A. Proof of Theorem 3.8

In order to proof Theorem 3.8, we need to make use of the following lemma, which is derived from the restricted strong convexity and smoothness property of $\mathcal{F}_n$. The proof of Lemma A.1 is presented in Section C.

Lemma A.1. Assume the sample loss function $\mathcal{F}_n$ satisfies Conditions 3.5 and 3.6. Then for all matrices $\mathbf{Y} \in \mathbb{R}^{d_1 \times d_2}$ with rank at most $2r$ and $\mathbf{W} \in \mathbb{R}^{d_1 \times d_2}$ with rank at most $4r$, we have

$$\mu \|\mathbf{W}\|_F^2 \leq \mathrm{vec}(\mathbf{W})^\top \nabla^2 \mathcal{F}_n(\mathbf{Y}) \mathrm{vec}(\mathbf{W}) \leq L \|\mathbf{W}\|_F^2.$$

Moreover, for all matrices $\mathbf{W}_1, \mathbf{W}_2 \in \mathbb{R}^{d_1 \times d_2}$ with rank at most $2r$, we have

$$\left|2\mathrm{vec}(\mathbf{W}_1)^\top \nabla^2 \mathcal{F}_n(\mathbf{Y}) \mathrm{vec}(\mathbf{W}_2) - (L+\mu)\langle \mathbf{W}_1, \mathbf{W}_2 \rangle\right| \leq \frac{L-\mu}{2}(\|\mathbf{W}_1\|_F^2 + \|\mathbf{W}_2\|_F^2).$$

Now we are ready to prove Theorem 3.8.

Proof of Theorem 3.8. Recall $\mathbf{Z} = [\mathbf{U}; \mathbf{V}]$ is the local minimizer of constrained optimization problem (3.1). Since $\nabla h_i(\mathbf{Z}) = 2\mathbf{e}_i\mathbf{e}_i^\top \mathbf{Z}$ are linearly independent for all $i \in [d_1 + d_2]$, thus there exists $\boldsymbol{\lambda} \geq 0$ such that $(\mathbf{Z}, \boldsymbol{\lambda})$ is a KKT pair, which satisfies the conditions listed in Lemma 3.3. Denote $\mathbf{X} = \mathbf{U}\mathbf{V}^\top$ and $\widetilde{\mathbf{Z}} = [\mathbf{U}; -\mathbf{V}]$. Then according to the Lagrangian function for optimization problem (3.1), we can calculate its gradient with respect to $\mathbf{Z}$ as follows

$$\begin{aligned}\nabla_{\mathbf{Z}} \mathcal{L}(\mathbf{Z}, \boldsymbol{\lambda}) &= \nabla_{\mathbf{Z}} \mathcal{F}_n(\mathbf{U}\mathbf{V}^\top) + \frac{\gamma}{4} \nabla_{\mathbf{Z}} \big[\|\mathbf{U}^\top\mathbf{U} - \mathbf{V}^\top\mathbf{V}\|_F^2\big] + \sum_{i=1}^{d_1+d_2} \lambda_i \nabla h_i(\mathbf{Z}) \\ &= \begin{bmatrix} \nabla \mathcal{F}_n(\mathbf{X})\mathbf{V} \\ \nabla \mathcal{F}_n(\mathbf{X})^\top \mathbf{U} \end{bmatrix} + \gamma \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top \mathbf{Z} + 2\sum_{i=1}^{d_1+d_2} \lambda_i \mathbf{e}_i\mathbf{e}_i^\top \mathbf{Z}. \end{aligned} \tag{A.1}$$

Moreover, for any matrix $\boldsymbol{\Delta} \in \mathbb{R}^{(d_1+d_2)\times r}$, denote $\boldsymbol{\Delta} = [\boldsymbol{\Delta}_U; \boldsymbol{\Delta}_V]$, where $\boldsymbol{\Delta}_U \in \mathbb{R}^{d_1 \times r}$, $\boldsymbol{\Delta}_V \in \mathbb{R}^{d_2 \times r}$, then we have

$$\begin{aligned}\mathrm{vec}(\boldsymbol{\Delta})^\top \nabla^2_{\mathbf{Z}} \mathcal{L}(\mathbf{Z}, \boldsymbol{\lambda}) \mathrm{vec}(\boldsymbol{\Delta}) &= \mathrm{vec}(\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top)^\top \nabla^2 \mathcal{F}_n(\mathbf{X}) \mathrm{vec}(\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top) \\ &\quad + 2\langle \nabla \mathcal{F}_n(\mathbf{X}), \boldsymbol{\Delta}_U\boldsymbol{\Delta}_V^\top \rangle + \gamma \langle \boldsymbol{\Delta}, \widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{Z}}^\top\mathbf{Z} + \widetilde{\mathbf{Z}}\widetilde{\boldsymbol{\Delta}}^\top\mathbf{Z} + \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\boldsymbol{\Delta} \rangle + 2\sum_{i=1}^{d_1+d_2} \lambda_i \langle \mathbf{e}_i\mathbf{e}_i^\top, \boldsymbol{\Delta}\boldsymbol{\Delta}^\top \rangle. \end{aligned} \tag{A.2}$$

Let $\mathbf{R}$ be the optimal rotation with respect to $\mathbf{Z}$ and $\mathbf{Z}^*$, i.e., $\mathbf{R} = \mathrm{argmin}_{\widetilde{\mathbf{R}} \in \mathbb{Q}_r} \|\mathbf{Z} - \mathbf{Z}^*\widetilde{\mathbf{R}}\|_F$, where $\mathbb{Q}_r$ is the set of r-by-r orthogonal matrices. For any $i \in [d_1 + d_2]$, if $h_i(\mathbf{Z}) = 0$, then we have

$$\langle \nabla h_i(\mathbf{Z}), \mathbf{Z}^*\mathbf{R} - \mathbf{Z} \rangle = 2\langle \mathbf{e}_i\mathbf{e}_i^\top\mathbf{Z}, \mathbf{Z}^*\mathbf{R} \rangle - 2\langle \mathbf{e}_i\mathbf{e}_i^\top\mathbf{Z}, \mathbf{Z} \rangle \leq 2\|\mathbf{Z}_{i,*}\|_2 \cdot \|\mathbf{Z}^*_{i,*}\|_2 - 2\|\mathbf{Z}_{i,*}\|_2^2 \leq 0,$$

where the first inequality follows from Cauchy-Schwarz inequality, and the second inequality holds because $\|\mathbf{Z}_{i,*}\|_2 = \alpha$, $\|\mathbf{Z}^*_{i,*}\|_2 \leq \|\mathbf{Z}^*\|_{2,\infty} \leq \alpha$. Thus according to Lemma 3.4, we obtain

$$\mathrm{vec}(\mathbf{Z}^*\mathbf{R} - \mathbf{Z})^\top \nabla^2_{\mathbf{Z}} \mathcal{L}(\mathbf{Z}, \boldsymbol{\lambda}) \mathrm{vec}(\mathbf{Z}^*\mathbf{R} - \mathbf{Z}) \geq 0. \tag{A.3}$$

Denote $\boldsymbol{\Delta} = \mathbf{Z} - \mathbf{Z}^*\mathbf{R}$, then according to (A.2), we further obtain the equivalent form of (A.3)

$$\begin{aligned}&\underbrace{\mathrm{vec}(\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top)^\top \nabla^2 \mathcal{F}_n(\mathbf{X}) \mathrm{vec}(\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top)}_{I_1} \\ &\quad + \underbrace{2\langle \nabla \mathcal{F}_n(\mathbf{X}), \boldsymbol{\Delta}_U\boldsymbol{\Delta}_V^\top \rangle + 2\textstyle\sum_{i=1}^{d_1+d_2} \lambda_i \langle \mathbf{e}_i\mathbf{e}_i^\top, \boldsymbol{\Delta}\boldsymbol{\Delta}^\top \rangle}_{I_2} + \underbrace{\gamma \langle \boldsymbol{\Delta}, \widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{Z}}^\top\mathbf{Z} + \widetilde{\mathbf{Z}}\widetilde{\boldsymbol{\Delta}}^\top\mathbf{Z} + \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\boldsymbol{\Delta} \rangle}_{I_3} \geq 0.\end{aligned}$$

Consider I_1 first. Since $\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top = \mathbf{X} - \mathbf{X}^* + \boldsymbol{\Delta}_U\boldsymbol{\Delta}_V^\top$, thus we have

$$\begin{aligned} I_1 &= \mathrm{vec}(\mathbf{X}-\mathbf{X}^*)^\top \nabla^2\mathcal{F}_n(\mathbf{X})\mathrm{vec}(\mathbf{X}-\mathbf{X}^*) + \mathrm{vec}(\boldsymbol{\Delta}_U\boldsymbol{\Delta}_V^\top)^\top\nabla^2\mathcal{F}_n(\mathbf{X})\mathrm{vec}(\boldsymbol{\Delta}_U\boldsymbol{\Delta}_V^\top) \\ &\quad + 2\mathrm{vec}(\mathbf{X}-\mathbf{X}^*)^\top\nabla^2\mathcal{F}_n(\mathbf{X})\mathrm{vec}(\boldsymbol{\Delta}_U\boldsymbol{\Delta}_V^\top) \\ &= -\mathrm{vec}(\mathbf{X}-\mathbf{X}^*)^\top \nabla^2\mathcal{F}_n(\mathbf{X})\mathrm{vec}(\mathbf{X}-\mathbf{X}^*) + \mathrm{vec}(\boldsymbol{\Delta}_U\boldsymbol{\Delta}_V^\top)^\top\nabla^2\mathcal{F}_n(\mathbf{X})\mathrm{vec}(\boldsymbol{\Delta}_U\boldsymbol{\Delta}_V^\top) \\ &\quad + 2\mathrm{vec}(\mathbf{X}-\mathbf{X}^*)^\top\nabla^2\mathcal{F}_n(\mathbf{X})\mathrm{vec}(\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top) \\ &\le -\mu\|\mathbf{X}-\mathbf{X}^*\|_F^2 + L\|\boldsymbol{\Delta}_U\boldsymbol{\Delta}_V^\top\|_F^2 + \underbrace{2\mathrm{vec}(\mathbf{X}-\mathbf{X}^*)^\top\nabla^2\mathcal{F}_n(\mathbf{X})\mathrm{vec}(\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top)}_{I_{11}}, \end{aligned} \tag{A.4}$$

where the inequality follows from Lemma A.1. Next, we are going to prove that I_{11} is close to $2\langle\nabla\mathcal{F}_n(\mathbf{X}) - \nabla\mathcal{F}_n(\mathbf{X}^*), \mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\rangle$. More specifically, according to Lemma A.1, we have

$$\big|I_{11} - (L+\mu)\langle\mathbf{X}-\mathbf{X}^*, \mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\rangle\big| \le \frac{L-\mu}{2}(\|\mathbf{X}-\mathbf{X}^*\|_F^2 + \|\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\|_F^2). \tag{A.5}$$

Besides, according to the integral form of the mean value theorem, we have

$$\begin{aligned} &\big|2\langle\nabla\mathcal{F}_n(\mathbf{X}) - \nabla\mathcal{F}_n(\mathbf{X}^*), \mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\rangle - (L+\mu)\langle\mathbf{X}-\mathbf{X}^*, \mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\rangle\big| \\ \le& \Big|\int_0^1 2\mathrm{vec}(\mathbf{X}-\mathbf{X}^*)^\top\nabla^2\mathcal{F}_n(t\mathbf{X} + (1-t)\mathbf{X}^*)\mathrm{vec}(\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top)dt \\ &- (L+\mu)\langle\mathbf{X}-\mathbf{X}^*, \mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\rangle\Big| \\ \le& \int_0^1 \frac{L-\mu}{2}(\|\mathbf{X}-\mathbf{X}^*\|_F^2 + \|\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\|_F^2)dt \\ =& \frac{L-\mu}{2}(\|\mathbf{X}-\mathbf{X}^*\|_F^2 + \|\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\|_F^2), \end{aligned} \tag{A.6}$$

where the second inequality follows from Lemma A.1. Combining (A.5) and (A.6), we obtain

$$\big|I_{11} - 2\langle\nabla\mathcal{F}_n(\mathbf{X}) - \nabla\mathcal{F}_n(\mathbf{X}^*), \mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\rangle\big| \le (L-\mu)\cdot(\|\mathbf{X}-\mathbf{X}^*\|_F^2 + \|\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\|_F^2), \tag{A.7}$$

which implies that I_{11} is close to $2\langle\nabla\mathcal{F}_n(\mathbf{X}) - \nabla\mathcal{F}_n(\mathbf{X}^*), \mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\rangle$, as long as $(L-\mu)$ is small enough. Noticing that $-\boldsymbol{\Delta}$ is a feasible direction for problem (3.1), according to Lemma 3.2, we have

$$\begin{aligned} \langle\nabla\mathcal{G}(\mathbf{Z}), \boldsymbol{\Delta}\rangle &= \langle\nabla_{\mathbf{Z}}\mathcal{F}_n(\mathbf{U}\mathbf{V}^\top) + \gamma\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\mathbf{Z}, \boldsymbol{\Delta}\rangle \\ &= \langle\nabla\mathcal{F}_n(\mathbf{X}), \mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\rangle + \gamma\langle\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\mathbf{Z}, \boldsymbol{\Delta}\rangle \le 0. \end{aligned} \tag{A.8}$$

Therefore, we further obtain the upper bound of I_{11} as follows

$$\begin{aligned} I_{11} &\le 2\langle\nabla\mathcal{F}_n(\mathbf{X}) - \nabla\mathcal{F}_n(\mathbf{X}^*), \mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\rangle + (L-\mu)\cdot(\|\mathbf{X}-\mathbf{X}^*\|_F^2 + \|\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\|_F^2) \\ &\le -2\gamma\langle\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\mathbf{Z}, \boldsymbol{\Delta}\rangle + 2\big|\langle\nabla\mathcal{F}_n(\mathbf{X}^*), \mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\big| + (L-\mu)\cdot(\|\mathbf{X}-\mathbf{X}^*\|_F^2 + \|\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\|_F^2) \\ &\le -2\gamma\langle\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\mathbf{Z}, \boldsymbol{\Delta}\rangle + (L-\mu)\cdot(\|\mathbf{X}-\mathbf{X}^*\|_F^2 + \|\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\|_F^2) \\ &\quad + 2\sqrt{2r}\|\nabla\mathcal{F}_n(\mathbf{X}^*)\|_2\cdot\|\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\|_F^2, \end{aligned} \tag{A.9}$$

where the first inequality follows from (A.7), the second inequality follows from (A.8), and the last inequality holds because $|\langle\mathbf{A}, \mathbf{B}\rangle| \le \|\mathbf{A}\|_2\cdot\|\mathbf{B}\|_*$ and $(\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top)$ has rank at most $2r$. Hence, combining (A.4) and (A.9), we obtain the upper bound of I_1 as follows

$$\begin{aligned} I_1 &\le (3L-4\mu)\|\mathbf{X}-\mathbf{X}^*\|_F^2 + (3L-2\mu)\|\boldsymbol{\Delta}_U\boldsymbol{\Delta}_V^\top\|_F^2 - 2\gamma\langle\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\mathbf{Z}, \boldsymbol{\Delta}\rangle \\ &\quad + 2\sqrt{2r}\|\nabla\mathcal{F}_n(\mathbf{X}^*)\|_2\cdot\|\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\|_F^2, \end{aligned} \tag{A.10}$$

where the inequality follows from the fact that $\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top = \mathbf{X} - \mathbf{X}^* + \boldsymbol{\Delta}_U\boldsymbol{\Delta}_V^\top$, and $\|\mathbf{A}+\mathbf{B}\|_F^2 \le 2\|\mathbf{A}\|_F^2 + 2\|\mathbf{B}\|_F^2$. Furthermore, we turn to upper bound I_2. To begin with, we have

$$
\begin{aligned}
I_2 &= 2\langle \nabla \mathcal{F}_n(\mathbf{X}), \mathbf{X}^* - \mathbf{X} + \mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\rangle + 2\sum_{i=1}^{d_1+d_2} \lambda_i \langle \mathbf{e}_i\mathbf{e}_i^\top, \mathbf{Z}^*\mathbf{Z}^{*\top} - \mathbf{Z}\mathbf{Z}^\top + \mathbf{Z}\boldsymbol{\Delta}^\top + \boldsymbol{\Delta}\mathbf{Z}^\top\rangle \\
&\le 2\underbrace{\langle \nabla \mathcal{F}_n(\mathbf{X}), \mathbf{X}^* - \mathbf{X}\rangle}_{I_{21}} + 2\underbrace{\sum_{i=1}^{d_1+d_2} \lambda_i [\mathbf{Z}^*\mathbf{Z}^{*\top} - \mathbf{Z}\mathbf{Z}^\top]_{ii}}_{I_{22}} + 2\underbrace{\Big\langle \nabla_{\mathbf{Z}} \mathcal{F}_n(\mathbf{U}\mathbf{V}^\top) + 2\sum_{i=1}^{d_1+d_2}\lambda_i \mathbf{e}_i\mathbf{e}_i^\top\mathbf{Z}, \boldsymbol{\Delta}\Big\rangle}_{I_{23}}.
\end{aligned}
$$

According to the restricted strong convexity Condition 3.5, we can upper bound I_{21} as follows

$$
\begin{aligned}
I_{21} &= -\langle \nabla \mathcal{F}_n(\mathbf{X}) - \nabla \mathcal{F}_n(\mathbf{X}^*), \mathbf{X} - \mathbf{X}^*\rangle - \langle \nabla \mathcal{F}_n(\mathbf{X}^*), \mathbf{X} - \mathbf{X}^*\rangle \\
&\le -\mu\|\mathbf{X} - \mathbf{X}^*\|_F^2 + |\langle \nabla \mathcal{F}_n(\mathbf{X}^*), \mathbf{X} - \mathbf{X}^*\rangle|. \qquad \text{(A.11)}
\end{aligned}
$$

Denote index set $\mathcal{I} = \{i \in [d_1+d_2] \mid h_i(\mathbf{Z}) = 0\}$, then according to the complimentary slacknees condition in Lemma 3.3, we have $\lambda_i h_i(\mathbf{Z}) = 0, \forall i \in [d_1+d_2]$, which implies that $\lambda_i = 0$, if $i \notin \mathcal{I}$. Therefore, we have

$$
I_{22} = \sum_{i\in\mathcal{I}} \lambda_i([\mathbf{Z}^*\mathbf{Z}^{*\top}]_{ii} - [\mathbf{Z}\mathbf{Z}^\top]_{ii}) \le \sum_{i\in\mathcal{I}} \lambda_i(\|\mathbf{Z}_{i,*}^*\|_{2,\infty}^2 - \alpha^2) \le 0. \qquad \text{(A.12)}
$$

According to the stationarity condition in Lemma 3.3, we have $I_{23} = -\gamma\langle \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\mathbf{Z}, \boldsymbol{\Delta}\rangle$. Combining (A.11) and (A.12), we obtain

$$
\begin{aligned}
I_2 &\le -2\mu\|\mathbf{X} - \mathbf{X}^*\|_F^2 + 2|\langle \nabla \mathcal{F}_n(\mathbf{X}^*), \mathbf{X} - \mathbf{X}^*\rangle| - 2\gamma\langle \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\mathbf{Z}, \boldsymbol{\Delta}\rangle \\
&\le -2\mu\|\mathbf{X} - \mathbf{X}^*\|_F^2 + 2\sqrt{2r}\|\nabla \mathcal{F}_n(\mathbf{X}^*)\|_2 \cdot \|\mathbf{X} - \mathbf{X}^*\|_F - 2\gamma\langle \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\mathbf{Z}, \boldsymbol{\Delta}\rangle, \qquad \text{(A.13)}
\end{aligned}
$$

where the last inequality is due to $|\langle \mathbf{A}, \mathbf{B}\rangle| \le \|\mathbf{A}\|_2 \cdot \|\mathbf{B}\|_*$ and the fact that $(\mathbf{X} - \mathbf{X}^*)$ has rank at most $2r$. Therefore, combining (A.10) and (A.13), we have

$$
\begin{aligned}
I_1 + I_2 &\le (3L - 6\mu)\|\mathbf{X} - \mathbf{X}^*\|_F^2 + (3L - 2\mu)\|\boldsymbol{\Delta}_U\boldsymbol{\Delta}_V^\top\|_F^2 - 4\gamma\langle \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\mathbf{Z}, \boldsymbol{\Delta}\rangle \\
&\quad + 2\sqrt{2r}\|\nabla \mathcal{F}_n(\mathbf{X}^*)\|_2 \cdot (\|\mathbf{U}\boldsymbol{\Delta}_V^\top + \boldsymbol{\Delta}_U\mathbf{V}^\top\|_F + \|\mathbf{X} - \mathbf{X}^*\|_F). \qquad \text{(A.14)}
\end{aligned}
$$

Finally, we are going to upper bound the remaining term $I_3 - 4\gamma\langle \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\mathbf{Z}, \boldsymbol{\Delta}\rangle$. Recall $\widetilde{\mathbf{Z}} = [\mathbf{U}; -\mathbf{V}]$, and denote $\widetilde{\boldsymbol{\Delta}} = [\boldsymbol{\Delta}_U; -\boldsymbol{\Delta}_V]$. According to the definition of I_3, we have

$$
\begin{aligned}
I_3 - 4\gamma\langle \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\mathbf{Z}, \boldsymbol{\Delta}\rangle &= \gamma\langle \widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{Z}}^\top\mathbf{Z} + \widetilde{\mathbf{Z}}\widetilde{\boldsymbol{\Delta}}^\top\mathbf{Z} + \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\boldsymbol{\Delta} - 4\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\mathbf{Z}, \boldsymbol{\Delta}\rangle \\
&= \frac{\gamma}{2}\langle \widetilde{\mathbf{Z}}\widetilde{\boldsymbol{\Delta}}^\top + \widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{Z}}^\top, \mathbf{Z}\boldsymbol{\Delta}^\top + \boldsymbol{\Delta}\mathbf{Z}^\top\rangle + \gamma\langle \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top, \boldsymbol{\Delta}\boldsymbol{\Delta}^\top\rangle - 2\gamma\langle \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top, \boldsymbol{\Delta}\mathbf{Z} + \mathbf{Z}\boldsymbol{\Delta}^\top\rangle. \qquad \text{(A.15)}
\end{aligned}
$$

Denote $\widetilde{\mathbf{Z}}^* = [\mathbf{U}^*; -\mathbf{V}^*]$. Given the fact that $\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}^* = 0$, we have

$$
\begin{aligned}
\langle \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top, \boldsymbol{\Delta}\boldsymbol{\Delta}^\top\rangle &= \langle \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top - \widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}, \boldsymbol{\Delta}\boldsymbol{\Delta}^\top\rangle + \langle \widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}, \boldsymbol{\Delta}\boldsymbol{\Delta}^\top\rangle \\
&= \langle \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top - \widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}, \boldsymbol{\Delta}\boldsymbol{\Delta}^\top\rangle + \langle \widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}, \mathbf{Z}\mathbf{Z}^\top\rangle. \qquad \text{(A.16)}
\end{aligned}
$$

Similarly, we have

$$
\langle \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top, \boldsymbol{\Delta}\mathbf{Z} + \mathbf{Z}\boldsymbol{\Delta}^\top\rangle = \langle \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top - \widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}, \boldsymbol{\Delta}\mathbf{Z} + \mathbf{Z}\boldsymbol{\Delta}^\top\rangle + 2\langle \widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}, \mathbf{Z}\mathbf{Z}^\top\rangle. \qquad \text{(A.17)}
$$

Thus, plugging (A.16) and (A.17) into (A.15), we obtain

$$
\begin{aligned}
I_3 - 4\gamma\langle \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\mathbf{Z}, \boldsymbol{\Delta}\rangle &= \frac{\gamma}{2}\underbrace{\langle \widetilde{\mathbf{Z}}\widetilde{\boldsymbol{\Delta}}^\top + \widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{Z}}^\top, \mathbf{Z}\boldsymbol{\Delta}^\top + \boldsymbol{\Delta}\mathbf{Z}^\top\rangle}_{I_{31}} + \gamma\underbrace{\langle \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top - \widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}, \boldsymbol{\Delta}\boldsymbol{\Delta}^\top\rangle}_{I_{32}} \\
&\quad - 2\gamma\underbrace{\langle \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top - \widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}, \boldsymbol{\Delta}\mathbf{Z} + \mathbf{Z}\boldsymbol{\Delta}^\top\rangle}_{I_{33}} - 3\gamma\underbrace{\langle \widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}, \mathbf{Z}\mathbf{Z}^\top\rangle}_{I_{34}}. \qquad \text{(A.18)}
\end{aligned}
$$

Moreover, we have

$$\begin{aligned}\frac{\gamma}{2}I_{31}-\gamma I_{33}&=\frac{\gamma}{2}\langle\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top-\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}+\widetilde{\boldsymbol{\Delta}}\widetilde{\boldsymbol{\Delta}}^\top,\boldsymbol{\Delta}\mathbf{Z}^\top+\mathbf{Z}\boldsymbol{\Delta}^\top\rangle-\gamma\langle\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top-\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top},\boldsymbol{\Delta}\mathbf{Z}^\top+\mathbf{Z}\boldsymbol{\Delta}^\top\rangle\\&=\frac{\gamma}{2}\langle\widetilde{\boldsymbol{\Delta}}\widetilde{\boldsymbol{\Delta}}^\top,\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\rangle+\frac{\gamma}{2}\langle\widetilde{\boldsymbol{\Delta}}\widetilde{\boldsymbol{\Delta}}^\top,\mathbf{Z}\mathbf{Z}^\top-\mathbf{Z}^*\mathbf{Z}^{*\top}\rangle-\frac{\gamma}{2}\langle\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top-\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top},\boldsymbol{\Delta}\mathbf{Z}^\top+\mathbf{Z}\boldsymbol{\Delta}^\top\rangle\\&=\frac{\gamma}{2}\langle\widetilde{\boldsymbol{\Delta}}\widetilde{\boldsymbol{\Delta}}^\top,\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\rangle-\frac{\gamma}{2}\langle\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top-\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top},\mathbf{Z}\mathbf{Z}^\top-\mathbf{Z}^*\mathbf{Z}^{*\top}\rangle,\end{aligned}\tag{A.19}$$

where the third equality holds because $\langle\widetilde{\boldsymbol{\Delta}}\widetilde{\boldsymbol{\Delta}}^\top,\mathbf{Z}\mathbf{Z}^\top-\mathbf{Z}^*\mathbf{Z}^{*\top}\rangle=\langle\boldsymbol{\Delta}\boldsymbol{\Delta}^\top,\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top-\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\rangle$. Besides, we have

$$\gamma I_{32}-\gamma I_{33}=-\gamma\langle\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top-\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top},\mathbf{Z}\mathbf{Z}^\top-\mathbf{Z}^*\mathbf{Z}^{*\top}\rangle.\tag{A.20}$$

Since $I_{34}=\|\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}\|_F^2\geq 0$, thus plugging (A.19) and (A.20) into (A.18), we obtain the upper bound of $I_3-4\gamma\langle\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\mathbf{Z},\boldsymbol{\Delta}\rangle$

$$\begin{aligned}I_3-4\gamma\langle\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top\mathbf{Z},\boldsymbol{\Delta}\rangle&\leq\frac{\gamma}{2}\langle\widetilde{\boldsymbol{\Delta}}\widetilde{\boldsymbol{\Delta}}^\top,\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\rangle-\frac{3\gamma}{2}\langle\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top-\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top},\mathbf{Z}\mathbf{Z}^\top-\mathbf{Z}^*\mathbf{Z}^{*\top}\rangle\\&=\frac{\gamma}{2}\|\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\|_F^2-2\gamma\|\boldsymbol{\Delta}_U\boldsymbol{\Delta}_V^\top\|_F^2-\frac{3\gamma}{2}\|\mathbf{Z}\mathbf{Z}^\top-\mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2+6\gamma\|\mathbf{X}-\mathbf{X}^*\|_F^2.\end{aligned}\tag{A.21}$$

Finally, combining (A.14) and (A.21), we conclude

$$\begin{aligned}0&\leq(3L-6\mu+6\gamma)\|\mathbf{X}-\mathbf{X}^*\|_F^2+(3L-2\mu-2\gamma)\|\boldsymbol{\Delta}_U\boldsymbol{\Delta}_V^\top\|_F^2+\frac{\gamma}{2}\|\boldsymbol{\Delta}\boldsymbol{\Delta}\|_F^2-\frac{3\gamma}{2}\|\mathbf{Z}\mathbf{Z}^\top-\mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2\\&\quad+2\sqrt{2r}\|\nabla\mathcal{F}_n(\mathbf{X}^*)\|_2\cdot(\|\mathbf{U}\boldsymbol{\Delta}_V^\top+\boldsymbol{\Delta}_U\mathbf{V}^\top\|_F+\|\mathbf{X}-\mathbf{X}^*\|_F)\\&\leq(3L-6\mu+6\gamma+\beta)\cdot\|\mathbf{X}-\mathbf{X}^*\|_F^2+(3L-2\mu-2\gamma+\beta)\cdot\|\boldsymbol{\Delta}_U\boldsymbol{\Delta}_V^\top\|_F^2+\frac{\gamma}{2}\|\boldsymbol{\Delta}\boldsymbol{\Delta}\|_F^2\\&\quad-\frac{3\gamma}{2}\|\mathbf{Z}\mathbf{Z}^\top-\mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2+\frac{10r}{\beta}\|\nabla\mathcal{F}_n(\mathbf{X}^*)\|_2^2,\end{aligned}\tag{A.22}$$

where the second inequality holds because of triangle inequality and $2ab\leq\beta a^2+b^2/\beta$, for any $\beta>0$. Choose γ and β such that $3L-6\mu+6\gamma+\beta\geq0$ and $3L-2\mu-2\gamma+\beta\geq0$, then according to (A.22), we have

$$\begin{aligned}0&\leq\frac{1}{2}(3L-6\mu+3\gamma+\beta)\cdot\|\mathbf{Z}\mathbf{Z}^\top-\mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2+\frac{1}{2}(3L-2\mu-\gamma+\beta)\cdot\|\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\|_F^2+\frac{10r}{\beta}\|\nabla\mathcal{F}_n(\mathbf{X}^*)\|_2^2\\&\leq\frac{1}{2}(9L-10\mu+\gamma+3\beta)\cdot\|\mathbf{Z}\mathbf{Z}^\top-\mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2+\frac{10r}{\beta}\|\nabla\mathcal{F}_n(\mathbf{X}^*)\|_2^2,\end{aligned}$$

where the first inequality holds because $2\|\mathbf{X}-\mathbf{X}^*\|_F^2\leq\|\mathbf{Z}\mathbf{Z}^\top-\mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2$ and $2\|\boldsymbol{\Delta}_U\boldsymbol{\Delta}_V^\top\|_F^2\leq\|\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\|_F^2$, and the second inequality is due to Lemma D.1 and the fact that $3L-2\mu-\gamma+\beta\geq0$. Therefore, under condition that $L/\mu<18/17$, set $\beta=(18\mu-17L)/12$, and choose γ such that $\mu-L/2\leq\gamma<\min\{(22\mu-19L)/4,(3L-2\mu)/2\}$, we have $9L-10\mu+\gamma+3\beta<0$. Thus, we conclude

$$\|\mathbf{X}-\mathbf{X}^*\|_F^2\leq\frac{1}{2}\|\mathbf{Z}\mathbf{Z}^\top-\mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2\leq\Gamma r\|\nabla\mathcal{F}_n(\mathbf{X}^*)\|_2^2\leq\Gamma r\epsilon^2(n,\delta),$$

with probability at least $1-\delta$, where Γ is a constant depending on L,μ and γ, and the last inequality follows from Condition 3.7. Thus we complete the proof. □

B. Proofs for Specific Examples

In this section, we present proofs for the specific models including matrix completion and one-bit matrix completion. In the following discussions, we denote $d=\max\{d_1,d_2\}$ for simplicity.

B.1. Proof for Matrix Completion

In order to prove the results for noisy matrix completion, we need to make use of the following lemmas, which are tailored for noisy matrix completion. In the following discussions, we let $|\Omega|=n$, and $\mathbf{A}_{jk}=\mathbf{e}_j\mathbf{e}_k^\top$, where $\mathbf{e}_j\in\mathbb{R}^{d_1},\mathbf{e}_k\in\mathbb{R}^{d_2}$ are basis vectors. Define $\mathcal{A}$ as the corresponding linear transformation operator such that $\mathcal{A}(\boldsymbol{\Delta})=[\langle\mathbf{A}_{j(1)k(1)},\boldsymbol{\Delta}\rangle,\ldots,\langle\mathbf{A}_{j(n)k(n)},\boldsymbol{\Delta}\rangle]^\top$, where $\big(j(i),k(i)\big)\in\Omega$ for any $i\in[n]$.

Lemma B.1. (Negahban & Wainwright, 2012) There are universal constants $\{c_i\}_{i=1}^6$ such that if $n \geq c_1 r d \log d$, and for all $\boldsymbol{\Delta} \in \mathbb{R}^{d_1 \times d_2}$ that satisfy the following condition

$$\sqrt{\frac{d_1 d_2}{r}} \frac{\|\boldsymbol{\Delta}\|_{\infty,\infty}}{\|\boldsymbol{\Delta}\|_F} \cdot \frac{\|\boldsymbol{\Delta}\|_*}{\|\boldsymbol{\Delta}\|_F} \leq \frac{1}{c_2} \sqrt{n/(d \log d)}, \tag{B.1}$$

with probability at least $1 - c_3/d$, we have

$$\left| \frac{\|\mathcal{A}(\boldsymbol{\Delta})\|_2}{\sqrt{n}} - \frac{\|\boldsymbol{\Delta}\|_F}{\sqrt{d_1 d_2}} \right| \leq c_4 \frac{\|\boldsymbol{\Delta}\|_F}{\sqrt{d_1 d_2}} \left(1 + \frac{c_5 \sqrt{d_1 d_2} \|\boldsymbol{\Delta}\|_{\infty,\infty}}{\sqrt{n} \|\boldsymbol{\Delta}\|_F} \right).$$

Lemma B.2. (Negahban & Wainwright, 2012) Consider noisy matrix completion with uniform sampling model. Suppose the noisy entry E_{jk} follows i.i.d. zero mean distribution with variance ν^2. Then, with probability at least $1 - c_6/d$, we have

$$\left\| \frac{1}{p} \sum_{(j,k)\in\Omega} E_{jk} \mathbf{A}_{jk} \right\|_2 \leq c_7 \nu \sqrt{\frac{d \log d}{p}},$$

where c_6, c_7 are universal constants, and $p = n/(d_1 d_2)$.

Proof of Corollary 4.1. In order to prove Corollary 4.1, we need to verify the restricted strong convexity and smoothness conditions in Condition 3.5, 3.6 for $\mathcal{F}_n(\mathbf{X})$. Moreover, we need to establish Condition 3.7.

To begin with, we recast the objective loss function for matrix completion as $\mathcal{F}_n(\mathbf{X}) = (2p)^{-1} \sum_{(j,k)\in\Omega} \big(\langle \mathbf{A}_{jk}, \mathbf{X} \rangle - Y_{jk} \big)^2$. Thus for all matrices $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^{d_1 \times d_2}$ with rank at most r we have

$$\mathcal{F}_n(\mathbf{X}_1) - \mathcal{F}_n(\mathbf{X}_2) - \langle \nabla \mathcal{F}_n(\mathbf{X}_2), \mathbf{X}_2 - \mathbf{X}_1 \rangle = (2p)^{-1} \|\mathcal{A}(\boldsymbol{\Delta})\|_2^2,$$

where $\boldsymbol{\Delta} = \mathbf{X}_1 - \mathbf{X}_2$. Next, we establish the restricted strong convexity and smoothness conditions for $\mathcal{F}_n(\mathbf{X})$ based on Lemma B.1.

Case 1**:** If $\boldsymbol{\Delta}$ violates condition (B.1), we have

$$\begin{aligned} \|\boldsymbol{\Delta}\|_F^2 &\leq c_0 \big(\sqrt{d_1 d_2} \|\boldsymbol{\Delta}\|_\infty \big) \|\boldsymbol{\Delta}\|_* \sqrt{\frac{d \log d}{nr}} \\ &\leq 2 c_0 \alpha' \sqrt{2 d_1 d_2} \|\boldsymbol{\Delta}\|_F \sqrt{\frac{d \log d}{n}}, \end{aligned}$$

where $\alpha' = \beta r \sigma_1^* / \sqrt{d_1 d_2}$, which comes from the incoherence condition of low rank matrices $\mathbf{X}_1$ and $\mathbf{X}_2$. Hence we can obtain

$$\|\boldsymbol{\Delta}\|_F^2 \leq c_1 \alpha'^2 \frac{d \log d}{p}. \tag{B.2}$$

Case 2**:** If $\boldsymbol{\Delta}$ satisfies condition (B.1), by Lemma B.1, we have

$$\left| \frac{\|\mathcal{A}(\boldsymbol{\Delta})\|_2}{\sqrt{p}} - \|\boldsymbol{\Delta}\|_F \right| \leq \frac{1}{90} \|\boldsymbol{\Delta}\|_F \left(1 + \frac{c_3 \sqrt{d_1 d_2} \|\boldsymbol{\Delta}\|_{\infty,\infty}}{\sqrt{n} \|\boldsymbol{\Delta}\|_F} \right).$$

If $c_3 \sqrt{d_1 d_2} \|\boldsymbol{\Delta}\|_{\infty,\infty} / (\sqrt{n} \|\boldsymbol{\Delta}\|_F) \geq 1/25$, we have

$$\|\boldsymbol{\Delta}\|_F^2 \leq c_5 \frac{\alpha'^2}{p}. \tag{B.3}$$

Otherwise, if $c_3 \sqrt{d_1 d_2} \|\boldsymbol{\Delta}\|_{\infty,\infty} / (\sqrt{n} \|\boldsymbol{\Delta}\|_F) \leq 1/25$, we have

$$\frac{42}{43} \|\boldsymbol{\Delta}\|_F^2 \leq \frac{\|\mathcal{A}(\boldsymbol{\Delta})\|_2^2}{p} \leq \frac{44}{43} \|\boldsymbol{\Delta}\|_F^2.$$

Thus we obtain the restricted strong convexity and smoothness conditions for $\mathcal{F}_n(\mathbf{X})$ with parameters $\mu = 42/43, L = 44/43$. Next, for Condition 3.7, we have $\nabla \mathcal{F}_n(\mathbf{X}^*) = p^{-1}\sum_{(j,k)\in\Omega} E_{jk}\mathbf{e}_j\mathbf{e}_k^\top$. Since each E_{jk} follows i.i.d. Gaussian distribution with variance $\nu^2/(d_1 d_2)$. Therefore, according to Lemma B.2, we can obtain $\|\nabla \mathcal{F}_n(\mathbf{X}^*)\|_2 \le c_5\nu\sqrt{d\log d/n}$ holds with probability at least $1 - c_6/d$. In addition, combining this result with the additional error bounds (B.2) and (B.3), with probability at least $1 - c_6/d$, we can establish Condition 3.7 with parameter $\epsilon^2 = c_7 \max\{\nu^2, r^2\beta^2\sigma_1^2\} d\log d/n$.

Therefore, for matrix completion (2.5), we set $\alpha^2 = \alpha'$ in all constraints $h_i(\mathbf{Z}) \le 0,\ i = 1, \ldots, d_1 + d_2$. Then we can apply our general framework to matrix completion, and for any local minima $\mathbf{Z}$ of matrix completion (2.5), we obtain the following standardized estimation error between $\mathbf{U}\mathbf{V}^\top$ and $\mathbf{X}^*$

$$\|\mathbf{U}\mathbf{V}^\top - \mathbf{X}^*\|_F^2 \le c_7 \max\{\nu^2, r\beta^2\sigma_1^2\}\frac{rd\log d}{n},$$

which completes the proof. □

B.2. Proof for One-Bit Matrix Completion

For one-bit matrix completion, we also consider the uniform sampling model as discussed in matrix completion. In the following discussion, let Ω denote the observed index set with cardinality $|\Omega| = n$, and $\mathbf{A}_{jk} = \mathbf{e}_j\mathbf{e}_k^\top$ with corresponding transformation operator $\mathcal{A}$. In addition, we define the following two quantities $\mu_{\beta'}, L_{\beta'}$, which control the quadratic lower and upper bounds of the second-order Taylor expansion of the sample loss function.

$$\mu_{\beta'} \le \min\left(\inf_{|x|\le\beta'}\left\{\frac{f'^2(x)}{f^2(x)} - \frac{f''(x)}{f(x)}\right\}, \inf_{|x|\le\beta'}\left\{\frac{f'^2(x)}{(1-f(x))^2} + \frac{f''(x)}{1-f(x)}\right\}\right), \tag{B.4}$$

$$L_{\beta'} \ge \max\left(\sup_{|x|\le\beta'}\left\{\frac{f'^2(x)}{f^2(x)} - \frac{f''(x)}{f(x)}\right\}, \sup_{|x|\le\beta'}\left\{\frac{f'^2(x)}{(1-f(x))^2} + \frac{f''(x)}{1-f(x)}\right\}\right). \tag{B.5}$$

Note that when $f(\cdot)$ and β' are given, $\mu_{\beta'}$ and $L_{\beta'}$ are fixed constants, which do not depend on the dimension of the unknonw low-rank matrix.

Proof of Corollary 4.3. In order to prove Corollary 4.3, we need to verify the restricted strong convexity and smoothness conditions in Conditions 3.5, 3.6 for $\mathcal{F}_n(\mathbf{X})$. Furthermore, we need to establish Condition 3.7. Note that we impose the constraint $\mathcal{D}$ to ensure the estimator $\mathbf{X}$ satisfies incoherence condition (2.4) such that $\|\mathbf{X}\|_{\infty,\infty} \le r\beta\sigma_1/\sqrt{d_1 d_2}$. Thus we should consider the twice differentiable function $f(x) = g(x/\tau)$, where $\tau = \nu/\sqrt{d_1 d_2}$ is a scale parameter. For example, one common used function is the Probit function $f(x) = \Phi(x/\sigma)$ with $\sigma = \nu/\sqrt{d_1 d_2}$, where Φ denotes the cumulative distribution function of standard Gaussian distribution. And this is equivalent to observation model (2.6) with Z_{jk} i.i.d. following normal distribution with variance $\nu^2/(d_1 d_2)$.

We can rewrite the objective function for one-bit matrix completion as follows

$$\mathcal{F}_n(\mathbf{X}) := -\frac{1}{n}\sum_{(j,k)\in\Omega}\left\{\mathbb{1}\big\{(Y_{jk}=1)\big\}\log\big(g(\langle\mathbf{A}_{jk},\mathbf{X}\rangle/\tau)\big) + \mathbb{1}\big\{(Y_{jk}=-1)\big\}\log\big(1 - g(\langle\mathbf{A}_{jk},\mathbf{X}\rangle/\tau)\big)\right\}.$$

Therefore, we obtain

$$\nabla^2\mathcal{F}_n(\mathbf{X}) = \frac{1}{p\nu^2}\sum_{(j,k)\in\Omega} B_{jk}(\mathbf{X})\mathrm{vec}(\mathbf{A}_{jk})\mathrm{vec}(\mathbf{A}_{jk})^\top, \tag{B.6}$$

where $B_{jk}(\mathbf{X})$ is defined as

$$\begin{aligned} B_{jk}(\mathbf{X}) = &\left(\frac{g'^2(\langle\mathbf{A}_{jk},\mathbf{X}\rangle/\tau)}{g^2(\langle\mathbf{A}_{jk},\mathbf{X}\rangle/\tau)} - \frac{g''(\langle\mathbf{A}_{jk},\mathbf{X}\rangle/\tau)}{g(\langle\mathbf{A}_{jk},\mathbf{X}\rangle/\tau)}\right)\mathbb{1}\{Y_{jk}=1\} \\ &+ \left(\frac{g''(\langle\mathbf{A}_{jk},\mathbf{X}\rangle/\tau)}{1-g(\langle\mathbf{A}_{jk},\mathbf{X}\rangle/\tau)} - \frac{g'^2(\langle\mathbf{A}_{jk},\mathbf{X}\rangle/\tau)}{(1-g(\langle\mathbf{A}_{jk},\mathbf{X}\rangle/\tau)^2}\right)\mathbb{1}\{Y_{jk}=-1\}. \end{aligned}$$

Therefore, using mean value theorem, for all matrices $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^{d_1\times d_2}$ with rank at most r, we can obtain

$$\mathcal{F}_n(\mathbf{X}_1) = \mathcal{F}_n(\mathbf{X}_2) + \langle\nabla\mathcal{F}_n(\mathbf{X}_2), \mathbf{X}_2 - \mathbf{X}_1\rangle + \frac{1}{2}(\mathbf{x}_2 - \mathbf{x}_1)^\top\nabla^2\mathcal{F}_n(\mathbf{W})(\mathbf{x}_2 - \mathbf{x}_1),$$

where $\mathbf{W} = \mathbf{X}_1 + t(\mathbf{X}_2 - \mathbf{X}_1)$ for some $t \in [0,1]$, and $\mathbf{x}_1 = \mathrm{vec}(\mathbf{X}_1)$, $\mathbf{x}_2 = \mathrm{vec}(\mathbf{X}_2)$. Thus according to (B.6), we have

$$\begin{aligned}(\mathbf{x}_2 - \mathbf{x}_1)^\top \nabla^2 \mathcal{F}_n(\mathbf{W})(\mathbf{x}_2 - \mathbf{x}_1) &= \frac{1}{p\nu^2} \sum_{(j,k)\in\Omega} B_{jk}(\mathbf{W}) \langle \mathrm{vec}(\mathbf{A}_{jk})^\top (\mathbf{x}_2 - \mathbf{x}_1), \mathrm{vec}(\mathbf{A}_{jk})^\top (\mathbf{x}_2 - \mathbf{x}_1)\rangle \\ &= \frac{1}{p\nu^2} \sum_{(j,k)\in\Omega} B_{jk}(\mathbf{W}) \langle \mathbf{A}_{jk}, \boldsymbol{\Delta}\rangle^2,\end{aligned}$$

where $\boldsymbol{\Delta} = \mathbf{X}_2 - \mathbf{X}_1$. This implies

$$c_1 \frac{\|\mathcal{A}(\boldsymbol{\Delta})\|_2^2}{p} \leq \frac{1}{p\nu^2} \sum_{(j,k)\in\Omega} B_{jk}(\mathbf{W}) \langle \mathbf{A}_{jk}, \boldsymbol{\Delta}\rangle^2 \leq c_2 \frac{\|\mathcal{A}(\boldsymbol{\Delta})\|_2^2}{p},$$

where the inequalities come from the definition of $\mu_{\beta'}, L_{\beta'}$. Next, for the term $\|\mathcal{A}(\boldsymbol{\Delta})\|_2^2/p$, we can follow the same proofs as in matrix completion. Therefore, if $n \geq c_3 rd\log d$, with probability at least $1 - c_4/d$, we can obtain the restricted strong convexity and smoothness conditions for $\mathcal{F}_n(\mathbf{X})$ with parameters $\mu = \mu_{\beta'} 42/(43\nu^2)$, $L = L_{\beta'} 44/(43\nu^2)$. Moreover, we will have an additional statistical error bound that $\|\mathbf{X} - \mathbf{X}^*\|_F^2 \leq c_6 \alpha'^2 d\log d/p$, where $\alpha' = \beta r \sigma_1^*/\sqrt{d_1 d_2}$.

Next, for Condition 3.7, we have $\nabla \mathcal{F}_n(\mathbf{X}^*) = \big(n\nu/\sqrt{d_1 d_2}\big)^{-1} \sum_{(j,k)\in\Omega} b_{jk} \mathbf{A}_{jk}$, where we have

$$b_{jk} = -\frac{g'(\langle \mathbf{A}_{jk}, \mathbf{X}^*\rangle/\tau)}{g(\langle \mathbf{A}_{jk}, \mathbf{X}^*\rangle/\tau)} \mathbb{1}\{Y_{jk} = 1\} + \frac{g'(\langle \mathbf{A}_{jk}, \mathbf{X}^*\rangle/\tau)}{1 - g(\langle \mathbf{A}_{jk}, \mathbf{X}^*\rangle/\tau)} \mathbb{1}\{Y_{jk} = -1\}.$$

Thus accordding to Lemma B.2, we can obtain $\|\nabla \mathcal{F}_n(\mathbf{X}^*)\|_2 \leq c_7 \gamma_{\alpha'} \sqrt{d\log d/n}$ holds with probability at least $1 - c_8/d$, where c_7, c_8 are some constants.

Therefore, for one-bit matrix completion problem (2.8), we set $\alpha^2 = \alpha'$ in all constraints $h_i(\mathbf{U}) \leq 0,\ i = 1, \ldots, d_1 + d_2$. Then we can apply our general framework to one-bit matrix completion, and for any local minima $\mathbf{Z}$ of one-bit matrix completion problem (2.8), we can obtain the following estimation error between $\mathbf{U}\mathbf{V}^\top$ and $\mathbf{X}^*$

$$\|\mathbf{U}\mathbf{V}^\top - \mathbf{X}^*\|_F^2 \leq c_9 \max\{\gamma_{\beta'}^2, r\beta^2\sigma_1^2\} \frac{rd\log d}{n},$$

which completes the proof. □

C. Proof of Lemma A.1

Proof. According to the restricted strong convexity and smoothness Conditions 3.5 and 3.6, for all matrices $\mathbf{Y}_1, \mathbf{Y}_2 \in \mathbb{R}^{d_1 \times d_2}$ with rank at most $6r$, we have

$$\mu \|\mathbf{Y}_2 - \mathbf{Y}_1\|_F^2 \leq \langle \nabla \mathcal{F}_n(\mathbf{Y}_2) - \nabla \mathcal{F}_n(\mathbf{Y}_1), \mathbf{Y}_2 - \mathbf{Y}_1\rangle \leq L\|\mathbf{Y}_2 - \mathbf{Y}_1\|_F^2. \tag{C.1}$$

According to the definition of Hessian, we have

$$\begin{aligned}\mathrm{vec}(\mathbf{W})^\top \nabla^2 \mathcal{F}_n(\mathbf{Y}) \mathrm{vec}(\mathbf{W}) &= \left\langle \mathbf{W}, \lim_{t\to 0} \frac{\nabla \mathcal{F}_n(\mathbf{Y} + t\mathbf{W}) - \nabla \mathcal{F}_n(\mathbf{Y})}{t} \right\rangle \\ &= \lim_{t\to 0} \langle (\mathbf{Y} + t\mathbf{W}) - \mathbf{Y}, \nabla \mathcal{F}_n(\mathbf{Y} + t\mathbf{W}) - \nabla \mathcal{F}_n(\mathbf{Y})\rangle / t^2. \end{aligned} \tag{C.2}$$

For all matrices $\mathbf{Y} \in \mathbb{R}^{d_1 \times d_2}$ with rank at most $2r$, and matrices $\mathbf{W} \in \mathbb{R}^{d_1 \times d_2}$ with rank at most $4r$, we have $\mathbf{Y} + t\mathbf{W}$ has rank at most $6r$, thus applying (C.1) to (C.2), we obtain

$$\mu \|\mathbf{W}\|_F^2 \leq \mathrm{vec}(\mathbf{W})^\top \nabla^2 \mathcal{F}_n(\mathbf{Y}) \mathrm{vec}(\mathbf{W}) \leq L\|\mathbf{W}\|_F^2. \tag{C.3}$$

Since $\mathbf{W}_1, \mathbf{W}_2$ has rank at most $2r$, we have $\mathbf{W}_1 + \mathbf{W}_2, \mathbf{W}_1 - \mathbf{W}_2$ has rank at most $4r$. Thus, by substituting $\mathbf{W}$ by $\mathbf{W}_1 + \mathbf{W}_2$ and $\mathbf{W}_1 - \mathbf{W}_2$ in (C.3) respectively, we obtain

$$\begin{aligned}\mu \|\mathbf{W}_1 + \mathbf{W}_2\|_F^2 &\leq \mathrm{vec}(\mathbf{W}_1 + \mathbf{W}_2)^\top \nabla^2 \mathcal{F}_n(\mathbf{Y}) \mathrm{vec}(\mathbf{W}_1 + \mathbf{W}_2) \leq L\|\mathbf{W}_1 + \mathbf{W}_2\|_F^2, \\ \mu \|\mathbf{W}_1 - \mathbf{W}_2\|_F^2 &\leq \mathrm{vec}(\mathbf{W}_1 - \mathbf{W}_2)^\top \nabla^2 \mathcal{F}_n(\mathbf{Y}) \mathrm{vec}(\mathbf{W}_1 - \mathbf{W}_2) \leq L\|\mathbf{W}_1 - \mathbf{W}_2\|_F^2.\end{aligned}$$

Therefore, by taking difference, we further obtain

$$\big|4\mathrm{vec}(\mathbf{W}_1)^\top \nabla^2 \mathcal{F}_n(\mathbf{Y}) \mathrm{vec}(\mathbf{W}_2) - 2(L + \mu)\langle \mathbf{W}_1, \mathbf{W}_2\rangle\big| \leq (L - \mu) \cdot (\|\mathbf{W}_1\|_F^2 + \|\mathbf{W}_2\|_F^2),$$

which completes the proof. □

D. Auxiliary Lemma

Lemma D.1. (Ge et al., 2017) Let $\mathbf{Z}, \mathbf{Z}^*$ be two $d \times r$ matrices. Let $\mathbf{R}$ be the optimal rotation with respect to $\mathbf{Z}$ and $\mathbf{Z}^*$ such that $\mathbf{R} = \mathrm{argmin}_{\widetilde{\mathbf{R}} \in \mathbb{Q}_r} \|\mathbf{Z} - \mathbf{Z}^* \widetilde{\mathbf{R}}\|_F$. Then we have that $\mathbf{Z}^\top \mathbf{Z}^* \mathbf{R}$ is positive semidefinite. Moreover, we have the following inequality

$$\|(\mathbf{Z} - \mathbf{Z}^* \mathbf{R})(\mathbf{Z} - \mathbf{Z}^* \mathbf{R})^\top\|_F^2 \leq 2\|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^* \mathbf{Z}^*\|_F^2.$$