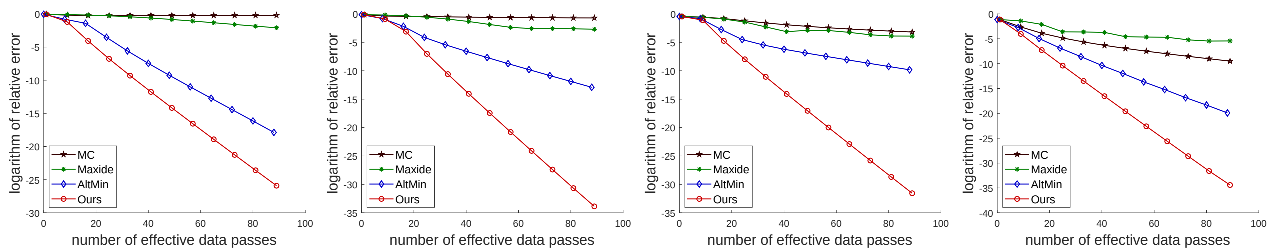
A. Additional Experiments

In this section, we provide additional experimental results, including simulations under the rectangular setting, the relative error versus CPU time plots and multi-label learning experiments on Yahoo datasets.

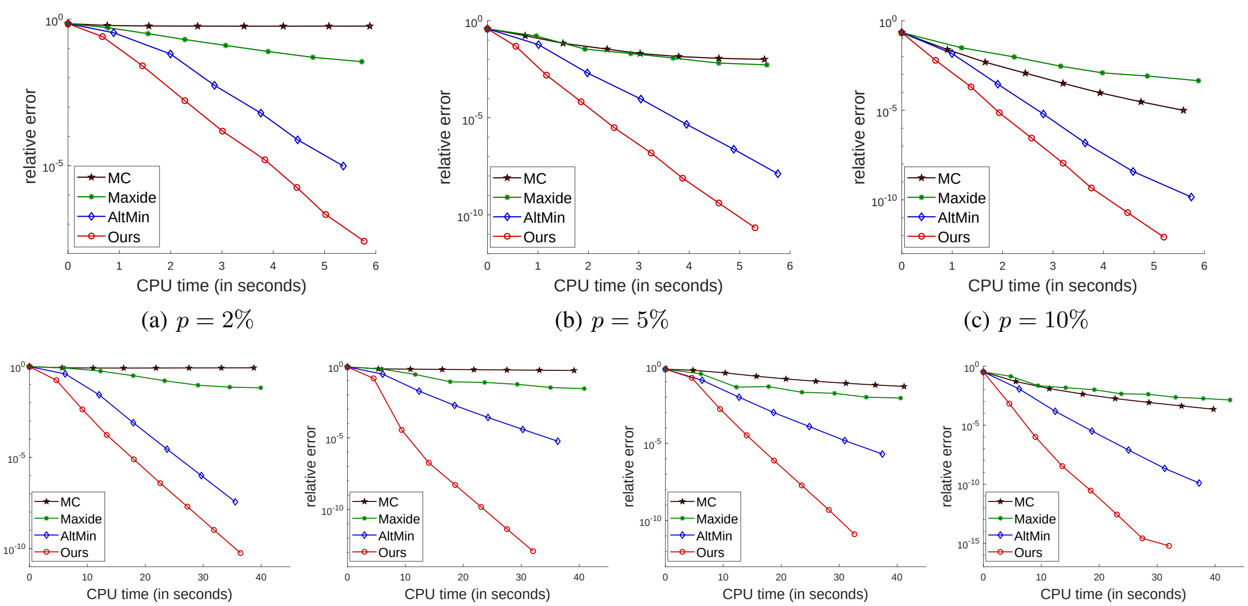
Simulations under Rectangular setting. The data generation approach for rectangular setting follows the same procedure in Section 6.1. Specifically, the unknown low-rank matrix $\mathbf{L}^* \in \mathbb{R}^{d_1 \times d_2}$ is modeled as $\mathbf{L}^* = \mathbf{X}_L \mathbf{M}^* \mathbf{X}_R^\top$ with dimensions $d_1 = 5000$, $d_2 = 2000$ and rank $r = 5$. The feature dimensions n_1, n_2 are specified as $n_1 = 100$ and $n_2 = 50$. We compare the performance of our proposed algorithm with existing (inductive) matrix completion methods, as explained in Section 6.1, in terms of relative error and effective number of data passes, which is illustrated in Figure 3. Note that the sampling rate p is chosen from the range $\{0.25\%, 0.5\%, 1\%, 2\%\}$. The results show that our proposed algorithm has better performance under the rectangular setting compared with existing methods.



(a) $p = 0.25\%$ (b) $p = 0.5\%$ (c) $p = 1\%$ (d) $p = 2\%$

Figure 3. Plots of logarithm relative error vs. number of effective data passes for different (inductive) matrix completion algorithms under the setting $d_1 = 5000$, $d_2 = 2000$, $n_1 = 100$, $n_2 = 50$ and $r = 5$ with sampling rate p varied in the range $\{0.25\%, 0.5\%, 1\%, 2\%\}$.

Relative Error versus CPU Time. To further demonstrate the computational advantage of the proposed method, we compare our algorithm with existing (inductive) matrix completion methods in terms of relative error and CPU time for all of the aforementioned simulation settings. It can be seen from Figure 4 that our proposed method achieves the lowest relative error with respect to the same CPU time under all settings, which again confirms the superiority of our proposed algorithm.



(a) $p = 2\%$ (b) $p = 5\%$ (c) $p = 10\%$

(d) $p = 0.25\%$ (e) $p = 0.5\%$ (f) $p = 1\%$ (g) $p = 2\%$

Figure 4. Plots of relative error vs. CPU time for different (inductive) matrix completion algorithms under the settings: (i) $d = 1000$, $n = 100$ and $r = 10$ with sampling rate p selected from $\{2\%, 5\%, 10\%\}$ in the top panel. (ii) $d_1 = 5000$, $d_2 = 2000$, $n_1 = 100$, $n_2 = 50$ and $r = 5$ with p varied in the range $\{0.25\%, 0.5\%, 1\%, 2\%\}$ in the bottom panel.

Multi-Label Learning on Yahoo Datasets. We provide the experimental results for multi-label learning on the Web page classification Yahoo datasets obtained from Ueda & Saito (2003), including Arts, Education, Health and Science. In particular, each dataset has 5000 different Web pages, which are associated with 462-743 different features and labeled with 26-40 different predefined categories. We extract the top-100 principal components to reconstruct the feature matrix so that our estimation model will not overfit. The comparison results are demonstrated in Table 3, which shows that our method achieves higher prediction accuracy and shorter computational time against existing inductive matrix completion approaches under all of the experimental settings (c.f. Section 6.2 for detailed information).

Table 3. Experimental results in terms of AP and total running time on Yahoo datasets for multi-label learning via different methods.

Dataest	Method	$p\% = 10\%$		$p\% = 25\%$		$p\% = 50\%$	
		averaged AP (std)	time (s)	averaged AP (std)	time (s)	averaged AP (std)	time (s)
Arts	BR-linear	0.4333 (0.0055)	1.63×10^1	0.4608 (0.0059)	4.99×10^1	0.4769 (0.0070)	1.22×10^2
	Maxide	0.5131 (0.0062)	3.53×10^0	0.5698 (0.0046)	3.08×10^0	0.5821 (0.0059)	2.78×10^0
	AltMin	0.5232 (0.0060)	3.04×10^0	0.5647 (0.0056)	2.72×10^0	0.5707 (0.0060)	2.20×10^0
	Ours	**0.5412** (0.0056)	1.16×10^0	**0.5768** (0.0050)	0.83×10^0	**0.5872** (0.0061)	1.41×10^0
Education	BR-linear	0.4515 (0.0057)	1.52×10^1	0.4637 (0.0073)	4.23×10^1	0.4775 (0.0073)	9.88×10^1
	Maxide	0.5451 (0.0054)	4.95×10^0	0.5851 (0.0052)	4.16×10^0	0.5844 (0.0052)	3.31×10^0
	AltMin	0.5338 (0.0057)	2.03×10^0	0.5731 (0.0075)	2.68×10^0	0.5848 (0.0059)	1.70×10^0
	Ours	**0.5555** (0.0042)	1.05×10^0	**0.5906** (0.0058)	1.01×10^0	**0.5963** (0.0055)	0.88×10^0
Health	BR-linear	0.6402 (0.0064)	1.39×10^1	0.6859 (0.0028)	3.46×10^1	0.6933 (0.0051)	8.45×10^1
	Maxide	0.7210 (0.0073)	6.10×10^0	0.7454 (0.0059)	4.29×10^0	0.7586 (0.0051)	4.03×10^0
	AltMin	0.6961 (0.0091)	1.73×10^0	0.7370 (0.0042)	2.82×10^0	0.7476 (0.0045)	2.86×10^0
	Ours	**0.7305** (0.0053)	0.76×10^0	**0.7496** (0.0036)	1.60×10^0	**0.7627** (0.0049)	1.94×10^0
Science	BR-linear	0.4517 (0.0113)	1.80×10^1	0.4671 (0.0089)	4.67×10^1	0.4769 (0.0083)	1.15×10^2
	Maxide	0.4960 (0.0075)	7.19×10^0	0.5297 (0.0042)	5.46×10^0	0.5413 (0.0059)	5.25×10^0
	AltMin	0.4888 (0.0081)	4.38×10^0	0.5341 (0.0052)	4.55×10^0	0.5477 (0.0056)	3.88×10^0
	Ours	**0.5095** (0.0069)	1.11×10^0	**0.5391** (0.0039)	0.94×10^0	**0.5536** (0.0065)	1.38×10^0

B. Proofs of the Main Results

In this section, we provide the proofs for our main theoretical results. To begin with, we introduce some notations to simplify our proof. Let $\mathcal{I} : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^{d_1 \times d_2}$ be the identity map, i.e., $\mathcal{I}(\mathbf{A}) = \mathbf{A}$. Denote the elementwise infinity norm of matrix $\mathbf{A}$ by $\|\mathbf{A}\|_{\infty,\infty}$. For any $\mathbf{Z} \in \mathbb{R}^{(n_1+n_2)\times r}$, we denote $\mathbf{Z} = [\mathbf{U}; \mathbf{V}]$, where $\mathbf{U} \in \mathbb{R}^{n_1 \times r}$ and $\mathbf{V} \in \mathbb{R}^{n_2 \times r}$. According to (4.1), our objective is equivalent to minimize the following regularized loss function in terms of $\mathbf{Z}$

$$\widetilde{f}_\Omega(\mathbf{Z}) := f_\Omega(\mathbf{U}, \mathbf{V}) = \frac{1}{2p}\|\mathcal{P}_\Omega(\mathbf{X}_L \mathbf{U}\mathbf{V}^\top \mathbf{X}_R^\top - \mathbf{L})\|_F^2 + \frac{1}{8}\|\mathbf{U}^\top\mathbf{U} - \mathbf{V}^\top\mathbf{V}\|_F^2. \tag{B.1}$$

Let Sym : $\mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^{(d_1+d_2)\times(d_1+d_2)}$ be the lifting operator, such that for any matrix $\mathbf{A} \in \mathbb{R}^{d_1 \times d_2}$

$$\text{Sym}(\mathbf{A}) = \begin{bmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^\top & \mathbf{0} \end{bmatrix}.$$

For any block matrices $\mathbf{A} \in \mathbb{R}^{(d_1+d_2)\times(d_1+d_2)}$ with partitions

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \text{ where } \mathbf{A}_{ij} \in \mathbb{R}^{d_i \times d_j}, i, j \in \{1, 2\},$$

define linear operators $\mathcal{P}_{\text{diag}}$ and $\mathcal{P}_{\text{off}} : \mathbb{R}^{(d_1+d_2)\times(d_1+d_2)} \to \mathbb{R}^{(d_1+d_2)\times(d_1+d_2)}$ as

$$\mathcal{P}_{\text{diag}}(\mathbf{A}) = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} \quad \text{and} \quad \mathcal{P}_{\text{off}}(\mathbf{A}) = \begin{bmatrix} \mathbf{0} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{0} \end{bmatrix}.$$

Similarly, for any block matrices $\mathbf{B} \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}$ with partitions

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}, \text{ where } \mathbf{B}_{ij} \in \mathbb{R}^{n_i \times n_j}, i, j \in \{1, 2\},$$

define operators $\overline{\mathcal{P}}_{\text{diag}}$ and $\overline{\mathcal{P}}_{\text{off}}: \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)} \to \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}$ as

$$\overline{\mathcal{P}}_{\text{diag}}(\mathbf{B}) = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{22} \end{bmatrix} \quad \text{and} \quad \overline{\mathcal{P}}_{\text{off}}(\mathbf{B}) = \begin{bmatrix} \mathbf{0} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{0} \end{bmatrix}.$$

In addition, let $\widetilde{\Omega} \in [d_1 + d_2] \times [d_1 + d_2]$ be the corresponding index set of observed entries in the lifted space, then the observed matrix in the lifted space can be expressed as

$$\mathcal{P}_{\widetilde{\Omega}}\big(\text{Sym}(\mathbf{L})\big) = \begin{bmatrix} \mathbf{0} & \mathcal{P}_\Omega(\mathbf{L}) \\ \big(\mathcal{P}_\Omega(\mathbf{L})\big)^\top & \mathbf{0} \end{bmatrix}.$$

And we let $\mathbf{X} \in \mathbb{R}^{(d_1+d_2)\times(n_1+n_2)}$ be the corresponding feature matrix in the lifted space, such that

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_L & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_R \end{bmatrix}.$$

Thus with the notations above, the regularized loss function $\widetilde{f}_\Omega$ in (B.1) can be rewritten as

$$\widetilde{f}_\Omega(\mathbf{Z}) = \frac{1}{4p}\big\|\mathcal{P}_{\widetilde{\Omega}}\big(\mathbf{X}\mathbf{Z}\mathbf{Z}^\top\mathbf{X}^\top - \text{Sym}(\mathbf{L})\big)\big\|_F^2 + \frac{1}{8}\|\mathbf{Z}^\top\mathbf{D}\mathbf{Z}\|_F^2, \tag{B.2}$$

where $\mathbf{D}$ is defined as

$$\mathbf{D} = \begin{bmatrix} \mathbf{I}_{n_1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_{n_2} \end{bmatrix}.$$

Recall that $\mathbf{Z}^* = [\mathbf{U}^*; \mathbf{V}^*]$ and $\mathcal{P}_\Omega(\mathbf{L}) = \mathcal{P}_\Omega(\mathbf{L}^*)$, then the gradient of $\widetilde{f}_\Omega$ can be formulated as

$$\begin{aligned} \nabla\widetilde{f}_\Omega(\mathbf{Z}) =& \frac{1}{p}\mathbf{X}^\top\mathcal{P}_{\widetilde{\Omega}}(\mathbf{X}[\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}]\mathbf{X}^\top)\mathbf{X}\mathbf{Z} \\ &+ \frac{1}{2}(\overline{\mathcal{P}}_{\text{diag}} - \overline{\mathcal{P}}_{\text{off}})(\mathbf{Z}\mathbf{Z}^T)\mathbf{Z}. \end{aligned} \tag{B.3}$$

B.1. Proof of Theorem 5.3

Proof. According to the initialization phase of Algorithm 1, we have

$$\begin{aligned} \|\mathbf{U}_{\text{init}}\mathbf{V}_{\text{init}}^\top - \mathbf{M}^*\|_2 &= \|p_0^{-1}\mathbf{X}_L^\top\mathcal{P}_{\Omega_0}(\mathbf{L}^*)\mathbf{X}_R - \mathbf{M}^*\|_2 \\ &= \|\mathbf{X}_L^\top(p_0^{-1}\mathcal{P}_{\Omega_0}(\mathbf{L}^*) - \mathbf{L}^*)\mathbf{X}_R\|_2 \\ &= \Bigg\|\sum_{(i,j)=(1,1)}^{(d_1,d_2)}(p_0^{-1}\xi_{ij} - 1)L_{ij}^* \cdot \mathbf{X}_L^\top\mathbf{e}_i\mathbf{e}_j^\top\mathbf{X}_R\Bigg\|_2 := \Bigg\|\sum_{(i,j)=(1,1)}^{(d_1,d_2)}\mathbf{S}_{ij}\Bigg\|_2, \end{aligned} \tag{B.4}$$

where $\xi_{ij} = 1$, if $(i,j) \in \Omega_0$; $\xi_{ij} = 0$, otherwise. Next we are going to apply Matrix Bernstein to the right hand side of (B.4). Note that $\mathbb{E}[\mathbf{S}_{ij}] = \mathbf{0}$, and we have

$$\|\mathbf{S}_{ij}\|_2 \le \frac{1}{p_0}\,|L_{ij}^*| \cdot \|\mathbf{X}_L^\top\mathbf{e}_i\|_2 \cdot \|\mathbf{X}_R^\top\mathbf{e}_j\|_2 \le \frac{1}{p_0}\,\|\mathbf{L}^*\|_{\infty,\infty} \cdot \|\mathbf{X}_L\|_{2,\infty} \cdot \|\mathbf{X}_R\|_{2,\infty}.$$

By Assumptions 3.1 and 3.2, we further obtain

$$\|\mathbf{S}_{ij}\|_2 \le \frac{1}{p_0}\,\|\mathbf{X}_L\overline{\mathbf{U}}^*\|_{2,\infty} \cdot \|\mathbf{\Sigma}^*\|_2 \cdot \|\mathbf{X}_R\overline{\mathbf{V}}^*\|_{2,\infty} \cdot \sqrt{\frac{\mu_1^2 n_1 n_2}{d_1 d_2}} \le \frac{\mu_0\mu_1 r\sqrt{n_1n_2}\sigma_1^*}{p_0 d_1 d_2}.$$

Denote $\mathbf{S} = \sum_{i,j} \mathbf{S}_{ij}$. To apply Matrix Bernstein, it remains to bound the terms $\|\mathbb{E}(\mathbf{S}\mathbf{S}^\top)\|_2$ and $\|\mathbb{E}(\mathbf{S}^\top\mathbf{S})\|_2$ respectively. Since $\mathbf{S}_{ij}$'s are independent, we have

$$\begin{aligned}\|\mathbb{E}(\mathbf{S}\mathbf{S}^\top)\|_2 &= \bigg\| \sum_{(i,j)=(1,1)}^{(d_1,d_2)} \mathbb{E}(\mathbf{S}_{ij}\mathbf{S}_{ij}^\top) \bigg\|_2 \\ &= \frac{1-p_0}{p_0} \bigg\| \sum_{(i,j)=(1,1)}^{(d_1,d_2)} L_{ij}^{*2}\, (\mathbf{X}_L^\top \mathbf{e}_i) \cdot \|\mathbf{X}_R^\top \mathbf{e}_j\|_2^2 \cdot (\mathbf{e}_i^\top \mathbf{X}_L) \bigg\|_2 \\ &\le \frac{1}{p_0} \bigg\| \sum_{i=1}^{d_1} \mathbf{e}_i \mathbf{e}_i^\top \cdot \Big[\sum_{j=1}^{d_2} L_{ij}^{*2} \cdot \|\mathbf{X}_R^\top \mathbf{e}_j\|_2^2 \Big] \bigg\|_2 \\ &\le \frac{1}{p_0} \cdot \max_{i\in[d_1]} \bigg(\sum_{j=1}^{d_2} L_{ij}^{*2} \cdot \|\mathbf{X}_R^\top \mathbf{e}_j\|_2^2 \bigg) \le \frac{\mu_1 n_1}{p_0 d_2} \cdot \|\mathbf{L}^*\|_{2,\infty}^2, \end{aligned} \tag{B.5}$$

where the first inequality is due to the fact that $\|\mathbf{A}\mathbf{B}\|_2 \le \|\mathbf{A}\|_2 \cdot \|\mathbf{B}\|_2$ and $\mathbf{X}_L$ is orthonormal, the last inequality follows from Assumption 3.2. According to the SVD of $\mathbf{L}^*$ and Assumption 3.1, we have

$$\|\mathbf{L}^*\|_{2,\infty} = \|(\mathbf{X}_L \overline{\mathbf{U}}^*) \cdot \mathbf{\Sigma}^*\|_{2,\infty} \le \|\mathbf{X}_L \overline{\mathbf{U}}^*\|_{2,\infty} \cdot \|\mathbf{\Sigma}^*\|_2 \le \sqrt{\frac{\mu_0 r}{d_1}} \sigma_1^*. \tag{B.6}$$

Therefore, plugging (B.6) into (B.5), we obtain

$$\|\mathbb{E}(\mathbf{S}\mathbf{S}^\top)\|_2 \le \frac{\mu_0 \mu_1 r n_1}{p_0 d_1 d_2} \sigma_1^{*2}.$$

Similarly, we can obtain the upper bound of $\|\mathbb{E}(\mathbf{S}^\top\mathbf{S})\|_2$, which implies

$$\max\Big\{ \big\|\mathbb{E}(\mathbf{S}^\top\mathbf{S})\big\|_2, \big\|\mathbb{E}(\mathbf{S}\mathbf{S}^\top)\big\|_2 \Big\} \le \frac{\mu_0 \mu_1 r (n_1+n_2)}{p_0 d_1 d_2} \sigma_1^{*2}.$$

Applying Matrix Bernstein Lemma E.1, under condition that $p_0 \ge c\mu_0\mu_1 r^2 \kappa^2 (n_1+n_2)\log d/(\gamma^2 d_1 d_2)$, we have

$$\mathbb{P}\bigg\{ \bigg\| \sum_{(i,j)=(1,1)}^{(d_1,d_2)} \mathbf{S}_{ij} \bigg\|_2 \ge \gamma \cdot \frac{\sigma_r^*}{\sqrt{r}} \bigg\} \le (d_1+d_2) \cdot \exp(-c' \log d) \le \frac{1}{d},$$

where $c, c', \gamma > 0$ are some constants. This further implies that with probability at least $1 - 1/d$, we have $\|\mathbf{U}_{\text{init}} \mathbf{V}_{\text{init}}^\top - \mathbf{M}^*\|_2 \le \gamma \sigma_r^* / \sqrt{r}$. Finally, according to Lemma 5.14 in Tu et al. (2015), we obtain

$$D^2(\mathbf{Z}_{\text{init}}, \mathbf{Z}^*) \le \frac{2}{\sqrt{2}-1} \cdot \frac{\|\mathbf{U}_{\text{init}} \mathbf{V}_{\text{init}}^\top - \mathbf{M}^*\|_F^2}{\sigma_r(\mathbf{M}^*)} \le \frac{10r \|\mathbf{U}_{\text{init}} \mathbf{V}_{\text{init}}^\top - \mathbf{M}^*\|_2^2}{\sigma_r^*} \le 10\gamma^2 \sigma_r^*,$$

where the second inequality holds because $\text{rank}(\mathbf{U}_{\text{init}} \mathbf{V}_{\text{init}}^\top - \mathbf{M}^*)$ is at most $2r$. □

B.2. Proof of Theorem 5.4

Before proceeding to the main proof, we introduce the following notations and facts. Recall that $\mathbf{M}^* = \overline{\mathbf{U}}^* \mathbf{\Sigma}^* \overline{\mathbf{V}}^{*\top}$ and $\mathbf{Z}^* = [\overline{\mathbf{U}}^*; \overline{\mathbf{V}}^*] \mathbf{\Sigma}^{*1/2}$, we denote $\widetilde{\mathbf{Z}}^* = [\overline{\mathbf{U}}^*; -\overline{\mathbf{V}}^*] \mathbf{\Sigma}^{*1/2}$. Note that $\mathbf{M}^*$ and $\mathbf{L}^*$ have the same set of singular values, thus for any $\ell \in [r]$, $\sigma_\ell^2(\mathbf{Z}^*) = \sigma_\ell^2(\widetilde{\mathbf{Z}}^*) = 2\sigma_\ell^*$. We further note that $\mathbf{Z}^{*\top} \widetilde{\mathbf{Z}}^* = \widetilde{\mathbf{Z}}^{*\top} \mathbf{Z}^* = \mathbf{0}$, and $\text{Sym}(\mathbf{M}^*) = (\mathbf{Z}^* \mathbf{Z}^{*\top} - \widetilde{\mathbf{Z}}^* \widetilde{\mathbf{Z}}^{*\top})/2$. Define reference function $G(\mathbf{Z})$ as $G(\mathbf{Z}) = \|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^* \mathbf{Z}^{*\top}\|_F^2 / 4$, then the gradient of G is given by

$$\nabla G(\mathbf{Z}) = (\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^* \mathbf{Z}^{*\top}) \mathbf{Z}. \tag{B.7}$$

Thus according to (B.3), we have

$$\nabla \widetilde{f}_\Omega(\mathbf{Z}) = \frac{1}{2} \nabla G(\mathbf{Z}) + \frac{1}{2} (\widetilde{\mathbf{Z}}^* \widetilde{\mathbf{Z}}^{*\top}) \mathbf{Z} + \mathbf{X}^\top \bigg(\frac{1}{p} \mathcal{P}_{\widetilde{\Omega}} - \mathcal{P}_{\text{off}} \bigg) (\mathbf{X}[\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^* \mathbf{Z}^{*\top}] \mathbf{X}^\top) \mathbf{X}\mathbf{Z}. \tag{B.8}$$

The following lemmas demonstrate the local curvature and local smoothness properties of $\widetilde{f}_\Omega$, which are proved in Sections C.1 and C.2, respectively. In both lemmas, for any $\mathbf{Z} \in \mathbb{R}^{(n_1+n_2)\times r}$, we let $\mathbf{R} = \mathrm{argmin}_{\widetilde{\mathbf{R}}\in\mathbb{Q}_r} \|\mathbf{Z} - \mathbf{Z}^*\widetilde{\mathbf{R}}\|_F$, and denote $\mathbf{H} = \mathbf{Z} - \mathbf{Z}^*\mathbf{R}$.

Lemma B.1 (Local curvature)**.** Under the previously stated assumptions in Theorem 5.4, for any fixed $\mathbf{Z} = [\mathbf{U}; \mathbf{V}] \in \mathbb{R}^{(n_1+n_2)\times r}$ such that $D(\mathbf{Z}, \mathbf{Z}^*) \le \sqrt{2\sigma_r^*/5}$ and $\|\mathbf{X}_L\mathbf{U}\|_{2,\infty} \le 2\sqrt{\mu_0 r\sigma_1^*/d_1}$, $\|\mathbf{X}_R\mathbf{V}\|_{2,\infty} \le 2\sqrt{\mu_0 r\sigma_1^*/d_2}$, there exists constants c_1, c_2 such that if $|\Omega| \ge c_1 \max\{\mu_0^2 r^2\kappa^2, \mu_0\mu_1 r\kappa n\}\log d$, then with probability at least $1 - c_2/d$, we have

$$\langle \nabla\widetilde{f}_\Omega(\mathbf{Z}), \mathbf{H}\rangle \ge \frac{1}{20}\|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2 + \frac{1}{4\sigma_1^*}\|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}\|_F^2 + \frac{\sigma_r^*}{8}\|\mathbf{H}\|_F^2 - 40\|\mathbf{H}\|_F^4,$$

where $\widetilde{\mathbf{Z}}^* = [\mathbf{U}^*; -\mathbf{V}^*]$.

Lemma B.2 (Local smoothness)**.** Under the previous stated assumptions as in Theorem 5.4, for any fixed $\mathbf{Z} = [\mathbf{U}; \mathbf{V}] \in \mathbb{R}^{(n_1+n_2)\times r}$ such that $D(\mathbf{Z}, \mathbf{Z}^*) \le \sqrt{\sigma_r^*}/4$ and $\|\mathbf{X}_L\mathbf{U}\|_{2,\infty} \le 2\sqrt{\mu_0 r\sigma_1^*/d_1}$, $\|\mathbf{X}_R\mathbf{V}\|_{2,\infty} \le 2\sqrt{\mu_0 r\sigma_1^*/d_2}$, there exist constants c_1, c_2 such that if $|\Omega| \ge c_1\mu_0\mu_1 r\kappa n\log d$, then with probability at least $1 - c_2/d$, we have

$$\|\nabla\widetilde{f}_\Omega(\mathbf{Z})\|_F^2 \le (16r+4)\sigma_1^*\|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2 + \|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}\|_F^2 + 4r\sigma_r^{*2}\|\mathbf{H}\|_F^2,$$

where $\widetilde{\mathbf{Z}}^* = [\mathbf{U}^*; -\mathbf{V}^*]$.

Now we are ready to prove Theorem 5.4.

Proof. Theorem 5.4 will be proved by induction. Consider the s-th iteration in Phase 2 of Algorithm 1, for any $s \ge 1$. Suppose the previous iterate $\mathbf{Z}_{s-1}$ is sufficiently close to $\mathbf{Z}^*$, i.e., $D(\mathbf{Z}_{s-1}, \mathbf{Z}^*) \le \alpha\sqrt{\sigma_r^*}$, where α is defined in Theorem 5.4. In the following discussions, we are going to show that the following contraction result with respect to the s-th iterate:

$$D^2(\mathbf{Z}_s, \mathbf{Z}^*) \le \left(1 - \frac{\eta\sigma_r^*}{16}\right)\cdot D^2(\mathbf{Z}_{s-1}, \mathbf{Z}^*) + 3\delta\cdot\alpha\sqrt{\sigma_r^*} + 2\delta^2 \tag{B.9}$$

holds with probability at least $1 - c_1/d$, where $\mathbf{Z}_s = [\mathbf{U}_s; \mathbf{V}_s]$ and $\mathbf{Z}_{s-1} = [\mathbf{U}_{s-1}; \mathbf{V}_{s-1}]$.

Denote the optimal rotation with respect to $\mathbf{Z}_{\text{init}}$ by $\mathbf{R}_{\text{init}}$ such that $\mathbf{R}_{\text{init}} = \mathrm{argmin}_{\mathbf{R}\in\mathbb{Q}_r}\|\mathbf{Z}_{\text{init}} - \mathbf{Z}^*\mathbf{R}\|_F$. Since the initial iterate $\mathbf{Z}_{\text{init}}$ satisfies $D(\mathbf{Z}_{\text{init}}, \mathbf{Z}^*) \le \sqrt{\sigma_r^*}/40$, we have $\|\mathbf{Z}_{\text{init}} - \mathbf{Z}^*\mathbf{R}_{\text{init}}\|_2 \le \sqrt{\sigma_r^*}/40$, which implies

$$\sqrt{\sigma_1^*} \le \|\mathbf{Z}^*\mathbf{R}_{\text{init}}\|_2 - \|\mathbf{Z}_{\text{init}} - \mathbf{Z}^*\mathbf{R}_{\text{init}}\|_2 \le \|\mathbf{Z}_{\text{init}}\|_2 \le \|\mathbf{Z}^*\mathbf{R}_{\text{init}}\|_2 + \|\mathbf{Z}_{\text{init}} - \mathbf{Z}^*\mathbf{R}_{\text{init}}\|_2 \le 2\sqrt{\sigma_1^*}.$$

Thus, according to the definition of $\mathcal{C}_1, \mathcal{C}_2$ in (4.2) and Assumption 3.1, we have

$$\|\mathbf{X}_L\mathbf{U}^*\|_{2,\infty} \le \|\mathbf{X}_L\overline{\mathbf{U}}^*\|_{2,\infty}\cdot\|\mathbf{\Sigma}^*\|_2^{1/2} \le \sqrt{\frac{\mu r\sigma_1^*}{d_1}} \le \sqrt{\frac{\mu r}{d_1}}\cdot\|\mathbf{Z}_{\text{init}}\|_2,$$

which implies that $\mathbf{U}^* \in \mathcal{C}_1$. Similarly, we can derive that $\mathbf{V}^* \in \mathcal{C}_2$. In addition, based on the definition of the constraint sets $\mathcal{C}_1$ and $\mathcal{C}_2$ in (4.2), we further have $\|\mathbf{U}\|_{2,\infty} \le 2\sqrt{\mu_0 r\sigma_1^*/d_1}$ and $\|\mathbf{V}\|_{2,\infty} \le 2\sqrt{\mu_0 r\sigma_1^*/d_2}$. For any $s \in \{1, 2, \ldots, S\}$, we denote $\mathbf{R}_s = \mathrm{argmin}_{\mathbf{R}\in\mathbb{Q}_r}\|\mathbf{Z}_s - \mathbf{Z}^*\mathbf{R}\|_F$ as the optimal rotation with respect to $\mathbf{Z}_s$, and we let $\mathbf{H}_s = \mathbf{Z}_s - \mathbf{Z}^*\mathbf{R}_s$. Consider the s-iteration of Phase 2 in Algorithm 1, we let $\widehat{\mathbf{U}}_s = \mathbf{U}_{s-1} - \eta\nabla_{\mathbf{U}} f_{\Omega_s}(\mathbf{U}_{s-1}, \mathbf{V}_{s-1})$ and $\widehat{\mathbf{V}}_s = \mathbf{V}_{s-1} - \eta\nabla_{\mathbf{V}} f_{\Omega_s}(\mathbf{U}_{s-1}, \mathbf{V}_{s-1})$. Thus based on the update rule, we have

$$\begin{aligned} D^2(\mathbf{Z}_s, \mathbf{Z}^*) &\le \|\mathbf{U}_s - \mathbf{U}^*\mathbf{R}_{s-1}\|_F^2 + \|\mathbf{V}_s - \mathbf{V}^*\mathbf{R}_{s-1}\|_F^2 \\ &= \underbrace{\|\mathcal{P}_{\mathcal{C}_1}(\widehat{\mathbf{U}}_s, \delta) - \mathbf{U}^*\mathbf{R}_{s-1}\|_F^2}_{I_1} + \underbrace{\|\mathcal{P}_{\mathcal{C}_2}(\widehat{\mathbf{V}}_s, \delta) - \mathbf{V}^*\mathbf{R}_{s-1}\|_F^2}_{I_2}, \end{aligned}$$

where the first inequality follows from Definition 5.1, and the second inequality follows from the update rule. As for the first term I_1, we have

$$\begin{aligned} I_1 &= \|\mathcal{P}_{\mathcal{C}_1}(\widehat{\mathbf{U}}_s) - \mathbf{U}^*\mathbf{R}_{s-1}\|_F^2 + 2\langle\mathcal{P}_{\mathcal{C}_1}(\widehat{\mathbf{U}}_s, \delta) - \mathcal{P}_{\mathcal{C}_1}(\widehat{\mathbf{U}}_s), \mathcal{P}_{\mathcal{C}_1}(\widehat{\mathbf{U}}_s) - \mathbf{U}^*\mathbf{R}_{s-1}\rangle \\ &\quad + \|\mathcal{P}_{\mathcal{C}_1}(\widehat{\mathbf{U}}_s, \delta) - \mathcal{P}_{\mathcal{C}_1}(\widehat{\mathbf{U}}_s)\|_F^2 \\ &\le \|\mathcal{P}_{\mathcal{C}_1}(\widehat{\mathbf{U}}_s) - \mathbf{U}^*\mathbf{R}_{s-1}\|_F^2 + 2\delta\cdot\|\mathcal{P}_{\mathcal{C}_1}(\widehat{\mathbf{U}}_s) - \mathbf{U}^*\mathbf{R}_{s-1}\|_F + \delta^2 \\ &\le \|\mathbf{U}_{s-1} - \eta\nabla_{\mathbf{U}} f_{\Omega_s}(\mathbf{U}_{s-1}, \mathbf{V}_{s-1}) - \mathbf{U}^*\mathbf{R}_{s-1}\|_F^2 + 2\delta\cdot\|\mathbf{U}_{s-1} - \eta\nabla_{\mathbf{U}} f_{\Omega_s}(\mathbf{U}_{s-1}, \mathbf{V}_{s-1}) - \mathbf{U}^*\mathbf{R}_{s-1}\|_F + \delta^2, \end{aligned}$$

where the first inequality holds because $\mathcal{P}_{\mathcal{C}_1}(\widehat{\mathbf{U}}_s, \delta)$ is the δ-approximate solution and $\mathcal{P}_{\mathcal{C}_1}(\widehat{\mathbf{U}}_s)$ is the exact solution to the same optimization problem, and the second inequality is due to the non-expansive property of projection $\mathcal{P}_{\mathcal{C}_1}$ and the fact that $\mathbf{U}^* \in \mathcal{C}_1$. Based on the similar technique, we can upper bound I_2 as follows

$$I_2 \leq \|\mathbf{V}_{s-1} - \eta\nabla_{\mathbf{V}} f_{\Omega_s}(\mathbf{U}_{s-1}, \mathbf{V}_{s-1}) - \mathbf{V}^*\mathbf{R}_{s-1}\|_F^2 + 2\delta \cdot \|\mathbf{V}_{s-1} - \eta\nabla_{\mathbf{V}} f_{\Omega_s}(\mathbf{U}_{s-1}, \mathbf{V}_{s-1}) - \mathbf{V}^*\mathbf{R}_{s-1}\|_F + \delta^2.$$

Therefore, combining the upper bounds of I_1 and I_2, we have

$$\begin{aligned} D^2(\mathbf{Z}_s, \mathbf{Z}^*) &\leq D^2(\mathbf{Z}_{s-1}, \mathbf{Z}^*) - 2\eta\langle\nabla\widetilde{f}_{\Omega_s}(\mathbf{Z}_{s-1}), \mathbf{H}_{s-1}\rangle + \eta^2\|\nabla\widetilde{f}_{\Omega_s}(\mathbf{Z}_{s-1})\|_F^2 \\ &\quad + 2\sqrt{2}\delta \cdot \Big(D^2(\mathbf{Z}_{s-1}, \mathbf{Z}^*) - 2\eta\langle\nabla\widetilde{f}_{\Omega_s}(\mathbf{Z}_{s-1}), \mathbf{H}_{s-1}\rangle + \eta^2\|\nabla\widetilde{f}_{\Omega_s}(\mathbf{Z}_{s-1})\|_F^2\Big)^{1/2} + 2\delta^2, \end{aligned}$$

where $\widetilde{f}_{\Omega_s}$ is defined in (B.3), and the inequality follows from the triangle inequality.

It is worth noting that the subsampling technique ensures the previous iterate $\mathbf{Z}_{s-1}$ is independent of the samples Ω_s used in the s-th iteration. According to the assumptions of Theorem 5.4, $|\Omega_s| = |\Omega|/(2S) \leq c_1 \max\{\mu_0^2 r^2\kappa^2, \mu_0\mu_1 r\kappa n\}\log d$, thus we can directly apply Lemmas B.1 and B.2. More specifically, with probability at least $1 - c_1/d$, we have

$$\langle\nabla\widetilde{f}_{\Omega_s}(\mathbf{Z}_{s-1}), \mathbf{H}_{s-1}\rangle \geq \frac{1}{20}\|\mathbf{Z}_{s-1}\mathbf{Z}_{s-1}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2 + \frac{1}{4\sigma_1^*}\|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}_{s-1}\|_F^2 + \frac{\sigma_r^*}{8}\|\mathbf{H}_{s-1}\|_F^2 - 40\|\mathbf{H}_{s-1}\|_F^4,$$

where $\widetilde{\mathbf{Z}}^* = [\mathbf{U}^*; -\mathbf{V}^*]$. In addition, we have

$$\|\nabla\widetilde{f}_{\Omega_s}(\mathbf{Z}_s)\|_F^2 \leq (16r + 4)\sigma_1^*\|\mathbf{Z}_{s-1}\mathbf{Z}_{s-1}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2 + \|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}_{s-1}\|_F^2 + 4r\sigma_r^{*2}\|\mathbf{H}_{s-1}\|_F^2.$$

Thus, by setting step size $\eta \leq 1/(200r\sigma_1^*)$, we obtain

$$-2\eta\langle\nabla\widetilde{f}_{\Omega_s}(\mathbf{Z}_{s-1}), \mathbf{H}_{s-1}\rangle + \eta^2\|\nabla\widetilde{f}_{\Omega_s}(\mathbf{Z}_s)\|_F^2 \leq -\frac{\eta\sigma_r^*}{8}\|\mathbf{H}_{s-1}\|_F^2 + 80\eta\,\|\mathbf{H}_{s-1}\|_F^4,$$

which implies that the induction hypothesis that $D(\mathbf{Z}_{s-1}, \mathbf{Z}^*) = \|\mathbf{H}_{s-1}\|_F \leq \alpha\sqrt{\sigma_r^*}$, we have

$$\begin{aligned} D^2(\mathbf{Z}_s, \mathbf{Z}^*) &\leq \left(1 - \frac{\eta\sigma_r^*}{16}\right) \cdot D^2(\mathbf{Z}_{s-1}, \mathbf{Z}^*) + 2\delta \cdot \sqrt{2\left(1 - \frac{\eta\sigma_r^*}{16}\right)} \cdot D(\mathbf{Z}_{s-1}, \mathbf{Z}^*) + 2\delta^2 \\ &\leq \left(1 - \frac{\eta\sigma_r^*}{16}\right) \cdot D^2(\mathbf{Z}_{s-1}, \mathbf{Z}^*) + 3\delta \cdot \alpha\sqrt{\sigma_r^*} + 2\delta^2, \end{aligned}$$

which completes the proof of (B.9).

Moreover, according to the assumption that $D(\mathbf{Z}_{\text{init}}, \mathbf{Z}^*) \leq \alpha\sqrt{\sigma_r^*}/2$, if we choose $\delta \leq \alpha\sqrt{\sigma_r^*}/(2\sqrt{2})$, then we have

$$\begin{aligned} D^2(\mathbf{Z}_0, \mathbf{Z}^*) &\leq \|\mathbf{U}_0 - \mathbf{U}^*\mathbf{R}_{\text{init}}\|_F^2 + \|\mathbf{V}_0 - \mathbf{V}^*\mathbf{R}_{\text{init}}\|_F^2 \\ &\leq \|\mathcal{P}_{\mathcal{C}_1}(\mathbf{U}_{\text{init}}) - \mathbf{U}^*\mathbf{R}_{\text{init}}\|_F^2 + \|\mathcal{P}_{\mathcal{C}_2}(\mathbf{V}_{\text{init}}) - \mathbf{V}^*\mathbf{R}_{\text{init}}\|_F^2 \\ &\qquad + 2\delta \cdot \big(\|\mathcal{P}_{\mathcal{C}_1}(\mathbf{U}_{\text{init}}) - \mathbf{U}^*\mathbf{R}_{\text{init}}\|_F + \|\mathcal{P}_{\mathcal{C}_2}(\mathbf{V}_{\text{init}}) - \mathbf{V}^*\mathbf{R}_{\text{init}}\|_F\big) + 2\delta^2 \\ &\leq D^2(\mathbf{Z}_{\text{init}}, \mathbf{Z}^*) + 2\sqrt{2}\delta \cdot D(\mathbf{Z}_{\text{init}}, \mathbf{Z}^*) + 2\delta^2 \leq \alpha^2\sigma_r^*, \end{aligned}$$

where we let $\mathbf{R}_{\text{init}}$ be the optimal rotation with respect to $\mathbf{Z}_{\text{init}}$, and the last inequality follows from the non-expansiveness property of projection onto convex set. Thus, we have shown that the induction hypothesis $D(\mathbf{Z}_{s-1}, \mathbf{Z}^*) \leq \alpha\sqrt{\sigma_r^*}$ holds for the first iterate.

To this end, it remains to verify the induction step, or more specifically, $D(\mathbf{Z}_{s-1}, \mathbf{Z}^*) \leq \alpha\sqrt{\sigma_r^*}$ implies $D(\mathbf{Z}_s, \mathbf{Z}^*) \leq \alpha\sqrt{\sigma_r^*}$, for any $s \geq 1$. This step can be proved based on (B.9): with high probability, we have

$$\begin{aligned} D^2(\mathbf{Z}_s, \mathbf{Z}^*) &\leq \left(1 - \frac{\eta\sigma_r^*}{16}\right) \cdot D^2(\mathbf{Z}_{s-1}, \mathbf{Z}^*) + 3\delta \cdot \alpha\sqrt{\sigma_r^*} + 2\delta^2 \\ &\leq \left(1 - \frac{\eta\sigma_r^*}{16}\right) \cdot \alpha^2\sigma_r^* + 3\delta \cdot \alpha\sqrt{\sigma_r^*} + 2\delta^2 \leq \alpha^2\sigma_r^*, \end{aligned}$$

provided that $\delta \leq c_2\sqrt{\sigma_r^*}/(r\kappa)$ with constant c_2 sufficiently small. Finally, by induction and union bound, we obtain

$$D^2(\mathbf{Z}_S, \mathbf{Z}^*) \leq \left(1 - \frac{\eta\sigma_r^*}{16}\right)^S \cdot D^2(\mathbf{Z}_0, \mathbf{Z}^*) + \frac{16}{\eta\sigma_r^*} \cdot (3\delta \cdot \alpha\sqrt{\sigma_r^*} + 2\delta^2)$$

holds with probability at least $1 - c_1 S/d$, which completes the proof. □

B.3. Proof of Theorem 5.5

The proof of Theorem 5.5 will be similar to the proof of Theorem 5.4. The only difference is that we do not require sample splitting in Phase 3, thus the iterates are no longer independent from the subset of samples. The following curvature and smoothness of $\widetilde{f}_\Omega$ are proved for all $\mathbf{Z} \in \mathbb{R}^{(n_1+n_2)\times r}$ satisfying $D(\mathbf{Z}, \mathbf{Z}^*) \leq c_0\sqrt{\sigma_r^*}/(\mu_1 n)$. The proofs are presented in Sections C.3 and C.4 respectively.

Lemma B.3 (Local curvature)**.** Under the previously stated assumptions in Theorem 5.5, for all $\mathbf{Z} = [\mathbf{U}; \mathbf{V}] \in \mathbb{R}^{(n_1+n_2)\times r}$ such that $D(\mathbf{Z}, \mathbf{Z}^*) \leq c_0\sqrt{\sigma_r^*}/(\mu_1 n)$ with constant c_0 small enough, there exists constants c_1, c_2 such that if $|\Omega| \geq c_1\mu_0\mu_1 rn\log d$, then with probability at least $1 - c_2/d$, we have

$$\langle \nabla \widetilde{f}_\Omega(\mathbf{Z}), \mathbf{H}\rangle \geq \frac{1}{20}\|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2 + \frac{1}{4\sigma_1^*}\|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}\|_F^2 + \frac{\sigma_r^*}{8}\|\mathbf{H}\|_F^2 - 20\|\mathbf{H}\|_F^4,$$

where $\widetilde{\mathbf{Z}}^* = [\mathbf{U}^*; -\mathbf{V}^*]$.

Lemma B.4 (Local smoothness)**.** Under the previous stated assumptions as in Theorem 5.5, for all $\mathbf{Z} = [\mathbf{U}; \mathbf{V}] \in \mathbb{R}^{(n_1+n_2)\times r}$ such that $D(\mathbf{Z}, \mathbf{Z}^*) \leq c_0\sqrt{\sigma_r^*}/(\mu_1 n)$ with constant c_0 small enough, there exist constants c_1, c_2 such that if $|\Omega| \geq c_1\mu_0\mu_1 rn\log d$, then with probability at least $1 - c_2/d$, we have

$$\|\nabla \widetilde{f}_\Omega(\mathbf{Z})\|_F^2 \leq 84\sigma_1^*\|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2 + \|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}\|_F^2 + 4\sigma_1^*\sigma_r^*\|\mathbf{H}\|_F^2 + 40\sigma_1^*\|\mathbf{H}\|_F^4,$$

where $\widetilde{\mathbf{Z}}^* = [\mathbf{U}^*; -\mathbf{V}^*]$.

Now we are ready to prove Theorem 5.5.

Proof. For any $t \in \{0, 1, \ldots, T\}$, we denote $\mathbf{R}^t = \operatorname{argmin}_{\mathbf{R}\in\mathbb{Q}_r} \|\mathbf{Z}^t - \mathbf{Z}^*\mathbf{R}\|_F$ as the optimal rotation with respect to $\mathbf{Z}^t$, and we let $\mathbf{H}^t = \mathbf{Z}^t - \mathbf{Z}^*\mathbf{R}^t$. Note that the initial iterate of Phase 3 in Algorithm 1 satisfying $D(\mathbf{Z}^0, \mathbf{Z}^*) \leq c_0\sqrt{\sigma_r^*}/(\mu n)$. Assume the induction hypothesis $D(\mathbf{Z}^s, \mathbf{Z}^*) \leq c_0\sqrt{\sigma_r^*}/(\mu n)$ holds for $s = 1, 2, \ldots, t$. Consider the t-th iteration, based on the update rule, we have

$$\begin{aligned} D^2(\mathbf{Z}^{t+1}, \mathbf{Z}^*) &\leq \|\mathbf{U}^{t+1} - \mathbf{U}^*\mathbf{R}^t\|_F^2 + \|\mathbf{V}^{t+1} - \mathbf{V}^*\mathbf{R}^t\|_F^2 \\ &= \|\mathbf{U}^t - \tau\nabla_{\mathbf{U}} f_\Omega(\mathbf{U}^t, \mathbf{V}^t) - \mathbf{U}^*\mathbf{R}^t\|_F^2 + \|\mathbf{V}^t - \tau\nabla_{\mathbf{V}} f_\Omega(\mathbf{U}^t, \mathbf{V}^t) - \mathbf{V}^*\mathbf{R}^t\|_F^2 \\ &= D^2(\mathbf{Z}^t, \mathbf{Z}^*) - 2\tau\langle \nabla \widetilde{f}_\Omega(\mathbf{Z}^t), \mathbf{H}^t\rangle + \tau^2\|\nabla \widetilde{f}_\Omega(\mathbf{Z}^t)\|_F^2, \end{aligned}$$

where the inequality follows from Definition 5.1. According to the assumptions of Theorem 5.5, we can directly apply Lemmas B.3 and B.4. More specifically, with probability at least $1 - c'/d$, we have

$$\langle \nabla \widetilde{f}_\Omega(\mathbf{Z}^t), \mathbf{H}^t\rangle \geq \frac{1}{20}\|\mathbf{Z}^t\mathbf{Z}^{t\top} - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2 + \frac{1}{4\sigma_1^*}\|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}^t\|_F^2 + \frac{\sigma_r^*}{8}\|\mathbf{H}^t\|_F^2 - 20\|\mathbf{H}^t\|_F^4,$$

where $\widetilde{\mathbf{Z}}^* = [\mathbf{U}^*; -\mathbf{V}^*]$. In addition, we have

$$\|\nabla \widetilde{f}_\Omega(\mathbf{Z}^t)\|_F^2 \leq 84\sigma_1^*\|\mathbf{Z}^t\mathbf{Z}^{t\top} - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2 + \|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}^t\|_F^2 + 4\sigma_1^*\sigma_r^*\|\mathbf{H}^t\|_F^2 + 40\sigma_1^*\|\mathbf{H}^t\|_F^4.$$

Thus, by setting step size $\tau \leq c_1/\sigma_1^*$ with $c_1 \leq 1/840$, we obtain

$$-2\tau\langle \nabla \widetilde{f}_{\Omega_s}(\mathbf{Z}^t), \mathbf{H}^t\rangle + \tau^2\|\nabla \widetilde{f}_\Omega(\mathbf{Z}^t)\|_F^2 \leq -\frac{\eta\sigma_r^*}{8}\|\mathbf{H}^t\|_F^2 + 50\eta\,\|\mathbf{H}^t\|_F^4,$$

which implies that under condition that $D(\mathbf{Z}^t, \mathbf{Z}^*) = \|\mathbf{H}^t\|_F \leq c_2\sqrt{\sigma_r^*}$ with $c_2 \leq 1/30$, we have

$$D^2(\mathbf{Z}^{t+1}, \mathbf{Z}^*) \leq \left(1 - \frac{\tau\sigma_r^*}{16}\right) \cdot D^2(\mathbf{Z}^t, \mathbf{Z}^*),$$

which implies the $(t+1)$-th iterate $\mathbf{Z}^{t+1}$ still satisfies $D^2(\mathbf{Z}^{t+1}, \mathbf{Z}^*) \leq c_0\sqrt{\sigma_r^*}/(\mu n)$. Thus by induction, we complete the proof. □

B.4. Proof of Theorem 5.2

Proof. To prove the overall theoretical guarantee of Algorithm 1, we need to examine the conditions required by previous main theorems regarding the three phases. More specifically, to ensure the $O(\sqrt{\sigma_r^*})$ initial ball assumption $D(\mathbf{Z}_{\text{init}}, \mathbf{Z}^*) \leq \alpha\sqrt{\sigma_r^*}/2$ in Theorem 5.4, it suffices to set $\delta \leq \alpha/8$ in Theorem 5.3, which implies the sample complexity required by Phase 1 is $|\Omega_0| = O(r^2\kappa^2 n \log d)$. In addition, according to Theorem 5.4, we have

$$D^2(\mathbf{Z}_S, \mathbf{Z}^*) \leq \left(1 - \frac{c_2}{16r\kappa}\right)^S \cdot \alpha^2\sigma_r^* + c_4\delta \cdot r\kappa\sqrt{\sigma_r^*}.$$

Thus, in order to guarantee the $O(\sqrt{\sigma_r^*}/n)$ initial ball assumption $D(\mathbf{Z}_S, \mathbf{Z}^*) \leq c_0\sqrt{\sigma_r^*}/(\mu_1 n)$ holds in Theorem 5.5, it is sufficient to perform $S = O(r\kappa \log n)$ iterations in Phase 2 of Algorithm 1 and let the approximation error $\delta = O(1/(r\kappa n^2))$. Based on Theorem 5.4, we derive the sample complexity required by Phase 2 is $c \cdot \max\{\mu_1 n, \mu_0 r\kappa\}\mu_0 r^2\kappa^2 \log n \log d$. Together with Theorem 5.5, we conclude that the overall sample complexity of Algorithm 1.

Finally, as for the reconstruction error $\|\mathbf{M}^T - \mathbf{M}^*\|_F$, let $\mathbf{R}^T$ be the optimal rotation between $\mathbf{Z}^T$ and $\mathbf{Z}^*$, then we have

$$\begin{aligned}\|\mathbf{U}^T(\mathbf{V}^T)^\top - \mathbf{U}^*\mathbf{V}^{*\top}\|_F &\leq \|\mathbf{U}^T(\mathbf{V}^T - \mathbf{V}^*\mathbf{R}^T)^\top\|_F + \|(\mathbf{U}^T - \mathbf{U}^*\mathbf{R}^T)(\mathbf{V}^*\mathbf{R}^T)^\top\|_F \\ &\leq \|\mathbf{U}^T\|_2 \cdot \|\mathbf{V}^T - \mathbf{V}^*\mathbf{R}^T\|_F + \|\mathbf{V}^*\mathbf{R}^T\|_2 \cdot \|\mathbf{U}^T - \mathbf{U}^*\mathbf{R}^T\|_F \\ &\leq 3\sqrt{\sigma_1^*} \cdot D(\mathbf{Z}^T, \mathbf{Z}^*),\end{aligned} \tag{B.10}$$

where the second inequality is due to $\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_2 \cdot \|\mathbf{B}\|_F$, and the last inequality follows from the fact that $\|\mathbf{Z}^T - \mathbf{Z}^*\mathbf{R}^T\|_F \leq D(\mathbf{Z}_S, \mathbf{Z}^*) \leq \alpha\sqrt{\sigma_r^*}/(\mu_1 n) \leq \sqrt{\sigma_1^*}$. Noticing that according to Theorem 5.5, Phase 3 achieves linear rate of convergence, which implies that with $T = O(\kappa \log(1/\epsilon))$ iterations, we have $D(\mathbf{Z}^T, \mathbf{Z}^*) \leq \epsilon$. Combining with (B.10), we complete the proof. □

C. Proofs of the Technical Lemmas in Section B

In this section, we provide the theoretical proofs of the technical lemmas used in Section B.

C.1. Proof of Lemma B.1

To prove Lemma B.1, we need to make use of the following auxiliary lemmas. Inspired by Tu et al. (2015), we show that the reference function $G(\mathbf{Z})$ has a similar local curvature property in Lemma C.1. Based on Matrix Bernstein Inequality, Lemma C.2 generalizes the results of Theorem 4.1 in Candès & Recht (2009) to inductive setting, while Lemma C.3 provides the high probability bound on the remaining term.

Lemma C.1. Let $\mathbf{Z}, \mathbf{Z}^* \in \mathbb{R}^{(n_1+n_2)\times r}$, and $G(\mathbf{Z}) = \|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2/4$. For any $\mathbf{Z}$ satisfying

$$\|\mathbf{Z} - \mathbf{Z}^*\mathbf{R}\|_2^2 \leq \sigma_r^2(\mathbf{Z}^*)/5, \text{ where } \mathbf{R} = \operatorname*{argmin}_{\widetilde{\mathbf{R}} \in \mathbb{Q}_r} \|\mathbf{Z} - \mathbf{Z}^*\widetilde{\mathbf{R}}\|_F,$$

we have

$$\langle \nabla G(\mathbf{Z}), \mathbf{Z} - \mathbf{Z}^*\mathbf{R}\rangle \geq \frac{\sigma_r^2(\mathbf{Z}^*)}{4}\|\mathbf{Z} - \mathbf{Z}^*\mathbf{R}\|_F^2 + \frac{1}{4}\|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2.$$

Lemma C.2. Assume the previously stated assumptions in Lemma B.1 hold. Define subspace

$$\mathcal{T} = \left\{\overline{\mathbf{U}}^*\mathbf{A}^\top + \mathbf{B}\overline{\mathbf{V}}^{*\top}, \text{ for some } \mathbf{A} \in \mathbb{R}^{n_2\times r} \text{ and } \mathbf{B} \in \mathbb{R}^{n_1 \times r}\right\}.$$

Let $\mathcal{P}_\mathcal{T} : \mathbb{R}^{n_1\times n_2} \to \mathbb{R}^{n_1\times n_2}$ be the Euclidean projection onto $\mathcal{T}$. Specifically, for any $\mathbf{Z} \in \mathbb{R}^{n_1\times n_2}$, we have

$$\mathcal{P}_\mathcal{T}(\mathbf{Z}) = \overline{\mathbf{U}}^*\overline{\mathbf{U}}^{*\top}\mathbf{Z} + \mathbf{Z}\overline{\mathbf{V}}^*\overline{\mathbf{V}}^{*\top} - \overline{\mathbf{U}}^*\overline{\mathbf{U}}^{*\top}\mathbf{Z}\overline{\mathbf{V}}^*\overline{\mathbf{V}}^{*\top}. \tag{C.1}$$

For any $\gamma \in (0,1)$, there exist constants c_1, c_2 such that if $|\Omega| \geq c_1\mu_0\mu_1 rn\log d/\gamma^2$, then for all $\mathbf{Z} \in \mathbb{R}^{n_1\times n_2}$, with probability at least $1 - c_2/d$, we have

$$\left\|\mathcal{P}_\mathcal{T}(\mathbf{Z}) - p^{-1}\mathcal{P}_\mathcal{T}\Big(\mathbf{X}_L^\top\mathcal{P}_\Omega\big(\mathbf{X}_L\mathcal{P}_\mathcal{T}(\mathbf{Z})\mathbf{X}_R^\top\big)\mathbf{X}_R\Big)\right\|_F \leq \gamma\|\mathbf{Z}\|_F. \tag{C.2}$$

Moreover for all $\mathbf{Z}_1, \mathbf{Z}_2 \in \mathcal{T}$, we have

$$\big|\langle \mathbf{X}_L\mathbf{Z}_1\mathbf{X}_R^\top - p^{-1}\mathcal{P}_\Omega\big(\mathbf{X}_L\mathbf{Z}_1\mathbf{X}_R^\top\big), \mathbf{X}_L\mathbf{Z}_2\mathbf{X}_R^\top\rangle\big| \leq \gamma\|\mathbf{Z}_1\|_F\cdot\|\mathbf{Z}_2\|_F,$$

and for all $\mathbf{Z} \in \mathcal{T}$, we have

$$p^{-1}\big\|\mathcal{P}_\Omega\big(\mathbf{X}_L\mathbf{Z}\mathbf{X}_R^\top\big)\big\|_F^2 \leq (1+\gamma)\|\mathbf{Z}\|_F^2.$$

Lemma C.3. Assume the previously stated assumptions in Lemma B.1 hold. For any fixed $\mathbf{U}\in\mathbb{R}^{n_1\times r}$, $\mathbf{V}\in\mathbb{R}^{n_2\times r}$ satisfying $\|\mathbf{X}_L\mathbf{U}\|_{2,\infty} \leq 3\sqrt{\mu_0 r\sigma_1^*/d_1}$ and $\|\mathbf{X}_R\mathbf{V}\|_{2,\infty} \leq 3\sqrt{\mu_0 r\sigma_1^*/d_2}$, there exist constants c_1, c_2 such that with probability $1-c_1/d$, we have

$$\frac{1}{p}\left\|\mathcal{P}_\Omega(\mathbf{X}_L\mathbf{U}\mathbf{V}^\top\mathbf{X}_R^\top)\right\|_F^2 \leq \|\mathbf{U}\mathbf{V}^\top\|_F^2 + \gamma\sigma_r^*\cdot(\|\mathbf{U}\|_F^2+\|\mathbf{V}\|_F^2),$$

provided that $|\Omega| \geq c_2\max\{\mu_0^2r^2\kappa^2, \mu_0\mu_1 r\kappa n\}\log d/\gamma^2$.

Now we are ready to prove Lemma B.1.

Proof of Lemma B.1. Recall that $\mathbf{R} = \operatorname{argmin}_{\widetilde{\mathbf{R}}\in\mathbb{Q}_r}\|\mathbf{Z}-\mathbf{Z}^*\widetilde{\mathbf{R}}\|_F$ and $\mathbf{H} = \mathbf{Z}-\mathbf{Z}^*\mathbf{R}$. According to the gradient of $\widetilde{f}_\Omega$ in (B.8), we have

$$\langle\nabla\widetilde{f}_\Omega(\mathbf{Z}),\mathbf{H}\rangle = \underbrace{\frac{1}{2}\langle\nabla G(\mathbf{Z}),\mathbf{H}\rangle + \frac{1}{2}\langle\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z},\mathbf{H}\rangle}_{I_1} + \underbrace{\langle(p^{-1}\mathcal{P}_{\widetilde{\Omega}}-\mathcal{P}_{\text{off}})\big(\mathbf{X}\big[\mathbf{Z}\mathbf{Z}^\top-\mathbf{Z}^*\mathbf{Z}^{*\top}\big]\mathbf{X}^\top\big),\mathbf{X}\mathbf{H}(\mathbf{X}\mathbf{Z})^\top\rangle}_{I_2}.$$

In the sequel, we are going to bound the terms I_1 and I_2, respectively.

Lower bound of I_1. According to the assumption $\|\mathbf{H}\|_F \leq \sqrt{2\sigma_r^*/5}$, we have $\|\mathbf{Z}-\mathbf{Z}^*\mathbf{R}\|_2^2 \leq \sigma_r^2(\mathbf{Z}^*)/5$. Thus, we can apply Lemma C.1 directly

$$\begin{aligned}
I_1 &= \frac{1}{2}\langle\nabla G(\mathbf{Z}),\mathbf{H}\rangle + \frac{1}{2}\operatorname{tr}(\mathbf{Z}^\top\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z})\\
&\geq \frac{\sigma_r^2(\mathbf{Z}^*)}{8}\|\mathbf{H}\|_F^2 + \frac{1}{8}\|\mathbf{Z}\mathbf{Z}^\top-\mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2 + \frac{1}{2\|\widetilde{\mathbf{Z}}^*\|_2^2}\|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}\|_F^2\\
&= \frac{\sigma_r^*}{4}\|\mathbf{H}\|_F^2 + \frac{1}{8}\|\mathbf{Z}\mathbf{Z}^\top-\mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2 + \frac{1}{4\sigma_1^*}\|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}\|_F^2,
\end{aligned} \tag{C.3}$$

where the first equality holds because $\mathbf{Z}^{*\top}\widetilde{\mathbf{Z}}^* = \mathbf{0}$, the second inequality follows from Lemma C.1 and the fact that $\|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}\|_F^2 \leq \|\widetilde{\mathbf{Z}}^*\|_2^2\cdot\|\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}\|_F^2$, and the third equality holds because $\sigma_r^2(\mathbf{Z}^*) = 2\sigma_r^*$ and $\|\widetilde{\mathbf{Z}}^*\|_2^2 = 2\sigma_1^*$.

Upper bound of $|I_2|$. Note that $\mathbf{Z} = [\mathbf{U};\mathbf{V}]$ and $\mathbf{M}^* = \mathbf{U}^*\mathbf{V}^{*\top}$, we denote $\mathbf{M} = \mathbf{U}\mathbf{V}^\top$, $\widetilde{\mathbf{U}} = \mathbf{U}^*\mathbf{R}$ and $\widetilde{\mathbf{V}} = \mathbf{V}^*\mathbf{R}$. Besides, let $\mathbf{H}_U\in\mathbb{R}^{n_1\times r}$, $\mathbf{H}_V\in\mathbb{R}^{n_2\times r}$ be the top n_1 and bottom n_2 rows of $\mathbf{H}$, respectively, then we have $\mathbf{U} = \widetilde{\mathbf{U}}+\mathbf{H}_U$ and $\mathbf{V} = \widetilde{\mathbf{V}}+\mathbf{H}_V$. Note that $\mathbf{U}\mathbf{V}^\top - \widetilde{\mathbf{U}}\widetilde{\mathbf{V}}^\top = \widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top + \mathbf{H}_U\mathbf{H}_V^\top$, and $\mathbf{H}_U\mathbf{V}^\top + \mathbf{U}\mathbf{H}_V^\top = \widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top + 2\mathbf{H}_U\mathbf{H}_V^\top$. Based on the above notations, we can reformulate I_2 as follows

$$\begin{aligned}
I_2 &= \langle(p^{-1}\mathcal{P}_\Omega-\mathcal{I})(\mathbf{X}_L[\mathbf{U}\mathbf{V}^\top-\widetilde{\mathbf{U}}\widetilde{\mathbf{V}}^\top]\mathbf{X}_R^\top),\mathbf{X}_L(\mathbf{H}_U\mathbf{V}^\top+\mathbf{U}\mathbf{H}_V^\top)\mathbf{X}_R^\top\rangle\\
&= \underbrace{\langle(p^{-1}\mathcal{P}_\Omega-\mathcal{I})(\mathbf{X}_L[\widetilde{\mathbf{U}}\mathbf{H}_V^\top+\mathbf{H}_U\widetilde{\mathbf{V}}^\top]\mathbf{X}_R^\top),\mathbf{X}_L(\widetilde{\mathbf{U}}\mathbf{H}_V^\top+\mathbf{H}_U\widetilde{\mathbf{V}}^\top)\mathbf{X}_R^\top\rangle}_{I_{21}}\\
&\quad + 3\underbrace{\langle(p^{-1}\mathcal{P}_\Omega-\mathcal{I})(\mathbf{X}_L[\widetilde{\mathbf{U}}\mathbf{H}_V^\top+\mathbf{H}_U\widetilde{\mathbf{V}}^\top]\mathbf{X}_R^\top),\mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top\rangle}_{I_{22}}\\
&\quad + 2\underbrace{\langle(p^{-1}\mathcal{P}_\Omega-\mathcal{I})(\mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top),\mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top\rangle}_{I_{23}}.
\end{aligned}$$

Note that $\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top$ falls into the subspace $\mathcal{T}$ defined in Lemma C.2. Thus according to Lemma C.2, we obtain

$$|I_{21}| \leq \gamma\|\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top\|_F^2 \leq 2\gamma\|\mathbf{M} - \mathbf{M}^*\|_F^2 + \frac{\gamma}{2}\|\mathbf{H}\|_F^4 \tag{C.4}$$

holds with probability at least $1 - c_2/d$, provided that $|\Omega| \geq c_1\mu_0\mu_1 rn\log d/\gamma^2$, where γ is a constant such that $\gamma \in (0,1)$. Here, the second inequality holds because $\|\mathbf{A}+\mathbf{B}\|_F^2 \leq 2\|\mathbf{A}\|_F^2 + 2\|\mathbf{B}\|_F^2$ and $\|\mathbf{H}_U\mathbf{H}_V^\top\|_F \leq \|\mathbf{H}\|_F^2/2$. As for the term I_{22}, we have

$$\begin{aligned}
|I_{22}| &\leq \frac{1}{p}\big|\langle \mathcal{P}_\Omega(\mathbf{X}_L[\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top]\mathbf{X}_R^\top), \mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top\rangle\big| + \big|\langle \mathbf{X}_L(\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top)\mathbf{X}_R^\top, \mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top\rangle\big| \\
&\leq \frac{1}{2\beta}\cdot p^{-1}\big\|\mathcal{P}_\Omega(\mathbf{X}_L[\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top]\mathbf{X}_R^\top)\big\|_F^2 + \frac{\beta}{2}\cdot p^{-1}\big\|\mathcal{P}_\Omega(\mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top)\big\|_F^2 \\
&\quad + \frac{1}{2\beta}\|\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top\|_F^2 + \frac{\beta}{2}\|\mathbf{H}_U\mathbf{H}_V^\top\|_F^2 \\
&\leq \frac{\beta}{2}\cdot p^{-1}\big\|\mathcal{P}_\Omega(\mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top)\big\|_F^2 + \frac{2+\gamma}{2\beta}\|\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top\|_F^2 + \frac{\beta}{8}\|\mathbf{H}\|_F^4,
\end{aligned} \tag{C.5}$$

where constant $\beta > 0$ will be specified later. Here, the second inequality follows from the Young's Inequality, and the last inequality is due to Lemma C.2 and the fact that $\|\mathbf{H}_U\mathbf{H}_V^\top\|_F \leq \|\mathbf{H}_U\|_F \cdot \|\mathbf{H}_V\|_F \leq \|\mathbf{H}\|_F^2/2$. Since we have $\|\mathbf{X}_L\mathbf{H}_U\|_{2,\infty} \leq \|\mathbf{X}_L\mathbf{U}\|_{2,\infty} + \|\mathbf{X}_L\widetilde{\mathbf{U}}\|_{2,\infty} \leq 3\sqrt{\mu_0 r\sigma_1^*/d_1}$, and similarly we have $\|\mathbf{X}_R\mathbf{H}_V\|_{2,\infty} \leq 3\sqrt{\mu_0 r\sigma_1^*/d_2}$, according to Lemma C.3, we further obtain

$$\frac{1}{p}\big\|\mathcal{P}_\Omega(\mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top)\big\|_F^2 \leq \|\mathbf{H}_U\mathbf{H}_V\|_F^2 + \gamma\sigma_r^*\|\mathbf{H}\|_F^2 \leq \frac{1}{4}\|\mathbf{H}\|_F^4 + \gamma\sigma_r^*\|\mathbf{H}\|_F^2 \tag{C.6}$$

holds with probability at least $1 - c_1/d$, provided that $|\Omega| \geq c_2\max\{\mu_0^2 r^2\kappa^2, \mu_0\mu_1 r\kappa n\}\log d/\gamma^2$. Thus, plugging (C.6) into (C.5), we have

$$\begin{aligned}
|I_{22}| &\leq \frac{\beta}{2}\gamma\sigma_r^*\|\mathbf{H}\|_F^2 + \frac{2+\gamma}{2\beta}\|\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top\|_F^2 + \frac{\beta}{4}\|\mathbf{H}\|_F^4 \\
&\leq \frac{2+\gamma}{\beta}\|\mathbf{M} - \mathbf{M}^*\|_F^2 + \frac{\beta}{2}\gamma\sigma_r^*\|\mathbf{H}\|_F^2 + \left(\frac{2+\gamma}{4\beta} + \frac{\beta}{4}\right)\|\mathbf{H}\|_F^4,
\end{aligned} \tag{C.7}$$

where $\mathbf{M} = \mathbf{U}\mathbf{V}^\top$. Similarly, according to Lemma C.3, we can upper bound $|I_{23}|$

$$\begin{aligned}
|I_{23}| &\leq \frac{1}{p}|\langle \mathcal{P}_\Omega(\mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top), \mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top\rangle| + |\langle \mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top, \mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top\rangle| \\
&\leq \frac{1}{p}\|\mathcal{P}_\Omega(\mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top)\|_F^2 + \|\mathbf{H}_U\mathbf{H}_V^\top\|_F^2 \\
&\leq \frac{1}{2}\|\mathbf{H}\|_F^4 + \gamma\sigma_r^*\|\mathbf{H}\|_F^2.
\end{aligned} \tag{C.8}$$

Therefore, combining (C.4),(C.7) and (C.8), we obtain the upper bound of $|I_2|$

$$\begin{aligned}
|I_2| &\leq |I_{21}| + 3|I_{22}| + 2|I_{23}| \\
&\leq \left(2\gamma + \frac{3(2+\gamma)}{\beta}\right)\|\mathbf{M} - \mathbf{M}^*\|_F^2 + \left(\frac{\gamma}{2} + \frac{3(2+\gamma)}{4\beta} + \frac{3\beta}{4} + 1\right)\|\mathbf{H}\|_F^4 + \left(\frac{3\beta}{2} + 2\right)\gamma\sigma_r^*\|\mathbf{H}\|_F^2.
\end{aligned} \tag{C.9}$$

Finally, set $\beta = 48$, then combining (C.3) and (C.9), we obtain

$$\langle \nabla\widetilde{f}_\Omega(\mathbf{Z}), \mathbf{H}\rangle \geq \left(\frac{1}{16} - \frac{3\gamma}{2}\right)\|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2 + \frac{1}{4\sigma_1^*}\|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}\|_F^2 + \left(\frac{\sigma_r^*}{4} - 75\gamma\sigma_r^*\right)\|\mathbf{H}\|_F^2 - 40\|\mathbf{H}\|_F^4,$$

where the inequality holds because $\|\mathbf{M} - \mathbf{M}^*\|_F^2 \leq \|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2/2$. Thus, by choosing constant γ to be sufficiently small, we complete the proof. □

C.2. Proof of Lemma B.2

Proof. Recall the gradient of $\widetilde{f}_\Omega$ in (B.3), we have

$$\|\nabla \widetilde{f}_\Omega(\mathbf{Z})\|_F^2 \leq 2 \underbrace{\left\| p^{-1}\mathbf{X}^\top \mathcal{P}_{\widetilde{\Omega}}\big(\mathbf{X}[\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}]\mathbf{X}^\top\big)\mathbf{X}\mathbf{Z}\right\|_F^2}_{I_1} + \frac{1}{2} \underbrace{\left\|(\overline{\mathcal{P}}_{\text{diag}} - \overline{\mathcal{P}}_{\text{off}})\mathbf{Z}\mathbf{Z}^\top\mathbf{Z}\right\|_F^2}_{I_2},$$

where the inequality holds because $\|\mathbf{A}+\mathbf{B}\|_F^2 \leq 2(\|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2)$. In the sequel, we will upper bound I_1 and I_2, respectively.

Upper bound of I_1. Recall that $\mathbf{Z} = [\mathbf{U};\mathbf{V}]$, $\mathbf{R} = \operatorname{argmin}_{\widetilde{\mathbf{R}}\in\mathbb{Q}_r} \|\mathbf{Z} - \mathbf{Z}^*\widetilde{\mathbf{R}}\|_F$, and $\mathbf{H} = \mathbf{Z} - \mathbf{Z}^*\mathbf{R}$. Denote $\mathbf{M} = \mathbf{U}\mathbf{V}^\top$, then based on the above notations, we have

$$\begin{aligned} I_1 &= \|p^{-1}\mathbf{X}_L^\top \mathcal{P}_\Omega(\mathbf{X}_L[\mathbf{M}-\mathbf{M}^*]\mathbf{X}_R^\top)\mathbf{X}_R\mathbf{V}\|_F^2 + \left\|p^{-1}\mathbf{X}_R^\top\big(\mathcal{P}_\Omega(\mathbf{X}_L[\mathbf{M}-\mathbf{M}^*]\mathbf{X}_R^\top)\big)^\top\mathbf{X}_L\mathbf{U}\right\|_F^2 \\ &\leq r\cdot \underbrace{\|p^{-1}\mathbf{X}_L^\top \mathcal{P}_\Omega(\mathbf{X}_L[\mathbf{M}-\mathbf{M}^*]\mathbf{X}_R^\top)\mathbf{X}_R\mathbf{V}\|_2^2}_{I_{11}} + r\cdot \underbrace{\left\|p^{-1}\mathbf{X}_R^\top\big(\mathcal{P}_\Omega(\mathbf{X}_L[\mathbf{M}-\mathbf{M}^*]\mathbf{X}_R^\top)\big)^\top\mathbf{X}_L\mathbf{U}\right\|_2^2}_{I_{12}}, \end{aligned} \tag{C.10}$$

where the inequality holds because both $\mathbf{U}$ and $\mathbf{V}$ have rank at most r. Consider the term I_{11} first, we observe

$$\begin{aligned} &\frac{1}{p}\mathbf{X}_L^\top \mathcal{P}_\Omega(\mathbf{X}_L[\mathbf{M}-\mathbf{M}^*]\mathbf{X}_R^\top)\mathbf{X}_R\mathbf{V} - (\mathbf{M}-\mathbf{M}^*)\mathbf{V} \\ &= \sum_{(i,j)=(1,1)}^{(d_1,d_2)} (p^{-1}\xi_{ij}-1)\cdot[\mathbf{X}_L(\mathbf{M}-\mathbf{M}^*)\mathbf{X}_R^\top]_{ij}\cdot[\mathbf{X}_L]_{i,*}^\top\cdot[\mathbf{X}_R\mathbf{V}]_{j,*} := \sum_{(i,j)=(1,1)}^{(d_1,d_2)} \mathbf{A}_{ij}, \end{aligned}$$

where $\xi_{ij}=1$, if $(i,j)\in\Omega$; $\xi_{ij}=0$, otherwise. We are going to apply matrix bernstein inequality to the above summation. Due to sample splitting, the randomness only comes from Ω, thus $\mathbf{A}_{ij}$'s are independent and $\mathbb{E}[\mathbf{A}_{ij}] = 0$. Denote $\widetilde{\mathbf{U}} = \mathbf{U}^*\mathbf{R}$ and $\widetilde{\mathbf{V}} = \mathbf{V}^*\mathbf{R}$, and let $\mathbf{H}_U \in \mathbb{R}^{n_1\times r}$, $\mathbf{H}_V \in \mathbb{R}^{n_2\times r}$ be the top n_1 and bottom n_2 rows of $\mathbf{H}$. Then we have $\mathbf{M}-\mathbf{M}^* = \widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top + \mathbf{H}_U\mathbf{H}_V^\top$. For any (i,j), we have the following upper bound of $\|\mathbf{A}_{ij}\|_2$

$$\begin{aligned} \|\mathbf{A}_{ij}\|_2 &\leq \frac{1}{p}\left|[\mathbf{X}_L(\mathbf{M}-\mathbf{M}^*)\mathbf{X}_R^\top]_{ij}\right|\cdot\|[\mathbf{X}_L]_{i,*}^\top\cdot[\mathbf{X}_R\mathbf{V}]_{j,*}\|_2 \\ &\leq \frac{1}{p}\|\mathbf{X}_L(\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top + \mathbf{H}_U\mathbf{H}_V^\top)\mathbf{X}_R^\top\|_{\infty,\infty}\cdot\|\mathbf{X}_L\|_{2,\infty}\cdot\|\mathbf{X}_R\mathbf{V}\|_{2,\infty} \\ &\leq \frac{10\mu_0\mu_1 rn\sigma_1^*}{pd_1d_2}\cdot\|\mathbf{H}\|_F, \end{aligned}$$

where the last inequality is due to Assumptions 3.1, 3.2 and the fact that $\|\mathbf{X}_L\mathbf{H}_U\|_{2,\infty} \leq \|\mathbf{X}_L\mathbf{U}\|_{2,\infty} + \|\mathbf{X}_L\widetilde{\mathbf{U}}\|_{2,\infty} \leq 3\sqrt{\mu_0 r\sigma_1^*/d_1}$ and $\|\mathbf{X}_R\mathbf{V}\|_{2,\infty} \leq 2\sqrt{\mu_0 r\sigma_1^*/d_2}$. To apply matrix bernstein, it remains to bound $\|\sum_{i,j}\mathbb{E}[\mathbf{A}_{ij}\mathbf{A}_{ij}^\top]\|_2$ and $\|\sum_{i,j}\mathbb{E}[\mathbf{A}_{ij}^\top\mathbf{A}_{ij}]\|_2$. In particular, we have

$$\begin{aligned} \Big\|\sum_{(i,j)=(1,1)}^{(d_1,d_2)}\mathbb{E}[\mathbf{A}_{ij}\mathbf{A}_{ij}^\top]\Big\|_2 &= \frac{1-p}{p}\Big\|\sum_{(i,j)=(1,1)}^{(d_1,d_2)}[\mathbf{X}_L(\mathbf{M}-\mathbf{M}^*)\mathbf{X}_R^\top]_{ij}^2\cdot[\mathbf{X}_L]_{i,*}^\top\cdot\big\|[\mathbf{X}_R\mathbf{V}]_{j,*}\big\|_2^2\cdot[\mathbf{X}_L]_{i,*}\Big\|_2 \\ &\leq \frac{1}{p}\Big\|\sum_{i=1}^{d_1}\mathbf{e}_i\mathbf{e}_i^\top\cdot\sum_{j=1}^{d_2}\Big([\mathbf{X}_L(\mathbf{M}-\mathbf{M}^*)\mathbf{X}_R^\top]_{ij}^2\cdot\big\|[\mathbf{X}_R\mathbf{V}]_{j,*}\big\|_2^2\Big)\Big\|_2 \\ &\leq \frac{1}{p}\max_{i\in[d_1]}\sum_{j=1}^{d_2}\Big([\mathbf{X}_L(\mathbf{M}-\mathbf{M}^*)\mathbf{X}_R^\top]_{ij}^2\cdot\|\mathbf{X}_R\mathbf{V}\|_{2,\infty}^2\Big) \\ &\leq \frac{4\mu_0 r\sigma_1^*}{pd_2}\cdot\big\|\mathbf{X}_L(\mathbf{M}-\mathbf{M}^*)\mathbf{X}_R^\top\big\|_{2,\infty}^2 \leq \frac{c\mu_0\mu_1 r\sigma_1^* n}{pd_1d_2}\cdot\|\mathbf{M}-\mathbf{M}^*\|_F^2, \end{aligned}$$

where the last inequality holds because $\|\mathbf{A}\mathbf{B}\|_{2,\infty} \leq \|\mathbf{A}\|_{2,\infty} \cdot \|\mathbf{B}\|_F$. Similarly, we have

$$\begin{aligned}
\bigg\| \sum_{(i,j)=(1,1)}^{(d_1,d_2)} \mathbb{E}[\mathbf{A}_{ij}^\top \mathbf{A}_{ij}] \bigg\|_2 &\leq \frac{1}{p} \bigg\| \sum_{(i,j)=(1,1)}^{(d_1,d_2)} [\mathbf{X}_L(\mathbf{M}-\mathbf{M}^*)\mathbf{X}_R^\top]_{ij}^2 \cdot [\mathbf{X}_R\mathbf{V}]_{j,*}^\top \cdot \big\|[\mathbf{X}_L]_{i,*}\big\|_2^2 \cdot [\mathbf{X}_R\mathbf{V}]_{j,*} \bigg\|_2 \\
&\leq \frac{1}{p} \|\mathbf{X}_L\|_{2,\infty}^2 \cdot \|\mathbf{X}_R\mathbf{V}\|_{2,\infty}^2 \cdot \sum_{(i,j)=(1,1)}^{(d_1,d_2)} [\mathbf{X}_L(\mathbf{M}-\mathbf{M}^*)\mathbf{X}_R^\top]_{ij}^2 \\
&\leq \frac{4\mu_0\mu_1 r\sigma_1^* n}{pd_1d_2} \cdot \|\mathbf{M}-\mathbf{M}^*\|_F^2,
\end{aligned}$$

where the second inequality follows from the definition of spectral norm. Therefore, according to Lemma E.1, with probability at least $1-c/d$, we have

$$\bigg\| \sum_{(i,j)=(1,1)}^{(d_1,d_2)} \mathbf{A}_{ij} \bigg\|_2 \leq \gamma\sigma_r^{*1/2}\|\mathbf{M}-\mathbf{M}^*\|_F + \gamma^2\sigma_r^*\|\mathbf{H}\|_F,$$

under condition that $|\Omega| \geq c'\mu_0\mu_1 rn\kappa/\gamma^2$, where c, c' are both constants. Hence, by triangle's inequality, we obtain the upper bound of I_{11}

$$I_{11} \leq \big(\|(\mathbf{M}-\mathbf{M}^*)\mathbf{V}\|_2 + \gamma\sigma_r^{*1/2}\|\mathbf{M}-\mathbf{M}^*\|_F + \gamma^2\sigma_r^*\|\mathbf{H}\|_F\big)^2 \leq 8\sigma_1^*\|\mathbf{M}-\mathbf{M}^*\|_F^2 + \sigma_r^{*2}\|\mathbf{H}\|_F^2, \tag{C.11}$$

where the last inequality holds because $\|\mathbf{V}\|_2 \leq \|\mathbf{V}^*\|_2 + \|\mathbf{V}-\mathbf{V}^*\mathbf{R}\|_2 \leq 2\sqrt{\sigma_1^*}$ and $\gamma \in (0, 1/2)$. Similarly, we obtain the upper bound of I_{12}

$$I_{12} \leq \big(\|\mathbf{U}^\top(\mathbf{M}-\mathbf{M}^*)\|_2 + \gamma\sigma_r^{*1/2}\|\mathbf{M}-\mathbf{M}^*\|_F + \gamma^2\sigma_r^{*2}\|\mathbf{H}\|_F\big)^2 \leq 8\sigma_1^*\|\mathbf{M}-\mathbf{M}^*\|_F^2 + \sigma_r^{*2}\|\mathbf{H}\|_F^2. \tag{C.12}$$

Plugging (C.11) and (C.12) into (C.10), we have

$$I_1 \leq 2r \cdot (8\sigma_1^*\|\mathbf{M}-\mathbf{M}^*\|_F^2 + \sigma_r^{*2}\|\mathbf{H}\|_F^2). \tag{C.13}$$

Upper bound of I_2. As for I_2, we obtain

$$\begin{aligned}
I_2 &= \|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z} - (\overline{\mathcal{P}}_{\text{diag}} - \overline{\mathcal{P}}_{\text{off}})(\mathbf{Z}^*\mathbf{Z}^{*\top} - \mathbf{Z}\mathbf{Z}^\top)\mathbf{Z}\|_F^2 \\
&\leq 2\|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}\|_F^2 + 2\|(\overline{\mathcal{P}}_{\text{diag}} - \overline{\mathcal{P}}_{\text{off}})(\mathbf{Z}^*\mathbf{Z}^{*\top} - \mathbf{Z}\mathbf{Z}^\top)\|_F^2 \cdot \|\mathbf{Z}\|_2^2 \\
&\leq 2\|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}\|_F^2 + 8\sigma_1^*\|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2,
\end{aligned} \tag{C.14}$$

where the first inequality holds because $\|\mathbf{A}-\mathbf{B}\|_F^2 \leq 2(\|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2)$ and $\|\mathbf{A}\mathbf{B}\|_F \leq \|\mathbf{A}\|_F \cdot \|\mathbf{B}\|_2$, and the second inequality holds because $\|\mathbf{Z}\|_2 \leq 2\sqrt{\sigma_1^*}$.

Finally, combining (C.13) and (C.14), we obtain

$$\begin{aligned}
\|\nabla\widetilde{f}_\Omega(\mathbf{Z})\|_F^2 &\leq 32r\sigma_1^*\|\mathbf{M}-\mathbf{M}^*\|_F^2 + 4r\sigma_r^{*2}\|\mathbf{H}\|_F^2 + \|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}\|_F^2 + 4\sigma_1^*\|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2 \\
&\leq (16r+4)\sigma_1^*\|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2 + \|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}\|_F^2 + 4r\sigma_r^{*2}\|\mathbf{H}\|_F^2,
\end{aligned}$$

where the second inequality holds because $\|\mathbf{M}-\mathbf{M}^*\|_F^2 \leq \|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2/2$, which completes the proof. □

C.3. Proof of Lemma B.3

Proof of Lemma B.3. Similar to the proof of Lemma B.1, we have

$$\langle\nabla\widetilde{f}_\Omega(\mathbf{Z}), \mathbf{H}\rangle = \underbrace{\frac{1}{2}\langle\nabla G(\mathbf{Z}), \mathbf{H}\rangle + \frac{1}{2}\langle\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}, \mathbf{H}\rangle}_{I_1} + \underbrace{\big\langle(p^{-1}\mathcal{P}_{\widetilde{\Omega}} - \mathcal{P}_{\text{off}})\big(\mathbf{X}\big[\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\big]\mathbf{X}^\top\big), \mathbf{X}\mathbf{H}(\mathbf{X}\mathbf{Z})^\top\big\rangle}_{I_2}.$$

According to Lemma C.1, we have

$$I_1 \geq \frac{\sigma_r^*}{4}\|\mathbf{H}\|_F^2 + \frac{1}{8}\|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2 + \frac{1}{4\sigma_1^*}\|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}\|_F^2. \tag{C.15}$$

Recall the notations in the proof of Lemma B.1, we let $\mathbf{Z} = [\mathbf{U}; \mathbf{V}]$ and $\mathbf{M}^* = \mathbf{U}^*\mathbf{V}^{*\top}$, we denote $\mathbf{M} = \mathbf{U}\mathbf{V}^\top$, $\widetilde{\mathbf{U}} = \mathbf{U}^*\mathbf{R}$ and $\widetilde{\mathbf{V}} = \mathbf{V}^*\mathbf{R}$. Besides, let $\mathbf{H}_U \in \mathbb{R}^{n_1\times r}$, $\mathbf{H}_V \in \mathbb{R}^{n_2\times r}$ be the top n_1 and bottom n_2 rows of $\mathbf{H}$. Then we can reformulate I_2 as follows

$$\begin{aligned} I_2 = &\underbrace{\langle (p^{-1}\mathcal{P}_\Omega - \mathcal{I})(\mathbf{X}_L[\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top]\mathbf{X}_R^\top), \mathbf{X}_L(\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top)\mathbf{X}_R^\top\rangle}_{I_{21}} \\ &+ 3\underbrace{\langle (p^{-1}\mathcal{P}_\Omega - \mathcal{I})(\mathbf{X}_L[\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top]\mathbf{X}_R^\top), \mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top\rangle}_{I_{22}} \\ &+ 2\underbrace{\langle (p^{-1}\mathcal{P}_\Omega - \mathcal{I})(\mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top), \mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top\rangle}_{I_{23}}. \end{aligned}$$

Note that $\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top$ falls into the subspace $\mathcal{T}$ defined in Lemma C.2. Thus according to Lemma C.2, we can still obtain the same upper bound of I_{21}

$$|I_{21}| \leq \gamma\|\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top\|_F^2 \leq 2\gamma\|\mathbf{M} - \mathbf{M}^*\|_F^2 + \frac{\gamma}{2}\|\mathbf{H}\|_F^4$$

holds with probability at least $1 - c_2/d$, provided that $|\Omega| \geq c_1\mu_0\mu_1 rn\log d/\gamma^2$. As for the term I_{22}, similarly we have

$$\begin{aligned} |I_{22}| &\leq \frac{1}{p}\big|\langle \mathcal{P}_\Omega(\mathbf{X}_L[\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top]\mathbf{X}_R^\top), \mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top\rangle\big| + \big|\langle \mathbf{X}_L(\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top)\mathbf{X}_R^\top, \mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top\rangle\big| \\ &\leq \frac{\beta}{2}\cdot p^{-1}\big\|\mathcal{P}_\Omega(\mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top)\big\|_F^2 + \frac{2+\gamma}{2\beta}\|\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top\|_F^2 + \frac{\beta}{8}\|\mathbf{H}\|_F^4, \end{aligned}$$

where constant $\beta > 0$ will be specified later. Here, the second inequality follows from the Young's Inequality and Lemma C.2. According to bernstein-type inequality for Bernoulli random variables, we further obtain

$$\frac{1}{p}\big\|\mathcal{P}_\Omega(\mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top)\big\|_F^2 \leq \frac{1}{p}|\Omega|\cdot\|\mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top\|_{\infty,\infty}^2 \leq \frac{3}{2}\mu_1^2 n_1 n_2\cdot\|\mathbf{H}_U\mathbf{H}_V^\top\|_F^2 \leq c_0^2\sigma_r^*\|\mathbf{H}\|_F^2$$

holds with probability at least $1 - c_1/d$, where the second inequality follows from Lemma E.2 and the incoherence Assumptions 3.2, and the last inequality holds because $\|\mathbf{H}\|_F^2 = D^2(\mathbf{Z}, \mathbf{Z}^*) \leq c_0^2\sigma_r^*/(\mu_1^2 n^2)$. Therefore, we obtain the upper bound of I_{22}

$$\begin{aligned} |I_{22}| &\leq \frac{2+\gamma}{2\beta}\|\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top\|_F^2 + \frac{\beta}{2}\cdot c_0^2\sigma_r^*\|\mathbf{H}\|_F^2 + \frac{\beta}{8}\|\mathbf{H}\|_F^4 \\ &\leq \frac{2+\gamma}{\beta}\|\mathbf{M} - \mathbf{M}^*\|_F^2 + \frac{\beta}{2}\cdot c_0^2\sigma_r^*\|\mathbf{H}\|_F^2 + \left(\frac{2+\gamma}{4\beta} + \frac{\beta}{8}\right)\|\mathbf{H}\|_F^4. \end{aligned}$$

Similarly, we can upper bound I_{23}

$$\begin{aligned} |I_{23}| &\leq \frac{1}{p}|\langle \mathcal{P}_\Omega(\mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top), \mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top\rangle| + |\langle \mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top, \mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top\rangle| \\ &= \frac{1}{p}\|\mathcal{P}_\Omega(\mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top)\|_F^2 + \|\mathbf{H}_U\mathbf{H}_V^\top\|_F^2 \leq c_0^2\sigma_r^*\|\mathbf{H}\|_F^2 + \frac{1}{4}\|\mathbf{H}\|_F^4. \end{aligned}$$

Hence, we obtain

$$\begin{aligned} |I_2| &\leq |I_{21}| + 3|I_{22}| + 2|I_{23}| \\ &\leq \left(2\gamma + \frac{3(2+\gamma)}{\beta}\right)\cdot\|\mathbf{M} - \mathbf{M}^*\|_F^2 + \left(2 + \frac{3\beta}{2}\right)\cdot c_0^2\sigma_r^*\|\mathbf{H}\|_F^2 + \left(\frac{3(2+\gamma)}{4\beta} + \frac{3\beta}{8} + \frac{1}{2}\right)\cdot\|\mathbf{H}\|_F^4. \end{aligned} \tag{C.16}$$

Set $\beta = 48$, and let constants c_0, γ be sufficiently small. Combining (C.15) and (C.16), we complete the proof. □

C.4. Proof of Lemma B.4

In order to prove Lemma B.4, we need to make use of the following lemma.

Lemma C.4. Let $\mathbf{X} \in \mathbb{R}^{d_1 \times n_1}$, $\mathbf{Y} \in \mathbb{R}^{d_2 \times n_2}$ be the feature matrices, which are orthonormal and self-incoherent with parameter μ_1, and $\Omega \subseteq [d_1] \times [d_2]$ be an index set followed Bernoulli Model (3.1) with $p = |\Omega|/(d_1 d_2)$. For any $\gamma \in (0, 1)$, there exist constants c_1 and c_2 such that, under condition $|\Omega| \geq c_1 \mu_1 n \log d/\gamma^2$, for all $\mathbf{Z} \in \mathbb{R}^{n_1 \times n_2}$

$$\|p^{-1}\mathbf{X}^\top \mathcal{P}_\Omega\big(\mathbf{X}\mathbf{Z}\mathbf{Y}^\top\big)\mathbf{Y}\|_F \leq (1 + \mu_1 n\gamma)\|\mathbf{Z}\|_F,$$

holds with probability at least $1 - c_2/d$.

Now we are ready to prove Lemma B.4.

Proof of Lemma B.4. According to the gradient of $\widetilde{f}_\Omega$ in (B.3), we have

$$\|\nabla \widetilde{f}_\Omega(\mathbf{Z})\|_F^2 \leq 2 \underbrace{\big\|p^{-1}\mathbf{X}^\top \mathcal{P}_{\widetilde{\Omega}}\big(\mathbf{X}[\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}]\mathbf{X}^\top\big)\mathbf{X}\mathbf{Z}\big\|_F^2}_{I_1} + \frac{1}{2}\underbrace{\big\|(\overline{\mathcal{P}}_{\text{diag}} - \overline{\mathcal{P}}_{\text{off}})\mathbf{Z}\mathbf{Z}^\top\mathbf{Z}\big\|_F^2}_{I_2}.$$

In the following discussions, we will upper bound I_1 and I_2 respectively. As for I_1, similar to the proof of Lemma B.2, we have

$$I_1 = \underbrace{\|p^{-1}\mathbf{X}_L^\top \mathcal{P}_\Omega(\mathbf{X}_L[\mathbf{M} - \mathbf{M}^*]\mathbf{X}_R^\top)\mathbf{X}_R\mathbf{V}\|_F^2}_{I_{11}} + \underbrace{\big\|p^{-1}\mathbf{X}_R^\top\big(\mathcal{P}_\Omega(\mathbf{X}_L[\mathbf{M} - \mathbf{M}^*]\mathbf{X}_R^\top)\big)^\top\mathbf{X}_L\mathbf{U}\big\|_F^2}_{I_{12}}. \tag{C.17}$$

Consider the first term I_{11}. Note that $\mathbf{M} - \mathbf{M}^* = \mathbf{H}_U\widetilde{\mathbf{V}}^\top + \widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\mathbf{H}_V^\top$. Recall the definition of $\mathcal{P}_\mathcal{T}$ in Lemma C.2, we have $\mathcal{P}_\mathcal{T}(\mathbf{M}) + \mathcal{P}_{\mathcal{T}^\perp}(\mathbf{M}) = \mathbf{M}$, for any $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$, where $\mathcal{P}_{\mathcal{T}^\perp}(\mathbf{M}) = (\mathbf{I} - \overline{\mathbf{U}}^*\overline{\mathbf{U}}^{*\top})\mathbf{M}(\mathbf{I} - \overline{\mathbf{V}}^*\overline{\mathbf{V}}^{*\top})$. Thus we obtain

$$\begin{aligned}
I_{11} &\leq 2\big\|p^{-1}\big(\mathbf{X}_L^\top\mathcal{P}_\Omega(\mathbf{X}_L[\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top]\mathbf{X}_R^\top)\mathbf{X}_R\big)\mathbf{V}\big\|_F^2 + 2\big\|p^{-1}\mathbf{X}_L^\top\mathcal{P}_\Omega(\mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top)\mathbf{X}_R\mathbf{V}\big\|_F^2 \\
&\leq 4\underbrace{\big\|p^{-1}\mathcal{P}_\mathcal{T}\big(\mathbf{X}_L^\top\mathcal{P}_\Omega(\mathbf{X}_L[\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top]\mathbf{X}_R^\top)\mathbf{X}_R\big)\mathbf{V}\big\|_F^2}_{J_1} \\
&\quad + 4\underbrace{\big\|p^{-1}\mathcal{P}_{\mathcal{T}^\perp}\big(\mathbf{X}_L^\top\mathcal{P}_\Omega(\mathbf{X}_L[\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top]\mathbf{X}_R^\top)\mathbf{X}_R\big)\mathbf{H}_V\big\|_F^2}_{J_2} + 2\underbrace{\big\|p^{-1}\mathbf{X}_L^\top\mathcal{P}_\Omega(\mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top)\mathbf{X}_R\mathbf{V}\big\|_F^2}_{J_3},
\end{aligned}$$

where the inequality follows from $\|\mathbf{A} + \mathbf{B}\|_F^2 \leq 2\|\mathbf{A}\|_F^2 + 2\|\mathbf{B}\|_F^2$, and the equality is due to the definition of $\mathcal{P}_{\mathcal{T}^\perp}$ and the fact that $(\mathbf{I} - \overline{\mathbf{V}}^*\overline{\mathbf{V}}^{*\top})\widetilde{\mathbf{V}} = \mathbf{0}$. In the sequel, we will upper bound the terms J_1, J_2 and J_3 respectively. Note that $\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top \in \mathcal{T}$, according to Lemma C.2, with probability at least $1 - c'/d$, we have

$$J_1 \leq (1+\gamma)^2\|\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top\|_F^2 \cdot \|\mathbf{V}\|_2^2 \leq 4(1+\gamma)^2\sigma_1^* \cdot \|\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top\|_F^2, \tag{C.18}$$

provided that $|\Omega| \geq \mu_0\mu_1 rn\log d/\gamma^2$, where the first inequality holds because $\|\mathbf{A}\mathbf{B}\|_F \leq \|\mathbf{A}\|_F \cdot \|\mathbf{B}\|_2$, and the second inequality follows from the fact that $\|\mathbf{V}\|_2 \leq \|\mathbf{V}^*\|_2 + \|\mathbf{H}_V\|_2 \leq 2\sqrt{\sigma_1^*}$. Consider the second term J_2, we have

$$\begin{aligned}
J_2 &\leq \|p^{-1}\mathcal{P}_{\mathcal{T}^\perp}\big(\mathbf{X}_L^\top\mathcal{P}_\Omega(\mathbf{X}_L[\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top]\mathbf{X}_R^\top)\mathbf{X}_R\big)\|_F^2 \cdot \|\mathbf{H}_V\|_F^2 \\
&\leq \|p^{-1}\mathbf{X}_L^\top\mathcal{P}_\Omega(\mathbf{X}_L[\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top]\mathbf{X}_R^\top)\mathbf{X}_R\|_F^2 \cdot \|\mathbf{H}_V\|_F^2 \\
&\leq (1 + \mu_1 n\gamma)^2 \cdot \|\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top\|_F^2 \cdot \|\mathbf{H}_V\|_F^2 \leq 2\gamma^2\sigma_r^*\|\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top\|_F^2,
\end{aligned}$$

where the first inequality holds because $\|\mathbf{A}\mathbf{B}\|_F \leq \|\mathbf{A}\|_F \cdot \|\mathbf{B}\|_F$, the second inequality is due to the non-expansiveness of $\mathcal{P}_{\mathcal{T}^\perp}$, the third inequality follows from Lemma C.4, and the last inequality is due to $\|\mathbf{H}\|_F \leq c_0\sqrt{\sigma_r^*}/(\mu_1 n)$. According to Lemma C.4, we can upper bound the last term J_3 as follows

$$\begin{aligned}
J_3 &\leq \big\|p^{-1}\mathbf{X}_L^\top\mathcal{P}_\Omega(\mathbf{X}_L\mathbf{H}_U\mathbf{H}_V^\top\mathbf{X}_R^\top)\mathbf{X}_R\big\|_F^2 \cdot \|\mathbf{V}\|_2^2 \\
&\leq (1 + \mu_1 n\gamma)^2 \cdot \|\mathbf{H}_U\mathbf{H}_V^\top\|_F^2 \cdot \|\mathbf{V}\|_2^2 \leq 2\gamma^2\sigma_1^*\sigma_r^*\|\mathbf{H}\|_F^2,
\end{aligned}$$

where the second inequality follows from Lemma C.4, and the last inequality is due to $\|\mathbf{H}_U\mathbf{H}_V^\top\|_F \leq \|\mathbf{H}\|_F^2/2$ and $\|\mathbf{H}\|_F \leq \alpha\sqrt{\sigma_r^*}/(\mu_1 n)$. Therefore, we obtain the upper bound of I_{11}

$$
\begin{aligned}
I_{11} \leq 4J_1 + 4J_2 + 2J_3 &\leq 20\sigma_1^*\|\widetilde{\mathbf{U}}\mathbf{H}_V^\top + \mathbf{H}_U\widetilde{\mathbf{V}}^\top\|_F^2 + \sigma_1^*\sigma_r^*\|\mathbf{H}\|_F^2 \\
&\leq 40\sigma_1^*\|\mathbf{M}-\mathbf{M}^*\|_F^2 + 10\sigma_1^*\|\mathbf{H}\|_F^4 + \sigma_1^*\sigma_r^*\|\mathbf{H}\|_F^2,
\end{aligned}
$$

where we set γ to be small enough in the first inequality, and the last inequality follows from $\|\mathbf{A}+\mathbf{B}\|_F^2 \leq 2\|\mathbf{A}\|_F^2 + 2\|\mathbf{B}\|_F^2$. By symmetry, we can use the same techniques to bound the term I_{12}, which will yields the same upper bound and implies

$$
I_1 \leq 80\sigma_1^*\|\mathbf{M}-\mathbf{M}^*\|_F^2 + 20\sigma_1^*\|\mathbf{H}\|_F^4 + 2\sigma_1^*\sigma_r^*\|\mathbf{H}\|_F^2.
$$

The upper bound of the remaining term I_2 is as follows

$$
I_2 \leq 2\|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}\|_F^2 + 8\sigma_1^*\|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2,
$$

where the detailed proof can be found in the proof of Lemma B.2. Hence, we obtain

$$
\|\nabla\widetilde{f}_\Omega(\mathbf{Z})\|_F^2 \leq 2I_1 + \frac{1}{2}I_2 \leq 84\sigma_1^*\|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}^*\mathbf{Z}^{*\top}\|_F^2 + \|\widetilde{\mathbf{Z}}^*\widetilde{\mathbf{Z}}^{*\top}\mathbf{Z}\|_F^2 + 4\sigma_1^*\sigma_r^*\|\mathbf{H}\|_F^2 + 40\sigma_1^*\|\mathbf{H}\|_F^4,
$$

which completes the proof. □

D. Proof of Technical Lemmas

In this section, we prove the technical lemmas used in Section C.

D.1. Proof of Lemma C.1

Proof. We begin our proof with some properties regarding the optimal solution. Let $\mathbf{H} = \mathbf{Z} - \mathbf{Z}^*\mathbf{R}$ and $\mathbf{A}\mathbf{\Sigma}\mathbf{B}^\top$ be the SVD of $\mathbf{Z}^{*\top}\mathbf{Z}$, then we have $\mathbf{R} = \mathbf{A}\mathbf{B}^\top$. Thus,

$$
\mathbf{Z}^\top\mathbf{Z}^*\mathbf{R} = \mathbf{B}\mathbf{\Sigma}\mathbf{B}^\top = (\mathbf{Z}^*\mathbf{R})^\top\mathbf{Z},
$$

which implies that $\mathbf{Z}^\top\mathbf{Z}^*\mathbf{R}$ is symmetric and positive definite. Moreover, we have

$$
\mathbf{H}^\top\mathbf{Z}^*\mathbf{R} = \mathbf{Z}^\top\mathbf{Z}^*\mathbf{R} - \mathbf{R}^\top\mathbf{Z}^{*\top}\mathbf{Z}^*\mathbf{R} = \mathbf{Z}^\top\mathbf{Z}^*\mathbf{R} - (\mathbf{Z}^*\mathbf{R})^\top\mathbf{Z}^*\mathbf{R},
$$

which implies $\mathbf{H}^\top\mathbf{Z}^*\mathbf{R}$ is also symmetric. Without loss of generality, we assume $\mathbf{R} = \mathbf{I}$, then $\mathbf{Z}^\top\mathbf{Z}^*$ is positive definite, and $\mathbf{H}^\top\mathbf{Z}^*$ is symmetric. Thus, to prove Lemma C.1, it is sufficient to prove

$$
\langle(\mathbf{Z}^*\mathbf{H}^\top + \mathbf{H}\mathbf{Z}^{*\top} + \mathbf{H}\mathbf{H}^\top)(\mathbf{Z}^* + \mathbf{H}), \mathbf{H}\rangle \geq \frac{\sigma_r^2(\mathbf{Z}^*)}{4}\|\mathbf{H}\|_F^2 + \frac{1}{4}\|\mathbf{Z}^*\mathbf{H}^\top + \mathbf{H}\mathbf{Z}^{*\top} + \mathbf{H}\mathbf{H}^*\|_F^2,
$$

which is equivalent to

$$
\begin{aligned}
0 \leq \mathrm{tr}\bigg(&(\mathbf{H}^\top\mathbf{H})^2 + 3\mathbf{H}^\top\mathbf{H}\mathbf{H}^\top\mathbf{Z}^* + (\mathbf{H}^\top\mathbf{Z}^*)^2 + \mathbf{H}^\top\mathbf{H}\mathbf{Z}^{*\top}\mathbf{Z}^* \\
&- \frac{1}{4}\big[(\mathbf{H}^\top\mathbf{H})^2 + 4\mathbf{H}^\top\mathbf{H}\mathbf{H}^\top\mathbf{Z}^* + 2(\mathbf{H}^\top\mathbf{Z}^*)^2 + 2\mathbf{H}^\top\mathbf{H}\mathbf{Z}^{*\top}\mathbf{Z}^*\big] - \frac{\sigma_r^2(\mathbf{Z}^*)}{4}\mathbf{H}^\top\mathbf{H}\bigg).
\end{aligned}
$$

Combining terms, we have

$$
0 \leq \mathrm{tr}\bigg(\frac{3}{4}(\mathbf{H}^\top\mathbf{H})^2 + 2\mathbf{H}^\top\mathbf{H}\mathbf{H}^\top\mathbf{Z}^* + \frac{1}{2}(\mathbf{H}^\top\mathbf{Z}^*)^2 + \frac{1}{2}\mathbf{H}^\top\mathbf{H}\mathbf{Z}^{*\top}\mathbf{Z}^* - \frac{\sigma_r^2(\mathbf{Z}^*)}{4}\mathbf{H}^\top\mathbf{H}\bigg),
$$

which is further equivalent to

$$
0 \leq \mathrm{tr}\bigg(\frac{1}{2}(\mathbf{H}^\top\mathbf{Z}^* + 2\mathbf{H}^\top\mathbf{H})^2 - \frac{5}{4}(\mathbf{H}^\top\mathbf{H})^2 + \frac{1}{2}\mathbf{H}^\top\mathbf{H}\mathbf{Z}^{*\top}\mathbf{Z}^* - \frac{\sigma_r^2(\mathbf{Z}^*)}{4}\mathbf{H}^\top\mathbf{H}\bigg). \tag{D.1}
$$

Note that $\mathrm{tr}\big((\mathbf{H}^\top\mathbf{H})^2\big) = \|\mathbf{H}^\top\mathbf{H}\|_F^2 \leq \|\mathbf{H}\|_F^2 \cdot \|\mathbf{H}\|_2^2$, and $\mathrm{tr}(\mathbf{H}^\top\mathbf{H}\mathbf{Z}^{*\top}\mathbf{Z}^*) \geq \sigma_r^2(\mathbf{Z}^*) \cdot \|\mathbf{H}\|_F^2$. Therefore, in order to prove (D.1), it is sufficient to require that

$$
\|\mathbf{H}\|_2^2 \leq \frac{\sigma_r^2(\mathbf{Z}^*)}{5},
$$

which completes the proof. □

D.2. Proof of Lemma C.2

Proof. For any $i \in [d_1]$, denote $\mathbf{X}_{i,*} \in \mathbb{R}^{1\times n_1}$ as the i-th row of $\mathbf{X}$. Similarly, for any $j \in [d_2]$, denote $\mathbf{Y}_{j,*} \in \mathbb{R}^{1\times n_2}$ as the j-th row of $\mathbf{Y}$. Thus, for any $\mathbf{Z} \in \mathbb{R}^{n_1\times n_2}$, we have

$$\begin{aligned}\mathbf{X}^\top\mathcal{P}_\Omega\big(\mathbf{X}\mathcal{P}_\mathcal{T}(\mathbf{Z})\mathbf{Y}^\top\big)\mathbf{Y} &= \sum_{(i,j)\in\Omega}\langle \mathbf{X}\mathcal{P}_\mathcal{T}(\mathbf{Z})\mathbf{Y}^\top, \mathbf{e}_i\mathbf{e}_j^\top\rangle \mathbf{X}^\top\mathbf{e}_i\mathbf{e}_j^\top\mathbf{Y}\\ &= \sum_{(i,j)\in\Omega}\langle \mathcal{P}_\mathcal{T}(\mathbf{Z}), \mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}\rangle\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}\\ &= \sum_{(i,j)\in\Omega}\langle \mathbf{Z}, \mathcal{P}_\mathcal{T}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*})\rangle\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}, \end{aligned}\tag{D.2}$$

where the last equality holds because $\langle\mathcal{P}_\mathcal{T}(\mathbf{A}),\mathbf{B}\rangle = \langle\mathbf{A},\mathcal{P}_\mathcal{T}(\mathbf{B})\rangle$. Besides, for any $(i,j)\in[d_1]\times[d_2]$, let $\xi_{ij}=1$, if $(i,j)\in\Omega$, and zero otherwise. Note that both $\mathbf{X}$ and $\mathbf{Y}$ are orthonormal, thus according to (D.2), we have

$$\begin{aligned}\mathcal{P}_\mathcal{T}(\mathbf{Z}) - p^{-1}\mathcal{P}_\mathcal{T}\Big(\mathbf{X}^\top\mathcal{P}_\Omega\big(\mathbf{X}\mathcal{P}_\mathcal{T}(\mathbf{Z})\mathbf{Y}^\top\big)\mathbf{Y}\Big) &= \mathcal{P}_\mathcal{T}\Big(\mathbf{X}^\top(\mathcal{I}-p^{-1}\mathcal{P}_\Omega)(\mathbf{X}\mathcal{P}_\mathcal{T}(\mathbf{Z})\mathbf{Y}^\top)\mathbf{Y}\Big)\\ &= \sum_{\substack{(i,j)\in\\ [d_1]\times[d_2]}}(1-p^{-1}\xi_{ij})\langle\mathbf{Z},\mathcal{P}_\mathcal{T}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*})\rangle\mathcal{P}_\mathcal{T}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}),\end{aligned}$$

where $\mathcal{I}:\mathbb{R}^{d_1\times d_2}\to\mathbb{R}^{d_1\times d_2}$ is the identity mapping. For any $(i,j)\in[d_1]\times[d_2]$, define linear operator $\mathcal{S}_{ij}:\mathbb{R}^{n_1\times n_2}\to\mathbb{R}^{n_1\times n_2}$, such that

$$\mathcal{S}_{ij}(\mathbf{Z}) = (1-p^{-1}\xi_{ij})\langle\mathbf{Z},\mathcal{P}_\mathcal{T}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*})\rangle\mathcal{P}_\mathcal{T}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}).$$

Define $\mathbf{S}_{ij}\in\mathbb{R}^{d_1d_2\times d_1d_2}$ as the corresponding matrix to the linear operator $\mathcal{S}_{ij}$ such that

$$\mathbf{S}_{ij} = (1-p^{-1}\xi_{ij})\mathrm{vec}\big(\mathcal{P}_\mathcal{T}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*})\big)\mathrm{vec}\big(\mathcal{P}_\mathcal{T}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*})\big)^\top,$$

then we can easily show that $\|\sum_{i,j}\mathcal{S}_{ij}(\mathbf{Z})\|_F^2 = \|\sum_{i,j}\mathbf{S}_{ij}\cdot\mathrm{vec}(\mathbf{Z})\|_F^2$, for any $\mathbf{Z}\in\mathbb{R}^{d_1\times d_2}$. Next, we are going to apply matrix bernstein to the summation $\sum_{i,j}\mathbf{S}_{ij}$. We note that $\mathbb{E}[\mathbf{S}_{ij}]=\mathbf{0}$, and $\mathbf{S}_{ij}$ is symmetric, for any $(i,j)\in[d_1]\times[d_2]$. Since $\mathbf{X}\mathbf{M}^*\mathbf{Y}^\top$ is μ_0-incoherent and $\mathbf{M}^* = \overline{\mathbf{U}}^*\boldsymbol{\Sigma}\overline{\mathbf{V}}^{*\top}$, we have $\|\mathbf{X}\overline{\mathbf{U}}^*\|_{2,\infty}\le\sqrt{\mu_0 r/d_1}$, $\|\mathbf{Y}\overline{\mathbf{V}}^*\|_{2,\infty}\le\sqrt{\mu_0 r/d_2}$. According to the definition of $\mathcal{P}_\mathcal{T}$ in (C.1), for any $(i,j)\in[d_1]\times[d_2]$, we obtain

$$\begin{aligned}\|\mathcal{P}_\mathcal{T}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*})\|_F^2 &= \langle\overline{\mathbf{U}}^*\overline{\mathbf{U}}^{*\top}\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*} + \mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}\overline{\mathbf{V}}^*\overline{\mathbf{V}}^{*\top} - \overline{\mathbf{U}}^*\overline{\mathbf{U}}^{*\top}\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}\overline{\mathbf{V}}^*\overline{\mathbf{V}}^{*\top}, \mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}\rangle\\ &= \|\overline{\mathbf{U}}^{*\top}\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}\|_F^2 + \|\overline{\mathbf{V}}^{*\top}\mathbf{Y}_{j,*}^\top\mathbf{X}_{i,*}\|_F^2 - \|\overline{\mathbf{U}}^{*\top}\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}\overline{\mathbf{V}}^*\|_F^2\\ &\le \|\mathbf{X}_{i,*}\overline{\mathbf{U}}^*\|_2^2\cdot\|\mathbf{Y}_{j,*}\|_2^2 + \|\mathbf{Y}_{j,*}\overline{\mathbf{V}}^*\|_2\cdot\|\mathbf{X}_{i,*}\|_2^2\\ &\le \|\mathbf{X}\overline{\mathbf{U}}^*\|_{2,\infty}^2\cdot\|\mathbf{Y}\|_{2,\infty}^2 + \|\mathbf{Y}\overline{\mathbf{V}}^*\|_{2,\infty}\cdot\|\mathbf{X}\|_{2,\infty}^2 \le \frac{\mu_0\mu_1 r(n_1+n_2)}{d_1d_2},\end{aligned}\tag{D.3}$$

where the first equality holds because $\|\mathcal{P}_\mathcal{T}(\mathbf{A})\|_F^2 = \langle\mathcal{P}_\mathcal{T}(\mathbf{A}),\mathbf{A}\rangle$, the first inequality holds because for any vectors $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$, $\|\mathbf{x}\mathbf{y}^\top\|_F\le\|\mathbf{x}\|_2\cdot\|\mathbf{y}\|_2$, and the last inequality holds because both $\mathbf{X},\mathbf{Y}$ are μ_1 self-incoherent. To apply matrix bernstein inequality, we need to bound $\|\mathbf{S}_{ij}\|_2$ and $\big\|\sum_{i,j}\mathbb{E}(\mathbf{S}_{ij}^2)\big\|_2$, respectively. To begin with, for any $(i,j)\in[d_1]\times[d_2]$, according to definition of $\mathbf{S}_{ij}$, we have

$$\|\mathbf{S}_{ij}\|_2 \le \frac{1}{p}\|\mathrm{vec}(\mathcal{P}_\mathcal{T}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}))\|_2^2 = \frac{1}{p}\|\mathcal{P}_\mathcal{T}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*})\|_F^2 \le \frac{\mu_0\mu_1 r(n_1+n_2)}{pd_1d_2},\tag{D.4}$$

where the first inequality holds because $|\langle \mathbf{A}, \mathbf{B}\rangle| \leq \|\mathbf{A}\|_F \cdot \|\mathbf{B}\|_F$, and the second inequality follows from (D.3). Similarly, for any $\mathbf{Z} \in \mathbb{R}^{n_1\times n_2}$, we have

$$
\begin{aligned}
\|\sum_{i,j}\mathbb{E}[\mathbf{S}_{ij}^2]\|_2 &= \frac{1-p}{p}\bigg\|\sum_{i,j}\text{vec}(\mathcal{P}_{\mathcal{T}}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}))\cdot\|\mathcal{P}_{\mathcal{T}}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*})\|_F^2\cdot\text{vec}(\mathcal{P}_{\mathcal{T}}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}))^\top\bigg\|_2\\
&\leq \frac{\mu_0\mu_1 r(n_1+n_2)}{pd_1d_2}\cdot\sup_{\|\mathbf{Z}\|_F=1}\bigg\|\sum_{i,j}\text{vec}(\mathcal{P}_{\mathcal{T}}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}))\text{vec}(\mathcal{P}_{\mathcal{T}}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}))^\top\text{vec}(\mathbf{Z})\bigg\|_2\\
&= \frac{\mu_0\mu_1 r(n_1+n_2)}{pd_1d_2}\cdot\sup_{\|\mathbf{Z}\|_F=1}\bigg\|\mathcal{P}_{\mathcal{T}}\bigg(\sum_{i,j}\langle\mathcal{P}_{\mathcal{T}}(\mathbf{Z}),\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}\rangle\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}\bigg)\bigg\|_F\\
&\leq \frac{\mu_0\mu_1 r(n_1+n_2)}{pd_1d_2}\cdot\sup_{\|\mathbf{Z}\|_F=1}\big\|\mathcal{P}_{\mathcal{T}}\big(\mathbf{X}^\top\mathbf{X}\mathcal{P}_{\mathcal{T}}(\mathbf{Z})\mathbf{Y}^\top\mathbf{Y}\big)\big\|_F \leq \frac{\mu_0\mu_1 r(n_1+n_2)}{pd_1d_2},
\end{aligned}
$$

where the first equality follows from the definition of $\mathbf{S}_{ij}$, the first inequality follows from (D.3) and the definition of spectral norm, the second equality holds because $\langle\mathcal{P}_{\mathcal{T}}(\mathbf{A}),\mathbf{B}\rangle = \langle\mathbf{A},\mathcal{P}_{\mathcal{T}}(\mathbf{B})\rangle$, and the second inequality holds because $\mathbf{X}$, $\mathbf{Y}$ are orthonormal and the projection operator $\mathcal{P}_{\mathcal{T}}$ is non-expansive. Thus, we obtain

$$
\big\|\mathbb{E}(\mathbf{S}_{ij}^2)\big\|_2 \leq \frac{\mu_0\mu_1 r(n_1+n_2)}{pd_1d_2}. \tag{D.5}
$$

Therefore, combining (D.4) and (D.5), according to Lemma E.1, for any $\gamma\in(0,1)$, we have

$$
\mathbb{P}\bigg\{\bigg\|\sum_{i,j}\mathbf{S}_{ij}\bigg\|_2 \geq \gamma\bigg\} \leq (n_1+n_2)\cdot\exp\bigg(\frac{-\gamma^2/2}{(1+\gamma/3)\mu_0\mu_1 r(n_1+n_2)/(pd_1d_2)}\bigg) \leq c'/d,
$$

under condition $p \geq c\mu_0\mu_1 rn\log(d)/(\gamma^2 d_1d_2)$, where c is a constant. Note that for all $\mathbf{Z}\in\mathbb{R}^{d_1\times d_2}$, we have

$$
\bigg\|\sum_{i,j}\mathcal{S}_{ij}(\mathbf{Z})\bigg\|_F = \bigg\|\sum_{i,j}\mathbf{S}_{ij}\cdot\text{vec}(\mathbf{Z})\bigg\|_F \leq \bigg\|\sum_{i,j}\mathbf{S}_{ij}\bigg\|_2\cdot\|\mathbf{Z}\|_F,
$$

we complete the proof of (C.2). Furthermore, for all $\mathbf{Z}_1,\mathbf{Z}_2\in\mathcal{T}$, we have

$$
\begin{aligned}
\big|\big\langle\mathbf{X}\mathbf{Z}_1\mathbf{Y}^\top - p^{-1}\mathcal{P}_\Omega\big(\mathbf{X}\mathbf{Z}_1\mathbf{Y}^\top\big),\mathbf{X}\mathbf{Z}_2\mathbf{Y}^\top\big\rangle\big| &= \Big|\Big\langle\mathcal{P}_{\mathcal{T}}(\mathbf{Z}_1) - p^{-1}\mathcal{P}_{\mathcal{T}}\Big(\mathbf{X}^\top\mathcal{P}_{\mathcal{T}}\big(\mathbf{X}\mathcal{P}_{\mathcal{T}}(\mathbf{Z}_1)\mathbf{Y}^\top\big)\mathbf{Y}^\top\Big),\mathbf{Z}_2\Big\rangle\Big|\\
&\leq \Big\|\mathcal{P}_{\mathcal{T}}(\mathbf{Z}_1) - p^{-1}\mathcal{P}_{\mathcal{T}}\Big(\mathbf{X}^\top\mathcal{P}_{\mathcal{T}}\big(\mathbf{X}\mathcal{P}_{\mathcal{T}}(\mathbf{Z}_1)\mathbf{Y}^\top\big)\mathbf{Y}^\top\Big)\Big\|_F\cdot\|\mathbf{Z}_2\|_F\\
&\leq \gamma\|\mathbf{Z}_1\|_F\cdot\|\mathbf{Z}_2\|_F,
\end{aligned}
$$

where the first equality holds because $\langle\mathbf{A},\mathcal{P}_{\mathcal{T}}(\mathbf{B})\rangle = \langle\mathcal{P}_{\mathcal{T}}(\mathbf{A}),\mathbf{B}\rangle$, the second inequality holds because $\langle\mathbf{A},\mathbf{B}\rangle \leq \|\mathbf{A}\|_F\cdot\|\mathbf{B}\|_F$, and the last inequality follows from (C.2). Finally, for all $\mathbf{Z}\in\mathcal{T}$, we have

$$
p^{-1}\big\|\mathcal{P}_\Omega\big(\mathbf{X}\mathbf{Z}\mathbf{Y}^\top\big)\big\|_F^2 \leq \big|\big\langle\mathbf{Z} - p^{-1}\mathbf{X}^\top\mathcal{P}_\Omega\big(\mathbf{X}\mathbf{Z}\mathbf{Y}^\top\big)\mathbf{Y},\mathbf{Z}\big\rangle\big| + \|\mathbf{Z}\|_F^2 \leq (1+\gamma)\|\mathbf{Z}\|_F^2,
$$

which complete the proof. □

D.3. Proof of Lemma C.3

Proof. Note that we have

$$
\begin{aligned}
\frac{1}{p}\big\|\mathcal{P}_\Omega(\mathbf{X}_L\mathbf{U}\mathbf{V}^\top\mathbf{X}_R^\top)\big\|_F^2 &= \|\mathbf{X}_L\mathbf{U}\mathbf{V}^\top\mathbf{X}_R^\top\|_F^2 + \langle p^{-1}\mathcal{P}_\Omega(\mathbf{X}_L\mathbf{U}\mathbf{V}^\top\mathbf{X}_R^\top) - \mathbf{X}_L\mathbf{U}\mathbf{V}\mathbf{X}_R^\top,\mathbf{X}_L\mathbf{U}\mathbf{V}^\top\mathbf{X}_R^\top\rangle\\
&= \|\mathbf{U}\mathbf{V}^\top\|_F^2 + \langle(p^{-1}\mathcal{P}_\Omega - \mathcal{I})(\mathbf{X}_L\mathbf{U}\mathbf{V}^\top\mathbf{X}_R^\top),\mathbf{X}_L\mathbf{U}\mathbf{V}^\top\mathbf{X}_R^\top\rangle\\
&= \|\mathbf{U}\mathbf{V}^\top\|_F^2 + \sum_{(i,j)=(1,1)}^{(d_1,d_2)}\bigg(\frac{\xi_{ij}}{p}-1\bigg)\cdot\big[\mathbf{X}_L\mathbf{U}\mathbf{V}^\top\mathbf{X}_R^\top\big]_{ij}^2,
\end{aligned}
$$

where $\xi_{ij}=1$, if $(i,j)\in\Omega$; $\xi_{ij}=0$, otherwise.Thus, it is sufficient to bound the second term on the right hand side. For simplicity, we let $\alpha_{ij}=(p^{-1}\xi_{ij}-1)\cdot[\mathbf{X}_L\mathbf{U}\mathbf{V}^\top\mathbf{X}_R^\top]_{ij}^2$. Note that $\mathbf{U},\mathbf{V}$ are fixed, we have $\mathbb{E}[\alpha_{ij}]=0$. In addition, we can upper bound $|\alpha_{ij}|$ by

$$|\alpha_{ij}|\leq\frac{1}{p}\|\mathbf{X}_L\mathbf{U}\mathbf{V}^\top\mathbf{X}_R^\top\|_{\infty,\infty}^2\leq\frac{1}{p}\max\big\{\|\mathbf{X}_L\|_{2,\infty}^2\cdot\|\mathbf{U}\|_F^2\cdot\|\mathbf{X}_R\mathbf{V}\|_{2,\infty}^2,\|\mathbf{X}_L\mathbf{U}\|_{2,\infty}^2\cdot\|\mathbf{V}\|_F^2\cdot\|\mathbf{X}_R\|_{2,\infty}^2\big\}$$
$$\leq\frac{9\mu_0\mu_1r\sigma_1^*(n_1+n_2)}{pd_1d_2}\cdot(\|\mathbf{U}\|_F^2+\|\mathbf{V}\|_F^2),$$

where the second inequality follows from Assumption 3.2. Next, we are going to bound the variance

$$\begin{aligned}\mathrm{Var}\Big(\sum_{(i,j)=(1,1)}^{(d_1,d_2)}\alpha_{ij}\Big)&=\sum_{(i,j)=(1,1)}^{(i,j)}\mathrm{Var}(\alpha_{ij})\\&\leq\frac{1}{p}\sum_{(i,j)=(1,1)}^{(d_1,d_2)}\big[\mathbf{X}_L\mathbf{U}\mathbf{V}^\top\mathbf{X}_R^\top\big]_{ij}^4\\&\leq\frac{1}{p}\|\mathbf{X}_L\mathbf{U}\mathbf{V}^\top\mathbf{X}_R^\top\|_{\infty,\infty}^2\cdot\|\mathbf{X}_L\mathbf{U}\mathbf{V}^\top\mathbf{X}_R^\top\|_F^2\leq\frac{9\mu_0^2r^2\sigma_1^{*2}}{pd_1d_2}(\|\mathbf{U}\|_F^2+\|\mathbf{V}\|_F^2)^2,\end{aligned}$$

where equality holds because α_{ij}'s are independent, and the last inequality follows from the assumptions $\|\mathbf{X}_L\mathbf{U}\|_{2,\infty}\leq 3\sqrt{\mu_0r\sigma_1^*/d_1}$, $\|\mathbf{X}_R\mathbf{V}\|_{2,\infty}\leq 3\sqrt{\mu_0r\sigma_1^*/d_2}$. Therefore, applying bernstein inequality for random variables, under condition that $|\Omega|\geq c\max\{\mu_0^2r^2\kappa^2,\mu_0\mu_1r\kappa n\}\log d/\gamma^2$, with probability at least $1-c'/d$, we have

$$\Big|\sum_{(i,j)=(1,1)}^{(d_1,d_2)}\alpha_{ij}\Big|\leq\gamma\sigma_r^*\cdot(\|\mathbf{U}\|_F^2+\|\mathbf{V}\|_F^2),$$

which completes the proof. □

D.4. Proof of Lemma C.4

Proof. For any $i\in[d_1]$, we denote $\mathbf{X}_{i,*}\in\mathbb{R}^{1\times n_1}$ as the i-th row vector of $\mathbf{X}$. Similarly, for any $j\in[d_2]$, we denote $\mathbf{Y}_{j,*}\in\mathbb{R}^{1\times n_2}$ as the j-th row vector of $\mathbf{Y}$. Besides, for any $(i,j)\in[d_1]\times[d_2]$, let $\xi_{ij}=1$, if $(i,j)\in\Omega$, and zero otherwise. Note that $\mathbf{X}$, $\mathbf{Y}$ are orthonormal, thus for any $\mathbf{Z}\in\mathbb{R}^{n_1\times n_2}$, we have

$$\begin{aligned}\mathbf{Z}-p^{-1}\mathbf{X}^\top\mathcal{P}_\Omega\big(\mathbf{X}\mathbf{Z}\mathbf{Y}^\top\big)\mathbf{Y}&=\mathbf{X}^\top(\mathcal{I}-p^{-1}\mathcal{P}_\Omega)(\mathbf{X}\mathbf{Z}\mathbf{Y}^\top)\mathbf{Y}\\&=\sum_{(i,j)\in[d_1]\times[d_2]}(1-p^{-1}\xi_{ij})\langle\mathbf{X}\mathbf{Z}\mathbf{Y}^\top,\mathbf{e}_i\mathbf{e}_j^\top\rangle\mathbf{X}^\top\mathbf{e}_i\mathbf{e}_j^\top\mathbf{Y}\\&=\sum_{(i,j)\in[d_1]\times[d_2]}(1-p^{-1}\xi_{ij})\langle\mathbf{Z},\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}\rangle\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*},\end{aligned}$$

where $\mathcal{I}:\mathbb{R}^{d_1\times d_2}\to\mathbb{R}^{d_1\times d_2}$ is the identity mapping, and $\mathbf{e}_i$ denotes the i-th standard basis. For any $(i,j)\in[d_1]\times[d_2]$, define linear operator $\mathcal{S}_{ij}:\mathbb{R}^{n_1\times n_2}\to\mathbb{R}^{n_1\times n_2}$, such that

$$\mathcal{S}_{ij}(\mathbf{Z})=(1-p^{-1}\xi_{ij})\langle\mathbf{Z},\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}\rangle\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}.$$

Define $\mathbf{S}_{ij}\in\mathbb{R}^{d_1d_2\times d_1d_2}$ as the corresponding matrix to the linear operator $\mathcal{S}_{ij}$ such that

$$\mathbf{S}_{ij}=(1-p^{-1}\xi_{ij})\mathrm{vec}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*})\mathrm{vec}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*})^\top,\tag{D.6}$$

then we can easily show that $\|\sum_{i,j}\mathcal{S}_{ij}(\mathbf{Z})\|_F^2=\|\sum_{i,j}\mathbf{S}_{ij}\cdot\mathrm{vec}(\mathbf{Z})\|_2^2$, for any $\mathbf{Z}\in\mathbb{R}^{d_1\times d_2}$. Obviously, we have $\mathbb{E}[\mathbf{S}_{ij}]=\mathbf{0}$ and $\mathbf{S}_{ij}$ is symmetric. For any $(i,j)\in[d_1]\times[d_2]$, according to the definition of $\mathbf{S}_{ij}$ in (D.6), we have

$$\|\mathbf{S}_{ij}\|_2\leq\frac{1}{p}\|\mathrm{vec}(\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*})\|_2^2=\frac{1}{p}\|\mathbf{X}_{i,*}^\top\mathbf{Y}_{j,*}\|_F^2\leq\frac{1}{p}\|\mathbf{X}\|_{2,\infty}^2\cdot\|\mathbf{Y}\|_{2,\infty}^2\leq\frac{\mu_1^2n_1n_2}{pd_1d_2},\tag{D.7}$$

where the second inequality holds because $\|\mathbf{X}_{i,*}^\top \mathbf{Y}_{j,*}\|_F = \|\mathbf{X}_{i,*}\|_2 \cdot \|\mathbf{Y}_{j,*}\|_2$, and the last inequality follows from the fact that $\mathbf{X}$,$\mathbf{Y}$ are μ_1 self-incoherent. Moreover, for any $\mathbf{Z} \in \mathbb{R}^{n_1 \times n_2}$, we have

$$
\begin{aligned}
\Big\| \sum_{i,j} \mathbb{E}[\mathbf{S}_{ij}^2] \Big\|_2 &= \frac{1-p}{p} \bigg\| \sum_{i,j} \mathrm{vec}(\mathbf{X}_{i,*}^\top \mathbf{Y}_{j,*}) \cdot \|\mathbf{X}_{i,*}^\top \mathbf{Y}_{j,*}\|_F^2 \cdot \mathrm{vec}(\mathbf{X}_{i,*}^\top \mathbf{Y}_{j,*})^\top \bigg\|_2 \\
&\le \frac{\mu_1^2 n_1 n_2}{p d_1 d_2} \cdot \sup_{\|\mathbf{Z}\|_F = 1} \bigg\| \sum_{i,j} \mathrm{vec}(\mathbf{X}_{i,*}^\top \mathbf{Y}_{j,*}) \mathrm{vec}(\mathbf{X}_{i,*}^\top \mathbf{Y}_{j,*})^\top \mathrm{vec}(\mathbf{Z}) \bigg\|_2 \\
&= \frac{\mu_1^2 n_1 n_2}{p d_1 d_2} \cdot \sup_{\|\mathbf{Z}\|_F = 1} \Big\| \sum_{i,j} \langle \mathbf{Z}, \mathbf{X}_{i,*}^\top \mathbf{Y}_{j,*} \rangle \mathbf{X}_{i,*}^\top \mathbf{Y}_{j,*} \Big\|_F \\
&= \frac{\mu_1^2 n_1 n_2}{p d_1 d_2} \cdot \sup_{\|\mathbf{Z}\|_F = 1} \big\| \mathbf{X}^\top \mathbf{X} \mathbf{Z} \mathbf{Y}^\top \mathbf{Y} \big\|_F \le \frac{\mu_1^2 n_1 n_2}{p d_1 d_2},
\end{aligned}
$$

where the first equality follows from the definition of $\mathbf{S}_{ij}$ in (D.6), the first inequality follows from the fact that $\|\mathbf{X}_{i,*}^\top \mathbf{Y}_{j,*}\|_F^2 \le \mu^2 n_1 n_2/(d_1 d_2)$, and the last equality holds because $\mathbf{X}$, $\mathbf{Y}$ are both orthonormal. Therefore, according to matrix bernstein inequality as in Lemma E.1, for any $\gamma \in (0,1)$, we have

$$
\mathbb{P}\bigg\{ \bigg\| \sum_{i,j} \mathbf{S}_{ij} \bigg\|_2 \ge \mu_1 n \gamma \bigg\} \le 2 n_1 n_2 \cdot \exp\bigg(\frac{-\gamma^2 \mu_1^2 n^2/2}{(1+\gamma\mu_1 n/3)\mu_1^2 n_1 n_2/(p d_1 d_2)} \bigg) \le c'/d,
$$

under condition $|\Omega| \ge c\mu_1 n \log d/\gamma^2$, where c is a constant. Thus according to the definition of $\mathcal{S}_{ij}$, for all $\mathbf{Z} \in \mathbb{R}^{n_1 \times n_2}$, with probability at least $1 - c'/d$, we have

$$
\|\mathbf{Z} - p^{-1}\mathbf{X}^\top \mathcal{P}_\Omega(\mathbf{X}\mathbf{Z}\mathbf{Y}^\top)\mathbf{Y}\|_F = \bigg\| \sum_{i,j} \mathcal{S}_{ij}(\mathbf{Z}) \bigg\|_F = \bigg\| \sum_{i,j} \mathbf{S}_{ij} \cdot \mathrm{vec}(\mathbf{Z}) \bigg\|_2 \le \bigg\| \sum_{i,j} \mathbf{S}_{ij} \bigg\|_2 \cdot \|\mathbf{Z}\|_F \le \mu_1 n \gamma \|\mathbf{Z}\|_F.
$$

By triangle's inequality, we complete the proof. □

E. Additional Auxiliary Lemmas

In this section,we provide the bernstein inequalities used in the proofs for our main results .

Lemma E.1. (Tropp, 2012) Consider a finite sequence $\{\mathbf{Z}_k\}$ of independent random matrices with dimension $d_1 \times d_2$. Assume that each random matrix satisfies

$$
\mathbb{E}(\mathbf{Z}_k) = \mathbf{0} \quad \text{and} \quad \|\mathbf{Z}_k\|_2 \le R \quad \text{almost surely.}
$$

Define

$$
\sigma^2 = \max\bigg\{ \bigg\| \sum_k \mathbb{E}(\mathbf{Z}_k \mathbf{Z}_k^\top) \bigg\|_2, \bigg\| \sum_k \mathbb{E}(\mathbf{Z}_k^\top \mathbf{Z}_k) \bigg\|_2 \bigg\}.
$$

Then, for all $t \ge 0$, we have

$$
\mathbb{P}\bigg\{ \bigg\| \sum_k \mathbf{Z}_k \bigg\|_2 \ge t \bigg\} \le (d_1 + d_2) \cdot \exp\bigg(\frac{-t^2/2}{\sigma^2 + Rt/3} \bigg).
$$

Lemma E.2. Assume the index set Ω follows Bernoulli model (3.1). There exists constants c, c' such that under condition that $|\Omega| \ge c \log d$, with probability at least $1 - c'/d$, we have

$$
\big| |\Omega| - p d_1 d_2 \big| \le \frac{1}{2} p d_1 d_2,
$$

where $p = |\Omega|/(d_1 d_2)$.

Lemma E.2 can be directly derived from the bernstein-type inequality for independent random variables.