A Proofs for phase retrieval

Before proceeding, we gather a few simple facts. By Lemma 33, the sampling vectors obey

$$\max_{1\le j\le m} \|\boldsymbol{a}_j\|_2 \le \sqrt{6n} \tag{95}$$

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

$$\max_{1\le j\le m} \left|\boldsymbol{a}_j^\top \boldsymbol{x}^\natural\right| \le 5\sqrt{\log n} \tag{96}$$

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

A.1 Proof of Lemma 1

We start with the smoothness bound, namely, $\nabla^2 f(\boldsymbol{x}) \preceq O(\log n)\boldsymbol{I}_n$. It suffices to prove that

$$\left\|\nabla^2 f(\boldsymbol{x})\right\| \lesssim \log n.$$

To this end, we first decompose the Hessian (cf. (42)) into three components as follows:

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

In the sequel, we control the above 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

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

- The second term $\boldsymbol{\Lambda}_2$ can be controlled by means of Lemma 35:

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

  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

$$\|\boldsymbol{\Lambda}_1\| \le \left\|\frac{3}{m}\sum_{j=1}^m \left|\boldsymbol{a}_j^\top\left(\boldsymbol{x}-\boldsymbol{x}^\natural\right)\right|\left|\boldsymbol{a}_j^\top\left(\boldsymbol{x}+\boldsymbol{x}^\natural\right)\right|\boldsymbol{a}_j\boldsymbol{a}_j^\top\right\|. \tag{97}$$

  Under the assumption $\max_{1\le j\le m}\left|\boldsymbol{a}_j^\top\left(\boldsymbol{x}-\boldsymbol{x}^\natural\right)\right| \le C_2\sqrt{\log n}$ and the condition (96), we can also obtain

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

  Substitution into (97) 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 an easy consequence of Lemma 34.

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

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

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

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

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

We begin with the simpler term $\mathbf{\Lambda}_5$. Lemma 35 implies that with probability at least $1-O(n^{-10})$,

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

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

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

To bound $\mathbf{\Lambda}_4$, invoke Lemma 36 to conclude that with probability at least $1-C_2e^{-c_2 m}$ (for some constants $c_2, C_2>0$),

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

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

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

with ξ being a standard Gaussian random variable. By the assumption that $\left\|\boldsymbol{x}-\boldsymbol{x}^\natural\right\|_2 \leq 2C_1$, we have

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

allowing one to obtain

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

This further implies

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

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

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

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

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

as claimed.

A.2 Proof of Lemma 2

Using the update rule (cf. (17)) as well as the mean value theorem for vector-valued functions [Lan93, Theorem 4.2], we get

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

where we denote $\boldsymbol{x}\left(\tau\right)=\boldsymbol{x}^{\natural}+\tau(\boldsymbol{x}^{t}-\boldsymbol{x}^{\natural})$. Here, the first identity makes use of the fact that $\nabla f(\boldsymbol{x}^{\natural})=\boldsymbol{0}$.

1. Under the condition (43), it is self-evident that for all $0\leq\tau\leq 1$,

$$\left\|\boldsymbol{x}\left(\tau\right)-\boldsymbol{x}^{\natural}\right\|_2=\left\|\tau(\boldsymbol{x}^{t}-\boldsymbol{x}^{\natural})\right\|_2\leq 2C_1 \qquad \text{and}$$
$$\max_{1\leq l\leq m}\left|\boldsymbol{a}_l^{\top}\left(\boldsymbol{x}(\tau)-\boldsymbol{x}^{\natural}\right)\right|\leq\max_{1\leq l\leq m}\left|\boldsymbol{a}_l^{\top}\tau\left(\boldsymbol{x}^{t}-\boldsymbol{x}^{\natural}\right)\right|\leq C_2\sqrt{\log n}.$$

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

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

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

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

   which immediately yields

$$\left\|\boldsymbol{x}^{t+1}-\boldsymbol{x}^{\natural}\right\|_2\leq\left\|\boldsymbol{I}_n-\eta\int_0^1\nabla^2 f\left(\boldsymbol{x}\left(\tau\right)\right)\mathrm{d}\tau\right\|\cdot\left\|\boldsymbol{x}^{t}-\boldsymbol{x}^{\natural}\right\|_2\leq\left(1-\eta/2\right)\left\|\boldsymbol{x}^{t}-\boldsymbol{x}^{\natural}\right\|_2. \tag{98}$$

2. On the other hand, if we know that $\|\boldsymbol{x}^{t}-\boldsymbol{x}^{\natural}\|_2\leq\frac{1}{n}\|\boldsymbol{x}^{\natural}\|_2$, then the Cauchy-Schwarz inequality and the fact (95) indicate that

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

   leading to the satisfaction of (43) and therefore, the contraction (98).

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 (45) to reach

$$\left\|\boldsymbol{x}^{t}-\boldsymbol{x}^{\natural}\right\|_2\leq(1-\eta/2)^t\left\|\boldsymbol{x}^{0}-\boldsymbol{x}^{\natural}\right\|_2\leq C_1(1-\eta/2)^t\left\|\boldsymbol{x}^{\natural}\right\|_2. \tag{99}$$

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

$$\left\|\boldsymbol{x}^{T_0}-\boldsymbol{x}^{\natural}\right\|_2\leq C_1(1-\eta/2)^{T_0}\left\|\boldsymbol{x}^{\natural}\right\|_2\ll\frac{1}{n}\left\|\boldsymbol{x}^{\natural}\right\|_2, \tag{100}$$

provided that $\eta\asymp 1/\log n$. Invoking Lemma 2 again with the condition that $\left\|\boldsymbol{x}^{T_0}-\boldsymbol{x}^{\natural}\right\|_2\ll\frac{1}{n}\left\|\boldsymbol{x}^{\natural}\right\|_2$ yields

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

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

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

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

$$\max_{1\leq j\leq m}\left|\boldsymbol{a}_j^{\top}\left(\boldsymbol{x}^{t}-\boldsymbol{x}^{\natural}\right)\right|\leq\max_{1\leq j\leq m}\left\|\boldsymbol{a}_j\right\|_2\left\|\boldsymbol{x}^{t}-\boldsymbol{x}^{\natural}\right\|_2\ll\sqrt{n}\cdot\frac{1}{n}\ll\sqrt{\log n}, \tag{102}$$

where the penultimate relation uses the conditions (95) and (101).

A.4 Proof of Lemma 4

Suppose that

$$\max_{1\le l\le m}\left\|\boldsymbol{x}^{t}-\boldsymbol{x}^{t,(l)}\right\|_{2}\le C_{3}\sqrt{\frac{\log n}{n}},\tag{103}$$

and we would like to establish the bound (103) for $t+1$. Going through the same derivation as in (52) will result in

$$\max_{1\le l\le m}\left|\boldsymbol{a}_{l}^{\top}\left(\boldsymbol{x}^{t,(l)}-\boldsymbol{x}^{\natural}\right)\right|\le C_{4}\sqrt{\log n}\tag{104}$$

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

We use the gradient update rule once again to decompose

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

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

1. We first control the term $\boldsymbol{\nu}_{2}^{(l)}$, which is easier to deal with. Specifically,

$$\begin{aligned}\|\boldsymbol{\nu}_{2}^{(l)}\|_{2}&\le\eta\frac{\|\boldsymbol{a}_{l}\|_{2}}{m}\left|\big(\boldsymbol{a}_{l}^{\top}\boldsymbol{x}^{t,(l)}\big)^{2}-\big(\boldsymbol{a}_{l}^{\top}\boldsymbol{x}^{\natural}\big)^{2}\right|\left|\boldsymbol{a}_{l}^{\top}\boldsymbol{x}^{t,(l)}\right|\\&\overset{\text{(i)}}{\lesssim}C_{4}(C_{4}+5)(C_{4}+10)\eta\frac{n\log n}{m}\sqrt{\frac{\log n}{n}}\overset{\text{(ii)}}{\le}c\eta\sqrt{\frac{\log n}{n}},\end{aligned}$$

for any small constant $c>0$. Here (i) follows since $\max_{1\le l\le m}\|\boldsymbol{a}_{l}\|_{2}\le\sqrt{6n}$ (see (95)) and, in view of (104),

$$\begin{aligned}\left|\big(\boldsymbol{a}_{l}^{\top}\boldsymbol{x}^{t,(l)}\big)^{2}-\big(\boldsymbol{a}_{l}^{\top}\boldsymbol{x}^{\natural}\big)^{2}\right|&\le\left|\boldsymbol{a}_{l}^{\top}\big(\boldsymbol{x}^{t,(l)}-\boldsymbol{x}^{\natural}\big)\right|\left(\left|\boldsymbol{a}_{l}^{\top}\big(\boldsymbol{x}^{t,(l)}-\boldsymbol{x}^{\natural}\big)\right|+2\left|\boldsymbol{a}_{l}^{\top}\boldsymbol{x}^{\natural}\right|\right)\le C_{4}(C_{4}+10)\log n,\\\text{and}\qquad\left|\boldsymbol{a}_{l}^{\top}\boldsymbol{x}^{t,(l)}\right|&\le\left|\boldsymbol{a}_{l}^{\top}\big(\boldsymbol{x}^{t,(l)}-\boldsymbol{x}^{\natural}\big)\right|+\left|\boldsymbol{a}_{l}^{\top}\boldsymbol{x}^{\natural}\right|\le(C_{4}+5)\sqrt{\log n}.\end{aligned}$$

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

2. For the term $\boldsymbol{\nu}_{1}^{(l)}$, the mean value theorem for vector-valued functions [Lan93, Theorem 4.2] tells us that

$$\boldsymbol{\nu}_{1}^{(l)}=\left[\boldsymbol{I}_{n}-\eta\int_{0}^{1}\nabla^{2}f\left(\boldsymbol{x}\left(\tau\right)\right)\mathrm{d}\tau\right]\big(\boldsymbol{x}^{t}-\boldsymbol{x}^{t,(l)}\big),$$

where we abuse the notation and denote $\boldsymbol{x}\left(\tau\right)=\boldsymbol{x}^{t,(l)}+\tau(\boldsymbol{x}^{t}-\boldsymbol{x}^{t,(l)})$. By the induction hypotheses (49) and the condition (104), one can verify that

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

$$\max_{1\le l\le m}\left|\boldsymbol{a}_{l}^{\top}\left(\boldsymbol{x}\left(\tau\right)-\boldsymbol{x}^{\natural}\right)\right|\le\tau\max_{1\le l\le m}\left|\boldsymbol{a}_{l}^{\top}\left(\boldsymbol{x}^{t}-\boldsymbol{x}^{\natural}\right)\right|+(1-\tau)\max_{1\le l\le m}\left|\boldsymbol{a}_{l}^{\top}\left(\boldsymbol{x}^{t,(l)}-\boldsymbol{x}^{\natural}\right)\right|\le C_{2}\sqrt{\log n}$$

for all $0\le\tau\le1$, as long as $C_4\le C_2$. The second line follows directly from (104). To see why (105) holds, we note that

$$\left\|\boldsymbol{x}^{t,(l)}-\boldsymbol{x}^{\natural}\right\|_{2}\le\left\|\boldsymbol{x}^{t,(l)}-\boldsymbol{x}^{t}\right\|_{2}+\left\|\boldsymbol{x}^{t}-\boldsymbol{x}^{\natural}\right\|_{2}\le C_{3}\sqrt{\frac{\log n}{n}}+C_{1},$$

where the second inequality follows from the induction hypotheses (103) and (49a). This combined with (49a) gives

$$\left\|\boldsymbol{x}\left(\tau\right)-\boldsymbol{x}^{\natural}\right\|_{2}\leq\tau\left\|\boldsymbol{x}^{t}-\boldsymbol{x}^{\natural}\right\|_{2}+\left(1-\tau\right)\left\|\boldsymbol{x}^{t,(l)}-\boldsymbol{x}^{\natural}\right\|_{2}$$
$$\leq\tau C_{1}+\left(1-\tau\right)\left(C_{3}\sqrt{\frac{\log n}{n}}+C_{1}\right)\leq 2C_{1}$$

as long as n is large enough, thus justifying (105). Hence by Lemma 1, $\nabla^{2}f\left(\boldsymbol{x}\left(\tau\right)\right)$ is positive definite and almost well-conditioned. By choosing $\eta\leq\frac{1}{5C_{2}(10+C_{2})\log n}$, we get

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

3. Combine the preceding bounds on $\boldsymbol{\nu}_{1}^{(l)}$ and $\boldsymbol{\nu}_{2}^{(l)}$ as well as the induction bound (103) to arrive at

$$\left\|\boldsymbol{x}^{t+1}-\boldsymbol{x}^{t+1,(l)}\right\|_{2}\leq\left(1-\eta/2\right)\left\|\boldsymbol{x}^{t}-\boldsymbol{x}^{t,(l)}\right\|_{2}+c\eta\sqrt{\frac{\log n}{n}}\leq C_{3}\sqrt{\frac{\log n}{n}}. \tag{106}$$

This establishes (51) for the $(t+1)$th iteration.

A.5 Proof of Lemma 5

We first assume that the leading eigenvector $\tilde{\boldsymbol{x}}^{0}$ of $\boldsymbol{Y}$ obeys

$$\left\|\tilde{\boldsymbol{x}}^{0}-\boldsymbol{x}^{\natural}\right\|_{2}\leq\left\|\tilde{\boldsymbol{x}}^{0}+\boldsymbol{x}^{\natural}\right\|_{2}.$$

One can then invoke the Davis-Kahan $\sin\Theta$ theorem [YWS15, Corollary 1] to obtain

$$\left\|\tilde{\boldsymbol{x}}^{0}-\boldsymbol{x}^{\natural}\right\|_{2}\leq 2\sqrt{2}\frac{\left\|\boldsymbol{Y}-\mathbb{E}\left[\boldsymbol{Y}\right]\right\|}{\lambda_{1}\left(\mathbb{E}\left[\boldsymbol{Y}\right]\right)-\lambda_{2}\left(\mathbb{E}\left[\boldsymbol{Y}\right]\right)}.$$

Note that (54) — $\|\boldsymbol{Y}-\mathbb{E}[\boldsymbol{Y}]\|\leq\delta$ — is a direct consequence of Lemma 35. Additionally, the fact that $\mathbb{E}\left[\boldsymbol{Y}\right]=\boldsymbol{I}+2\boldsymbol{x}^{\natural}\boldsymbol{x}^{\natural\top}$ gives

$$\lambda_{1}\left(\mathbb{E}\left[\boldsymbol{Y}\right]\right)=3,\qquad\lambda_{2}\left(\mathbb{E}\left[\boldsymbol{Y}\right]\right)=1,\qquad\text{and}\qquad\lambda_{1}\left(\mathbb{E}\left[\boldsymbol{Y}\right]\right)-\lambda_{2}\left(\mathbb{E}\left[\boldsymbol{Y}\right]\right)=2.$$

Combining this spectral gap and the inequality $\|\boldsymbol{Y}-\mathbb{E}[\boldsymbol{Y}]\|\leq\delta$, we arrive at

$$\left\|\tilde{\boldsymbol{x}}^{0}-\boldsymbol{x}^{\natural}\right\|_{2}\leq\sqrt{2}\delta.$$

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 (54) 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, $\lambda_{1}\left(\boldsymbol{Y}\right)\geq 3-\delta\geq 0$ when $\delta\leq 1$. This further implies that

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

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

$$\left\|\boldsymbol{x}^{0}-\boldsymbol{x}^{\natural}\right\|_{2}=\left\|\sqrt{\frac{\lambda_{1}\left(\boldsymbol{Y}\right)}{3}}\tilde{\boldsymbol{x}}^{0}-\boldsymbol{x}^{\natural}\right\|_{2}=\left\|\sqrt{\frac{\lambda_{1}\left(\boldsymbol{Y}\right)}{3}}\tilde{\boldsymbol{x}}^{0}-\tilde{\boldsymbol{x}}^{0}+\tilde{\boldsymbol{x}}^{0}-\boldsymbol{x}^{\natural}\right\|_{2}$$

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

Notably, the above results are consistent with our assumption (40) that $\left\|\boldsymbol{x}^0-\boldsymbol{x}^\natural\right\|_2 \leq \left\|\boldsymbol{x}^0+\boldsymbol{x}^\natural\right\|_2$, as long as δ is sufficiently small. If instead one has

$$
\left\|\tilde{\boldsymbol{x}}^0-\boldsymbol{x}^\natural\right\|_2 \geq \left\|\tilde{\boldsymbol{x}}^0+\boldsymbol{x}^\natural\right\|_2,
$$

then repeating the same analysis gives $\left\|\boldsymbol{x}^0+\boldsymbol{x}^\natural\right\|_2 \leq 2\delta$, contradicting our assumption $\left\|\boldsymbol{x}^0-\boldsymbol{x}^\natural\right\|_2 \leq \left\|\boldsymbol{x}^0+\boldsymbol{x}^\natural\right\|_2$ when $\delta>0$ is small.

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 $\delta>0$,

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

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

1. We start by controlling $\left\|\tilde{\boldsymbol{x}}^0-\tilde{\boldsymbol{x}}^{0,(l)}\right\|_2$. Combining (55) and (108) yields

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

For δ sufficiently small, this implies that $\left\|\tilde{\boldsymbol{x}}^0-\tilde{\boldsymbol{x}}^{0,(l)}\right\|_2 \leq \left\|\tilde{\boldsymbol{x}}^0+\tilde{\boldsymbol{x}}^{0,(l)}\right\|_2$, and hence the Davis-Kahan $\sin\Theta$ theorem [DK70] gives

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

Here, the second inequality uses Weyl's inequality:

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

2. We now connect $\|\boldsymbol{x}^0-\boldsymbol{x}^{0,(l)}\|_2$ with $\|\tilde{\boldsymbol{x}}^0-\tilde{\boldsymbol{x}}^{0,(l)}\|_2$. Applying the Weyl's inequality and (54) yields

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

and, similarly, $\lambda_1(\boldsymbol{Y}^{(l)}), \|\boldsymbol{Y}\|, \|\boldsymbol{Y}^{(l)}\| \in [2,4]$. Invoke Lemma 37 to arrive at

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

where the last inequality comes from (109).

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

$$
\begin{aligned}
\max_{1\leq l\leq m}\left\|\big(\boldsymbol{Y}-\boldsymbol{Y}^{(l)}\big)\tilde{\boldsymbol{x}}^{0,(l)}\right\|_2 &= \max_{1\leq l\leq m}\frac{1}{m}\left\|\big(\boldsymbol{a}_l^\top\boldsymbol{x}^\natural\big)^2\boldsymbol{a}_l\boldsymbol{a}_l^\top\tilde{\boldsymbol{x}}^{0,(l)}\right\|_2 \\
&\leq \max_{1\leq l\leq m}\frac{\big(\boldsymbol{a}_l^\top\boldsymbol{x}^\natural\big)^2\left|\boldsymbol{a}_l^\top\tilde{\boldsymbol{x}}^{0,(l)}\right|\left\|\boldsymbol{a}_l\right\|_2}{m}
\end{aligned}
$$

$$
\begin{aligned}
&\overset{\text{(i)}}{\lesssim} \frac{\log n\cdot\sqrt{\log n}\cdot\sqrt{n}}{m} \\
&\asymp \sqrt{\frac{\log n}{n}}\cdot\frac{n\log n}{m}. \qquad (112)
\end{aligned}
$$

The inequality (i) makes use of the fact $\max_l \left|\boldsymbol{a}_l^\top \boldsymbol{x}^\natural\right| \leq 5\sqrt{\log n}$ (cf. (96)), the bound $\max_l \|\boldsymbol{a}_l\|_2 \leq 6\sqrt{n}$ (cf. (95)), and $\max_l \left|\boldsymbol{a}_l^\top \tilde{\boldsymbol{x}}^{0,(l)}\right| \leq 5\sqrt{\log n}$ (due to statistical independence and standard Gaussian concentration). As long as $m/(n\log n)$ is sufficiently large, substituting the above bound into (111) leads us to conclude that

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

for any constant $C_3>0$.

B Proofs for matrix completion

Before proceeding to the proofs, we make note of an immediate consequence of the incoherence property (24):

$$
\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} \leq \sqrt{\frac{\kappa\mu}{n}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}} \leq \sqrt{\frac{\kappa\mu r}{n}}\left\|\boldsymbol{X}^\natural\right\|. \qquad (114)
$$

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

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

In addition, the problem of finding $\hat{\boldsymbol{H}}^t$ (see (25)) is called the *orthogonal Procrustes problem* [tB77]. It is well-known that the minimizer $\hat{\boldsymbol{H}}^t$ always exists and is given by

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

where for any matrix $\boldsymbol{B}$ with singular value decomposition $\boldsymbol{B}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$, the sign matrix $\mathrm{sgn}(\boldsymbol{B})$ is defined to be

$$
\mathrm{sgn}(\boldsymbol{B}) := \boldsymbol{U}\boldsymbol{V}^\top. \qquad (115)
$$

B.1 Proof of Lemma 7

By the expression of the Hessian in (59), one can decompose

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

The basic idea to upper / lower bound $\mathrm{vec}\left(\boldsymbol{V}\right)^\top \nabla^2 f_{\mathrm{clean}}\left(\boldsymbol{X}\right)\mathrm{vec}\left(\boldsymbol{V}\right)$ 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

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

When it comes to the lower bound, one discovers that

$$\begin{aligned}
\alpha_4 &= \frac{1}{2}\left\{\left\|\boldsymbol{V}\boldsymbol{X}^{\natural\top}\right\|_{\mathrm{F}}^2 + \left\|\boldsymbol{X}^{\natural}\boldsymbol{V}^{\top}\right\|_{\mathrm{F}}^2 + 2\mathrm{Tr}\left(\boldsymbol{X}^{\natural\top}\boldsymbol{V}\boldsymbol{X}^{\natural\top}\boldsymbol{V}\right)\right\} \\
&\geq \sigma_{\min}\|\boldsymbol{V}\|_{\mathrm{F}}^2 + \mathrm{Tr}\left[\left(\boldsymbol{Z}+\boldsymbol{X}^{\natural}-\boldsymbol{Z}\right)^{\top}\boldsymbol{V}\left(\boldsymbol{Z}+\boldsymbol{X}^{\natural}-\boldsymbol{Z}\right)^{\top}\boldsymbol{V}\right] \\
&\geq \sigma_{\min}\|\boldsymbol{V}\|_{\mathrm{F}}^2 + \mathrm{Tr}\left(\boldsymbol{Z}^{\top}\boldsymbol{V}\boldsymbol{Z}^{\top}\boldsymbol{V}\right) - 2\left\|\boldsymbol{Z}-\boldsymbol{X}^{\natural}\right\|\|\boldsymbol{Z}\|\|\boldsymbol{V}\|_{\mathrm{F}}^2 - \left\|\boldsymbol{Z}-\boldsymbol{X}^{\natural}\right\|^2\|\boldsymbol{V}\|_{\mathrm{F}}^2 \\
&\geq \left(\sigma_{\min}-5\delta\sigma_{\max}\right)\|\boldsymbol{V}\|_{\mathrm{F}}^2 + \mathrm{Tr}\left(\boldsymbol{Z}^{\top}\boldsymbol{V}\boldsymbol{Z}^{\top}\boldsymbol{V}\right), \qquad (116)
\end{aligned}$$

where the last line comes from the assumption that

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

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 48 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 (116) 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 $\kappa\delta \ll 1$.

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 $\left\|\mathcal{P}_{\Omega}\left(\boldsymbol{V}\boldsymbol{X}^{\natural\top}+\boldsymbol{X}^{\natural}\boldsymbol{V}^{\top}\right)\right\|_{\mathrm{F}}^2$ as well. These identities yield

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

For β_2, one has

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

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

$$|\beta_2| \leq \frac{1}{p}\left\|\mathcal{P}_\Omega\left(\boldsymbol{V}\left(\boldsymbol{X}-\boldsymbol{X}^\natural\right)^\top\right)\right\|_{\mathrm{F}}^2 + \frac{2}{p}\left\|\mathcal{P}_\Omega\left(\boldsymbol{V}\left(\boldsymbol{X}-\boldsymbol{X}^\natural\right)^\top\right)\right\|_{\mathrm{F}}\left\|\mathcal{P}_\Omega\left(\boldsymbol{X}^\natural\boldsymbol{V}^\top\right)\right\|_{\mathrm{F}}. \tag{117}$$

This then calls for upper bounds on the following two terms

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

The injectivity of $\mathcal{P}_\Omega$ (i.e. [CR09, Section 4.2] or Lemma 44)—when restricted to the tangent space of $\boldsymbol{M}^\natural$—gives

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

for any fixed constant $\gamma > 0$, provided that $n^2 p \gtrsim \mu n r \log n$. In addition,

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

which holds as long as $np/\log n$ is sufficiently large. Taken collectively, the above bounds yield that for any small constant $\gamma > 0$,

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

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

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

where (i) utilizes the incoherence condition (cf. (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

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

In view of Lemma 39,

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

as soon as $\epsilon\sqrt{\kappa^{3}\mu r}\log n\ll 1$. This in turn implies that

$$|\alpha_{2}|\leq\frac{1}{10}\sigma_{\min}\left\|\boldsymbol{V}\right\|_{\mathrm{F}}^{2}.$$

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

4. The last term α_3 can also be controlled using the injectivity of $\mathcal{P}_{\Omega}$ when restricted to the tangent space of $\boldsymbol{M}^{\natural}$. Specifically, it follows from the bounds in [CR09, Section 4.2] or Lemma 44 that

$$|\alpha_{3}|\leq\gamma\left\|\boldsymbol{V}\boldsymbol{X}^{\natural\top}+\boldsymbol{X}^{\natural}\boldsymbol{V}^{\top}\right\|_{\mathrm{F}}^{2}\leq 4\gamma\sigma_{\max}\left\|\boldsymbol{V}\right\|_{\mathrm{F}}^{2}\leq\frac{1}{10}\sigma_{\min}\left\|\boldsymbol{V}\right\|_{\mathrm{F}}^{2}$$

for any small constant $\gamma>0$, as soon as $n^{2}p\gg\mu rn\log n$.

5. Taking all the preceding bounds collectively yields

$$\begin{aligned}\mathrm{vec}\left(\boldsymbol{V}\right)^{\top}\nabla^{2}f_{\mathrm{clean}}\left(\boldsymbol{X}\right)\mathrm{vec}\left(\boldsymbol{V}\right) &\geq\alpha_{4}-|\alpha_{1}|-|\alpha_{2}|-|\alpha_{3}| \\ &\geq\left(\frac{9}{10}-\frac{3}{10}\right)\sigma_{\min}\left\|\boldsymbol{V}\right\|_{\mathrm{F}}^{2}\geq\frac{1}{2}\sigma_{\min}\left\|\boldsymbol{V}\right\|_{\mathrm{F}}^{2}\end{aligned}$$

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

$$\begin{aligned}\left|\mathrm{vec}\left(\boldsymbol{V}\right)^{\top}\nabla^{2}f_{\mathrm{clean}}\left(\boldsymbol{X}\right)\mathrm{vec}\left(\boldsymbol{V}\right)\right| &\leq\alpha_{4}+|\alpha_{1}|+|\alpha_{2}|+|\alpha_{3}| \\ &\leq\left(2\sigma_{\max}+\frac{3}{10}\sigma_{\min}\right)\left\|\boldsymbol{V}\right\|_{\mathrm{F}}^{2}\leq 2.5\sigma_{\max}\left\|\boldsymbol{V}\right\|_{\mathrm{F}}^{2}\end{aligned}$$

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

$$\left\|\nabla^{2}f_{\mathrm{clean}}\left(\boldsymbol{X}\right)\right\|\leq 2.5\sigma_{\max}$$

as claimed.

B.2 Proof of Lemma 8

Given that $\hat{\boldsymbol{H}}^{t+1}$ is chosen to minimize the error in the Frobenius norm sense (cf. (25)), we have

$$\begin{aligned}&\left\|\boldsymbol{X}^{t+1}\hat{\boldsymbol{H}}^{t+1}-\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}}\leq\left\|\boldsymbol{X}^{t+1}\hat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}}=\left\|\left[\boldsymbol{X}^{t}-\eta\nabla f\left(\boldsymbol{X}^{t}\right)\right]\hat{\boldsymbol{H}}^{t}-\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}} \\ &\quad=\left\|\boldsymbol{X}^{t}\hat{\boldsymbol{H}}^{t}-\eta\nabla f\big(\boldsymbol{X}^{t}\hat{\boldsymbol{H}}^{t}\big)-\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}} \qquad (118)\\ &\quad=\left\|\boldsymbol{X}^{t}\hat{\boldsymbol{H}}^{t}-\eta\left[\nabla f_{\mathrm{clean}}\big(\boldsymbol{X}^{t}\hat{\boldsymbol{H}}^{t}\big)-\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{E}\right)\boldsymbol{X}^{t}\hat{\boldsymbol{H}}^{t}\right]-\boldsymbol{X}^{\natural}\right\| \\ &\quad\leq\underbrace{\left\|\boldsymbol{X}^{t}\hat{\boldsymbol{H}}^{t}-\eta\nabla f_{\mathrm{clean}}\big(\boldsymbol{X}^{t}\hat{\boldsymbol{H}}^{t}\big)-\big(\boldsymbol{X}^{\natural}-\eta\nabla f_{\mathrm{clean}}\big(\boldsymbol{X}^{\natural}\big)\big)\right\|_{\mathrm{F}}}_{:=I_{1}}+\eta\underbrace{\left\|\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{E}\right)\boldsymbol{X}^{t}\hat{\boldsymbol{H}}^{t}\right\|_{\mathrm{F}}}_{:=I_{2}}, \qquad (119)\end{aligned}$$

where (118) follows from the identity $\nabla f(\boldsymbol{X}^{t}\boldsymbol{R})=\nabla f\left(\boldsymbol{X}^{t}\right)\boldsymbol{R}$ for any rotation matrix $\boldsymbol{R}\in\mathbb{R}^{r\times r}$, and the last inequality utilizes the fact that $\nabla f_{\mathrm{clean}}(\boldsymbol{X}^{\natural})=\boldsymbol{0}$. It thus suffices to control I_1 and I_2.

1. For the second term in (119), it is easy to see that

$$I_2 \le \eta \left\| \frac{1}{p} \mathcal{P}_\Omega \left(\boldsymbol{E} \right) \right\| \left\| \boldsymbol{X}^t \hat{\boldsymbol{H}}^t \right\|_{\mathrm{F}} \le 2\eta \left\| \frac{1}{p} \mathcal{P}_\Omega \left(\boldsymbol{E} \right) \right\| \left\| \boldsymbol{X}^\natural \right\|_{\mathrm{F}} \le 2\eta C \sigma \sqrt{\frac{n}{p}} \| \boldsymbol{X}^\natural \|_{\mathrm{F}}.$$

Here, the second inequality holds because

$$\left\| \boldsymbol{X}^t \hat{\boldsymbol{H}}^t \right\|_{\mathrm{F}} \le \left\| \boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \right\|_{\mathrm{F}} + \left\| \boldsymbol{X}^\natural \right\|_{\mathrm{F}} \overset{\text{(i)}}{\le} 2 \left\| \boldsymbol{X}^\natural \right\|_{\mathrm{F}},$$

while (i) occurs under the hypothesis (67a) together with our assumptions on the noise and the sample complexity. The last inequality makes use of Lemma 46.

2. With regard to the first term I_1 in (119), it follows from the mean value theorem for vector-valued functions [Lan93, Theorem 4.2] that

$$\begin{aligned}
&\mathrm{vec}\left[\boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \eta \nabla f_{\mathrm{clean}}\big(\boldsymbol{X}^t \hat{\boldsymbol{H}}^t \big) - \big(\boldsymbol{X}^\natural - \eta \nabla f_{\mathrm{clean}}\big(\boldsymbol{X}^\natural \big) \big) \right] \\
&= \mathrm{vec}\left[\boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \right] - \eta \mathrm{vec}\left[\nabla f_{\mathrm{clean}}\big(\boldsymbol{X}^t \hat{\boldsymbol{H}}^t \big) - \nabla f_{\mathrm{clean}} \left(\boldsymbol{X}^\natural \right) \right] \\
&= \Bigg(\boldsymbol{I}_{nr} - \eta \underbrace{\int_0^1 \nabla^2 f_{\mathrm{clean}} \left(\boldsymbol{X}(\tau) \right) \mathrm{d}\tau}_{:= \boldsymbol{A}} \Bigg) \mathrm{vec}\left(\boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \right),
\end{aligned} \tag{120}$$

where we denote $\boldsymbol{X}(\tau) := \boldsymbol{X}^\natural + \tau(\boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural)$. Taking the squared Euclidean norm of both sides of the identity (120) leads to

$$\begin{aligned}
(I_1)^2 &= \mathrm{vec}\big(\boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \big)^\top \left(\boldsymbol{I}_{nr} - \eta \boldsymbol{A} \right)^2 \mathrm{vec}\big(\boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \big) \\
&= \mathrm{vec}\big(\boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \big)^\top \left[\boldsymbol{I}_{nr} - 2\eta \boldsymbol{A} + \eta^2 \boldsymbol{A}^2 \right] \mathrm{vec}\big(\boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \big) \\
&\le \left\| \boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \right\|_{\mathrm{F}}^2 + \eta^2 \left\| \boldsymbol{A} \right\|^2 \left\| \boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \right\|_{\mathrm{F}}^2 \\
&\qquad - 2\eta \, \mathrm{vec}\big(\boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \big)^\top \boldsymbol{A} \, \mathrm{vec}\big(\boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \big),
\end{aligned} \tag{121}$$

where in (121) we have used the fact that

$$\mathrm{vec}\big(\boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \big)^\top \boldsymbol{A}^2 \mathrm{vec}\big(\boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \big) \le \left\| \boldsymbol{A} \right\|^2 \left\| \mathrm{vec}\big(\boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \big) \right\|_2^2.$$

Based on (67b), it is easily seen that

$$\left\| \boldsymbol{X}\left(\tau \right) - \boldsymbol{X}^\natural \right\|_{2,\infty} \le \left(C_5 \mu r \sqrt{\frac{\log n}{np}} + \frac{C_8}{\sigma_{\min}} \sigma \sqrt{\frac{n \log n}{p}} \right) \left\| \boldsymbol{X}^\natural \right\|_{2,\infty}, \qquad \forall \tau \in [0,1].$$

Invoking Lemma 7 (by taking $\boldsymbol{X} = \boldsymbol{X}\left(\tau \right), \boldsymbol{Y} = \boldsymbol{X}^t$ and $\boldsymbol{Z} = \boldsymbol{X}^\natural$ in the lemma) gives

$$\mathrm{vec}\big(\boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \big)^\top \boldsymbol{A} \, \mathrm{vec}\big(\boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \big) \ge \frac{1}{2} \sigma_{\min} \left\| \boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \right\|_{\mathrm{F}}^2$$

$$\text{and} \qquad \left\| \boldsymbol{A} \right\| = \left\| \int_0^1 \nabla^2 f_{\mathrm{clean}} \left(\boldsymbol{X}(\tau) \right) \mathrm{d}\tau \right\| \le \frac{5}{2} \sigma_{\max}.$$

Substituting these two inequalities into (121) yields

$$\begin{aligned}
(I_1)^2 &\le \left(1 + \frac{25}{4} \eta^2 \sigma_{\max}^2 - \sigma_{\min} \eta \right) \left\| \boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \right\|_{\mathrm{F}}^2 \\
&\le \left(1 - \frac{1}{2} \sigma_{\min} \eta \right) \left\| \boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \right\|_{\mathrm{F}}^2
\end{aligned}$$

as long as $\eta \le \frac{2\sigma_{\min}}{25\sigma_{\max}^2}$, which further implies that

$$I_1 \le \left(1 - \frac{\sigma_{\min}}{4} \eta \right) \left\| \boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \right\|_{\mathrm{F}}.$$

3. Combining the preceding bounds on both I_1 and I_2 and making use of the hypothesis (67a), we conclude that

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

as long as $\eta \leq \frac{2}{5\sigma_{\max}}$, $\rho \geq 1-\frac{\sigma_{\min}}{4}\eta$ and C_1 is sufficiently large. This completes the proof.

B.3 Proof of Lemma 9

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

$$\tilde{\boldsymbol{X}}^{t+1} := \boldsymbol{X}^t\hat{\boldsymbol{H}}^t-\eta\frac{1}{p}\mathcal{P}_{\Omega}\left[\boldsymbol{X}^t\boldsymbol{X}^{t\top}-\left(\boldsymbol{M}^{\natural}+\boldsymbol{E}\right)\right]\boldsymbol{X}^{\natural}. \tag{122}$$

As we will justify shortly, this auxiliary matrix satisfies

$$\arg\min_{\boldsymbol{R}\in\mathbb{R}^{r\times r}:\boldsymbol{R}\boldsymbol{R}^{\top}=\boldsymbol{I}_r}\left\|\tilde{\boldsymbol{X}}^{t+1}\boldsymbol{R}-\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}}=\boldsymbol{I}_r,$$

meaning that it is already rotated to the direction that is most aligned with the truth. This important property facilitates analysis.

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

$$\left\|\boldsymbol{X}^{t+1}\hat{\boldsymbol{H}}^{t+1}-\boldsymbol{X}^{\natural}\right\| \leq \underbrace{\left\|\boldsymbol{X}^{t+1}\hat{\boldsymbol{H}}^{t+1}-\tilde{\boldsymbol{X}}^{t+1}\right\|}_{:=\alpha_1}+\underbrace{\left\|\tilde{\boldsymbol{X}}^{t+1}-\boldsymbol{X}^{\natural}\right\|}_{:=\alpha_2}.$$

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

$$\begin{aligned}
\alpha_2 &= \left\|\boldsymbol{X}^t\hat{\boldsymbol{H}}^t-\eta\frac{1}{p}\mathcal{P}_{\Omega}\left[\boldsymbol{X}^t\boldsymbol{X}^{t\top}-\left(\boldsymbol{M}^{\natural}+\boldsymbol{E}\right)\right]\boldsymbol{X}^{\natural}-\boldsymbol{X}^{\natural}\right\| \\
&\leq \eta\left\|\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{E}\right)\right\|\left\|\boldsymbol{X}^{\natural}\right\|+\left\|\boldsymbol{X}^t\hat{\boldsymbol{H}}^t-\eta\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{X}^t\boldsymbol{X}^{t\top}-\boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\boldsymbol{X}^{\natural}-\boldsymbol{X}^{\natural}\right\| \qquad (123)\\
&\leq \eta\left\|\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{E}\right)\right\|\left\|\boldsymbol{X}^{\natural}\right\|+\underbrace{\left\|\boldsymbol{X}^t\hat{\boldsymbol{H}}^t-\eta\left(\boldsymbol{X}^t\boldsymbol{X}^{t\top}-\boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\boldsymbol{X}^{\natural}-\boldsymbol{X}^{\natural}\right\|}_{:=\beta_1} \\
&\quad \underbrace{+\,\eta\left\|\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{X}^t\boldsymbol{X}^{t\top}-\boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\boldsymbol{X}^{\natural}-\left(\boldsymbol{X}^t\boldsymbol{X}^{t\top}-\boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\boldsymbol{X}^{\natural}\right\|}_{:=\beta_2}, \qquad (124)
\end{aligned}$$

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

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

which, by Lemma 48, satisfies the following symmetry property

$$\left(\hat{\boldsymbol{H}}^t\right)^{\top}\boldsymbol{X}^{t\top}\boldsymbol{X}^{\natural}=\boldsymbol{X}^{\natural\top}\boldsymbol{X}^t\hat{\boldsymbol{H}}^t \qquad \Longrightarrow \qquad \boldsymbol{\Delta}^{t\top}\boldsymbol{X}^{\natural}=\boldsymbol{X}^{\natural\top}\boldsymbol{\Delta}^t. \tag{125}$$

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

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

The leading term γ_1 can be upper bounded by

$$\begin{aligned}\gamma_1 &= \left\|\boldsymbol{\Delta}^t - \eta\boldsymbol{\Delta}^t\boldsymbol{\Sigma}^{\natural} - \eta\boldsymbol{X}^{\natural}\boldsymbol{\Delta}^{t\top}\boldsymbol{X}^{\natural}\right\| \\ &= \left\|\boldsymbol{\Delta}^t - \eta\boldsymbol{\Delta}^t\boldsymbol{\Sigma}^{\natural} - \eta\boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\boldsymbol{\Delta}^t\right\| \qquad (126) \\ &= \left\|\frac{1}{2}\boldsymbol{\Delta}^t\left(1 - 2\eta\boldsymbol{\Sigma}^{\natural}\right) + \frac{1}{2}\left(1 - 2\eta\boldsymbol{M}^{\natural}\right)\boldsymbol{\Delta}^t\right\| \\ &\leq \left(\frac{1}{2}\left\|1 - 2\eta\boldsymbol{\Sigma}^{\natural}\right\| + \frac{1}{2}\left\|1 - 2\eta\boldsymbol{M}^{\natural}\right\|\right)\left\|\boldsymbol{\Delta}^t\right\|,\end{aligned}$$

where in (126) follows from the symmetry property (125). By choosing $\eta \leq \frac{1}{2\sigma_{\max}}$, one can ensure the contraction

$$\gamma_1 \leq \left[\frac{1}{2}\left(1 - 2\eta\sigma_{\min}\right) + \frac{1}{2}\right]\left\|\boldsymbol{\Delta}^t\right\| = \left(1 - \eta\sigma_{\min}\right)\left\|\boldsymbol{\Delta}^t\right\|. \qquad (127)$$

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

$$\gamma_2 \leq \eta\left\|\boldsymbol{\Delta}^t\right\|^2\left\|\boldsymbol{X}^{\natural}\right\|. \qquad (128)$$

The bounds (127) and (128) taken collectively give

$$\beta_1 \leq \left(1 - \eta\sigma_{\min}\right)\left\|\boldsymbol{\Delta}^t\right\| + \eta\left\|\boldsymbol{\Delta}^t\right\|^2\left\|\boldsymbol{X}^{\natural}\right\|. \qquad (129)$$

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

$$\begin{aligned}\frac{1}{\eta}\beta_2 &= \left\|\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{\Delta}^t\boldsymbol{\Delta}^{t\top} + \boldsymbol{\Delta}^t\boldsymbol{X}^{\natural\top} + \boldsymbol{X}^{\natural}\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^{\natural} - \left[\boldsymbol{\Delta}^t\boldsymbol{\Delta}^{t\top} + \boldsymbol{\Delta}^t\boldsymbol{X}^{\natural\top} + \boldsymbol{X}^{\natural}\boldsymbol{\Delta}^{t\top}\right]\boldsymbol{X}^{\natural}\right\| \\ &\leq \underbrace{\left\|\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{\Delta}^t\boldsymbol{X}^{\natural\top}\right)\boldsymbol{X}^{\natural} - \left(\boldsymbol{\Delta}^t\boldsymbol{X}^{\natural\top}\right)\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}}}_{:=\theta_1} + \underbrace{\left\|\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{X}^{\natural}\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^{\natural} - \left(\boldsymbol{X}^{\natural}\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}}}_{:=\theta_2} \\ &\quad + \underbrace{\left\|\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{\Delta}^t\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^{\natural} - \left(\boldsymbol{\Delta}^t\boldsymbol{\Delta}^{t\top}\right)\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}}}_{:=\theta_3}, \qquad (130)\end{aligned}$$

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

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

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

where, as usual, $\delta_{l,j} = \mathbb{1}_{\{(l,j)\in\Omega\}}$. Lemma 38 together with the union bound reveals that

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

$$\lesssim \sqrt{\frac{\|\boldsymbol{X}^{\natural}\|_{2,\infty}^{2}\sigma_{\max}\log n}{p}} + \frac{\|\boldsymbol{X}^{\natural}\|_{2,\infty}^{2}\log n}{p}$$

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

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

which further reveals that

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

for arbitrarily small $\gamma > 0$. Here, (i) follows from the relation between $\|\cdot\|$ and $\|\cdot\|_{\mathrm{F}}$, that is $\|\boldsymbol{\Delta}^{t}\|_{\mathrm{F}} \leq \sqrt{r}\,\|\boldsymbol{\Delta}^{t}\|$, (ii) holds owing to the incoherence condition (114), while (iii) arises as long as

$$n^{2}p \gg \kappa^{3}\mu r^{2}\log n.$$

- For the second term θ_2 in (130), denote

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

whose lth row is given by

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

It is self-evident that

$$\theta_2 = \frac{1}{p}\left\|\boldsymbol{A}\right\|_{\mathrm{F}} = \frac{1}{p}\sqrt{\sum_{l=1}^{n}\left\|\boldsymbol{A}_{l,\cdot}\right\|_{2}^{2}}.$$

From the induction hypotheses (67b) and (67c), we have

$$\left\|\boldsymbol{\Delta}^{t}\right\|_{2,\infty} \leq C_{5}\rho^{t}\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} + C_{8}\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} := \xi \tag{131}$$

$$\left\|\boldsymbol{\Delta}^{t}\right\| \leq C_{9}\rho^{t}\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^{\natural}\right\| + C_{10}\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^{\natural}\right\| := \psi. \tag{132}$$

We also introduce a "truncation level"

$$\omega := 2p \frac{\xi}{\|\boldsymbol{X}^\natural\|_{2,\infty}} \sigma_{\max} \|\boldsymbol{X}^\natural_{l,\cdot}\|_2 \tag{133}$$

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

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

We will apply different strategies when upper bounding the terms ϕ_1 and ϕ_2.

***Lemma* 21.** *Under the sample complexity condition in Lemma 9, there exist some constants $C_0, C_1 > 0$ such that with probability exceeding $1 - C_0 \exp(-C_1 nr \log n)$,*

$$\phi_1 \lesssim \sqrt{p \xi^2 \sigma_{\max} \|\boldsymbol{X}^\natural\|_{2,\infty}^2 nr \log^2 n} \tag{134}$$

holds simultaneously for all $\boldsymbol{\Delta}^t$ obeying (131) and (132). Here, ξ is defined in (131).

The second term ϕ_2 is controlled via the following lemma.

***Lemma* 22.** *Under the sample complexity condition in Lemma 9, there exist some constants $c_1, C_1 > 0$ such that with probability at least $1 - C_1 \exp(-c_1 nr \log n)$,*

$$\phi_2 \lesssim \sqrt{\kappa \mu r^2 p \log^2 n} \xi \left\|\boldsymbol{X}^\natural\right\|^2, \tag{135}$$

holds simultaneously for all $\boldsymbol{\Delta}^t$ obeying (131) and (132). Here, ξ is defined in (131).

The bounds (134) and (135) taken collectively yield

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

- Next, we assert that the third term θ_3 in (130) should be smaller than θ_2. The proof follows by repeating the same proof 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

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

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

(c) Combining the preceding bounds on β_1 and β_2 and (124), we reach

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

$$\overset{\text{(iii)}}{\leq} \left(1 - \frac{\sigma_{\min}}{2}\eta\right) \left\|\boldsymbol{\Delta}^t\right\| + C\eta\sigma\sqrt{\frac{n}{p}} \left\|\boldsymbol{X}^\natural\right\|$$

$$+ \tilde{C}\eta\sqrt{\frac{\kappa^2\mu^2 r^3 \log^3 n}{np}}\sigma_{\max}\left(C_5\rho^t\mu r\sqrt{\frac{1}{np}} + C_8\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right)\left\|\boldsymbol{X}^\natural\right\|, \tag{136}$$

where (i) uses the definition of ξ (cf. (131)), (ii) holds if γ is small enough and if

$$\left\|\boldsymbol{\Delta}^t\right\| \left\|\boldsymbol{X}^\natural\right\| \ll \sigma_{\min},$$

and (iii) follows from Lemma 46 as well as the incoherence condition (114). An immediate consequence of (136) is that under the sample size condition and the noise condition of this lemma, one has

$$\alpha_2 = \left\|\tilde{\boldsymbol{X}}^{t+1} - \boldsymbol{X}^\natural\right\| \left\|\boldsymbol{X}^\natural\right\| \leq \frac{1}{2}\sigma_{\min} \tag{137}$$

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

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

$$\alpha_1 = \left\|\boldsymbol{X}^{t+1}\hat{\boldsymbol{H}}^t\boldsymbol{R}_1 - \tilde{\boldsymbol{X}}^{t+1}\boldsymbol{R}_2\right\|,$$

with

$$\boldsymbol{R}_1 = \left(\hat{\boldsymbol{H}}^t\right)^{-1}\hat{\boldsymbol{H}}^{t+1} := \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \left\|\boldsymbol{X}^{t+1}\hat{\boldsymbol{H}}^t\boldsymbol{R} - \boldsymbol{X}^\natural\right\|_{\mathrm{F}} \qquad \text{and} \qquad \boldsymbol{R}_2 = \boldsymbol{I}_r. \tag{138}$$

(a) First, we claim that

$$\boldsymbol{R}_2 = \boldsymbol{I}_r = \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \left\|\tilde{\boldsymbol{X}}^{t+1}\boldsymbol{R} - \boldsymbol{X}^\natural\right\|_{\mathrm{F}}. \tag{139}$$

Clearly, the fact that $\boldsymbol{R}_2 = \boldsymbol{I}_r$ needs explanation. In fact, in view of Lemma 48, the fact $\boldsymbol{R}_2 = \boldsymbol{I}_r$ follows if one can show that $\boldsymbol{X}^{\natural\top}\tilde{\boldsymbol{X}}^{t+1}$ is symmetric and positive semidefinite. First of all, it follows from Lemma 48 that $\boldsymbol{X}^{\natural\top}\boldsymbol{X}^t\hat{\boldsymbol{H}}^t$ is symmetric and, hence, by definition,

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

is also symmetric. Additionally,

$$\left\|\boldsymbol{X}^{\natural\top}\tilde{\boldsymbol{X}}^{t+1} - \boldsymbol{\Sigma}^\natural\right\| \leq \left\|\tilde{\boldsymbol{X}}^{t+1} - \boldsymbol{X}^\natural\right\| \left\|\boldsymbol{X}^\natural\right\| \leq \frac{1}{2}\sigma_{\min},$$

where the second relation holds according to (137). Applying Weyl's inequality guarantees

$$\boldsymbol{X}^{\natural\top}\tilde{\boldsymbol{X}}^{t+1} \succeq \boldsymbol{0},$$

thus justifying that $\boldsymbol{R}_2 = \boldsymbol{I}_r$ via Lemma 48.

(b) With (138) and (139) in place, we resort to Lemma 41 to establish the bound. Specifically, take $\boldsymbol{X}_1 = \tilde{\boldsymbol{X}}^{t+1}$ and $\boldsymbol{X}_2 = \boldsymbol{X}^{t+1}\hat{\boldsymbol{H}}^t$, and it comes from (137) that

$$\left\|\boldsymbol{X}_1 - \boldsymbol{X}^\natural\right\| \left\|\boldsymbol{X}^\natural\right\| \leq \frac{1}{2}\sigma_{\min}.$$

Moreover, we have

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

in which

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

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

This allows one to derive

$$
\begin{aligned}
\left\| \boldsymbol{X}^{t+1} \hat{\boldsymbol{H}}^t - \tilde{\boldsymbol{X}}^{t+1} \right\| &\le \eta \left\| \frac{1}{p} \mathcal{P}_\Omega \left[\boldsymbol{X}^t \boldsymbol{X}^{t\top} - \boldsymbol{M} \right] \left(\boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \right) \right\| + \eta \left\| \frac{1}{p} \mathcal{P}_\Omega \left[\boldsymbol{E} \right] \left(\boldsymbol{X}^t \hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural \right) \right\| \\
&\le \left(2n \left\| \boldsymbol{\Delta}^t \right\|_{2,\infty}^2 + 2\sqrt{n} \log n \left\| \boldsymbol{\Delta}^t \right\|_{2,\infty} \left\| \boldsymbol{X}^\natural \right\| + C\sigma \sqrt{\frac{n}{p}} \right) \left\| \boldsymbol{\Delta}^t \right\| \qquad (140)
\end{aligned}
$$

where the last line follows from Lemma 39 and Lemma 46. As a consequence,

$$
\begin{aligned}
\left\| \boldsymbol{X}_1 - \boldsymbol{X}_2 \right\| \left\| \boldsymbol{X}^\natural \right\| &= \left\| \boldsymbol{X}^{t+1} \hat{\boldsymbol{H}}^t - \tilde{\boldsymbol{X}}^{t+1} \right\| \left\| \boldsymbol{X}^\natural \right\| \\
&\le \eta \left(2n \left\| \boldsymbol{\Delta}^t \right\|_{2,\infty}^2 + 2\sqrt{n} \log n \left\| \boldsymbol{\Delta}^t \right\|_{2,\infty} \left\| \boldsymbol{X}^\natural \right\| + C\sigma \sqrt{\frac{n}{p}} \right) \left\| \boldsymbol{\Delta}^t \right\| \left\| \boldsymbol{X}^\natural \right\|
\end{aligned}
$$

Under our sample size condition and the noise condition assumed in this lemma, we have

$$
\left\| \boldsymbol{X}_1 - \boldsymbol{X}_2 \right\| \left\| \boldsymbol{X}^\natural \right\| \le \frac{1}{4} \sigma_{\min}.
$$

Apply Lemma 41 and (140) to reach

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

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

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

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

Proof of Lemma 21. Regarding ϕ_1, we will first show the concentration bound for any $\boldsymbol{\Delta}^t$ independent of Ω, and then invoke the standard covering argument to extend it to all $\boldsymbol{\Delta}^t$. For any $\boldsymbol{\Delta}^t$ independent of Ω, one has

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

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

$$
\begin{aligned}
&\leq p\left\|\boldsymbol{X}_{l,\cdot}^{\natural}\right\|_{2}^{2}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}^{2}\psi^{2}\\
&\leq 2p\left\|\boldsymbol{X}_{l,\cdot}^{\natural}\right\|_{2}^{2}\xi^{2}\sigma_{\max},
\end{aligned}
$$

where ξ and ψ are defined respectively in (131) and (132). Here, the last line makes use of the fact that

$$
\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\psi \ll \xi\left\|\boldsymbol{X}^{\natural}\right\| = \xi\sqrt{\sigma_{\max}}. \tag{141}
$$

Apply the matrix Bernstein inequality [Tro15b, Theorem 6.1.1] to get

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

for some constant $c>0$, provided that

$$
t \leq 2p\sigma_{\max}\frac{\xi}{\|\boldsymbol{X}^{\natural}\|_{2,\infty}}\|\boldsymbol{X}_{l,\cdot}^{\natural}\|_{2}.
$$

This upper bound on t is exactly the truncation level ω we introduce in (133). 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 $O\left(p\xi^{2}\sigma_{\max}\|\boldsymbol{X}_{l,\cdot}^{\natural}\|_{2}^{2}\log r\right)$. Therefore, invoking the concentration bounds for quadratic functions [HKZ12, Theorem 1] yields that with probability at least $1-C_{0}e^{-Cnr\log n}$,

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

Now that we have established an upper bound on any fixed matrix $\boldsymbol{\Delta}^{t}$ (which holds with exponentially high probability), we can proceed to invoke the standard epsilon-net argument to establish a uniform bound over all feasible $\boldsymbol{\Delta}^{t}$. This argument is fairly standard, and is thus omitted; see [Tao12, Section 2.3.1] or the proof of Lemma 40. In conclusion, we have that with probability exceeding $1-C_{0}e^{-\frac{1}{2}Cnr\log n}$,

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

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

Proof of Lemma 22. Notice that

$$
\begin{aligned}
\|\boldsymbol{A}_{l,\cdot}\|_{2} &\leq \left\|\boldsymbol{X}_{l,\cdot}^{\natural}\right\|_{2}\left\|\sum_{j=1}^{n}\left(\delta_{l,j}-p\right)\boldsymbol{\Delta}_{j,\cdot}^{t\top}\boldsymbol{X}_{j,\cdot}^{\natural}\right\| \qquad (142)\\
&\leq \left\|\boldsymbol{X}_{l,\cdot}^{\natural}\right\|_{2}\left(\left\|\sum_{j=1}^{n}\delta_{l,j}\boldsymbol{\Delta}_{j,\cdot}^{t\top}\boldsymbol{X}_{j,\cdot}^{\natural}\right\|+p\left\|\boldsymbol{\Delta}^{t}\right\|\left\|\boldsymbol{X}^{\natural}\right\|\right)\\
&\leq \left\|\boldsymbol{X}_{l,\cdot}^{\natural}\right\|_{2}\left(\left\|\left[\delta_{l,1}\boldsymbol{\Delta}_{1,\cdot}^{t\top},\cdots,\delta_{l,n}\boldsymbol{\Delta}_{n,\cdot}^{t\top}\right]\right\|\left\|\begin{bmatrix}\delta_{l,1}\boldsymbol{X}_{1,\cdot}^{\natural}\\ \vdots\\ \delta_{l,n}\boldsymbol{X}_{n,\cdot}^{\natural}\end{bmatrix}\right\|+p\psi\left\|\boldsymbol{X}^{\natural}\right\|\right)
\end{aligned}
$$

$$\overset{\text{(i)}}{\leq} \left\|\boldsymbol{X}^{\natural}_{l,\cdot}\right\|_2 \left(\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}^t\right)\right\| \cdot 1.2\sqrt{p}\left\|\boldsymbol{X}^{\natural}\right\| + p\psi\left\|\boldsymbol{X}^{\natural}\right\|\right), \tag{143}$$

where (i) follows from Lemma 38, namely,

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

under the sample complexity condition $np \gg \kappa\mu r\log n$. Hence the event

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

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

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

$$\begin{aligned}
\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}^t\right)\right\| &\geq \frac{\frac{2p\sigma_{\max}\frac{\xi}{\|\boldsymbol{X}^{\natural}\|_{2,\infty}}\|\boldsymbol{X}^{\natural}_{l,\cdot}\|_2}{\|\boldsymbol{X}^{\natural}\|\left\|\boldsymbol{X}^{\natural}_{l,\cdot}\right\|_2} - p\psi}{1.2\sqrt{p}} \geq \frac{\frac{2\sqrt{p}\left\|\boldsymbol{X}^{\natural}\right\|\xi}{\|\boldsymbol{X}^{\natural}\|_{2,\infty}} - \sqrt{p}\psi}{1.2} \\
&\geq 1.5\sqrt{p}\frac{\xi}{\|\boldsymbol{X}^{\natural}\|_{2,\infty}}\left\|\boldsymbol{X}^{\natural}\right\|,
\end{aligned}$$

where the last inequality follows from the bound (141). As a result, we can upper bound ϕ_2 by

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

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

Next, we divide into multiple cases based on the levels of $\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}^t\right)\|$. By Lemma 40, with probability at least $1 - c_1\exp\left(-c_2 nr\log n\right)$,

$$\sum_{l=1}^{n}\mathbb{1}_{\left\{\|\boldsymbol{G}_l(\boldsymbol{\Delta}^t)\|\geq 2\sqrt{p}\psi+\sqrt{2^k r}\xi\right\}} \leq \frac{\alpha n}{2^k} \tag{144}$$

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

$$\sqrt{2^k r}\xi \geq 2\sqrt{p}\psi \quad \text{or equivalently} \quad k \geq \log\frac{4p\psi^2}{r\xi^2}; \tag{145}$$

otherwise we can use (141) to obtain

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

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

$$S_k = \left\{1 \leq l \leq n \mid \left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}^t\right)\right\| \in \left(\sqrt{2^k r}\xi, \sqrt{2^{k+1} r}\xi\right]\right\} \tag{146}$$

defined for each integer k obeying (145). Under the condition (145), it follows from (144) that

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

meaning that the cardinality of S_k satisfies

$$|S_k| \leq \frac{\alpha n}{2^{k-2}}$$

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

$$\begin{aligned}
\sqrt{\sum_{l \in S_k} \|\boldsymbol{A}_{l,\cdot}\|_2^2} &\overset{\text{(i)}}{\leq} \sqrt{|S_k| \cdot \|\boldsymbol{X}^\natural\|_{2,\infty}^2 \left(1.2\sqrt{2^{k+1}r}\xi\sqrt{p}\,\|\boldsymbol{X}^\natural\| + p\psi\,\|\boldsymbol{X}^\natural\|\right)^2} \\
&\leq \sqrt{\frac{\alpha n}{2^{k-2}}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}\left(2\sqrt{2^{k+1}r}\xi\sqrt{p}\left\|\boldsymbol{X}^\natural\right\| + p\psi\left\|\boldsymbol{X}^\natural\right\|\right) \\
&\overset{\text{(ii)}}{\leq} 4\sqrt{\frac{\alpha n}{2^{k-2}}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}\sqrt{2^{k+1}r}\xi\sqrt{p}\left\|\boldsymbol{X}^\natural\right\| \\
&\leq 16\sqrt{\alpha\kappa\mu r^2 p}\xi\left\|\boldsymbol{X}^\natural\right\|^2,
\end{aligned}$$

where (i) results from the bound (143) and the constraint (146) in S_k, and (ii) is a consequence of (145).

Now that we have develop 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_0 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

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

if $k \geq \log n$. One can thus conclude that

$$\begin{aligned}
&\phi_2^2 \leq \sum_{k=\log\frac{4p\psi^2}{r\xi^2}}^{c_1 \log n} \sum_{l \in S_k} \|\boldsymbol{A}_{l,\cdot}\|_2^2 \lesssim \left(16\sqrt{\alpha\kappa\mu r^2 p}\xi\left\|\boldsymbol{X}^\natural\right\|^2\right)^2 \cdot \log n \\
\Longrightarrow \qquad &\phi_2 \lesssim \sqrt{\alpha\kappa\mu r^2 p \log n}\xi\left\|\boldsymbol{X}^\natural\right\|^2,
\end{aligned}$$

which finishes the proof by taking $\alpha = c \log n$ for some sufficiently large constant $c > 0$. $\square$

B.4 Proof of Lemma 10

The first consequence is concerned with the estimation error of $\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}$ with respect to the Frobenius norm. Combining (67a) , (67d) and the triangle inequality yields

$$\begin{aligned}
&\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^\natural\right\|_{\mathrm{F}} \leq \left\|\boldsymbol{X}^t\hat{\boldsymbol{H}}^t - \boldsymbol{X}^\natural\right\|_{\mathrm{F}} + \left\|\boldsymbol{X}^t\hat{\boldsymbol{H}}^t - \boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\|_{\mathrm{F}} \\
&\quad\leq C_4\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}} + \frac{C_1\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}} + C_3\rho^t\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} + \frac{C_7\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} \\
&\quad\leq C_4\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}} + \frac{C_1\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}} + C_3\rho^t\mu r\sqrt{\frac{\log n}{np}}\sqrt{\frac{\kappa\mu}{n}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}} + \frac{C_7\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\sqrt{\frac{\kappa\mu}{n}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}} \\
&\quad\leq 2C_4\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}} + \frac{2C_1\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{\mathrm{F}},
\end{aligned} \tag{147}$$

where the last step holds true as long as $n \gg \kappa\mu \log n$. The first inequality in (70a) follows directly from the definition of $\hat{\boldsymbol{H}}^{t,(l)}$.

With regard to the estimation error of $\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}$ when measured in $\ell_{2,\infty}$ norm, we use (67d) and (67b) to get

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

Regarding the error in the operator norm, one can take the triangle inequality

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

where the second line follows from (74). Combine (67d) and (67c) to yield

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

where the second inequality uses the incoherence of $\boldsymbol{X}^\natural$ (cf. (114)) and the last one holds as long as

$$n \gg \kappa^3\mu r\log n.$$

B.5 Proof of Lemma 11

From the definition of $\boldsymbol{R}^{t+1,(l)}$ (cf. (69)), we must have

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

The gradient update rule allows one to express

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

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 rotation matrix $\boldsymbol{R} \in \mathcal{O}^{r\times r}$ (similarly for $\nabla f^{(l)}\left(\boldsymbol{X}\right)$). Relate the right-hand side of the above equation with $\nabla f_{\mathrm{clean}}\left(\boldsymbol{X}\right)$ as follows

$$\boldsymbol{X}^{t+1}\hat{\boldsymbol{H}}^t - \boldsymbol{X}^{t+1,(l)}\boldsymbol{R}^{t,(l)} = \underbrace{\big(\boldsymbol{X}^t\hat{\boldsymbol{H}}^t - \boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\big) - \eta\left[\nabla f_{\mathrm{clean}}\big(\boldsymbol{X}^t\hat{\boldsymbol{H}}^t\big) - \nabla f_{\mathrm{clean}}\big(\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\big)\right]}_{:=\boldsymbol{\nu}_1^{(l)}}$$

$$
\underbrace{-\eta\left[\frac{1}{p}\mathcal{P}_{\Omega_l}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top}-\boldsymbol{M}^{\natural}\right)-\mathcal{P}_l\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top}-\boldsymbol{M}^{\natural}\right)\right]\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}}_{:=\boldsymbol{\nu}_2^{(l)}}
$$

$$
\underbrace{+\eta\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{E}\right)\left(\boldsymbol{X}^t\hat{\boldsymbol{H}}^t-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right)}_{:=\boldsymbol{\nu}_3^{(l)}}\underbrace{+\eta\frac{1}{p}\mathcal{P}_{\Omega_l}\left(\boldsymbol{E}\right)\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}}_{:=\boldsymbol{\nu}_4^{(l)}}, \tag{148}
$$

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

$$
\nabla f^{(l)}\left(\boldsymbol{X}\right)=\nabla f\left(\boldsymbol{X}\right)-\frac{1}{p}\mathcal{P}_{\Omega_l}\left[\boldsymbol{X}\boldsymbol{X}^{\top}-\left(\boldsymbol{M}^{\natural}+\boldsymbol{E}\right)\right]\boldsymbol{X}+\mathcal{P}_l\left(\boldsymbol{X}\boldsymbol{X}^{\top}-\boldsymbol{M}^{\natural}\right)\boldsymbol{X} \tag{149}
$$

for all $\boldsymbol{X}\in\mathbb{R}^{n\times r}$ with $\mathcal{P}_{\Omega_l}$ and $\mathcal{P}_l$ defined respectively in (63) and (64). In the sequel, we control the four terms in the reverse order.

1. The last term $\boldsymbol{\nu}_4^{(l)}$ is controlled via the following lemma.

   ***Lemma** 23. Suppose that the sample size obeys $np > C\log n$ for some sufficiently large constant $C>0$. Then with probability $1-O(n^{-10})$, the vector $\boldsymbol{\nu}_4^{(l)}$ as defined in (148) satisfies*

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

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

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

   where the second inequality comes from Lemma 46.

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

   ***Lemma** 24. Suppose that the sample size obeys $n^2p\gg\mu^2r^2\log n$. Then with probability $1-O(n^{-10})$, the vector $\boldsymbol{\nu}_2^{(l)}$ as defined in (148) satisfies*

   $$
   \left\|\boldsymbol{\nu}_2^{(l)}\right\|_{\mathrm{F}}\leq\eta\sqrt{\frac{\kappa^2\mu^2r^2\log n}{np}}\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}-\boldsymbol{X}^{\natural}\right\|_{2,\infty}\sigma_{\max}. \tag{150}
   $$

4. Regarding the first term $\boldsymbol{\nu}_1^{(l)}$, apply the mean value theorem for vector-valued functions [Lan93, Theorem 4.2] to get

   $$
   \mathrm{vec}\big(\boldsymbol{\nu}_1^{(l)}\big)=\left(\boldsymbol{I}_{nr}-\eta\int_0^1\nabla^2 f_{\mathrm{clean}}\left(\boldsymbol{X}(\tau)\right)\mathrm{d}\tau\right)\mathrm{vec}\left(\boldsymbol{X}^t\hat{\boldsymbol{H}}^t-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right), \tag{151}
   $$

   where we abuse the notation and denote $\boldsymbol{X}(\tau):=\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}+\tau\left(\boldsymbol{X}^t\hat{\boldsymbol{H}}^t-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right)$. Going through the same derivations as in the proof of Lemma 8 (Appendix B.2), we get

   $$
   \|\boldsymbol{\nu}_1^{(l)}\|_{\mathrm{F}}\leq\left(1-\frac{\sigma_{\min}}{4}\eta\right)\left\|\boldsymbol{X}^t\hat{\boldsymbol{H}}^t-\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\|_{\mathrm{F}} \tag{152}
   $$

   with the proviso that $\eta\leq\frac{2\sigma_{\min}}{25\sigma_{\max}^2}$.

Putting the preceding four bounds together and using (148), we arrive at

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

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

for some absolute constant $\tilde{C}>0$, where (i) holds as long as $\sigma\sqrt{\frac{n}{p}}\ll\sigma_{\min}$. This taken collectively with the hypotheses (67d) and (70b) leads to

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

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

B.6 Proof of Lemma 12

We first introduce an auxiliary matrix

$$
\tilde{\boldsymbol{X}}^{t+1,(l)}:=\boldsymbol{X}^{t,(l)}\hat{\boldsymbol{H}}^{t,(l)}-\eta\left[\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left[\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top}-\left(\boldsymbol{M}^\natural+\boldsymbol{E}\right)\right]+\mathcal{P}_l\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top}-\boldsymbol{M}^\natural\right)\right]\boldsymbol{X}^\natural. \tag{153}
$$

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

$$
\left\|\left(\boldsymbol{X}^{t+1,(l)}\hat{\boldsymbol{H}}^{t+1,(l)}-\boldsymbol{X}^\natural\right)_{l,\cdot}\right\|_2\le\underbrace{\left\|\left(\boldsymbol{X}^{t+1,(l)}\hat{\boldsymbol{H}}^{t+1,(l)}-\tilde{\boldsymbol{X}}^{t+1,(l)}\right)_{l,\cdot}\right\|_2}_{:=\alpha_1}+\underbrace{\left\|\left(\tilde{\boldsymbol{X}}^{t+1,(l)}-\boldsymbol{X}^\natural\right)_{l,\cdot}\right\|_2}_{:=\alpha_2}. \tag{154}
$$

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

1. Regarding the second term α_2 of (154), we see from the definition of $\tilde{\boldsymbol{X}}^{t+1,(l)}$ (cf. (153)) that

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

where we utilize the definition of $\mathcal{P}_{\Omega^{-l}}$ and $\mathcal{P}_l$ in (64). This further gives

$$\left(\tilde{\boldsymbol{X}}^{t+1,(l)}-\boldsymbol{X}^\natural\right)_{l,\cdot}=\left[\boldsymbol{X}^{t,(l)}\hat{\boldsymbol{H}}^{t,(l)}-\eta\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top}-\boldsymbol{X}^\natural\boldsymbol{X}^{\natural\top}\right)\boldsymbol{X}^\natural-\boldsymbol{X}^\natural\right]_{l,\cdot}. \tag{155}$$

For notational convenience, we denote

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

This allows us to rewrite (155) as

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

which further implies that

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

Here, the last line follows from the induction hypothesis (67e) and our assumptions on the sample size and the noise, namely,

$$\left\|\boldsymbol{\Delta}^{t,(l)}_{l,\cdot}\right\|_2\le C_2\rho^t\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}+C_6\frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}\ll\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} \tag{156}$$

as long as $np\gg\mu^2r^2$ and $\sigma\sqrt{\frac{n\log n}{p}}\ll\sigma_{\min}$. By taking $\eta\le\frac{1}{\sigma_{\max}}$, we can obtain

$$\alpha_2\le\left(1-\eta\sigma_{\min}\right)\left\|\boldsymbol{\Delta}^{t,(l)}_{l,\cdot}\right\|_2+2\eta\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\left\|\boldsymbol{X}^\natural\right\|. \tag{157}$$

An immediate consequence of the above two inequalities is

$$\alpha_2\le\|\boldsymbol{X}^\natural\|_{2,\infty}. \tag{158}$$

2. The first term α_1 of (154) can be equivalently written as

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

where

$$\begin{aligned}&\boldsymbol{R}_1=\left(\hat{\boldsymbol{H}}^{t,(l)}\right)^{-1}\hat{\boldsymbol{H}}^{t+1,(l)}:=\arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\left\|\boldsymbol{X}^{t+1,(l)}\hat{\boldsymbol{H}}^{t,(l)}\boldsymbol{R}-\boldsymbol{X}^\natural\right\|_{\mathrm{F}},\\\text{and}\qquad&\boldsymbol{R}_2=\boldsymbol{I}_r:=\arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\left\|\tilde{\boldsymbol{X}}^{t+1,(l)}\boldsymbol{R}-\boldsymbol{X}^\natural\right\|_{\mathrm{F}}.\end{aligned} \tag{159}$$

We will explain (159) shortly. Simple algebra yields

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

Here, the second relation uses

$$\left\|\tilde{\boldsymbol{X}}^{t+1,(l)}_{l,\cdot}\right\|_2\le\left\|\tilde{\boldsymbol{X}}^{t+1,(l)}_{l,\cdot}-\boldsymbol{X}^\natural_{l,\cdot}\right\|_2+\left\|\boldsymbol{X}^\natural_{l,\cdot}\right\|_2=\alpha_2+\left\|\boldsymbol{X}^\natural_{l,\cdot}\right\|_2\le2\left\|\boldsymbol{X}^\natural\right\|_{2,\infty},$$

where the last inequality follows from (158). It remains to upper bound β_1 and β_2.

(a) For both cases, the central quantity to control is $\boldsymbol{X}^{t+1,(l)}\hat{\boldsymbol{H}}^{t,(l)} - \tilde{\boldsymbol{X}}^{t+1,(l)}$. By the definition of $\tilde{\boldsymbol{X}}^{t+1,(l)}$ in (153) and the gradient update rule for $\boldsymbol{X}^{t+1,(l)}$, one has

$$\begin{aligned}
&\boldsymbol{X}^{t+1,(l)}\hat{\boldsymbol{H}}^{t,(l)} - \tilde{\boldsymbol{X}}^{t+1,(l)} \\
&= \left\{ \boldsymbol{X}^{t,(l)}\hat{\boldsymbol{H}}^{t,(l)} - \eta \left[\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left[\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \left(\boldsymbol{M}^{\natural} + \boldsymbol{E}\right)\right] + \mathcal{P}_l\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{M}^{\natural}\right)\right] \boldsymbol{X}^{t,(l)}\hat{\boldsymbol{H}}^{t,(l)} \right\} \\
&\quad - \left\{ \boldsymbol{X}^{t,(l)}\hat{\boldsymbol{H}}^{t,(l)} - \eta \left[\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left[\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \left(\boldsymbol{M}^{\natural} + \boldsymbol{E}\right)\right] + \mathcal{P}_l\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{M}^{\natural}\right)\right] \boldsymbol{X}^{\natural} \right\} \\
&= -\eta \left[\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right) + \mathcal{P}_l\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\right] \boldsymbol{\Delta}^{t,(l)} \\
&\quad + \eta\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left(\boldsymbol{E}\right)\boldsymbol{\Delta}^{t,(l)}. \qquad (160)
\end{aligned}$$

It is easy to verify that

$$\left\|\frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left(\boldsymbol{E}\right)\right\| \le \left\|\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{E}\right)\right\| \lesssim \sigma\sqrt{\frac{n}{p}},$$

where the last relation uses Lemma 46. In order to control (160), we need to upper bound the following quantity

$$\gamma := \left\| \frac{1}{p}\mathcal{P}_{\Omega^{-l}}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right) + \mathcal{P}_l\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\right\|. \qquad (161)$$

To this end, we make the observation that

$$\begin{aligned}
\gamma &\le \underbrace{\left\| \frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\right\|}_{:=\gamma_1} \\
&\quad + \underbrace{\left\| \frac{1}{p}\mathcal{P}_{\Omega_l}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right) - \mathcal{P}_l\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\right\|}_{:=\gamma_2}. \qquad (162)
\end{aligned}$$

An application of Lemma 39 reveals that

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

Letting $\boldsymbol{C} = \boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}$, one can bound the other term γ_2 by taking advantage of the triangle inequality and the symmetry property:

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

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 following inequality

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

with the proviso that $\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}-\boldsymbol{X}^{\natural}\right\|_{2,\infty}\leq\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}$. The previous two bounds taken collectively give

$$\gamma\leq 2n\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}-\boldsymbol{X}^{\natural}\right\|_{2,\infty}^{2}+2\sqrt{n}\log n\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}-\boldsymbol{X}^{\natural}\right\|_{2,\infty}\left\|\boldsymbol{X}^{\natural}\right\| \\ +\tilde{C}\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}-\boldsymbol{X}^{\natural}\right\|_{2,\infty}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} \tag{163}$$

for some constant $\tilde{C}>0$. In summary, we obtain

$$\begin{aligned}\left\|\boldsymbol{X}^{t+1,(l)}\hat{\boldsymbol{H}}^{t,(l)}-\tilde{\boldsymbol{X}}^{t+1,(l)}\right\| &\leq\eta\left(\gamma+\left\|\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{E}\right)\right\|\right)\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \\ &\leq\eta\delta\sigma_{\min}\left\|\boldsymbol{\Delta}^{t,(l)}\right\|,\end{aligned} \tag{164}$$

for δ sufficiently small. With this in place, we can continue our derivation on β_1 and β_2.

(b) With regard to β_1, in view of (160) we can obtain

$$\begin{aligned}\beta_1 &\overset{\text{(i)}}{=}\eta\left\|\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top}-\boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)_{l,\cdot}\boldsymbol{\Delta}^{t,(l)}\right\|_2 \\ &\leq\eta\left\|\left(\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top}-\boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)_{l,\cdot}\right\|_2\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \\ &\overset{\text{(ii)}}{=}\eta\left\|\left(\boldsymbol{\Delta}^{t,(l)}\boldsymbol{X}^{t,(l)\top}+\boldsymbol{X}^{\natural}\boldsymbol{\Delta}^{t,(l)\top}\right)_{l,\cdot}\right\|_2\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \\ &\leq\eta\left(\left\|\boldsymbol{\Delta}_{l,\cdot}^{t,(l)}\right\|_2\left\|\boldsymbol{X}^{t,(l)}\right\|+\left\|\boldsymbol{X}_{l,\cdot}^{\natural}\right\|_2\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\right)\left\|\boldsymbol{\Delta}^{t,(l)}\right\| \\ &\leq\eta\left\|\boldsymbol{\Delta}_{l,\cdot}^{t,(l)}\right\|_2\left\|\boldsymbol{X}^{t,(l)}\right\|\left\|\boldsymbol{\Delta}^{t,(l)}\right\|+\eta\left\|\boldsymbol{X}_{l,\cdot}^{\natural}\right\|_2\left\|\boldsymbol{\Delta}^{t,(l)}\right\|^2,\end{aligned} \tag{165}$$

where (i) follows from the definition of $\mathcal{P}_{\Omega^{-l}}$ and $\mathcal{P}_l$ (note that all entries in the lth row of $\mathcal{P}_{\Omega^{-l}}(\boldsymbol{E})$ are zero), and (ii) arises since

$$\begin{aligned}\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top}-\boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}&=\boldsymbol{X}^{t,(l)}\boldsymbol{X}^{t,(l)\top}-\boldsymbol{X}^{\natural}\boldsymbol{X}^{t,(l)\top}+\boldsymbol{X}^{\natural}\boldsymbol{X}^{t,(l)\top}-\boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top} \\ &=\big(\boldsymbol{X}^{t,(l)}-\boldsymbol{X}^{\natural}\big)\boldsymbol{X}^{t,(l)\top}+\boldsymbol{X}^{\natural}\big(\boldsymbol{X}^{t,(l)}-\boldsymbol{X}^{\natural}\big).\end{aligned}$$

(c) For β_2, letting $\boldsymbol{A}:=\big(\tilde{\boldsymbol{X}}^{t+1,(l)}\big)^{\top}\boldsymbol{X}^{\natural}$ and $\boldsymbol{E}=\big(\boldsymbol{X}^{t+1,(l)}\hat{\boldsymbol{H}}^{t,(l)}-\tilde{\boldsymbol{X}}^{t+1,(l)}\big)^{\top}\boldsymbol{X}^{\natural}$ in Lemma 47, we claim that

$$\begin{aligned}\sigma_r\left(\boldsymbol{A}\right)&\geq\frac{1}{2}\sigma_{\min}, \\ \left\|\boldsymbol{E}\right\|&\leq\frac{1}{2}\sigma_{\min}\leq\sigma_r\left(\boldsymbol{A}\right).\end{aligned}$$

The first inequality would hold as long as

$$\tilde{\boldsymbol{X}}^{t+1,(l)\top}\boldsymbol{X}^{\natural}\succeq\frac{1}{2\sigma_{\min}}\boldsymbol{I}_r. \tag{166}$$

The second inequality arises from (164), namely,

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

where (i) holds since $\left\|\boldsymbol{\Delta}^{t,(l)}\right\|\leq\left\|\boldsymbol{X}^{\natural}\right\|$ and (ii) holds true for δ sufficiently small and $\eta\leq\frac{1}{\sigma_{\max}}$. Invoke Lemma 47 to obtain

$$\beta_2=\left\|\boldsymbol{R}_1-\boldsymbol{R}_2\right\|\leq\frac{2}{\sigma_{r-1}\left(\boldsymbol{A}\right)+\sigma_r\left(\boldsymbol{A}\right)}\left\|\boldsymbol{E}\right\|$$

$$\leq \frac{2}{\sigma_{\min}} \left\| \boldsymbol{X}^{t+1,(l)} \hat{\boldsymbol{H}}^{t,(l)} - \tilde{\boldsymbol{X}}^{t+1,(l)} \right\| \left\| \boldsymbol{X}^{\natural} \right\| \tag{167}$$

$$\leq 2\delta\eta \left\| \boldsymbol{\Delta}^{t,(l)} \right\| \left\| \boldsymbol{X}^{\natural} \right\|, \tag{168}$$

where (167) makes use of the fact (166) (and hence $\sigma_{r-1}(\boldsymbol{A}) \geq \sigma_r(\boldsymbol{A}) \geq \frac{1}{2}\sigma_{\min}$), and the last line comes from (164). We will justify (166) shortly.

(d) Putting the previous bounds (165) and (168) together yields

$$\alpha_1 \leq \eta \left\| \boldsymbol{\Delta}_{l,\cdot}^{t,(l)} \right\|_2 \left\| \boldsymbol{X}^{t,(l)} \right\| \left\| \boldsymbol{\Delta}^{t,(l)} \right\| + \eta \left\| \boldsymbol{X}_{l,\cdot}^{\natural} \right\|_2 \left\| \boldsymbol{\Delta}^{t,(l)} \right\|^2 + 4\delta\eta \left\| \boldsymbol{X}^{\natural} \right\|_{2,\infty} \left\| \boldsymbol{\Delta}^{t,(l)} \right\| \left\| \boldsymbol{X}^{\natural} \right\|. \tag{169}$$

(e) We still need to prove the claims (159) and (166). The definitions (153) and (161) together with the triangle inequality give

$$\begin{aligned} \left\| \tilde{\boldsymbol{X}}^{t+1,(l)} - \boldsymbol{X}^{\natural} \right\| \left\| \boldsymbol{X}^{\natural} \right\| &\leq \left\{ \left\| \boldsymbol{\Delta}^{t,(l)} \right\| + \eta\gamma \left\| \boldsymbol{X}^{\natural} \right\| + \eta \left\| \frac{1}{p} \mathcal{P}_{\Omega}(\boldsymbol{E}) \right\| \left\| \boldsymbol{X}^{\natural} \right\| \right\} \left\| \boldsymbol{X}^{\natural} \right\| \\ &\leq \left\| \boldsymbol{\Delta}^{t,(l)} \right\| \left\| \boldsymbol{X}^{\natural} \right\| + \eta\gamma\sigma_{\max} + \eta C\sigma \sqrt{\frac{n}{p}} \sigma_{\max} \\ &\ll \sigma_{\min}, \end{aligned}$$

where the second inequality uses Lemma 46, and the last line follows from $\eta \leq \frac{1}{\sigma_{\max}}$, the inequality (164), the noise condition $\sigma \ll \sqrt{\frac{n}{p}}\sigma_{\min}$, and the fact that $\left\| \boldsymbol{\Delta}^{t,(l)} \right\| \left\| \boldsymbol{X}^{\natural} \right\| \ll \sigma_{\min}$, as well as the assumption $\gamma \ll 1$. In particular, the upper bound on γ can be seen from (163) together with our induction hypotheses and sample complexity condition. These in turn imply that

$$\left\| \tilde{\boldsymbol{X}}^{t+1,(l)\top} \boldsymbol{X}^{\natural} - \boldsymbol{\Sigma}^{\natural} \right\| \leq \left\| \tilde{\boldsymbol{X}}^{t+1,(l)} - \boldsymbol{X}^{\natural} \right\| \left\| \boldsymbol{X}^{\natural} \right\| \ll \sigma_{\min}. \tag{170}$$

Moreover, Lemma 48 reveals that $\boldsymbol{X}^{\natural\top} \boldsymbol{X}^{t,(l)} \hat{\boldsymbol{H}}^{t,(l)}$ is symmetric, and hence it is easily seen from the definition of $\tilde{\boldsymbol{X}}^{(t+1),l}$ that $\boldsymbol{X}^{\natural\top} \tilde{\boldsymbol{X}}^{(t+1),l}$ is symmetric. This symmetry property combined with (170) indicates that

$$\tilde{\boldsymbol{X}}^{t+1,(l)\top} \boldsymbol{X}^{\natural} \succeq \frac{1}{2\sigma_{\min}} \boldsymbol{I}_r, \tag{171}$$

which justifies (166). Finally, invoke Lemma 48 to prove the claim (159).

3. Combine (154), (157) and (169) to reach

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

Here, (i) follows since $\left\|\mathbf{\Delta}^{t,(l)}\right\| \leq \left\|\boldsymbol{X}^{\natural}\right\|$ and δ is sufficiently small, (ii) invokes the hypotheses (67e) and (70c) and sees that

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

holds under the sample size and noise condition, while (iii) is valid as long as $C_2 \gg \kappa C_9$, $\rho \geq 1 - \frac{\sigma_{\min}}{3}\eta$ and $C_6 \gg \kappa C_{10}$.

Proof of Lemma 23. By the unitary invariance of the Frobenius norm, one has

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

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

$$\frac{p}{\eta}\left\|\boldsymbol{\nu}_4^{(l)}\right\|_{\mathrm{F}} \leq \left\|\sum_{j=1}^{n}\underbrace{\delta_{l,j}E_{l,j}\boldsymbol{X}_{j,\cdot}^{t,(l)}}_{:=\boldsymbol{u}_j}\right\|_2 + \underbrace{\left\|\sum_{j:j\neq l}\delta_{l,j}E_{l,j}\boldsymbol{X}_{l,\cdot}^{t,(l)}\right\|_2}_{:=\alpha}, \tag{172}$$

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 $\boldsymbol{\delta}_{l,\cdot}$ and $\boldsymbol{E}_{l,\cdot}$, we can treat the first term as a sum of independent vectors. It is easy to verify that

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

where $\|\cdot\|_{\psi_1}$ denotes the sub-exponential norm [KLT11, Section 6]. Further, one can calculate

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

The matrix Bernstein inequality [KLT11, Proposition 2] then tells us that with probability at least $1 - O\left(n^{-10}\right)$,

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

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

Additionally, the remaining term α in (172) can be controlled using the same argument, given that

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

We then complete the proof by observing that

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

□

Proof of Lemma 24. For notational simplicity, we denote

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

Since the Frobenius norm is unitarily invariant, we have

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

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

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

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

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

Here, (i) is a consequence of the following inequality

$$\left\|\boldsymbol{X}^{t,(l)}\right\|_{2,\infty} = \left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\|_{2,\infty} \le \left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^{\natural}\right\|_{2,\infty} + \left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} \le 2\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}, \tag{173}$$

where the last inequality follows by combining (70a), the sample complexity condition $n^2p \gg \mu^2r^2\log n$, and the assumption on the noise. In addition, (ii) follows from

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

where the last inequality comes from (70a), the sample complexity and the noise assumptions. The matrix Bernstein inequality [Tro15b, Theorem 6.1.1] reveals that

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

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

$$\begin{aligned}\frac{p}{\eta}\left\|\boldsymbol{\nu}_2^{(l)}\right\|_{\mathrm{F}} &\lesssim \left\|\sum_{j=1}^{n}\left(\delta_{l,j}-p\right)C_{l,j}\boldsymbol{X}_{j,\cdot}^{t,(l)}\right\|_2 + \sqrt{np}\left\|\boldsymbol{C}\right\|_\infty\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\\ &\lesssim \sqrt{p\log n}\left\|\boldsymbol{C}\right\|_\infty\left\|\boldsymbol{X}^{\natural}\right\|_{\mathrm{F}} + \sqrt{np}\left\|\boldsymbol{C}\right\|_\infty\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\end{aligned} \tag{174}$$

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

To finish up, we make the observation that

$$
\begin{aligned}
\|\boldsymbol{C}\|_\infty &= \left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\left(\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right)^\top - \boldsymbol{X}^\natural\boldsymbol{X}^{\natural\top}\right\|_\infty \\
&\le \left\|\left(\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^\natural\right)\left(\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right)^\top\right\|_\infty + \left\|\boldsymbol{X}^\natural\left(\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^\natural\right)^\top - \boldsymbol{X}^\natural\boldsymbol{X}^{\natural\top}\right\|_\infty \\
&\le \left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^\natural\right\|_{2,\infty}\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)}\right\|_{2,\infty} + \left\|\boldsymbol{X}^\natural\right\|_{2,\infty}\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^\natural\right\|_{2,\infty} \\
&\le 3\left\|\boldsymbol{X}^{t,(l)}\boldsymbol{R}^{t,(l)} - \boldsymbol{X}^\natural\right\|_{2,\infty}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}, \qquad (175)
\end{aligned}
$$

where the last line arises from (173). This combined with (174) gives

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

where (i) comes from (175), and (ii) makes use of the incoherence condition (114).

□

B.7 Proof of Lemma 13

For notational convenience, we define the following two rotation matrices in this section

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

Before proceeding, we make note of the following two perturbation bounds for $\boldsymbol{M}^0$ and $\boldsymbol{M}^{(l)}$ as defined in Algorithm 2 and Algorithm 5, respectively, that hold under our sample complexity and noise conditions:

$$
\begin{aligned}
\left\|\boldsymbol{M}^0 - \boldsymbol{M}^\natural\right\| &\le \left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{M}^\natural\right) - \boldsymbol{M}^\natural\right\| + \left\|\frac{1}{p}\mathcal{P}_\Omega\left(\boldsymbol{E}\right)\right\| \lesssim \left\{\mu r\sqrt{\frac{1}{np}}\sqrt{\sigma_{\max}} + \frac{\sigma}{\sqrt{\sigma_{\min}}}\sqrt{\frac{n}{p}}\right\}\left\|\boldsymbol{X}^\natural\right\| \\
&\ll \sigma_{\min}, \qquad (176)
\end{aligned}
$$

where the first line arises from the triangle inequality, Lemma 45, Lemma 46 and the incoherence assumption (114). Similarly, we have

$$
\left\|\boldsymbol{M}^{(l)} - \boldsymbol{M}^\natural\right\| \lesssim \left\{\mu r\sqrt{\frac{1}{np}}\sqrt{\sigma_{\max}} + \frac{\sigma}{\sqrt{\sigma_{\min}}}\sqrt{\frac{n}{p}}\right\}\left\|\boldsymbol{X}^\natural\right\| \ll \sigma_{\min}. \qquad (177)
$$

These two bounds are immediate consequences of Lemma 45, Lemma 46 as well as the incoherence condition on $\boldsymbol{M}^\natural$.

We start by proving (67a), (67b) and (67c). The central decomposition we need is the following

$$
\boldsymbol{X}^0\hat{\boldsymbol{H}}^0 - \boldsymbol{X}^\natural = \boldsymbol{U}^0\left(\boldsymbol{\Sigma}^0\right)^{1/2}\left(\hat{\boldsymbol{H}}^0 - \boldsymbol{Q}\right) + \boldsymbol{U}^0\left[\left(\boldsymbol{\Sigma}^0\right)^{1/2}\boldsymbol{Q} - \boldsymbol{Q}\left(\boldsymbol{\Sigma}^\natural\right)^{1/2}\right] + \left(\boldsymbol{U}^0\boldsymbol{Q} - \boldsymbol{U}^\natural\right)\left(\boldsymbol{\Sigma}^\natural\right)^{1/2}. \qquad (178)
$$

1. For the spectral norm error bound in (67c), the triangle inequality together with (178) and the fact $\|\boldsymbol{U}^0\| = 1$ yields

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

Note that $\left\|\left(\boldsymbol{\Sigma}^0\right)^{1/2}\right\| \le 2\sqrt{\sigma_{\max}}$, and, from our assumptions, $\sigma_{\max}/\sigma_{\min} \lesssim 1$. Recognizing that $\left\|\boldsymbol{M}^0 - \boldsymbol{M}^\natural\right\| \ll \sigma_{\min}$ (cf. (176)), we can apply Lemma 43, Lemma 42 and Lemma 49 to yield

$$\left\|\hat{\boldsymbol{H}}^0 - \boldsymbol{Q}\right\| \lesssim \frac{1}{\sigma_{\min}}\left\|\boldsymbol{M}^0 - \boldsymbol{M}^\natural\right\|, \tag{179}$$

$$\left\|\left(\boldsymbol{\Sigma}^0\right)^{1/2}\boldsymbol{Q} - \boldsymbol{Q}\left(\boldsymbol{\Sigma}^\natural\right)^{1/2}\right\| \lesssim \frac{1}{\sqrt{\sigma_{\min}}}\left\|\boldsymbol{M}^0 - \boldsymbol{M}^\natural\right\|, \tag{180}$$

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

These taken collectively imply the advertised upper bound

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

2. Multiplying both sides of (67c) by $\sqrt{r}$ and observing that $\sigma_{\max}/\sigma_{\min} \asymp 1$ complete the proof of (67a).

3. The proof of (67b) follows from similar argument as in the first part. Combine (178) and the triangle inequality to reach

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

Plugging in the estimates (179), (180) results in

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

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

$$\left\|\boldsymbol{U}^0\boldsymbol{Q} - \boldsymbol{U}^\natural\right\|_{2,\infty} \lesssim \left(\mu r\sqrt{\frac{1}{np}} + \frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right)\left\|\boldsymbol{U}^\natural\right\|_{2,\infty} \qquad \text{and} \qquad \left\|\boldsymbol{U}^0\right\|_{2,\infty} \lesssim \left\|\boldsymbol{U}^\natural\right\|_{2,\infty}. \tag{181}$$

These combined with the previous inequality give

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

4. We now move on to proving (67e). Recall that $\boldsymbol{Q}^{(l)} = \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\left\|\boldsymbol{U}^{0,(l)}\boldsymbol{R} - \boldsymbol{U}^\natural\right\|_{\mathrm{F}}$. By the triangle inequality,

$$\begin{aligned}\left\|\left(\boldsymbol{X}^{0,(l)}\hat{\boldsymbol{H}}^{0,(l)} - \boldsymbol{X}^\natural\right)_{l,\cdot}\right\|_2 &\le \left\|\boldsymbol{X}^{0,(l)}_{l,\cdot}\left(\hat{\boldsymbol{H}}^{0,(l)} - \boldsymbol{Q}^{(l)}\right)\right\|_2 + \left\|\left(\boldsymbol{X}^{0,(l)}\boldsymbol{Q}^{(l)} - \boldsymbol{X}^\natural\right)_{l,\cdot}\right\|_2 \\ &\le \left\|\boldsymbol{X}^{0,(l)}_{l,\cdot}\right\|_2\left\|\hat{\boldsymbol{H}}^{0,(l)} - \boldsymbol{Q}^{(l)}\right\| + \left\|\left(\boldsymbol{X}^{0,(l)}\boldsymbol{Q}^{(l)} - \boldsymbol{X}^\natural\right)_{l,\cdot}\right\|_2.\end{aligned} \tag{182}$$

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

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

We can thus decompose

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

which further implies

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

In order to control this, we first see that

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

where the last inequality arises from Lemma 42 and the bound (177). Additionally, Lemma 49 gives

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

Putting the previous three bounds together and recognizing that $\left\|\boldsymbol{M}^{\natural}\right\|_{2,\infty} \leq \sqrt{\sigma_{\max}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}$, we reach

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

This also implies that $\left\|\boldsymbol{X}^{0,(l)}_{l,\cdot}\right\|_2 \leq 2\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}$. Substituting the bound back to (182) results in

$$\begin{aligned}\left\|\left(\boldsymbol{X}^{0,(l)}\hat{\boldsymbol{H}}^{0,(l)} - \boldsymbol{X}^{\natural}\right)_{l,\cdot}\right\|_2 &\lesssim \left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}\left\|\hat{\boldsymbol{H}}^{0,(l)} - \boldsymbol{Q}^{(l)}\right\| + \left\{\mu r\sqrt{\frac{1}{np}} + \frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right\}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty} \\ &\lesssim \left\{\mu r\sqrt{\frac{1}{np}} + \frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\right\}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty},\end{aligned}$$

where the second line relies on Lemma 43. This establishes (67e).

5. Our final goal is to justify (67d). 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 $\boldsymbol{R}^{0,(l)}$, one has

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

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 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 42 and the bounds (176) and (177), one has

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

and Davis-Kahan's sinΘ theorem [DK70] yields

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

These estimates taken collectively 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 $\left\|\left(\boldsymbol{M}^{0}-\boldsymbol{M}^{(l)}\right)\boldsymbol{U}^{0,(l)}\right\|_{\mathrm{F}}$. We have already encountered this type of quantity multiple times, and hence we omit the proof details for conciseness. The conclusion is this: 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

$$\left\|\boldsymbol{U}^{0,(l)}\right\|_{2,\infty}\lesssim\left\|\boldsymbol{U}^{\natural}\right\|_{2,\infty}\lesssim\frac{1}{\sqrt{\sigma_{\min}}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty},\tag{183}$$

then taking the previous bounds collectively established the desired bound

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

Proof of the claim (183). Denote by $\boldsymbol{M}^{(l),\mathrm{zero}}$ the matrix derived by zeroing out the lth row/column of $\boldsymbol{M}^{(l)}$, and $\boldsymbol{U}^{(l),\mathrm{zero}}$ the $n\times r$ matrix containing the leading r eigenvectors of $\boldsymbol{M}^{(l),\mathrm{zero}}$. On the one hand, [AFWZ17, Lemmas 4 and 14] demonstrates that

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

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

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

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 zero by construction. In addition,

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

The claim (183) then follows by combining the estimates above. □

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

$$\max_{1\leq l\leq m}\left|\boldsymbol{a}_l^*\boldsymbol{x}^\natural\right| \leq 5\sqrt{\log m} \tag{184}$$

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

$$\max_{1\leq l\leq m}\left\|\boldsymbol{a}_l\right\|_2 \leq 3\sqrt{K}. \tag{185}$$

For $c>0$, we also use $\lfloor c \rfloor$ to denote the largest integer that does not exceeds c.

C.1 Proof of Lemma 14

First, we find it convenient to decompose the Wirtinger Hessian (cf. (79)) into the population component (when evaluated at the truth) and the perturbation part as follows:

$$\nabla^2 f(\boldsymbol{z}) = \nabla^2 F\left(\boldsymbol{z}^\natural\right) + \left(\nabla^2 f(\boldsymbol{z}) - \nabla^2 F\left(\boldsymbol{z}^\natural\right)\right), \tag{186}$$

where the population Hessian (or the expected Hessian) at the truth $\boldsymbol{z}^\natural$ is given by

$$\nabla^2 F\left(\boldsymbol{z}^\natural\right) = \begin{bmatrix} \boldsymbol{I}_K & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{h}^\natural\boldsymbol{x}^{\natural\top} \\ \boldsymbol{0} & \boldsymbol{I}_K & \boldsymbol{x}^\natural\boldsymbol{h}^{\natural\top} & \boldsymbol{0} \\ \boldsymbol{0} & \left(\boldsymbol{x}^\natural\boldsymbol{h}^{\natural\top}\right)^* & \boldsymbol{I}_K & \boldsymbol{0} \\ \left(\boldsymbol{h}^\natural\boldsymbol{x}^{\natural\top}\right)^* & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_K \end{bmatrix}. \tag{187}$$

The proof then proceeds by showing that (i) the population Hessian $\nabla^2 F\left(\boldsymbol{z}^\natural\right)$ is well-conditioned, and that (ii) the perturbation $\nabla^2 f(\boldsymbol{z}) - \nabla^2 F\left(\boldsymbol{z}^\natural\right)$ is well-controlled under our assumptions.

We start be controlling the the population Hessian as follows.

***Lemma* 25.** *Instate the hypotheses and notation in Lemma 14. We have*

$$\left\|\nabla^2 F\left(\boldsymbol{z}^\natural\right)\right\| = 2 \qquad \text{and} \qquad \boldsymbol{u}^*\left[\boldsymbol{D}\nabla^2 F\left(\boldsymbol{z}^\natural\right) + \nabla^2 F\left(\boldsymbol{z}^\natural\right)\boldsymbol{D}\right]\boldsymbol{u} \geq \left\|\boldsymbol{u}\right\|_2^2.$$

It is clear from Lemma 25 that the population component satisfies the restricted strong convexity and smoothness conditions as advertised.

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

$$\begin{aligned}\mathcal{S} := \Bigg\{ &(\boldsymbol{h},\boldsymbol{x}) \in \mathbb{C}^K\times\mathbb{C}^K : \max\left\{\left\|\boldsymbol{h}-\boldsymbol{h}^\natural\right\|_2, \left\|\boldsymbol{x}-\boldsymbol{x}^\natural\right\|_2\right\} \leq \delta, \\ &\max_{1\leq j\leq m}\left|\boldsymbol{a}_j^*\left(\boldsymbol{x}-\boldsymbol{x}^\natural\right)\right| \leq 2C_3\frac{1}{\log^{1.5} m}, \ \max_{1\leq j\leq m}\left|\boldsymbol{b}_j^*\boldsymbol{h}\right| \leq 2C_4\frac{\mu}{\sqrt{m}}\log^2 m\Bigg\},\end{aligned} \tag{188}$$

where C_3 and C_4 are some positive constants. This corresponds to the set of $(\boldsymbol{h},\boldsymbol{x})$ satisfying the first two conditions in Lemma 14. Our perturbation bound is this:

***Lemma* 26.** *Let $\mathcal{S}$ be a set as defined in (188) for some sufficiently small constant $\delta>0$. With probability at least $1-O\left(m^{-10}\right)$, one has*

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

as long as $m \gg \mu^2 K\log^5 m$.

With the above two lemmas at hand, it is straightforward to see that

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

which verifies the smoothness upper bound. These can also be used to show that

$$\begin{aligned}
&\boldsymbol{u}^*\left[\boldsymbol{D}\nabla^2 f\left(\boldsymbol{z}\right) + \nabla^2 f\left(\boldsymbol{z}\right)\boldsymbol{D}\right]\boldsymbol{u} \\
&\quad \geq \boldsymbol{u}^*\left[\boldsymbol{D}\nabla^2 F\left(\boldsymbol{z}^\natural\right) + \nabla^2 F\left(\boldsymbol{z}^\natural\right)\boldsymbol{D}\right]\boldsymbol{u} - 2\left\|\boldsymbol{D}\right\|\left\|\nabla^2 f\left(\boldsymbol{z}\right) - \nabla^2 F\left(\boldsymbol{z}^\natural\right)\right\|\left\|\boldsymbol{u}\right\|_2^2 \\
&\quad \geq \left\|\boldsymbol{u}\right\|_2^2 - 2\left(1+\delta\right)\cdot\frac{1}{4}\left\|\boldsymbol{u}\right\|_2^2 \\
&\quad \geq \frac{1}{4}\left\|\boldsymbol{u}\right\|_2^2
\end{aligned}$$

for $\delta \leq 1/2$, proving the claim on restricted strong convexity.

It then suffices to establish Lemma 25 and Lemma 26.

Proof of Lemma 25. We start with proving the identity $\left\|\nabla^2 F\left(\boldsymbol{z}^\natural\right)\right\| = 2$. Let

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

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

$$\nabla^2 F\left(\boldsymbol{z}^\natural\right) = \boldsymbol{I}_{4K} + \boldsymbol{u}_1\boldsymbol{u}_1^* + \boldsymbol{u}_2\boldsymbol{u}_2^* - \boldsymbol{u}_3\boldsymbol{u}_3^* - \boldsymbol{u}_4\boldsymbol{u}_4^*,$$

which immediately implies

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

We now switch attention to the quadratic lower bound. The focus will be on the term $\boldsymbol{u}^*\boldsymbol{D}\nabla^2 F\left(\boldsymbol{z}^\natural\right)\boldsymbol{u}$, since the other term $\boldsymbol{u}^*\nabla^2 F\left(\boldsymbol{z}^\natural\right)\boldsymbol{D}\boldsymbol{u}$ is simply the conjugate of this (since $\nabla^2 F(\boldsymbol{z}^\natural)$ is Hermitian). We can rewrite the quadratic form as

$$\begin{aligned}
\boldsymbol{u}^*\boldsymbol{D}\nabla^2 F\left(\boldsymbol{z}^\natural\right)\boldsymbol{u} &\overset{\text{(i)}}{=} \left[\left(\boldsymbol{h}_1-\boldsymbol{h}_2\right)^*, \left(\boldsymbol{x}_1-\boldsymbol{x}_2\right)^*, \left(\overline{\boldsymbol{h}_1-\boldsymbol{h}_2}\right)^*, \left(\overline{\boldsymbol{x}_1-\boldsymbol{x}_2}\right)^*\right]\boldsymbol{D}\cdot \\
&\qquad \begin{bmatrix}\boldsymbol{I}_K & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{h}^\natural\boldsymbol{x}^{\natural\top} \\ \boldsymbol{0} & \boldsymbol{I}_K & \boldsymbol{x}^\natural\boldsymbol{h}^{\natural\top} & \boldsymbol{0} \\ \boldsymbol{0} & \left(\boldsymbol{x}^\natural\boldsymbol{h}^{\natural\top}\right)^* & \boldsymbol{I}_K & \boldsymbol{0} \\ \left(\boldsymbol{h}^\natural\boldsymbol{x}^{\natural\top}\right)^* & \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\left(\boldsymbol{h}_1-\boldsymbol{h}_2\right)^*, \gamma_2\left(\boldsymbol{x}_1-\boldsymbol{x}_2\right)^*, \gamma_1\left(\overline{\boldsymbol{h}_1-\boldsymbol{h}_2}\right)^*, \gamma_2\left(\overline{\boldsymbol{x}_1-\boldsymbol{x}_2}\right)^*\right]\begin{bmatrix}\boldsymbol{h}_1-\boldsymbol{h}_2+\boldsymbol{h}^\natural\boldsymbol{x}^{\natural\top}\overline{\boldsymbol{x}_1-\boldsymbol{x}_2} \\ \boldsymbol{x}_1-\boldsymbol{x}_2+\boldsymbol{x}^\natural\boldsymbol{h}^{\natural\top}\overline{\boldsymbol{h}_1-\boldsymbol{h}_2} \\ \left(\boldsymbol{x}^\natural\boldsymbol{h}^{\natural\top}\right)^*\left(\boldsymbol{x}_1-\boldsymbol{x}_2\right)+\overline{\boldsymbol{h}_1-\boldsymbol{h}_2} \\ \left(\boldsymbol{h}^\natural\boldsymbol{x}^{\natural\top}\right)^*\left(\boldsymbol{h}_1-\boldsymbol{h}_2\right)+\overline{\boldsymbol{x}_1-\boldsymbol{x}_2}\end{bmatrix} \\
&= \left[\gamma_1\left(\boldsymbol{h}_1-\boldsymbol{h}_2\right)^*, \gamma_2\left(\boldsymbol{x}_1-\boldsymbol{x}_2\right)^*, \gamma_1\left(\overline{\boldsymbol{h}_1-\boldsymbol{h}_2}\right)^*, \gamma_2\left(\overline{\boldsymbol{x}_1-\boldsymbol{x}_2}\right)^*\right]\begin{bmatrix}\boldsymbol{h}_1-\boldsymbol{h}_2+\boldsymbol{h}^\natural\left(\boldsymbol{x}_1-\boldsymbol{x}_2\right)^*\boldsymbol{x}^\natural \\ \boldsymbol{x}_1-\boldsymbol{x}_2+\boldsymbol{x}^\natural\left(\boldsymbol{h}_1-\boldsymbol{h}_2\right)^*\boldsymbol{h}^\natural \\ \overline{\boldsymbol{h}_1-\boldsymbol{h}_2}+\overline{\boldsymbol{h}^\natural\left(\boldsymbol{x}_1-\boldsymbol{x}_2\right)^*\boldsymbol{x}^\natural} \\ \overline{\boldsymbol{x}_1-\boldsymbol{x}_2}+\overline{\boldsymbol{x}^\natural\left(\boldsymbol{h}_1-\boldsymbol{h}_2\right)^*\boldsymbol{h}^\natural}\end{bmatrix} \\
&= 2\gamma_1\left\|\boldsymbol{h}_1-\boldsymbol{h}_2\right\|_2^2 + 2\gamma_2\left\|\boldsymbol{x}_1-\boldsymbol{x}_2\right\|_2^2 \\
&\quad + \left(\gamma_1+\gamma_2\right)\underbrace{\left(\boldsymbol{h}_1-\boldsymbol{h}_2\right)^*\boldsymbol{h}^\natural\left(\boldsymbol{x}_1-\boldsymbol{x}_2\right)^*\boldsymbol{x}^\natural}_{:=\beta} + \left(\gamma_1+\gamma_2\right)\underbrace{\overline{\left(\boldsymbol{h}_1-\boldsymbol{h}_2\right)^*\boldsymbol{h}^\natural\left(\boldsymbol{x}_1-\boldsymbol{x}_2\right)^*\boldsymbol{x}^\natural}}_{=\overline{\beta}}.
\end{aligned} \tag{189}$$

where (i) uses the definitions of $\boldsymbol{u}$ and $\nabla^2 F\left(\boldsymbol{z}^\natural\right)$, and (ii) follows from the definition of $\boldsymbol{D}$. In view of the assumptions that $\gamma_1, \gamma_2 \in \mathbb{R}$ and $\max\{|\gamma_1 - 1|, |\gamma_2 - 1|\} \leq \delta$, we can obtain

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

where the last inequality utilizes the identity that

$$2\|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2^2 + 2\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2 = \|\boldsymbol{u}\|_2^2. \tag{190}$$

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

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

Since $\left\|\boldsymbol{h}_2 - \boldsymbol{h}^\natural\right\|$ and $\left\|\boldsymbol{x}_2 - \boldsymbol{x}^\natural\right\|_2$ are both small by our assumption, β_2, β_3 and β_4 are easy to control. Regarding β_2, we discover that

$$\begin{aligned}|\beta_2| &\leq \left\|\boldsymbol{h}^\natural - \boldsymbol{h}_2\right\|_2 \left\|\boldsymbol{x}^\natural - \boldsymbol{x}_2\right\|_2 \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2 \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 \leq \delta^2 \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2 \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 \\ &\leq \delta \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2 \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2,\end{aligned}$$

where the second inequality is a consequence of the assumption

$$\max\left\{\left\|\boldsymbol{h}^\natural - \boldsymbol{h}_2\right\|_2, \left\|\boldsymbol{x}^\natural - \boldsymbol{x}_2\right\|_2\right\} \leq \delta \leq 1.$$

Similarly, we can obtain

$$\begin{aligned}&|\beta_3| \leq \delta \|\boldsymbol{x}_2\|_2 \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2 \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 \leq 2\delta \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2 \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2, \\ \text{and} \qquad &|\beta_4| \leq \delta \|\boldsymbol{h}_2\|_2 \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2 \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 \leq 2\delta \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2 \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2,\end{aligned}$$

where the second relations in both lines make use of the facts that

$$\|\boldsymbol{x}_2\|_2 \leq \left\|\boldsymbol{x}_2 - \boldsymbol{x}^\natural\right\|_2 + \left\|\boldsymbol{x}^\natural\right\|_2 \leq 1 + \delta \leq 2 \qquad \text{and} \qquad \|\boldsymbol{h}_2\|_2 \leq \left\|\boldsymbol{h}_2 - \boldsymbol{h}^\natural\right\|_2 + \left\|\boldsymbol{h}^\natural\right\|_2 \leq 1 + \delta \leq 2. \tag{191}$$

Combine the previous three bounds to reach

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

where we have invoked the AM-GM inequality and the identity (190).

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 (i.e. the 4th condition of Lemma 14), we can invoke Lemma 59 to obtain

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

which allows one to rewrite β_1 as

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

Consequently,

$$\beta_1 + \overline{\beta_1} = 2\left|(\boldsymbol{x}_1 - \boldsymbol{x}_2)^* \boldsymbol{x}_2\right|_2^2 + 2\mathrm{Re}\left[(\boldsymbol{x}_1 - \boldsymbol{x}_2)^* \boldsymbol{x}_2\right] \left(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2 - \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2^2\right)$$

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

Here, (i) arises from the triangle inequality that

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

and (ii) occurs since $\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2\leq\|\boldsymbol{x}_1-\boldsymbol{x}^\natural\|_2+\|\boldsymbol{x}_2-\boldsymbol{x}^\natural\|_2\leq 2\delta$ and $\|\boldsymbol{x}_2\|_2\leq 2$ (cf. (191)).

To finish up, note that $\gamma_1+\gamma_2\leq 2(1+\delta)\leq 3$ for $\delta<1/2$. Substitute these bounds into (189) to obtain

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

as long as δ is small enough. □

Proof of Lemma 26. In view of the expressions of $\nabla^2 f\left(\boldsymbol{z}\right)$ and $\nabla^2 F\left(\boldsymbol{z}^\natural\right)$ (see (79) and (187)) and the triangle inequality, we can obtain

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

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

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

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, we can use the triangle inequality to derive

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

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

$$
\beta_1=\left\|\sum_{j=1}^m\left(\left|\boldsymbol{a}_j^*\boldsymbol{x}\right|^2-\left|\boldsymbol{a}_j^*\boldsymbol{x}^\natural\right|^2\right)\boldsymbol{b}_j\boldsymbol{b}_j^*\right\|\leq\left\|\sum_{j=1}^m\left|\left|\boldsymbol{a}_j^*\boldsymbol{x}\right|^2-\left|\boldsymbol{a}_j^*\boldsymbol{x}^\natural\right|^2\right|\boldsymbol{b}_j\boldsymbol{b}_j^*\right\|, \tag{192}
$$

   where

$$
\left|\left|\boldsymbol{a}_j^*\boldsymbol{x}\right|^2-\left|\boldsymbol{a}_j^*\boldsymbol{x}^\natural\right|^2\right|=\left|\left[\boldsymbol{a}_j^*\left(\boldsymbol{x}-\boldsymbol{x}^\natural\right)\right]^*\boldsymbol{a}_j^*\left(\boldsymbol{x}-\boldsymbol{x}^\natural\right)+\left[\boldsymbol{a}_j^*\left(\boldsymbol{x}-\boldsymbol{x}^\natural\right)\right]^*\boldsymbol{a}_j^*\boldsymbol{x}^\natural+\left(\boldsymbol{a}_j^*\boldsymbol{x}^\natural\right)^*\boldsymbol{a}_j^*\left(\boldsymbol{x}-\boldsymbol{x}^\natural\right)\right|
$$

$$
\begin{aligned}
&\leq \left|\boldsymbol{a}_j^* \left(\boldsymbol{x}-\boldsymbol{x}^\natural\right)\right|^2 + 2\left|\boldsymbol{a}_j^* \left(\boldsymbol{x}-\boldsymbol{x}^\natural\right)\right| \left|\boldsymbol{a}_j^*\boldsymbol{x}^\natural\right| \\
&\leq 4C_3^2 \frac{1}{\log^3 m} + 4C_3 \frac{1}{\log^{3/2} m} \cdot 5\sqrt{\log m} \\
&\lesssim \frac{1}{\log m}.
\end{aligned}
$$

Here, the first relation comes from the identity

$$
u^*u - v^*v = (u-v)^*(u-v) + (u-v)^*v + v^*(u-v),
$$

the second line uses the triangle inequality, and the third line follows from (184) and the assumption that

$$
\max_{1\leq j\leq m} \left|\boldsymbol{a}_j^* \left(\boldsymbol{x}-\boldsymbol{x}^\natural\right)\right| \leq 2C_3 \frac{1}{\log^{1.5} m}.
$$

Substitution into (192) gives

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

where the last inequality comes from the property of $\boldsymbol{b}_j$, i.e. $\sum_{j=1}^m \boldsymbol{b}_j\boldsymbol{b}_j^* = \boldsymbol{I}_K$.

- The other term β_2 can be bounded through Lemma 60, which reveals that with probability $1-O\left(m^{-10}\right)$,

$$
\beta_2 \lesssim \sqrt{\frac{K}{m}\log^2 m}.
$$

Taken collectively, the preceding two bounds give

$$
\sup_{\boldsymbol{z}\in\mathcal{S}} \alpha_1 \lesssim \sqrt{\frac{K}{m}\log^2 m} + \frac{1}{\log m}.
$$

2. To control α_2, α_3 and α_4, we define a new set

$$
\begin{aligned}
\mathcal{S}\left(\delta, 2C_4\mu\log^2 m\right) := \Bigg\{ &(\boldsymbol{h},\boldsymbol{x}) \in \mathbb{C}^K \times \mathbb{C}^K : \ \max\left\{\|\boldsymbol{h}-\boldsymbol{h}^\natural\|_2, \|\boldsymbol{x}-\boldsymbol{x}^\natural\|_2\right\} \leq \delta \\
&\text{and } \max_{1\leq j\leq m} \left|\boldsymbol{b}_j^*\boldsymbol{h}\right| \leq 2C_4 \frac{\mu}{\sqrt{m}} \log^2 m \Bigg\}.
\end{aligned}
$$

It is easily seen that $\mathcal{S} \subseteq \mathcal{S}\left(\delta, 2C_4\mu\log^2 m\right)$ and, as a result,

$$
\sup_{\boldsymbol{z}\in\mathcal{S}} \alpha_j \leq \sup_{\boldsymbol{z}\,\in\,\mathcal{S}(\delta,2C_4\mu\log^2 m)} \alpha_j, \qquad j=2,3,4.
$$

It thus suffices to control $\sup_{\boldsymbol{z}\in\mathcal{S}(\delta,2C_4\mu\log^2 m)} \alpha_j$ for $j=2,3,4$. Note that this new set has been studied in Lemma 63, which forms the basis for our proof.

(a) The triangle inequality allows us to bound α_2 as

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

The second term θ_2 is easy to control. To see this, we have

$$
\left\| \|\boldsymbol{h}\|_2^2 \boldsymbol{I}_K - \boldsymbol{I}_K \right\| = \left| \|\boldsymbol{h}\|_2^2 - 1 \right| = \left|\|\boldsymbol{h}\|_2 - 1\right| \left(\|\boldsymbol{h}\|_2 + 1\right) \leq 3\delta,
$$

where the last relation uses the assumption that $\left\|\boldsymbol{h}-\boldsymbol{h}^{\natural}\right\|_2 \leq \delta$ and hence

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

Regarding the first term θ_1, we plan to apply Lemma 63 to show that with probability $1-O(m^{-10})$,

$$\sup_{(\boldsymbol{h},\boldsymbol{x})\,\in\,\mathcal{S}(\delta,2C_4\mu\log^2 m)} \theta_1 \leq 4\delta. \tag{193}$$

To this end, we need to introduce a set of matrices $\{\boldsymbol{A}_j(\boldsymbol{h},\boldsymbol{x})\}$ satisfying the requirement in Lemma 63. Specifically, we define $\boldsymbol{A}_j(\boldsymbol{h},\boldsymbol{x}) = |\boldsymbol{b}_j^*\boldsymbol{h}|^2\boldsymbol{I}_K$. It can be easily seen that

$$\theta_1 = \left\|\sum_{j=1}^m \boldsymbol{A}_j\left(\boldsymbol{h},\boldsymbol{x}\right)\left(\boldsymbol{a}_j\boldsymbol{a}_j^* - \boldsymbol{I}_K\right)\right\|, \tag{194}$$

which suits the form studied in Lemma 63. Next, we need to determine two bounds M_1 and M_2 as in (285) of Lemma 63.

i. When $(\boldsymbol{h},\boldsymbol{x}) \in \mathcal{S}\left(\delta, 2C_4\mu\log^2 m\right)$ for δ sufficiently small, we have $\|\boldsymbol{h}\|_2 \leq 2$ and

$$\begin{aligned}\left\|\frac{1}{m}\sum_{j=1}^m \boldsymbol{A}_j\boldsymbol{A}_j^*\right\| &= \frac{1}{m}\sum_{j=1}^m |\boldsymbol{b}_j^*\boldsymbol{h}|^4 \leq \frac{1}{m}\left\{\max_j |\boldsymbol{b}_j^*\boldsymbol{h}|^2\right\}\sum_{j=1}^m |\boldsymbol{b}_j^*\boldsymbol{h}|^2 \overset{\text{(a)}}{=} \frac{1}{m}\left\{\max_j |\boldsymbol{b}_j^*\boldsymbol{h}|^2\right\}\|\boldsymbol{h}\|_2^2 \\ &\leq \frac{4}{m}\left(\frac{2C_4\mu\log^2 m}{\sqrt{m}}\right)^2,\end{aligned}$$

where (a) has used the fact $\sum_{j=1}^m \boldsymbol{b}_j\boldsymbol{b}_j^* = \boldsymbol{I}_K$ and hence

$$\sum_{j=1}^m |\boldsymbol{b}_j^*\boldsymbol{h}|^2 = \boldsymbol{h}^*\left(\sum_{j=1}^m \boldsymbol{b}_j\boldsymbol{b}_j^*\right)\boldsymbol{h} = \|\boldsymbol{h}\|_2^2,$$

and the last inequality comes from the definition of $\mathcal{S}\left(\delta, 2C_4\mu\log^2 m\right)$. Hence, we can take $M_1 = 4C_4\mu\frac{\log^2 m}{m}$ as in Lemma 63.

ii. In addition, one can take $M_2 = 4K/m$, since

$$\begin{aligned}\|\boldsymbol{A}_j(\boldsymbol{h}_2,\boldsymbol{x}_2) - \boldsymbol{A}_j(\boldsymbol{h}_1,\boldsymbol{x}_1)\| &= \left||\boldsymbol{b}_j^*\boldsymbol{h}_2|^2 - |\boldsymbol{b}_j^*\boldsymbol{h}_1|^2\right| = \left|(\boldsymbol{h}_2-\boldsymbol{h}_1)^*\boldsymbol{b}_j\boldsymbol{b}_j^*\boldsymbol{h}_2 + \boldsymbol{h}_1^*\boldsymbol{b}_j\boldsymbol{b}_j^*(\boldsymbol{h}_2-\boldsymbol{h}_1)\right| \\ &\leq 2\max\{\|\boldsymbol{h}_1\|_2, \|\boldsymbol{h}_2\|_2\}\|\boldsymbol{h}_2-\boldsymbol{h}_1\|_2\|\boldsymbol{b}_j\|_2^2 \\ &\leq 4\|\boldsymbol{h}_2-\boldsymbol{h}_1\|_2\|\boldsymbol{b}_j\|_2^2 \\ &\leq \frac{4K}{m}\max\left\{\|\boldsymbol{h}_2-\boldsymbol{h}_1\|_2, \|\boldsymbol{x}_2-\boldsymbol{x}_1\|_2\right\}.\end{aligned}$$

Here, the last line holds since $\|\boldsymbol{b}_j\|_2^2 = K/m$.

In order to invoke Lemma 63, it remains to check whether M_1 and M_2 obey the conditions therein. In fact, we see that

$$m \gg 4K^2/m = M_2K \qquad \text{and}$$

$$\left(\min\left\{\frac{\delta}{mM_1}, 1\right\}\right)^2 m = \left(\frac{\delta}{4C_4\mu\log^2 m}\right)^2 m \gg K\log m$$

as long as $m \gg \left(\mu^2/\delta^2\right)K\log^5 m$, and hence they satisfy the conditions required in Lemma 63. As a result, we can apply Lemma 63 on (194) to conclude the assertion (193).

(b) Once again, we are going to use Lemma 63 to derive that with probability exceeding $1 - O\left(m^{-10}\right)$,

$$\sup_{\boldsymbol{z} \in \mathcal{S}(\delta, 2C_4\mu \log^2 m)} \alpha_3 \le 7\delta.$$

To this end, we first let $\boldsymbol{\Delta}(\boldsymbol{h},\boldsymbol{x}) := \boldsymbol{h}\boldsymbol{x}^* - \boldsymbol{h}^\natural \boldsymbol{x}^{\natural *}$ and $\boldsymbol{Q}(\boldsymbol{h},\boldsymbol{x}) := \sum_{j=1}^m \boldsymbol{b}_j \boldsymbol{b}_j^* \boldsymbol{\Delta}(\boldsymbol{h},\boldsymbol{x}) \boldsymbol{a}_j \boldsymbol{a}_j^*$. What we wish to establish can be reformulated as

$$\mathbb{P}\left(\sup_{(\boldsymbol{h},\boldsymbol{x}) \in \mathcal{S}(\delta, 2C_4\mu\log^2 m)} \left\|\boldsymbol{Q}\left(\boldsymbol{h},\boldsymbol{x}\right)\right\| \ge 7\delta \right) \lesssim m^{-10}.$$

Note that when $\boldsymbol{h}$ and $\boldsymbol{x}$ are independent of the $\boldsymbol{a}_j$'s, one has $\mathbb{E}\left[\boldsymbol{Q}\left(\boldsymbol{h},\boldsymbol{x}\right)\right] = \boldsymbol{\Delta}\left(\boldsymbol{h},\boldsymbol{x}\right)$ and

$$\begin{aligned} \left\|\mathbb{E}\left[\boldsymbol{Q}\left(\boldsymbol{h},\boldsymbol{x}\right)\right]\right\| = \left\|\boldsymbol{\Delta}\left(\boldsymbol{h},\boldsymbol{x}\right)\right\| = \left\|\boldsymbol{h}\boldsymbol{x}^* - \boldsymbol{h}^\natural\boldsymbol{x}^{\natural *}\right\| &= \left\|\boldsymbol{h}\left(\boldsymbol{x}-\boldsymbol{x}^\natural\right)^* + \left(\boldsymbol{h}-\boldsymbol{h}^\natural\right)\boldsymbol{x}^{\natural *}\right\| \\ &\le \left\|\boldsymbol{h}\right\|_2 \left\|\boldsymbol{x}-\boldsymbol{x}^\natural\right\|_2 + \left\|\boldsymbol{h}-\boldsymbol{h}^\natural\right\|_2 \left\|\boldsymbol{x}^\natural\right\|_2 \le 3\delta, \end{aligned}$$

where we have used the facts that $\max\left\{\left\|\boldsymbol{h}-\boldsymbol{h}^\natural\right\|_2, \left\|\boldsymbol{x}-\boldsymbol{x}^\natural\right\|_2\right\} \le \delta$ and $\left\|\boldsymbol{h}\right\|_2 \le 1+\delta \le 2$. It then suffices to show that

$$\mathbb{P}\left(\sup_{(\boldsymbol{h},\boldsymbol{x}) \in \mathcal{S}(\delta, 2C_4\mu\log^2 m)} \left\|\boldsymbol{Q}(\boldsymbol{h},\boldsymbol{x}) - \mathbb{E}\left[\boldsymbol{Q}(\boldsymbol{h},\boldsymbol{x})\right]\right\| \ge 4\delta \right) \lesssim m^{-10}. \tag{195}$$

Now we are positioned to describe how to invoke Lemma 63 for this purpose. Set $N = K$ and $\boldsymbol{A}_j(\boldsymbol{h},\boldsymbol{x}) = \boldsymbol{b}_j\boldsymbol{b}_j^*\boldsymbol{\Delta}(\boldsymbol{h},\boldsymbol{x})$ as in Lemma 63, and hence

$$\boldsymbol{Q}(\boldsymbol{h},\boldsymbol{x}) - \mathbb{E}\left[\boldsymbol{Q}(\boldsymbol{h},\boldsymbol{x})\right] = \sum_{j=1}^m \boldsymbol{A}_j(\boldsymbol{h},\boldsymbol{x})(\boldsymbol{a}_j\boldsymbol{a}_j^* - \boldsymbol{I}_K). \tag{196}$$

We then choose M_1 and M_2 (required in Lemma 63) appropriately as follows.

i. Take $M_1 \le 5C_4\mu\log^2 m/m$ since

$$\left\|\sum_{j=1}^m \boldsymbol{A}_j\boldsymbol{A}_j^*\right\| = \left\|\sum_{j=1}^m \|\boldsymbol{b}_j^*\boldsymbol{\Delta}\|_2^2 \boldsymbol{b}_j\boldsymbol{b}_j^*\right\| \le \left\{\max_j \|\boldsymbol{b}_j^*\boldsymbol{\Delta}\|_2^2\right\} \left\|\sum_{j=1}^m \boldsymbol{b}_j\boldsymbol{b}_j^*\right\| = \max_j \|\boldsymbol{b}_j^*\boldsymbol{\Delta}\|_2^2,$$

and

$$\begin{aligned} \left\|\boldsymbol{b}_j^*\boldsymbol{\Delta}\right\|_2 = \left\|\boldsymbol{b}_j^*\left(\boldsymbol{h}\boldsymbol{x}^* - \boldsymbol{h}^\natural\boldsymbol{x}^{\natural *}\right)\right\|_2 &\le \left\|\boldsymbol{b}_j^*\boldsymbol{h}\boldsymbol{x}^*\right\|_2 + \left\|\boldsymbol{b}_j^*\boldsymbol{h}^\natural\boldsymbol{x}^{\natural *}\right\|_2 \\ &\le \left|\boldsymbol{b}_j^*\boldsymbol{h}\right| \left\|\boldsymbol{x}\right\|_2 + \left|\boldsymbol{b}_j^*\boldsymbol{h}^\natural\right| \left\|\boldsymbol{x}^\natural\right\|_2 \\ &\le 2C_4\frac{\mu}{\sqrt{m}}\log^2 m \cdot 2 + \frac{\mu}{\sqrt{m}} \cdot 1 \\ &\le 5C_4\frac{\mu}{\sqrt{m}}\log^2 m, \end{aligned}$$

where the penultimate inequality follows from the assumption that $\left|\boldsymbol{b}_j^*\boldsymbol{h}\right| \le 2C_4\frac{\mu}{\sqrt{m}}\log^2 m$ (see the definition of $\mathcal{S}\left(\delta, 2C_4\mu\log^2 m\right)$) and the fact that $\left\|\boldsymbol{x}\right\|_2 \le 1+\delta \le 2$. Hence

$$\left(\min\left\{\frac{\delta}{mM_1}, 1\right\}\right)^2 m \ge \frac{\delta^2}{(5C_4)^2\mu^2\log^4 m} m \gg K\log m$$

as long as $m \gg \left(\mu^2/\delta^2\right) K\log^5 m$.

ii. Pick $M_2 \leq 4K/m$, since for any $(\boldsymbol{h}_1, \boldsymbol{x}_1), (\boldsymbol{h}_2, \boldsymbol{x}_2) \in \mathcal{S}\left(\delta, \mu \log^2 m\right)$:

$$\|\boldsymbol{A}_j(\boldsymbol{h}_2, \boldsymbol{x}_2) - \boldsymbol{A}_j(\boldsymbol{h}_1, \boldsymbol{x}_1)\| = \|\boldsymbol{b}_j \boldsymbol{b}_j^* (\boldsymbol{\Delta}(\boldsymbol{h}_1, \boldsymbol{x}_1) - \boldsymbol{\Delta}(\boldsymbol{h}_2, \boldsymbol{x}_2))\| \leq \frac{K}{m} \|\boldsymbol{\Delta}(\boldsymbol{h}_1, \boldsymbol{x}_1) - \boldsymbol{\Delta}(\boldsymbol{h}_2, \boldsymbol{x}_2)\|$$

and

$$\begin{aligned}\|\boldsymbol{\Delta}(\boldsymbol{h}_1, \boldsymbol{x}_1) - \boldsymbol{\Delta}(\boldsymbol{h}_2, \boldsymbol{x}_2)\| = \|\boldsymbol{h}_1 \boldsymbol{x}_1^* - \boldsymbol{h}_2 \boldsymbol{x}_2^*\| &= \left\|(\boldsymbol{h}_1 - \boldsymbol{h}_2)\, \boldsymbol{x}_1^* + \boldsymbol{h}_2 \,(\boldsymbol{x}_1 - \boldsymbol{x}_2)^*\right\| \\ &\leq (\|\boldsymbol{x}_1\|_2 + \|\boldsymbol{h}_2\|_2) \max\left\{\|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2, \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2\right\} \\ &\leq 4 \max\left\{\|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2, \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2\right\},\end{aligned}$$

where we use the simple facts $\|\boldsymbol{b}_j \boldsymbol{b}_j^*\| = \frac{K}{m}$ and $\max\left\{\|\boldsymbol{x}_1\|_2, \|\boldsymbol{h}_2\|_2\right\} \leq 2$ for all $(\boldsymbol{h}_1, \boldsymbol{x}_1), (\boldsymbol{h}_2, \boldsymbol{x}_2) \in \mathcal{S}\left(\delta, \mu \log^2 m\right)$. Therefore,

$$m \gg 4K^2/m \geq M_2 K.$$

The above bounds taken collectively allow us to apply Lemma 63 on (196) to establish (195).

(c) It remains to control α_4, for which we make note of the following inequality

$$\alpha_4 \leq \left\|\sum_{j=1}^m \boldsymbol{b}_j \boldsymbol{b}_j^* (\boldsymbol{h}\boldsymbol{x}^\top - \boldsymbol{h}^\natural \boldsymbol{x}^{\natural\top}) \overline{\boldsymbol{a}_j}\, \overline{\boldsymbol{a}_j}^*\right\| + \left\|\sum_{j=1}^m \boldsymbol{b}_j \boldsymbol{b}_j^* \boldsymbol{h}^\natural \boldsymbol{x}^{\natural\top} (\overline{\boldsymbol{a}_j}\, \overline{\boldsymbol{a}_j}^* - \boldsymbol{I}_K)\right\|$$

with $\overline{\boldsymbol{a}_j}$ denoting the entrywise conjugate of $\boldsymbol{a}_j$. Since $\{\overline{\boldsymbol{a}_j}\}$ have 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,

$$\mathbb{P}\left(\sup_{(\boldsymbol{h},\boldsymbol{x}) \in \mathcal{S}(\delta, 2C_4 \mu \log^2 m)} \left\|\sum_{j=1}^m \boldsymbol{b}_j \boldsymbol{b}_j^* (\boldsymbol{h}\boldsymbol{x}^\top - \boldsymbol{h}^\natural \boldsymbol{x}^{\natural\top}) \overline{\boldsymbol{a}_j}\, \overline{\boldsymbol{a}_j}^*\right\| \geq 7\delta\right) \lesssim m^{-10}.$$

The second term can again be bounded using Lemma 63. Let $\boldsymbol{A}_j(\boldsymbol{h}, \boldsymbol{x}) = \boldsymbol{b}_j \boldsymbol{b}_j^* \boldsymbol{h}^\natural \boldsymbol{x}^{\natural\top}$ as required in that lemma. Since $\|\boldsymbol{x}^\natural\|_2 = 1$ and $\max_j |\boldsymbol{b}_j^* \boldsymbol{h}^\natural| \leq \mu/\sqrt{m}$, we have

$$\left\|\frac{1}{m} \sum_{j=1}^m \boldsymbol{A}_j \boldsymbol{A}_j^*\right\| = \left\|\frac{1}{m} \sum_{j=1}^m |\boldsymbol{b}_j^* \boldsymbol{h}^\natural|^2 \boldsymbol{b}_j \boldsymbol{b}_j^*\right\| \leq \left\{\max_j \left|\boldsymbol{b}_j^* \boldsymbol{h}^\natural\right|^2\right\} \left\|\frac{1}{m} \sum_{j=1}^m \boldsymbol{b}_j \boldsymbol{b}_j^*\right\| \leq \frac{\mu^2}{m^2},$$

suggesting a bound $M_1 \leq \mu/m$. Obviously, $\boldsymbol{A}_j(\boldsymbol{h}, \boldsymbol{x})$ is invariant and hence $M_2 = 0$. Therefore, we are ready to invoke Lemma 63 to get

$$\mathbb{P}\left(\left\|\sum_{j=1}^m \boldsymbol{b}_j \boldsymbol{b}_j^* \boldsymbol{h}^\natural \boldsymbol{x}^{\natural\top} (\overline{\boldsymbol{a}_j}\, \overline{\boldsymbol{a}_j}^* - \boldsymbol{I}_K)\right\| \geq 4\delta\right) \lesssim m^{-10}.$$

Put together the above bounds to reach $\alpha_4 \leq 11\delta$ with probability $1 - O(m^{-10})$.

3. Combining all the previous bounds, we deduce that with probability $1 - O(m^{-10})$,

$$\left\|\nabla^2 f(\boldsymbol{z}) - \nabla^2 F\left(\boldsymbol{z}^\natural\right)\right\| \lesssim \left(\sqrt{\frac{K}{m} \log^2 m} + \frac{1}{\log m}\right) + \delta \lesssim \delta$$

as long as $\delta > 0$ is a small constant and $m \gg \mu^2 K \log^5 m$, as claimed.

□

C.2 Proof of Lemma 15

By the definition of α^{t+1}, one has

$$\left[\mathrm{dist}\left(\boldsymbol{z}^{t+1},\boldsymbol{z}^{\natural}\right)\right]^{2}=\left\|\frac{1}{\overline{\alpha^{t+1}}}\boldsymbol{h}^{t+1}-\boldsymbol{h}^{\natural}\right\|_{2}^{2}+\left\|\alpha^{t+1}\boldsymbol{x}^{t+1}-\boldsymbol{x}^{\natural}\right\|_{2}^{2}\le\left\|\frac{1}{\overline{\alpha^{t}}}\boldsymbol{h}^{t+1}-\boldsymbol{h}^{\natural}\right\|_{2}^{2}+\left\|\alpha^{t}\boldsymbol{x}^{t+1}-\boldsymbol{x}^{\natural}\right\|_{2}^{2}.$$

The gradient update rules (77) and (78) imply that

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

where we denote $\tilde{\boldsymbol{h}}^{t}=\frac{1}{\overline{\alpha^{t}}}\boldsymbol{h}^{t}$ and $\tilde{\boldsymbol{x}}^{t}=\alpha^{t}\boldsymbol{x}^{t}$ as usual. This further gives

$$\begin{bmatrix}\frac{1}{\overline{\alpha^{t}}}\boldsymbol{h}^{t+1}-\boldsymbol{h}^{\natural}\\ \alpha^{t}\boldsymbol{x}^{t+1}-\boldsymbol{x}^{\natural}\\ \overline{\frac{1}{\overline{\alpha^{t}}}\boldsymbol{h}^{t+1}-\boldsymbol{h}^{\natural}}\\ \overline{\alpha^{t}\boldsymbol{x}^{t+1}-\boldsymbol{x}^{\natural}}\end{bmatrix}=\begin{bmatrix}\tilde{\boldsymbol{h}}^{t}-\boldsymbol{h}^{\natural}\\ \tilde{\boldsymbol{x}}^{t}-\boldsymbol{x}^{\natural}\\ \overline{\tilde{\boldsymbol{h}}^{t}-\boldsymbol{h}^{\natural}}\\ \overline{\tilde{\boldsymbol{x}}^{t}-\boldsymbol{x}^{\natural}}\end{bmatrix}-\eta\underbrace{\begin{bmatrix}\|\tilde{\boldsymbol{x}}^{t}\|_{2}^{-2}\boldsymbol{I}_{K}&&&\\&\|\tilde{\boldsymbol{h}}^{t}\|_{2}^{-2}\boldsymbol{I}_{K}&&\\&&\|\tilde{\boldsymbol{x}}^{t}\|_{2}^{-2}\boldsymbol{I}_{K}&\\&&&\|\tilde{\boldsymbol{h}}^{t}\|_{2}^{-2}\boldsymbol{I}_{K}\end{bmatrix}}_{:=\boldsymbol{D}}\begin{bmatrix}\nabla_{\boldsymbol{h}}f\left(\tilde{\boldsymbol{z}}^{t}\right)\\ \nabla_{\boldsymbol{x}}f\left(\tilde{\boldsymbol{z}}^{t}\right)\\ \overline{\nabla_{\boldsymbol{h}}f\left(\tilde{\boldsymbol{z}}^{t}\right)}\\ \overline{\nabla_{\boldsymbol{x}}f\left(\tilde{\boldsymbol{z}}^{t}\right)}\end{bmatrix}. \tag{197}$$

The mean value theorem for vector-valued functions (see Appendix D.3.4) together with the fact that $\nabla f\left(\boldsymbol{z}^{\natural}\right)=\boldsymbol{0}$ tells us

$$\begin{bmatrix}\nabla_{\boldsymbol{h}}f\left(\tilde{\boldsymbol{z}}^{t}\right)\\ \nabla_{\boldsymbol{x}}f\left(\tilde{\boldsymbol{z}}^{t}\right)\\ \overline{\nabla_{\boldsymbol{h}}f\left(\tilde{\boldsymbol{z}}^{t}\right)}\\ \overline{\nabla_{\boldsymbol{x}}f\left(\tilde{\boldsymbol{z}}^{t}\right)}\end{bmatrix}=\begin{bmatrix}\nabla_{\boldsymbol{h}}f\left(\tilde{\boldsymbol{z}}^{t}\right)-\nabla_{\boldsymbol{h}}f\left(\boldsymbol{z}^{\natural}\right)\\ \nabla_{\boldsymbol{x}}f\left(\tilde{\boldsymbol{z}}^{t}\right)-\nabla_{\boldsymbol{x}}f\left(\boldsymbol{z}^{\natural}\right)\\ \overline{\nabla_{\boldsymbol{h}}f\left(\tilde{\boldsymbol{z}}^{t}\right)-\nabla_{\boldsymbol{h}}f\left(\boldsymbol{z}^{\natural}\right)}\\ \overline{\nabla_{\boldsymbol{x}}f\left(\tilde{\boldsymbol{z}}^{t}\right)-\nabla_{\boldsymbol{x}}f\left(\boldsymbol{z}^{\natural}\right)}\end{bmatrix}=\underbrace{\int_{0}^{1}\nabla^{2}f\left(\tilde{\boldsymbol{z}}\left(\tau\right)\right)\mathrm{d}\tau}_{:=\boldsymbol{A}}\begin{bmatrix}\tilde{\boldsymbol{h}}^{t}-\boldsymbol{h}^{\natural}\\ \tilde{\boldsymbol{x}}^{t}-\boldsymbol{x}^{\natural}\\ \overline{\tilde{\boldsymbol{h}}^{t}-\boldsymbol{h}^{\natural}}\\ \overline{\tilde{\boldsymbol{x}}^{t}-\boldsymbol{x}^{\natural}}\end{bmatrix}, \tag{198}$$

where we denote $\tilde{\boldsymbol{z}}\left(\tau\right):=\boldsymbol{z}^{\natural}+\tau\left(\tilde{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\natural}\right)$ and $\nabla^{2}f$ is the Wirtinger Hessian. Combining (197) and (198), we immediately get a formula for the gradient update rule

$$\begin{bmatrix}\frac{1}{\overline{\alpha^{t}}}\boldsymbol{h}^{t+1}-\boldsymbol{h}^{\natural}\\ \alpha^{t}\boldsymbol{x}^{t+1}-\boldsymbol{x}^{\natural}\\ \overline{\frac{1}{\overline{\alpha^{t}}}\boldsymbol{h}^{t+1}-\boldsymbol{h}^{\natural}}\\ \overline{\alpha^{t}\boldsymbol{x}^{t+1}-\boldsymbol{x}^{\natural}}\end{bmatrix}=\left(\boldsymbol{I}-\eta\boldsymbol{D}\boldsymbol{A}\right)\begin{bmatrix}\tilde{\boldsymbol{h}}^{t}-\boldsymbol{h}^{\natural}\\ \tilde{\boldsymbol{x}}^{t}-\boldsymbol{x}^{\natural}\\ \overline{\tilde{\boldsymbol{h}}^{t}-\boldsymbol{h}^{\natural}}\\ \overline{\tilde{\boldsymbol{x}}^{t}-\boldsymbol{x}^{\natural}}\end{bmatrix}. \tag{199}$$

To further simplify notation, denote $\hat{\boldsymbol{z}}^{t+1}=\left(\hat{\boldsymbol{h}}^{t+1*},\hat{\boldsymbol{x}}^{t+1*}\right)^{*}$, where $\hat{\boldsymbol{h}}^{t+1}=\frac{1}{\overline{\alpha^{t}}}\boldsymbol{h}^{t+1}$ and $\hat{\boldsymbol{x}}^{t+1}=\alpha^{t}\boldsymbol{x}^{t+1}$. Take the squared Euclidean norm on both sides of (199) to reach

$$\begin{aligned}\left\|\hat{\boldsymbol{z}}^{t+1}-\boldsymbol{z}^{\natural}\right\|_{2}^{2}&=\frac{1}{2}\begin{bmatrix}\tilde{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\natural}\\ \overline{\tilde{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\natural}}\end{bmatrix}^{*}\left(\boldsymbol{I}-\eta\boldsymbol{D}\boldsymbol{A}\right)^{*}\left(\boldsymbol{I}-\eta\boldsymbol{D}\boldsymbol{A}\right)\begin{bmatrix}\tilde{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\natural}\\ \overline{\tilde{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\natural}}\end{bmatrix}\\ &=\frac{1}{2}\begin{bmatrix}\tilde{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\natural}\\ \overline{\tilde{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\natural}}\end{bmatrix}^{*}\left(\boldsymbol{I}+\eta^{2}\boldsymbol{A}\boldsymbol{D}^{2}\boldsymbol{A}-\eta\left(\boldsymbol{D}\boldsymbol{A}+\boldsymbol{A}\boldsymbol{D}\right)\right)\begin{bmatrix}\tilde{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\natural}\\ \overline{\tilde{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\natural}}\end{bmatrix}\\ &\le\left(1+\eta^{2}\|\boldsymbol{A}\|^{2}\|\boldsymbol{D}\|^{2}\right)\left\|\tilde{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\natural}\right\|_{2}^{2}-\frac{\eta}{2}\begin{bmatrix}\tilde{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\natural}\\ \overline{\tilde{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\natural}}\end{bmatrix}^{*}\left(\boldsymbol{D}\boldsymbol{A}+\boldsymbol{A}\boldsymbol{D}\right)\begin{bmatrix}\tilde{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\natural}\\ \overline{\tilde{\boldsymbol{z}}^{t}-\boldsymbol{z}^{\natural}}\end{bmatrix}.\end{aligned} \tag{200}$$

Since $\boldsymbol{z}\left(\tau\right)$ lies between $\tilde{\boldsymbol{z}}^{t}$ and $\boldsymbol{z}^{\natural}$, we conclude from the assumptions (81a)-(81c) that for all $\tau\in[0,1]$,

$$\max\left\{\left\|\boldsymbol{h}\left(\tau\right)-\boldsymbol{h}^{\natural}\right\|_{2},\left\|\boldsymbol{x}\left(\tau\right)-\boldsymbol{x}^{\natural}\right\|_{2}\right\}\le\mathrm{dist}\left(\boldsymbol{z}^{t},\boldsymbol{z}^{\natural}\right)\le\xi\le\delta;$$

$$\max_{1\leq j\leq m}\left|\boldsymbol{a}_j^*\left(\boldsymbol{x}\left(\tau\right)-\boldsymbol{x}^\natural\right)\right|\leq C_3\frac{1}{\log^{3/2}m};$$
$$\max_{1\leq j\leq m}\left|\boldsymbol{b}_j^*\boldsymbol{h}\left(\tau\right)\right|\leq C_4\frac{\mu}{\sqrt{m}}\log^2 m$$

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

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

satisfy

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

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

$$\|\boldsymbol{A}\|\,\|\boldsymbol{D}\|\leq 4\qquad\text{and}$$

$$\begin{bmatrix}\tilde{\boldsymbol{z}}^t-\boldsymbol{z}^\natural\\ \overline{\tilde{\boldsymbol{z}}^t-\boldsymbol{z}^\natural}\end{bmatrix}^*(\boldsymbol{DA}+\boldsymbol{AD})\begin{bmatrix}\tilde{\boldsymbol{z}}^t-\boldsymbol{z}^\natural\\ \overline{\tilde{\boldsymbol{z}}^t-\boldsymbol{z}^\natural}\end{bmatrix}\geq\frac{1}{4}\left\|\begin{bmatrix}\tilde{\boldsymbol{z}}^t-\boldsymbol{z}^\natural\\ \overline{\tilde{\boldsymbol{z}}^t-\boldsymbol{z}^\natural}\end{bmatrix}\right\|_2^2=\frac{1}{2}\left\|\tilde{\boldsymbol{z}}^t-\boldsymbol{z}^\natural\right\|_2^2.$$

Substitution into (200) indicates that

$$\left\|\hat{\boldsymbol{z}}^{t+1}-\boldsymbol{z}^\natural\right\|_2^2\leq\left(1+16\eta^2-\frac{\eta}{4}\right)\left\|\tilde{\boldsymbol{z}}^t-\boldsymbol{z}^\natural\right\|_2^2.$$

When $\eta\leq\frac{1}{128}$, this implies that

$$\left\|\hat{\boldsymbol{z}}^t-\boldsymbol{z}^\natural\right\|_2^2\leq(1-\eta/8)\left\|\tilde{\boldsymbol{z}}^t-\boldsymbol{z}^\natural\right\|_2^2,$$

and hence

$$\left\|\tilde{\boldsymbol{z}}^{t+1}-\boldsymbol{z}^\natural\right\|_2\leq\left\|\hat{\boldsymbol{z}}^{t+1}-\boldsymbol{z}^\natural\right\|_2\leq(1-\eta/8)^{1/2}\left\|\tilde{\boldsymbol{z}}^t-\boldsymbol{z}^\natural\right\|_2\leq(1-\eta/16)\,\mathrm{dist}(\boldsymbol{z}^t,\boldsymbol{z}^\natural).\tag{201}$$

This completes the proof of Lemma 15.

Before we conclude this section, we make note of a byproduct regarding the alignment parameter α^{t+1}, which will be useful in the subsequent analysis.

Lemma 27. *Suppose that $m\gg 1$. The following two hold true.*

- *If $\left||\alpha^t|-1\right|<1/2$ and $\mathrm{dist}(\boldsymbol{z}^t,\boldsymbol{z}^\natural)\leq C_1/\log^2 m$, then*

$$\left|\frac{\alpha^{t+1}}{\alpha^t}-1\right|\leq c\,\mathrm{dist}(\boldsymbol{z}^t,\boldsymbol{z}^\natural)\leq\frac{cC_1}{\log^2 m}$$

  for some absolute constant $c>0$;

- *If $\left||\alpha^0|-1\right|<1/4$ and $\mathrm{dist}(\boldsymbol{z}^s,\boldsymbol{z}^\natural)\leq C_1(1-\eta/16)^s/\log^2 m$ for all $0\leq s\leq t$, then one has*

$$\left||\alpha^{s+1}|-1\right|<1/2,\qquad 0\leq s\leq t.$$

Proof. Reuse the notation in this appendix, namely, $\hat{\boldsymbol{z}}^{t+1}=\begin{bmatrix}\hat{\boldsymbol{h}}^{t+1}\\ \hat{\boldsymbol{x}}^{t+1}\end{bmatrix}$ with $\hat{\boldsymbol{h}}^{t+1}=\frac{1}{\overline{\alpha^t}}\boldsymbol{h}^{t+1}$ and $\hat{\boldsymbol{x}}^{t+1}=\alpha^t\boldsymbol{x}^{t+1}$. From the proof in this section, one can tell (see (201))

$$\left\|\tilde{\boldsymbol{z}}^{t+1}-\boldsymbol{z}^\natural\right\|_2\leq\left\|\hat{\boldsymbol{z}}^{t+1}-\boldsymbol{z}^\natural\right\|_2\leq\mathrm{dist}(\boldsymbol{z}^t,\boldsymbol{z}^\natural).$$

Invoke Lemma 55 with $\beta=\alpha^t$ to get

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

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

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

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 $\mathrm{dist}(\boldsymbol{z}^s,\boldsymbol{z}^\natural) \leq C_1(1-\eta/16)^s/\log^2 m$ hold for all $0 \leq s \leq \tau \leq t$, then using our result in the first part gives

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

for m sufficiently large. The proof is then complete by induction. $\square$

C.3 Proof of Lemma 16

We first make the observation that for any $\alpha \neq 0$, one has

$$\begin{aligned}&\left\|\begin{bmatrix}\frac{1}{\overline{\alpha}}\boldsymbol{I}_K & \\ & \alpha\boldsymbol{I}_K\end{bmatrix}\begin{bmatrix}\boldsymbol{h}^{t+1,(l)}\\ \boldsymbol{x}^{t+1,(l)}\end{bmatrix} - \begin{bmatrix}\frac{1}{\overline{\alpha^{t+1}}}\boldsymbol{I}_K & \\ & \alpha^{t+1}\boldsymbol{I}_K\end{bmatrix}\begin{bmatrix}\boldsymbol{h}^{t+1}\\ \boldsymbol{x}^{t+1}\end{bmatrix}\right\|_2\\
&= \left\|\begin{bmatrix}\frac{1}{\overline{\alpha}}\boldsymbol{I}_K & \\ & \alpha\boldsymbol{I}_K\end{bmatrix}\begin{bmatrix}\boldsymbol{h}^{t+1,(l)}\\ \boldsymbol{x}^{t+1,(l)}\end{bmatrix} - \begin{bmatrix}\frac{\overline{\alpha^t}}{\overline{\alpha^{t+1}}}\boldsymbol{I}_K & \\ & \frac{\alpha^{t+1}}{\alpha^t}\boldsymbol{I}_K\end{bmatrix}\begin{bmatrix}\frac{1}{\overline{\alpha^t}}\boldsymbol{I}_K & \\ & \alpha^t\boldsymbol{I}_K\end{bmatrix}\begin{bmatrix}\boldsymbol{h}^{t+1}\\ \boldsymbol{x}^{t+1}\end{bmatrix}\right\|_2\\
&= \left\|\begin{bmatrix}\frac{\overline{\alpha^t}}{\overline{\alpha^{t+1}}}\boldsymbol{I}_K & \\ & \frac{\alpha^{t+1}}{\alpha^t}\boldsymbol{I}_K\end{bmatrix}\right\| \cdot \left\|\begin{bmatrix}\frac{\overline{\alpha^{t+1}}}{\overline{\alpha^t}}\frac{1}{\overline{\alpha}}\boldsymbol{I}_K & \\ & \frac{\alpha^t}{\alpha^{t+1}}\alpha\boldsymbol{I}_K\end{bmatrix}\begin{bmatrix}\boldsymbol{h}^{t+1,(l)}\\ \boldsymbol{x}^{t+1,(l)}\end{bmatrix} - \begin{bmatrix}\frac{1}{\overline{\alpha^t}}\boldsymbol{I}_K & \\ & \alpha^t\boldsymbol{I}_K\end{bmatrix}\begin{bmatrix}\boldsymbol{h}^{t+1}\\ \boldsymbol{x}^{t+1}\end{bmatrix}\right\|_2.\end{aligned} \tag{202}$$

Specifically, by taking

$$\alpha = \frac{\alpha^{t+1}}{\alpha^t}\alpha_{\mathrm{mutual}}^{t,(l)}$$

in (202) with

$$\alpha_{\mathrm{mutual}}^{t,(l)} := \arg\min_{\alpha}\left\|\begin{bmatrix}\frac{1}{\overline{\alpha}}\boldsymbol{I}_K & \\ & \alpha\boldsymbol{I}_K\end{bmatrix}\begin{bmatrix}\boldsymbol{h}^{t,(l)}\\ \boldsymbol{x}^{t,(l)}\end{bmatrix} - \begin{bmatrix}\frac{1}{\overline{\alpha^t}}\boldsymbol{I}_K & \\ & \alpha^t\boldsymbol{I}_K\end{bmatrix}\begin{bmatrix}\boldsymbol{h}^t\\ \boldsymbol{x}^t\end{bmatrix}\right\|_2,$$

we can bound the quantity of interest by

$$\begin{aligned}\mathrm{dist}\big(\boldsymbol{z}^{t+1,(l)},\tilde{\boldsymbol{z}}^{t+1}\big) &= \min_{\alpha}\left\|\begin{bmatrix}\frac{1}{\overline{\alpha}}\boldsymbol{I}_K & \\ & \alpha\boldsymbol{I}_K\end{bmatrix}\begin{bmatrix}\boldsymbol{h}^{t+1,(l)}\\ \boldsymbol{x}^{t+1,(l)}\end{bmatrix} - \begin{bmatrix}\frac{1}{\overline{\alpha^{t+1}}}\boldsymbol{I}_K & \\ & \alpha^{t+1}\boldsymbol{I}_K\end{bmatrix}\begin{bmatrix}\boldsymbol{h}^{t+1}\\ \boldsymbol{x}^{t+1}\end{bmatrix}\right\|_2\\
&\leq \left\|\begin{bmatrix}\frac{\overline{\alpha^t}}{\overline{\alpha^{t+1}}}\boldsymbol{I}_K & \\ & \frac{\alpha^{t+1}}{\alpha^t}\boldsymbol{I}_K\end{bmatrix}\right\| \cdot \left\|\begin{bmatrix}\frac{1}{\overline{\alpha_{\mathrm{mutual}}^{t,(l)}}}\boldsymbol{I}_K & \\ & \alpha_{\mathrm{mutual}}^{t,(l)}\boldsymbol{I}_K\end{bmatrix}\begin{bmatrix}\boldsymbol{h}^{t+1,(l)}\\ \boldsymbol{x}^{t+1,(l)}\end{bmatrix} - \begin{bmatrix}\frac{1}{\overline{\alpha^t}}\boldsymbol{I}_K & \\ & \alpha^t\boldsymbol{I}_K\end{bmatrix}\begin{bmatrix}\boldsymbol{h}^{t+1}\\ \boldsymbol{x}^{t+1}\end{bmatrix}\right\|_2\\
&\leq \max\left\{\left|\frac{\alpha^{t+1}}{\alpha^t}\right|,\left|\frac{\alpha^t}{\alpha^{t+1}}\right|\right\}\left\|\begin{bmatrix}\frac{1}{\overline{\alpha_{\mathrm{mutual}}^{t,(l)}}}\boldsymbol{h}^{t+1,(l)} - \frac{1}{\overline{\alpha^t}}\boldsymbol{h}^{t+1}\\ \alpha_{\mathrm{mutual}}^{t,(l)}\boldsymbol{x}^{t+1,(l)} - \alpha^t\boldsymbol{x}^{t+1}\end{bmatrix}\right\|_2.\end{aligned} \tag{203}$$

The reason that we connect this with $\alpha_{\mathrm{mutual}}^{t,(l)}$ lies in the fact that $\alpha_{\mathrm{mutual}}^{t,(l)}$ is the optimal complex scaling to align $\boldsymbol{z}^{t,(l)}$ and $\tilde{\boldsymbol{z}}^t$. This is more convenient to work with when controlling the gap between $\boldsymbol{z}^{t,(l)}$ and $\boldsymbol{z}^t$.

We now focus on bounding (203). Recall our notation

$$\tilde{\boldsymbol{h}}^t := \frac{1}{\overline{\alpha^t}}\boldsymbol{h}^t \qquad \text{and} \qquad \tilde{\boldsymbol{x}}^t := \alpha^t\boldsymbol{x}^t.$$

For simplicity of presentation, we further define two related vectors

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

Thus, we can apply the gradient update rules (77) and (78) to get

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

By construction, we can decompose

$$\begin{aligned}
\nabla_{\boldsymbol{h}}f^{(l)}\left(\boldsymbol{h},\boldsymbol{x}\right) &= \nabla_{\boldsymbol{h}}f\left(\boldsymbol{h},\boldsymbol{x}\right) - \left(\boldsymbol{b}_l^*\boldsymbol{h}\boldsymbol{x}^*\boldsymbol{a}_l - y_l\right)\boldsymbol{b}_l\boldsymbol{a}_l^*\boldsymbol{x} \qquad \text{and} \\
\nabla_{\boldsymbol{x}}f^{(l)}\left(\boldsymbol{h},\boldsymbol{x}\right) &= \nabla_{\boldsymbol{h}}f\left(\boldsymbol{h},\boldsymbol{x}\right) - \overline{\left(\boldsymbol{b}_l^*\boldsymbol{h}\boldsymbol{x}^*\boldsymbol{a}_l - y_l\right)}\boldsymbol{a}_l\boldsymbol{b}_l^*\boldsymbol{h},
\end{aligned}$$

allowing us to continue the derivation and obtain

$$\begin{aligned}
\begin{bmatrix} \frac{1}{\overline{\alpha_{\text{mutual}}^{t,(l)}}}\boldsymbol{h}^{t+1,(l)} - \frac{1}{\overline{\alpha^t}}\boldsymbol{h}^{t+1} \\ \alpha_{\text{mutual}}^{t,(l)}\boldsymbol{x}^{t+1,(l)} - \alpha^t\boldsymbol{x}^{t+1} \end{bmatrix} &= \begin{bmatrix} \hat{\boldsymbol{h}}^{t,(l)} - \frac{\eta}{\left\|\hat{\boldsymbol{x}}^{t,(l)}\right\|_2^2}\nabla_{\boldsymbol{h}}f\left(\hat{\boldsymbol{h}}^{t,(l)},\hat{\boldsymbol{x}}^{t,(l)}\right) - \left(\tilde{\boldsymbol{h}}^t - \frac{\eta}{\|\tilde{\boldsymbol{x}}^t\|_2^2}\nabla_{\boldsymbol{h}}f\left(\tilde{\boldsymbol{h}}^t,\tilde{\boldsymbol{x}}^t\right)\right) \\ \hat{\boldsymbol{x}}^{t,(l)} - \frac{\eta}{\left\|\hat{\boldsymbol{h}}^{t,(l)}\right\|_2^2}\nabla_{\boldsymbol{x}}f\left(\hat{\boldsymbol{h}}^{t,(l)},\hat{\boldsymbol{x}}^{t,(l)}\right) - \left(\tilde{\boldsymbol{x}}^t - \frac{\eta}{\|\tilde{\boldsymbol{h}}^t\|_2^2}\nabla_{\boldsymbol{x}}f\left(\tilde{\boldsymbol{h}}^t,\tilde{\boldsymbol{x}}^t\right)\right) \end{bmatrix} \\
&\quad - \eta\underbrace{\begin{bmatrix} \frac{1}{\left\|\hat{\boldsymbol{x}}^{t,(l)}\right\|_2^2}\left(\boldsymbol{b}_l^*\hat{\boldsymbol{h}}^{t,(l)}\hat{\boldsymbol{x}}^{t,(l)*}\boldsymbol{a}_l - y_l\right)\boldsymbol{b}_l\boldsymbol{a}_l^*\hat{\boldsymbol{x}}^{t,(l)} \\ \frac{1}{\left\|\hat{\boldsymbol{h}}^{t,(l)}\right\|_2^2}\overline{\left(\boldsymbol{b}_l^*\hat{\boldsymbol{h}}^{t,(l)}\hat{\boldsymbol{x}}^{t,(l)*}\boldsymbol{a}_l - y_l\right)}\boldsymbol{a}_l\boldsymbol{b}_l^*\hat{\boldsymbol{h}}^{t,(l)} \end{bmatrix}}_{:=\boldsymbol{I}_3}.
\end{aligned}$$

This further gives

$$\begin{aligned}
\begin{bmatrix} \frac{1}{\overline{\alpha_{\text{mutual}}^{t,(l)}}}\boldsymbol{h}^{t+1,(l)} - \frac{1}{\overline{\alpha^t}}\boldsymbol{h}^{t+1} \\ \alpha_{\text{mutual}}^{t,(l)}\boldsymbol{x}^{t+1,(l)} - \alpha^t\boldsymbol{x}^{t+1} \end{bmatrix} &= \underbrace{\begin{bmatrix} \hat{\boldsymbol{h}}^{t,(l)} - \frac{\eta}{\left\|\hat{\boldsymbol{x}}^{t,(l)}\right\|_2^2}\nabla_{\boldsymbol{h}}f\left(\hat{\boldsymbol{h}}^{t,(l)},\hat{\boldsymbol{x}}^{t,(l)}\right) - \left(\tilde{\boldsymbol{h}}^t - \frac{\eta}{\left\|\hat{\boldsymbol{x}}^{t,(l)}\right\|_2^2}\nabla_{\boldsymbol{h}}f\left(\tilde{\boldsymbol{h}}^t,\tilde{\boldsymbol{x}}^t\right)\right) \\ \hat{\boldsymbol{x}}^{t,(l)} - \frac{\eta}{\left\|\hat{\boldsymbol{h}}^{t,(l)}\right\|_2^2}\nabla_{\boldsymbol{x}}f\left(\hat{\boldsymbol{h}}^{t,(l)},\hat{\boldsymbol{x}}^{t,(l)}\right) - \left(\tilde{\boldsymbol{x}}^t - \frac{\eta}{\left\|\hat{\boldsymbol{h}}^{t,(l)}\right\|_2^2}\nabla_{\boldsymbol{x}}f\left(\tilde{\boldsymbol{h}}^t,\tilde{\boldsymbol{x}}^t\right)\right) \end{bmatrix}}_{:=\boldsymbol{I}_1} \\
&\quad + \eta\underbrace{\begin{bmatrix} \left(\frac{1}{\|\tilde{\boldsymbol{x}}^t\|_2^2} - \frac{1}{\left\|\hat{\boldsymbol{x}}^{t,(l)}\right\|_2^2}\right)\nabla_{\boldsymbol{h}}f\left(\tilde{\boldsymbol{h}}^t,\tilde{\boldsymbol{x}}^t\right) \\ \left(\frac{1}{\|\tilde{\boldsymbol{h}}^t\|_2^2} - \frac{1}{\left\|\hat{\boldsymbol{h}}^{t,(l)}\right\|_2^2}\right)\nabla_{\boldsymbol{x}}f\left(\tilde{\boldsymbol{h}}^t,\tilde{\boldsymbol{x}}^t\right) \end{bmatrix}}_{:=\boldsymbol{I}_2} - \eta\boldsymbol{I}_3. \qquad (204)
\end{aligned}$$

In what follows, we bound the three terms separately.

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

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

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

The vector-valued mean value theorem (see Appendix D.3.4) reveals that

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

where we abuse the notation and denote $\boldsymbol{z}\left(\tau\right) = \tilde{\boldsymbol{z}}^t + \tau\left(\hat{\boldsymbol{z}}^{t,(l)} - \tilde{\boldsymbol{z}}^t\right)$. In order to invoke Lemma 14 to analyze the Hessian properties, we need to verify the conditions required therein. Recall the induction hypothesis (86b) that

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

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

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

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

where we have used the induction hypotheses (86a) and (86b);

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

$$\begin{aligned} \max_{1\leq j\leq m} \left|\boldsymbol{a}_j^*\left(\boldsymbol{x}\left(\tau\right) - \boldsymbol{x}^\natural\right)\right| &\leq \max_{1\leq j\leq m} \left|\boldsymbol{a}_j^*\left(\hat{\boldsymbol{x}}^{t,(l)} - \tilde{\boldsymbol{x}}^t\right)\right| + \max_{1\leq j\leq m} \left|\boldsymbol{a}_j^*\left(\tilde{\boldsymbol{x}}^t - \boldsymbol{x}^\natural\right)\right| \\ &\leq \max_{1\leq j\leq m} \left\|\boldsymbol{a}_j\right\|_2 \left\|\hat{\boldsymbol{z}}^{t,(l)} - \tilde{\boldsymbol{z}}^t\right\|_2 + C_3 \frac{1}{\log^{3/2} m} \\ &\leq 3\sqrt{K} \cdot C_2 \frac{\mu}{\sqrt{m}} \sqrt{\frac{\mu^2 K \log^9 m}{m}} + C_3 \frac{1}{\log^{3/2} m} \leq 2C_3 \frac{1}{\log^{3/2} m}, \end{aligned}$$

which follows from the bound $\|\boldsymbol{a}_j\|_2 \leq 3\sqrt{K}$ and the induction hypotheses (86b) and (86c);

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

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

which makes use of the fact $\|\boldsymbol{b}_j\|_2 = \sqrt{K/m}$ as well as the induction hypotheses (86b) and (86d).

These properties satisfy the conditions required in Lemma 14. Thus, we can repeat the argument used in Appendix C.2 to obtain

$$\|\boldsymbol{I}_1\|_2 \le \left(1 - \frac{\eta}{16}\right) \big\|\hat{\boldsymbol{z}}^{t,(l)} - \tilde{\boldsymbol{z}}^t\big\|_2.$$

2. In terms of the second term $\boldsymbol{I}_2$, it is easy to see that

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

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

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

where the last inequality comes from the induction hypothesis (86a). Additionally, one can easily verify that

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

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

$$\|\boldsymbol{I}_2\|_2 \lesssim \frac{1}{\log^2 m} \big\|\hat{\boldsymbol{z}}^{t,(l)} - \tilde{\boldsymbol{z}}^t\big\|_2.$$

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

$$\left\| \left(\boldsymbol{b}_l^* \hat{\boldsymbol{h}}^{t,(l)} \hat{\boldsymbol{x}}^{t,(l)*} \boldsymbol{a}_l - y_l \right) \boldsymbol{b}_l \boldsymbol{a}_l^* \hat{\boldsymbol{x}}^{t,(l)} \right\|_2 \le \left| \boldsymbol{b}_l^* \hat{\boldsymbol{h}}^{t,(l)} \hat{\boldsymbol{x}}^{t,(l)*} \boldsymbol{a}_l - y_l \right| \|\boldsymbol{b}_l\|_2 \left|\boldsymbol{a}_l^* \hat{\boldsymbol{x}}^{t,(l)}\right|. \tag{205}$$

Invoke the fact $\|\boldsymbol{a}_l\|_2 \lesssim \sqrt{K}$ and the induction hypotheses (86b) and (86c) to arrive at

$$\begin{aligned}
\left|\boldsymbol{a}_l^* \big(\hat{\boldsymbol{x}}^{t,(l)} - \boldsymbol{x}^\natural\big)\right| &\le \left|\boldsymbol{a}_l^* \big(\tilde{\boldsymbol{x}}^t - \boldsymbol{x}^\natural\big)\right| + \left|\boldsymbol{a}_l^* \big(\hat{\boldsymbol{x}}^{t,(l)} - \tilde{\boldsymbol{x}}^t\big)\right| \\
&\lesssim C_3 \frac{1}{\log^{1.5} m} + \|\boldsymbol{a}_l\|_2 \|\hat{\boldsymbol{x}}^{t,(l)} - \tilde{\boldsymbol{x}}^t\|_2 \\
&\lesssim C_3 \frac{1}{\log^{1.5} m} + \sqrt{K} \operatorname{dist}\big(\boldsymbol{z}^{t,(l)}, \tilde{\boldsymbol{z}}^t\big) \\
&\lesssim C_3 \frac{1}{\log^{1.5} m} + C_2 \frac{\mu\sqrt{K}}{\sqrt{m}} \sqrt{\frac{\mu^2 K \log^9 m}{m}} \lesssim C_3 \frac{1}{\log^{1.5} m}
\end{aligned} \tag{206}$$

as long as $m \gg \mu^2 K \log^6 m$, which together with the fact $\left|\boldsymbol{a}_l^* \boldsymbol{x}^\natural\right| \lesssim \sqrt{\log m}$ yields

$$\left|\boldsymbol{a}_l^* \hat{\boldsymbol{x}}^{t,(l)}\right| \le \left|\boldsymbol{a}_l^* \boldsymbol{x}^\natural\right| + \left|\boldsymbol{a}_l^* \big(\hat{\boldsymbol{x}}^{t,(l)} - \boldsymbol{x}^\natural\big)\right|$$

$$\lesssim \sqrt{\log m} + C_3 \frac{1}{\log^{1.5} m} \asymp \sqrt{\log m}.$$

In addition, using the fact $\|\boldsymbol{b}_l\|_2 = \sqrt{K/m}$ and the induction hypotheses (86b) and (86d), we get

$$\begin{aligned}
\left|\boldsymbol{b}_l^* (\hat{\boldsymbol{h}}^{t,(l)} - \boldsymbol{h}^\natural)\right| &\le \left|\boldsymbol{b}_l^* \tilde{\boldsymbol{h}}^t\right| + \left|\boldsymbol{b}_l^* \boldsymbol{h}^\natural\right| + \left|\boldsymbol{b}_l^* (\hat{\boldsymbol{h}}^{t,(l)} - \tilde{\boldsymbol{h}}^t)\right| \\
&\le C_4 \frac{\mu}{\sqrt{m}} \log^2 m + \frac{\mu}{\sqrt{m}} + \|\boldsymbol{b}_l\|_2 \left\|\hat{\boldsymbol{h}}^{t,(l)} - \tilde{\boldsymbol{h}}^t\right\|_2 \\
&\lesssim C_4 \frac{\mu}{\sqrt{m}} \log^2 m + \sqrt{\frac{K}{m}} \operatorname{dist}\left(\boldsymbol{z}^{t,(l)}, \tilde{\boldsymbol{z}}^t\right) \\
&\lesssim C_4 \frac{\mu}{\sqrt{m}} \log^2 m + C_2 \sqrt{\frac{K}{m}} \frac{\mu}{\sqrt{m}} \sqrt{\frac{\mu^2 K \log^9 m}{m}} \lesssim C_4 \frac{\mu}{\sqrt{m}} \log^2 m, \qquad (207)
\end{aligned}$$

as long as $m \gg \mu K \log^{5/2} m$. Therefore, make use of (206) and (207) to obtain

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

Substitution into (205) gives

$$\begin{aligned}
\left\|\left(\boldsymbol{b}_l^* \hat{\boldsymbol{h}}^{t,(l)} \hat{\boldsymbol{x}}^{t,(l)*} \boldsymbol{a}_l - y_l\right) \boldsymbol{b}_l \boldsymbol{a}_l^* \hat{\boldsymbol{x}}^{t,(l)}\right\|_2 &\le \left|\boldsymbol{b}_l^* \hat{\boldsymbol{h}}^{t,(l)} \hat{\boldsymbol{x}}^{t,(l)*} \boldsymbol{a}_l - y_l\right| \|\boldsymbol{b}_l\|_2 \left|\boldsymbol{a}_l^* \hat{\boldsymbol{x}}^{t,(l)}\right| \\
&\lesssim C_4 \mu \frac{\log^{5/2} m}{\sqrt{m}} \cdot \sqrt{\frac{K}{m}} \cdot \sqrt{\log m}. \qquad (208)
\end{aligned}$$

Similarly, we can also derive

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

Putting these bounds together indicates that

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

The above bounds taken together with (203) and (204) ensure the existence of a constant $C > 0$ such that

$$\begin{aligned}
\operatorname{dist}\left(\boldsymbol{z}^{t+1,(l)}, \tilde{\boldsymbol{z}}^{t+1}\right) &\le \max\left\{\left|\frac{\alpha^{t+1}}{\alpha^t}\right|, \left|\frac{\alpha^t}{\alpha^{t+1}}\right|\right\} \left\{\left(1 - \frac{\eta}{16} + C\eta \frac{1}{\log^2 m}\right) \left\|\hat{\boldsymbol{z}}^{t,(l)} - \tilde{\boldsymbol{z}}^t\right\|_2 + C\left(C_4\right)^2 \eta \frac{\mu}{\sqrt{m}} \sqrt{\frac{\mu^2 K \log^9 m}{m}}\right\} \\
&\le \frac{1 - \eta/21}{1 - \eta/20} \left\{\left(1 - \frac{\eta}{20}\right) \left\|\hat{\boldsymbol{z}}^{t,(l)} - \tilde{\boldsymbol{z}}^t\right\|_2 + C\left(C_4\right)^2 \eta \frac{\mu}{\sqrt{m}} \sqrt{\frac{\mu^2 K \log^9 m}{m}}\right\} \\
&\le \left(1 - \frac{\eta}{21}\right) \left\|\hat{\boldsymbol{z}}^{t,(l)} - \tilde{\boldsymbol{z}}^t\right\|_2 + 2C\left(C_4\right)^2 \eta \frac{\mu}{\sqrt{m}} \sqrt{\frac{\mu^2 K \log^9 m}{m}}
\end{aligned}$$

$$
\begin{aligned}
&= \left(1-\frac{\eta}{21}\right)\operatorname{dist}\big(\boldsymbol{z}^{t,(l)},\tilde{\boldsymbol{z}}^{t}\big) + 2C\left(C_4\right)^2\eta\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^2 K\log^9 m}{m}} \\
&\le C_2\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^2 K\log^9 m}{m}},
\end{aligned}
$$

with the proviso that m is sufficiently large, $C_2 \gg (C_4)^2$ and

$$
\max\left\{\left|\frac{\alpha^{t+1}}{\alpha^{t}}\right|,\left|\frac{\alpha^{t}}{\alpha^{t+1}}\right|\right\} < \frac{1-\eta/21}{1-\eta/20}. \tag{209}
$$

Here, the last inequality invokes the induction hypothesis (86b) and assumes that $C_2 > 0$ is sufficiently large. According to Lemma 27, the condition (209) is guaranteed in the presence of a sufficiently large m.

Finally we establish the second inequality claimed in the lemma. Take $(\boldsymbol{h}_1,\boldsymbol{x}_1) = (\tilde{\boldsymbol{h}}^{t+1},\tilde{\boldsymbol{x}}^{t+1})$ and $(\boldsymbol{h}_2,\boldsymbol{x}_2) = (\hat{\boldsymbol{h}}^{t+1,(l)},\hat{\boldsymbol{x}}^{t+1,(l)})$ in Lemma 58. Since both $(\boldsymbol{h}_1,\boldsymbol{x}_1)$ and $(\boldsymbol{h}_2,\boldsymbol{x}_2)$ are close enough to $(\boldsymbol{h}^{\natural},\boldsymbol{x}^{\natural})$, we deduce that

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

as claimed.

C.4 Proof of Lemma 17

Before going forward, we make note of the following inequality

$$
\max_{1\le j\le m}\left|\boldsymbol{b}_j^*\frac{1}{\overline{\alpha^{t+1}}}\boldsymbol{h}^{t+1}\right| \le \left|\frac{\alpha^{t}}{\alpha^{t+1}}\right|\max_{1\le j\le m}\left|\boldsymbol{b}_j^*\frac{1}{\overline{\alpha^{t}}}\boldsymbol{h}^{t+1}\right| \le (1+\delta)\max_{1\le j\le m}\left|\boldsymbol{b}_j^*\frac{1}{\overline{\alpha^{t}}}\boldsymbol{h}^{t+1}\right|
$$

for some small $\delta \asymp \frac{1}{\log^2 m}$, where the last relation follows from Lemma 27 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 $\boldsymbol{b}_j^*\frac{1}{\overline{\alpha^t}}\boldsymbol{h}^{t+1}$.

The gradient update rule for $\boldsymbol{h}^{t+1}$ (cf. (77)) gives

$$
\frac{1}{\overline{\alpha^{t}}}\boldsymbol{h}^{t+1} = \tilde{\boldsymbol{h}}^{t} - \eta\xi\sum_{j=1}^{m}\boldsymbol{b}_j\boldsymbol{b}_j^*\big(\tilde{\boldsymbol{h}}^{t}\tilde{\boldsymbol{x}}^{t*}-\boldsymbol{h}^{\natural}\boldsymbol{x}^{\natural *}\big)\boldsymbol{a}_j\boldsymbol{a}_j^*\tilde{\boldsymbol{x}}^{t}.
$$

Here and below, we denote

$$
\xi = 1/\left\|\tilde{\boldsymbol{x}}^{t}\right\|_2^2
$$

out of notational convenience. The above formula can be further decomposed into the following terms

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

which relies on the fact that $\sum_{j=1}^{m}\boldsymbol{b}_j\boldsymbol{b}_j^* = \boldsymbol{I}_K$. In the sequel, we shall control each term separately.

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

$$\frac{1}{\eta\xi} \left| \boldsymbol{b}_l^* \boldsymbol{v}_1 \right| = \left| \sum_{j=1}^m \boldsymbol{b}_l^* \boldsymbol{b}_j \boldsymbol{b}_j^* \tilde{\boldsymbol{h}}^t \left[\boldsymbol{a}_j^* \left(\tilde{\boldsymbol{x}}^t - \boldsymbol{x}^\natural \right) \left(\boldsymbol{a}_j^* \tilde{\boldsymbol{x}}^t \right)^* + \boldsymbol{a}_j^* \boldsymbol{x}^\natural \left(\boldsymbol{a}_j^* \left(\tilde{\boldsymbol{x}}^t - \boldsymbol{x}^\natural \right) \right)^* \right] \right|$$
$$\leq \sum_{j=1}^m \left| \boldsymbol{b}_l^* \boldsymbol{b}_j \right| \left\{ \max_{1 \leq j \leq m} \left| \boldsymbol{b}_j^* \tilde{\boldsymbol{h}}^t \right| \right\} \left\{ \max_{1 \leq j \leq m} \left| \boldsymbol{a}_j^* \left(\tilde{\boldsymbol{x}}^t - \boldsymbol{x}^\natural \right) \right| \left(\left| \boldsymbol{a}_j^* \tilde{\boldsymbol{x}}^t \right| + \left| \boldsymbol{a}_j^* \boldsymbol{x}^\natural \right| \right) \right\}. \tag{210}$$

Combining the induction hypothesis (86c) and the condition (184) yields

$$\max_{1 \leq j \leq m} \left| \boldsymbol{a}_j^* \tilde{\boldsymbol{x}}^t \right| \leq \max_{1 \leq j \leq m} \left| \boldsymbol{a}_j^* \left(\tilde{\boldsymbol{x}}^t - \boldsymbol{x}^\natural \right) \right| + \max_{1 \leq j \leq m} \left| \boldsymbol{a}_j^* \boldsymbol{x}^\natural \right| \leq C_3 \frac{1}{\log^{1.5} m} + 5\sqrt{\log m} \leq 6\sqrt{\log m}.$$

This further implies

$$\max_{1 \leq j \leq m} \left| \boldsymbol{a}_j^* \left(\tilde{\boldsymbol{x}}^t - \boldsymbol{x}^\natural \right) \right| \left(\left| \boldsymbol{a}_j^* \tilde{\boldsymbol{x}}^t \right| + \left| \boldsymbol{a}_j^* \boldsymbol{x}^\natural \right| \right) \leq C_3 \frac{1}{\log^{1.5} m} \cdot 11\sqrt{\log m} \leq 11 C_3 \frac{1}{\log m}.$$

Substituting it into (210) and taking Lemma 51, we arrive at

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

with the proviso that C_3 is sufficiently small.

2. We then move on to $\boldsymbol{v}_3$, which obeys

$$\frac{1}{\eta\xi} \left| \boldsymbol{b}_l^* \boldsymbol{v}_3 \right| \leq \left| \sum_{j=1}^m \boldsymbol{b}_l^* \boldsymbol{b}_j \boldsymbol{b}_j^* \boldsymbol{h}^\natural \boldsymbol{x}^{\natural *} \boldsymbol{a}_j \boldsymbol{a}_j^* \boldsymbol{x}^\natural \right| + \left| \sum_{j=1}^m \boldsymbol{b}_l^* \boldsymbol{b}_j \boldsymbol{b}_j^* \boldsymbol{h}^\natural \boldsymbol{x}^{\natural *} \boldsymbol{a}_j \boldsymbol{a}_j^* \left(\tilde{\boldsymbol{x}}^t - \boldsymbol{x}^\natural \right) \right|. \tag{211}$$

Regarding the first term, we have the following lemma.

Lemma 28. *Suppose $m \gg K \log m$. Then with probability at least $1 - O\left(m^{-10}\right)$, one has*

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

For the remaining term, we apply the same strategy as in the previous one (i.e. $\boldsymbol{b}_l^* \boldsymbol{v}_1$) to get

$$\left| \sum_{j=1}^m \boldsymbol{b}_l^* \boldsymbol{b}_j \boldsymbol{b}_j^* \boldsymbol{h}^\natural \boldsymbol{x}^{\natural *} \boldsymbol{a}_j \boldsymbol{a}_j^* \left(\tilde{\boldsymbol{x}}^t - \boldsymbol{x}^\natural \right) \right| \leq \sum_{j=1}^m \left| \boldsymbol{b}_l^* \boldsymbol{b}_j \right| \left\{ \max_{1 \leq j \leq m} \left| \boldsymbol{b}_j^* \boldsymbol{h}^\natural \right| \right\} \left\{ \max_{1 \leq j \leq m} \left| \boldsymbol{a}_j^* \left(\tilde{\boldsymbol{x}}^t - \boldsymbol{x}^\natural \right) \right| \right\} \left\{ \max_{1 \leq j \leq m} \left| \boldsymbol{a}_j^* \boldsymbol{x}^\natural \right| \right\}$$
$$\leq 4 \log m \cdot \frac{\mu}{\sqrt{m}} \cdot C_3 \frac{1}{\log^{1.5} m} \cdot 5\sqrt{\log m}$$
$$\lesssim C_3 \frac{\mu}{\sqrt{m}},$$

where the second relation holds because of the incoherence (34), the induction hypothesis (86c), the condition (184) and Lemma 51. Combining the above three inequalities and the incoherence (34) yields

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

3. Finally, we need to control $|\boldsymbol{b}_l^* \boldsymbol{v}_2|$. For convenience of presentation, we will only bound $|\boldsymbol{b}_1^* \boldsymbol{v}_2|$ in the sequel, but the argument easily extends to all other $\boldsymbol{b}_l$'s. The idea is to separate $\{\boldsymbol{b}_l \mid 1 \leq l \leq m\}$ into groups of

size τ, and to look at each group separately. Here, $\tau \asymp \mathrm{poly}\log(n)$ represents some integer to be specified later. For each $1 < l \leq m - \tau$, it is seen that

$$\begin{aligned}
&\boldsymbol{b}_1^* \sum_{j=1}^{\tau} \boldsymbol{b}_{l+j} \boldsymbol{b}_{l+j}^* \tilde{\boldsymbol{h}}^t \left(\left| \boldsymbol{a}_{l+j}^* \boldsymbol{x}^\natural \right|^2 - \left\| \boldsymbol{x}^\natural \right\|_2^2 \right) \\
&= \boldsymbol{b}_1^* \sum_{j=1}^{\tau} \boldsymbol{b}_{l+1} \boldsymbol{b}_{l+1}^* \tilde{\boldsymbol{h}}^t \left(\left| \boldsymbol{a}_{l+j}^* \boldsymbol{x}^\natural \right|^2 - \left\| \boldsymbol{x}^\natural \right\|_2^2 \right) + \boldsymbol{b}_1^* \sum_{j=1}^{\tau} \left(\boldsymbol{b}_{l+j} \boldsymbol{b}_{l+j}^* - \boldsymbol{b}_{l+1} \boldsymbol{b}_{l+1}^* \right) \tilde{\boldsymbol{h}}^t \left(\left| \boldsymbol{a}_{l+j}^* \boldsymbol{x}^\natural \right|^2 - \left\| \boldsymbol{x}^\natural \right\|_2^2 \right) \\
&= \left\{ \sum_{j=1}^{\tau} \left(\left| \boldsymbol{a}_{l+j}^* \boldsymbol{x}^\natural \right|^2 - \left\| \boldsymbol{x}^\natural \right\|_2^2 \right) \right\} \boldsymbol{b}_1^* \boldsymbol{b}_{l+1} \boldsymbol{b}_{l+1}^* \tilde{\boldsymbol{h}}^t + \boldsymbol{b}_1^* \sum_{j=1}^{\tau} \left(\boldsymbol{b}_{l+j} - \boldsymbol{b}_{l+1} \right) \boldsymbol{b}_{l+j}^* \tilde{\boldsymbol{h}}^t \left(\left| \boldsymbol{a}_{l+j}^* \boldsymbol{x}^\natural \right|^2 - \left\| \boldsymbol{x}^\natural \right\|_2^2 \right) \\
&\quad + \boldsymbol{b}_1^* \sum_{j=1}^{\tau} \boldsymbol{b}_{l+1} \left(\boldsymbol{b}_{l+j} - \boldsymbol{b}_{l+1} \right)^* \tilde{\boldsymbol{h}}^t \left(\left| \boldsymbol{a}_{l+j}^* \boldsymbol{x}^\natural \right|^2 - \left\| \boldsymbol{x}^\natural \right\|_2^2 \right). \qquad (212)
\end{aligned}$$

- Before bounding the first term in (212), we first state the following lemma related to the quantity $\left| \sum_{j=1}^{\tau} \left(|\boldsymbol{a}_{l+j}^* \boldsymbol{x}^\natural|^2 - \|\boldsymbol{x}^\natural\|_2^2 \right) \right|$.

  Lemma 29. *Suppose $\tau \geq C \log m$ for some sufficiently large constant $C > 0$, then*

  $$\left| \sum_{j=1}^{\tau} \left(\left| \boldsymbol{a}_{l+j}^* \boldsymbol{x}^\natural \right|^2 - \|\boldsymbol{x}^\natural\|_2^2 \right) \right| \leq C_1 \sqrt{\tau \log m},$$

  with probability at least $1 - O\left(m^{-10}\right)$. Here $C_1 > 0$ is some absolute constant.

  Notably, this quantity is the pre-factor in the first term of (212), whose fluctuation does not grow fast as it is the sum of i.i.d. random variables over a group. With this result in place, we can bound the first term in (212) as

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

  Taking the summation over all groups gives

  $$\sum_{k=0}^{\lfloor \frac{m}{\tau} \rfloor} \left| \left\{ \sum_{j=1}^{\tau} \left(\left| \boldsymbol{a}_{k\tau+j}^* \boldsymbol{x}^\natural \right|^2 - \|\boldsymbol{x}^\natural\|_2^2 \right) \right\} \boldsymbol{b}_1^* \boldsymbol{b}_{k\tau+1} \boldsymbol{b}_{k\tau+1}^* \tilde{\boldsymbol{h}}^t \right| \lesssim \sqrt{\tau \log m} \sum_{k=0}^{\lfloor \frac{m}{\tau} \rfloor} \left| \boldsymbol{b}_1^* \boldsymbol{b}_{k\tau+1} \right| \max_{1 \leq l \leq m} \left| \boldsymbol{b}_l^* \tilde{\boldsymbol{h}}^t \right|. \qquad (213)$$

  It is straightforward to see from the proof of Lemma 51 that

  $$\sum_{k=0}^{\lfloor \frac{m}{\tau} \rfloor} \left| \boldsymbol{b}_1^* \boldsymbol{b}_{k\tau+1} \right| = \|\boldsymbol{b}_1\|_2^2 + \sum_{k=1}^{\lfloor \frac{m}{\tau} \rfloor} \left| \boldsymbol{b}_1^* \boldsymbol{b}_{k\tau+1} \right| \leq \frac{K}{m} + O\left(\frac{\log m}{\tau} \right). \qquad (214)$$

  Substituting this display (214) into the previous inequality (213) gives

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

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

- The second term of (212) obeys

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

where the first relation is due to Cauchy-Schwarz, while the second one holds because of the following lemma.

Lemma 30. *Suppose $\tau\ge C\log^4 m$ for some sufficiently large constant $C>0$, then there exists some constant $C_1>0$ such that with probability exceeding $1-O\left(m^{-10}\right)$,*

$$\sum_{j=1}^{\tau}\left(\left|\boldsymbol{a}_j^*\boldsymbol{x}^\natural\right|^2-\left\|\boldsymbol{x}^\natural\right\|_2^2\right)^2\le C_1\tau.$$

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

$$\begin{aligned}
&\left|\boldsymbol{b}_1^*\sum_{k=0}^{\lfloor\frac{m}{\tau}\rfloor}\sum_{j=1}^{\tau}\left(\boldsymbol{b}_{k\tau+j}-\boldsymbol{b}_{k\tau+1}\right)\boldsymbol{b}_{k\tau+j}^*\tilde{\boldsymbol{h}}^t\left\{\left|\boldsymbol{a}_{l+j}^*\boldsymbol{x}^\natural\right|^2-\left\|\boldsymbol{x}^\natural\right\|_2^2\right\}\right|\\
&\quad\lesssim\left\{\sqrt{\tau}\sum_{k=0}^{\lfloor\frac{m}{\tau}\rfloor}\sqrt{\sum_{j=1}^{\tau}\left|\boldsymbol{b}_1^*\left(\boldsymbol{b}_{k\tau+j}-\boldsymbol{b}_{k\tau+1}\right)\right|^2}\right\}\max_{1\le l\le m}\left|\boldsymbol{b}_l^*\tilde{\boldsymbol{h}}^t\right|\\
&\quad\le 0.1\max_{1\le l\le m}\left|\boldsymbol{b}_l^*\tilde{\boldsymbol{h}}^t\right|.
\end{aligned}$$

Here, the last line arises from Lemma 54, which says that as long as $m\gg\tau K\log m$,

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

holds for any small constant $c>0$.

- The third term of (212) obeys

$$\begin{aligned}
&\left|\boldsymbol{b}_1^*\sum_{j=1}^{\tau}\boldsymbol{b}_{l+1}\left(\boldsymbol{b}_{l+j}-\boldsymbol{b}_{l+1}\right)^*\tilde{\boldsymbol{h}}^t\left\{\left|\boldsymbol{a}_{l+j}^*\boldsymbol{x}^\natural\right|^2-\left\|\boldsymbol{x}^\natural\right\|_2^2\right\}\right|\\
&\quad\le\left|\boldsymbol{b}_1^*\boldsymbol{b}_{l+1}\right|\left\{\sum_{j=1}^{\tau}\left|\left|\boldsymbol{a}_{l+j}^*\boldsymbol{x}^\natural\right|^2-\left\|\boldsymbol{x}^\natural\right\|_2^2\right|\right\}\max_{1\le l\le m-\tau,\,0\le j<\tau}\left|\left(\boldsymbol{b}_{l+j}-\boldsymbol{b}_l\right)^*\tilde{\boldsymbol{h}}^t\right|\\
&\quad\lesssim\tau\left|\boldsymbol{b}_1^*\boldsymbol{b}_{l+1}\right|\max_{1\le l\le m-\tau,\,0\le j<\tau}\left|\left(\boldsymbol{b}_{l+j}-\boldsymbol{b}_l\right)^*\tilde{\boldsymbol{h}}^t\right|,
\end{aligned}$$

where the last line relies on the inequality

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

owing to Lemma 30 and the Cauchy-Schwarz inequality. Summing over all groups gives

$$\begin{aligned}
&\sum_{k=0}^{\lfloor\frac{m}{\tau}\rfloor}\left|\boldsymbol{b}_1^*\sum_{j=1}^{\tau}\boldsymbol{b}_{k\tau+1}\left(\boldsymbol{b}_{k\tau+j}-\boldsymbol{b}_{k\tau+1}\right)^*\tilde{\boldsymbol{h}}^t\left\{\left|\boldsymbol{a}_{k\tau+j}^*\boldsymbol{x}^\natural\right|^2-\left\|\boldsymbol{x}^\natural\right\|_2^2\right\}\right| \\
&\quad\lesssim \tau\sum_{k=0}^{\lfloor\frac{m}{\tau}\rfloor}\left|\boldsymbol{b}_1^*\boldsymbol{b}_{k\tau+1}\right|\max_{1\le l\le m-\tau,\,0\le j<\tau}\left|\left(\boldsymbol{b}_{l+j}-\boldsymbol{b}_l\right)^*\tilde{\boldsymbol{h}}^t\right| \\
&\quad\lesssim \log m\max_{1\le l\le m-\tau,\,0\le j<\tau}\left|\left(\boldsymbol{b}_{l+j}-\boldsymbol{b}_l\right)^*\tilde{\boldsymbol{h}}^t\right|,
\end{aligned}$$

where the last relation makes use of (214). It then boils down to bounding $\max_{1\le l\le m-\tau,\,0\le j<\tau}\left|\left(\boldsymbol{b}_{l+j}-\boldsymbol{b}_l\right)^*\tilde{\boldsymbol{h}}^t\right|$. Without loss of generality, it suffices to look at $\left|(\boldsymbol{b}_j-\boldsymbol{b}_1)^*\tilde{\boldsymbol{h}}^t\right|$ for all $0<j\le\tau$. Specifically, we claim for the moment that

$$\left|\left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^*\tilde{\boldsymbol{h}}^t\right|\lesssim cC_4\frac{\mu}{\sqrt{m}}\log m,\qquad 0<j\le\tau. \tag{215}$$

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

$$\sum_{k=0}^{\lfloor\frac{m}{\tau}\rfloor}\left|\boldsymbol{b}_1^*\sum_{j=1}^{\tau}\boldsymbol{b}_{k\tau+1}\left(\boldsymbol{b}_{k\tau+j}-\boldsymbol{b}_{k\tau+1}\right)^*\tilde{\boldsymbol{h}}^t\left\{\left|\boldsymbol{a}_{k\tau+j}^*\boldsymbol{x}^\natural\right|^2-\left\|\boldsymbol{x}^\natural\right\|_2^2\right\}\right|\lesssim cC_4\frac{\mu}{\sqrt{m}}\log^2 m.$$

- Putting the above results together, we get

$$\sum_{k=0}^{\lfloor\frac{m}{\tau}\rfloor}\left|\boldsymbol{b}_1^*\sum_{j=1}^{\tau}\boldsymbol{b}_{k\tau+j}\boldsymbol{b}_{k\tau+j}^*\tilde{\boldsymbol{h}}^t\left\{\left|\boldsymbol{a}_{k\tau+j}^*\boldsymbol{x}^\natural\right|^2-\left\|\boldsymbol{x}^\natural\right\|_2^2\right\}\right|\le 0.2\max_{1\le l\le m}\left|\boldsymbol{b}_l^*\tilde{\boldsymbol{h}}^t\right|+O\left(cC_4\frac{\mu}{\sqrt{m}}\log^2 m\right).$$

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

$$\begin{aligned}
\left|\boldsymbol{b}_l^*\tilde{\boldsymbol{h}}^{t+1}\right| &\le (1+\delta)\left\{(1-\eta\xi)\left|\boldsymbol{b}_l^*\tilde{\boldsymbol{h}}^t\right|+0.3\eta\xi\max_{1\le l\le m}\left|\boldsymbol{b}_l^*\tilde{\boldsymbol{h}}^t\right|+C_8C_3\eta\xi\frac{\mu}{\sqrt{m}}+C_8\eta\xi cC_4\frac{\mu}{\sqrt{m}}\log^2 m\right\} \\
&\overset{\text{(i)}}{\le}\left(1+O\left(\frac{1}{\log^2 m}\right)\right)\left\{(1-0.7\eta\xi)\,C_4\frac{\mu}{\sqrt{m}}\log^2 m+C_8C_3\eta\xi\frac{\mu}{\sqrt{m}}+C_8\eta\xi cC_4\frac{\mu}{\sqrt{m}}\log^2 m\right\} \\
&\overset{\text{(ii)}}{\le} C_4\frac{\mu}{\sqrt{m}}\log^2 m.
\end{aligned}$$

Here, (i) uses the induction hypothesis (86d), 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

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

Finally, it remains to verify the claim (215), which we accomplish as follows.

Proof of Claim (215). We will prove this claim by induction. Again, observe that

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

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

Use the gradient update rule for $\boldsymbol{h}^t$ (cf. (77)) once again to get

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

where we denote

$$\tilde{\boldsymbol{h}}^{t-1} := \frac{1}{\overline{\alpha^{t-1}}}\boldsymbol{h}^{t-1} \qquad \tilde{\boldsymbol{x}}^{t-1} := \alpha^{t-1}\boldsymbol{x}^{t-1} \qquad \text{and} \qquad \theta := \frac{1}{\|\tilde{\boldsymbol{x}}^{t-1}\|_2^2}.$$

This further gives rise to

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

For β_1, one can get

$$\frac{1}{\eta\theta}|\beta_1| \le \left|(\boldsymbol{b}_j - \boldsymbol{b}_1)^*\boldsymbol{h}^\natural\right|\|\boldsymbol{x}^\natural\|_2\|\tilde{\boldsymbol{x}}^{t-1}\|_2 \le 4\frac{\mu}{\sqrt{m}},$$

where we utilize the incoherence condition (34) and the fact that $\tilde{\boldsymbol{x}}^{t-1}$ and $\boldsymbol{x}^\natural$ are extremely close, i.e.

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

Regarding the second term β_2, we have

$$\frac{1}{\eta\theta}|\beta_2| \le \left\{\sum_{l=1}^m \left|(\boldsymbol{b}_j - \boldsymbol{b}_1)^*\boldsymbol{b}_l\right|\right\}\underbrace{\max_{1\le l\le m}\left|\boldsymbol{b}_l^*\left(\tilde{\boldsymbol{h}}^{t-1}(\tilde{\boldsymbol{x}}^{t-1})^* - \boldsymbol{h}^\natural\boldsymbol{x}^{\natural*}\right)(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})\tilde{\boldsymbol{x}}^{t-1}\right|}_{:=\psi}.$$

The term ψ can be bounded as follows

$$\begin{aligned}
\psi &\le \max_{1\le l\le m}\left|\boldsymbol{b}_l^*\tilde{\boldsymbol{h}}^{t-1}(\tilde{\boldsymbol{x}}^{t-1})^*(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})\tilde{\boldsymbol{x}}^{t-1}\right| + \max_{1\le l\le m}\left|\boldsymbol{b}_l^*\boldsymbol{h}^\natural\boldsymbol{x}^{\natural*}(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})\tilde{\boldsymbol{x}}^{t-1}\right|\\
&\le \max_{1\le l\le m}\left|\boldsymbol{b}_l^*\tilde{\boldsymbol{h}}^{t-1}\right|\max_{1\le l\le m}\left|(\tilde{\boldsymbol{x}}^{t-1})^*(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})\tilde{\boldsymbol{x}}^{t-1}\right|\\
&\qquad + \max_{1\le l\le m}\left|\boldsymbol{b}_l^*\boldsymbol{h}^\natural\right|\max_{1\le l\le m}\left|\boldsymbol{x}^{\natural*}(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})\tilde{\boldsymbol{x}}^{t-1}\right|\\
&\lesssim \log m\left\{\max_{1\le l\le m}\left|\boldsymbol{b}_l^*\tilde{\boldsymbol{h}}^{t-1}\right| + \frac{\mu}{\sqrt{m}}\right\}.
\end{aligned}$$

Here, we have used the fact that

$$\left|(\tilde{\boldsymbol{x}}^{t-1})^*(\boldsymbol{a}_l\boldsymbol{a}_l^* - \boldsymbol{I})\tilde{\boldsymbol{x}}^{t-1}\right| \le \left\|\boldsymbol{a}_l^*\tilde{\boldsymbol{x}}^{t-1}\right\|_2^2 + \left\|\tilde{\boldsymbol{x}}^{t-1}\right\|_2^2 \lesssim \log m,$$

$$\left|\boldsymbol{x}^{\natural*}\left(\boldsymbol{a}_l\boldsymbol{a}_l^*-\boldsymbol{I}\right)\tilde{\boldsymbol{x}}^{t-1}\right|\leq\left\|\boldsymbol{a}_l^*\tilde{\boldsymbol{x}}^{t-1}\right\|_2\left\|\boldsymbol{a}_l^*\boldsymbol{x}^{\natural}\right\|_2+\left\|\tilde{\boldsymbol{x}}^{t-1}\right\|_2\|\boldsymbol{x}^{\natural}\|_2\lesssim\log m,$$

which are immediate consequences of (86c) and (184). Combining this with Lemma 53, we see that there exists some sufficiently small constant $c>0$ such that

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

as long as $m\gtrsim\tau K\log^4 m$.

To summarize, we arrive at

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

Making use of the induction hypothesis (81a) and the fact that $\left\|\tilde{\boldsymbol{x}}^{t-1}\right\|_2^2\geq 0.9$, we can reach

$$\left|\left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^*\tilde{\boldsymbol{h}}^t\right|\leq(1+\delta)\left\{(1-0.9\eta\theta)\left|\left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^*\tilde{\boldsymbol{h}}^{t-1}\right|+cC_4\eta\theta\frac{\mu}{\sqrt{m}}\log m+\frac{c\mu}{\sqrt{m}\log m}\right\}.$$

Recall that $\delta\asymp 1/\log^2 m$. As a result, if $\eta>0$ is some sufficiently small constant and if

$$\left|\left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^*\tilde{\boldsymbol{h}}^{t-1}\right|\leq 10c\left(C_4\frac{\mu}{\sqrt{m}}\log m+\frac{\mu}{\eta\theta\sqrt{m}\log m}\right)\leq 20cC_4\frac{\mu}{\sqrt{m}}\log m$$

hold, then

$$\left|\left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^*\tilde{\boldsymbol{h}}^t\right|\leq 20cC_4\frac{\mu}{\sqrt{m}}\log m.$$

Therefore, this concludes the proof of the claim (215) by induction, provided that the base case is true, i.e. for some $c>0$ sufficiently small

$$\left|\left(\boldsymbol{b}_j-\boldsymbol{b}_1\right)^*\tilde{\boldsymbol{h}}^0\right|\leq 20cC_4\frac{\mu}{\sqrt{m}}\log m. \tag{216}$$

We defer the proof of (216) to Appendix C.6 (see Lemma 32). □

Proof of Lemma 28. Denote

$$w_j=\boldsymbol{b}_l^*\boldsymbol{b}_j\boldsymbol{b}_j^*\boldsymbol{h}^{\natural}\boldsymbol{x}^{\natural*}\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x}^{\natural}.$$

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

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

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

$$\left\|w_j-\mathbb{E}\left[w_j\right]\right\|_{\psi_1}\leq 2\left\|w_j\right\|_{\psi_1}\leq 4\left|\boldsymbol{b}_l^*\boldsymbol{b}_j\right|\left|\boldsymbol{b}_j^*\boldsymbol{h}^{\natural}\right|\left\|\boldsymbol{a}_j^*\boldsymbol{x}^{\natural}\right\|_{\psi_2}^2\lesssim\left|\boldsymbol{b}_l^*\boldsymbol{b}_j\right|\frac{\mu}{\sqrt{m}},$$

where we utilize the incoherence assumption (34) and the fact that $\left\|\boldsymbol{a}_j^*\boldsymbol{x}^{\natural}\right\|_{\psi_2}\lesssim 1$. Let

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

which follows since $\sum_j\left|\boldsymbol{b}_l^*\boldsymbol{b}_j\right|^2=\boldsymbol{b}_l^*\left(\sum_j\boldsymbol{b}_j\boldsymbol{b}_j^*\right)\boldsymbol{b}_l=\|\boldsymbol{b}_l\|_2^2=K/m$. We can invoke [FWWZ17, Corollary 1] to obtain

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{j=1}^m\left(w_j-\mathbb{E}\left[w_j\right]\right)\right|\geq t\right)\leq\exp\left(1-\frac{m}{8}\min\left\{\frac{t}{2V},\left(\frac{t}{2V}\right)^2\right\}\right).$$

By taking $t = 2\epsilon V$ for $\epsilon \in (0,1)$, we see that with probability at least $1 - \exp\left(1 - m\epsilon^2/8\right)$,

$$\left|\sum_{j=1}^{m} \boldsymbol{b}_l^* \boldsymbol{b}_j \boldsymbol{b}_j^* \boldsymbol{h}^\natural \boldsymbol{x}^{\natural *} \boldsymbol{a}_j \boldsymbol{a}_j^* \boldsymbol{x}^\natural - \boldsymbol{b}_l^* \boldsymbol{h}^\natural\right| \le 2\epsilon V m \lesssim \epsilon \sqrt{K} \frac{\mu}{\sqrt{m}}.$$

Finally, picking $\epsilon \asymp 1/\sqrt{K}$, we conclude that with probability at least $1 - \exp\left(1 - cm/K\right)$ for some constant $c > 0$,

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

We conclude the proof by observing that $m \gg K \log m$. $\square$

Proof of Lemma 29. By unitary invariance, we may take $\boldsymbol{x}^\natural = \boldsymbol{e}_1$. Then $2\left|\boldsymbol{a}_j^* \boldsymbol{x}^\natural\right|^2$ follows a χ_2^2 distribution. Standard χ^2 concentration reveals

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

with probability exceeding $1 - O\left(m^{-10}\right)$. $\square$

Proof of Lemma 30. From the elementary inequality $(a+b)^2 \le 2\left(a^2 + b^2\right)$, we see that

$$\sum_{j=1}^{\tau}\left(\left|\boldsymbol{a}_j^* \boldsymbol{x}^\natural\right|^2 - \left\|\boldsymbol{x}^\natural\right\|_2^2\right)^2 \le 2\sum_{j=1}^{\tau}\left(\left|\boldsymbol{a}_j^* \boldsymbol{x}^\natural\right|^4 + \left\|\boldsymbol{x}^\natural\right\|_2^4\right) = 2\sum_{j=1}^{\tau}\left|\boldsymbol{a}_j^* \boldsymbol{x}^\natural\right|^4 + 2\tau, \tag{217}$$

where the last identity holds true since $\left\|\boldsymbol{x}^\natural\right\|_2 = 1$. It thus suffices to control $\sum_{j=1}^{\tau}\left|\boldsymbol{a}_j^* \boldsymbol{x}^\natural\right|^4$. Let $\xi_i = \boldsymbol{a}_j^* \boldsymbol{x}^\natural$, 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 $1 - O(m^{-10})$,

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

which together with (217) concludes the proof. $\square$

C.5 Proof of Lemma 18

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

$$\min_{\alpha\in\mathbb{C},|\alpha|=1}\left\{\left\|\alpha\check{\boldsymbol{h}}^0-\boldsymbol{h}^\natural\right\|_2+\left\|\alpha\check{\boldsymbol{x}}^0-\boldsymbol{x}^\natural\right\|_2\right\}\le\frac{c_1\left\|\boldsymbol{M}-\mathbb{E}\left[\boldsymbol{M}\right]\right\|}{\sigma_1\left(\mathbb{E}\left[\boldsymbol{M}\right]\right)-\sigma_2\left(\boldsymbol{M}\right)},\tag{218}$$

for some universal constant $c_1>0$. Regarding the numerator in (218), it has been shown in [LLSW16, Lemma 5.20] that for any $\xi>0$,

$$\left\|\boldsymbol{M}-\mathbb{E}\left[\boldsymbol{M}\right]\right\|\le\xi\tag{219}$$

with probability exceeding $1-O(m^{-10})$, provided that

$$m\ge\frac{c_2\mu^2K\log^2 m}{\xi^2}$$

for some universal constant $c_2>0$. For the denominator in (218), we can take (219) together with Weyl's inequality to demonstrate that

$$\begin{aligned}\sigma_1\left(\mathbb{E}\left[\boldsymbol{M}\right]\right)-\sigma_2\left(\boldsymbol{M}\right)&\ge\sigma_1\left(\mathbb{E}\left[\boldsymbol{M}\right]\right)-\sigma_2\left(\mathbb{E}\left[\boldsymbol{M}\right]\right)-\left\|\boldsymbol{M}-\mathbb{E}\left[\boldsymbol{M}\right]\right\|\\&\ge1-\xi,\end{aligned}$$

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 (218) reveal that

$$\min_{\alpha\in\mathbb{C},|\alpha|=1}\left\{\left\|\alpha\check{\boldsymbol{h}}^0-\boldsymbol{h}^\natural\right\|_2+\left\|\alpha\check{\boldsymbol{x}}^0-\boldsymbol{x}^\natural\right\|_2\right\}\le\frac{c_1\xi}{1-\xi}\le2c_1\xi\tag{220}$$

as long as $\xi<1/2$.

The preceding bound is concerned with the singular vectors $\check{\boldsymbol{h}}^0$ and $\check{\boldsymbol{x}}^0$. In order to further connect them with the scaled singular vectors $\boldsymbol{h}^0=\sqrt{\sigma_1\left(\boldsymbol{M}\right)}\,\check{\boldsymbol{h}}^0$ and $\boldsymbol{x}^0=\sqrt{\sigma_1\left(\boldsymbol{M}\right)}\,\check{\boldsymbol{x}}^0$, we resort to the following lemma.

***Lemma* 31.** *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 $\sigma_1(\tilde{\boldsymbol{A}})$, $\tilde{\boldsymbol{u}}$ and $\tilde{\boldsymbol{v}}$ be the leading singular value, left and right singular vectors of $\tilde{\boldsymbol{A}}$, respectively. Then,*

$$\left|\sigma_1(\boldsymbol{A})-\sigma_1(\tilde{\boldsymbol{A}})\right|\le\left\|(\boldsymbol{A}-\tilde{\boldsymbol{A}})\boldsymbol{v}\right\|_2+\left(\|\boldsymbol{u}-\tilde{\boldsymbol{u}}\|_2+\|\boldsymbol{v}-\tilde{\boldsymbol{v}}\|_2\right)\|\tilde{\boldsymbol{A}}\|;$$

$$\left\|\sqrt{\sigma_1(\boldsymbol{A})}\boldsymbol{u}-\sqrt{\sigma_1(\tilde{\boldsymbol{A}})}\tilde{\boldsymbol{u}}\right\|_2+\left\|\sqrt{\sigma_1(\boldsymbol{A})}\boldsymbol{v}-\sqrt{\sigma_1(\tilde{\boldsymbol{A}})}\tilde{\boldsymbol{v}}\right\|_2\le\sqrt{\sigma_1(\boldsymbol{A})}\left(\|\boldsymbol{u}-\tilde{\boldsymbol{u}}\|_2+\|\boldsymbol{v}-\tilde{\boldsymbol{v}}\|_2\right)+\frac{2\left|\sigma_1(\boldsymbol{A})-\sigma_1(\tilde{\boldsymbol{A}})\right|}{\sqrt{\sigma_1(\boldsymbol{A})}+\sqrt{\sigma_1(\tilde{\boldsymbol{A}})}}.$$

Observe that for any $\alpha\in\mathbb{C}$ with $|\alpha|=1$, by the definition of $\boldsymbol{h}^0$ and $\boldsymbol{x}^0$ we have

$$\left\|\alpha\boldsymbol{h}^0-\boldsymbol{h}^\natural\right\|_2+\left\|\alpha\boldsymbol{x}^0-\boldsymbol{x}^\natural\right\|_2=\left\|\sqrt{\sigma_1\left(\boldsymbol{M}\right)}\left(\alpha\check{\boldsymbol{h}}^0\right)-\boldsymbol{h}^\natural\right\|_2+\left\|\sqrt{\sigma_1\left(\boldsymbol{M}\right)}\left(\alpha\check{\boldsymbol{x}}^0\right)-\boldsymbol{x}^\natural\right\|_2.$$

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 31 to get

$$\begin{aligned}\left\|\alpha\boldsymbol{h}^0-\boldsymbol{h}^\natural\right\|_2+\left\|\alpha\boldsymbol{x}^0-\boldsymbol{x}^\natural\right\|_2&\le\sqrt{\sigma_1(\mathbb{E}[\boldsymbol{M}])}\left(\left\|\alpha\check{\boldsymbol{h}}^0-\boldsymbol{h}^\natural\right\|_2+\left\|\alpha\check{\boldsymbol{x}}^0-\boldsymbol{x}^\natural\right\|_2\right)+\frac{2\left|\sigma_1(\boldsymbol{M})-\sigma_1\left(\mathbb{E}\left[\boldsymbol{M}\right]\right)\right|}{\sqrt{\sigma_1(\boldsymbol{M})}+\sqrt{\sigma_1\left(\mathbb{E}\left[\boldsymbol{M}\right]\right)}}\\&=\left\|\alpha\check{\boldsymbol{h}}^0-\boldsymbol{h}^\natural\right\|_2+\left\|\alpha\check{\boldsymbol{x}}^0-\boldsymbol{x}^\natural\right\|_2+\frac{2\left|\sigma_1(\boldsymbol{M})-1\right|}{\sqrt{\sigma_1(\boldsymbol{M})}+1}.\end{aligned}\tag{221}$$

In addition, we can use Weyl's inequality to deduce that

$$\left|\sigma_1(\boldsymbol{M})-1\right|=\left|\sigma_1(\boldsymbol{M})-\sigma_1(\mathbb{E}[\boldsymbol{M}])\right|\le\left\|\boldsymbol{M}-\mathbb{E}[\boldsymbol{M}]\right\|\le\xi,\tag{222}$$

where the last inequality comes from (219). Substitute (222) back into (221) to obtain

$$\left\|\alpha \boldsymbol{h}^0 - \boldsymbol{h}^\natural\right\|_2 + \left\|\alpha \boldsymbol{x}^0 - \boldsymbol{x}^\natural\right\|_2 \le \left\|\alpha \check{\boldsymbol{h}}^0 - \boldsymbol{h}^\natural\right\|_2 + \left\|\alpha \check{\boldsymbol{x}}^0 - \boldsymbol{x}^\natural\right\|_2 + 2\xi. \tag{223}$$

Taking the minimum over α, one can thus conclude that

$$\min_{\alpha\in\mathbb{C},|\alpha|=1} \left\{\left\|\alpha \boldsymbol{h}^0 - \boldsymbol{h}^\natural\right\|_2 + \left\|\alpha \boldsymbol{x}^0 - \boldsymbol{x}^\natural\right\|_2\right\} \le \min_{\alpha\in\mathbb{C},|\alpha|=1} \left\{\left\|\alpha \check{\boldsymbol{h}}^0 - \boldsymbol{h}^\natural\right\|_2 + \left\|\alpha \check{\boldsymbol{x}}^0 - \boldsymbol{x}^\natural\right\|_2\right\} + 2\xi \le 2c_1\xi + 2\xi,$$

where the last inequality comes from (220). Since ξ is arbitrary, this finishes the proof for (89).

Carrying out similar arguments (which we omit here), we can also establish (90).

Proof of Lemma 31. The first claim follows since

$$\begin{aligned}\left|\sigma_1(\boldsymbol{A}) - \sigma_1(\tilde{\boldsymbol{A}})\right| &= \left|\boldsymbol{u}^*\boldsymbol{A}\boldsymbol{v} - \tilde{\boldsymbol{u}}^*\tilde{\boldsymbol{A}}\tilde{\boldsymbol{v}}\right| \\ &\le \left|\boldsymbol{u}^*\big(\boldsymbol{A} - \tilde{\boldsymbol{A}}\big)\boldsymbol{v}\right| + \left|\boldsymbol{u}^*\tilde{\boldsymbol{A}}\boldsymbol{v} - \tilde{\boldsymbol{u}}^*\tilde{\boldsymbol{A}}\boldsymbol{v}\right| + \left|\tilde{\boldsymbol{u}}^*\tilde{\boldsymbol{A}}\boldsymbol{v} - \tilde{\boldsymbol{u}}^*\tilde{\boldsymbol{A}}\tilde{\boldsymbol{v}}\right| \\ &\le \left\|\big(\boldsymbol{A} - \tilde{\boldsymbol{A}}\big)\boldsymbol{v}\right\|_2 + \|\boldsymbol{u} - \tilde{\boldsymbol{u}}\|_2\,\|\tilde{\boldsymbol{A}}\| + \|\tilde{\boldsymbol{A}}\|\,\|\boldsymbol{v} - \tilde{\boldsymbol{v}}\|_2\,. \end{aligned} \tag{224}$$

With regards to the second claim, we see that

$$\begin{aligned}\left\|\sqrt{\sigma_1(\boldsymbol{A})}\boldsymbol{u} - \sqrt{\sigma_1(\tilde{\boldsymbol{A}})}\tilde{\boldsymbol{u}}\right\|_2 &\le \left\|\sqrt{\sigma_1(\boldsymbol{A})}\boldsymbol{u} - \sqrt{\sigma_1(\boldsymbol{A})}\tilde{\boldsymbol{u}}\right\|_2 + \left\|\sqrt{\sigma_1(\boldsymbol{A})}\tilde{\boldsymbol{u}} - \sqrt{\sigma_1(\tilde{\boldsymbol{A}})}\tilde{\boldsymbol{u}}\right\|_2 \\ &= \sqrt{\sigma_1(\boldsymbol{A})}\,\|\boldsymbol{u} - \tilde{\boldsymbol{u}}\|_2 + \left|\sqrt{\sigma_1(\boldsymbol{A})} - \sqrt{\sigma_1(\tilde{\boldsymbol{A}})}\right| \\ &= \sqrt{\sigma_1(\boldsymbol{A})}\,\|\boldsymbol{u} - \tilde{\boldsymbol{u}}\|_2 + \frac{\left|\sigma_1(\boldsymbol{A}) - \sigma_1(\tilde{\boldsymbol{A}})\right|}{\sqrt{\sigma_1(\boldsymbol{A})} + \sqrt{\sigma_1(\tilde{\boldsymbol{A}})}}.\end{aligned}$$

Similarly, one can obtain

$$\left\|\sqrt{\sigma_1(\boldsymbol{A})}\boldsymbol{v} - \sqrt{\sigma_1(\tilde{\boldsymbol{A}})}\tilde{\boldsymbol{v}}\right\|_2 \le \sqrt{\sigma_1(\boldsymbol{A})}\,\|\boldsymbol{v} - \tilde{\boldsymbol{v}}\|_2 + \frac{\left|\sigma_1(\boldsymbol{A}) - \sigma_1(\tilde{\boldsymbol{A}})\right|}{\sqrt{\sigma_1(\boldsymbol{A})} + \sqrt{\sigma_1(\tilde{\boldsymbol{A}})}}.$$

Add these two inequalities to complete the proof. □

C.6 Proof of Lemma 19

The proof is composed of three steps:

- In the first step, we aim to show that the normalized singular vectors of $\boldsymbol{M}$ and $\boldsymbol{M}^{(l)}$ are close enough; see (232).
- We proceed by passing this proximity result to the scaled singular vectors (see (235)).
- In the end, we translate the usual ℓ_2 distance metric to the distance function we defined in (32); see (237). Along the way, we also prove the incoherence of with respect to $\{\boldsymbol{b}_j\}$.

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

$$\min_{\alpha\in\mathbb{C},|\alpha|=1} \left\{\left\|\alpha \check{\boldsymbol{h}}^0 - \check{\boldsymbol{h}}^{0,(l)}\right\|_2 + \left\|\alpha \check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(l)}\right\|_2\right\} \le c_1 \frac{\left\|\big(\boldsymbol{M} - \boldsymbol{M}^{(l)}\big)\check{\boldsymbol{x}}^{0,(l)}\right\|_2 + \left\|\check{\boldsymbol{h}}^{0,(l)*}\big(\boldsymbol{M} - \boldsymbol{M}^{(l)}\big)\right\|_2}{\sigma_1\big(\boldsymbol{M}^{(l)}\big) - \sigma_2\left(\boldsymbol{M}\right)}$$

for some constant $c_1 > 0$. Using the Weyl's inequality we get

$$\begin{aligned}\sigma_1\big(\boldsymbol{M}^{(l)}\big) - \sigma_2\left(\boldsymbol{M}\right) &\geq \sigma_1\big(\mathbb{E}[\boldsymbol{M}^{(l)}]\big) - \|\boldsymbol{M}^{(l)} - \mathbb{E}[\boldsymbol{M}^{(l)}]\| - \sigma_2\left(\mathbb{E}[\boldsymbol{M}]\right) - \|\boldsymbol{M} - \mathbb{E}[\boldsymbol{M}]\| \\ &\geq 3/4 - \|\boldsymbol{M}^{(l)} - \mathbb{E}[\boldsymbol{M}^{(l)}]\| - \|\boldsymbol{M} - \mathbb{E}[\boldsymbol{M}]\| \geq 1/2,\end{aligned}$$

where the penultimate inequality follows from

$$\sigma_1\big(\mathbb{E}[\boldsymbol{M}^{(l)}]\big) \geq 3/4$$

for m sufficiently large, and the last inequality comes from [LLSW16, 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

$$\beta^{0,(l)} := \arg\min_{\alpha\in\mathbb{C},|\alpha|=1} \left\|\alpha\check{\boldsymbol{h}}^0 - \check{\boldsymbol{h}}^{0,(l)}\right\|_2 + \left\|\alpha\check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(l)}\right\|_2 \tag{225}$$

allows us to obtain

$$\left\|\beta^{0,(l)}\check{\boldsymbol{h}}^0 - \check{\boldsymbol{h}}^{0,(l)}\right\|_2 + \left\|\beta^{0,(l)}\check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(l)}\right\|_2 \leq 2c_1\left\{\left\|\big(\boldsymbol{M} - \boldsymbol{M}^{(l)}\big)\check{\boldsymbol{x}}^{0,(l)}\right\|_2 + \left\|\check{\boldsymbol{h}}^{0,(l)*}\big(\boldsymbol{M} - \boldsymbol{M}^{(l)}\big)\right\|_2\right\}. \tag{226}$$

It then boils down to controlling the two terms on the right-hand side of (226). By construction,

$$\boldsymbol{M} - \boldsymbol{M}^{(l)} = \boldsymbol{b}_l\boldsymbol{b}_l^*\boldsymbol{h}^\natural\boldsymbol{x}^{\natural*}\boldsymbol{a}_l\boldsymbol{a}_l^*.$$

- On the one hand,

$$\begin{aligned}\left\|\big(\boldsymbol{M} - \boldsymbol{M}^{(l)}\big)\check{\boldsymbol{x}}^{0,(l)}\right\|_2 &= \left\|\boldsymbol{b}_l\boldsymbol{b}_l^*\boldsymbol{h}^\natural\boldsymbol{x}^{\natural*}\boldsymbol{a}_l\boldsymbol{a}_l^*\check{\boldsymbol{x}}^{0,(l)}\right\|_2 = \|\boldsymbol{b}_l\|_2\left|\boldsymbol{b}_l^*\boldsymbol{h}^\natural\right|\left|\boldsymbol{a}_l^*\boldsymbol{x}^\natural\right| \cdot \left|\boldsymbol{a}_l^*\check{\boldsymbol{x}}^{0,(l)}\right| \\ &\leq 30\frac{\mu}{\sqrt{m}} \cdot \sqrt{\frac{K\log^2 m}{m}},\end{aligned} \tag{227}$$

  where we use the fact that $\|\boldsymbol{b}_l\|_2 = \sqrt{K/m}$, the incoherence condition (34), the bound (184) and the fact that with probability exceeding $1 - O\left(m^{-10}\right)$,

$$\max_{1\leq l\leq m}\left|\boldsymbol{a}_l^*\check{\boldsymbol{x}}^{0,(l)}\right| \leq 5\sqrt{\log m},$$

  which originates from the independence between $\check{\boldsymbol{x}}^{0,(l)}$ and $\boldsymbol{a}_l$.

- On the other hand, for any $\tilde{\alpha}$ obeying $|\tilde{\alpha}| = 1$ one has

$$\begin{aligned}\left\|\check{\boldsymbol{h}}^{0,(l)*}\big(\boldsymbol{M} - \boldsymbol{M}^{(l)}\big)\right\|_2 &= \left\|\check{\boldsymbol{h}}^{0,(l)*}\boldsymbol{b}_l\boldsymbol{b}_l^*\boldsymbol{h}^\natural\boldsymbol{x}^{\natural*}\boldsymbol{a}_l\boldsymbol{a}_l^*\right\|_2 = \|\boldsymbol{a}_l\|_2\left|\boldsymbol{b}_l^*\boldsymbol{h}^\natural\right|\left|\boldsymbol{a}_l^*\boldsymbol{x}^\natural\right| \cdot \left|\boldsymbol{b}_l^*\check{\boldsymbol{h}}^{0,(l)}\right| \\ &\overset{\text{(i)}}{\leq} 3\sqrt{K} \cdot \frac{\mu}{\sqrt{m}} \cdot 5\sqrt{\log m} \cdot \left|\boldsymbol{b}_l^*\check{\boldsymbol{h}}^{0,(l)}\right| \\ &\overset{\text{(ii)}}{\leq} 15\sqrt{\frac{\mu^2 K\log m}{m}}\left|\tilde{\alpha}\boldsymbol{b}_l^*\check{\boldsymbol{h}}^0\right| + 15\sqrt{\frac{\mu^2 K\log m}{m}}\left|\boldsymbol{b}_l^*\left(\tilde{\alpha}\check{\boldsymbol{h}}^0 - \check{\boldsymbol{h}}^{0,(l)}\right)\right| \\ &\overset{\text{(iii)}}{\leq} 15\sqrt{\frac{\mu^2 K\log m}{m}}\left|\boldsymbol{b}_l^*\check{\boldsymbol{h}}^0\right| + 15\sqrt{\frac{\mu^2 K\log m}{m}} \cdot \sqrt{\frac{K}{m}}\left\|\tilde{\alpha}\check{\boldsymbol{h}}^0 - \check{\boldsymbol{h}}^{0,(l)}\right\|_2.\end{aligned}$$

  Here, (i) arises from the incoherence condition (34) together with the bounds (184) and (185), the inequality (ii) comes from the triangle inequality, and the last line (iii) holds since $\|\boldsymbol{b}_l\|_2 = \sqrt{K/m}$ and $|\tilde{\alpha}| = 1$.

Substitution into (226) yields

$$\left\|\beta^{0,(l)}\check{\boldsymbol{h}}^0 - \check{\boldsymbol{h}}^{0,(l)}\right\|_2 + \left\|\beta^{0,(l)}\check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(l)}\right\|_2$$

$$\leq 2c_1 \left\{ 30\frac{\mu}{\sqrt{m}} \cdot \sqrt{\frac{K\log^2 m}{m}} + 15\sqrt{\frac{\mu^2 K\log m}{m}} \left|\boldsymbol{b}_l^* \check{\boldsymbol{h}}^0\right| + 15\sqrt{\frac{\mu^2 K\log m}{m}} \cdot \sqrt{\frac{K}{m}} \left\|\tilde{\alpha}\check{\boldsymbol{h}}^0 - \check{\boldsymbol{h}}^{0,(l)}\right\|_2 \right\}.$$

Since the previous inequality holds for all $|\tilde{\alpha}| = 1$, we can choose $\tilde{\alpha} = \beta^{0,(l)}$ and rearrange terms to get

$$\begin{aligned}
&\left(1 - 30c_1\sqrt{\frac{\mu^2 K\log m}{m}}\sqrt{\frac{K}{m}}\right)\left(\left\|\beta^{0,(l)}\check{\boldsymbol{h}}^0 - \check{\boldsymbol{h}}^{0,(l)}\right\|_2 + \left\|\beta^{0,(l)}\check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(l)}\right\|_2\right) \\
&\quad \leq 60c_1\frac{\mu}{\sqrt{m}} \cdot \sqrt{\frac{K\log^2 m}{m}} + 30c_1\sqrt{\frac{\mu^2 K\log m}{m}} \left|\boldsymbol{b}_l^* \check{\boldsymbol{h}}^0\right|.
\end{aligned}$$

Under the condition that $m \gg \mu K \log^{1/2} m$, one has $1 - 30c_1\sqrt{\frac{\mu^2 K\log m}{m}}\sqrt{\frac{K}{m}} \geq \frac{1}{2}$, and therefore

$$\left\|\beta^{0,(l)}\check{\boldsymbol{h}}^0 - \check{\boldsymbol{h}}^{0,(l)}\right\|_2 + \left\|\beta^{0,(l)}\check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(l)}\right\|_2 \leq 120c_1\frac{\mu}{\sqrt{m}} \cdot \sqrt{\frac{K\log^2 m}{m}} + 60c_1\sqrt{\frac{\mu^2 K\log m}{m}} \left|\boldsymbol{b}_l^* \check{\boldsymbol{h}}^0\right|,$$

which immediately implies that

$$\begin{aligned}
&\max_{1\leq l\leq m}\left\{\left\|\beta^{0,(l)}\check{\boldsymbol{h}}^0 - \check{\boldsymbol{h}}^{0,(l)}\right\|_2 + \left\|\beta^{0,(l)}\check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(l)}\right\|_2\right\} \\
&\quad \leq 120c_1\frac{\mu}{\sqrt{m}} \cdot \sqrt{\frac{K\log^2 m}{m}} + 60c_1\sqrt{\frac{\mu^2 K\log m}{m}} \max_{1\leq l\leq m}\left|\boldsymbol{b}_l^* \check{\boldsymbol{h}}^0\right|.
\end{aligned} \tag{228}$$

We move on to $\left|\boldsymbol{b}_l^* \check{\boldsymbol{h}}^0\right|$, and aim to show that $\max_{1\leq l\leq m}\left|\boldsymbol{b}_l^* \check{\boldsymbol{h}}^0\right|$ can also be upper bounded by the left-hand side of (228). By construction, we have

$$\boldsymbol{M}\check{\boldsymbol{x}}^0 = \sigma_1(\boldsymbol{M})\check{\boldsymbol{h}}^0,$$

which further leads to

$$\begin{aligned}
\left|\boldsymbol{b}_l^* \check{\boldsymbol{h}}^0\right| &= \frac{1}{\sigma_1(\boldsymbol{M})}\left|\boldsymbol{b}_l^*\boldsymbol{M}\check{\boldsymbol{x}}^0\right| \\
&\overset{\text{(i)}}{\leq} 2\left|\sum_{j=1}^m (\boldsymbol{b}_l^*\boldsymbol{b}_j)\boldsymbol{b}_j^*\boldsymbol{h}^\natural\boldsymbol{x}^{\natural *}\boldsymbol{a}_j\boldsymbol{a}_j^*\check{\boldsymbol{x}}^0\right| \\
&\leq 2\left(\sum_{j=1}^m |\boldsymbol{b}_l^*\boldsymbol{b}_j|\right)\max_{1\leq j\leq m}\left\{\left|\boldsymbol{b}_j^*\boldsymbol{h}^\natural\right|\left|\boldsymbol{a}_j^*\boldsymbol{x}^\natural\right|\left|\boldsymbol{a}_j^*\check{\boldsymbol{x}}^0\right|\right\} \\
&\overset{\text{(ii)}}{\leq} 8\log m \cdot \frac{\mu}{\sqrt{m}} \cdot \left(5\sqrt{\log m}\right)\max_{1\leq j\leq m}\left\{\left|\boldsymbol{a}_j^*\check{\boldsymbol{x}}^{0,(j)}\right| + \|\boldsymbol{a}_j\|_2\left\|\beta^{0,(j)}\check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(j)}\right\|_2\right\} \\
&\leq 200\frac{\mu\log^2 m}{\sqrt{m}} + 120\sqrt{\frac{\mu^2 K\log^3 m}{m}}\max_{1\leq j\leq m}\left\|\beta^{0,(j)}\check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(j)}\right\|_2,
\end{aligned} \tag{229}$$

where $\beta^{0,(j)}$ is as defined in (225). Here, (i) comes from the estimates $\sigma_1(\boldsymbol{M}) \geq \frac{1}{2}$. The bound (ii) follows by combining the incoherence condition (34), the bound (184), the triangle inequality, as well as the estimate

$$\sum_{j=1}^m |\boldsymbol{b}_l^*\boldsymbol{b}_j| \leq 4\log m,$$

which arises from Lemma 51. The last line uses the estimate $\max_{1\leq j\leq m}\left|\boldsymbol{a}_j^*\check{\boldsymbol{x}}^{0,(j)}\right| \leq 5\sqrt{\log m}$ and (185). Our bound (229) further implies

$$\max_{1\leq l\leq m}\left|\boldsymbol{b}_l^* \check{\boldsymbol{h}}^0\right| \leq 200\frac{\mu\log^2 m}{\sqrt{m}} + 120\sqrt{\frac{\mu^2 K\log^3 m}{m}}\max_{1\leq j\leq m}\left\|\beta^{0,(j)}\check{\boldsymbol{x}}^0 - \check{\boldsymbol{x}}^{0,(j)}\right\|_2. \tag{230}$$

The above bound (230) taken together with (228) gives

$$\max_{1\le l\le m}\left\{\left\|\beta^{0,(l)}\check{\boldsymbol{h}}^{0}-\check{\boldsymbol{h}}^{0,(l)}\right\|_{2}+\left\|\beta^{0,(l)}\check{\boldsymbol{x}}^{0}-\check{\boldsymbol{x}}^{0,(l)}\right\|_{2}\right\}\le 120c_{1}\frac{\mu}{\sqrt{m}}\cdot\sqrt{\frac{K\log^{2}m}{m}}$$
$$+60c_{1}\sqrt{\frac{\mu^{2}K\log m}{m}}\left(200\frac{\mu\log^{2}m}{\sqrt{m}}+120\sqrt{\frac{\mu^{2}K\log^{3}m}{m}}\max_{1\le j\le m}\left\|\beta^{0,(j)}\check{\boldsymbol{x}}^{0}-\check{\boldsymbol{x}}^{0,(j)}\right\|_{2}\right). \tag{231}$$

As long as $m\gg\mu^{2}K\log^{2}m$ we have $60c_{1}\sqrt{\frac{\mu^{2}K\log m}{m}}\cdot120\sqrt{\frac{\mu^{2}K\log^{3}m}{m}}\le1/2$. Rearranging terms, we are left with

$$\max_{1\le l\le m}\left\{\left\|\beta^{0,(l)}\check{\boldsymbol{h}}^{0}-\check{\boldsymbol{h}}^{0,(l)}\right\|_{2}+\left\|\beta^{0,(l)}\check{\boldsymbol{x}}^{0}-\check{\boldsymbol{x}}^{0,(l)}\right\|_{2}\right\}\le C_{4}\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^{2}K\log^{5}m}{m}} \tag{232}$$

for some constant $C_4>0$. Further, this bound combined with (230) yields

$$\max_{1\le l\le m}\left|\boldsymbol{b}_{l}^{*}\check{\boldsymbol{h}}^{0}\right|\le 200\frac{\mu\log^{2}m}{\sqrt{m}}+120\sqrt{\frac{\mu^{2}K\log^{3}m}{m}}\cdot C_{4}\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^{2}K\log^{5}m}{m}}\le c_{2}\frac{\mu\log^{2}m}{\sqrt{m}} \tag{233}$$

for some constant $c_2>0$, with the proviso that $m\gg\mu^{2}K\log^{2}m$.

We now translate the preceding bounds to the scaled version. Recall from the bound (222) that

$$\frac{1}{2}\le1-\xi\le\|\boldsymbol{M}\|=\sigma_{1}(\boldsymbol{M})\le1+\xi\le2, \tag{234}$$

as long as $\xi\le1/2$. 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 31 to derive that

$$\begin{aligned}&\left|\sigma_{1}(\boldsymbol{M})-\sigma_{1}(\boldsymbol{M}^{(l)})\right|\\&\quad\le\left\|\left(\boldsymbol{M}-\boldsymbol{M}^{(l)}\right)\check{\boldsymbol{x}}^{0,(l)}\right\|_{2}+\left\{\left\|\alpha\check{\boldsymbol{h}}^{0}-\check{\boldsymbol{h}}^{0,(l)}\right\|_{2}+\left\|\alpha\check{\boldsymbol{x}}^{0}-\check{\boldsymbol{x}}^{0,(l)}\right\|_{2}\right\}\|\boldsymbol{M}\|\\&\quad\le\left\|\left(\boldsymbol{M}-\boldsymbol{M}^{(l)}\right)\check{\boldsymbol{x}}^{0,(l)}\right\|_{2}+2\left\{\left\|\alpha\check{\boldsymbol{h}}^{0}-\check{\boldsymbol{h}}^{0,(l)}\right\|_{2}+\left\|\alpha\check{\boldsymbol{x}}^{0}-\check{\boldsymbol{x}}^{0,(l)}\right\|_{2}\right\}\end{aligned}$$

and

$$\begin{aligned}&\left\|\alpha\boldsymbol{h}^{0}-\boldsymbol{h}^{0,(l)}\right\|_{2}+\left\|\alpha\boldsymbol{x}^{0}-\boldsymbol{x}^{0,(l)}\right\|_{2}\\&\quad=\left\|\sqrt{\sigma_{1}\left(\boldsymbol{M}\right)}\left(\alpha\check{\boldsymbol{h}}^{0}\right)-\sqrt{\sigma_{1}\left(\boldsymbol{M}^{(l)}\right)}\check{\boldsymbol{h}}^{0,(l)}\right\|_{2}+\left\|\sqrt{\sigma_{1}\left(\boldsymbol{M}\right)}\alpha\check{\boldsymbol{x}}^{0}-\sqrt{\sigma_{1}\left(\boldsymbol{M}^{(l)}\right)}\check{\boldsymbol{x}}^{0,(l)}\right\|_{2}\\&\quad\le\sqrt{\sigma_{1}(\boldsymbol{M})}\left\{\left\|\alpha\check{\boldsymbol{h}}^{0}-\check{\boldsymbol{h}}^{0,(l)}\right\|_{2}+\left\|\alpha\check{\boldsymbol{x}}^{0}-\check{\boldsymbol{x}}^{0,(l)}\right\|_{2}\right\}+\frac{2\left|\sigma_{1}(\boldsymbol{M})-\sigma_{1}(\boldsymbol{M}^{(l)})\right|}{\sqrt{\sigma_{1}(\boldsymbol{M})}+\sqrt{\sigma_{1}(\boldsymbol{M}^{(l)})}}\\&\quad\le\sqrt{2}\left\{\left\|\alpha\check{\boldsymbol{h}}^{0}-\check{\boldsymbol{h}}^{0,(l)}\right\|_{2}+\left\|\alpha\check{\boldsymbol{x}}^{0}-\check{\boldsymbol{x}}^{0,(l)}\right\|_{2}\right\}+\sqrt{2}\left|\sigma_{1}(\boldsymbol{M})-\sigma_{1}(\boldsymbol{M}^{(l)})\right|.\end{aligned}$$

Taking the previous two bounds collectively yields

$$\left\|\alpha\boldsymbol{h}^{0}-\boldsymbol{h}^{0,(l)}\right\|_{2}+\left\|\alpha\boldsymbol{x}^{0}-\boldsymbol{x}^{0,(l)}\right\|_{2}\le\left\|\left(\boldsymbol{M}-\boldsymbol{M}^{(l)}\right)\check{\boldsymbol{x}}^{0,(l)}\right\|_{2}+6\left\{\left\|\alpha\check{\boldsymbol{h}}^{0}-\check{\boldsymbol{h}}^{0,(l)}\right\|_{2}+\left\|\alpha\check{\boldsymbol{x}}^{0}-\check{\boldsymbol{x}}^{0,(l)}\right\|_{2}\right\},$$

which together with (227) and (232) implies

$$\min_{\alpha\in\mathbb{C},|\alpha|=1}\left\{\left\|\alpha\boldsymbol{h}^{0}-\boldsymbol{h}^{0,(l)}\right\|_{2}+\left\|\alpha\boldsymbol{x}^{0}-\boldsymbol{x}^{0,(l)}\right\|_{2}\right\}\le c_{5}\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^{2}K\log^{5}m}{m}} \tag{235}$$

for some constant $c_5 > 0$, as long as ξ is sufficiently small. Moreover, we have

$$\left\|\frac{1}{\overline{\alpha^0}}\boldsymbol{h}^0 - \frac{\alpha}{\overline{\alpha^0}}\boldsymbol{h}^{0,(l)}\right\|_2 + \left\|\alpha^0\boldsymbol{x}^0 - \alpha\alpha^0\boldsymbol{x}^{0,(l)}\right\|_2 \le 2\left\{\left\|\boldsymbol{h}^0 - \alpha\boldsymbol{h}^{0,(l)}\right\|_2 + \left\|\boldsymbol{x}^0 - \alpha\boldsymbol{x}^{0,(l)}\right\|_2\right\}$$

for any $|\alpha| = 1$, where α^0 is as defined in (88) and satisfies

$$1/2 \le \alpha^0 \le 2. \tag{236}$$

This property follows directly from Lemma 18 and Lemma 55. Therefore,

$$\begin{aligned}
&\min_{\alpha\in\mathbb{C},|\alpha|=1}\sqrt{\left\|\frac{1}{\overline{\alpha^0}}\boldsymbol{h}^0 - \frac{\alpha}{\overline{\alpha^0}}\boldsymbol{h}^{0,(l)}\right\|_2^2 + \left\|\alpha^0\boldsymbol{x}^0 - \alpha\alpha^0\boldsymbol{x}^{0,(l)}\right\|_2^2} \\
&\le \min_{\alpha\in\mathbb{C},|\alpha|=1}\left\{\left\|\frac{1}{\overline{\alpha^0}}\boldsymbol{h}^0 - \frac{\alpha}{\overline{\alpha^0}}\boldsymbol{h}^{0,(l)}\right\|_2 + \left\|\alpha^0\boldsymbol{x}^0 - \alpha\alpha^0\boldsymbol{x}^{0,(l)}\right\|_2\right\} \\
&\le 2\min_{\alpha\in\mathbb{C},|\alpha|=1}\left\{\left\|\boldsymbol{h}^0 - \alpha\boldsymbol{h}^{0,(l)}\right\|_2 + \left\|\boldsymbol{x}^0 - \alpha\boldsymbol{x}^{0,(l)}\right\|_2\right\} \\
&\le 2c_5\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^2 K\log^5 m}{m}}.
\end{aligned}$$

Furthermore, define a new alignment parameter that accommodates a larger set:

$$\alpha_{\text{mutual}}^{0,(l)} := \operatorname*{argmin}_{\alpha\in\mathbb{C}}\left\{\left\|\frac{1}{\overline{\alpha}}\boldsymbol{h}^{0,(l)} - \frac{1}{\overline{\alpha^0}}\boldsymbol{h}^0\right\|_2^2 + \left\|\alpha\boldsymbol{x}^{0,(l)} - \alpha^0\boldsymbol{x}^0\right\|_2^2\right\}.$$

Since $\alpha_{\text{mutual}}^{0,(l)}$ is the minimizer over a larger feasible set, we have

$$\begin{aligned}
\operatorname{dist}\left(\boldsymbol{z}^{0,(l)}, \tilde{\boldsymbol{z}}^0\right) &= \sqrt{\left\|\frac{1}{\overline{\alpha^0}}\boldsymbol{h}^0 - \frac{1}{\overline{\alpha_{\text{mutual}}^{0,(l)}}}\boldsymbol{h}^{0,(l)}\right\|_2^2 + \left\|\alpha^0\boldsymbol{x}^0 - \alpha_{\text{mutual}}^{0,(l)}\boldsymbol{x}^{0,(l)}\right\|_2^2} \\
&\le \min_{\alpha\in\mathbb{C},|\alpha|=1}\sqrt{\left\|\frac{1}{\overline{\alpha^0}}\boldsymbol{h}^0 - \frac{\alpha}{\overline{\alpha^0}}\boldsymbol{h}^{0,(l)}\right\|_2^2 + \left\|\alpha^0\boldsymbol{x}^0 - \alpha\alpha^0\boldsymbol{x}^{0,(l)}\right\|_2^2} \\
&\le 2c_5\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^2 K\log^5 m}{m}}.
\end{aligned} \tag{237}$$

This completes the proof for the claim (93).

Regarding $\left|\boldsymbol{b}_l^*\tilde{\boldsymbol{h}}^0\right|$, one first sees that

$$\left|\boldsymbol{b}_l^*\boldsymbol{h}^0\right| = \left|\sqrt{\sigma_1(\boldsymbol{M})}\boldsymbol{b}_l^*\check{\boldsymbol{h}}^0\right| \le \sqrt{2}c_2\frac{\mu\log^2 m}{\sqrt{m}},$$

where the last relation holds due to (233) and (234). Hence, using the property (236), we have

$$\left|\boldsymbol{b}_l^*\tilde{\boldsymbol{h}}^0\right| = \left|\boldsymbol{b}_l^*\frac{1}{\overline{\alpha^0}}\boldsymbol{h}^0\right| \le \left|\frac{1}{\overline{\alpha^0}}\right|\left|\boldsymbol{b}_l^*\boldsymbol{h}^0\right| \le 2\sqrt{2}c_2\frac{\mu\log^2 m}{\sqrt{m}},$$

which finishes the proof of the claim (94).

Before concluding this section, we note a byproduct of the proof. Specifically, we can establish the claim required in (216) using many results derived in this section. This is formally stated in the following lemma.

Lemma 32. *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*

$$\max_{1\le j\le\tau}\left|(\boldsymbol{b}_j - \boldsymbol{b}_1)^*\tilde{\boldsymbol{h}}^0\right| \le c\frac{\mu}{\sqrt{m}}\log m,$$

where $c > 0$ is any sufficiently small constant.

Proof. Instate the notation and hypothesis in Appendix C.6. Recognize that

$$\begin{aligned}\left|(\boldsymbol{b}_j-\boldsymbol{b}_1)^*\tilde{\boldsymbol{h}}^0\right| &= \left|(\boldsymbol{b}_j-\boldsymbol{b}_1)^*\frac{1}{\overline{\alpha^0}}\boldsymbol{h}^0\right| = \left|(\boldsymbol{b}_j-\boldsymbol{b}_1)^*\frac{1}{\overline{\alpha^0}}\sqrt{\sigma_1(\boldsymbol{M})}\check{\boldsymbol{h}}^0\right| \\ &\leq \left|\frac{1}{\overline{\alpha^0}}\right|\sqrt{\sigma_1(\boldsymbol{M})}\left|(\boldsymbol{b}_j-\boldsymbol{b}_1)^*\check{\boldsymbol{h}}^0\right| \\ &\leq 4\left|(\boldsymbol{b}_j-\boldsymbol{b}_1)^*\check{\boldsymbol{h}}^0\right|,\end{aligned}$$

where the last inequality comes from (234) and (236). It thus suffices to prove that $\left|(\boldsymbol{b}_j-\boldsymbol{b}_1)^*\check{\boldsymbol{h}}^0\right| \leq c\frac{\mu}{\sqrt{m}}\log m$ for some $c>0$ small enough. To this end, it is seen that

$$\begin{aligned}\left|(\boldsymbol{b}_j-\boldsymbol{b}_1)^*\check{\boldsymbol{h}}^0\right| &= \frac{1}{\sigma_1(\boldsymbol{M})}\left|(\boldsymbol{b}_j-\boldsymbol{b}_1)^*\boldsymbol{M}\check{\boldsymbol{x}}^0\right| \\ &\leq 2\left|\sum_{k=1}^m (\boldsymbol{b}_j-\boldsymbol{b}_1)^*\boldsymbol{b}_k\boldsymbol{b}_k^*\boldsymbol{h}^\natural\boldsymbol{x}^{\natural *}\boldsymbol{a}_k\boldsymbol{a}_k^*\check{\boldsymbol{x}}^0\right| \\ &\leq 2\left(\sum_{k=1}^m\left|(\boldsymbol{b}_j-\boldsymbol{b}_1)^*\boldsymbol{b}_k\right|\right)\max_{1\leq k\leq m}\left\{\left|\boldsymbol{b}_k^*\boldsymbol{h}^\natural\right|\left|\boldsymbol{a}_k^*\boldsymbol{x}^\natural\right|\left|\boldsymbol{a}_k^*\check{\boldsymbol{x}}^0\right|\right\} \\ &\overset{\text{(i)}}{\leq} c\frac{1}{\log^2 m}\cdot\frac{\mu}{\sqrt{m}}\cdot\left(5\sqrt{\log m}\right)\max_{1\leq j\leq m}\left\{\left|\boldsymbol{a}_j^*\check{\boldsymbol{x}}^{0,(j)}\right| + \|\boldsymbol{a}_j\|_2\left\|\alpha^{0,(j)}\check{\boldsymbol{x}}^0-\check{\boldsymbol{x}}^{0,(j)}\right\|_2\right\} \\ &\overset{\text{(ii)}}{\lesssim} c\frac{\mu}{\sqrt{m}}\frac{1}{\log m} \leq c\frac{\mu}{\sqrt{m}}\log m, \end{aligned}\tag{238}$$

where (i) comes from Lemma 53, the incoherence condition, and (184). The last line (ii) holds since we have already established (see (229) and (232))

$$\max_{1\leq j\leq m}\left\{\left|\boldsymbol{a}_j^*\check{\boldsymbol{x}}^{0,(j)}\right| + \|\boldsymbol{a}_j\|_2\left\|\alpha^{0,(j)}\check{\boldsymbol{x}}^0-\check{\boldsymbol{x}}^{0,(j)}\right\|_2\right\} \lesssim \sqrt{\log m}.$$

The proof is then complete. □

C.7 Proof of Lemma 20

Recall that α^0 and $\alpha^{0,(l)}$ are the alignment parameters between $\boldsymbol{z}^0$ and $\boldsymbol{z}^\natural$, and between $\boldsymbol{z}^{0,(l)}$ and $\boldsymbol{z}^\natural$, respectively, that is,

$$\begin{aligned}\alpha^0 &:= \operatorname*{argmin}_{\alpha\in\mathbb{C}}\left\{\left\|\frac{1}{\overline{\alpha}}\boldsymbol{h}^0-\boldsymbol{h}^\natural\right\|_2^2 + \left\|\alpha\boldsymbol{x}^0-\boldsymbol{x}^\natural\right\|_2^2\right\}, \\ \alpha^{0,(l)} &:= \operatorname*{argmin}_{\alpha\in\mathbb{C}}\left\{\left\|\frac{1}{\overline{\alpha}}\boldsymbol{h}^{0,(l)}-\boldsymbol{h}^\natural\right\|_2^2 + \left\|\alpha\boldsymbol{x}^{0,(l)}-\boldsymbol{x}^\natural\right\|_2^2\right\}.\end{aligned}$$

Also, we let

$$\alpha_{\text{mutual}}^{0,(l)} := \operatorname*{argmin}_{\alpha\in\mathbb{C}}\left\{\left\|\frac{1}{\overline{\alpha}}\boldsymbol{h}^{0,(l)}-\frac{1}{\overline{\alpha^0}}\boldsymbol{h}^0\right\|_2^2 + \left\|\alpha\boldsymbol{x}^{0,(l)}-\alpha^0\boldsymbol{x}^0\right\|_2^2\right\}$$

as before. The triangle inequality together with (91) and (237) then tells us that

$$\begin{aligned}&\sqrt{\left\|\frac{1}{\overline{\alpha_{\text{mutual}}^{0,(l)}}}\boldsymbol{h}^{0,(l)}-\boldsymbol{h}^\natural\right\|_2^2 + \left\|\alpha_{\text{mutual}}^{0,(l)}\boldsymbol{x}^{0,(l)}-\boldsymbol{x}^\natural\right\|_2^2} \\ &\quad\leq \sqrt{\left\|\frac{1}{\overline{\alpha^0}}\boldsymbol{h}^0-\frac{1}{\overline{\alpha_{\text{mutual}}^{0,(l)}}}\boldsymbol{h}^{0,(l)}\right\|_2^2 + \left\|\alpha^0\boldsymbol{x}^0-\alpha_{\text{mutual}}^{0,(l)}\boldsymbol{x}^{0,(l)}\right\|_2^2} + \sqrt{\left\|\frac{1}{\overline{\alpha^0}}\boldsymbol{h}^0-\boldsymbol{h}^\natural\right\|_2^2 + \left\|\alpha^0\boldsymbol{x}^0-\boldsymbol{x}^\natural\right\|_2^2} \\ &\quad\leq 2c_5\frac{\mu}{\sqrt{m}}\sqrt{\frac{\mu^2 K\log^5 m}{m}} + C_1\frac{1}{\log^2 m}\end{aligned}$$

$$\leq 2C_1 \frac{1}{\log^2 m},$$

where the last relation holds as long as $m \gg \mu^2 \sqrt{K} \log^{9/2} m$.

Letting

$$\boldsymbol{x}_1 = \alpha^0 \boldsymbol{x}^0, \quad \boldsymbol{h}_1 = \frac{1}{\overline{\alpha^0}} \boldsymbol{h}_0 \qquad \text{and} \qquad \boldsymbol{x}_2 = \alpha_{\text{mutual}}^{0,(l)} \boldsymbol{x}^{0,(l)}, \quad \boldsymbol{h}_2 = \frac{1}{\overline{\alpha_{\text{mutual}}^{0,(l)}}} \boldsymbol{h}^{0,(l)},$$

it is easy to see that $\boldsymbol{x}_1, \boldsymbol{h}_1, \boldsymbol{x}_2, \boldsymbol{h}_2$ satisfy the assumptions in Lemma 58, which implies

$$\begin{aligned}
&\sqrt{\left\| \frac{1}{\overline{\alpha^{0,(l)}}} \boldsymbol{h}^{0,(l)} - \frac{1}{\overline{\alpha^0}} \boldsymbol{h}^0 \right\|_2^2 + \left\| \alpha^{0,(l)} \boldsymbol{x}^{0,(l)} - \alpha^0 \boldsymbol{x}^0 \right\|_2^2} \\
&\quad \lesssim \sqrt{\left\| \frac{1}{\overline{\alpha^0}} \boldsymbol{h}^0 - \frac{1}{\overline{\alpha_{\text{mutual}}^{0,(l)}}} \boldsymbol{h}^{0,(l)} \right\|_2^2 + \left\| \alpha^0 \boldsymbol{x}^0 - \alpha_{\text{mutual}}^{0,(l)} \boldsymbol{x}^{0,(l)} \right\|_2^2} \\
&\quad \lesssim \frac{\mu}{\sqrt{m}} \sqrt{\frac{\mu^2 K \log^5 m}{m}}, \qquad (239)
\end{aligned}$$

where the last line comes from (237).With this upper estimate at hand, we are now ready to show that with high probability,

$$\begin{aligned}
\left| \boldsymbol{a}_l^* \left(\alpha^0 \boldsymbol{x}^0 - \boldsymbol{x}^\natural \right) \right| &\overset{\text{(i)}}{\leq} \left| \boldsymbol{a}_l^* \left(\alpha^{0,(l)} \boldsymbol{x}^{0,(l)} - \boldsymbol{x}^\natural \right) \right| + \left| \boldsymbol{a}_l^* \left(\alpha^0 \boldsymbol{x}^0 - \alpha^{0,(l)} \boldsymbol{x}^{0,(l)} \right) \right| \\
&\overset{\text{(ii)}}{\leq} 5\sqrt{\log m} \left\| \alpha^{0,(l)} \boldsymbol{x}^{0,(l)} - \boldsymbol{x}^\natural \right\|_2 + \|\boldsymbol{a}_l\|_2 \left\| \alpha^0 \boldsymbol{x}^0 - \alpha^{0,(l)} \boldsymbol{x}^{0,(l)} \right\|_2 \\
&\overset{\text{(iii)}}{\lesssim} \sqrt{\log m} \cdot \frac{1}{\log^2 m} + \sqrt{K} \frac{\mu}{\sqrt{m}} \sqrt{\frac{\mu^2 K \log^5 m}{m}} \\
&\overset{\text{(iv)}}{\lesssim} \frac{1}{\log^{1.5} m},
\end{aligned}$$

where (i) follows from the triangle inequality, (ii) uses the Cauchy-Schwarz inequality and the independence between $\boldsymbol{x}^{0,(l)}$ and $\boldsymbol{a}_l$, (iii) holds because of (92) and (239), 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

In this subsection, we collect several technical lemmas used throughout the proof for phase retrieval. The proofs are given if needed.

***Lemma* 33.** *Suppose $\boldsymbol{a}_j \overset{i.i.d.}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$ for each $1 \leq j \leq m$. With probability at least $1 - me^{-1.5n}$, one has*

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

Proof. This directly follows from the standard concentration inequality for χ^2 random variables and the union bound. □

***Lemma* 34.** *Suppose that $\boldsymbol{a}_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$ for every $1 \leq j \leq m$. Fix any small constant $\delta > 0$. With probability at least $1 - C_2 e^{-c_2 m}$, one has*

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

as long as $m \geq c_0 n$ 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]. $\square$

Lemma 35. *Suppose that $\boldsymbol{a}_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$, for every $1 \leq j \leq m$. With probability at least $1 - O(n^{-10})$, we have*

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

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

Proof. This is adapted from [CLS15, Lemma 7.4]. $\square$

Lemma 36. *Suppose that $\boldsymbol{a}_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$, for every $1 \leq j \leq m$. Fix any small constant $\delta > 0$ and any constant $C > 0$. Suppose $m \geq c_0 n$ for some sufficiently large constant $c_0 > 0$. Then with probability at least $1 - C_2 e^{-c_2 m}$,*

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

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

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

with ξ being a standard Gaussian random variable.

Proof. This is supplied in [CC17, supplementary material]. $\square$

We also need the following perturbation bound on the leading eigenvalues.

Lemma 37. *Let $\lambda_1(\boldsymbol{A})$, $\boldsymbol{u}$ be the leading eigenvalue and eigenvector of $\boldsymbol{A}$, respectively, and $\lambda_1(\tilde{\boldsymbol{A}})$, $\tilde{\boldsymbol{u}}$ be the leading eigenvalue and eigenvector of $\tilde{\boldsymbol{A}}$, respectively. Suppose that $\lambda_1(\boldsymbol{A}), \lambda_1(\tilde{\boldsymbol{A}}), \|\boldsymbol{A}\|, \|\tilde{\boldsymbol{A}}\| \in [C_1, C_2]$ for some $C_1, C_2 > 0$. Then,*

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

Proof. Observe that

$$\begin{aligned} \left\| \sqrt{\lambda_1(\boldsymbol{A})}\boldsymbol{u} - \sqrt{\lambda_1(\tilde{\boldsymbol{A}})}\tilde{\boldsymbol{u}} \right\|_2 &\leq \left\| \sqrt{\lambda_1(\boldsymbol{A})}\boldsymbol{u} - \sqrt{\lambda_1(\tilde{\boldsymbol{A}})}\boldsymbol{u} \right\|_2 + \left\| \sqrt{\lambda_1(\tilde{\boldsymbol{A}})}\boldsymbol{u} - \sqrt{\lambda_1(\tilde{\boldsymbol{A}})}\tilde{\boldsymbol{u}} \right\|_2 \\ &\leq \left| \sqrt{\lambda_1(\boldsymbol{A})} - \sqrt{\lambda_1(\tilde{\boldsymbol{A}})} \right| + \sqrt{\lambda_1(\tilde{\boldsymbol{A}})} \, \|\boldsymbol{u} - \tilde{\boldsymbol{u}}\|_2 , \end{aligned} \tag{240}$$

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

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

where the last inequality comes from our assumptions on $\lambda_1(\boldsymbol{A})$ and $\lambda_1(\tilde{\boldsymbol{A}})$. This combined with (240) yields

$$\left\| \sqrt{\lambda_1(\boldsymbol{A})}\boldsymbol{u} - \sqrt{\lambda_1(\tilde{\boldsymbol{A}})}\tilde{\boldsymbol{u}} \right\|_2 \leq \frac{\left| \lambda_1(\boldsymbol{A}) - \lambda_1(\tilde{\boldsymbol{A}}) \right|}{2\sqrt{C_1}} + \sqrt{C_2} \, \|\boldsymbol{u} - \tilde{\boldsymbol{u}}\|_2 . \tag{241}$$

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

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

which together with (241) gives

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

as claimed. □

D.2 Technical lemmas for matrix completion

In this section, we gather some technical lemmas used in Appendix B.

***Lemma* 38.** *Suppose $\{\delta_{l,j}\}_{1\le j\le n}$ are independent Bernoulli random variables with mean p, and $\boldsymbol{A} \in \mathbb{R}^{n\times r}$ is some fixed matrix independent of the $\delta_{l,j}$'s. 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

$$\mathsf{Median}\left[\|\boldsymbol{G}_l(\boldsymbol{A})\|\right] \le \sqrt{p\|\boldsymbol{A}\|^2 + \sqrt{p\|\boldsymbol{A}\|_{2,\infty}^2\|\boldsymbol{A}\|^2\log(4r)} + \frac{\|\boldsymbol{A}\|_{2,\infty}^2}{3}\log(4r)}$$

and with probability exceeding $1 - n^{-(1.5C-1)}$,

$$\|\boldsymbol{G}_l(\boldsymbol{A})\| \le \sqrt{p\|\boldsymbol{A}\|^2 + C\left(\sqrt{p\|\boldsymbol{A}\|_{2,\infty}^2\|\boldsymbol{A}\|^2\log n} + \|\boldsymbol{A}\|_{2,\infty}^2\log n\right)}$$

$$\left\|\sum_{j=1}^n(\delta_{l,j} - p)\boldsymbol{A}_{1,\cdot}^\top\boldsymbol{A}_{1,\cdot}\right\| \le C\left(\sqrt{p\|\boldsymbol{A}\|_{2,\infty}^2\|\boldsymbol{A}\|^2\log n} + \|\boldsymbol{A}\|_{2,\infty}^2\log n\right).$$

Proof. Fix some $1\le l\le n$, by definition

$$\begin{aligned}
\|\boldsymbol{G}_l(\boldsymbol{A})\|^2 = \left\|\boldsymbol{G}_l(\boldsymbol{A})\boldsymbol{G}_l(\boldsymbol{A})^\top\right\| &= \left\|\sum_{j=1}^n \delta_{l,j}\boldsymbol{A}_{j,\cdot}^\top\boldsymbol{A}_{j,\cdot}\right\| \\
&\le \left\|\sum_{j=1}^n(\delta_{l,j} - p)\boldsymbol{A}_{j,\cdot}^\top\boldsymbol{A}_{j,\cdot}\right\| + p\|\boldsymbol{A}\|^2.
\end{aligned}$$

Therefore, it suffices to control the first term. Letting

$$L := \max_{1\le j\le n}\left\|(\delta_{l,j} - p)\boldsymbol{A}_{j,\cdot}^\top\boldsymbol{A}_{j,\cdot}\right\| \le \|\boldsymbol{A}\|_{2,\infty}^2$$

$$\text{and}\qquad V := \left\|\sum_{j=1}^n\mathbb{E}\left[(\delta_{l,j} - p)^2\boldsymbol{A}_{j,\cdot}^\top\boldsymbol{A}_{j,\cdot}\boldsymbol{A}_{j,\cdot}^\top\boldsymbol{A}_{j,\cdot}\right]\right\| \le \mathbb{E}\left[(\delta_{l,j} - p)^2\right]\|\boldsymbol{A}\|_{2,\infty}^2\left\|\sum_{j=1}^n\boldsymbol{A}_{j,\cdot}^\top\boldsymbol{A}_{j,\cdot}\right\|$$

$$\leq p \left\| \boldsymbol{A} \right\|_{2,\infty}^{2} \left\| \boldsymbol{A} \right\|^{2}$$

and invoking matrix Bernstein's inequality [Tro15b, Theorem 6.1.1], one has for all $t \geq 0$,

$$\mathbb{P}\left\{ \left\| \sum_{j=1}^{n} \left(\delta_{l,j} - p\right) \boldsymbol{A}_{j,\cdot}^{\top} \boldsymbol{A}_{j,\cdot} \right\| \geq t \right\} \leq 2r \cdot \exp\left(\frac{-t^2/2}{p \left\| \boldsymbol{A} \right\|_{2,\infty}^{2} \left\| \boldsymbol{A} \right\|^{2} + \left\| \boldsymbol{A} \right\|_{2,\infty}^{2} t/3} \right). \tag{242}$$

We can thus find an upper bound on $\mathsf{Median}\left[\left\| \sum_{j=1}^{n} \left(\delta_{l,j} - p\right) \boldsymbol{A}_{j,\cdot}^{\top} \boldsymbol{A}_{j,\cdot} \right\| \right]$ by finding a value t that ensures the right-hand side of (242) to be smaller than $1/2$. Using this strategy and some simple calculation, we get

$$\mathsf{Median}\left[\left\| \sum_{j=1}^{n} \left(\delta_{l,j} - p\right) \boldsymbol{A}_{j,\cdot}^{\top} \boldsymbol{A}_{j,\cdot} \right\| \right] \leq \sqrt{p \left\| \boldsymbol{A} \right\|_{2,\infty}^{2} \left\| \boldsymbol{A} \right\|^{2} \log\left(4r\right)} + \frac{\left\| \boldsymbol{A} \right\|_{2,\infty}^{2}}{3} \log\left(4r\right)$$

and for any $C \geq 3$,

$$\left\| \sum_{j=1}^{n} \left(\delta_{l,j} - p\right) \boldsymbol{A}_{j,\cdot}^{\top} \boldsymbol{A}_{j,\cdot} \right\| \leq C \left(\sqrt{p \left\| \boldsymbol{A} \right\|_{2,\infty}^{2} \left\| \boldsymbol{A} \right\|^{2} \log n} + \left\| \boldsymbol{A} \right\|_{2,\infty}^{2} \log n \right)$$

holds with probability at least $1 - n^{-(1.5C-1)}$. As consequence, we have

$$\mathsf{Median}\left[\left\| \boldsymbol{G}_l\left(\boldsymbol{A}\right) \right\| \right] \leq \sqrt{p \left\| \boldsymbol{A} \right\|^{2} + \sqrt{p \left\| \boldsymbol{A} \right\|_{2,\infty}^{2} \left\| \boldsymbol{A} \right\|^{2} \log\left(4r\right)} + \frac{\left\| \boldsymbol{A} \right\|_{2,\infty}^{2}}{3} \log\left(4r\right)},$$

and with probability exceeding $1 - n^{-(1.5C-1)}$,

$$\left\| \boldsymbol{G}_l\left(\boldsymbol{A}\right) \right\|^{2} \leq p \left\| \boldsymbol{A} \right\|^{2} + C \left(\sqrt{p \left\| \boldsymbol{A} \right\|_{2,\infty}^{2} \left\| \boldsymbol{A} \right\|^{2} \log n} + \left\| \boldsymbol{A} \right\|_{2,\infty}^{2} \log n \right).$$

This completes the proof. $\square$

***Lemma* 39.** *Suppose the sample size exceeds $n^2 p \geq C\kappa\mu r n \log n$ for some sufficiently large constant $C > 0$. Then with probability at least $1 - O\left(n^{-10}\right)$,*

$$\left\| \frac{1}{p} \mathcal{P}_{\Omega}\left(\boldsymbol{X}\boldsymbol{X}^{\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top} \right) \right\| \leq 2n\epsilon^{2} \left\| \boldsymbol{X}^{\natural} \right\|_{2,\infty}^{2} + 2\epsilon\sqrt{n}\log n \left\| \boldsymbol{X}^{\natural} \right\|_{2,\infty} \left\| \boldsymbol{X}^{\natural} \right\|$$

holds simultaneously for all $\boldsymbol{X} \in \mathbb{R}^{n \times r}$ satisfying

$$\left\| \boldsymbol{X} - \boldsymbol{X}^{\natural} \right\|_{2,\infty} \leq \epsilon \left\| \boldsymbol{X}^{\natural} \right\|_{2,\infty}, \tag{243}$$

where $\epsilon > 0$ is some absolute constant.

Proof. To simplify the notations hereafter, we denote

$$\boldsymbol{\Delta} := \boldsymbol{X} - \boldsymbol{X}^{\natural}.$$

With this notation in place, one can decompose

$$\boldsymbol{X}\boldsymbol{X}^{\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top} = \boldsymbol{\Delta}\boldsymbol{X}^{\natural\top} + \boldsymbol{X}^{\natural}\boldsymbol{\Delta}^{\top} + \boldsymbol{\Delta}\boldsymbol{\Delta}^{\top},$$

which together with the triangle inequality implies that

$$\begin{aligned} \left\| \frac{1}{p} \mathcal{P}_{\Omega}\left(\boldsymbol{X}\boldsymbol{X}^{\top} - \boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top} \right) \right\| &\leq \left\| \frac{1}{p} \mathcal{P}_{\Omega}\left(\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top} \right) \right\| + \left\| \frac{1}{p} \mathcal{P}_{\Omega}\left(\boldsymbol{X}^{\natural}\boldsymbol{\Delta}^{\top} \right) \right\| + \left\| \frac{1}{p} \mathcal{P}_{\Omega}\left(\boldsymbol{\Delta}\boldsymbol{\Delta}^{\top} \right) \right\| \\ &= \underbrace{\left\| \frac{1}{p} \mathcal{P}_{\Omega}\left(\boldsymbol{\Delta}\boldsymbol{\Delta}^{\top} \right) \right\|}_{:=\alpha_1} + 2 \underbrace{\left\| \frac{1}{p} \mathcal{P}_{\Omega}\left(\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top} \right) \right\|}_{:=\alpha_2}. \end{aligned} \tag{244}$$

In the sequel, we bound α_1 and α_2 separately.

1. Recall from [Mat90, Theorem 2.5] the elementary inequality that

$$\|\boldsymbol{C}\| \leq \big\||\boldsymbol{C}|\big\|, \tag{245}$$

where $|\boldsymbol{C}| := [|c_{i,j}|]_{1\leq i,j\leq n}$ for any matrix $\boldsymbol{C} = [c_{i,j}]_{1\leq i,j\leq n}$. Consequently, for any matrix $\boldsymbol{D} := [d_{i,j}]_{1\leq i,j}$ such that $|d_{i,j}| \geq |c_{i,j}|$ for all i and j, one has $\big\||\boldsymbol{C}|\big\| \leq \big\||\boldsymbol{D}|\big\|$. Therefore

$$\alpha_1 \leq \left\|\frac{1}{p}\mathcal{P}_\Omega\left(\left|\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\right|\right)\right\| \leq \|\boldsymbol{\Delta}\|_{2,\infty}^2 \left\|\frac{1}{p}\mathcal{P}_\Omega\left(\mathbf{1}\mathbf{1}^\top\right)\right\|.$$

Lemma 45 then tells us that with probability at least $1 - O(n^{-10})$,

$$\left\|\frac{1}{p}\mathcal{P}_\Omega\left(\mathbf{1}\mathbf{1}^\top\right) - \mathbf{1}\mathbf{1}^\top\right\| \leq C\sqrt{\frac{n}{p}} \tag{246}$$

for some universal constant $C > 0$, as long as $p \gtrsim \log n/n$. This yields

$$\left\|\frac{1}{p}\mathcal{P}_\Omega\left(\mathbf{1}\mathbf{1}^\top\right)\right\| \leq \left\|\frac{1}{p}\mathcal{P}_\Omega\left(\mathbf{1}\mathbf{1}^\top\right) - \mathbf{1}\mathbf{1}^\top\right\| + \left\|\mathbf{1}\mathbf{1}^\top\right\| \leq C\sqrt{\frac{n}{p}} + n \leq 2n, \tag{247}$$

provided that $p \gg 1/n$. Putting together the previous bounds, we arrive at

$$\alpha_1 \leq 2n\,\|\boldsymbol{\Delta}\|_{2,\infty}^2. \tag{248}$$

2. Regarding the second term α_2, apply the elementary inequality (245) to get

$$\left\|\mathcal{P}_\Omega\left(\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right)\right\| \leq \left\|\mathcal{P}_\Omega\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right|\right)\right\|,$$

which motivates us to look at $\left\|\mathcal{P}_\Omega\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right|\right)\right\|$ instead. A key step of this part is to take advantage of the $\ell_{2,\infty}$ norm constraint of $\mathcal{P}_\Omega\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right|\right)$. Specifically, we claim for the moment that with probability exceeding $1 - O(n^{-10})$,

$$\left\|\mathcal{P}_\Omega\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right|\right)\right\|_{2,\infty}^2 \leq 2p\sigma_{\max}\,\|\boldsymbol{\Delta}\|_{2,\infty}^2 := \theta \tag{249}$$

holds under our sample size condition. In addition, we also have the following trivial ℓ_∞ norm bound

$$\left\|\mathcal{P}_\Omega\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right|\right)\right\|_\infty \leq \|\boldsymbol{\Delta}\|_{2,\infty}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} := \gamma. \tag{250}$$

In what follows, we will denote

$$\boldsymbol{A} := \mathcal{P}_\Omega\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right|\right)$$

for simplicity of presentation.

(a) To facilitate the analysis of $\|\boldsymbol{A}\|$, we first introduce $k_0 = \frac{1}{2}\log(\kappa\mu r) - 1$ auxiliary matrices $\boldsymbol{B}_s \in \mathbb{R}^{n\times n}$ that satisfy

$$\|\boldsymbol{A}\| \leq \|\boldsymbol{B}_{k_0}\| + \sum_{s=0}^{k_0-1}\|\boldsymbol{B}_s\|. \tag{251}$$

To be precise, each $\boldsymbol{B}_s$ is defined such that

$$[\boldsymbol{B}_s]_{j,k} = \begin{cases} \frac{1}{2^s}\gamma, & \text{if} \quad A_{j,k} \in (\frac{1}{2^{s+1}}\gamma, \frac{1}{2^s}\gamma], \\ 0, & \text{else}, \end{cases} \qquad \forall 0 \leq s \leq k_0 - 1 \qquad \text{and}$$

$$[\boldsymbol{B}_{k_0}]_{j,k} = \begin{cases} \frac{1}{2^{k_0}}\gamma, & \text{if} \quad A_{j,k} \leq \frac{1}{2^{k_0}}\gamma, \\ 0, & \text{else}, \end{cases}$$

which clearly satisfy (251); 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

$$\|\boldsymbol{B}_{k_0}\| \leq \|\boldsymbol{B}_{k_0}\|_\infty\sqrt{(2np)^2} \leq 4np\frac{1}{\sqrt{\kappa\mu r}}\|\boldsymbol{\Delta}\|_{2,\infty}\left\|\boldsymbol{X}^\natural\right\|_{2,\infty} \leq 4\sqrt{n}p\,\|\boldsymbol{\Delta}\|_{2,\infty}\left\|\boldsymbol{X}^\natural\right\|,$$

where the first inequality arises from Lemma 50, with $2np$ being a crude upper bound on the number of nonzero entries in each row and each column. The last inequality follows from the incoherence condition (114). For any $0 \leq s < k_0$, 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 (249). 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

$$k_{\mathrm{c}} = 2np$$

entries in each column of $\boldsymbol{B}_s$, where k_{c} is derived from the standard Chernoff bound on Ω. Utilizing Lemma 50 once more, we discover that

$$\|\boldsymbol{B}_s\| \leq \|\boldsymbol{B}_s\|_\infty \sqrt{k_{\mathrm{r}} k_{\mathrm{c}}} = \frac{1}{2^s}\gamma\sqrt{k_{\mathrm{r}} k_{\mathrm{c}}} = \sqrt{16np^2\sigma_{\max} \|\boldsymbol{\Delta}\|_{2,\infty}^2} = 4\sqrt{np}\, \|\boldsymbol{\Delta}\|_{2,\infty} \left\|\boldsymbol{X}^\natural\right\|$$

for each $0 \leq s < k_0$. Combining all, we arrive at

$$\begin{aligned}
\|\boldsymbol{A}\| &\leq \sum_{s=0}^{k_0-1} \|\boldsymbol{B}_s\| + \|\boldsymbol{B}_{k_0}\| \leq (k_0+1)\, 4\sqrt{np}\, \|\boldsymbol{\Delta}\|_{2,\infty} \left\|\boldsymbol{X}^\natural\right\| \\
&\leq 2\sqrt{np}\log(\kappa\mu r)\, \|\boldsymbol{\Delta}\|_{2,\infty} \left\|\boldsymbol{X}^\natural\right\| \\
&\leq 2\sqrt{np}\log n\, \|\boldsymbol{\Delta}\|_{2,\infty} \left\|\boldsymbol{X}^\natural\right\|,
\end{aligned}$$

where the last relation holds because $n \gg \kappa\mu r$. This further gives

$$\alpha_2 \leq \frac{1}{p}\|\boldsymbol{A}\| \leq 2\sqrt{n}\log n\, \|\boldsymbol{\Delta}\|_{2,\infty} \left\|\boldsymbol{X}^\natural\right\|. \tag{252}$$

(b) In order to finish the proof of this part, we need to justify the claim (249). Observe that

$$\begin{aligned}
\left\|\left[\mathcal{P}_\Omega\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right|\right)\right]_{l,\cdot}\right\|_2^2 &= \sum\nolimits_{j=1}^n \left(\boldsymbol{\Delta}_{l,\cdot}\boldsymbol{X}_{j,\cdot}^{\natural\top}\delta_{l,j}\right)^2 \\
&= \boldsymbol{\Delta}_{l,\cdot}\left(\sum\nolimits_{j=1}^n \delta_{l,j}\boldsymbol{X}_{j,\cdot}^{\natural\top}\boldsymbol{X}_{j,\cdot}^\natural\right)\boldsymbol{\Delta}_{l,\cdot}^\top \\
&\leq \|\boldsymbol{\Delta}\|_{2,\infty}^2 \left\|\sum\nolimits_{j=1}^n \delta_{l,j}\boldsymbol{X}_{j,\cdot}^{\natural\top}\boldsymbol{X}_{j,\cdot}^\natural\right\|
\end{aligned} \tag{253}$$

for every $1 \leq l \leq n$. Invoke Lemma 38 to yield

$$\begin{aligned}
\left\|\sum\nolimits_{j=1}^n \delta_{l,j}\boldsymbol{X}_{j,\cdot}^{\natural\top}\boldsymbol{X}_{j,\cdot}^\natural\right\| &= \left\|\left[\delta_{l,1}\boldsymbol{X}_{1,\cdot}^{\natural\top}, \delta_{l,2}\boldsymbol{X}_{2,\cdot}^{\natural\top}, \cdots, \delta_{l,n}\boldsymbol{X}_{n,\cdot}^{\natural\top}\right]\right\|^2 \\
&\leq p\sigma_{\max} + C\left(\sqrt{p\left\|\boldsymbol{X}^\natural\right\|_{2,\infty}^2 \left\|\boldsymbol{X}^\natural\right\|^2 \log n} + \left\|\boldsymbol{X}^\natural\right\|_{2,\infty}^2 \log n\right) && (254) \\
&\leq \left(p + C\sqrt{\frac{p\mu r\log n}{n}} + C\frac{\mu r\log n}{n}\right)\sigma_{\max} \\
&\leq 2p\sigma_{\max}, && (255)
\end{aligned}$$

with high probability, as soon as $np \gg \mu r\log n$. Combining (253) and (255) yields

$$\left\|\left[\mathcal{P}_\Omega\left(\left|\boldsymbol{\Delta}\boldsymbol{X}^{\natural\top}\right|\right)\right]_{l,\cdot}\right\|_2^2 \leq 2p\sigma_{\max}\|\boldsymbol{\Delta}\|_{2,\infty}^2, \qquad 1 \leq l \leq n$$

as claimed in (249).

3. Taken together, the preceding bounds (244), (248) and (252) yield

$$\left\|\frac{1}{p}\mathcal{P}_{\Omega}\left(\boldsymbol{X}\boldsymbol{X}^{\top}-\boldsymbol{X}^{\natural}\boldsymbol{X}^{\natural\top}\right)\right\|\leq 2n\left\|\boldsymbol{\Delta}\right\|_{2,\infty}^{2}+2\sqrt{n}\log n\left\|\boldsymbol{\Delta}\right\|_{2,\infty}\left\|\boldsymbol{X}^{\natural}\right\|,$$

thus establishing the final bound claimed in the lemma.

□

Lemma 40. *Let $\{\delta_{l,j}\}_{1\leq l\leq j\leq n}$ 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*

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

Then with probability at least $1-c_{1}e^{-\frac{\alpha C}{2}nr}$,

$$\sum_{l=1}^{n}\mathbb{1}_{\left\{\|\boldsymbol{G}_{l}(\boldsymbol{\Delta})\|\geq4\sqrt{p}\psi+2\sqrt{kr}\xi\right\}}\leq\frac{2\alpha n\log n}{k}$$

holds simultaneously for all $\boldsymbol{\Delta}\in\mathbb{R}^{n\times r}$ obeying

$$\begin{aligned}
&\left\|\boldsymbol{\Delta}\right\|_{2,\infty}\leq C_{5}\rho^{t}\mu r\sqrt{\frac{\log n}{np}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}+C_{8}\sigma\sqrt{\frac{n\log n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|_{2,\infty}:=\xi\\
\text{and}\qquad&\left\|\boldsymbol{\Delta}\right\|\leq C_{9}\rho^{t}\mu r\frac{1}{\sqrt{np}}\left\|\boldsymbol{X}^{\natural}\right\|+C_{10}\sigma\sqrt{\frac{n}{p}}\left\|\boldsymbol{X}^{\natural}\right\|:=\psi,
\end{aligned}$$

where $c_{1},C,C_{5},C_{9}>0$ are absolute constants.

Proof. For simplicity of presentation, we will prove the claim for the asymmetric case where $\{\delta_{l,j}\}_{1\leq l,j\leq n}$ 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

$$\boldsymbol{G}_{l}(\boldsymbol{\Delta})=\boldsymbol{G}_{l}^{\mathrm{upper}}(\boldsymbol{\Delta})+\boldsymbol{G}_{l}^{\mathrm{lower}}(\boldsymbol{\Delta}),$$

where all nonzero components of $\boldsymbol{G}_{l}^{\mathrm{upper}}(\boldsymbol{\Delta})$ come from the upper triangular part (those blocks with $l\leq j$), while all nonzero components of $\boldsymbol{G}_{l}^{\mathrm{lower}}(\boldsymbol{\Delta})$ are from the lower triangular part (those blocks with $l>j$). We can then look at $\{\boldsymbol{G}_{l}^{\mathrm{upper}}(\boldsymbol{\Delta})\mid1\leq l\leq n\}$ and $\{\boldsymbol{G}_{l}^{\mathrm{upper}}(\boldsymbol{\Delta})\mid1\leq l\leq n\}$ separately using the argument we develop for the asymmetric case. From now on, we we assume that $\{\delta_{l,j}\}_{1\leq l,j\leq n}$ are independent.

Suppose for the moment that $\boldsymbol{\Delta}$ is statistically independent of $\{\delta_{l,j}\}_{1\leq l,j\leq n}$. Clearly, for any $\boldsymbol{\Delta},\tilde{\boldsymbol{\Delta}}\in\mathbb{R}^{n\times r}$,

$$\begin{aligned}
\left|\left\|\boldsymbol{G}_{l}\left(\boldsymbol{\Delta}\right)\right\|-\left\|\boldsymbol{G}_{l}(\tilde{\boldsymbol{\Delta}})\right\|\right|&\leq\left\|\boldsymbol{G}_{l}\left(\boldsymbol{\Delta}\right)-\boldsymbol{G}_{l}(\tilde{\boldsymbol{\Delta}})\right\|\leq\left\|\boldsymbol{G}_{l}\left(\boldsymbol{\Delta}\right)-\boldsymbol{G}_{l}(\tilde{\boldsymbol{\Delta}})\right\|_{\mathrm{F}}\\
&\leq\sqrt{\sum_{j=1}^{n}\left\|\boldsymbol{\Delta}_{j,\cdot}-\tilde{\boldsymbol{\Delta}}_{j,\cdot}\right\|_{2}^{2}}\\
&:=d\big(\boldsymbol{\Delta},\tilde{\boldsymbol{\Delta}}\big),
\end{aligned}$$

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,

$$\max_{1\leq j\leq n}\left\|\delta_{l,j}\boldsymbol{\Delta}_{j,\cdot}\right\|_{2}\leq\left\|\boldsymbol{\Delta}\right\|_{2,\infty}\leq\xi$$

according to our assumption. Hence, Talagrand's inequality [CC16, Proposition 1] reveals the existence of some absolute constants $C,c>0$ such that

$$\mathbb{P}\left\{\left\|\boldsymbol{G}_{l}\left(\boldsymbol{\Delta}\right)\right\|-\mathsf{Median}\left[\left\|\boldsymbol{G}_{l}\left(\boldsymbol{\Delta}\right)\right\|\right]\geq\lambda\xi\right\}\leq C\exp\left(-c\lambda^{2}\right).\tag{256}$$

We then proceed to control $\mathsf{Median}\left[\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right)\right\|\right]$. A direct application of Lemma 38 yields

$$\begin{aligned}\mathsf{Median}\left[\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right)\right\|\right] &\leq \sqrt{p\psi^2+\sqrt{p\log\left(4r\right)}\xi\psi+\frac{\xi^2}{3}\log\left(4r\right)}\\ &\leq 2\sqrt{p}\psi,\end{aligned}$$

where the last relation follows if $p\psi^2 \gg \xi^2 \log r$, 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 (256) and taking $\lambda = \sqrt{kr}$ give

$$\mathbb{P}\left\{\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right)\right\| \geq 2\sqrt{p}\psi+\sqrt{kr}\xi\right\} \leq C\exp\left(-ckr\right) \tag{257}$$

for any $k \geq 0$. Furthermore, invoking [AS08, Corollary A.1.14] and using the bound (257), one has

$$\mathbb{P}\left(\sum_{l=1}^{n}\mathbb{1}_{\left\{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta})\right\|\geq 2\sqrt{p}\psi+\sqrt{kr}\xi\right\}} \geq tnC\exp\left(-ckr\right)\right) \leq 2\exp\left(-\frac{t\log t}{2}nC\exp\left(-ckr\right)\right)$$

for any $t \geq 6$. Choose $t = \frac{\alpha\log n}{kC\exp(-ckr)} \geq 6$ to obtain

$$\mathbb{P}\left(\sum_{l=1}^{n}\mathbb{1}_{\left\{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta})\right\|\geq 2\sqrt{p}\psi+\sqrt{kr}\xi\right\}} \geq \frac{\alpha n\log n}{k}\right) \leq 2\exp\left(-\frac{\alpha C}{2}nr\log n\right). \tag{258}$$

So far we have demonstrated that for any fixed $\boldsymbol{\Delta}$ obeying our assumptions, $\sum_{l=1}^{n}\mathbb{1}_{\left\{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta})\right\|\geq 2\sqrt{p}\psi+\sqrt{kr}\xi\right\}}$ is well controlled with exponentially high probability. In order to extend the results to all feasible $\boldsymbol{\Delta}$, we resort to the standard ϵ-net argument. Clearly, due to the homogeneity property of $\left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right)\right\|$, it suffices to restrict attention to the following set:

$$\mathcal{S} = \left\{\boldsymbol{\Delta} \mid \min\left\{\xi,\psi\right\} \leq \left\|\boldsymbol{\Delta}\right\| \leq \psi\right\}. \tag{259}$$

We then proceed with the following steps.

1. Introduce the auxiliary function

$$\chi_l(\boldsymbol{\Delta}) = \begin{cases}1, & \text{if } \left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right)\right\| \geq 4\sqrt{p}\psi+2\sqrt{kr}\xi,\\ \frac{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta})\right\|-2\sqrt{p}\psi-\sqrt{kr}\xi}{2\sqrt{p}\psi-\sqrt{kr}\xi}, & \text{if } \left\|\boldsymbol{G}_l\left(\boldsymbol{\Delta}\right)\right\| \in [2\sqrt{p}\psi+\sqrt{kr}\xi,\ 4\sqrt{p}\psi+2\sqrt{kr}\xi],\\ 0, & \text{else.}\end{cases}$$

   Clearly, this function is sandwiched between two indicator functions

$$\mathbb{1}_{\left\{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta})\right\|\geq 4\sqrt{p}\psi+2\sqrt{kr}\xi\right\}} \leq \chi_l(\boldsymbol{\Delta}) \leq \mathbb{1}_{\left\{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta})\right\|\geq 2\sqrt{p}\psi+\sqrt{kr}\xi\right\}}.$$

   Note that χ_l is more convenient to work with due to continuity.

2. Consider an ϵ-net $\mathcal{N}_\epsilon$ of the set $\mathcal{S}$ as defined in (259) [Tao12, Section 2.3.1]. For any $\epsilon = 1/n^{O(1)}$, one can find such a net with cardinality $\log\left|\mathcal{N}_\epsilon\right| \lesssim nr\log n$. Apply the union bound and (258) to yield

$$\begin{aligned}\mathbb{P}\left(\sum_{l=1}^{n}\chi_l(\boldsymbol{\Delta}) \geq \frac{\alpha n\log n}{k},\ \forall\boldsymbol{\Delta}\in\mathcal{N}_\epsilon\right) &\leq \mathbb{P}\left(\sum_{l=1}^{n}\mathbb{1}_{\left\{\left\|\boldsymbol{G}_l(\boldsymbol{\Delta})\right\|\geq 2\sqrt{p}\psi+\sqrt{kr}\xi\right\}} \geq \frac{\alpha n\log n}{k},\ \forall\boldsymbol{\Delta}\in\mathcal{N}_\epsilon\right)\\ &\leq 2\left|\mathcal{N}_\epsilon\right|\exp\left(-\frac{\alpha C}{2}nr\log n\right) \leq 2\exp\left(-\frac{\alpha C}{4}nr\log n\right),\end{aligned}$$

   as long as α is chosen to be sufficiently large.

3. One can then use the continuity argument to extend the bound to all $\mathbf{\Delta}$ outside the ϵ-net, i.e. with exponentially high probability,

$$\sum_{l=1}^{n} \chi_l(\mathbf{\Delta}) \le \frac{2\alpha n \log n}{k}, \ \forall \mathbf{\Delta} \in \mathcal{S}$$

$$\Longrightarrow \qquad \sum_{l=1}^{n} \mathbb{1}_{\left\{\|\boldsymbol{G}_l(\mathbf{\Delta})\| \ge 4\sqrt{p}\psi + 2\sqrt{kr}\xi\right\}} \le \sum_{l=1}^{n} \chi_l(\mathbf{\Delta}) \le \frac{2\alpha n \log n}{k}, \ \forall \mathbf{\Delta} \in \mathcal{S}$$

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* 41.** *Suppose $\boldsymbol{X}_1, \boldsymbol{X}_2 \in \mathbb{R}^{n\times r}$ are two matrices such that*

$$\left\|\boldsymbol{X}_1 - \boldsymbol{X}^\natural\right\| \left\|\boldsymbol{X}^\natural\right\| \le \sigma_{\min}/2, \tag{260}$$

$$\left\|\boldsymbol{X}_1 - \boldsymbol{X}_2\right\| \left\|\boldsymbol{X}^\natural\right\| \le \sigma_{\min}/4. \tag{261}$$

Denote

$$\boldsymbol{R}_1 := \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \left\|\boldsymbol{X}_1\boldsymbol{R} - \boldsymbol{X}^\natural\right\|_{\mathrm{F}},$$

$$\boldsymbol{R}_2 := \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \left\|\boldsymbol{X}_2\boldsymbol{R} - \boldsymbol{X}^\natural\right\|_{\mathrm{F}}.$$

Then the following two inequalities hold true:

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

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

$$\left\|\boldsymbol{A} - \boldsymbol{X}^{\natural\top}\boldsymbol{X}^\natural\right\| \le \left\|\boldsymbol{X}_1 - \boldsymbol{X}^\natural\right\| \left\|\boldsymbol{X}^\natural\right\| \overset{\text{(i)}}{\le} \sigma_{\min}/2 \qquad \overset{\text{(ii)}}{\Longrightarrow} \qquad \sigma_r\left(\boldsymbol{A}\right) \ge \sigma_{\min}/2, \tag{262}$$

where (i) follows from the assumption (260) and (ii) is a direct application of Weyl's inequality. In addition, $\boldsymbol{A} + \boldsymbol{E} = \boldsymbol{X}_2^\top \boldsymbol{X}^\natural$ is also invertible since

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

where (i) arises from the assumption (261) and (ii) holds given (262). When both $\boldsymbol{A}$ and $\boldsymbol{A}+\boldsymbol{E}$ are invertible, the rotation matrices $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ admit closed-form expressions as follows

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

Moreover,

$$\left\|\boldsymbol{X}_1\right\| \le \left\|\boldsymbol{X}_1 - \boldsymbol{X}^\natural\right\| + \left\|\boldsymbol{X}^\natural\right\| \le \frac{\sigma_{\min}}{2\left\|\boldsymbol{X}^\natural\right\|} \le \frac{\sigma_{\max}}{2\left\|\boldsymbol{X}^\natural\right\|} = \frac{1}{2}\left\|\boldsymbol{X}^\natural\right\|. \tag{263}$$

With these in place, we turn to establishing the claimed bounds. We will focus on the upper bound on $\left\|\boldsymbol{X}_1\boldsymbol{R}_1 - \boldsymbol{X}_2\boldsymbol{R}_2\right\|_{\mathrm{F}}$, as the inequality $\left\|\boldsymbol{X}_1\boldsymbol{R}_1 - \boldsymbol{X}_2\boldsymbol{R}_2\right\| \le \left\|\boldsymbol{X}_1\boldsymbol{R}_1 - \boldsymbol{X}_2\boldsymbol{R}_2\right\|_{\mathrm{F}}$ is immediate. Simple algebra reveals that

$$\begin{aligned}
\left\|\boldsymbol{X}_1\boldsymbol{R}_1 - \boldsymbol{X}_2\boldsymbol{R}_2\right\|_{\mathrm{F}} &= \left\|\left(\boldsymbol{X}_1 - \boldsymbol{X}_2\right)\boldsymbol{R}_2 + \boldsymbol{X}_1\left(\boldsymbol{R}_1 - \boldsymbol{R}_2\right)\right\|_{\mathrm{F}} \\
&\le \left\|\boldsymbol{X}_1 - \boldsymbol{X}_2\right\|_{\mathrm{F}} + \left\|\boldsymbol{X}_1\right\| \left\|\boldsymbol{R}_1 - \boldsymbol{R}_2\right\|_{\mathrm{F}} \\
&\le \left\|\boldsymbol{X}_1 - \boldsymbol{X}_2\right\|_{\mathrm{F}} + 2\left\|\boldsymbol{X}^\natural\right\| \left\|\boldsymbol{R}_1 - \boldsymbol{R}_2\right\|_{\mathrm{F}},
\end{aligned} \tag{264}$$

where the last inequality comes from (263). An application of Lemma 47 leads us to conclude that

$$\left\|\boldsymbol{R}_1 - \boldsymbol{R}_2\right\|_{\mathrm{F}} \le \frac{2}{\sigma_r\left(\boldsymbol{A}\right) + \sigma_{r-1}\left(\boldsymbol{A}\right)}\left\|\boldsymbol{E}\right\|_{\mathrm{F}}$$

$$
\begin{aligned}
&\leq \frac{2}{\sigma_{\min}} \left\| \left(\boldsymbol{X}_2 - \boldsymbol{X}_1 \right)^\top \boldsymbol{X}^\natural \right\|_{\mathrm{F}} && (265) \\
&\leq \frac{2}{\sigma_{\min}} \left\| \boldsymbol{X}_2 - \boldsymbol{X}_1 \right\|_{\mathrm{F}} \left\| \boldsymbol{X}^\natural \right\|, && (266)
\end{aligned}
$$

where (265) utilizes (262). Combine (264) and (266) to reach

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

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

□

Lemma 42. *Let $\boldsymbol{M}, \tilde{\boldsymbol{M}} \in \mathbb{R}^{n\times n}$ be two symmetric matrices with top-r eigendecomposition $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^\top, \tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{U}}^\top$, respectively. Assume $\left\| \boldsymbol{M} - \boldsymbol{M}^\natural \right\| \leq \sigma_{\min}/4$ and $\left\| \tilde{\boldsymbol{M}} - \boldsymbol{M}^\natural \right\| \leq \sigma_{\min}/4$, and suppose $\sigma_{\max}/\sigma_{\min}$ is bounded by some constant $c_1 > 0$, with $\sigma_{\max}$ and $\sigma_{\min}$ the largest and the smallest singular values of $\boldsymbol{M}^\natural$, respectively. If we denote*

$$
\boldsymbol{Q} := \arg \min_{\boldsymbol{R} \in \mathcal{O}^{r\times r}} \left\| \boldsymbol{U}\boldsymbol{R} - \tilde{\boldsymbol{U}} \right\|_{\mathrm{F}},
$$

then there exists some numerical constant $c_3 > 0$ such that

$$
\left\| \boldsymbol{\Sigma}^{1/2} \boldsymbol{Q} - \boldsymbol{Q} \tilde{\boldsymbol{\Sigma}}^{1/2} \right\| \leq \frac{c_3}{\sqrt{\sigma_{\min}}} \left\| \tilde{\boldsymbol{M}} - \boldsymbol{M} \right\| \qquad \text{and} \qquad \left\| \boldsymbol{\Sigma}^{1/2} \boldsymbol{Q} - \boldsymbol{Q} \tilde{\boldsymbol{\Sigma}}^{1/2} \right\|_{\mathrm{F}} \leq \frac{c_3}{\sqrt{\sigma_{\min}}} \left\| (\tilde{\boldsymbol{M}} - \boldsymbol{M}) \boldsymbol{U} \right\|_{\mathrm{F}}.
$$

Proof. Here, we focus on the Frobenius norm; the bound on the operator norm follows from the same argument, and hence we omit the proof.

Since $\|\cdot\|_{\mathrm{F}}$ is unitarily invariant, we have

$$
\left\| \boldsymbol{\Sigma}^{1/2} \boldsymbol{Q} - \boldsymbol{Q} \tilde{\boldsymbol{\Sigma}}^{1/2} \right\|_{\mathrm{F}} = \left\| \boldsymbol{Q}^\top \boldsymbol{\Sigma}^{1/2} \boldsymbol{Q} - \tilde{\boldsymbol{\Sigma}}^{1/2} \right\|_{\mathrm{F}},
$$

where $\boldsymbol{Q}^\top \boldsymbol{\Sigma}^{1/2} \boldsymbol{Q}$ and $\tilde{\boldsymbol{\Sigma}}^{1/2}$ are the matrix square roots of $\boldsymbol{Q}^\top \boldsymbol{\Sigma} \boldsymbol{Q}$ and $\tilde{\boldsymbol{\Sigma}}$, respectively. In view of the matrix square root perturbation bound [Sch92, Lemma 2.1],

$$
\left\| \boldsymbol{\Sigma}^{1/2} \boldsymbol{Q} - \boldsymbol{Q} \tilde{\boldsymbol{\Sigma}}^{1/2} \right\|_{\mathrm{F}} \leq \frac{1}{\sigma_{\min}\left[(\boldsymbol{\Sigma})^{1/2} \right] + \sigma_{\min}\left[(\tilde{\boldsymbol{\Sigma}})^{1/2} \right]} \left\| \boldsymbol{Q}^\top \boldsymbol{\Sigma} \boldsymbol{Q} - \tilde{\boldsymbol{\Sigma}} \right\|_{\mathrm{F}} \leq \frac{1}{\sqrt{\sigma_{\min}}} \left\| \boldsymbol{Q}^\top \boldsymbol{\Sigma} \boldsymbol{Q} - \tilde{\boldsymbol{\Sigma}} \right\|_{\mathrm{F}}, \quad (267)
$$

where the last inequality follows from the lower estimates $\sigma_{\min}(\boldsymbol{\Sigma}) \geq \frac{1}{4}\sigma_{\min}$ and $\sigma_{\min}(\tilde{\boldsymbol{\Sigma}}) \geq \frac{1}{4}\sigma_{\min}$. Recognizing that $\boldsymbol{\Sigma} = \boldsymbol{U}^\top \boldsymbol{M} \boldsymbol{U}$ and $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{U}}^\top \tilde{\boldsymbol{M}} \tilde{\boldsymbol{U}}$, one gets

$$
\begin{aligned}
\left\| \boldsymbol{Q}^\top \boldsymbol{\Sigma} \boldsymbol{Q} - \tilde{\boldsymbol{\Sigma}} \right\|_{\mathrm{F}} &= \left\| (\boldsymbol{U}\boldsymbol{Q})^\top \boldsymbol{M} (\boldsymbol{U}\boldsymbol{Q}) - \tilde{\boldsymbol{U}}^\top \tilde{\boldsymbol{M}} \tilde{\boldsymbol{U}} \right\|_{\mathrm{F}} \\
&\leq \left\| (\boldsymbol{U}\boldsymbol{Q})^\top \boldsymbol{M} (\boldsymbol{U}\boldsymbol{Q}) - (\boldsymbol{U}\boldsymbol{Q})^\top \tilde{\boldsymbol{M}} (\boldsymbol{U}\boldsymbol{Q}) \right\|_{\mathrm{F}} + \left\| (\boldsymbol{U}\boldsymbol{Q})^\top \tilde{\boldsymbol{M}} (\boldsymbol{U}\boldsymbol{Q}) - \tilde{\boldsymbol{U}}^\top \tilde{\boldsymbol{M}} (\boldsymbol{U}\boldsymbol{Q}) \right\|_{\mathrm{F}} \\
&\qquad + \left\| \tilde{\boldsymbol{U}}^\top \tilde{\boldsymbol{M}} (\boldsymbol{U}\boldsymbol{Q}) - \tilde{\boldsymbol{U}}^\top \tilde{\boldsymbol{M}} \tilde{\boldsymbol{U}} \right\|_{\mathrm{F}} \\
&\leq \left\| (\tilde{\boldsymbol{M}} - \boldsymbol{M}) \boldsymbol{U} \right\|_{\mathrm{F}} + 2 \left\| \boldsymbol{U}\boldsymbol{Q} - \tilde{\boldsymbol{U}} \right\|_{\mathrm{F}} \left\| \tilde{\boldsymbol{M}} \right\|. && (268)
\end{aligned}
$$

Invoke the Davis-Kahan sinΘ theorem [DK70] to obtain

$$
\left\| \boldsymbol{U}\boldsymbol{Q} - \tilde{\boldsymbol{U}} \right\|_{\mathrm{F}} \leq \frac{c_2}{\sigma_r(\boldsymbol{M}) - \sigma_{r+1}(\tilde{\boldsymbol{M}})} \left\| (\tilde{\boldsymbol{M}} - \boldsymbol{M}) \boldsymbol{U} \right\|_{\mathrm{F}} \leq \frac{2c_2}{\sigma_{\min}} \left\| (\tilde{\boldsymbol{M}} - \boldsymbol{M}) \boldsymbol{U} \right\|_{\mathrm{F}}, \quad (269)
$$

for some constant $c_2 > 0$, where the last inequality follows from the bounds $\sigma_r(\boldsymbol{M}) \geq \frac{3}{4}\sigma_{\min}$ and $\sigma_{r+1}(\tilde{\boldsymbol{M}}) \leq \frac{1}{4}\sigma_{\min}$. Combine (267), (268), (269) and the fact $\sigma_{\max}/\sigma_{\min} \leq c_1$ to reach

$$
\left\| \boldsymbol{\Sigma}^{1/2} \boldsymbol{Q} - \boldsymbol{Q} \tilde{\boldsymbol{\Sigma}}^{1/2} \right\|_{\mathrm{F}} \leq \frac{c_3}{\sqrt{\sigma_{\min}}} \left\| (\tilde{\boldsymbol{M}} - \boldsymbol{M}) \boldsymbol{U} \right\|_{\mathrm{F}}
$$

for some constant $c_3 > 0$. □

The following lemma quantifies the difference between the rotation matrices for the eigenvectors $\boldsymbol{U}$ and for the scaled eigenvectors $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}^{1/2}$.

***Lemma* 43.** *Let $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ be a symmetric matrix with top-r eigendecomposition $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^\top$. Denote $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}^{1/2}$ and $\boldsymbol{X}^\natural = \boldsymbol{U}^\natural(\boldsymbol{\Sigma}^\natural)^{1/2}$, and define*

$$\hat{\boldsymbol{Q}} := \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \left\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\natural\right\|_{\mathrm{F}} \qquad \text{and} \qquad \hat{\boldsymbol{H}} := \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \left\|\boldsymbol{X}\boldsymbol{R} - \boldsymbol{X}^\natural\right\|_{\mathrm{F}}.$$

Assume $\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\| \le \sigma_{\min}/2$, 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

$$\left\|\hat{\boldsymbol{Q}} - \hat{\boldsymbol{H}}\right\| \le \frac{c_3}{\sigma_{\min}} \left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\|.$$

Proof. We first collect several useful facts about the spectrum of $\boldsymbol{\Sigma}$. Weyl's inequality tells us that $\left\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^\natural\right\| \le \left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\| \le \sigma_{\min}/2$, which further implies that

$$\sigma_r\left(\boldsymbol{\Sigma}\right) \ge \sigma_r\left(\boldsymbol{\Sigma}^\natural\right) - \left\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^\natural\right\| \ge \sigma_{\min}/2 \qquad \text{and} \qquad \left\|\boldsymbol{\Sigma}\right\| \le \left\|\boldsymbol{\Sigma}^\natural\right\| + \left\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^\natural\right\| \le 2\sigma_{\max}.$$

Denote

$$\boldsymbol{Q} = \boldsymbol{U}^\top\boldsymbol{U}^\natural \qquad \text{and} \qquad \boldsymbol{H} = \boldsymbol{X}^\top\boldsymbol{X}^\natural.$$

Simple algebra yields

$$\boldsymbol{H} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{Q}\left(\boldsymbol{\Sigma}^\natural\right)^{1/2} = \underbrace{\boldsymbol{\Sigma}^{1/2}(\boldsymbol{Q} - \hat{\boldsymbol{Q}})\left(\boldsymbol{\Sigma}^\natural\right)^{1/2} + \left(\boldsymbol{\Sigma}^{1/2}\hat{\boldsymbol{Q}} - \hat{\boldsymbol{Q}}\boldsymbol{\Sigma}^{1/2}\right)\left(\boldsymbol{\Sigma}^\natural\right)^{1/2}}_{:=\boldsymbol{E}} + \underbrace{\hat{\boldsymbol{Q}}\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\natural\right)^{1/2}}_{:=\boldsymbol{A}}.$$

It can be easily seen that $\sigma_{r-1}\left(\boldsymbol{A}\right) \ge \sigma_r\left(\boldsymbol{A}\right) \ge \sigma_{\min}/2$, and

$$\begin{aligned}
\|\boldsymbol{E}\| &\le \left\|\boldsymbol{\Sigma}^{1/2}\right\| \cdot \left\|\boldsymbol{Q} - \hat{\boldsymbol{Q}}\right\| \left\|\left(\boldsymbol{\Sigma}^\natural\right)^{1/2}\right\| + \left\|\boldsymbol{\Sigma}^{1/2}\hat{\boldsymbol{Q}} - \hat{\boldsymbol{Q}}\boldsymbol{\Sigma}^{1/2}\right\| \left\|\left(\boldsymbol{\Sigma}^\natural\right)^{1/2}\right\| \\
&\le 2\sigma_{\max} \underbrace{\left\|\boldsymbol{Q} - \hat{\boldsymbol{Q}}\right\|}_{:=\alpha} + \sqrt{\sigma_{\max}} \underbrace{\left\|\boldsymbol{\Sigma}^{1/2}\hat{\boldsymbol{Q}} - \hat{\boldsymbol{Q}}\boldsymbol{\Sigma}^{1/2}\right\|}_{:=\beta},
\end{aligned}$$

which can be controlled as follows.

- Regarding α, use [AFWZ17, Lemma 3] to reach

$$\alpha = \left\|\boldsymbol{Q} - \hat{\boldsymbol{Q}}\right\| \le 4\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\|^2 / \sigma_{\min}^2. \tag{270}$$

- For β, one has

$$\beta \overset{\text{(i)}}{=} \left\|\hat{\boldsymbol{Q}}^\top\boldsymbol{\Sigma}^{1/2}\hat{\boldsymbol{Q}} - \boldsymbol{\Sigma}^{1/2}\right\| \overset{\text{(ii)}}{\le} \frac{1}{2\sigma_r\left(\boldsymbol{\Sigma}^{1/2}\right)} \left\|\hat{\boldsymbol{Q}}^\top\boldsymbol{\Sigma}\hat{\boldsymbol{Q}} - \boldsymbol{\Sigma}\right\| \overset{\text{(iii)}}{=} \frac{1}{2\sigma_r\left(\boldsymbol{\Sigma}^{1/2}\right)} \left\|\boldsymbol{\Sigma}\hat{\boldsymbol{Q}} - \hat{\boldsymbol{Q}}\boldsymbol{\Sigma}\right\|,$$

  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

$$\begin{aligned}
\left\|\boldsymbol{\Sigma}\hat{\boldsymbol{Q}} - \hat{\boldsymbol{Q}}\boldsymbol{\Sigma}\right\| &\le \left\|\boldsymbol{\Sigma}\boldsymbol{Q} - \boldsymbol{Q}\boldsymbol{\Sigma}\right\| + 2\left\|\boldsymbol{\Sigma}\right\|\left\|\boldsymbol{Q} - \hat{\boldsymbol{Q}}\right\| \\
&= \left\|\boldsymbol{U}\left(\boldsymbol{M} - \boldsymbol{M}^\natural\right)\boldsymbol{U}^{\natural\top} + \boldsymbol{Q}\left(\boldsymbol{\Sigma}^\natural - \boldsymbol{\Sigma}\right)\right\| + 2\left\|\boldsymbol{\Sigma}\right\|\left\|\boldsymbol{Q} - \hat{\boldsymbol{Q}}\right\| \\
&\le \left\|\boldsymbol{U}\left(\boldsymbol{M} - \boldsymbol{M}^\natural\right)\boldsymbol{U}^{\natural\top}\right\| + \left\|\boldsymbol{Q}\left(\boldsymbol{\Sigma}^\natural - \boldsymbol{\Sigma}\right)\right\| + 2\left\|\boldsymbol{\Sigma}\right\|\left\|\boldsymbol{Q} - \hat{\boldsymbol{Q}}\right\| \\
&\le 2\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\| + 4\sigma_{\max}\alpha,
\end{aligned}$$

  where the last inequality uses the Weyl's inequality $\|\boldsymbol{\Sigma}^\natural - \boldsymbol{\Sigma}\| \le \|\boldsymbol{M} - \boldsymbol{M}^\natural\|$ and the fact that $\|\boldsymbol{\Sigma}\| \le 2\sigma_{\max}$.

- Rearrange the previous bounds to arrive at

$$\|\boldsymbol{E}\| \leq 2\sigma_{\max}\alpha + \sqrt{\sigma_{\max}}\frac{1}{\sqrt{\sigma_{\min}}}\left(2\left\|\boldsymbol{M} - \boldsymbol{M}^{\natural}\right\| + 4\sigma_{\max}\alpha\right) \leq c_2\left\|\boldsymbol{M} - \boldsymbol{M}^{\natural}\right\|$$

  for some numerical constant $c_2 > 0$.

Recognizing that $\hat{\boldsymbol{Q}} = \mathrm{sgn}\left(\boldsymbol{A}\right)$ (see definition in (115)), we are ready to invoke Lemma 47 to deduce that

$$\left\|\hat{\boldsymbol{Q}} - \hat{\boldsymbol{H}}\right\| \leq \frac{2}{\sigma_{r-1}\left(\boldsymbol{A}\right) + \sigma_r\left(\boldsymbol{A}\right)}\|\boldsymbol{E}\| \leq \frac{c_3}{\sigma_{\min}}\left\|\boldsymbol{M} - \boldsymbol{M}^{\natural}\right\|$$

for some constant $c_3 > 0$. □

Finally, we collect several useful results used throughout our analysis for matrix completion.

***Lemma* 44.** *Fix any small constant $\delta > 0$, and suppose that $m \gg \mu n r \log n$. Then with probability exceeding $1 - O(n^{-10})$, one has*

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

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

Proof. This result has been established in [CR09, Section 4.2] for asymmetric sampling patterns (so that each (i,j), $i \neq j$ is included in Ω independently). It is straightforward to extend the result and proof to symmetric sampling patterns (where each (i,j), $i \geq j$ is included in Ω independently). We omit the proof for conciseness. □

***Lemma* 45.** *Suppose $n^2 p \geq c_0 n \log n$ for some sufficiently large constant $c_0 > 0$. With probability at least $1 - O\left(n^{-10}\right)$, one has for some fixed matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$*

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

where $C > 0$ is some absolute constant.

Proof. See [KMO10a, Lemma 3.2]; similar to the last lemma, the result therein was provided for the asymmetric sampling patterns but can be easily extended to the symmetric case. □

***Lemma* 46.** *Suppose the sample size obeys $n^2 p \geq c_0 n \log^2 n$ for some sufficiently large constant $c_0 > 0$. With probability at least $1 - O\left(n^{-10}\right)$, one has*

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

where $C > 0$ is some universal constant.

Proof. See [CW15, Lemma 11]. □

***Lemma* 47.** *Let $\boldsymbol{A} \in \mathbb{R}^{r\times r}$ be a nonsingular matrix. Then for any $\boldsymbol{E} \in \mathbb{R}^{r\times r}$ with $\|\boldsymbol{E}\| \leq \sigma_{\min}\left(\boldsymbol{A}\right)$ and any unitarily invariant norm $\|\cdot\|$, one has*

$$\left\|\hat{\boldsymbol{H}}\left(\boldsymbol{A} + \boldsymbol{E}\right) - \hat{\boldsymbol{H}}\left(\boldsymbol{A}\right)\right\| \leq \frac{2}{\sigma_{r-1}\left(\boldsymbol{A}\right) + \sigma_r\left(\boldsymbol{A}\right)}\|\boldsymbol{E}\|,$$

where the function $\hat{\boldsymbol{H}} : \mathbb{R}^{r\times r} \to \mathbb{R}^{r\times r}$ is defined as $\hat{\boldsymbol{H}}\left(\boldsymbol{A}\right) = \boldsymbol{A}\left(\boldsymbol{A}^{\top}\boldsymbol{A}\right)^{-1/2}$.

Proof. This is an immediate consequence of [Mat93, Theorem 2.3]. □

Lemma 48 (Orthogonal procrustes problem). *Suppose $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times r}$ and define*

$$\hat{\boldsymbol{R}} := \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \|\boldsymbol{A}\boldsymbol{R} - \boldsymbol{B}\|_{\mathrm{F}}\,.$$

$\hat{\boldsymbol{R}}$ is the minimizer if and only if $\hat{\boldsymbol{R}}^\top \boldsymbol{A}^\top \boldsymbol{B}$ is symmetric and positive semidefinite.

Proof. This is an immediate consequence of [tB77, Theorem 2]. ☐

Lemma 49. *Let $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ be a symmetric matrix with top-r eigendecomposition $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^\top$. Denote*

$$\hat{\boldsymbol{Q}} := \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \left\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\natural\right\|_{\mathrm{F}},$$

and assume $\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\| \le \sigma_{\min}/2$. Then there is some numerical constant $c_3 > 0$ such that

$$\left\|\boldsymbol{U}\hat{\boldsymbol{Q}} - \boldsymbol{U}^\natural\right\| \le \frac{c_3}{\sigma_{\min}} \left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\|.$$

Proof. Define $\boldsymbol{Q} = \boldsymbol{U}^\top\boldsymbol{U}^\natural$. The triangle inequality gives

$$\left\|\boldsymbol{U}\hat{\boldsymbol{Q}} - \boldsymbol{U}^\natural\right\| \le \left\|\boldsymbol{U}\big(\hat{\boldsymbol{Q}} - \boldsymbol{Q}\big)\right\| + \left\|\boldsymbol{U}\boldsymbol{Q} - \boldsymbol{U}^\natural\right\| \le \left\|\hat{\boldsymbol{Q}} - \boldsymbol{Q}\right\| + \left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\natural - \boldsymbol{U}^\natural\right\|. \tag{271}$$

[AFWZ17, Lemma 3] asserts that

$$\left\|\hat{\boldsymbol{Q}} - \boldsymbol{Q}\right\| \le 4\left(\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\|/\sigma_{\min}\right)^2$$

as long as $\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\| \le \sigma_{\min}/2$. For the remaining term in (271), one can use $\boldsymbol{U}^{\natural\top}\boldsymbol{U}^\natural = \boldsymbol{I}_r$ to obtain

$$\left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\natural - \boldsymbol{U}^\natural\right\| = \left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\natural - \boldsymbol{U}^\natural\boldsymbol{U}^{\natural\top}\boldsymbol{U}^\natural\right\| \le \left\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\natural\boldsymbol{U}^{\natural\top}\right\|,$$

which together with the Davis-Kahan sinΘ theorem [DK70] reveals that

$$\left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\natural - \boldsymbol{U}^\natural\right\| \le \frac{c_2}{\sigma_{\min}} \left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\|$$

for some constant $c_2 > 0$. Combine the estimates on $\left\|\hat{\boldsymbol{Q}} - \boldsymbol{Q}\right\|$, $\left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\natural - \boldsymbol{U}^\natural\right\|$ and (271) to reach

$$\left\|\boldsymbol{U}\hat{\boldsymbol{Q}} - \boldsymbol{U}^\natural\right\| \le \left(\frac{4}{\sigma_{\min}}\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\|\right)^2 + \frac{c_2}{\sigma_{\min}}\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\| \le \frac{c_3}{\sigma_{\min}}\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\|$$

for some numerical constant $c_3 > 0$, where we utilizes the fact that $\left\|\boldsymbol{M} - \boldsymbol{M}^\natural\right\|/\sigma_{\min} \le \frac{1}{2}$. ☐

Lemma 50 (Spectral norm of sparse binary matrices). *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*

$$\|\boldsymbol{A}\| \le \sqrt{k_{\mathrm{c}}k_{\mathrm{r}}}.$$

Proof. From the definition of the spectral norm, one has

$$\begin{aligned}
\|\boldsymbol{A}\|^2 = \max_{\|\boldsymbol{x}\|_2=1}\|\boldsymbol{A}\boldsymbol{x}\|_2^2 &= \max_{\|\boldsymbol{x}\|_2=1}\sum_{j=1}^n\left(\sum_{k=1}^n x_k A_{j,k}\right)^2 \\
&\overset{\text{(i)}}{\le} \max_{\|\boldsymbol{x}\|_2=1}\sum_{j=1}^n k_{\mathrm{r}}\left(\sum_{k=1}^n x_k^2 A_{j,k}^2\right) \\
&\overset{\text{(ii)}}{=} \max_{\|\boldsymbol{x}\|_2=1}\sum_{j=1}^n k_{\mathrm{r}}\left(\sum_{k=1}^n x_k^2 A_{j,k}\right),
\end{aligned}$$

where (i) follows from Cauchy-Schwarz and the row sparsity constraint, and (ii) uses the identity $A_{j,k}^2 = A_{j,k}$ since $\boldsymbol{A}$ is a binary matrix. Exchanging the summation gives

$$\|\boldsymbol{A}\|^2 \le k_{\mathrm{r}}\max_{\|\boldsymbol{x}\|_2=1}\sum_{j=1}^n\left(\sum_{k=1}^n x_k^2 A_{j,k}\right) = k_{\mathrm{r}}\max_{\|\boldsymbol{x}\|_2=1}\sum_{k=1}^n x_k^2\left(\sum_{j=1}^n A_{j,k}\right) \le k_{\mathrm{r}}k_{\mathrm{c}}$$

as claimed, where the last line holds since $\sum_{j=1}^n A_{j,k} \le k_{\mathrm{c}}$ and $\sum_{k=1}^n x_k^2 = 1$. ☐

D.3 Technical lemmas for phase retrieval

D.3.1 Discrete Fourier transform matrices

Let $\boldsymbol{B} \in \mathbb{C}^{m\times K}$ be the first K columns of a DFT matrix $\boldsymbol{F} \in \mathbb{C}^{m\times m}$, and denote by $\boldsymbol{b}_l$ the lth column of the matrix $\boldsymbol{B}^*$. By definition,

$$\boldsymbol{b}_l = \frac{1}{\sqrt{m}}\left(1, \omega^{(l-1)}, \omega^{2(l-1)}, \cdots, \omega^{(K-1)(l-1)}\right)^*,$$

where $\omega := e^{-i\frac{2\pi}{m}}$. It is seen that for any $j \neq l$,

$$\boldsymbol{b}_l^*\boldsymbol{b}_j = \frac{1}{m}\sum_{k=0}^{K-1}\omega^{k(l-1)}\cdot\overline{\omega^{k(j-1)}} \overset{\text{(i)}}{=} \frac{1}{m}\sum_{k=0}^{K-1}\omega^{k(l-1)}\cdot\omega^{k(1-j)} = \frac{1}{m}\sum_{k=0}^{K-1}\left(\omega^{l-j}\right)^k \overset{\text{(ii)}}{=} \frac{1}{m}\frac{1-\omega^{K(l-j)}}{1-\omega^{l-j}}. \tag{272}$$

Here, (i) uses $\overline{\omega^{\alpha}} = \omega^{-\alpha}$ for all α, while the last identity (ii) follows from the formula for the sum of a finite geometric series. This leads to the following lemma.

***Lemma* 51.** *For any $m \geq 3$ and any $1 \leq l \leq m$, we have*

$$\sum_{j=1}^{m}|\boldsymbol{b}_l^*\boldsymbol{b}_j| \leq 4\log m.$$

Proof. We first make use of the identity (272) to obtain

$$\sum_{j=1}^{m}|\boldsymbol{b}_l^*\boldsymbol{b}_j| = \|\boldsymbol{b}_l\|_2^2 + \frac{1}{m}\sum_{j:j\neq l}^{m}\left|\frac{1-\omega^{K(l-j)}}{1-\omega^{l-j}}\right| = \frac{K}{m} + \frac{1}{m}\sum_{j:j\neq l}^{m}\left|\frac{\sin\left[K(l-j)\frac{\pi}{m}\right]}{\sin\left[(l-j)\frac{\pi}{m}\right]}\right|,$$

where the last identity follows since $\|\boldsymbol{b}_l\|_2^2 = \frac{K}{m}$ and, for all α,

$$|1-\omega^{\alpha}| = \left|1-e^{-i\frac{2\pi}{m}\alpha}\right| = \left|e^{-i\frac{\pi}{m}\alpha}\left(e^{i\frac{\pi}{m}\alpha} - e^{-i\frac{\pi}{m}\alpha}\right)\right| = 2\left|\sin\left(\alpha\frac{\pi}{m}\right)\right|. \tag{273}$$

Without loss of generality, we focus on the case when $l = 1$ in the sequel. Continue the derivation to get

$$\begin{aligned}
\sum_{j=1}^{m}|\boldsymbol{b}_1^*\boldsymbol{b}_j| &= \frac{K}{m} + \frac{1}{m}\sum_{j=2}^{m}\left|\frac{\sin\left[K(1-j)\frac{\pi}{m}\right]}{\sin\left[(1-j)\frac{\pi}{m}\right]}\right| \overset{\text{(i)}}{\leq} \frac{1}{m}\sum_{j=2}^{m}\left|\frac{1}{\sin\left[(j-1)\frac{\pi}{m}\right]}\right| + \frac{K}{m} \\
&= \frac{1}{m}\left(\sum_{j=2}^{\lfloor\frac{m}{2}\rfloor+1}\left|\frac{1}{\sin\left[(j-1)\frac{\pi}{m}\right]}\right| + \sum_{j=\lfloor\frac{m}{2}\rfloor+2}^{m}\left|\frac{1}{\sin\left[(j-1)\frac{\pi}{m}\right]}\right|\right) + \frac{K}{m} \\
&\overset{\text{(ii)}}{=} \frac{1}{m}\left(\sum_{j=2}^{\lfloor\frac{m}{2}\rfloor+1}\left|\frac{1}{\sin\left[(j-1)\frac{\pi}{m}\right]}\right| + \sum_{j=\lfloor\frac{m}{2}\rfloor+2}^{m}\left|\frac{1}{\sin\left[(m+1-j)\frac{\pi}{m}\right]}\right|\right) + \frac{K}{m},
\end{aligned}$$

where (i) follows from $\left|\sin\left(K(1-j)\frac{\pi}{m}\right)\right| \leq 1$ and $|\sin(x)| = |\sin(-x)|$, and (ii) relies on the fact that $\sin(x) = \sin(\pi - x)$. The property that $\sin(x) \geq \frac{1}{2}x$ for any $x \in \left[0, \frac{\pi}{2}\right]$ allows one to further upper bound it as

$$\begin{aligned}
\sum_{j=1}^{m}|\boldsymbol{b}_1^*\boldsymbol{b}_j| &\leq \frac{1}{m}\left(\sum_{j=2}^{\lfloor\frac{m}{2}\rfloor+1}\frac{2m}{(j-1)\pi} + \sum_{j=\lfloor\frac{m}{2}\rfloor+2}^{m}\frac{2m}{(m+1-j)\pi}\right) + \frac{K}{m} = \frac{2}{\pi}\left(\sum_{k=1}^{\lfloor\frac{m}{2}\rfloor}\frac{1}{k} + \sum_{k=1}^{\lfloor\frac{m+1}{2}\rfloor-1}\frac{1}{k}\right) + \frac{K}{m} \\
&\overset{\text{(i)}}{\leq} \frac{4}{\pi}\sum_{k=1}^{m}\frac{1}{k} + \frac{K}{m} \overset{\text{(ii)}}{\leq} \frac{4}{\pi}(1+\log m) + \frac{K}{m} \overset{\text{(iii)}}{\leq} 4\log m,
\end{aligned}$$

where in (i) we extend the range of the summation, (ii) uses the elementary inequality $\sum_{k=1}^{m}\frac{1}{k} \leq 1 + \log m$ and (iii) holds true as long as $m \geq 3$. $\square$

Next we consider the difference of two inner products. Given (272), we can obtain for $j \neq l$ and $j \neq 1$

$$\begin{aligned}
\left|(\boldsymbol{b}_l-\boldsymbol{b}_1)^*\boldsymbol{b}_j\right| &= \frac{1}{m}\left|\frac{1-\omega^{K(l-j)}}{1-\omega^{l-j}}-\frac{1-\omega^{K(1-j)}}{1-\omega^{1-j}}\right| \\
&= \frac{1}{m}\left|\frac{1-\omega^{K(l-j)}}{1-\omega^{l-j}}-\frac{1-\omega^{K(1-j)}}{1-\omega^{l-j}}+\frac{1-\omega^{K(1-j)}}{1-\omega^{l-j}}-\frac{1-\omega^{K(1-j)}}{1-\omega^{1-j}}\right| \\
&= \frac{1}{m}\left|\frac{\omega^{K(1-j)}-\omega^{K(l-j)}}{1-\omega^{l-j}}+\left(\omega^{l-j}-\omega^{1-j}\right)\frac{1-\omega^{K(1-j)}}{\left(1-\omega^{l-j}\right)\left(1-\omega^{1-j}\right)}\right| \\
&\leq \frac{1}{m}\left|\frac{\omega^{K(1-j)}-\omega^{K(l-j)}}{1-\omega^{l-j}}\right|+\frac{1}{m}\left|\left(\omega^{l-j}-\omega^{1-j}\right)\frac{1-\omega^{K(1-j)}}{\left(1-\omega^{l-j}\right)\left(1-\omega^{1-j}\right)}\right|,
\end{aligned}$$

where the last relation is the triangle inequality. Observing that $|\omega^{\alpha}|=1$ for all α, we can proceed to upper bound $\left|(\boldsymbol{b}_l-\boldsymbol{b}_1)^*\boldsymbol{b}_j\right|$ as

$$\left|(\boldsymbol{b}_l-\boldsymbol{b}_1)^*\boldsymbol{b}_j\right| \leq \frac{1}{m}\left|\frac{1-\omega^{K(l-1)}}{1-\omega^{l-j}}\right|+\frac{1}{m}\left|\left(1-\omega^{1-l}\right)\frac{1-\omega^{K(1-j)}}{\left(1-\omega^{l-j}\right)\left(1-\omega^{1-j}\right)}\right|.$$

The identity (273) allows us to rewrite this bound as

$$\left|(\boldsymbol{b}_l-\boldsymbol{b}_1)^*\boldsymbol{b}_j\right| \leq \frac{1}{m}\left|\frac{1}{\sin\left[(l-j)\frac{\pi}{m}\right]}\right|\left\{\left|\sin\left[K(l-1)\frac{\pi}{m}\right]\right|+\left|\frac{\sin\left[(1-l)\frac{\pi}{m}\right]\sin\left[K(1-j)\frac{\pi}{m}\right]}{\sin\left[(1-j)\frac{\pi}{m}\right]}\right|\right\}. \tag{274}$$

We can further obtain the following simpler upper bounds on $\left|(\boldsymbol{b}_l-\boldsymbol{b}_1)^*\boldsymbol{b}_j\right|$.

Lemma 52. *For all $0 \leq l-1 \leq \tau \leq \left\lfloor \frac{m}{10} \right\rfloor$, we have*

$$\left|(\boldsymbol{b}_l-\boldsymbol{b}_1)^*\boldsymbol{b}_j\right| \leq \begin{cases} \frac{4\tau}{(j-l)}\frac{K}{m}+\frac{8\tau/\pi}{(j-l)^2} & \text{for} \quad l+\tau \leq j \leq \left\lfloor \frac{m}{2}\right\rfloor+1, \\ \frac{4\tau}{m-(j-l)}\frac{K}{m}+\frac{8\tau/\pi}{[m-(j-1)]^2} & \text{for} \quad \left\lfloor \frac{m}{2}\right\rfloor+l \leq j \leq m-\tau. \end{cases}$$

In addition, for any j and l, the following uniform upper bound holds

$$\left|(\boldsymbol{b}_l-\boldsymbol{b}_1)^*\boldsymbol{b}_j\right| \leq 2\frac{K}{m}.$$

Proof. Combined with the facts that $|\sin x| \leq 2|x|$ for all x and $|\sin x| \leq 1$, we can upper bound (274) as

$$\left|(\boldsymbol{b}_l-\boldsymbol{b}_1)^*\boldsymbol{b}_j\right| \leq \frac{1}{m}\left|\frac{1}{\sin\left[(l-j)\frac{\pi}{m}\right]}\right|\left\{2K\tau\frac{\pi}{m}+\left|\frac{2\tau\frac{\pi}{m}}{\sin\left[(1-j)\frac{\pi}{m}\right]}\right|\right\},$$

where we also utilize the assumption that $0 \leq l-1 \leq \tau$. Then for $l+\tau \leq j \leq \left\lfloor \frac{m}{2}\right\rfloor+1$, 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 $\sin(x) \geq \frac{1}{2}x$ for any $x \in \left[0,\frac{\pi}{2}\right]$, we arrive at

$$\left|(\boldsymbol{b}_l-\boldsymbol{b}_1)^*\boldsymbol{b}_j\right| \leq \frac{2}{(j-l)\pi}\left(2K\tau\frac{\pi}{m}+\frac{4\tau}{j-1}\right)=\frac{4\tau}{j-l}\frac{K}{m}+\frac{8\tau/\pi}{(j-l)^2}.$$

Similarly we can obtain the upper bound for $\left\lfloor \frac{m}{2}\right\rfloor+l \leq j \leq m-\tau$ using nearly identical argument (which is omitted).

The uniform upper bound can be justified as follows

$$\left|(\boldsymbol{b}_l-\boldsymbol{b}_1)^*\boldsymbol{b}_j\right| \leq \left(\|\boldsymbol{b}_l\|_2+\|\boldsymbol{b}_1\|_2\right)\|\boldsymbol{b}_j\|_2 \leq 2\frac{K}{m}.$$

The last relation follows from the identity $\|\boldsymbol{b}_l\|_2^2=\frac{K}{m}$ for all $1 \leq l \leq m$. □

With these detailed estimates in place, we are ready to list the following two consequences.

***Lemma* 53.** *Suppose $m \geq C\tau K \log^4 m$ for some sufficiently large constant $C > 0$. If $0 \leq l - 1 \leq \tau$, then one has*

$$\sum_{j=1}^{m} \left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^* \boldsymbol{b}_j\right| \leq c \frac{1}{\log^2 m},$$

where $c > 0$ is any sufficiently small constant independent of m and K.

Proof. For some $c_0 > 0$, we can split the index set $[m]$ into three disjoint sets, i.e. $[m] = \mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3$, where

$$\begin{aligned} \mathcal{A}_1 &= \left\{ j : l + c_0 \tau \log^2 m \leq j \leq \left\lfloor \frac{m}{2} \right\rfloor + 1 \right\}, \\ \mathcal{A}_2 &= \left\{ j : \left\lfloor \frac{m}{2} \right\rfloor + l \leq j \leq m - c_0 \tau \log^2 m \right\}, \\ \text{and} \qquad \mathcal{A}_3 &= [m] \setminus (\mathcal{A}_1 \cup \mathcal{A}_2). \end{aligned}$$

With this decomposition at hand, we can write

$$\sum_{j=1}^{m} \left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^* \boldsymbol{b}_j\right| = \sum_{j \in \mathcal{A}_1} \left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^* \boldsymbol{b}_j\right| + \sum_{j \in \mathcal{A}_2} \left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^* \boldsymbol{b}_j\right| + \sum_{j \in \mathcal{A}_3} \left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^* \boldsymbol{b}_j\right|.$$

We first look at $\mathcal{A}_1$. By Lemma 52, one has for any $j \in \mathcal{A}_1$,

$$\left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^* \boldsymbol{b}_j\right| \leq \frac{4\tau}{j - l} \frac{K}{m} + \frac{8\tau/\pi}{(j - l)^2},$$

and hence

$$\begin{aligned} \sum_{j \in \mathcal{A}_1} \left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^* \boldsymbol{b}_j\right| &\leq \sum_{j = l + c_0 \tau \log^2 m}^{\left\lfloor \frac{m}{2} \right\rfloor + 1} \left(\frac{4\tau}{j - l} \frac{K}{m} + \frac{8\tau/\pi}{(j - l)^2} \right) \leq \frac{4\tau K}{m} \sum_{k=1}^{m} \frac{1}{k} + \frac{8\tau}{\pi} \sum_{k = c_0 \tau \log^2 m}^{m} \frac{1}{k^2} \\ &\leq 8\tau \frac{K}{m} \log m + \frac{16\tau}{\pi} \frac{1}{c_0 \tau \log^2 m}, \end{aligned}$$

where the last inequality arises from the elementary inequality

$$\sum_{k=1}^{m} \frac{1}{k} \leq 1 + \log m \leq 2 \log m \qquad \text{and} \qquad \sum_{k=c}^{m} \frac{1}{k^2} \leq \frac{2}{c}.$$

Similarly, for $j \in \mathcal{A}_2$, we have

$$\left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^* \boldsymbol{b}_j\right| \leq \frac{4\tau}{m - (j - l)} \frac{K}{m} + \frac{8\tau/\pi}{[m - (j - 1)]^2},$$

which in turn implies

$$\sum_{j \in \mathcal{A}_2} \left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^* \boldsymbol{b}_j\right| \leq 8\tau \frac{K}{m} \log m + \frac{16\tau}{\pi} \frac{1}{c_0 \tau \log^2 m}.$$

Regarding $j \in \mathcal{A}_3$, we observe that

$$|\mathcal{A}_3| \leq 2 \left(c_0 \tau \log^2 m + l\right) \leq 2 \left(c_0 \tau \log^2 m + \tau\right) \leq 4 c_0 \tau \log^2 m.$$

This together with the naive bound $\left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^* \boldsymbol{b}_j\right| \leq 2\frac{K}{m}$ gives

$$\sum_{j \in \mathcal{A}_3} \left|(\boldsymbol{b}_l - \boldsymbol{b}_1)^* \boldsymbol{b}_j\right| \leq 2 \frac{K}{m} |\mathcal{A}_3| \leq \frac{8 c_0 \tau K \log^2 m}{m}.$$

The previous three estimates taken collectively yield

$$\sum_{j=1}^{m}\left|\left(\boldsymbol{b}_l-\boldsymbol{b}_1\right)^*\boldsymbol{b}_j\right| \le \frac{16\tau K\log m}{m}+\frac{32\tau}{\pi}\frac{1}{c_0\tau\log^2 m}+\frac{8c_0\tau K\log^2 m}{m} \le c\frac{1}{\log^2 m}$$

for any small constant $c>0$, as long as $c_0>0$ and $m/(c_0\tau K\log^4 m)$ are both taken to be sufficiently large. $\square$

Lemma 54. *Suppose that $m \ge C\tau K\log m$ for some sufficiently large constant $C>0$. Then we have*

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

where $c>0$ is any sufficiently small constant independent of m and K.

Proof. The proof strategy is similar to the one used for Lemma 53. First notice that

$$\left|\boldsymbol{b}_1^*\left(\boldsymbol{b}_{k\tau+j}-\boldsymbol{b}_{k\tau+1}\right)\right| = \left|\left(\boldsymbol{b}_m-\boldsymbol{b}_{m+1-j}\right)^*\boldsymbol{b}_{k\tau}\right|.$$

As before, for some $c_1>0$, we can split the index set $\left\{1,\cdots,\lfloor\frac{m}{\tau}\rfloor\right\}$ into three disjoint sets, i.e. $\mathcal{B}_1\cup\mathcal{B}_2\cup\mathcal{B}_3$, where

$$\begin{aligned}
\mathcal{B}_1 &= \left\{k: c_1\le k\le \left\lfloor\left(\left\lfloor\frac{m}{2}\right\rfloor+1-j\right)/\tau\right\rfloor\right\},\\
\mathcal{B}_2 &= \left\{k: \left\lfloor\left(\left\lfloor\frac{m}{2}\right\rfloor+1-j\right)/\tau\right\rfloor+1\le k\le \left\lfloor\left(m+1-j\right)/\tau\right\rfloor-c_1\right\},\\
\text{and}\qquad \mathcal{B}_3 &= \left\{1,\cdots,\left\lfloor\frac{m}{\tau}\right\rfloor\right\}\setminus\left(\mathcal{B}_1\cup\mathcal{B}_2\right).
\end{aligned}$$

By Lemma 52, one has

$$\left|\left(\boldsymbol{b}_m-\boldsymbol{b}_{m+1-j}\right)^*\boldsymbol{b}_{k\tau}\right| \le \frac{4\tau}{k\tau}\frac{K}{m}+\frac{8\tau/\pi}{(k\tau)^2},\qquad k\in\mathcal{B}_1.$$

Hence for any $k\in\mathcal{B}_1$,

$$\sqrt{\sum_{j=1}^{\tau}\left|\boldsymbol{b}_1^*\left(\boldsymbol{b}_{k\tau+j}-\boldsymbol{b}_{k\tau+1}\right)\right|^2} \le \sqrt{\tau}\left(\frac{4\tau}{k\tau}\frac{K}{m}+\frac{8\tau/\pi}{(k\tau)^2}\right),$$

which further implies that

$$\sum_{k\in\mathcal{B}_1}\sqrt{\sum_{j=1}^{\tau}\left|\boldsymbol{b}_1^*\left(\boldsymbol{b}_{k\tau+j}-\boldsymbol{b}_{k\tau+1}\right)\right|^2} \le \sqrt{\tau}\sum_{k=c_1}^{m}\left(\frac{4\tau}{k\tau}\frac{K}{m}+\frac{8\tau/\pi}{(k\tau)^2}\right) \le 8\sqrt{\tau}\frac{K\log m}{m}+\frac{16}{\pi}\frac{1}{\sqrt{\tau}}\frac{1}{c_1}.$$

A similar bound can be obtained for $k\in\mathcal{B}_2$.

For the remaining set $\mathcal{B}_3$, observe that

$$|\mathcal{B}_3|\le 2c_1.$$

This together with the crude upper bound $\left|\left(\boldsymbol{b}_l-\boldsymbol{b}_1\right)^*\boldsymbol{b}_j\right|\le 2\frac{K}{m}$ gives

$$\sum_{k\in\mathcal{B}_3}\sqrt{\sum_{j=1}^{\tau}\left|\boldsymbol{b}_1^*\left(\boldsymbol{b}_{k\tau+j}-\boldsymbol{b}_{k\tau+1}\right)\right|^2} \le |\mathcal{B}_3|\sqrt{\tau\max_j\left|\boldsymbol{b}_1^*\left(\boldsymbol{b}_{k\tau+j}-\boldsymbol{b}_{k\tau+1}\right)\right|^2} \le |\mathcal{B}_3|\sqrt{\tau}\cdot\frac{2K}{m} \le \frac{4c_1\sqrt{\tau}K}{m}.$$

The previous three estimates taken collectively yield

$$\sum_{k=0}^{\lfloor\frac{m}{\tau}\rfloor}\sqrt{\sum_{j=1}^{\tau}\left|\boldsymbol{b}_1^*\left(\boldsymbol{b}_{k\tau+j}-\boldsymbol{b}_{k\tau+1}\right)\right|^2} \le 2\left(8\sqrt{\tau}\frac{K\log m}{m}+\frac{16}{\pi}\frac{1}{\sqrt{\tau}}\frac{1}{c_1}\right)+\frac{4c_1\sqrt{\tau}K}{m} \le c\frac{1}{\sqrt{\tau}},$$

as long as c_1 and $m/(c_1\tau K\log m)$ are both taken to be sufficiently large. $\square$

D.3.2 Complex-valued alignment

Let $g_{\boldsymbol{h},\boldsymbol{x}}(\cdot) : \mathbb{C} \to \mathbb{R}$ be a real-valued function defined as

$$g_{\boldsymbol{h},\boldsymbol{x}}(\alpha) := \left\| \frac{1}{\overline{\alpha}}\boldsymbol{h} - \boldsymbol{h}^{\natural} \right\|_2^2 + \left\| \alpha \boldsymbol{x} - \boldsymbol{x}^{\natural} \right\|_2^2,$$

which is the key function when defining $\mathrm{dist}(\cdot,\cdot)$ in (32). The following lemma reveals that the minimizer of $g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)$ cannot be far away from β as long as $\left(\frac{1}{\overline{\beta}}\boldsymbol{h}, \beta\boldsymbol{x}\right)$ is sufficiently close to $(\boldsymbol{h}^{\natural}, \boldsymbol{x}^{\natural})$.

***Lemma* 55.** *Assume theres exists $\beta \in \mathbb{C}$ and $\frac{1}{2} \le |\beta| \le \frac{3}{2}$ such that $\max\left\{ \left\| \frac{1}{\overline{\beta}}\boldsymbol{h} - \boldsymbol{h}^{\natural} \right\|_2, \left\| \beta\boldsymbol{x} - \boldsymbol{x}^{\natural} \right\|_2 \right\} \le \delta \le \frac{1}{4}$. Denote by $\hat{\alpha}$ the minimizer of $g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)$, then we must have*

$$\left| |\hat{\alpha}| - |\beta| \right| \le |\hat{\alpha} - \beta| \le 18\delta.$$

Proof. The first inequality is a direct consequence of the triangle inequality. Hence we concentrate on the second one. Notice that

$$g_{\boldsymbol{h},\boldsymbol{x}}(\beta) = \left\| \frac{1}{\overline{\beta}}\boldsymbol{h} - \boldsymbol{h}^{\natural} \right\|_2^2 + \left\| \beta\boldsymbol{x} - \boldsymbol{x}^{\natural} \right\|_2^2 \le 2\delta^2, \tag{275}$$

where we use the assumption that $\max\left\{ \left\| \frac{1}{\overline{\beta}}\boldsymbol{h} - \boldsymbol{h}^{\natural} \right\|_2, \left\| \beta\boldsymbol{x} - \boldsymbol{x}^{\natural} \right\|_2 \right\} \le \delta$. This immediately implies that $g_{\boldsymbol{h},\boldsymbol{x}}(\hat{\alpha}) \le 2\delta^2$. 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 can lower bound $g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)$ as follows:

$$\begin{aligned} g_{\boldsymbol{h},\boldsymbol{x}}(\alpha) &\ge \left\| \alpha\boldsymbol{x} - \boldsymbol{x}^{\natural} \right\|_2^2 = \left\| (\alpha - \beta)\boldsymbol{x} + \left(\beta\boldsymbol{x} - \boldsymbol{x}^{\natural}\right) \right\|_2^2 \\ &= |\alpha - \beta|^2 \|\boldsymbol{x}\|_2^2 + \left\| \beta\boldsymbol{x} - \boldsymbol{x}^{\natural} \right\|_2^2 + 2\mathrm{Re}\left[(\alpha - \beta)\left(\beta\boldsymbol{x} - \boldsymbol{x}^{\natural}\right)^* \boldsymbol{x} \right] \\ &\ge |\alpha - \beta|^2 \|\boldsymbol{x}\|_2^2 - 2|\alpha - \beta| \left| \left(\beta\boldsymbol{x} - \boldsymbol{x}^{\natural}\right)^* \boldsymbol{x} \right|. \end{aligned}$$

Given that $\left\| \beta\boldsymbol{x} - \boldsymbol{x}^{\natural} \right\|_2 \le \delta \le \frac{1}{4}$ and $\left\| \boldsymbol{x}^{\natural} \right\|_2 = 1$, we have

$$\|\beta\boldsymbol{x}\|_2 \ge \|\boldsymbol{x}^{\natural}\|_2 - \left\| \beta\boldsymbol{x} - \boldsymbol{x}^{\natural} \right\|_2 \ge 1 - \delta \ge \frac{3}{4},$$

which together with the fact that $\frac{1}{2} \le |\beta| \le \frac{3}{2}$ implies

$$\|\boldsymbol{x}\|_2 \ge 1/2 \qquad \text{and} \qquad \|\boldsymbol{x}\|_2 \le 2$$

and

$$\left| \left(\beta\boldsymbol{x} - \boldsymbol{x}^{\natural}\right)^* \boldsymbol{x} \right| \le \left\| \beta\boldsymbol{x} - \boldsymbol{x}^{\natural} \right\|_2 \|\boldsymbol{x}\|_2 \le 2\delta.$$

Taking the previous estimates collectively yields

$$g_{\boldsymbol{h},\boldsymbol{x}}(\alpha) \ge \frac{1}{4}|\alpha - \beta|^2 - 4\delta|\alpha - \beta|.$$

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 (275). 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* 56.** *Assume that $\max\left\{ \left\| \boldsymbol{h} - \boldsymbol{h}^{\natural} \right\|_2, \left\| \boldsymbol{x} - \boldsymbol{x}^{\natural} \right\|_2 \right\} \le \delta$ for some sufficiently small constant $\delta > 0$. Then, for any α satisfying $|\alpha - 1| \le 18\delta$ and any $u, v \in \mathbb{C}$, one has*

$$[u^*, v^*]\, \nabla^2 g_{\boldsymbol{h},\boldsymbol{x}}(\alpha) \begin{bmatrix} u \\ v \end{bmatrix} \ge \frac{1}{2}\left(|u|^2 + |v|^2 \right),$$

where $\nabla^2 g_{\boldsymbol{h},\boldsymbol{x}}(\cdot)$ stands for the Wirtinger Hessian of $g_{\boldsymbol{h},\boldsymbol{x}}(\cdot)$.

Proof. Use $g_{\boldsymbol{h},\boldsymbol{x}}(\alpha,\overline{\alpha})$ and $g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)$ interchangeably. The Wirtinger gradient and Hessian of $g_{\boldsymbol{h},\boldsymbol{x}}(\cdot)$ can be calculated as

$$\nabla g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)=\begin{bmatrix}\frac{\partial g_{\boldsymbol{h},\boldsymbol{x}}(\alpha,\overline{\alpha})}{\partial\alpha}\\ \frac{\partial g_{\boldsymbol{h},\boldsymbol{x}}(\alpha,\overline{\alpha})}{\partial\overline{\alpha}}\end{bmatrix}=\begin{bmatrix}\alpha\|\boldsymbol{x}\|_2^2-\boldsymbol{x}^*\boldsymbol{x}^\natural-\alpha^{-1}(\overline{\alpha})^{-2}\|\boldsymbol{h}\|_2^2+(\overline{\alpha})^{-2}\boldsymbol{h}^{\natural*}\boldsymbol{h}\\ \overline{\alpha}\|\boldsymbol{x}\|_2^2-\boldsymbol{x}^{\natural*}\boldsymbol{x}-(\overline{\alpha})^{-1}\alpha^{-2}\|\boldsymbol{h}\|_2^2+\alpha^{-2}\boldsymbol{h}^*\boldsymbol{h}^\natural\end{bmatrix};\tag{276}$$

$$\nabla^2 g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)=\begin{bmatrix}\|\boldsymbol{x}\|_2^2+|\alpha|^{-4}\|\boldsymbol{h}\|_2^2 & 2\alpha^{-1}(\overline{\alpha})^{-3}\|\boldsymbol{h}\|_2^2-2(\overline{\alpha})^{-3}\boldsymbol{h}^{\natural*}\boldsymbol{h}\\ 2(\overline{\alpha})^{-1}\alpha^{-3}\|\boldsymbol{h}\|_2^2-2\alpha^{-3}\boldsymbol{h}^*\boldsymbol{h}^\natural & \|\boldsymbol{x}\|_2^2+|\alpha|^{-4}\|\boldsymbol{h}\|_2^2\end{bmatrix}.$$

For any $u,v\in\mathbb{C}$, one has

$$[u^*,v^*]\nabla^2 g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)\begin{bmatrix}u\\ v\end{bmatrix}=\underbrace{\left(\|\boldsymbol{x}\|_2^2+|\alpha|^{-4}\|\boldsymbol{h}\|_2^2\right)}_{:=\beta_1}\left(|u|^2+|v|^2\right)+\underbrace{2\operatorname{Re}\left[u^*v\left(2\alpha^{-1}(\overline{\alpha})^{-3}\|\boldsymbol{h}\|_2^2-2(\overline{\alpha})^{-3}\boldsymbol{h}^{\natural*}\boldsymbol{h}\right)\right]}_{:=\beta_2}.$$

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 $\max\left\{\|\boldsymbol{h}-\boldsymbol{h}^\natural\|_2,\|\boldsymbol{x}-\boldsymbol{x}^\natural\|_2\right\}\le\delta$, one necessarily has

$$1-\delta\le\|\boldsymbol{x}\|_2\le 1+\delta\qquad\text{and}\qquad 1-\delta\le\|\boldsymbol{h}\|_2\le 1+\delta.$$

Thus, for any α obeying $|\alpha-1|\le 18\delta$, one has

$$\beta_1\ge\left(1+|\alpha|^{-4}\right)(1-\delta)^2\ge\left(1+(1+18\delta)^{-4}\right)(1-\delta)^2\ge 1$$

as long as $\delta>0$ is sufficiently small. Regarding the second term β_2, we rely on the assumptions $|\alpha-1|\le 18\delta$, $\|\boldsymbol{x}\|_2\le 1+\delta$ and $\|\boldsymbol{h}\|_2\le 1+\delta$ to get

$$\begin{aligned}|\beta_2|&\le 2|u||v||\alpha|^{-3}\left|\left(\alpha^{-1}-1\right)\|\boldsymbol{h}\|_2^2-(\boldsymbol{h}^\natural-\boldsymbol{h})^*\boldsymbol{h}\right|\\ &\le 2|u||v||\alpha|^{-3}\left(\left|\alpha^{-1}-1\right|\|\boldsymbol{h}\|_2^2+\|\boldsymbol{h}-\boldsymbol{h}^\natural\|_2\|\boldsymbol{h}\|_2\right)\\ &\le 2|u||v|(1-18\delta)^{-3}\left(\frac{18\delta}{1-18\delta}(1+\delta)^2+\delta(1+\delta)\right)\\ &\lesssim\delta\left(|u|^2+|v|^2\right)\end{aligned}$$

for $\delta>0$ sufficiently small. Combining the previous bounds on β_1 and β_2, we arrive at

$$[u^*,v^*]\nabla^2 g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)\begin{bmatrix}u\\ v\end{bmatrix}\ge(1-O(\delta))\left(|u|^2+|v|^2\right)\ge\frac{1}{2}\left(|u|^2+|v|^2\right)$$

as long as δ is sufficiently small. This completes the proof. $\square$

***Lemma* 57.** *Suppose that the vectors $\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{h}_1,\boldsymbol{h}_2\in\mathbb{C}^K$ satisfy*

$$\max\left\{\|\boldsymbol{x}_1-\boldsymbol{x}^\natural\|_2,\|\boldsymbol{h}_1-\boldsymbol{h}^\natural\|_2,\|\boldsymbol{x}_2-\boldsymbol{x}^\natural\|_2,\|\boldsymbol{h}_2-\boldsymbol{h}^\natural\|_2\right\}\le\delta\tag{277}$$

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

$$|\alpha_1-\alpha_2|\lesssim\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2+\|\boldsymbol{h}_1-\boldsymbol{h}_2\|_2.$$

Proof. Since α_1 minimizes $g_{\boldsymbol{h}_1,\boldsymbol{x}_1}(\alpha)$, we must have by the mean value theorem (see Appendix D.3.4)

$$\begin{aligned}&g_{\boldsymbol{h}_1,\boldsymbol{x}_1}(\alpha_2)\ge g_{\boldsymbol{h}_1,\boldsymbol{x}_1}(\alpha_1)\\ &\qquad=g_{\boldsymbol{h}_1,\boldsymbol{x}_1}(\alpha_2)+\nabla g_{\boldsymbol{h}_1,\boldsymbol{x}_1}(\alpha_2)^*\begin{bmatrix}\alpha_1-\alpha_2\\ \overline{\alpha_1-\alpha_2}\end{bmatrix}+\frac{1}{2}\left(\overline{\alpha_1-\alpha_2},\alpha_1-\alpha_2\right)\nabla^2 g_{\boldsymbol{h}_1,\boldsymbol{x}_1}(\tilde{\alpha})\begin{bmatrix}\alpha_1-\alpha_2\\ \overline{\alpha_1-\alpha_2}\end{bmatrix},\end{aligned}$$

where $\tilde{\alpha}$ 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 $g_{\boldsymbol{h}_1,\boldsymbol{x}_1}(\cdot)$, respectively. Rearrange the previous inequality to obtain

$$|\alpha_1-\alpha_2|\lesssim\frac{\|\nabla g_{\boldsymbol{h}_1,\boldsymbol{x}_1}(\alpha_2)\|_2}{\lambda_{\min}\left(\nabla^2 g_{\boldsymbol{h}_1,\boldsymbol{x}_1}(\tilde{\alpha})\right)}.$$

This requires us to examine the Wirtinger gradient and Hessian of $g_{\boldsymbol{h}_1,\boldsymbol{x}_1}(\cdot)$.

1. Regarding the Wirtinger Hessian, by the assumption (277), we can invoke Lemma 55 with $\beta = 1$ to reach $\max\{|\alpha_1 - 1|, |\alpha_2 - 1|\} \le 18\delta$. This together with Lemma 56 implies

$$\lambda_{\min}\left(\nabla^2 g_{\boldsymbol{h}_1,\boldsymbol{x}_1}(\tilde{\alpha})\right) \ge 1/2,$$

since $\tilde{\alpha}$ lies between α_1 and α_2.

2. For the Wirtinger gradient, since α_2 is the minimizer of $g_{\boldsymbol{h}_2,\boldsymbol{x}_2}(\alpha)$, one necessarily has

$$\nabla g_{\boldsymbol{h}_2,\boldsymbol{x}_2}(\alpha_2) = \boldsymbol{0}.$$

This further gives

$$\left\|\nabla g_{\boldsymbol{h}_1,\boldsymbol{x}_1}(\alpha_2)\right\|_2 = \left\|\nabla g_{\boldsymbol{h}_1,\boldsymbol{x}_1}(\alpha_2) - \nabla g_{\boldsymbol{h}_2,\boldsymbol{x}_2}(\alpha_2)\right\|_2.$$

Substitute into the gradient expression (cf. (276)) to reach

$$\begin{aligned}
&\left\|\nabla g_{\boldsymbol{h}_1,\boldsymbol{x}_1}(\alpha_2) - \nabla g_{\boldsymbol{h}_2,\boldsymbol{x}_2}(\alpha_2)\right\|_2 \\
&\quad = \sqrt{2}\Big|\left[\alpha_2 \|\boldsymbol{x}_1\|_2^2 - \boldsymbol{x}_1^*\boldsymbol{x}^\natural - \alpha_2^{-1}(\overline{\alpha_2})^{-2}\|\boldsymbol{h}_1\|_2^2 + (\overline{\alpha_2})^{-2}\boldsymbol{h}^{\natural *}\boldsymbol{h}_1\right] \\
&\qquad - \left[\alpha_2 \|\boldsymbol{x}_2\|_2^2 - \boldsymbol{x}_2^*\boldsymbol{x}^\natural - \alpha_2^{-1}(\overline{\alpha_2})^{-2}\|\boldsymbol{h}_2\|_2^2 + (\overline{\alpha_2})^{-2}\boldsymbol{h}^{\natural *}\boldsymbol{h}_2\right]\Big| \\
&\quad \lesssim |\alpha_2|\left|\|\boldsymbol{x}_1\|_2^2 - \|\boldsymbol{x}_2\|_2^2\right| + \left|\boldsymbol{x}_1^*\boldsymbol{x}^\natural - \boldsymbol{x}_2^*\boldsymbol{x}^\natural\right| + \frac{1}{|\alpha_2|^3}\left|\|\boldsymbol{h}_1\|_2^2 - \|\boldsymbol{h}_2\|_2^2\right| + \frac{1}{|\alpha_2|^2}\left|\boldsymbol{h}^{\natural *}\boldsymbol{h}_1 - \boldsymbol{h}^{\natural *}\boldsymbol{h}_2\right| \\
&\quad \lesssim |\alpha_2|\left|\|\boldsymbol{x}_1\|_2^2 - \|\boldsymbol{x}_2\|_2^2\right| + \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 + \frac{1}{|\alpha_2|^3}\left|\|\boldsymbol{h}_1\|_2^2 - \|\boldsymbol{h}_2\|_2^2\right| + \frac{1}{|\alpha_2|^2}\|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2,
\end{aligned}$$

where the last line follows from the triangle inequality. It is straightforward to see that

$$\frac{1}{2} \le |\alpha_2| \le 2, \qquad \left|\|\boldsymbol{x}_1\|_2^2 - \|\boldsymbol{x}_2\|_2^2\right| \lesssim \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2, \qquad \left|\|\boldsymbol{h}_1\|_2^2 - \|\boldsymbol{h}_2\|_2^2\right| \lesssim \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2.$$

Taking these estimates together reveals

$$\left\|\nabla g_{\boldsymbol{h}_1,\boldsymbol{x}_1}(\alpha_2) - \nabla g_{\boldsymbol{h}_2,\boldsymbol{x}_2}(\alpha_2)\right\|_2 \lesssim \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 + \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2.$$

3. Combining the previous two bounds on the gradient and the Hessian, we conclude that

$$|\alpha_1 - \alpha_2| \lesssim \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 + \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2.$$

□

***Lemma* 58.** *Suppose that the vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathbb{C}^K$ satisfy*

$$\max\left\{\left\|\boldsymbol{x}_1 - \boldsymbol{x}^\natural\right\|_2, \left\|\boldsymbol{h}_1 - \boldsymbol{h}^\natural\right\|_2, \left\|\boldsymbol{x}_2 - \boldsymbol{x}^\natural\right\|_2, \left\|\boldsymbol{h}_2 - \boldsymbol{h}^\natural\right\|_2\right\} \le \delta \tag{278}$$

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 55 together with the assumption (278) guarantees that

$$\frac{1}{2} \le |\alpha_1| \le 2 \qquad \text{and} \qquad \frac{1}{2} \le |\alpha_2| \le 2.$$

Now we can prove the lemma. The triangle inequality gives

$$\|\alpha_1\boldsymbol{x}_1 - \alpha_2\boldsymbol{x}_2\|_2 = \|\alpha_1(\boldsymbol{x}_1 - \boldsymbol{x}_2) + (\alpha_1 - \alpha_2)\boldsymbol{x}_2\|_2$$

$$\begin{aligned}
&\leq |\alpha_1| \left\|\boldsymbol{x}_1 - \boldsymbol{x}_2\right\|_2 + |\alpha_1 - \alpha_2| \left\|\boldsymbol{x}_2\right\|_2 \\
&\overset{\text{(i)}}{\leq} 2\left\|\boldsymbol{x}_1 - \boldsymbol{x}_2\right\|_2 + 2\left|\alpha_1 - \alpha_2\right| \\
&\overset{\text{(ii)}}{\lesssim} \left\|\boldsymbol{x}_1 - \boldsymbol{x}_2\right\|_2 + \left\|\boldsymbol{h}_1 - \boldsymbol{h}_2\right\|_2,
\end{aligned}$$

where (i) holds since $|\alpha_1| \leq 2$ and $\|\boldsymbol{x}_2\|_2 \leq 1 + \delta \leq 2$, and (ii) arises from Lemma 57 that $|\alpha_1 - \alpha_2| \lesssim \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 + \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2$. Similarly,

$$\begin{aligned}
\left\|\frac{1}{\overline{\alpha_1}}\boldsymbol{h}_1 - \frac{1}{\overline{\alpha_2}}\boldsymbol{h}_2\right\|_2 &= \left\|\frac{1}{\overline{\alpha_1}}\left(\boldsymbol{h}_1 - \boldsymbol{h}_2\right) + \left(\frac{1}{\overline{\alpha_1}} - \frac{1}{\overline{\alpha_2}}\right)\boldsymbol{h}_2\right\|_2 \\
&\leq \left|\frac{1}{\overline{\alpha_1}}\right| \left\|\boldsymbol{h}_1 - \boldsymbol{h}_2\right\|_2 + \left|\frac{1}{\overline{\alpha_1}} - \frac{1}{\overline{\alpha_2}}\right| \left\|\boldsymbol{h}_2\right\|_2 \\
&\leq 2\left\|\boldsymbol{h}_1 - \boldsymbol{h}_2\right\|_2 + 2\frac{|\alpha_1 - \alpha_2|}{|\alpha_1\alpha_2|} \\
&\lesssim \left\|\boldsymbol{x}_1 - \boldsymbol{x}_2\right\|_2 + \left\|\boldsymbol{h}_1 - \boldsymbol{h}_2\right\|_2,
\end{aligned}$$

where the last inequality comes from Lemma 57 as well as the fact that $|\alpha_1| \geq 1/2$ and $|\alpha_2| \geq 1/2$ as shown above. Combining all of the above bounds and recognizing that $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 + \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2 \leq \sqrt{2\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2 + 2\|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2^2}$, we conclude the proof. □

Finally, we can extend $g_{\boldsymbol{h},\boldsymbol{x}}(\alpha)$ to account for any pairs $(\boldsymbol{h}_1, \boldsymbol{x}_1)$ and $(\boldsymbol{h}_2, \boldsymbol{x}_2)$. Specifically, for any $\boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{C}^K$, we introduce a real-valued function $\tilde{g}(\cdot) : \mathbb{C} \to \mathbb{R}$ as follows

$$\tilde{g}(\alpha) := \left\|\frac{1}{\overline{\alpha}}\boldsymbol{h}_1 - \boldsymbol{h}_2\right\|_2^2 + \left\|\alpha\boldsymbol{x}_1 - \boldsymbol{x}_2\right\|_2^2.$$

Then we have the following property for the minimizer.

Lemma 59. *For any $\boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{C}^K$, denote*

$$\alpha^\sharp := \arg\min \tilde{g}(\alpha), \qquad \tilde{\boldsymbol{x}}_1 = \alpha^\sharp \boldsymbol{x}_1 \qquad \text{and} \qquad \tilde{\boldsymbol{h}}_1 = \frac{1}{\overline{\alpha^\sharp}}\boldsymbol{h}_1.$$

Then one has

$$\left\|\tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2\right\|_2^2 + \boldsymbol{x}_2^*\left(\tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2\right) = \left\|\tilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2\right\|_2^2 + \left(\tilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2\right)^*\boldsymbol{h}_2.$$

Proof. We can rewrite the function as

$$\begin{aligned}
\tilde{g}(\alpha) &= |\alpha|^2 \left\|\boldsymbol{x}_1\right\|_2^2 + \left\|\boldsymbol{x}_2\right\|_2^2 - (\alpha\boldsymbol{x}_1)^*\boldsymbol{x}_2 - \boldsymbol{x}_2^*(\alpha\boldsymbol{x}_1) + \left|\frac{1}{\overline{\alpha}}\right|^2 \left\|\boldsymbol{h}_1\right\|_2^2 + \left\|\boldsymbol{h}_2\right\|_2^2 - \left(\frac{1}{\overline{\alpha}}\boldsymbol{h}_1\right)^*\boldsymbol{h}_2 - \boldsymbol{h}_2^*\left(\frac{1}{\overline{\alpha}}\boldsymbol{h}_1\right) \\
&= \overline{\alpha}\alpha\left\|\boldsymbol{x}_1\right\|_2^2 + \left\|\boldsymbol{x}_2\right\|_2^2 - \overline{\alpha}\boldsymbol{x}_1^*\boldsymbol{x}_2 - \alpha\boldsymbol{x}_2^*\boldsymbol{x}_1 + \frac{1}{\overline{\alpha}\alpha}\left\|\boldsymbol{h}_1\right\|_2^2 + \left\|\boldsymbol{h}_2\right\|_2^2 - \frac{1}{\alpha}\boldsymbol{h}_1^*\boldsymbol{h}_2 - \frac{1}{\overline{\alpha}}\boldsymbol{h}_2^*\boldsymbol{h}_1.
\end{aligned}$$

The first order optimality condition [KD09, Equation (38)] requires

$$\left.\frac{\partial\tilde{g}}{\partial\overline{\alpha}}\right|_{\alpha=\alpha^\sharp} = \alpha^\sharp\left\|\boldsymbol{x}_1\right\|_2^2 - \boldsymbol{x}_1^*\boldsymbol{x}_2 + \frac{1}{\alpha^\sharp}\left(-\frac{1}{\overline{\alpha^\sharp}^2}\right)\left\|\boldsymbol{h}_1\right\|_2^2 - \left(-\frac{1}{\overline{\alpha^\sharp}^2}\right)\boldsymbol{h}_2^*\boldsymbol{h}_1 = 0,$$

which further simplifies to

$$\overline{\alpha^\sharp}^2\alpha^\sharp\left\|\boldsymbol{x}_1\right\|_2^2 - \overline{\alpha^\sharp}^2\boldsymbol{x}_1^*\boldsymbol{x}_2 = \frac{1}{\alpha^\sharp}\left\|\boldsymbol{h}_1\right\|_2^2 - \boldsymbol{h}_2^*\boldsymbol{h}_1.$$

Recalling that $\tilde{\boldsymbol{x}}_1 = \alpha^\sharp\boldsymbol{x}_1$ and $\tilde{\boldsymbol{h}}_1 = \frac{1}{\overline{\alpha^\sharp}}\boldsymbol{h}_1$ gives

$$\overline{\alpha^\sharp}\left\|\tilde{\boldsymbol{x}}_1\right\|_2^2 - \overline{\alpha^\sharp}\tilde{\boldsymbol{x}}_1^*\boldsymbol{x}_2 = \overline{\alpha^\sharp}\left\|\tilde{\boldsymbol{h}}_1\right\|_2^2 - \overline{\alpha^\sharp}\boldsymbol{h}_2^*\tilde{\boldsymbol{h}}_1.$$

Since $\alpha^\sharp \neq 0$ (otherwise $\tilde{g}(\alpha^\sharp) = \infty$ and cannot be the minimizer), we arrive at

$$\|\tilde{\boldsymbol{x}}_1\|_2^2 - \tilde{\boldsymbol{x}}_1^* \boldsymbol{x}_2 = \big\|\tilde{\boldsymbol{h}}_1\big\|_2^2 - \boldsymbol{h}_2^* \tilde{\boldsymbol{h}}_1.$$

Furthermore, this condition is equivalent to

$$\tilde{\boldsymbol{x}}_1^* \left(\tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2\right) = \big(\tilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2\big)^* \tilde{\boldsymbol{h}}_1.$$

Recognizing that

$$\begin{aligned}
\tilde{\boldsymbol{x}}_1^* \left(\tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2\right) &= \boldsymbol{x}_2^* \left(\tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2\right) + \left(\tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2\right)^* \left(\tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2\right) = \boldsymbol{x}_2^* \left(\tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2\right) + \|\tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2\|_2^2, \\
\tilde{\boldsymbol{h}}_1^* \big(\tilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2\big) &= \boldsymbol{h}_2^* \big(\tilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2\big) + \big(\tilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2\big)^* \big(\tilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2\big) = \boldsymbol{h}_2^* \big(\tilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2\big) + \|\tilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2\|_2^2,
\end{aligned}$$

we get

$$\|\tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2\|_2^2 + \boldsymbol{x}_2^* \left(\tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_2\right) = \big\|\tilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2\big\|_2^2 + \big(\tilde{\boldsymbol{h}}_1 - \boldsymbol{h}_2\big)^* \boldsymbol{h}_2.$$

□

D.3.3 Matrix concentration

Lemma 60. *Suppose $m \gg K \log^2 m$. With probability exceeding $1 - O\left(m^{-10}\right)$, we have*

$$\left\| \sum_{j=1}^m \left|\boldsymbol{a}_j^* \boldsymbol{x}^\natural\right|^2 \boldsymbol{b}_j \boldsymbol{b}_j^* - \boldsymbol{I}_K \right\| \lesssim \sqrt{\frac{K}{m} \log^2 m}.$$

Proof. The property that $\sum_{j=1}^m \boldsymbol{b}_j \boldsymbol{b}_j^* = \boldsymbol{I}_K$ allows us to rewrite the quantity on the left-hand side as

$$\left\| \sum_{j=1}^m \left|\boldsymbol{a}_j^* \boldsymbol{x}^\natural\right|^2 \boldsymbol{b}_j \boldsymbol{b}_j^* - \boldsymbol{I}_K \right\| = \left\| \sum_{j=1}^m \underbrace{\left(\left|\boldsymbol{a}_j^* \boldsymbol{x}^\natural\right|^2 - 1\right) \boldsymbol{b}_j \boldsymbol{b}_j^*}_{:= \boldsymbol{Z}_j} \right\|,$$

where the $\boldsymbol{Z}_j$'s are independent zero-mean random matrices. To control the above spectral norm, we would like to apply the (unbounded) 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.

$$\big\| \|\boldsymbol{Z}_j\| \big\|_{\psi_1} = \|\boldsymbol{b}_j\|_2^2 \left\| \left|\boldsymbol{a}_j^* \boldsymbol{x}^\natural\right|^2 - 1 \right\|_{\psi_1} \lesssim \|\boldsymbol{b}_j\|_2^2 \left\| \left|\boldsymbol{a}_j^* \boldsymbol{x}^\natural\right|^2 \right\|_{\psi_1} \lesssim \frac{K}{m},$$

where we make use of the facts that

$$\|\boldsymbol{b}_j\|_2^2 = \frac{K}{m} \qquad \text{and} \qquad \left\| \left|\boldsymbol{a}_j^* \boldsymbol{x}^\natural\right|^2 \right\|_{\psi_1} \lesssim 1.$$

We further need to bound the variance parameter, that is

$$\begin{aligned}
\sigma_0^2 = \left\| \mathbb{E}\left[\sum_{j=1}^m \boldsymbol{Z}_j \boldsymbol{Z}_j^*\right] \right\| &= \left\| \mathbb{E}\left[\sum_{j=1}^m \left(\left|\boldsymbol{a}_j^* \boldsymbol{x}^\natural\right|^2 - 1\right)^2 \boldsymbol{b}_j \boldsymbol{b}_j^* \boldsymbol{b}_j \boldsymbol{b}_j^*\right] \right\| \\
&\lesssim \left\| \sum_{j=1}^m \boldsymbol{b}_j \boldsymbol{b}_j^* \boldsymbol{b}_j \boldsymbol{b}_j^* \right\| = \frac{K}{m} \left\| \sum_{j=1}^m \boldsymbol{b}_j \boldsymbol{b}_j^* \right\| = \frac{K}{m},
\end{aligned}$$

where the second line arises since $\mathbb{E}\left[\left|\boldsymbol{a}_j^* \boldsymbol{x}^\natural\right|^2 - 1\right] \asymp 1$, $\|\boldsymbol{b}_j\|_2^2 = K/m$, and $\sum_{j=1}^m \boldsymbol{b}_j \boldsymbol{b}_j^* = \boldsymbol{I}_K$. A direct application of the matrix Bernstein inequality [Kol11, Theorem 2.7] leads us to conclude that with probability exceeding $1 - O\left(m^{-10}\right)$,

$$\left\| \sum_{j=1}^m \boldsymbol{Z}_j \right\| \lesssim \max\left\{ \sqrt{\frac{K}{m} \log m}, \frac{K}{m} \log^2 m \right\} \leq \sqrt{\frac{K}{m} \log^2 m},$$

where the last relation holds under the assumption that $m \gg K \log^2 m$. □

***Lemma* 61.** *For any two sub-Gaussian random variables X and Y, $\|XY\|_{\psi_1} \le 2\|X\|_{\psi_2}\|Y\|_{\psi_2}$, where $\|\cdot\|_{\psi_1}$ and $\|\cdot\|_{\psi_2}$ are the sub-exponential and sub-Gaussian norm, respectively (see [Ver12]).*

Proof. For any integer $p \ge 1$, it follows from the definition of $\|\cdot\|_{\psi_2}$ that

$$\left(\mathbb{E}\left[|XY|^p\right]\right)^{\frac{1}{p}} \le \left(\mathbb{E}\left[|X|^{2p}\right]\right)^{\frac{1}{2p}} \cdot \left(\mathbb{E}\left[|Y|^{2p}\right]\right)^{\frac{1}{2p}} \le \sqrt{2p}\|X\|_{\psi_2} \cdot \sqrt{2p}\|Y\|_{\psi_2} = 2p\|X\|_{\psi_2}\|Y\|_{\psi_2}.$$

The claim then follows since $\|XY\|_{\psi_1} = \sup_{p\ge 1} \frac{1}{p}\left(\mathbb{E}\left[|XY|^p\right]\right)^{\frac{1}{p}}$. $\square$

***Lemma* 62.** *Suppose that $\{\boldsymbol{A}_j\}_{1\le j\le m}$ is a set of fixed matrices in $\mathbb{C}^{N\times K}$. We have*

$$\mathbb{P}\left(\left\|\frac{1}{m}\sum_{j=1}^{m}\boldsymbol{A}_j(\boldsymbol{a}_j\boldsymbol{a}_j^* - \boldsymbol{I}_K)\right\| \ge 2\gamma\left\|\frac{1}{m}\sum_{j=1}^{m}\boldsymbol{A}_j\boldsymbol{A}_j^*\right\|^{1/2}\right) \le \exp\left(c(N+K) - \gamma^2 m/C\right), \qquad \forall\gamma\in(0,1).$$

Here, $c, C > 0$ are some absolute constant.

Proof. For notational simplicity, define

$$\boldsymbol{R} = \sum_{j=1}^{m}\boldsymbol{A}_j\left(\boldsymbol{a}_j\boldsymbol{a}_j^* - \boldsymbol{I}_K\right).$$

First we show that

$$\mathbb{P}\left(\frac{1}{m}\left|\boldsymbol{u}^*\boldsymbol{R}\boldsymbol{v}\right| \ge \gamma\left\|\frac{1}{m}\sum_{j=1}^{m}\boldsymbol{A}_j\boldsymbol{A}_j^*\right\|^{1/2}\right) \le \exp\left(1 - \gamma^2 m/C\right), \qquad \forall\gamma\in(0,1) \tag{279}$$

holds for any fixed $\boldsymbol{u}\in\mathbb{C}^N$, $\boldsymbol{v}\in\mathbb{C}^K$ with $\|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = 1$.To this end, denote

$$w_j = \boldsymbol{u}_j^*\boldsymbol{A}_j\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{v} - \boldsymbol{u}_j^*\boldsymbol{A}_j\boldsymbol{v},$$

which is a zero-mean random variable. In view of [FWWZ17, Corollary 1], one has for any $t \ge 0$ and $m \ge 4$,

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{j=1}^{m}w_j\right| \ge t\right) \le \exp\left\{1 - \frac{m}{8}\min\left\{\frac{t}{2M}, \left(\frac{t}{2M}\right)^2\right\}\right\},$$

where $M = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\|w_j\|_{\psi_1}^2}$. By taking $t = 2\gamma M$ in the inequality above with $\gamma\in(0,1)$, we deduce that

$$\mathbb{P}\left(\frac{1}{m}\left|\boldsymbol{u}^*\boldsymbol{R}\boldsymbol{v}\right| \ge 2\gamma M\right) \le \exp\left(1 - \frac{\gamma^2 m}{8}\right).$$

Comparing this bound to the claimed concentration (279), we need

$$M \lesssim \left\|\frac{1}{m}\sum_{j=1}^{m}\boldsymbol{A}_j\boldsymbol{A}_j^*\right\|^{1/2}.$$

To justify this, note that

$$\|\boldsymbol{u}_j^*\boldsymbol{A}_j\boldsymbol{a}_j\|_{\psi_2} \lesssim \|\boldsymbol{A}_j^*\boldsymbol{u}\|_2 \qquad \text{and} \qquad \|\boldsymbol{a}_j^*\boldsymbol{v}\|_{\psi_2} \lesssim \|\boldsymbol{v}\|_2 = 1.$$

As a result, invoke Lemma 61 to get

$$\|\boldsymbol{u}_j^*\boldsymbol{A}_j\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{v}\|_{\psi_1} \le 2\|\boldsymbol{u}_j^*\boldsymbol{A}_j\boldsymbol{a}_j\|_{\psi_2}\|\boldsymbol{a}_j^*\boldsymbol{v}\|_{\psi_2} \lesssim \|\boldsymbol{A}_j^*\boldsymbol{u}\|_2. \tag{280}$$

This inequality can be used to upper bound $\|w_j\|_{\psi_1}$, as it follows from [Ver12, Remark 5.18] and (280) that

$$\|w_j\|_{\psi_1} \leq 2\|\boldsymbol{u}_j^* \boldsymbol{A}_j \boldsymbol{a}_j \boldsymbol{a}_j^* \boldsymbol{v}\|_{\psi_1} \lesssim \|\boldsymbol{A}_j^* \boldsymbol{u}\|_2.$$

This means that

$$M^2 m = \sum_{j=1}^{m} \|w_j\|_{\psi_1}^2 \lesssim \sum_{j=1}^{m} \|\boldsymbol{A}_j^* \boldsymbol{u}\|_2^2 = \boldsymbol{u}^* \left(\sum_{j=1}^{m} \boldsymbol{A}_j \boldsymbol{A}_j^* \right) \boldsymbol{u} \leq \left\| \sum_{j=1}^{m} \boldsymbol{A}_j \boldsymbol{A}_j^* \right\|,$$

$$\Longrightarrow \qquad M \lesssim \left\| \frac{1}{m} \sum_{j=1}^{m} \boldsymbol{A}_j \boldsymbol{A}_j^* \right\|^{1/2}.$$

We are now ready to extend the result in (279) to a uniform bound that simultaneously cover all $\boldsymbol{u}$ and $\boldsymbol{v}$, towards which we adopt the standard covering argument. Define

$$\mathcal{C} = \left\{ (\boldsymbol{u}, \boldsymbol{v}) \in \mathbb{C}^N \times \mathbb{C}^K : \|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = 1 \right\}$$

and let $\mathcal{C}_\varepsilon$ be an ε-net of $\mathcal{C}$ (see definition in [Tao12, Section 2.3]) with respect to the metric

$$\chi\left((\boldsymbol{u}, \boldsymbol{v}), (\boldsymbol{u}', \boldsymbol{v}') \right) := \max\{\|\boldsymbol{u} - \boldsymbol{u}'\|_2, \|\boldsymbol{v} - \boldsymbol{v}'\|_2\}.$$

Note that for any $(\boldsymbol{u}_1, \boldsymbol{v}_1), (\boldsymbol{u}_2, \boldsymbol{v}_2) \in \mathcal{C}$,

$$|\boldsymbol{u}_1^* \boldsymbol{R} \boldsymbol{v}_1 - \boldsymbol{u}_2^* \boldsymbol{R} \boldsymbol{v}_2| \leq \|\boldsymbol{R}\|_2 (\|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_2 + \|\boldsymbol{v}_1 - \boldsymbol{v}_2\|_2).$$

Taking $\varepsilon = 1/4$, the standard epsilon-net argument (which we omit the details but refer the readers to [Tao12, Section 2.3] and [Ver12]) yields

$$\|\boldsymbol{R}\|_2 \leq \frac{1}{1 - 2\varepsilon} \sup_{(\boldsymbol{u}, \boldsymbol{v}) \in \mathcal{C}_\varepsilon} |\boldsymbol{u}^* \boldsymbol{R} \boldsymbol{v}| = 2 \sup_{(\boldsymbol{u}, \boldsymbol{v}) \in \mathcal{C}_\varepsilon} |\boldsymbol{u}^* \boldsymbol{R} \boldsymbol{v}|, \tag{281}$$

and guarantees the existence of an ε-net with cardinality at most

$$|\mathcal{C}_\varepsilon| \leq e^{c_1 (N+K)}$$

for some constant $c_1 > 0$. On the other hand, (279) together with the union bound reveals that

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{m} \sup_{(\boldsymbol{u}, \boldsymbol{v}) \in \mathcal{C}_\varepsilon} |\boldsymbol{u}^* \boldsymbol{R} \boldsymbol{v}| \geq \gamma \left\| \frac{1}{m} \sum_{j=1}^{m} \boldsymbol{A}_j \boldsymbol{A}_j^* \right\|^{1/2} \right) &\leq |\mathcal{C}_\varepsilon| \exp\left(1 - \gamma^2 m / C\right) \\
&\leq \exp\left(c(N+K) - \gamma^2 m / C \right),
\end{aligned}$$

for some constant $c > c_1$. This combined with (281) concludes the proof.

□

The next technical lemma is concerned with the supremum of the spectral norm of random matrices over an "incoherence" region. In order to state this lemma, we introduce a metric on $\mathbb{C}^K \times \mathbb{C}^K$ as

$$\chi\big((\boldsymbol{h}_1, \boldsymbol{x}_1), (\boldsymbol{h}_2, \boldsymbol{x}_2) \big) := \max\left\{ \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|_2, \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 \right\}. \tag{282}$$

Correspondingly, define a set $\mathcal{R}_\delta$ to be

$$\mathcal{R}_\delta := \left\{ (\boldsymbol{h}, \boldsymbol{x}) \in \mathbb{C}^K \times \mathbb{C}^K : \chi\left((\boldsymbol{h}, \boldsymbol{x}), \left(\boldsymbol{h}^\natural, \boldsymbol{x}^\natural \right) \right) \leq \delta \right\} \tag{283}$$

and set

$$\mathcal{S}(\delta, \alpha) := \mathcal{R}_\delta \cap \left\{ (\boldsymbol{h}, \boldsymbol{x}) \in \mathbb{C}^K \times \mathbb{C}^K : \max_{1 \leq j \leq m} \left| \boldsymbol{b}_j^* \boldsymbol{h} \right| \leq \frac{\alpha}{\sqrt{m}} \right\}. \tag{284}$$

In words, $\mathcal{S}(\delta, \alpha)$ is a collection of points that is not far away from the truth and that is incoherence w.r.t. the $\boldsymbol{b}_j$'s, as long as δ and α are not large. With theses definitions in place, our result is as follows.

***Lemma* 63.** *Suppose that* $\{\boldsymbol{A}_j(\boldsymbol{h},\boldsymbol{x})\}_{1\le j\le m}$ *is a set of* $\mathbb{C}^{N\times K}$*-valued functions defined on* $\mathbb{C}^K\times\mathbb{C}^K$*, such that for all* $(\boldsymbol{h},\boldsymbol{x}),(\boldsymbol{h}_1,\boldsymbol{x}_1),(\boldsymbol{h}_2,\boldsymbol{x}_2)\in\mathcal{S}(\delta,\alpha)$ *the following two conditions hold*

$$\left\|\frac{1}{m}\sum_{j=1}^m \boldsymbol{A}_j(\boldsymbol{h},\boldsymbol{x})\boldsymbol{A}_j^*(\boldsymbol{h},\boldsymbol{x})\right\|^{1/2} \le M_1; \tag{285a}$$

$$\max_{1\le j\le m}\left\|\boldsymbol{A}_j(\boldsymbol{h}_2,\boldsymbol{x}_2)-\boldsymbol{A}_j(\boldsymbol{h}_1,\boldsymbol{x}_1)\right\| \le M_2\cdot\chi\big((\boldsymbol{h}_2,\boldsymbol{x}_2),(\boldsymbol{h}_1,\boldsymbol{x}_1)\big). \tag{285b}$$

Define $\boldsymbol{R}(\boldsymbol{h},\boldsymbol{x})=\sum_{j=1}^m\boldsymbol{A}_j(\boldsymbol{h},\boldsymbol{x})(\boldsymbol{a}_j\boldsymbol{a}_j^*-\boldsymbol{I}_K)$*. If the parameters* δ*,* M_1 *and* M_2 *satisfy*

$$\left(\min\left\{\frac{\delta}{mM_1},1\right\}\right)^2 m\gg(K+N)\log m \qquad \text{and} \qquad m\gg M_2K, \tag{286}$$

then with probability at least $1-O\left(m^{-10}\right)$*, one has*

$$\sup_{(\boldsymbol{h},\boldsymbol{x})\in\mathcal{S}(\delta,\alpha)}\|\boldsymbol{R}(\boldsymbol{h},\boldsymbol{x})\|\le 4\delta.$$

Proof. The lemma is proved using the standard covering argument. Note that by definition, one has $\mathcal{S}(\delta,\alpha)\subseteq\mathcal{R}_\delta$. For any $\varepsilon>0$, standard covering argument ensures the existence of an ε-cover $\mathcal{S}_\varepsilon(\delta,\alpha)$ of $\mathcal{S}(\delta,\alpha)$ — defined w.r.t. the metric χ — such that

$$|\mathcal{S}_\varepsilon(\delta,\alpha)|\le(1+\delta/\varepsilon)^{c_1K} \tag{287}$$

for some constant $c_1>0$.

First, suppose that the event $\mathcal{E}_0:=\left\{\sum_{j=1}^m\|\boldsymbol{a}_j\|_2^2\le 2mK\right\}$ happens. For any $(\boldsymbol{h}_1,\boldsymbol{x}_1),(\boldsymbol{h}_2,\boldsymbol{x}_2)\in\mathcal{S}(\delta,\alpha)$, if $\chi\big((\boldsymbol{h}_1,\boldsymbol{x}_1),(\boldsymbol{h}_2,\boldsymbol{x}_2)\big)\le\varepsilon$, the function $\boldsymbol{R}$ defined in the lemma obeys

$$\begin{aligned}
\Big|\|\boldsymbol{R}(\boldsymbol{h}_1,\boldsymbol{x}_1)\|-\|\boldsymbol{R}(\boldsymbol{h}_2,\boldsymbol{x}_2)\|\Big| &\le \|\boldsymbol{R}(\boldsymbol{h}_1,\boldsymbol{x}_1)-\boldsymbol{R}(\boldsymbol{h}_2,\boldsymbol{x}_2)\|\\
&\le\left\|\sum_{j=1}^m\left[\boldsymbol{A}_j(\boldsymbol{h}_1,\boldsymbol{x}_1)-\boldsymbol{A}_j(\boldsymbol{h},\boldsymbol{x}_2)\right]\boldsymbol{a}_j\boldsymbol{a}_j^*\right\|+\left\|\sum_{j=1}^m\left[\boldsymbol{A}_j(\boldsymbol{h}_1,\boldsymbol{x}_1)-\boldsymbol{A}_j(\boldsymbol{h},\boldsymbol{x}_2)\right]\right\|\\
&\le\max_{1\le j\le m}\|\boldsymbol{A}_j(\boldsymbol{h}_2,\boldsymbol{x}_2)-\boldsymbol{A}_j(\boldsymbol{h}_1,\boldsymbol{x}_1)\|\left(\sum_{j=1}^m\|\boldsymbol{a}_j\|_2^2+m\right)\le 4KmM_2\varepsilon,
\end{aligned}$$

where the last inequality uses the condition (285b), and definition of $\mathcal{E}_0$, and the distance bound on $\chi\big((\boldsymbol{h}_1,\boldsymbol{x}_1),(\boldsymbol{h}_2,\boldsymbol{x}_2)\big)$. Taking $\varepsilon=\frac{\delta}{2KmM_2}$ and invoking the standard covering argument lead to

$$\mathbb{P}\left(\sup_{(\boldsymbol{h},\boldsymbol{x})\in\mathcal{S}(\delta,\alpha)}\|\boldsymbol{R}(\boldsymbol{h},\boldsymbol{x})\|\le\sup_{(\boldsymbol{h},\boldsymbol{x})\in\mathcal{S}_\varepsilon(\delta,\alpha)}\|\boldsymbol{R}(\boldsymbol{h},\boldsymbol{x})\|+2\delta\right)\ge\mathbb{P}(\mathcal{E}_0). \tag{288}$$

It then suffices to control $\|\boldsymbol{R}(\boldsymbol{h},\boldsymbol{x})\|$ over the ε-net. Combining Lemma 62, the assumption (285a), the cardinality bound (287) and the union bound yields

$$\begin{aligned}
\mathbb{P}\left(\sup_{(\boldsymbol{h},\boldsymbol{x})\,\in\,\mathcal{S}_\varepsilon(\delta,\alpha)}\|\boldsymbol{R}(\boldsymbol{h},\boldsymbol{x})\|\ge 2\gamma mM_1\right)&\le|\mathcal{S}_\varepsilon(\delta,\alpha)|\exp\left(c(N+K)-\gamma^2m/C\right)\\
&\le(1+\delta/\varepsilon)^{c_1K}\exp\left(c(N+K)-\gamma^2m/C\right)\\
&\le\exp\left[c_1K\log(1+2KmM_2)+c(N+K)-\gamma^2m/C\right],
\end{aligned}\tag{289}$$

where the last inequality holds for the choice $\varepsilon=\frac{\delta}{2KmM_2}$. Taking $\gamma=\min\left\{\frac{\delta}{mM_1},1\right\}$ in (289) and combining it with (288), we get

$$\mathbb{P}\left(\sup_{(\boldsymbol{h},\boldsymbol{x})\,\in\,\mathcal{S}(\delta,\alpha)}\|\boldsymbol{R}(\boldsymbol{h},\boldsymbol{x})\|\ge 4\delta\right)$$

$$\leq \exp\left[c_1 K\log(1+2KmM_2)+c(N+K)-\frac{m}{C}\left(\min\left\{\frac{\delta}{mM_1},1\right\}\right)^2\right]+\mathbb{P}(\mathcal{E}_0^c).$$

Notably, when $M_2 K \lesssim m$ (which is satisfied due to the assumption (286)) we have

$$c_1 K\log(1+2KmM_2)+c(N+K)\lesssim K\log m+(K+N)\lesssim (K+N)\log m.$$

Also, by our assumption (286), there exists a constant $C_1>1$ such that

$$\left(\min\left\{\frac{\delta}{mM_1},1\right\}\right)^2 m\geq C_1(K+N)\log m \qquad \text{and} \qquad C_1 m\geq M_2 K.$$

If C_1 is sufficiently large, we have

$$c_1 K\log(1+2KmM_2)+c(N+K)\leq \frac{m}{2C}\left(\min\left\{\frac{\delta}{mM_1},1\right\}\right)^2,$$

and

$$\frac{m}{C}\left(\min\left\{\frac{\delta}{mM_1},1\right\}\right)^2\gg \log m.$$

Given that $\mathbb{P}(\mathcal{E}_0^c)$ is exponentially small, we conclude that

$$\mathbb{P}\left(\sup_{(\boldsymbol{h},\boldsymbol{x})\,\in\,\mathcal{S}(\delta,\alpha)}\|\boldsymbol{R}(\boldsymbol{h},\boldsymbol{x})\|\geq 4\delta\right)\lesssim m^{-10}.$$

□

D.3.4 Wirtinger calculus

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 derivative and the conventional derivative [KD09]:

$$\left[\begin{array}{c}\boldsymbol{x}\\ \boldsymbol{y}\end{array}\right]=\boldsymbol{J}^{-1}\left[\begin{array}{c}\boldsymbol{z}\\ \overline{\boldsymbol{z}}\end{array}\right], \tag{290a}$$

$$\nabla_{\mathrm{r}} f\left(\left[\begin{array}{c}\boldsymbol{x}\\ \boldsymbol{y}\end{array}\right]\right)=\boldsymbol{J}^{*}\nabla_{\mathrm{c}} f\left(\left[\begin{array}{c}\boldsymbol{z}_1\\ \overline{\boldsymbol{z}}_1\end{array}\right]\right), \tag{290b}$$

$$\nabla_{\mathrm{r}}^2 f\left(\left[\begin{array}{c}\boldsymbol{x}\\ \boldsymbol{y}\end{array}\right]\right)=\boldsymbol{J}^{*}\nabla_{\mathrm{c}}^2 f\left(\left[\begin{array}{c}\boldsymbol{z}\\ \overline{\boldsymbol{z}}\end{array}\right]\right)\boldsymbol{J}, \tag{290c}$$

where the subscripts r and c represent calculus in the real (conventional) sense and in the complex (Wirtinger) sense, respectively, and

$$\boldsymbol{J}=\left[\begin{array}{cc}\boldsymbol{I}_n & i\boldsymbol{I}_n\\ \boldsymbol{I}_n & -i\boldsymbol{I}_n\end{array}\right].$$

With these relationships in place, we are ready to prove the mean value theorem under Wirtinger calculus. Recall from the mean value theorem in conventional calculus that [Lan93, Theorem 4.2]

$$\nabla_{\mathrm{r}} f\left(\left[\begin{array}{c}\boldsymbol{x}_1\\ \boldsymbol{y}_1\end{array}\right]\right)-\nabla_{\mathrm{r}} f\left(\left[\begin{array}{c}\boldsymbol{x}_2\\ \boldsymbol{y}_2\end{array}\right]\right)=\left[\int_0^1\nabla_{\mathrm{r}}^2 f\left(\left[\begin{array}{c}\boldsymbol{x}(\tau)\\ \boldsymbol{y}(\tau)\end{array}\right]\right)\mathrm{d}\tau\right]\left(\left[\begin{array}{c}\boldsymbol{x}_1\\ \boldsymbol{y}_1\end{array}\right]-\left[\begin{array}{c}\boldsymbol{x}_2\\ \boldsymbol{y}_2\end{array}\right]\right), \tag{291}$$

where

$$\left[\begin{array}{c}\boldsymbol{x}(\tau)\\ \boldsymbol{y}(\tau)\end{array}\right]=\left[\begin{array}{c}\boldsymbol{x}_2\\ \boldsymbol{y}_2\end{array}\right]+\tau\left(\left[\begin{array}{c}\boldsymbol{x}_1\\ \boldsymbol{y}_1\end{array}\right]-\left[\begin{array}{c}\boldsymbol{x}_2\\ \boldsymbol{y}_2\end{array}\right]\right).$$

Substitute the identities (290) into (291) to arrive at

$$\begin{aligned}\boldsymbol{J}^{*}\nabla_{\mathrm{c}}f\left(\left[\begin{array}{c}\boldsymbol{z}_1\\\overline{\boldsymbol{z}_1}\end{array}\right]\right)-\boldsymbol{J}^{*}\nabla_{\mathrm{c}}f\left(\left[\begin{array}{c}\boldsymbol{z}_2\\\overline{\boldsymbol{z}}_2\end{array}\right]\right)&=\boldsymbol{J}^{*}\left[\int_0^1\nabla_{\mathrm{c}}^2f\left(\left[\begin{array}{c}\boldsymbol{z}(\tau)\\\overline{\boldsymbol{z}}(\tau)\end{array}\right]\right)\mathrm{d}\tau\right]\boldsymbol{J}\boldsymbol{J}^{-1}\left(\left[\begin{array}{c}\boldsymbol{z}_1\\\overline{\boldsymbol{z}_1}\end{array}\right]-\left[\begin{array}{c}\boldsymbol{z}_2\\\overline{\boldsymbol{z}_2}\end{array}\right]\right)\\&=\boldsymbol{J}^{*}\left[\int_0^1\nabla_{\mathrm{c}}^2f\left(\left[\begin{array}{c}\boldsymbol{z}(\tau)\\\overline{\boldsymbol{z}}(\tau)\end{array}\right]\right)\mathrm{d}\tau\right]\left(\left[\begin{array}{c}\boldsymbol{z}_1\\\overline{\boldsymbol{z}_1}\end{array}\right]-\left[\begin{array}{c}\boldsymbol{z}_2\\\overline{\boldsymbol{z}_2}\end{array}\right]\right),\end{aligned}\tag{292}$$

where $\boldsymbol{z}_1=\boldsymbol{x}_1+i\boldsymbol{y}_1$, $\boldsymbol{z}_2=\boldsymbol{x}_2+i\boldsymbol{y}_2$ and

$$\left[\begin{array}{c}\boldsymbol{z}(\tau)\\\overline{\boldsymbol{z}}(\tau)\end{array}\right]=\left[\begin{array}{c}\boldsymbol{z}_2\\\overline{\boldsymbol{z}}_2\end{array}\right]+\tau\left(\left[\begin{array}{c}\boldsymbol{z}_1\\\overline{\boldsymbol{z}_1}\end{array}\right]-\left[\begin{array}{c}\boldsymbol{z}_2\\\overline{\boldsymbol{z}_2}\end{array}\right]\right).$$

Simplification of (292) gives

$$\nabla_{\mathrm{c}}f\left(\left[\begin{array}{c}\boldsymbol{z}_1\\\overline{\boldsymbol{z}_1}\end{array}\right]\right)-\nabla_{\mathrm{c}}f\left(\left[\begin{array}{c}\boldsymbol{z}_2\\\overline{\boldsymbol{z}_2}\end{array}\right]\right)=\left[\int_0^1\nabla_{\mathrm{c}}^2f\left(\left[\begin{array}{c}\boldsymbol{z}(\tau)\\\overline{\boldsymbol{z}}(\tau)\end{array}\right]\right)\mathrm{d}\tau\right]\left(\left[\begin{array}{c}\boldsymbol{z}_1\\\overline{\boldsymbol{z}_1}\end{array}\right]-\left[\begin{array}{c}\boldsymbol{z}_2\\\overline{\boldsymbol{z}_2}\end{array}\right]\right).$$

Repeating the above arguments, one can also show that

$$f(\boldsymbol{z}_1)-f(\boldsymbol{z}_2)=\nabla_{\mathrm{c}}f(\boldsymbol{z}_2)^{*}\left[\begin{array}{c}\boldsymbol{z}_1-\boldsymbol{z}_2\\\overline{\boldsymbol{z}_1-\boldsymbol{z}_2}\end{array}\right]+\frac{1}{2}\left[\begin{array}{c}\boldsymbol{z}_1-\boldsymbol{z}_2\\\overline{\boldsymbol{z}_1-\boldsymbol{z}_2}\end{array}\right]^{*}\nabla_{\mathrm{c}}^2f(\tilde{\boldsymbol{z}})\left[\begin{array}{c}\boldsymbol{z}_1-\boldsymbol{z}_2\\\overline{\boldsymbol{z}_1-\boldsymbol{z}_2}\end{array}\right].$$