6 SUPPLEMENTARY MATERIAL

6.1 Proof of Lemma 1

Proof: From (3), we know that $\mathbf{X}$ and $\mathbf{Y}$ are deterministic. With $\mathbf{X}\mathbf{Y}^T \triangleq \widehat{\mathbf{Z}}$, the optimization problem (3) becomes

$$\begin{aligned}
&\arg\min_{\mathbf{X}=\mathbf{Y},0\le\mathbf{Y}\le\tau} \mathbb{E}_{\mathbf{Z}}[\|\widehat{\mathbf{Z}}-\mathbf{Z}\|_F^2] \quad (13)\\
=&\arg\min_{\mathbf{X}=\mathbf{Y},0\le\mathbf{Y}\le\tau} \mathbb{E}_{\mathbf{Z}}\left[\mathrm{Tr}\left[(\widehat{\mathbf{Z}}-\mathbf{Z})^T(\widehat{\mathbf{Z}}-\mathbf{Z})\right]\right]\\
=&\arg\min_{\mathbf{X}=\mathbf{Y},0\le\mathbf{Y}\le\tau} \mathbb{E}_{\mathbf{Z}}[\mathrm{Tr}[\widehat{\mathbf{Z}}^T\widehat{\mathbf{Z}}-2\widehat{\mathbf{Z}}\mathbf{Z}^T+\mathbf{Z}^T\mathbf{Z}]]\\
\overset{(a)}{=}&\arg\min_{\mathbf{X},\mathbf{Y}} \mathrm{Tr}\left[\widehat{\mathbf{Z}}^T\widehat{\mathbf{Z}}-2\mathbb{E}_{\mathbf{Z}}[\widehat{\mathbf{Z}}\mathbf{Z}^T]+\mathbb{E}_{\mathbf{Z}}[\mathbf{Z}^T]\mathbb{E}_{\mathbf{Z}}[\mathbf{Z}]\right]\\
=&\arg\min_{\mathbf{X}=\mathbf{Y},0\le\mathbf{Y}\le\tau} \|\mathbf{X}\mathbf{Y}^T-\mathbb{E}_{\mathbf{Z}}[\mathbf{Z}]\|_F^2 \quad (14)
\end{aligned}$$

where (a) because that $\mathbb{E}$ and Tr are both linear operators and $\mathsf{Tr}[\mathbb{E}[\mathbf{Z}\mathbf{Z}^T]]$ is certain positive number which is independent on the optimization variables, then this term can be replaced by another positive constant $\mathsf{Tr}[\mathbb{E}_{\mathbf{Z}}[\mathbf{Z}^T]\mathbb{E}_{\mathbf{Z}}[\mathbf{Z}]]$. Based on the equivalence between (13) and (14), we can apply the result shown in [28, Lemma 2]. Since the constraint set is polyhedral in the formulation (3), the linear constraint qualification (LCQ) condition is satisfied. Therefore, a Karush-Kuhn-Tucker (KKT) point always exists, and it corresponds to the points that satisfy the first-order stationary solution. Then, the claim is true.

A Sketch of the Proofs. Here, we only give a sketch of the proofs for proving the theoretical results about the convergence of the algorithm. Specially, Lemma 2 proves the boundness of $\mathbf{X}$-iterate. Then, Lemma 3 shows the boudness of the size of the difference of two primal variables, generated using independent data samples. This lemma is useful in the sense that it bounds the randomness of primal variables by that of the samples. Thirdly, the magnitude of the dual difference is bounded by the norm of the successive difference of primal variables and sampled data in Lemma 4.

In the next step, we bound the successive decrease of the augmented Lagrangian in Lemma 5 and show that the augmented Lagrangian is lower bounded by some constant in Lemma 6. Combining the upper bound of $\mathcal{P}(\mathbf{X}^{(t)},\mathbf{Y}^{(t)},\mathbf{\Lambda}^{(t)})$ shown in Lemma 7, we can characterize the convergence rate of the stochastic algorithms with different ways of sample aggregation.

6.2 Proof of Lemma 2

First we show that the size of the $\mathbf{X}$-iterate is bounded from above.

Lemma 2 *If $\rho\ge 8NK\tau^2$, we have an upper bound of $\|\mathbf{X}\|_F$, which is*

$$\begin{aligned}
&\min\left\{\frac{1}{7\tau\sqrt{NK}}\left(2\mathcal{Z}+8NK\tau^2\right),\frac{1}{6\tau\sqrt{NK}}\left(2\mathcal{Z}+8NK\tau^2+\|\mathbf{X}^{(1)}\|_F\tau\sqrt{NK}\right),\|\mathbf{X}^{(1)}\|_F\right\}\\
&\triangleq\gamma\sqrt{NK} \quad (15)
\end{aligned}$$

Note that the constant $\gamma>0$ is defined in a way that the bound of $\|\mathbf{X}\|_F$ can be expressed in a consistent way as the size of $\mathbf{Y}$ (which is bounded above by $\tau\sqrt{NK}$).

Proof: The optimality condition of the $\mathbf{X}$ subproblem (6c) is given by

$$(\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T-\mathbf{Z}_2^{(t)})\mathbf{Y}^{(t+1)}+\rho(\mathbf{X}^{(t+1)}-\mathbf{Y}^{(t+1)}+\mathbf{\Lambda}^{(t)}/\rho)=0. \quad (16)$$

Then, we have

$$\begin{aligned}
\mathbf{X}^{(t+1)}=&(\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}+\rho\mathbf{Y}^{(t+1)}-\mathbf{\Lambda}^{(t)})\frac{1}{\rho}\left(\mathbf{I}-\underbrace{\left(-\frac{(\mathbf{Y}^{(t+1)})^T\mathbf{Y}^{(t+1)}}{\rho}\right)}_{\triangleq\mathbf{T}}\right)^{-1}\\
\overset{(a)}{=}&(\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}+\rho\mathbf{Y}^{(t+1)}-\mathbf{\Lambda}^{(t)})\frac{1}{\rho}\sum_{i=0}^{\infty}\mathbf{T}^i \quad (17)
\end{aligned}$$

where (a) we use the geometric series and it is true due to the fact that $\|(\mathbf{Y}^{(t+1)})^T\mathbf{Y}^{(t+1)}\|_F\le NK\tau^2$ and $\rho\ge 8NK\tau^2$ which yield $|\lambda_{\max}(\mathbf{T})|\le\|\mathbf{T}\|_F\le 1/8<1$.

Applying matrix triangle inequality of both sides of (17), we have

$$\begin{aligned}\|\mathbf{X}^{(t+1)}\|_F &\overset{(a)}{\leq} \frac{8}{7\rho}\left((\|\mathbf{Z}_2^{(t)}\|_F+\rho)\tau\sqrt{NK}+\|\mathbf{\Lambda}^{(t)}\|_F\right)\\ &\overset{(b)}{\leq} \frac{8}{7\rho}\left((\|\mathbf{Z}_2^{(t)}\|_F+\rho)\tau\sqrt{NK}+\|\mathbf{X}^{(t)}\|_F NK\tau^2+\|\mathbf{Z}_2^{(t)}\|_F\tau\sqrt{NK}\right) \qquad (18)\\ &= \frac{8}{7\rho}\left((2\|\mathbf{Z}_2^{(t)}\|_F+\rho)\tau\sqrt{NK}+\|\mathbf{X}^{(t)}\|_F NK\tau^2\right) \qquad (19)\end{aligned}$$

where (a) is true due to the fact that

$$\|\sum_{i=0}^{\infty}\mathbf{T}^i\|_F \leq \sum_{i=0}^{\infty}\|\mathbf{T}^i\|_F \leq \sum_{i=0}^{\infty}\|\mathbf{T}\|_F^i \leq \sum_{i=0}^{\infty}\left(\frac{1}{8}\right)^i = \frac{8}{7}; \qquad (20)$$

(b) is true because after substituting (6d) into (16), we have

$$\mathbf{\Lambda}^{(t+1)} = -(\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_2^{(t)})\mathbf{Y}^{(t+1)}, \qquad (21)$$

and further substituting (21) into (18) we know that

$$\begin{aligned}\|\mathbf{\Lambda}^{(t)}\|_F \leq \|\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T - \mathbf{Z}_2^{(t)}\|_F\|\mathbf{Y}^{(t)}\|_F &\leq (\|\mathbf{X}^{(t)}\|_F\|\mathbf{Y}^{(t)}\|_F + \|\mathbf{Z}_2^{(t)}\|_F)\|\mathbf{Y}^{(t)}\|_F\\ &\leq \tau\sqrt{NK}(\|\mathbf{X}^{(t)}\|_F\tau\sqrt{NK}+\|\mathbf{Z}_2^{(t)}\|_F). \qquad (22)\end{aligned}$$

In the following, we use the mathematical induction to prove the boundness of $\mathbf{X}$-iterate.

When $t = 1$, we have

$$\|\mathbf{X}^{(2)}\|_F \leq \begin{cases} \frac{8}{7\rho}\left((2\|\mathbf{Z}_2^{(1)}\|_F+\rho)\tau\sqrt{NK}+\|\mathbf{X}^{(1)}\|_F NK\tau^2\right), & \text{if } \|\mathbf{X}^{(2)}\|_F \leq \|\mathbf{X}^{(1)}\|_F,\\ \frac{8}{7\rho}\left((2\|\mathbf{Z}_2^{(1)}\|_F+\rho)\tau\sqrt{NK}+\|\mathbf{X}^{(2)}\|_F NK\tau^2\right), & \text{if } \|\mathbf{X}^{(1)}\|_F \leq \|\mathbf{X}^{(2)}\|_F.\end{cases}$$

For the case $\|\mathbf{X}^{(1)}\|_F \leq \|\mathbf{X}^{(2)}\|_F$, using (19) we can further have

$$\begin{aligned}\|\mathbf{X}^{(2)}\|_F &\leq \left(1-\frac{8N\tau}{7\rho}\right)^{-1}\frac{8\tau\sqrt{NK}}{7\rho}(2\|\mathbf{Z}_2^{(1)}\|_F+\rho)\\ &= \frac{8\tau\sqrt{NK}}{7\rho-8NK\tau^2}\left(2\|\mathbf{Z}_2^{(1)}\|_F+\rho\right)\\ &\overset{(a)}{\leq} \frac{1}{6\tau\sqrt{NK}}\left(2\|\mathbf{Z}_2^{(1)}\|_F+8NK\tau^2\right) \qquad (23)\end{aligned}$$

where (a) is due to $\rho \geq 8NK\tau^2$.

Thus, we have

$$\|\mathbf{X}^{(2)}\|_F \leq \begin{cases} \min\left\{\frac{1}{7\tau\sqrt{NK}}\left((2\|\mathbf{Z}_2^{(1)}\|_F+8NK\tau^2)+\|\mathbf{X}^{(1)}\|_F\tau\sqrt{NK}\right), \|\mathbf{X}^{(1)}\|_F\right\}, & \text{if } \|\mathbf{X}^{(2)}\|_F \leq \|\mathbf{X}^{(1)}\|_F,\\ \frac{1}{6\tau\sqrt{NK}}\left(2\|\mathbf{Z}_2^{(1)}\|_F+8NK\tau^2\right), & \text{if } \|\mathbf{X}^{(1)}\|_F \leq \|\mathbf{X}^{(2)}\|_F.\end{cases}$$

When $t \geq 2$, from (19) we know that

$$\|\mathbf{X}^{(t+1)}\|_F \leq \frac{8}{7\rho}\left((2\max\{\|\mathbf{Z}_2^{(l)}\|_F|l=1,\ldots,t\}+\rho)\tau\sqrt{NK}+\max\{\|\mathbf{X}^{(l)}\|_F|l=1,\ldots,t\}NK\tau^2\right),$$

which implies

$$\max\{\|\mathbf{X}^{(2)}\|_F,\ldots,\|\mathbf{X}^{(t+1)}\|_F\} \leq \frac{8}{7\rho}\left((2\max\{\|\mathbf{Z}_2^{(l)}\|_F|l=1,\ldots,t\}+\rho)\tau\sqrt{NK}+\max\{\|\mathbf{X}^{(1)}\|_F,\ldots,\|\mathbf{X}^{(t+1)}\|_F\}NK\tau^2\right). \qquad (24)$$

With $\zeta \triangleq \max\{\|\mathbf{X}^{(2)}\|_F,\ldots,\|\mathbf{X}^{(t+1)}\|_F\}$ and substituting ζ into (24), we can get

$$\zeta \leq \begin{cases} \min\left\{\frac{1}{7\tau\sqrt{NK}}\left((2\max\{\|\mathbf{Z}_2^{(l)}\|_F|l=1,\ldots,t\}+8NK\tau^2)+\|\mathbf{X}^{(1)}\|_F\tau\sqrt{NK}\right), \|\mathbf{X}^{(1)}\|_F\right\}, & \text{if } \zeta \leq \|\mathbf{X}^{(1)}\|_F,\\ \frac{1}{6\tau\sqrt{NK}}\left(2\max\{\|\mathbf{Z}_2^{(l)}\|_F|l=1,\ldots,t\}+8NK\tau^2\right), & \text{if } \|\mathbf{X}^{(1)}\|_F \leq \zeta.\end{cases}$$

According to A3, we have $\max\{\|\mathbf{Z}_2^{(l)}\|_F | l = 1, \ldots, t\} \leq \mathcal{Z}$. Further, we can arrive at

$$\|\mathbf{X}^{(t)}\|_F \leq \min\left\{\frac{1}{7\tau\sqrt{NK}}\left(2\mathcal{Z} + 8NK\tau^2\right), \frac{1}{6\tau\sqrt{NK}}\left(2\mathcal{Z} + 8NK\tau^2 + \|\mathbf{X}^{(1)}\|_F\tau\sqrt{NK}\right), \|\mathbf{X}^{(1)}\|_F\right\} \triangleq \gamma\sqrt{NK} \tag{25}$$

6.3 Proof of Lemma 3

Lemma 3 *Consider using the update rules* (6b) (6c) (6d) *to solve* (3)*. We have*

$$\|\mathbf{Y}^{(t+1)} - \mathbf{Y}_2^{(t+1)}\|_F \leq \frac{\gamma\sqrt{NK}}{\rho}\|\mathbf{Z}_1^{(t)} - \mathbf{Z}_2^{(t)}\|_F \tag{26}$$

where

$$\mathbf{Y}_2^{(t+1)} \triangleq \arg\min_{0\leq\mathbf{Y}\leq\tau} \widehat{\mathcal{L}}_{\mathbf{Y}}(\mathbf{X}^{(t)}, \mathbf{Y}; \mathbf{\Lambda}^{(t)}; \mathbf{Z}_2^{(t)}) \tag{27}$$

and $\mathbf{Y}^{(t+1)}$ *is defined in* (6b)*.*
Similarly, we have

$$\|\mathbf{X}^{(t+1)} - \mathbf{X}_1^{(t+1)}\|_F \leq \frac{\tau\sqrt{NK}}{\rho}\|\mathbf{Z}_1^{(t)} - \mathbf{Z}_2^{(t)}\|_F \tag{28}$$

where $\mathbf{X}_1^{(t+1)} \triangleq \arg\min_{\mathbf{X}} \widehat{\mathcal{L}}_{\mathbf{X}}(\mathbf{X}, \mathbf{Y}^{(t+1)}; \mathbf{\Lambda}^{(t)}; \mathbf{Z}_1^{(t)})$ *and* $\mathbf{X}^{(t+1)}$ *is defined in* (6c)*.*

Proof: According to the optimality condition of problem (6b) and (27), when $0 \leq \mathbf{Y} \leq \tau$, we have

$$\langle(\mathbf{X}^{(t)})^T(\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_1^{(t)}) - \rho(\mathbf{X}^{(t)} - \mathbf{Y}^{(t+1)} + \mathbf{\Lambda}^{(t)}/\rho)^T + \beta^{(t)}(\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T, (\mathbf{Y} - \mathbf{Y}^{(t+1)})^T\rangle \geq 0, \tag{29}$$

and

$$\langle(\mathbf{X}^{(t)})^T(\mathbf{X}^{(t)}(\mathbf{Y}_2^{(t+1)})^T - \mathbf{Z}_2^{(t)}) - \rho(\mathbf{X}^{(t)} - \mathbf{Y}_2^{(t+1)} + \mathbf{\Lambda}^{(t)}/\rho)^T + \beta^{(t)}(\mathbf{Y}_2^{(t+1)} - \mathbf{Y}^{(t)})^T, (\mathbf{Y} - \mathbf{Y}_2^{(t+1)})^T\rangle \geq 0. \tag{30}$$

Letting $\mathbf{Y} = \mathbf{Y}_2^{(t+1)}$ in (29) and $\mathbf{Y} = \mathbf{Y}^{(t+1)}$ in (30), we have

$$\begin{aligned}\langle(\mathbf{X}^{(t)})^T(\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_1^{(t)}), (\mathbf{Y}_2^{(t+1)} - \mathbf{Y}^{(t+1)})^T\rangle \geq \\ \langle\rho(\mathbf{X}^{(t)} - \mathbf{Y}^{(t+1)} + \mathbf{\Lambda}^{(t)}/\rho)^T - \beta^{(t)}(\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T, (\mathbf{Y}_2^{(t+1)} - \mathbf{Y}^{(t+1)})^T\rangle,\end{aligned} \tag{31}$$

and

$$\begin{aligned}\langle(\mathbf{X}^{(t)})^T(\mathbf{X}^{(t)}(\mathbf{Y}_2^{(t+1)})^T - \mathbf{Z}_2^{(t)}), (\mathbf{Y}^{(t+1)} - \mathbf{Y}_2^{(t+1)})^T\rangle \geq \\ \langle\rho(\mathbf{X}^{(t)} - \mathbf{Y}_2^{(t+1)} + \mathbf{\Lambda}^{(t)}/\rho)^T - \beta^{(t)}(\mathbf{Y}_2^{(t+1)} - \mathbf{Y}^{(t)})^T, (\mathbf{Y}^{(t+1)} - \mathbf{Y}_2^{(t+1)})^T\rangle.\end{aligned} \tag{32}$$

Adding (31) and (32), we can have

$$\langle(\mathbf{X}^{(t)})^T(\mathbf{Z}_2^{(t)} - \mathbf{Z}_1^{(t)}), (\mathbf{Y}_2^{(t+1)} - \mathbf{Y}^{(t+1)})^T\rangle \geq \|\mathbf{X}^{(t)}(\mathbf{Y}_2^{(t+1)} - \mathbf{Y}^{(t+1)})^T\|_F^2 + (\rho + \beta^{(t)})\|\mathbf{Y}_2^{(t+1)} - \mathbf{Y}^{(t+1)}\|_F^2. \tag{33}$$

According to Cauchy-Schwarz inequality, we know that

$$\|\mathbf{X}^{(t)}\|_F\|\mathbf{Z}_2^{(t)} - \mathbf{Z}_1^{(t)}\|_F\|\mathbf{Y}_2^{(t+1)} - \mathbf{Y}^{(t+1)}\|_F \geq \langle(\mathbf{X}^{(t)})^T(\mathbf{Z}_2^{(t)} - \mathbf{Z}_1^{(t)}), (\mathbf{Y}_2^{(t+1)} - \mathbf{Y}^{(t+1)})^T\rangle. \tag{34}$$

Combining with (33), we can obtain

$$\|\mathbf{Z}_2^{(t)} - \mathbf{Z}_1^{(t)}\|_F \geq \frac{\rho + \beta^{(t)}}{\|\mathbf{X}^{(t)}\|_F}\|\mathbf{Y}_2^{(t+1)} - \mathbf{Y}^{(t+1)}\|_F \geq \frac{\rho}{\|\mathbf{X}^{(t)}\|_F}\|\mathbf{Y}_2^{(t+1)} - \mathbf{Y}^{(t+1)}\|_F. \tag{35}$$

Applying Lemma 2, we have

$$\|\mathbf{Y}^{(t+1)} - \mathbf{Y}_2^{(t+1)}\|_F \leq \frac{\gamma\sqrt{NK}}{\rho}\|\mathbf{Z}_1^{(t)} - \mathbf{Z}_2^{(t)}\|_F. \tag{36}$$

Similar result is also applied to $\mathbf{X}$-iterate.

6.4 Proof of Lemma 4

Lemma 4 *Consider using the update rules* (6b) (6c) (6d) *to solve* (3)*. Then we have*

$$\|\mathbf{\Lambda}^{(t+1)}-\mathbf{\Lambda}^{(t)}\|_F^2 \le 4NK\tau^2\|\mathbf{Z}_2^{(t)}-\mathbf{Z}_2^{(t-1)}\|_F^2+4(NK)^2\tau^4\|\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)}\|_F^2 \\ +4\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T-\mathbf{Z}_2^{(t-1)}\|_F^2\|\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)}\|_F^2+4NK\tau^2\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})^T\|_F^2. \quad (37)$$

Proof:

According to (21), we have

$$\begin{aligned}
&\mathbf{\Lambda}^{(t+1)}-\mathbf{\Lambda}^{(t)}\\
=&-(\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T\mathbf{Y}^{(t+1)}-\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T\mathbf{Y}^{(t)}-(\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}-\mathbf{Z}_2^{(t-1)}\mathbf{Y}^{(t)})) \quad (38)\\
=&-\big(\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T\mathbf{Y}^{(t+1)}-\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T\mathbf{Y}^{(t)}-(\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}-\mathbf{Z}_2^{(t-1)}\mathbf{Y}^{(t)})\\
&+\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)})^T\mathbf{Y}^{(t+1)}-\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)})^T\mathbf{Y}^{(t+1)}\big)\\
=&-\big((\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)})(\mathbf{Y}^{(t+1)})^T\mathbf{Y}^{(t+1)}+\mathbf{X}^{(t)}\big((\mathbf{Y}^{(t+1)})^T\mathbf{Y}^{(t+1)}-(\mathbf{Y}^{(t)})^T\mathbf{Y}^{(t)}\big)-(\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}-\mathbf{Z}_2^{(t-1)}\mathbf{Y}^{(t)})\big)\\
=&(\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}-\mathbf{Z}_2^{(t-1)}\mathbf{Y}^{(t)})-(\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)})(\mathbf{Y}^{(t+1)})^T\mathbf{Y}^{(t+1)}\\
&-\underbrace{\frac{1}{2}\big(\mathbf{X}^{(t)}\big((\mathbf{Y}^{(t+1)}+\mathbf{Y}^{(t)})^T(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})+(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})^T(\mathbf{Y}^{(t+1)}+\mathbf{Y}^{(t)})\big)\big)}_{\triangleq\mathcal{Q}}. \quad (39)
\end{aligned}$$

Note that the following is true

$$\begin{aligned}
\mathcal{Q}=&\frac{1}{2}\big(\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})^T(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})+2\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})\big)\\
&+\frac{1}{2}\mathbf{X}^{t}(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})^T(\mathbf{Y}^{(t+1)}+\mathbf{Y}^{(t)})\\
=&\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})+\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})^T\mathbf{Y}^{(t+1)}. \quad (40)
\end{aligned}$$

Plugging (40) into (39), we have

$$\begin{aligned}
&\mathbf{\Lambda}^{(t+1)}-\mathbf{\Lambda}^{(t)}\\
=&(\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}-\mathbf{Z}_2^{(t-1)}\mathbf{Y}^{(t)})-(\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)})(\mathbf{Y}^{(t+1)})^T\mathbf{Y}^{(t+1)}-\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})\\
&-\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})\mathbf{Y}^{(t+1)}\\
=&(\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}-\mathbf{Z}_2^{(t-1)}\mathbf{Y}^{(t+1)}+\mathbf{Z}_2^{(t-1)}\mathbf{Y}^{(t+1)}-\mathbf{Z}_2^{(t-1)}\mathbf{Y}^{(t)})-(\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)})(\mathbf{Y}^{(t+1)})^T\mathbf{Y}^{(t+1)}\\
&-\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})-\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})\mathbf{Y}^{(t+1)}\\
=&(\mathbf{Z}_2^{(t)}-\mathbf{Z}_2^{(t-1)})\mathbf{Y}^{(t+1)}+(\mathbf{Z}_2^{(t-1)}-\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T)(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})-(\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)})(\mathbf{Y}^{(t+1)})^T\mathbf{Y}^{(t+1)}\\
&-\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})^T\mathbf{Y}^{(t+1)}. \quad (41)
\end{aligned}$$

Applying triangle inequality on both sides of (41), we have

$$\|\mathbf{\Lambda}^{(t+1)}-\mathbf{\Lambda}^{(t)}\|_F \le \|\mathbf{Z}_2^{(t)}-\mathbf{Z}_2^{(t-1)}\|_F\|\mathbf{Y}^{(t+1)}\|_F+\|\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)}\|_F\|(\mathbf{Y}^{(t+1)})^T\mathbf{Y}^{(t+1)}\|_F \\ +\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T-\mathbf{Z}_2^{(t-1)}\|_F\|\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)}\|_F+\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})^T\|_F\|\mathbf{Y}^{(t+1)}\|_F. \quad (42)$$

Squaring both sides of (42), we obtain

$$\|\mathbf{\Lambda}^{(t+1)}-\mathbf{\Lambda}^{(t)}\|_F^2 \overset{(a)}{\le} 4NK\tau^2\|\mathbf{Z}_2^{(t)}-\mathbf{Z}_2^{(t-1)}\|_F^2+4(NK)^2\tau^4\|\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)}\|_F^2 \\ +4\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T-\mathbf{Z}_2^{(t-1)}\|_F^2\|\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)}\|_F^2+4NK\tau^2\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})^T\|_F^2. \quad (43)$$

where (a) is true due to $\mathbf{Y}\le\tau$, i.e., $\|\mathbf{Y}\|_F\le\tau\sqrt{NK}$. The claim is proved.

6.5 Proof of Lemma 5

Lemma 5 *Consider using the update rules* (6b) (6c) (6d)*. If*

$$\rho > 8NK\tau^2 \quad \textit{and} \quad \beta^{(t)} > \frac{8}{\rho}\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T - \mathbf{Z}_2^{(t-1)}\|_F^2 - \rho, \tag{44}$$

are satisfied, we have

$$\begin{aligned}
&\mathcal{L}(\mathbf{X}^{(t+1)}, \mathbf{Y}^{(t+1)}, \mathbf{\Lambda}^{(t+1)}) - \mathcal{L}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)}) \\
&\le -c_1\|\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)}\|_F^2 - c_2\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T\|_F^2 - c_3\|\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)}\|_F^2 \\
&\quad + \langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}_o^{(t)} - \mathbf{Y}_o^{(t+1)})^T\rangle + \langle\delta\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}, (\mathbf{X}_o^{(t)} - \mathbf{X}_o^{(t+1)})\rangle \\
&\quad + \frac{4NK\tau^2}{\rho}\|\mathbf{Z}_2^{(t)} - \mathbf{Z}_2^{(t-1)}\|_F^2 + \frac{2NK}{\rho}(\gamma^2\|\delta\mathbf{Z}_1^{(t)}\|_F^2 + \tau^2\|\delta\mathbf{Z}_2^{(t)}\|_F^2)
\end{aligned}$$

where $c_1, c_2, c_3 > 0$ *are some positive constants,* $\delta\mathbf{Z}_1^{(t)} \triangleq \overline{\mathbf{Z}} - \mathbf{Z}_1^{(t)}$*,* $\delta\mathbf{Z}_2^{(t)} \triangleq \overline{\mathbf{Z}} - \mathbf{Z}_2^{(t)}$*, and*

$$\mathbf{Y}_o^{(t+1)} \triangleq \arg\min_{0\le\mathbf{Y}\le\tau} \widehat{\mathcal{L}}_{\mathbf{Y}}(\mathbf{X}^{(t)}, \mathbf{Y}; \mathbf{\Lambda}^{(t)}; \overline{\mathbf{Z}}), \tag{45}$$

$$\mathbf{X}_o^{(t+1)} \triangleq \arg\min_{\mathbf{X}} \widehat{\mathcal{L}}_{\mathbf{Y}}(\mathbf{X}, \mathbf{Y}^{(t+1)}; \mathbf{\Lambda}^{(t)}; \overline{\mathbf{Z}}). \tag{46}$$

Proof: We have the following descent estimate

$$\begin{aligned}
&\mathcal{L}(\mathbf{X}^{(t+1)}, \mathbf{Y}^{(t+1)}, \mathbf{\Lambda}^{(t+1)}) - \mathcal{L}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)}) \\
=&\underbrace{\mathcal{L}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t+1)}, \mathbf{\Lambda}^{(t)}) - \mathcal{L}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)})}_{\triangleq\mathcal{A}} + \underbrace{\mathcal{L}(\mathbf{X}^{(t+1)}, \mathbf{Y}^{(t+1)}, \mathbf{\Lambda}^{(t)}) - \mathcal{L}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t+1)}, \mathbf{\Lambda}^{(t)})}_{\triangleq\mathcal{B}} \qquad (47)\\
&+ \underbrace{\mathcal{L}(\mathbf{X}^{(t+1)}, \mathbf{Y}^{(t+1)}, \mathbf{\Lambda}^{(t+1)}) - \mathcal{L}(\mathbf{X}^{(t+1)}, \mathbf{Y}^{(t+1)}, \mathbf{\Lambda}^{(t)})}_{\triangleq\mathcal{C}} \\
\le&\underbrace{\widehat{\mathcal{L}}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t+1)}, \mathbf{\Lambda}^{(t)}) - \mathcal{L}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)})}_{\triangleq\widehat{\mathcal{A}}} + \mathcal{B} + \mathcal{C} \qquad (48)
\end{aligned}$$

where

$$\widehat{\mathcal{L}}(\mathbf{X}^{(t)}, \mathbf{Y}, \mathbf{\Lambda}^{(t)}) = \frac{1}{2}\|\mathbf{X}^{(t)}\mathbf{Y}^T - \overline{\mathbf{Z}}\|_F^2 + \frac{\rho}{2}\|\mathbf{X}^{(t)} - \mathbf{Y} + \mathbf{\Lambda}^{(t)}/\rho\|_F^2 + \frac{\beta^{(t)}}{2}\|\mathbf{Y} - \mathbf{Y}^{(t)}\|_F^2, \tag{49}$$

and

$$\begin{aligned}
\widehat{\mathcal{A}} =& \frac{1}{2}\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)})^T - \overline{\mathbf{Z}}\|_F^2 - \frac{1}{2}\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T - \overline{\mathbf{Z}}\|_F^2 \\
&+ \frac{\rho}{2}\|\mathbf{X}^{(t)} - \mathbf{Y}^{(t+1)} + \mathbf{\Lambda}^{(t)}/\rho\|_F^2 - \frac{\rho}{2}\|\mathbf{X}^{(t)} - \mathbf{Y}^{(t)} + \mathbf{\Lambda}^{(t)}/\rho\|_F^2 + \frac{\beta^{(t)}}{2}\|\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)}\|_F^2 \\
\overset{(a)}{=}& \langle((\mathbf{X}^{(t)})^T\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)})^T - \overline{\mathbf{Z}}), (\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T\rangle - \frac{1}{2}\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T\|_F^2 \\
&- \rho\langle(\mathbf{X}^{(t)} - \mathbf{Y}^{(t+1)} + \mathbf{\Lambda}^{(t)}/\rho)^T, (\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T\rangle - \frac{\rho}{2}\|\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)}\|_F^2 + \frac{\beta^{(t)}}{2}\|\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)}\|_F^2
\end{aligned}$$

where (a) is due to the fact that Taylor expansion for quadratic problems is exact.

According to the optimality condition for problem (6b), we know that

$$\langle(\mathbf{X}^{(t)})^T(\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_1^{(t)}) - \rho(\mathbf{X}^{(t)} - \mathbf{Y}^{(t+1)} + \mathbf{\Lambda}^{(t)}/\rho)^T + \beta^{(t)}(\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T, (\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T\rangle \le 0. \tag{50}$$

Substituting $\delta\mathbf{Z}_1^{(t)} = \overline{\mathbf{Z}} - \mathbf{Z}_1^{(t)}$ into (50), we have

$$\begin{aligned}
&\langle(\mathbf{X}^{(t)})^T(\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)})^T - \overline{\mathbf{Z}}) - \rho(\mathbf{X}^{(t)} - \mathbf{Y}^{(t+1)} + \mathbf{\Lambda}^{(t)}/\rho)^T + \beta^{(t)}(\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T, (\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T\rangle \\
&\qquad \le -\langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T\rangle \qquad (51)
\end{aligned}$$

such that $\widehat{\mathcal{A}}$ can be further bounded as follows,

$$\begin{aligned}\widehat{\mathcal{A}} \leq& -\frac{1}{2}\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})^T\|_F^2 - \frac{\rho+\beta^{(t)}}{2}\|\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)}\|_F^2 + \langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}^{(t)}-\mathbf{Y}^{(t+1)})^T\rangle \\ \overset{(a)}{\leq}& -\frac{1}{2}\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})^T\|_F^2 - \frac{\rho+\beta^{(t)}}{2}\|\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)}\|_F^2 \\ &+ \langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}_o^{(t)}-\mathbf{Y}_o^{(t+1)})^T\rangle + \frac{2NK\gamma^2}{\rho}\|\delta\mathbf{Z}_1^{(t)}\|_F^2 \end{aligned} \tag{52}$$

where (a) is due to the fact that

$$\begin{aligned}&\langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}^{(t)}-\mathbf{Y}^{(t+1)})^T\rangle \\ =&\langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}^{(t)}-\mathbf{Y}^{(t+1)})^T - (\mathbf{Y}_o^{(t)}-\mathbf{Y}_o^{(t+1)})^T + (\mathbf{Y}_o^{(t)}-\mathbf{Y}_o^{(t+1)})^T\rangle \\ =&\langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}_o^{(t)}-\mathbf{Y}_o^{(t+1)})^T\rangle + \langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}^{(t)}-\mathbf{Y}^{(t+1)})^T - (\mathbf{Y}_o^{(t)}-\mathbf{Y}_o^{(t+1)})^T\rangle \end{aligned} \tag{53}$$

$$\begin{aligned} \overset{(a.1)}{\leq}&\langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}_o^{(t)}-\mathbf{Y}_o^{(t+1)})^T\rangle + \|\mathbf{X}^{(t)}\|_F\|\delta\mathbf{Z}_1^{(t)}\|_F(\|\mathbf{Y}^{(t)}-\mathbf{Y}_o^{(t)}\|_F + \|\mathbf{Y}^{(t+1)}-\mathbf{Y}_o^{(t+1)}\|_F) \\ \overset{(a.2)}{\leq}&\langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}_o^{(t)}-\mathbf{Y}_o^{(t+1)})^T\rangle + \frac{2\|\mathbf{X}^{(t)}\|_F\gamma\sqrt{NK}}{\rho}\|\delta\mathbf{Z}_1^{(t)}\|_F^2 \\ \overset{(a.3)}{\leq}&\langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}_o^{(t)}-\mathbf{Y}_o^{(t+1)})^T\rangle + \frac{2NK\gamma^2}{\rho}\|\delta\mathbf{Z}_1^{(t)}\|_F^2 \end{aligned} \tag{54}$$

where

$(a.1)$ is true according to Cauchy-Schwarz inequality;

$(a.2)$ we apply Lemma 3;

$(a.3)$ we apply Lemma 2.

Similarly, we have

$$\mathcal{B} \leq -\frac{1}{2}\|(\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)})(\mathbf{Y}^{(t+1)})^T\|_F^2 - \frac{\rho}{2}\|\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)}\|_F^2 - \langle\delta\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}, \mathbf{X}^{(t+1)}-\mathbf{X}^{(t)}\rangle \tag{55}$$

$$\begin{aligned}&\leq -\frac{1}{2}\|(\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)})(\mathbf{Y}^{(t+1)})^T\|_F^2 - \frac{\rho}{2}\|\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)}\|_F^2 \\ &\quad + \langle\delta\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}, (\mathbf{X}_o^{(t)}-\mathbf{X}_o^{(t+1)})\rangle + \frac{2NK\tau^2}{\rho}\|\delta\mathbf{Z}_2^{(t)}\|_F^2, \end{aligned} \tag{56}$$

$$\begin{aligned}\mathcal{C} =& \langle\mathbf{X}^{(t+1)}-\mathbf{Y}^{(t+1)}, \mathbf{\Lambda}^{(t+1)}-\mathbf{\Lambda}^{(t)}\rangle \\ \overset{(a)}{=}& \frac{1}{\rho}\|\mathbf{\Lambda}^{(t+1)}-\mathbf{\Lambda}^{(t)}\|_F^2 \end{aligned} \tag{57}$$

where (a) is according to (6d).

Substituting the result of Lemma 4 into (57), from (48) we can obtain

$$\begin{aligned}&\mathcal{L}(\mathbf{X}^{(t+1)}, \mathbf{Y}^{(t+1)}, \mathbf{\Lambda}^{(t+1)}) - \mathcal{L}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)}) \\ &\leq -\left(\frac{\rho}{2} - \frac{4N^2K^2\tau^4}{\rho}\right)\|\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)}\|_F^2 - \left(\frac{1}{2} - \frac{4NK\tau^2}{\rho}\right)\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})^T\|_F^2 \\ &- \left(\frac{\rho}{2} + \frac{\beta^{(t)}}{2} - \frac{4\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T - \mathbf{Z}_2^{(t-1)}\|_F^2}{\rho}\right)\|\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)}\|_F^2 - \frac{1}{2}\|(\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)})(\mathbf{Y}^{(t+1)})^T\|_F^2. \\ &+ \langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}_o^{(t)}-\mathbf{Y}_o^{(t+1)})^T\rangle + \langle\delta\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}, (\mathbf{X}_o^{(t)}-\mathbf{X}_o^{(t+1)})\rangle \\ &+ \frac{4NK\tau^2}{\rho}\|\mathbf{Z}_2^{(t)}-\mathbf{Z}_2^{(t-1)}\|_F^2 + \frac{2NK}{\rho}\left(\gamma^2\|\delta\mathbf{Z}_1^{(t)}\|_F^2 + \tau^2\|\delta\mathbf{Z}_2^{(t)}\|_F^2\right). \end{aligned} \tag{58}$$

Therefore, according to (58) if

$$c_1 = \frac{\rho}{2} - \frac{4N^2K^2\tau^4}{\rho} > 0, \tag{59a}$$

$$c_2 = \frac{1}{2} - \frac{4NK\tau^2}{\rho} > 0, \tag{59b}$$

$$c_3 = \frac{\rho+\beta^{(t)}}{2} - \frac{4\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T - \mathbf{Z}_2^{(t-1)}\|_F^2}{\rho} > 0, \tag{59c}$$

which are equivalent to

$$\rho > 8NK\tau^2 \quad \text{and} \quad \beta^{(t)} > \frac{8\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T - \mathbf{Z}_2^{(t-1)}\|_F^2 - \rho^2}{\rho}, \tag{60}$$

we can have the successive difference of the augmented Lagrangian between two iterations as follows

$$\begin{aligned}\mathcal{L}(\mathbf{X}^{(t+1)}, \mathbf{Y}^{(t+1)}, \mathbf{\Lambda}^{(t+1)}) \le& \mathcal{L}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)}) - c_1\|\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)}\|_F^2 - c_2\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T\|_F^2 \\ &- c_3\|\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)}\|_F^2 + \langle (\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}_o^{(t)} - \mathbf{Y}_o^{(t+1)})^T\rangle + \langle \delta\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}, (\mathbf{X}_o^{(t)} - \mathbf{X}_o^{(t+1)})\rangle \\ &+ \frac{4NK\tau^2}{\rho}\|\mathbf{Z}_2^{(t)} - \mathbf{Z}_2^{(t-1)}\|_F^2 + \frac{2NK}{\rho}\left(\gamma^2\|\delta\mathbf{Z}_1^{(t)}\|_F^2 + \tau^2\|\delta\mathbf{Z}_2^{(t)}\|_F^2\right)\end{aligned} \tag{61}$$

where $c_1, c_2, c_3 > 0$.

6.6 Proof of Lemma 6

Lemma 6 *Consider using the update rules* (6b) (6c) (6d)*. If $\rho \ge NK\tau^2$ is satisfied, we have*

$$\mathcal{L}(\mathbf{X}^{(t+1)}, \mathbf{Y}^{(t+1)}, \mathbf{\Lambda}^{(t+1)}) \ge -((NK\gamma\tau)^2 + \|\overline{\mathbf{Z}}\|_F^2 + \|\delta\mathbf{Z}_2^{(t)}\|_F^2). \tag{62}$$

Proof: At iteration $t+1$, the augmented Lagrangian can be lower bounded as

$$\begin{aligned}&\mathcal{L}(\mathbf{X}^{(t+1)}, \mathbf{Y}^{(t+1)}, \mathbf{\Lambda}^{(t+1)}) \\ \overset{(a)}{=}&\frac{1}{2}\|\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \overline{\mathbf{Z}}\|_F^2 + \langle \mathbf{X}^{(t+1)} - \mathbf{Y}^{(t+1)}, \mathbf{\Lambda}^{(t+1)}\rangle + \frac{\rho}{2}\|\mathbf{X}^{(t+1)} - \mathbf{Y}^{(t+1)}\|_F^2 \\ \overset{(b)}{=}&\frac{1}{2}\|\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \overline{\mathbf{Z}}\|_F^2 + \langle \mathbf{X}^{(t+1)} - \mathbf{Y}^{(t+1)}, -(\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_2^{(t)})\mathbf{Y}^{(t+1)}\rangle \\ &+ \frac{\rho}{2}\|\mathbf{X}^{(t+1)} - \mathbf{Y}^{(t+1)}\|_F^2 \\ \overset{(c)}{\ge}&\frac{1}{2}(\rho - NK\tau^2)\|\mathbf{X}^{(t+1)} - \mathbf{Y}^{(t+1)}\|_F^2 - \frac{1}{2}\|\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \overline{\mathbf{Z}})\|_F^2 - \|\delta\mathbf{Z}_2^{(t)})\|_F^2 \end{aligned} \tag{63}$$

$$\begin{aligned}\overset{(d)}{\ge}& - \frac{1}{2}\|\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \overline{\mathbf{Z}})\|_F^2 - \|\delta\mathbf{Z}_2^{(t)})\|_F^2 \\ \overset{(e)}{\ge}& - (NK\gamma\tau)^2 - \|\overline{\mathbf{Z}}\|_F^2 - \|\delta\mathbf{Z}_2^{(t)})\|_F^2\end{aligned} \tag{64}$$

where (a) is according to (5); (b) is due to (21); (c) is true because the fact that

$$\begin{aligned}0 \le&\|(\mathbf{X}^{(t+1)} - \mathbf{Y}^{(t+1)})(\mathbf{Y}^{(t+1)})^T - (\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_2^{(t)})\|_F^2 \\ =&\|(\mathbf{X}^{(t+1)} - \mathbf{Y}^{(t+1)})(\mathbf{Y}^{(t+1)})^T\|_F^2 - 2\langle (\mathbf{Y}^{(t+1)})^T(\mathbf{X}^{(t+1)} - \mathbf{Y}^{(t+1)}), \mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_2^{(t)}\rangle \\ &+ \|\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_2^{(t)})\|_F^2 \\ \le&\|(\mathbf{X}^{(t+1)} - \mathbf{Y}^{(t+1)})(\mathbf{Y}^{(t+1)})^T\|_F^2 - 2\langle (\mathbf{Y}^{(t+1)})^T(\mathbf{X}^{(t+1)} - \mathbf{Y}^{(t+1)}), \mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_2^{(t)}\rangle \\ &+ 2\|\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \overline{\mathbf{Z}})\|_F^2 + 2\|\delta\mathbf{Z}_2^{(t)}\|_F^2,\end{aligned}$$

and $\|\mathbf{Y}\|_F^2 \le NK\tau^2$; (d) is true because we chose $\rho \ge NK\tau^2$, and (e) is due to the fact

$$\|\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \overline{\mathbf{Z}}\|_F^2 \le 2\|\mathbf{X}^{(t+1)}(\mathbf{Y}^{t+1})^T\|_F^2 + 2\|\overline{\mathbf{Z}}\|_F^2 \tag{65}$$

and boundness of $\mathbf{X}$ and $\mathbf{Y}$.

From (63) and (64), we know that if $\rho \ge NK\tau^2$, we have $\mathcal{L}(\mathbf{X}^{(t+1)}, \mathbf{Y}^{(t+1)}, \mathbf{\Lambda}^{(t+1)}) \ge -((NK\gamma\tau)^2 + \|\overline{\mathbf{Z}}\|_F^2 + \|\delta\mathbf{Z}_2^{(t)}\|_F^2)$.

6.7 Proof of Lemma 7

Lemma 7 *Consider using the update rules* (6b) (6c) (6d) *to solve* (1)*. Then there exist some constants $\sigma_{\boldsymbol{p}}, \sigma_{\boldsymbol{d}} > 0$ such that*

$$\begin{aligned}\|\widetilde{\nabla}\mathcal{L}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)})\|_F^2 + \|\mathbf{X}^{(t)} - \mathbf{Y}^{(t)}\|_F^2 \le& (\sigma_{\boldsymbol{p}} + \sigma_{\boldsymbol{d}})\mathcal{G} \\ &+ 2\sigma_{\boldsymbol{z}_2}(\|\delta\mathbf{Z}_2^{(t)}\|_F^2 + \|\delta\mathbf{Z}_2^{(t-1)}\|_F^2) + 8NK\gamma^2(\|\delta\mathbf{Z}_1^{(t)}\|_F^2 + \|\delta\mathbf{Z}_1^{(t-1)}\|_F^2)\end{aligned} \tag{66}$$

where

$$\mathcal{G}^{(t)} \triangleq \|\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)}\|_F^2 + \|\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)}\|_F^2 + \|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T\|_F^2, \ \text{and}\ \sigma_{\boldsymbol{z}_2} \triangleq 20NK\tau^2 + 12NK\tau^2/\rho^2. \tag{67}$$

Proof: From the optimality condition of $\mathbf{Y}$ in (6b), we have

$$(\mathbf{Y}^{(t+1)})^T = \text{proj}_{\mathcal{Y}}\big[(\mathbf{Y}^{(t+1)})^T - ((\mathbf{X}^{(t)})^T(\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_1^{(t)}) \\ - \rho(\mathbf{X}^{(t)} - \mathbf{Y}^{(t+1)} + \mathbf{\Lambda}^{(t)}/\rho)^T + \beta^{(t)}(\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T)\big]. \quad (68)$$

Then, we have

$$\begin{aligned}
&\left\|(\mathbf{Y}^{(t)})^T - \text{proj}_{\mathcal{Y}}\left[(\mathbf{Y}^{(t)})^T - ((\mathbf{X}^{(t)})^T(\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T - \mathbf{Z}_1^{(t-1)}) - \rho(\mathbf{X}^{(t)} - \mathbf{Y}^{(t)} + \mathbf{\Lambda}^{(t)}/\rho)^T)\right]\right\|_F \\
=&\Big\|(\mathbf{Y}^{(t)})^T - (\mathbf{Y}^{(t+1)})^T + (\mathbf{Y}^{(t+1)})^T \\
&\quad - \text{proj}_{\mathcal{Y}}\left[(\mathbf{Y}^{(t)})^T - ((\mathbf{X}^{(t)})^T(\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T - \mathbf{Z}_1^{(t-1)}) - \rho(\mathbf{X}^{(t)} - \mathbf{Y}^{(t)} + \mathbf{\Lambda}^{(t)}/\rho)^T)\right]\Big\|_F \\
\overset{(a)}{\le}&\|\mathbf{Y}^{(t)} - \mathbf{Y}^{(t+1)}\|_F \\
&+ \Big\|\text{proj}_{\mathcal{Y}}\Big[(\mathbf{Y}^{(t+1)})^T - ((\mathbf{X}^{(t)})^T(\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_1^{(t)}) - \rho(\mathbf{X}^{(t)} - \mathbf{Y}^{(t+1)} + \mathbf{\Lambda}^{(t)}/\rho)^T \\
&+ \beta^{(t)}(\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T)\Big] \\
&- \text{proj}_{\mathcal{Y}}\left[(\mathbf{Y}^{(t)})^T - ((\mathbf{X}^{(t)})^T(\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T - \mathbf{Z}_1^{(t-1)}) - \rho(\mathbf{X}^{(t)} - \mathbf{Y}^{(t)} + \mathbf{\Lambda}^{(t)}/\rho)^T)\right]\Big\|_F \\
\overset{(b)}{\le}&(2+\rho)\|\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)}\|_F + \beta^{(t)}\|\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)}\|_F + \|(\mathbf{X}^{(t)})^T\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T\|_F \\
&+ \|(\mathbf{X}^{(t)})^T(\mathbf{Z}_1^{(t)} - \mathbf{Z}_1^{(t-1)})\|_F \\
\overset{(c)}{\le}&(2+\rho)\|\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)}\|_F + \beta^{(t)}\|\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)}\|_F + \gamma\sqrt{NK}\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T\|_F \\
&+ \gamma\sqrt{NK}\|(\mathbf{Z}_1^{(t)} - \mathbf{Z}_1^{(t-1)})\|_F
\end{aligned} \quad (69)$$

where $\text{proj}_{\mathcal{Y}}$ denotes the projection of $\mathbf{Y}$ to the feasible set;

(a) we use triangle inequality and (68);

(b) is true due to the nonexpansiveness of the projection operator;

(c) is because of Cauchy-Schwarz inequality and Lemma 2.

Similarly, we can bound the size of the gradient of the augmented Lagrangian with respect to $\mathbf{X}$ by the following series of inequalities

$$\begin{aligned}
&\|\nabla_{\mathbf{X}}\mathcal{L}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)})\|_F = \|(\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T - \mathbf{Z}_2^{(t)})\mathbf{Y}^{(t)} + \rho(\mathbf{X}^{(t)} - \mathbf{Y}^{(t)} + \mathbf{\Lambda}^{(t)}/\rho)\|_F \\
\overset{(a)}{=}&\Big\|(\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T - \mathbf{Z}_2^{(t)})\mathbf{Y}^{(t)} + \rho(\mathbf{X}^{(t)} - \mathbf{Y}^{(t)} + \mathbf{\Lambda}^{(t)}/\rho) - ((\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_2^{(t)})\mathbf{Y}^{(t+1)} \\
&\quad + \rho(\mathbf{X}^{(t+1)} - \mathbf{Y}^{(t+1)} + \mathbf{\Lambda}^{(t)}/\rho))\Big\|_F \\
\le&\|(\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T - \mathbf{Z}_2^{(t)})\mathbf{Y}^{(t)} - ((\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_2^{(t)})\mathbf{Y}^{(t+1)})\|_F \\
&\quad + \rho\|\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)}\|_F + \rho\|\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)}\|_F \\
\overset{(b)}{=}&\|\mathbf{\Lambda}^{(t+1)} - \mathbf{\Lambda}^{(t)}\|_F + \tau\sqrt{NK}\|\mathbf{Z}_2^{(t)} - \mathbf{Z}_2^{(t-1)}\|_F + \rho\|\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)}\|_F + \rho\|\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)}\|_F
\end{aligned} \quad (70)$$

where

(a) is from the optimality condition of the $\mathbf{X}$ subproblem (16);

(b) is true due to the fact that

$$\begin{aligned}
&\|(\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T - \mathbf{Z}_2^{(t)})\mathbf{Y}^{(t)} - ((\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_2^{(t)})\mathbf{Y}^{(t+1)}) + \mathbf{Z}_2^{(t-1)}\mathbf{Y}^{(t)} - \mathbf{Z}_2^{(t-1)}\mathbf{Y}^{(t)}\|_F \\
\overset{(c)}{\le}&\|\mathbf{\Lambda}^{(t+1)} - \mathbf{\Lambda}^{(t)}\|_F + \tau\sqrt{NK}\|\mathbf{Z}_2^{(t)} - \mathbf{Z}_2^{(t-1)}\|_F
\end{aligned} \quad (71)$$

where (c) we use (38) and triangle inequality.

Squaring both sides of (70) and applying Lemma 4, we have

$$\|\nabla_{\mathbf{X}}\mathcal{L}(\mathbf{X}^{(t)},\mathbf{Y}^{(t)},\mathbf{\Lambda}^{(t)})\|_F^2 \le 4(4N^2K^2\tau^4+\rho^2)\|\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)}\|_F^2 \\ + 4(4\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T-\mathbf{Z}_2^{(t-1)}\|_F^2+\rho^2)\|\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)}\|_F^2 + 16NK\tau^2\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})^T\|_F^2 \\ + 20NK\tau^2\|\mathbf{Z}_2^{(t)}-\mathbf{Z}_2^{(t-1)}\|_F^2. \tag{72}$$

Therefore, combining (69) and (72), there must exists a finite positive number σ_{p} such that

$$\begin{aligned}&\|\widetilde{\nabla}\mathcal{L}(\mathbf{X}^{(t)},\mathbf{Y}^{(t)},\mathbf{\Lambda}^{(t)})\|_F^2\\ \le&\sigma_{\mathsf{p}}^{(t)}\mathcal{G}^{(t)}+20NK\tau^2\|\mathbf{Z}_2^{(t)}-\mathbf{Z}_2^{(t-1)}\|_F^2+4NK\gamma^2\|\mathbf{Z}_1^{(t)}-\mathbf{Z}_1^{(t-1)}\|_F^2.\\ \le&\sigma_{\mathsf{p}}\mathcal{G}^{(t)}+20NK\tau^2\|\mathbf{Z}_2^{(t)}-\mathbf{Z}_2^{(t-1)}\|_F^2+4NK\gamma^2\|\mathbf{Z}_1^{(t)}-\mathbf{Z}_1^{(t-1)}\|_F^2\end{aligned} \tag{73}$$

where

$$\mathcal{G}^{(t)} \triangleq \|\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)}\|_F^2+\|\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)}\|_F^2+\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})^T\|_F^2. \tag{74}$$

In particular, we have

$$\begin{aligned}\sigma_{\mathsf{p}}^{(t)} \triangleq& \max\{4(4N^2K^2\tau^4+\rho^2),4(2+\rho)^2+4\rho^2)+4(\beta^{(t)})^2+16\beta^{(t)},4NK\gamma^2+16NK\tau^2\}\\ \overset{(a)}{\le}& \max\big\{4(4N^2K^2\tau^4+\rho^2),4(2+\rho)^2+4\rho^2)+8((NK\tau\gamma)^2+\mathcal{Z}^2)^2+16((NK\tau\gamma)^2+\mathcal{Z}^2),\\ &4NK\gamma^2+16NK\tau^2\big\}\triangleq\sigma_{\mathsf{p}}\end{aligned} \tag{75}$$

where (a) is true due to the fact that according to Cauchy-Schwarz inequality we have

$$\beta^{(t)}\le 2(\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t)})\|_F^2+\|\mathbf{Z}_2^{(t-1)}\|_F^2)\le 2(N\tau\|\mathbf{X}^{(t)}\|_F^2+\|\mathbf{Z}_2^{(t-1)}\|_F^2), \tag{76}$$

and based on Lemma 2 and A2, we can obtain

$$\beta^{(t)}\le 2((NK\tau\gamma)^2+\mathcal{Z}^2). \tag{77}$$

Also, we have

$$\begin{aligned}\|\mathbf{X}^{(t)}-\mathbf{Y}^{(t)}\|_F =&\|\mathbf{X}^{(t)}-\mathbf{X}^{(t+1)}+\mathbf{X}^{(t+1)}-\mathbf{Y}^{(t+1)}+\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)}\|_F\\ \le&\|\mathbf{X}^{(t)}-\mathbf{X}^{(t+1)}\|_F+\|\mathbf{X}^{(t+1)}-\mathbf{Y}^{(t+1)}\|_F+\|\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)}\|_F\\ \overset{(a)}{=}&\|\mathbf{X}^{(t)}-\mathbf{X}^{(t+1)}\|_F+\frac{1}{\rho}\|\mathbf{\Lambda}^{(t+1)}-\mathbf{\Lambda}^{(t)}\|_F+\|\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)}\|_F,\end{aligned} \tag{78}$$

where (a) is true due to (6d). According to Lemma 4, there exists a finite positive constant σ_{d} such that

$$\|\mathbf{X}^{(t)}-\mathbf{Y}^{(t)}\|_F^2\le\sigma_{\mathsf{d}}\mathcal{G}^{(t)}+12NK\tau^2/\rho^2\|\mathbf{Z}_2^{(t)}-\mathbf{Z}_2^{(t-1)}\|_F^2 \tag{79}$$

where $\sigma_{\mathsf{d}}\triangleq\max\{12N^2K^2\tau^4/\rho^2+3,24((NK\tau\gamma)^2+\mathcal{Z}^2)/\rho^2+3,12NK\tau^2/\rho^2\}$.

The inequalities (73) and (79) imply that

$$\begin{aligned}&\|\widetilde{\nabla}\mathcal{L}(\mathbf{X}^{(t)},\mathbf{Y}^{(t)},\mathbf{\Lambda}^{(t)})\|_F^2+\|\mathbf{X}^{(t)}-\mathbf{Y}^{(t)}\|_F^2\\ \le&(\sigma_{\mathsf{p}}+\sigma_{\mathsf{d}})\mathcal{G}^{(t)}+\underbrace{(20NK\tau^2+12NK\tau^2/\rho^2)}_{\triangleq\sigma_{\mathbf{z}_2}}\|\mathbf{Z}_2^{(t)}-\mathbf{Z}_2^{(t-1)}\|_F^2+4NK\gamma^2\|\mathbf{Z}_1^{(t)}-\mathbf{Z}_1^{(t-1)}\|_F^2\end{aligned} \tag{80}$$

$$\overset{(a)}{\le}(\sigma_{\mathsf{p}}+\sigma_{\mathsf{d}})\mathcal{G}^{(t)}+2\sigma_{\mathbf{z}_2}(\|\delta\mathbf{Z}_2^{(t)}\|_F^2+\|\delta\mathbf{Z}_2^{(t-1)}\|_F^2)+8NK\gamma^2(\|\delta\mathbf{Z}_1^{(t)}\|_F^2+\|\delta\mathbf{Z}_1^{(t-1)}\|_F^2) \tag{81}$$

where (a) we use triangle inequality, i.e.,

$$\|\mathbf{Z}_2^{(t)}-\overline{\mathbf{Z}}+\overline{\mathbf{Z}}-\mathbf{Z}_2^{(t-1)}\|_F\le\|\delta\mathbf{Z}_2^{(t)}\|_F+\|\delta\mathbf{Z}_2^{(t-1)}\|_F. \tag{82}$$

Squaring both sides of (82), we have

$$\|\mathbf{Z}_2^{(t)}-\mathbf{Z}_2^{(t-1)}\|_F^2\le 2\|\delta\mathbf{Z}_2^{(t)}\|_F^2+2\|\delta\mathbf{Z}_2^{(t-1)}\|_F^2. \tag{83}$$

Similar result is also applied to $\|\mathbf{Z}_1^{(t)}-\mathbf{Z}_1^{(t-1)}\|_F^2$.

6.8 Proof of Theorem 1

Proof: According to (58), there exists a constant $\sigma_{\mathsf{m}} \triangleq \min\{c_1, c_2, c_3\}$ such that

$$\mathcal{L}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)}) - \mathcal{L}(\mathbf{X}^{(t+1)}, \mathbf{Y}^{(t+1)}, \mathbf{\Lambda}^{(t+1)}) \geq \sigma_{\mathsf{m}}\mathcal{G}^{(t)} - \langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}_o^{(t)} - \mathbf{Y}_o^{(t+1)})^T\rangle$$
$$- \langle\delta\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}, (\mathbf{X}_o^{(t)} - \mathbf{X}_o^{(t+1)})\rangle - \frac{4NK\tau^2}{\rho}\|\mathbf{Z}_2^{(t)} - \mathbf{Z}_2^{(t-1)}\|_F^2 - \frac{2NK}{\rho}\left(\gamma^2\|\delta\mathbf{Z}_1^{(t)}\|_F^2 + \tau^2\|\delta\mathbf{Z}_2^{(t)}\|_F^2\right). \quad (84)$$

Combining (80) and (84), we have

$$\|\widetilde{\nabla}\mathcal{L}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)})\|_F^2 + \|\mathbf{X}^{(t)} - \mathbf{Y}^{(t)}\|_F^2$$
$$\leq \frac{\sigma_{\mathsf{p}} + \sigma_{\mathsf{d}}}{\sigma_{\mathsf{m}}}\Big(\mathcal{L}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)}) - \mathcal{L}(\mathbf{X}^{(t+1)}, \mathbf{Y}^{(t+1)}, \mathbf{\Lambda}^{(t+1)}) + \langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}_o^{(t)} - \mathbf{Y}_o^{(t+1)})^T\rangle$$
$$+ \langle\delta\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}, (\mathbf{X}_o^{(t)} - \mathbf{X}_o^{(t+1)})\rangle + \frac{4NK\tau^2}{\rho}\|\mathbf{Z}_2^{(t)} - \mathbf{Z}_2^{(t-1)}\|_F^2 + \frac{2NK}{\rho}(\gamma^2\|\delta\mathbf{Z}_1^{(t)}\|_F^2 + \tau^2\|\delta\mathbf{Z}_2^{(t)}\|_F^2)\Big)$$
$$+ 8NK\gamma^2(\|\delta\mathbf{Z}_1^{(t)}\|_F^2 + \|\delta\mathbf{Z}_1^{(t-1)}\|_F^2) + 2\sigma_{\mathbf{z}_2}(\|\delta\mathbf{Z}_2^{(t)}\|_F^2 + \|\delta\mathbf{Z}_2^{(t-1)}\|_F^2). \quad (85)$$

Summing both sides of (85) over $t = 1, \ldots, T$, we have

$$\sum_{t=1}^{T}\|\widetilde{\nabla}\mathcal{L}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)})\|_F^2 + \|\mathbf{X}^{(t)} - \mathbf{Y}^{(t)}\|_F^2$$
$$\leq \mathcal{C}\big(\mathcal{L}(\mathbf{X}^{(1)}, \mathbf{Y}^{(1)}, \mathbf{\Lambda}^{(1)}) - \mathcal{L}(\mathbf{X}^{(T+1)}, \mathbf{Y}^{(T+1)}, \mathbf{\Lambda}^{(T+1)})$$
$$+ \sum_{t=1}^{T}\mathcal{C}\big(\big\langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}_o^{(t)} - \mathbf{Y}_o^{(t+1)})^T\big\rangle + \big\langle\delta\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}, (\mathbf{X}_o^{(t)} - \mathbf{X}_o^{(t+1)})\big\rangle$$
$$+ \frac{8NK\tau^2}{\rho}(\|\delta\mathbf{Z}_2^{(t)}\|_F^2 + \|\delta\mathbf{Z}_2^{(t-1)}\|_F^2) + \frac{2NK}{\rho}(\gamma^2\|\delta\mathbf{Z}_1^{(t)}\|_F^2 + \tau^2\|\delta\mathbf{Z}_2^{(t)}\|_F^2)\big)$$
$$+ \sum_{t=1}^{T}8NK\gamma^2(\|\delta\mathbf{Z}_1^{(t)}\|_F^2 + \|\delta\mathbf{Z}_1^{(t-1)}\|_F^2) + 2\sigma_{\mathbf{z}_2}(\|\delta\mathbf{Z}_2^{(t)}\|_F^2 + \|\delta\mathbf{Z}_2^{(t-1)}\|_F^2) \quad (86)$$

where $\mathcal{C} \triangleq (\sigma_{\mathsf{p}} + \sigma_{\mathsf{d}})/\sigma_{\mathsf{m}}$, and the constant $\mathcal{C}$ is only dependent on τ, N and $\mathcal{Z}$.

Taking expectations with respect to $\mathbf{Z}^{(1,\ldots,T)} \triangleq (\mathbf{Z}_1^{(1)}, \mathbf{Z}_2^{(1)}, \ldots, \mathbf{Z}_1^{(T)}, \mathbf{Z}_2^{(T)})$ on both sides of (86), we have

$$\sum_{t=1}^{T}\mathbb{E}[\mathcal{P}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)})]$$
$$\overset{(a)}{\leq} \mathbb{E}\Big[\mathcal{C}\big(\mathcal{L}(\mathbf{X}^{(1)}, \mathbf{Y}^{(1)}, \mathbf{\Lambda}^{(1)}) + \mathcal{C}(N^2\gamma\tau + \|\bar{\mathbf{Z}}\|_F^2 + \|\delta\mathbf{Z}_2^{(T)}\|_F^2)$$
$$+ \mathcal{C}\sum_{t=1}^{T}\Big(\big\langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}_o^{(t)} - \mathbf{Y}_o^{(t+1)})^T\big\rangle + \big\langle\delta\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}, (\mathbf{X}_o^{(t)} - \mathbf{X}_o^{(t+1)})\big\rangle$$
$$+ \frac{8NK\tau^2}{\rho}(\|\delta\mathbf{Z}_2^{(t)}\|_F^2 + \|\delta\mathbf{Z}_2^{(t-1)}\|_F^2) + \frac{2NK}{\rho}(\gamma^2\|\delta\mathbf{Z}_1^{(t)}\|_F^2 + \tau^2\|\delta\mathbf{Z}_2^{(t)}\|_F^2)\Big)$$
$$+ \sum_{t=1}^{T}8NK\gamma^2(\|\delta\mathbf{Z}_1^{(t)}\|_F^2 + \|\delta\mathbf{Z}_1^{(t-1)}\|_F^2) + 2\sigma_{\mathbf{z}_2}(\|\delta\mathbf{Z}_2^{(t)}\|_F^2 + \|\delta\mathbf{Z}_2^{(t-1)}\|_F^2)\Big] \quad (87)$$

where (a) is true due to Lemma 6.

Based on the independence of the sequence $\mathbf{Z}^{(1,\ldots,T)}$ over t, we have the following results: i) since $\mathbf{X}^{(t)}$ is only dependent on $\mathbf{Z}_2^{(t-1)}$, we have $\mathbb{E}_{\mathbf{Z}_1^{(t)}}[\langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}_o^{(t)} - \mathbf{Y}_o^{(t+1)})^T\rangle|\mathbf{Z}^{(1,\ldots,t-1)}] = 0, \forall t$; ii) similarly $\mathbf{Y}^{(t+1)}$ is dependent on $\mathbf{Z}_1^{(t)}$ and $\mathbf{Z}_2^{(t-1)}$, we have $\mathbb{E}_{\mathbf{Z}_2^{(t)}}[\langle\delta\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}, (\mathbf{X}_o^{(t)} - \mathbf{X}_o^{(t+1)})\rangle|\mathbf{Z}^{(1,\ldots,t-1)}] = 0, \forall t$; iii) according to A2 and Table 1, we have

$$\sum_{t=1}^{T}\mathbb{E}[\mathcal{P}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)})] \leq \mathcal{C}\mathcal{L}(\mathbf{X}^{(1)}, \mathbf{Y}^{(1)}, \mathbf{\Lambda}^{(1)}) + \mathcal{C}((NK\gamma\tau)^2 + \|\bar{\mathbf{Z}}\|_F^2 + \frac{\sigma^2}{L})$$
$$+ \frac{2\mathcal{C}TNK(\gamma^2 + 5\tau^2)\sigma^2}{\rho L} + 16NKT\gamma^2\frac{\sigma^2}{L} + 4T\sigma_{\mathbf{z}_2}\frac{\sigma^2}{L}. \quad (88)$$

Further, from the definition of the output, we have

$$\mathbb{E}[\mathcal{P}(\mathbf{X}^{(r)},\mathbf{Y}^{(r)},\mathbf{\Lambda}^{(r)})] \overset{(a)}{\leq} \frac{1}{T}\mathcal{C}\left(\mathcal{L}(\mathbf{X}^{(1)},\mathbf{Y}^{(1)},\mathbf{\Lambda}^{(1)}) + (NK\gamma\tau)^2 + \|\bar{\mathbf{Z}}\|_F^2 + \frac{\sigma^2}{L}\right) + \left(\frac{\mathcal{C}(\gamma^2+5\tau^2)}{4\tau^2} + \frac{3}{4NK\tau^2} + 20NK(4\tau^2+\gamma^2)\right)\frac{\sigma^2}{L} \quad (89)$$

where (a) is true due to $\rho \geq 8NK\tau^2$ and $\sigma_{\mathbf{z}_2} = 20NK\tau^2 + 12NK\tau^2/\rho^2$.

Let $\mathcal{U} \triangleq \mathcal{L}(\mathbf{X}^{(1)},\mathbf{Y}^{(1)},\mathbf{\Lambda}^{(1)}) + (NK\gamma\tau)^2 + \|\bar{\mathbf{Z}}\|_F^2$ and $\mathcal{W} \triangleq \mathcal{C}(\gamma^2+5\tau^2)/(4\tau^2) + 3/(4NK\tau^2) + 20N(4\tau^2+\gamma^2)$. We have

$$\mathbb{E}[\mathcal{P}(\mathbf{X}^{(r)},\mathbf{Y}^{(r)},\mathbf{\Lambda}^{(r)})] \leq \frac{1}{T}\mathcal{C}(\mathcal{U}+\frac{\sigma^2}{L}) + \frac{\mathcal{W}\sigma^2}{L}. \quad (90)$$

6.9 Proof of Theorem 2

Proof: For the stochastic algorithm, we have the following results. According to Lemma 5 and (83), we know that

$$\mathcal{L}(\mathbf{X}^{(t+1)},\mathbf{Y}^{(t+1)},\mathbf{\Lambda}^{(t+1)}) \leq \mathcal{L}(\mathbf{X}^{(t)},\mathbf{Y}^{(t)},\mathbf{\Lambda}^{(t)}) - c_1\|\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)}\|_F^2 - c_2\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})^T\|_F^2 - c_3\|\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)}\|_F^2 + \langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)},(\mathbf{Y}_o^{(t)}-\mathbf{Y}_o^{(t+1)})^T\rangle + \langle\delta\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)},(\mathbf{X}_o^{(t)}-\mathbf{X}_o^{(t+1)})\rangle + \frac{8NK\tau^2}{\rho}\left(\|\delta\mathbf{Z}_2^{(t)}\|_F^2 + \|\delta\mathbf{Z}_2^{(t-1)}\|_F^2\right) + \frac{2NK}{\rho}\left(\gamma^2\|\delta\mathbf{Z}_1^{(t)}\|_F^2 + \tau^2\|\delta\mathbf{Z}_2^{(t)}\|_F^2\right). \quad (91)$$

Since the samples are not *i.i.d.*, we use $\mathcal{F}^{(t)}$ to denote the history of the algorithm until time t. Let

$$\mathcal{F}^{(t)} \triangleq \{\mathbf{X}^{(1)},\ldots,\mathbf{X}^{(t)},\mathbf{Y}^{(1)},\ldots,\mathbf{Y}^{(t)},\mathbf{Z}_1^{(1)},\mathbf{Z}_2^{(1)},\ldots,\mathbf{Z}_1^{(t-1)},\mathbf{Z}_2^{(t-1)}\}. \quad (92)$$

Taking expectation with respective to $\mathbf{Z}_1^{(t)}$ and $\mathbf{Z}_2^{(t)}$ on both sides of (91) conditioned on $\mathcal{F}^{(t)}$, we obtain

$$\mathbb{E}[\mathcal{L}(\mathbf{X}^{(t+1)},\mathbf{Y}^{(t+1)},\mathbf{\Lambda}^{(t+1)})|\mathcal{F}^{(t)}] \overset{(a)}{\leq} \mathcal{L}(\mathbf{X}^{(t)},\mathbf{Y}^{(t)},\mathbf{\Lambda}^{(t)}) - \underbrace{\mathbb{E}\left[c_1\|\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)}\|_F^2 + c_2\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})^T\|_F^2 + c_3\|\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)}\|_F^2|\mathcal{F}^{(t)}\right]}_{\triangleq\mathcal{H}^{(t)}} + \underbrace{\frac{2NK(\gamma^2+5\tau^2)}{\rho}(\sigma^2)^{(t)}}_{\triangleq\mathcal{I}^{(t)}} \quad (93)$$

where (a) is due to the fact that

$$\mathsf{Tr}[\mathsf{Var}[\mathbf{Z}_1^{(t)}|\mathcal{F}^{(t)}]] = \mathsf{Tr}[\mathsf{Var}[\mathbf{Z}_2^{(t)}|\mathcal{F}^{(t)}]] = \frac{1}{t^2}\sigma^2 \triangleq (\sigma^2)^{(t)}, \quad \mathsf{Tr}[\mathsf{Var}[\mathbf{Z}_2^{(t-1)}|\mathcal{F}^{(t)}]] = 0. \quad (94)$$

Note that $\mathcal{H}^{(t)}$ and $\mathcal{I}^{(t)}$ are nonnegative. Furthermore, according to (6b) and (6c), $\mathcal{H}^{(t)}$ is a function in $\mathcal{F}^{(t)}$.

Also, according to the definition of $(\sigma^2)^{(t)}$ in (94), we have

$$\sum_{t=1}^{\infty}(\sigma^2)^{(t)} < \infty. \quad (95)$$

Applying the Supermartingale Convergence Theorem [R1, Proposition 4.2], we have

$$\lim_{t\to\infty}\mathbb{E}[\|\mathbf{X}^{(t+1)}-\mathbf{X}^{(t)}\|_F^2|\mathcal{F}^{(t)}] = 0, \quad (96)$$
$$\lim_{t\to\infty}\mathbb{E}[\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)})^T\|_F^2|\mathcal{F}^{(t)}] = 0,$$
$$\lim_{t\to\infty}\mathbb{E}[\|\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)}\|_F^2|\mathcal{F}^{(t)}] = 0,$$

which imply

$$\lim_{t\to\infty} \mathbb{E}_{\mathcal{F}^{(t)}}[\|\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)}\|_F^2] = 0, \tag{97}$$
$$\lim_{t\to\infty} \mathbb{E}_{\mathcal{F}^{(t)}}[\|\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T\|_F^2] = 0, \tag{98}$$
$$\lim_{t\to\infty} \mathbb{E}_{\mathcal{F}^{(t)}}[\|\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)}\|_F^2] = 0, \tag{99}$$

and we can conclude that,

$$\mathbf{X}^{(t+1)} \xrightarrow{m.s.} \mathbf{X}^{(t)}, \quad \mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)})^T \xrightarrow{m.s.} \mathbf{X}^{(t)}(\mathbf{Y}^{(t)})^T, \quad \mathbf{Y}^{(t+1)} \xrightarrow{m.s.} \mathbf{Y}^{(t)}. \tag{100}$$

By Lemma 4, we have

$$\lim_{t\to\infty} \mathbb{E}_{\mathcal{F}^{(t)}}[\|\mathbf{\Lambda}^{(t+1)} - \mathbf{\Lambda}^{(t)}\|_F^2] = 0, \tag{101}$$

which implies $\lim_{t\to\infty} \mathbb{E}_{\mathcal{F}^{(t)}}[\|\mathbf{X}^{(t)} - \mathbf{Y}^{(t)}\|_F^2] = 0$, i.e.,

$$\mathbf{\Lambda}^{(t+1)} \xrightarrow{m.s.} \mathbf{\Lambda}^{(t)}, \quad \mathbf{X}^{(t)} \xrightarrow{m.s.} \mathbf{Y}^{(t)}. \tag{102}$$

The optimality condition of (6b) is given by

$$\langle(\mathbf{X}^{(t)})^T(\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_1^{(t)}) - \rho(\mathbf{X}^{(t)} - \mathbf{Y}^{(t+1)} + \mathbf{\Lambda}^{(t)}/\rho)^T + \beta^{(t)}(\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T, (\mathbf{Y} - \mathbf{Y}^{(t+1)})^T\rangle \geq 0, \quad \forall\, \mathbf{Y} \geq 0 \quad \text{and} \quad \mathbf{Y} \leq \tau. \tag{103}$$

Substituting (21) into (103), we have

$$\Big\langle(\mathbf{X}^{(t)})^T(\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_1^{(t)}) - \rho(\mathbf{X}^{(t)} - \mathbf{Y}^{(t+1)})^T + (\mathbf{Y}^{(t+1)})^T(\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_2^{(t)})^T + \beta^{(t)}(\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T, (\mathbf{Y} - \mathbf{Y}^{(t+1)})^T\Big\rangle \geq 0, \forall\, \mathbf{Y} \geq 0 \quad \text{and} \quad \mathbf{Y} \leq \tau. \tag{104}$$

Taking expectation on $\mathbf{Z}_1^{(t)}, \mathbf{Z}_2^{(t)}$ conditioned on $\mathcal{F}^{(t)}$, based on A1 we have

$$\lim_{t\to\infty} \mathbb{E}_{\mathbf{Z}_1^{(t)},\mathbf{Z}_2^{(t)}}[\langle(\mathbf{X}^{(t)})^T(\mathbf{X}^{(t)}(\mathbf{Y}^{(t+1)})^T - \bar{\mathbf{Z}}) - \rho(\mathbf{X}^{(t)} - \mathbf{Y}^{(t+1)})^T + (\mathbf{Y}^{(t+1)})^T(\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \bar{\mathbf{Z}})^T + \beta^{(t)}(\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)})^T, (\mathbf{Y} - \mathbf{Y}^{(t+1)})^T\rangle|\mathcal{F}^{(t)}] \geq 0, \quad \forall\, \mathbf{Y} \geq 0 \quad \text{and} \quad \mathbf{Y} \leq \tau. \tag{105}$$

Taking expectation with respective to $\mathcal{F}^{(t)}$ on both sides of (105) and passing limit over any converging subsequence of $\{\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)}\}$, we have

$$\langle(\mathbf{X}^*)^T(\mathbf{X}^*(\mathbf{Y}^*)^T - \bar{\mathbf{Z}}) + (\mathbf{Y}^*)^T(\mathbf{X}^*(\mathbf{Y}^*)^T - \bar{\mathbf{Z}})^T - \rho(\mathbf{X}^* - \mathbf{Y}^*)^T, (\mathbf{Y} - \mathbf{Y}^*)^T\rangle \geq 0, \quad \forall\, \mathbf{Y} \geq 0 \quad \text{and} \quad \mathbf{Y} \leq \tau. \tag{106}$$

The optimality condition of (6c) is given by

$$(\mathbf{X}^{(t+1)}(\mathbf{Y}^{(t+1)})^T - \mathbf{Z}_2^{(t)})(\mathbf{Y}^{(t+1)}) + \rho(\mathbf{X}^{(t+1)} - \mathbf{Y}^{(t+1)} + \mathbf{\Lambda}^{(t)}/\rho) = 0. \tag{107}$$

Similarly, taking expectation on (107) over $\mathbf{Z}_2^{(t)}$ and then $\mathcal{F}^{(t)}$ on both sides and passing limit over the same subsequence, we have

$$(\mathbf{X}^*(\mathbf{Y}^*)^T - \bar{\mathbf{Z}})\mathbf{Y}^* + \rho(\mathbf{X}^* - \mathbf{Y}^* + \mathbf{\Lambda}^*/\rho) = 0. \tag{108}$$

Using the fact $\mathbf{X}^* = \mathbf{Y}^*$ by (102), we have

$$\langle(\mathbf{X}^*(\mathbf{X}^*)^T - \bar{\mathbf{Z}})\mathbf{X}^*, \mathbf{X} - \mathbf{X}^*\rangle \geq 0, \quad \forall\, \mathbf{X} \geq 0 \quad \text{and} \quad \mathbf{X} \leq \tau, \tag{109}$$
$$(\mathbf{X}^*(\mathbf{X}^*)^T - \bar{\mathbf{Z}})\mathbf{X}^* + \mathbf{\Lambda}^* = 0, \tag{110}$$

which are the KKT conditions of problem (3). From (109), we know that $\mathbf{X}^*$ stratifies the stationary condition of problem (1).

6.10 Proof of Theorem 3

Proof: Summing both sides of (85) over $t = 1, \ldots, T$ and taking expectation with respect to $\mathbf{Z}_1^{(t)}$ and $\mathbf{Z}_2^{(t)}$ conditioned on $\mathcal{F}^{(t)}, \forall t$, we have

$$\sum_{t=1}^{T} \mathbb{E}_{\mathbf{Z}_1^{(t)},\mathbf{Z}_2^{(t)}}[\mathcal{P}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)})|\mathcal{F}^{(t)}] \overset{(a)}{\le} \mathcal{C}\Big(\mathcal{L}(\mathbf{X}^{(1)}, \mathbf{Y}^{(1)}, \mathbf{\Lambda}^{(1)}) + \mathcal{C}((NK\gamma\tau)^2 + \|\overline{\mathbf{Z}}\|_F^2 + \sigma^2/T^2) + \left(\frac{2NK(5\tau^2+\gamma^2)}{\rho} + 8NK\gamma^2 + 2\sigma_{\mathbf{z}_2}\right)\sum_{t=1}^{T}(\sigma^2)^{(t)} \tag{111}$$

where (a) is due to the independence of the sequence $\mathbf{Z}_1^{(t)}$ and $\mathbf{Z}_2^{(t)}$ over t and independence of the data given $\mathcal{F}^{(t)}$, we have the following results: i) since $\mathbf{X}^{(t)}$ is only dependent on $\mathbf{Z}_2^{(t-1)}$ conditioned on $\mathcal{F}^{(t)}$, we have $\mathbb{E}[\langle(\mathbf{X}^{(t)})^T\delta\mathbf{Z}_1^{(t)}, (\mathbf{Y}_o^{(t)} - \mathbf{Y}_o^{(t+1)})^T\rangle|\mathcal{F}^{(t)}] = 0, \forall t$; ii) similarly $\mathbf{Y}^{(t+1)}$ is only dependent on $\mathbf{Z}_1^{(t)}$ and $\mathbf{Z}_2^{(t-1)}$ conditioned on $\mathcal{F}^{(t)}$, we have $\mathbb{E}[\langle\delta\mathbf{Z}_2^{(t)}\mathbf{Y}^{(t+1)}, (\mathbf{X}_o^{(t)} - \mathbf{X}_o^{(t+1)})\rangle|\mathcal{F}^{(t)}] = 0, \forall t$.

Taking expectation over $\mathcal{F}^{(t)}, \forall t$ on (111) and using $\rho \ge 8NK\tau^2$ and $\sigma_{\mathbf{z}_2} = 20NK\tau^2 + 12NK\tau^2/\rho^2$, we have

$$\begin{aligned}\sum_{t=1}^{T} \mathbb{E}_{\mathcal{F}^{(t)}}[\mathcal{P}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)})] \le& \mathcal{C}\mathcal{L}(\mathbf{X}^{(1)}, \mathbf{Y}^{(1)}, \mathbf{\Lambda}^{(1)}) + \mathcal{C}\left((NK\gamma\tau)^2 + \|\overline{\mathbf{Z}}\|_F^2 + \sigma^2/T^2\right) \\ &+ \left(\frac{(5\tau^2+\gamma^2)}{4\tau^2} + \frac{3}{16NK\tau^2} + 8NK(\gamma^2+5\tau^2)\right)\sum_{t=1}^{T}(\sigma^2)^{(t)} \\ \overset{(a)}{\le}& \mathcal{C}\left(\mathcal{L}(\mathbf{X}^{(1)}, \mathbf{Y}^{(1)}, \mathbf{\Lambda}^{(1)}) + (NK\gamma\tau)^2 + \|\overline{\mathbf{Z}}\|_F^2 + \sigma^2\right) \\ &+ \left(\frac{(5\tau^2+\gamma^2)}{4\tau^2} + \frac{3}{16NK\tau^2} + 8NK(\gamma^2+5\tau^2)\right)2\sigma^2\end{aligned} \tag{112}$$

where (a) is due to $T \ge 1$ and $\sum_{t=1}^{T}(\sigma^2)^{(t)} \le \sum_{t=1}^{\infty}(\sigma^2)^{(t)} \le 2\sigma^2$.

Further, from the definition of the output, we have

$$\mathbb{E}[\mathcal{P}(\mathbf{X}^{(r)}, \mathbf{Y}^{(r)}, \mathbf{\Lambda}^{(r)})] \le \frac{\mathcal{C}\mathcal{S} + \mathcal{C}\sigma^2 + \mathcal{K}\sigma^2}{T} \tag{113}$$

where $\mathcal{S} = \mathcal{L}(\mathbf{X}^{(1)}, \mathbf{Y}^{(1)}, \mathbf{\Lambda}^{(1)}) + (NK\gamma\tau)^2 + \|\overline{\mathbf{Z}}\|_F^2$ and $\mathcal{K} = 2((5\tau^2+\gamma^2)/(4\tau^2) + 3/(16NK\tau^2) + 8N(\gamma^2+5\tau^2))$.

6.11 Proof of Corollary 1

Proof: The proof is similar as Theorem 2 and Theorem 3. Using the weighted aggregate samples, (94) becomes

$$\mathsf{Tr}[\mathsf{Var}[\mathbf{Z}_1^{(t)}|\mathcal{F}^{(t)}]] = \mathsf{Tr}[\mathsf{Var}[\mathbf{Z}_2^{(t)}|\mathcal{F}^{(t)}]] = \left(\frac{2}{t+1}\right)^2\sigma^2 \triangleq (\sigma_{\mathsf{a}}^2)^{(t)}, \quad \mathsf{Tr}[\mathsf{Var}[\mathbf{Z}_2^{(t-1)}|\mathcal{F}^{(t)}]] = 0.$$

Consequently, we have

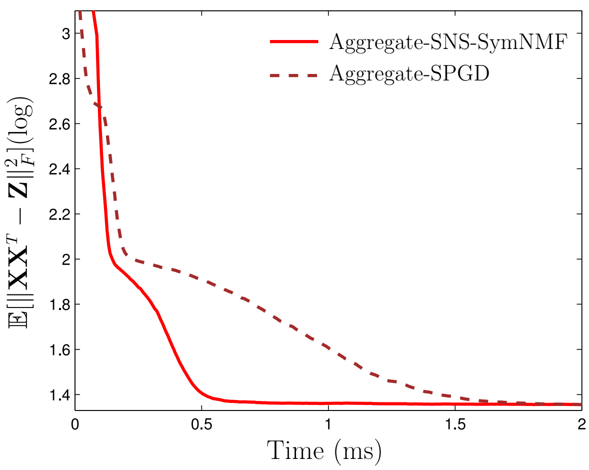
$$\sum_{t=1}^{T}(\sigma_{\mathsf{a}}^2)^{(t)} = \sum_{t=1}^{T}\left(\frac{2}{t+1}\right)^2\sigma^2 \ge \sum_{t=1}^{T}\left(\frac{1}{t}\right)^2\sigma^2 = \sum_{t=1}^{T}(\sigma^2)^{(t)}.$$

According to (112), the claim is proved.

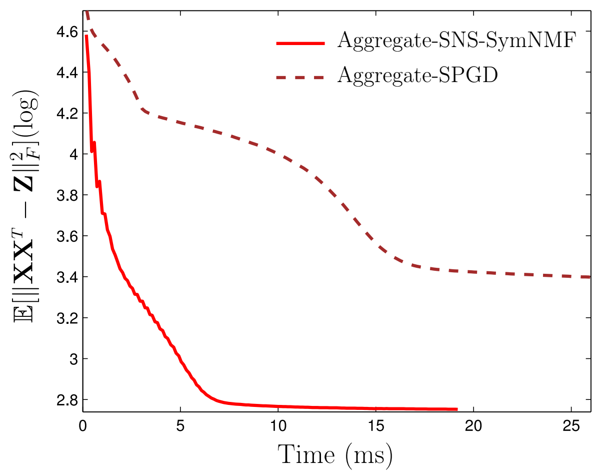
7 Additional Numerical Results

Based on the data sets described in the main context, we further provide more numerical results.

7.1 Computational Time



(a) Objective Value $\alpha = 0.01$, $N = 400$.

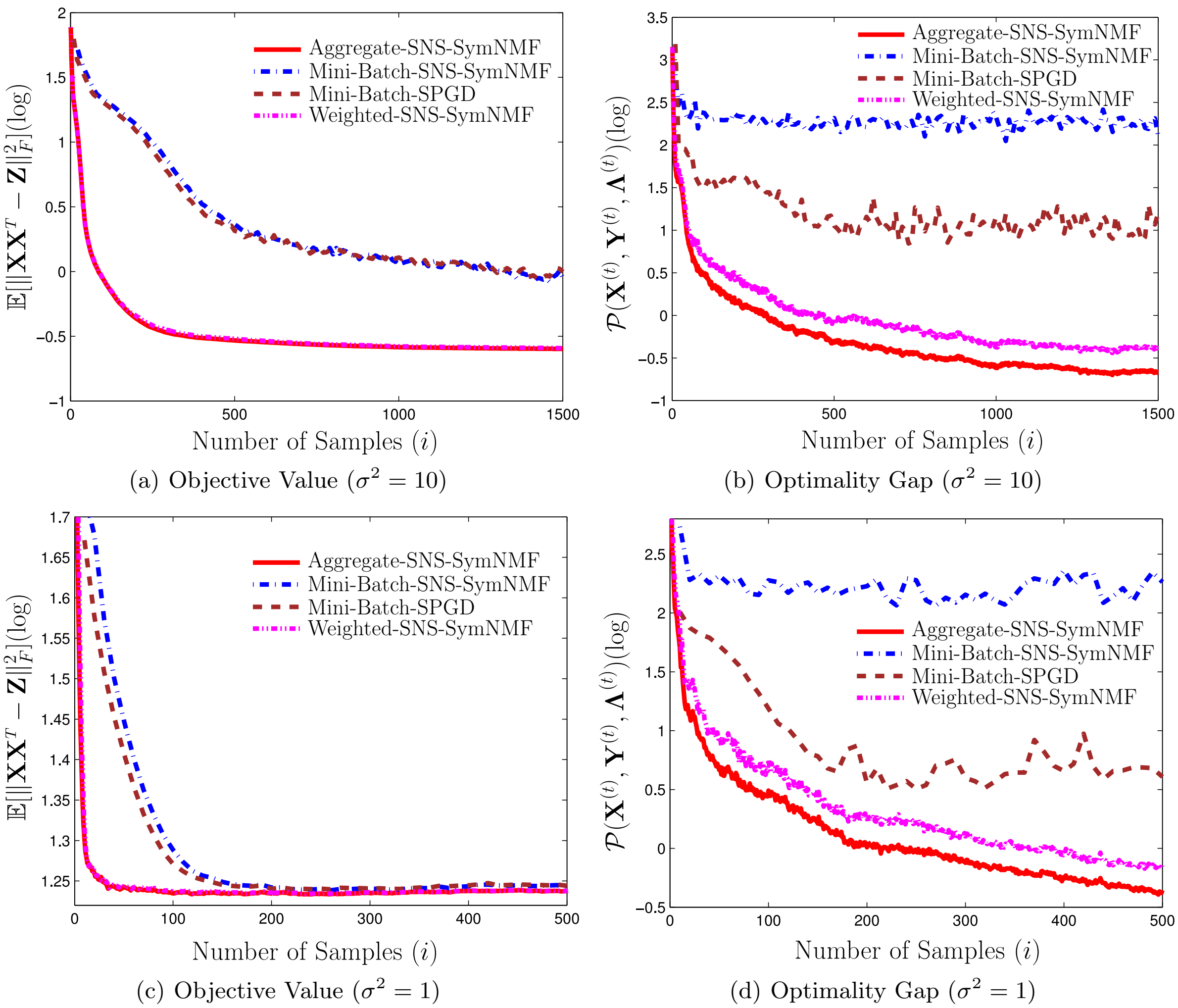


(b) Objective Value $\alpha = 0.001$, $N = 2000$.

Figure 4: The convergence behaviors for the static network; $K = 4$.

We can also apply the aggregate data for SPGD. Because of the efficient implementation of the proposed SNS-SymNMF algorithm, it can be observed that in Figure 4 when the dimension of the matrix is large, the aggregate SNS-SymNMF algorithm is faster than aggregate-SPGD in terms of computational time. Aggregate-SPGD is referred as the algorithm which we use the aggregate samples instead of Mini-Batch for SPGD. Here, the ratio of the numbers of the nodes within each cluster is kept as $1:2:4:3$ for the cases when $N = 400$ and $N = 2000$.

7.2 Effect of Variance of Samples



(a) Objective Value ($\sigma^2 = 10$)

(b) Optimality Gap ($\sigma^2 = 10$)

(c) Objective Value ($\sigma^2 = 1$)

(d) Optimality Gap ($\sigma^2 = 1$)

Figure 5: The convergence behavior; $K = 4$, $\alpha = 0.02$, $N = 50$, $L = 10$.

From Figure 5, it can be seen that when σ^2 is small, the Mini-Batch-SNS-SymNMF and Mini-Batch-SPGD algorithms can converge closely to aggregate-SNS-SymNMF.

7.3 Sensitivity of Parameters

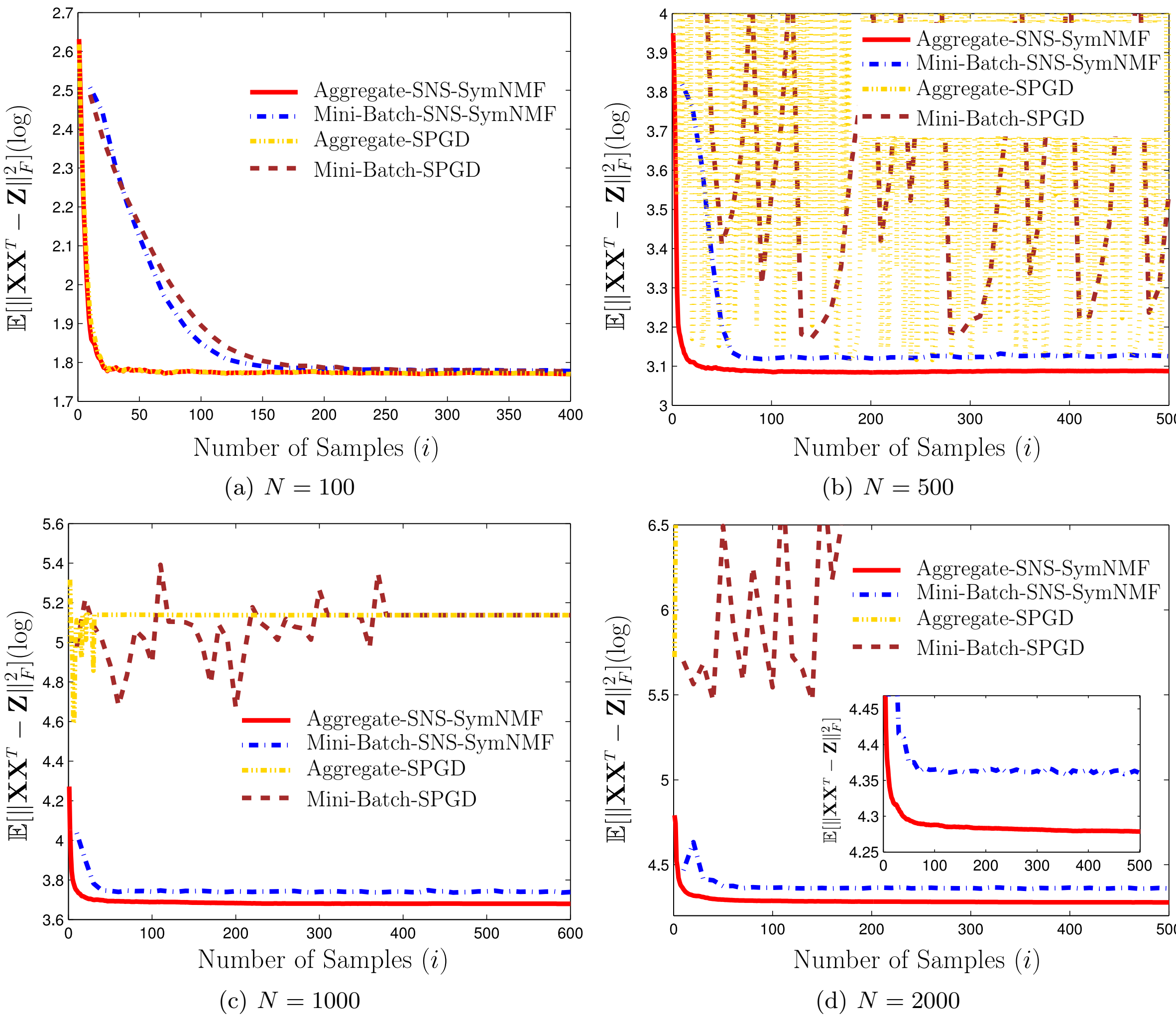


Figure 6: The convergence behavior; $\alpha = 0.01$, $K = 4$, $\sigma^2 = 1$, $L = 10$.

Fixing the parameters α in SPGD and ρ in SNS-SymNMF and increasing N, from Figure 6 we can observe that the SPGD algorithm diverges when $N = 500, 1000, 2000$. Here, we also keep the ratio of the numbers of the nodes within each cluster as $1:2:4:3$. Based on the numerical results, we know that the convergence and convergence rate of the SNS-SymNMF algorithm are much more robust against the parameter than SPGD. In general, the step-size tuning of SPGD is difficult in practice even for the case that the feasible set of the problem is bounded.

Secondary Literatures

[R1] Dimitri P Bertsekas & John N Tsitsiklis, *Neuro-dynamic programming.* Athena Scientific, Belmont, MA, 1996.