

# Supplementary Material

## Contents

<b>A Preliminaries: On the In-Network RIP and its Consequences</b>	<b>14</b>
<b>B Proof of Theorem 11</b>	<b>15</b>
B.1 Phase I	15
B.1.1 Step 1: Approximating the trajectory	16
B.1.2 Step 2: Bounding the signal- and noise-terms	17
B.1.3 Step 3: From deterministic to random initialization	19
B.2 Phase II	19
B.2.1 Step 1: Growing alignment phase	21
B.2.2 Step 2: Refinement phase	23
B.2.3 Summary	26
<b>C Convergence of Algorithm 3 in the setting <math>r \in [r^*, 2r^*]</math></b>	<b>26</b>
C.1 Proof of Theorem 2	27
<b>D Proof of Corollary 1</b>	<b>28</b>
<b>E Proofs of the Intermediate Results in Phase I (Appendix B.1)</b>	<b>34</b>
E.1 Proof of Lemma 5	34
E.2 Proof of Lemma 6	35
E.3 Proof of Lemma 7	36
E.4 Proof of Lemma 8	36
E.5 Proof of Propositions 2 and 4	38
E.6 Technical lemmas: auxiliary to proving Lemma 7	39
<b>F Proofs of the Intermediate Results in Phase II (Appendix B.2)</b>	<b>41</b>
F.1 Proofs of Lemma 9	41
F.2 Technical lemmas: auxiliary to proving Lemma 9	50
<b>G Proofs of the RIP Properties and Consequences</b>	<b>52</b>
G.1 Proof of Lemma 1	52
G.2 Proof of Lemma 2	53
G.3 Proof of Lemma 3	54
G.4 Proof of Lemma 4	54

## A Preliminaries: On the In-Network RIP and its Consequences

This section collects some preliminary results on the introduced In-network RIP (Definition 3) and related properties/consequences, which will be used through the rest of the appendix. The proof of these intermediate results can be found in Appendix G.

**Lemma 2.** Suppose that  $\bar{\mathcal{A}}$  satisfies the  $(\delta_{2(r+1)}, 2(r+1))$ -RIP, and the gossip matrix is chosen according to Assumption [1](#) with

$$\rho^2 \leq \frac{2\delta_{2(r+1)}^2}{4m^5(1+2\delta_{2(r+1)})^2}.$$

Then, for any matrix  $Z \in \mathbb{R}^{md \times md}$  such that each of its  $d \times d$  blocks  $[Z]_{ij}$  and its block average  $\bar{Z} = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m Z_{ij}$  are of rank at most  $r$ , it holds

$$\|\mathcal{WZ}\mathcal{W} - \mathcal{A}^* \mathcal{A}(Z)\| \leq 2\delta_{2r} \|\mathcal{JZ}\mathcal{J}\|_F + 2 \left( \rho + (1 + 2\delta_{2(r+1)}) \hat{\Delta}_{2(r+1)} \right) \|Z - \mathcal{JZ}\mathcal{J}\|_F,$$

where

$$\hat{\Delta}_{2(r+1)} \triangleq 4m^5 \rho^2 \frac{(1 + 2\delta_{2(r+1)})^2}{\delta_{2(r+1)}}.$$

**Lemma 3.** Suppose that  $\bar{\mathcal{A}}$  satisfies the  $(\delta_4, 4)$ -RIP, and the gossip matrix is chosen according to Assumption [1](#) with

$$\rho^2 \leq \frac{2\delta_4^2}{4m^4(1+2\delta_4)^2}.$$

Then,

$$\|\mathcal{WZ}\mathcal{W} - \mathcal{A}^* \mathcal{A}(Z)\| \leq \left( 2\delta_4 + 2 \left( \rho + (1 + 2\delta_4) \hat{\Delta}_4 \right) \right) \|Z\|_*,$$

for any matrix  $Z \in \mathbb{R}^{md \times md}$ . Here,  $\|Z\|_*$  denotes the nuclear norm of  $Z$ .

**Lemma 4.** Suppose that  $\bar{\mathcal{A}}$  satisfies the  $(\delta_{2r^*}, 2r^*)$ -RIP, and the gossip matrix is chosen according to Assumption [1](#) with

$$\rho \leq \frac{2\delta_{2r^*}^2}{4m^5(1+2\delta_{2r^*})^2}.$$

Then, the following hold:

$$\begin{aligned} \|(\mathcal{I} - \mathcal{J})\mathcal{A}^* \mathcal{A}(Z^*)\| &\leq \rho^{1/2} (\|Z^*\| + \delta_{2r^*} \|Z^*\|_F), \\ \|(\mathcal{I} - \mathcal{J})\mathcal{A}^* \mathcal{A}(Z^*)(\mathcal{I} - \mathcal{J})\| &\leq \rho (\|Z^*\| + \delta_{2r^*} \|Z^*\|_F). \end{aligned}$$

## B Proof of Theorem [1](#)

As preliminary sketched in Sec. [3.1](#), the proof of the theorem is organized in two phases. Phase I (Appendix [B.1](#)) shows that the iterates  $U^t$  stay in close proximity to a carefully selected reference sequence,  $U^t$ . This reference sequence exhibits desirable alignment with the signal space, a key factor to ensure progresses of  $U^t$  toward the true solution. Phase II (Appendix [B.2](#)) demonstrates that this established alignment of the iterates to the signal space remains steady while the generalization error progressively lessens.

### B.1 Phase I

This section is devoted to the analysis of the spectral phase of the algorithm, with detailed proof of the supporting lemmata deferred to Appendix [E](#).

Recalling the signal plus noise decomposition of the iterates  $U^t$  as in [\(18\)](#), our primary objective is to demonstrate that, after a certain elapsed time  $t$  post-initialization ( $t = 0$ ), the signal components  $\sigma_{\min}(U^t Q^t)$  and  $\|V_{Z^*}^\top U^t Q^t\|$  “outweigh” the noise-related terms  $\|U^t Q^{t,\perp}\|$  and  $\|(V_{Z^*}^\perp)^\top U^t Q^{t,\perp}\|$ , respectively. Essentially, this indicates a significant alignment of the iterates  $U^t$  with the signal eigenspace. The proof is organized in the following three steps:

• **Step 1 (approximating the trajectory):** We show that, for any given initialization  $U^0 = \mu U$  with small  $\mu$ , and sufficiently connected network ( $\rho$  small), the iterates  $U^t$  can be closely approximated by the following dynamics in the first few iterations  $t$ :

$$\tilde{U}^t = \left( \mathcal{W}^2 + \frac{\alpha}{m} \mathcal{A}^* \mathcal{A}(Z^*) \right)^t U^0 =: \mathcal{M}^t U^0, \quad \text{with } \mathcal{M} := \mathcal{W}^2 + \frac{\alpha}{m} \mathcal{A}^* \mathcal{A}(Z^*). \quad (28)$$

We refer to Sec. 3.1 for further insight behind the choice of such a reference trajectory;

• **Step 2 (bounding the signal- and noise-terms):** Employing the eigen-properties of the mapping  $\mathcal{M}$  in (28) and the signal-plus-noise decomposition (18) of  $U^t$ , we provide suitable bounds of the quantities  $\sigma_{\min}(U^t Q^t)$ ,  $\|U^t Q^{t,\perp}\|$ , and  $\|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\|$ . From these, we determine conditions on the free parameters to achieve the desired alignment of the iterates  $U^t$  with the signal eigenspace.

• **Step 3 (from deterministic to random initialization):** The derivations in Steps 1 and 2 hold for any given initialization matrix  $U$ . In this step, we conclude the analysis by stating the final result when  $U$  is randomly selected, with Gaussian entries.

### B.1.1 Step 1: Approximating the trajectory

Employing (28) and following the same steps as in the proof of [34, Lemma 8.1] applied to  $\tilde{U}^t$ , we can express  $U^t$  as a function of  $\tilde{U}^t$  as follows:

$$U^t = \tilde{U}^t + E^t, \quad (29)$$

where the error sequence  $E^t$  reads

$$E^t = \sum_{i=1}^t \left( W^2 + \frac{\alpha}{m} \mathcal{A}^* \mathcal{A}(Z^*) \right)^{t-i} \frac{\alpha}{m} (\mathcal{A}^* \mathcal{A}(U^{i-1} (U^{i-1})^\top) U^{i-1}). \quad (30)$$

Next we quantify the quality of the approximation of  $U^t$  by  $\tilde{U}^t$ . Specifically, defining

$$t^* \triangleq \min \left\{ t \in \mathbb{N}_+ : \|\tilde{U}^{t-1} - U^{t-1}\| > \|\tilde{U}^{t-1}\| \right\},$$

Lemma 5 below shows that  $\tilde{U}^t$  is a good approximation of  $U^t$  ( $\|E^t\|$  is uniformly bounded) as long as  $t \in [t^*]$ . Lemma 6 shows that  $t^*$  is bounded away from zero, providing thus an estimate of the time interval within which  $U^t$  is “well approximated” by  $\tilde{U}^t$ .

**Lemma 5.** *Suppose that (i)  $\bar{\mathcal{A}}$  satisfies the  $(\delta_{2(r^*+1)}, 2(r^*+1))$ -RIP property and the  $(\delta_4, 4)$ -RIP; and (ii) the gossip matrix  $W$  is chosen according to Assumption 7 with*

$$\rho^2 \leq \frac{2\delta_{2(r^*+1)}^2}{4m^5(1+2\delta_{2(r^*+1)})^2}.$$

Then,  $E^t$  can be bounded as

$$\begin{aligned} \|E^t\| &\leq \frac{4m^{-1}}{(1+2\sqrt{r^*}\delta_{2(r^*+1)})\lambda_1(\bar{Z}^*)} \mu^3 \min\{r, md\} \left( 1 + 2\delta_4 + 2 \left( \rho + (1+2\delta_4)\hat{\Delta}_4 \right) \right) \times \\ &\times \left( 1 + \alpha(1+2\sqrt{r^*}\delta_{2(r^*+1)})\lambda_1(\bar{Z}^*) \right)^{3t} \left( \|\mathcal{J}U\| + \rho^{1/2}\sqrt{r^*}\|(\mathcal{I}-\mathcal{J})U\| \right)^3, \quad \forall t \in [t^*]. \end{aligned} \quad (31)$$

**Lemma 6.** *In the setting of Lemma 5 the following lower bound of  $t^*$  holds:*

$$t^* \geq \left\lfloor \frac{\ln \left( \frac{(1+2\sqrt{r^*}\delta_{2(r^*+1)})\lambda_1(\bar{Z}^*)}{4(1+2\delta_4+2(\rho+(1+2\delta_4)\hat{\Delta}_4))\mu^2 \frac{(\|\mathcal{J}U\| + \rho^{1/2}\sqrt{r^*}\|(\mathcal{I}-\mathcal{J})U\|)^3}{m\|U^\top v_1\|}} \right)}{\ln \left( \frac{(1+\alpha\lambda_1(\bar{X}^*)(1+2\sqrt{r^*}\delta_{r^*+1}))^3}{1+\alpha\lambda_1(\bar{X}^*)(1-2\sqrt{r^*}\delta_{2(r^*+1)})} \right)} \right\rfloor.$$

### B.1.2 Step 2: Bounding the signal- and noise-terms

Through the section we will use the following notation, which simplifies the statement of our results:

$$\begin{aligned} \nu &\triangleq \lambda_1(\bar{Z}^*)(1 + \sqrt{r^*}\delta_{2r^*}), \\ \delta_\rho &\triangleq \frac{\sqrt{r^*}\rho^{1/2}\nu}{\lambda_{r^*}(\bar{Z}^*) - 3\delta_{2(r^*+1)}\sqrt{r^*}\lambda_1(\bar{Z}^*)}, \\ \eta_r &\triangleq 2\delta_r + 2(\rho + (1 + 2\delta_r)\hat{\Delta}_r), \\ \sigma &\triangleq 2\kappa^2 \frac{(1 + \alpha 2\sqrt{r^*}\delta_{2(r^*+1)})\lambda_1(\bar{Z}^*)\|\mathcal{J}U\| + (\rho^2 + \alpha\rho^{1/2}\nu)\|(\mathcal{I} - \mathcal{J})U\|}{c\sigma_{\min}(V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\top U) (1 + \alpha 2\sqrt{r^*}\delta_{2(r^*+1)}) \lambda_1(\bar{Z}^*)}, \end{aligned} \quad (32)$$

$$\gamma_t \triangleq \frac{\mu\sigma_{r^*+1}(\mathcal{M}^{t-1}) (\mu\sigma_{r^*+1}(\mathcal{M})\|\mathcal{J}U\| + (\rho^2 + \alpha\rho^{1/2}\nu)\|(\mathcal{I} - \mathcal{J})U\|) + \|E^t\|}{(1 - \delta_\rho)\mu\sigma_{r^*}(\mathcal{M}^t)\sigma_{\min}(V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\top U)}. \quad (33)$$

The upcoming lemma delves into the characteristics of the dynamics  $\tilde{U}^t$  (through the eigen-properties of  $\mathcal{M}$ ) to bound the signal and noise quantities of interest  $\sigma_{\min}(U^t Q^t)$ ,  $\|U^t Q^{t,\perp}\|$ , and  $\|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\|$ . These bounds will shed light on the conditions ensuring that  $U^t$  exhibits sufficient alignment to the signal space (in the sense previously discussed).

**Lemma 7.** *Suppose that (i)  $\bar{A}$  satisfies the  $(\delta_{2(r^*+1)}, 2(r^*+1))$ -RIP, with*

$$\delta_{2(r^*+1)} \leq \frac{1}{128\sqrt{r^*}}\kappa^{-4};$$

(ii)  $W$  is chosen according to Assumption [7](#) with

$$\rho \leq \min \left\{ \frac{\delta_{2(r^*+1)}}{2m^2\sqrt{m}(1 + 2\delta_{2(r^*+1)})}, \frac{(\lambda_{r^*}(\bar{Z}^*) - 3\delta_{2(r^*+1)}\sqrt{r^*}\lambda_1(\bar{Z}^*))^2}{3r^*\lambda_1(\bar{Z}^*)^2(1 + \sqrt{r^*}\delta_{2r^*})^2} \right\}; \quad (34)$$

and (iii)  $\gamma_t$  defined in [\(33\)](#) satisfies

$$\gamma_t \leq \frac{\kappa^{-2}}{17}. \quad (35)$$

Then,  $\sigma_{\min}(U^t Q^t)$ ,  $\|U^t Q^{t,\perp}\|$  and  $\|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\|$  can be bounded respectively as

$$\sigma_{\min}(U^t Q^t) \geq \frac{\mu(1 - \delta_\rho)}{2}\sigma_{r^*}(\mathcal{M}^t)\sigma_{\min}(V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\top U) - \|E^t\| \quad (36a)$$

$$\|U^t Q^{t,\perp}\| \leq \frac{\kappa^{-2}}{8}\mu\sigma_{r^*}(\mathcal{M}^t)\sigma_{\min}(V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\top U), \quad (36b)$$

$$\|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\| \leq 56 \left( \sqrt{r^*}\kappa^2\delta_{2(r^*+1)} + \gamma_t \right). \quad (36c)$$

The desiderata is a sufficient large  $\sigma_{\min}(U^t Q^t)$  (persistent signal component) along with small noise-terms  $\|U^t Q^{t,\perp}\|$  and  $\|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\|$ . A close look at equations [\(36\)](#) suggests that this is achievable by guaranteeing sufficiently small  $\rho$  (quantified by [\(34\)](#)),  $\|E^t\|$  and  $\gamma_t$ . The latter represents a suitable noise-to-signal ratio measure, which accounts for the deviation of  $U^t$  from  $\tilde{U}^t$  as well as the consensus discrepancies in  $U$ .

In Step 1 (Lemma [5](#)) we showed that  $\|E^t\|$  remains “small” as long as  $t \in [t^*]$ . The subsequent lemma proves that, after a certain number of iterations nested within  $[t^*]$ , one can confidently ensure that  $\gamma_t$  also is small, so that [\(36\)](#) will hold.

**Lemma 8.** *Reinstate Assumptions (i)-(ii) of Lemma [7](#). Choose  $\mu$  such that*

$$\begin{aligned} \mu^2 &\leq \frac{(1 + 2\sqrt{r^*}\delta_{2(r^*+1)})\lambda_1(\bar{Z}^*)}{4(1 + \eta)\min\{r, md\}} \min \left\{ \frac{m\|U^\top v_1\|}{(\|\mathcal{J}U\| + \sqrt{r^*}\rho^{1/2}\|(\mathcal{I} - \mathcal{J})U\|)^3} (\sigma)^{-96\kappa^2}, \right. \\ &\quad \left. \frac{c(1 - \delta_\rho)m\sigma_{\min}(V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\top U)}{8\kappa^2 (\|\mathcal{J}U\| + \rho^{1/2}\sqrt{r^*}\|(\mathcal{I} - \mathcal{J})U\|)^3} (\sigma)^{-12\kappa^2} \right\}, \end{aligned} \quad (37)$$

where  $c$  is a strictly positive constant  $c < 1/2$  and  $\sigma$  is defined in (32).

Then after

$$t_* \triangleq \left[ \ln(\sigma) \left( \ln \left( \frac{1 + \alpha \lambda_{r^*}(\bar{Z}^*) - \alpha 2\sqrt{r^*} \delta_{2(r^*+1)} \lambda_1(\bar{Z}^*)}{1 + \alpha 2\sqrt{r^*} \delta_{2(r^*+1)} \lambda_1(\bar{Z}^*)} \right) \right)^{-1} \right]$$

iterations, it holds

$$\gamma_{t_*} \leq c\kappa^{-2}.$$

Lemma 8 establishes the existence of a time  $t_*$  at which  $\gamma_{t_*}$  is sufficiently small. Notice that, in the setting of the lemma,  $t_* \leq t^*$ . As such, we can apply also Lemma 5 at  $t = t_*$ , ensuring that  $\|E^{t_*}\|$  concurrently resides under the desired limit. This, coupled with a judicious choice of  $\mu$ , results in the ultimate bounds on the quantities in (36), as detailed next.

Using Lemma 8 in (36), yields

$$\sigma_{\min}(U^{t_*} Q^{t_*}) \geq (1 - \delta_\rho) \frac{\mu}{4} \beta - \|E^{t_*}\|, \quad (38)$$

$$\|U^{t_*} Q^{t_*, \perp}\| \leq \frac{\kappa^{-2}}{8} \mu \beta, \quad (39)$$

$$\left\| (V_{\bar{Z}^*}^\perp)^\top V_{U^{t_*} Q^{t_*}} \right\| \leq 56 \left( \sqrt{r^*} \kappa^2 \delta_{2(r^*+2)} + c\kappa^{-2} \right),$$

where for convenience, we defined

$$\beta \triangleq \sigma_{r^*}(\mathcal{M}^{t_*}) \sigma_{\min}(V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\top U).$$

We proceed to further bound the RHS of (38) and (39).

By definition of  $\gamma_t$  and choosing  $c \leq 1/17$ , it holds

$$\|E^t\| \leq \frac{1}{17} (1 - \delta_\rho) \mu \beta.$$

Consequently

$$\sigma_{\min}(U^{t_*} Q^{t_*}) \geq \frac{\mu}{2} (1 - \delta_\rho) \beta - \frac{\mu}{17} (1 - \delta_\rho) \mu \beta \stackrel{(a)}{\geq} \frac{\mu \beta}{4},$$

whereby in (a) we used  $\delta_\rho \leq 1/3$ , which holds under (34).

It remains to upper and lower bound  $\beta$ . Under the postulated assumption on  $\delta_{2r^*}$ ,  $\lambda_{r^*}(\mathcal{M}^t) > 1$ . Therefore,

$$\beta \geq \sigma_{\min}(V_{\mathcal{J}\mathcal{M}\mathcal{J}}).$$

Similarly, using Weyl's inequality,

$$\begin{aligned} \sigma_{r^*}(\mathcal{M}^{t_*}) &\leq \left( 1 + \alpha \lambda_{r^*}(\bar{Z}^*) + \frac{\alpha}{m} \|\mathcal{W} Z^* \mathcal{W} - \mathcal{A}^* \mathcal{A}(Z^*)\| \right)^{t_*} \\ &\stackrel{(a)}{\leq} \left( 1 + \alpha \lambda_{r^*}(\bar{Z}) + \alpha 2\sqrt{r^*} \delta_{2r^*} \lambda_1(\bar{Z}^*) \right)^{t_*} \\ &\leq \left( 1 + \alpha \lambda_{r^*}(\bar{Z}^*) + \alpha 2\sqrt{r^*} \delta_{2r^*} \lambda_1(\bar{Z}^*) \right)^{\frac{\ln(\sigma)}{\ln(1 + \alpha \lambda_{r^*}(\bar{Z}^*) + \alpha 2\sqrt{r^*} \delta_{2r^*} \lambda_1(\bar{Z}^*))}} \times \frac{\ln(1 + \alpha \lambda_{r^*}(\bar{Z}^*) + \alpha 2\sqrt{r^*} \delta_{2r^*} \lambda_1(\bar{Z}^*))}{\ln\left(\frac{1 + \alpha \lambda_{r^*}(\bar{Z}^*) - \alpha 2\sqrt{r^*} \delta_{2r^*} \lambda_1(\bar{Z}^*)}{1 + \alpha 2\sqrt{r^*} \delta_{2(r^*+1)} \lambda_1(\bar{Z}^*)}\right)} + 1 \\ &\stackrel{(b)}{\leq} 2\sigma \ln\left(\frac{\ln(1 + \alpha \lambda_{r^*}(\bar{Z}^*) + \alpha 2\sqrt{r^*} \delta_{2r^*} \lambda_1(\bar{Z}^*))}{\ln\left(\frac{1 + \alpha \lambda_{r^*}(\bar{Z}^*) - \alpha 2\sqrt{r^*} \delta_{2r^*} \lambda_1(\bar{Z}^*)}{1 + \alpha 2\sqrt{r^*} \delta_{2(r^*+1)} \lambda_1(\bar{Z}^*)}\right)}\right) \end{aligned}$$

where in (a) we used Lemma 2 and (b) follows from  $1 + \alpha \lambda_{r^*}(\bar{Z}^*) + 2\alpha \sqrt{r^*} \delta_{2r^*} \lambda_1(\bar{Z}^*) \leq 2$ , due to the assumption on  $\alpha$  and  $\delta_{2r^*}$ . Finally, we bound the exponent as

$$\begin{aligned} \frac{\ln(1 + \alpha(\lambda_{r^*}(\bar{Z}^*) + \alpha \sqrt{r^*} \delta_{2r^*} \lambda_1(\bar{Z}^*)))}{\ln\left(\frac{1 + \alpha \lambda_{r^*}(\bar{Z}^*) - \alpha 2\sqrt{r^*} \delta_{2r^*} \lambda_1(\bar{Z}^*)}{1 + \alpha 2\sqrt{r^*} \delta_{2(r^*+1)} \lambda_1(\bar{Z}^*)}\right)} &\leq \frac{\alpha(\lambda_{r^*}(\bar{Z}^*) + 2\sqrt{r^*} \delta_{2r^*} \lambda_1(\bar{Z}^*))}{\frac{\alpha(\lambda_{r^*}(\bar{Z}^*) - 4\sqrt{r^*} \delta_{2r^*} \lambda_1(\bar{Z}^*))}{1 + \alpha \lambda_{r^*}(\bar{Z}^*) - 2\sqrt{r^*} \delta_{2(r^*+1)} \lambda_1(\bar{Z}^*)}} \\ &\leq \frac{(1 + \alpha \lambda_{r^*}(\bar{Z})) (1 + 1/64)}{(\bar{Z}^*) (1 - 1/32)} \stackrel{(a)}{\leq} \frac{(1 + 3/4) (1 + 1/64)}{1 - 1/32} \stackrel{(b)}{\leq} 2, \end{aligned}$$

where in (a) we used  $\alpha \leq \frac{3}{4} \|\bar{Z}^*\|^{-1} \kappa^{-2}$ .

We conclude the analysis, proving the following bound for  $U^{t_*}$ :  $\|U^{t_*}\| \leq 3\sqrt{m} \|\bar{X}\|$ .

Following a similar procedure as in the proof of Lemma 5, we can write

$$\begin{aligned} \|U^{t_*}\| &\leq \mu \|M^{t_*} U\| + \|E^{t_*}\| \leq \mu \left(1 + \alpha(1 + 2\sqrt{r^*} \delta_{2(r^*+1)}) \lambda_1(\bar{Z}^*)\right)^{t_*} \\ &\quad \times (\|\mathcal{J}U\| + \rho^{1/2} \sqrt{r^*} \|(\mathcal{I} - \mathcal{J})U\|) + \frac{\tilde{c}_3}{2} (1 - \delta_\rho) \mu \sigma_{r^*}(M^{t_*}) \sigma_{\min}(V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\top U) \kappa^{-2}. \end{aligned}$$

The bound  $\|U^{t_*}\| \leq 3\sqrt{m} \|\bar{X}\|$  is obtained by using the bound of  $\mu$  in (37).

We summarize next the obtained properties of the signal and noise component at time  $t_*$ , which represents the ‘‘exist’’ of Phase I.

**Proposition 1.** *Restate the conditions of Lemma 8 choosing a strictly positive constant  $c \leq \frac{1}{17}$ , and assume  $\alpha \leq \frac{3}{4} \kappa^{-2} \|\bar{Z}^*\|^{-1}$ . The following hold:*

$$\sigma_{\min}(U^{t_*} Q^{t_*}) \geq \frac{\mu\beta}{4}, \quad (40a)$$

$$\|U^{t_*} Q^{t_*, \perp}\| \leq \frac{\kappa^{-2}}{8} \mu\beta, \quad (40b)$$

$$\|(V_{\bar{Z}^*}^\perp)^\top V_{U^{t_*} Q^{t_*}}\| \leq 56 \left(\sqrt{r^*} \kappa^2 \delta_{2(r^*+1)} + c\kappa^{-2}\right), \quad (40c)$$

$$\|U^{t_*}\| \leq 3\sqrt{m} \|\bar{X}\|, \quad (40d)$$

with

$$\sigma_{\min}(V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\top U) \leq \beta \leq 2\sigma_{\min}(V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\top) \sigma^2. \quad (41)$$

### B.1.3 Step 3: From deterministic to random initialization

We conclude the analysis of Phase I allowing for  $U$  to be random with Gaussian entries. We obtain the following counterpart of Proposition 1

**Proposition 2.** *Reinstate the conditions of Proposition 1, and choose a strictly positive constant  $c \leq \frac{1}{17}$ . Additionally, suppose that  $U \in \mathbb{R}^{md \times r}$  is a random matrix such that each of its elements is i.i.d. and distributed according to  $\mathcal{N}(0, \sqrt{m}/\sqrt{r})$ . Then, (40) holds with probability at least  $1 - \exp(-r/4) - \exp(-c_2 r)$ , wherein  $\mu$  and  $\beta$  satisfy*

$$\begin{aligned} \mu^2 &\leq \frac{(1 + 2\alpha\sqrt{r^*} \delta_{2(r^*+1)}) \lambda_1(\bar{Z}^*)}{4(1 + \eta) \min\{r, md\}} \min \left\{ \frac{r\sqrt{r}}{2d\sqrt{d}(1 + (\rho r^* m)^{1/2})} \sigma^{-96\kappa^2}, \right. \\ &\quad \left. \frac{c(1 - \delta_\rho) e^{-1/2} (2 - \sqrt{2}) r \sqrt{r}}{16c_2 \kappa^2 d \sqrt{d} (1 + (\rho r^* m)^{1/2})} \sigma^{-12\kappa^2} \right\}, \end{aligned}$$

with

$$\sigma \leq 4c_2 \kappa^2 \frac{(1 + 2\sqrt{r^*} \alpha \delta_{2(r^*+1)}) \lambda_1(\bar{Z}) \sqrt{\frac{d}{r}} + (\rho^2 + \alpha \rho^{1/2} \Delta) \sqrt{\frac{dm}{r}}}{ce^{-1/2} (2 - \sqrt{2}) (1 + 2\alpha\sqrt{r^*} \delta_{2(r^*+1)}) \lambda_1(\bar{Z}^*)},$$

and

$$\sqrt{m} \frac{e^{-1/2} (2 - \sqrt{2})}{2c_2} \leq \beta \leq 2\sqrt{m} \sigma^2.$$

## B.2 Phase II

We exploit now the properties of the iterates at time  $t_*$  to establish further refinements on the dynamics of the signal and noise terms.

We begin establishing some bounds of the quantities of interest at time  $t + 1$ , conditioned to the following events at time  $t$ :

(a)  $\|U^t\| \leq 3\sqrt{m} \|\bar{X}\|;$

- (b)  $\|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| \leq c_1 \kappa^{-1}$ ;  
(c)  $\|\mathcal{W} \Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t)\| \leq c_2 m \lambda_{r^*}(\bar{Z}^*)$ .

We then proceed by induction to show that such events hold.

In the following we define for convenience

$$\Delta^t \triangleq Z^* - U^t (U^t)^\top. \quad (42)$$

Also  $\|\cdot\|$  denotes any matrix norm, thus fulfilling  $\|ABC\| \leq \|A\| \|B\| \|C\|$ .

**Lemma 9.** Assume that  $\alpha \leq \frac{1}{81000} \|\bar{Z}^*\|^{-1} \kappa^{-2}$ . Then we have the following:

(i) If (a) holds, with  $c_2 \leq 1$ , then

$$\|U^{t+1}\| \leq 3\sqrt{m} \|\bar{X}\|. \quad (43)$$

(ii) If (a)-(c) hold, with  $c_1, c_2 \leq 1/110$ , and  $V_{Z^*}^\top U^t$  has full rank, then

$$\sigma_{\min}(V_{Z^*}^\top U^{t+1} Q^t) \geq \left(1 + \frac{\alpha}{4} \lambda_{r^*}(\bar{Z}^*) - \frac{\alpha}{m} \sigma_{\min}^2(V_{Z^*}^\top U^t)\right) \sigma_{\min}(V_{Z^*}^\top U^t). \quad (44)$$

(iii) If (a), (b) hold, with  $c_1 \leq 1/1200$ ,  $V_{Z^*}^\top U^{t+1} W^t$  is full row rank, and

$$\alpha \leq (1/1200)m \|\mathcal{W} \Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t)\|^{-1}, \quad (45)$$

then

$$\begin{aligned} \|U^{t+1} Q^{t+1, \perp}\| &\leq \left(1 - \frac{2\alpha}{m} \left(\frac{\|U^t Q^{t, \perp}\|}{4} - \|\mathcal{W} \Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t)\|\right) + \right. \\ &\quad \left. + 9\alpha \|\bar{Z}^*\| \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| \right) \|U^t Q^{t, \perp}\|. \end{aligned} \quad (46)$$

(iv) If (a), (b) hold, with  $c_1 \leq 1/(24 \cdot 800)$ , and

$$\|U^t Q^{t, \perp}\| \leq \min\{2\sigma_{\min}(U^t Q^t), c_0 \kappa^{-2} \sqrt{m} \|\bar{X}\|\}, \quad (47)$$

then,

$$\begin{aligned} \|(V_{Z^*}^\perp)^\top V_{U^{t+1} Q^{t+1}}\| &\leq \left(1 - \alpha \frac{\lambda_{r^*}(\bar{Z}^*)}{4}\right) \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| \\ &\quad + \frac{100\alpha}{m} \|\mathcal{W} \Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t)\| + \frac{500\alpha^2}{m^2} \|\Delta^t\|^2. \end{aligned} \quad (48)$$

(v) If (a), (b) hold, with  $c_1 \leq 1/(81000)$ , and further

$$\sigma_{\min}(U^t Q^t) \geq \sqrt{\frac{m}{10}} \sigma_{r^*}(\bar{X}), \quad (49)$$

$$\begin{aligned} \max\{&\|V_{Z^*}^\top (\mathcal{W} \Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t))\|, \|V_{U^t Q^t}^\top (\mathcal{W} \Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t))\|, \\ &\|\mathcal{W} \Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t)\|\} \leq c_0 \kappa^{-2} \|\Delta^t\|, \end{aligned} \quad (50)$$

then,

$$\|V_{Z^*}^\top (\Delta^{t+1})\| \leq \left(1 - \frac{\alpha}{200} \lambda_{r^*}(\bar{Z}) + \alpha 9 \|\bar{Z}^*\| \rho\right) \|V_{Z^*}^\top \Delta^t\| + \frac{\alpha \lambda_{r^*}(\bar{Z}^*)}{100} \|U^t Q^{t, \perp} (U^t Q^{t, \perp})^\top\|. \quad (51)$$

*Proof.* See Appendix [F](#) □

Now, we strive to establish, via induction, that the events outlined in (a)-(c), alongside the additional conditions utilized in Lemma [9](#) are upheld. The proof is organized in two steps. **Step 1** delves into the dynamics of the quantities under scrutiny, initiating from the iterates at time  $t_*$  (which marks the exit from Phase I), hence possessing a favorable alignment with the signal subspace (in the sense of Proposition [1](#)). Consequently, it is proven that throughout the progression of iterations, the alignment, as quantified by  $\sigma_{\min}(V_{Z^*}^\top U^t)$ , accelerates at a pace that outstrips the misalignment, gauged via  $\|U^t Q^{t, \perp}\|$  and  $\|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\|$ , eventually attaining a suitably elevated value. Upon reaching an adequate growth in  $\sigma_{\min}(V_{Z^*}^\top U^t)$ , we demonstrate in **Step 2** that (i)  $\sigma_{\min}(V_{Z^*} \bar{Z}^*)$  establishes a stable lower limit, while (ii)  $\|U^t Q^{t, \perp}\|$  exhibits a gradual increase, (iii)  $\|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\|$  remains stable, and (iv) the generalization error continually diminishes to yield the final result.

### B.2.1 Step 1: Growing alignment phase

Select a time  $t_*$  such that there holds

$$\begin{aligned} \|U^{t_*} Q^{t_*, \perp}\| &\leq 2\tau, \\ \|U^{t_*}\| &\leq 3\sqrt{m}\|\bar{X}\|, \\ \|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\| &\leq \bar{c}_2 \kappa^{-2}, \\ \sigma_{\min}(U^{t_*} Q^{t_*}) &\geq \tau. \end{aligned} \tag{52}$$

We will anticipate that we will later identify such a  $t_*$  as that in Lemma 8 with  $\sigma$  as in Proposition 2 and  $\nu = \frac{\mu\beta}{4}$  as in Proposition 1 with bounds on  $\beta$  as in Proposition 2.

We prove next by induction that the following hold:

$$\sigma_{\min}(V_{\bar{Z}^*}^\top U^t) \geq \frac{1}{2} \left(1 + \frac{\alpha}{8} \lambda_{r^*}(\bar{Z}^*)\right)^{t-t_*} \tau, \tag{53a}$$

$$\|U^t Q^{t, \perp}\| \leq 2 \left(1 + 80\bar{c}_1 \lambda_{r^*}(\bar{Z})\right)^{t-t_*} \tau, \tag{53b}$$

$$\|U^t\| \leq 3\sqrt{m}\|\bar{X}\|, \tag{53c}$$

$$\|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\| \leq \bar{c}_2 \kappa^{-2}, \tag{53d}$$

for all  $t \in [t_*, t_1]$ , where

$$t_1 \triangleq \min \left\{ t \geq t_* : \sigma_{\min}(V_{\bar{Z}^*}^\top U^t) \geq \frac{\sigma_{r^*}(\bar{X})\sqrt{m}}{\sqrt{10}} \right\}.$$

Notice that (53) holds at time  $t = t_*$ . It is sufficient to note for (53a) that

$$\sigma_{\min}(V_{\bar{Z}^*}^\top U^{t_*}) \geq \sigma_{\min}(V_{\bar{Z}^*}^\top V_{U^{t_*} Q^{t_*}}) \sigma_{\min}(U^{t_*} Q^{t_*}) \stackrel{(a)}{\geq} (1 - \bar{c}_2 \kappa^{-2}) \sigma_{\min}(U^{t_*} Q^{t_*}) \stackrel{(b)}{\geq} \frac{\tau}{2},$$

where in (a) we used (52) and in (b) we set  $\bar{c}_2 \leq 1/2$ .

We begin proving that  $\|\mathcal{W}\Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t)\|$  is sufficiently small (event (c)) for any time  $t$  within the interval of interest. Using

$$\Delta_S^t \triangleq Z^* - U^t Q^t (U^t Q^t)^\perp \quad \text{and} \quad \Delta_\perp^t \triangleq -U^t Q^{t, \perp} (U^t Q^t)^\perp,$$

we have

$$\begin{aligned} \|\mathcal{A}^* \mathcal{A}(\Delta^t) - \mathcal{W}\Delta^t \mathcal{W}\| &= \|\mathcal{A}^* \mathcal{A}(\Delta_S^t + \Delta_\perp^t) - \mathcal{W}(\Delta_S^t + \Delta_\perp^t) \mathcal{W}\| \\ &\leq \|\mathcal{A}^* \mathcal{A}(\Delta_S^t) - \mathcal{W}\Delta_S^t \mathcal{W}\| + \|\mathcal{A}^* \mathcal{A}(\Delta_\perp^t) - \mathcal{W}\Delta_\perp^t \mathcal{W}\|. \end{aligned}$$

Observe that each of the  $d \times d$  blocks of  $\Delta_S^t$  is of rank at most  $2r^*$  and so is its average block. Hence, under the assumption that  $\bar{\mathcal{A}}$  satisfies the  $(\delta_{4(r^*+1)}, 4(r^*+1))$ -RIP, we can invoke Lemma 2 and obtain

$$\|\mathcal{A}^* \mathcal{A}(\Delta_S^t) - \mathcal{W}\Delta_S^t \mathcal{W}\| \leq 2\delta_{4(r^*+1)} \|\mathcal{J}\Delta_S^t \mathcal{J}\|_f + 2\eta_{4(r^*+1)} \|\Delta_S^t - \mathcal{J}\Delta_S^t \mathcal{J}\|_F,$$

where

$$\eta_{4(r^*+1)} \triangleq 2 \left( \rho + (1 + 2\delta_{4(r^*+1)}) \hat{\Delta}_{4(r^*+1)} \right).$$

Consequently,

$$\|\mathcal{A}^* \mathcal{A}(\Delta_S^t) - \mathcal{W}\Delta_S^t \mathcal{W}\| \leq \left( 2\sqrt{2r^*} \delta_{4(r^*+1)} + 2\sqrt{2r^* m} \eta_{4(r^*+1)} \right) \|\Delta_S^t\|.$$

Further,  $-\Delta_\perp^t$  is a p.s.d matrix with rank at most  $r - r^*$ . Thus, we may invoke Lemma 3

$$\|\mathcal{W}\Delta_\perp^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta_\perp^t)\| \leq (2\delta_4 + 2\eta_4) \|\Delta_\perp^t\|_*,$$

where

$$\eta_4 \triangleq 2\rho + 2(\rho + (1 + 2\delta_4)) \hat{\Delta}_4.$$



Combining we obtain

$$\|\mathcal{A}^* \mathcal{A}(\Delta^t) - \mathcal{W} \Delta^t \mathcal{W}\| \leq 2\sqrt{2r^*} (\delta_{4(r^*+1)} + m\eta_{4(r^*+1)}) \|\Delta_S^t\| + 2(r - r^*)(\delta_4 + \eta_4) \|\Delta_\perp^t\|.$$

Using the induction hypothesis (53c), we can write

$$\begin{aligned} \|\mathcal{A}^* \mathcal{A}(\Delta^t) - \mathcal{W} \Delta^t \mathcal{W}\| &\leq 2\sqrt{2r^*} (\delta_{4(r^*+1)} + m\eta_{4(r^*+1)}) 10m \|\bar{Z}^*\| \\ &\quad + 2(r - r^*)(\delta_4 + \eta_4) \|U^t Q^{t,\perp}\|^2. \end{aligned}$$

Further, under the induction hypothesis (53b), it holds

$$\begin{aligned} \|\mathcal{A}^* \mathcal{A}(\Delta^t) - \mathcal{W} \Delta^t \mathcal{W}\| &\leq 2\sqrt{2r^*} (\delta_{4(r^*+1)} + m\eta_{4(r^*+1)}) 10m \|\bar{Z}^*\| \\ &\quad + 8(r - r^*)(\delta_4 + \eta_4) (1 + 80\alpha\bar{c}_1\lambda_{r^*}(\bar{Z}^*))^{2(t-t_*)} \tau^2. \end{aligned}$$

By (53a), there must hold

$$t_1 - t_* \leq \frac{16}{\alpha\lambda_{r^*}(\bar{Z}^*)} \ln \left( \sqrt{\frac{5}{2}} \frac{\sigma_{r^*}(\bar{X})\sqrt{m}}{\tau} \right). \quad (54)$$

Given that

$$\max \left\{ 2\sqrt{2}(\delta_{4(r^*+1)} + m\eta_{4(r^*+1)}), (\delta_4 + \eta_4) \right\} \leq \frac{\bar{c}_3\kappa^{-4}}{\sqrt{r^*}},$$

we have that for all  $t \in [t_*, t_1]$

$$\|\mathcal{A}^* \mathcal{A}(\Delta^t) - \mathcal{W} \Delta^t \mathcal{W}\| \leq 10\bar{c}_3 m \kappa^{-4} \|\bar{Z}^*\| + (r - r^*) \frac{\bar{c}_3}{\sqrt{r^*}} \kappa^{-4} (1 + \alpha 80\bar{c}_1 \lambda_{r^*}(\bar{Z}^*))^{2(t_1-t_*)} \tau^2.$$

Assuming  $r \leq md$ , using (54), with

$$\tau \leq \bar{c}_4 \frac{\sigma_{r^*}(\bar{X})\sqrt{m}}{\min\{r, md\}\kappa^2},$$

and choosing  $8\bar{c}_4^{7/4} \leq 30$ , it follows that

$$\|\mathcal{A}^* \mathcal{A}(\Delta^t) - \mathcal{W} \Delta^t \mathcal{W}\| \leq (10\bar{c}_3 + 8\bar{c}_3\bar{c}_4^{7/4}) m \kappa^{-2} \lambda_{r^*}(\bar{Z}^*) \leq 40\bar{c}_3 m \kappa^{-2} \lambda_{r^*}(\bar{Z}^*). \quad (55)$$

Under the assumption that  $\alpha \leq \bar{c}_5 \|\bar{Z}^*\|^{-1} \kappa^{-4}$  and  $\bar{c}_5 < 1/110$ ,  $\bar{c}_2 < 1/110$  and  $40\bar{c}_3 < 1/110$ , we can invoke Lemma 9(ii), and obtain

$$\begin{aligned} \sigma_{\min}(V_{Z^*}^\top U^{t+1} Q^{t+1}) &= \sigma_{\min}(V_{Z^*}^\top U^{t+1}) \geq \sigma_{\min}(V_{Z^*}^\top U^{t+1} Q^t) \\ &\geq \left(1 + \frac{\alpha}{4} \lambda_{r^*}(\bar{Z}^*) - \frac{\alpha}{m} \sigma_{\min}^2(V_{Z^*}^\top U^t)\right) \sigma_{\min}(V_{Z^*}^\top U^t). \end{aligned}$$

Because  $t \leq t_1$ ,

$$\sigma_{\min}(V_{Z^*}^\top U^t) \leq \frac{\sqrt{m} \sigma_{\min}(\bar{X})}{\sqrt{10}},$$

and therefore

$$\sigma_{\min}(V_{Z^*}^\top U^{t+1}) \geq \left(1 + \frac{3\alpha}{20} \lambda_{r^*}(\bar{Z}^*)\right) \sigma_{\min}(V_{Z^*}^\top U^t).$$

Using the induction hypothesis (53a) there holds

$$\sigma_{\min}(V_{Z^*}^\top U^{t+1}) \geq \frac{\tau}{2} \left(1 + \frac{\alpha}{8} \lambda_{r^*}(\bar{Z}^*)\right)^{t+1-t_*}, \quad (56)$$

and thus (53a) holds.

Observe that (56) implies that  $V_{Z^*}^\top U^{t+1} Q^t$  is full rank. Consequently, we can invoke Lemma 9(iii): under  $\bar{c}_5 \leq \frac{\bar{c}_3}{40}, \frac{\bar{c}_3}{40} < \frac{1}{1200}$ , and  $\bar{c}_2 < 1/1200$ , it holds

$$\begin{aligned} \|U^{t+1} Q^{t+1,\perp}\| &\leq \left(1 - \frac{\alpha}{m} \left(\frac{\|U^t Q^{t,\perp}\|}{2} - 2\|\mathcal{A}^* \mathcal{A}(\Delta^t) - \mathcal{W} \Delta^t \mathcal{W}\|\right) + 9\alpha \|\bar{Z}^*\| \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\|\right) \|U^t Q^{t,\perp}\| \\ &\stackrel{(a)}{\leq} (1 + 80\alpha\bar{c}_1\lambda_{r^*}(\bar{Z}^*)) \|U^t Q^{t,\perp}\|, \end{aligned}$$

where in (a) we used  $80\bar{c}_3 + 9\bar{c}_2 < 80\bar{c}_1$ .

Under the induction hypothesis (53b), we then have

$$\|U^{t+1}Q^{t+1,\perp}\| \leq 2(1 + \alpha 80\bar{c}_1\lambda_{r^*}(\bar{Z}^*))^{t+1-t_*} \tau,$$

and thus (53b) is established.

Under the induction hypothesis (53c) and (55) we can invoke Lemma 9(i) yielding

$$\|U^{t+1}\| \leq 3\sqrt{m}\|\bar{X}\|.$$

Next, we invoke Lemma 9(iv) under

$$\bar{c}_5 \leq 1/(24 \cdot 800), \quad 40\bar{c}_3 \leq 1/(24 \cdot 800), \quad \text{and} \quad \bar{c}_2 \leq 1/(24 \cdot 800).$$

In fact, since

$$\|U^tQ^{t,\perp}\| \leq 2\sqrt{2}c_4^{7/8} \frac{\sigma_{r^*}(\bar{X})\sqrt{m}}{r^{7/8}\kappa^{14/8}}, \quad \forall t \in [t_1, t_*],$$

choosing  $2\sqrt{2}c_4^{7/4} \leq 1/(24 \cdot 800)$ , all conditions of Lemma 9(iv) are satisfied. Therefore,

$$\begin{aligned} \|(V_{\bar{Z}^*}^\perp)^\top V_{U^{t+1}Q^{t+1}}\| &\leq \left(1 - \frac{\alpha}{4}\lambda_{r^*}(\bar{Z}^*)\right) \|(V_{\bar{Z}^*}^\perp)^\top V_{U^tQ^t}\| \\ &\quad + \frac{100\alpha}{m} \|\mathcal{W}\Delta^t\mathcal{W} - \mathcal{A}^*\mathcal{A}(\Delta^t)\| + \frac{500\alpha^2}{m^2} \|\Delta^t\|^2, \end{aligned}$$

whereby (55), (53c) and induction hypothesis (53d) yields

$$\|(V_{\bar{Z}^*}^\perp)^\top V_{U^{t+1}Q^{t+1}}\| \leq (\bar{c}_2(1 - \alpha/4\lambda_{r^*}(\bar{Z})) + 4000\bar{c}_3\alpha\lambda_{r^*}(\bar{Z}^*) + 5000\bar{c}_5\alpha\lambda_{r^*}(\bar{Z}^*)) \kappa^{-2}.$$

Thus, requesting  $\bar{c}_2/4 \geq 4000\bar{c}_3 + 5000\bar{c}_5$  yields

$$\|(V_{\bar{Z}^*}^\perp)^\top V_{U^{t+1}Q^{t+1}}\| \leq \bar{c}_2\kappa^{-2}$$

concluding the proof of (53).

## B.2.2 Step 2: Refinement phase

Define

$$\begin{aligned} t_a &\triangleq t_1 + \left\lfloor \frac{300}{\alpha\lambda_{r^*}(\bar{Z}^*)} \ln \left( \frac{\kappa^{1/4}}{8} \sqrt{\frac{r^*}{r-r^*}} \frac{m^{7/8}\|\bar{X}\|^{7/4}}{\tau^{7/4}} \right) \right\rfloor, \\ t_b &\triangleq \min\{t : (\sqrt{r-r^*} + 1)\|\Delta_\perp^t\|_F \geq \|\Delta^t\|_F, t \geq t_1\}, \\ t_2 &\triangleq \min\{t_a, t_b\}. \end{aligned}$$

The proof follows in a similar fashion as Step 1. More specifically, it is shown inductively that

$$\sigma_{\min}(U^tQ^t) \geq \sigma_{\min}(V_{\bar{Z}^*}^\top U^t) \geq \frac{\sigma_{r^*}(\bar{X})\sqrt{m}}{\sqrt{10}}, \quad (57a)$$

$$\|U^tQ^{t,\perp}\| \leq (1 + 80\alpha\bar{c}_1\lambda_{r^*}(\bar{Z}^*))^{t-t_1} \|U^{t_1}Q^{t_1,\perp}\|, \quad (57b)$$

$$\|U^t\| \leq 3\sqrt{m}\|\bar{X}\| \quad (57c)$$

$$\|(V_{\bar{Z}^*}^\perp)^\top V_{U^tQ^t}\| \leq \bar{c}_2\kappa^{-2}, \quad (57d)$$

$$\|V_{\bar{Z}^*}^\top \Delta^t\|_F \leq 10\sqrt{r^*} \left(1 - \frac{\alpha}{300}\lambda_{r^*}(\bar{Z}^*) + 9\alpha\rho\|\bar{Z}^*\|\right)^{t-t_1} m\|\bar{Z}^*\|, \quad (57e)$$

for all  $t \in [t_1, t_2]$ .

Observe that the above inequalities but (57e) readily hold at  $t = t_1$ . Thus, we start showing that (57e) holds as well at time  $t_1$ .

Observe that

$$\|V_{\bar{Z}^*}^\top \Delta^{t_1}\|_F \leq \|V_{\bar{Z}^*}^\top \Delta_S^{t_1}\|_F + \|V_{\bar{Z}^*}^\top \Delta_\perp^{t_1}\|_F.$$

By definition  $V_{Z^*}^\top \Delta_\perp^{t_1} = 0$ . Therefore,

$$\|V_{Z^*}^\top \Delta^{t_1}\|_F \leq \|V_{Z^*}^\top \Delta_S^{t_1}\|_F \leq \|Z^*\|_F + \|U^{t_1} Q^{t_1} (U^{t_1} Q^{t_1})^\top\|_F,$$

whereby (57c) at  $t = t_1$  (57e) at  $t = t_1$  follows.

We prove now (57), following a similar approach as in Step 1. We need to invoke Lemma 9. For this, we proceed similarly, and obtain

$$\begin{aligned} \|\mathcal{A}^* \mathcal{A}(\Delta^t) - \mathcal{W} \Delta^t \mathcal{W}\| &\leq \left(2\sqrt{2}\delta_{4(r^*+1)} + 2\sqrt{2}m\eta_{4(r^*+1)}\right) 10m\sqrt{r^*} \|\bar{Z}^*\| \\ &\quad + 2(r - r^*)(\delta_4 + \eta_4) \|U^t Q^{t,\perp}\|^2. \end{aligned}$$

At evaluating  $t_1$  the end of Phase II it holds that

$$\|U^{t_1} Q^{t_1,\perp}\|^2 \leq 8\sigma_{\min}^{1/4}(\bar{X}) \tau^{7/4} m^{1/8}.$$

Consequently, under the assumption that

$$\max\{2\sqrt{2}\delta_{4(r^*+1)} + 2\sqrt{2}m\eta_{4(r^*+1)}, 2(\delta_4 + \eta_4)\} \leq \frac{\bar{c}_3 \kappa^{-4}}{\sqrt{r^*}}, \quad (58)$$

there holds

$$\begin{aligned} \|\mathcal{A}^* \mathcal{A}(\Delta^t) - \mathcal{W} \Delta^t \mathcal{W}\| &\leq 10m\bar{c}_3 \kappa^{-2} \lambda_{r^*}(\bar{Z}^*) \\ &\quad + \frac{r - r^*}{\sqrt{r^*}} \bar{c}_3 \kappa^{-4} (1 + 80\alpha\bar{c}_1 \sigma_{\min}^2(\bar{X}))^{t_a - t_1} 8\sigma_{\min}^{1/4}(\bar{X}) \tau^{7/4} m^{1/8}, \end{aligned}$$

for all  $t \leq t_2$ . Given that

$$\frac{1}{8} \kappa^{1/4} \sqrt{\frac{r^*}{r - r^*}} \frac{m^{7/8} \|\bar{X}\|^{7/4}}{\tau^{7/4}} > 1,$$

if  $\bar{c}_1 \leq \bar{c}_3$ , we have

$$(1 + \alpha\bar{c}_1 80\sigma_{\min}^2(\bar{X}))^{t_a - t_1} \leq \left(\frac{\kappa^{1/4}}{8} \sqrt{\frac{r^*}{r - r^*}} \frac{m^{7/8} \|\bar{X}\|^{7/4}}{\tau^{7/4}}\right)^{24000\bar{c}_3}.$$

Further requesting  $c_3 \leq 1/(48000)$ , yields

$$\begin{aligned} \|\mathcal{A}^* \mathcal{A}(\Delta^t) - \mathcal{W} \Delta^t \mathcal{W}\| &\leq 10m\bar{c}_3 \kappa^{-2} \lambda_{r^*}(\bar{Z}^*) \\ &\quad + \frac{r - r^*}{\sqrt{r^*}} \bar{c}_3 \kappa^{-4} \sqrt{\frac{1}{8}} \frac{(r^*)^{1/4}}{(r - r^*)^{1/4}} \kappa^{-6/8} \|\bar{X}\|^{7/8} \sigma_{\min}^{9/8}(\bar{X}) m. \end{aligned}$$

Using  $r \geq 1$  and  $r^* \geq 1$ , yields

$$\|\mathcal{A}^* \mathcal{A}(\Delta^t) - \mathcal{W} \Delta^t \mathcal{W}\| \leq 40m\bar{c}_3 \kappa^{-2} \sigma_{r^*}^2(\bar{X}). \quad (59)$$

Given (59) and the induction hypotheses (57c), (57d) we can invoke Lemma 9(ii) and infer

$$\sigma_{\min}(U^{t+1} Q^{t+1}) \geq \sigma_{\min}(V_{Z^*}^\top U^{t+1}) \geq \sigma_{\min}(V_{Z^*}^\top) \left(1 + \frac{\alpha}{4} \lambda_{r^*}(\bar{Z}^*) - \frac{\alpha}{m} \sigma_{r^*}^2(V_{Z^*}^\top U^t)\right).$$

Following [34, Theorem 9.6], where we consider the case in which  $\sigma_{\min}(V_{Z^*}^\top U^{t+1} Q^{t+1}) \geq \frac{1}{2} \sigma_{\min}(X)$  and the complement to yield

$$\sigma_{\min}(V_{Z^*}^\top U^{t+1} Q^{t+1}) \geq \frac{\sigma_{\min}(\bar{X}) \sqrt{m}}{\sqrt{10}}$$

as long as  $\bar{c}_5 \leq 1/45$ . Observe that this implies that  $V_{Z^*}^\top U^{t+1} Q^{t+1}$  is invertible. Consequently, (57b) can be established in the same way as in Step 1. The same approach applies also to prove (57c) and (57d).

We have left to establish (57e). For this, we invoke Lemmas 3 and 2, and write

$$\begin{aligned} \|V_{Z^*}^\top (\mathcal{W} \Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t))\|_F &\leq \|V_{Z^*}^\top (\mathcal{W} \Delta_S^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta_S^t))\|_F + \|V_{Z^*}^\top (\mathcal{W} \Delta_\perp^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta_\perp^t))\|_F \\ &\leq (2\delta_{4(r^*+1)} + 2\eta_{4(r^*+1)}) \|\Delta_S^t\|_F + (2\delta_4 + 2\eta_4) \|\Delta_\perp^t\|_*, \end{aligned}$$

whereby if  $t \leq t_2$ , it holds

$$(\sqrt{r - r^*} + 1) \|U^t W^t\|_F^2 \leq \|\Delta^t\|_F.$$

Therefore, under the assumptions on the RIP that we have imposed in (58), it holds

$$\|V_{Z^*}^\top (\mathcal{W} \Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t))\|_F \leq \frac{2\bar{c}_3}{\kappa^2} \|\Delta^t\|_F.$$

Following a similar procedure we can establish the same bound with the difference in  $V_{Z^*}$  to  $V_{U^t W^t}^\top$  or no projection at all. Given that  $\bar{c}_3 \leq 1/(2 \cdot 81000)$ ,  $\bar{c}_5 \leq 1/81000$ , and  $\bar{c}_2 \leq 1/81000$ , we can invoke Lemma 9(v), yielding

$$\|V_{Z^*}^\top \Delta^{t+1}\|_F \leq \left(1 - \frac{\alpha}{200} \lambda_{r^*}(\bar{Z}^*) + 9\alpha \|\bar{Z}^*\| \rho\right) \|V_{Z^*}^\top \Delta^t\|_F + \frac{\alpha \lambda_{r^*}(\bar{Z}^*)}{100} \|U^t Q^{t,\perp} (U^t Q^{t,\perp})^\top\|_F.$$

Using the induction hypothesis (57e) and  $\|U^{t_1} Q^{t_1,\perp}\| \leq 8\sigma_{\min}^{1/4} \tau^{7/4} m^{1/8}$ , we obtain

$$\begin{aligned} \|U^t Q^{t,\perp} (U^t Q^{t,\perp})^\top\|_F &\leq \sqrt{r - r^*} \|U^t Q^{t,\perp}\|^2 \\ &\leq \sqrt{r - r^*} (1 + \alpha \bar{c}_1 80 \lambda_{r^*}(\bar{Z}^*))^{2(t_2 - t_1)} 8\sigma_{\min}^{1/4}(\bar{X}) \tau^{7/4} m^{1/8}. \end{aligned}$$

This is implied by choosing

$$80\bar{c}_2\bar{c}_5 \leq \frac{\sqrt{400} - \sqrt{399}}{\sqrt{399}};$$

which yields

$$\begin{aligned} \|V_{Z^*}^\top \Delta^{t+1}\|_F &\leq \left(1 - \frac{\alpha \lambda_{r^*}(\bar{Z}^*)}{200} + 9\alpha \|\bar{Z}^*\| \rho\right) \|V_{Z^*}^\top \Delta^t\|_F \\ &\quad + \frac{\alpha \lambda_{r^*}(\bar{Z}^*)}{100} \sqrt{r^*} \left(1 - \frac{\alpha}{400} \lambda_{r^*}(\bar{Z}^*)\right)^{t-t_1} m \|\bar{Z}^*\|, \end{aligned}$$

whereby invoking the induction hypothesis we have

$$\begin{aligned} \|V_{Z^*}^\top \Delta^{t+1}\| &\leq 10\sqrt{r^*} m \|\bar{Z}^*\| \left(1 - \frac{\alpha}{300} \lambda_{r^*}(\bar{Z}^*) + 9\alpha \|\bar{Z}^*\| \rho\right)^{t-t_1} \times \\ &\quad \times \left(1 - \frac{\alpha}{200} \lambda_{r^*}(\bar{Z}^*) + 9\alpha \|\bar{Z}^*\| \rho\right) + \frac{\alpha \lambda_{r^*}(\bar{Z}^*)}{1000} \sqrt{r^*} m 10 \|\bar{Z}^*\| \left(1 - \frac{\alpha}{400} \lambda_{r^*}(\bar{Z}^*)\right)^{t-t_1}. \end{aligned}$$

This concludes the proof by induction.

We have now all the ingredients to show that by time  $t_2$  the generalization error is small.

From [34] Lemma B.4] it follows that

$$\|U^{t_2} (U^{t_2})^\top - Z^*\|_F \leq 4 \|V_{Z^*}^\top (Z^* - U^{t_2} (U^{t_2})^\top)\|_F + \|U^{t_2} Q^{t_2,\perp} (U^{t_2} Q^{t_2,\perp})^\top\|_F.$$

To upper bound the RHS, we look separately at the two cases in which  $t_2 = t_a$  and  $t_2 = t_b$ .

We start with  $t_2 = t_a$ : by setting  $\rho \leq \frac{\lambda_{r^*}(\bar{Z}^*)}{1200 \cdot 9 \|\bar{Z}^*\|}$  we obtain

$$\|\Delta^{t_2}\|_F \leq 40\sqrt{r^*} \left(1 - \frac{\alpha \lambda_{r^*}(\bar{Z}^*)}{400}\right)^{t_2 - t_1} m \|\bar{Z}^*\| + \sqrt{r^*} \left(1 - \frac{\alpha \lambda_{r^*}(\bar{Z}^*)}{400}\right)^{t_2 - t_1} m \|\bar{Z}^*\|,$$

evaluating

$$\|\Delta^{\hat{t}}\|_F \leq 8(r^*)^{1/8} (r - r^*)^{3/8} \|\bar{X}\|^{11/16} m^{11/32} \kappa^{-3/16} \tau^{21/16}.$$

The case  $t_2 = t_b$ : Requesting  $\bar{c}_1 \leq 1/(19200)$  and following the procedure in [35] Theorem 9.6, Phase III] yields the same result for the case  $t_2 = t_b$ .

Summing up the times  $t_1$  and  $t_2$  yields the following result.

### B.2.3 Summary

We summarize the established result in the following proposition.

**Proposition 3.** Consider the matrix sensing problem [\(1\)](#), with augmented ground-truth  $Z^*$ , with  $r^* + 1 \leq r \leq md$ , and the measurement operator  $\bar{A}$  satisfying the  $4(r^* + 1, \delta)$ -RIP, with  $\delta \lesssim \frac{\kappa^{-4}}{\sqrt{r^*}}$ . Let  $\{U^t\}_t$  be the (augmented) sequence generated by Algorithm [3](#) under the following tuning: (i) the stepsize  $\alpha \lesssim \kappa^{-4} \|\bar{X}\|^{-2}$ , (ii) the gossip matrix is chosen to satisfy Assumption [7](#) with

$$\rho \lesssim \frac{\kappa^{-2} \delta}{\sqrt{r^* m^2} \sqrt{m}}.$$

Then, assume the existence of an iteration count  $t_*$  such that for  $\tau > 0$  the following holds

$$\begin{aligned} \|U^{t_*}\| &\leq 3\sqrt{m} \|\bar{X}\|, & \sigma_{\min}(U^{t_*} Q^{t_*}) &\geq \tau, \\ \|U^{t_*} Q^{t_*, \perp}\| &\leq 2\tau, & \|(V_{Z^*}^\perp)^\top V_{U^{t_*} Q^{t_*}}\| &\leq \bar{c}_2 \kappa^{-2}, \end{aligned}$$

with

$$\tau \lesssim \frac{\sigma_{r^*}(\bar{X}) \sqrt{m}}{\min\{r, md\} \kappa^2}.$$

Then, after

$$t_2 - t_* \leq \frac{2}{\alpha \lambda_{r^*}(\bar{Z}^*)} \left( 8 \ln \left( \sqrt{\frac{5}{2}} \frac{\sigma_{r^*}(\bar{X}) \sqrt{m}}{\tau} \right) + 150 \ln \left( \frac{\kappa^{1/4}}{8} \sqrt{\frac{r^*}{r - r^*}} \frac{m^{7/8} \|\bar{X}\|^{7/4}}{\tau^{7/4}} \right) \right)$$

iterations,

$$\|\Delta^{t_2}\|_F \leq 50 (1 + \sqrt{r - r^*}) (r^*)^{1/8} (r - r^*)^{3/8} \|\bar{X}\|^{11/16} m^{11/32} \tau^{21/16} \kappa^{-3/16}$$

**Remark 1.** The statement in Proposition [3](#) is made only for  $r > r^*$ . The case  $r = r^*$  can be obtained by setting  $r = r^* + 1$  and setting the last column of each  $\bar{U}_i^0 = 0$ . Observe that this will column remain zero for all  $t$ , and is thus equivalent to the case of  $r = r^*$ .

We are now ready to establish that Theorem [1](#) holds. For this we identify the  $t_*$  in Proposition [3](#) with  $t_*$  in Lemma [8](#). This implies through Proposition [1](#) that the time  $t_*$  exists with  $\tau = \frac{\mu\beta}{4}$ . Consequently,  $\sigma$ , and  $\beta$ ,  $\rho$  and  $\mu$  behave as stated and required in Proposition [2](#), and the statement holds with the same probability.

## C Convergence of Algorithm [3](#) in the setting $r \in [r^*, 2r^*]$

We provide the counterpart of Theorem [1](#) in the setting  $r \in [r^*, 2r^*]$ .

**Theorem 2.** Consider the matrix sensing problem [\(1\)](#), with augmented ground-truth  $Z^*$ , under  $r^* \leq r \leq 2r^*$ , and the measurement operator  $\bar{A}$  satisfying the  $4(r^* + 1, \delta)$ -RIP with  $\delta \lesssim \kappa^{-4} (r^*)^{-1/2}$ . Let  $\{U^t\}_t$  be the (augmented) sequence generated by Algorithm [9](#) under the following tuning: (i) the stepsize  $\alpha \lesssim \kappa^{-4} \|\bar{X}\|^{-2}$ ; (ii) the gossip matrix  $W$  chosen to satisfy Assumption [7](#) with

$$\rho \lesssim \frac{\delta^2}{m^6 \kappa^4 r^*}$$

and (iii) the initialization  $U^0$  is chosen as  $U^0 = \mu U$ , where  $U \in \mathbb{R}^{md \times r}$  has i.i.d.  $\mathcal{N}(0, \sqrt{m/r})$  distributed entries, and  $\mu$  satisfies

$$\mu^2 \lesssim \min \left\{ \frac{\varepsilon^2 \sigma_{r^*}(\bar{X})}{\kappa^6 dr \min\{r, md\}}, \frac{\sqrt{r}}{d\sqrt{d}} \left( \kappa^2 \frac{\sqrt{dr}}{\varepsilon} \right)^{-96\kappa^2} \right\}.$$

Then, after

$$\hat{t} \lesssim \frac{1}{\alpha \sigma_{r^*}(\bar{X})} \left( \ln \left( \frac{\kappa^2 \sqrt{rd}}{\varepsilon} \right) + \ln \left( \frac{\sigma_{r^*}(\bar{X})}{\varepsilon \mu} \right) + \ln \left( \max \left\{ 1, \kappa \frac{r^*}{\max\{1, r - r^*\}} \right\} \frac{r \|\bar{X}\|}{\mu \varepsilon} \right) \right)$$

iterations, there holds

$$\frac{\|Z^* - U^{\hat{t}}(U^{\hat{t}})\|_F}{\|Z^*\|} \lesssim (\max\{r - r^*, 1\})^{7/8} (r^*)^{1/8} \|\bar{X}\|^{-21/16} \mu^{21/16} \left( \kappa^4 \frac{dr}{\varepsilon} \right)^{21/16}.$$

with probability at least  $1 - (c_1 \varepsilon)^{r - r^* + 1} - \exp(-c_2 r)$  where  $c_1 > 0$ ,  $c_2 > 0$  are universal constants.

Similar to Theorem 1, Theorem 2 can be seen as an extension to the centralized counterpart in [34], with a degradation on the allowed initialization size that stems from effectively having a larger RIP due to the network. Observe that in opposition to Theorem 1 we allow  $\hat{t}$  to be increasing with  $r$  as per assumption,  $r \leq 2r^*$ . Comparing the statements in Theorems 1 and 2 allows us to conclude that overparametrizing does not yield a performance loss in terms of computational complexity but will incur additional communication costs due to the exchange of additional variables.

### C.1 Proof of Theorem 2

**Proposition 4.** Assume that  $2r^* \geq r \geq r^*$ , and instantiate here the assumptions in Proposition 1. Then, if  $U$  is initialized such that each element of  $U$  is i.i.d. and distributed according to  $\mathcal{N}(0, \sqrt{m}/\sqrt{r})$ , with probability at least  $1 - (c_1\varepsilon)^{r-r^*+1} - \exp(-c_2r)$  the statements in Proposition 1 hold with

$$\mu^2 \leq \frac{(1 + 2\alpha\sqrt{r^*}\delta_{2(r^*+1)}\lambda_1(\bar{Z}^*))}{4(1 + \eta) \min\{r, md\}} \min \left\{ \frac{r\sqrt{r}}{2d\sqrt{d}(1 + (\rho r^* m)^{1/2})} \sigma^{-96\kappa^2} \right. \\ \left. \frac{\tilde{c}_3(1 - \delta_\rho)\varepsilon\sqrt{r}}{8c_2\kappa^2 d\sqrt{d}(1 + (\rho r^* m)^{1/2})} \sigma^{-12\kappa^2} \right\}$$

with

$$\sigma \leq 2\kappa^2 \frac{\sqrt{dr} \left( (1 + 2\alpha\sqrt{r^*}\delta_{2(r^*+1)}\lambda_1(\bar{Z}^*)) + (\rho^2 + \alpha\rho^{1/2}\Delta)m^{1/2} \right)}{\tilde{c}_3\varepsilon(1 + 2\alpha\sqrt{r^*}\delta_{2(r^*+1)}\lambda_1(\bar{Z}^*))}$$

and

$$\sqrt{m}\frac{\varepsilon}{r} \leq \beta \leq 2\sqrt{m}\sigma^2.$$

*Proof.* See Appendix E. □

Observe that Proposition 5 already accommodates  $r > r^*$ . Consequently, we are only missing the choice  $r = r^*$ , which we cover in the following Proposition.

**Proposition 5.** Consider the matrix sensing problem (1), with augmented ground-truth  $Z^*$ , with  $r = r^*$  and the measurement operator  $\bar{A}$  satisfying the  $4(r^* + 1, \delta)$ -RIP with  $\delta \lesssim \frac{\kappa^{-4}}{\sqrt{r^*}}$ . Let  $\{U^t\}_t$  be the (augmented) sequence generated by Algorithm 3 where  $\bar{U}_i^t \in \mathbb{R}^{md \times r^*+1}$  where w.l.o.g. the last column is set to all zeros for all  $i \in [m]$ . Under the following tuning: (i) the stepsize  $\alpha \lesssim \kappa^{-4}\|\bar{X}\|^{-2}$ , (ii) the gossip matrix is chosen to satisfy Assumption 1 with

$$\rho \lesssim \frac{\kappa^{-2}\delta}{\sqrt{r^*}m^2\sqrt{m}}.$$

Then, choose  $\tau > 0$  for which

$$\|U^{t_*}\| \leq 3\sqrt{m}\|\bar{X}\|, \quad \sigma_{\min}(U^{t_*}Q^{t_*}) \geq \tau, \\ \|U^{t_*}Q^{t_*,\perp}\| \leq 2\tau, \quad \|(V_{\bar{Z}^*}^\perp)^\top V_{U^{t_*}Q^{t_*}}\| \leq \bar{c}_2\kappa^{-2},$$

and

$$\tau \lesssim \frac{\sigma_{r^*}(\bar{X})\sqrt{m}}{\min\{r, md\}\kappa^2}.$$

Then, after

$$t_2 - t_* \leq \frac{2}{\alpha\lambda_{r^*}(\bar{Z}^*)} \left( 8 \ln \left( \sqrt{\frac{5}{2}} \frac{\sigma_{r^*}(\bar{X})\sqrt{m}}{\tau} \right) + 150 \ln \left( \frac{\kappa^{1/4}}{8} \sqrt{\frac{r^*}{\max\{r - r^*, 1\}}} \frac{m^{7/8}\|\bar{X}\|^{7/4}}{\tau^{7/4}} \right) \right)$$

iterations, it holds

$$\|\Delta^{t_2}\|_F \leq 50 \left( 1 + \max\{\sqrt{r - r^*}, 1\} \right) (r^*)^{1/8} (r - r^*)^{3/8} \|\bar{X}\|^{11/16} m^{11/32} \tau^{21/16} \kappa^{-3/16}$$

*Proof.* The proof follows by applying Proposition 5 to the iterates defined with the extra column of zeros. This yields an equivalent scheme to using Algorithm 3 with  $U^t \in \mathbb{R}^{md \times r^*}$ .  $\square$

We are now ready to establish that Theorem 2 holds. For this we identify the  $t_*$  in Proposition 5 with  $t_*$  in Lemma 8. This implies through Proposition 1 that the time  $t_*$  exists with  $\tau = \frac{\mu\beta}{4}$ . Consequently,  $\sigma$ , and  $\beta$ ,  $\rho$  and  $\mu$  behave as stated and required in Proposition 4, and the statement holds with the same probability.

## D Proof of Corollary 1

We begin with the following Corollary.

**Corollary 2.** *Under the conditions of Theorem 3 it holds*

$$\frac{\|U^t(U^t)^\top - Z^*\|_F}{\|Z^*\|^2} \lesssim \left( \max\{(r - r^*)^{5/8}, 1\} (r^*)^{1/8} \|\bar{X}\|^{-21/16} \left(\frac{\tau}{\sqrt{m}}\right)^{13/16} \right), \quad (60)$$

for any  $t \in [T, \hat{t}]$ , where

$$T \gtrsim \frac{1 - \frac{\sqrt{m}}{\tau}}{\alpha\kappa \|\bar{Z}^*\| \max\{r - r^*, 1\} \left(\frac{\tau}{\sqrt{m}}\right)^{1/4}}. \quad (61)$$

Observe that then, we may choose  $\tau = \frac{\mu\beta}{4}$  with  $\mu$  and  $\beta$  as in Proposition 1 and Lemma 8, yielding the result in Corollary 1. We now proceed to prove Corollary 2. The proof consists of two parts. In **Step I** we start off by the iterates' properties at time  $\hat{t}$ . We then proceed to establish the result by induction. More specifically, we establish that the alignment  $\sigma_{\min}(U^t Q^t)$  remains bounded away from zero while the error  $\|U^t Q^{t,\perp}\|$  grows slowly and the error  $\|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\|$  shrinks so as to become  $\mathcal{O}(\tau^k)$ , for some  $k > 0$ . This is a critical requirement for **Step II** as the error  $\|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\|$  together with the estimation error are what dictate the *rate of growth* of  $\|U^t Q^{t,\perp}\|$  (c.f. Lemma 9 (iii)). Consequently, in **Step II** we part from the iterates' properties at the end of **Step I**, and exploit the fact that both the estimation error and  $\|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| \leq \mathcal{O}(\tau^k)$  to establish that the rate of growth of the remaining error is  $\mathcal{O}(\tau^k)$ . This allows us to claim that the estimation error will grow extremely slowly, thus providing with a large window throughout which it is small.

Recall that we may write

$$\|\mathcal{W}\Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t)\| \leq \|\mathcal{W}\Delta_S^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta_S^t)\| + \|\mathcal{W}\Delta_\perp^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta_\perp^t)\|.$$

Invoking Lemmas 2 and 3 yields

$$\|\mathcal{W}\Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t)\| \leq (2\delta_4(r^{*+1}) + \eta_4(r^{*+1})) \|\Delta_S^t\|_F + (2\delta_4 + \eta_4) \|\Delta_\perp^t\|_*.$$

For convenience we define

$$\Delta \triangleq \max\{2\delta_4(r^{*+1}) + \eta_4(r^{*+1}), 2\delta_4 + \eta_4\}.$$

Under the assumption that

$$\begin{aligned} \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| &\leq \kappa^{-2}, \\ \sigma_{\min}(V_{Z^*}^\top U^t Q^t) &\geq \frac{\sigma_{r^*}(\bar{X})\sqrt{m}}{\sqrt{10}}, \\ \|U^t\| &\leq 3\sqrt{m}\|\bar{X}\| \end{aligned}$$

hold at time  $t$ , we can invoke Lemma B.4 in [34] and claim

$$\|U^t(U^t)^\top - Z^*\| \leq 4\|V_{Z^*}(U^t(U^t)^\top - Z^*)\| + \|U^t Q^{t,\perp}(U^t Q^{t,\perp})^\top\|.$$

For further convenience, we define the quantities

$$\begin{aligned} A_1 &\triangleq 50\sqrt{1 + \sqrt{r - r^*}} (r^*)^{1/8} (r - r^*)^{3/8} \|\bar{X}\|^{11/16} m^{11/32} \kappa^{-3/16}, \\ A_2 &\triangleq \|\bar{X}\|^{11/32} m^{11/64} \left(\frac{r^*}{r - r^*}\right)^{1/16}. \end{aligned}$$

**Step I:** In this phase we prove by induction that the following is true:

$$\|U^t\| \leq 3\sqrt{m}\|\bar{X}\| \quad (62a)$$

$$\sigma_{\min}(U^t Q^t) \geq \frac{\sigma_{r^*}(\bar{X})\sqrt{m}}{\sqrt{10}} \quad (62b)$$

$$\|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\| \leq \left(1 - \alpha \frac{\lambda_{r^*}(\bar{Z}^*)}{4}\right)^{t-t_2} \bar{c}_2 \kappa^2 + A_3 \tau^{1/2} \quad (62c)$$

$$\|U^t Q^{t,\perp}\| \leq (1 + \alpha(1/6)\lambda_{r^*}(\bar{Z}^*))^{t-t_2} A_2 \tau^{21/32} \quad (62d)$$

$$\|V_{\bar{Z}^*}^\top (Z^* - U^t (U^t)^\top)\|_F \leq \left(1 - \frac{\alpha}{400} \lambda_{r^*}(\bar{Z}^*)\right)^{t-t_2} A_1 \tau^{21/16} + A_4 \tau^{1/2} \quad (62e)$$

for all  $t \in [t_2, t_3]$  where

$$t_3 \triangleq t_2 + \left\lfloor \frac{4 \log(\sqrt{m}\tau^{-1}(r - r^*)^{-1})}{16 \log(1 + \alpha(1/6)\lambda_{r^*}(\bar{Z}^*))} \right\rfloor.$$

Observe that (62) hold at time  $t = t_2$ . Recall that Lemma 9 requires a bound on  $\|\mathcal{W}\Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t)\|$  to invoke most of its consequences. Consequently we start off by establishing that given that (62) holds at time  $t$ ,  $\|\mathcal{W}\Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t)\|$  is sufficiently small.

As argued prior,

$$\begin{aligned} \|\mathcal{W}\Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t)\| &\leq 4\Delta \|V_{\bar{Z}^*}^\top (Z^* - U^t (U^t)^\top)\|_F + (4\Delta + 1)\sqrt{r - r^*} \|U^t Q^{t,\perp}\|^2 \\ \|\Delta^t\| &\leq 4\|V_{\bar{Z}^*}^\top (Z^* - U^t (U^t)^\top)\|_F + \sqrt{r - r^*} \|U^t Q^{t,\perp}\|^2 \end{aligned}$$

Invoking the induction hypothesis

$$\begin{aligned} \|\mathcal{W}\Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t)\| &\leq \left(4\Delta A_1 + (4\Delta + 1)A_2^2 \tau^{-1/2} m^{1/4}\right) \tau^{21/16} + 4\Delta A_4 \tau^{1/2} \\ \|\Delta^t\| &\leq \left(4\Delta + A_2^2 \tau^{-1/2} m^{1/4}\right) \tau^{21/16} + 4A_4 \tau^{1/2}. \end{aligned}$$

Thus, if

$$\left(4\Delta A_1 + (4\Delta + 1)A_2^2 \tau^{-1/2} m^{1/4}\right) \tau^{21/16} + 4\Delta A_4 \tau^{1/2} \leq \frac{m\lambda_{r^*}(\bar{z}^*)}{24 \cdot 800}$$

the condition on the quantity  $\|\mathcal{A}^* \mathcal{A}(\Delta^t) - \mathcal{W}\Delta^t \mathcal{W}\|$  holds for all statements that can be invoked in Lemma 9.

Consequently, we can invoke Lemma 9(i) yielding

$$\|U^{t+1}\| \leq 3\sqrt{m}\|\bar{X}\|.$$

To invoke Lemma 9(ii) we require that  $\|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\| \leq \frac{\kappa^{-1}}{110}$  which is fulfilled by the induction hypothesis as long as

$$\bar{c}_2 \kappa^{-2} + A_3 \tau^{1/2} \leq \frac{\kappa^{-2}}{110}.$$

If the above holds, we may invoke Lemma 9(ii) and following the same argument as in Phase II we obtain

$$\sigma_{\min}(U^{t+1} Q^{t+1}) \geq \frac{\sigma_{r^*}(\bar{X})\sqrt{m}}{\sqrt{10}}.$$

Next, our goal is to invoke Lemma 9(iv). For this we require that

$$(1 + \alpha(1/6)\lambda_{r^*}(\bar{Z}^*))^{t-t_2} A_2 \tau^{21/32} \leq \frac{\sqrt{m}\|\bar{X}\|}{\kappa^2 24 \cdot 800}.$$



Then, we may invoke Lemma 9(iv) yielding

$$\begin{aligned}
\|(V_{\bar{Z}^*}^\perp)^\top V_{U^{t+1}Q^{t+1}}\| &\leq \left(1 - \alpha \frac{\lambda_{r^*}(\bar{Z}^*)}{4}\right)^{t-t_2} \|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\| \\
&\quad + \frac{100\alpha}{m} \|\mathcal{W}\Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t)\| + \frac{500\alpha^2}{m^2} \|\Delta^t\|^2 \\
&\leq \left(1 - \alpha \frac{\lambda_{r^*}(\bar{Z}^*)}{4}\right)^{t-t_2} \|(V_{\bar{Z}^*}^\perp)^\top V_{U^{t+1}Q^{t+1}}\| \\
&\quad + \frac{100\alpha}{m} \left( (4\Delta A_1 + (4\Delta + 1)A_2^2 \tau^{-1/2} m^{1/4}) \tau^{21/16} + 4\Delta A_4 \tau^{1/2} \right) \\
&\quad + \frac{\alpha^2 500}{m^2} \left( \left( (4\Delta + A_2^2 \tau^{-1/2} m^{1/4}) \tau^{21/16} + 4A_4 \tau^{1/2} \right)^2 \right).
\end{aligned}$$

Applying the induction hypothesis yields

$$\begin{aligned}
\|(V_{\bar{Z}^*}^\perp)^\top V_{U^{t+1}Q^{t+1}}\| &\leq \left(1 - \alpha \frac{\lambda_{r^*}(\bar{Z}^*)}{4}\right)^{t+1-t_2} \bar{c}_2 \kappa^{-2} + \left(1 - \alpha \frac{\lambda_{r^*}(\bar{Z}^*)}{4}\right) A_3 \tau^{1/2} \\
&\quad + \frac{100\alpha}{m} \left( (4\Delta A_1 + (4\Delta + 1)A_2^2 \tau^{-1/2} m^{1/4}) \tau^{21/16} + 4\Delta A_4 \tau^{1/2} \right) \\
&\quad + \frac{\alpha^2 500}{m^2} \left( \left( (4\Delta + A_2^2 \tau^{-1/2} m^{1/4}) \tau^{21/16} + 4A_4 \tau^{1/2} \right)^2 \right),
\end{aligned}$$

and thus for the bound to hold for  $t + 1$  we require that

$$\begin{aligned}
&\quad + \frac{100}{m} \left( (4\Delta A_1 + (4\Delta + 1)A_2^2 \tau^{-1/2} m^{1/4}) \tau^{21/16} + 4\Delta A_4 \tau^{1/2} \right) \\
&\quad + \frac{\alpha 500}{m^2} \left( \left( (4\Delta + A_2^2 \tau^{-1/2} m^{1/4}) \tau^{21/16} + 4A_4 \tau^{1/2} \right)^2 \right) \leq \frac{\lambda_{r^*}(\bar{Z}^*)}{4} A_3 \tau^{1/2}.
\end{aligned}$$

All conditions to invoke Lemma 9(iii) hold and thus

$$\begin{aligned}
\|U^{t+1}Q^{t+1,\perp}\| &\leq \\
&\left(1 - \frac{\alpha}{2m} \|U^t Q^{t,\perp}\|^2 + 9\alpha \|\bar{Z}^*\| \|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\| + 2\frac{\alpha}{m} \|\mathcal{W}\Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t)\| \right) \|U^t Q^{t,\perp}\|.
\end{aligned}$$

Observe that we have established conditions for which  $\|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\| \leq \frac{\kappa^{-2}}{110}$  and consequently

$$\begin{aligned}
\|U^{t+1}Q^{t+1,\perp}\| &\leq \\
&\left(1 + 9\alpha \|\bar{Z}^*\| \frac{\kappa^{-2}}{110} + 2\frac{\alpha}{m} \left( (4\Delta A_1 + (4\Delta + 1)A_2^2 \tau^{-1/2} m^{1/4}) \tau^{21/16} + 4\Delta A_4 \tau^{1/2} \right) \right) \|U^t Q^{t,\perp}\|,
\end{aligned}$$

where we apply the induction hypothesis to yield

$$\begin{aligned}
\|U^{t+1}Q^{t+1,\perp}\| &\leq (1 + \alpha 80 \lambda_{r^*}(\bar{Z}^*))^{t-t_2} A_2 \tau^{21/32} \times \\
&\quad \times \left(1 + 9\alpha \|\bar{Z}^*\| \frac{\kappa^{-2}}{110} + 2\frac{\alpha}{m} \left( (4\Delta A_1 + (4\Delta + 1)A_2^2 \tau^{-1/2} m^{1/4}) \tau^{21/16} + 4\Delta A_4 \tau^{1/2} \right) \right)
\end{aligned}$$

where we request that

$$\frac{9}{110} \kappa^{-2} \|\bar{Z}^*\| + \frac{2}{m} \left( (4\Delta A_1 + (4\Delta + 1)A_2^2 \tau^{-1/2} m^{1/4}) \tau^{21/16} + 4\Delta A_4 \tau^{1/2} \right) \leq 1/6 \lambda_{r^*}(\bar{Z}^*)$$

to establish that the statement holds for  $t + 1$ . Finally, we have left to bound (62e). For this, we can invoke Lemma 9(v) yielding

$$\begin{aligned}
\|V_{\bar{Z}^*}^\top (\Delta^{t+1})\|_F &\leq \left(1 - \frac{\alpha}{200} \lambda_{r^*}(\bar{Z}^*) + 9\alpha \|\bar{Z}^*\| \rho\right) \|V_{\bar{Z}^*}^\top \Delta^t\| + \frac{\alpha \lambda_{r^*}(\bar{Z}^*)}{100} \sqrt{r - r^*} \|U^t Q^{t,\perp}\| \\
&\leq \left(1 - \frac{\alpha \lambda_{r^*}(\bar{Z}^*)}{200} + 9\alpha \|\bar{Z}^*\| \rho\right) \|V_{\bar{Z}^*}^\top \Delta^t\|_F + \frac{\alpha \lambda_{r^*}(\bar{Z}^*)}{100} A_2^2 \tau^{21/16} \tau^{-1/2} m^{1/4}.
\end{aligned}$$

Invoking the induction hypothesis yields

$$\begin{aligned} \|V_{Z^*}^\top \Delta^{t+1}\|_F &\leq \left(1 - \frac{\alpha \lambda_{r^*}(\bar{Z})}{200}\right)^{t-t_2} \left(1 - \frac{\alpha \lambda_{r^*}(\bar{Z}^*)}{200} + 9\alpha \|\bar{Z}^*\|_\rho\right) A_1 \tau^{21/16} \\ &+ \left(1 - \frac{\alpha}{200} \lambda_{r^*}(\bar{Z}^*) + 9\alpha \|\bar{Z}^*\|_\rho\right) A_4 \tau^{1/2} + \frac{\alpha \lambda_{r^*}(\bar{Z}^*)}{100} A_2^2 \tau^{21/16} \tau^{-1/2} m^{1/4}. \end{aligned}$$

Thus, by requesting

$$\begin{aligned} \left(\frac{1}{200} \lambda_{r^*}(\bar{Z}^*) - 9\|\bar{Z}^*\|_\rho\right) A_4 \tau^{1/2} &\geq \frac{\lambda_{r^*}(\bar{Z}^*)}{100} A_2^2 \tau^{21/16} \tau^{-1/2} m^{1/4} \\ \rho &\leq \frac{\lambda_{r^*}(\bar{Z}^*)}{3600 \|\bar{Z}^*\|} \end{aligned}$$

yields the desired result, and consequently the proof of this part is concluded, by setting  $A_4 = 4A_2^2 \tau^{5/16} m^{1/4}$ , and  $A_3 = \frac{\bar{c}_2}{2m^{1/4}} \kappa^{-2}$ , with  $\bar{c}_2 \leq \frac{1}{220}$ , with the assumption that  $\tau < 1$ . A sufficient condition for all conditions to hold is given by

$$\begin{aligned} \rho &\leq \frac{1}{3600 \kappa^{-2}} \\ \left(4\Delta A_1 \tau^{1/2} + A_2^2 m^{1/4} (20\Delta + 1)\right) \tau^{6/16} &+ \frac{5}{m} \alpha \left(4A_1 \tau^{1/2} + 17m^{1/2} A_2^2\right)^2 \tau^{9/8} \leq \frac{\lambda_{r^*}(\bar{Z}^*) m}{\kappa^2 8 \cdot 10^4}. \end{aligned}$$

**Step II:** We summarize what we obtain from Step I when  $t = t_3$ .

$$\begin{aligned} \|U^{t_3}\| &\leq 3\sqrt{m} \|\bar{X}\| \\ \sigma_{\min}(U^{t_3} Q^{t_3}) &\geq \frac{\sigma_{r^*}(\bar{X}) \sqrt{m}}{\sqrt{10}} \\ \|(V_{Z^*}^\perp)^\top V_{U^{t_3} Q^{t_3}}\| &\leq \frac{1}{1 - \alpha \frac{\lambda_{r^*}(\bar{Z}^*)}{4}} \tau^{1/2} m^{-1/4} \sqrt{r - r^*} + \frac{\bar{c}_2}{2} \tau^{1/2} m^{-1/4} \\ \|U^{t_3} Q^{t_3, \perp}\| &\leq \tau^{-1/4} m^{1/8} (r - r^*)^{-1/4} A_2 \tau^{21/32} \\ \|V_{Z^*}^\top (\Delta^{t_3})\| &\leq A_1 \tau^{21/16} + 4A_2^2 \tau^{13/16} \end{aligned}$$

From here we establish by induction that

$$\begin{aligned} \|U^t\| &\leq 3\sqrt{m} \|\bar{X}\| \\ \sigma_{\min}(U^t Q^t) &\geq \frac{\sigma_{r^*}(\bar{X}) \sqrt{m}}{\sqrt{10}} \\ \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| &\leq \left(1 - \frac{\alpha}{4} \lambda_{r^*}(\bar{Z}^*)\right)^{t-t_3} \left(\frac{\tau^{1/2} \sqrt{r - r^*}}{1 - \alpha \frac{\lambda_{r^*}(\bar{Z}^*)}{4}} + \frac{\bar{c}_2}{2} \tau^{1/2}\right) m^{-1/4} + C_1 \tau^{1/8} \\ \|U^t Q^{t, \perp}\| &\leq \left(1 + C_2 \left(\frac{\tau}{m^{1/2}}\right)^{1/8}\right)^{t-t_3} \frac{A_2 m^{1/8}}{(r - r^*)^{1/4}} \tau^{13/32} \\ \|V_{Z^*}^\top (\Delta^t)\|_F &\leq \left(1 - \frac{\alpha \lambda_{r^*}(\bar{Z}^*)}{400}\right)^{t-t_3} \left(A_1 \tau^{21/16} + 4A_2^2 \tau^{13/16}\right) + C_3 \tau^{1/8}, \end{aligned}$$

for all  $t \in [t_3, t_4]$

$$t_4 \triangleq t_3 + \left\lfloor \frac{\frac{1}{4} \log\left(\frac{\sqrt{m}}{\tau}\right)}{16 \log\left(1 + C_2 \left(\frac{\tau}{\sqrt{m}}\right)^{1/8}\right)} \right\rfloor.$$

We start off in a similar way as in the previous phase. By induction hypothesis we can invoke Lemma B.4 in [34] and thus

$$\begin{aligned} \|\mathcal{W} \Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t)\| &\leq 4\Delta \|V_{Z^*}^\top (\Delta^t)\|_F + (4\Delta + 1) \sqrt{r - r^*} \|U^t Q^{t, \perp}\|^2 \\ \|\Delta^t\| &\leq \|V_{Z^*}^\top \Delta^t\|_F + \sqrt{r - r^*} \|U^t Q^{t, \perp}\|^2, \end{aligned}$$

whereby induction hypothesis it follows that

$$\begin{aligned} \|\mathcal{W}\Delta^t\mathcal{W} - \mathcal{A}^*\mathcal{A}\| &\leq 4\Delta \left( A_1\tau^{21/16} + 4A_2^2\tau^{13/16} + C_3\tau^{1/8} \right) \\ &\quad + (4\Delta + 1)A_2^2\tau^{13/16} \left( 1 + C_2 \left( \frac{\tau}{\sqrt{m}} \right)^{1/8} \right)^{2(t_4-t_3)} m^{1/4} \\ \|\Delta^t\| &\leq \left( A_1\tau^{21/16} + 4A_2^2\tau^{13/16} + C_3\tau^{1/8} \right) + A_2^2\tau^{13/16} \left( 1 + C_2 \left( \frac{\tau}{\sqrt{m}} \right)^{1/8} \right)^{2(t_4-t_3)} m^{1/4}, \end{aligned}$$

whereby using the expression for  $t_4$  is that

$$\begin{aligned} \|\mathcal{W}\Delta^t\mathcal{W} - \mathcal{A}^*\mathcal{A}\| &\leq 4\Delta \left( A_1\tau^{21/16} + 4A_2^2\tau^{13/16} + C_3\tau^{1/8} \right) + (4\Delta + 1)A_2^2\tau^{25/32} (m)^{1/64} m^{1/4} \\ \|\Delta^t\| &\leq \left( A_1\tau^{21/16} + 4A_2^2\tau^{13/16} + C_3\tau^{1/8} \right) + A_2^2\tau^{25/32} m^{1/64} m^{1/4}. \end{aligned}$$

A requirement for us to invoke elements of Lemma 9 is

$$4\Delta \left( A_1\tau^{21/16} + 4A_2^2\tau^{13/16} + C_3\tau^{1/8} \right) + (4\Delta + 1)A_2^2\tau^{25/32} m^{1/64} m^{1/4} \leq \frac{m\lambda_{r^*}(\bar{Z}^*)}{24 \cdot 800}.$$

We follow the same steps as in the previous part and invoke Lemma 9(i) yielding

$$\|U^{t+1}\| \leq 3\sqrt{m}\|\bar{X}\|.$$

To invoke Lemma 9(ii) we request that  $\|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\| \leq \frac{\kappa^{-2}}{110}$  which is fulfilled by induction hypothesis as long as

$$\frac{\bar{c}_2}{2}\tau^{1/2}m^{-1/4} + \left( \frac{\sqrt{r-r^*}}{1 - \alpha \frac{\lambda_{r^*}(\bar{Z}^*)}{4}} \right) \tau^{1/2}m^{-1/4} + C_1\tau^{1/8} \leq \frac{\kappa^{-1}}{100}$$

from which we obtain

$$\sigma_{\min}(U^{t+1}Q^t) \geq \frac{\sigma_{r^*}(\bar{X})\sqrt{m}}{\sqrt{10}},$$

following the same procedure as in Phase II. Similar as in Step I, to invoke Lemma 9(iv) we require that

$$\frac{A_2}{(r-r^*)^{1/4}}\tau^{12/32} \leq \frac{\kappa^{-2}m\|\bar{X}\|}{24 \cdot 800}$$

yielding

$$\begin{aligned} \|(V_{\bar{Z}^*}^\perp)^\top V_{U^{t+1}Q^{t+1}}\| &\leq \left( 1 - \frac{\alpha}{4}\lambda_{r^*}(\bar{Z}^*) \right) \|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\| \\ &\quad + \frac{100}{m}\alpha\|\mathcal{W}\Delta^t\mathcal{W} - \mathcal{A}^*\mathcal{A}(\Delta^t)\| + \frac{500\alpha^2}{m^2}\|\Delta^t\|^2. \end{aligned}$$

By induction hypothesis we have

$$\begin{aligned} \|(V_{\bar{Z}^*}^\perp)^\top V_{U^{t+1}Q^{t+1}}\| &\leq \left( 1 - \frac{\alpha}{4}\lambda_{r^*}(\bar{Z}^*) \right)^{t+1-t_3} \left( \frac{\tau^{1/2}\sqrt{r-r^*}}{1 - \alpha \frac{\lambda_{r^*}(\bar{Z}^*)}{4}} + \frac{\bar{c}_2}{2}\tau^{1/2} \right) m^{-1/4} + C_1\tau^{1/8} \\ &\quad - \frac{\alpha}{4}\lambda_{r^*}(\bar{Z}^*)C_1\tau^{1/8} + \frac{500\alpha^2}{m^2} \left( A_1\tau^{21/16} + 4A_2^2\tau^{13/16} + C_3\tau^{1/8} + A_2^2\tau^{25/32}m^{1/64}m^{1/4} \right)^2 \\ &\quad + \alpha \frac{100}{m} \left( 4\Delta \left( A_1\tau^{21/16} + 4A_2^2\tau^{13/16} + C_3\tau^{1/8} \right) + (4\Delta + 1)A_2^2\tau^{25/32}m^{1/64}m^{1/4} \right). \end{aligned}$$

Consequently, for the statement to hold for  $t+1$  it must hold that

$$\begin{aligned} \frac{\lambda_{r^*}(\bar{Z}^*)}{4}C_1\tau^{1/8} &\geq \frac{500\alpha}{m^2} \left( A_1\tau^{21/16} + 4A_2^2\tau^{13/16} + C_3\tau^{1/8} + A_2^2\tau^{25/32}m^{1/64}m^{1/4} \right)^2 \\ &\quad + \frac{100}{m} \left( 4\Delta \left( A_1\tau^{21/16} + 4A_2^2\tau^{13/16} + C_3\tau^{1/8} \right) + (4\Delta + 1)A_2^2\tau^{25/32}m^{1/64}m^{1/4} \right). \end{aligned}$$

At this state, all conditions hold to invoke Lemma 9(iii) yielding

$$\|U^{t+1}Q^{t+1,\perp}\| \leq \left(1 + \frac{2\alpha}{m}\|\mathcal{W}\Delta^{t+1}\mathcal{W} - \mathcal{A}^*\mathcal{A}^*(\Delta^t)\| + 9\alpha\|\bar{Z}^*\| \|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\|\right) \|U^t Q^{t,\perp}\|,$$

whereby we request that

$$\begin{aligned} & \frac{2\alpha}{m} \left(4\Delta \left(A_1\tau^{21/16} + 4A_2^2\tau^{13/16} + C_3\tau^{1/8}\right) + (4\Delta + 1)A_2^2\tau^{25/32}M^{1/64}m^{1/4}\right) \\ & + 9\alpha\|\bar{Z}^*\| \left(\frac{\tau^{1/2}\sqrt{r-r^*}}{1 - \alpha\frac{\lambda_{r^*}(\bar{Z})}{4}}m^{-1/4} + \frac{\bar{c}_2}{2}\tau^{1/2}m^{-1/4} + C_1\tau^{1/8}\right) \leq C_2 \left(\frac{\tau}{\sqrt{m}}\right)^{1/8} \end{aligned}$$

to yield

$$\|U^{t+1}Q^{t+1,\perp}\| \leq \left(1 + C_2 \left(\frac{\tau}{\sqrt{m}}\right)^{1/8}\right) \|U^t Q^{t,\perp}\|,$$

and thus invoking the induction hypothesis we obtain the desired result. Finally, we can invoke Lemma 9(v) to yield

$$\begin{aligned} & \|V_{\bar{Z}^*}^\top \Delta^{t+1}\|_F \leq \left(1 - \frac{\alpha\lambda_{r^*}(\bar{Z}^*)}{200} + \alpha 9\|\bar{Z}^*\|\rho\right) \|V_{\bar{Z}^*}^\top \Delta^t\|_F + \frac{\alpha\lambda_{r^*}(\bar{Z}^*)}{100}\sqrt{r-r^*}\|U^t Q^{t,\perp}\|^2 \\ & \leq \left(1 - \frac{\alpha\lambda_{r^*}(\bar{Z}^*)}{200} + \alpha 9\|\bar{Z}^*\|\rho\right) \|V_{\bar{Z}^*}^\top \Delta^t\|_F + \frac{\alpha\lambda_{r^*}(\bar{Z}^*)}{100}m^{1/64}m^{1/4}A_2^2\tau^{25/32} \\ & \leq \left(1 - \frac{\alpha\lambda_{r^*}(\bar{Z}^*)}{400}\right)^{t-t_3} \left(1 - \frac{\alpha}{200}\lambda_{r^*}(\bar{Z}^*) + \alpha 9\|\bar{Z}^*\|\rho\right) \left(A_1\tau^{21/16} + 4A_2^2\tau^{13/16}\right) \\ & + C_3\tau^{1/8} - \left(\frac{\alpha}{200}\lambda_{r^*}(\bar{Z}^*) - 9\alpha\rho\|\bar{Z}^*\|\right) C_3\tau^{1/8} + \frac{\alpha\lambda_{r^*}(\bar{Z}^*)}{100}m^{1/64}m^{1/4}A_2^2\tau^{25/32}. \end{aligned}$$

Thus, requesting that

$$\rho \leq \frac{1}{3600\kappa^2} \quad \frac{C_3}{4}\tau^{1/8} \geq m^{1/64}m^{1/4}A_2^2\tau^{25/32},$$

and the proof by induction follows with adequate choices of  $C_1$ ,  $C_2$  and  $C_3$  which we provide shortly. For convenience define

$$\begin{aligned} B & \triangleq 4(\Delta + 1) \left( A_1 \left( \frac{\tau}{\sqrt{m}} \right)^{21/16} m^{21/32} + 4A_2^2 \left( \frac{\tau}{\sqrt{m}} \right)^{13/16} m^{13/32} \right. \\ & \left. + C_3 + A_2^2 \left( \frac{\tau}{\sqrt{m}} \right)^{25/32} m^{25/62} m^{17/64} \right) \end{aligned}$$

Observe that we may choose

$$\begin{aligned} C_3 & = 4m^{17/64}A_2^2\tau^{21/32} \\ C_1\tau^{1/8} & \leq \frac{\kappa^{-1}}{100} - \frac{\bar{c}_2}{2}\tau^{1/2}m^{-1/4} - \left(\frac{\sqrt{r-r^*}}{1 - \alpha\frac{\lambda_{r^*}(\bar{Z}^*)}{4}}\right)\tau^{1/2}m^{-1/4} \\ C_1\frac{\lambda_{r^*}(\bar{Z}^*)}{4}\tau^{1/8} & \geq \frac{500\alpha}{m}B^2 + \frac{100}{m}B \\ C_2\left(\frac{\tau}{\sqrt{m}}\right)^{1/8} & \geq \frac{2\alpha}{m}(4\Delta + 1)B + 9\alpha\|\bar{Z}^*\| \left(\frac{\sqrt{r-r^*}}{1 - \alpha\frac{\lambda_{r^*}(\bar{Z}^*)}{4}} + \frac{\bar{c}_2}{2}\right)\tau^{1/2} + \alpha 9\|\bar{Z}^*\|C_1\tau^{1/2}. \end{aligned}$$

Under the assumption that

$$\frac{500\alpha}{m}B \leq 1,$$

we set

$$C_1 = \frac{800}{m^{1/4} \lambda_{r^*}(\bar{Z}^*)},$$

and

$$C_2 = \alpha \left( \frac{8(\Delta + 1)}{m} B \left( \frac{\tau}{\sqrt{m}} \right)^{-1/8} + 9 \|\bar{Z}^*\| \left( \frac{\sqrt{r - r^*} \bar{c}_2}{1 - \alpha \frac{\lambda_{r^*}(\bar{Z}^*)}{4}} \right) (\tau/\sqrt{m})^{3/8} + 9 \|\bar{Z}^*\| \frac{800}{\lambda_{r^*}(\bar{Z}^*)} (\tau/\sqrt{m})^{3/8} \right).$$

For some universal constant

$$\begin{aligned} C_2 &\leq \alpha c \|\bar{Z}^*\| \sqrt{r - r^*} (\tau/\sqrt{m})^{3/8} + \alpha c \kappa (\tau/\sqrt{m})^{3/8} \\ &\quad + c \alpha \left( (r - r^*)^{5/8} (r^*)^{1/8} \|\bar{X}\|^{11/16} \left( \frac{\tau}{\sqrt{m}} \right)^{19/16} \right. \\ &\quad \left. + \left( \frac{r^*}{4 - r^*} \right)^{1/8} \|\bar{X}\|^{11/16} \left( \left( \frac{\tau}{\sqrt{m}} \right)^{11/16} + \left( \frac{\tau}{\sqrt{m}} \right)^{21/32} \right) \right) \end{aligned}$$

and consequently the window is of length at least

$$\left\lfloor \frac{1}{64} \left( \frac{1 - \sqrt{m}}{\tau} \right)^{1/8} \right\rfloor$$

Observe that throughout both parts

$$\begin{aligned} \|\Delta^t\|_F &\leq c_3 \left( A_1 \tau^{21/16} + 4A_2^2 \tau^{13/16} + 4A_2^2 \tau^{1/4} \tau^{1/8} + A_2^2 \tau^{25/32} m^{1/64} m^{1/4} \right) \\ &\leq c_3 (r - r^*)^{5/8} (r^*)^{1/8} \|\bar{X}\|^{11/16} m \left( \frac{\tau}{\sqrt{m}} \right)^{21/16} + c_3 \|\bar{X}\|^{22/32} \left( \frac{r^*}{r - r^*} \right)^{1/8} \times \\ &\quad \times \left( m^{24/32} \left( \frac{\tau}{\sqrt{m}} \right)^{13/16} + \left( \frac{\tau}{\sqrt{m}} \right)^{25/32} m^{47/64} m^{17/64} \right) \end{aligned}$$

Consequently,

$$\begin{aligned} \frac{\|\Delta^t\|_F}{m \|\bar{Z}^*\|} &\leq c_3 (r - r^*)^{5/8} (r^*)^{1/8} \|\bar{X}\|^{-21/16} \left( \frac{\tau}{\sqrt{m}} \right)^{21/16} \\ &\quad + c_3 \|\bar{X}\|^{-21/16} \left( \frac{r^*}{r - r^*} \right)^{1/8} \left( \frac{\tau}{\sqrt{m}} \right)^{13/16}. \end{aligned}$$

## E Proofs of the Intermediate Results in Phase I (Appendix [B.1](#))

### E.1 Proof of Lemma [5](#)

Let

$$\tilde{E}^i \triangleq \frac{\alpha}{m} \mathcal{A}^* \mathcal{A} \left( U^{i-1} (U^{i-1})^\top \right) U^{i-1}.$$

Then, it can be established by induction that for  $t \geq 1$

$$\tilde{U}^t - U^t = -E^t = \sum_{i=1}^t \left( \mathcal{W}^2 + \frac{\alpha}{m} \mathcal{A}^* \mathcal{A}(\bar{Z}^*) \right)^{t-i} \tilde{E}^i$$

Invoking Lemma [3](#) there holds that

$$\begin{aligned} \|\tilde{E}^i\| &\leq \frac{\alpha}{m} \left( 1 + 2\delta_4 + 2 \left( \rho + (1 + 2\delta_4) \hat{\Delta}_4 \right) \right) \|U^{i-1} (U^{i-1})^\top\|_* \|U^{i-1}\| \\ &\leq \frac{\alpha}{m} \left( 1 + 2\delta_4 + 2 \left( \rho + (1 + 2\delta_4) \hat{\Delta}_4 \right) \right) \|U^{i-1}\|_F^2 \|U^{i-1}\|. \end{aligned}$$

Further, observe that

$$\left\| \left( \mathcal{W}^2 + \frac{\alpha}{m} \mathcal{A}^* \mathcal{A}(Z^*) \right)^{t-i} \tilde{E}^i \right\| = \left\| \left( \mathcal{W}^2 + \frac{\alpha}{m} Z^* + \frac{\alpha}{m} (\mathcal{A}^* \mathcal{A}(Z) - \mathcal{W} Z^* \mathcal{W}) \right) \right\|^{t-i} \|\tilde{E}^i\|.$$

Invoking Weyl's inequality and Lemma [1](#) we have

$$\lambda_1 \left( \mathcal{W}^2 + \frac{\alpha}{m} \mathcal{A}^* \mathcal{A}(Z^*) \right) \leq 1 + \alpha \lambda_1(\bar{Z}^*) + 2\sqrt{r^*} \alpha \delta_{2(r^*+1)} \lambda_1(\bar{Z}^*).$$

Consequently we have

$$\begin{aligned} \|E^t\| &\leq \sum_{i=1}^t (1 + \alpha(1 + 2\delta_{2(r^*+1)}) \lambda_1(\bar{Z}^*))^{t-i} \times \\ &\quad \times \frac{\alpha}{m} \left( 1 + 2\delta_4 + 2 \left( \rho + (1 + 2\delta_4) \hat{\Delta}_4 \right) \right) \|U^{i-1}\|_F^2 \|U^{i-1}\|. \end{aligned}$$

We upper bound

$$\begin{aligned} \|U^{i-1}\|_F^2 \|U^{i-1}\| &\leq \min\{r, md\} \|U^{i-1}\|^3 \leq \min\{r, md\} \left( \|\tilde{U}^{i-1}\| + \|\tilde{U}^{i-1} - U^{i-1}\| \right)^3 \\ &\leq 8 \min\{r, md\} \|\tilde{U}^{i-1}\|^3 \leq 8 \min\{r, md\} \left( 1 + \alpha(1 + 2\sqrt{r^*} \delta_{2(r^*+1)}) \lambda_1(\bar{Z}^*) \right)^{3(i-2)} \|\tilde{U}^1\|^3. \end{aligned}$$

Observe that

$$\begin{aligned} \|\tilde{U}^1\| &\leq \|\mathcal{J}U^0 + \frac{\alpha}{m} \mathcal{A}^* \mathcal{A}(Z^*) \mathcal{J}U^0\| + \|(\mathcal{W}^2 - \mathcal{J})U^0 + \frac{\alpha}{m} \mathcal{A}^* \mathcal{A}(Z^*) (\mathcal{I} - \mathcal{J})U^0\| \\ &\stackrel{(a)}{\leq} \left( 1 + \alpha(1 + 2\sqrt{r^*} \delta_{2(r^*+1)}) \lambda_1(\bar{Z}^*) \right) \|\mathcal{J}U^0\| + \left( \rho^2 + \alpha \rho^{1/2} (1 + \delta_{2r^*} \sqrt{r^*}) \lambda_1(\bar{Z}^*) \right) \|(\mathcal{I} - \mathcal{J})U^0\| \\ &\leq \left( 1 + \alpha(1 + 2\sqrt{r^*} \delta_{2(r^*+1)}) \lambda_1(\bar{Z}^*) \right) \left( \|\mathcal{J}U^0\|^2 + \rho^{1/2} \sqrt{r^*} \|(\mathcal{I} - \mathcal{J})U^0\|^2 \right) \end{aligned}$$

where in (a) we invoke Lemmas [1](#) and [4](#). Thus,

$$\begin{aligned} \|U^{i-1}\|_F^2 \|U^{i-1}\| &\leq 8 \min\{r, md\} \left( 1 + \alpha(1 + 2\sqrt{r^*} \delta_{2(r^*+1)}) \lambda_1(\bar{Z}^*) \right)^{3(i-1)} \times \\ &\quad \times \left( \|\mathcal{J}U^0\| + \rho^{1/2} \sqrt{r^*} \|(\mathcal{I} - \mathcal{J})U^0\| \right). \end{aligned}$$

Combining we obtain

$$\begin{aligned} \|E^t\| &\leq \sum_{i=1}^t \left( 1 + \alpha(1 + 2\sqrt{r^*} \delta_{2(r^*+1)}) \lambda_1(\bar{Z}^*) \right)^{t+2i-3} \frac{\alpha}{m} \left( 1 + 2\delta_4 + 2 \left( \rho + (1 + 2\delta_4) \hat{\Delta}_4 \right) \right) \times \\ &\quad \times 8 \min\{r, md\} \left( \|\mathcal{J}U^0\|^2 + \rho^{1/2} \sqrt{r^*} \|(\mathcal{I} - \mathcal{J})U^0\| \right)^3 \\ &\leq \frac{4m^{-1}}{(1 + 2\sqrt{r^*} \delta_{2(r^*+1)}) \lambda_1(\bar{Z}^*)} \mu^3 \min\{r, md\} \left( 1 + 2\delta_4 + 2 \left( \rho + (1 + 2\delta_4) \hat{\Delta}_4 \right) \right) \times \\ &\quad \times \left( 1 + \alpha(1 + 2\sqrt{r^*} \delta_{2(r^*+1)}) \lambda_1(\bar{Z}^*) \right)^{3t} \left( \|\mathcal{J}U^0\| + \rho^{1/2} \sqrt{r^*} \|(\mathcal{I} - \mathcal{J})U^0\| \right)^3. \end{aligned}$$

## E.2 Proof of Lemma [6](#)

Denote by  $v_1$  the eigenvector corresponding to the largest eigenvalue of  $\mathcal{W}^2 + \frac{\alpha}{m} \mathcal{A}^* \mathcal{A}(Z^*)$ . Then,

$$(\tilde{U}^t)^\top v_1 \geq (U^0)^\top \left( \lambda_1 \left( \mathcal{W}^2 + \frac{\alpha}{m} \mathcal{A}^* \mathcal{A}(Z^*) \right) \right)^t v_1.$$

Invoking Weyl's inequality and Lemma [2](#) yields

$$\begin{aligned} \lambda_1 \left( \mathcal{W}^2 + \frac{\alpha}{m} \mathcal{A}^* \mathcal{A}(Z^*) \right) &\geq \lambda_1 \left( \mathcal{W}^2 + \frac{\alpha}{m} Z^* \right) - \frac{\alpha}{m} \|\mathcal{W} Z^* \mathcal{W} - \mathcal{A}^* \mathcal{A}(Z^*)\| \\ &\geq 1 + \alpha(1 - 2\sqrt{r^*} \delta_{2(r^*+1)}) \lambda_1(\bar{Z}^*). \end{aligned}$$

Consequently, combined with Lemma 5 we obtain

$$\begin{aligned} \frac{\|E^t\|}{\|\tilde{U}^t\|} &\leq \frac{4m^{-1}}{\lambda_1(\bar{Z}^*)(1+2\sqrt{r^*}\delta_{2(r^*+1)})} \mu^3 \min\{r, md\} \left(1+2\delta_4+2\left(\rho+(1+2\delta_4)\hat{\Delta}_4\right)\right) \times \\ &\times \frac{\left(1+\alpha\lambda_1(\bar{Z}^*)(1+2\sqrt{r^*}\delta_{2(r^*+1)})\right)^{3t} \left(\|\mathcal{J}U^0\|+\rho^{1/2}\sqrt{r^*}\|(\mathcal{I}-\mathcal{J})U^0\|\right)^3}{\left(1+\alpha\lambda_1(\bar{Z}^*)(1-2\sqrt{r^*}\delta_{2(r^*+1)})\right)^t \| (U^0)^\top v_1 \|.} \end{aligned}$$

Therefore,

$$t^* \geq \left\lceil \frac{\ln \left( \frac{(1+2\sqrt{r^*}\delta_{2(r^*+1)})\lambda_1(\bar{Z}^*)}{4(1+2\delta_4+2(\rho+(1+2\delta_4)\hat{\Delta}_4))\mu^2 \frac{(\|\mathcal{J}U\|+\rho^{1/2}\sqrt{r^*}\|(\mathcal{I}-\mathcal{J})U\|)^3}{m\|U^\top v_1\|}} \right)}{\ln \left( \frac{(1+\alpha\lambda_1(\bar{X}^*)(1+2\sqrt{r^*}\delta_{r^*+1}))^3}{1+\alpha\lambda_1(\bar{X}^*)(1-2\sqrt{r^*}\delta_{2(r^*+1)})} \right)} \right\rceil.$$

### E.3 Proof of Lemma 7

The condition on  $\gamma$  implies, invoking Lemma 10 because  $\gamma \leq \frac{\kappa^{-2}}{17}$  that

$$\sigma_{r^*}(M^t U^0 + E^t) \geq (1 - \delta_\rho) \mu \sigma_{r^*}(M^t) \sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U) - \|E^t\|,$$

and

$$\sigma_{r^*}(M^t U^0 + E^t) \geq \frac{1 - \delta_\rho}{2} \mu \sigma_{r^*}(M^t) \sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U).$$

Further,

$$\|(V_{\mathcal{J}M\mathcal{J}}^\perp)^\top V_{U^t}\| \leq \frac{\gamma}{(1 - \delta_\rho)(1 - \gamma)}$$

and consequently because

$$\|(V_{\bar{Z}^*}^\perp)^\top V_{U^t}\| \leq \|(V_{\bar{Z}^*}^\perp)^\top V_{\mathcal{J}M\mathcal{J}}\| + \|(V_{\mathcal{J}M\mathcal{J}}^\perp)^\top V_{U^t}\|.$$

To bound  $\|(V_{\bar{Z}^*}^\perp)^\top V_{\mathcal{J}M\mathcal{J}}\|$  we invoke the Davis-Kahan  $\sin(\Theta)$  theorem yielding

$$\|(V_{\bar{Z}^*}^\perp)^\top V_{\mathcal{J}M\mathcal{J}}\| \leq \frac{\|Z^* - \mathcal{J}\mathcal{A}^*\mathcal{A}(Z^*)\mathcal{J}\|}{m\lambda_{r^*}(\bar{Z}^*) - \|Z^* - \mathcal{J}\mathcal{A}^*\mathcal{A}(Z^*)\mathcal{J}\|} \stackrel{(a)}{\leq} \frac{\sqrt{r^*}\delta_{r^*}\lambda_1(\bar{Z}^*)}{\lambda_{r^*}(\bar{Z}) - \sqrt{r^*}\delta_{r^*}\lambda_1(\bar{Z}^*)},$$

where in (a) have assumed that  $\lambda_{r^*}(\bar{Z}) > \sqrt{r^*}\lambda_1(\bar{Z}^*)\delta_{r^*}$  and used that  $\bar{\mathcal{A}}$  fulfills the RIP. Therefore, we conclude

$$\|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\| \leq \frac{\sqrt{r^*}\delta_{r^*}\lambda_1(\bar{Z}^*)}{\lambda_{r^*}(\bar{Z}) - \sqrt{r^*}\delta_{r^*}\lambda_1(\bar{Z}^*)} + \frac{\gamma}{(1 - \delta_\rho)(1 - \gamma)}.$$

Further, from Lemma 11 and our assumption on  $\gamma$  it follows that

$$\|U^t Q^{t,\perp}\| \leq \frac{\kappa^{-2}}{8} \mu \sigma_{r^*}(M^t) \sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U).$$

Finally, under our conditions on  $\delta_{r^*}$ ,  $\rho$  and  $\gamma$ , it holds that  $\|(V_{\bar{Z}^*}^\perp)^\top V_{U^t}\| \leq \frac{1}{8}$  implying that we can invoke Lemma 11 which yields

$$\sigma_{\min}(U^t Q^t) \geq \frac{1}{2} \sigma_{r^*}(U^t) \geq \frac{1 - \delta_\rho}{2} \mu \sigma_{r^*}(M^t) \sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U) - \|E^t\|.$$

### E.4 Proof of Lemma 8

Our goal is to find conditions under which for some  $t$  there holds that for  $c < \frac{1}{2}$

$$\gamma_t = \frac{\mu \sigma_{r^*+1}(M^{t-1}) (\sigma_{r^*+1}(M) \|\mathcal{J}U\| + (\rho^2 + \alpha \rho^{1/2} \nu) \|(\mathcal{I} - \mathcal{J})U\|) + \|E^t\|}{(1 - \delta_\rho) \mu \sigma_{r^*}(M^t) \sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U)} \leq c \kappa^{-2} \quad (63)$$

hold. Sufficient conditions for the above to hold are

$$\mu\sigma_{r^{*+1}}(M^{t-1}) \left( \sigma_{r^{*+1}}(M) \|\mathcal{J}U\| + (\rho^2 + \alpha\rho^{1/2}\nu) \|(\mathcal{I} - \mathcal{J})U\| \right) \quad (64)$$

$$\begin{aligned} &\leq \frac{c}{2} \kappa^{-2} (1 - \delta_\rho) \mu \sigma_{r^*}(M^t) \sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U) \\ \|E^t\| &\leq \frac{c}{2} \kappa^{-2} (1 - \delta_\rho) \mu \sigma_{r^*}(M^t) \sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U). \end{aligned} \quad (65)$$

Using Weyl's inequality we have

$$\sigma_{r^{*+1}}(M) \leq 1 + \frac{\alpha}{m} \|\mathcal{W}Z^*\mathcal{W} - \mathcal{A}^*\mathcal{A}(Z^*)\| \stackrel{(a)}{\leq} 1 + \alpha 2\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*) \quad (66)$$

$$\sigma_{r^*}(M) \stackrel{(a)}{\geq} 1 + \alpha \lambda_{r^*}(\bar{Z}^*) - \alpha 2\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*), \quad (67)$$

where in (a) we have used Lemma 2. Thus, sufficient conditions for (64) and (69) are

$$\mu \left( 1 + \alpha 2\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*) \right)^{t-1} \left( (1 + \alpha 2\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*)) \|\mathcal{J}U\| \right) \quad (68)$$

$$\begin{aligned} &+ \mu \left( 1 + \alpha 2\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*) \right)^{t-1} \left( \rho^2 + \alpha\rho^{1/2}\nu \right) \|(\mathcal{I} - \mathcal{J})U\| \\ &\leq \frac{\tilde{c}_3}{2} \kappa^{-2} (1 - \delta_\rho) \mu \left( 1 + \alpha \lambda_{r^*}(\bar{Z}^*) - \alpha 2\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*) \right)^t \sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U). \end{aligned}$$

$$\|E\|^t \leq \frac{\tilde{c}_3}{2} \kappa^{-2} (1 - \delta_\rho) \mu \left( 1 + \alpha \lambda_{r^*}(\bar{Z}^*) - \alpha 2\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*) \right)^t \sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U). \quad (69)$$

A sufficient condition for (64) is

$$\begin{aligned} &\underbrace{2\kappa^2 \frac{(1 + \alpha 2\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*)) \|\mathcal{J}U\| + (\rho^2 + \alpha\rho^{1/2}\nu) \|(\mathcal{I} - \mathcal{J})U\|}{\tilde{c}_3 \sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U) (1 + \alpha 2\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*))}}_{=\sigma} \\ &\leq \left( \frac{1 + \alpha \lambda_{r^*}(\bar{Z}^*) - \alpha 2\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*)}{1 + \alpha 2\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*)} \right)^t. \end{aligned}$$

Thus, by definition of  $t_*$  the above holds for all  $t \leq t_*$ .

We find next conditions under which (69) hold. Observe that (31) holds for  $t \in [t^*]$ . Aiming at using Lemma 5, we find next conditions under which  $t_* \leq t^*$ . To this end, we ask

$$\begin{aligned} &\frac{1}{8} \ln \left( \frac{(1 + 2\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*)) \lambda_1(\bar{Z}^*)}{4(1 + \eta) \mu^2 \frac{\min\{r, md\} (\|\mathcal{J}U\| + \rho^{1/2} \sqrt{r^*} \|(\mathcal{I} - \mathcal{J})U\|)^3}{m \|U^\top v_1\|}} \right) \geq \ln(\sigma) \times \\ &\times \underbrace{\frac{\ln(1 + \alpha(1 + 2\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*))) + \frac{1}{2} \ln \left( \frac{1 + \alpha \lambda_1(\bar{Z}^*) (1 + 2\alpha \sqrt{r^*} \delta_{2(r^{*+1})})}{1 + \alpha \lambda_1(\bar{Z}^*) (1 - 2\alpha \sqrt{r^*} \delta_{2(r^{*+1})})} \right)}{\ln \left( \frac{1 + \alpha \lambda_{r^*}(\bar{Z}^*) - \alpha 2\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*)}{1 + \alpha 2\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*)} \right)}}_{\triangleq A}, \end{aligned}$$

where for convenience we defined  $\eta \triangleq 2\delta_4 + 2(\rho + (1 + 2\delta_4)\hat{\Delta}_4)$ . Using  $\delta_{2(r^{*+1})} < \frac{11}{51\sqrt{r^*}} \kappa^{-4}$  and  $\alpha \leq \frac{1}{2} \kappa^{-2} \|\bar{Z}^*\|^{-1}$ , we have

$$A \leq \frac{\lambda_1(\bar{Z}^*) + 6\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*)}{\lambda_{r^*}(\bar{Z}^*) - 4\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*)} (1 + \alpha \lambda_{r^*}(\bar{Z}^*)) \leq 12\kappa^2.$$

Thus, for  $t_* \leq t^*$  it suffices that

$$\ln \left( \frac{(1 + 2\sqrt{r^*} \delta_{2(r^{*+1})} \lambda_1(\bar{Z}^*)) \lambda_1(\bar{Z}^*)}{4(1 + \eta) \mu^2 \frac{\min\{r, md\} (\|\mathcal{J}U\| + \sqrt{r^*} \rho^{1/2} \|(\mathcal{I} - \mathcal{J})U\|)^3}{m \|U^\top v_1\|}} \right) \geq 96\kappa^2 \ln(\sigma),$$



which can be rearranged to yield

$$\mu^2 \leq \frac{m(1 + 2\sqrt{r^*}\delta_{2(r^*+1)})\lambda_1(\bar{Z}^*)\|U^\top v_1\|}{4(1 + \eta)\min\{r, md\}(\|\mathcal{J}U\| + \sqrt{r^*}\rho^{1/2}\|(\mathcal{I} - \mathcal{J})U\|)^3}(\sigma)^{-96\kappa^2}.$$

We have thus established that up until  $t_*$   $\|E^t\|$  can be bounded by invoking Lemma 5. Thus, we evaluate the upper bound on  $\|E^t\|$  from Lemma 5 on the RHS of (69) to establish a sufficient condition for (69) to hold. By evaluating and rearranging we obtain

$$\begin{aligned} & \frac{4\kappa^2}{m(1 + 2\sqrt{r^*}\delta_{2(r^*+1)})\lambda_1(\bar{Z}^*)}\mu^3 \min\{r, md\}(1 + \eta) \times \\ & \times \left(1 + \alpha \left(1 + 2\sqrt{r^*}\delta_{2(r^*+1)}\right)\lambda_1(\bar{Z}^*)\right)^{3t} \left(\|\mathcal{J}U\| + \rho^{1/2}\sqrt{r^*}\|(\mathcal{I} - \mathcal{J})U\|\right)^3 \\ & \leq \frac{c}{2}(1 - \delta_\rho)\mu \left(1 + \alpha\lambda_{r^*}(\bar{Z}^*) - \alpha 2\sqrt{r^*}\delta_{2(r^*+1)}\lambda_1(\bar{Z}^*)\right)^t \sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U). \end{aligned}$$

A sufficient conditions for the inequality above to hold is

$$\begin{aligned} \mu^2 & \leq \frac{c(1 - \delta_\rho)(1 + 2\sqrt{r^*}\delta_{2(r^*+1)})\lambda_1(\bar{Z}^*)m\sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U)}{8\kappa^2 \min\{r, md\}(1 + \Delta)(\|\mathcal{J}U\| + \rho^{1/2}\sqrt{r^*}\|(\mathcal{I} - \mathcal{J})U\|)^3} \times \\ & \times \left(\frac{1 + \alpha\lambda_{r^*}(\bar{Z}^*) - 2\alpha\sqrt{r^*}\delta_{2(r^*+1)}\lambda_1(\bar{Z}^*)}{(1 + \alpha(1 + 2\sqrt{r^*}\delta_{2(r^*+1)})\lambda_1(\bar{Z}^*))^3}\right)^{t_*}, \end{aligned}$$

whereby algebra there holds that

$$\left(\frac{1 + \alpha\lambda_{r^*}(\bar{Z}^*) - 2\alpha\sqrt{r^*}\delta_{2(r^*+1)}\lambda_1(\bar{Z}^*)}{(1 + \alpha(1 + 2\sqrt{r^*}\delta_{2(r^*+1)})\lambda_1(\bar{Z}^*))^3}\right)^{t_*} \geq \exp(-12\kappa^2 \ln(\sigma)).$$

Therefore, a sufficient requirement on  $\mu^2$  for (69) is

$$\mu^2 \leq \frac{c(1 - \delta_\rho)(1 + 2\sqrt{r^*}\delta_{2(r^*+1)})\lambda_1(\bar{Z}^*)m\sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U)}{8\kappa^2 \min\{r, md\}(1 + \eta)(\|\mathcal{J}U\| + \rho^{1/2}\sqrt{r^*}\|(\mathcal{I} - \mathcal{J})U\|)^3} \times (\sigma)^{-12\kappa^2}.$$

## E.5 Proof of Propositions 2 and 4

We start with the part that is common to both propositions and then particularize for each of the two.

With probability at least  $1 - \mathcal{O}(\exp(-c_0d))$  it holds that

$$\|\mathcal{J}U\| \leq \sqrt{\frac{dm}{r}} \quad \text{and} \quad \rho^{1/2}\|U\| \leq \sqrt{\frac{d}{r}}m\rho^{1/2}.$$

Further, with probability  $1 - \mathcal{O}(e^{-r})$  there holds

$$\|U^\top v_1\| \geq \frac{\sqrt{m}}{2}.$$

Finally, observe that  $\sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U) = \frac{1}{\sqrt{m}}\sigma_{\min}(V_M^\top \bar{U})$  where the elements of  $\bar{U}$  are i.i.d. and distributed as  $\mathcal{N}(0, 1/\sqrt{r})$ , and consequently, we Theorem 8.8 in [34] from which we conclude that for every  $\varepsilon > 0$  there holds with probability at least  $1 - (c_1\varepsilon)^{r-r^*+1} - \exp(-c_2r)$  that

$$\sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U) \geq \sqrt{m}\varepsilon \frac{\sqrt{r} - \sqrt{r^* - 1}}{\sqrt{r}} \geq \begin{cases} \sqrt{m}\varepsilon \frac{2-\sqrt{2}}{2} & \text{if } r \geq 2r^* \\ \sqrt{m}\frac{\varepsilon}{r} & \text{otherwise.} \end{cases}$$

**Case 1 (Proposition 2)**  $r \geq 2r^*$  Setting  $\varepsilon = \frac{e^{-1/2}}{c_1}$  there yields with probability at least  $1 - \exp(-r/4) - \exp(c_2r)$  that

$$\sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U) \geq \sqrt{m} \frac{e^{-1/2}(2 - \sqrt{2})}{2c_2}.$$

Consequently,

$$\mu^2 \leq \frac{(1 + 2\alpha\sqrt{r^*}\delta_{2(r^*+1)}\lambda_1(\bar{Z}^*))}{4(1 + \eta)\min\{r, md\}} \min \left\{ \frac{m\sqrt{m}}{2m\sqrt{m}\frac{d}{r}\sqrt{\frac{d}{r}}(1 + \sqrt{\rho r^* m})^3} \sigma^{-96\kappa^2}, \right. \\ \left. \frac{c(1 - \delta_\rho)m\sqrt{m}e^{-1/2}(2 - \sqrt{2})}{16c_2\kappa^2 m\sqrt{m}\frac{d}{r}\sqrt{\frac{d}{r}}(1 + \sqrt{\rho r^* m})^3} \sigma^{-12\kappa^2} \right\},$$

with

$$\sigma \leq 4c_2\kappa^2 \frac{(1 + \alpha 2\sqrt{r^*}\delta_{2(r^*+1)}\lambda_1(\bar{Z}^*))\sqrt{m}\sqrt{\frac{d}{r}} + (\rho^2 + \alpha\rho^{1/2}\nu)\sqrt{\frac{d}{r}}m}{c(1 + \alpha 2\sqrt{r^*}\delta_{2(r^*+1)}\lambda_1(\bar{Z}^*))\sqrt{m}e^{-1/2}(2 - \sqrt{2})},$$

and

$$\frac{e^{-1/2}(2 - \sqrt{2})}{2c_2} \sqrt{m} \leq \beta \leq 2\sqrt{m}\sigma^2.$$

**Case 2 (Proposition 4)**  $r^* \leq r \leq 2r^*$ . With probability  $1 - (c_1\varepsilon)^{r-r^*+1} - \exp(-c_2r)$ , it holds

$$\sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U) \geq \sqrt{m}\frac{\varepsilon}{r}.$$

Consequently,

$$\mu^2 \leq \frac{(1 + \alpha 2\sqrt{r^*}\delta_{2(r^*+1)}\lambda_1(\bar{Z}^*))}{4(1 + \eta)\min\{r, md\}} \min \left\{ \frac{m\sqrt{m}}{2m\sqrt{m}\frac{d}{r}\sqrt{\frac{d}{r}}(1 + \sqrt{\rho r^* m})^3} \sigma^{-96\kappa^2}, \right. \\ \left. \frac{c(1 - \delta_\rho)m\sqrt{m}\varepsilon}{8c_2\kappa^2 m\sqrt{m}\frac{d}{r}\sqrt{\frac{d}{r}}(1 + \sqrt{\rho r^* m})^3} \sigma^{-12\kappa^2} \right\},$$

with

$$\sigma \leq 2\kappa^2 r \frac{(1 + \alpha 2\sqrt{r^*}\delta_{2(r^*+1)}\lambda_1(\bar{Z}^*))\sqrt{\frac{md}{r}} + (\rho^2 + \alpha\rho^{1/2}\nu)m\sqrt{\frac{d}{r}}}{c(1 + 2\alpha\sqrt{r^*}\delta_{2(r^*+1)}\lambda_1(\bar{Z}^*))\sqrt{m}\varepsilon},$$

and

$$\sqrt{m}\frac{\varepsilon}{r} \leq \beta \leq 2\sqrt{m}\sigma^2.$$

## E.6 Technical lemmas: auxiliary to proving Lemma 7

**Lemma 10.** Suppose that  $\bar{A}$  satisfies the  $(\delta_{2(r^*+1)}, 2(r^*+1))$ -RIP property. Then, if

$$\delta_\rho < 1,$$

and  $W$  satisfies Assumption 1 with  $\rho$  such that

$$\rho^2 \leq \frac{2\delta_{2(r^*+1)}^2}{4m^5(1 + 2\delta_{2(r^*+1)})^2},$$

then,

$$\sigma_{r^*}(U^t) \geq (1 - \delta_\rho)\mu\sigma_{r^*}(\mathcal{M}^t)\sigma_{\min}(V_{\mathcal{J}M\mathcal{J}}^\top U) - \|E^t\|$$

and

$$\sigma_{r^*+1}(U^t) \leq \mu\sigma_{r^*+1}(\mathcal{M}^{t-1})(1 + \alpha 2\sqrt{r^*}\delta_{2(r^*+1)}\lambda_1(\bar{Z}^*))\|\mathcal{J}U\| \\ + \mu\sigma_{r^*+1}(\mathcal{M}^{t-1})\left(\rho^2 + \alpha\rho^{1/2}\nu\right)\|(\mathcal{I} - \mathcal{J})U\| + \|E^t\|.$$

Then, if

$$A \triangleq \mu\sigma_{r^*}(\mathcal{M}^t)(1 - \delta_\rho)\sigma_{\min}(V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\top U) - \mu\sigma_{r^*+1}(\mathcal{M}^{t-1}) \left(1 + 2\alpha\sqrt{r^*}\delta_{2(r^*+1)}\lambda_1(\bar{Z}^*)\right) \|\mathcal{J}U\| \\ - \mu\sigma_{r^*+1}(\mathcal{M}^{t-1}) \left(\rho^2 + \alpha\rho^{1/2}\nu\right) \|(\mathcal{I} - \mathcal{J})U\| - \|E^t\| > 0,$$

it holds

$$\|(V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\perp)^\top V_{U^t}\| \leq \frac{\mu\sigma_{r^*+1}(\mathcal{M}^t)\|\mathcal{J}U\| + \mu\sigma_{r^*+1}(\mathcal{M}^{t-1}) \left(\rho^2 + \alpha\rho^{1/2}\nu\right) \|(\mathcal{I} - \mathcal{J})U\| + \|E^t\|}{(1 - \delta_\rho)A}.$$

*Proof.* From Weyl's inequality it follows that

$$\sigma_{r^*}(\mathcal{M}^t U^0 + E^t) \geq \sigma_{r^*}(\mathcal{M}^t U^0) - \|E^t\| \geq \sigma_{r^*}(V_{\mathcal{M}}^\top \mathcal{M}^t U^0) - \|E^t\|, \\ \sigma_{r^*}(V_{\mathcal{M}}^\top \mathcal{M}^t U^0) \geq \sigma_{\min}(V_{\mathcal{M}}^\top \mathcal{M}^t V_{\mathcal{M}} V_{\mathcal{M}}^\top U^0) \geq \sigma_{r^*}(\mathcal{M}^t) \sigma_{\min}(V_{\mathcal{M}}^\top U^0).$$

Observe that

$$\sigma_{\min}(V_{\mathcal{M}}^\top U^0) \geq \sigma_{\min}(V_{\mathcal{M}}^\top V_{\mathcal{J}\mathcal{M}\mathcal{J}} V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\top U^0) \geq \sigma_{\min}(V_{\mathcal{M}}^\top V_{\mathcal{J}\mathcal{M}\mathcal{J}}) \sigma_{\min}(V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\top U^0).$$

Further,

$$\sigma_{\min}(V_{\mathcal{M}}^\top V_{\mathcal{J}\mathcal{M}\mathcal{J}}) \geq 1 - \|(V_{\mathcal{M}}^\perp)^\top V_{\mathcal{J}\mathcal{M}\mathcal{J}}\|.$$

Invoking the David-Kahan  $\sin(\Theta)$ -theorem we have

$$\|(V_{\mathcal{M}}^\perp)^\top V_{\mathcal{J}\mathcal{M}\mathcal{J}}\|_F \leq \frac{\sqrt{r^*} \frac{\alpha}{m} \|(V_{\mathcal{M}}^\perp)^\top (\mathcal{I} - \mathcal{J}) \mathcal{A}^* \mathcal{A}(\bar{Z}^*) \mathcal{J} V_{\mathcal{J}\mathcal{M}\mathcal{J}}\|}{\min_{i \in [r^*], r^*+1 \leq j \leq md} \{|\lambda_i(\mathcal{J}\mathcal{M}\mathcal{J}) - \lambda_j(\mathcal{M})|\}}.$$

Observe that from Weyl's inequality and Lemma 2

$$\lambda_{r^*}(\mathcal{J}\mathcal{M}\mathcal{J}) \geq 1 + \alpha\lambda_{r^*}(\bar{Z}^*) - \alpha\delta_{r^*}\sqrt{r^*}\lambda_1(\bar{Z}^*), \\ \lambda_{r^*+1}(\mathcal{M}) \leq 1 + \alpha 2\delta_{2r^*}\sqrt{r^*}\lambda_1(\bar{Z}^*).$$

Further, invoking Lemma 4 we obtain

$$\|(V_{\mathcal{M}}^\perp)^\top V_{\mathcal{J}\mathcal{M}\mathcal{J}}\|_F \leq \frac{\sqrt{r^*}\rho^{1/2}\lambda_1(\bar{Z}^*)(1 + \sqrt{r^*}\delta_{2r^*})}{\underbrace{\lambda_{r^*}(\bar{Z}^*) - 3\delta_{2(r^*+1)}\sqrt{r^*}\lambda_1(\bar{Z}^*)}_{\triangleq \delta_\rho}}.$$

With the choice of  $\rho$  selected by the lemma there holds  $\delta_\rho < 1$  and therefore

$$\sigma_{\min}(V_{\mathcal{M}}^\top U^0) \geq (1 - \delta_\rho)\sigma_{\min}(V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\top U^0).$$

Consequently, it follows that

$$\sigma_{r^*}(\mathcal{M}^t U^0 + E^t) \geq (1 - \delta_\rho)\mu\sigma_{r^*}(\mathcal{M}^t)\sigma_{\min}(V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\top U) - \|E^t\|.$$

Observe that

$$\sigma_{r^*+1}(\mathcal{M}^t U^0) \leq \sigma_{r^*+1}(\mathcal{M}^{t-1} \mathcal{M} U^0) \leq \sigma_{r^*}(\mathcal{M}^{t-1}) (\sigma_1(\mathcal{M}\mathcal{J}U) + \sigma_1(\mathcal{M}(\mathcal{I} - \mathcal{J})U^0)).$$

Consequently invoking Lemma 4 we obtain

$$\sigma_{r^*+1}(\mathcal{M}^t U^0 + E^t) \leq \mu\sigma_{r^*+1}(\mathcal{M}^{t-1}) \left(1 + \alpha 2\sqrt{r^*}\delta_{2(r^*+1)}\lambda_1(\bar{Z}^*)\right) \|\mathcal{J}U\| \\ + \mu\sigma_{r^*+1}(\mathcal{M}^{t-1}) \left(\left(\rho^2 + \alpha\rho^{1/2}\lambda_1(\bar{Z}^*)(1 + \sqrt{r^*}\delta_{2r^*})\right)\|(\mathcal{I} - \mathcal{J})U\| + \|E^t\|\right).$$

Further, we write

$$\mathcal{M}^t U^0 + E^t = \mu Z^t V_{\mathcal{M}} V_{\mathcal{M}}^\top U + \underbrace{Z^t V_{\mathcal{M}}^\perp (V_{\mathcal{M}}^\perp)^\top U^0 + E^t}_{\triangleq H}.$$

Under the assumption that  $V_{\mathcal{M}}^\top U$  is of rank  $r^*$ ,  $\mathcal{M}^t V_{\mathcal{M}} V_{\mathcal{M}}^\top U$  is full ranked. Then,  $V_{\mathcal{M}}^\top V_{\mathcal{M}} \mathcal{M}^t V_{\mathcal{M}} V_{\mathcal{M}}^\top U$  is of rank  $r^*$  and  $V_{\mathcal{M}}$  is spanned by the left singular vectors of  $Z^t V_{\mathcal{M}} V_{\mathcal{M}}^\top U$ . Under the assumption that

$$A > 0,$$

it follows from the Davis-Kayan  $\sin(\Theta)$  theorem that

$$\|(V_{\mathcal{M}}^\perp)^\top V_{U^t}\| \leq \frac{\|H\|}{A},$$

whereby using  $\sigma_{\min}(V_{\mathcal{M}}^\top U) \geq (1 - \delta_\rho) \sigma_{\min}(V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\top U)$  and that

$$\begin{aligned} \|H\| &\leq \mu \|\mathcal{M}^t V_{\mathcal{M}}^\perp (V_{\mathcal{M}}^\perp)^\top \mathcal{J}\| + \mu \|\mathcal{M}^t V_{\mathcal{M}}^\perp (V_{\mathcal{M}}^\perp)^\top (\mathcal{I} - \mathcal{J})U\| + \|E^t\| \\ &\leq \mu \sigma_{r^*+1}(\mathcal{M}^t) \|\mathcal{J}U\| + \mu \|V_{\mathcal{M}}^\perp (V_{\mathcal{M}}^\perp)^\top \mathcal{M}^t (\mathcal{I} - \mathcal{J})U\| + \|E^t\|. \end{aligned}$$

In this way,

$$\mu \|V_{\mathcal{M}}^\perp (V_{\mathcal{M}}^\perp)^\top \mathcal{M}^t (\mathcal{I} - \mathcal{J})U\| \leq \mu \|V_{\mathcal{M}}^\perp (V_{\mathcal{M}}^\perp)^\top \mathcal{M}^{t-1} \|\mathcal{M}(\mathcal{I} - \mathcal{J})U\|.$$

Invoking Lemma 4, we have

$$\mu \|V_{\mathcal{M}}^\perp (V_{\mathcal{M}}^\perp)^\top \mathcal{M}^t (\mathcal{I} - \mathcal{J})U\| \leq \mu \sigma_{r^*+1}(\mathcal{M}^{t-1}) \times \left( \rho^2 + \alpha \rho^{1/2} \lambda_1(\bar{Z}^*) \left( 1 + \sqrt{r^*} \delta_{2r^*} \right) \right) \|(\mathcal{I} - \mathcal{J})U\|,$$

yielding

$$\begin{aligned} \|(V_{\mathcal{M}}^\perp)^\top V_{U^t}\| &\leq \\ &\frac{\mu \sigma_{r^*+1}(\mathcal{M}^t) \|\mathcal{J}U\| + \mu \sigma_{r^*+1}(\mathcal{M}^{t-1}) (\rho^2 + \alpha \rho^{1/2} \lambda_1(\bar{Z}^*) (1 + \sqrt{r^*} \delta_{2r^*})) \|(\mathcal{I} - \mathcal{J})U\| + \|E^t\|}{A}. \end{aligned}$$

Observe that

$$\|(V_{\mathcal{M}}^\perp)^\top V_{U^t}\| \geq \|(V_{\mathcal{M}}^\perp)^\top V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\perp (V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\perp)^\top V_{U^t}\| \geq \sigma_{\min}((V_{\mathcal{M}}^\perp)^\top V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\perp) \|(V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\perp)^\top V_{U^t}\|,$$

where

$$\sigma_{\min}((V_{\mathcal{M}}^\perp)^\top V_{\mathcal{J}\mathcal{M}\mathcal{J}}^\perp) \geq 1 - \|(V_{\mathcal{M}}^\perp)^\top V_{\mathcal{J}\mathcal{M}\mathcal{J}}\| \geq 1 - \delta_\rho,$$

yielding the result.  $\square$

**Lemma 11.** Assume that  $\|(V_{Z^*}^\perp)^\top V_{U^t}\| \leq \frac{1}{8}$  for some  $t \geq 1$ . Then, the following hold:

$$\begin{aligned} \sigma_{r^*}(U^t Q^t) &\geq \frac{\sigma_{r^*}(U^t)}{2}, \\ \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| &\leq 7 \|(V_{Z^*}^\perp)^\top V_{U^t}\|, \\ \|U^t Q^{t,\perp}\| &\leq 2 \sigma_{r^*+1}(U^t). \end{aligned}$$

The proof of Lemma 11 follows readily from that of Lemma 8.4 in [34].

## F Proofs of the Intermediate Results in Phase II (Appendix B.2)

### F.1 Proofs of Lemma 9

(i) The proof of this statement follows readily from that of Lemma 9.4 in [34] and thus is omitted.

(ii)

$$\begin{aligned} V_{Z^*}^\top U^{t+1} Q^t &= V_{Z^*}^\top \left( \mathcal{W}^2 + \frac{\alpha}{m} \mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top) \right) U^t Q^t \\ &= V_{Z^*}^\top \left( \mathcal{W}^2 + \frac{\alpha}{m} (Z^* - \mathcal{W} U^t (\mathcal{W} U^t)^\top) \right) U^t Q^t \\ &\quad + V_{Z^*}^\top \left( \frac{\alpha}{m} (\mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top) - \mathcal{W} (Z^* - U^t (U^t)^\top) \mathcal{W}) \right) U^t Q^t. \end{aligned}$$

Observe that

$$V_{Z^*}^\top \mathcal{W}^2 = \left( \frac{1}{\sqrt{m}} \mathbf{1}_m^\top \otimes V_{Z^*}^\top \right) (W^2 \otimes I_d) = \left( \frac{1}{\sqrt{m}} \mathbf{1}_m^\top \otimes V_{Z^*}^\top \right).$$

Further,  $X = \sqrt{m} V_{Z^*} \Sigma_{\bar{X}} Q_X^\top$  and consequently

$$\begin{aligned} V_{Z^*}^\top U^{t+1} Q^t &= (I + \alpha \Sigma_{\bar{X}}^2) (V_{Z^*}^\top)^\top U^t Q^t - \frac{\alpha}{m} V_{Z^*}^\top U^t Q^t (Q^t)^\top (\mathcal{W} U^t)^\top U^t Q^t \\ &\quad + V_{Z^*}^\top \frac{\alpha}{m} (\mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top) - \mathcal{W} (Z^* - U^t (U^t)^\top) \mathcal{W}) U^t Q^t \\ &= (I + \alpha \Sigma_{\bar{X}}^2) V_{Z^*}^\top U^t Q^t \left( I - \frac{\alpha}{m} (Q^t)^\top (U^t)^\top V_{Z^*} V_{Z^*}^\top U^t Q^t \right) \\ &\quad - \frac{\alpha}{m} \underbrace{V_{Z^*}^\top U^t (U^t)^\top \mathcal{W} (V_{Z^*}^\perp (V_{Z^*}^\perp)^\top)}_{\triangleq A_1} U^t Q^t \\ &\quad + \frac{\alpha}{m} \underbrace{V_{Z^*}^\top (\mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top) - \mathcal{W} (Z^* - U^t (U^t)^\top) \mathcal{W})}_{\triangleq A_2} U^t Q^t \\ &\quad + \frac{\alpha^2}{m} \underbrace{\Sigma_{\bar{X}}^2 V_{Z^*}^\top U^t (U^t)^\top V_{Z^*} V_{Z^*}^\top U^t Q^t}_{\triangleq A_3}. \end{aligned}$$

We now rewrite each  $A_i$  as

$$A_i = P_i V_{Z^*}^\top U^t Q^t \left( I - \frac{\alpha}{m} (Q^t)^\top (U^t)^\top V_{Z^*} V_{Z^*}^\top U^t Q^t \right).$$

In order to proceed with  $A_1$  we focus on the term  $(V_{Z^*}^\perp)^\top U^t Q^t$ . Under the assumption that  $V_{Z^*}^\top U^t$  is of rank  $r^*$  we have

$$(V_{Z^*}^\perp)^\top U^t Q^t = (V_{Z^*}^\perp)^\top U^t Q^t (V_{Z^*}^\top U^t Q^t)^{-1} (V_{Z^*}^\top U^t Q^t).$$

Using the singular value decomposition  $U^t Q^t = V_{U^t Q^t} \Sigma_{U^t Q^t} Q_{U^t Q^t}^\top$ , yields

$$\begin{aligned} (V_{Z^*}^\perp)^\top U^t Q^t &= (V_{Z^*}^\perp)^\top V_{U^t Q^t} \Sigma_{U^t Q^t} Q_{U^t Q^t}^\top (V_{Z^*}^\top V_{U^t Q^t} \Sigma_{U^t Q^t} Q_{U^t Q^t}^\top)^{-1} (V_{Z^*}^\top V_{U^t Q^t} \Sigma_{U^t Q^t} Q_{U^t Q^t}^\top) \\ &= (V_{Z^*}^\perp)^\top V_{U^t Q^t} (V_{Z^*}^\top V_{U^t Q^t})^{-1} V_{Z^*}^\top U^t Q^t. \end{aligned}$$

Therefore, we write  $A_1$

$$V_{Z^*}^\top U^t (U^t)^\top \mathcal{W} (V_{Z^*}^\perp (V_{Z^*}^\perp)^\top) U^t Q^t = V_{Z^*}^\top U^t (U^t)^\top \mathcal{W} V_{Z^*}^\perp (V_{Z^*}^\perp)^\top V_{U^t Q^t} (V_{Z^*}^\top V_{U^t Q^t})^{-1} V_{Z^*}^\top U^t Q^t$$

post-multiplying by  $(I - \frac{\alpha}{m} (U^t Q^t)^\top V_{Z^*} V_{Z^*}^\top U^t Q^t)^{-1}$  and its inverse (a sufficient condition for the inverse to exist is given by  $\alpha 9 \|\bar{X}\|^2 \leq 1$  which holds under the assumption on  $\alpha$ )

$$\begin{aligned} P_1 &= V_{Z^*}^\top U^t (U^t)^\top \mathcal{W} (V_{Z^*}^\perp (V_{Z^*}^\perp)^\top) V_{U^t Q^t} (V_{Z^*}^\top V_{U^t Q^t})^{-1} \times \\ &\quad \left( I - \frac{\alpha}{m} V_{Z^*}^\top U^t Q^t (U^t Q^t)^\top V_{Z^*} \right) V_{Z^*}^\top U^t Q^t. \end{aligned}$$

We proceed now similarly with  $A_2$  yielding

$$A_2 = \frac{\alpha}{m} V_{Z^*}^\top (\mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top) - \mathcal{W} (Z^* - U^t (U^t)^\top) \mathcal{W}) V_{U^t Q^t} (V_{Z^*}^\top U^t Q^t)^{-1} V_{Z^*}^\top U^t Q^t$$

following the steps as for the case of  $A_1$  we have

$$\begin{aligned} P_2 &= V_{Z^*}^\top (\mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top) - \mathcal{W} (Z^* - U^t (U^t)^\top) \mathcal{W}) V_{U^t Q^t} (V_{Z^*}^\top V_{U^t Q^t})^{-1} \times \\ &\quad \times \left( I - \frac{\alpha}{m} V_{Z^*}^\top U^t Q^t (U^t Q^t)^\top V_{Z^*} \right)^{-1}. \end{aligned}$$

Similarly,

$$\begin{aligned} A_3 &= \underbrace{\Sigma_{\bar{X}}^2 V_{Z^*}^\top U^t Q^t \left( I - \frac{\alpha}{m} (Q^t)^\top (U^t)^\top V_{Z^*} V_{Z^*}^\top U^t Q^t \right)^{-1} (Q^t)^\top (U^t)^\top V_{Z^*}}_{\triangleq P_3} \times \\ &\quad \times V_{Z^*}^\top U^t Q^t \left( I - \frac{\alpha}{m} (Q^t)^\top (U^t)^\top V_{Z^*} V_{Z^*}^\top U^t Q^t \right). \end{aligned}$$

Consequently, under the assumption that

$$\alpha \leq \frac{1}{8100} \|\bar{Z}^*\|^{-1} \kappa^{-2} \text{ and } \|U^t\| \leq 3\sqrt{m} \|\bar{X}\|$$

and using Weyl's inequality we have

$$\begin{aligned} \sigma_{\min}(V_{Z^*}^\top U^{t+1} Q^t) &\geq \sigma_{\min} \left( I + \alpha \Sigma_{\bar{X}}^2 - \frac{\alpha}{m} P_1 + \frac{\alpha}{m} P_2 + \frac{\alpha^2}{m} P_3 \right) \times \\ &= \sigma_{\min} \left( I + \alpha \Sigma_{\bar{X}}^2 - \frac{\alpha}{m} P_1 + \frac{\alpha}{m} P_2 + \frac{\alpha^2}{m} P_3 \right) \sigma_{\min}(V_{Z^*}^\top U^t Q^t) \left( 1 - \frac{\alpha}{m} \sigma_{\min}^2(V_{Z^*}^\top U^t Q^t) \right) \\ &\geq \left( 1 + \alpha \lambda_{r^*}(\bar{Z}^*) - \frac{\alpha}{m} \|P_1\| - \frac{\alpha}{m} \|P_2\| - \frac{\alpha^2}{m} \|P_3\| \right) \sigma_{\min}(V_{Z^*}^\top U^t) \left( 1 - \frac{\alpha}{m} \sigma_{\min}^2(V_{Z^*}^\top U^t) \right). \end{aligned}$$

We now bound each of the quantities  $\|P_i\|$  with

$$\|P_1\| = \left\| V_{Z^*}^\top U^t Q^t (U^t Q^t)^\top \mathcal{W} (V_{Z^*}^\perp (V_{Z^*}^\perp)^\top) V_{U^t Q^t} (V_{Z^*}^\top V_{U^t Q^t})^{-1} \left( I - \frac{\alpha}{m} V_{Z^*}^\top U^t (U^t)^\top V_{Z^*} \right)^{-1} \right\|.$$

By assumption  $\|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| \leq c_1 \kappa^{-1}$  implying  $\sigma_{\min}((V_{Z^*}^\perp)^\top V_{U^t Q^t}) \geq 1 - c_1 \kappa^{-1}$  and consequently observing that  $Z^* \mathcal{W} = \mathcal{W} Z^*$

$$\|(V_{Z^*}^\perp)^\top \mathcal{W} V_{U^t Q^t}\| \leq \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\|$$

there holds

$$\begin{aligned} \|P_1\| &\leq \frac{\|U^t Q^t\|^2 \|V_{Z^*}^\top V_{U^t Q^t}\| \|V_{Z^*}^\top \mathcal{W} V_{U^t Q^t}\|}{\sigma_{\min}(V_{Z^*}^\top V_{U^t Q^t}) (1 - \frac{\alpha}{m} \|V_{Z^*}^\top U^t\|^2)} \leq \frac{\|U^t Q^t\|^2 \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\|^2}{(1 - c_1 \kappa^{-1}) (1 - \frac{\alpha}{m} 9 \|\bar{X}\|^2)} \\ &\leq \frac{\|U^t Q^t\|^2 \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\|^2}{(1 - c_1 \kappa^{-1}) (1 - 9 \kappa^{-2} (1/81000))} \leq 4 \|U^t Q^t\|^2 \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\|^2 \\ &\leq 4 \|U^t Q^t\|^2 c_1 \kappa^{-1} \leq 36m \|\bar{X}\|^2 c_1 \kappa^{-2}. \end{aligned}$$

Similarly,

$$\begin{aligned} \|P_2\| &\leq \frac{\|\mathcal{A}^* \mathcal{A}(Z^* - U^t (U^t)^\top) - \mathcal{W}(Z^* - U^t (U^t)^\top) \mathcal{W}\|}{\sigma_{\min}(V_{Z^*}^\top V_{U^t Q^t}) (1 - \frac{\alpha}{m} \|V_{Z^*}^\top U^t\|^2)} \\ &\leq \frac{\|\mathcal{A}^* \mathcal{A}(Z^* - U^t (U^t)^\top) - \mathcal{W}(Z^* - U^t (U^t)^\top) \mathcal{W}\|}{(1 - c \kappa^{-1}) (1 - 9 \kappa^{-2} 1/81000)} \\ &\leq 4 \|\mathcal{A}^* \mathcal{A}(Z^* - U^t (U^t)^\top) - \mathcal{W}(Z^* - U^t (U^t)^\top) \mathcal{W}\| \end{aligned}$$

and

$$\|P_3\| \leq \frac{\|\bar{X}\|^2 \|V_{Z^*}^\top U^t Q^t\|^2}{1 - \frac{\alpha}{m} \|V_{Z^*}^\top U^t Q^t\|^2} \leq \frac{\|\bar{X}\|^2 \|V_{Z^*}^\top U^t Q^t\|^2}{1 - \frac{\alpha}{9} \|\bar{X}\|^2} \leq \frac{\|\bar{X}\|^2 \|V_{Z^*}^\top U^t Q^t\|^2}{1 - 9(1/81000) \kappa^{-1}},$$

under the assumption that  $\alpha$  there holds

$$\|P_3\| \leq 2 \|\bar{X}\|^2 \|V_{Z^*}^\top U^t Q^t\|^2 \leq 18m \|\bar{X}\|^4.$$

Combining all estimates together with  $\|U^t Q^t\|^2 \leq 9m \|\bar{X}\|^2$  yields

$$\begin{aligned} \sigma_{\min}(V_{Z^*}^\top U^{t+1} Q^t) &\geq (1 + \alpha \lambda_{r^*}(\bar{Z}^*) - \alpha 36 c_1 \kappa^{-1} \|\bar{X}\|^2 - \alpha^2 18 \|\bar{X}\|^4 \\ &\quad - \frac{\alpha}{m} 4 \|\mathcal{A}^* \mathcal{A}(Z^* - U^t (U^t)^\top) - \mathcal{W}(Z^* - U^t (U^t)^\top) \mathcal{W}\|) \times \\ &\quad \times \sigma_{\min}(V_{Z^*}^\top U^t) \left( 1 - \frac{\alpha}{m} \sigma_{\min}^2(V_{Z^*}^\top U^t) \right). \end{aligned}$$

Observe that  $\alpha \leq \frac{\lambda_{r^*}(\bar{Z}^*)}{81000 \|\bar{X}\|^2} \|\bar{X}\|^{-2}$  and therefore

$$\begin{aligned} \sigma_{\min}(V_{Z^*}^\top U^{t+1} Q^t) &\geq \left( 1 + \alpha \lambda_{r^*}(\bar{Z}^*) - \alpha c_1 36 \kappa^{-2} \|\bar{X}\|^2 - \alpha \frac{18}{81000} \lambda_{r^*}(\bar{Z}^*) \right. \\ &\quad \left. - \frac{\alpha}{m} 4 \|\mathcal{A}^* \mathcal{A}(Z^* - U^t (U^t)^\top) - \mathcal{W}(Z^* - U^t (U^t)^\top) \mathcal{W}\| \right) \sigma_{\min}(V_{Z^*}^\top U^t) \left( 1 - \frac{\alpha}{m} \sigma_{\min}^2(V_{Z^*}^\top U^t) \right). \end{aligned}$$

Therefore

$$\sigma_{\min}(V_{Z^*}^\top U^{t+1} Q^t) \geq (1 + \alpha(1 - 55c_1) \sigma_{\min}^2(\bar{X})) \sigma_{\min}(V_{Z^*}^\top U^t) \left(1 - \frac{\alpha}{m} \sigma_{\min}^2(V_{Z^*}^\top U^t)\right).$$

Under the assumption that  $c_1 \leq \frac{1}{110}$ , there holds

$$\begin{aligned} \sigma_{\min}(V_{Z^*}^\top U^{t+1} Q^t) &\geq \left(1 + \frac{\alpha}{2} \sigma_{\min}^2(\bar{X})\right) \sigma_{\min}(V_{Z^*}^\top U^t) \left(1 - \frac{\alpha}{m} \sigma_{\min}^2(V_{Z^*}^\top U^t)\right) \\ &= \sigma_{\min}(V_{Z^*}^\top U^t) \left(1 + \frac{\alpha}{2} \sigma_{\min}^2(\bar{X}) \left(1 - \frac{\alpha}{m} \sigma_{\min}^2(V_{Z^*}^\top U^t)\right) - \frac{\alpha}{m} \sigma_{\min}^2(V_{Z^*}^\top U^t)\right) \\ &\geq \sigma_{\min}(V_{Z^*}^\top U^t) \left(1 + \frac{\alpha}{4} \sigma_{\min}^2(\bar{X}) - \frac{\alpha}{m} \sigma_{\min}^2(V_{Z^*}^\top U^t)\right). \end{aligned}$$

where we have used that

$$1 - \frac{\alpha}{m} \sigma_{\min}^2(V_{Z^*}^\top U^t) \geq 1 - \frac{\kappa^{-1} \|\bar{X}\|^{-2}}{81000m} 9m \|\bar{X}\|^2 \geq \frac{1}{2}.$$

The final result follows by noticing that  $\sigma_{\min}(V_{Z^*}^\top U^{t+1}) \geq \sigma_{\min}(V_{Z^*}^\top U^{t+1} Q^t)$ .

(iii) The proof of this statement is practically identical to that of [34, Lemma 9.2], and thus is omitted. However, [34, Lemma 9.2] builds upon [34, Lemma B.1] of which we provide a suitable version for our purpose in Lemma 12.

(iv) For convenience, define

$$U^{t+1} Q^{t+1} = \left(\mathcal{W}^2 + \frac{\alpha}{m} \mathcal{M}\right) U^t Q^{t+1} \quad (70)$$

$$= \left(\mathcal{W}^2 + \frac{\alpha}{m} \mathcal{M}\right) (V_{U^t Q^t} V_{U^t Q^t}^\top U^t Q^t (Q^t)^\top Q^{t+1} + U^t Q^{t,\perp} (Q^{t,\perp})^\top Q^{t+1}). \quad (71)$$

Due to the assumption on  $\|U^t Q^{t,\perp}\|$  and because  $(Q^t)^\top Q^{t+1}$  is invertible which follows from Lemma 13, we have that the matrix  $V_{U^t Q^t}^\top U^t Q^t (Q^t)^\top Q^{t+1}$  is invertible and we may consequently write

$$\begin{aligned} &U^t Q^{t,\perp} (Q^{t,\perp})^\top Q^{t+1} \\ &= \underbrace{U^t Q^{t,\perp} (Q^{t,\perp})^\top Q^{t+1} (V_{U^t Q^t}^\top U^t Q^t (Q^t)^\top Q^{t+1})^{-1} V_{U^t Q^t}^\top}_{\triangleq P} \times \\ &\quad \times V_{U^t Q^t} V_{U^t Q^t}^\top U^t Q^t (Q^t)^\top Q^{t+1}. \end{aligned}$$

Therefore, we have

$$U^{t+1} Q^{t+1} = \left(\mathcal{W}^2 + \frac{\alpha}{m} \mathcal{M}\right) (I + P) V_{U^t Q^t} V_{U^t Q^t}^\top V_{U^t Q^t}^\top U^t Q^t (Q^t)^\top Q^{t+1}.$$

Define the matrices

$$\begin{aligned} Z' &\triangleq \left(\mathcal{W}^2 + \frac{\alpha}{m} \mathcal{M}\right) (I + P) V_{U^t Q^t} \\ Z &\triangleq \left(I + \frac{\alpha}{m} \mathcal{M}\right) (I + P) V_{U^t Q^t}. \end{aligned}$$

Because the matrix  $U^t Q^t$  is full rank, there holds  $\|(V_{Z^*}^\perp)^\top V_{U^{t+1} Q^{t+1}}\| = \|(V_{Z^*}^\perp)^\top V_{Z'}\|$ . We now argue that  $\text{span}(V_{Z'}^\perp) \subseteq \text{span}(V_Z)$  and consequently  $\|(V_{Z^*}^\perp)^\top V_{Z'}\| \leq \|(V_{Z^*}^\perp)^\top V_Z\|$ . We observe that the main deviating factor between the matrices  $Z$  and  $Z'$  are the terms  $\mathcal{W}^2 + \frac{\alpha}{m} \mathcal{M}$  and  $I + \frac{\alpha}{m} \mathcal{M}$  respectively. It suffices then to establish that  $\text{span}(\mathcal{W}^2 + \frac{\alpha}{m} \mathcal{M}) \subseteq \text{span}(I + \frac{\alpha}{m} \mathcal{M})$  as the term  $(I + P) V_{U^t Q^t}$  is common to both matrices. Observe that because  $\mathcal{M}$  is symmetric by construction, it suffices to choose  $\alpha$  sufficiently small such that

$$\left\| \frac{\alpha}{m} \mathcal{M} \right\| \leq \frac{1}{2} \quad (72)$$

for  $I + \frac{\alpha}{m} \mathcal{M} \succeq \frac{1}{2} I$ . This implies that  $\text{span}(I + \frac{\alpha}{m} \mathcal{M}) = \mathbb{R}^{md}$ . This implies that

$$\|(V_{Z^*}^\perp)^\top V_{U^{t+1} Q^{t+1}}\| \leq \|(V_{Z^*}^\perp)^\top V_Z\| = \|(V_{Z^*}^\perp)^\top V_Z Q_Z^\top\| = \|V_\perp^\top Z (Z^\top Z)^{-1/2}\|. \quad (73)$$

From here we proceed similarly to the proof of [34, Lemma 9.3] with the following definitions of  $B_i$

$$B_1 \triangleq \frac{\alpha}{m} (X X^\top - \mathcal{W} U^t (\mathcal{W} U^t)^\top), \quad (74)$$

$$B_2 \triangleq \frac{\alpha}{m} (\mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top) - \mathcal{W} (Z^* - U^t (U^t)^\top) \mathcal{W}), \quad (75)$$

$$B_3 \triangleq U^t Q^{t,\perp} (Q^{t,\perp})^\top Q^{t+1} (V_{U^t Q^t}^\top U^t Q^t (Q^t)^\top Q^{t+1})^{-1} V_{U^t Q^t}^\top, \quad (76)$$

$$B_4 \triangleq \frac{\alpha}{m} \mathcal{M} P, \quad (77)$$

$$B \triangleq \frac{\alpha}{m} \mathcal{M} + P + \frac{\alpha}{m} \mathcal{M} P. \quad (78)$$

Suppose the following holds (a fact that will be proved later)

$$-\frac{1}{2} I \preceq V_{U^t Q^t}^\top B V_{U^t Q^t} + V_{U^t Q^t}^\top B^\top V_{U^t Q^t} + V_{U^t Q^t}^\top B^\top B V_{U^t Q^t} \preceq \frac{1}{2} I. \quad (79)$$

Following an identical procedure to that in [34, Lemma 9.3], we obtain

$$\begin{aligned} Z(Z^\top Z)^{-1/2} &= V_{U^t Q^t} + B V_{U^t Q^t} - \frac{1}{2} (I + B) V_{U^t Q^t} V_{U^t Q^t}^\top (B + B^\top) V_{U^t Q^t} - D \\ D &\triangleq (I + B) V_{U^t Q^t} \left( \frac{1}{2} V_{U^t Q^t}^\top B^\top B V_{U^t Q^t} - C \right). \end{aligned}$$

The expansion of the term  $(V_{Z^*}^\perp)^\top Z(Z^\top Z)^{-1/2}$  becomes analogous to that in [34, Lemma 9.3]. To upper bound the term  $(I)$  defined analogously as in [34, Lemma 9.3] we have

$$\begin{aligned} (I) &= V_\perp^\top \left( I + B_1 - \frac{1}{2} V_{U^t Q^t} V_{U^t Q^t}^\top (B_1 + B_1^\top) \right) V_{U^t Q^t} \\ &= (V_{Z^*}^\perp)^\top \left( I + \frac{\alpha}{m} (I - V_{U^t Q^t} V_{U^t Q^t}^\top) (Z^* - \mathcal{W} U^t (\mathcal{W} U^t)^\top) \right) V_{U^t Q^t} \\ &= (V_{Z^*}^\perp)^\top V_{U^t Q^t} - \frac{\alpha}{m} (V_{Z^*}^\perp)^\top V_{U^t Q^t} V_{U^t Q^t}^\top Z^* \\ &\quad - \frac{\alpha}{m} (V_{Z^*}^\perp)^\top (I - V_{U^t Q^t} V_{U^t Q^t}^\top) \mathcal{W} U^t (Q^t (Q^t)^\top + Q^{t,\perp} (Q^{t,\perp})^\top) (\mathcal{W} U^t)^\top V_{U^t Q^t}. \end{aligned}$$

Observe that  $\text{span}(\mathcal{W} U^t Q^t (Q^t)^\top) \subseteq \text{span}(V_{U^t Q^t})$  and consequently,

$$(I - V_{U^t Q^t} V_{U^t Q^t}^\top) \mathcal{W} U^t Q^t = 0.$$

Therefore

$$\begin{aligned} (I) &= (V_{Z^*}^\perp)^\top V_{U^t Q^t} - \frac{\alpha}{m} (V_{Z^*}^\perp)^\top V_{U^t Q^t} V_{U^t Q^t}^\top Z^* \\ &\quad - \frac{\alpha}{m} (V_{Z^*}^\perp)^\top (I - V_{U^t Q^t} V_{U^t Q^t}^\top) \mathcal{W} U^t Q^{t,\perp} (\mathcal{W} U^t Q^{t,\perp})^\top V_{U^t Q^t} \\ &= (V_{Z^*}^\perp)^\top V_{U^t Q^t} \left( I - \frac{\alpha}{m} V_{U^t Q^t}^\top Z^* V_{U^t Q^t} \right) \\ &\quad - \frac{\alpha}{m} (V_{Z^*}^\perp)^\top (I - V_{U^t Q^t} V_{U^t Q^t}^\top) \mathcal{W} U^t Q^{t,\perp} (\mathcal{W} U^t Q^{t,\perp})^\top (V_{Z^*} V_{Z^*}^\top + V_{Z^*}^\perp (V_{Z^*}^\perp)^\top) V_{U^t Q^t}. \end{aligned}$$

Observe that  $V_{Z^*}^\top \mathcal{W} U^t Q^{t,\perp} = V_{Z^*}^\top U^t Q^{t,\perp} = V^t \Sigma^t (Q^t)^\top Q^{t,\perp} = 0$ . Therefore,

$$\|(I)\| \leq \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| \left( 1 - \frac{\alpha}{m} \sigma_{\min}(V_{U^t Q^t}^\top Z^* V_{U^t Q^t}) \right) + \frac{\alpha}{m} \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| \|U^t Q^{t,\perp}\|^2 \quad (80)$$

where  $\sigma_{\min}(V_{U^t Q^t}^\top Z^* V_{U^t Q^t}) = \sigma_{\min}(X) \sigma_{\min}(V_{U^t Q^t}^\top V_{Z^*})$ . By assumption  $\sigma_{\min}(V_{Z^*}^\top V_{U^t Q^t}) \geq 1 - c\kappa^{-1}$  setting  $c \leq \frac{1}{2}$ . We thus obtain

$$\|(I)\| \leq \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| \left( 1 - \frac{\alpha}{2m} \lambda_{r^*}(\bar{Z}^*) m + \frac{\alpha}{m} \|U^t Q^{t,\perp}\|^2 \right) \quad (81)$$

$$\leq \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| \left( 1 - \frac{\alpha}{2} \lambda_{r^*}(\bar{Z}^*) + \alpha c \lambda_{r^*}(\bar{Z}^*) \right). \quad (82)$$



Setting  $c \leq \frac{1}{6}$ ,

$$\|(I)\| \leq \|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\| \left(1 - \frac{\alpha}{3} \lambda_{r^*}(\bar{Z}^*)\right). \quad (83)$$

For (II) it follows in the same way as in [34] (c.f. page 59)

$$\|(II)\| \leq \frac{\alpha}{m} \|\mathcal{A}^* \mathcal{A}(Z^* - U^t(U^t)^\top) - \mathcal{W}(Z^* - U^t(U^t)^\top) \mathcal{W}\|. \quad (84)$$

$$\|(III)\| \leq \left\| (V_{\bar{Z}^*}^\perp)^\top \left( B_3 - \frac{1}{2} V_{U^t Q^t} V_{U^t Q^t}^\top (B_3 + B_3^\top) \right) V_{U^t Q^t} \right\|, \quad (85)$$

with

$$B_3 = U^t Q^{t,\perp} (Q^{t,\perp})^\top Q^{t+1} (V_{U^t Q^t}^\top U^t Q^t (Q^t)^\top Q^{t+1})^{-1} V_{U^t Q^t}^\top \quad (86)$$

$$= U^t Q^{t,\perp} (Q^{t,\perp})^\top Q^{t+1} ((Q^t)^\top Q^{t+1})^{-1} (V_{U^t Q^t}^\top U^t Q^t)^{-1} V_{U^t Q^t}^\top, \quad (87)$$

and

$$\|B_3\| \leq \|U^t Q^{t,\perp}\| \| (Q^{t,\perp})^\top Q^{t+1} \| \| ((Q^t)^\top Q^{t+1})^{-1} \| \| (V_{U^t Q^t}^\top U^t Q^t)^{-1} \|. \quad (88)$$

Using Lemma 13 under which it holds that  $\sigma_{\min}((Q^t)^\top Q^{t+1}) \geq \frac{1}{2}$

$$\|B_3\| \leq \frac{\| (Q^{t,\perp})^\top Q^{t+1} \| \| U^t Q^{t,\perp} \|}{\sigma_{\min}((Q^t)^\top Q^{t+1}) \sigma_{\min}(U^t Q^t)} \leq 4 \| (Q^{t,\perp})^\top Q^{t+1} \| \quad (89)$$

From the same lemma it follows that

$$\| (Q^{t,\perp})^\top Q^{t+1} \| \leq \alpha \left( \frac{\lambda_{r^*}(\bar{Z}^*)}{6400} + \frac{\| U^t Q^t \| \| U^t Q^{t,\perp} \|}{m} \right) \| (V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t} \| \quad (90)$$

$$+ \frac{4\alpha}{m} \|\mathcal{A}^* \mathcal{A}(Z^* - U^t(U^t)^\top) - \mathcal{W}(Z^* - U^t(U^t)^\top) \mathcal{W}\| \quad (91)$$

therefore, using the above and that  $\| U^t Q^{t,\perp} \| \leq c\kappa^{-2} \sqrt{m} \lambda_{r^*}(\bar{Z}^*)$ ,  $\| U^t Q^t \| \leq 3\sqrt{m} \|\bar{X}\|$  we have for  $c \leq \frac{1}{24 \cdot 800}$

$$\|(III)\| \leq 2 \|B_3\| \leq 8 \| (Q^{t,\perp})^\top Q^{t+1} \| \leq \frac{\alpha}{400} \sigma_{\min}^2(\bar{X}) \| (V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t} \| \quad (92)$$

$$+ 32 \frac{\alpha}{m} \|\mathcal{A}^* \mathcal{A}(Z^* - U^t(U^t)^\top) - \mathcal{W}(Z^* - U^t(U^t)^\top) \mathcal{W}\|. \quad (93)$$

Similarly,

$$\|(IV)\| \leq 2 \|B_4\| = \frac{2\alpha}{m} \|\mathcal{M}P\| \leq \frac{2\alpha}{m} \|\mathcal{A}^* \mathcal{A}(Z^* - U^t(U^t)^\top)\| \|B_3\|. \quad (94)$$

We have already obtained a bound on  $\|B_3\|$  meaning that we are to bound the remaining matrix yielding under the lemma's assumptions that

$$\|\mathcal{A}^* \mathcal{A}(Z^* - U^t(U^t)^\top)\| \leq 11c\kappa^{-2},$$

and therefore with  $c \leq \frac{1}{24 \cdot 800}$

$$\|(IV)\| \leq \frac{\alpha}{50} \kappa^{-2} \lambda_{r^*}(\bar{Z}^*) \| (V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t} \| + 352c \frac{\alpha}{m} \|\mathcal{A}^* \mathcal{A}(Z^* - U^t(U^t)^\top) - \mathcal{W}(Z^* - U^t(U^t)^\top) \mathcal{W}\|$$

and

$$\|(V)\| \leq \frac{1}{2} \| (V_{\bar{Z}^*}^\perp)^\top B V_{U^t Q^t} V_{U^t Q^t}^\top (B + B^\top) V_{U^t Q^t} \| \leq \frac{1}{2} \|B\| \|B + B^\top\| \leq \|B\|^2.$$

Observe that

$$\begin{aligned} \|B\| &\leq \left\| \frac{\alpha}{m} \mathcal{M} + P + \frac{\alpha}{m} \mathcal{M}P \right\| \leq \|B_4\| + \frac{\alpha}{m} \|\mathcal{M}\| + \|B_3\| \\ &\leq \left(1 + \frac{\alpha}{m} \|\mathcal{A}^* \mathcal{A}(Z^* - U^t(U^t)^\top) - \mathcal{W}(Z^* - U^t(U^t)^\top) \mathcal{W}\|\right) \|B_3\| + \frac{\alpha}{m} \|\mathcal{M}\| \\ &\leq \frac{\alpha}{800} \lambda_{r^*}(\bar{Z}^*) \| (V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t} \| + 17 \frac{\alpha}{m} \|\mathcal{A}^* \mathcal{A}(Z^* - U^t(U^t)^\top) - \mathcal{W}(Z^* - U^t(U^t)^\top) \mathcal{W}\| \\ &\quad + \frac{\alpha}{m} \|Z^* - (\mathcal{W}U^t)(\mathcal{W}U^t)^\top\|. \end{aligned}$$

Therefore,

$$\begin{aligned}
\|(V)\| &\leq 3 \frac{\alpha^2}{(800)^2} \lambda_{r^*}^2(\bar{Z}^*) \|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\|^2 \\
&\quad + \frac{3(17)^2 \alpha^2}{m^2} \|\mathcal{W}(Z^* - U^t(U^t)^\top) \mathcal{W} - \mathcal{A}^* \mathcal{A}(Z^\top - U^t(U^t)^\top)\|^2 \\
&\quad + \frac{3\alpha^2}{m^2} \|Z^\top - (\mathcal{W}U^t)(\mathcal{W}U^t)^\top\|^2 \leq \frac{3\alpha^2}{m^2} \|Z^* - (\mathcal{W}U^t)(\mathcal{W}U^t)^\top\|^2 \\
&\quad + \frac{3\alpha c \kappa^{-4}}{(800)^2} \lambda_{r^*}(\bar{Z}^*) \|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\|^2 \\
&\quad + \frac{3(17)^2 c \kappa^{-4}}{24 \cdot 800 m} \alpha \|\mathcal{W}(Z^* - U^t(U^t)^\top) \mathcal{W} - \mathcal{A}^* \mathcal{A}(Z^* - U^t(U^t)^\top)\|.
\end{aligned}$$

Similarly as in [34] we have

$$\begin{aligned}
\|(VI)\| &\leq \|(V_{\bar{Z}^*}^\perp)^\top D\| \leq \left\| (V_{\bar{Z}^*}^\perp)^\top (I + B) V_{U^t Q^t} \left( \frac{1}{2} V_{U^t Q^t}^\top B^\top B V_{U^t Q^t} - C \right) \right\| \\
&\leq 2 (\|B\|^2 + \|C\|) \\
\|C\| &\leq 3 \|V_{U^t Q^t}^\top B V_{U^t Q^t} + V_{U^t Q^t}^\top B^\top V_{U^t Q^t} + V_{U^t Q^t}^\top B^\top B V_{U^t Q^t}\|.
\end{aligned}$$

Following [34, Lemma 9.3] there holds that

$$\begin{aligned}
\|(VI)\| &\leq 56 \left( 3 \frac{\alpha^2}{m^2} \|Z^* - (\mathcal{W}U^t)(\mathcal{W}U^t)^\top\|^2 + 3c \frac{\alpha \kappa^{-4}}{2 \cdot 400^2} \lambda_{r^*}(\bar{Z}^*) \|(V_{\bar{Z}^*}^\perp)^\top V_{U^t Q^t}\|^2 \right. \\
&\quad \left. + \frac{3\alpha}{2m} c \kappa^{-4} \|\mathcal{A}^* \mathcal{A}(Z^* - U^t(U^t)^\top) - \mathcal{W}(Z^* - U^t(U^t)^\top) \mathcal{W}\| \right).
\end{aligned}$$

The final results follows readily combining all the above bounds.

We are left to establish that (79) holds.

Observe that by definition the matrix is symmetric. Consequently it suffices to prove that

$$2\|B\| + \|B^\top B\| \leq \frac{1}{2}.$$

From the upper bound on (VI) it follows from the derivations in [34, Lemma 9.3] that

$$\|B\| \leq 10\alpha \|\bar{X}\|^2 + \frac{\alpha}{400} \lambda_{r^*}(\bar{Z}^*) + 33c\alpha \lambda_{r^*}(\bar{Z}^*).$$

Consequently, for the above to hold it suffices to chose  $c$  such that

$$10c\kappa^{-2} + \frac{c}{400} \kappa^{-2} + 33c^2 \kappa^{-2} \leq \frac{1}{6},$$

which follows under the restriction on  $c$  as stated in the lemma.

(v) We write

$$\begin{aligned}
Z^* - U^{t+1}(U^{t+1})^\top &= Z^* - \mathcal{W}^2 U^t (\mathcal{W}^2 U^t)^\top - \frac{\alpha}{m} \mathcal{W}^2 U^t (U^t)^\top (Z^* - \mathcal{W} U^t (\mathcal{W} U^t)^\top) \\
&\quad - \frac{\alpha}{m} (Z^* - \mathcal{W} U^t (\mathcal{W} U^t)^\top) U^t (U^t)^\top \mathcal{W}^2 \\
&\quad + \frac{\alpha}{m} \mathcal{W}^2 U^t (U^t)^\top (\mathcal{W} (Z^* - U^t (U^t)^\top) \mathcal{W} - \mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top)) \\
&\quad + \frac{\alpha}{m} (\mathcal{W} (Z^* - U^t (U^t)^\top) \mathcal{W} - \mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top)) U^t (U^t)^\top \mathcal{W}^2 \\
&\quad - \frac{\alpha^2}{m^2} (\mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top)) U^t (U^t)^\top (\mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top)) \\
&= \mathcal{W} \left( I - \frac{\alpha}{m} \mathcal{W} U^t (U^t)^\top \mathcal{W} \right) (Z^* - U^t (U^t)^\top) \left( I - \frac{\alpha}{m} \mathcal{W} U^t (U^t)^\top \mathcal{W} \right) \mathcal{W} + \mathcal{W} U^t (U^t)^\top \mathcal{W} \\
&\quad - \mathcal{W}^2 U^t (U^t)^\top \mathcal{W}^2 - \frac{\alpha^2}{m^2} \mathcal{W}^2 U^t (U^t)^\top \mathcal{W} (Z^* - U^t (U^t)^\top) \mathcal{W} U^t (U^t)^\top \mathcal{W} \\
&\quad + \frac{\alpha}{m} \mathcal{W}^2 U^t (U^t)^\top (\mathcal{W} (Z^* - U^t (U^t)^\top) \mathcal{W} - \mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top)) \\
&\quad + \frac{\alpha}{m} (\mathcal{W} (Z^* - U^t (U^t)^\top) \mathcal{W} - \mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top)) U^t (U^t)^\top \mathcal{W}^2 \\
&\quad - \frac{\alpha^2}{m^2} \mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top) U^t (U^t)^\top \mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top).
\end{aligned}$$

We are interested in bounding the following terms

$$\begin{aligned}
&V_{Z^*}^\top (Z^* - U^{t+1}(U^{t+1})^\top) = \\
&\underbrace{V_{Z^*}^\top \left( I - \frac{\alpha}{m} U^t (U^t)^\top \right) (Z^* - \mathcal{W} U^t (\mathcal{W} U^t)^\top) \left( I - \frac{\alpha}{m} U^t (U^t)^\top \right) \mathcal{W}}_{\triangleq (I)} \\
&\quad + \underbrace{\frac{\alpha}{m} V_{Z^*}^\top (Z^* - \mathcal{W} U^t (\mathcal{W} U^t)^\top) U^t (U^t)^\top (\mathcal{W} - \mathcal{W}^2)}_{\triangleq (II)} \\
&\quad - \underbrace{\frac{\alpha^2}{m^2} V_{Z^*}^\top U^t (U^t)^\top \mathcal{W} (Z^* - U^t (U^t)^\top) \mathcal{W} U^t (U^t)^\top \mathcal{W}}_{\triangleq (III)} \\
&\quad + \underbrace{\frac{\alpha}{m} V_{Z^*}^\top U^t (U^t)^\top (\mathcal{W} (Z^* - U^t (U^t)^\top) \mathcal{W} - \mathcal{A}^* \mathcal{A} (X X^\top - U^t (U^t)^\top))}_{\triangleq (IV)} \\
&\quad + \underbrace{\frac{\alpha}{m} V_{Z^*}^\top ((\mathcal{W} (Z^* - U^t (U^t)^\top) \mathcal{W} - \mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top)) U^t (U^t)^\top) \mathcal{W}^2}_{\triangleq (V)} \\
&\quad - \underbrace{\frac{\alpha^2}{m^2} V_{Z^*}^\top \mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top) U^t (U^t)^\top \mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top)}_{\triangleq (VI)}.
\end{aligned}$$

We begin with term (I):

$$\begin{aligned}
(I) &= V_{Z^*}^\top \left( I - \frac{\alpha}{m} U^t (U^t)^\top \right) (Z^* - (U^t)(U^t)^\top) \left( I - \frac{\alpha}{m} U^t (U^t)^\top \right) \mathcal{W} \\
&= V_{Z^*}^\top \left( I - \frac{\alpha}{m} U^t (U^t)^\top \right) V_{Z^*} V_{Z^*}^\top (Z^* - (U^t)(U^t)^\top) \left( I - \frac{\alpha}{m} U^t (U^t)^\top \right) \mathcal{W} \\
&+ V_{Z^*}^\top \left( I - \frac{\alpha}{m} U^t (U^t)^\top \right) (V_{Z^*}^\perp (V_{Z^*}^\perp)^\top) (Z^* - (U^t)(U^t)^\top) \left( I - \frac{\alpha}{m} U^t (U^t)^\top \right) \mathcal{W} \\
&= \underbrace{V_{Z^*}^\top \left( I - \frac{\alpha}{m} U^t (U^t)^\top \right) V_{Z^*} V_{Z^*}^\top (Z^* - (U^t)(U^t)^\top) \left( I - \frac{\alpha}{m} U^t (U^t)^\top \right) \mathcal{W}}_{\triangleq (I,a)} \\
&- \underbrace{V_{Z^*}^\top \left( I - \frac{\alpha}{m} U^t (U^t)^\top \right) (V_{Z^*}^\perp (V_{Z^*}^\perp)^\top) U^t (U^t)^\top \left( I - \frac{\alpha}{m} U^t (U^t)^\top \right) \mathcal{W}}_{\triangleq (I,b)}.
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
\| (I, a) \| &\leq \left\| I - \frac{\alpha}{m} V_{Z^*}^\top U^t (U^t)^\top V_{Z^*} \right\| \left\| V_{Z^*}^\top (Z^* - U^t (U^t)^\top) \right\| \left\| \left( I - \frac{\alpha}{m} U^t (U^t)^\top \right) \mathcal{W} \right\| \\
&\leq \left( 1 - \frac{\alpha}{m} \sigma_{\min}^2 (V_{Z^*}^\top U^t Q^t) \right) \left\| V_{Z^*}^\top (Z^* - U^t (U^t)^\top) \right\|.
\end{aligned}$$

By assumption  $\sigma_{\min}^2 (U^t Q^t) \geq \frac{m}{10} \lambda_{r^*} (\bar{Z}^*)$ , which leads to

$$\| (I, a) \| \leq \left( 1 - \frac{\alpha}{20} \lambda_{r^*} (\bar{Z}^*) \right) \left\| V_{Z^*}^\top (Z^* - U^t (U^t)^\top) \right\|.$$

We now proceed with the quantity

$$\begin{aligned}
\| (I, b) \| &\leq \frac{\alpha}{m} \left\| V_{Z^*}^\top U^t (U^t)^\top \mathcal{W} (V_{Z^*}^\perp (V_{Z^*}^\perp)^\top) U^t (U^t)^\top \right\| \\
&= \frac{\alpha}{m} \left\| V_{Z^*}^\top U^t Q^t (U^t Q^t)^\top \mathcal{W} (V_{Z^*}^\perp (V_{Z^*}^\perp)^\top) U^t (U^t)^\top \right\| \\
&\leq \frac{\alpha}{m} \left\| V_{Z^*}^\top U^t Q^t \right\| \left\| (V_{Z^*}^\perp)^\top \mathcal{W} U^t Q^t \right\| \left\| (V_{Z^*}^\perp)^\top U^t (U^t)^\top \right\|,
\end{aligned}$$

where because  $Z^* \mathcal{W} = \mathcal{W} Z^*$ , there holds

$$\| (V_{Z^*}^\perp)^\top \mathcal{W} U^t Q^t \| \leq \| (V_{Z^*}^\perp)^\top U^t Q^t \|.$$

We can now use the same procedure as in the proof of [34, Lemma 9.5] and obtain, under  $c \leq \frac{1}{9 \cdot 400}$ , that

$$\| (I, b) \| \leq \frac{\lambda_{r^*} (\bar{Z}^*)}{100} \left\| V_{Z^*}^\top (Z^* - U^t (U^t)^\top) \right\| + \frac{\lambda_{r^*} (\bar{Z}^*)}{400} \left\| U^t Q^{t,\perp} (U^t Q^{t,\perp})^\top \right\|.$$

Proceeding to (II), we have

$$\begin{aligned}
\| (II) \| &\leq \frac{\alpha}{m} \left\| V_{Z^*}^\top (Z^* - U^t (U^t)^\top) \right\| \left\| U^t (U^t) (\mathcal{W} - \mathcal{W}^2) \right\| \\
&\leq \frac{\alpha 9 \|\bar{X}\|^2 m}{m} \|\mathcal{W} - \mathcal{W}^2\| \left\| V_{Z^*}^\top (Z^* - U^t (U^t)^\top) \right\| \leq 9\alpha \|\bar{X}\|^2 \rho \left\| V_{Z^*}^\top (Z^* - U^t (U^t)^\top) \right\|.
\end{aligned}$$

Combining

$$\begin{aligned}
\| I \| + \| II \| &\leq \left( 1 - \frac{\alpha \lambda_{r^*} (\bar{Z}^*)}{40} + 9\alpha \|\bar{X}\|^2 \rho \right) \left\| V_{Z^*}^\top (Z^* - U^t (U^t)^\top) \right\| \\
&+ \frac{\alpha \lambda_{r^*} (\bar{Z}^*)}{400} \left\| U^t Q^{t,\perp} (U^t Q^{t,\perp})^\top \right\|.
\end{aligned}$$

Following the proof of [34, Lemma 9.5], we obtain, under  $c \leq \frac{1}{81000}$ ,

$$\| (III) \| \leq \alpha \lambda_{r^*} (\bar{Z}^*) \left( \frac{1}{200} \left\| V_{Z^*}^\top (Z^* - U^t (U^t)^\top) \right\| + \frac{1}{1000} \left\| U^t Q^{t,\perp} (U^t Q^{t,\perp})^\top \right\| \right).$$

Similarly, we have

$$\| (IV) \| + \| (V) \| \leq \frac{\alpha \lambda_{r^*}(\bar{Z}^*)}{50} (\| V_{Z^*}^\top (U^t (U^t)^\top - Z^*) \| + \| U^t Q^{t,\perp} (U^t Q^{t,\perp})^\top \|).$$

Lastly,

$$\begin{aligned} \| (VI) \| &\leq \frac{\alpha^2}{m^2} \| U^t \|^2 \| \mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)) \| \leq \frac{9\alpha^2}{m} \| \bar{X} \|^2 c^2 \kappa^{-4} \| Z^* - U^t (U^t) \|^2 \\ &\leq \frac{9\alpha^2}{m} \| \bar{X} \|^2 c \kappa^{-4} \| Z^* - U^t (U^t)^\top \| (\| X \|^2 + \| U^t \|^2) \\ &\leq \frac{9}{m} \| \bar{X} \|^2 10m \| \bar{X} \|^2 c \kappa^{-4} \| Z^* - U^t (U^t)^\top \| \\ &\leq 90\alpha \lambda_{r^*}^2(\bar{Z}^*) \| Z^* - U^t (U^t)^\top \|. \end{aligned}$$

Using [34, Lemma B.4], yields

$$\| (VI) \| \leq 90\alpha \lambda_{r^*}^2(\bar{Z}^*) (4 \| V_{Z^*}^\top (Z^* - U^t (U^t)^\top) \| + \| U^t Q^{t,\perp} (U^t Q^{t,\perp})^\top \|).$$

Combining all bounds, we finally obtain

$$\begin{aligned} \| V_{Z^*}^\top (Z^* - U^{t+1} (U^{t+1})^\top) \| &\leq \left( 1 - \frac{\alpha}{200} \lambda_{r^*}(\bar{Z}^*) + 9\alpha \| \bar{X} \|^2 \rho \right) \| V_{Z^*}^\top (Z^* - U^t (U^t)^\top) \| \\ &\quad + \alpha \frac{\lambda_{r^*}(\bar{Z}^*)}{100} \| U^t Q^{t,\perp} (U^t Q^{t,\perp})^\top \|. \end{aligned}$$

## F.2 Technical lemmas: auxiliary to proving Lemma 9

**Lemma 12.** Assume that  $\alpha \leq \alpha \min\{\|\bar{Z}^*\|^{-1}, m\|\mathcal{W}\Delta^t\mathcal{W} - \mathcal{A}^*\mathcal{A}(\Delta^t)\|\}$  for some  $c \in (0, 1/1200)$  and that  $\|U^t\| \leq 3\sqrt{m}\|\bar{X}\|$ . Moreover assume that  $V_{Z^*}^\top U^{t+1} Q^t$  is full rank. Then, if  $\|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| \leq c\kappa^{-1}$  there holds that

$$\|(V_{Z^*}^\perp)^\top V_{U^{t+1} Q^t}\| \leq 2 \left( \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| + \frac{\alpha}{m} \|\mathcal{A}^* \mathcal{A}(\Delta^t)\| \right) \quad (95)$$

$$\|(V_{Z^*}^\perp)^\top V_{U^{t+1} Q^t}\| \leq \frac{1}{50}. \quad (96)$$

*Proof.*

$$U^{t+1} Q^t = \left( \mathcal{W}^2 + \frac{\alpha}{m} \mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top) \right) U^t Q^t, \quad (97)$$

and denote by  $V_{U^t Q^t} \Sigma_{U^t Q^t} Q_{U^t Q^t}^\top$  the singular value decomposition of  $U^t Q^t$ . Denote by

$$Y \triangleq \left( \mathcal{W}^2 + \frac{\alpha}{m} \mathcal{A}^* \mathcal{A} (X X^\top - U^t (U^t)^\top) \right) V_{U^t Q^t}. \quad (98)$$

By assumption the matrix  $\Sigma_{U^t Q^t} Q_{U^t Q^t}^\top$  is full rank and the matrix  $Y = V_Y \Sigma_Y Q_Y^\top$  has the same column space as  $U^{t+1} Q^t$ . It follows that

$$\|(V_{Z^*}^\perp)^\top V_{U^{t+1} Q^t}\| = \|(V_{Z^*}^\perp)^\top V_Y\| \leq \|(V_{Z^*}^\perp)^\top V_Y \Sigma_Y Q_Y^\top\| \|(\Sigma_Y Q_Y^\top)^{-1}\| = \frac{\|(V_{Z^*}^\perp)^\top Y\|}{\sigma_{\min}(Y)} \quad (99)$$

Then,

$$\sigma_{\min}(Y) \geq \sigma_{\min}(V_{Z^*}^\top Y) = \sigma_{\min}(V_{Z^*}^\top V_{U^t Q^t}) - \frac{\alpha}{m} \| (V_{Z^*}^\top (\mathcal{A}^* \mathcal{A} (Z^* - U^t (U^t)^\top)) V_{U^t Q^t} \| \quad (100)$$

Observe that by assumption  $\|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| \leq c\kappa^{-1}$  if  $c < 1$  there holds that  $\sigma_{\min}(V_{Z^*}^\top V_{U^t Q^t}) \geq 1 - c\kappa^{-1}$  implying that the matrix  $V_{Z^*}^\top V_{U^t Q^t}$  is of rank  $r^*$  (full-rank). Observe that there holds that

$$\|\mathcal{A}^* \mathcal{A}(\Delta^t)\| \leq \|\mathcal{A}^* \mathcal{A}(\Delta^t) - \mathcal{W}\Delta^t\mathcal{W}\| + \|Z^* - \mathcal{W}U^t(\mathcal{W}U^t)^\top\| \quad (101)$$

$$\frac{\alpha}{m} \|\mathcal{A}^* \mathcal{A}(\Delta^t)\| \leq c + \frac{\alpha}{m} \|Z^*\| + \alpha 9 \|\bar{X}\|^2 \leq 11c \quad (102)$$

where we have used that

$$\alpha \leq c \min\{\|\bar{X}\|^{-2}, m\|(\mathcal{W}^2 - \mathcal{A}^* \mathcal{A})(\Delta^t)\|^{-1}\} \quad (103)$$

and

$$\|U^t\| \leq 3\sqrt{m}\|\bar{X}\|. \quad (104)$$

Thus, there holds

$$\sigma_{\min}(Y) \geq 1 - c\kappa^{-1} - 11c. \quad (105)$$

Under the assumption that  $c \leq \frac{1}{24}$  there holds that

$$\sigma_{\min}(Y) \geq \frac{1}{2}. \quad (106)$$

In this way we Further,

$$\|(V_{Z^*}^\perp)^\top Y\| \leq \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| + \frac{\alpha}{m} \|\mathcal{A}^* \mathcal{A}(\Delta^t)\|. \quad (107)$$

Observe that by construction

$$\|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| = \|(V_{Z^*}^\perp)^\top V_{U^{t+1} Q^t}\|, \quad (108)$$

and therefore

$$\|(V_{Z^*}^\perp)^\top V_{U^{t+1} Q^t}\| \leq 2 \left( \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| + \frac{\alpha}{m} \|\mathcal{A}^* \mathcal{A}(\Delta^t)\| \right). \quad (109)$$

Similarly,

$$\|(V_{Z^*}^\perp)^\top Y\| \leq \|(V_{Z^*}^\perp)^\top \mathcal{W}^2 V_{U^t Q^t}\| + \frac{\alpha}{m} \|\mathcal{A}^* \mathcal{A}(\Delta^t)\| \quad (110)$$

where using that  $\mathcal{W}^2 Z^* = Z^* \mathcal{W}^2$  we obtain

$$\|(V_{Z^*}^\perp)^\top \mathcal{W}^2 V_{U^t Q^t}\| \leq \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\|, \quad (111)$$

consequently,

$$\|(V_{Z^*}^\perp)^\top V_{U^{t+1} Q^t}\| \leq \frac{\|(V_{Z^*}^\perp)^\top Y\|}{\sigma_{\min}(Y)} \leq 2 \left( \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| + \frac{\alpha}{m} \|\mathcal{A}^* \mathcal{A}(Z^* - U^t (U^t)^\top)\| \right) \quad (112)$$

We can further bound

$$\frac{\alpha}{m} \|\mathcal{A}^* \mathcal{A}(Z^* - U^t (U^t)^\top)\| \leq \frac{\alpha}{m} \|Z^* - U^t (U^t)^\top\| + \frac{\alpha}{m} \|\mathcal{A}^* \mathcal{A}(\Delta^t) - \mathcal{W} \Delta^t \mathcal{W}\| \quad (113)$$

$$\leq \frac{\alpha}{m} (m\|\bar{X}\|^2 + 9m\|\bar{X}\|^2) + c \leq 11c. \quad (114)$$

Consequently,

$$\|(V_{Z^*}^\perp)^\top V_{U^{t+1} Q^t}\| \leq 2c\kappa^{-1} + 22c \leq \frac{1}{50} \quad (115)$$

under the assumption that  $c < \frac{1}{1200}$ .

□

**Lemma 13.** Assume that  $\|U^t Q^{t,\perp}\| \leq 2\sigma_{\min}(U^t Q^t)$  and  $\|U^t\| \leq 3\sqrt{m}\|\bar{X}\|$  holds. Further, assume that

$$\|\mathcal{W} \Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t)\| \leq c m \sigma_{\min}^2(\bar{X}) \quad (116)$$

$$\|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| \leq c \quad (117)$$

$$\alpha \leq c\kappa^{-1} \|\bar{Z}^*\|^{-1} \quad (118)$$

$$\|U^t Q^{t,\perp}\| \leq c\kappa^{-2} \sqrt{m} \|\bar{X}\| \quad (119)$$

for some  $c \leq \frac{1}{240}$ . Then, there holds that

$$\|(Q^{t,\perp})^\top Q^{t+1}\| \leq \alpha \left( \frac{\lambda_{r^*}(\bar{Z}^*)}{6400} + \frac{\|U^t Q^t\| \|U^t Q^{t,\perp}\|}{m} \right) \|(V_{Z^*}^\perp)^\top V_{U^t Q^t}\| + 4 \frac{\alpha}{m} \|\mathcal{W} \Delta^t \mathcal{W} - \mathcal{A}^* \mathcal{A}(\Delta^t)\| \quad (120)$$

and

$$\sigma_{\min}((Q^{t,\perp})^\top Q^{t+1}) \geq 1/2 \quad (121)$$

*Proof.* Follows identically to that in [34]. □

## G Proofs of the RIP Properties and Consequences

### G.1 Proof of Lemma 1

By definition of the map  $\mathcal{A}$  we have

$$\|\mathcal{A}(Z)\|_2^2 = \frac{1}{mn} \sum_{\ell=1}^N \left( \left\langle \left( mw_{\mathcal{V}(\ell)} w_{\mathcal{V}(\ell)}^\top - \frac{1}{m} \mathbf{1}_m \mathbf{1}_m^\top \right) \otimes A_\ell, Z \right\rangle + m \langle A_\ell, \bar{Z} \rangle \right)^2 \quad (122)$$

$$\stackrel{(a)}{\leq} (1 + \varepsilon_u) m^2 \|\bar{\mathcal{A}}(\bar{Z})\|_2^2 + \underbrace{\left( 1 + \frac{1}{\varepsilon_u} \right) \frac{1}{mn} \sum_{\ell=1}^N \left( \left\langle A_\ell, \sum_{j=1}^m \sum_{k=1}^m \left( mw_{\mathcal{V}(\ell)j} w_{\mathcal{V}(\ell)k} - \frac{1}{m} \right) [Z]_{jk} \right\rangle \right)^2}_{\triangleq A}, \quad (123)$$

where in (a) we use Young's inequality for any  $\varepsilon_u > 0$ . Similarly, for any  $\varepsilon_\ell \in (0, 1)$

$$\|\mathcal{A}(Z)\|_2^2 \geq (1 - \varepsilon_\ell) m^2 \|\bar{\mathcal{A}}(\bar{Z})\|_2^2 + \left( 1 - \frac{1}{\varepsilon_\ell} \right) A. \quad (124)$$

We now proceed by upper bounding  $A$ . Because the matrix  $W$  is doubly stochastic we can write

$$A = \frac{1}{mn} \sum_{\ell=1}^N \left( \left\langle A_\ell, \sum_{j=1}^m \sum_{k=1}^m \left( mw_{\mathcal{V}(\ell)j} w_{\mathcal{V}(\ell)k} - \frac{1}{m} \right) ([Z]_{jk} - \bar{Z}) \right\rangle \right)^2. \quad (125)$$

consequently, we write

$$A \leq \frac{1}{mn} \sum_{\ell=1}^N \left( \sum_{j=1}^m \sum_{k=1}^m \left| mw_{\mathcal{V}(\ell)j} w_{\mathcal{V}(\ell)k} - \frac{1}{m} \right| \left| \langle A_\ell, [Z]_{jk} - \bar{Z} \rangle \right| \right)^2. \quad (126)$$

Because  $w_{ij} \geq 0$  for all  $i$  and  $j$  there holds

$$\max_{i,l,k} \left| w_{il} w_{ik} m - \frac{1}{m} \right| \leq m \max_{ik} \left| w_{ik}^2 - \frac{1}{m^2} \right| = m \max_{ik} \left( w_{ik} + \frac{1}{m} \right) \left| w_{ik} - \frac{1}{m} \right| \quad (127)$$

$$\leq m \left( 1 + \frac{1}{m} \right) \|W - J\|_\infty \leq 2\sqrt{m}\rho. \quad (128)$$

Thus,

$$A \leq \frac{1}{mn} \sum_{\ell=1}^N 4m^3 \rho^2 \left( \sum_{j=1}^m \sum_{k=1}^m \left| \langle A_\ell, [Z]_{jk} - \bar{Z} \rangle \right| \right)^2 \quad (129)$$

$$\leq \frac{1}{mn} \sum_{\ell=1}^N \sum_{j=1}^m \sum_{k=1}^m 4m^5 \rho^2 \left( \langle A_\ell, [Z]_{jk} - \bar{Z} \rangle \right)^2 \leq 4m^5 \rho^2 \sum_{j=1}^m \sum_{k=1}^m \|\bar{\mathcal{A}}([Z]_{jk} - \bar{Z})\|_2^2. \quad (130)$$

Using that  $\bar{\mathcal{A}}$  fulfills the  $(\delta_{2r}, 2r)$  RIP we obtain

$$\|\mathcal{A}(Z)\|_2^2 \leq (1 + \varepsilon_u) (1 + \delta_{2r}) m^2 \|\bar{Z}\|_F^2 + \left( 1 + \frac{1}{\varepsilon_u} \right) 4m^5 \rho^2 (1 + \delta_{2r}) \|Z - \mathcal{J}Z\|_F^2 \quad (131)$$

$$\|\mathcal{A}(Z)\|_2^2 \geq (1 - \varepsilon_\ell) (1 - \delta_{2r}) m^2 \|\bar{Z}\|_F^2 - \left( \frac{1}{\varepsilon_\ell} - 1 \right) 4m^5 \rho^2 (1 + \delta_{2r}) \|Z - \mathcal{J}Z\|_F^2. \quad (132)$$

Set  $\varepsilon_u = \varepsilon_\ell = \frac{\delta_{2r}}{1 + \delta_{2r}}$ , then

$$(1 + \varepsilon_u)(1 + \delta_{2r}) = 1 + 2\delta_{2r} \quad (133)$$

$$(1 - \varepsilon)(1 - \delta_{2r}) \geq 1 - 2\delta_{2r}, \quad (134)$$

and

$$\left(1 + \frac{1}{\varepsilon_u}\right) = \frac{1 + 2\delta_{2r}}{\delta_{2r}} \quad (135)$$

$$\left(\frac{1}{\varepsilon_\ell} - 1\right) = \frac{1}{\delta_{2r}} \leq \frac{1 + 2\delta_{2r}}{\delta_{2r}}, \quad (136)$$

and consequently

$$\|\mathcal{A}(Z)\|_2^2 \leq (1 + 2\delta_{2r}) \|\mathcal{J}Z\mathcal{J}\|_F^2 + 4m^5 \rho^2 \frac{1 + 2\delta_{2r}}{\delta_{2r}} (1 + \delta_{2r}) \|Z - \mathcal{J}Z\mathcal{J}\|^2 \quad (137)$$

$$\|\mathcal{A}(Z)\|_2^2 \geq (1 - 2\delta_{2r}) \|\mathcal{J}Z\mathcal{J}\|_F^2 - 4m^5 \rho^2 \frac{1 + 2\delta_{2r}}{\delta_{2r}} (1 + \delta_{2r}) \|Z - \mathcal{J}Z\mathcal{J}\|^2. \quad (138)$$

## G.2 Proof of Lemma 2

due to linearity of the operators we have

$$\|\mathcal{W}Z\mathcal{W} - \mathcal{A}^* \mathcal{A}(Z)\| \leq \underbrace{\|\mathcal{J}Z\mathcal{J} - \mathcal{A}^* \mathcal{A}(\mathcal{J}Z\mathcal{J})\|}_{\triangleq (I)} + \underbrace{\|\mathcal{W}(Z - \mathcal{J}Z\mathcal{J})\mathcal{W} - \mathcal{A}^* \mathcal{A}(Z - \mathcal{J}Z\mathcal{J})\|}_{\triangleq (II)}. \quad (139)$$

We start off with (I) which can be rewritten as

$$(I) = \max_{v: \|v\|_2 \leq 1} \langle \mathcal{J}vv^\top \mathcal{J}, \mathcal{J}Z\mathcal{J} \rangle - \langle \mathcal{A}(vv^\top), \mathcal{A}(\mathcal{J}Z\mathcal{J}) \rangle \quad (140)$$

$$= \max_{v: \|v\|_2 \leq 1} \left( \frac{1}{4} \|\mathcal{J}(Z + vv^\top)\mathcal{J}\|_F^2 - \frac{1}{4} \|\mathcal{J}(Z - vv^\top)\mathcal{J}\|_F^2 - \frac{1}{4} \|\mathcal{A}(vv^\top + \mathcal{J}Z\mathcal{J})\|_F^2 \right) \quad (141)$$

$$+ \frac{1}{4} \|\mathcal{A}(vv^\top - \mathcal{J}Z\mathcal{J})\|_F^2 \leq \max_{v: \|v\|_2 \leq 1} \left( \frac{1}{4} \|\mathcal{J}(Z + vv^\top)\mathcal{J}\|_F^2 - \frac{1}{4} \|\mathcal{J}(Z - vv^\top)\mathcal{J}\|_F^2 \right) \quad (142)$$

$$- \frac{(1 + 2\delta_{2(r+1)})}{4} \|\mathcal{J}(Z + vv^\top)\mathcal{J}\|_F^2 + \frac{1 - 2\delta_{2(r+1)}}{4} \|\mathcal{J}(Z - vv^\top)\mathcal{J}\|_F^2 \quad (143)$$

$$+ \frac{\hat{\Delta}_{2(r+1)}}{2} \|vv^\top - \mathcal{J}vv^\top \mathcal{J}\|_F^2 \leq \max_{v: \|v\|_2 \leq 1} (\delta_{2(r+1)} \|\mathcal{J}(Z + vv^\top)\mathcal{J}\|_F^2 \quad (144)$$

$$- \delta_{2(r+1)} \|\mathcal{J}(Z - vv^\top)\mathcal{J}\|_F^2 + \frac{\hat{\Delta}_{2(r+1)}}{2} \|vv^\top - \mathcal{J}vv^\top \mathcal{J}\|_F^2) \quad (145)$$

$$= \max_{v: \|v\|_2 \leq 1} \left( \delta_{2(r+1)} (\|\mathcal{J}Z\mathcal{J}\|_F^2 + \|\mathcal{J}vv^\top \mathcal{J}\|_F^2) + \frac{\hat{\Delta}_{2(r+1)}}{2} \|vv^\top - \mathcal{J}vv^\top \mathcal{J}\|_F^2 \right) \quad (146)$$

where in (a) we have used Lemma 1. Under the assumption that  $\hat{\Delta}_{2(r+1)} \leq \delta_{2(r+1)}$  there holds

$$(I) \leq \delta_{2(r+1)} \|\mathcal{J}Z\mathcal{J}\|_F^2 + \delta_{2(r+1)} \stackrel{(a)}{\leq} 2\delta_{2(r+1)}, \quad (147)$$

where we have used in (a) that  $\|\mathcal{J}Z\mathcal{J}\|_f^2 \leq 1$ . Consequently, due to the linearity of the operators involved, there holds for any  $Z$  fulfilling the conditions of the lemma that

$$\|\mathcal{W}Z\mathcal{W} - \mathcal{A}^* \mathcal{A}(Z)\| \leq 2\delta_{2(r+1)} \|Z\|_F. \quad (148)$$

We now proceed with term (II) and write

$$(II) = \max_{v: \|v\|_2 \leq 1} \left( \underbrace{\langle vv^\top, \mathcal{W}(Z - \mathcal{J}Z\mathcal{J})\mathcal{W} \rangle}_{\triangleq (II.a)} - \underbrace{\langle \mathcal{A}(vv^\top), \mathcal{A}(Z - \mathcal{J}Z\mathcal{J}) \rangle}_{\triangleq (II.b)} \right), \quad (149)$$



where

$$- (II.b) = -\langle \mathcal{A}(vv^\top - \mathcal{J}vv^\top \mathcal{J}), \mathcal{A}(Z - \mathcal{J}Z\mathcal{J}) \rangle - \langle \mathcal{A}(\mathcal{J}vv^\top \mathcal{J}), \mathcal{A}(Z - \mathcal{J}Z\mathcal{J}) \rangle \quad (150)$$

$$\leq \|\mathcal{A}(vv^\top - \mathcal{J}vv^\top \mathcal{J})\| \|\mathcal{A}(Z - \mathcal{J}Z\mathcal{J})\| + \|\mathcal{A}(\mathcal{J}vv^\top \mathcal{J})\| \|\mathcal{A}(Z - \mathcal{J}Z\mathcal{J})\| \quad (151)$$

$$\stackrel{(a)}{\leq} \hat{\Delta}_{2(r+1)} \|vv^\top - \mathcal{J}vv^\top \mathcal{J}\|_F \|Z - \mathcal{J}Z\mathcal{J}\|_F \quad (152)$$

$$+ \sqrt{(1 + 2\delta_{2(r+1)})\hat{\Delta}_{2(r+1)}} \|\mathcal{J}vv^\top \mathcal{J}\|_F \|Z - \mathcal{J}Z\mathcal{J}\|_F \leq \frac{\hat{\Delta}_{2(r+1)}}{2} \|vv^\top - \mathcal{J}vv^\top \mathcal{J}\|_F^2 \quad (153)$$

$$+ \frac{\hat{\Delta}_{2(r+1)}}{2} \|Z - \mathcal{J}Z\mathcal{J}\|_F^2 + \sqrt{(1 + 2\delta_{2(r+1)})\hat{\Delta}_{2(r+1)}} \|\mathcal{J}vv^\top \mathcal{J}\|_F \|Z - \mathcal{J}Z\mathcal{J}\|_F, \quad (154)$$

where in (a) we have invoked Lemma 1 and

$$(II.a) = \langle vv^\top, \mathcal{W}Z(\mathcal{I} - \mathcal{J})\mathcal{W} + \mathcal{J}Z(\mathcal{I} - \mathcal{J})\mathcal{W} \rangle \leq 2\rho \|(\mathcal{I} - \mathcal{J})Z\|_F \leq 2\rho \|Z - \mathcal{J}Z\mathcal{J}\|_F. \quad (155)$$

combining

$$(II) \leq 2\rho \|Z - \mathcal{J}Z\mathcal{J}\|_F + (1 + 2\delta_{2(r+1)})\hat{\Delta}_{2(r+1)} \|Z - \mathcal{J}Z\mathcal{J}\|_F^2 \quad (156)$$

$$+ \max_{v: \|v\|_2 \leq 1} \left( \frac{\hat{\Delta}_{2(r+1)}}{2} \|vv^\top - \mathcal{J}vv^\top \mathcal{J}\|_F^2 + \frac{(1 + 2\delta_{2(r+1)})\hat{\Delta}_{2(r+1)}}{2} \|\mathcal{J}vv^\top \mathcal{J}\|_F^2 \right). \quad (157)$$

Under the assumption that  $\|Z - \mathcal{J}Z\mathcal{J}\|_F^2 \leq 1$  there holds

$$(II) \leq 2\rho + 2(1 + 2\delta_{2(r+1)})\hat{\Delta}_{2(r+1)}, \quad (158)$$

and using the linearity of the operator, there holds for all matrices fulfilling the constraints of the lemma that

$$(II) \leq 2 \left( \rho + (1 + 2\delta_{2(r+1)})\hat{\Delta}_{2(r+1)} \right) \|Z - \mathcal{J}Z\mathcal{J}\|_F. \quad (159)$$

Gathering both results the desired results follows.

### G.3 Proof of Lemma 3

Denote by  $Z = \sum_{i=1}^{md} \lambda_i v_i v_i^\top$

$$\|(\mathcal{W}Z\mathcal{W} - \mathcal{A}^* \mathcal{A})(Z)\|_F \leq \sum_{i=1}^{md} \lambda_i \|\mathcal{W}v_i v_i^\top \mathcal{W} - \mathcal{A}^* \mathcal{A}(v_i v_i^\top)\|_F. \quad (160)$$

Invoking Lemma 1 there holds

$$\|(\mathcal{W}Z\mathcal{W} - \mathcal{A}^* \mathcal{A})(Z)\|_F \leq \quad (161)$$

$$\leq \sum_{i=1}^{md} \lambda_i \left( 2\delta_4 + 2 \left( \rho + (1 + 2\delta_4)\hat{\Delta}_4 \right) \right) = \left( 2\delta_4 + 2 \left( \rho + (1 + 2\delta_4)\hat{\Delta}_4 \right) \right) \|Z\|_*. \quad (162)$$

### G.4 Proof of Lemma 4

$$\|(\mathcal{I} - \mathcal{J})\mathcal{A}^* \mathcal{A}(Z^*)\| = \|\mathcal{A}^* \mathcal{A}(Z^*)(\mathcal{I} - \mathcal{J})\|, \quad (163)$$

and

$$\mathcal{A}^* \mathcal{A}(Z^*) = \frac{1}{mn} \sum_{s=1}^N \left( w_{\mathcal{V}(s)} w_{\mathcal{V}(s)}^\top \right) \otimes A_s \langle A_s, \bar{Z}^* \rangle \quad (164)$$

$$= \mathcal{W} \frac{1}{mn} \sum_{s=1}^N \left( e_{\mathcal{V}(s)} e_{\mathcal{V}(s)}^\top \right) \otimes A_s \langle A_s, \bar{Z}^* \rangle \mathcal{W}. \quad (165)$$

Thus, because  $\mathcal{W}$  is p.s.d. there holds

$$(\mathcal{I} - \mathcal{J})\mathcal{A}^*\mathcal{A}(Z^*) = (\mathcal{W} - \mathcal{J})\frac{1}{mn}\sum_{s=1}^N \left(m^2 e_{\mathcal{V}(s)} e_{\mathcal{V}(s)}^\top\right) \otimes A_s \langle A_s, \bar{Z}^* \rangle \mathcal{W} \quad (166)$$

$$= (\mathcal{W}^{1/2} - \mathcal{J})\frac{1}{mn}\sum_{s=1}^m \left(m^2 w_{\mathcal{V}(s)}^{1/2} \left(w_{\mathcal{V}(s)}^{1/2}\right)^\top\right) \otimes A_s \langle A_s, \bar{Z}^* \rangle \mathcal{W}^{1/2}. \quad (167)$$

Consequently, we have

$$\|(\mathcal{I} - \mathcal{J})\mathcal{A}^*\mathcal{A}(Z^*)\| \leq \rho^{1/2} \left\| \frac{1}{mn} \sum_{s=1}^m \left(m^2 w_{\mathcal{V}(s)}^{1/2} \left(w_{\mathcal{V}(s)}^{1/2}\right)^\top\right) \otimes A_s \langle A_s, \bar{Z}^* \rangle \right\| \quad (168)$$

Under the assumptions of the lemma we can invoke Lemma 2 yielding

$$\|(\mathcal{I} - \mathcal{J})\mathcal{A}^*\mathcal{A}(Z^*)\| \leq \rho^{1/2} (\|Z^*\| + \delta_{2r^*} \|Z^*\|_F) \quad (169)$$

Following the same procedure

$$\|(\mathcal{I} - \mathcal{J})\mathcal{A}^*\mathcal{A}(Z^*)(\mathcal{I} - \mathcal{J})\| \leq \rho (\|Z^*\| + \delta_{2r^*} \|Z^*\|_F). \quad (170)$$