Supplementary Materials

1 Preliminary lemmas

The following lemma gives some fundamental results for $\sin\Theta(U,V)$, which can be easily verified via definition.

Lemma 1. *Let* $[U,\, U_{\rm c}]$ *and* $[V,\, V_{\rm c}]$ *be two orthogonal matrices with* $U, V \in \mathbb{R}^{n\times k}$*. Then*

$$\|\sin\Theta(U,V)\|_{\rm ui} = \|U_{\rm c}^{\rm T} V\|_{\rm ui} = \|U^{\rm T} V_{\rm c}\|_{\rm ui}.$$

Here $\|\cdot\|_{\rm ui}$ *denotes any unitarily invariant norm, including the spectral norm and Frobenius norm. In particular, for the spectral norm, it holds* $\|\sin\Theta(U,V)\| = \|UU^{\rm T} - VV^{\rm T}\|$*; for the Frobenius norm, it holds* $\|\sin\Theta(U,V)\|_F = \frac{1}{\sqrt{2}}\|UU^{\rm T} - VV^{\rm T}\|_F$*.*

The following lemma is the well-known Weyl theorem, which gives the perturbation bound for eigenvalues of Hermitian matrix.

Lemma 2. *(Stewart and Sun, 1990, p.203) For two Hermitian matrices* $A, \widetilde{A} \in \mathbb{C}^{n\times n}$*, let* $\lambda_1 \le \cdots \le \lambda_n$*,* $\tilde\lambda_1 \le \cdots \le \tilde\lambda_n$ *be eigenvalues of* A*,* $\widetilde{A}$*, respectively. Then*

$$|\lambda_j - \tilde\lambda_j| \le \|A - \widetilde{A}\|, \quad \text{for } 1 \le j \le n.$$

The following lemma is used to establish the perturbation bound for the invariant subspace of a Hermitian matrix, which is due to Davis and Kahan.

Lemma 3. *(Davis and Kahan, 1970, Theorem 5.1) Let* H *and* M *be two Hermitian matrices, and let* S *be a matrix of a compatible size as determined by the Sylvester equation*

$$HY - YM = S.$$

If either all eigenvalues of H *are contained in a closed interval that contains no eigenvalue of* M *or vice versa, then the Sylvester equation has a unique solution* Y*, and moreover*

$$\|Y\|_{\rm ui} \le \frac{1}{\delta}\|S\|_{\rm ui},$$

where $\delta = \min|\lambda - \omega|$ *over all eigenvalues* ω *of* M *and all eigenvalues* λ *of* H*.*

For a rectangular matrix $A \in \mathbb{R}^{m\times n}$ (without loss of generality, assume $m \ge n$), let the SVD of A be $A = U\Sigma V^{\rm T}$, where $U = [U_1\,|\,U_2\,|\,U_3] = [u_1,\ldots,u_k\,|\,u_{k+1},\ldots,u_r\,|\,u_{r+1},\ldots,u_m] \in \mathbb{R}^{m\times m}$, $V = [V_1\,|\,V_2\,|\,V_3] = [v_1,\ldots,v_k\,|\,v_{k+1},\ldots,v_r\,|\,v_{r+1},\ldots,v_n] \in \mathbb{R}^{n\times n}$ are orthogonal matrices, and $\Sigma = \begin{bmatrix} \operatorname{diag}(\Sigma_1,\Sigma_2) & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}$, $\Sigma_1 = \operatorname{diag}(\sigma_1,\ldots,\sigma_k)$, $\Sigma_2 = \operatorname{diag}(\sigma_{k+1},\ldots,\sigma_r)$ with $\sigma_1 \ge \cdots \ge \sigma_r > 0$, $k \le r = \operatorname{rank}(A)$. Then the spectral decomposition of $\begin{bmatrix} 0 & A \\ A^{\rm T} & 0 \end{bmatrix}$ can be given by

$$\begin{bmatrix} 0 & A \\ A^{\rm T} & 0 \end{bmatrix} = X \operatorname{diag}(\Sigma_1, \Sigma_2, -\Sigma_1, -\Sigma_2, 0_{n-r}, 0_{m-r}) X^{\rm T}, \tag{1}$$

where $X = \frac{1}{\sqrt{2}}\begin{bmatrix} U & -U & 0 & \sqrt{2}U_3 \\ V & V & \sqrt{2}V_3 & 0 \end{bmatrix}$ is an orthogonal matrix.

With the help of (1) and Lemmas 2 and 3, we are able to prove Lemma 4, which established an error bound for singular vectors.

Lemma 4. *Given* $A \in \mathbb{R}^{m\times n}$ $(m \ge n)$*, let the SVD of* A *be given as above. Let* $\hat\sigma_j$*,* $\hat u_j$*,* $\hat v_j$ *be respectively the approximate singular values, right and left singular vectors of* A *satisfying that* $\widehat{U} = [\hat u_1,\ldots,\hat u_k] \in \mathbb{R}^{m\times k}$ *and* $\widehat{V} = [\hat v_1,\ldots,\hat v_k] \in \mathbb{R}^{n\times k}$ *are both orthonormal ,* $\widehat{\Sigma} = \widehat{U}^{\rm T} A \widehat{V} = \operatorname{diag}(\hat\sigma_1,\ldots,\hat\sigma_k)$ *with* $\hat\sigma_1 \ge \cdots \ge \hat\sigma_k > 0$*. Let*

$$E = A\widehat{V} - \widehat{U}\widehat{\Sigma}, \qquad F = A^{\rm T}\widehat{U} - \widehat{V}\widehat{\Sigma}. \tag{2}$$

If

$$\|(I_m - \widehat{U}\widehat{U}^{\mathrm{T}})A(I_n - \widehat{V}\widehat{V}^{\mathrm{T}})\| < \hat{\sigma}_k, \qquad \max\{\|E\|, \|F\|\} < \sigma_k - \sigma_{k+1},$$

then

$$\max\{\Theta_u, \Theta_v\} \leq \eta, \qquad \frac{\|U_1\Sigma_1 V_1^{\mathrm{T}} - \widehat{U}\widehat{\Sigma}\widehat{V}^{\mathrm{T}}\|_{\max}}{\|A\|} \leq (\|U_1\|_{2,\infty}\Theta_v + \|V_1\|_{2,\infty}\Theta_u) + (1 + 3\|U_1\|_{2,\infty}\|V_1\|_{2,\infty})\Theta_u\Theta_v,$$

where $\Theta_u = \|\sin\Theta(U_1, \widehat{U})\|$, $\Theta_v = \|\sin\Theta(V_1, \widehat{V})\|$, $\eta = \frac{\max\{\|E\|,\|F\|\}}{\sigma_k - \sigma_{k+1} - \max\{\|E\|,\|F\|\}}$.

Proof. Let

$$H = \begin{bmatrix} 0 & A \\ A^{\mathrm{T}} & 0 \end{bmatrix}, \qquad X_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} U_1 & -U_1 \\ V_1 & V_1 \end{bmatrix},$$

$$\widehat{X}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} \widehat{U} & -\widehat{U} \\ \widehat{V} & \widehat{V} \end{bmatrix}, \qquad X_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} U_2 & -U_2 & 0 & \sqrt{2}U_3 \\ V_2 & V_2 & \sqrt{2}V_3 & 0 \end{bmatrix}.$$

By calculations, we have

$$\|X_1X_1^{\mathrm{T}} - \widehat{X}_1\widehat{X}_1^{\mathrm{T}}\|_{\mathrm{ui}} = \|\operatorname{diag}(U_1U_1^{\mathrm{T}} - \widehat{U}\widehat{U}^{\mathrm{T}}, V_1V_1^{\mathrm{T}} - \widehat{V}\widehat{V}^{\mathrm{T}})\|_{\mathrm{ui}} \tag{3}$$

By simple calculations, we have

$$H\widehat{X}_1 - \widehat{X}_1\operatorname{diag}(\widehat{\Sigma}, -\widehat{\Sigma}) = \frac{1}{\sqrt{2}}\begin{bmatrix} A\widehat{V} - \widehat{U}\widehat{\Sigma} & A\widehat{V} - \widehat{U}\widehat{\Sigma} \\ A^{\mathrm{T}}\widehat{U} - \widehat{V}\widehat{\Sigma} & -A^{\mathrm{T}}\widehat{U} + \widehat{V}\widehat{\Sigma} \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} E & E \\ F & -F \end{bmatrix} \triangleq R, \tag{4a}$$

$$HX_2 - X_2\operatorname{diag}(\Sigma_2, -\Sigma_2, 0, 0) = 0, \tag{4b}$$

where (4a) uses (2), (4b) uses the SVD of A. Then it follows from (4a) that

$$\|R\| = \left\|\operatorname{diag}(E, F)\frac{1}{\sqrt{2}}\begin{bmatrix} I_k & I_k \\ I_k & -I_k \end{bmatrix}\right\| = \|\operatorname{diag}(E, F)\| = \max\{\|E\|, \|F\|\}. \tag{5}$$

Pre-multiplying (4a) by X_2^{T} and using (4b), we have

$$X_2^{\mathrm{T}}R = X_2^{\mathrm{T}}H\widehat{X}_1 - X_2^{\mathrm{T}}\widehat{X}_1\operatorname{diag}(\widehat{\Sigma}, -\widehat{\Sigma}) = \operatorname{diag}(\Sigma_2, -\Sigma_2, 0, 0)X_2^{\mathrm{T}}\widehat{X}_1 - X_2^{\mathrm{T}}\widehat{X}_1\operatorname{diag}(\widehat{\Sigma}, -\widehat{\Sigma}). \tag{6}$$

To apply Lemma 3 to (6), we need to estimate the gap between the eigenvalues of $\operatorname{diag}(\widehat{\Sigma}, -\widehat{\Sigma})$ and those of $\operatorname{diag}(\Sigma_2, -\Sigma_2, 0, 0)$. Using (4a) and $\widehat{U}^{\mathrm{T}}A\widehat{V} = \widehat{\Sigma}$, we have

$$(H - R\widehat{X}_1^{\mathrm{T}} - \widehat{X}_1R^{\mathrm{T}})\widehat{X}_1 = H\widehat{X}_1 - R = \widehat{X}_1\operatorname{diag}(\widehat{\Sigma}, -\widehat{\Sigma}), \tag{7}$$

which implies that $\pm\hat{\sigma}_j$ are eigenvalues of $H - R\widehat{X}_1^{\mathrm{T}} - \widehat{X}_1R^{\mathrm{T}}$, and the corresponding eigenvectors are $\frac{1}{\sqrt{2}}\begin{bmatrix} \pm\hat{u}_j \\ \hat{v}_j \end{bmatrix}$, for $j = 1, \ldots, k$. Next, we declare that $\hat{\sigma}_1, \ldots, \hat{\sigma}_k$ are the k largest eigenvalues of $H - R\widehat{X}_1^{\mathrm{T}} - \widehat{X}_1R^{\mathrm{T}}$. This is because

$$\begin{aligned}
&\max_{\widehat{X}_1^{\mathrm{T}}x = 0} \frac{x^{\mathrm{T}}(H - R\widehat{X}_1^{\mathrm{T}} - \widehat{X}_1R^{\mathrm{T}})x}{x^{\mathrm{T}}x} \\
\leq &\|(I - \widehat{X}_1\widehat{X}_1^{\mathrm{T}})(H - R\widehat{X}_1^{\mathrm{T}} - \widehat{X}_1R^{\mathrm{T}})(I - \widehat{X}_1\widehat{X}_1^{\mathrm{T}})\| \\
= &\|(I - \widehat{X}_1\widehat{X}_1^{\mathrm{T}})H(I - \widehat{X}_1\widehat{X}_1^{\mathrm{T}})\| \\
= &\left\|\begin{bmatrix} I_m - \widehat{U}\widehat{U}^{\mathrm{T}} & 0 \\ 0 & I_n - \widehat{V}\widehat{V}^{\mathrm{T}} \end{bmatrix}\begin{bmatrix} 0 & A \\ A^{\mathrm{T}} & 0 \end{bmatrix}\begin{bmatrix} I_m - \widehat{U}\widehat{U}^{\mathrm{T}} & 0 \\ 0 & I_n - \widehat{V}\widehat{V}^{\mathrm{T}} \end{bmatrix}\right\| \\
= &\|(I_m - \widehat{U}\widehat{U}^{\mathrm{T}})A(I_n - \widehat{V}\widehat{V}^{\mathrm{T}})\| < \hat{\sigma}_k.
\end{aligned}$$

Therefore, by Lemma 2, we have

$$|\sigma_j - \hat{\sigma}_j| \le \|R\widehat{X}_1^{\mathrm{T}} + \widehat{X}_1 R^{\mathrm{T}}\|, \quad \text{for } j = 1, \ldots, k. \tag{8}$$

Together with (5), we get

$$\begin{aligned} |\sigma_j - \hat{\sigma}_j| \le \|R\widehat{X}_1^{\mathrm{T}} + \widehat{X}_1 R^{\mathrm{T}}\| &= \max_j |\lambda_j([R, \widehat{X}_1] \left[\begin{smallmatrix} \widehat{X}_1^{\mathrm{T}} \\ R^{\mathrm{T}} \end{smallmatrix}\right])| = \max_j |\lambda_j(\left[\begin{smallmatrix} \widehat{X}_1^{\mathrm{T}} \\ R^{\mathrm{T}} \end{smallmatrix}\right] [R, \widehat{X}_1])| \\ &= \max_j |\lambda_j(\left[\begin{smallmatrix} 0 & I_k \\ R^{\mathrm{T}}R & 0 \end{smallmatrix}\right])| = \|R\| = \max\{\|E\|, \|F\|\}. \end{aligned} \tag{9}$$

Here we uses the property that for any two matrix $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times m}$, the nonzero eigenvalues of AB and BA are the same.

Now by the assumption that $\max\{\|E\|, \|F\|\} < \sigma_k - \sigma_{k+1}$, we have

$$\hat{\sigma}_k - \sigma_{k+1} = \sigma_k - \sigma_{k+1} + \hat{\sigma}_k - \sigma_k \ge \sigma_k - \sigma_{k+1} - \max\{\|E\|, \|F\|\} > 0, \tag{10}$$

therefore, the eigenvalues of $\mathrm{diag}(\Sigma_2, -\Sigma_2, 0, 0)$ lie in $[-\sigma_{k+1}, \sigma_{k+1}]$, which has no eigenvalues of $\mathrm{diag}(\widehat{\Sigma}, -\widehat{\Sigma})$. We are able to apply Lemma 3 to (6), which yields

$$\|X_2^{\mathrm{T}}\widehat{X}_1\|_{\mathrm{ui}} \le \frac{\|X_2^{\mathrm{T}}R\|_{\mathrm{ui}}}{\sigma_k - \sigma_{k+1} - \max\{\|E\|, \|F\|\}}. \tag{11}$$

Using (3), Lemma 1, (10) and (11), we get

$$\max\{\Theta_u, \Theta_v\} = \|\sin\Theta(X_1, \widehat{X}_1)\| = \|X_2^{\mathrm{T}}\widehat{X}_1\| \le \frac{\|X_2^{\mathrm{T}}R\|}{\sigma_k - \sigma_{k+1} - \max\{\|E\|, \|F\|\}} \le \eta. \tag{12}$$

Let

$$\widehat{U} = U\Gamma_u = [U_1, U_2, U_3] \begin{bmatrix} \Gamma_{u1} \\ \Gamma_{u2} \\ \Gamma_{u3} \end{bmatrix}, \quad \widehat{V} = V\Gamma_v = [V_1, V_2, V_3] \begin{bmatrix} \Gamma_{v1} \\ \Gamma_{v2} \\ \Gamma_{v3} \end{bmatrix}, \tag{13}$$

where $\Gamma_{u1} \in \mathbb{R}^{k\times k}$, $\Gamma_{u2} \in \mathbb{R}^{(r-k)\times k}$, $\Gamma_{u3} \in \mathbb{R}^{(m-r)\times k}$, $\Gamma_{v1} \in \mathbb{R}^{k\times k}$, $\Gamma_{v2} \in \mathbb{R}^{(r-k)\times k}$, $\Gamma_{v3} \in \mathbb{R}^{(n-r)\times k}$, and $\begin{bmatrix} \Gamma_{u1} \\ \Gamma_{u2} \\ \Gamma_{u3} \end{bmatrix}$, $\begin{bmatrix} \Gamma_{v1} \\ \Gamma_{v3} \\ \Gamma_{v3} \end{bmatrix}$ are both orthonormal. By (12), we have

$$\left\|\left[\begin{smallmatrix} \Gamma_{u2} \\ \Gamma_{u3} \end{smallmatrix}\right]\right\| = \Theta_u, \quad \sigma_{\min}(\Gamma_{u1}) = \sqrt{1 - \Theta_u^2}, \quad \left\|\left[\begin{smallmatrix} \Gamma_{u2} \\ \Gamma_{u3} \end{smallmatrix}\right]\right\| = \Theta_v, \quad \sigma_{\min}(\Gamma_{v1}) = \sqrt{1 - \Theta_v^2}. \tag{14}$$

Substituting (13) into $\widehat{U}^{\mathrm{T}}A\widehat{V} = \widehat{\Sigma}$ and using the SVD of A, we have

$$\widehat{\Sigma} = [\Gamma_{u1}^{\mathrm{T}}, \Gamma_{u2}^{\mathrm{T}}, \Gamma_{u3}^{\mathrm{T}}]\, \mathrm{diag}(\Sigma_1, \Sigma_2, 0_{(m-r)\times(n-r)}) \begin{bmatrix} \Gamma_{v1} \\ \Gamma_{v2} \\ \Gamma_{v3} \end{bmatrix} = \Gamma_{u1}^{\mathrm{T}}\Sigma_1\Gamma_{v1} + \Gamma_{u2}^{\mathrm{T}}\Sigma_2\Gamma_{v2}. \tag{15}$$

Then it follows that

$$\begin{aligned} \|\Sigma_1 - \Gamma_{u1}\widehat{\Sigma}\Gamma_{v1}^{\mathrm{T}}\| =& \|(\Sigma_1 - \Gamma_{u1}\Gamma_{u1}^{\mathrm{T}}\Sigma_1) + (\Gamma_{u1}\Gamma_{u1}^{\mathrm{T}}\Sigma_1 - \Gamma_{u1}\Gamma_{u1}^{\mathrm{T}}\Sigma_1\Gamma_{v1}\Gamma_{v1}^{\mathrm{T}}) - \Gamma_{u1}\Gamma_{u2}^{\mathrm{T}}\Sigma_2\Gamma_{v2}\Gamma_{v1}^{\mathrm{T}}\| \\ \le& \|I - \Gamma_{u1}\Gamma_{u1}^{\mathrm{T}}\|\|\Sigma_1\| + \|\Gamma_{u1}\Gamma_{u1}^{\mathrm{T}}\|\|I - \Gamma_{v1}\Gamma_{v1}^{\mathrm{T}}\|\|\Sigma_1\| + \|\Gamma_{u2}\|\|\Gamma_{v2}\|\|\Sigma_2\| \\ \le& (\Theta_u^2 + \Theta_v^2 + \Theta_u\Theta_v)\|\Sigma_1\|. \end{aligned} \tag{16}$$

Finally, using (14), (15), (16) and $\|\Gamma_{u1}\| \le 1$, $\|\Gamma_{v1}\| \le 1$, $\|\widehat{\Sigma}\| \le \|A\|$, we have

$$\begin{aligned}
\|U_1\Sigma_1V_1^{\mathrm{T}} - \widehat{U}\widehat{\Sigma}\widehat{V}^{\mathrm{T}}\|_{\max} &= \max_{i,j} |e_i^{\mathrm{T}}(U_1\Sigma_1V_1^{\mathrm{T}} - \widehat{U}\widehat{\Sigma}\widehat{V}^{\mathrm{T}})e_j| \\
&= \max_{i,j} |e_i^{\mathrm{T}}(U_1\Sigma_1V_1^{\mathrm{T}} - U\Gamma_u\widehat{\Sigma}\Gamma_v^{\mathrm{T}}V^{\mathrm{T}})e_j| \\
&\le \max_{i,j} |e_i^{\mathrm{T}}(U_1\Sigma_1V_1^{\mathrm{T}} - U_1\Gamma_{u1}\widehat{\Sigma}\Gamma_{v1}^{\mathrm{T}}V_1^{\mathrm{T}})e_j| + \|[U_2, U_3]\left[\begin{smallmatrix}\Gamma_{u2}\\ \Gamma_{u3}\end{smallmatrix}\right]\widehat{\Sigma}\left[\begin{smallmatrix}\Gamma_{v2}\\ \Gamma_{v3}\end{smallmatrix}\right]^{\mathrm{T}}[V_2, V_3]^{\mathrm{T}}\| \\
&+ \max_{i,j}\left(|e_i^{\mathrm{T}}[U_2, U_3]\left[\begin{smallmatrix}\Gamma_{u2}\\ \Gamma_{u3}\end{smallmatrix}\right]\widehat{\Sigma}\Gamma_{v1}^{\mathrm{T}}V_1^{\mathrm{T}}e_j| + |e_i^{\mathrm{T}}U_1\Gamma_{u1}\widehat{\Sigma}\left[\begin{smallmatrix}\Gamma_{v2}\\ \Gamma_{v3}\end{smallmatrix}\right]^{\mathrm{T}}[V_2, V_3]^{\mathrm{T}}e_j|\right) \\
&\le \max_{i,j}\left(3\|e_i^{\mathrm{T}}U_1\|\|e_j^{\mathrm{T}}V_1\|\|A\|\Theta_u\Theta_v + \|A\|\Theta_u\Theta_v + \|e_j^{\mathrm{T}}V_1\|\|A\|\Theta_u + \|e_i^{\mathrm{T}}U_1\|\|A\|\Theta_v\right) \\
&\le \|A\|\left((\|U_1\|_{2,\infty}\Theta_v + \|V_1\|_{2,\infty}\Theta_u) + (1 + 3\|U_1\|_{2,\infty}\|V_1\|_{2,\infty})\Theta_u\Theta_v\right),
\end{aligned}$$

completing the proof. □

Lemma 5. *(Tropp, 2015, Corollary 6.1.2) Let $\mathbf{S}_1, \dots, \mathbf{S}_n$ be independent random matrices with common dimension $d_1 \times d_2$, and assume that each matrix has uniformly bounded deviation from its mean:*

$$\|\mathbf{S_k} - \mathbb{E}(\mathbf{S}_k)\| \le L, \quad \textit{for each } k = 1, \dots, n.$$

Let $\mathbf{Z} = \sum_{k=1}^n \mathbf{S}_k$, $v(\mathbf{Z})$ denote the matrix covariance statistic of the sum:

$$\begin{aligned}
v(\mathbf{Z}) &= \max\{\|\mathbb{E}[(\mathbf{Z} - \mathbb{E}(Z))(\mathbf{Z} - \mathbb{E}(Z))^{\mathrm{H}}]\|, \|\mathbb{E}[(\mathbf{Z} - \mathbb{E}(Z))^{\mathrm{H}}(\mathbf{Z} - \mathbb{E}(Z))]\|\} \\
&= \max\{\|\mathbb{E}[\sum_{k=1}^n(\mathbf{S}_k - \mathbb{E}(\mathbf{S}_k))(\mathbf{S}_k - \mathbb{E}(\mathbf{S}_k))^{\mathrm{H}}]\|, \|\mathbb{E}[\sum_{k=1}^n(\mathbf{S}_k - \mathbb{E}(\mathbf{S}_k))^{\mathrm{H}}(\mathbf{S}_k - \mathbb{E}(\mathbf{S}_k))]\|\}.
\end{aligned}$$

Then for all $t \ge 0$,

$$\mathbb{P}\{\|\mathbf{Z} - \mathbb{E}(\mathbf{Z})\| \ge t\} \le (d_1 + d_2) \cdot \exp\left(\frac{-t^2/2}{v(\mathbf{Z}) + Lt/3}\right).$$

Lemma 6. *For any linear homogeneous function $F : \mathbb{R}^k \to \mathbb{R}^{m\times n}$, assume that the linear system of equations $F(x) = C$ either has a unique solution or has no solution at all. Then it holds*

$$\operatorname{argmin}_x \|F(x) - C\| = \operatorname{argmin}_x \|F(x) - C\|_F.$$

Proof. For any $A, B \in \mathbb{R}^{m\times n}$, define $\langle A, B\rangle = \operatorname{trace}(A^{\mathrm{T}}B)$. It is easy to see that $\langle \cdot, \cdot \rangle$ is an inner product over $\mathbb{R}^{m\times n}$. Denote the range space of $F(\cdot)$ by $\mathcal{F}$, and its orthogonal complement space by $\mathcal{F}^{\perp}$. Write $C = C_{\mathrm{LS}} + C$ such that $C_{\mathrm{LS}} \in \mathcal{F}$, and $C \in \mathcal{F}^{\perp}$. Then the solutions to $\min \|F(x) - C\|$ and $\min \|F(x) - C\|_F$ are nothing but the solutions to $F(x) = C_{\mathrm{LS}}$. Since $C_{\mathrm{LS}} \in \mathcal{F}$, $F(x) = C_{\mathrm{LS}}$ has at least a solution. By the assumption, the solution should be unique. The proof is completed. □

Lemma 7. *Let $L_* \in \mathbb{R}^{m\times n}$ with $m \ge n$, let the SVD of L_* be $L_* = U_*\Sigma_*V_*^{\mathrm{T}}$, where $U_* \in \mathbb{R}^{m\times r}$, $V_* \in \mathbb{R}^{n\times r}$ are orthonormal, $\Sigma_* = \operatorname{diag}(\sigma_{1*}, \dots, \sigma_{r*})$ with $\sigma_{1*} \ge \dots \ge \sigma_{r*} > 0$. Let $G \in \mathbb{R}^{m\times n}$ be a perturbation to L_*, $X \in \mathbb{R}^{m\times r}$, $Y \in \mathbb{R}^{n\times r}$ have full column rank. Denote $\theta_x = \|\sin\Theta(U_*, X)\|$, $\theta_y = \|\sin\Theta(V_1, Y)\|$. Then*

$$\min_{X,Y}\|L_* - G - XY^{\mathrm{T}}\| \ge \sigma_{r*}\max\{\sqrt{1-\theta_x^2}\theta_y, \sqrt{1-\theta_y^2}\theta_x\}\sqrt{1-\theta_x^2}\sqrt{1-\theta_y^2} - \|G\|.$$

Proof. Let $U_{*,c}$, $V_{*,c}$ be such that $U = [U_*, U_{*,c}]$, $V = [V_*, V_{*,c}]$ are orthogonal. Let $\widehat{X} = U_*C_x + U_{*,c}S_x$, $\widehat{Y} = V_*C_y + V_{*,c}S_y$, where the columns of $\widehat{X}$, $\widehat{Y}$ form the orthonormal basis for $\mathcal{R}(X)$ and $\mathcal{R}(Y)$, respectively, $C_x^{\mathrm{T}}C_x + S_x^{\mathrm{T}}S_x = I_r$, $C_y^{\mathrm{T}}C_y + S_y^{\mathrm{T}}S_y = I_r$. By Lemma 1, we know that $\|S_x\| = \theta_x$, $\|S_y\| = \theta_y$.

Noticing that

$$\begin{aligned}
\min_{X,Y}\|L_* - XY^{\mathrm{T}}\|^2 &= \min_D \|U^{\mathrm{T}}L_*V - U^{\mathrm{T}}\widehat{X}D\widehat{Y}^{\mathrm{T}}V\|^2 = \min_D \left\|\begin{bmatrix}\Sigma_1 & 0\\ 0 & 0\end{bmatrix} - \begin{bmatrix}C_x\\ S_x\end{bmatrix}D[C_y^{\mathrm{T}}, S_y^{\mathrm{T}}]\right\|^2 \\
&= \left\|\begin{bmatrix}\Sigma_1 & 0\\ 0 & 0\end{bmatrix} - \begin{bmatrix}C_x\\ S_x\end{bmatrix}[C_x, S_x]^{\mathrm{T}}\begin{bmatrix}\Sigma_1 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}C_y\\ S_y\end{bmatrix}[C_y^{\mathrm{T}}, S_y^{\mathrm{T}}]\right\|^2,
\end{aligned}$$

we have

$$\begin{aligned}\min_{X,Y}\|L_*-XY^{\mathrm{T}}\|^2 &\geq \max\left\{\left\|C_x[C_x,S_x]^{\mathrm{T}}\begin{bmatrix}\Sigma_1 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}C_y\\ S_y\end{bmatrix}S_y^{\mathrm{T}}\right\|^2, \left\|S_x[C_x,S_x]^{\mathrm{T}}\begin{bmatrix}\Sigma_1 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}C_y\\ S_y\end{bmatrix}C_y^{\mathrm{T}}\right\|^2\right\}\\ &\geq \max\{\sigma_{\min}^2(C_x)\|S_y\|^2, \sigma_{\min}^2(C_y)\|S_x\|^2\}\sigma_{\min}^2\left([C_x,S_x]^{\mathrm{T}}\begin{bmatrix}\Sigma_1 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}C_y\\ S_y\end{bmatrix}\right)\\ &\geq \max\{(1-\theta_x^2)\theta_y^2, (1-\theta_y^2)\theta_x^2\}(\sigma_{r*}\sqrt{1-\theta_x^2}\sqrt{1-\theta_y^2})^2\end{aligned}$$

Combining it with the fact that $\|L_*-G-XY^{\mathrm{T}}\|\geq\|L_*-XY^{\mathrm{T}}\|-\|G\|$ for any X, Y, we get the conclusion. □

Lemma 8. *Let L_*, G be the same as in Lemma 7. Let $X=(L_*-G)Y$, where $Y\in\mathbb{R}^{n\times r}$ is orthonormal. Denote $\theta_x=\|\sin\Theta(U_*,X)\|$, $\theta_y=\|\sin\Theta(V_*,Y)\|$. If $\|G\|<\sigma_{r*}\sqrt{1-\theta_y^2}$, then*

$$\sigma_r(X)\geq\sigma_{r*}\sqrt{1-\theta_y^2}-\|G\|,\qquad \theta_x\leq\frac{\|G\|}{\sigma_r\sqrt{1-\theta_y^2}-\|G\|}.$$

Proof. By Lemma 2 and Lemma 1, we have

$$\begin{aligned}\sigma_r(X)&=\sigma_r((L_*-G)Y)\geq\sigma_r(L_*Y)-\|GY\|\geq\sigma_r(\Sigma_*V_*^{\mathrm{T}}Y)-\|G\|\geq\sigma_{r_*}\sigma_{\min}(V_*^{\mathrm{T}}Y)-\|G\|\\ &=\sigma_{r*}\sigma_{\min}^{\frac{1}{2}}(Y^{\mathrm{T}}V_*V_*^{\mathrm{T}}Y)-\|G\|\geq\sigma_{r*}\sigma_{\min}^{\frac{1}{2}}(I_r-Y^{\mathrm{T}}(I-V_*V_*^{\mathrm{T}})Y)-\|G\|\\ &=\sigma_{r*}\sqrt{1-\|(I-V_*V_*^{\mathrm{T}})Y\|^2}-\|G\|=\sigma_{r*}\sqrt{1-\theta_y^2}-\|G\|>0.\end{aligned}\tag{17}$$

Therefore, X has full column rank. Denote $G_x=(X^{\mathrm{T}}X)^{-\frac{1}{2}}$, $\widehat{X}=XG_x$. Then $\widehat{X}$ and $X=AY$ can be rewritten as $\widehat{X}=AYG_x$. Using Lemma 1 and (17), we have

$$\|\theta_x\|=\|U_{*,c}^{\mathrm{T}}\widehat{X}\|=\|U_{*,c}^{\mathrm{T}}(L_*-G)YG_x\|\leq\|GYG_x\|\leq\|G\|\|G_x\|\leq\frac{\|G\|}{\sigma_r(X)}\leq\frac{\|G\|}{\sigma_{r*}\sqrt{1-\theta_y^2}-\|G\|}.$$

The proof is completed. □

Lemma 9. *Let U, $X\in\mathbb{R}^{m\times r}$ both have orthonormal columns. It holds $\|X\|_{2,\infty}\leq\|U\|_{2,\infty}+\|\sin\Theta(U,X)\|$.*

Proof. Let U_c be such that $[U,U_c]$ is an orthogonal matrix. We can write $X=UC_x+U_cS_x$, where $C_x^{\mathrm{T}}C_x+S_x^{\mathrm{T}}S_x=I_r$. By Lemma 1, we have $\|\sin\Theta(U,X)\|=\|U_c^{\mathrm{T}}X\|=\|S_x\|$. Then for any $1\leq i\leq m$, we have

$$\|e_i^{\mathrm{T}}X\|=\|e_i^{\mathrm{T}}UC_x+e_i^{\mathrm{T}}U_cS_x\|\leq\|e_i^{\mathrm{T}}U\|+\|S_x\|,$$

the conclusion follows. □

Lemma 10. *(Jain and Netrapalli, 2015, Lemmas 8,10) Let $A\in\mathbb{R}^{m\times n}$ with $m\geq n$. Suppose Ω is obtained by sampling each entry of A with probability $p\in[\frac{1}{4m},0.5]$. Then w.p. $\geq 1-1/m^{10+\log\alpha}$,*

$$\|\frac{1}{p}\Pi_\Omega(A)-A\|\leq\frac{6\sqrt{\alpha m}}{\sqrt{p}}\|A\|_{\max}.$$

2 Proof for Main Theorems

2.1 Proof of Theorem 1

Proof of Theorem 1. First, it holds $\|(I-U_*U_*^{\mathrm{T}})M(I-V_*V_*^{\mathrm{T}})\|=\|(I-U_*U_*^{\mathrm{T}})S_*(I-V_*V_*^{\mathrm{T}})\|$. Then by assumption, we have $\|(I-U_*U_*^{\mathrm{T}})M(I-V_*V_*^{\mathrm{T}})\|<\sigma_{r*}$.

Second, we have

$$\|E\| = \|MV_* - U_*\Sigma_*\| = \|L_*V_* - U_*\Sigma_* + S_*V_*\| = \|S_*V_*\|,$$
$$\|F\| = \|M^{\mathrm{T}}U_* - V_*\Sigma_*\| = \|L_*^{\mathrm{T}}U_* - V_*\Sigma_* + S_*^{\mathrm{T}}U_*\| = \|S_*^{\mathrm{T}}U_*\|.$$

It follows

$$\max\{\|E\|, \|F\|\} = \max\{\|S_*V_*\|, \|S_*^{\mathrm{T}}U_*\|\} < \sigma_r - \sigma_{r+1}.$$

Then applying Lemma 4 gives the conclusion. □

2.2 Proof of Theorem 2

Throughout the rest of this section, we follow the notations in Algorithm 1. Besides that, we will also adopt the following notations. Denote

$$r = \operatorname{rank}(L_*), \qquad \kappa_* = \kappa_2(L_*), \qquad p' = p(1-\varrho), \qquad \Omega_t = \Omega/\operatorname{supp}(S_t), \qquad G_t = S_t - S_*. \tag{18}$$

The SVDs of L_* is given by

$$L_* = [U_*, U_{*,c}] \operatorname{diag}(\Sigma_*, 0)[V_*, V_{*,c}]^{\mathrm{T}}, \tag{19}$$

where $[U_*, U_{*,c}]$ and $[V_*, V_{*,c}]$ are orthogonal matrices $U_* \in \mathbb{R}^{m\times r}$ and $V_* \in \mathbb{R}^{n\times r}$, $\Sigma_* = \operatorname{diag}(\sigma_{1*}, \dots, \sigma_{r*})$ with $\sigma_{1*} \geq \cdots \geq \sigma_{r*} > 0$. Further denote

$$\theta_{x,t} = \|\sin\Theta(U_*, X_t)\|, \qquad \theta_{y,t} = \|\sin\Theta(V_*, Y_t)\|. \tag{20}$$

Lemma 11. $\|S_t - S_*\|_{\max} \leq 2\|\Pi_\Omega(X_t\Sigma_tY_t^{\mathrm{T}} - L_*)\|_{\max}$ *for* $t = 0, 1, \dots$.

Proof. Denote $\Phi_* = \operatorname{supp}(S_*)$, $\Phi_t = \operatorname{supp}(S_t)$, it is obvious that $S_t - S_*$ is supported on $\Phi_t \cup \Phi_*$ and $\Phi_t \cup \Phi_* \subset \Omega$. Now we claim that

$$\|\Pi_\Omega(S_t - S_*)\|_{\max} \leq 2\|\Pi_\Omega(X_t\Sigma_tY_t^{\mathrm{T}} - L_*)\|_{\max}.$$

To show the claim, it suffices to consider the following two cases.

Case (1) For any $(i,j) \in \Phi_t$, it holds $(S_t)_{(i,j)} = (L_* + S_* - X_t\Sigma_tY_t^{\mathrm{T}})_{(i,j)}$. Then it follows that

$$|(S_t - S_*)_{(i,j)}| = |(L_* - X_t\Sigma_tY_t^{\mathrm{T}})_{(i,j)}| \leq \|\Pi_\Omega(X_t\Sigma_tY_t^{\mathrm{T}} - L_*)\|_{\max}.$$

Case (2) For any $(i,j) \in \Phi_* \setminus \Phi_t$, it holds $(S_t)_{(i,j)} = 0$. If $|(S_t - S_*)_{(i,j)}| = |(S_*)_{(i,j)}| > 2\|\Pi_\Omega(X_t\Sigma_tY_t^{\mathrm{T}} - L_*)\|_{\max}$, then

$$|(L_* + S_* - X_t\Sigma_tY_t^{\mathrm{T}})_{(i,j)}| > \|\Pi_\Omega(L_* - X_t\Sigma_tY_t^{\mathrm{T}})\|_{\max}.$$

Noticing that S_* only changes s entries of $\Pi_\Omega(L_* - X_t\Sigma_tY_t^{\mathrm{T}})$, we know that the (i,j) entry of $|\Pi_\Omega(L_* + S_* - X_t\Sigma_tY_t^{\mathrm{T}})|$ is larger than the $(s{+}1)$st largest entry of $|\Pi_\Omega(L_* + S_* - X_t\Sigma_tY_t^{\mathrm{T}})|$. This contradicts with $(i,j) \notin \Phi_t$. □

Lemma 12. *Assume* **(A1)**. *Denote* $r_s' = \frac{\|S_0 - S_*\|_F^2}{\|S_0 - S_*\|^2}$. *Let* S_0 *be obtained as in Algorithm 1. It holds*

$$\|S_0 - S_*\| \leq 2\sqrt{\frac{2\varrho p}{r_s'}}\mu r\|L_*\|.$$

Proof. First, for any i, j, we have $L_{ij} = e_i^{\mathrm{T}}U_*\Sigma_*V_*^{\mathrm{T}}e_j$. Using **(A1)**, we have

$$|L_{ij}| \leq \|e_i^{\mathrm{T}}U_*\|\|\Sigma_*\|\|e_j^{\mathrm{T}}V_*\| \leq \frac{\mu r}{\sqrt{mn}}\|L_*\|,$$

and hence

$$\|L_*\|_{\max} \leq \frac{\mu r}{\sqrt{mn}}\|L_*\|. \tag{21}$$

By Lemma 11, we have

$$\|S_0 - S_*\|_{\max} \le 2\|\Pi_\Omega(L_*)\|_{\max} \le \frac{2\mu r}{\sqrt{mn}}\|L_*\|. \tag{22}$$

Therefore, using (21), (22) and **(A2)**, we have

$$\|S_0 - S_*\|_F \le \sqrt{2s}\|S_0 - S_*\|_{\max} \le 2\sqrt{2s}\|\Pi_\Omega(L_*)\|_{\max} \le 2\sqrt{2s}\|L_*\|_{\max} \le 2\sqrt{2\varrho p}\mu r\|L_*\|. \tag{23}$$

By the definition of r_s', it follows that

$$\|S_0 - S_*\| \le \frac{\|S_0 - S_*\|_F}{\sqrt{r_s'}} \le 2\sqrt{\frac{2\varrho p}{r_s'}}\mu r\|L_*\|.$$

The proof is completed. □

Proof of Theorem 2. By **(A3)**, Lemma 10 and (21), w.p. $\ge 1 - 1/m^{10+\log\alpha}$, it holds

$$\|\frac{1}{p'}\Pi_{\Omega_0}(L_*) - L_*\| \le \frac{6\sqrt{\alpha m}}{\sqrt{p'}}\|L_*\|_{\max} \le \xi\mu r\|L_*\|. \tag{24}$$

Using Lemma 12 and (24), we have w.p. $\ge 1 - 1/m^{10+\log\alpha}$,

$$\|\frac{1}{p'}\Pi_{\Omega_0}(M - S_0) - L_*\| \le \|\frac{1}{p'}\Pi_{\Omega_0}(L_*) - L_*\| + \frac{1}{p'}\|S_0 - S_*\| \le (\xi + \gamma)\mu r\|L_*\|. \tag{25}$$

Let the SVD of $X_1^{\mathrm{T}} L_* Y_1$ be $X_1^{\mathrm{T}} L_* Y_1 = P\widetilde{\Sigma}Q^{\mathrm{T}}$, where P, Q are orthogonal matrices, $\widetilde{\Sigma} = \mathrm{diag}(\tilde{\sigma}_1, \dots, \tilde{\sigma}_r)$. Denote $\widetilde{U} = X_1 P$, $\widetilde{V} = Y_1 Q$, and let

$$E = L_*\widetilde{V} - \widetilde{U}\widetilde{\Sigma}, \qquad F = L_*^{\mathrm{T}}\widetilde{U} - \widetilde{V}\widetilde{\Sigma}. \tag{26}$$

Then it follows that

$$\|\Sigma_1 - X_1^{\mathrm{T}} L_* Y_1\| = \|X_1^{\mathrm{T}}[\frac{1}{p'}\Pi_{\Omega_0}(M - S_0) - L_*]Y_1\| \le \|\frac{1}{p'}\Pi_{\Omega_0}(M - S_0) - L_*\|. \tag{27}$$

Using (25) and (27), by calculations, we get

$$\begin{aligned}
\|E\| &= \|L_*\widetilde{V} - \widetilde{U}\widetilde{\Sigma}\| = \|L_* Y_1 - X_1 P\widetilde{\Sigma}Q^{\mathrm{T}}\| = \|L_* Y_1 - X_1 X_1^{\mathrm{T}} L_* Y_1\| \\
&\le \|L_* Y_1 - X_1\Sigma_1\| + \|X_1(\Sigma_1 - X_1^{\mathrm{T}} L_* Y_1)\| \\
&= \|L_* Y_1 - \frac{1}{p'}\Pi_{\Omega_0}(M - S_0)Y_1\| + \|\Sigma_1 - X_1^{\mathrm{T}} L_* Y_1\| \\
&\le 2\|\frac{1}{p'}\Pi_{\Omega_0}(M - S_0) - L_*\| \le 2(\xi + \gamma)\mu r\|L_*\|, \quad \text{w.p. } \ge 1 - 1/m^{10+\log\alpha}.
\end{aligned} \tag{28}$$

Similarly, we get

$$\|F\| \le 2(\xi + \gamma)\mu r\|L_*\|, \quad \text{w.p. } \ge 1 - 1/m^{10+\log\alpha}. \tag{29}$$

Next, we only need to show $\max\{\|E\|, \|F\|\} \le \sigma_{r*}$ and $\|(I_m - \widetilde{U}\widetilde{U}^{\mathrm{T}})L_*(I_n - \widetilde{V}\widetilde{V}^{\mathrm{T}})\| < \tilde{\sigma}_r$. Once these two inequalities hold, we may apply Lemma 4.

For the first inequality, using (28), (29) and the assumption $(\xi + \gamma)\mu\kappa r < \frac{1}{6}$, we get

$$\max\{\|E\|, \|F\|\} \le 2(\xi + \gamma)\mu r\|L_*\| < \sigma_{r*}, \quad \text{w.p. } \ge 1 - 1/m^{10+\log\alpha}. \tag{30}$$

For the second inequality, using (23) and (24), we have

$$\|\frac{1}{p'}\Pi_\Omega(M-S_0)-L_*\| \le \|\frac{1}{p'}\Pi_\Omega(L_*)-L_*\| + \frac{1}{p'}\|S_*-S_0\| \le (\xi+\gamma)\mu r\|L_*\|. \tag{31}$$

Then using Lemma 2, (25),(27) and (31), we have

$$|\tilde\sigma_r-\sigma_{r*}| \le |\tilde\sigma_r-\hat\gamma_{r,0}| + |\hat\gamma_{r,0}-\sigma_{r*}| \le \|X_1^{\mathrm{T}}L_*Y_1-\Sigma_1\| + \|\frac{1}{p'}\Pi_{\Omega_0}(M-S_0)-L_*\| \le 2(\xi+\gamma)\mu r\|L_*\|.$$

It follows that

$$\tilde\sigma_r \ge \sigma_{r*} - 2(\xi+\gamma)\mu r\|L_*\|. \tag{32}$$

Then using the assumption $(\xi+\gamma)\mu\kappa r < \frac{1}{6}$, (25) and (32), we have

$$\begin{aligned}&\|(I_m-\widetilde U\widetilde U^{\mathrm{T}})L_*(I_n-\widetilde V\widetilde V^{\mathrm{T}})\| = \|(I_m-X_1X_1^{\mathrm{T}})[L_*-\frac{1}{p'}\Pi_{\Omega_0}(M-S_0)](I_n-Y_1Y_1^{\mathrm{T}})\|\\ \le&\|L_*-\frac{1}{p'}\Pi_{\Omega_0}(M-S_0)\| \le (\xi+\gamma)\mu r\|L_*\| < \sigma_{r*}-2(\xi+\gamma)\mu r\|L_*\| \le \tilde\sigma_r.\end{aligned}$$

Now using (28), (29), the assumption $(\xi+\gamma)\mu\kappa r < \frac{1}{6}$ and Lemma 4, we have

$$\max\{\theta_{x,1},\theta_{y,1}\} = \max\{\|\sin\Theta(U_*,\widetilde U)\|,\|\sin\Theta(V_*,\widetilde V)\|\} \le \frac{2(\xi+\gamma)\mu r\kappa}{1-1/3} = 3(\xi+\gamma)\mu r\kappa, \tag{33a}$$

$$\|L_*-\widetilde U\widetilde\Sigma\widetilde V^{\mathrm{T}}\|_{\max}/\|L_*\| \le (\|U_*\|_{2,\infty}\theta_{y,1}+\|V_*\|_{2,\infty}\theta_{x,1}) + (1+3\|U_*\|_{2,\infty}\|V_*\|_{2,\infty})\theta_{x,1}\theta_{y,1}. \tag{33b}$$

Using the assumption $(\xi+\gamma)\mu\kappa r < \frac{1}{3}\sqrt{\frac{\mu_1' r}{m}}$, by (33a), we have $\max\{\theta_{x,1},\theta_{y,1}\} \le \sqrt{\frac{\mu_1' r}{m}}$. On the other hand, assumption **(A1)** implies that

$$\|U_*\|_{2,\infty} \le \sqrt{\frac{\mu r}{m}}, \qquad \|V_*\|_{2,\infty} \le \sqrt{\frac{\mu r}{n}}. \tag{34}$$

Then it follows from Lemma 9 that

$$\|X_1\|_{2,\infty} \le \|U_*\|_{2,\infty} + \|\sin\Theta(X_1,U_*)\| \le \sqrt{\frac{\mu r}{m}} + \sqrt{\frac{\mu_1' r}{m}} \le \sqrt{\frac{\mu_1 r}{m}}, \tag{35a}$$

$$\|Y_1\|_{2,\infty} \le \|V_*\|_{2,\infty} + \|\sin\Theta(Y_1,V_*)\| \le \sqrt{\frac{\mu r}{n}} + \sqrt{\frac{\mu_1' r}{m}} \le \sqrt{\frac{\mu_1 r}{n}}. \tag{35b}$$

Using the assumption $(\xi+\gamma)\mu\kappa r < \frac{1}{3}\sqrt{\frac{\mu_1' r}{m}}$, (25), (27), (33b), (34) and (35), by calculations, we have

$$\begin{aligned}\|L_*-X_1\Sigma_1Y_1^{\mathrm{T}}\|_{\max}/\|L_*\| \le&\|L_*-\widetilde U\widetilde\Sigma\widetilde V^{\mathrm{T}}\|_{\max}/\|L_*\| + \|\widetilde U\widetilde\Sigma\widetilde V^{\mathrm{T}}-X_1\Sigma_1Y_1^{\mathrm{T}}\|_{\max}/\|L_*\|\\ =&\|L_*-\widetilde U\widetilde\Sigma\widetilde V^{\mathrm{T}}\|_{\max}/\|L_*\| + \|X_1(X_1^{\mathrm{T}}L_*Y_1-\Sigma_1)Y_1^{\mathrm{T}}\|_{\max}/\|L_*\|\\ \le&\|L_*-\widetilde U\widetilde\Sigma\widetilde V^{\mathrm{T}}\|_{\max}/\|L_*\| + \|X_1\|_{2,\infty}\|X_1^{\mathrm{T}}L_*Y_1-\Sigma_1\|\|Y_1\|_{2,\infty}/\|L_*\|\\ \le&\|L_*-\widetilde U\widetilde\Sigma\widetilde V^{\mathrm{T}}\|_{\max}/\|L_*\| + \|X_1\|_{2,\infty}\|Y_1\|_{2,\infty}(\xi+\gamma)\mu r\kappa,\\ \le&(\|U_*\|_{2,\infty}\theta_{y,1}+\|V_*\|_{2,\infty}\theta_{x,1}) + (1+3\|U_*\|_{2,\infty}\|V_*\|_{2,\infty})\theta_{x,1}\theta_{y,1}\\ &\qquad + \|X_1\|_{2,\infty}\|Y_1\|_{2,\infty}\frac{1}{3}\sqrt{\frac{\mu_1' r}{m}}\\ \le&\left(\sqrt{\frac{\mu r}{m}}\theta_{y,1}+\sqrt{\frac{\mu r}{n}}\theta_{x,1}+\theta_{x,1}\theta_{y,1}\right) + \mathcal{O}(n^{-3/2}),\end{aligned}$$

which completes the proof. □

2.3 Proof of Theorem 3

Proof of Theorem 3. First, we give an upper bound for $\sup_{X\in\mathbb{R}^{m\times r}}\|\Pi_{\Omega_t}(R)\Pi_{\Omega_t}(XY_t^{\mathrm{T}})^{\mathrm{T}}\|/\|X\|$. Let $\{\delta_{ij}\}$ be an independent family of $\textsc{Bernoulli}(p')$ random variables, $X^{\mathrm{T}}=[x_1,\dots,x_m]\in\mathbb{R}^{r\times m}$ be arbitrary nonzero matrix with $\|X\|=1$, and $Y_t^{\mathrm{T}}=[y_1,\dots,y_n]$. Denote $E_{ij}=e_ie_j^{\mathrm{T}}$, $R=[r_{ij}]$, $\mathbf{W}_{il}=\sum_{j,k}\delta_{ij}r_{ij}E_{ij}\delta_{lk}x_k^{\mathrm{T}}y_lE_{kl}^{\mathrm{T}}$, $\mathbf{Z}=\sum_{i,l}\mathbf{Z}_{i,l}$. By calculations, we have

$$
\begin{aligned}
&\mathbb{E}(\mathbf{W}_{il})=p'^2\sum_{j,k}r_{ij}E_{ij}y_k^{\mathrm{T}}x_lE_{kl},=p'^2\sum_{j}r_{ij}E_{ij}y_j^{\mathrm{T}}x_lE_{jl}=p'^2R_{(i,:)}Y_tx_lE_{il}=0,\\
&\|\mathbf{W}_{il}\|\le\sqrt{p'n}\max|r_{ij}x_j^{\mathrm{T}}y_l|\le\sqrt{\mu'rp'}\|R\|_{\max},\\
&\|\mathbb{E}[\sum_{i,l}\mathbf{W}_{il}\mathbf{W}_{il}^{\mathrm{T}}]\|=\|\mathbb{E}[\sum_{i,l}(\sum_{j,k}\delta_{ij}r_{ij}E_{ij}\delta_{lk}x_k^{\mathrm{T}}y_lE_{kl}^{\mathrm{T}})(\sum_{j',k'}\delta_{ij'}r_{ij'}E_{ij'}\delta_{lk'}x_{k'}^{\mathrm{T}}y_lE_{k'l}^{\mathrm{T}})^{\mathrm{T}}]\|=0,\\
&\|\mathbb{E}[\sum_{i,l}\mathbf{W}_{il}^{\mathrm{T}}\mathbf{W}_{il}]\|=0.
\end{aligned}
$$

By Lemma 5, we have $\mathbb{P}\{\|\mathbf{W}\|>t\}\le(m+n)\exp\left(-\frac{3t/2}{\sqrt{\mu'rp'}\|R\|_{\max}}\right)$. Let $t=\frac{2}{3}(\log(m+n)+5)\sqrt{\mu'rp'}\|R\|_{\max}$, then w.p. ≥ 0.99, it holds

$$\|\mathbf{W}\|\le\frac{2}{3}(\log(m+n)+5)\sqrt{\mu'rp'}\|R\|_{\max}. \tag{36}$$

Second, It is easy to see that $X_{\text{opt}}=(M-S_t)Y_t$. By calculations, we have

$$
\begin{aligned}
&\min_X\|\Pi_{\Omega_t}(XY_t^{\mathrm{T}})-\Pi_{\Omega_t}(M-S_t)\|^2 &(37)\\
=&\min_{\Delta X}\|\Pi_{\Omega_t}((X_{\text{opt}}+\Delta X)Y_t^{\mathrm{T}})-\Pi_{\Omega_t}\big((M-S_t)Y_tY_t^{\mathrm{T}}+(M-S_t)(I-Y_tY_t^{\mathrm{T}})\big)\|^2\\
=&\min_{\Delta X}\|\Pi_{\Omega_t}(\Delta XY_t^{\mathrm{T}})-\Pi_{\Omega_t}(R)\|^2. &(38)
\end{aligned}
$$

Then we declare that (38) is minimized when $\Delta X=\widetilde{X}_{\text{opt}}-X_{\text{opt}}$. This is because (37) is minimized when $X=\widetilde{X}_{\text{opt}}$ and $X=X_{\text{opt}}+\Delta X$. Thus, we have

$$\|\widetilde{X}_{\text{opt}}-X_{\text{opt}}\|=\|\Delta X\|\le\frac{\sup_{X\in\mathbb{R}^{m\times r}}\|\Pi_{\Omega_t}(R)\Pi_{\Omega_t}(XY_t^{\mathrm{T}})^{\mathrm{T}}\|}{\sigma^2}. \tag{39}$$

Substituting (36) into (39), we get the conclusion. □

2.4 Proof of Theorem 4

Lemma 13. *Denote* $r_s=\inf_t\frac{\|S_t-S_*\|_F^2}{\|S_t-S_*\|^2}$, $\zeta=\sqrt{\frac{2s\mu r}{mr_s}}$. *If* $\|L_*-X_t\Sigma_tY_t^{\mathrm{T}}\|_{\max}\le c_t\|L_*\|\sqrt{\frac{\mu r}{m}}$ *for some positive parameter* c_t, *then*

$$\|S_t-S_*\|\le 2c_t\|L_*\|\zeta,\qquad |\gamma_{j,t}-\sigma_{j*}|\le 2c_t\|L_*\|\zeta.$$

Proof. Using Lemma 11, by simple calculations, we have

$$\|S_t-S_*\|\le\frac{\|S_t-S_*\|_F}{\sqrt{r_s}}\le\sqrt{\frac{2s}{r_s}}\|S_t-S_*\|_{\max}\le 2\sqrt{\frac{2s}{r_s}}\|\Pi_\Omega(L_*-X_t\Sigma_tY_t^{\mathrm{T}})\|_{\max}\le 2c_t\|L_*\|\sqrt{\frac{2s\mu r}{mr_s}}=2c_t\|L_*\|\zeta.$$

Then by Lemma 2, we know that

$$|\gamma_{j,t}-\sigma_{j*}|\le\|(M-S_t)-L_*\|=\|S_t-S_*\|\le 2c_t\|L_*\|\zeta.$$

The proof is completed. □

Proof of Theorem 4. First, denote $\bar{X}_{t+1} = (M - S_t)Y_t$, then we know that $\bar{X}_{t+1}$ is the solution to $\min_X \|M - S_t - XY_t^{\mathrm{T}}\|$. Also note that $\widetilde{X}_{t+1}$ on line 8 of Algorithm 1 is the solution to $\min_X \|\Pi_{\Omega_t}(M - S_t - XY_t^{\mathrm{T}})\|$. Then by Theorem 3, we have

$$\|\bar{X}_{t+1} - \widetilde{X}_{t+1}\| \leq C_{\mathrm{LS}} \|(M - S_t)(I - Y_t Y_t^{\mathrm{T}})\|_{\max}, \quad \text{w.p. } \geq 0.99.$$

Then it follows that from Lemma 1, Lemma 11 and Lemma 13 that

$$\begin{aligned}
\|\bar{X}_{t+1} - \widetilde{X}_{t+1}\| &\leq C_{\mathrm{LS}}(\|L_*(I - Y_t Y_t^{\mathrm{T}})\|_{\max} + \|(S_t - S_*)(I - Y_t Y_t^{\mathrm{T}})\|_{\max}) \\
&\leq C_{\mathrm{LS}}(\|L_*\|\sqrt{\frac{\mu r}{m}}\theta_{y,t} + \|S_t - S_*\|_{2,\infty}) \leq C_{\mathrm{LS}}(\|L_*\|\sqrt{\frac{\mu r}{m}}\theta_{y,t} + \sqrt{2p\varrho n}\|S_t - S_*\|_{\max}) \\
&\leq C_{\mathrm{LS}}(\|L_*\|\sqrt{\frac{\mu r}{m}}\theta_{y,t} + 2\sqrt{2p\varrho n}\|L_* - X_t \Sigma_t Y_t^{\mathrm{T}}\|_{\max}) \leq \frac{C}{\sqrt{m}}\|L_*\|\theta_{y,t}.
\end{aligned} \tag{40}$$

Second, using Lemma 13 and $4c\kappa\zeta < 1$, we have

$$\|S_t - S_*\| < 2c\theta_{y,t}\|L_*\|\zeta \leq \sqrt{2}c\|L_*\|\zeta < \frac{\sigma_{r*}}{\sqrt{2}} \leq \sigma_{r*}\sqrt{1 - \theta_{y,t}^2}, \tag{41}$$

Then by Lemma 8, we know that

$$\|\sin\Theta(\bar{X}_{t+1}, U_*)\| \leq \frac{\|S_t - S_*\|}{\sigma_{r*}\sqrt{1 - \theta_{y,t}^2} - \|S_t - S_*\|}. \tag{42}$$

Using (42), the assumption $\|L_* - X_t\Sigma_t Y_t^{\mathrm{T}}\|_{\max} \leq c\|L_*\|\ \theta_{y,t}\sqrt{\frac{\mu r}{m}}$, Lemma 13 and $\theta_{y,t} \leq \frac{1}{\sqrt{2}}$, we get

$$\|\sin\Theta(\bar{X}_{t+1}, U_*)\| \leq \frac{2c\|L_*\|\zeta\theta_{y,t}}{\frac{\sigma_{r*}}{\sqrt{2}} - 2c\|L_*\|\zeta\theta_{y,t}} \leq \frac{2\sqrt{2}c\kappa\zeta\theta_{y,t}}{1 - 2c\kappa\zeta} < 4\sqrt{2}c\kappa\zeta\theta_{y,t}. \tag{43}$$

Therefore, using Lemma 1, (40) and (43), we have

$$\begin{aligned}
\|\theta_{x,t+1}\| &= \|U_{*,c}^{\mathrm{T}} X_{t+1}\| = \|U_{*,c}^{\mathrm{T}}\widetilde{X}_{t+1}R_{x,t+1}^{-1}\| \leq \|U_{*,c}^{\mathrm{T}}\bar{X}_{t+1}R_{x,t+1}^{-1}\| + \|U_{*,c}^{\mathrm{T}}(\widetilde{X}_{t+1} - \bar{X}_{t+1})R_{x,t+1}^{-1}\| \\
&\leq \|R_{x,t+1}^{-1}\|(\|\sin\Theta(\bar{X}_{t+1}, U_*)\|\|\bar{X}_{t+1}\| + \|\widetilde{X}_t - \bar{X}_t\|) \\
&\leq \frac{1}{\sigma_r(\widetilde{X}_{t+1})}\Big(4\sqrt{2}c\kappa\zeta\|\bar{X}_{t+1}\| + \frac{C}{\sqrt{m}}\|L_*\|\Big)\theta_{y,t}.
\end{aligned} \tag{44}$$

Now using Lemma 2, (40), $\theta_{y,t} \leq \frac{1}{\sqrt{2}}$ and (41), we have

$$\begin{aligned}
&\|\bar{X}_{t+1}\| = \|(M - S_t)Y_t\| \leq \|L_* Y\| + \|S_t - S_*\| \leq \|L_*\| + \sqrt{2}c\|L_*\|\zeta, \\
&\sigma_r(\widetilde{X}_{t+1}) \geq \sigma_r(\bar{X}_{t+1}) - \frac{C}{\sqrt{m}}\|L_*\|\theta_{y,t} \geq \sigma_r((M - S_t)Y_t) - \frac{C}{\sqrt{2m}}\|L_*\| \geq \sigma_r(L_* Y_t) - \|S_t - S_*\| - \frac{C}{\sqrt{2m}}\|L_*\| \\
&\qquad \geq \sigma_{r*}\sqrt{1 - \theta_{y,t}^2} - \sqrt{2}c\|L_*\|\zeta - \frac{C}{\sqrt{2m}}\|L_*\| \geq \frac{\sigma_{r*}}{\sqrt{2}} - \sqrt{2}c\|L_*\|\zeta - \frac{C}{\sqrt{2m}}\|L_*\|.
\end{aligned}$$

Substituting them into (44), we get the conclusion. □

2.5 Proof of Theorem 5

Lemma 14. *Follow the notations and assumptions in Lemma 1. Then*

$$\|L_* - \widehat{X}_{t+1}R_{x,t+1}Y_t^{\mathrm{T}}\|_{\max} \leq \Big((1 + C_{\mathrm{LS}}\sqrt{\frac{\mu' r}{n}})\sqrt{\frac{\mu r}{m}} + (1 + C_{\mathrm{LS}}\sqrt{2\varrho p n})\sqrt{\frac{\mu' r}{n}}2c\zeta\Big)\|L_*\|\theta_{y,t}.$$

Proof. Direct calculations give rise to

$$\begin{aligned}
\|L_* - \widetilde{X}_{t+1}Y_t^{\mathrm{T}}\|_{\max} &\leq \|L_* - (M-S_t)Y_tY_t^{\mathrm{T}}\|_{\max} + \|(M-S_t)Y_tY_t^{\mathrm{T}} - \widetilde{X}_{t+1}Y_t^{\mathrm{T}}\|_{\max} \\
&\leq \|L_* - (M-S_t)Y_tY_t^{\mathrm{T}}\|_{\max} + \|(M-S_t)Y_t - \widetilde{X}_{t+1}\|\sqrt{\frac{\mu' r}{n}} \quad (45a)\\
&\leq \|L_* - (M-S_t)Y_tY_t^{\mathrm{T}}\|_{\max} + C_{\mathrm{LS}}\sqrt{\frac{\mu' r}{n}}\|(M-S_t)(I-Y_tY_t^{\mathrm{T}})\|_{\max} \quad (45b)\\
&\leq (1+C_{\mathrm{LS}}\sqrt{\frac{\mu' r}{n}})\|L_*(I-Y_tY_t^{\mathrm{T}})\|_{\max} + (1+C_{\mathrm{LS}}\sqrt{2\varrho pn})\sqrt{\frac{\mu' r}{n}}\|S_t - S_*\|_{\max} \\
&\leq \Big((1+C_{\mathrm{LS}}\sqrt{\frac{\mu' r}{n}})\sqrt{\frac{\mu r}{m}} + (1+C_{\mathrm{LS}}\sqrt{2\varrho pn})\sqrt{\frac{\mu' r}{n}}2c\zeta\Big)\|L_*\|\theta_{y,t} \quad (45c)
\end{aligned}$$

where (45a) uses $\|Y_t\|_{2,\infty} \leq \sqrt{\frac{\mu' r}{m}}$, (45b) uses Theorem 3, (45c) uses the SVD of L_*, $\|U_*\|_{2,\infty} \leq \sqrt{\frac{\mu r}{m}}$ and Lemme 1. □

Proof of Theorem 5. First, by Lemma 7, we have

$$\|M - S_t - X_t\Sigma_tY_t^{\mathrm{T}}\| \geq \sigma_{r*}\max\{\sqrt{1-\theta_{x,t}^2}\theta_{y,t}, \sqrt{1-\theta_{y,t}^2}\theta_{x,t}\}\sqrt{1-\theta_{x,t}^2}\sqrt{1-\theta_{y,t}^2} - \|S_t - S_*\|$$

Then using (41), $\theta_{x,t} \leq \frac{1}{\sqrt{2}}$ and $\theta_{y,t} \leq \frac{1}{\sqrt{2}}$, we get

$$\|M - S_t - X_t\Sigma_tY_t^{\mathrm{T}}\| \geq \frac{\sigma_{r*}}{2\sqrt{2}}\theta_{y,t} - 2c\|L_*\|\sqrt{\frac{\mu r}{m}}\theta_{y,t}. \quad (46)$$

Second, by calculations, we have

$$\begin{aligned}
\|M - S_t - \widehat{X}_{t+1}\widetilde{Y}_{t+1}^{\mathrm{T}}\| &\leq \|(I - \widehat{X}_{t+1}\widehat{X}_{t+1}^{\mathrm{T}})(M-S_t)\| + \|\widehat{X}_{t+1}\widehat{X}_{t+1}^{\mathrm{T}}(M-S_t) - \widehat{X}_{t+1}\widetilde{Y}_{t+1}^{\mathrm{T}}\| \\
&\leq \|(I - \widehat{X}_{t+1}\widehat{X}_{t+1}^{\mathrm{T}})L_*\| + \|S_t - S_*\| + \|\widehat{X}_{t+1}^{\mathrm{T}}(M-S_t) - \widetilde{Y}_{t+1}^{\mathrm{T}}\| \\
&\leq \|L_*\|\theta_{x,t+1} + 2c\|L_*\|\zeta\sqrt{\frac{\mu r}{m}}\theta_{x,t+1} + \frac{C}{\sqrt{m}}\|L_*\|\theta_{x,t+1} \quad (47)\\
&\leq (1 + 2c\zeta\sqrt{\frac{\mu r}{m}} + \frac{C}{\sqrt{m}})\phi\|L_*\|\theta_{y,t}, \quad (48)
\end{aligned}$$

where the first two terms of (47) use Lemma 1 and (41), respectively, and the last term can obtained similar to (40), with the help of Lemma 14.

Then it follows that

$$\begin{aligned}
\|M - S_{t+1} - X_{t+1}\Sigma_{t+1}Y_{t+1}^{\mathrm{T}}\| &\leq \|M - S_t - X_{t+1}\Sigma_{t+1}Y_{t+1}^{\mathrm{T}}\| \quad (49a)\\
&\leq (1 + 2c\sqrt{\frac{\mu r}{m}} + \frac{C}{\sqrt{m}})\phi\|L_*\|\theta_{y,t} \quad (49b)\\
&\leq \frac{(1 + 2c\zeta\sqrt{\frac{\mu r}{m}} + \frac{C}{\sqrt{m}})\phi\|L_*\|}{\frac{\sigma_{r*}}{2\sqrt{2}} - 2c\zeta\|L_*\|\sqrt{\frac{\mu r}{m}}}\|M - S_t - X_t\Sigma_tY_t^{\mathrm{T}}\| \quad (49c)\\
&= \psi\|M - S_t - X_t\Sigma_tY_t^{\mathrm{T}}\|,
\end{aligned}$$

where (49a) uses Lemma 6, (49b) uses (48), (49c) uses (46). The proof is completed. □

References

Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.*, 7(1):1–46, 1970.

Prateek Jain and Praneeth Netrapalli. Fast exact matrix completion with finite samples. In *Proceedings of The 28th Conference on Learning Theory (COLT)*, pages 1007–1034, Paris, France, 2015.

Gilbert W Stewart and Ji-Guang Sun. *Matrix Perturbation Theory*. Academic Press, Boston, 1990.

Joel A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015. ISSN 1935-8237. doi: 10.1561/2200000048. URL http://dx.doi.org/10.1561/2200000048.