsf{H}}\check{\boldsymbol{h}}^0\right| + 15\sqrt{\frac{\mu^2K\log m}{m}}\cdot\sqrt{\frac{K}{m}}\left\|\widetilde{\alpha}\check{\boldsymbol{h}}^0-\check{\boldsymbol{h}}^{0,(l)}\right\|_2. \end{aligned}$$

  Here, (i) arises from the incoherence condition (36) together with the bounds (189) and (190), the inequality (ii) comes from the triangle inequality, and the last line (iii) holds since $\|\boldsymbol{b}_l\|_2=\sqrt{K/m}$ and $|\widetilde{\alpha}|=1$.

Substitution of the above bounds into (234) yields

$$\begin{aligned}&\left\|\beta^{0,(l)}\check{\boldsymbol{h}}^0-\check{\boldsymbol{h}}^{0,(l)}\right\|_2+\left\|\beta^{0,(l)}\check{\boldsymbol{x}}^0-\check{\boldsymbol{x}}^{0,(l)}\right\|_2 \\ &\quad\le 2c_1\left\{30\frac{\mu}{\sqrt{m}}\cdot\sqrt{\frac{K\log^2 m}{m}}+15\sqrt{\frac{\mu^2K\log m}{m}}\left|\boldsymbol{b}_l^{\mathsf{H}}\check{\boldsymbol{h}}^0\right|+15\sqrt{\frac{\mu^2K\log m}{m}}\cdot\sqrt{\frac{K}{m}}\left\|\widetilde{\alpha}\check{\boldsymbol{h}}^0-\check{\boldsymbol{h}}^{0,(l)}\right\|_2\right\}.\end{aligned}$$

Since the previous inequality holds for all $|\widetilde{\alpha}|=1$, we can choose $\widetilde{\alpha}=\beta^{0,(l)}$ and rearrange terms to get

$$\begin{aligned}&\left(1-30c_1\sqrt{\frac{\mu^2K\log m}{m}}\sqrt{\frac{K}{m}}\right)\left(\left\|\beta^{0,(l)}\check{\boldsymbol{h}}^0-\check{\boldsymbol{h}}^{0,(l)}\right\|_2+\left\|\beta^{0,(l)}\check{\boldsymbol{x}}^0-\check{\boldsymbol{x}}^{0,(l)}\right\|_2\right) \\ &\quad\le 60c_1\frac{\mu}{\sqrt{m}}\cdot\sqrt{\frac{K\log^2 m}{m}}+30c_1\sqrt{\frac{\mu^2K\log m}{m}}\left|\boldsymbol{b}_l^{\mathsf{H}}\check{\boldsymbol{h}}^0\right|.\end{aligned}$$

Under the condition that $m\gg\mu K\log^{1/2}m$, one has $1-30c_1\sqrt{\mu^2K\log m/m}\cdot\sqrt{K/m}\ge\frac{1}{2}$, and therefore

$$\left\|\beta^{0,(l)}\check{\boldsymbol{h}}^0-\check{\boldsymbol{h}}^{0,(l)}\right\|_2+\left\|\beta^{0,(l)}\check{\boldsymbol{x}}^0-\check{\boldsymbol{x}}^{0,(l)}\right\|_2\le 120c_1\frac{\mu}{\sqrt{m}}\cdot\sqrt{\frac{K\log^2 m}{m}}+60c_1\sqrt{\frac{\mu^2K\log m}{m}}\left|\boldsymbol{b}_l^{\mathsf{H}}\check{\boldsymbol{h}}^0\right|,$$

which immediately implies that

$$\begin{aligned}&\max_{1\le l\le m}\left\{\left\|\beta^{0,(l)}\check{\boldsymbol{h}}^0-\check{\boldsymbol{h}}^{0,(l)}\right\|_2+\left\|\beta^{0,(l)}\check{\boldsymbol{x}}^0-\check{\boldsymbol{x}}^{0,(l)}\right\|_2\right\} \\ &\quad\le 120c_1\frac{\mu}{\sqrt{m}}\cdot\sqrt{\frac{K\log^2 m}{m}}+60c_1\sqrt{\frac{\mu^2K\log m}{m}}\max_{1\le l\le m}\left|\boldsymbol{b}_l^{\mathsf{H}}\check{\boldsymbol{h}}^0\right|.\end{aligned} \tag{236}$$

We then move on to $\left|\boldsymbol{b}_l^{\mathsf{H}}\check{\boldsymbol{h}}^0\right|$. The aim is to show that $\max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\check{\boldsymbol{h}}^0\right|$ can also be upper bounded by the left-hand side of (236). By construction, we have $\boldsymbol{M}\check{\boldsymbol{x}}^0=\sigma_1\left(\boldsymbol{M}\right)\check{\boldsymbol{h}}^0$, which further leads to

$$\begin{aligned}
\left|\boldsymbol{b}_l^{\mathsf{H}}\check{\boldsymbol{h}}^0\right| &= \frac{1}{\sigma_1\left(\boldsymbol{M}\right)}\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{M}\check{\boldsymbol{x}}^0\right| \\
&\overset{\text{(i)}}{\leq} 2\left|\sum_{j=1}^{m}\left(\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{b}_j\right)\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}^{\star}\boldsymbol{x}^{\star\mathsf{H}}\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\check{\boldsymbol{x}}^0\right| \\
&\leq 2\left(\sum_{j=1}^{m}\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{b}_j\right|\right)\max_{1\leq j\leq m}\left\{\left|\boldsymbol{b}_j^{\mathsf{H}}\boldsymbol{h}^{\star}\right|\left|\boldsymbol{a}_j^{\mathsf{H}}\boldsymbol{x}^{\star}\right|\left|\boldsymbol{a}_j^{\mathsf{H}}\check{\boldsymbol{x}}^0\right|\right\} \\
&\overset{\text{(ii)}}{\leq} 8\log m\cdot\frac{\mu}{\sqrt{m}}\cdot\left(5\sqrt{\log m}\right)\max_{1\leq j\leq m}\left\{\left|\boldsymbol{a}_j^{\mathsf{H}}\check{\boldsymbol{x}}^{0,(j)}\right|+\left\|\boldsymbol{a}_j\right\|_2\left\|\beta^{0,(j)}\check{\boldsymbol{x}}^0-\check{\boldsymbol{x}}^{0,(j)}\right\|_2\right\} \\
&\leq 200\frac{\mu\log^2 m}{\sqrt{m}}+120\sqrt{\frac{\mu^2K\log^3 m}{m}}\max_{1\leq j\leq m}\left\|\beta^{0,(j)}\check{\boldsymbol{x}}^0-\check{\boldsymbol{x}}^{0,(j)}\right\|_2, \qquad (237)
\end{aligned}$$

where $\beta^{0,(j)}$ is as defined in (233). Here, (i) comes from the lower bound $\sigma_1\left(\boldsymbol{M}\right)\geq 1/2$. The bound (ii) follows by combining the incoherence condition (36), the bound (189), the triangle inequality, as well as the estimate $\sum_{j=1}^{m}\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{b}_j\right|\leq 4\log m$ from Lemma 48. The last line uses the upper estimate $\max_{1\leq j\leq m}\left|\boldsymbol{a}_j^{\mathsf{H}}\check{\boldsymbol{x}}^{0,(j)}\right|\leq 5\sqrt{\log m}$ and (190). Our bound (237) further implies

$$\max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\check{\boldsymbol{h}}^0\right|\leq 200\frac{\mu\log^2 m}{\sqrt{m}}+120\sqrt{\frac{\mu^2K\log^3 m}{m}}\max_{1\leq j\leq m}\left\|\beta^{0,(j)}\check{\boldsymbol{x}}^0-\check{\boldsymbol{x}}^{0,(j)}\right\|_2. \qquad (238)$$

The above bound (238) taken together with (236) gives

$$\begin{aligned}
\max_{1\leq l\leq m}\left\{\left\|\beta^{0,(l)}\check{\boldsymbol{h}}^0-\check{\boldsymbol{h}}^{0,(l)}\right\|_2+\left\|\beta^{0,(l)}\check{\boldsymbol{x}}^0-\check{\boldsymbol{x}}^{0,(l)}\right\|_2\right\} &\leq 120c_1\frac{\mu}{\sqrt{m}}\cdot\sqrt{\frac{K\log^2 m}{m}} \\
+60c_1\sqrt{\frac{\mu^2K\log m}{m}}&\left(200\frac{\mu\log^2 m}{\sqrt{m}}+120\sqrt{\frac{\mu^2K\log^3 m}{m}}\max_{1\leq j\leq m}\left\|\beta^{0,(j)}\check{\boldsymbol{x}}^0-\check{\boldsymbol{x}}^{0,(j)}\right\|_2\right). \qquad (239)
\end{aligned}$$

As long as $m\gg\mu^2K\log^2 m$ we have $60c_1\sqrt{\mu^2K\log m/m}\cdot 120\sqrt{\mu^2K\log^3 m/m}\leq 1/2$. Rearranging terms, we are left with

$$\max_{1\leq l\leq m}\left\{\left\|\beta^{0,(l)}\check{\boldsymbol{h}}^0-\check{\boldsymbol{h}}^{0,(l)}\right\|_2+\left\|\beta^{0,(l)}\check{\boldsymbol{x}}^0-\check{\boldsymbol{x}}^{0,(l)}\right\|_2\right\}\leq c_3\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^2K\log^5 m}{m}} \qquad (240)$$

for some constant $c_3>0$. Further, this bound combined with (238) yields

$$\max_{1\leq l\leq m}\left|\boldsymbol{b}_l^{\mathsf{H}}\check{\boldsymbol{h}}^0\right|\leq 200\frac{\mu\log^2 m}{\sqrt{m}}+120\sqrt{\frac{\mu^2K\log^3 m}{m}}\cdot c_3\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^2K\log^5 m}{m}}\leq c_2\frac{\mu\log^2 m}{\sqrt{m}} \qquad (241)$$

for some constant $c_2>0$, with the proviso that $m\gg\mu^2K\log^2 m$.

We now translate the preceding bounds to the scaled version. Recall from the bound (231) that

$$1/2\leq 1-\xi\leq\|\boldsymbol{M}\|=\sigma_1(\boldsymbol{M})\leq 1+\xi\leq 2, \qquad (242)$$

as long as $\xi\leq 1/2$. For any $\alpha\in\mathbb{C}$ with $|\alpha|=1$, $\alpha\check{\boldsymbol{h}}^0,\alpha\check{\boldsymbol{x}}^0$ are still the leading left and right singular vectors of $\boldsymbol{M}$. Hence, we can use Lemma 60 to derive that

$$\left|\sigma_1(\boldsymbol{M})-\sigma_1(\boldsymbol{M}^{(l)})\right|\leq\left\|\left(\boldsymbol{M}-\boldsymbol{M}^{(l)}\right)\check{\boldsymbol{x}}^{0,(l)}\right\|_2+\left\{\left\|\alpha\check{\boldsymbol{h}}^0-\check{\boldsymbol{h}}^{0,(l)}\right\|_2+\left\|\alpha\check{\boldsymbol{x}}^0-\check{\boldsymbol{x}}^{0,(l)}\right\|_2\right\}\|\boldsymbol{M}\|$$

$$\leq \left\| \left(\boldsymbol{M} - \boldsymbol{M}^{(l)} \right) \check{\boldsymbol{x}}^{0,(l)} \right\|_2 + 2 \left\{ \left\| \alpha \check{\boldsymbol{h}}^0 - \check{\boldsymbol{h}}^{0,(l)} \right\|_2 + \left\| \alpha \check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(l)} \right\|_2 \right\}$$

and

$$\begin{aligned}
&\left\| \alpha \boldsymbol{h}^0 - \boldsymbol{h}^{0,(l)} \right\|_2 + \left\| \alpha \boldsymbol{x}^0 - \boldsymbol{x}^{0,(l)} \right\|_2 \\
&\quad = \left\| \sqrt{\sigma_1(\boldsymbol{M})} \left(\alpha \check{\boldsymbol{h}}^0 \right) - \sqrt{\sigma_1\left(\boldsymbol{M}^{(l)} \right)} \check{\boldsymbol{h}}^{0,(l)} \right\|_2 + \left\| \sqrt{\sigma_1(\boldsymbol{M})} \alpha \check{\boldsymbol{x}}^0 - \sqrt{\sigma_1\left(\boldsymbol{M}^{(l)} \right)} \check{\boldsymbol{x}}^{0,(l)} \right\|_2 \\
&\quad \leq \sqrt{\sigma_1(\boldsymbol{M})} \left\{ \left\| \alpha \check{\boldsymbol{h}}^0 - \check{\boldsymbol{h}}^{0,(l)} \right\|_2 + \left\| \alpha \check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(l)} \right\|_2 \right\} + \frac{2 \left| \sigma_1(\boldsymbol{M}) - \sigma_1(\boldsymbol{M}^{(l)}) \right|}{\sqrt{\sigma_1(\boldsymbol{M})} + \sqrt{\sigma_1(\boldsymbol{M}^{(l)})}} \\
&\quad \leq \sqrt{2} \left\{ \left\| \alpha \check{\boldsymbol{h}}^0 - \check{\boldsymbol{h}}^{0,(l)} \right\|_2 + \left\| \alpha \check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(l)} \right\|_2 \right\} + \sqrt{2} \left| \sigma_1(\boldsymbol{M}) - \sigma_1(\boldsymbol{M}^{(l)}) \right|.
\end{aligned}$$

Taking the previous two bounds collectively yields

$$\left\| \alpha \boldsymbol{h}^0 - \boldsymbol{h}^{0,(l)} \right\|_2 + \left\| \alpha \boldsymbol{x}^0 - \boldsymbol{x}^{0,(l)} \right\|_2 \leq \sqrt{2} \left\| \left(\boldsymbol{M} - \boldsymbol{M}^{(l)} \right) \check{\boldsymbol{x}}^{0,(l)} \right\|_2 + 6 \left\{ \left\| \alpha \check{\boldsymbol{h}}^0 - \check{\boldsymbol{h}}^{0,(l)} \right\|_2 + \left\| \alpha \check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(l)} \right\|_2 \right\},$$

which together with (235) and (240) implies

$$\min_{\alpha \in \mathbb{C}, |\alpha|=1} \left\{ \left\| \alpha \boldsymbol{h}^0 - \boldsymbol{h}^{0,(l)} \right\|_2 + \left\| \alpha \boldsymbol{x}^0 - \boldsymbol{x}^{0,(l)} \right\|_2 \right\} \leq c_5 \frac{\mu}{\sqrt{m}} \sqrt{\frac{\mu^2 K \log^5 m}{m}} \tag{243}$$

for some constant $c_5 > 0$, as long as ξ is sufficiently small. Moreover, we have

$$\left\| \frac{1}{\overline{\alpha^0}} \boldsymbol{h}^0 - \frac{\alpha}{\overline{\alpha^0}} \boldsymbol{h}^{0,(l)} \right\|_2 + \left\| \alpha^0 \boldsymbol{x}^0 - \alpha \alpha^0 \boldsymbol{x}^{0,(l)} \right\|_2 \leq 2 \left\{ \left\| \boldsymbol{h}^0 - \alpha \boldsymbol{h}^{0,(l)} \right\|_2 + \left\| \boldsymbol{x}^0 - \alpha \boldsymbol{x}^{0,(l)} \right\|_2 \right\}$$

for any $|\alpha| = 1$, where α^0 is defined in (38) and, according to Lemma 19, satisfies

$$1/2 \leq |\alpha^0| \leq 2. \tag{244}$$

Therefore,

$$\begin{aligned}
&\min_{\alpha \in \mathbb{C}, |\alpha|=1} \sqrt{ \left\| \frac{1}{\overline{\alpha^0}} \boldsymbol{h}^0 - \frac{\alpha}{\overline{\alpha^0}} \boldsymbol{h}^{0,(l)} \right\|_2^2 + \left\| \alpha^0 \boldsymbol{x}^0 - \alpha \alpha^0 \boldsymbol{x}^{0,(l)} \right\|_2^2 } \\
&\quad \leq \min_{\alpha \in \mathbb{C}, |\alpha|=1} \left\{ \left\| \frac{1}{\overline{\alpha^0}} \boldsymbol{h}^0 - \frac{\alpha}{\overline{\alpha^0}} \boldsymbol{h}^{0,(l)} \right\|_2 + \left\| \alpha^0 \boldsymbol{x}^0 - \alpha \alpha^0 \boldsymbol{x}^{0,(l)} \right\|_2 \right\} \\
&\quad \leq 2 \min_{\alpha \in \mathbb{C}, |\alpha|=1} \left\{ \left\| \boldsymbol{h}^0 - \alpha \boldsymbol{h}^{0,(l)} \right\|_2 + \left\| \boldsymbol{x}^0 - \alpha \boldsymbol{x}^{0,(l)} \right\|_2 \right\} \\
&\quad \leq 2 c_5 \frac{\mu}{\sqrt{m}} \sqrt{\frac{\mu^2 K \log^5 m}{m}}.
\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
\operatorname{dist}\left(\boldsymbol{z}^{0,(l)}, \widetilde{\boldsymbol{z}}^0 \right) &= \min_{\alpha \in \mathbb{C}} \sqrt{ \left\| \frac{1}{\overline{\alpha}} \boldsymbol{h}^{0,(l)} - \frac{1}{\overline{\alpha^0}} \boldsymbol{h}^0 \right\|_2^2 + \left\| \alpha \boldsymbol{x}^{0,(l)} - \alpha^0 \boldsymbol{x}^0 \right\|_2^2 } \\
&\leq \min_{\alpha \in \mathbb{C}, |\alpha|=1} \sqrt{ \left\| \frac{1}{\overline{\alpha^0}} \boldsymbol{h}^0 - \frac{\alpha}{\overline{\alpha^0}} \boldsymbol{h}^{0,(l)} \right\|_2^2 + \left\| \alpha^0 \boldsymbol{x}^0 - \alpha \alpha^0 \boldsymbol{x}^{0,(l)} \right\|_2^2 } \\
&\leq 2 c_5 \frac{\mu}{\sqrt{m}} \sqrt{\frac{\mu^2 K \log^5 m}{m}},
\end{aligned} \tag{245}$$

where the second line follows since the latter is minimizing over a smaller feasible set. This completes the proof for the claim (96).

Regarding $\left|\boldsymbol{b}_l^{\mathsf{H}}\widetilde{\boldsymbol{h}}^0\right|$, one first sees that

$$\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{h}^0\right| = \left|\sqrt{\sigma_1\left(\boldsymbol{M}\right)}\boldsymbol{b}_l^{\mathsf{H}}\check{\boldsymbol{h}}^0\right| \leq \sqrt{2}c_2\frac{\mu\log^2 m}{\sqrt{m}},$$

where the last relation holds due to (241) and (242). Hence, using the property (244), we have

$$\left|\boldsymbol{b}_l^{\mathsf{H}}\widetilde{\boldsymbol{h}}^0\right| = \left|\boldsymbol{b}_l^{\mathsf{H}}\frac{1}{\overline{\alpha^0}}\boldsymbol{h}^0\right| \leq \left|\frac{1}{\overline{\alpha^0}}\right|\left|\boldsymbol{b}_l^{\mathsf{H}}\boldsymbol{h}^0\right| \leq 2\sqrt{2}c_2\frac{\mu\log^2 m}{\sqrt{m}},$$

which finishes the proof of the claim (97).

Before concluding this section, we note a byproduct of the proof. Specifically, we can establish the claim required in (226) using many results derived in this section. This is formally stated in the following lemma.

Lemma 30. *Fix any small constant $c > 0$. Suppose the number of samples obeys $m \gg \tau K \log^4 m$. Then with probability at least $1 - O\left(m^{-10}\right)$, we have*

$$\max_{1\leq j\leq\tau}\left|\left(\boldsymbol{b}_j - \boldsymbol{b}_1\right)^{\mathsf{H}}\widetilde{\boldsymbol{h}}^0\right| \leq c\frac{\mu}{\sqrt{m}}\log m.$$

Proof. Instate the notation and hypotheses in Appendix C.6. Recognize that

$$\begin{aligned}
\left|\left(\boldsymbol{b}_j - \boldsymbol{b}_1\right)^{\mathsf{H}}\widetilde{\boldsymbol{h}}^0\right| &= \left|\left(\boldsymbol{b}_j - \boldsymbol{b}_1\right)^{\mathsf{H}}\frac{1}{\overline{\alpha^0}}\boldsymbol{h}^0\right| = \left|\left(\boldsymbol{b}_j - \boldsymbol{b}_1\right)^{\mathsf{H}}\frac{1}{\overline{\alpha^0}}\sqrt{\sigma_1\left(\boldsymbol{M}\right)}\check{\boldsymbol{h}}^0\right| \\
&\leq \left|\frac{1}{\overline{\alpha^0}}\right|\sqrt{\sigma_1\left(\boldsymbol{M}\right)}\left|\left(\boldsymbol{b}_j - \boldsymbol{b}_1\right)^{\mathsf{H}}\check{\boldsymbol{h}}^0\right| \\
&\leq 4\left|\left(\boldsymbol{b}_j - \boldsymbol{b}_1\right)^{\mathsf{H}}\check{\boldsymbol{h}}^0\right|,
\end{aligned}$$

where the last inequality comes from (242) and (244). It thus suffices to prove that $\left|\left(\boldsymbol{b}_j - \boldsymbol{b}_1\right)^{\mathsf{H}}\check{\boldsymbol{h}}^0\right| \leq c\mu\log m/\sqrt{m}$ for some $c > 0$ small enough. To this end, it can be seen that

$$\begin{aligned}
\left|\left(\boldsymbol{b}_j - \boldsymbol{b}_1\right)^{\mathsf{H}}\check{\boldsymbol{h}}^0\right| &= \frac{1}{\sigma_1\left(\boldsymbol{M}\right)}\left|\left(\boldsymbol{b}_j - \boldsymbol{b}_1\right)^{\mathsf{H}}\boldsymbol{M}\check{\boldsymbol{x}}^0\right| \\
&\leq 2\left|\sum_{k=1}^m\left(\boldsymbol{b}_j - \boldsymbol{b}_1\right)^{\mathsf{H}}\boldsymbol{b}_k\boldsymbol{b}_k^{\mathsf{H}}\boldsymbol{h}^\star\boldsymbol{x}^{\star\mathsf{H}}\boldsymbol{a}_k\boldsymbol{a}_k^{\mathsf{H}}\check{\boldsymbol{x}}^0\right| \\
&\leq 2\left(\sum_{k=1}^m\left|\left(\boldsymbol{b}_j - \boldsymbol{b}_1\right)^{\mathsf{H}}\boldsymbol{b}_k\right|\right)\max_{1\leq k\leq m}\left\{\left|\boldsymbol{b}_k^{\mathsf{H}}\boldsymbol{h}^\star\right|\left|\boldsymbol{a}_k^{\mathsf{H}}\boldsymbol{x}^\star\right|\left|\boldsymbol{a}_k^{\mathsf{H}}\check{\boldsymbol{x}}^0\right|\right\} \\
&\overset{\text{(i)}}{\leq} c\frac{1}{\log^2 m}\cdot\frac{\mu}{\sqrt{m}}\cdot\left(5\sqrt{\log m}\right)\max_{1\leq j\leq m}\left\{\left|\boldsymbol{a}_j^{\mathsf{H}}\check{\boldsymbol{x}}^{0,(j)}\right| + \left\|\boldsymbol{a}_j\right\|_2\left\|\alpha^{0,(j)}\check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(j)}\right\|_2\right\} \\
&\overset{\text{(ii)}}{\lesssim} c\frac{\mu}{\sqrt{m}}\frac{1}{\log m} \leq c\frac{\mu}{\sqrt{m}}\log m,
\end{aligned} \tag{246}$$

where (i) comes from Lemma 50, the incoherence condition (36), and the estimate (189). The last line (ii) holds since we have already established (see (237) and (240))

$$\max_{1\leq j\leq m}\left\{\left|\boldsymbol{a}_j^{\mathsf{H}}\check{\boldsymbol{x}}^{0,(j)}\right| + \left\|\boldsymbol{a}_j\right\|_2\left\|\alpha^{0,(j)}\check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(j)}\right\|_2\right\} \lesssim \sqrt{\log m}.$$

The proof is then complete. □

C.7 Proof of Lemma 21

Recall that α^0 and $\alpha^{0,(l)}$ are the alignment parameters between $\boldsymbol{z}^0$ and $\boldsymbol{z}^\star$, and between $\boldsymbol{z}^{0,(l)}$ and $\boldsymbol{z}^\star$, respectively, that is,

$$\alpha^0 := \operatorname*{argmin}_{\alpha\in\mathbb{C}}\left\{\left\|\frac{1}{\overline{\alpha}}\boldsymbol{h}^0 - \boldsymbol{h}^\star\right\|_2^2 + \left\|\alpha\boldsymbol{x}^0 - \boldsymbol{x}^\star\right\|_2^2\right\},$$

$$\alpha^{0,(l)} := \underset{\alpha\in\mathbb{C}}{\operatorname{argmin}} \left\{ \left\| \frac{1}{\overline{\alpha}} \boldsymbol{h}^{0,(l)} - \boldsymbol{h}^{\star} \right\|_2^2 + \left\| \alpha \boldsymbol{x}^{0,(l)} - \boldsymbol{x}^{\star} \right\|_2^2 \right\}.$$

Also, we let

$$\alpha_{\text{mutual}}^{0,(l)} := \underset{\alpha\in\mathbb{C}}{\operatorname{argmin}} \left\{ \left\| \frac{1}{\overline{\alpha}} \boldsymbol{h}^{0,(l)} - \frac{1}{\overline{\alpha^0}} \boldsymbol{h}^{0} \right\|_2^2 + \left\| \alpha \boldsymbol{x}^{0,(l)} - \alpha^0 \boldsymbol{x}^{0} \right\|_2^2 \right\}.$$

The triangle inequality together with (94) and (245) then tells us that

$$\begin{aligned}
&\sqrt{\left\| \frac{1}{\overline{\alpha_{\text{mutual}}^{0,(l)}}} \boldsymbol{h}^{0,(l)} - \boldsymbol{h}^{\star} \right\|_2^2 + \left\| \alpha_{\text{mutual}}^{0,(l)} \boldsymbol{x}^{0,(l)} - \boldsymbol{x}^{\star} \right\|_2^2} \\
&\quad \le \sqrt{\left\| \frac{1}{\overline{\alpha^0}} \boldsymbol{h}^{0} - \frac{1}{\overline{\alpha_{\text{mutual}}^{0,(l)}}} \boldsymbol{h}^{0,(l)} \right\|_2^2 + \left\| \alpha^0 \boldsymbol{x}^{0} - \alpha_{\text{mutual}}^{0,(l)} \boldsymbol{x}^{0,(l)} \right\|_2^2} + \sqrt{\left\| \frac{1}{\overline{\alpha^0}} \boldsymbol{h}^{0} - \boldsymbol{h}^{\star} \right\|_2^2 + \left\| \alpha^0 \boldsymbol{x}^{0} - \boldsymbol{x}^{\star} \right\|_2^2} \\
&\quad \le 2c_5 \frac{\mu}{\sqrt{m}} \sqrt{\frac{\mu^2 K \log^5 m}{m}} + C_1 \frac{1}{\log^2 m} \\
&\quad \le 2C_1 \frac{1}{\log^2 m},
\end{aligned}$$

where the last relation holds as long as $m \gg \mu^2 \sqrt{K} \log^{9/2} m$.

Let

$$\boldsymbol{x}_1 = \alpha^0 \boldsymbol{x}^0, \quad \boldsymbol{h}_1 = \frac{1}{\overline{\alpha^0}} \boldsymbol{h}_0 \qquad \text{and} \qquad \boldsymbol{x}_2 = \alpha_{\text{mutual}}^{0,(l)} \boldsymbol{x}^{0,(l)}, \quad \boldsymbol{h}_2 = \frac{1}{\overline{\alpha_{\text{mutual}}^{0,(l)}}} \boldsymbol{h}^{0,(l)}.$$

It is easy to see that $\boldsymbol{x}_1, \boldsymbol{h}_1, \boldsymbol{x}_2, \boldsymbol{h}_2$ satisfy the assumptions in Lemma 55, which implies

$$\begin{aligned}
\sqrt{\left\| \frac{1}{\overline{\alpha^{0,(l)}}} \boldsymbol{h}^{0,(l)} - \frac{1}{\overline{\alpha^0}} \boldsymbol{h}^{0} \right\|_2^2 + \left\| \alpha^{0,(l)} \boldsymbol{x}^{0,(l)} - \alpha^0 \boldsymbol{x}^{0} \right\|_2^2} &\lesssim \sqrt{\left\| \frac{1}{\overline{\alpha^0}} \boldsymbol{h}^{0} - \frac{1}{\overline{\alpha_{\text{mutual}}^{0,(l)}}} \boldsymbol{h}^{0,(l)} \right\|_2^2 + \left\| \alpha^0 \boldsymbol{x}^{0} - \alpha_{\text{mutual}}^{0,(l)} \boldsymbol{x}^{0,(l)} \right\|_2^2} \\
&\lesssim \frac{\mu}{\sqrt{m}} \sqrt{\frac{\mu^2 K \log^5 m}{m}}, \qquad (247)
\end{aligned}$$

where the last line comes from (245). With this upper estimate at hand, we are now ready to show that with high probability,

$$\begin{aligned}
\left| \boldsymbol{a}_l^{\mathsf{H}} \left(\alpha^0 \boldsymbol{x}^0 - \boldsymbol{x}^{\star} \right) \right| &\overset{\text{(i)}}{\le} \left| \boldsymbol{a}_l^{\mathsf{H}} \left(\alpha^{0,(l)} \boldsymbol{x}^{0,(l)} - \boldsymbol{x}^{\star} \right) \right| + \left| \boldsymbol{a}_l^{\mathsf{H}} \left(\alpha^0 \boldsymbol{x}^0 - \alpha^{0,(l)} \boldsymbol{x}^{0,(l)} \right) \right| \\
&\overset{\text{(ii)}}{\le} 5\sqrt{\log m} \left\| \alpha^{0,(l)} \boldsymbol{x}^{0,(l)} - \boldsymbol{x}^{\star} \right\|_2 + \|\boldsymbol{a}_l\|_2 \left\| \alpha^0 \boldsymbol{x}^0 - \alpha^{0,(l)} \boldsymbol{x}^{0,(l)} \right\|_2 \\
&\overset{\text{(iii)}}{\lesssim} \sqrt{\log m} \cdot \frac{1}{\log^2 m} + \sqrt{K} \frac{\mu}{\sqrt{m}} \sqrt{\frac{\mu^2 K \log^5 m}{m}} \\
&\overset{\text{(iv)}}{\lesssim} \frac{1}{\log^{3/2} m},
\end{aligned}$$

where (i) follows from the triangle inequality, (ii) uses Cauchy-Schwarz and the independence between $\boldsymbol{x}^{0,(l)}$ and $\boldsymbol{a}_l$, (iii) holds because of (95) and (247) under the condition $m \gg \mu^2 K \log^6 m$, and (iv) holds true as long as $m \gg \mu^2 K \log^4 m$.

D Technical lemmas

D.1 Technical lemmas for phase retrieval

D.1.1 Matrix concentration inequalities

Lemma 31. *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 [Ver12, Corollary 5.35]. □

Lemma 32. *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*

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

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

Proof. This is adapted from [CLS15, Lemma 7.4]. □

Lemma 33. *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*

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

with ξ *being a standard Gaussian random variable.*

Proof. This is supplied in [CC17, supplementary material]. □

D.1.2 Matrix perturbation bounds

Lemma 34. *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

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

$$\leq \left|\sqrt{\lambda_1(\boldsymbol{A})} - \sqrt{\lambda_1(\widetilde{\boldsymbol{A}})}\right| + \sqrt{\lambda_1(\widetilde{\boldsymbol{A}})}\,\|\boldsymbol{u}-\widetilde{\boldsymbol{u}}\|_2\,, \tag{248}$$

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 (248) yields

$$\left\|\sqrt{\lambda_1(\boldsymbol{A})}\,\boldsymbol{u} - \sqrt{\lambda_1(\widetilde{\boldsymbol{A}})}\,\widetilde{\boldsymbol{u}}\right\|_2 \leq \frac{\left|\lambda_1(\boldsymbol{A}) - \lambda_1(\widetilde{\boldsymbol{A}})\right|}{2\sqrt{C_1}} + \sqrt{C_2}\,\|\boldsymbol{u}-\widetilde{\boldsymbol{u}}\|_2\,. \tag{249}$$

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 (249) 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. □

D.2 Technical lemmas for matrix completion

D.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

$$\text{minimize}_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \quad \|\boldsymbol{A}\boldsymbol{R}-\boldsymbol{B}\|_{\mathrm{F}}\,. \tag{250}$$

The first lemma is concerned with the characterization of the minimizer $\widehat{\boldsymbol{R}}$ of (250).

Lemma 35. *For $\boldsymbol{A},\boldsymbol{B} \in \mathbb{R}^{n\times r}$, $\widehat{\boldsymbol{R}}$ is the minimizer of (250) 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 [TB77, 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 35, $\widehat{\boldsymbol{R}} = \boldsymbol{U}\boldsymbol{V}^\top$ is the minimizer of (250). In the special case when $\boldsymbol{C} := \boldsymbol{A}^\top\boldsymbol{B}$ is invertible, $\widehat{\boldsymbol{R}}$ enjoys the following equivalent form:

$$\widehat{\boldsymbol{R}} = \widehat{\boldsymbol{H}}(\boldsymbol{C}) := \boldsymbol{C}\left(\boldsymbol{C}^\top\boldsymbol{C}\right)^{-1/2}, \tag{251}$$

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 36. *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 $\|\boldsymbol{E}\| \leq \sigma_{\min}(\boldsymbol{C})$ and any unitarily invariant norm $|||\cdot|||$, one has*

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

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

Proof. This is an immediate consequence of [Mat93, Theorem 2.3]. $\square$

With Lemma 36 in place, we are ready to present the following bounds on two matrices after "aligning" them with $\boldsymbol{X}^\star$.

Lemma 37. *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*

$$\|\boldsymbol{X}_1 - \boldsymbol{X}^\star\| \|\boldsymbol{X}^\star\| \leq \sigma_{\min}/2, \tag{252a}$$
$$\|\boldsymbol{X}_1 - \boldsymbol{X}_2\| \|\boldsymbol{X}^\star\| \leq \sigma_{\min}/4. \tag{252b}$$

Denote

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

Then the following two inequalities hold true:

$$\|\boldsymbol{X}_1\boldsymbol{R}_1 - \boldsymbol{X}_2\boldsymbol{R}_2\| \leq 5\kappa \|\boldsymbol{X}_1 - \boldsymbol{X}_2\| \qquad \text{and} \qquad \|\boldsymbol{X}_1\boldsymbol{R}_1 - \boldsymbol{X}_2\boldsymbol{R}_2\|_{\mathrm{F}} \leq 5\kappa \|\boldsymbol{X}_1 - \boldsymbol{X}_2\|_{\mathrm{F}}.$$

Proof. Before proving the claims, we first gather some immediate consequences of the assumptions (252). Denote $\boldsymbol{C} = \boldsymbol{X}_1^\top \boldsymbol{X}^\star$ and $\boldsymbol{E} = (\boldsymbol{X}_2 - \boldsymbol{X}_1)^\top \boldsymbol{X}^\star$. It is easily seen that $\boldsymbol{C}$ is invertible since

$$\left\|\boldsymbol{C} - \boldsymbol{X}^{\star\top}\boldsymbol{X}^\star\right\| \leq \|\boldsymbol{X}_1 - \boldsymbol{X}^\star\| \|\boldsymbol{X}^\star\| \overset{\text{(i)}}{\leq} \sigma_{\min}/2 \qquad \overset{\text{(ii)}}{\Longrightarrow} \qquad \sigma_r(\boldsymbol{C}) \geq \sigma_{\min}/2, \tag{253}$$

where (i) follows from the assumption (252a) and (ii) is a direct application of Weyl's inequality. In addition, $\boldsymbol{C} + \boldsymbol{E} = \boldsymbol{X}_2^\top \boldsymbol{X}^\star$ is also invertible since

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

where (i) arises from the assumption (252b) and (ii) holds because of (253). 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 $\|\boldsymbol{X}_1\|$:

$$\|\boldsymbol{X}_1\| \overset{\text{(i)}}{\leq} \|\boldsymbol{X}_1 - \boldsymbol{X}^\star\| + \|\boldsymbol{X}^\star\| \overset{\text{(ii)}}{\leq} \frac{\sigma_{\min}}{2\|\boldsymbol{X}^\star\|} + \|\boldsymbol{X}^\star\| \leq \frac{\sigma_{\max}}{2\|\boldsymbol{X}^\star\|} + \|\boldsymbol{X}^\star\| \overset{\text{(iii)}}{\leq} 2\|\boldsymbol{X}^\star\|, \tag{254}$$

where (i) is the triangle inequality, (ii) uses the assumption (252a) and (iii) arises from the fact that $\|\boldsymbol{X}^\star\| = \sqrt{\sigma_{\max}}$.

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

$$\begin{aligned}
\|\boldsymbol{X}_1\boldsymbol{R}_1 - \boldsymbol{X}_2\boldsymbol{R}_2\|_{\mathrm{F}} &= \|(\boldsymbol{X}_1 - \boldsymbol{X}_2)\boldsymbol{R}_2 + \boldsymbol{X}_1(\boldsymbol{R}_1 - \boldsymbol{R}_2)\|_{\mathrm{F}} \\
&\leq \|\boldsymbol{X}_1 - \boldsymbol{X}_2\|_{\mathrm{F}} + \|\boldsymbol{X}_1\| \|\boldsymbol{R}_1 - \boldsymbol{R}_2\|_{\mathrm{F}} \\
&\leq \|\boldsymbol{X}_1 - \boldsymbol{X}_2\|_{\mathrm{F}} + 2\|\boldsymbol{X}^\star\| \|\boldsymbol{R}_1 - \boldsymbol{R}_2\|_{\mathrm{F}},
\end{aligned} \tag{255}$$

where the first inequality uses the fact that $\|\boldsymbol{R}_2\| = 1$ and the last inequality comes from (254). An application of Lemma 36 leads us to conclude that

$$
\begin{aligned}
\|\boldsymbol{R}_1 - \boldsymbol{R}_2\|_{\mathrm{F}} &\le \frac{2}{\sigma_r(\boldsymbol{C}) + \sigma_{r-1}(\boldsymbol{C})} \|\boldsymbol{E}\|_{\mathrm{F}} \\
&\le \frac{2}{\sigma_{\min}} \left\| (\boldsymbol{X}_2 - \boldsymbol{X}_1)^\top \boldsymbol{X}^\star \right\|_{\mathrm{F}} \qquad (256) \\
&\le \frac{2}{\sigma_{\min}} \|\boldsymbol{X}_2 - \boldsymbol{X}_1\|_{\mathrm{F}} \|\boldsymbol{X}^\star\|, \qquad (257)
\end{aligned}
$$

where (256) utilizes (253). Combine (255) and (257) to reach

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

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

D.2.2 Matrix concentration inequalities

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

Lemma 38. *Fix any small constant $\delta > 0$, and suppose that $m \gg \delta^{-2}\mu n r \log n$. Then with probability exceeding $1 - O(n^{-10})$, one has*

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

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

Proof. This result has been established in [CR09, Section 4.2] for asymmetric sampling patterns (where each (i,j), $i \ne j$ is included in Ω independently). It is straightforward to extend the proof and the result to symmetric sampling patterns (where each (i,j), $i \ge j$ is included in Ω independently). We omit the proof for conciseness. □

Lemma 39. *Fix a matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$. Suppose $n^2 p \ge c_0 n \log n$ for some sufficiently large constant $c_0 > 0$. With probability at least $1 - O(n^{-10})$, one has*

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

where $C > 0$ is some absolute constant.

Proof. See [KMO10a, Lemma 3.2]. Similar to Lemma 38, the result therein was provided for the asymmetric sampling patterns but can be easily extended to the symmetric case. □

Lemma 40. *Recall from Section 3.2 that $\boldsymbol{E} \in \mathbb{R}^{n\times n}$ is the symmetric noise matrix. Suppose the sample size obeys $n^2 p \ge c_0 n \log^2 n$ for some sufficiently large constant $c_0 > 0$. With probability at least $1 - O(n^{-10})$, one has*

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

where $C > 0$ is some universal constant.

Proof. See [CW15, Lemma 11]. □

Lemma 41. *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 [Tro15b, Theorem 6.1.1], one has for all $t \geq 0$,

$$\mathbb{P}\left\{\left\|\sum_{j=1}^{n}(\delta_{l,j} - p_j)\boldsymbol{A}_{j,\cdot}^{\top}\boldsymbol{A}_{j,\cdot}\right\| \geq t\right\} \leq 2r\cdot\exp\left(\frac{-t^2/2}{p\|\boldsymbol{A}\|_{2,\infty}^2\|\boldsymbol{A}\|^2 + \|\boldsymbol{A}\|_{2,\infty}^2\cdot t/3}\right). \tag{258}$$

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 (258) 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 $1-n^{-(1.5C-1)}$,

$$\left\|\boldsymbol{G}_l\left(\boldsymbol{A}\right)\right\|^2 \le p\left\|\boldsymbol{A}\right\|^2 + C\left(\sqrt{p\left\|\boldsymbol{A}\right\|_{2,\infty}^2\left\|\boldsymbol{A}\right\|^2\log n} + \left\|\boldsymbol{A}\right\|_{2,\infty}^2\log n\right).$$

This completes the proof. $\square$

Lemma 42. *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*

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

Suppose the sample size obeys $n^2p \gg \kappa\mu r n\log^2 n$. Then for any $k>0$ and $\alpha>0$ large enough, with probability at least $1-c_1 e^{-\alpha Cnr\log n/2}$,

$$\sum_{l=1}^{n}\mathbb{1}_{\left\{\|\boldsymbol{G}_l(\boldsymbol{\Delta})\|\ge 4\sqrt{p}\psi+2\sqrt{kr}\xi\right\}} \le \frac{2\alpha n\log n}{k}$$

holds simultaneously for all $\boldsymbol{\Delta}\in\mathbb{R}^{n\times r}$ obeying

$$\begin{aligned}
&\left\|\boldsymbol{\Delta}\right\|_{2,\infty} \le C_5\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^\star\right\|_{2,\infty} + C_8\sigma\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\star\right\|_{2,\infty} := \xi \\
\text{and}\qquad &\left\|\boldsymbol{\Delta}\right\| \le C_9\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\star\right\| + C_{10}\sigma\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\star\right\| := \psi,
\end{aligned}$$

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 $\boldsymbol{\Delta},\widetilde{\boldsymbol{\Delta}}\in\mathbb{R}^{n\times r}$,

$$\begin{aligned}
\left|\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right)\right\| - \left\|\boldsymbol{G}_l\big(\widetilde{\boldsymbol{\Delta}}\big)\right\|\right| &\le \left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right) - \boldsymbol{G}_l\big(\widetilde{\boldsymbol{\Delta}}\big)\right\| \le \left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right) - \boldsymbol{G}_l\big(\widetilde{\boldsymbol{\Delta}}\big)\right\|_{\mathrm{F}} \\
&\le \sqrt{\sum_{j=1}^{n}\left\|\boldsymbol{\Delta}_{j,\cdot} - \widetilde{\boldsymbol{\Delta}}_{j,\cdot}\right\|_2^2} \\
&:= d\big(\boldsymbol{\Delta},\widetilde{\boldsymbol{\Delta}}\big),
\end{aligned}$$

which implies that $\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right)\right\|$ is 1-Lipschitz with respect to the metric $d\left(\cdot,\cdot\right)$. Moreover,

$$\max_{1\le j\le n}\left\|\delta_{l,j}\boldsymbol{\Delta}_{j,\cdot}\right\|_2 \le \left\|\boldsymbol{\Delta}\right\|_{2,\infty} \le \xi$$

according to our assumption. Hence, Talagrand's inequality [CC18, Proposition 1] reveals the existence of some absolute constants $C,c>0$ such that for all $\lambda>0$

$$\mathbb{P}\left\{\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right)\right\| - \mathsf{Median}\left[\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right)\right\|\right] \ge \lambda\xi\right\} \le C\exp\left(-c\lambda^2\right). \tag{259}$$

We then proceed to control $\mathsf{Median}\left[\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right)\right\|\right]$. A direct application of Lemma 41 yields

$$\mathsf{Median}\left[\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right)\right\|\right] \le \sqrt{2p\psi^2 + \sqrt{p\log\left(4r\right)}\xi\psi + \frac{2\xi^2}{3}\log\left(4r\right)} \le 2\sqrt{p}\psi,$$

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 (114). Thus, substitution into (259) and taking $\lambda = \sqrt{kr}$ give

$$\mathbb{P}\left\{\|\boldsymbol{G}_l(\boldsymbol{\Delta})\| \geq 2\sqrt{p}\psi + \sqrt{kr}\xi\right\} \leq C\exp(-ckr) \tag{260}$$

for any $k \geq 0$. Furthermore, invoking [AS08, Corollary A.1.14] and using the bound (260), one has

$$\mathbb{P}\left(\sum_{l=1}^{n} \mathbb{1}_{\left\{\|\boldsymbol{G}_l(\boldsymbol{\Delta})\| \geq 2\sqrt{p}\psi + \sqrt{kr}\xi\right\}} \geq tnC\exp(-ckr)\right) \leq 2\exp\left(-\frac{t\log t}{2} nC\exp(-ckr)\right)$$

for any $t \geq 6$. Choose $t = \alpha \log n / [kC\exp(-ckr)] \geq 6$ to obtain

$$\mathbb{P}\left(\sum_{l=1}^{n} \mathbb{1}_{\left\{\|\boldsymbol{G}_l(\boldsymbol{\Delta})\| \geq 2\sqrt{p}\psi + \sqrt{kr}\xi\right\}} \geq \frac{\alpha n \log n}{k}\right) \leq 2\exp\left(-\frac{\alpha C}{2} nr\log n\right). \tag{261}$$

So far we have demonstrated that for any fixed $\boldsymbol{\Delta}$ obeying our assumptions, $\sum_{l=1}^{n} \mathbb{1}_{\left\{\|\boldsymbol{G}_l(\boldsymbol{\Delta})\| \geq 2\sqrt{p}\psi + \sqrt{kr}\xi\right\}}$ 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 $\|\boldsymbol{G}_l(\boldsymbol{\Delta})\|$, it suffices to restrict attention to the following set:

$$\mathcal{S} = \left\{\boldsymbol{\Delta} \mid \min\{\xi, \psi\} \leq \|\boldsymbol{\Delta}\| \leq \psi\right\}, \tag{262}$$

where $\psi/\xi \lesssim \|\boldsymbol{X}^\star\|/\|\boldsymbol{X}^\star\|_{2,\infty} \lesssim \sqrt{n}$. We then proceed with the following steps.

1. Introduce the auxiliary function

$$\chi_l(\boldsymbol{\Delta}) = \begin{cases} 1, & \text{if } \|\boldsymbol{G}_l(\boldsymbol{\Delta})\| \geq 4\sqrt{p}\psi + 2\sqrt{kr}\xi, \\ \frac{\|\boldsymbol{G}_l(\boldsymbol{\Delta})\| - 2\sqrt{p}\psi - \sqrt{kr}\xi}{2\sqrt{p}\psi + \sqrt{kr}\xi}, & \text{if } \|\boldsymbol{G}_l(\boldsymbol{\Delta})\| \in [2\sqrt{p}\psi + \sqrt{kr}\xi,\ 4\sqrt{p}\psi + 2\sqrt{kr}\xi], \\ 0, & \text{else.} \end{cases}$$

   Clearly, this function is sandwiched between two indicator functions

$$\mathbb{1}_{\left\{\|\boldsymbol{G}_l(\boldsymbol{\Delta})\| \geq 4\sqrt{p}\psi + 2\sqrt{kr}\xi\right\}} \leq \chi_l(\boldsymbol{\Delta}) \leq \mathbb{1}_{\left\{\|\boldsymbol{G}_l(\boldsymbol{\Delta})\| \geq 2\sqrt{p}\psi + \sqrt{kr}\xi\right\}}.$$

   Note that χ_l is more convenient to work with due to continuity.

2. Consider an ϵ-net $\mathcal{N}_\epsilon$ [Tao12, Section 2.3.1] of the set $\mathcal{S}$ as defined in (262). 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 (261) to yield

$$\begin{aligned} \mathbb{P}\left(\sum_{l=1}^{n} \chi_l(\boldsymbol{\Delta}) \geq \frac{\alpha n \log n}{k},\ \forall \boldsymbol{\Delta} \in \mathcal{N}_\epsilon\right) &\leq \mathbb{P}\left(\sum_{l=1}^{n} \mathbb{1}_{\left\{\|\boldsymbol{G}_l(\boldsymbol{\Delta})\| \geq 2\sqrt{p}\psi + \sqrt{kr}\xi\right\}} \geq \frac{\alpha n \log n}{k},\ \forall \boldsymbol{\Delta} \in \mathcal{N}_\epsilon\right) \\ &\leq 2|\mathcal{N}_\epsilon|\exp\left(-\frac{\alpha C}{2} nr\log n\right) \leq 2\exp\left(-\frac{\alpha C}{4} nr\log n\right), \end{aligned}$$

   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,

$$\sum_{l=1}^{n} \chi_l(\boldsymbol{\Delta}) \leq \frac{2\alpha n \log n}{k}, \quad \forall \boldsymbol{\Delta} \in \mathcal{S}$$

$$\Longrightarrow \qquad \sum_{l=1}^{n} \mathbb{1}_{\left\{\|\boldsymbol{G}_l(\boldsymbol{\Delta})\| \geq 4\sqrt{p}\psi + 2\sqrt{kr}\xi\right\}} \leq \sum_{l=1}^{n} \chi_l(\boldsymbol{\Delta}) \leq \frac{2\alpha n \log n}{k}, \quad \forall \boldsymbol{\Delta} \in \mathcal{S}$$

   This is fairly standard (see, e.g. [Tao12, Section 2.3.1]) and is thus omitted here.

We have thus concluded the proof. $\square$

Lemma 43. *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 $1 - O\left(n^{-10}\right)$,*

$$\left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{X}\boldsymbol{X}^{\top} - \boldsymbol{X}^{\star}\boldsymbol{X}^{\star\top}\right) \right\| \leq 2n\epsilon^2 \left\|\boldsymbol{X}^{\star}\right\|_{2,\infty}^2 + 4\epsilon\sqrt{n}\log n \left\|\boldsymbol{X}^{\star}\right\|_{2,\infty} \left\|\boldsymbol{X}^{\star}\right\|$$

holds simultaneously for all $\boldsymbol{X} \in \mathbb{R}^{n\times r}$ satisfying

$$\left\|\boldsymbol{X} - \boldsymbol{X}^{\star}\right\|_{2,\infty} \leq \epsilon \left\|\boldsymbol{X}^{\star}\right\|_{2,\infty}, \tag{263}$$

where $\epsilon > 0$ is any fixed constant.

Proof. To simplify the notations hereafter, we denote $\boldsymbol{\Delta} := \boldsymbol{X} - \boldsymbol{X}^{\star}$. With this notation in place, one can decompose

$$\boldsymbol{X}\boldsymbol{X}^{\top} - \boldsymbol{X}^{\star}\boldsymbol{X}^{\star\top} = \boldsymbol{\Delta}\boldsymbol{X}^{\star\top} + \boldsymbol{X}^{\star}\boldsymbol{\Delta}^{\top} + \boldsymbol{\Delta}\boldsymbol{\Delta}^{\top},$$

which together with the triangle inequality implies that

$$\begin{aligned}\left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{X}\boldsymbol{X}^{\top} - \boldsymbol{X}^{\star}\boldsymbol{X}^{\star\top}\right) \right\| &\leq \left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{\Delta}\boldsymbol{X}^{\star\top}\right) \right\| + \left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{X}^{\star}\boldsymbol{\Delta}^{\top}\right) \right\| + \left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{\Delta}\boldsymbol{\Delta}^{\top}\right) \right\| \\ &= \underbrace{\left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{\Delta}\boldsymbol{\Delta}^{\top}\right) \right\|}_{:=\alpha_1} + 2\underbrace{\left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{\Delta}\boldsymbol{X}^{\star\top}\right) \right\|}_{:=\alpha_2}.\end{aligned} \tag{264}$$

In the sequel, we bound α_1 and α_2 separately.

1. Recall from [Mat90, Theorem 2.5] the elementary inequality that

$$\|\boldsymbol{C}\| \leq \left\| |\boldsymbol{C}| \right\|, \tag{265}$$

   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 $\left\| |\boldsymbol{C}| \right\| \leq \left\| |\boldsymbol{D}| \right\|$. Therefore

$$\alpha_1 \leq \left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\left|\boldsymbol{\Delta}\boldsymbol{\Delta}^{\top}\right|\right) \right\| \leq \left\|\boldsymbol{\Delta}\right\|_{2,\infty}^2 \left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\mathbf{1}\mathbf{1}^{\top}\right) \right\|.$$

   Lemma 39 then tells us that with probability at least $1 - O(n^{-10})$,

$$\left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\mathbf{1}\mathbf{1}^{\top}\right) - \mathbf{1}\mathbf{1}^{\top} \right\| \leq C\sqrt{\frac{n}{p}} \tag{266}$$

   for some universal constant $C > 0$, as long as $p \gg \log n/n$. This together with the triangle inequality yields

$$\left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\mathbf{1}\mathbf{1}^{\top}\right) \right\| \leq \left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\mathbf{1}\mathbf{1}^{\top}\right) - \mathbf{1}\mathbf{1}^{\top} \right\| + \left\|\mathbf{1}\mathbf{1}^{\top}\right\| \leq C\sqrt{\frac{n}{p}} + n \leq 2n, \tag{267}$$

   provided that $p \gg 1/n$. Putting together the previous bounds, we arrive at

$$\alpha_1 \leq 2n \left\|\boldsymbol{\Delta}\right\|_{2,\infty}^2. \tag{268}$$

2. Regarding the second term α_2, apply the elementary inequality (265) once again to get

$$\left\|\mathcal{P}_{\Omega}\left(\boldsymbol{\Delta}\boldsymbol{X}^{\star\top}\right)\right\| \leq \left\|\mathcal{P}_{\Omega}\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\star\top}\right|\right)\right\|,$$

   which motivates us to look at $\left\|\mathcal{P}_{\Omega}\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\star\top}\right|\right)\right\|$ instead. A key step of this part is to take advantage of the $\ell_{2,\infty}$ norm constraint of $\mathcal{P}_{\Omega}\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\star\top}\right|\right)$. Specifically, we claim for the moment that with probability exceeding $1 - O(n^{-10})$,

$$\left\|\mathcal{P}_{\Omega}\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\star\top}\right|\right)\right\|_{2,\infty}^2 \leq 2p\sigma_{\max}\left\|\boldsymbol{\Delta}\right\|_{2,\infty}^2 := \theta \tag{269}$$

holds under our sample size condition. In addition, we also have the following trivial ℓ_∞ norm bound

$$\left\|\mathcal{P}_\Omega\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\star\top}\right|\right)\right\|_\infty \leq \|\boldsymbol{\Delta}\|_{2,\infty}\|\boldsymbol{X}^\star\|_{2,\infty} := \gamma. \tag{270}$$

In what follows, for simplicity of presentation, we will denote

$$\boldsymbol{A} := \mathcal{P}_\Omega\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\star\top}\right|\right). \tag{271}$$

(a) To facilitate the analysis of $\|\boldsymbol{A}\|$, we first introduce $k_0+1=\frac{1}{2}\log(\kappa\mu r)$ auxiliary matrices[9] $\boldsymbol{B}_s \in \mathbb{R}^{n\times n}$ that satisfy

$$\|\boldsymbol{A}\| \leq \|\boldsymbol{B}_{k_0}\| + \sum_{s=0}^{k_0-1}\|\boldsymbol{B}_s\|. \tag{272}$$

To be precise, each $\boldsymbol{B}_s$ is defined such that

$$[\boldsymbol{B}_s]_{j,k} = \begin{cases} \frac{1}{2^s}\gamma, & \text{if} \quad A_{j,k}\in(\frac{1}{2^{s+1}}\gamma, \frac{1}{2^s}\gamma], \\ 0, & \text{else}, \end{cases} \qquad \text{for } 0\leq s\leq k_0-1 \qquad \text{and}$$

$$[\boldsymbol{B}_{k_0}]_{j,k} = \begin{cases} \frac{1}{2^{k_0}}\gamma, & \text{if} \quad A_{j,k}\leq \frac{1}{2^{k_0}}\gamma, \\ 0, & \text{else}, \end{cases}$$

which clearly satisfy (272); 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

$$\|\boldsymbol{B}_{k_0}\| \overset{\text{(i)}}{\leq} \|\boldsymbol{B}_{k_0}\|_\infty\sqrt{(2np)^2} \overset{\text{(ii)}}{\leq} 4np\frac{1}{\sqrt{\kappa\mu r}}\|\boldsymbol{\Delta}\|_{2,\infty}\|\boldsymbol{X}^\star\|_{2,\infty} \overset{\text{(iii)}}{\leq} 4\sqrt{np}\|\boldsymbol{\Delta}\|_{2,\infty}\|\boldsymbol{X}^\star\|,$$

where (i) arises from Lemma 44, 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 (114). Besides, for any $0\leq s\leq k_0-1$, by construction one has

$$\|\boldsymbol{B}_s\|_{2,\infty}^2 \leq 4\theta = 8p\sigma_{\max}\|\boldsymbol{\Delta}\|_{2,\infty}^2 \qquad \text{and} \qquad \|\boldsymbol{B}_s\|_\infty = \frac{1}{2^s}\gamma,$$

where θ is as defined in (269). 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

$$\frac{\|\boldsymbol{B}_s\|_{2,\infty}^2}{\|\boldsymbol{B}_s\|_\infty^2} \leq \frac{8p\sigma_{\max}\|\boldsymbol{\Delta}\|_{2,\infty}^2}{\left(\frac{1}{2^s}\gamma\right)^2} := k_{\mathrm{r}}$$

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 44 once more, we discover that

$$\|\boldsymbol{B}_s\| \leq \|\boldsymbol{B}_s\|_\infty\sqrt{k_{\mathrm{r}}k_{\mathrm{c}}} = \frac{1}{2^s}\gamma\sqrt{k_{\mathrm{r}}k_{\mathrm{c}}} = \sqrt{16np^2\sigma_{\max}\|\boldsymbol{\Delta}\|_{2,\infty}^2} = 4\sqrt{np}\|\boldsymbol{\Delta}\|_{2,\infty}\|\boldsymbol{X}^\star\|$$

for each $0\leq s\leq k_0-1$. Combining all, we arrive at

$$\|\boldsymbol{A}\| \leq \sum_{s=0}^{k_0-1}\|\boldsymbol{B}_s\| + \|\boldsymbol{B}_{k_0}\| \leq (k_0+1)\,4\sqrt{np}\|\boldsymbol{\Delta}\|_{2,\infty}\|\boldsymbol{X}^\star\|$$

[9]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.

$$
\begin{aligned}
&\leq 2\sqrt{n}p\log(\kappa\mu r)\|\boldsymbol{\Delta}\|_{2,\infty}\|\boldsymbol{X}^\star\| \\
&\leq 2\sqrt{n}p\log n\|\boldsymbol{\Delta}\|_{2,\infty}\|\boldsymbol{X}^\star\|,
\end{aligned}
$$

where the last relation holds under the condition $n \geq \kappa\mu r$. This further gives

$$
\alpha_2 \leq \frac{1}{p}\|\boldsymbol{A}\| \leq 2\sqrt{n}\log n\|\boldsymbol{\Delta}\|_{2,\infty}\|\boldsymbol{X}^\star\|. \tag{273}
$$

(b) In order to finish the proof of this part, we need to justify the claim (269). Observe that

$$
\begin{aligned}
\left\|\left[\mathcal{P}_\Omega\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\star\top}\right|\right)\right]_{l,\cdot}\right\|_2^2 &= \sum\nolimits_{j=1}^n \left(\boldsymbol{\Delta}_{l,\cdot}\boldsymbol{X}_{j,\cdot}^{\star\top}\delta_{l,j}\right)^2 \\
&= \boldsymbol{\Delta}_{l,\cdot}\left(\sum\nolimits_{j=1}^n \delta_{l,j}\boldsymbol{X}_{j,\cdot}^{\star\top}\boldsymbol{X}_{j,\cdot}^\star\right)\boldsymbol{\Delta}_{l,\cdot}^\top \\
&\leq \|\boldsymbol{\Delta}\|_{2,\infty}^2\left\|\sum\nolimits_{j=1}^n \delta_{l,j}\boldsymbol{X}_{j,\cdot}^{\star\top}\boldsymbol{X}_{j,\cdot}^\star\right\|
\end{aligned} \tag{274}
$$

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 41 to yield

$$
\begin{aligned}
\left\|\sum\nolimits_{j=1}^n \delta_{l,j}\boldsymbol{X}_{j,\cdot}^{\star\top}\boldsymbol{X}_{j,\cdot}^\star\right\| &= \left\|\left[\delta_{l,1}\boldsymbol{X}_{1,\cdot}^{\star\top}, \delta_{l,2}\boldsymbol{X}_{2,\cdot}^{\star\top}, \cdots, \delta_{l,n}\boldsymbol{X}_{n,\cdot}^{\star\top}\right]\right\|^2 \\
&\leq p\sigma_{\max} + C\left(\sqrt{p\|\boldsymbol{X}^\star\|_{2,\infty}^2\|\boldsymbol{X}^\star\|^2\log n} + \|\boldsymbol{X}^\star\|_{2,\infty}^2\log n\right) \\
&\leq \left(p + C\sqrt{\frac{p\kappa\mu r\log n}{n}} + C\frac{\kappa\mu r\log n}{n}\right)\sigma_{\max} \\
&\leq 2p\sigma_{\max},
\end{aligned} \tag{275}
$$

with high probability, as soon as $np \gg \kappa\mu r\log n$. Combining (274) and (275) yields

$$
\left\|\left[\mathcal{P}_\Omega\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\star\top}\right|\right)\right]_{l,\cdot}\right\|_2^2 \leq 2p\sigma_{\max}\|\boldsymbol{\Delta}\|_{2,\infty}^2, \qquad 1 \leq l \leq n
$$

as claimed in (269).

3. Taken together, the preceding bounds (264), (268) and (273) yield

$$
\left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{X}\boldsymbol{X}^\top - \boldsymbol{X}^\star\boldsymbol{X}^{\star\top}\right)\right\| \leq \alpha_1 + 2\alpha_2 \leq 2n\|\boldsymbol{\Delta}\|_{2,\infty}^2 + 4\sqrt{n}\log n\|\boldsymbol{\Delta}\|_{2,\infty}\|\boldsymbol{X}^\star\|.
$$

The proof is completed by substituting the assumption $\|\boldsymbol{\Delta}\|_{2,\infty} \leq \epsilon\|\boldsymbol{X}^\star\|_{2,\infty}$. □

In the end of this subsection, we record a useful lemma to bound the spectral norm of a sparse Bernoulli matrix.

Lemma 44. *Let $\boldsymbol{A} \in \{0,1\}^{n_1\times n_2}$ be a binary matrix, and suppose that there are at most k_{r} and k_{c} nonzero entries in each row and column of $\boldsymbol{A}$, respectively. Then one has $\|\boldsymbol{A}\| \leq \sqrt{k_{\mathrm{c}}k_{\mathrm{r}}}$.*

Proof. This immediately follows from the elementary inequality $\|\boldsymbol{A}\|^2 \leq \|\boldsymbol{A}\|_{1\to1}\|\boldsymbol{A}\|_{\infty\to\infty}$ (see [Hig92, equation (1.11)]), where $\|\boldsymbol{A}\|_{1\to1}$ and $\|\boldsymbol{A}\|_{\infty\to\infty}$ are the induced 1-norm (or maximum absolute column sum norm) and the induced ∞-norm (or maximum absolute row sum norm), respectively. □

D.2.3 Matrix perturbation bounds

Lemma 45. *Let $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ be a symmetric matrix with the top-r eigendecomposition $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^\top$. Assume $\|\boldsymbol{M} - \boldsymbol{M}^\star\| \leq \sigma_{\min}/2$ and denote*

$$
\widehat{\boldsymbol{Q}} := \underset{\boldsymbol{R}\in\mathcal{O}^{r\times r}}{\operatorname{argmin}}\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|_{\mathrm{F}}.
$$

Then there is some numerical constant $c_3 > 0$ such that

$$
\left\|\boldsymbol{U}\widehat{\boldsymbol{Q}} - \boldsymbol{U}^\star\right\| \leq \frac{c_3}{\sigma_{\min}}\|\boldsymbol{M} - \boldsymbol{M}^\star\|.
$$

Proof. Define $\boldsymbol{Q} = \boldsymbol{U}^\top \boldsymbol{U}^\star$. The triangle inequality gives

$$\left\|\boldsymbol{U}\widehat{\boldsymbol{Q}} - \boldsymbol{U}^\star\right\| \le \left\|\boldsymbol{U}\big(\widehat{\boldsymbol{Q}} - \boldsymbol{Q}\big)\right\| + \left\|\boldsymbol{U}\boldsymbol{Q} - \boldsymbol{U}^\star\right\| \le \left\|\widehat{\boldsymbol{Q}} - \boldsymbol{Q}\right\| + \left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\star - \boldsymbol{U}^\star\right\|. \tag{276}$$

[AFWZ17, Lemma 3] asserts that

$$\left\|\widehat{\boldsymbol{Q}} - \boldsymbol{Q}\right\| \le 4\left(\left\|\boldsymbol{M} - \boldsymbol{M}^\star\right\| / \sigma_{\min}\right)^2$$

as long as $\|\boldsymbol{M} - \boldsymbol{M}^\star\| \le \sigma_{\min}/2$. For the remaining term in (276), one can use $\boldsymbol{U}^{\star\top}\boldsymbol{U}^\star = \boldsymbol{I}_r$ to obtain

$$\left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\star - \boldsymbol{U}^\star\right\| = \left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\star - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\boldsymbol{U}^\star\right\| \le \left\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right\|,$$

which together with the Davis-Kahan $\sin\Theta$ theorem [DK70] reveals that

$$\left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\star - \boldsymbol{U}^\star\right\| \le \frac{c_2}{\sigma_{\min}}\left\|\boldsymbol{M} - \boldsymbol{M}^\star\right\|$$

for some constant $c_2 > 0$. Combine the estimates on $\left\|\widehat{\boldsymbol{Q}} - \boldsymbol{Q}\right\|$, $\left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\star - \boldsymbol{U}^\star\right\|$ and (276) to reach

$$\left\|\boldsymbol{U}\widehat{\boldsymbol{Q}} - \boldsymbol{U}^\star\right\| \le \left(\frac{4}{\sigma_{\min}}\left\|\boldsymbol{M} - \boldsymbol{M}^\star\right\|\right)^2 + \frac{c_2}{\sigma_{\min}}\left\|\boldsymbol{M} - \boldsymbol{M}^\star\right\| \le \frac{c_3}{\sigma_{\min}}\left\|\boldsymbol{M} - \boldsymbol{M}^\star\right\|$$

for some numerical constant $c_3 > 0$, where we have utilized the fact that $\left\|\boldsymbol{M} - \boldsymbol{M}^\star\right\| / \sigma_{\min} \le 1/2$. □

Lemma 46. *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 $\|\boldsymbol{M} - \boldsymbol{M}^\star\| \le \sigma_{\min}/4$ and $\|\widetilde{\boldsymbol{M}} - \boldsymbol{M}^\star\| \le \sigma_{\min}/4$, 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 $\boldsymbol{M}^\star$, respectively. If we denote*

$$\boldsymbol{Q} := \underset{\boldsymbol{R}\in\mathcal{O}^{r\times r}}{\operatorname{argmin}} \left\|\boldsymbol{U}\boldsymbol{R} - \widetilde{\boldsymbol{U}}\right\|_{\mathrm{F}},$$

then there exists some numerical constant $c_3 > 0$ such that

$$\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{Q} - \boldsymbol{Q}\widetilde{\boldsymbol{\Sigma}}^{1/2}\right\| \le \frac{c_3}{\sqrt{\sigma_{\min}}}\left\|\widetilde{\boldsymbol{M}} - \boldsymbol{M}\right\| \qquad \text{and} \qquad \left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{Q} - \boldsymbol{Q}\widetilde{\boldsymbol{\Sigma}}^{1/2}\right\|_{\mathrm{F}} \le \frac{c_3}{\sqrt{\sigma_{\min}}}\left\|\big(\widetilde{\boldsymbol{M}} - \boldsymbol{M}\big)\boldsymbol{U}\right\|_{\mathrm{F}}.$$

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 [Sch92, Lemma 2.1],

$$\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{Q} - \boldsymbol{Q}\widetilde{\boldsymbol{\Sigma}}^{1/2}\right\|_{\mathrm{F}} \le \frac{1}{\sigma_{\min}\big[(\boldsymbol{\Sigma})^{1/2}\big] + \sigma_{\min}\big[(\widetilde{\boldsymbol{\Sigma}})^{1/2}\big]}\left\|\boldsymbol{Q}^\top\boldsymbol{\Sigma}\boldsymbol{Q} - \widetilde{\boldsymbol{\Sigma}}\right\|_{\mathrm{F}} \le \frac{1}{\sqrt{\sigma_{\min}}}\left\|\boldsymbol{Q}^\top\boldsymbol{\Sigma}\boldsymbol{Q} - \widetilde{\boldsymbol{\Sigma}}\right\|_{\mathrm{F}}, \tag{277}$$

where the last inequality follows from the lower estimates

$$\sigma_{\min}(\boldsymbol{\Sigma}) \ge \sigma_{\min}(\boldsymbol{\Sigma}^\star) - \|\boldsymbol{M} - \boldsymbol{M}^\star\| \ge \sigma_{\min}/4$$

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

$$\begin{aligned}
\left\|\boldsymbol{Q}^\top\boldsymbol{\Sigma}\boldsymbol{Q} - \widetilde{\boldsymbol{\Sigma}}\right\|_{\mathrm{F}} &= \left\|\big(\boldsymbol{U}\boldsymbol{Q}\big)^\top\boldsymbol{M}\big(\boldsymbol{U}\boldsymbol{Q}\big) - \widetilde{\boldsymbol{U}}^\top\widetilde{\boldsymbol{M}}\widetilde{\boldsymbol{U}}\right\|_{\mathrm{F}} \\
&\le \left\|\big(\boldsymbol{U}\boldsymbol{Q}\big)^\top\boldsymbol{M}\big(\boldsymbol{U}\boldsymbol{Q}\big) - \big(\boldsymbol{U}\boldsymbol{Q}\big)^\top\widetilde{\boldsymbol{M}}\big(\boldsymbol{U}\boldsymbol{Q}\big)\right\|_{\mathrm{F}} + \left\|\big(\boldsymbol{U}\boldsymbol{Q}\big)^\top\widetilde{\boldsymbol{M}}\big(\boldsymbol{U}\boldsymbol{Q}\big) - \widetilde{\boldsymbol{U}}^\top\widetilde{\boldsymbol{M}}\big(\boldsymbol{U}\boldsymbol{Q}\big)\right\|_{\mathrm{F}} \\
&\quad + \left\|\widetilde{\boldsymbol{U}}^\top\widetilde{\boldsymbol{M}}\big(\boldsymbol{U}\boldsymbol{Q}\big) - \widetilde{\boldsymbol{U}}^\top\widetilde{\boldsymbol{M}}\widetilde{\boldsymbol{U}}\right\|_{\mathrm{F}}
\end{aligned}$$

$$
\leq \left\| (\widetilde{\boldsymbol{M}} - \boldsymbol{M})\boldsymbol{U} \right\|_{\mathrm{F}} + 2\|\boldsymbol{U}\boldsymbol{Q} - \widetilde{\boldsymbol{U}}\|_{\mathrm{F}} \|\widetilde{\boldsymbol{M}}\| \leq \left\| (\widetilde{\boldsymbol{M}} - \boldsymbol{M})\boldsymbol{U} \right\|_{\mathrm{F}} + 4\sigma_{\max}\|\boldsymbol{U}\boldsymbol{Q} - \widetilde{\boldsymbol{U}}\|_{\mathrm{F}}, \tag{278}
$$

where the last relation holds due to the upper estimate

$$
\|\widetilde{\boldsymbol{M}}\| \leq \|\boldsymbol{M}^\star\| + \|\widetilde{\boldsymbol{M}} - \boldsymbol{M}^\star\| \leq \sigma_{\max} + \sigma_{\min}/4 \leq 2\sigma_{\max}.
$$

Invoke the Davis-Kahan sinΘ theorem [DK70] to obtain

$$
\|\boldsymbol{U}\boldsymbol{Q} - \widetilde{\boldsymbol{U}}\|_{\mathrm{F}} \leq \frac{c_2}{\sigma_r(\boldsymbol{M}) - \sigma_{r+1}(\widetilde{\boldsymbol{M}})} \left\| (\widetilde{\boldsymbol{M}} - \boldsymbol{M})\boldsymbol{U} \right\|_{\mathrm{F}} \leq \frac{2c_2}{\sigma_{\min}} \left\| (\widetilde{\boldsymbol{M}} - \boldsymbol{M})\boldsymbol{U} \right\|_{\mathrm{F}}, \tag{279}
$$

for some constant $c_2 > 0$, where the last inequality follows from the bounds

$$
\begin{aligned}
\sigma_r(\boldsymbol{M}) &\geq \sigma_r(\boldsymbol{M}^\star) - \|\boldsymbol{M} - \boldsymbol{M}^\star\| \geq 3\sigma_{\min}/4, \\
\sigma_{r+1}(\widetilde{\boldsymbol{M}}) &\leq \sigma_{r+1}(\boldsymbol{M}^\star) + \|\widetilde{\boldsymbol{M}} - \boldsymbol{M}^\star\| \leq \sigma_{\min}/4.
\end{aligned}
$$

Combine (277), (278), (279) 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 47. *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 $\boldsymbol{X}^\star = \boldsymbol{U}^\star(\boldsymbol{\Sigma}^\star)^{1/2}$, and define*

$$
\widehat{\boldsymbol{Q}} := \operatorname*{argmin}_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|_{\mathrm{F}} \qquad \text{and} \qquad \widehat{\boldsymbol{H}} := \operatorname*{argmin}_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \|\boldsymbol{X}\boldsymbol{R} - \boldsymbol{X}^\star\|_{\mathrm{F}}.
$$

Assume $\|\boldsymbol{M} - \boldsymbol{M}^\star\| \leq \sigma_{\min}/2$, 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

$$
\|\widehat{\boldsymbol{Q}} - \widehat{\boldsymbol{H}}\| \leq \frac{c_3}{\sigma_{\min}} \|\boldsymbol{M} - \boldsymbol{M}^\star\|.
$$

Proof. We first collect several useful facts about the spectrum of $\boldsymbol{\Sigma}$. Weyl's inequality tells us that $\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^\star\| \leq \|\boldsymbol{M} - \boldsymbol{M}^\star\| \leq \sigma_{\min}/2$, which further implies that

$$
\sigma_r(\boldsymbol{\Sigma}) \geq \sigma_r(\boldsymbol{\Sigma}^\star) - \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^\star\| \geq \sigma_{\min}/2 \qquad \text{and} \qquad \|\boldsymbol{\Sigma}\| \leq \|\boldsymbol{\Sigma}^\star\| + \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^\star\| \leq 2\sigma_{\max}.
$$

Denote

$$
\boldsymbol{Q} = \boldsymbol{U}^\top\boldsymbol{U}^\star \qquad \text{and} \qquad \boldsymbol{H} = \boldsymbol{X}^\top\boldsymbol{X}^\star.
$$

Simple algebra yields

$$
\boldsymbol{H} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{Q}(\boldsymbol{\Sigma}^\star)^{1/2} = \underbrace{\boldsymbol{\Sigma}^{1/2}(\boldsymbol{Q} - \widehat{\boldsymbol{Q}})(\boldsymbol{\Sigma}^\star)^{1/2} + (\boldsymbol{\Sigma}^{1/2}\widehat{\boldsymbol{Q}} - \widehat{\boldsymbol{Q}}\boldsymbol{\Sigma}^{1/2})(\boldsymbol{\Sigma}^\star)^{1/2}}_{:=\boldsymbol{E}} + \underbrace{\widehat{\boldsymbol{Q}}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\star)^{1/2}}_{:=\boldsymbol{A}}.
$$

It can be easily seen that $\sigma_{r-1}(\boldsymbol{A}) \geq \sigma_r(\boldsymbol{A}) \geq \sigma_{\min}/2$, and

$$
\begin{aligned}
\|\boldsymbol{E}\| &\leq \|\boldsymbol{\Sigma}^{1/2}\| \cdot \|\boldsymbol{Q} - \widehat{\boldsymbol{Q}}\| \cdot \|(\boldsymbol{\Sigma}^\star)^{1/2}\| + \left\|\boldsymbol{\Sigma}^{1/2}\widehat{\boldsymbol{Q}} - \widehat{\boldsymbol{Q}}\boldsymbol{\Sigma}^{1/2}\right\| \cdot \|(\boldsymbol{\Sigma}^\star)^{1/2}\| \\
&\leq 2\sigma_{\max} \underbrace{\|\boldsymbol{Q} - \widehat{\boldsymbol{Q}}\|}_{:=\alpha} + \sqrt{\sigma_{\max}} \underbrace{\left\|\boldsymbol{\Sigma}^{1/2}\widehat{\boldsymbol{Q}} - \widehat{\boldsymbol{Q}}\boldsymbol{\Sigma}^{1/2}\right\|}_{:=\beta},
\end{aligned}
$$

which can be controlled as follows.

- Regarding α, use [AFWZ17, Lemma 3] to reach

$$
\alpha = \|\boldsymbol{Q} - \widehat{\boldsymbol{Q}}\| \leq 4\|\boldsymbol{M} - \boldsymbol{M}^\star\|^2/\sigma_{\min}^2.
$$

- 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 [Sch92, Lemma 2.1]. We can further take the triangle inequality to obtain

$$\begin{aligned}
\left\| \boldsymbol{\Sigma} \widehat{\boldsymbol{Q}} - \widehat{\boldsymbol{Q}} \boldsymbol{\Sigma} \right\| &= \left\| \boldsymbol{\Sigma} \boldsymbol{Q} - \boldsymbol{Q} \boldsymbol{\Sigma} + \boldsymbol{\Sigma}(\widehat{\boldsymbol{Q}} - \boldsymbol{Q}) - (\widehat{\boldsymbol{Q}} - \boldsymbol{Q}) \boldsymbol{\Sigma} \right\| \\
&\leq \left\| \boldsymbol{\Sigma} \boldsymbol{Q} - \boldsymbol{Q} \boldsymbol{\Sigma} \right\| + 2 \left\| \boldsymbol{\Sigma} \right\| \left\| \boldsymbol{Q} - \widehat{\boldsymbol{Q}} \right\| \\
&= \left\| \boldsymbol{U}\left(\boldsymbol{M} - \boldsymbol{M}^{\star}\right) \boldsymbol{U}^{\star\top} + \boldsymbol{Q}\left(\boldsymbol{\Sigma}^{\star} - \boldsymbol{\Sigma}\right) \right\| + 2 \left\| \boldsymbol{\Sigma} \right\| \left\| \boldsymbol{Q} - \widehat{\boldsymbol{Q}} \right\| \\
&\leq \left\| \boldsymbol{U}\left(\boldsymbol{M} - \boldsymbol{M}^{\star}\right) \boldsymbol{U}^{\star\top} \right\| + \left\| \boldsymbol{Q}\left(\boldsymbol{\Sigma}^{\star} - \boldsymbol{\Sigma}\right) \right\| + 2 \left\| \boldsymbol{\Sigma} \right\| \left\| \boldsymbol{Q} - \widehat{\boldsymbol{Q}} \right\| \\
&\leq 2 \left\| \boldsymbol{M} - \boldsymbol{M}^{\star} \right\| + 4\sigma_{\max}\alpha,
\end{aligned}$$

where the last inequality uses the Weyl's inequality $\|\boldsymbol{\Sigma}^{\star} - \boldsymbol{\Sigma}\| \leq \|\boldsymbol{M} - \boldsymbol{M}^{\star}\|$ and the fact that $\|\boldsymbol{\Sigma}\| \leq 2\sigma_{\max}$.

- Rearrange the previous bounds to arrive at

$$\|\boldsymbol{E}\| \leq 2\sigma_{\max}\alpha + \sqrt{\sigma_{\max}} \frac{1}{\sqrt{\sigma_{\min}}} \left(2 \left\| \boldsymbol{M} - \boldsymbol{M}^{\star} \right\| + 4\sigma_{\max}\alpha\right) \leq c_2 \left\| \boldsymbol{M} - \boldsymbol{M}^{\star} \right\|$$

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 (177)), we are ready to invoke Lemma 36 to deduce that

$$\left\| \widehat{\boldsymbol{Q}} - \widehat{\boldsymbol{H}} \right\| \leq \frac{2}{\sigma_{r-1}(\boldsymbol{A}) + \sigma_r(\boldsymbol{A})} \|\boldsymbol{E}\| \leq \frac{c_3}{\sigma_{\min}} \left\| \boldsymbol{M} - \boldsymbol{M}^{\star} \right\|$$

for some constant $c_3 > 0$. $\square$

D.3 Technical lemmas for blind deconvolution

D.3.1 Wirtinger calculus

In this section, we formally prove the fundamental theorem of calculus and the mean-value form of Taylor's theorem under the Wirtinger calculus; see (283) and (284), respectively.

Let $f : \mathbb{C}^n \to \mathbb{R}$ be a real-valued function. Denote $\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y} \in \mathbb{C}^n$, then $f(\cdot)$ can alternatively be viewed as a function $\mathbb{R}^{2n} \to \mathbb{R}$. There is a one-to-one mapping connecting the Wirtinger derivatives and the conventional derivatives [KD09]:

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} = \boldsymbol{J}^{-1} \begin{bmatrix} \boldsymbol{z} \\ \overline{\boldsymbol{z}} \end{bmatrix}, \tag{280a}$$

$$\nabla_{\mathbb{R}} f\left(\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}\right) = \boldsymbol{J}^{\mathsf{H}} \nabla_{\mathbb{C}} f\left(\begin{bmatrix} \boldsymbol{z} \\ \overline{\boldsymbol{z}} \end{bmatrix}\right), \tag{280b}$$

$$\nabla_{\mathbb{R}}^2 f\left(\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}\right) = \boldsymbol{J}^{\mathsf{H}} \nabla_{\mathbb{C}}^2 f\left(\begin{bmatrix} \boldsymbol{z} \\ \overline{\boldsymbol{z}} \end{bmatrix}\right) \boldsymbol{J}, \tag{280c}$$

where the subscripts $\mathbb{R}$ and $\mathbb{C}$ represent calculus in the real (conventional) sense and in the complex (Wirtinger) sense, respectively, and

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{I}_n & i\boldsymbol{I}_n \\ \boldsymbol{I}_n & -i\boldsymbol{I}_n \end{bmatrix}.$$

With these relationships in place, we are ready to verify the fundamental theorem of calculus using the Wirtinger derivatives. Recall from [Lan93, Chapter XIII, Theorem 4.2] that

$$\nabla_{\mathbb{R}} f\left(\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{y}_1 \end{bmatrix}\right) - \nabla_{\mathbb{R}} f\left(\begin{bmatrix} \boldsymbol{x}_2 \\ \boldsymbol{y}_2 \end{bmatrix}\right) = \left[\int_0^1 \nabla_{\mathbb{R}}^2 f\left(\begin{bmatrix} \boldsymbol{x}(\tau) \\ \boldsymbol{y}(\tau) \end{bmatrix}\right) \mathrm{d}\tau\right] \left(\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{y}_1 \end{bmatrix} - \begin{bmatrix} \boldsymbol{x}_2 \\ \boldsymbol{y}_2 \end{bmatrix}\right), \tag{281}$$

where

$$\begin{bmatrix} \boldsymbol{x}(\tau) \\ \boldsymbol{y}(\tau) \end{bmatrix} := \begin{bmatrix} \boldsymbol{x}_2 \\ \boldsymbol{y}_2 \end{bmatrix} + \tau \left(\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{y}_1 \end{bmatrix} - \begin{bmatrix} \boldsymbol{x}_2 \\ \boldsymbol{y}_2 \end{bmatrix} \right).$$

Substitute the identities (280) into (281) to arrive at

$$\begin{aligned} \boldsymbol{J}^{\mathsf{H}} \nabla_{\mathbb{C}} f\left(\begin{bmatrix} \boldsymbol{z}_1 \\ \overline{\boldsymbol{z}_1} \end{bmatrix} \right) - \boldsymbol{J}^{\mathsf{H}} \nabla_{\mathbb{C}} f\left(\begin{bmatrix} \boldsymbol{z}_2 \\ \overline{\boldsymbol{z}_2} \end{bmatrix} \right) &= \boldsymbol{J}^{\mathsf{H}} \left[\int_0^1 \nabla^2_{\mathbb{C}} f\left(\begin{bmatrix} \boldsymbol{z}(\tau) \\ \overline{\boldsymbol{z}(\tau)} \end{bmatrix} \right) \mathrm{d}\tau \right] \boldsymbol{J}\boldsymbol{J}^{-1} \left(\begin{bmatrix} \boldsymbol{z}_1 \\ \overline{\boldsymbol{z}_1} \end{bmatrix} - \begin{bmatrix} \boldsymbol{z}_2 \\ \overline{\boldsymbol{z}_2} \end{bmatrix} \right) \\ &= \boldsymbol{J}^{\mathsf{H}} \left[\int_0^1 \nabla^2_{\mathbb{C}} f\left(\begin{bmatrix} \boldsymbol{z}(\tau) \\ \overline{\boldsymbol{z}(\tau)} \end{bmatrix} \right) \mathrm{d}\tau \right] \left(\begin{bmatrix} \boldsymbol{z}_1 \\ \overline{\boldsymbol{z}_1} \end{bmatrix} - \begin{bmatrix} \boldsymbol{z}_2 \\ \overline{\boldsymbol{z}_2} \end{bmatrix} \right), \end{aligned} \tag{282}$$

where $\boldsymbol{z}_1 = \boldsymbol{x}_1 + i\boldsymbol{y}_1$, $\boldsymbol{z}_2 = \boldsymbol{x}_2 + i\boldsymbol{y}_2$ and

$$\begin{bmatrix} \boldsymbol{z}(\tau) \\ \overline{\boldsymbol{z}(\tau)} \end{bmatrix} := \begin{bmatrix} \boldsymbol{z}_2 \\ \overline{\boldsymbol{z}_2} \end{bmatrix} + \tau \left(\begin{bmatrix} \boldsymbol{z}_1 \\ \overline{\boldsymbol{z}_1} \end{bmatrix} - \begin{bmatrix} \boldsymbol{z}_2 \\ \overline{\boldsymbol{z}_2} \end{bmatrix} \right).$$

Simplification of (282) gives

$$\nabla_{\mathbb{C}} f\left(\begin{bmatrix} \boldsymbol{z}_1 \\ \overline{\boldsymbol{z}_1} \end{bmatrix} \right) - \nabla_{\mathbb{C}} f\left(\begin{bmatrix} \boldsymbol{z}_2 \\ \overline{\boldsymbol{z}_2} \end{bmatrix} \right) = \left[\int_0^1 \nabla^2_{\mathbb{C}} f\left(\begin{bmatrix} \boldsymbol{z}(\tau) \\ \overline{\boldsymbol{z}(\tau)} \end{bmatrix} \right) \mathrm{d}\tau \right] \left(\begin{bmatrix} \boldsymbol{z}_1 \\ \overline{\boldsymbol{z}_1} \end{bmatrix} - \begin{bmatrix} \boldsymbol{z}_2 \\ \overline{\boldsymbol{z}_2} \end{bmatrix} \right). \tag{283}$$

Repeating the above arguments, one can also show that

$$f(\boldsymbol{z}_1) - f(\boldsymbol{z}_2) = \nabla_{\mathbb{C}} f(\boldsymbol{z}_2)^{\mathsf{H}} \begin{bmatrix} \boldsymbol{z}_1 - \boldsymbol{z}_2 \\ \overline{\boldsymbol{z}_1 - \boldsymbol{z}_2} \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \boldsymbol{z}_1 - \boldsymbol{z}_2 \\ \overline{\boldsymbol{z}_1 - \boldsymbol{z}_2} \end{bmatrix}^{\mathsf{H}} \nabla^2_{\mathbb{C}} f(\widetilde{\boldsymbol{z}}) \begin{bmatrix} \boldsymbol{z}_1 - \boldsymbol{z}_2 \\ \overline{\boldsymbol{z}_1 - \boldsymbol{z}_2} \end{bmatrix}, \tag{284}$$

where $\widetilde{\boldsymbol{z}}$ is some point lying on the vector connecting $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$. This is the mean-value form of Taylor's theorem under the Wirtinger calculus.

D.3.2 Discrete Fourier transform matrices

Let $\boldsymbol{B} \in \mathbb{C}^{m \times K}$ be the first K columns of a discrete Fourier transform (DFT) matrix $\boldsymbol{F} \in \mathbb{C}^{m \times m}$, and denote by $\boldsymbol{b}_l$ the lth column of the matrix $\boldsymbol{B}^{\mathsf{H}}$. By definition,

$$\boldsymbol{b}_l = \frac{1}{\sqrt{m}} \left(1, \omega^{(l-1)}, \omega^{2(l-1)}, \cdots, \omega^{(K-1)(l-1)} \right)^{\mathsf{H}},$$

where $\omega := e^{-i\frac{2\pi}{m}}$ with i representing the imaginary unit. It is seen that for any $j \neq l$,

$$\boldsymbol{b}_l^{\mathsf{H}} \boldsymbol{b}_j = \frac{1}{m} \sum_{k=0}^{K-1} \omega^{k(l-1)} \cdot \overline{\omega^{k(j-1)}} \overset{\text{(i)}}{=} \frac{1}{m} \sum_{k=0}^{K-1} \omega^{k(l-1)} \cdot \omega^{k(1-j)} = \frac{1}{m} \sum_{k=0}^{K-1} \left(\omega^{l-j} \right)^k \overset{\text{(ii)}}{=} \frac{1}{m} \frac{1 - \omega^{K(l-j)}}{1 - \omega^{l-j}}. \tag{285}$$

Here, (i) uses $\overline{\omega^{\alpha}} = \omega^{-\alpha}$ for all $\alpha \in \mathbb{R}$, while the last identity (ii) follows from the formula for the sum of a finite geometric series when $\omega^{l-j} \neq 1$. This leads to the following lemma.

Lemma 48. *For any $m \geq 3$ and any $1 \leq l \leq m$, we have*

$$\sum_{j=1}^{m} \left| \boldsymbol{b}_l^{\mathsf{H}} \boldsymbol{b}_j \right| \leq 4 \log m.$$

Proof. We first make use of the identity (285) to obtain

$$\sum_{j=1}^{m} \left| \boldsymbol{b}_l^{\mathsf{H}} \boldsymbol{b}_j \right| = \|\boldsymbol{b}_l\|_2^2 + \frac{1}{m} \sum_{j:j \neq l}^{m} \left| \frac{1 - \omega^{K(l-j)}}{1 - \omega^{l-j}} \right| = \frac{K}{m} + \frac{1}{m} \sum_{j:j \neq l}^{m} \left| \frac{\sin\left[K(l-j)\frac{\pi}{m} \right]}{\sin\left[(l-j)\frac{\pi}{m} \right]} \right|,$$

where the last identity follows since $\|\boldsymbol{b}_l\|_2^2 = K/m$ and, for all $\alpha \in \mathbb{R}$,

$$|1 - \omega^{\alpha}| = \left| 1 - e^{-i\frac{2\pi}{m}\alpha} \right| = \left| e^{-i\frac{\pi}{m}\alpha} \left(e^{i\frac{\pi}{m}\alpha} - e^{-i\frac{\pi}{m}\alpha} \right) \right| = 2 \left| \sin\left(\alpha \frac{\pi}{m} \right) \right|. \tag{286}$$

Without loss of generality, we focus on the case when $l=1$ in the sequel. Recall that for $c>0$, we denote by $\lfloor c \rfloor$ the largest integer that does not exceed c. We can continue the derivation to get

$$\begin{aligned}
\sum_{j=1}^{m}\left|\boldsymbol{b}_1^{\mathsf{H}}\boldsymbol{b}_j\right| &= \frac{K}{m}+\frac{1}{m}\sum_{j=2}^{m}\left|\frac{\sin\left[K\left(1-j\right)\frac{\pi}{m}\right]}{\sin\left[\left(1-j\right)\frac{\pi}{m}\right]}\right| \overset{\text{(i)}}{\leq} \frac{1}{m}\sum_{j=2}^{m}\left|\frac{1}{\sin\left[\left(j-1\right)\frac{\pi}{m}\right]}\right|+\frac{K}{m}\\
&= \frac{1}{m}\left(\sum_{j=2}^{\lfloor\frac{m}{2}\rfloor+1}\left|\frac{1}{\sin\left[\left(j-1\right)\frac{\pi}{m}\right]}\right|+\sum_{j=\lfloor\frac{m}{2}\rfloor+2}^{m}\left|\frac{1}{\sin\left[\left(j-1\right)\frac{\pi}{m}\right]}\right|\right)+\frac{K}{m}\\
&\overset{\text{(ii)}}{=} \frac{1}{m}\left(\sum_{j=2}^{\lfloor\frac{m}{2}\rfloor+1}\left|\frac{1}{\sin\left[\left(j-1\right)\frac{\pi}{m}\right]}\right|+\sum_{j=\lfloor\frac{m}{2}\rfloor+2}^{m}\left|\frac{1}{\sin\left[\left(m+1-j\right)\frac{\pi}{m}\right]}\right|\right)+\frac{K}{m},
\end{aligned}$$

where (i) follows from $\left|\sin\left(K\left(1-j\right)\frac{\pi}{m}\right)\right|\leq 1$ and $\left|\sin\left(x\right)\right|=\left|\sin\left(-x\right)\right|$, and (ii) relies on the fact that $\sin\left(x\right)=\sin\left(\pi-x\right)$. The property that $\sin\left(x\right)\geq x/2$ for any $x\in\left[0,\pi/2\right]$ allows one to further derive

$$\begin{aligned}
\sum_{j=1}^{m}\left|\boldsymbol{b}_1^{\mathsf{H}}\boldsymbol{b}_j\right| &\leq \frac{1}{m}\left(\sum_{j=2}^{\lfloor\frac{m}{2}\rfloor+1}\frac{2m}{\left(j-1\right)\pi}+\sum_{j=\lfloor\frac{m}{2}\rfloor+2}^{m}\frac{2m}{\left(m+1-j\right)\pi}\right)+\frac{K}{m} = \frac{2}{\pi}\left(\sum_{k=1}^{\lfloor\frac{m}{2}\rfloor}\frac{1}{k}+\sum_{k=1}^{\lfloor\frac{m+1}{2}\rfloor-1}\frac{1}{k}\right)+\frac{K}{m}\\
&\overset{\text{(i)}}{\leq} \frac{4}{\pi}\sum_{k=1}^{m}\frac{1}{k}+\frac{K}{m} \overset{\text{(ii)}}{\leq} \frac{4}{\pi}\left(1+\log m\right)+1 \overset{\text{(iii)}}{\leq} 4\log m,
\end{aligned}$$

where in (i) we extend the range of the summation, (ii) uses the elementary inequality $\sum_{k=1}^{m}k^{-1}\leq 1+\log m$ and (iii) holds true as long as $m\geq 3$. $\square$

The next lemma considers the difference of two inner products, namely, $\left(\boldsymbol{b}_l-\boldsymbol{b}_1\right)^{\mathsf{H}}\boldsymbol{b}_j$.

Lemma 49. *For all $0\leq l-1\leq\tau\leq\lfloor\frac{m}{10}\rfloor$, we have*

$$\left|\left(\boldsymbol{b}_l-\boldsymbol{b}_1\right)^{\mathsf{H}}\boldsymbol{b}_j\right|\leq\begin{cases}\frac{4\tau}{(j-l)}\frac{K}{m}+\frac{8\tau/\pi}{(j-l)^2} & \textit{for}\quad l+\tau\leq j\leq\lfloor\frac{m}{2}\rfloor+1,\\ \frac{4\tau}{m-(j-l)}\frac{K}{m}+\frac{8\tau/\pi}{[m-(j-1)]^2} & \textit{for}\quad \lfloor\frac{m}{2}\rfloor+l\leq j\leq m-\tau.\end{cases}$$

In addition, for any j and l, the following uniform upper bound holds

$$\left|\left(\boldsymbol{b}_l-\boldsymbol{b}_1\right)^{\mathsf{H}}\boldsymbol{b}_j\right|\leq 2\frac{K}{m}.$$

Proof. Given (285), we can obtain for $j\neq l$ and $j\neq 1$,

$$\begin{aligned}
\left|\left(\boldsymbol{b}_l-\boldsymbol{b}_1\right)^{\mathsf{H}}\boldsymbol{b}_j\right| &= \frac{1}{m}\left|\frac{1-\omega^{K(l-j)}}{1-\omega^{l-j}}-\frac{1-\omega^{K(1-j)}}{1-\omega^{1-j}}\right|\\
&= \frac{1}{m}\left|\frac{1-\omega^{K(l-j)}}{1-\omega^{l-j}}-\frac{1-\omega^{K(1-j)}}{1-\omega^{l-j}}+\frac{1-\omega^{K(1-j)}}{1-\omega^{l-j}}-\frac{1-\omega^{K(1-j)}}{1-\omega^{1-j}}\right|\\
&= \frac{1}{m}\left|\frac{\omega^{K(1-j)}-\omega^{K(l-j)}}{1-\omega^{l-j}}+\left(\omega^{l-j}-\omega^{1-j}\right)\frac{1-\omega^{K(1-j)}}{\left(1-\omega^{l-j}\right)\left(1-\omega^{1-j}\right)}\right|\\
&\leq \frac{1}{m}\left|\frac{1-\omega^{K(l-1)}}{1-\omega^{l-j}}\right|+\frac{2}{m}\left|\left(1-\omega^{1-l}\right)\frac{1}{\left(1-\omega^{l-j}\right)\left(1-\omega^{1-j}\right)}\right|,
\end{aligned}$$

where the last line is due to the triangle inequality and $\left|\omega^{\alpha}\right|=1$ for all $\alpha\in\mathbb{R}$. The identity (286) allows us to rewrite this bound as

$$\left|\left(\boldsymbol{b}_l-\boldsymbol{b}_1\right)^{\mathsf{H}}\boldsymbol{b}_j\right|\leq\frac{1}{m}\left|\frac{1}{\sin\left[\left(l-j\right)\frac{\pi}{m}\right]}\right|\left\{\left|\sin\left[K\left(l-1\right)\frac{\pi}{m}\right]\right|+\left|\frac{\sin\left[\left(1-l\right)\frac{\pi}{m}\right]}{\sin\left[\left(1-j\right)\frac{\pi}{m}\right]}\right|\right\}. \tag{287}$$

Combined with the fact that $|\sin x| \leq 2|x|$ for all $x \in \mathbb{R}$, we can upper bound (287) as

$$\left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^{\mathsf{H}} \boldsymbol{b}_j\right| \leq \frac{1}{m}\left|\frac{1}{\sin\left[(l-j)\frac{\pi}{m}\right]}\right|\left\{2K\tau\frac{\pi}{m} + \left|\frac{2\tau\frac{\pi}{m}}{\sin\left[(1-j)\frac{\pi}{m}\right]}\right|\right\},$$

where we also utilize the assumption $0 \leq l-1 \leq \tau$. Then for $l + \tau \leq j \leq \lfloor m/2 \rfloor + 1$, one has

$$\left|(l-j)\frac{\pi}{m}\right| \leq \frac{\pi}{2} \qquad \text{and} \qquad \left|(1-j)\frac{\pi}{m}\right| \leq \frac{\pi}{2}.$$

Therefore, utilizing the property $\sin(x) \geq x/2$ for any $x \in [0, \pi/2]$, we arrive at

$$\left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^{\mathsf{H}} \boldsymbol{b}_j\right| \leq \frac{2}{(j-l)\pi}\left(2K\tau\frac{\pi}{m} + \frac{4\tau}{j-1}\right) \leq \frac{4\tau}{(j-l)}\frac{K}{m} + \frac{8\tau/\pi}{(j-l)^2},$$

where the last inequality holds since $j-1 > j-l$. Similarly we can obtain the upper bound for $\lfloor m/2 \rfloor + l \leq j \leq m - \tau$ using nearly identical argument (which is omitted for brevity).

The uniform upper bound can be justified as follows

$$\left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^{\mathsf{H}} \boldsymbol{b}_j\right| \leq (\|\boldsymbol{b}_l\|_2 + \|\boldsymbol{b}_1\|_2)\|\boldsymbol{b}_j\|_2 \leq 2K/m.$$

The last relation holds since $\|\boldsymbol{b}_l\|_2^2 = K/m$ for all $1 \leq l \leq m$. □

Next, we list two consequences of the above estimates in Lemma 50 and Lemma 51.

Lemma 50. *Fix any constant $c > 0$ that is independent of m and K. Suppose $m \geq C\tau K \log^4 m$ for some sufficiently large constant $C > 0$, which solely depends on c. If $0 \leq l-1 \leq \tau$, then one has*

$$\sum_{j=1}^{m}\left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^{\mathsf{H}} \boldsymbol{b}_j\right| \leq \frac{c}{\log^2 m}.$$

Proof. For some constant $c_0 > 0$, we can split the index set $[m]$ into the following three disjoint sets

$$\begin{aligned}
&\mathcal{A}_1 = \left\{j : l + c_0\tau\log^2 m \leq j \leq \left\lfloor\frac{m}{2}\right\rfloor\right\},\\
&\mathcal{A}_2 = \left\{j : \left\lfloor\frac{m}{2}\right\rfloor + l \leq j \leq m - c_0\tau\log^2 m\right\},\\
\text{and} \qquad &\mathcal{A}_3 = [m] \setminus (\mathcal{A}_1 \cup \mathcal{A}_2).
\end{aligned}$$

With this decomposition in place, we can write

$$\sum_{j=1}^{m}\left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^{\mathsf{H}} \boldsymbol{b}_j\right| = \sum_{j\in\mathcal{A}_1}\left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^{\mathsf{H}} \boldsymbol{b}_j\right| + \sum_{j\in\mathcal{A}_2}\left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^{\mathsf{H}} \boldsymbol{b}_j\right| + \sum_{j\in\mathcal{A}_3}\left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^{\mathsf{H}} \boldsymbol{b}_j\right|.$$

We first look at $\mathcal{A}_1$. By Lemma 49, one has for any $j \in \mathcal{A}_1$,

$$\left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^{\mathsf{H}} \boldsymbol{b}_j\right| \leq \frac{4\tau}{j-l}\frac{K}{m} + \frac{8\tau/\pi}{(j-l)^2},$$

and hence

$$\begin{aligned}
\sum_{j\in\mathcal{A}_1}\left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^{\mathsf{H}} \boldsymbol{b}_j\right| &\leq \sum_{j=l+c_0\tau\log^2 m}^{\lfloor\frac{m}{2}\rfloor+1}\left(\frac{4\tau}{j-l}\frac{K}{m} + \frac{8\tau/\pi}{(j-l)^2}\right) \leq \frac{4\tau K}{m}\sum_{k=1}^{m}\frac{1}{k} + \frac{8\tau}{\pi}\sum_{k=c_0\tau\log^2 m}^{m}\frac{1}{k^2}\\
&\leq 8\tau\frac{K}{m}\log m + \frac{16\tau}{\pi}\frac{1}{c_0\tau\log^2 m},
\end{aligned}$$

where the last inequality arises from $\sum_{k=1}^{m} k^{-1} \leq 1 + \log m \leq 2\log m$ and $\sum_{k=c}^{m} k^{-2} \leq 2/c$.

Similarly, for $j \in \mathcal{A}_2$, we have

$$\left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^{\mathsf{H}} \boldsymbol{b}_j\right| \leq \frac{4\tau}{m-(j-l)}\frac{K}{m} + \frac{8\tau/\pi}{[m-(j-1)]^2},$$

which in turn implies

$$\sum_{j\in\mathcal{A}_2} \left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^{\mathsf{H}} \boldsymbol{b}_j\right| \leq 8\tau\frac{K}{m}\log m + \frac{16\tau}{\pi}\frac{1}{c_0\tau\log^2 m}.$$

Regarding $j \in \mathcal{A}_3$, we observe that

$$|\mathcal{A}_3| \leq 2\left(c_0\tau\log^2 m + l\right) \leq 2\left(c_0\tau\log^2 m + \tau + 1\right) \leq 4c_0\tau\log^2 m.$$

This together with the simple bound $\left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^{\mathsf{H}} \boldsymbol{b}_j\right| \leq 2K/m$ gives

$$\sum_{j\in\mathcal{A}_3} \left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^{\mathsf{H}} \boldsymbol{b}_j\right| \leq 2\frac{K}{m}\left|\mathcal{A}_3\right| \leq \frac{8c_0\tau K\log^2 m}{m}.$$

The previous three estimates taken collectively yield

$$\sum_{j=1}^{m} \left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^{\mathsf{H}} \boldsymbol{b}_j\right| \leq \frac{16\tau K\log m}{m} + \frac{32\tau}{\pi}\frac{1}{c_0\tau\log^2 m} + \frac{8c_0\tau K\log^2 m}{m} \leq c\frac{1}{\log^2 m}$$

as long as $c_0 \geq (32/\pi)\cdot(1/c)$ and $m \geq 8c_0\tau K\log^4 m/c$. $\square$

Lemma 51. *Fix any constant $c > 0$ that is independent of m and K. Consider an integer $\tau > 0$, and suppose that $m \geq C\tau K\log m$ for some large constant $C > 0$, which depends solely on c. Then we have*

$$\sum_{k=0}^{\lfloor m/\tau\rfloor} \sqrt{\sum_{j=1}^{\tau}\left|\boldsymbol{b}_1^{\mathsf{H}}\left(\boldsymbol{b}_{k\tau+j} - \boldsymbol{b}_{k\tau+1}\right)\right|^2} \leq \frac{c}{\sqrt{\tau}}.$$

Proof. The proof strategy is similar to the one used in Lemma 50. First notice that

$$\left|\boldsymbol{b}_1^{\mathsf{H}}\left(\boldsymbol{b}_{k\tau+j} - \boldsymbol{b}_{k\tau+1}\right)\right| = \left|(\boldsymbol{b}_m - \boldsymbol{b}_{m+1-j})^{\mathsf{H}} \boldsymbol{b}_{k\tau}\right|.$$

As before, for some $c_1 > 0$, we can split the index set $\{1,\cdots,\lfloor m/\tau\rfloor\}$ into three disjoint sets

$$\begin{aligned}
&\mathcal{B}_1 = \left\{k : c_1 \leq k \leq \left\lfloor\left(\left\lfloor\frac{m}{2}\right\rfloor + 1 - j\right)/\tau\right\rfloor\right\},\\
&\mathcal{B}_2 = \left\{k : \left\lfloor\left(\left\lfloor\frac{m}{2}\right\rfloor + 1 - j\right)/\tau\right\rfloor + 1 \leq k \leq \left\lfloor(m+1-j)/\tau\right\rfloor - c_1\right\},\\
\text{and}\qquad &\mathcal{B}_3 = \left\{1,\cdots,\left\lfloor\frac{m}{\tau}\right\rfloor\right\}\setminus(\mathcal{B}_1\cup\mathcal{B}_2),
\end{aligned}$$

where $1 \leq j \leq \tau$.

By Lemma 49, one has

$$\left|(\boldsymbol{b}_m - \boldsymbol{b}_{m+1-j})^{\mathsf{H}} \boldsymbol{b}_{k\tau}\right| \leq \frac{4\tau}{k\tau}\frac{K}{m} + \frac{8\tau/\pi}{(k\tau)^2}, \qquad k \in \mathcal{B}_1.$$

Hence for any $k \in \mathcal{B}_1$,

$$\sqrt{\sum_{j=1}^{\tau}\left|\boldsymbol{b}_1^{\mathsf{H}}\left(\boldsymbol{b}_{k\tau+j} - \boldsymbol{b}_{k\tau+1}\right)\right|^2} \leq \sqrt{\tau}\left(\frac{4\tau}{k\tau}\frac{K}{m} + \frac{8\tau/\pi}{(k\tau)^2}\right) = \sqrt{\tau}\left(\frac{4}{k}\frac{K}{m} + \frac{8/\pi}{k^2\tau}\right),$$

which further implies that

$$\sum_{k\in\mathcal{B}_1}\sqrt{\sum_{j=1}^{\tau}\left|\boldsymbol{b}_1^{\mathsf{H}}\left(\boldsymbol{b}_{k\tau+j}-\boldsymbol{b}_{k\tau+1}\right)\right|^2}\leq\sqrt{\tau}\sum_{k=c_1}^{m}\left(\frac{4}{k}\frac{K}{m}+\frac{8/\pi}{k^2\tau}\right)\leq 8\sqrt{\tau}\frac{K\log m}{m}+\frac{16}{\pi}\frac{1}{\sqrt{\tau}}\frac{1}{c_1},$$

where the last inequality follows since $\sum_{k=1}^{m}k^{-1}\leq 2\log m$ and $\sum_{k=c_1}^{m}k^{-2}\leq 2/c_1$. A similar bound can be obtained for $k\in\mathcal{B}_2$.

For the remaining set $\mathcal{B}_3$, observe that

$$|\mathcal{B}_3|\leq 2c_1.$$

This together with the crude upper bound $\left|\left(\boldsymbol{b}_l-\boldsymbol{b}_1\right)^{\mathsf{H}}\boldsymbol{b}_j\right|\leq 2K/m$ gives

$$\sum_{k\in\mathcal{B}_3}\sqrt{\sum_{j=1}^{\tau}\left|\boldsymbol{b}_1^{\mathsf{H}}\left(\boldsymbol{b}_{k\tau+j}-\boldsymbol{b}_{k\tau+1}\right)\right|^2}\leq|\mathcal{B}_3|\sqrt{\tau\max_j\left|\boldsymbol{b}_1^{\mathsf{H}}\left(\boldsymbol{b}_{k\tau+j}-\boldsymbol{b}_{k\tau+1}\right)\right|^2}\leq|\mathcal{B}_3|\sqrt{\tau}\cdot\frac{2K}{m}\leq\frac{4c_1\sqrt{\tau}K}{m}.$$

The previous estimates taken collectively yield

$$\sum_{k=0}^{\lfloor m/\tau\rfloor}\sqrt{\sum_{j=1}^{\tau}\left|\boldsymbol{b}_1^{\mathsf{H}}\left(\boldsymbol{b}_{k\tau+j}-\boldsymbol{b}_{k\tau+1}\right)\right|^2}\leq 2\left(8\sqrt{\tau}\frac{K\log m}{m}+\frac{16}{\pi}\frac{1}{\sqrt{\tau}}\frac{1}{c_1}\right)+\frac{4c_1\sqrt{\tau}K}{m}\leq c\frac{1}{\sqrt{\tau}},$$

as long as $c_1\gg 1/c$ and $m/(c_1\tau K\log m)\gg 1/c$. □

D.3.3 Complex-valued alignment

Let $g_{\boldsymbol{h},\boldsymbol{x}}(\cdot):\mathbb{C}\to\mathbb{R}$ be a real-valued function defined as

$$g_{\boldsymbol{h},\boldsymbol{x}}(\alpha):=\left\|\frac{1}{\overline{\alpha}}\boldsymbol{h}-\boldsymbol{h}^{\star}\right\|_2^2+\left\|\alpha\boldsymbol{x}-\boldsymbol{x}^{\star}\right\|_2^2,$$

which is the key function in the definition (34). Therefore, the alignment parameter of $(\boldsymbol{h},\boldsymbol{x})$ to $(\boldsymbol{h}^{\star},\boldsymbol{x}^{\star})$ is the minimizer of $g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)$. This section is devoted to studying various properties of $g_{\boldsymbol{h},\boldsymbol{x}}(\cdot)$. To begin with, the Wirtinger gradient and Hessian of $g_{\boldsymbol{h},\boldsymbol{x}}(\cdot)$ can be calculated as

$$\nabla g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)=\begin{bmatrix}\frac{\partial g_{\boldsymbol{h},\boldsymbol{x}}(\alpha,\overline{\alpha})}{\partial\alpha}\\ \frac{\partial g_{\boldsymbol{h},\boldsymbol{x}}(\alpha,\overline{\alpha})}{\partial\overline{\alpha}}\end{bmatrix}=\begin{bmatrix}\alpha\left\|\boldsymbol{x}\right\|_2^2-\boldsymbol{x}^{\mathsf{H}}\boldsymbol{x}^{\star}-\alpha^{-1}\left(\overline{\alpha}\right)^{-2}\left\|\boldsymbol{h}\right\|_2^2+\left(\overline{\alpha}\right)^{-2}\boldsymbol{h}^{\star\mathsf{H}}\boldsymbol{h}\\ \overline{\alpha}\left\|\boldsymbol{x}\right\|_2^2-\boldsymbol{x}^{\star\mathsf{H}}\boldsymbol{x}-\left(\overline{\alpha}\right)^{-1}\alpha^{-2}\left\|\boldsymbol{h}\right\|_2^2+\alpha^{-2}\boldsymbol{h}^{\mathsf{H}}\boldsymbol{h}^{\star}\end{bmatrix};\tag{288}$$

$$\nabla^2 g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)=\begin{bmatrix}\left\|\boldsymbol{x}\right\|_2^2+|\alpha|^{-4}\left\|\boldsymbol{h}\right\|_2^2 & 2\alpha^{-1}\left(\overline{\alpha}\right)^{-3}\left\|\boldsymbol{h}\right\|_2^2-2\left(\overline{\alpha}\right)^{-3}\boldsymbol{h}^{\star\mathsf{H}}\boldsymbol{h}\\ 2\left(\overline{\alpha}\right)^{-1}\alpha^{-3}\left\|\boldsymbol{h}\right\|_2^2-2\alpha^{-3}\boldsymbol{h}^{\mathsf{H}}\boldsymbol{h}^{\star} & \left\|\boldsymbol{x}\right\|_2^2+|\alpha|^{-4}\left\|\boldsymbol{h}\right\|_2^2\end{bmatrix}.\tag{289}$$

The first lemma reveals that, as long as $\left(\frac{1}{\overline{\beta}}\boldsymbol{h},\beta\boldsymbol{x}\right)$ is sufficiently close to $(\boldsymbol{h}^{\star},\boldsymbol{x}^{\star})$, the minimizer of $g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)$ cannot be far away from β.

Lemma 52. *Assume theres exists $\beta\in\mathbb{C}$ with $1/2\leq|\beta|\leq 3/2$ such that $\max\left\{\left\|\frac{1}{\overline{\beta}}\boldsymbol{h}-\boldsymbol{h}^{\star}\right\|_2,\left\|\beta\boldsymbol{x}-\boldsymbol{x}^{\star}\right\|_2\right\}\leq\delta\leq 1/4$. Denote by $\widehat{\alpha}$ the minimizer of $g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)$, then we necessarily have*

$$\left||\widehat{\alpha}|-|\beta|\right|\leq|\widehat{\alpha}-\beta|\leq 18\delta.$$

Proof. The first inequality is a direct consequence of the triangle inequality. Hence we concentrate on the second one. Notice that by assumption,

$$g_{\boldsymbol{h},\boldsymbol{x}}(\beta)=\left\|\frac{1}{\overline{\beta}}\boldsymbol{h}-\boldsymbol{h}^{\star}\right\|_2^2+\left\|\beta\boldsymbol{x}-\boldsymbol{x}^{\star}\right\|_2^2\leq 2\delta^2,\tag{290}$$

which immediately implies that $g_{\boldsymbol{h},\boldsymbol{x}}(\widehat{\alpha}) \leq 2\delta^2$. It thus suffices to show that for any α obeying $|\alpha - \beta| > 18\delta$, one has $g_{\boldsymbol{h},\boldsymbol{x}}(\alpha) > 2\delta^2$, and hence it cannot be the minimizer. To this end, we lower bound $g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)$ as follows:

$$\begin{aligned} g_{\boldsymbol{h},\boldsymbol{x}}(\alpha) &\geq \|\alpha\boldsymbol{x} - \boldsymbol{x}^\star\|_2^2 = \|(\alpha-\beta)\boldsymbol{x} + (\beta\boldsymbol{x} - \boldsymbol{x}^\star)\|_2^2 \\ &= |\alpha-\beta|^2\|\boldsymbol{x}\|_2^2 + \|\beta\boldsymbol{x} - \boldsymbol{x}^\star\|_2^2 + 2\mathrm{Re}\left[(\alpha-\beta)(\beta\boldsymbol{x}-\boldsymbol{x}^\star)^{\mathsf{H}}\boldsymbol{x}\right] \\ &\geq |\alpha-\beta|^2\|\boldsymbol{x}\|_2^2 - 2|\alpha-\beta|\left|(\beta\boldsymbol{x}-\boldsymbol{x}^\star)^{\mathsf{H}}\boldsymbol{x}\right|. \end{aligned}$$

Given that $\|\beta\boldsymbol{x} - \boldsymbol{x}^\star\|_2 \leq \delta \leq 1/4$ and $\|\boldsymbol{x}^\star\|_2 = 1$, we have

$$\|\beta\boldsymbol{x}\|_2 \geq \|\boldsymbol{x}^\star\|_2 - \|\beta\boldsymbol{x} - \boldsymbol{x}^\star\|_2 \geq 1 - \delta \geq 3/4,$$

which together with the fact that $1/2 \leq |\beta| \leq 3/2$ implies

$$\|\boldsymbol{x}\|_2 \geq 1/2 \qquad \text{and} \qquad \|\boldsymbol{x}\|_2 \leq 2$$

and

$$\left|(\beta\boldsymbol{x} - \boldsymbol{x}^\star)^{\mathsf{H}}\boldsymbol{x}\right| \leq \|\beta\boldsymbol{x} - \boldsymbol{x}^\star\|_2\|\boldsymbol{x}\|_2 \leq 2\delta.$$

Taking the previous estimates collectively yields

$$g_{\boldsymbol{h},\boldsymbol{x}}(\alpha) \geq \frac{1}{4}|\alpha-\beta|^2 - 4\delta|\alpha-\beta|.$$

It is self-evident that once $|\alpha-\beta| > 18\delta$, one gets $g_{\boldsymbol{h},\boldsymbol{x}}(\alpha) > 2\delta^2$, and hence α cannot be the minimizer as $g_{\boldsymbol{h},\boldsymbol{x}}(\alpha) > g_{\boldsymbol{h},\boldsymbol{x}}(\beta)$ according to (290). This concludes the proof. □

The next lemma reveals the local strong convexity of $g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)$ when α is close to one.

Lemma 53. *Assume that $\max\{\|\boldsymbol{h}-\boldsymbol{h}^\star\|_2, \|\boldsymbol{x}-\boldsymbol{x}^\star\|_2\} \leq \delta$ for some sufficiently small constant $\delta > 0$. Then, for any α satisfying $|\alpha - 1| \leq 18\delta$ and any $u, v \in \mathbb{C}$, one has*

$$\left[u^{\mathsf{H}}, v^{\mathsf{H}}\right]\nabla^2 g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)\begin{bmatrix} u \\ v \end{bmatrix} \geq \frac{1}{2}\left(|u|^2 + |v|^2\right),$$

where $\nabla^2 g_{\boldsymbol{h},\boldsymbol{x}}(\cdot)$ stands for the Wirtinger Hessian of $g_{\boldsymbol{h},\boldsymbol{x}}(\cdot)$.

Proof. For simplicity of presentation, we use $g_{\boldsymbol{h},\boldsymbol{x}}(\alpha,\overline{\alpha})$ and $g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)$ interchangeably. By (289), for any $u, v \in \mathbb{C}$, one has

$$\left[u^{\mathsf{H}}, v^{\mathsf{H}}\right]\nabla^2 g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)\begin{bmatrix} u \\ v \end{bmatrix} = \underbrace{\left(\|\boldsymbol{x}\|_2^2 + |\alpha|^{-4}\|\boldsymbol{h}\|_2^2\right)}_{:=\beta_1}\left(|u|^2+|v|^2\right) + 2\underbrace{\mathrm{Re}\left[u^{\mathsf{H}}v\left(2\alpha^{-1}(\overline{\alpha})^{-3}\|\boldsymbol{h}\|_2^2 - 2(\overline{\alpha})^{-3}\boldsymbol{h}^{\star\mathsf{H}}\boldsymbol{h}\right)\right]}_{:=\beta_2}.$$

We would like to demonstrate that this is at least on the order of $|u|^2 + |v|^2$. We first develop a lower bound on β_1. Given the assumption that $\max\{\|\boldsymbol{h}-\boldsymbol{h}^\star\|_2, \|\boldsymbol{x}-\boldsymbol{x}^\star\|_2\} \leq \delta$, one necessarily has

$$1-\delta \leq \|\boldsymbol{x}\|_2 \leq 1+\delta \qquad \text{and} \qquad 1-\delta \leq \|\boldsymbol{h}\|_2 \leq 1+\delta.$$

Thus, for any α obeying $|\alpha - 1| \leq 18\delta$, one has

$$\beta_1 \geq \left(1 + |\alpha|^{-4}\right)(1-\delta)^2 \geq \left(1 + (1+18\delta)^{-4}\right)(1-\delta)^2 \geq 1$$

as long as $\delta > 0$ is sufficiently small. Regarding the second term β_2, we utilizes the conditions $|\alpha - 1| \leq 18\delta$, $\|\boldsymbol{x}\|_2 \leq 1+\delta$ and $\|\boldsymbol{h}\|_2 \leq 1+\delta$ to get

$$|\beta_2| \leq 2|u||v||\alpha|^{-3}\left|\alpha^{-1}\|\boldsymbol{h}\|_2^2 - \boldsymbol{h}^{\star\mathsf{H}}\boldsymbol{h}\right|$$

$$
\begin{aligned}
&= 2\left|u\right|\left|v\right|\left|\alpha\right|^{-3}\left|\left(\alpha^{-1}-1\right)\left\|\boldsymbol{h}\right\|_2^2-\left(\boldsymbol{h}^{\star}-\boldsymbol{h}\right)^{\mathsf{H}}\boldsymbol{h}\right| \\
&\le 2\left|u\right|\left|v\right|\left|\alpha\right|^{-3}\left(\left|\alpha^{-1}-1\right|\left\|\boldsymbol{h}\right\|_2^2+\left\|\boldsymbol{h}-\boldsymbol{h}^{\star}\right\|_2\left\|\boldsymbol{h}\right\|_2\right) \\
&\le 2\left|u\right|\left|v\right|\left(1-18\delta\right)^{-3}\left(\frac{18\delta}{1-18\delta}\left(1+\delta\right)^2+\delta\left(1+\delta\right)\right) \\
&\lesssim \delta\left(\left|u\right|^2+\left|v\right|^2\right),
\end{aligned}
$$

where the last relation holds since $2\left|u\right|\left|v\right|\le\left|u\right|^2+\left|v\right|^2$ and $\delta>0$ is sufficiently small. Combining the previous bounds on β_1 and β_2, we arrive at

$$
\left[u^{\mathsf{H}},v^{\mathsf{H}}\right]\nabla^2 g_{\boldsymbol{h},\boldsymbol{x}}\left(\alpha\right)\left[\begin{array}{c}u\\v\end{array}\right]\ge\left(1-O(\delta)\right)\left(\left|u\right|^2+\left|v\right|^2\right)\ge\frac{1}{2}\left(\left|u\right|^2+\left|v\right|^2\right)
$$

as long as δ is sufficiently small. This completes the proof. □

Additionally, in a local region surrounding the optimizer, the alignment parameter is Lipschitz continuous, namely, the difference of the alignment parameters associated with two distinct vector pairs is at most proportional to the ℓ_2 distance between the two vector pairs involved, as demonstrated below.

Lemma 54. *Suppose that the vectors $\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{h}_1,\boldsymbol{h}_2\in\mathbb{C}^K$ satisfy*

$$
\max\left\{\left\|\boldsymbol{x}_1-\boldsymbol{x}^{\star}\right\|_2,\left\|\boldsymbol{h}_1-\boldsymbol{h}^{\star}\right\|_2,\left\|\boldsymbol{x}_2-\boldsymbol{x}^{\star}\right\|_2,\left\|\boldsymbol{h}_2-\boldsymbol{h}^{\star}\right\|_2\right\}\le\delta\le 1/4 \tag{291}
$$

for some sufficiently small constant $\delta>0$. Denote by α_1 and α_2 the minimizers of $g_{\boldsymbol{h}_1,\boldsymbol{x}_1}\left(\alpha\right)$ and $g_{\boldsymbol{h}_2,\boldsymbol{x}_2}\left(\alpha\right)$, respectively. Then we have

$$
\left|\alpha_1-\alpha_2\right|\lesssim\left\|\boldsymbol{x}_1-\boldsymbol{x}_2\right\|_2+\left\|\boldsymbol{h}_1-\boldsymbol{h}_2\right\|_2.
$$

Proof. Since α_1 minimizes $g_{\boldsymbol{h}_1,\boldsymbol{x}_1}\left(\alpha\right)$, the mean-value form of Taylor's theorem (see Appendix D.3.1) gives

$$
\begin{aligned}
&g_{\boldsymbol{h}_1,\boldsymbol{x}_1}\left(\alpha_2\right)\ge g_{\boldsymbol{h}_1,\boldsymbol{x}_1}\left(\alpha_1\right)\\
&\qquad= g_{\boldsymbol{h}_1,\boldsymbol{x}_1}\left(\alpha_2\right)+\nabla g_{\boldsymbol{h}_1,\boldsymbol{x}_1}\left(\alpha_2\right)^{\mathsf{H}}\left[\begin{array}{c}\alpha_1-\alpha_2\\\overline{\alpha_1-\alpha_2}\end{array}\right]+\frac{1}{2}\left(\overline{\alpha_1-\alpha_2},\alpha_1-\alpha_2\right)\nabla^2 g_{\boldsymbol{h}_1,\boldsymbol{x}_1}\left(\widetilde{\alpha}\right)\left[\begin{array}{c}\alpha_1-\alpha_2\\\overline{\alpha_1-\alpha_2}\end{array}\right],
\end{aligned}
$$

where $\widetilde{\alpha}$ is some complex number lying between α_1 and α_2, and $\nabla g_{\boldsymbol{h}_1,\boldsymbol{x}_1}$ and $\nabla^2 g_{\boldsymbol{h}_1,\boldsymbol{x}_1}$ are the Wirtinger gradient and Hessian of $g_{\boldsymbol{h}_1,\boldsymbol{x}_1}\left(\cdot\right)$, respectively. Rearrange the previous inequality to obtain

$$
\left|\alpha_1-\alpha_2\right|\lesssim\frac{\left\|\nabla g_{\boldsymbol{h}_1,\boldsymbol{x}_1}\left(\alpha_2\right)\right\|_2}{\lambda_{\min}\left(\nabla^2 g_{\boldsymbol{h}_1,\boldsymbol{x}_1}\left(\widetilde{\alpha}\right)\right)} \tag{292}
$$

as long as $\lambda_{\min}\left(\nabla^2 g_{\boldsymbol{h}_1,\boldsymbol{x}_1}\left(\widetilde{\alpha}\right)\right)>0$. This calls for evaluation of the Wirtinger gradient and Hessian of $g_{\boldsymbol{h}_1,\boldsymbol{x}_1}\left(\cdot\right)$.

Regarding the Wirtinger Hessian, by the assumption (291), we can invoke Lemma 52 with $\beta=1$ to reach $\max\left\{\left|\alpha_1-1\right|,\left|\alpha_2-1\right|\right\}\le 18\delta$. This together with Lemma 53 implies

$$
\lambda_{\min}\left(\nabla^2 g_{\boldsymbol{h}_1,\boldsymbol{x}_1}\left(\widetilde{\alpha}\right)\right)\ge 1/2,
$$

since $\widetilde{\alpha}$ lies between α_1 and α_2.

For the Wirtinger gradient, since α_2 is the minimizer of $g_{\boldsymbol{h}_2,\boldsymbol{x}_2}\left(\alpha\right)$, the first-order optimality condition [KD09, equation (38)] requires $\nabla g_{\boldsymbol{h}_2,\boldsymbol{x}_2}\left(\alpha_2\right)=\boldsymbol{0}$, which gives

$$
\left\|\nabla g_{\boldsymbol{h}_1,\boldsymbol{x}_1}\left(\alpha_2\right)\right\|_2=\left\|\nabla g_{\boldsymbol{h}_1,\boldsymbol{x}_1}\left(\alpha_2\right)-\nabla g_{\boldsymbol{h}_2,\boldsymbol{x}_2}\left(\alpha_2\right)\right\|_2.
$$

Plug in the gradient expression (288) to reach

$$
\begin{aligned}
&\left\|\nabla g_{\boldsymbol{h}_1,\boldsymbol{x}_1}\left(\alpha_2\right)-\nabla g_{\boldsymbol{h}_2,\boldsymbol{x}_2}\left(\alpha_2\right)\right\|_2\\
&\quad=\sqrt{2}\left|\left[\alpha_2\left\|\boldsymbol{x}_1\right\|_2^2-\boldsymbol{x}_1^{\mathsf{H}}\boldsymbol{x}^{\star}-\alpha_2^{-1}\left(\overline{\alpha_2}\right)^{-2}\left\|\boldsymbol{h}_1\right\|_2^2+\left(\overline{\alpha_2}\right)^{-2}\boldsymbol{h}^{\star\mathsf{H}}\boldsymbol{h}_1\right]\right.
\end{aligned}
$$

$$- \left[\alpha_2 \|\boldsymbol{x}_2\|_2^2 - \boldsymbol{x}_2^{\mathsf{H}}\boldsymbol{x}^\star - \alpha_2^{-1}\left(\overline{\alpha_2}\right)^{-2}\|\boldsymbol{h}_2\|_2^2 + \left(\overline{\alpha_2}\right)^{-2}\boldsymbol{h}^{\star\mathsf{H}}\boldsymbol{h}_2\right]\Big|$$

$$\lesssim |\alpha_2| \left|\|\boldsymbol{x}_1\|_2^2 - \|\boldsymbol{x}_2\|_2^2\right| + \left|\boldsymbol{x}_1^{\mathsf{H}}\boldsymbol{x}^\star - \boldsymbol{x}_2^{\mathsf{H}}\boldsymbol{x}^\star\right| + \frac{1}{|\alpha_2|^3}\left|\|\boldsymbol{h}_1\|_2^2 - \|\boldsymbol{h}_2\|_2^2\right| + \frac{1}{|\alpha_2|^2}\left|\boldsymbol{h}^{\star\mathsf{H}}\boldsymbol{h}_1 - \boldsymbol{h}^{\star\mathsf{H}}\boldsymbol{h}_2\right|$$

$$\lesssim |\alpha_2| \left|\|\boldsymbol{x}_1\|_2^2 - \|\boldsymbol{x}_2\|_2^2\right| + \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 + \frac{1}{|\alpha_2|^3}\left|\|\boldsymbol{h}_1\|_2^2 - \|\boldsymbol{h}_2\|_2^2\right| + \frac{1}{|\alpha_2|^2}\|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2,$$

where the last line follows from the triangle inequality. It is straightforward to see that

$$1/2 \le |\alpha_2| \le 2, \qquad \left|\|\boldsymbol{x}_1\|_2^2 - \|\boldsymbol{x}_2\|_2^2\right| \lesssim \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2, \qquad \left|\|\boldsymbol{h}_1\|_2^2 - \|\boldsymbol{h}_2\|_2^2\right| \lesssim \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2$$

under the condition (291) and the assumption $\|\boldsymbol{x}^\star\|_2 = \|\boldsymbol{h}^\star\|_2 = 1$, where the first inequality follows from Lemma 52. Taking these estimates together reveals that

$$\|\nabla g_{\boldsymbol{h}_1,\boldsymbol{x}_1}(\alpha_2) - \nabla g_{\boldsymbol{h}_2,\boldsymbol{x}_2}(\alpha_2)\|_2 \lesssim \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 + \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2.$$

The proof is accomplished by substituting the two bounds on the gradient and the Hessian into (292). □

Further, if two vector pairs are both close to the optimizer, then their distance after alignement (w.r.t. the optimizer) cannot be much larger than their distance without alignment, as revealed by the following lemma.

Lemma 55. *Suppose that the vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathbb{C}^K$ satisfy*

$$\max\left\{\|\boldsymbol{x}_1 - \boldsymbol{x}^\star\|_2, \|\boldsymbol{h}_1 - \boldsymbol{h}^\star\|_2, \|\boldsymbol{x}_2 - \boldsymbol{x}^\star\|_2, \|\boldsymbol{h}_2 - \boldsymbol{h}^\star\|_2\right\} \le \delta \le 1/4 \tag{293}$$

for some sufficiently small constant $\delta > 0$. Denote by α_1 and α_2 the minimizers of $g_{\boldsymbol{h}_1,\boldsymbol{x}_1}(\alpha)$ and $g_{\boldsymbol{h}_2,\boldsymbol{x}_2}(\alpha)$, respectively. Then we have

$$\|\alpha_1\boldsymbol{x}_1 - \alpha_2\boldsymbol{x}_2\|_2^2 + \left\|\frac{1}{\overline{\alpha_1}}\boldsymbol{h}_1 - \frac{1}{\overline{\alpha_2}}\boldsymbol{h}_2\right\|_2^2 \lesssim \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2 + \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2^2.$$

Proof. To start with, we control the magnitudes of α_1 and α_2. Lemma 52 together with the assumption (293) guarantees that

$$1/2 \le |\alpha_1| \le 2 \qquad \text{and} \qquad 1/2 \le |\alpha_2| \le 2.$$

Now we can prove the lemma. The triangle inequality gives

$$\begin{aligned}
\|\alpha_1\boldsymbol{x}_1 - \alpha_2\boldsymbol{x}_2\|_2 &= \|\alpha_1(\boldsymbol{x}_1 - \boldsymbol{x}_2) + (\alpha_1 - \alpha_2)\boldsymbol{x}_2\|_2 \\
&\le |\alpha_1|\,\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 + |\alpha_1 - \alpha_2|\,\|\boldsymbol{x}_2\|_2 \\
&\overset{\text{(i)}}{\le} 2\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 + 2|\alpha_1 - \alpha_2| \\
&\overset{\text{(ii)}}{\lesssim} \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 + \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2,
\end{aligned}$$

where (i) holds since $|\alpha_1| \le 2$ and $\|\boldsymbol{x}_2\|_2 \le 1 + \delta \le 2$, and (ii) arises from Lemma 54 that $|\alpha_1 - \alpha_2| \lesssim \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 + \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2$. Similarly,

$$\begin{aligned}
\left\|\frac{1}{\overline{\alpha_1}}\boldsymbol{h}_1 - \frac{1}{\overline{\alpha_2}}\boldsymbol{h}_2\right\|_2 &= \left\|\frac{1}{\overline{\alpha_1}}(\boldsymbol{h}_1 - \boldsymbol{h}_2) + \left(\frac{1}{\overline{\alpha_1}} - \frac{1}{\overline{\alpha_2}}\right)\boldsymbol{h}_2\right\|_2 \\
&\le \left|\frac{1}{\overline{\alpha_1}}\right|\|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2 + \left|\frac{1}{\overline{\alpha_1}} - \frac{1}{\overline{\alpha_2}}\right|\|\boldsymbol{h}_2\|_2 \\
&\le 2\|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2 + 2\frac{|\alpha_1 - \alpha_2|}{|\alpha_1\alpha_2|} \\
&\lesssim \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 + \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2,
\end{aligned}$$

where the last inequality comes from Lemma 54 as well as the facts that $|\alpha_1| \ge 1/2$ and $|\alpha_2| \ge 1/2$ as shown above. Combining all of the above bounds and recognizing that $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 + \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2 \le \sqrt{2\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2 + 2\|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2^2}$, we conclude the proof. □

Finally, there is a useful identity associated with the minimizer of $\widetilde{g}(\alpha)$ as defined below.

Lemma 56. *For any $\boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{C}^K$, denote*

$$\alpha^\sharp := \arg\min_\alpha \widetilde{g}(\alpha), \quad \textit{where} \quad \widetilde{g}(\alpha) := \left\| \frac{1}{\overline{\alpha}} \boldsymbol{h}_1 - \boldsymbol{h}_2 \right\|_2^2 + \|\alpha \boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2 .$$

Let $\widetilde{\boldsymbol{x}}_1 = \alpha^\sharp \boldsymbol{x}_1$ and $\widetilde{\boldsymbol{h}}_1 = \frac{1}{\overline{\alpha^\sharp}} \boldsymbol{h}_1$, then we have

$$\left\| \widetilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2 \right\|_2^2 + \boldsymbol{x}_2^{\mathsf{H}} \left(\widetilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2 \right) = \left\| \widetilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2 \right\|_2^2 + \left(\widetilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2 \right)^{\mathsf{H}} \boldsymbol{h}_2.$$

Proof. We can rewrite the function $\widetilde{g}(\alpha)$ as

$$\begin{aligned} \widetilde{g}(\alpha) &= |\alpha|^2 \|\boldsymbol{x}_1\|_2^2 + \|\boldsymbol{x}_2\|_2^2 - (\alpha \boldsymbol{x}_1)^{\mathsf{H}} \boldsymbol{x}_2 - \boldsymbol{x}_2^{\mathsf{H}} (\alpha \boldsymbol{x}_1) + \left| \frac{1}{\overline{\alpha}} \right|^2 \|\boldsymbol{h}_1\|_2^2 + \|\boldsymbol{h}_2\|_2^2 - \left(\frac{1}{\overline{\alpha}} \boldsymbol{h}_1 \right)^{\mathsf{H}} \boldsymbol{h}_2 - \boldsymbol{h}_2^{\mathsf{H}} \left(\frac{1}{\overline{\alpha}} \boldsymbol{h}_1 \right) \\ &= \overline{\alpha} \alpha \|\boldsymbol{x}_1\|_2^2 + \|\boldsymbol{x}_2\|_2^2 - \overline{\alpha} \boldsymbol{x}_1^{\mathsf{H}} \boldsymbol{x}_2 - \alpha \boldsymbol{x}_2^{\mathsf{H}} \boldsymbol{x}_1 + \frac{1}{\overline{\alpha} \alpha} \|\boldsymbol{h}_1\|_2^2 + \|\boldsymbol{h}_2\|_2^2 - \frac{1}{\alpha} \boldsymbol{h}_1^{\mathsf{H}} \boldsymbol{h}_2 - \frac{1}{\overline{\alpha}} \boldsymbol{h}_2^{\mathsf{H}} \boldsymbol{h}_1. \end{aligned}$$

The first-order optimality condition [KD09, equation (38)] requires

$$\left. \frac{\partial \widetilde{g}}{\partial \overline{\alpha}} \right|_{\alpha = \alpha^\sharp} = \alpha^\sharp \|\boldsymbol{x}_1\|_2^2 - \boldsymbol{x}_1^{\mathsf{H}} \boldsymbol{x}_2 + \frac{1}{\alpha^\sharp} \left(- \frac{1}{\overline{\alpha^\sharp}^2} \right) \|\boldsymbol{h}_1\|_2^2 - \left(- \frac{1}{\overline{\alpha^\sharp}^2} \right) \boldsymbol{h}_2^{\mathsf{H}} \boldsymbol{h}_1 = 0,$$

which further simplifies to

$$\|\widetilde{\boldsymbol{x}}_1\|_2^2 - \widetilde{\boldsymbol{x}}_1^{\mathsf{H}} \boldsymbol{x}_2 = \left\| \widetilde{\boldsymbol{h}}_1 \right\|_2^2 - \boldsymbol{h}_2^{\mathsf{H}} \widetilde{\boldsymbol{h}}_1$$

since $\widetilde{\boldsymbol{x}}_1 = \alpha^\sharp \boldsymbol{x}_1$, $\widetilde{\boldsymbol{h}}_1 = \frac{1}{\overline{\alpha^\sharp}} \boldsymbol{h}_1$, and $\alpha^\sharp \neq 0$ (otherwise $\widetilde{g}(\alpha^\sharp) = \infty$ and cannot be the minimizer). Furthermore, this condition is equivalent to

$$\widetilde{\boldsymbol{x}}_1^{\mathsf{H}} \left(\widetilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2 \right) = \left(\widetilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2 \right)^{\mathsf{H}} \widetilde{\boldsymbol{h}}_1.$$

Recognizing that

$$\begin{aligned} \widetilde{\boldsymbol{x}}_1^{\mathsf{H}} \left(\widetilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2 \right) &= \boldsymbol{x}_2^{\mathsf{H}} \left(\widetilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2 \right) + \left(\widetilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2 \right)^{\mathsf{H}} \left(\widetilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2 \right) = \boldsymbol{x}_2^{\mathsf{H}} \left(\widetilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2 \right) + \|\widetilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2\|_2^2, \\ \widetilde{\boldsymbol{h}}_1^{\mathsf{H}} \big(\widetilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2 \big) &= \boldsymbol{h}_2^{\mathsf{H}} \big(\widetilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2 \big) + \big(\widetilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2 \big)^{\mathsf{H}} \big(\widetilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2 \big) = \boldsymbol{h}_2^{\mathsf{H}} \big(\widetilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2 \big) + \|\widetilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2\|_2^2, \end{aligned}$$

we arrive at the desired identity. □

D.3.4 Matrix concentration inequalities

The proof for blind deconvolution is largely built upon the concentration of random matrices that are functions of $\left\{ \boldsymbol{a}_j \boldsymbol{a}_j^{\mathsf{H}} \right\}$. In this subsection, we collect the measure concentration results for various forms of random matrices that we encounter in the analysis.

Lemma 57. *Suppose $\boldsymbol{a}_j \overset{i.i.d.}{\sim} \mathcal{N}\left(\boldsymbol{0}, \frac{1}{2} \boldsymbol{I}_K \right) + i \mathcal{N}\left(\boldsymbol{0}, \frac{1}{2} \boldsymbol{I}_K \right)$ for every $1 \leq j \leq m$, and $\{c_j\}_{1 \leq j \leq m}$ are a set of fixed numbers. Then there exist some universal constants $\widetilde{C}_1, \widetilde{C}_2 > 0$ such that for all $t \geq 0$*

$$\mathbb{P}\left(\left\| \sum_{j=1}^m c_j (\boldsymbol{a}_j \boldsymbol{a}_j^{\mathsf{H}} - \boldsymbol{I}_K) \right\| \geq t \right) \leq 2 \exp\left(\widetilde{C}_1 K - \widetilde{C}_2 \min\left\{ \frac{t}{\max_j |c_j|}, \frac{t^2}{\sum_{j=1}^m c_j^2} \right\} \right).$$

Proof. This is a simple variant of [Ver12, Theorem 5.39], which uses the Bernstein inequality and the standard covering argument. Hence we omit its proof. □

Lemma 58. *Suppose* $\boldsymbol{a}_j \overset{i.i.d.}{\sim} \mathcal{N}\left(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_K\right) + i\mathcal{N}\left(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_K\right)$ *for every* $1 \leq j \leq m$. *Then there exist some absolute constants* $\widetilde{C}_1, \widetilde{C}_2, \widetilde{C}_3 > 0$ *such that for all* $\max\{1, 3\widetilde{C}_1 K/\widetilde{C}_2\}/m \leq \varepsilon \leq 1$, *one has*

$$\mathbb{P}\left(\sup_{|J|\leq \varepsilon m}\left\|\sum_{j\in J}\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\right\| \geq 4\widetilde{C}_3\varepsilon m \log\frac{e}{\varepsilon}\right) \leq 2\exp\left(-\frac{\widetilde{C}_2\widetilde{C}_3}{3}\varepsilon m\log\frac{e}{\varepsilon}\right),$$

where $J \subseteq [m]$ *and* $|J|$ *denotes its cardinality.*

Proof. The proof relies on Lemma 57 and the union bound. First, invoke Lemma 57 to see that for any fixed $J \subseteq [m]$ and for all $t \geq 0$, we have

$$\mathbb{P}\left(\left\|\sum_{j\in J}(\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}} - \boldsymbol{I}_K)\right\| \geq |J|\, t\right) \leq 2\exp\left(\widetilde{C}_1 K - \widetilde{C}_2|J|\min\left\{t, t^2\right\}\right), \tag{294}$$

for some constants $\widetilde{C}_1, \widetilde{C}_2 > 0$, and as a result,

$$\begin{aligned}
\mathbb{P}\left(\sup_{|J|\leq \varepsilon m}\left\|\sum_{j\in J}\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\right\| \geq \lceil\varepsilon m\rceil(1+t)\right) &\overset{\text{(i)}}{\leq} \mathbb{P}\left(\sup_{|J|=\lceil\varepsilon m\rceil}\left\|\sum_{j\in J}\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\right\| \geq \lceil\varepsilon m\rceil(1+t)\right)\\
&\leq \mathbb{P}\left(\sup_{|J|=\lceil\varepsilon m\rceil}\left\|\sum_{j\in J}(\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}} - \boldsymbol{I}_K)\right\| \geq \lceil\varepsilon m\rceil t\right)\\
&\overset{\text{(ii)}}{\leq} \binom{m}{\lceil\varepsilon m\rceil}\cdot 2\exp\left(\widetilde{C}_1 K - \widetilde{C}_2\lceil\varepsilon m\rceil\min\left\{t, t^2\right\}\right),
\end{aligned}$$

where $\lceil c\rceil$ denotes the smallest integer that is no smaller than c. Here, (i) holds since we take the supremum over a larger set and (ii) results from (294) and the union bound. Apply the elementary inequality $\binom{n}{k} \leq (en/k)^k$ for any $0 \leq k \leq n$ to obtain

$$\begin{aligned}
\mathbb{P}\left(\sup_{|J|\leq \varepsilon m}\left\|\sum_{j\in J}\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\right\| \geq \lceil\varepsilon m\rceil(1+t)\right) &\leq 2\left(\frac{em}{\lceil\varepsilon m\rceil}\right)^{\lceil\varepsilon m\rceil}\exp\left(\widetilde{C}_1 K - \widetilde{C}_2\lceil\varepsilon m\rceil\min\left\{t, t^2\right\}\right)\\
&\leq 2\left(\frac{e}{\varepsilon}\right)^{2\varepsilon m}\exp\left(\widetilde{C}_1 K - \widetilde{C}_2\varepsilon m\min\left\{t, t^2\right\}\right)\\
&= 2\exp\left[\widetilde{C}_1 K - \varepsilon m\left(\widetilde{C}_2\min\left\{t, t^2\right\} - 2\log(e/\varepsilon)\right)\right].
\end{aligned} \tag{295}$$

where the second inequality uses $\varepsilon m \leq \lceil\varepsilon m\rceil \leq 2\varepsilon m$ whenever $1/m \leq \varepsilon \leq 1$.

The proof is then completed by taking $\widetilde{C}_3 \geq \max\{1, 6/\widetilde{C}_2\}$ and $t = \widetilde{C}_3\log(e/\varepsilon)$. To see this, it is easy to check that $\min\{t, t^2\} = t$ since $t \geq 1$. In addition, one has $\widetilde{C}_1 K \leq \widetilde{C}_2\varepsilon m/3 \leq \widetilde{C}_2\varepsilon m t/3$, and $2\log(e/\varepsilon) \leq \widetilde{C}_2 t/3$. Combine the estimates above with (295) to arrive at

$$\begin{aligned}
\mathbb{P}\left(\sup_{|J|\leq \varepsilon m}\left\|\sum_{j\in J}\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\right\| \geq 4\widetilde{C}_3\varepsilon m\log(e/\varepsilon)\right) &\overset{\text{(i)}}{\leq} \mathbb{P}\left(\sup_{|J|\leq \varepsilon m}\left\|\sum_{j\in J}\boldsymbol{a}_j\boldsymbol{a}_j^{\mathsf{H}}\right\| \geq \lceil\varepsilon m\rceil(1+t)\right)\\
&\leq 2\exp\left[\widetilde{C}_1 K - \varepsilon m\left(\widetilde{C}_2\min\left\{t, t^2\right\} - 2\log(e/\varepsilon)\right)\right]\\
&\overset{\text{(ii)}}{\leq} 2\exp\left(-\varepsilon m\widetilde{C}_2 t/3\right) = 2\exp\left(-\frac{\widetilde{C}_2\widetilde{C}_3}{3}\varepsilon m\log(e/\varepsilon)\right)
\end{aligned}$$

as claimed. Here (i) holds due to the facts that $\lceil\varepsilon m\rceil \leq 2\varepsilon m$ and $1 + t \leq 2t \leq 2\widetilde{C}_3\log(e/\varepsilon)$. The inequality (ii) arises from the estimates listed above. $\square$

Lemma 59. *Suppose $m \gg K \log^3 m$. With probability exceeding $1 - O\left(m^{-10}\right)$, we have*

$$\left\| \sum_{j=1}^{m} \left|\boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^{\star}\right|^2 \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} - \boldsymbol{I}_K \right\| \lesssim \sqrt{\frac{K}{m} \log m}.$$

Proof. The identity $\sum_{j=1}^{m} \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} = \boldsymbol{I}_K$ allows us to rewrite the quantity on the left-hand side as

$$\left\| \sum_{j=1}^{m} \left|\boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^{\star}\right|^2 \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} - \boldsymbol{I}_K \right\| = \left\| \sum_{j=1}^{m} \underbrace{\left(\left|\boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^{\star}\right|^2 - 1 \right) \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}}}_{:=\boldsymbol{Z}_j} \right\|,$$

where the $\boldsymbol{Z}_j$'s are independent zero-mean random matrices. To control the above spectral norm, we resort to the matrix Bernstein inequality [Kol11, Theorem 2.7]. To this end, we first need to upper bound the sub-exponential norm $\|\cdot\|_{\psi_1}$ (see definition in [Ver12]) of each summand $\boldsymbol{Z}_j$, i.e.

$$\big\| \|\boldsymbol{Z}_j\| \big\|_{\psi_1} = \|\boldsymbol{b}_j\|_2^2 \left\| \left|\boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^{\star}\right|^2 - 1 \right\|_{\psi_1} \lesssim \|\boldsymbol{b}_j\|_2^2 \left\| \left|\boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^{\star}\right|^2 \right\|_{\psi_1} \lesssim \frac{K}{m},$$

where we make use of the facts that

$$\|\boldsymbol{b}_j\|_2^2 = K/m \qquad \text{and} \qquad \left\| \left|\boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^{\star}\right|^2 \right\|_{\psi_1} \lesssim 1.$$

We further need to bound the variance parameter, that is,

$$\begin{aligned}
\sigma_0^2 &:= \left\| \mathbb{E}\left[\sum_{j=1}^{m} \boldsymbol{Z}_j \boldsymbol{Z}_j^{\mathsf{H}} \right] \right\| = \left\| \mathbb{E}\left[\sum_{j=1}^{m} \left(\left|\boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^{\star}\right|^2 - 1 \right)^2 \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} \right] \right\| \\
&\lesssim \left\| \sum_{j=1}^{m} \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} \right\| = \frac{K}{m} \left\| \sum_{j=1}^{m} \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} \right\| = \frac{K}{m},
\end{aligned}$$

where the second line arises since $\mathbb{E}\big[\big(|\boldsymbol{a}_j^{\mathsf{H}} \boldsymbol{x}^{\star}|^2 - 1 \big)^2 \big] \asymp 1$, $\|\boldsymbol{b}_j\|_2^2 = K/m$, and $\sum_{j=1}^{m} \boldsymbol{b}_j \boldsymbol{b}_j^{\mathsf{H}} = \boldsymbol{I}_K$. A direct application of the matrix Bernstein inequality [Kol11, Theorem 2.7] leads us to conclude that with probability exceeding $1 - O\left(m^{-10}\right)$,

$$\left\| \sum\nolimits_{j=1}^{m} \boldsymbol{Z}_j \right\| \lesssim \max\left\{ \sqrt{\frac{K}{m} \log m}, \frac{K}{m} \log^2 m \right\} \asymp \sqrt{\frac{K}{m} \log m},$$

where the last relation holds under the assumption that $m \gg K \log^3 m$. $\square$

D.3.5 Matrix perturbation bounds

We also need the following perturbation bound on the top singular vectors of a given matrix. The following lemma is parallel to Lemma 34.

Lemma 60. *Let $\sigma_1(\boldsymbol{A})$, $\boldsymbol{u}$ and $\boldsymbol{v}$ be the leading singular value, left and right singular vectors of $\boldsymbol{A}$, respectively, and let $\sigma_1(\widetilde{\boldsymbol{A}})$, $\widetilde{\boldsymbol{u}}$ and $\widetilde{\boldsymbol{v}}$ be the leading singular value, left and right singular vectors of $\widetilde{\boldsymbol{A}}$, respectively. Suppose $\sigma_1(\boldsymbol{A})$ and $\sigma_1(\widetilde{\boldsymbol{A}})$ are not identically zero, then one has*

$$\left| \sigma_1(\boldsymbol{A}) - \sigma_1(\widetilde{\boldsymbol{A}}) \right| \le \left\| \big(\boldsymbol{A} - \widetilde{\boldsymbol{A}}\big) \boldsymbol{v} \right\|_2 + \left(\|\boldsymbol{u} - \widetilde{\boldsymbol{u}}\|_2 + \|\boldsymbol{v} - \widetilde{\boldsymbol{v}}\|_2 \right) \|\widetilde{\boldsymbol{A}}\|;$$

$$\left\| \sqrt{\sigma_1(\boldsymbol{A})}\, \boldsymbol{u} - \sqrt{\sigma_1(\widetilde{\boldsymbol{A}})}\, \widetilde{\boldsymbol{u}} \right\|_2 + \left\| \sqrt{\sigma_1(\boldsymbol{A})}\, \boldsymbol{v} - \sqrt{\sigma_1(\widetilde{\boldsymbol{A}})}\, \widetilde{\boldsymbol{v}} \right\|_2 \le \sqrt{\sigma_1(\boldsymbol{A})} \left(\|\boldsymbol{u} - \widetilde{\boldsymbol{u}}\|_2 + \|\boldsymbol{v} - \widetilde{\boldsymbol{v}}\|_2 \right) + \frac{2 \left| \sigma_1(\boldsymbol{A}) - \sigma_1(\widetilde{\boldsymbol{A}}) \right|}{\sqrt{\sigma_1(\boldsymbol{A})} + \sqrt{\sigma_1(\widetilde{\boldsymbol{A}})}}.$$

Proof. The first claim follows since

$$\begin{aligned}
\left|\sigma_1(\boldsymbol{A}) - \sigma_1(\widetilde{\boldsymbol{A}})\right| &= \left|\boldsymbol{u}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{v} - \widetilde{\boldsymbol{u}}^{\mathsf{H}}\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{v}}\right| \\
&\leq \left|\boldsymbol{u}^{\mathsf{H}}(\boldsymbol{A} - \widetilde{\boldsymbol{A}})\boldsymbol{v}\right| + \left|\boldsymbol{u}^{\mathsf{H}}\widetilde{\boldsymbol{A}}\boldsymbol{v} - \widetilde{\boldsymbol{u}}^{\mathsf{H}}\widetilde{\boldsymbol{A}}\boldsymbol{v}\right| + \left|\widetilde{\boldsymbol{u}}^{\mathsf{H}}\widetilde{\boldsymbol{A}}\boldsymbol{v} - \widetilde{\boldsymbol{u}}^{\mathsf{H}}\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{v}}\right| \\
&\leq \left\|(\boldsymbol{A} - \widetilde{\boldsymbol{A}})\boldsymbol{v}\right\|_2 + \|\boldsymbol{u} - \widetilde{\boldsymbol{u}}\|_2 \left\|\widetilde{\boldsymbol{A}}\right\| + \left\|\widetilde{\boldsymbol{A}}\right\| \|\boldsymbol{v} - \widetilde{\boldsymbol{v}}\|_2 .
\end{aligned}$$

With regards to the second claim, we see that

$$\begin{aligned}
\left\|\sqrt{\sigma_1(\boldsymbol{A})}\,\boldsymbol{u} - \sqrt{\sigma_1(\widetilde{\boldsymbol{A}})}\,\widetilde{\boldsymbol{u}}\right\|_2 &\leq \left\|\sqrt{\sigma_1(\boldsymbol{A})}\,\boldsymbol{u} - \sqrt{\sigma_1(\boldsymbol{A})}\,\widetilde{\boldsymbol{u}}\right\|_2 + \left\|\sqrt{\sigma_1(\boldsymbol{A})}\,\widetilde{\boldsymbol{u}} - \sqrt{\sigma_1(\widetilde{\boldsymbol{A}})}\,\widetilde{\boldsymbol{u}}\right\|_2 \\
&= \sqrt{\sigma_1(\boldsymbol{A})}\,\|\boldsymbol{u} - \widetilde{\boldsymbol{u}}\|_2 + \left|\sqrt{\sigma_1(\boldsymbol{A})} - \sqrt{\sigma_1(\widetilde{\boldsymbol{A}})}\right| \\
&= \sqrt{\sigma_1(\boldsymbol{A})}\,\|\boldsymbol{u} - \widetilde{\boldsymbol{u}}\|_2 + \frac{\left|\sigma_1(\boldsymbol{A}) - \sigma_1(\widetilde{\boldsymbol{A}})\right|}{\sqrt{\sigma_1(\boldsymbol{A})} + \sqrt{\sigma_1(\widetilde{\boldsymbol{A}})}}.
\end{aligned}$$

Similarly, one can obtain

$$\left\|\sqrt{\sigma_1(\boldsymbol{A})}\,\boldsymbol{v} - \sqrt{\sigma_1(\widetilde{\boldsymbol{A}})}\,\widetilde{\boldsymbol{v}}\right\|_2 \leq \sqrt{\sigma_1(\boldsymbol{A})}\,\|\boldsymbol{v} - \widetilde{\boldsymbol{v}}\|_2 + \frac{\left|\sigma_1(\boldsymbol{A}) - \sigma_1(\widetilde{\boldsymbol{A}})\right|}{\sqrt{\sigma_1(\boldsymbol{A})} + \sqrt{\sigma_1(\widetilde{\boldsymbol{A}})}}.$$

Add these two inequalities to complete the proof. □