Supplemental material for "Blind Demixing via Wirtinger Flow with Random initialization"

A Establish Approximate State Evolution

A.1 Establishing Approximate State Evolution for Phase 1 of Stage I

We are moving to prove that if the induction hypotheses (41) hold for the t^{th} iteration, then $\alpha_{\boldsymbol{h}_i}$ (21a), $\beta_{\boldsymbol{h}_i}$ (21b), $\alpha_{\boldsymbol{x}_i}$ (20a) and $\beta_{\boldsymbol{x}_i}$ (20b) obey the approximate state evolution (23). This is demonstrated in Lemma 2.

Lemma 2. *Suppose $m \geq Cs^2\mu^2 \max\{K, N\} \log^{10} m$ for some sufficiently large constant $C > 0$. For any $0 \leq t \leq T_1$ (28), if the t^{th} iterate satisfies the induction hypotheses (41) , then for $i = 1, \cdots, s$, with probability at least $1 - c_1 m^{-\nu} - c_1 m e^{-c_2 K}$ for some constants $\nu, c_1, c_2 > 0$, the approximate evolution state (23) holds for some $|\psi_{\boldsymbol{h}_i^t}|, |\psi_{\boldsymbol{x}_i^t}|, |\varphi_{\boldsymbol{h}_i^t}|, |\varphi_{\boldsymbol{x}_i^t}|, |\rho_{\boldsymbol{h}_i^t}|, |\rho_{\boldsymbol{x}_i^t}| \ll 1/\log m$, $i = 1, \cdots, s$.*

Proof. Please refer to Appendix D for details. □

In the sequel, we will prove the hypotheses (41) hold for Phase 1 of Stage I via inductive arguments. Before moving forward, we first investigate the incoherence between $\{\boldsymbol{x}_i^t\}$, $\{\boldsymbol{x}_i^{t,\text{sgn}}\}$ (resp. $\{\boldsymbol{h}_i^t\}$, $\{\boldsymbol{h}_i^{t,\text{sgn}}\}$) and $\{\boldsymbol{a}_{ij}\}$, $\{\boldsymbol{a}_{ij}^{\text{sgn}}\}$ (resp. $\{\boldsymbol{b}_j\}$, $\{\boldsymbol{b}_j^{\text{sgn}}\}$).

Lemma 3. *Suppose that $m \geq Cs^2\mu^2 \max\{K, N\} \log^8 m$ for some sufficiently large constant $C > 0$ and the t^{th} iterate satisfies the induction hypotheses (41) for $t \leq T_0$ (28), then with probability at least $1 - c_1 m^{-\nu} - c_1 m e^{-c_2 K}$ for some constants $\nu, c_1, c_2 > 0$,*

$$\max_{1\leq i\leq s, 1\leq l\leq m} \left|\boldsymbol{a}_{il}^* \widetilde{\boldsymbol{x}}_i^t\right| \cdot \|\widetilde{\boldsymbol{x}}_i^t\|_2^{-1} \lesssim \sqrt{\log m}, \tag{A.1a}$$

$$\max_{1\leq i\leq s, 1\leq l\leq m} \left|\boldsymbol{a}_{il,\perp}^* \widetilde{\boldsymbol{x}}_{i\perp}^t\right| \cdot \|\widetilde{\boldsymbol{x}}_{i\perp}^t\|_2^{-1} \lesssim \sqrt{\log m}, \tag{A.1b}$$

$$\max_{1\leq i\leq s, 1\leq l\leq m} \left|\boldsymbol{a}_{il}^* \check{\boldsymbol{x}}_i^{t,\text{sgn}}\right| \cdot \|\check{\boldsymbol{x}}_i^{t,\text{sgn}}\|_2^{-1} \lesssim \sqrt{\log m}, \tag{A.1c}$$

$$\max_{1\leq i\leq s, 1\leq l\leq m} \left|\boldsymbol{a}_{il,\perp}^* \check{\boldsymbol{x}}_{i\perp}^{t,\text{sgn}}\right| \cdot \|\check{\boldsymbol{x}}_{i\perp}^{t,\text{sgn}}\|_2^{-1} \lesssim \sqrt{\log m}, \tag{A.1d}$$

$$\max_{1\leq i\leq s, 1\leq l\leq m} \left|\boldsymbol{a}_{il}^{\text{sgn}*} \check{\boldsymbol{x}}_i^{t,\text{sgn}}\right| \cdot \|\check{\boldsymbol{x}}_i^{t,\text{sgn}}\|_2^{-1} \lesssim \sqrt{\log m}, \tag{A.1e}$$

$$\max_{1\leq i\leq s, 1\leq l\leq m} \left|\boldsymbol{b}_l^* \widetilde{\boldsymbol{h}}_i^t\right| \cdot \|\widetilde{\boldsymbol{h}}_i^t\|_2^{-1} \lesssim \frac{\mu}{\sqrt{m}} \log^2 m, \tag{A.2a}$$

$$\max_{1\leq i\leq s, 1\leq l\leq m} \left|\boldsymbol{b}_l^* \check{\boldsymbol{h}}_i^{t,\text{sgn}}\right| \cdot \|\check{\boldsymbol{h}}_i^{t,\text{sgn}}\|_2^{-1} \lesssim \frac{\mu}{\sqrt{m}} \log^2 m, \tag{A.2b}$$

$$\max_{1\leq i\leq s, 1\leq l\leq m} \left|\boldsymbol{b}_l^{\text{sgn}*} \check{\boldsymbol{h}}_i^{t,\text{sgn}}\right| \cdot \|\check{\boldsymbol{h}}_i^{t,\text{sgn}}\|_2^{-1} \lesssim \frac{\mu}{\sqrt{m}} \log^2 m. \tag{A.2c}$$

Proof. Based on the induction hypotheses (41), we can prove the claim (A.1) in Lemma 3 by invoking the triangle inequality, Cauchy-Schwarz inequality and standard Gaussian concentration. Furthermore, based on the induction hypotheses (41), the claim (A.2) can be identified according to the definition of the incoherence parameter (9) and the fact $\|\boldsymbol{b}_j\|_2 = \sqrt{K/M}$. □

Now we are ready to specify that the hypotheses (41) hold for $0 \le t \le T_1$ (28). We aim to demonstrate that if the hypotheses (41) hold up to the t^{th} iteration for some $0 \le t \le T_1$, then they hold for the $(t+1)^{\text{th}}$ iteration. Since the case for $t = 0$ can be easily justified due to the equivalent initial points, we mainly focus the inductive step.

Lemma 4. *Suppose the induction hypotheses (41) hold true up to the t^{th} iteration for some $t \le T_1$ (28), then for $i = 1, \cdots, s$, with probability at least $1 - c_1 m^{-\nu} - c_1 m e^{-c_2 K}$ for some constants $\nu, c_1, c_2 > 0$,*

$$\max_{1\le l\le m} \operatorname{dist}\left(\boldsymbol{z}_i^{t+1,(l)}, \widetilde{\boldsymbol{z}}_i^{t+1}\right) \le (\beta_{\boldsymbol{h}_i^{t+1}} + \beta_{\boldsymbol{x}_i^{t+1}}) \left(1 + \frac{1}{s\log m}\right)^{t+1} C_1 \cdot \frac{s\mu^2\kappa\sqrt{\max\{K,N\}\log^8 m}}{m} \tag{A.3}$$

holds $m \ge Cs\mu^2\kappa\sqrt{\max\{K,N\}\log^8 m}$ with some sufficiently large constant $C > 0$ as long as the stepsize $\eta > 0$ obeys $\eta \asymp s^{-1}$ and $C_1 > 0$ is sufficiently large.

In terms of the difference between $\boldsymbol{x}_i^t$ and $\boldsymbol{x}_i^{t,(l)}$ (resp. $\boldsymbol{h}_i^t$ and $\boldsymbol{h}_i^{t,(l)}$) along with the signal direction, i.e., (41b) and (41c), we reach the following lemma.

Lemma 5. *Suppose the induction hypotheses (41) hold true up to the t^{th} iteration for some $t \le T_1$ (28), then with probability at least $1 - c_1 m^{-\nu} - c_1 m e^{-c_2 K}$ for some constants $\nu, c_1, c_2 > 0$,*

$$\max_{1\le l\le m} \operatorname{dist}\left(\boldsymbol{h}_i^{\natural *}\boldsymbol{h}_i^{t+1,(l)}, \boldsymbol{h}_i^{\natural *}\widetilde{\boldsymbol{h}}_i^{t+1}\right) \cdot \|\boldsymbol{h}_i^{\natural}\|_2^{-1} \le \alpha_{\boldsymbol{h}_i^{t+1}} \left(1 + \frac{1}{s\log m}\right)^{t+1} C_2 \frac{s\mu^2\kappa\sqrt{K\log^{13} m}}{m} \tag{A.4}$$

$$\max_{1\le l\le m} \operatorname{dist}\left(x_{i1}^{t+1,(l)}, \widetilde{x}_{i1}^{t+1}\right) \le \alpha_{\boldsymbol{x}_i^{t+1}} \left(1 + \frac{1}{s\log m}\right)^{t+1} C_2 \frac{s\mu^2\kappa\sqrt{N\log^{13} m}}{m} \tag{A.5}$$

holds for some sufficiently large $C_2 > 0$ with $C_2 \gg C_4$, provided that $m \ge Cs\mu^2\kappa \max\{K,N\}\log^{12} m$ for some sufficiently large constant $C > 0$ and the stepsize $\eta > 0$ obeys $\eta \asymp s^{-1}$.

Proof. Please refer to Appendix E for details. □

The next lemma concerns the relation between $\boldsymbol{h}_i^t$ and $\boldsymbol{h}_i^{t,\text{sgn}}$, i.e., (41d), and the relation between $\boldsymbol{x}_i^t$ and $\boldsymbol{x}_i^{t,\text{sgn}}$, i.e., (41e).

Lemma 6. *Suppose the induction hypotheses (41) hold true up to the t^{th} iteration for some $t \le T_1$ (28), then with probability at least $1 - c_1 m^{-\nu} - c_1 m e^{-c_2 K}$ for some constants $\nu, c_1, c_2 > 0$,*

$$\max_{1\le i\le s} \operatorname{dist}\left(\boldsymbol{h}_i^{t+1,\text{sgn}}, \widetilde{\boldsymbol{h}}_i^{t+1}\right) \le \alpha_{\boldsymbol{h}_i^{t+1}} \left(1 + \frac{1}{s\log m}\right)^{t+1} C_3 \sqrt{\frac{s\mu^2\kappa^2 K\log^8 m}{m}} \tag{A.6a}$$

$$\max_{1\le i\le s} \operatorname{dist}\left(\boldsymbol{x}_i^{t+1,\text{sgn}}, \widetilde{\boldsymbol{x}}_i^{t+1}\right) \le \alpha_{\boldsymbol{x}_i^{t+1}} \left(1 + \frac{1}{s\log m}\right)^{t+1} C_3 \sqrt{\frac{s\mu^2\kappa^2 N\log^8 m}{m}} \tag{A.6b}$$

holds for some sufficiently large $C_3 > 0$, provided that $m \ge Cs\mu^2\kappa^2 \max\{K,N\}\log^8 m$ for some sufficiently large constant $C > 0$ and the stepsize $\eta > 0$ obeys $\eta \asymp s^{-1}$.

We still need to characterize the difference $\widetilde{\boldsymbol{h}}_i^t - \widehat{\boldsymbol{h}}_i^{t,(l)} - \widetilde{\boldsymbol{h}}_i^{t,\text{sgn}} + \widehat{\boldsymbol{h}}_i^{t,,\text{sgn},(l)}$ (41f) and the difference $\widetilde{\boldsymbol{x}}_i^t - \widehat{\boldsymbol{x}}_i^{t,(l)} - \widetilde{\boldsymbol{x}}_i^{t,\text{sgn}} + \widehat{\boldsymbol{x}}_i^{t,\text{sgn},(l)}$ (41g) in the following lemma.

Lemma 7. *Suppose the induction hypotheses (41) hold true up to the* t^{th} *iteration for some* $t \le T_1$ *(28), then with probability at least* $1 - c_1 m^{-\nu} - c_1 m e^{-c_2 K}$ *for some constants* $\nu, c_1, c_2 > 0$,

$$\max_{1\le l\le m} \left\| \widetilde{\boldsymbol{h}}_i^{t+1} - \widehat{\boldsymbol{h}}_i^{t+1,(l)} - \widetilde{\boldsymbol{h}}_i^{t+1,\mathrm{sgn}} + \widehat{\boldsymbol{h}}_i^{t+1,\mathrm{sgn},(l)} \right\|_2 \le \alpha_{\boldsymbol{h}_i^{t+1}} \left(1 + \frac{1}{s\log m}\right)^{t+1} C_4 \frac{s\mu^2\sqrt{K\log^{16} m}}{m} \tag{A.7a}$$

$$\max_{1\le l\le m} \left\| \widetilde{\boldsymbol{x}}_i^{t+1,} - \widehat{\boldsymbol{x}}_i^{t+1,,(l)} - \widetilde{\boldsymbol{x}}_i^{t+1,\mathrm{sgn}} + \widehat{\boldsymbol{x}}_i^{t+1,\mathrm{sgn},(l)} \right\|_2 \le \alpha_{\boldsymbol{x}_i^{t+1}} \left(1 + \frac{1}{s\log m}\right)^{t+1} C_4 \frac{s\mu^2\sqrt{N\log^{16} m}}{m} \tag{A.7b}$$

holds for some sufficiently large $C_4 > 0$*, provided that* $m \ge Cs\mu^2 \max\{K, N\}\log^8 m$ *for some sufficiently large constant* $C > 0$ *and the stepsize* $\eta > 0$ *obeys* $\eta \asymp s^{-1}$.

Remark 1. *The arguments applied to prove Lemma 4-Lemma 7 are similar to each other. We thus mainly focus on the proof of (A.5) in Lemma 5 in Appendix E.*

A.2 Establishing Approximate State Evolution for Phase 2 of Stage I

In this subsection, we move to prove that the approximate state evolution (23) holds for $T_1 < t \le T_\gamma$ (T_γ and T_1 are defined in (27) and (28) respectively) via inductive argument. Different from the analysis in Phase 1, only $\{\boldsymbol{z}^{t,(l)}\}$ is sufficient to establish the "near-independence" between iterates and design vectors when the sizes of the signal component follow $\alpha_{\boldsymbol{h}_i^t}$, $\alpha_{\boldsymbol{x}_i} \gtrsim 1/\log m$ in Phase 2 (according to the definition of T_1). As in Phase 1, we begin with specifying the induction hypotheses: for $1 \le i \le s$,

$$\begin{aligned} &\max_{1\le l\le m} \operatorname{dist}\left(\boldsymbol{z}_i^{t,(l)}, \widetilde{\boldsymbol{z}}_i^t\right) \\ &\le (\beta_{\boldsymbol{h}_i^t} + \beta_{\boldsymbol{x}_i^t}) \left(1 + \frac{1}{s\log m}\right)^t C_6 \frac{s\mu^2\kappa\sqrt{\max\{K,N\}\log^{18} m}}{m} \quad \text{(A.8a)} \\ &c_5 \le \left\|\boldsymbol{h}_i^t\right\|_2, \left\|\boldsymbol{x}_i^t\right\|_2 \le C_5, \quad \text{(A.8b)} \end{aligned}$$

From (A.8), we can conclude that one has

$$\max_{1\le i\le s, 1\le l\le m} \left|\boldsymbol{a}_{il}^* \widetilde{\boldsymbol{x}}_i^t\right| \cdot \|\widetilde{\boldsymbol{x}}_i^t\|_2^{-1} \lesssim \sqrt{\log m}, \tag{A.9}$$

$$\max_{1\le i\le s, 1\le l\le m} \left|\boldsymbol{b}_l^* \widetilde{\boldsymbol{h}}_i^t\right| \cdot \|\widetilde{\boldsymbol{h}}_i^t\|_2^{-1} \lesssim \frac{\mu}{\sqrt{m}} \log^2 m, \tag{A.10}$$

with probability at least $1 - c_1 m^{-\nu} - c_1 m e^{-c_2 K}$ for some constants $\nu, c_1, c_2 > 0$ during $T_1 < t \le T_\gamma$ as long as $m \gg Cs\mu^2\kappa K \log^8 m$.

We then move to prove that if the induction hypotheses (41) hold for the t^{th} iteration, then $\alpha_{\boldsymbol{h}_i}$ (21a), $\beta_{\boldsymbol{h}_i}$ (21b), $\alpha_{\boldsymbol{x}_i}$ (20a) and $\beta_{\boldsymbol{x}_i}$ (20b) obey the approximate state evolution (41). This is demonstrated in Lemma 8.

Lemma 8. *Suppose* $m \ge Cs^2\mu^2\kappa^4 \max\{K, N\}\log^{12} m$ *for some sufficiently large constant* $C > 0$*. For any* $T_1 \le t \le T_\gamma$ *(*T_1 *and* T_γ *are defined in (27) and (28) respectively), if the* t^{th} *iterate satisfies the induction hypotheses (41) , then for* $i = 1, \cdots, s$*, with probability at least* $1 - c_1 m^{-\nu} - c_1 m e^{-c_2 K}$ *for some constants* $\nu, c_1, c_2 > 0$*, the approximate evolution state (23) hold for some* $|\psi_{\boldsymbol{h}_i^t}|, |\psi_{\boldsymbol{x}_i^t}|, |\varphi_{\boldsymbol{h}_i^t}|, |\varphi_{\boldsymbol{x}_i^t}|, |\rho_{\boldsymbol{h}_i^t}|, |\rho_{\boldsymbol{x}_i^t}| \ll 1/\log m$*,* $i = 1, \cdots, s$.

It remains to proof the induction step on the difference between leave-one-out sequences $\{\boldsymbol{z}^{t,(l)}\}$ and the original sequences $\{\boldsymbol{z}^t\}$, which is demonstrated in the following lemma.

Lemma 9. *Suppose the induction hypotheses (41) are valid during Phase 1 and the induction hypotheses (A.8) hold true from* T_1^{th} *to the* t^{th} *for some* $t \le T_\gamma$ *(27), then for* $i = 1, \cdots, s$*, with probability at least* $1 - c_1 m^{-\nu} -$

$c_1 m e^{-c_2 K}$ *for some constants* $\nu, c_1, c_2 > 0$,

$$\max_{1 \le l \le m} \operatorname{dist}\left(\boldsymbol{z}_i^{t,(l)}, \widetilde{\boldsymbol{z}}_i^t\right) \le \left(\beta_{\boldsymbol{h}_i^{t+1}} + \beta_{\boldsymbol{x}_i^{t+1}}\right)\left(1 + \frac{1}{s \log m}\right)^{t+1} C_6 \frac{s\mu^2\kappa\sqrt{K \log^{18} m}}{m} \tag{A.11}$$

holds $m \ge C s \mu^2 \kappa K \log^8 m$ *with some sufficiently large constant* $C > 0$ *as long as the stepsize* $\eta > 0$ *obeys* $\eta \asymp s^{-1}$ *and* $C_6 > 0$ *is sufficiently large.*

Remark 2. *The proof of Lemma 8 and Lemma 9 is inspired by the arguments used in Section H and Section I in [16].*

A.3 Analysis for Stage II

Combining the analyses in Phase 1 and Phase 2, we complete the proof of Theorem 1 for Stage I, i.e. $0 \le t \le T_\gamma$ (27). Consider the definition of T_γ (27) and the incoherence between iterates and design vectors given in (A.9) and (A.10), we arrive at

$$\left\|\widetilde{\boldsymbol{x}}_i^{T_\gamma} - \boldsymbol{x}_i^\natural\right\|_2 \le \frac{\gamma}{\sqrt{s}} \tag{A.12}$$

$$\operatorname{dist}(\boldsymbol{z}^{T_\gamma}, \boldsymbol{z}^\natural) \le \gamma \tag{A.13}$$

$$\max_{1 \le i \le s, 1 \le j \le m} \left|\boldsymbol{a}_{ij}^* \widetilde{\boldsymbol{x}}_i^{T_\gamma}\right| \cdot \|\widetilde{\boldsymbol{x}}_i^{T_\gamma}\|_2^{-1} \lesssim \sqrt{\log m}, \tag{A.14}$$

$$\max_{1 \le i \le s, 1 \le j \le m} \left|\boldsymbol{b}_j^* \widetilde{\boldsymbol{h}}_i^{T_\gamma}\right| \cdot \|\widetilde{\boldsymbol{h}}_i^{T_\gamma}\|_2^{-1} \lesssim \frac{\mu}{\sqrt{m}} \log^2 m, \tag{A.15}$$

which further implies that

$$\max_{1 \le i \le s, 1 \le j \le m} \left|\boldsymbol{a}_{ij}^* \left(\widetilde{\boldsymbol{x}}^{T_\gamma} - \boldsymbol{x}^\natural\right)\right| \lesssim \frac{\gamma\sqrt{\log m}}{\sqrt{s}}, \tag{A.16}$$

based on the inductive hypothesis (A.8a). Based on these properties, we can exploit the techniques applied in [18, Section IV] to prove that for $t \ge T_\gamma + 1$,

$$\begin{aligned} \operatorname{dist}\left(\boldsymbol{z}^t, \boldsymbol{z}^\natural\right) \le & \left(1 - \frac{\eta}{16\kappa}\right)^{t - T_\gamma} \operatorname{dist}\left(\boldsymbol{z}^{T_\gamma}, \boldsymbol{z}^\natural\right) \\ \le & \gamma \left(1 - \frac{\eta}{16\kappa}\right)^{t - T_\gamma}, \end{aligned} \tag{A.17}$$

where the stepsize $\eta > 0$ obeys $\eta \asymp s^{-1}$ as long as $m \gg s^2 \mu^2 \kappa^4 \max\{K, N\} \log^8 m$. It remains to prove the claim (15) for Stage II. Since we have already demonstrate that the ratio $\alpha_{\boldsymbol{h}_i^t} / \beta_{\boldsymbol{h}_i^t}$ increases exponentially fast in Stage I, there is

$$\frac{\alpha_{\boldsymbol{h}_i^{T_1}}}{\beta_{\boldsymbol{h}_i^{T_1}}} \ge \frac{1}{\sqrt{2K \log K}} (1 + c_3 \eta)^{T_1}.$$

By the definition of T_1 (see (28)) and Lemma 1, one has $\alpha_{\boldsymbol{h}_i^{T_1}} \asymp \beta_{\boldsymbol{h}_i^{T_1}} \asymp 1$ and thus

$$\frac{\alpha_{\boldsymbol{h}_i^{T_1}}}{\beta_{\boldsymbol{h}_i^{T_1}}} \asymp 1. \tag{A.18}$$

When it comes to $t > T_\gamma$, based on (A.17), we have

$$\begin{aligned} \frac{\alpha_{\boldsymbol{h}_i^t}}{\beta_{\boldsymbol{h}_i^t}} \ge & \frac{1 - \operatorname{dist}(\boldsymbol{h}_i^t, \boldsymbol{h}_i^\natural)}{\operatorname{dist}(\boldsymbol{h}_i^t, \boldsymbol{h}_i^\natural)} \ge \frac{1 - \operatorname{dist}\left(\boldsymbol{z}^t, \boldsymbol{z}^\natural\right)}{\operatorname{dist}\left(\boldsymbol{z}^t, \boldsymbol{z}^\natural\right)} \\ \ge & \frac{1 - \gamma/\sqrt{2}}{\gamma/\sqrt{2}} \left(1 - \frac{\eta}{16\kappa}\right)^{t - T_\gamma} \overset{\text{(i)}}{\asymp} \frac{\alpha_{\boldsymbol{h}_i^{T_1}}}{\beta_{\boldsymbol{h}_i^{T_1}}} \left(1 - \frac{\eta}{16\kappa}\right)^{t - T_\gamma} \end{aligned}$$

$$\begin{aligned}
&\gtrsim \frac{1}{\sqrt{K\log K}}\left(1+c_3\eta\right)^{T_1}\left(1-\frac{\eta}{16\kappa}\right)^{t-T_\gamma}\\
&\overset{\text{(ii)}}{\gtrsim} \frac{1}{\sqrt{K\log K}}\left(1+c_3\eta\right)^{T_\gamma}\left(1-\frac{\eta}{16\kappa}\right)^{t-T_\gamma}\\
&\gtrsim \frac{1}{\sqrt{K\log K}}\left(1+c_3\eta\right)^{t},
\end{aligned}$$

where (i) is derived from (A.18) and the fact that γ is a constant, (ii) arises from $T_\gamma - T_1 \asymp s^{-1}$ based on Lemma 1, and the last inequality is satisfied as long as $c_3 > 0$ and $\eta \asymp s^{-1}$. Likewise, we can apply the same arguments to the ratio $\alpha_{\boldsymbol{x}_i^t}/\beta_{\boldsymbol{x}_i^t}$, thereby concluding that

$$\frac{\alpha_{\boldsymbol{x}_i^t}}{\beta_{\boldsymbol{x}_i^t}} \gtrsim \frac{1}{\sqrt{N\log N}}\left(1+c_4\eta\right)^{t}. \tag{A.19}$$

B Preliminaries

For $\boldsymbol{a}_{ij} \in \mathbb{C}^N$, the standard concentration inequality gives that, for $i = 1, \cdots, s$,

$$\max_{1\le j\le m} |a_{ij,1}| = \max_{1\le j\le m} \left|\boldsymbol{a}_{ij}^* \boldsymbol{x}^\natural\right| \le 5\sqrt{\log m} \tag{B.1}$$

with probability $1 - \mathcal{O}\left(m^{-10}\right)$ [18]. In addition, by applying the standard concentration inequality, we arrive at, for $i = 1, \cdots, s$,

$$\max_{1\le j\le m} \|\boldsymbol{a}_{ij}\|_2 \le 3\sqrt{N} \tag{B.2}$$

with probability $1 - C' \exp\left(me^{-cK}\right)$ for some constants, $c, C' > 0$ [18].

Lemma 10. *Fix any constant $c_0 > 1$. Define the population matrix $\nabla^2_{\boldsymbol{z}_i} F(\boldsymbol{z})$ as*

$$\begin{bmatrix}
\|\boldsymbol{x}_i\|_2^2 \boldsymbol{I}_K & \boldsymbol{h}_i\boldsymbol{x}_i^* - \boldsymbol{h}_i^\natural\boldsymbol{x}_i^{\natural *} & \boldsymbol{0} & \boldsymbol{h}_i^\natural\boldsymbol{x}_i^{\natural\top}\\
\boldsymbol{x}_i\boldsymbol{h}_i^* - \boldsymbol{x}_i^\natural\boldsymbol{h}_i^{\natural *} & \|\boldsymbol{h}_i\|_2^2 \boldsymbol{I}_K & \boldsymbol{x}_i^\natural\boldsymbol{h}_i^{\natural\top} & \boldsymbol{0}\\
\boldsymbol{0} & \left(\boldsymbol{x}_i^\natural\boldsymbol{h}_i^{\natural\top}\right)^* & \overline{\|\boldsymbol{x}_i\|_2^2 \boldsymbol{I}_K} & \overline{\boldsymbol{h}_i\boldsymbol{x}_i^* - \boldsymbol{h}_i^\natural\boldsymbol{x}_i^{\natural *}}\\
\left(\boldsymbol{h}_i^\natural\boldsymbol{x}_i^{\natural\top}\right)^* & \boldsymbol{0} & \overline{\boldsymbol{x}_i\boldsymbol{h}_i^* - \boldsymbol{x}_i^\natural\boldsymbol{h}_i^{\natural *}} & \|\boldsymbol{h}_i\|_2^2 \boldsymbol{I}_K
\end{bmatrix}$$

Suppose that $m > c_1 s^2\mu^2 K \log^3 m$ for some sufficiently large constant $c_1 > 0$. Then with probability exceeding $1 - \mathcal{O}\left(m^{-10}\right)$,

$$\left\|\left(\boldsymbol{I}_{4K} - \eta\nabla^2 f(\boldsymbol{z})\right) - \left(\boldsymbol{I}_{4K} - \eta\nabla^2 F(\boldsymbol{z})\right)\right\| \lesssim \sqrt{\frac{s^2\mu^2 K\log m}{m}} \max\left\{\|\boldsymbol{z}\|_2^2, 1\right\}$$

$$\text{and} \qquad \left\|\nabla^2 f(\boldsymbol{z})\right\| \le 5\|\boldsymbol{z}\|_2^2 + 2$$

hold simultaneously for all $\boldsymbol{z}$ obeying $\max_{1\le i\le s, 1\le l\le m} |\boldsymbol{a}_{il}^*\boldsymbol{x}_i| \cdot \left\|\boldsymbol{x}_i\right\|_2^{-1} \lesssim \sqrt{\log m}$ and $\max_{1\le i\le s, 1\le l\le m} |\boldsymbol{b}_l^*\boldsymbol{h}_i| \cdot \left\|\boldsymbol{h}_i\right\|_2^{-1} \lesssim \frac{\mu}{\sqrt{m}}\log^2 m$, provided that $0 < \eta < \frac{c_2}{\max\{\|\boldsymbol{z}\|_2^2, 1\}}$ for some sufficiently small constant $c_2 > 0$.

C Proof of Lemma 1

To prove Lemma 1, we divide Stage I into several substages and analyze them separately. For simplification, we focus on the case when the initialization obeys (26), which can be generalized to other cases.

- **Stage I-a**. Consider the iterations $0 \le t \le T_0$ with $T_0 = \min\left\{t \mid \max_i \beta_{\boldsymbol{h}_i^{t+1}}/q_i \le \sqrt{0.6},\ \max_i \beta_{\boldsymbol{x}_i^{t+1}}/q_i \le \sqrt{0.6}\right\}$, we have the following claim.

Claim 1. *Assume that the stepsize $\eta > 0$ is sufficiently small, for $i = 1, \cdots, s$, we have*

$$\beta_{\boldsymbol{h}_i^{t+1}} \leq (1 - \frac{\eta}{2})\beta_{\boldsymbol{h}_i^t}, \qquad 0 \leq t \leq T_0, \tag{C.1a}$$

$$\alpha_{\boldsymbol{h}_i^{t+1}} \leq (1 + 2\eta)\alpha_{\boldsymbol{h}_i^t}, \qquad 0 \leq t \leq T_0, \tag{C.1b}$$

$$\alpha_{\boldsymbol{h}_i^{t+1}} \geq (1 + \frac{\eta}{2})\alpha_{\boldsymbol{h}_i^t}, \qquad 1 \leq t \leq T_0, \tag{C.1c}$$

$$\alpha_{\boldsymbol{h}_i^1} \geq \alpha_{\boldsymbol{h}_i^0}/2, \tag{C.1d}$$

$$\alpha_{\boldsymbol{h}_i^{T_0+1}} \geq (1 - 2\eta)\sqrt{0.6}q_i, \tag{C.1e}$$

$$T_0 \lesssim \frac{1}{\eta}. \tag{C.1f}$$

In addition, there is $T_0 < T_2$ (recalling the definition of T_2 (29)) since $\max_i \alpha_{\boldsymbol{h}_i^{T_0}} \ll c_8$. Similarity, the condition (C.1) is satisfied in the case with respect to $\boldsymbol{x}_i^t$ for $i = 1, \cdots, s$, $0 \leq t \leq T_0$.

In consequence, we conclude from Claim 1 that for $0 \leq t \leq T_0$:

$$\begin{aligned}
&c_8 q_i > \alpha_{\boldsymbol{h}_i^t} \geq \frac{\alpha_{\boldsymbol{h}_i^0}}{2} \geq \frac{q_i}{2\sqrt{K \log K}}, \\
&c_8' q_i > \alpha_{\boldsymbol{x}_i^t} \geq \frac{\alpha_{\boldsymbol{x}_i^0}}{2} \geq \frac{q_i}{2\sqrt{N \log N}}, \\
&1.5 q_i > \beta_{\boldsymbol{h}_i^0} \geq \beta_{\boldsymbol{h}_i^t} \geq \beta_{T_0+1} \geq (1 - 2\eta)\sqrt{0.6}q_i, \\
&1.5 q_i > \beta_{\boldsymbol{z}_i^0} \geq \beta_{\boldsymbol{z}_i^t} \geq \beta_{T_0+1} \geq (1 - 2\eta)\sqrt{0.6}q_i, \\
&\frac{\alpha_{\boldsymbol{h}_i^{t+1}}/\alpha_{\boldsymbol{h}_i^t}}{\beta_{\boldsymbol{h}_i^{t+1}}/\beta_{\boldsymbol{h}_i^t}} \geq 1 + \eta \quad \text{and} \quad \frac{\alpha_{\boldsymbol{x}_i^{t+1}}/\alpha_{\boldsymbol{x}_i^t}}{\beta_{\boldsymbol{x}_i^{t+1}}/\beta_{\boldsymbol{x}_i^t}} \geq 1 + \eta,
\end{aligned}$$

which justifies (30) and (31).

- **Stage I-b**. The second substage is consist of the iterations obeying $T_0 < t \leq T_2$ (recalling the definition of T_2 (29)).

Claim 2. *Assume that the stepsize $\eta > 0$ is sufficiently small, for $i = 1, \cdots, s$, $T_0 < t \leq T_2$, we have*

$$\beta_{\boldsymbol{h}_i^t} \in \left[(1 - 2\eta)^2\sqrt{0.6}q_i, (1 + 0.1\eta)\sqrt{0.6}q_i\right] \tag{C.2a}$$

$$\beta_{\boldsymbol{h}_i^{t+1}} \leq (1 + 0.1\eta)\beta_{\boldsymbol{h}_i^t} \tag{C.2b}$$

$$\alpha_{\boldsymbol{h}_i^{t+1}} \leq (1 + 2.2\eta)\alpha_{\boldsymbol{h}_i^t}. \tag{C.2c}$$

Similarity, the condition (C.2) is satisfied in the case with respect to $\boldsymbol{x}_i^t$ for $i = 1, \cdots, s$, $T_0 < t \leq T_2$.

Hence, recall the definition of T_1 (28), we arrive at

$$\begin{aligned}
T_2 - T_0 &\lesssim \frac{\log \frac{\max\{c_8, c_8'\}}{\alpha_0}}{\log(1 + 2.2\eta)} \lesssim \frac{\log \max\{K, N\}}{\eta}, \\
T_2 - T_1 &\lesssim \frac{\log \frac{\max\{c_8, c_8'\}}{\frac{\max\{c_7, c_7'\}}{\log^5 m}}}{\log(1 + 2.2\eta)} \lesssim \frac{\log \log m}{\eta}, \\
&\frac{\alpha_{\boldsymbol{h}_i^{t+1}}/\alpha_{\boldsymbol{h}_i^t}}{\beta_{\boldsymbol{h}_i^{t+1}}/\beta_{\boldsymbol{h}_i^t}} \geq 1 + 0.1\eta \quad \text{and} \quad \frac{\alpha_{\boldsymbol{x}_i^{t+1}}/\alpha_{\boldsymbol{x}_i^t}}{\beta_{\boldsymbol{x}_i^{t+1}}/\beta_{\boldsymbol{x}_i^t}} \geq 1 + 0.1\eta.
\end{aligned}$$

- **Stage I-c**. Consider the iteration in $T_2 \leq t \leq T_\gamma$, we have the following results.

Claim 3. *Assume that the stepsize $\eta > 0$ is sufficiently small, for $i = 1, \cdots, s$, $T_2 < t \leq T_\gamma$, we have*

$$\alpha_{\boldsymbol{h}_i^t} + \beta_{\boldsymbol{h}_i^t} \leq \frac{\gamma}{\kappa\sqrt{s}} q_i^2, \tag{C.3a}$$

$$\frac{\alpha_{\boldsymbol{h}_i^{t+1}}/\alpha_{\boldsymbol{h}_i^t}}{\beta_{\boldsymbol{h}_i^{t+1}}/\beta_{\boldsymbol{h}_i^t}} \geq 1 + c_9\eta, \tag{C.3b}$$

$$\alpha_{\boldsymbol{h}_i^{t+1}} \geq (1 - 1.1\eta + \eta\kappa\sqrt{s}/\gamma)\alpha_{\boldsymbol{h}_i^t}, \tag{C.3c}$$

$$\beta_{\boldsymbol{h}_i^{t+1}} \geq (1 - 0.9\eta)\beta_{\boldsymbol{h}_i^t}, \tag{C.3d}$$

$$T_\gamma - T_2 \lesssim \frac{1}{\eta}, \tag{C.3e}$$

for some constant $c_9 > 0$. Similarity, the condition (C.3) is satisfied in the case with respect to $\boldsymbol{x}_i^t$ for $i = 1, \cdots, s$, $T_2 < t \leq T_\gamma$.

D Proof of Lemma 2

D.1 Proof of (23c)

According to the Wirtinger flow gradient update rule (5b), the signal component x_{i1}^{t+1} can be represented as follows

$$\widetilde{x}_{i1}^{t+1} = \widetilde{x}_{i1}^t - \frac{\eta}{\|\widetilde{\boldsymbol{h}}_i^t\|_2^2} \sum_{j=1}^m \left(\sum_{k=1}^s \boldsymbol{h}_k^{t*} \boldsymbol{b}_j \boldsymbol{a}_{kj}^* \boldsymbol{x}_k^t - \boldsymbol{h}_k^{\natural *} \boldsymbol{b}_j \boldsymbol{a}_{kj}^* \boldsymbol{x}_k^\natural \right) \boldsymbol{b}_j^* \widetilde{\boldsymbol{h}}_i^t a_{ij,1}$$

Expanding this expression using $\boldsymbol{a}_{kj}^* \boldsymbol{x}_k^t = x_{k\|}^t \overline{a_{kj,1}} + \boldsymbol{a}_{kj,\perp}^* \boldsymbol{x}_{k\perp}^t$ and reformulate terms, we arrive at

$$\widetilde{x}_{i1}^{t+1} = \widetilde{x}_{i1}^t + \eta' J_{i1} - \eta' J_{i2} - \eta' J_{i3}, \tag{D.1}$$

where

$$J_{i1} = \sum_{j=1}^m \sum_{k=1}^s \boldsymbol{h}_k^{\natural *} \boldsymbol{b}_j \boldsymbol{b}_j^* \widetilde{\boldsymbol{h}}_i^t \overline{a_{kj,1}} q_k a_{ij,1}$$

$$J_{i2} = \sum_{j=1}^m \sum_{k=1}^s \widetilde{\boldsymbol{h}}_k^{t*} \boldsymbol{b}_j \boldsymbol{b}_j^* \widetilde{\boldsymbol{h}}_i^t \overline{a_{kj,1}} \widetilde{x}_{k\|}^t a_{ij,1}$$

$$J_{i3} = \sum_{j=1}^m \sum_{k=1}^s \widetilde{\boldsymbol{h}}_k^{t*} \boldsymbol{b}_j \boldsymbol{b}_j^* \widetilde{\boldsymbol{h}}_i^t \boldsymbol{a}_{kj,\perp}^* \boldsymbol{x}_{i\perp}^t a_{ij,1}$$

$$\eta' = \eta/\|\widetilde{\boldsymbol{h}}_i^t\|_2^2.$$

We will control the above three terms J_{i1}, J_{i2} and J_{i3} separately in the following.

- With regard to the first term J_{i1}, it has

$$\sum_{j=1}^m \sum_{k=1}^s q_k \boldsymbol{h}_k^{\natural *} \boldsymbol{b}_j \boldsymbol{b}_j^* \widetilde{\boldsymbol{h}}_i^t \overline{a_{kj,1}} a_{ij,1} = \sum_{k=1}^s q_k \boldsymbol{h}_k^{\natural *} \left(\sum_{j=1}^m \overline{a_{kj,1}} a_{ij,1} \boldsymbol{b}_j \boldsymbol{b}_j^* \right) \widetilde{\boldsymbol{h}}_i^t.$$

  According to Lemma 11 and Lemma 12, there is

$$J_{i1} = q_i \boldsymbol{h}_i^{\natural *} \widetilde{\boldsymbol{h}}_i^t + r_1, \tag{D.2}$$

  where the size of the remaining term r_1 satisfies

$$|r_1| \lesssim \sum_{k=1}^s q_k \boldsymbol{h}_k^{\natural *} \widetilde{\boldsymbol{h}}_i^t \sqrt{\frac{K}{m} \log m} \lesssim \sqrt{\frac{s^2 K}{m} \log m} \cdot \boldsymbol{h}_i^{\natural *} \widetilde{\boldsymbol{h}}_i^t, \tag{D.3}$$

  based on the fact that $\|\boldsymbol{h}_k^\natural\|^2 \lesssim 1$ and $\|\widetilde{\boldsymbol{h}}_k^t\|^2 \lesssim 1$ for $k = 1, \cdots, s$.

- Similar to the first term, the term J_{i2} can be represented as

$$J_{i2} = \left\| \widetilde{\boldsymbol{h}}_i^t \right\|_2^2 \widetilde{x}_{i1}^t + r_2, \tag{D.4}$$

where the term r_{i2} obeys

$$|r_2| \lesssim \left| \widetilde{x}_{i1}^t \right| \sum_{k=1}^{s} \widetilde{\boldsymbol{h}}_k^{t*} \widetilde{\boldsymbol{h}}_i^t \sqrt{\frac{K}{m} \log m} \lesssim \sqrt{\frac{s^2 K}{m} \log m} \left| \widetilde{x}_{i1}^t \right|. \tag{D.5}$$

- For the last term J_{i3}, it follows that

$$\sum_{j=1}^{m} \sum_{k=1}^{s} \widetilde{\boldsymbol{h}}_k^{t*} \boldsymbol{b}_j \boldsymbol{b}_j^* \widetilde{\boldsymbol{h}}_i^t \boldsymbol{a}_{kj,\perp}^* \widetilde{\boldsymbol{x}}_{i\perp}^t a_{ij,1} = \sum_{k=1}^{s} \widetilde{\boldsymbol{h}}_k^{t*} \left(\sum_{j=1}^{m} a_{ij,1} \boldsymbol{a}_{kj,\perp}^* \boldsymbol{x}_{i\perp}^t \boldsymbol{b}_j \boldsymbol{b}_j^* \right) \widetilde{\boldsymbol{h}}_i^t. \tag{D.6}$$

By exploiting the random-sign sequence $\left\{ \boldsymbol{x}_i^{t,\mathrm{sgn}} \right\}$, one can decompose

$$\sum_{j=1}^{m} a_{ij,1} \boldsymbol{a}_{kj,\perp}^* \widetilde{\boldsymbol{x}}_{i\perp}^t \boldsymbol{b}_j \boldsymbol{b}_j^* = \sum_{j=1}^{m} a_{ij,1} \boldsymbol{a}_{kj\perp}^* \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}} \boldsymbol{b}_j \boldsymbol{b}_j^* + \sum_{j=1}^{m} a_{ij,1} \boldsymbol{a}_{kj,\perp}^* \left(\widetilde{\boldsymbol{x}}_{i\perp}^t - \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}} \right) \boldsymbol{b}_j \boldsymbol{b}_j^*. \tag{D.7}$$

Note that $a_{ij,1} \boldsymbol{a}_{kj\perp}^* \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}} \boldsymbol{b}_j \boldsymbol{b}_j^*$ in (D.7) is statistically independent of ξ_{ij} (35) and $\boldsymbol{b}_j^{\mathrm{sgn}} \boldsymbol{b}_j^{\mathrm{sgn}*} = \boldsymbol{b}_j \boldsymbol{b}_j^*$. Hence we can consider $\sum_{j=1}^{m} a_{ij,1} \boldsymbol{a}_{kj\perp}^* \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}} \boldsymbol{b}_j \boldsymbol{b}_j^*$ as a weighted sum of the ξ_{ij}'s and exploit the Bernstein inequality to derive that

$$\left\| \sum_{j=1}^{m} a_{ij,1} \boldsymbol{a}_{kj\perp}^* \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}} \boldsymbol{b}_j \boldsymbol{b}_j^* \right\| = \left\| \sum_{j=1}^{m} \xi_{ij} \left(a_{ij,1} \boldsymbol{a}_{kj\perp}^* \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}} \boldsymbol{b}_j \boldsymbol{b}_j^* \right) \right\| \lesssim \sqrt{V_1 \log m} + B_1 \log m \tag{D.8}$$

with probability exceeding $1 - \mathcal{O}\left(m^{-10}\right)$, where

$$V_1 := \sum_{j=1}^{m} \left| a_{ij,1} \right|^2 \left| \boldsymbol{a}_{kj\perp}^* \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}} \right|^2 \left| \boldsymbol{b}_j \boldsymbol{b}_j^* \right|^2,$$

$$B_1 := \max_{1 \le j \le m} \left| a_{ij,1} \right| \left| \boldsymbol{a}_{kj\perp}^* \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}} \right| \left| \boldsymbol{b}_j \boldsymbol{b}_j^* \right|.$$

In view of Lemma 17 and the incoherence condition (A.1d) to deduce that with probability at least $1 - \mathcal{O}\left(m^{-10}\right)$,

$$V_1 \lesssim \left\| \sum_{j=1}^{m} \left| a_{i,1} \right|^2 \left| \boldsymbol{a}_{kj\perp}^* \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}} \right|^2 \boldsymbol{b}_j \boldsymbol{b}_j^* \right\| \left\| \boldsymbol{b}_j \right\|_2^2 \lesssim \frac{K}{m} \left\| \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}} \right\|_2^2$$

with the proviso that $m \gg \max\{K, N\} \log^3 m$. Furthermore, the incoherence condition (A.1d) together with the fact (B.1) implies that

$$B_1 \lesssim \frac{K}{m} \log m \left\| \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}} \right\|_2.$$

Substitute the bounds on V_1 and B_1 back to (D.8) to obtain

$$\left\| \sum_{j=1}^{m} a_{ij,1} \boldsymbol{a}_{kj\perp}^* \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}} \boldsymbol{b}_j \boldsymbol{b}_j^* \right\| \lesssim \sqrt{\frac{K \log m}{m}} \left\| \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}} \right\|_2 \tag{D.9}$$

as long as $m \gtrsim K \log^3 m$. In addition, we move to the second term on the right-hand side of (D.7). Let $\boldsymbol{\iota} = \sum_{j=1}^{m} a_{ij,1} \boldsymbol{a}_{kj}^* \boldsymbol{z} \boldsymbol{b}_j \boldsymbol{b}_j^*$, where $\boldsymbol{z} \in \mathbb{C}^{N-1}$ is independent with $\{\boldsymbol{a}_{kj}\}$ and $\|\boldsymbol{z}\|_2 = 1$. Hence, we have

$$\left\| \sum_{j=1}^{m} a_{ij,1} \boldsymbol{a}_{kj,\perp}^* \left(\widetilde{\boldsymbol{x}}_{i\perp}^t - \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}} \right) \boldsymbol{b}_j \boldsymbol{b}_j^* \right\| \le \|\boldsymbol{\iota}\|_2 \left\| \widetilde{\boldsymbol{x}}_{i\perp}^t - \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}} \right\|_2 \lesssim \sqrt{\frac{K \log m}{m}} \left\| \widetilde{\boldsymbol{x}}_{i\perp}^t - \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}} \right\|_2, \tag{D.10}$$

with probability exceeding $1-\mathcal{O}\left(m^{-10}\right)$, as long as that $m \gg K\log^3 m$. Here, the last inequality of (D.10) comes from Lemma 13. Substituting the above two bounds (D.9) and (D.10) into (D.7), it yields

$$\left\|\sum_{j=1}^{m} a_{ij,1}\boldsymbol{a}_{kj,\perp}^{*}\widetilde{\boldsymbol{x}}_{i\perp}^{t}\boldsymbol{b}_j\boldsymbol{b}_j^{*}\right\| \lesssim \sqrt{\frac{K\log m}{m}}\left\|\check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}}\right\|_2 + \sqrt{\frac{K\log m}{m}}\left\|\widetilde{\boldsymbol{x}}_{i\perp}^{t} - \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}}\right\|_2. \tag{D.11}$$

Combining (D.6) and (D.11), we arrive at

$$|J_{i3}| \lesssim \sqrt{\frac{s^2K\log m}{m}}\left\|\widetilde{\boldsymbol{x}}_{i\perp}^{t}\right\|_2 + \sqrt{\frac{s^2K\log m}{m}}\left\|\widetilde{\boldsymbol{x}}_{i\perp}^{t} - \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}}\right\|_2, \tag{D.12}$$

by exploiting the fact that $\|\widetilde{\boldsymbol{h}}_k^t\|^2 \lesssim 1$ for $k=1,\cdots,s$ and the triangle inequality $\left\|\check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}}\right\|_2 \leq \|\widetilde{\boldsymbol{x}}_{i\perp}^{t}\|_2 + \left\|\widetilde{\boldsymbol{x}}_{i\perp}^{t} - \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}}\right\|_2$.

- Collecting the bounds for J_{i1}, J_{i2} and J_{i3}, we arrive at

$$\begin{aligned}\widetilde{x}_{i1}^{t+1} &= \widetilde{x}_{i1}^{t} + \eta' J_{i1} - \eta' J_{i2} - \eta' J_{i3}\\ &= \widetilde{x}_{i1}^{t} + \eta q_i \boldsymbol{h}_i^{\natural *}\boldsymbol{h}_i^t/\|\widetilde{\boldsymbol{h}}_i^t\|_2^2 - \eta\widetilde{x}_{i1}^{t} + R\\ &= (1-\eta)\,x_{i1}^{t} + \eta q_i \boldsymbol{h}_i^{\natural *}\boldsymbol{h}_i^t/\|\widetilde{\boldsymbol{h}}_i^t\|_2^2 + R,\end{aligned} \tag{D.13}$$

where the residual term R follows that

$$|R| \lesssim \frac{\eta}{\|\widetilde{\boldsymbol{h}}_i^t\|_2^2}\sqrt{\frac{s^2K}{m}\log m}\Big(\boldsymbol{h}_i^{\natural *}\boldsymbol{h}_i^t + \left|\widetilde{x}_{i1}^{t}\right| + \left\|\widetilde{\boldsymbol{x}}_{i\perp}^{t}\right\|_2 + \left\|\widetilde{\boldsymbol{x}}_{i\perp}^{t} - \check{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}}\right\|_2\Big) \tag{D.14}$$

Substituting the hypotheses (41) into (D.13) and in view of the fact $\alpha_{\boldsymbol{x}_i^t} = \langle \boldsymbol{x}^t, \boldsymbol{x}^{\natural}\rangle/\|\boldsymbol{x}_i^{\natural}\|_2$ and the assumption that $\|\boldsymbol{h}_i^{\natural}\|_2 = \|\boldsymbol{x}_i^{\natural}\|_2 = q_i$ for $i=1,\cdots,s$, one has

$$\begin{aligned}&\alpha_{\boldsymbol{x}_i^{t+1}}\\ &= (1-\eta)\,\alpha_{\boldsymbol{x}_i^t} + \eta'' q_i \boldsymbol{h}_i^{\natural *}\widetilde{\boldsymbol{h}}_i^t + \mathcal{O}\left(\eta''\sqrt{\frac{s^2K}{m}\log m}\,\alpha_{\boldsymbol{x}_i^t}\right) + \mathcal{O}\left(\eta''\sqrt{\frac{s^2K}{m}\log m}\,\beta_{\boldsymbol{x}_i^t}\right) + \mathcal{O}\left(\eta''\sqrt{\frac{s^2K}{m}\log m}\cdot\alpha_{\boldsymbol{h}_i^t}\right)\\ &\quad + \mathcal{O}\left(\eta''\alpha_{\boldsymbol{x}_i^t}\left(1+\frac{1}{s\log m}\right)^t C_3\sqrt{\frac{s\mu^2N\log^8 m}{m}}\right)\\ &= (1-\eta + \frac{\eta q_i\psi_{\boldsymbol{x}_i^t}}{\alpha_{\boldsymbol{x}_i^t}^2 + \beta_{\boldsymbol{x}_i^t}^2})\alpha_{\boldsymbol{x}_i^t} + \eta(1-\rho_{\boldsymbol{x}_i^t})\frac{q_i\alpha_{\boldsymbol{h}_i^t}}{\alpha_{\boldsymbol{h}_i^t}^2 + \beta_{\boldsymbol{h}_i^t}^2},\end{aligned} \tag{D.15}$$

where $\eta'' = \eta/(q_i\|\boldsymbol{h}_i^t\|_2^2)$, for some $|\psi_{\boldsymbol{x}_i^t}|, |\rho_{\boldsymbol{x}_i^t}| \ll \frac{1}{\log m}$, provided that

$$\sqrt{\frac{s^2K\log m}{q_i^2 m}} \ll \frac{q_i}{\log m} \tag{D.16a}$$

$$\sqrt{\frac{s^2K\log m}{q_i^2 m}}\beta_{\boldsymbol{x}_i^t} \ll \frac{q_i}{\log m}\alpha_{\boldsymbol{x}_i^t} \tag{D.16b}$$

$$\left(1+\frac{1}{s\log m}\right)^t C_3\sqrt{\frac{s\mu^2N\log^8 m}{q_i^2 m}} \ll \frac{q_i}{\log m}, \tag{D.16c}$$

where the parameter q_i is assumed to be $0 < q_i \leq 1$. Therein, the first condition (D.16a) naturally holds as long as $m \gg s^2K\log^3 m$. In addition, the second condition (D.16b) holds true since $\beta_{\boldsymbol{x}_i^t} \leq$

$\|\boldsymbol{x}_i^t\|_2 \lesssim \alpha_{\boldsymbol{x}_i^t}\sqrt{\log^5 m}$ (based on (41j)) and $m \gg s^2 K \log^8 m$. For the last condition (D.16c), we have for $t \leq T_1 = \mathcal{O}\left(s \log \max\{K, N\}\right)$,

$$\left(1 + \frac{1}{s\log m}\right)^t = \mathcal{O}\left(1\right),$$

which further implies

$$\left(1 + \frac{1}{s\log m}\right)^t C_3\sqrt{\frac{s\mu^2 N \log^8 m}{q_i^2 m}} \lesssim C_3\sqrt{\frac{s\mu^2 N \log^8 m}{q_i^2 m}} \ll \frac{q_i}{\log m}$$

as long as the number of samples obeys $m \gg s\mu^2 N \log^{10} m$. This concludes the proof.

Despite it turns to be more tedious when proving (23a), similar arguments used above can be applied to the proof of (23a). Specifically, according to the Wirtinger flow gradient update rule (5a), the signal component $\langle \boldsymbol{h}_i^\natural, \widetilde{\boldsymbol{h}}_i^t \rangle$ can be represented as follows

$$\boldsymbol{h}_i^{\natural *}\widetilde{\boldsymbol{h}}_i^{t+1} = \boldsymbol{h}_i^{\natural *}\widetilde{\boldsymbol{h}}_i^{t} - \frac{\eta}{\|\widetilde{\boldsymbol{x}}_i^t\|_2^2}\sum_{j=1}^m\left(\sum_{k=1}^s \boldsymbol{b}_j^*\widetilde{\boldsymbol{h}}_k^t\widetilde{\boldsymbol{x}}_k^{t*}\boldsymbol{a}_{kj} - y_j\right)\boldsymbol{h}_i^{\natural *}\boldsymbol{b}_j\boldsymbol{a}_{ij}^*\widetilde{\boldsymbol{x}}_i^t$$

Expanding this expression using $\boldsymbol{a}_{kj}^*\boldsymbol{x}_k^t = x_{k\|}^t\overline{a_{kj,1}} + \boldsymbol{a}_{kj,\perp}^*\boldsymbol{x}_{k\perp}^t$ and rearranging terms, we are left with

$$\boldsymbol{h}_i^{\natural *}\widetilde{\boldsymbol{h}}_i^{t+1} = \boldsymbol{h}_i^{\natural *}\widetilde{\boldsymbol{h}}_i^{t} - \eta_i' L_{i1} + \eta_i' L_{i2} + \eta_i' L_{i3}, \tag{D.17}$$

where

$$\begin{aligned}
L_{i1} &= \sum_{j=1}^m\sum_{k=1}^s \boldsymbol{h}_i^{\natural *}\boldsymbol{b}_j\boldsymbol{b}_j^*\widetilde{\boldsymbol{h}}_k^t\widetilde{\boldsymbol{x}}_k^{t*}\boldsymbol{a}_{kj}\boldsymbol{a}_{ij}^*\boldsymbol{x}_i\\
L_{i2} &= \sum_{j=1}^m\sum_{k=1}^s \boldsymbol{h}_i^{\natural *}\boldsymbol{b}_j\boldsymbol{b}_j^*\boldsymbol{h}_k^{\natural} a_{kj,1} q_k \overline{a_{ij,1}}\widetilde{x}_{i1}^t\\
L_{i3} &= \sum_{j=1}^m\sum_{k=1}^s \boldsymbol{h}_i^{\natural *}\boldsymbol{b}_j\boldsymbol{b}_j^*\boldsymbol{h}_k^{\natural}\boldsymbol{a}_{ij,\perp}^*\boldsymbol{x}_{i\perp}^t a_{kj,1} q_k,\\
\eta_i' &= \eta/\|\widetilde{\boldsymbol{x}}_i^t\|_2^2.
\end{aligned}$$

Here, L_{i1}, L_{i2} and L_{i3} can be controlled via the strategies exploited to control J_{i1}, J_{i2} and J_{i3}.

D.2 Proof of (23d)

In view of Lemma 15 and Lemma 16, by utilizing similar arguments as in Section D.1, it yields that with probability exceeding $1 - O(m^{-10})$,

$$\widetilde{\boldsymbol{x}}_{i\perp}^{t+1} = (1-\eta)\,\widetilde{\boldsymbol{x}}_{i\perp}^t - \eta'\boldsymbol{r}_1, \tag{D.18}$$

where

$$\begin{aligned}
\eta' =& \eta/\|\widetilde{\boldsymbol{h}}_i^t\|_2^2\\
\|\boldsymbol{r}_1\|_2 \lesssim& \frac{\mu}{\sqrt{m}}\sqrt{\frac{s^2\mu^2K}{m}\log^9 m}\left(\left\|\widetilde{\boldsymbol{x}}_{i\perp}^t\right\|_2 + |\widetilde{x}_{i1}^t|\right) + (\boldsymbol{h}_i^{\natural *}\boldsymbol{h}_i^t)q_i\frac{\mu}{\sqrt{m}}\sqrt{\frac{s^2\mu^2K}{m}\log^9 m}
\end{aligned}$$

According to the definitions of $\alpha_{\boldsymbol{x}_i}$ (13) and $\beta_{\boldsymbol{x}_i}$ (14), we arrive at

$$\begin{aligned}
\beta_{\boldsymbol{x}_i^{t+1}} =& (1-\eta)\,\beta_{\boldsymbol{x}_i^t} + \mathcal{O}\left(\eta' q_i^2\frac{\mu}{\sqrt{m}}\sqrt{\frac{s^2\mu^2K}{m}\log^9 m}\cdot\alpha_{\boldsymbol{h}_i^t}\right) + \mathcal{O}\left(\eta'\frac{\mu}{\sqrt{m}}\sqrt{\frac{s^2\mu^2N}{m}\log^9 m}\left(\alpha_{\boldsymbol{x}_i^t} + \beta_{\boldsymbol{x}_i^t}\right)\right)\\
=&(1-\eta+\frac{\eta q_i\varphi_{\boldsymbol{x}_i^t}}{\alpha_{\boldsymbol{h}_i^t}^2+\beta_{\boldsymbol{h}_i^t}^2})\beta_{\boldsymbol{x}_i^t},
\end{aligned} \tag{D.19}$$

for some $|\varphi_{\boldsymbol{x}_i^t}| \ll \frac{1}{\log m}$, with the proviso that $m \gg s^2\mu^2 \max\{K,N\}\log^3 m$ and

$$\frac{\mu}{\sqrt{m}}\sqrt{\frac{s^2\mu^2 K}{m}\log^9 m}\cdot(\alpha_{\boldsymbol{x}_i^t}+\beta_{\boldsymbol{x}_i^t}) \ll \frac{q_i}{\log m}\beta_{\boldsymbol{x}_i^t}, \tag{D.20}$$

$$q_i^2\frac{\mu}{\sqrt{m}}\sqrt{\frac{s^2\mu^2 N}{m}\log^9 m}\cdot\alpha_{\boldsymbol{h}_i^t} \ll \frac{q_i}{\log m}\beta_{\boldsymbol{x}_i^t}. \tag{D.21}$$

Here, according to the assumption $\alpha_{\boldsymbol{h}_i^t} \lesssim 1/\log^5 m$ (see definition of T_1 (28)) and the induction hypothesis $\beta_t \geq c_5$ (see (41h)), the condition (D.20) and (D.21) are satisfied as long as $m \gg s^2\mu^2\max\{K,N\}\log^{11/2} m$.

E Proof of (A.5) in Lemma 5

By applying the arguments in [2, Appendix F], it yields that

$$\begin{aligned}&\operatorname{dist}\left(\boldsymbol{x}_i^{t+1,(l)}, \widetilde{\boldsymbol{x}}_i^{t+1}\right)\\ \leq&\kappa\sqrt{\sum_{k=1}^{s}\max\left\{\left|\frac{\omega_i^{t+1}}{\omega_i^t}\right|,\left|\frac{\omega_i^t}{\omega_i^{t+1}}\right|\right\}^2\|\boldsymbol{J}_k\|^2},\end{aligned} \tag{E.1}$$

where ω_i^t is the alignment parameter and

$$\boldsymbol{J}_k = \omega_k^t\boldsymbol{x}_k^{t+1} - \omega_{k,\text{mutual}}^{t,(l)}\boldsymbol{x}_k^{t+1,(l)} \tag{E.2}$$

where $\omega_{k,\text{mutual}}^{t,(l)}$ is defined in (38). According to (10) and (38), we arrive at

$$\begin{aligned}&\omega_i^t x_{i1}^{t+1} - \omega_{i,\text{mutual}}^{t,(l)} x_{i1}^{t+1,(l)}\\ =&\boldsymbol{e}_1^\top\left(\widetilde{\boldsymbol{x}}_i^{t+1} - \widehat{\boldsymbol{x}}_i^{t+1,(l)}\right)\\ =&\widetilde{x}_{i1}^t - \widehat{x}_{i1}^{t,(l)} - \eta'\boldsymbol{e}_1^\top\left(\nabla_{\boldsymbol{x}_i} f\left(\widetilde{\boldsymbol{z}}^t\right) - \nabla_{\boldsymbol{x}_i} f^{(l)}\left(\widehat{\boldsymbol{z}}_i^{t,(l)}\right)\right) - \eta'\left(\sum_{k=1}^{s}\widehat{\boldsymbol{h}}_i^{t,(l)*}\boldsymbol{b}_l\boldsymbol{a}_{kl}^*\widehat{\boldsymbol{x}}_i^{t,(l)} - \boldsymbol{h}_k^{\natural*}\boldsymbol{b}_l\boldsymbol{a}_{kl}^*\boldsymbol{x}_k^\natural\right)\boldsymbol{b}_l^*\widehat{\boldsymbol{h}}_i^{t,(l)}a_{il,1},\end{aligned}$$

where the stepsize $\eta' = \eta/\|\widetilde{\boldsymbol{h}}_i^t\|_2^2$. It follows from the fundamental theorem of calculus [19, Theorem 4.2] that

$$\begin{aligned}&\widetilde{x}_{i1}^{t+1} - \widehat{x}_{i1}^{t+1,(l)}\\ =&\left\{\widetilde{x}_{i1}^t - \widehat{x}_{i1}^{t,(l)} - \eta'\left(\int_0^1 \boldsymbol{e}_1^\top\nabla_{\boldsymbol{x}_i}^2 f\left(\boldsymbol{z}\left(\tau\right)\right)\mathrm{d}\tau\right)\left[\begin{array}{c}\widetilde{\boldsymbol{x}}_i^t - \widehat{\boldsymbol{x}}_i^{t,(l)}\\ \overline{\widetilde{\boldsymbol{x}}_i^t - \widehat{\boldsymbol{x}}_i^{t,(l)}}\end{array}\right]\right\}\\ &-\eta'\left[\left(\sum_{k=1}^{s}\widehat{\boldsymbol{h}}_i^{t,(l)*}\boldsymbol{b}_l\boldsymbol{a}_{kl}^*\widehat{\boldsymbol{x}}_i^{t,(l)} - \boldsymbol{h}_k^{\natural*}\boldsymbol{b}_l\boldsymbol{a}_{kl}^*\boldsymbol{x}_k^\natural\right)\boldsymbol{b}_l^*\widehat{\boldsymbol{h}}_i^{t,(l)}a_{il,1}\right],\end{aligned} \tag{E.3}$$

where $\boldsymbol{z}(\tau) = \widetilde{\boldsymbol{z}}^t + \tau\left(\widehat{\boldsymbol{z}}^{t,(l)} - \widetilde{\boldsymbol{z}}^t\right)$ with $0 \leq \tau \leq 1$ and the Wirtinger Hessian with respect to $\boldsymbol{x}_i$ is

$$\nabla_{\boldsymbol{x}_i}^2 f(\boldsymbol{z}) = \left[\begin{array}{cc}\boldsymbol{D} & \boldsymbol{E}\\ \boldsymbol{E}^* & \overline{\boldsymbol{D}}\end{array}\right], \tag{E.4}$$

with $\boldsymbol{D} = \sum_{j=1}^m |\boldsymbol{b}_j^*\boldsymbol{h}_i|^2\boldsymbol{a}_{ij}\boldsymbol{a}_{ij}^*$ and $\boldsymbol{E} = \sum_{j=1}^m \boldsymbol{b}_j\boldsymbol{b}_j^*\boldsymbol{h}_i(\boldsymbol{a}_{ij}\boldsymbol{a}_{ij}^*\boldsymbol{x}_i)^\top$.

- We begin by controlling the second term of (E.3). Based on (A.2a) and the hypothesis (41a), we obtain

$$\max_{1\leq i\leq s, 1\leq l\leq m}\left|\boldsymbol{b}_l^*\widehat{\boldsymbol{h}}_i^{t,(l)}\right|\cdot\|\widehat{\boldsymbol{h}}_i^{t,(l)}\|_2^{-1} \lesssim \frac{\mu}{\sqrt{m}}\log^2 m$$

Along with the standard concentration results

$$\left|\boldsymbol{a}_{il}^*\boldsymbol{x}_i^{t,(l)}\right| \lesssim \sqrt{\log m}\|\boldsymbol{x}_i^{t,(l)}\|_2,$$

one has

$$\left|\left(\sum_{k=1}^{s}\widehat{\boldsymbol{h}}_i^{t,(l)*}\boldsymbol{b}_l\boldsymbol{a}_{kl}^*\widehat{\boldsymbol{x}}_i^{t,(l)}-\boldsymbol{h}_k^{\natural*}\boldsymbol{b}_l\boldsymbol{a}_{kl}^*\boldsymbol{x}_k^{\natural}\right)\boldsymbol{b}_l^*\widehat{\boldsymbol{h}}_i^{t,(l)}a_{il,1}\right|\lesssim\frac{s\mu^2\log^5 m}{m}\left\|\widehat{\boldsymbol{x}}_i^{t,(l)}\right\|_2 \tag{E.5}$$

- It remains to bound the first term in (E.3). To achieve this, we first utilize the decomposition

$$\boldsymbol{a}_{ij}^*\left(\widetilde{\boldsymbol{x}}_i^t-\widehat{\boldsymbol{x}}_i^{t,(l)}\right)=\overline{a_{ij,1}}\left(\widetilde{x}_{i1}^t-\widehat{x}_{i1}^{t,(l)}\right)+\boldsymbol{a}_{ij,\perp}^*\left(\widetilde{\boldsymbol{x}}_{i\perp}^t-\widehat{\boldsymbol{x}}_{i\perp}^{t,(l)}\right)$$

to obtain that

$$\boldsymbol{e}_1^\top\left(\nabla_{\boldsymbol{x}_i}^2 f\left(\boldsymbol{z}\left(\tau\right)\right)\mathrm{d}\tau\right)\left[\begin{array}{c}\widetilde{\boldsymbol{x}}_i^t-\widehat{\boldsymbol{x}}_i^{t,(l)}\\ \overline{\widetilde{\boldsymbol{x}}_i^t-\widehat{\boldsymbol{x}}_i^{t,(l)}}\end{array}\right]=\omega_1\left(\tau\right)+\omega_2\left(\tau\right)+\omega_3\left(\tau\right),$$

where

$$\omega_1\left(\tau\right)=\sum_{j=1}^{m}|\boldsymbol{b}_j^*\boldsymbol{h}_i(\tau)|^2 a_{ij,1}\overline{a_{ij,1}}\left(\widetilde{x}_{i1}^t-\widehat{x}_{i1}^{t,(l)}\right)$$
$$\omega_2\left(\tau\right)=\sum_{j=1}^{m}|\boldsymbol{b}_j^*\boldsymbol{h}_i(\tau)|^2 a_{ij,1}\boldsymbol{a}_{ij,\perp}^*\left(\widetilde{\boldsymbol{x}}_{i\perp}^t-\widehat{\boldsymbol{x}}_{i\perp}^{t,(l)}\right)$$
$$\omega_3\left(\tau\right)=\sum_{j=1}^{m}\boldsymbol{b}_j^*\boldsymbol{h}_i(\tau)\boldsymbol{a}_{ij}^*\boldsymbol{x}_i(\tau)b_{j,1}\boldsymbol{a}_{ij}^\top\left(\overline{\widetilde{\boldsymbol{x}}_i^t-\widehat{\boldsymbol{x}}_i^{t,(l)}}\right).$$

Based on Lemma 10, Lemma 14 and the fact $\|\boldsymbol{b}_j\|_2=\sqrt{K/m}$, by exploiting the techniques in Section D, $\omega_1(\tau)$, $\omega_2(\tau)$ and $\omega_3(\tau)$ can be bounded as follows:

$$\omega_1\left(\tau\right)=\|\boldsymbol{h}_i(\tau)\|_2^2\left(\widetilde{x}_{i1}^t-\widehat{x}_{i1}^{t,(l)}\right)+\mathcal{O}\left(\sqrt{\frac{s^2\mu^2K\log m}{m}}\left(\widetilde{x}_{i1}^t-\widehat{x}_{i1}^{t,(l)}\right)\right) \tag{E.6}$$

$$|\omega_2(\tau)|\lesssim\sqrt{\frac{K\log^2 m}{m}}\left(\left\|\widetilde{\boldsymbol{x}}_{i\perp}^t-\widehat{\boldsymbol{x}}_{i\perp}^{t,(l)}\right\|_2+\left\|\widetilde{\boldsymbol{x}}_{i\perp}^t-\widehat{\boldsymbol{x}}_{i\perp}^{t,(l)}-\widetilde{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}}-\widehat{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn},(l)}\right\|_2\right) \tag{E.7}$$

$$\omega_3\left(\tau\right)=|h_{i1}(\tau)|\left(\widetilde{\boldsymbol{x}}_i^t-\widehat{\boldsymbol{x}}_i^{t,(l)}\right)^*\boldsymbol{x}_i(\tau)+\mathcal{O}\left(\frac{1}{\log^5 m}\left\|\widetilde{\boldsymbol{x}}_i^t-\widehat{\boldsymbol{x}}_i^{t,(l)}\right\|_2\right) \tag{E.8}$$

with probability at least $1-\mathcal{O}(m^{-10})$, provided that $m\gg\mu^2K\log^{13}m$.

- Combining the bounds (E.5) (E.6), (E.7) and (E.8), one has

$$\begin{aligned}&\widetilde{x}_{i1}^{t+1}-\widehat{x}_{i1}^{t+1,(l)}\\&=\left(1-\eta\frac{\int_0^1\|\boldsymbol{h}_i(\tau)\|_2^2\mathrm{d}\tau}{\|\widetilde{\boldsymbol{h}}_i^t\|_2^2}+\mathcal{O}\left(\eta'\sqrt{\frac{s^2\mu^2K\log m}{m}}\right)\right)\cdot\left(\widetilde{x}_{i1}^t-\widehat{x}_{i1}^{t,(l)}\right)+\mathcal{O}\left(\eta'\frac{s\mu^2\log^5 m}{m}\left\|\widehat{\boldsymbol{x}}_i^{t,(l)}\right\|_2\right)\\&\quad+\mathcal{O}\left(\eta'\sqrt{\frac{K\log^2 m}{m}}\left(\left\|\widetilde{\boldsymbol{x}}_{i\perp}^t-\widehat{\boldsymbol{x}}_{i\perp}^{t,(l)}\right\|_2+\left\|\widetilde{\boldsymbol{x}}_{i\perp}^t-\widehat{\boldsymbol{x}}_{i\perp}^{t,(l)}-\widetilde{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}}-\widehat{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn},(l)}\right\|_2\right)\right),\\&\quad+\mathcal{O}\left(\eta'\frac{1}{\log^5 m}\left\|\widetilde{\boldsymbol{x}}_i^t-\widehat{\boldsymbol{x}}_i^{t,(l)}\right\|_2\right)+\eta'\sup_{0\le\tau\le1}|h_{i1}(\tau)|\left(\widetilde{\boldsymbol{x}}_i^t-\widehat{\boldsymbol{x}}_i^{t,(l)}\right)^*\boldsymbol{x}_i(\tau)\end{aligned}$$

For simplification, note that for the last term, for any $t<T_1\lesssim s\log\max\{K,N\}$, $0\le\tau\le1$ and $1\le i\le s$, one has

$$\begin{aligned}|x_{i1}\left(\tau\right)|&\lesssim\left|\widetilde{x}_{i1}^t\right|+\left|\widetilde{x}_{i1}^t-\widehat{x}_{i1}^{t,(l)}\right|\\&\le\alpha_{\boldsymbol{x}_i^t}+\alpha_{\boldsymbol{x}_i^t}\left(1+\frac{1}{s\log m}\right)^t C_2\frac{s\mu^2\kappa\sqrt{N\log^{13}m}}{m}\\&\lesssim\alpha_{\boldsymbol{x}_i^t},\end{aligned}$$

as long as $m \gg s\mu^2\kappa\sqrt{N\log^{13} m}$. In addition, there is

$$\begin{aligned}
&\left| |h_{i1}(\tau)| \cdot \left(\widetilde{\boldsymbol{x}}_i^t - \widehat{\boldsymbol{x}}_i^{t,(l)}\right)^* \boldsymbol{x}_i(\tau) \right| \\
\lesssim & |x_{i1}(\tau)| \cdot \left\| \widetilde{\boldsymbol{x}}_i^t - \widehat{\boldsymbol{x}}_i^{t,(l)} \right\|_2 \cdot \|\boldsymbol{x}_i(\tau)\|_2 \\
\lesssim & \alpha_{\boldsymbol{x}_i^t} \left\| \widetilde{\boldsymbol{x}}_i^t - \widehat{\boldsymbol{x}}_i^{t,(l)} \right\|_2
\end{aligned} \tag{E.9}$$

based on the fact $\|\boldsymbol{x}_i(\tau)\|_2 \lesssim 1$. Furthermore, we have

$$\begin{aligned}
\sqrt{\frac{K\log^2 m}{m}} \left\| \widetilde{\boldsymbol{x}}_{i\perp}^t - \widehat{\boldsymbol{x}}_{i\perp}^{t,(l)} \right\|_2 \leq & \sqrt{\frac{K\log^2 m}{m}} \left\| \widetilde{\boldsymbol{x}}_i^t - \widehat{\boldsymbol{x}}_i^{t,(l)} \right\|_2 \\
\lesssim & \alpha_{\boldsymbol{x}_i^t} \left\| \widetilde{\boldsymbol{x}}_i^t - \widehat{\boldsymbol{x}}_i^{t,(l)} \right\|_2,
\end{aligned}$$

as long as $m \gg K\log^{12} m$, in view of the assumption $\alpha_{\boldsymbol{x}_i^t} \ll 1/\log^5 m$. Therefore, we can further obtain

$$\begin{aligned}
&\left| \widetilde{x}_{i1}^{t+1} - \widehat{x}_{i1}^{t+1,(l)} \right| \\
\leq & (1-\eta+\eta'\varrho_1) \left| \widetilde{x}_{i1}^t - \widehat{x}_{i1}^{t,(l)} \right| \\
& + \mathcal{O}\left(\eta' \frac{s\mu^2\log^5 m}{m} \|\widetilde{\boldsymbol{x}}_i^t\|_2\right) + \mathcal{O}\left(\eta' \alpha_{\boldsymbol{x}_i^t} \left\| \widetilde{\boldsymbol{x}}_{i\perp}^t - \widehat{\boldsymbol{x}}_{i\perp}^{t,(l)} \right\|_2\right) + \mathcal{O}\left(\eta' \sqrt{\frac{K\log^2 m}{m}} \left\| \widetilde{\boldsymbol{x}}_{i\perp}^t - \widehat{\boldsymbol{x}}_{i\perp}^{t,(l)} - \widetilde{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn}} - \widehat{\boldsymbol{x}}_{i\perp}^{t,\mathrm{sgn},(l)} \right\|_2\right),
\end{aligned}$$

where $\eta' = \eta/\|\widetilde{\boldsymbol{h}}_i^t\|_2^2$, for some $|\varrho_1| \ll \frac{1}{\log m}$. Here the last inequality comes from the sample complexity $m \gg sK\log^5 m$ and the assumption $\alpha_t \ll \frac{1}{\log^5 m}$. Given the inductive hypotheses (41), we can conclude

$$\begin{aligned}
&\mathrm{dist}\left(x_{i1}^{t+1,(l)}, \widetilde{x}_{i1}^{t+1}\right) = \left| \widetilde{x}_{i1}^{t+1} - \widehat{x}_{i1}^{t+1,(l)} \right| \cdot \|\boldsymbol{x}_i^\natural\|_2^{-1} \\
\leq & \kappa \left| \widetilde{x}_{i1}^{t+1} - \widehat{x}_{i1}^{t+1,(l)} \right| \\
\leq & (1-\eta+\eta'\varrho_1)\, \alpha_{\boldsymbol{x}_i^t} \left(1 + \frac{1}{s\log m}\right)^t \cdot \\
& C_2 \frac{s\kappa\mu^2\sqrt{N\log^{13} m}}{m} \quad + \mathcal{O}\left(\eta' \frac{s\mu^2\kappa\log^5 m}{m} \left\{\alpha_{\boldsymbol{x}_i^t} + \beta_{\boldsymbol{x}_i^t}\right\}\right) \\
& + \mathcal{O}\left(\eta' \alpha_{\boldsymbol{x}_i^t} \beta_{\boldsymbol{x}_i^t} \left(1 + \frac{1}{s\log m}\right)^t C_1 \frac{s\kappa\mu^2\sqrt{N\log^8 m}}{m}\right) \\
& + \mathcal{O}\Bigg(\eta' \sqrt{\frac{K\log^2 m}{m}} \alpha_{\boldsymbol{x}_i^t} \left(1 + \frac{1}{s\log m}\right)^t \\
& \quad C_4 \frac{s\kappa\mu^2\sqrt{N\log^{16} m}}{m}\Bigg) \\
\overset{\text{(i)}}{\leq} & (1-\eta+\varrho_2)\, \alpha_{\boldsymbol{x}_i^t} \left(1 + \frac{1}{s\log m}\right)^t C_2 \frac{s\mu^2\kappa\sqrt{N\log^{13} m}}{m} \\
\overset{\text{(ii)}}{\leq} & \alpha_{\boldsymbol{x}_i^{t+1}} \left(1 + \frac{1}{s\log m}\right)^{t+1} C_2 \frac{s\mu^2\kappa\sqrt{N\log^{13} m}}{m}
\end{aligned}$$

for some $|\varrho_2| \ll \frac{1}{\log m}$. Here, the inequality (i) holds true as long as

$$\frac{s\mu^2\kappa\log^5 m}{m} \left(\alpha_{\boldsymbol{x}_i^t} + \beta_{\boldsymbol{x}_i^t}\right) \ll \frac{1}{\log m} \alpha_{\boldsymbol{x}_i^t} C_2 \frac{s\mu^2\kappa\sqrt{N\log^{13} m}}{m} \tag{E.10a}$$

$$\beta_{\boldsymbol{x}_i^t} C_1 \frac{s\mu^2\kappa\sqrt{N\log^8 m}}{m} \ll \frac{1}{\log m} \alpha_{\boldsymbol{x}_i^t} C_2 \frac{s\mu^2\kappa\sqrt{N\log^{13} m}}{m}, \tag{E.10b}$$

$$\sqrt{\frac{K\log^2 m}{m}} C_4 \frac{s\mu^2\kappa\sqrt{N\log^{16} m}}{m} \ll \frac{1}{\log m} C_2 \frac{s\mu^2\kappa\sqrt{N\log^{13} m}}{m}. \tag{E.10c}$$

Here, the first condition (E.10a) is satisfied since (according to Lemma 1)

$$\left(\alpha_{\boldsymbol{x}_i^t} + \beta_{\boldsymbol{x}_i^t}\right) \lesssim \beta_{\boldsymbol{x}_i^t} \lesssim \alpha_{\boldsymbol{x}_i^t} \sqrt{N\log m}.$$

The second condition (E.10b) holds based on $\beta_{\boldsymbol{x}_i^t} \lesssim \alpha_{\boldsymbol{x}_i^t}\sqrt{N\log m}$. The third one (E.10c) holds as long as $m \gg K\log^7 m$. Moreover, we get that for some $|\varrho_1| \ll \frac{1}{\log m}$,

$$\begin{aligned}
(1-\eta+\varrho_2)\,\alpha_{\boldsymbol{x}_i^t} &= \left\{ \frac{\alpha_{\boldsymbol{x}_i^{t+1}}}{\alpha_{\boldsymbol{x}_i^t}} + \eta\varrho_3 \right\} \alpha_{\boldsymbol{x}_i^t} \\
&\le \left\{ \frac{\alpha_{\boldsymbol{x}_i^{t+1}}}{\alpha_{\boldsymbol{x}_i^t}} + \eta\mathcal{O}\left(\frac{\alpha_{\boldsymbol{x}_i^{t+1}}}{\alpha_{\boldsymbol{x}_i^t}} \varrho_3 \right) \right\} \alpha_{\boldsymbol{x}_i^t} \\
&\le \alpha_{\boldsymbol{x}_i^{t+1}} \left(1 + \frac{1}{s\log m} \right),
\end{aligned}$$

as long as $\alpha_{\boldsymbol{x}_i^{t+1}} / \alpha_{\boldsymbol{x}_i^t} \asymp 1$.

F Technical Lemmas

Lemma 11. *Suppose $m \gg K\log^3 m$. With probability exceeding $1-\mathcal{O}\left(m^{-10}\right)$, we have*

$$\left\| \sum_{j=1}^m \overline{a_{ij,1}} a_{ij,1} \boldsymbol{b}_j \boldsymbol{b}_j^* - \boldsymbol{I}_K \right\| \lesssim \sqrt{\frac{K}{m}\log m}.$$

Lemma 12. *Suppose $m \gg K\log^3 m$. For $k \neq i$, we have*

$$\left\| \sum_{j=1}^m \overline{a_{kj,1}} a_{ij,1} \boldsymbol{b}_j \boldsymbol{b}_j^* \right\| \lesssim \sqrt{\frac{K}{m}\log m},$$

$$\left\| \sum_{j=1}^m |a_{kj,1}|\, |a_{ij,1}|\, \boldsymbol{b}_j \boldsymbol{b}_j^* \right\| \lesssim \sqrt{\frac{K}{m}\log m},$$

with probability exceeding $1-\mathcal{O}\left(m^{-10}\right)$.

Lemma 13. *Suppose $m \gg K\log^3 m$ and $\boldsymbol{z} \in \mathbb{C}^{N-1}$ with $\|\boldsymbol{z}\|_2 = 1$ is independent with $\{\boldsymbol{a}_{kj}\}$. With probability exceeding $1-\mathcal{O}\left(m^{-10}\right)$, we have*

$$\left\| \sum_{j=1}^m a_{ij,1} \boldsymbol{a}_{kj,\perp}^* \boldsymbol{z} \boldsymbol{b}_j \boldsymbol{b}_j^* \right\| \lesssim \sqrt{\frac{K}{m}\log m}.$$

Remark 3. *Lemma 12, Lemma 13 and Lemma 11 can be proven by applying the arguments in [18, Section D.3.3].*

Lemma 14. *Suppose $m \gg (\mu^2/\delta^2) N\log^5 m$. With probability exceeding $1-\mathcal{O}\left(m^{-10}\right)$, we have*

$$\left\| \sum_{j=1}^m \left|\boldsymbol{b}_j^* \boldsymbol{h}_i\right|^2 \boldsymbol{a}_{ij,\perp} \boldsymbol{a}_{ij,\perp}^* - \|\boldsymbol{h}_i\|_2^2 \boldsymbol{I}_{N-1} \right\| \lesssim \delta \|\boldsymbol{h}_i\|_2^2,$$

obeying $\max_{1\le l\le m} |\boldsymbol{b}_l^* \boldsymbol{h}_i| \cdot \left\|\boldsymbol{h}_i\right\|_2^{-1} \lesssim \frac{\mu}{\sqrt{m}} \log^2 m$*. Furthermore, there is*

$$\left\| \sum_{j=1}^{m} \sum_{k=1}^{s} b_{j,1} \boldsymbol{b}_j^* \boldsymbol{h}_i \boldsymbol{a}_{ij} \boldsymbol{a}_{kj}^* - h_{i1} \boldsymbol{I}_N \right\| \lesssim \delta \left\|\boldsymbol{h}_i\right\|_2,$$

with probability exceeding $1 - \mathcal{O}\left(m^{-10}\right)$*, provided* $m \gg (\mu/\delta^2) s^2 N \log^3 m$*.*

Proof. Please refer to Lemma 11 and Lemma 12 in [2]. □

Lemma 15. *Suppose* $m \gg s\mu^2 \sqrt{N \log^9 m}$*, then with probability exceeding* $1 - \mathcal{O}\left(m^{-10}\right)$*, we have*

$$\left\| \sum_{j=1}^{m} \sum_{k=1}^{s} \boldsymbol{h}_k^* \boldsymbol{b}_j \boldsymbol{b}_j^* \boldsymbol{h}_i \boldsymbol{a}_{ij} \boldsymbol{a}_{kj}^* - \left\|\boldsymbol{h}_i\right\|_2^2 \boldsymbol{I}_N \right\| \lesssim \frac{s\mu^2 \sqrt{K \log^9 m}}{m} \left\|\boldsymbol{h}_i\right\|_2^2, \tag{F.1}$$

obeying $\max_{1\le i\le s, 1\le j\le m} \left|\boldsymbol{b}_j^* \boldsymbol{h}_i\right| \cdot \left\|\boldsymbol{h}_i\right\|_2^{-1} \lesssim \frac{\mu}{\sqrt{m}} \log^2 m$*.*

Lemma 16. *Suppose* $m \gg s\mu^2 \sqrt{N \log^5 m}$*. With probability exceeding* $1 - \mathcal{O}\left(m^{-10}\right)$*, we have*

$$\left\| \sum_{j=1}^{m} \sum_{k=1}^{s} \boldsymbol{h}_k^{\natural *} \boldsymbol{b}_j \boldsymbol{b}_j^* \boldsymbol{h}_i \boldsymbol{a}_{ij} \boldsymbol{a}_{kj}^* - (\boldsymbol{h}_i^{\natural *} \boldsymbol{h}_i) \boldsymbol{I}_N \right\| \lesssim \frac{s\mu^2 \sqrt{K \log^5 m}}{m} \left|\boldsymbol{h}_i^{\natural *} \boldsymbol{h}_i\right|, \tag{F.2}$$

obeying $\max_{1\le l\le m} \left|\boldsymbol{b}_l^* \boldsymbol{h}_i^{\natural}\right| \cdot \left\|\boldsymbol{h}_i^{\natural}\right\|_2^{-1} \le \frac{\mu}{\sqrt{m}}$ *and* $\max_{1\le l\le m} |\boldsymbol{b}_l^* \boldsymbol{h}_i| \cdot \left\|\boldsymbol{h}_i\right\|_2^{-1} \lesssim \frac{\mu}{\sqrt{m}} \log^2 m$*.*

Remark 4. *The proof of Lemma 15 and 16 exploits the same strategy as [16, Section K] does.*

Lemma 17. *Suppose that* $\boldsymbol{a}_{ij}$ *and* $\boldsymbol{b}_j$ *follows the definition in the main text.* $1 \le i \le s, 1 \le j \le m$*. Consider any* $\epsilon > 3/n$ *where* $n = \max\{K, N\}$*. Let*

$$\mathcal{S} := \left\{ \boldsymbol{z} \in \mathbb{C}^{N-1} \middle| \max_{1\le j\le m} \left|\boldsymbol{a}_{ij,\perp}^* \boldsymbol{z}\right| \le \beta \left\|\boldsymbol{z}\right\|_2 \right\},$$

where β *is any value obeying* $\beta \ge c_1 \sqrt{\log m}$ *for some sufficiently large constant* $c_1 > 0$*. Then with probability exceeding* $1 - \mathcal{O}\left(m^{-10}\right)$*, one has*

1. $\left|\sum_{j=1}^{m} |a_{ij,1}|^2 \left|\boldsymbol{a}_{kj\perp}^* \boldsymbol{z}\right|^2 \boldsymbol{b}_j \boldsymbol{b}_j^* - \left\|\boldsymbol{z}\right\|_2 \boldsymbol{I}_K\right| \le \epsilon \left\|\boldsymbol{z}\right\|_2$ *for all* $\boldsymbol{z} \in \mathcal{S}$*, provided that* $m \ge c_0 \max\left\{\frac{1}{\epsilon^2} n \log n,\ \frac{1}{\epsilon} \beta^2 n \log^2 m\right\}$*;*
2. $\left|\sum_{j=1}^{m} |a_{ij,1}| \left|\boldsymbol{a}_{kj\perp}^* \boldsymbol{z}\right| \boldsymbol{b}_j \boldsymbol{b}_j^*\right| \le \epsilon \left\|\boldsymbol{z}\right\|_2$ *for all* $\boldsymbol{z} \in \mathcal{S}$*, provided that* $m \ge c_0 \max\left\{\frac{1}{\epsilon^2} n \log n,\ \frac{1}{\epsilon} \beta n \log^{\frac{1}{2}} m\right\}$*;*

Here, $c_0 > 0$ *is some sufficiently large constant.*

Proof. Please refer to Lemma 12 in [16]. □