Supplementary Material for "Dynamic Matrix Recovery"

S.1 Proof of Theorem 1

We need the following lemma for proving Theorem 1.

Lemma S.1 *Let $\Sigma \in \mathbb{S}_+^{m\times m}$ be a positive definite matrix with the smallest eigenvalue $\mu > 0$. Then the following inequality*

$$\alpha^\top \Sigma(\alpha+\beta) \geq \mu\left(\|\alpha\|_2^2 - \|\alpha\|_2\|\beta\|_2\right) \tag{S.1}$$

holds for any $\alpha, \beta \in \mathbb{R}^m$ with $\|\alpha\|_2 \geq \|\beta\|_2$.

Proof of Lemma S.1: Note that for a fixed α,

$$\alpha^\top \Sigma(\alpha+\beta) = \alpha^\top\Sigma\alpha + \langle \Sigma\alpha, \beta\rangle \geq \alpha^\top\Sigma\alpha - \|\beta\|_2\|\Sigma\alpha\|_2.$$

Let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues of Σ satisfying $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m = \mu > 0$. Using the representation $\alpha = \sum_{s=1}^m a_s\psi_s$ where ψ_s are orthogonal eigenvectors corresponding to λ_s, we have

$$\alpha^\top\Sigma\alpha - \|\beta\|_2\|\Sigma\alpha\|_2 = \sum_{s=1}^m a_s^2\lambda_s - \|\beta\|_2\left(\sum_{s=1}^m a_s^2\lambda_s^2\right)^{1/2}.$$

Consider the minimization problem

$$\begin{aligned} \min \quad & \sum_{s=1}^m a_s^2\lambda_s - \|\beta\|_2\left(\sum_{s=1}^m a_s^2\lambda_s^2\right)^{1/2} \\ \text{s.t.} \quad & \sum_{s=1}^m a_s^2 = \|a\|_2^2. \end{aligned}$$

This is a convex problem with

$$\mathcal{L}(a_1,\ldots,a_m,\eta) = \sum_{s=1}^m a_s^2\lambda_s - \|\beta\|_2\left(\sum_{s=1}^m a_s^2\lambda_s^2\right)^{1/2} - \eta\left(\sum_{s=1}^m a_s^2 - \|a\|_2^2\right)$$

and

$$\frac{\partial \mathcal{L}}{\partial a_s} = 2a_s \left(\lambda_s - \frac{1}{2}\|\beta\|_2 \lambda_s^2 \left(\sum_{s=1}^{m} a_s^2 \lambda_s^2 \right)^{-1/2} - \eta \right).$$

With Karush-Kuhn-Tucker condition, we have minimum values $a_s, s = 1, 2, \ldots, m$ satisfying that only one of a_s is equal to $\|a\|_2$ and others are all zero, which means that

$$\alpha^\top \Sigma(\alpha + \beta) \geq \min_{s=1,2,\ldots,m} \left\{ \lambda_s \left(\|\alpha\|_2^2 - \|\beta\|_2 \|\alpha\|_2 \right) \right\} \geq \mu(\|\alpha\|_2^2 - \|\alpha\|_2 \|\beta\|_2).$$

□

Proof of Theorem 1: For brevity, we denote

$$\omega_j = \omega_h(j - t), \quad \mathcal{C}_j(M, N) = \frac{1}{n_j} \sum_{i=1}^{n_j} \mathbb{E}\langle M, X_{ji} \rangle \langle N, X_{ji} \rangle$$

for all $M, N \in \mathbb{R}_{m_1 \times m_2}$. With the definition of $\widetilde{M_t^\lambda}$,

$$\begin{aligned}
&\sum_{j=1}^{T} \omega_j \mathcal{C}_j(\widetilde{M_t^\lambda}, \widetilde{M_t^\lambda}) - 2\sum_{j=1}^{T} \omega_j \left\langle \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ji} X_{ji}, \widetilde{M_t^\lambda} \right\rangle + \lambda \|\widetilde{M_t^\lambda}\|_1 \\
\leq &\sum_{j=1}^{T} \omega_j \mathcal{C}_j(M, M) - 2\sum_{j=1}^{T} \omega_j \left\langle \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ji} X_{ji}, M \right\rangle + \lambda \|M\|_1
\end{aligned}$$

holds for any $M \in \mathbb{R}^{m_1 \times m_2}$. With the identity

$$\left\langle \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ji} X_{ji}, M \right\rangle = \langle \Delta_j, M \rangle + \mathcal{C}_j(M_j^0, M),$$

we have

$$\begin{aligned}
&\sum_{j=1}^{T} \omega_j \mathcal{C}_j(\widetilde{M_t^\lambda}, \widetilde{M_t^\lambda}) - 2\sum_{j=1}^{T} \omega_j \mathcal{C}_j(M_j^0, \widetilde{M_t^\lambda}) \\
\leq &\sum_{j=1}^{T} \omega_j \mathcal{C}_j(M, M) - 2\sum_{j=1}^{T} \omega_j \mathcal{C}_j(M_j^0, M) + 2\sum_{j=1}^{T} \omega_j \langle \Delta_j, \widetilde{M_t^\lambda} - M \rangle + \lambda(\|M\|_1 - \|\widetilde{M_t^\lambda}\|_1).
\end{aligned}$$

Using $\langle \cdot, * \rangle \leq \| \cdot \|_\infty \| * \|_1$, one can obtain

$$\begin{aligned}
&\sum_{j=1}^{T} \omega_j \mathcal{C}_j(\widetilde{M_t^\lambda} - M_j^0, \widetilde{M_t^\lambda} - M_j^0) \\
\leq &\sum_{j=1}^{T} \omega_j \mathcal{C}_j(M - M_j^0, M - M_j^0) + 2\|\sum_{j=1}^{T} \omega_j \Delta_j\|_\infty \|\widetilde{M_t^\lambda} - M\|_1 + \lambda(\|M\|_1 - \|\widetilde{M_t^\lambda}\|_1).
\end{aligned}$$

Under the assumption that $2\|\sum_{j=1}^{T}\omega_j\Delta_j\|_\infty \le \lambda$, we have

$$\sum_{j=1}^{T}\omega_j\mathcal{C}_j(\widetilde{M}_t^\lambda - M_j^0, \widetilde{M}_t^\lambda - M_j^0) \le \sum_{j=1}^{T}\omega_j\mathcal{C}_j(M - M_j^0, M - M_j^0) + 2\lambda\|M\|_1.$$

Set $M = M_t^0$, then

$$\sum_{j=1}^{T}\omega_j\mathcal{C}_j(\widetilde{M}_t^\lambda - M_j^0, \widetilde{M}_t^\lambda - M_j^0) \le \sum_{j=1}^{T}\omega_j\mathcal{C}_j(M_t^0 - M_j^0, M_t^0 - M_j^0) + 2\lambda\|M_t^0\|_1.$$

With the identity

$$\begin{aligned}
&\mathcal{C}_j(\widetilde{M}_t^\lambda - M_j^0, \widetilde{M}_t^\lambda - M_j^0)\\
=&\mathcal{C}_j(\widetilde{M}_t^\lambda - M_t^0, \widetilde{M}_t^\lambda - M_t^0) + \mathcal{C}_j(M_t^0 - M_j^0, M_t^0 - M_j^0) + 2\mathcal{C}_j(\widetilde{M}_t^\lambda - M_t^0, M_t^0 - M_j^0),
\end{aligned}$$

we have

$$\sum_{j=1}^{T}\omega_j\mathcal{C}_j(\widetilde{M}_t^\lambda - M_t^0, \widetilde{M}_t^\lambda - M_t^0) + 2\sum_{j=1}^{T}\omega_j\mathcal{C}_j(\widetilde{M}_t^\lambda - M_t^0, M_t^0 - M_j^0) \le 2\lambda\|M_t^0\|_1. \tag{S.2}$$

If $\left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2 < 2\left\|M_t^0 - \sum_{j=1}^{T}\omega_j M_j^0\right\|_2$, we can maintain the bound

$$\left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2 < 2\delta_h M_t. \tag{S.3}$$

If $\left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2 \ge 2\left\|M_t^0 - \sum_{j=1}^{T}\omega_j M_j^0\right\|_2$, with Assumption 1, we know that

$$\mathcal{C}_j(M, N) = \mathrm{vec}(M)^\top \Sigma\, \mathrm{vec}(N). \tag{S.4}$$

So (S.2) can be expressed as

$$\begin{aligned}
&\left(\mathrm{vec}(\widehat{M}_t^\lambda) - \mathrm{vec}(M_t^0)\right)^\top \Sigma \left[\left(\mathrm{vec}(\widehat{M}_t^\lambda) - \mathrm{vec}(M_t^0)\right) + \left(2\,\mathrm{vec}(M_t^0) - 2\sum_{j=1}^{T}\omega_j\,\mathrm{vec}(M_j^0)\right)\right]\\
\le&2\lambda\|M_t^0\|_1.
\end{aligned}$$

From Lemma S.1,

$$\begin{aligned}
&\left(\mathrm{vec}(\widehat{M}_t^\lambda) - \mathrm{vec}(M_t^0)\right)^\top \Sigma \left[\left(\mathrm{vec}(\widehat{M}_t^\lambda) - \mathrm{vec}(M_t^0)\right) + \left(2\,\mathrm{vec}(M_t^0) - 2\sum_{j=1}^{T}\omega_j\,\mathrm{vec}(M_j^0)\right)\right]\\
\ge&\mu\left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2^2 - \mu\left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2\left\|M_t^0 - \sum_{j=1}^{T}\omega_j M_j^0\right\|_2.
\end{aligned} \tag{S.5}$$

Thus we have

$$\mu \left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2^2 - \mu \left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2 \left\|M_t^0 - \sum_{j=1}^T \omega_j M_j^0\right\|_2 - 2\lambda\|M_t^0\|_1 \leq 0.$$

By solving this inequality, we have

$$\left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2 \leq \frac{\mu\delta_h M_t + \sqrt{\mu^2(\delta_h M_t)^2 + 8\mu\lambda\|M_t^0\|_1}}{2\mu} \leq \delta_h M_t + \sqrt{\frac{2\lambda\|M_t^0\|_1}{\mu}}. \tag{S.6}$$

Combining (S.3) and (S.6),

$$\left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2 \leq \delta_h M_t + \max\left\{\delta_h M_t, \sqrt{\frac{2\lambda\|M_t^0\|_1}{\mu}}\right\} \leq \delta_h M_t + \left((\delta_h M_t)^2 + 2\mu^{-1}\lambda\|M_t^0\|_1\right)^{1/2}. \tag{S.7}$$

Because $\widetilde{M}_t^\lambda$ is the minimizer of (6), there exists a sub-gradient matrix B satisfying that for all $M \in \mathbb{R}^{m_1\times m_2}$,

$$\left\langle B, \widetilde{M}_t^\lambda - M\right\rangle \leq 0,$$

which equals to that there exists $\widehat{V}_t \in \partial\|\widetilde{M}_t^\lambda\|_1$ satisfying that

$$2\sum_{j=1}^T \omega_j \left(\mathcal{C}_j(\widetilde{M}_t^\lambda, \widetilde{M}_t^\lambda - M) - \left\langle \frac{1}{n_j}\sum_{i=1}^{n_j} Y_{ji}X_{ji}, \widetilde{M}_t^\lambda - M\right\rangle\right) + \lambda\langle \widehat{V}_t, \widetilde{M}_t^\lambda - M\rangle \leq 0.$$

It is equivalent to

$$\begin{aligned}
&2\sum_{j=1}^T \omega_j \mathcal{C}_j(\widetilde{M}_t^\lambda - M_j^0, \widetilde{M}_t^\lambda - M) + \lambda\langle \widehat{V}_t - V, \widetilde{M}_t^\lambda - M\rangle \\
\leq& -\lambda\langle V, \widetilde{M}_t^\lambda - M\rangle + 2\sum_{j=1}^T \omega_j\langle \Delta_j, \widetilde{M}_t^\lambda - M\rangle,
\end{aligned}$$

where $V \in \partial\|M\|_1$.

Denote that $M = \sum_{j=1}^r \sigma_j u_j v_j^\top$ with support S_1, S_2. V have representation from Watson (1992):

$$V = \sum_{j=1}^r u_j v_j^\top + P_{S_1^\perp} W P_{S_2^\perp},\ \|W\|_\infty \leq 1.$$

We choose a W subject to

$$\langle P_{S_1^\perp} W P_{S_2^\perp}, \widetilde{M}_t^\lambda - M\rangle = \langle W, P_{S_1^\perp}\widehat{M}_t^\lambda P_{S_2^\perp}\rangle = \|P_{S_1^\perp}\widehat{M}_t^\lambda P_{S_2^\perp}\|_1.$$

With the monotonicity of sub-gradients, we have $\langle \widehat{V}_t - V, \widetilde{M}_t^\lambda - M\rangle \geq 0$. So

$$
\begin{aligned}
&2\sum_{j=1}^{T}\omega_j\mathcal{C}_j(\widetilde{M}_t^\lambda - M_j^0, \widetilde{M}_t^\lambda - M) + \lambda\|P_{S_1^\perp}\widetilde{M}_t^\lambda P_{S_2^\perp}\|_1 \\
\leq & -\lambda\langle\sum_{j=1}^{r}u_jv_j^\top, \widetilde{M}_t^\lambda - M\rangle + 2\sum_{j=1}^{T}\omega_j\langle\Delta_j, \widetilde{M}_t^\lambda - M\rangle.
\end{aligned}
$$

Using the identity

$$
\begin{aligned}
&2\sum_{j=1}^{T}\omega_j\mathcal{C}_j(\widetilde{M}_t^\lambda - M_j^0, \widetilde{M}_t^\lambda - M) \\
=&\sum_{j=1}^{T}\omega_j\left(\mathcal{C}_j(\widetilde{M}_t^\lambda - M_j^0, \widetilde{M}_t^\lambda - M_j^0) + \mathcal{C}_j(\widetilde{M}_t^\lambda - M, \widetilde{M}_t^\lambda - M) - \mathcal{C}_j(M - M_j^0, M - M_j^0)\right)
\end{aligned}
$$

and the inequality

$$
-\langle\sum_{j=1}^{r}u_jv_j^\top, \widetilde{M}_t^\lambda - M\rangle = -\langle\sum_{j=1}^{r}u_jv_j^\top, P_{S_1}(\widetilde{M}_t^\lambda - M)P_{S_2}\rangle \leq \|P_{S_1}(\widetilde{M}_t^\lambda - M)P_{S_2}\|_1,
$$

we have

$$
\begin{aligned}
&\sum_{j=1}^{T}\omega_j\mathcal{C}_j(\widetilde{M}_t^\lambda - M_j^0, \widetilde{M}_t^\lambda - M_j^0) + \sum_{j=1}^{T}\omega_j\mathcal{C}_j(\widetilde{M}_t^\lambda - M, \widetilde{M}_t^\lambda - M) + \lambda\|P_{S_1^\perp}\widetilde{M}_t^\lambda P_{S_2^\perp}\|_1 \\
\leq&\sum_{j=1}^{T}\omega_j\mathcal{C}_j(M - M_j^0, M - M_j^0) + \lambda\|P_{S_1}(\widetilde{M}_t^\lambda - M)P_{S_2}\|_1 + 2\sum_{j=1}^{T}\omega_j\langle\Delta_j, \widetilde{M}_t^\lambda - M\rangle.
\end{aligned}
\tag{S.8}
$$

With identity

$$
\begin{aligned}
\langle\Delta_j, \widetilde{M}_t^\lambda - M\rangle &= \langle P_M(\Delta_j), \widehat{M}_t^\lambda - M\rangle + \langle P_{S_1^\perp}\Delta_j P_{S_2^\perp}, \widehat{M}_t^\lambda - M\rangle \\
&= \langle P_M(\Delta_j), \widehat{M}_t^\lambda - M\rangle + \langle P_{S_1^\perp}\Delta_j P_{S_2^\perp}, \widehat{M}_t^\lambda\rangle,
\end{aligned}
$$

we have that

$$
\begin{aligned}
&\sum_{j=1}^{T}\omega_j\langle\Delta_j, \widetilde{M}_t^\lambda - M\rangle \\
=&\left\langle P_M\left(\sum_{j=1}^{T}\omega_j\Delta_j\right), \widehat{M}_t^\lambda - M\right\rangle + \left\langle P_{S_1^\perp}\left(\sum_{j=1}^{T}\omega_j\Delta_j\right)P_{S_2^\perp}, \widehat{M}_t^\lambda\right\rangle \\
\leq&\left\|P_M\left(\sum_{j=1}^{T}\omega_j\Delta_j\right)\right\|_2\left\|\widetilde{M}_t^\lambda - M\right\|_2 + \left\|P_{S_1^\perp}\left(\sum_{j=1}^{T}\omega_j\Delta_j\right)P_{S_2^\perp}\right\|_\infty\left\|P_{S_1^\perp}\widetilde{M}_t^\lambda P_{S_2^\perp}\right\|_1.
\end{aligned}
$$

With

$$\left\|P_M\left(\sum_{j=1}^T \omega_j\Delta_j\right)\right\|_2 \le \sqrt{\operatorname{rank}\left[P_A\left(\sum_{j=1}^T \omega_j\Delta_j\right)\right]}\left\|P_M\left(\sum_{j=1}^T \omega_j\Delta_j\right)\right\|_\infty$$
$$\le \sqrt{2\operatorname{rank}(M)}\|\sum_{j=1}^T \omega_j\Delta_j\|_\infty$$

and

$$\left\|P_{S_1^\perp}\left(\sum_{j=1}^T \omega_j\Delta_j\right)P_{S_2^\perp}\right\|_\infty \le \|\sum_{j=1}^T \omega_j\Delta_j\|_\infty,$$

we have

$$\sum_{j=1}^T \omega_j\langle\Delta_j, \widetilde{M}_t^\lambda - M\rangle \le \sqrt{2\operatorname{rank}(M)}\|\sum_{j=1}^T \omega_j\Delta_j\|_\infty\|\widetilde{M}_t^\lambda - M\|_2 + \|\sum_{j=1}^T \omega_j\Delta_j\|_\infty\|P_{S_1^\perp}\widetilde{M}_t^\lambda P_{S_2^\perp}\|_1. \tag{S.9}$$

Also we have

$$\begin{aligned}\|P_{S_1}(\widetilde{M}_t^\lambda - M)P_{S_2}\|_1 &\le \sqrt{\operatorname{rank}(M)}\|P_{S_1}(\widetilde{M}_t^\lambda - M)P_{S_2}\|_2\\ &\le \sqrt{\operatorname{rank}(M)}\|\widetilde{M}_t^\lambda - M\|_2.\end{aligned} \tag{S.10}$$

With (S.8), (S.9) and (S.10),

$$\begin{aligned}&\sum_{j=1}^T \omega_j\left[\mathcal{C}_j(\widetilde{M}_t^\lambda - M_j^0, \widetilde{M}_t^\lambda - M_j^0) + \mathcal{C}_j(\widetilde{M}_t^\lambda - M, \widetilde{M}_t^\lambda - M)\right]\\ &+\left(\lambda - 2\|\sum_{j=1}^T \omega_j\Delta_j\|_\infty\right)\|P_{S_1^\perp}\widetilde{M}_t^\lambda P_{S_2^\perp}\|_1\\ &\le \sum_{j=1}^T \omega_j\mathcal{C}_j(M - M_j^0, M - M_j^0) + \sqrt{\operatorname{rank}(M)}\left(\lambda + 2\sqrt{2}\|\sum_{j=1}^T \omega_j\Delta_j\|_\infty\right)\|\widetilde{M}_t^\lambda - M\|_2.\end{aligned}$$

With $\lambda \ge 2\|\sum_{j=1}^T \omega_j\Delta_j\|_\infty$,

$$\begin{aligned}&\sum_{j=1}^T \omega_j\mathcal{C}_j(\widetilde{M}_t^\lambda - M_j^0, \widetilde{M}_t^\lambda - M_j^0) + \sum_{j=1}^T \omega_j\mathcal{C}_j(\widetilde{M}_t^\lambda - M, \widetilde{M}_t^\lambda - M)\\ &\le \sum_{j=1}^T \omega_j\mathcal{C}_j(M - M_j^0, M - M_j^0) + \lambda(1+\sqrt{2})\sqrt{\operatorname{rank}(M)}\|\widetilde{M}_t^\lambda - M\|_2.\end{aligned}$$

Set $M = M_t^0$, so

$$\begin{aligned}&\sum_{j=1}^T \omega_j \mathcal{C}_j(\widetilde{M}_t^\lambda - M_j^0, \widetilde{M}_t^\lambda - M_j^0) + \sum_{j=1}^T \omega_j \mathcal{C}_j(\widetilde{M}_t^\lambda - M_t^0, \widetilde{M}_t^\lambda - M_t^0) \\ \leq &\sum_{j=1}^T \omega_j \mathcal{C}_j(M_t^0 - M_j^0, M_t^0 - M_j^0) + \lambda(1+\sqrt{2})\sqrt{\mathrm{rank}(M)}\|\widetilde{M}_t^\lambda - M_t^0\|_2.\end{aligned}$$

With the identity

$$\mathcal{C}_j(\widetilde{M}_t^\lambda - M_j^0, \widetilde{M}_t^\lambda - M_j^0) = \mathcal{C}_j(\widetilde{M}_t^\lambda - M_t^0) + 2\mathcal{C}_j(\widetilde{M}_t^\lambda - M_t^0, M_t^0 - M_j^0) + \mathcal{C}_j(M_t^0 - M_j^0, M_t^0 - M_j^0),$$

we have

$$\begin{aligned}&\sum_{j=1}^T \omega_j \left[\mathcal{C}_j(\widetilde{M}_t^\lambda - M_t^0, \widetilde{M}_t^\lambda - M_t^0) + \mathcal{C}_j(\widetilde{M}_t^\lambda - M_t^0, M_t^0 - M_j^0)\right] \\ \leq &\frac{1+\sqrt{2}}{2}\lambda\sqrt{\mathrm{rank}(M_t^0)}\|\widetilde{M}_t^\lambda - M_t^0\|_2.\end{aligned}$$

If $\|\widetilde{M}_t^\lambda - M_t^0\|_2 < \left\|M_t^0 - \sum_{j=1}^T \omega_j M_j^0\right\|_2$, we maintain the bound

$$\|\widetilde{M}_t^\lambda - M_t^0\|_2 < \delta_h M_t. \tag{S.11}$$

If $\|\widetilde{M}_t^\lambda - M_t^0\|_2 \geq \left\|M_t^0 - \sum_{j=1}^T \omega_j M_j^0\right\|_2$, using (S.4) and Lemma S.1, we know that

$$\begin{aligned}&\sum_{j=1}^T \omega_j \left[\mathcal{C}_j(\widetilde{M}_t^\lambda - M_t^0, \widetilde{M}_t^\lambda - M_t^0) + \mathcal{C}_j(\widetilde{M}_t^\lambda - M_t^0, M_t^0 - M_j^0)\right] \\ = &\left(\mathrm{vec}(\widehat{M}_t^\lambda) - \mathrm{vec}(M_t^0)\right)^\top \Sigma \left(\left(\mathrm{vec}(\widehat{M}_t^\lambda) - \mathrm{vec}(M_t^0)\right) + \left(M_t^0 - \sum_{j=1}^T \omega_j M_j^0\right)\right) \\ \geq &\mu\left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2^2 - \mu\left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2 \left\|M_t^0 - \sum_{j=1}^T \omega_j M_j^0\right\|_2.\end{aligned} \tag{S.12}$$

So we have

$$\mu\left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2^2 - \mu\left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2 \left\|M_t^0 - \sum_{j=1}^T \omega_j M_j^0\right\|_2 \leq \frac{1+\sqrt{2}}{2}\lambda\sqrt{\mathrm{rank}(M_t^0)}\left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2. \tag{S.13}$$

Combining (S.11) and (S.13), we have

$$\left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2 \leq \delta_h M_t + \frac{1+\sqrt{2}}{2}\frac{\lambda}{\mu}\sqrt{r_t}, \tag{S.14}$$

where $r_t = \mathrm{rank}(M_t^0)$. With (S.7) and (S.14), the proof is completed.

□

S.2 Proofs of Theorem 2 and Corollary 1, 2

Proposition 1 *Under Assumption 2 and 3, if $h \to 0$ and $(m_1 m_2)^{1/2} T h^2 \to \infty$ as $m_1, m_2, T \to \infty$, then*

$$\delta_h M_t \le \frac{1}{2}\alpha(K)(m_1 m_2)^{1/2} D_2 h^2 + o\left((m_1 m_2)^{1/2} h^2\right).$$

Proof of Proposition 1: Note that $T^{-1} \ll (m_1 m_2)^{-1/2} h^2$. The conclusion follows from the classical nonparametric results (of fixed design) that

$$\begin{aligned}
\left\| M_t^0 - \sum_{j=1}^T \omega_j M_j^0 \right\|_2 &= \left\| \sum_{j=1}^T \frac{1}{Th} K\left(\frac{j-t}{Th}\right) \left\{ M\left(\frac{j}{T}\right) - M\left(\frac{t}{T}\right) \right\} \right\|_2 \\
&= \left\| \int \frac{1}{h} K\left(\frac{u - t/T}{h}\right) \left\{ M(u) - M\left(\frac{t}{T}\right) \right\} du \right\|_2 + O\left(T^{-1}\right) \\
&= \left\| \int K(z) \left\{ M\left(zh + \frac{t}{T}\right) - M\left(\frac{t}{T}\right) \right\} dz \right\|_2 + O\left(T^{-1}\right) \\
&= \frac{1}{2}\alpha(K) \left\| \nabla^2 M(\frac{t}{T}) \right\|_2 h^2 + o\left((m_1 m_2)^{1/2} h^2\right) \\
&\le \frac{1}{2}\alpha(K)(m_1 m_2)^{1/2} D_2 h^2 + o\left((m_1 m_2)^{1/2} h^2\right).
\end{aligned}$$

□

Now we prove two lemmas which are needed to finish the proofs Lemma S.4 and Theorem 2. For brevity, define that $\|\cdot\| \triangleq \|\cdot\|_\infty$. We extend proposition 2 in Koltchinskii (2011b) to obtain the following Lemma S.2.

Lemma S.2 *Let $Z_1, Z_2, \ldots, Z_n \in (\mathcal{B}(\mathbb{R}^{m_1 \times m_2}), \mathbb{R}^{m_1 \times m_2}, \mu^{\otimes m_1 m_2})$ be independent, mean-zero random variables satisfying that*

$$\|Z_i\|_{L^2} \le \sigma, \quad \|Z_i\|_{\psi(\alpha)} \le K, \quad \|\cdot\|_{\psi(\alpha)} := \inf\left\{ t > 0, \mathbb{E}\left[\exp\left(\frac{\|\cdot\|^\alpha}{t^\alpha}\right)\right] \le 2 \right\}.$$

Let $\alpha > 1$ and $a = (a_1, a_2, \ldots, a_n)^\top \in \mathbb{R}^n$. There exists a constant $C > 0$ such that, for all $t > 0$, with probability at least $1 - e^{-t}$

$$\left\|\sum_{i=1}^{n} a_i Z_i\right\| \leq C \max\left(\sqrt{\|a\|_2^2\sigma^2(t + \log(2m))}, \|a\|K\left(\log\frac{K}{\sigma}\right)^{1/\alpha}(t + \log(2m))\right). \quad \text{(S.15)}$$

Proof of Lemma S.2: We follow the proof of proposition 2 in Koltchinskii (2011b) here.

Let

$$S_n := \sum_{i=1}^{n} a_i \widetilde{Z}_i, \quad \widetilde{Z} = \begin{pmatrix} 0 & Z \\ Z^\top & 0 \end{pmatrix} \in \mathbb{R}^{m\times m}.$$

We know that $\|S_n\| < s$ if and only if $-sI_m \prec S_n \prec sI_m$. Therefore,

$$P(\|S_n\| \geq s) = P(S_n \not\preceq sI_m) + P(S_n \not\succeq -sI_m)$$

and

$$P(S_n \not\preceq sI_m) = P\left(e^{\lambda S_n} \not\preceq e^{\lambda s I_m}\right) \leq P\left(\mathrm{Tr}(e^{\lambda S_n}) \geq e^{\lambda s}\right) \leq e^{-\lambda s}\mathbb{E}\,\mathrm{Tr}(e^{\lambda S_n}).$$

To bound $\mathbb{E}\,\mathrm{Tr}(e^{\lambda S_n})$, we use independence and Golden-Thompson inequality

$$\begin{aligned}
\mathbb{E}\,\mathrm{Tr}(e^{\lambda S_n}) &= \mathbb{E}\,\mathrm{Tr}\left(e^{\lambda S_{n-1} + \lambda a_n Z_n}\right) \\
&\leq \mathbb{E}\,\mathrm{Tr}\left(e^{\lambda S_{n_1}} e^{\lambda a_n Z_n}\right) \\
&= \mathrm{Tr}\left(\mathbb{E}e^{\lambda S_{n-1}}\mathbb{E}e^{\lambda a_n Z_n}\right) \\
&\leq \mathbb{E}\,\mathrm{Tr}\left(e^{\lambda S_{n-1}}\right)\left\|\mathbb{E}e^{\lambda a_n Z_n}\right\|.
\end{aligned}$$

By induction, we have

$$\mathbb{E}\,\mathrm{Tr}(e^{\lambda S_n}) \leq m\prod_{i=1}^{n}\left\|\mathbb{E}e^{\lambda a_i Z_i}\right\|.$$

It remains to bound the norm $\|\mathbb{E}e^{\lambda a_i Z_i}\|$. Using Taylor expansion and $\mathbb{E}(Z_i) = 0$, we have

$$\begin{aligned}
\mathbb{E}e^{\lambda a_i Z_i} &= I_m + \mathbb{E}\lambda^2 a_i^2 Z_i^2\left[\frac{1}{2!} + \frac{\lambda a_i Z_i}{3!} + \frac{\lambda^2 a_i^2 Z_i^2}{4!} + \ldots\right] \\
&\leq I_m + \lambda^2 a_i^2 \mathbb{E}Z_i^2\left[\frac{1}{2!} + \frac{\lambda a_i\|Z_i\|}{3!} + \frac{\lambda^2 a_i^2\|Z_i\|^2}{4!} + \ldots\right] \\
&= I_m + \lambda^2 a_i^2 \mathbb{E}Z_i^2\left[\frac{e^{\lambda a_i\|Z_i\|} - 1 - \lambda a_i\|Z_i\|}{\lambda^2 a_i^2\|Z_i\|^2}\right].
\end{aligned}$$

Therefore, for all $\tau > 0$ we have

$$\begin{aligned}
&\left\|\mathbb{E}e^{\lambda a_i Z_i}\right\| \\
&\leq 1 + \lambda^2 a_i^2 \left\|\mathbb{E}Z_i^2 \left[\frac{e^{\lambda a_i \|Z_i\|} - 1 - \lambda a_i \|Z_i\|}{\lambda^2 a_i^2 \|Z_i\|^2}\right]\right\| \\
&\leq 1 + \lambda^2 \left\|\mathbb{E}a_i^2 Z_i^2\right\| \left[\frac{e^{\lambda\tau} - 1 - \lambda\tau}{\lambda^2\tau^2}\right] + \lambda^2 \mathbb{E}\|a_i Z_i\|^2 \left[\frac{e^{\lambda a_i \|Z_i\|} - 1 - \lambda a_i \|Z_i\|}{\lambda^2 a_i^2 \|Z_i\|^2}\right] I(\|a_i Z_i\| \geq \tau).
\end{aligned}$$

Let

$$\tau = C_2 \|a_i\|_\infty K (\log \frac{K}{\sigma})^{1/\alpha}, \quad \lambda\tau \leq 1,$$

we have

$$\left\|\mathbb{E}e^{\lambda a_i Z_i}\right\| \leq 1 + C_1 \lambda^2 a_i^2 \sigma^2 \leq \exp\{C_1 \lambda^2 a_i^2 \sigma^2\}.$$

So when

$$P(\|S_n\| \geq s) \leq 2m \exp\{-\lambda s + C_1 \lambda^2 \|a\|_2^2 \sigma^2\}$$

and λ is chosen to be

$$\min\left\{\frac{1}{C_2 \|a\|_\infty K \left(\log K/\sigma\right)^{1/\alpha}}, \frac{s}{2C_1 \|a\|_2^2 \sigma^2}\right\},$$

which can directly deduce (S.15).

$\square$

Lemma S.3 *If X, Y are random variables with $\|X\|_{\psi(\alpha)} \leq K_1$, $\|Y\|_{\psi(\beta)} \leq K_2$, and there exists $\gamma \geq 1$ that $\frac{1}{\alpha} + \frac{1}{\beta} = \frac{1}{\gamma} \leq 1$, then we have*

$$\|XY\|_{\psi(\gamma)} \leq K_1 K_2. \tag{S.16}$$

Proof of Lemma S.3: Using Young's inequality we know that for all $\theta \in (0, 1)$

$$\frac{(\|X\|^\gamma)^{\alpha/\gamma}}{t^{\alpha\theta} \alpha/\gamma} + \frac{(\|Y\|^\gamma)^{\beta/\gamma}}{t^{\beta(1-\theta)} \beta/\gamma} \geq \frac{(\|XY\|)^\gamma}{t^\gamma}.$$

Thus we have

$$\mathbb{E}\left[\exp\left(\frac{\|XY\|^\gamma}{t^\gamma}\right)\right] \le \mathbb{E}\left[\exp\left(\frac{(\|X\|^\gamma)^{\alpha/\gamma}}{t^{\alpha\theta}\alpha/\gamma} + \frac{(\|Y\|^\gamma)^{\beta/\gamma}}{t^{\beta(1-\theta)}\beta/\gamma}\right)\right]$$
$$\le \frac{\gamma}{\alpha}\mathbb{E}\left[\exp\left(\frac{(\|X\|^\alpha}{t^{\alpha\theta}}\right)\right] + \frac{\gamma}{\beta}\mathbb{E}\left[\exp\left(\frac{(\|Y\|^\beta}{t^{\beta(1-\theta)}}\right)\right] \le 2,$$

where we choose t, θ by $t^\theta = K_1, t^{1-\theta} = K_2$.

So we have

$$\mathbb{E}\left[\exp\left(\frac{\|XY\|^\gamma}{(K_1K_2)^\gamma}\right)\right] \le 2.$$

□

We construct the following theorem to bound $\|\sum_{j=1}^T \omega_h(j-t)\Delta_j\|$ and Lemma S.4 is an immediate result of the following theorem. For brevity, we denote $\omega_j = \omega_h(j-t)$.

Theorem S.1 *Under Assumption 4-5, for all $t > 0$, with probability at least $1 - 3e^{-t}$,*

$$\left\|\sum_{j=1}^T \omega_j\Delta_j\right\| \le C_1 \max\left(\sigma\sqrt{\frac{(t+\log(2m))}{N}}, \frac{K\left(\log\frac{K}{\sigma}\right)^{1/\gamma}(t+\log(2m))}{W}\right)$$
$$+ C_2\mu_X \max\left(\sigma_\xi\sqrt{\frac{t}{N}}, \frac{K_1(\log\frac{K_1}{\sigma_\xi})^{1/\alpha}t}{W}\right)$$
$$+ C_3 \max\left(\varsigma\sqrt{\frac{(t+\log(2m))}{N}}, \frac{\mathcal{K}\left(\log\frac{\mathcal{K}}{\varsigma}\right)^{2/\beta}(t+\log(2m))}{W}\right),$$

where

$$N = \frac{1}{\sum_{j=1}^T \omega_j^2/n_j}, \quad W = \frac{1}{\max_j \omega_j/n_j}$$

and constants with respect to ω_j, n_j such that

$$K = \max\|\xi_{ji}(X_{ji} - \mathbb{E}X_{ji})\|_{\psi(\gamma)} \lesssim K_1K_2,$$

$$\sigma = \max\|\xi_{ji}(X_{ji} - \mathbb{E}X_{ji})\|_{L^2} \lesssim K_1K_2 2^{1/\gamma},$$

$$\mu_X = \max\|\mathbb{E}X_{ji}\| \lesssim K_2,$$

$$\sigma_\xi = \max\|\xi_{ji}\|_{L^2} \lesssim K_1 2^{1/\alpha},$$

$$\mathcal{K} = \max\|\langle M_j^0, X_{ji}\rangle X_{ji} - \mathbb{E}(\langle M_j^0, X_{ji}\rangle X_{ji})\|_{\psi(\beta/2)} \lesssim C_* K_2^2,$$

$$\varsigma = \max\|\langle M_j^0, X_{ji}\rangle X_{ji} - \mathbb{E}(\langle M_j^0, X_{ji}\rangle X_{ji})\|_{L^2} \lesssim C_* K_2^2 2^{2/\beta}.$$

Proof of Theorem S.1: Because that

$$
\begin{aligned}
\left\|\sum_{j=1}^{T}\omega_j\Delta_j\right\| &= \left\|\sum_{j=1}^{T}\sum_{i=1}^{n_j}\frac{1}{n_j}\omega_j\left(Y_{ji}X_{ji}-EY_{ji}X_{ji}\right)\right\| \\
&\leq \left\|\sum_{j=1}^{T}\sum_{i=1}^{n_j}\frac{1}{n_j}\omega_j\xi_{ji}X_{ij}\right\| \\
&+ \left\|\sum_{j=1}^{T}\sum_{i=1}^{n_j}\frac{1}{n_j}\omega_j\left\{\langle M_0(j),X_{ji}\rangle X_{ji}-E\left(\langle M_0(j),X_{ji}\rangle X_{ji}\right)\right\}\right\| \\
&\leq \left\|\sum_{j=1}^{T}\sum_{i=1}^{n_j}\frac{1}{n_j}\omega_j\xi_{ji}(X_{ij}-EX_{ij})\right\| + \left\|\sum_{j=1}^{T}\sum_{i=1}^{n_j}\frac{1}{n_j}\omega_j\xi_{ji}\mathbb{E}X_{ij}\right\| \\
&+ \left\|\sum_{j=1}^{T}\sum_{i=1}^{n_j}\frac{1}{n_j}\omega_j\left\{\langle M_0(j),X_{ji}\rangle X_{ji}-E\left(\langle M_0(j),X_{ji}\rangle X_{ji}\right)\right\}\right\| \\
&:= \mathcal{D}_1+\mathcal{D}_2+\mathcal{D}_3.
\end{aligned} \tag{S.17}
$$

a. For $\mathcal{D}_1$.

We denote that $Z_{ji} := \xi_{ji}(X_{ji}-\mathbb{E}X_{ji})$. Using Lemma S.3, Z_{ji} are independent, mean-zero random variables satisfying that

$$\|Z_{ji}\|_{\psi(\gamma)} \leq K, \quad \|Z_{ji}\|_{L^2} \leq \sigma(2)^{1/\gamma}.$$

Set $a_{ji} = \omega_j/n_j$. Using Lemma S.2, we have that for all $t>0$, with probability at least $1-e^{-t}$

$$\mathcal{D}_1 = \left\|\sum_{j=1}^{T}\sum_{i=1}^{n_j}a_{ji}Z_{ji}\right\| \leq C\max\left(\sigma\sqrt{\frac{(t+\log(2m))}{N}}, \frac{K\left(\log\frac{K}{\sigma}\right)^{1/\gamma}(t+\log(2m))}{W}\right). \tag{S.18}$$

b. For $\mathcal{D}_2$.

Directly using Lemma S.2 with $\|\mathbb{E}X_{ji}\| \leq \mu_X, \|\xi_{ji}\|_{L^2} \leq \sigma_\xi$ and $a_{ji} = \frac{\omega_j}{n_j}$, we know that for all $t \geq 0$, with probability at least $1-e^{-t}$

$$\mathcal{D}_2 = \left\|\sum_{j=1}^{T}\sum_{i=1}^{n_j}a_{ji}\mu_X\xi_{ji}\right\| \leq C\mu\max\left(\sigma_\xi\sqrt{\frac{t}{N}}, \frac{K_1(\log\frac{K_1}{\sigma_\xi})^{1/\alpha}t}{W}\right). \tag{S.19}$$

c. For $\mathcal{D}_3$. Denote that $Z_{ji} = \langle M_j^0, X_{ji}\rangle X_{ji} - \mathbb{E}(\langle M_j^0, X_{ji}\rangle X_{ji})$ and $a_{ji} = \frac{\omega_j}{n_j}$ with

$$\|Z_{ji}\|_{\psi(\beta/2)} \leq \mathcal{K}, \quad \|Z_{ji}\|_{L^2} \leq \varsigma.$$

Using Lemma S.2, we know that for all $t \geq 0$, with probability at least $1 - e^{-t}$

$$\mathcal{D}_3 = \left\| \sum_{j=1}^{T} \sum_{i=1}^{n_j} a_{ji} Z_{ji} \right\| \leq C \max\left(\varsigma \sqrt{\frac{(t + \log(2m))}{N}}, \frac{\mathcal{K}\left(\log \frac{\mathcal{K}}{\varsigma}\right)^{2/\beta}(t + \log(2m))}{W} \right). \tag{S.20}$$

From (S.18), (S.19), (S.20), we know that for all $t > 0$, with probability at least $1 - 3e^{-t}$,

$$\begin{aligned}
\left\| \sum_{j=1}^{T} \omega_j \Delta_j \right\| \leq\ & C_1 \max\left(\sigma \sqrt{\frac{(t + \log(2m))}{N}}, \frac{K\left(\log \frac{K}{\sigma}\right)^{1/\gamma}(t + \log(2m))}{W} \right) \\
& + C_2 \mu_X \max\left(\sigma_\xi \sqrt{\frac{t}{N}}, \frac{K_1(\log \frac{K_1}{\sigma_\xi})^{1/\alpha} t}{W} \right) \\
& + C_3 \max\left(\varsigma \sqrt{\frac{(t + \log(2m))}{N}}, \frac{\mathcal{K}\left(\log \frac{\mathcal{K}}{\varsigma}\right)^{2/\beta}(t + \log(2m))}{W} \right).
\end{aligned}$$

□

Lemma S.4 *Under Assumption 4 and 5, if* $h \to 0$ *and* $(m_1 m_2)^{1/2} T h^2 \to \infty$ *as* $m_1, m_2, T \to \infty$, $nTh \gg (K_*/\sigma_*)^2 \log(m_1 + m_2)$, *then with probability at least* $1 - 3/(m_1 + m_2)$,

$$\mathcal{W}_h \Delta_t \leq C_1 \sigma_* \sqrt{\frac{\log(m_1 + m_2)}{nTh}}, \tag{S.21}$$

where $C_1 > 0$ *is a constant independent to* $\sigma_*, K_*, T, h, m_1, m_2$ *such that*

$$\begin{aligned}
\sigma_* &= \max\left\{\sigma, \mu_X \sigma_\xi, \varsigma\right\}, \\
K_* &= \max\left\{ \mathcal{K}\left(\log \frac{\mathcal{K}}{\varsigma}\right)^{2/\beta}, \mu_X K_1 \left(\log \frac{K_1}{\sigma_\xi}\right)^{1/\alpha}, K\left(\log \frac{K}{\sigma}\right)^{1/\gamma} \right\},
\end{aligned}$$

in which $\mu_X = \max_{i,j} \|\mathbb{E}X_{ji}\|_\infty$, $\sigma_\xi = \max_{i,j} \|\xi_{ji}\|_{L^2}$ *and*

$$K = \max_{i,j} \|\xi_{ji}(X_{ji} - \mathbb{E}X_{ji})\|_{\psi(\gamma)}, \quad \mathcal{K} = \max_{i,j} \|\langle M_j^0, X_{ji}\rangle X_{ji} - \mathbb{E}(\langle M_j^0, X_{ji}\rangle X_{ji})\|_{\psi(\beta/2)},$$

$$\sigma = \max_{i,j} \|\xi_{ji}(X_{ji} - \mathbb{E}X_{ji})\|_{L^2}, \quad \varsigma = \max_{i,j} \|\langle M_j^0, X_{ji}\rangle X_{ji} - \mathbb{E}(\langle M_j^0, X_{ji}\rangle X_{ji})\|_{L^2},$$

with $\|X\|_{L^2} = \|\mathbb{E}XX^T\|_\infty^{1/2} \vee \|\mathbb{E}X^T X\|_\infty^{1/2}$.

Proof of Lemma S.4:

Define that

$$\mathcal{N}_h(t) = \{j : t - Th < j < t + Th\}, \quad \bar{n}_t = \frac{1}{|\mathcal{N}_h(t)|} \sum_{j \in \mathcal{N}_h(t)} n_j, \quad \check{n}_t = \min_{j \in \mathcal{N}_h(t)} n_j.$$

Recall the definition of ω_j that $\omega_j = \omega_h(j-t) = \frac{K\left(\frac{j-t}{Th}\right)}{\sum_{k=1}^T K\left(\frac{k-t}{Th}\right)}$, we have that

$$\begin{aligned}
N &= \frac{1}{\sum_{j=1}^T \omega_h(j-t)^2/n_j} = \frac{\left\{\sum_{k=1}^T \frac{1}{Th} K\left(\frac{k-t}{Th}\right)\right\}^2}{\sum_{j=1}^T \frac{1}{T^2h^2} K\left(\frac{j-t}{Th}\right)^2/n_j} \\
&\asymp \frac{1}{\sum_{j=1}^T \frac{1}{n_j T^2 h^2} K\left(\frac{j-t}{Th}\right)^2} \asymp \left\{\frac{1}{\bar{n}_t Th} \sum_{j=1}^T \frac{1}{Th} K\left(\frac{j-t}{Th}\right)^2\right\}^{-1} \\
&= \bar{n}_t Th R(K)^{-1},
\end{aligned}$$

and

$$\begin{aligned}
W &= \frac{1}{\max_j \omega_h(j-t)/n_j} = \left\{\max_j \frac{\frac{1}{Th} K\left(\frac{j-t}{Th}\right)}{n_j \frac{1}{Th} \sum_{k=1}^T K\left(\frac{k-t}{Th}\right)}\right\}^{-1} \\
&\asymp \left\{\max_j \frac{1}{n_j Th} K\left(\frac{j-t}{Th}\right)\right\}^{-1} \asymp \check{n}_t Th K(0)^{-1}.
\end{aligned}$$

With the assumption $\bar{n}_t, \check{n}_t \asymp n$, for $t \leq \log(m_1+m_2)$, when $nTh \gg \{K_*/\sigma_*\}^2 \log(m_1+m_2)$, the first terms in (S.18), (S.19) and (S.20) dominate those bounds which gives

$$\mathcal{W}_h \Delta_t \leq C_1 \sigma_* \sqrt{\frac{s + \log(m_1+m_2)}{nTh}} \ll 1.$$

By setting $t = \log(m_1+m_2)$, we can deduce Lemma S.4 directly.

$\square$

Proof of Theorem 2: With Lemma S.4 and Theorem 1, we have

$$\left(\delta_h M_t^2 + 2\mu^{-1} \mathcal{W}_h \Delta_t \left\|M_t^0\right\|_1\right)^{1/2} \gg (1+\sqrt{2})\mu^{-1} \mathcal{W}_h \Delta_t \sqrt{r_t},$$

which leads to the following conclusion. From Lemma S.4, we know that for all $t > 0$, with probability at least $1 - 3e^{-t}$,

$$\left\|\sum_{j=1}^T \omega_j \Delta_j\right\|_\infty \leq C_1 \max\left\{\sigma^* \sqrt{\frac{t + \log(m_1+m_2)}{\bar{n}_t Th}}, K^* \frac{t + \log(m_1+m_2)}{\check{n}_t Th}\right\}.$$

Choose that

$$t = \log(m_1 + m_2), \quad \lambda = 2\sqrt{2} C_1 \sigma^* \sqrt{\frac{\log(m_1 + m_2)}{\bar{n}_t T h}}.$$

If $(\check{n}_t^2 / \bar{n}_t) T h \gg \{K^*/\sigma^*\}^2 \log(m_1 + m_2)$, than

$$\sigma^* \sqrt{\frac{t + \log(m_1 + m_2)}{\bar{n}_t T h}} \gg K^* \frac{t + \log(m_1 + m_2)}{\check{n}_t T h}.$$

From Proposition 1

$$(m_1 m_2)^{-1/2} \left\| M_t^0 - \sum_{j=1}^{T} \omega_j M_j^0 \right\|_2 \le \frac{1}{2} \alpha(K) D_2 h^2.$$

With Theorem 1, so

$$(m_1 m_2)^{-1/2} \left\| \widetilde{M}_t^\lambda - M_t^0 \right\|_2 \le \frac{1}{2} C_2 \alpha(K) D_2 h^2 + \left(2 + \sqrt{2}\right) C_1 \sigma^* \left(\frac{r_t \log(m_1 + m_2)}{\mu^2 m_1 m_2 \bar{n}_t T h} \right)^{1/2}$$

holds with probability at least $1 - 3/(m_1 + m_2)$.

□

Proof of Corollary 1: Plug in the upper bound of $\mathcal{W}_h \Delta_t$ in Lemma S.4, Corollary 1 can be obtained immediately. □

Proof of Corollary 2:

From the distribution of X_{ji} and Assumption 4, we have that

$$\gamma = \alpha, \quad \beta = \infty, \quad \mu = (m_1 m_2)^{-1},$$

and

$$\sigma \le \sigma_\xi \sqrt{\frac{1}{m_1 \wedge m_2}}, \quad \mu_X \sigma_\xi = \sigma_\xi \sqrt{\frac{1}{m_1 \wedge m_2}}, \quad \varsigma \le M \sqrt{\frac{1}{m_1 \wedge m_2}},$$
$$K \asymp \sigma_\xi, \quad K_1 \asymp \sigma_\xi, \quad \mathcal{K} \asymp M.$$

Then it can be obtained that

$$\sigma^* = (C_M \vee \sigma_\xi) \sqrt{\frac{1}{m_1 \wedge m_2}},$$
$$K^*/\sigma^* \asymp (m_1 \wedge m_2)^{1/2} \log^{1/\alpha}(m_1 \wedge m_2).$$

Then $(K^*/\sigma^*)^2\log(m_1+m_2)\asymp(m_1\wedge m_2)\log^{2/\alpha}(m_1\wedge m_2)\log(m_1+m_2)$. Note that $\log(m_1+m_2)>\log(m_1\wedge m_2)$, when $(\check{n}_t^2/\bar{n}_t)Th\gg(m_1\wedge m_2)\log^{1+2/\alpha}(m_1+m_2)$, the first term in (S.21) dominates the bound. Apply Theorem 2 and Corollary 1 and the proof is finished.

□

S.3 Heterogeneous Scenario

Consider the heterogeneous assumption first.

Assumption S.1 *(Heterogeneous Assumption) The second moment matrix Σ_t of X_t is positive definite with smallest eigenvalue $\mu_t>0$ for $t=1,2,\cdots,T$.*

Here we provide the corresponding results of Theorem 1, Theorem 2 and Corollary 1 under the heterogeneous scenario. The key difference between homogeneous and heterogeneous scenarios is that the covariance matrix used for time points t is Σ in homogeneous case and $\sum_{j=1}^T\omega_h(j-t)\Sigma_j$ in heterogeneous case. Define $\nu_t=\sum_{j=1}^T\omega_h(j-t)\mu_t$, and the difference is using ν_t to replace μ in each theorems and corollaries for the upper bound of $\widetilde{M}_t^\lambda$.

Theorem S.2 *(Heterogeneous Theorem 1) Under Assumption S.1, if $\lambda\geq 2\mathcal{W}_h\Delta_t$, then*

$$\left\|\widetilde{M}_t^\lambda-M_t^0\right\|_2\leq\delta_hM_t+\min\left\{\left((\delta_hM_t)^2+\frac{2\lambda}{\nu_t}\left\|M_t^0\right\|_1\right)^{1/2},\frac{1+\sqrt{2}}{2}\frac{\lambda}{\nu_t}\sqrt{r_t}\right\}. \tag{S.22}$$

When selecting $\lambda=2\mathcal{W}_h\Delta_t$, we have

$$\left\|\widetilde{M}_t^\lambda-M_t^0\right\|_2\leq\delta_hM_t+\min\left\{\left((\delta_hM_t)^2+2\nu_t^{-1}\mathcal{W}_h\Delta_t\left\|M_t^0\right\|_1\right)^{1/2},(1+\sqrt{2})\nu_t^{-1}\mathcal{W}_h\Delta_t\sqrt{r_t}\right\}. \tag{S.23}$$

Proof In the Proof of Theorem 1, the whole proof is not changed but changing (S.5) as

$$\left(\operatorname{vec}(\widehat{M}_t^\lambda)-\operatorname{vec}(M_t^0)\right)^\top \Sigma\left[\left(\operatorname{vec}(\widehat{M}_t^\lambda)-\operatorname{vec}(M_t^0)\right)+\left(2\operatorname{vec}(M_t^0)-2\sum_{j=1}^T\omega_j\operatorname{vec}(M_j^0)\right)\right]$$

$$\geq \nu_t\left\|\widetilde{M}_t^\lambda-M_t^0\right\|_2^2-\nu_t\left\|\widetilde{M}_t^\lambda-M_t^0\right\|_2\left\|M_t^0-\sum_{j=1}^T\omega_jM_j^0\right\|_2.$$

and changing (S.12) as

$$\begin{aligned}&\sum_{j=1}^T\omega_j\left[\mathcal{C}_j(\widetilde{M}_t^\lambda-M_t^0,\widetilde{M}_t^\lambda-M_t^0)+\mathcal{C}_j(\widetilde{M}_t^\lambda-M_t^0,M_t^0-M_j^0)\right]\\
=&\left(\operatorname{vec}(\widehat{M}_t^\lambda)-\operatorname{vec}(M_t^0)\right)^\top\Sigma\left(\left(\operatorname{vec}(\widehat{M}_t^\lambda)-\operatorname{vec}(M_t^0)\right)+\left(M_t^0-\sum_{j=1}^T\omega_jM_j^0\right)\right)\\
\geq&\nu_t\left\|\widetilde{M}_t^\lambda-M_t^0\right\|_2^2-\nu_t\left\|\widetilde{M}_t^\lambda-M_t^0\right\|_2\left\|M_t^0-\sum_{j=1}^T\omega_jM_j^0\right\|_2.\end{aligned}$$

□

Theorem S.3 *(Heterogeneous Theorem 2) Under Assumption 1-5, let* $nTh\gg(K_*/\sigma_*)^2\log(m_1+m_2)$, $h\to0$ *and* $(m_1m_2)^{1/2}Th^2\to\infty$ *as* $n,m_1,m_2,T\to\infty$, *when*

$$\lambda=2C_1\sigma_*\sqrt{\frac{\log(m_1+m_2)}{nTh}},$$

where C_1 *and* σ_* *are defined in Lemma S.4 of Supplementary Material, then with probability at least* $1-3/(m_1+m_2)$,

$$(m_1m_2)^{-1/2}\left\|\widetilde{M}_t^\lambda-M_t^0\right\|_2\leq\frac{1}{2}\alpha(K)D_2h^2+\left(1+\sqrt{2}\right)C_1\sigma_*\left(\frac{r_t\log(m_1+m_2)}{\nu_t^2m_1m_2nTh}\right)^{1/2}+o(h^2),\tag{S.24}$$

where $\alpha(K)$ *is defined in (10) and* D_2 *in Assumption 2.*

Corollary S.1 *(Heterogeneous Corollary 1) Under assumptions of Theorem S.3, when*

$$nT\gg\max\{K_*^{5/2}\sigma_*^{-3}(\nu_t^2m_1m_2)^{1/4}\log(m_1+m_2),\ \sigma_*^2(\nu_t^2m_1m_2)^{-1}\log(m_1+m_2)\},$$

with probability at least $1-3/(m_1+m_2)$,

$$(m_1m_2)^{-1/2}\left\|\widetilde{M}_t^\lambda-M_t^0\right\|_2\leq C_2\left(\frac{\sigma_*^2r_t\log(m_1+m_2)}{\nu_t^2m_1m_2nT}\right)^{2/5},\tag{S.25}$$

where $C_2=1/2\left[(2+2\sqrt{2})C_1\right]^{4/5}[\alpha(K)D_2]^{1/5}$.

S.4 Definition of $\phi-$mixing

The dependence of two σ-field $\mathcal{A}$ and $\mathcal{B}$ is measured by

$$\phi(\mathcal{A},\mathcal{B}) = \sup\left\{\left|P(A) - \frac{P(A\cap B)}{P(B)}\right|; A\in\mathcal{A}, B\in\mathcal{B}, P(B)\neq 0\right\}.$$

And for a sequence of σ-field $\mathcal{A}_j, j\in\mathbb{N}^*$, the ϕ-coefficients $\phi_{\mathcal{A}}(k), k\in\mathbb{N}$ is defined as

$$\phi_{\mathcal{A}}(k) = \sup_{|j-s|\leq k}\phi(\mathcal{A}_j,\mathcal{A}_s).$$

If $\lim_{k\to\infty}\phi_{\mathcal{A}}(k)=0$, the sequence of σ-field $\mathcal{A}_j, j\in\mathbb{N}^*$ is said to be ϕ-mixing.

S.5 Proofs of Theorem 3 and Corollary 3, S.2

Lemma S.5 *Under Assumption 4 and 6, if $h\to 0$ as $n,T\to\infty$, $Th\gg\log(m_1+m_2)$ and $nTh\gg(K_*/\sigma_*)^2\log^3(m_1+m_2)$, with probability at least $1-3/(m_1+m_2)$,*

$$\mathcal{W}_h\Delta_t\leq\mathcal{C}_1(\Phi_{\mathcal{X}}\vee\Phi_{\mathcal{Y}})\sigma_*\sqrt{\frac{\log(m_1+m_2)}{nTh}},$$

where $\mathcal{C}_1>0$ is a constant independent to $\sigma_*, K_*, T, h, m_1, m_2$.

Proof of Lemma S.5 and Theorem 3: Denote $\mathbf{X}_j=(X_{j1},X_{j2},\ldots,X_{jn_j})^\top, \xi_j=(\xi_{j1},\xi_{j2},\ldots,\xi_{jn_j})^\top$. We know that $\sigma(\mathbf{X}_j-\mathbb{E}\mathbf{X}_j)\subseteq\mathcal{X}_j$. Define $\mathcal{Z}_j$ such that $\mathcal{Z}_j=\sigma((\mathbf{X}_j^\top,\xi_j^\top)^\top)$, then

$$\begin{aligned}
&\left|P(A_x\otimes A_y)-\frac{P(A_x\otimes A_y\cap B_x\otimes B_y)}{P(B_x\otimes B_y)}\right|\\
=&\left|P(A_x)\left(P(A_y)-\frac{P(A_y\cap B_y)}{P(B_y)}\right)+\frac{P(A_y\cap B_y)}{P(B_y)}\left(P(A_x)-\frac{P(A_x\cap B_x)}{P(B_x)}\right)\right|\\
\leq&\phi_{\mathcal{X}}(|j-l|)+\phi_{\mathcal{Y}}(|j-l|)-\phi_{\mathcal{X}}(|j-l|)\phi_{\mathcal{Y}}(|j-l|),
\end{aligned}$$

where $A_x\otimes A_y\in\mathcal{Z}_j,\ B_x\otimes B_y\in\mathcal{Z}_l$. So we know that $\phi_{\mathcal{Z}}\leq\phi_{\mathcal{X}}+\phi_{\mathcal{Y}}-\phi_{\mathcal{X}}\phi_{\mathcal{Y}}$, which means that

$$\Phi_{\mathcal{Z}}=\sum_{k=1}^{\infty}\sqrt{\phi_{\mathcal{Z}}(k)}\leq\Phi_{\mathcal{X}}+\Phi_{\mathcal{Y}}<\infty.$$

Let $Z_{ji} = \frac{\omega_j}{n_j}\xi_{ji}\left(X_{ji} - \mathbb{E}X_{ji}\right)$ and $Z_j = \sum_{i=1}^{n_j} Z_{ji}$. Using Lemma S.2, we know that

$$P\left(\|Z_j\| \geq s\right) \leq 2m\exp\left\{-C\min\left\{\frac{n_j s^2}{w_j^2\sigma^2}, \frac{n_j s}{\omega_j K(\log\frac{K}{\sigma})^{1/\gamma}}\right\}\right\}.$$

Denote the bounded function

$$g((\mathbf{X}_j^\top, \xi_j^\top)^\top) = \frac{\omega_j}{n_j}(\mathbf{X}_j - \mathbb{E}\mathbf{X}_j)^\top \xi_j \mathbf{1}_{\left\|\frac{\omega_j}{n_j}(\mathbf{X}_j - \mathbb{E}\mathbf{X}_j)^\top\xi_j\right\| \leq s},$$

we know that

$$\begin{aligned} P\left(\left\|\sum_{j=1}^T Z_j\right\| \geq t\right) &\leq \sum_{j=1}^T P(\|Z_j\| \geq s) + P\left(\left\|\sum_{j=1}^T Z_j \mathbf{1}_{\|Z_j\| \leq s}\right\| \geq t\right) \\ &= \sum_{j=1}^T P(\|Z_j\| \geq s) + P\left(\left\|\sum_{j=1}^\top g((\mathbf{X}_j^\top, \xi_j^\top)^\top)\right\| \geq t\right). \end{aligned}$$

With the tail distribution assumptions for X_j, ξ_j, when $s \leq \frac{\omega_j\sigma^2}{K(\log\frac{K}{\sigma})^{1/\gamma}}$, we have the inequalities

$$\begin{aligned} \left\|\mathbb{E}\sum_{j=1}^T g((\mathbf{X}_j^\top, \xi_j^\top)^\top)\right\| &= \left\|\sum_{j=1}^T Z_j \mathbf{1}_{\|Z_j\| \geq s}\right\| \\ &\leq \sum_{j=1}^T \left|\int_s^\infty P(\|Z_j\| \geq s)dx + sP(\|Z_j\| \geq s)\right| \\ &\leq \left(s + \frac{c}{snT^2h^2}\right)\sum_{j=1}^T P(\|Z_j\| \geq s). \end{aligned}$$

And when $s \geq \frac{\omega_j\sigma^2}{K(\log\frac{K}{\sigma})^{1/\gamma}}$, we have that

$$\left\|\mathbb{E}\sum_{j=1}^T g((\mathbf{X}_j^\top, \xi_j^\top)^\top)\right\| \leq \left(s + \frac{cK(\log\frac{K}{\sigma})^{1/\gamma}}{nTh}\right)\sum_{j=1}^T P(\|Z_j\| \geq s).$$

Meanwhile we have the bound

$$\left\|\mathbb{E}\sum_{j=1}^T g^2((\mathbf{X}_j^\top, \xi_j^\top)^\top)\right\| \leq \sum_{j=1}^T \frac{\omega_j^2}{n_j}\sigma_\xi^2\sigma_X^2 = \frac{\sigma^2}{N}.$$

Similar to Theorem 3 in Samson (2000), with the matrix value function $g(\cdot)$ and norm $\|\cdot\|_\infty$

$$\begin{aligned}
&P\left(\left\|\sum_{j=1}^{T} g((\mathbf{X}_j^\top, \xi_j^\top)^\top)\right\| \geq t\right) \\
\leq &P\left(\left\|\sum_{j=1}^{T} g((\mathbf{X}_j^\top, \xi_j^\top)^\top) - \mathbb{E}\sum_{j=1}^{T} g((\mathbf{X}_j^\top, \xi_j^\top)^\top)\right\| \geq t - \left\|\mathbb{E}\sum_{j=1}^{T} g((\mathbf{X}_j^\top, \xi_j^\top)^\top)\right\|\right) \\
\leq &2m \exp\left\{-\frac{1}{8\Phi_{\mathcal{Z}}^2} \min\left\{\frac{t}{2s}, \frac{Nt^2}{16\sigma^2}\right\}\right\},
\end{aligned}$$

when $t \geq \left\|\mathbb{E}\sum_{j=1}^{T} g((\mathbf{X}_j^\top, \xi_j^\top)^\top)\right\|$.

Thus we have

$$\begin{aligned}
&P\left(\left\|\sum_{j=1}^{T} Z_j\right\| \geq t\right) \\
\leq &\sum_{j=(t-h/2)\vee 0}^{(t+h/2)\wedge T} 2m \exp\left\{-C \min\left\{\frac{n_j s^2}{w_j^2\sigma^2}, \frac{n_j s}{\omega_j K (\log\frac{K}{\sigma})^{1/\gamma}}\right\}\right\} \\
&+2m \exp\left\{-\frac{1}{8\Phi_{\mathcal{Z}}^2} \min\left\{\frac{t}{2s}, \frac{Nt^2}{16\sigma^2}\right\}\right\} \\
\leq &\sum_{j=(t-h/2)\vee 0}^{(t+h/2)\wedge T} 2m \exp\left\{-\frac{C n_j s^2}{\omega_j^2\sigma^2}\right\} + 2m \exp\left\{-\frac{1}{8\Phi_{\mathcal{Z}}^2} \min\left\{\frac{t}{2s}, \frac{Nt^2}{16\sigma^2}\right\}\right\} \\
\lesssim &2m \exp\left\{-\frac{CnT^2h^2s^2}{\sigma^2} + \log Th\right\} + 2m \exp\left\{-\frac{1}{8\Phi_{\mathcal{Z}}^2} \min\left\{\frac{t}{2s}, \frac{Nt^2}{16\sigma^2}\right\}\right\}
\end{aligned}$$

when $t \geq \left\|\mathbb{E}\sum_{j=1}^{T} g((\mathbf{X}_j^\top, \xi_j^\top)^\top)\right\|$ and $s \leq \frac{\omega_j\sigma^2}{K(\log\frac{K}{\sigma})^{1/\gamma}}$.

And similarly

$$\begin{aligned}
&P\left(\left\|\sum_{j=1}^{T} Z_j\right\| \geq t\right) \\
\lesssim &\ 2m \exp\left\{-C\frac{nThs}{K(\log\frac{K}{\sigma})^{1/\gamma}} + \log Th\right\} + 2m \exp\left\{-\frac{1}{8\Phi_{\mathcal{Z}}^2} \min\left\{\frac{t}{2s}, \frac{Nt^2}{16\sigma^2}\right\}\right\}
\end{aligned}$$

when $t \geq \left\|\mathbb{E}\sum_{j=1}^{T} g((\mathbf{X}_j^\top, \xi_j^\top)^\top)\right\|$ and $s \geq \frac{\omega_j\sigma^2}{K(\log\frac{K}{\sigma})^{1/\gamma}}$.

When

$$n \gtrsim \left(K(\log K/\sigma)^{1/\gamma}/\sigma\right)^2 \log(Thm),$$

choose

$$s \asymp n^{-1/2}(Th)^{-1}\sigma \log^{1/2}(Thm), \quad t \asymp n^{-1/2}(Th)^{-1/2}\sigma\Phi_{\mathcal{Z}} \log^{1/2} m.$$

If

$$Th \gg \Phi_{\mathcal{Z}}^2 \log(Thm) \log m,$$

we have that $t \gg s\Phi_{\mathcal{Z}}^2 \log m$.

Meanwhile, when

$$n \lesssim \left(K(\log K/\sigma)^{1/\gamma}/\sigma\right)^2 \log(Thm),$$

choose

$$s \asymp n^{-1}(Th)^{-1}K(\log K/\sigma)^{1/\gamma} \log(Thm), \quad t \asymp n^{-1/2}(Th)^{-1/2}\sigma\Phi_{\mathcal{Z}} \log^{1/2} m.$$

If

$$nTh \gg (K(\log K/\sigma)^{1/\gamma}/\sigma)^2\Phi_{\mathcal{Z}}^2 \log^2(Thm) \log m,$$

we have that $t \gg s\Phi_{\mathcal{Z}}^2 \log m$.

Thus when

$$Th \gg \Phi_{\mathcal{Z}}^2 \log(Thm) \log m$$

and

$$nTh \gg (K(\log K/\sigma)^{1/\gamma}/\sigma)^2\Phi_{\mathcal{Z}}^2 \log^2(Thm) \log m,$$

the inequality

$$P\left(\left\|\sum_{j=1}^{T}\sum_{i=1}^{n_j}\frac{\omega_j}{n_j}\xi_{ji}(X_{ji} - \mathbb{E}(X_{ji}))\right\| \geq t\right) \leq \frac{1}{m}.$$

holds where $t \asymp n^{-1/2}(Th)^{-1/2}\sigma\Phi_{\mathcal{Z}} \log^{1/2} m$.

Similarly, we choose $t \asymp n^{-1/2}(Th)^{-1/2}\sigma_\xi\mu_X\Phi_{\mathcal{Y}} \log^{1/2} m$ and

$$P\left(\left\|\sum_{j=1}^{T}\sum_{i=1}^{n_j}\frac{\omega_j}{n_j}\xi_{ji}\mathbb{E}(X_{ji})\right\| \geq t\right) \leq \frac{1}{m}$$

holds when $Th \gg \Phi_{\mathcal{Y}}^2 \log(Thm) \log m$ and

$$nTh \gg (K_1(\log K_1/\sigma_\xi)^{1/\alpha}/\sigma_\xi)^2\Phi_{\mathcal{Y}}^2 \log^2(Thm) \log m.$$

And choose $t \asymp n^{-1/2}(Th)^{-1/2}\varsigma\Phi_{\mathcal{X}}\log^{1/2} m$, we have

$$P\left(\left\|\sum_{j=1}^{T}\sum_{i=1}^{n_j}\frac{\omega_j}{n_j}\left\{\langle M_0(j), X_{ji}\rangle X_{ji} - E\left(\langle M_0(j), X_{ji}\rangle X_{ji}\right)\right\}\right\| \geq t\right) \leq \frac{1}{m}$$

holds when $Th \gg \Phi_{\mathcal{X}}^2 \log(Thm)\log m$ and $nTh \gg (\mathcal{K}(\log\mathcal{K}/\varsigma)^{2/\beta}/\varsigma)^2\Phi_{\mathcal{X}}^2\log^2(Thm)\log m$.

Summarizing the above results with $\Phi_{\mathcal{Z}} \leq \Phi_{\mathcal{X}} + \Phi_{\mathcal{Y}}$, we know that when $Th \gg (\Phi_{\mathcal{X}}^2 \vee \Phi_{\mathcal{Y}}^2)\log(Thm)$ and $nTh \gg (K^*/\sigma^*)^2(\Phi_{\mathcal{X}}^2 \vee \Phi_{\mathcal{Y}}^2)\log^2(Thm)\log m$, there exists a constant $\mathcal{C}_1 > 0$, with probability at least $1 - 3/m$,

$$\left\|\sum_{j=1}^{T}\omega_j\Delta_j\right\| \leq \mathcal{C}_1(\Phi_{\mathcal{X}} \vee \Phi_{\mathcal{Y}})\sigma^*\sqrt{\frac{\log(m_1+m_2)}{nTh}}.$$

Choose

$$\lambda = 2\mathcal{C}_1(\Phi_{\mathcal{X}} \vee \Phi_{\mathcal{Y}})\sigma^*\sqrt{\frac{\log(m_1+m_2)}{nTh}}$$

and using Theorem 1 and proof of Theorem 2, we have that

$$(m_1m_2)^{-1/2}\left\|\widetilde{M}_t^{\lambda} - M_t^0\right\|_2 \leq \frac{1}{2}\alpha(K)D_2h^2 + (1+\sqrt{2})\mathcal{C}_1(\Phi_{\mathcal{X}} \vee \Phi_{\mathcal{Y}})\sigma^*\left(\frac{r_t\log(m_1+m_2)}{\mu^2 m_1m_2nTh}\right)^{1/2}.$$

Note that $Th \gg \log(Th)$ when $Th \to \infty$, we can represent the condition $Th \gg (\Phi_{\mathcal{X}}^2 \vee \Phi_{\mathcal{Y}}^2)\log(Thm)$ and $nTh \gg (K^*/\sigma^*)^2(\Phi_{\mathcal{X}}^2 \vee \Phi_{\mathcal{Y}}^2)\log^2(Thm)\log m$ as $Th \gg (\Phi_{\mathcal{X}}^2 \vee \Phi_{\mathcal{Y}}^2)\log m$ and $nTh \gg (K^*/\sigma^*)^2(\Phi_{\mathcal{X}}^2 \vee \Phi_{\mathcal{Y}}^2)\log^3 m$ respectively.

□

Corollary S.2 *Under assumptions of Theorem 3, when X_{ji} are i.i.d. uniformly distributed on $\mathcal{E}$, ξ_{ji} are independently follow sub-exponential mean-zero distributions, $Th \gg \log(m_1+m_2)$ and $nTh \gg (m_1 \wedge m_2)\log^{3+2/\alpha}(m_1+m_2)$, then with probability at least $1-3/(m_1+m_2)$,*

$$(m_1m_2)^{-1/2}\left\|\widetilde{M}_t^{\lambda} - M_t^0\right\|_2 \leq \frac{1}{2}\alpha(K)D_2h^2 + (1+\sqrt{2})\mathcal{C}_1(\Phi_{\mathcal{X}} \vee \Phi_{\mathcal{Y}})(C_M \vee \sigma_\xi)\left(\frac{r_t(m_1 \vee m_2)\log(m_1+m_2)}{nTh}\right)^{1/2}.$$

When $nT \gg (m_1 \vee m_2)\log^{(1+5/2\alpha)}(m_1+m_2)$ *and* $n^{-1}T^4 \gg (m_1 \vee m_2)^{-1}\log^4(m_1+m_2)$*, we select*

$$h = \mathcal{C}_h\left(\frac{(C_M \vee \sigma_\xi)^2(\Phi_{\mathcal{X}} \vee \Phi_{\mathcal{Y}})^2 r_t(m_1 \vee m_2)\log(m_1+m_2)}{nT}\right)^{1/5},$$

then

$$(m_1m_2)^{-1/2}\left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2 \le \mathcal{C}_2(\Phi_{\mathcal{X}} \vee \Phi_{\mathcal{Y}})^{4/5}(C_M \vee \sigma_\xi)^{4/5}\left(\frac{r_t(m_1 \vee m_2)\log(m_1+m_2)}{nT}\right)^{2/5}, \tag{S.26}$$

where $\mathcal{C}_1, \mathcal{C}_h, \mathcal{C}_2$ *are the same constants as above.*

The proofs of Corollary 3 and S.2 are similar to those of Corollary 1 and 2, which we do not repeat.

S.6 Application to Compressed Sensing

Corollary S.3 *(Independent case) Under Assumption 1-5, when* X_{ji} *are random matrices with independent mean-zero sub-gaussian elements with variance* σ_X^2*,* ξ_{ji} *are independently follow sub-gaussian distributions,* $nTh \gg m_1 \vee m_2$*,* $h \to 0$ *and* $(m_1m_2)^{1/2}Th^2 \to \infty$ *as* $n, m_1, m_2, T \to \infty$*, with probability at least* $1 - 2/(m_1+m_2)$*,*

$$\begin{aligned}&(m_1m_2)^{-1/2}\left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2 \\ &\le \frac{1}{2}\alpha(K)D_2h^2 + \left(1+\sqrt{2}\right)C_1\left[\eta \vee \{M(m_1m_2)^{1/2}\}\right]\left(\frac{r_t\log(m_1+m_2)}{(m_1 \wedge m_2)nTh}\right)^{1/2} + o(h^2),\end{aligned}$$

When $nT \gg (m_1 \vee m_2)\log(m_1+m_2)$*, let*

$$h = C_h\left(\frac{\{\eta^2 \vee (M^2m_1m_2)\}\, r_t\log(m_1+m_2)}{(m_1 \wedge m_2)nT}\right)^{1/5},$$

then

$$(m_1m_2)^{-1/2}\left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2 \le C_2\left(\frac{\{\eta^2 \vee (M^2m_1m_2)\}\, r_t\log(m_1+m_2)}{(m_1 \wedge m_2)nT}\right)^{2/5}, \tag{S.27}$$

where $\eta = \sigma_\xi/\sigma_X$ *and* C_1, C_h, C_2 *are the same constants as above.*

Theorem 6 in Koltchinskii (2011b) presented the error bound for static matrix compressed sensing which is

$$(m_1m_2)^{-1/2}\left\|\widehat{M}_t^\lambda - M_t^0\right\|_2 \le C_4\eta\left(\frac{r_t}{(m_1\wedge m_2)n}\right)^{1/2}.$$

Similarly to the matrix completion, when $T \gg \eta^{-1}\{\eta^2\vee(M^2m_1m_2)\}\{n(m_1\wedge m_2)\}^{1/4}\log(m_1+m_2)$, our dynamic method gives a sharper bound than the static method.

From Corollary 2 and Corollary S.3, when the variances of observation errors ξ_{ji} are small enough comparing to the variances of X_{ji}, i.e., $\sigma_\xi \le M$ in matrix completion setting and $\eta \le M(m_1m_2)^{1/2}$ in compressed sensing setting, the upper bound (15) and (S.27) have the same order such that

$$(m_1m_2)^{1/2}\left\|\widetilde{M}_t^\lambda - M_t^0\right\|_2 \le C_2M^{4/5}\left(\frac{r_t(m_1\vee m_2)\log(m_1+m_2)}{nT}\right)^{2/5}.$$

Proof of Corollary S.3: Because $\mathbb{E}X_{ji} = 0$, we know that $\mu_X = 0$ and (b) in (S.17) becomes zero. From the distribution of X_{ji} and Assumption 4, we know that $\beta = 2$ and $K_2 \asymp (m_1\vee m_2)^{1/2}\sigma_X$ by Theorem 4.4.5 in Vershynin (2018). Also, it is easy to check that $K \asymp \sigma_\xi\sigma_X(m_1\vee m_2)^{1/2},\ \sigma = \sigma_\xi\sigma_X(m_1\vee m_2)^{1/2}$. With that

$$\begin{aligned}&\left\|\mathbb{E}(\langle M_j^0, X_{ji}\rangle X_{ji} - \mathbb{E}(\langle M_j^0, X_{ji}\rangle X_{ji})(\langle M_j^0, X_{ji}\rangle X_{ji} - \mathbb{E}(\langle M_j^0, X_{ji}\rangle X_{ji})^T\right\|_\infty\\ \le& M^2m_1m_2(m_1\vee m_2)\sigma_X^4\end{aligned}$$

and with Lemma S.3

$$\begin{aligned}\left\|\langle M_j^0, X_{ji}\rangle X_{ji} - \mathbb{E}\langle M_j^0, X_{ji}\rangle X_{ji}\right\|_{\psi(1)} &\asymp \left\|\langle M_j^0, X_{ji}\rangle X_{ji}\right\|_{\psi(1)}\\ &\le \left\|\langle M_j^0, X_{ji}\rangle\right\|_{\psi(2)}\left\|X_{ji}\right\|_{\psi(2)}\\ &\le M(m_1m_2)^{1/2}\sigma_X^2,\end{aligned}$$

we have that $\varsigma \le M(m_1m_2)^{1/2}(m_1\vee m_2)^{1/2}\sigma_X^2, \mathcal{K} \lesssim M(m_1m_2)^{1/2}\sigma_X^2$.

Then it can be obtained that

$$\sigma^* = \left[\sigma_\xi \vee M\sigma_X (m_1 m_2)^{1/2}\right](m_1 \vee m_2)^{1/2}\sigma_X$$

$$K^*/\sigma^* \asymp (m_1 \vee m_2)^{-1/2} \log(1/(m_1 \vee m_2)).$$

Apply Theorem 2 and Corollary 1, the proof is completed. □

Corollary S.4 *(Dependent case.) Under assumptions of Theorem 3, when X_{ji} are random matrices with independent mean-zero sub-gaussian elements with variance σ_X^2 and ξ_{ji} are independently follow sub-gaussian distributions, $Th \gg \log(m_1+m_2)$ and $nTh \gg (m_1 \vee m_2)$, with probability at least $1 - 2/(m_1+m_2)$,*

$$\begin{aligned}
&(m_1 m_2)^{-1/2} \left\| \widetilde{M}_t^\lambda - M_t^0 \right\|_2 \\
&\le \frac{1}{2}\alpha(K) D_2 h^2 + \left(1 + \sqrt{2}\right) \mathcal{C}_1 (\Phi_\mathcal{X} \vee \Phi_\mathcal{Y}) \left[\eta \vee M (m_1 m_2)^{1/2}\right] \left(\frac{r_t \log(m_1+m_2)}{(m_1 \wedge m_2) nTh}\right)^{1/2}.
\end{aligned}$$

When $nT \gg (m_1 \vee m_2)\log(m_1+m_2)$ and $n^{-1}T^4 \gg (m_1 \vee m_2)^{-1}\log^4(m_1+m_2)$, we select

$$h = \mathcal{C}_h \left(\frac{(\eta^2 \vee M^2 m_1 m_2)(\Phi_\mathcal{X} \vee \Phi_\mathcal{Y})^2 r_t \log(m_1+m_2)}{(m_1 \wedge m_2) nT}\right)^{1/5},$$

then

$$(m_1 m_2)^{1/2} \left\| \widetilde{M}_t^\lambda - M_t^0 \right\|_2 \le \mathcal{C}_2 (\Phi_\mathcal{X} \vee \Phi_\mathcal{Y})^{4/5} \left(\frac{(\eta^2 \vee M^2 m_1 m_2)\, r_t \log(m_1+m_2)}{(m_1 \wedge m_2) nT}\right)^{2/5},$$

where $\eta = \sigma_\xi/\sigma_X$ and $\mathcal{C}_1, \mathcal{C}_h, \mathcal{C}_2$ are the same constants as above.

The proof of Corollary S.4 is similar to that of Corollary S.3, which we do not repeat.

S.7 Proof of Theorem 4

Proof of Theorem 4: First we know that

$$\nabla f_t(M) = 2\sum_{j=1}^{T} \omega_j \left[\frac{1}{n_j}\sum_{i=1}^{n_j} \mathbb{E}\langle M, X_{ji}\rangle X_{ji} - \frac{1}{n_j}\sum_{i=1}^{n_j} Y_{ji} X_{ji}\right].$$

This gives

$$\|\nabla f_t(M) - \nabla f_t(N)\|_2 = \left\| 2\sum_{j=1}^{T} \frac{\omega_j}{n_j} \sum_{i=1}^{n_j} \mathbb{E}\langle M - N, X_{ji}\rangle X_{ji} \right\|_2$$
$$\leq 2 \left\| \sum_{j=1}^{T} \frac{\omega_j}{n_j} \sum_{i=1}^{n_t} (\|X_{ji}\|_2 X_{ji}) \right\|_2 \|M - N\|_2.$$

Define that $2L_f = 4 \left\| \sum_{j=1}^{T} \omega_h(j-t)/n_j \sum_{i=1}^{n_j} (\|X_{ji}\|_2 X_{ji}) \right\|_2$, Toh & Yun (2010) proved that for any $k > 1$,

$$F_t(M_t^{(k)}) - F_t(\widetilde{M}_t^{\lambda}) \leq \frac{2L_f \|M_t^{(0)} - \widetilde{M}_t^{\lambda}\|_2^2}{(k+1)^2}. \tag{S.28}$$

Note that (S.28) focuses on the convergence rate of object function $F_t(M_t^{(k)})$ instead of $M_t^{(k)}$.from (S.28) we know

$$\sum_{j=1}^{T} \omega_j \mathcal{C}_j(M_t^{(k)}, M_t^{(k)}) - 2\sum_{j=1}^{T} \omega_j \left\langle \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ji} X_{ji}, M_t^{(k)} \right\rangle + \lambda \|M_t^{(k)}\|_1$$
$$\leq \sum_{j=1}^{T} \omega_j \mathcal{C}_j(\widehat{M}_t^{\lambda}, \widehat{M}_t^{\lambda}) - 2\sum_{j=1}^{T} \omega_j \left\langle \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ji} X_{ji}, \widehat{M}_t^{\lambda} \right\rangle + \lambda \|\widehat{M}_t^{\lambda}\|_1 + \frac{2L_f \gamma_t^2}{(k+1)^2}$$
$$\leq \sum_{j=1}^{T} \omega_j \mathcal{C}_j(M, M) - 2\sum_{j=1}^{T} \omega_j \left\langle \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ji} X_{ji}, M \right\rangle + \lambda \|M\|_1 + \frac{2L_f \gamma_t^2}{(k+1)^2}$$

for any $M \in \mathbb{M}$. Therefore, we have

$$\sum_{j=1}^{T} \omega_j \mathcal{C}_j(M_t^{(k)} - M_j^0, M_t^{(k)} - M_j^0)$$
$$\leq \sum_{j=1}^{T} \omega_j \mathcal{C}_j(M - M_j^0, M - M_j^0) + \left\langle 2\sum_{j=1}^{T} \omega_j \Delta_j, M_t^{(k)} - M \right\rangle$$
$$+\lambda(\|M\|_1 - \|M_t^{(k)}\|_1) + \frac{2L_f \gamma_t^2}{(k+1)^2}.$$

With the assumption that $\lambda > 2\|\sum_{j=1}^{T} \omega_j \Delta_j\|_\infty$, we use the similar methods in proof of Theorem 1 and immediately know that

$$\left\| M_t^{(k)} - M_t^0 \right\|_2$$
$$\leq \left\| M_t^0 - \sum_{j=1}^{T} \omega_h(j-t) M_j^0 \right\|_2 + \left(\left\| M_t^0 - \sum_{j=1}^{T} \omega_h(j-t) M_j^0 \right\|_2^2 + \frac{2\lambda}{\mu} \|M_t^0\|_1 + \frac{2L_f \gamma_t^2}{\mu(k+1)^2} \right)^{1/2}.$$

Next, we consider $F_t(\widetilde{M}_t^\lambda + M) - F_t(\widetilde{M}_t^\lambda)$ with the decomposition

$$
\begin{aligned}
M &= P_{S_1^\perp} M P_{S_2^\perp} \oplus P_{S_1} M P_{S_2} \oplus P_{S_1} M P_{S_2^\perp} \oplus P_{S_1^\perp} M P_{S_2} \\
&:= M^\perp \oplus M^{//} \oplus M^{//1} \oplus M^{//2} \\
&= \oplus_{j=1}^{l_1} \sigma_j^\perp M_j^\perp \oplus_{j=1}^{l_2} \sigma_j^{//} M_j^{//} \oplus_{j=1}^{l_3} \sigma_j^{//1} M_j^{//1} \oplus_{j=1}^{l_4} \sigma_j^{//2} M_j^{//2},
\end{aligned}
$$

where $\widetilde{M}_t^\lambda = \sum_{i=1}^r \sigma_i u_i v_i^\top$ with support S_1, S_2, $P_{S_1^\perp} M P_{S_2^\perp} = \sum_{i=1}^{l_1} \sigma_i^\perp u_i^\perp (v_j^\perp)^\top$ and

$$
\begin{aligned}
&\widetilde{M}_t^\lambda + \theta M_j^\perp = \sum_{i=1}^r \sigma_i u_i v_i^\top + \theta u_j^\perp (v_j^\perp)^\top, \\
&\widetilde{M}_t^\lambda + \theta M_j^{//} = \sum_{i=1}^r \sigma_i u_i v_i^\top + \theta u_{\alpha_j} v_{\beta_j}^\top, \\
&\widetilde{M}_t^\lambda + \theta M_j^{//1} = \sum_{i \neq \alpha_{1,j}} \sigma_i u_i v_i^\top + \sqrt{\sigma_{\alpha_{1,j}}^2 + \theta^2}\, u_{\alpha_{1,j}} \left(\frac{\sigma_{\alpha_{1,j}}}{\sqrt{\sigma_{\alpha_{1,j}}^2 + \theta^2}} v_{\alpha_{1,j}} + \frac{\theta}{\sqrt{\sigma_{\alpha_{1,j}}^2 + \theta^2}} v_{\beta_{1,j}}^\perp \right)^\top, \\
&\widetilde{M}_t^\lambda + \theta M_j^{//2} = \sum_{i \neq \alpha_{2,j}} \sigma_i u_i v_i^\top + \sqrt{\sigma_{\alpha_{2,j}}^2 + \theta^2} \left(\frac{\sigma_{\alpha_{2,j}}}{\sqrt{\sigma_{\alpha_{2,j}}^2 + \theta^2}} u_{\alpha_{2,j}} + \frac{\theta}{\sqrt{\sigma_{\alpha_{2,j}}^2 + \theta^2}} u_{\beta_{2,j}}^\perp \right) v_{\alpha_{2,j}}^\top.
\end{aligned}
$$

Using the convexity of $F_t(\cdot)$, we have that for all $M, N \in \mathbb{M}$ and $B \in \partial F_t(M)$,

$$
F_t(N) \geq F_t(M) + \langle B, N - M \rangle \tag{S.29}
$$

and

$$
F_t(\widetilde{M}_t^\lambda + M) - F_t(\widetilde{M}_t^\lambda) \geq \sum_{\cdot \in \mathcal{A}} \left(F_t(\widetilde{M}_t^\lambda + M^\cdot) - F_t(\widetilde{M}_t^\lambda) \right) \tag{S.30}
$$

where $\mathcal{A} = \{\perp, //, //1, //2\}$. Because $\widetilde{M}_t^\lambda$ is the minimizer of $F_t(M)$, there exists a matrix $\widehat{B} \in \partial F_t(\widetilde{M}_t^\lambda)$ satisfies that for all $M \in \mathbb{M}$,

$$
\langle \widehat{B}, \widetilde{M}_t^\lambda - M \rangle \leq 0,
$$

which means that there exists

$$
\widehat{V}_t \in \partial \|\widetilde{M}_t^\lambda\|_1 = \sum_{j=1}^r u_j v_j^\top + P_{S_1^\perp} \widehat{W} P_{S_2^\perp}, \quad \|\widehat{W}\|_\infty \leq 1 \tag{S.31}
$$

satisfying that

$$2\sum_{j=1}^{T}\omega_j\left(\mathcal{C}_j(\widetilde{M}_t^{\lambda},\widetilde{M}_t^{\lambda}-M)-\left\langle\frac{1}{n_j}\sum_{i=1}^{n_j}Y_{ji}X_{ji},\widetilde{M}_t^{\lambda}-M\right\rangle\right)+\lambda\langle\widehat{V}_t,\widetilde{M}_t^{\lambda}-M\rangle\leq 0. \quad \text{(S.32)}$$

Set $\varepsilon>\xi>0$.

a. We know that for all $B_\xi\in\partial F_t(\widetilde{M}_t^{\lambda}+\xi M^{\perp})$,

$$\begin{aligned}&F_t(\widetilde{M}_t^{\lambda}+\varepsilon M^{\perp})-F_t(\widetilde{M}_t^{\lambda}+\xi M^{\perp})\\ \geq&(\varepsilon-\xi)\langle B_\xi,M^{\perp}\rangle\\ =&(\varepsilon-\xi)\left[2\sum_{j=1}^{T}\omega_j\left(\mathcal{C}_j(\widetilde{M}_t^{\lambda}+\xi M^{\perp},M^{\perp})-\left\langle\frac{1}{n_j}\sum_{i=1}^{n_j}Y_{ji}X_{ji},M^{\perp}\right\rangle\right)+\lambda\langle V_\xi,M^{\perp}\rangle\right]\end{aligned}$$

holds for all

$$\begin{aligned}&V_\xi\in\partial\|\widetilde{M}_t^{\lambda}+\xi M^{\perp}\|_1\\ =&\left\{\sum_{j=1}^{r}u_jv_j^{\top}+\sum_{j=1}^{l_1}u_j^{\perp}(v_j^{\perp})^{\top}+P_{(S_1\oplus_{j=1}^{l_1}u_j)^{\perp}}WP_{(S_2\oplus_{j=1}^{l_1}v_j)^{\perp}},\|W\|_\infty\leq 1\right\}.\end{aligned}$$

With (S.31) and (S.32), setting $M=\widetilde{M}_t^{\lambda}+M^{\perp}$, we know that

$$\begin{aligned}&2\sum_{j=1}^{T}\omega_j\left(\mathcal{C}_j(\widetilde{M}_t^{\lambda}+\xi M^{\perp},M^{\perp})-\left\langle\frac{1}{n_j}\sum_{i=1}^{n_j}Y_{ji}X_{ji},M^{\perp}\right\rangle\right)+\lambda\langle V_\xi,M^{\perp}\rangle\\ \geq&2\sum_{j=1}^{T}\omega_j\mathcal{C}_j(\xi M^{\perp},M^{\perp})+\lambda\langle V_\xi-\widehat{V},M^{\perp}\rangle\\ =&2\xi\sum_{j=1}^{T}\omega_j\mathcal{C}_j(M^{\perp},M^{\perp})\\ &+\lambda\left\langle\sum_{j=1}^{l_1}u_j^{\perp}(v_j^{\perp})^{\top}+P_{(S_1\oplus_{j=1}^{l_1}u_j)^{\perp}}WP_{(S_2\oplus_{j=1}^{l_1}v_j)^{\perp}}-P_{S_1^{\perp}}\widehat{W}P_{S_2^{\perp}},\sum_{j=1}^{l_1}\sigma_j^{\perp}u_j^{\perp}(v_j^{\perp})^{\top}\right\rangle\\ \geq&2\xi\mu\|M^{\perp}\|_2^2,\end{aligned}$$

in which we use that

$$\begin{aligned}&\left\langle\sum_{j=1}^{l_1}u_j^{\perp}(v_j^{\perp})^{\top}+P_{(S_1\oplus_{j=1}^{l_1}u_j)^{\perp}}WP_{(S_2\oplus_{j=1}^{l_1}v_j)^{\perp}}-P_{S_1^{\perp}}\widehat{W}P_{S_2^{\perp}},\sum_{j=1}^{l_1}\sigma_j^{\perp}u_j^{\perp}(v_j^{\perp})^{\top}\right\rangle\\ =&\sum_{j=1}^{l_1}\sigma_j^{\perp}\left(1-P_{u_j^{\perp}}P_{S_1^{\perp}}\widehat{W}P_{S_2^{\perp}}P_{v_j^{\perp}}\right)\geq 0.\end{aligned}$$

So we have

$$F_t(\widetilde{M}_t^\lambda + \varepsilon M^\perp) - F_t(\widetilde{M}_t^\lambda + \xi M^\perp) \geq 2(\varepsilon - \xi)\xi\mu\|M^\perp\|_2^2,$$

which means that

$$F_t(\widetilde{M}_t^\lambda + \varepsilon M^\perp) - F_t(\widetilde{M}_t^\lambda) \geq \mu\|M^\perp\|_2^2 \int_0^\varepsilon 2\xi d\xi = \varepsilon^2\mu\|M^\perp\|_2^2. \tag{S.33}$$

b. Similarly, we know that for all $B_\xi \in \partial F_t(\widetilde{M}_t^\lambda + \xi M^{//})$

$$\begin{aligned} & F_t(\widetilde{M}_t^\lambda + \varepsilon M^{//}) - F_t(\widetilde{M}_t^\lambda + \xi M^{//}) \\ \geq & (\varepsilon - \xi)\left[2\sum_{j=1}^T \omega_j \left(\mathcal{C}_j(\widetilde{M}_t^\lambda + \xi M^{//}, M^{//}) - \left\langle \frac{1}{n_j}\sum_{i=1}^{n_j} Y_{ji}X_{ji}, M^{//}\right\rangle\right) + \lambda\langle V_\xi, M^{//}\rangle\right] \end{aligned}$$

holds for all

$$V_\xi \in \partial\|\widetilde{M}_t^\lambda + \xi M^{//}\|_1 \supset \partial\|\widetilde{M}_t^\lambda\|_1.$$

Choose $V_\xi = \widehat{V}$, so

$$2\sum_{j=1}^T \omega_j \left(\mathcal{C}_j(\widetilde{M}_t^\lambda + \xi M^{//}, M^{//}) - \left\langle \frac{1}{n_j}\sum_{i=1}^{n_j} Y_{ji}X_{ji}, M^{//}\right\rangle\right) + \lambda\langle V_\xi, M^{//}\rangle \geq 2\xi\mu\|M^{//}\|_2^2,$$

which means that

$$F_t(\widetilde{M}_t^\lambda + \varepsilon M^{//}) - F_t(\widetilde{M}_t^\lambda) \geq \varepsilon^2\mu\|M^{//}\|_2^2. \tag{S.34}$$

c. Similarly, we have that for all $B_\xi \in \partial F_t(\widetilde{M}_t^\lambda + \xi M^{//1})$

$$\begin{aligned} & F_t(\widetilde{M}_t^\lambda + \varepsilon M^{//1}) - F_t(\widetilde{M}_t^\lambda + \xi M^{//1}) \\ \geq & (\varepsilon - \xi)\left[2\sum_{j=1}^T \omega_j \left(\mathcal{C}_j(\widetilde{M}_t^\lambda + \xi M^{//1}, M^{//1}) - \left\langle \frac{1}{n_j}\sum_{i=1}^{n_j} Y_{ji}X_{ji}, M^{//1}\right\rangle\right) + \lambda\langle V_\xi, M^{//1}\rangle\right] \end{aligned}$$

holds for all

$$\begin{aligned} & V_\xi \in \partial\|\widetilde{M}_t^\lambda + \xi M^{//1}\|_1 \\ = & \left\{\sum_{j\notin\alpha_1} u_j v_j^\top + \sum_{j=1}^{l_3} u_{\alpha_{1,j}}\widetilde{v}_j^\top + P_{S_1^\perp} W P_{(S_2/_{j=1}^{l_3} v_{\alpha_{1,j}} \oplus_{j=1}^{l_3} \widetilde{v}_j)^\perp}, \|W\|_\infty \leq 1\right\} \end{aligned}$$

with

$$\widetilde{v}_j = \frac{\sigma_{\alpha_{1,j}}}{\sqrt{\sigma^2_{\alpha_{1,j}} + \xi^2}} v_{\alpha_{1,j}} + \frac{\xi}{\sqrt{\sigma^2_{\alpha_{1,j}} + \xi^2}} v^{\perp}_{\beta_{1,j}}.$$

So we have

$$\begin{aligned}
&2\sum_{j=1}^{T}\omega_j\left(\mathcal{C}_j(\widetilde{M}_t^{\lambda} + \xi M^{//1}, M^{//1}) - \left\langle \frac{1}{n_j}\sum_{i=1}^{n_j} Y_{ji}X_{ji}, M^{//1}\right\rangle\right) + \lambda\langle V_\xi, M^{//1}\rangle \\
\geq &2\sum_{j=1}^{T}\omega_j\mathcal{C}_j(\xi M^{//1}, M^{//1}) + \lambda\langle V_\xi - \widehat{V}, M^{//1}\rangle \\
= &2\xi\sum_{j=1}^{T}\omega_j\mathcal{C}_j(M^{//1}, M^{//1}) \\
+ &\lambda\left\langle \sum_{j=1}^{l_3}\left(u_{\alpha_{1,j}}\widetilde{v}_j^T - u_{\alpha_{1,j}}v^T_{\alpha_{1,j}}\right) + P_{S_1^{\perp}}WP_{(S_2/_{j=1}^{l_3}v_{\alpha_{1,j}}\oplus_{j=1}^{l_3}\widetilde{v}_j)^{\perp}} - P_{S_1^{\perp}}\widehat{W}P_{S_2^{\perp}}\ ,\right. \\
&\left.\sum_{j=1}^{l_3}\sigma_j^{//1}u_{\alpha_{1,j}}v^{\perp}_{\beta_{1,j}}\right\rangle \\
\geq &2\xi\mu\|M^{//1}\|_2^2 + \lambda\sum_{j=1}^{l_3}\frac{\sigma_j^{//1}\xi}{\sqrt{\sigma^2_{\alpha_{1,j}} + \xi^2}}.
\end{aligned}$$

Then

$$F_t(\widetilde{M}_t^{\lambda} + \varepsilon M^{//1}) - F_t(\widetilde{M}_t^{\lambda} + \xi M^{//2}) \geq 2(\varepsilon - \xi)\xi\mu\|M^{//1}\|_2^2 + (\varepsilon - \xi)\lambda\sum_{j=1}^{l_3}\frac{\sigma_j^{//1}\xi}{\sqrt{\sigma^2_{\alpha_{1,j}} + \xi^2}},$$

which means that

$$\begin{aligned}
F_t(\widetilde{M}_t^{\lambda} + \varepsilon M^{//1}) - F_t(\widetilde{M}_t^{\lambda}) &\geq \mu\|M^{//1}\|_2^2\int_0^{\varepsilon}2\xi d\xi + \lambda\sum_{j=1}^{l_3}\int_0^{\varepsilon}\frac{\sigma_j^{//1}\xi}{\sqrt{\sigma^2_{\alpha_{1,j}} + \xi^2}}d\xi \\
&= \varepsilon^2\mu\|M^{//1}\|_2^2 + \lambda\sum_{j=1}^{l_3}\sigma_j^{//1}\left(\sqrt{\sigma^2_{\alpha_{1,j}} + \varepsilon^2} - \sigma_{\alpha_{1,j}}\right).
\end{aligned} \tag{S.35}$$

d. The result is similar to (S.35) as

$$F_t(\widetilde{M}_t^{\lambda} + \varepsilon M^{//2}) - F_t(\widetilde{M}_t^{\lambda}) \geq \varepsilon^2\mu\|M^{//2}\|_2^2 + \lambda\sum_{j=1}^{l_4}\sigma_j^{//2}\left(\sqrt{\sigma^2_{\alpha_{2,j}} + \varepsilon^2} - \sigma_{\alpha_{2,j}}\right). \tag{S.36}$$

Using (S.30), (S.33), (S.34), (S.35), (S.36), we know that

$$
\begin{aligned}
F_t(\widetilde{M}_t^\lambda + M) - F_t(\widetilde{M}_t^\lambda) &\geq \sum_{\cdot\in\mathcal{A}} \left(F_t(\widetilde{M}_t^\lambda + M^\cdot) - F_t(\widetilde{M}_t^\lambda) \right) \\
&\geq \mu \sum_{\cdot\in\mathcal{A}} \|M^\cdot\|_2^2 + \lambda \sum_{j=1}^{l_3} \sigma_j^{//1} \left(\sqrt{\sigma_{\alpha_{1,j}}^2 + 1} - \sigma_{\alpha_{1,j}} \right) \\
&+ \lambda \sum_{j=1}^{l_4} \sigma_j^{//2} \left(\sqrt{\sigma_{\alpha_{2,j}}^2 + 1} - \sigma_{\alpha_{2,j}} \right) \\
&\geq \mu \sum_{\cdot\in\mathcal{A}} \|M^\cdot\|_2^2.
\end{aligned} \tag{S.37}
$$

Meanwhile, we can easily check that $\langle M^\cdot, M^* \rangle = 0$ hold for all $\cdot, * \in \mathcal{A},\ \cdot \neq *$, so we have

$$
\|M\|_2^2 = \left\| \sum_{\cdot\in\mathcal{A}} M^\cdot \right\|_2^2 = \sum_{\cdot\in\mathcal{A}} \|M^\cdot\|_2^2. \tag{S.38}
$$

With (S.28), (S.37), (S.38) and set $M = M_t^{(k)} - \widetilde{M}_t^\lambda$, we have that

$$
\|M_t^{(k)} - \widetilde{M}_t^\lambda\|_2^2 \leq \frac{2L_f\gamma_t^2}{(k+1)^2\mu}.
$$

Finally, with Theorem 1, the proof is finished. □

S.8 Proof of Corollary 4

Proof of Corollary 4: From (18), under the conditions in Corollary 3, with probability at least $1 - 3/(m_1 + m_2)$,

$$
(m_1m_2)^{-1/2} \left\| \widetilde{M}_t^\lambda - M_t^0 \right\|_2 \leq \mathcal{C}_2 \left(\frac{\sigma^{*2}(\Phi_\mathcal{X} \vee \Phi_\mathcal{Y})^2 r_t \log(m_1+m_2)}{\mu^2 m_1 m_2 nT} \right)^{2/5}.
$$

When k_1 satisfies

$$
k_1 \geq \mathcal{C}_2^{-1} \left(\frac{\sigma^{*2}(\Phi_\mathcal{X} \vee \Phi_\mathcal{Y})^2 r_t \log(m_1+m_2)}{\mu^2 m_1 m_2 nT} \right)^{-2/5} \left(\frac{2L_f\gamma_1^2}{\mu m_1 m_2} \right)^{1/2}.
$$

then

$$
(m_1m_2)^{-1/2} \left(\frac{2L_f\gamma_t^2}{\mu(k_1+1)^2} \right)^{1/2} \leq \mathcal{C}_2 \left(\frac{\sigma^{*2}(\Phi_\mathcal{X} \vee \Phi_\mathcal{Y})^2 r_t \log(m_1+m_2)}{\mu^2 m_1 m_2 nT} \right)^{2/5}
$$

and thus

$$(m_1 m_2)^{-1/2} \left\| M_1^{(k_1)} - M_1^0 \right\|_2 \le 2\mathcal{C}_2 \left(\frac{\sigma^{*2}(\Phi_{\mathcal{X}} \vee \Phi_{\mathcal{Y}})^2 r_t \log(m_1+m_2)}{\mu^2 m_1 m_2 n T} \right)^{2/5}.$$

Similarly, for $t = 2, 3, \cdots, T$ in the random initial, we have

$$K_0 = Tk_1 \ge \mathcal{C}_2^{-1} \left(\frac{\sigma^{*2}(\Phi_{\mathcal{X}} \vee \Phi_{\mathcal{Y}})^2 r_t \log(m_1+m_2)}{\mu^2 m_1 m_2 n T} \right)^{-2/5} \left(\frac{2L_f \gamma_1^2}{\mu m_1 m_2} \right)^{1/2} T.$$

In the proposed initial strategy, for $t = 2, 3, \ldots, T$, when

$$k_t \ge \mathcal{C}_2^{-1} \left(\frac{\sigma^{*2}(\Phi_{\mathcal{X}} \vee \Phi_{\mathcal{Y}})^2 r_t \log(m_1+m_2)}{\mu^2 m_1 m_2 n T} \right)^{-2/5} \left(\frac{2L_f \gamma_t^2}{\mu m_1 m_2} \right)^{1/2},$$

we have

$$\gamma_t \le 3\mathcal{C}_2 (m_1 m_2)^{1/2} \left(\frac{\sigma^{*2}(\Phi_{\mathcal{X}} \vee \Phi_{\mathcal{Y}})^2 r_t \log(m_1+m_2)}{\mu^2 m_1 m_2 n T} \right)^{2/5} + \frac{D_1}{T},$$

which gives

$$(m_1 m_2)^{-1/2} \left\| M_t^{(k_t)} - M_1^0 \right\|_2 \le 2\mathcal{C}_2 \left(\frac{\sigma^{*2}(\Phi_{\mathcal{X}} \vee \Phi_{\mathcal{Y}})^2 r_t \log(m_1+m_2)}{\mu^2 m_1 m_2 n T} \right)^{2/5}.$$

Choosing

$$\begin{aligned}
K_1 = \sum_{t=1}^{T} k_t &= \mathcal{C}_2^{-1} \left(\frac{\sigma^{*2}(\Phi_{\mathcal{X}} \vee \Phi_{\mathcal{Y}})^2 r_t \log(m_1+m_2)}{\mu^2 m_1 m_2 n T} \right)^{-2/5} \left(\frac{2L_f}{\mu m_1 m_2} \right)^{1/2} \sum_{t=1}^{T} \gamma_t \\
&\le \mathcal{C}_2^{-1} \left(\frac{\sigma^{*2}(\Phi_{\mathcal{X}} \vee \Phi_{\mathcal{Y}})^2 r_t \log(m_1+m_2)}{\mu^2 m_1 m_2 n T} \right)^{-2/5} \left(\frac{2L_f}{\mu m_1 m_2} \right)^{1/2} \times \\
&\left(\gamma_1 + T \left(3\mathcal{C}_2 (m_1 m_2)^{1/2} \left(\frac{\sigma^{*2}(\Phi_{\mathcal{X}} \vee \Phi_{\mathcal{Y}})^2 r_t \log(m_1+m_2)}{\mu^2 m_1 m_2 n T} \right)^{2/5} + \frac{D_1}{T} \right) \right) \\
&= \mathcal{C}_2^{-1} \left(\frac{\sigma^{*2}(\Phi_{\mathcal{X}} \vee \Phi_{\mathcal{Y}})^2 r_t \log(m_1+m_2)}{\mu^2 m_1 m_2 n T} \right)^{-2/5} \left(\frac{2L_f}{\mu m_1 m_2} \right)^{1/2} (\gamma_1 + D_1) \\
&+ 3\sqrt{2}(\frac{L_f}{\mu})^{1/2} T,
\end{aligned}$$

and the proof is finished. □

For independent case, we need to obtain the error bound

$$(m_1 m_2)^{-1/2} \left\| M_t^{(k_t)} - M_t^0 \right\|_2 \le 2C_2 \left(\frac{\sigma_*^2 r_t \log(m_1+m_2)}{\mu^2 m_1 m_2 n T} \right)^{2/5} \tag{S.39}$$

and the parallel result is

Corollary S.5 *Under assumptions of Theorem 2, to attain* (S.39) *for each* $t = 1, \ldots, T$, *the total iteration step* K_0 *of random initial choice satisfies*

$$K_0 \geq \frac{1}{2} C_2^{-1} \left(\frac{\sigma_*^2 r_t \log(m_1 + m_2)}{\mu^2 m_1 m_2} \right)^{-2/5} \left(\frac{2L_f}{\mu m_1 m_2} \right)^{1/2} \gamma_1 n^{2/5} T^{7/5},$$

and total iteration step K_1 *of the proposed initial strategy satisfies*

$$K_1 \geq \frac{1}{2} C_2^{-1} \left(\frac{\sigma_*^2 r_t \log(m_1 + m_2)}{\mu^2 m_1 m_2} \right)^{-2/5} \left(\frac{2L_f}{\mu m_1 m_2} \right)^{1/2} (\gamma_1 + D_1)(nT)^{2/5} + 3\sqrt{2} \left(\frac{L_f}{\mu} \right)^{1/2} T.$$

The proof of Corollary S.5 and is similar to that of Corollary 4 and hence we omit it here.

S.9 Additional Numerical Results

S.9.1 Netflix Dataset with varying the choosing of number of time intervals T

We conduct a numerical analysis using Netflix data, encompassing 1034 movies that were viewed more than 25000 times. We sample 3000 users from those whose ratings occurred more than 30 times in the dataset. The analysis focus on the first 500000 ratings recorded from October 1998 to November 2005. These ratings are randomly divided, with 80% assigned as training data and the remaining 20% as test data. To explore the impact of varying T while keeping nT constant, we consider 10 different values of T ranging from 1 to 1000 and split the observations into T time intervals in chronological order. After applying our method, results are presented in Figure S.1 and Table S.1.

For sufficiently large T ($T \geq 50$), the variation in T exerts a negligible influence on precision, but it results in an increase in time consumption. When T is too small ($T < 50$), merging data across a substantial temporal range into a single time point can result in a loss of temporal information, rendering the utilization of data inefficient. This observation also corresponds to the dynamic nature inherent in the underlying low-rank matrix, signifying its dynamic essence rather than a static state.

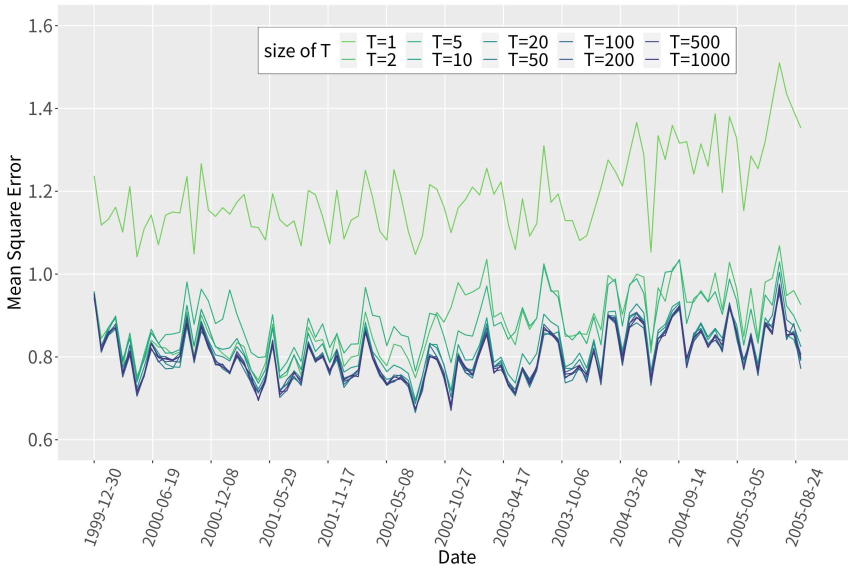

Figure S.1: Mean square error for different T in each time interval.

Table S.1: Average mean square error for different T.

T	1	2	5	10	20	50	100	200	500	1000
MSE*	1.194	0.881	0.886	0.825	0.812	0.800	0.803	0.802	0.801	0.802

S.9.2 Netflix Dataset with Link Function

In the real data experiment with Netflix dataset, each rating Y_i is a discrete integer from 1 to 5. Here we add experiment results with link function for the Netflix dataset. We assume that each user possesses a latent rating to each movie. Given the rating system's constraint of integer values ranging from 1 to 5, we consider the observed rating Y_i derived from a transformation φ on the corresponding latent rating Z_i, and assume that users typically tend to choose integer values in proximity to their latent ratings with a certain probability as their final ratings, i.e.,

$$Y_i = \varphi(Z_i) = \begin{cases} 1 & Z_i < 0.5 \\ \lfloor Z_i + 0.5 \rfloor + \delta_i & 0.5 \leq Z_i < 5.5 \\ 5 & Z_i \geq 5.5 \end{cases}$$

where $\delta_i, i = 1, 2, \cdots, n$ follow independent Bernoulli distributions with probability $Z_i - \lfloor Z_i + 0.5 \rfloor$ to take value 1 and $\lfloor \cdot \rfloor$ is the floor operator. And we estimate M_t^0 by

$$\widetilde{M}_t^\lambda = \underset{M \in \mathbb{M}}{\operatorname{argmin}} \, \mathbb{E}_\delta \sum_{j=1}^{T} \frac{\omega_h(j-t)}{n_j} \sum_{i=1}^{n_j} \left[\mathbb{E}(\varphi(\langle M, X_{ji} \rangle))^2 - 2Y_{ji}\varphi\left(\langle X_{ji}, M \rangle\right) \right] + \lambda \|M\|_1.$$

We followed a similar procedure and conducted new experiments with results presented in Figure S.2. It is interesting to note that the inclusion of this link function performs worse. One possible reason might be that the link function may not accurately reflect the true mechanism of users' ratings. Thus, in this example, we opted to not use such a link function to illustrate the proposed method. However, this does not necessarily imply a link function is not needed in general, e.g., a more flexible form of link function by nonparametric techniques can be considered for future research.

S.9.3 Image Classification

An experiment was conducted utilizing the CIFAR-10 dataset to assess the efficacy of the trace regression framework in improving subsequent image-related tasks, such as classification or regression. The CIFAR-10 dataset is publicly available at https://www.cs.toronto.edu/~kriz/cifar.html. This evaluation is particularly relevant when the original dataset is plagued by incomplete information and noise. Employing a similar approach as detailed in the real-world example involving video data processing, we treated each image's individual channel as a matrix denoted by M. Subsequently, we employed the Robust Principal Component

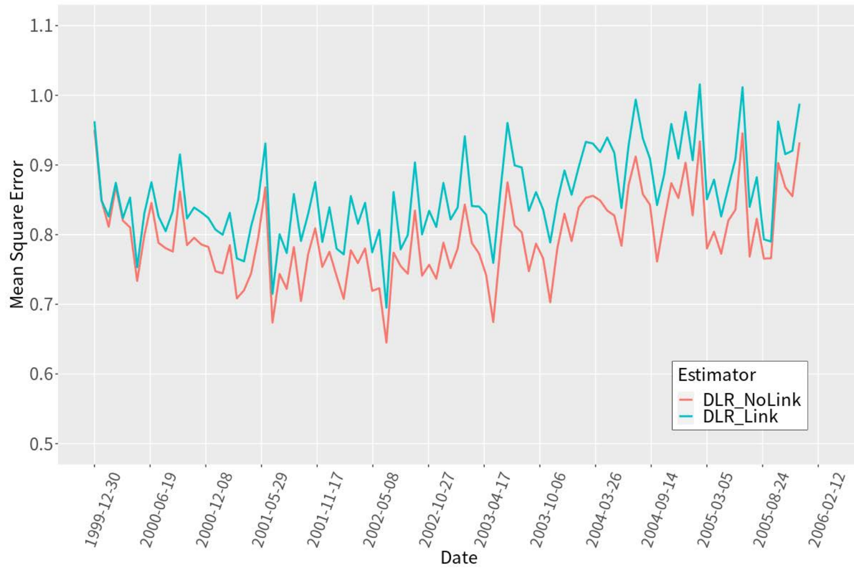

Figure S.2: MSE with and without link functions

Analysis (RPCA) method to acquire the sparse (S) and low-rank (L) components, respectively. For the purpose of dataset generation, we randomly preserved a subset of elements from the low-rank component (L), subjecting these elements to perturbation using i.i.d. normal noises. The recovery process involved applying our method to retrieve the low-rank component from the saved subset of elements in L and subsequently adding the sparse component S. We compared the performance of directly using the compressed data with utilizing the data obtained after image recovery based on the trace regression framework. For the classification tasks, we utilized LeNet and ResNet18 models. The corresponding results are presented in Table S.2. In the table, ρ represents the compression rate, and Acc denotes the classification accuracy in the test dataset. The findings suggest that prior image recovery before classification can lead to an enhancement in classification accuracy by approximately 3% to 5% for the aforementioned compressed datasets.

Table S.2: Comparison of classification results

ρ	Network	No recovery Acc	Recovery Acc
18.5%	LeNet	50.59%	55.94%
18.5%	ResNet	74.57%	75.75%
39.1%	LeNet	52.86%	58.7%
39.1%	ResNet	76.83%	80.40%
58.6%	LeNet	54.64%	61.03%
58.6%	ResNet	77.24%	82.1%
100%	LeNet	59.09%	61.56%
100%	ResNet	80.69%	83.41%

S.9.4 The selection of λ and other parameters in the FISTA algorithm

We first remark that the optimal tuning parameter λ can be chosen for each time points t. The penalty term can be relaxed to depend on t, and the theoretical analysis is still valid with slight modification. In practice, based on our experience with extensive numerical studies, the universal chosen λ would not produce similarity in the low-rank structure in terms of ranks and estimated eigen-space, and can substantially reduce the computational cost associated with tuning hyper-parameters. Theoretically, considering the heterogeneous case in Assumption S.1 as an example and assuming that the change of ν_t over t is not very large, i.e. there exist ν to bound ν_t such that $(1-c)\nu < \nu_t < (1+C)\nu$ with constants $c, C > 0$. With the same order assumption for sample size $n_t \asymp n$, λ and h has the optimal selection from Theorem 2 and (17) for each t by

$$h_t = C_{h,t}\left(\frac{\sigma_*^2 r_t \log(m_1+m_2)}{\nu_t^2 m_1 m_2 nT}\right)^{1/5}, \quad \lambda_t = 2C_{1,t}\sigma_*\sqrt{\frac{\log(m_1+m_2)}{nTh}}, \tag{S.40}$$

where $C_{h,t}$ and $C_{1,t}$ are related to M_t^0. With regularity and smoothness assumption 2, there exist $C_h, C_1 > 0$ to bound $C_{h,t}, c_{1,t}$ as their change is also not very large. So we can choose a universal bandwidth h and tuning parameter λ with the error bound for $\widetilde{M}_t^\lambda$ optimal up to a constant scalar.

Next, we present the detailed parameters in the FISTA algorithm of the simulation. In our proposed method, we set the tolerance (tor) to around 1×10^{-3}, which is a commonly used for the termination of optimization. For the initial time point $t = 1$, we employ the following strategy,

$$[M_1^{(0)}]_{ij} = \begin{cases} 0, & \text{if the } (i,j) \text{ element is not observed at time 1,} \\ Y_{1k}, & \text{if the } (i,j) \text{ element is observed at time 1, using the observation } Y_{1k}. \end{cases}$$

Then the output of last time $(t-1)$ is used as the initial matrix for current time t, i.e.,

$$M_t^{(0)} = M_{t-1}^{(k_{t-1})}.$$

The Lipchitz constant L_f here has an upper bound which can be calculated directly as

$$L_f = 2 \left\| \sum_{j=1}^{T} \frac{\omega_h(j-t)}{n_j} \sum_{i=1}^{n_j} (\|X_{ji}\|_2 X_{ji}) \right\|_2,$$

and the detail proof can be found in Supplementary S.6. These parameters are set in the same way for the benchmark, independent and dependent cases.

We employ a 5-fold cross-validation (CV) approach to select the tuning parameter λ in both independent and dependent settings. In our empirical findings, we observed that the results exhibit minimal differences when λ falls within a certain suitable range. To mitigate computational complexity in tuning the parameter, we suggest employing a coarse grid to estimate a rough range for λ and subsequently using a finer grid to make the final selection. When altering the matrix size, sample size, or the number of simulation points, it is not necessary to re-select λ using CV. Instead, we can rely on the theoretical relationship

$$\lambda = 2C_1\sigma_* \sqrt{\frac{\log(m_1+m_2)}{nTh}}$$

to adjust the new λ accordingly.

S.9.5 Rank Changes in Simulation and Real Data Experiment

For the simulation, we show the rank of recovered matrices in independent setting with $\rho = 0.2$ and $\sigma = 1$ in Table S.3 after truncating those dimensions with single values smaller than σ. For the Netflix dataset, the rank for Filter 1 is in Table S.4 and for Filter 2 is in Table S.5 after truncating those dimensions with single values smaller than $3\sqrt{1035}$. For the lions video dataset, the rank for low rank part L is in Table S.6 after truncating those dimensions with single values smaller than $0.5\sqrt{480}$.

Table S.3: Rank for Simulation in independent setting with $\rho = 0.2$ and $\sigma = 1$

T	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Rank	12	10	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
T	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Rank	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
T	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Rank	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
T	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
Rank	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
T	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Rank	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	11

Table S.4: Rank for Netflix dataset using Filter 1

T	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Rank	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
T	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Rank	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
T	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Rank	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
T	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
Rank	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
T	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Rank	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

Table S.5: Rank for Netflix dataset using Filter 2

T	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Rank	3	3	3	3	4	4	3	3	4	4	4	4	4	3	4	4	3	3	4	3
T	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Rank	3	3	3	3	4	3	3	4	3	3	3	3	3	3	3	3	3	3	3	3
T	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Rank	4	3	3	3	3	3	4	3	3	3	3	3	3	3	4	4	3	4	3	3
T	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
Rank	3	3	4	3	4	3	4	4	4	3	3	3	3	3	3	4	4	4	4	3
T	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Rank	4	4	3	4	4	4	3	3	4	3	4	4	4	4	4	4	4	3	3	3

Table S.6: Rank for lions video dataset

T	5	25	45	65	85
Rank (Channel Red)	6	5	6	5	6
Rank (Channel Green)	6	4	5	4	6
Rank (Channel Blue)	4	4	4	4	4

References

Koltchinskii, V. (2011b), 'Von neumann entropy penalization and low-rank matrix estimation', The Annals of Statistics **39**(6), 2936–2973.

Samson, P.-M. (2000), 'Concentration of measure inequalities for markov chains and phi-mixing processes', The Annals of Probability **28**(1), 416–461.

Toh, K.-C. & Yun, S. (2010), 'An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems', Pacific Journal of optimization **6**(615-640), 15.

Vershynin, R. (2018), High-dimensional probability: An introduction with applications in data science, Vol. 47, Cambridge university press.

Watson, G. A. (1992), 'Characterization of the subdifferential of some matrix norms', Linear algebra and its applications **170**(0), 33–45.