
Matrix Completion under Low-Rank Missing Mechanism

Fudan University, Texas A&M University, Peking University

Supplementary Material

This document provides supplementary material to the article "Matrix Completion under Low-Rank Missing Mechanism" written by the same authors.

S1 Appendix

S1.1 Estimation of $\Theta_\star$ (Cont')

To solve the optimization (3.4), we perform a easier way than the optimization of (3.1). Given the update $(\widehat{\mu}, \boldsymbol{Z}_{\text{old}})$ from previous iteration, a quadratic approximation of the objective function $-\ell_{\boldsymbol{W}}(\mu\boldsymbol{J} + \boldsymbol{Z}) + \lambda'\|\boldsymbol{Z}\|_*$ is formed:

$$\begin{aligned} P'_L\{\boldsymbol{Z}, \boldsymbol{Z}_{\text{old}}\} = &- \ell_{\boldsymbol{W}}(\widehat{\mu}\boldsymbol{J} + \boldsymbol{Z}_{\text{old}}) \\ &+ \langle \boldsymbol{Z} - \boldsymbol{Z}_{\text{old}}, -\nabla_{\boldsymbol{Z}}\ell_{\boldsymbol{W}}(\widehat{\mu}\boldsymbol{J} + \boldsymbol{Z}_{\text{old}})\rangle + \frac{L}{2}\|\boldsymbol{Z} - \boldsymbol{Z}_{\text{old}}\|_F^2 + \lambda'\|\boldsymbol{Z}\|_*, \end{aligned}$$

where $L > 0$ is an algorithmic parameter determining the step size of the proximal gradient algorithm, and is chosen by a backtracking method (Beck and Teboulle, 2009).

In this iterative algorithm, an successive update of $(\widehat{\mu}, \boldsymbol{Z})$ can be obtained by

$$\underset{\|\boldsymbol{Z}\|_\infty \leq \beta}{\arg\min} \quad P_L' \left\{ \boldsymbol{Z}, \boldsymbol{Z}_{\text{old}} \right\},$$

where the optimization is only required for $\boldsymbol{Z}$. Thus we need to solve the following optimization.

$$\underset{\|\boldsymbol{Z}\|_\infty \leq \beta}{\arg\min} \quad \langle \boldsymbol{Z} - \boldsymbol{Z}_{\text{old}}, -\nabla_{\boldsymbol{Z}} \ell_{\boldsymbol{W}} \left(\widehat{\mu} \boldsymbol{J} + \boldsymbol{Z}_{\text{old}} \right) \rangle + \frac{L}{2} \left\| \boldsymbol{Z} - \boldsymbol{Z}_{\text{old}} \right\|_F^2 + \lambda' \left\| \boldsymbol{Z} \right\|_*,$$

which is equivalent to

$$\underset{\|\boldsymbol{Z}\|_\infty \leq \beta}{\arg\min} \quad \frac{1}{2} \left\| \boldsymbol{Z} - \boldsymbol{Z}_{\text{old}} - \frac{1}{L} \nabla_{\boldsymbol{Z}} \ell_{\boldsymbol{W}} \left(\widehat{\mu} \boldsymbol{J} + \boldsymbol{Z}_{\text{old}} \right) \right\|_F^2 + \frac{\lambda'}{L} \left\| \boldsymbol{Z} \right\|_*. \tag{S1.1}$$

We apply a two-block alternative direction method of multipliers (ADMM) to an equivalent form of (S1.1):

$$\underset{\boldsymbol{Z} = \boldsymbol{G}, \|\boldsymbol{G}\|_\infty \leq \beta}{\arg\min} \quad \frac{\lambda'}{L} \left\| \boldsymbol{Z} \right\|_* + \frac{1}{2} \left\| \boldsymbol{G} - \boldsymbol{Z}_{\text{old}} - \frac{1}{L} \nabla_{\boldsymbol{Z}} \ell_{\boldsymbol{W}} \left(\widehat{\mu} \boldsymbol{J} + \boldsymbol{Z}_{\text{old}} \right) \right\|_F^2. \tag{S1.2}$$

The augmented Lagrangian for (S1.2) is

$$\begin{aligned} \mathcal{L}_u \left(\boldsymbol{Z}, \boldsymbol{G}; \boldsymbol{H} \right) = & \frac{\lambda'}{L} \left\| \boldsymbol{Z} \right\|_* + \frac{1}{2} \left\| \boldsymbol{G} - \boldsymbol{Z}_{\text{old}} - \frac{1}{L} \nabla_{\boldsymbol{Z}} \ell_{\boldsymbol{W}} \left(\widehat{\mu} \boldsymbol{J} + \boldsymbol{Z}_{\text{old}} \right) \right\|_F^2 \\ & - \langle \boldsymbol{H}, \boldsymbol{Z} - \boldsymbol{G} \rangle + \frac{u}{2} \left\| \boldsymbol{Z} - \boldsymbol{G} \right\|_F^2 + \mathbb{I}_{\left[\|\boldsymbol{G}\|_\infty \leq \beta \right]}, \end{aligned}$$

where $u > 0$ is an algorithmic parameter. The detailed algorithm to solve (S1.2) is summarized in Algorithm 2.

Proof of Theorem 1. It has been proved in (Chen et al., 2016) that the orthogonality of any two coefficients lie in the linear constraints will lead to the con-

vergence of the direct extension 3-block ADMM. In our case, this assumption is fulfilled due to the constraint $\boldsymbol{Z} = \boldsymbol{G}_1 = \boldsymbol{G}_2$. It has been shown in Theorem 2.4 (ii) of Chen et al. (2016) that it converges to a KKT point of (3.3). Since slater's condition is satisfied, we can conclude that it converges to a global optimum. See, e.g., Page 244 of Chapter 5 of Boyd and Vandenberghe (2004). □

Proof of Theorem 2. Theorem 4.4 of Beck and Teboulle (2009) shows the convergence analysis of FISTA with both constant and backtracking step sizes. In addition, they provide a detailed convergence rate of the algorithm. Here we adopt the FISTA with backtracking step sizes. □

S1.2 Estimation of $\boldsymbol{A}_\star$ (Cont')

S1.3 General n_1, n_2 Cases of the Theorems

For any threshold level $\beta > 0$, we adopt the two quantities L_{α_0} and γ_{α_0} defined respectively:

$$L_{\alpha_0} = \sup_{|m| \leq \alpha_0} \frac{|f'(m)|}{f(m)\{1 - f(m)\}} \quad \text{and} \quad \gamma_{\alpha_0} = \sup_{|m| \leq \alpha_0} \frac{f(m)\{1 - f(m)\}}{\{f'(m)\}^2}. \tag{S1.3}$$

As discussed in Davenport et al. (2014) L_{α_0} and γ_{α_0} control the "steepness" and "flatness" of link function f respectively. Under the two natural link function f, we have $\gamma_{\alpha_0} = e^{-\alpha_0}(1 + e^{\alpha_0})^2 \simeq e^{\alpha_0}$, $L_{\alpha_0} = 1$ and $\gamma_{\alpha_0} \leq \pi e^{\alpha_0^2/2}$, $L_{\alpha_0} \leq$

$8(\alpha_0+1)$ respectively.

Theorem S1. *Assume that Conditions C1-C2 hold, and* $(\mu_\star, \boldsymbol{Z}_\star) \in \widetilde{\mathcal{C}}_{n_1,n_2}(\alpha_1,\alpha_2)$. *Let* L_{α_0} *be the quantities as in* (S1.3). *Consider* $\widehat{\boldsymbol{M}} = \widehat{\mu}\boldsymbol{J} + \widehat{\boldsymbol{Z}}$ *where* $(\widehat{\mu}, \widehat{\boldsymbol{Z}})$ *is the solution to* (5.1). *There exist some positive constants* C_1, C_2, *such that for* $C_{L_{\alpha_0},n_1,n_2} = (16e+1)L_{\alpha_0}(\alpha_1+\alpha_2 r_{\boldsymbol{Z}_\star}^{1/2})(n_1n_2)^{1/2}(n_1\vee n_2)^{1/2}$, *we have with probability at least* $1-4C_1/(n_1+n_2)$,

$$(\mu_\star-\widehat{\mu})^2 \le C_2\left(\alpha_1^2 \wedge \Gamma_{n_1,n_2}\right), \quad \frac{1}{n_1n_2}\left\|\widehat{\boldsymbol{Z}}-\boldsymbol{Z}_\star\right\|_F^2 \le C_2\left(\alpha_2^2 \wedge \Gamma_{n_1,n_2}\right)$$
$$\text{and} \quad \frac{1}{n_1n_2}\left\|\widehat{\boldsymbol{M}}-\boldsymbol{M}_\star\right\|_F^2 \le C_2\left(\alpha_0^2 \wedge \Gamma_{n_1,n_2}\right). \tag{S1.4}$$

where $\Gamma_{n_1,n_2} = \frac{\gamma_{\alpha_0}C_{L_{\alpha_0},n_1,n_2}}{n_1n_2}$.

Theorem S2. *Assume that Conditions C1-C2 hold, and* $(\mu_\star, \boldsymbol{Z}_\star) \in \widetilde{\mathcal{C}}_{n_1,n_2}(\alpha_1,\alpha_2)$. *Let* L_{α_0} *and* γ_β *be the quantities as in* (S1.3). *Consider* $\widehat{\boldsymbol{M}}_\beta = \widehat{\mu}\boldsymbol{J} + \widehat{\boldsymbol{Z}}_\beta$ *where* $\widehat{\boldsymbol{Z}}_\beta$ *is the solution to* (5.3) *and* $\beta \ge 0$, *there exist some positive constants* C_1, C_2 *and* C_3, *such that for* $C_{L_{\alpha_0},n_1,n_2} = (16e+1)L_{\alpha_0}(\alpha_1+\alpha_2 r_{\boldsymbol{Z}_\star}^{1/2})(n_1n_2)^{1/2}(n_1\vee n_2)^{1/2}$ *and* $C_{L_{\alpha_1+\beta},n_1,n_2} = (16e+1)L_{\alpha_1+\beta}(\alpha_1+\beta r_{\mathcal{T}_\beta(\boldsymbol{Z}_\star)}^{1/2})(n_1n_2)^{1/2}(n_1\vee n_2)^{1/2}$, *we have with probability at least* $1-4C_1/(n_1+n_2)$,

$$d^2\left(\widehat{\boldsymbol{M}}_\beta, \boldsymbol{M}_\star\right) \le C_2\left(\alpha_1^2 \wedge \Gamma_{n_1,n_2}\right) + C_3\Lambda_{n_1,n_2} + \frac{2(\alpha_2-\beta)_+^2 N_\beta}{n_1n_2}, \tag{S1.5}$$

$$d^2\left(\widehat{\boldsymbol{\Theta}}_\beta^\dagger, \boldsymbol{\Theta}_\star^\dagger\right) \le \frac{C_2}{h_{\alpha_1,\beta}^2}\left(\alpha_1^2 \wedge \Gamma_{n_1,n_2}\right) + \frac{C_3\Lambda_{n_1,n_2}}{h_{\alpha_1,\beta}^2} + \frac{8N_\beta}{n_1n_2\theta_L^2}, \tag{S1.6}$$

where

$$\Lambda_{n_1,n_2} = \min\left[\beta^2, \widetilde{\Gamma}_{n_1,n_2} + \frac{\gamma_{\beta+\alpha_1} L_{\alpha_1+\beta}\beta\left\{8N_\beta + L_{\alpha_1+\beta}\left(n_1n_2 - N_\beta\right)|\mu_\star - \widehat{\mu}|\right\}}{n_1n_2}\right],$$

$$\textit{and} \quad \widetilde{\Gamma}_{n_1,n_2} = \frac{\gamma_{\beta+\alpha_1} C_{L_{\alpha_1+\beta},n_1,n_2}}{n_1n_2}.$$

Define the parameters $k_{\alpha_1,\alpha_2,\beta,n_1,n_2}$ and $k'_{\alpha_1,\alpha_2,n_1,n_2}$ such that

$$k_{\alpha_1,\alpha_2,\beta,n_1,n_2} = \min\Big[\beta^2, \gamma_{\beta+\alpha_1}\Big[L_{\alpha_1+\beta}\left(\alpha_1 + \beta r^{1/2}_{\mathcal{T}_\beta(\boldsymbol{Z}_\star)}\right)(n_1n_2)^{1/2}(n_1 \vee n_2)^{1/2}$$
$$+L_{\alpha_1+\beta}\beta\left\{8N_\beta + L_{\alpha_1+\beta}\left(n_1n_2 - N_\beta\right)k'^{1/2}_{\alpha_1,\alpha_2,n_1,n_2}\right\}\Big](n_1n_2)^{-1}\Big],$$

$$k'_{\alpha_1,\alpha_2,n_1,n_2} = \min\left\{\alpha_1^2, \frac{\left(\alpha_1 + \alpha_2 r^{1/2}_{\boldsymbol{Z}_\star}\right)\gamma_{\alpha_0}L_{\alpha_0}\left(n_1 \vee n_2\right)^{1/2}}{(n_1n_2)^{1/2}}\right\}.$$

In additional to $h_{(1),\beta}$, we need two more $h_{(2),\beta} = \max(\theta^{-1}_{\star,ij,\beta}\theta^{1/2}_{\star,ij})$, and $h_{(3),\beta} = \max(\theta^{-1}_{\star,ij,\beta}\theta_{\star,ij})$ to complete the general form of Theorem (4) in the following. Also let

$$\tilde{\Delta} = \max\left[\frac{(c_\sigma \vee a)\, e^{-\mu_\star/2+\alpha_2-\beta+|\alpha_2/2-\beta|}\left\{(n_1 \vee n_2)\log\left(n_1+n_2\right)\right\}^{1/2}}{n_1n_2},\right.$$
$$\left.\frac{\eta e^{\mu_\star/2+\alpha_1+|\alpha_2/2-\beta|}k^{1/2}_{\alpha_1,\alpha_2,\beta,n_1,n_2}\log^{3/2}n}{h_{\alpha_1,\beta}n}\right]. \tag{S1.7}$$

Theorem S3. *Assume Conditions C1-C4 hold and logit inverse link function f. There exist some positive constants C_4, C_5, C_6 and C_7 such that for $\tau \geq C_4\tilde{\Delta}$, we have with probability at least $1 - 3/(n_1+n_2)$,*

$$d^2\left(\widehat{\boldsymbol{A}}_\beta, \boldsymbol{A}_\star\right) \leq \max\left\{C_6n_1n_2h^2_{(1),\beta}r_{\boldsymbol{A}_\star}\tau^2 + \frac{C_7h^2_{(1),\beta}h^2_{(2),\beta}r_{\boldsymbol{A}_\star}\log\left(n_1+n_2\right)}{(n_1 \wedge n_2)},\right.$$
$$\left.\frac{C_5h_{(1),\beta}h_{(3),\beta}\log^{1/2}(n_1+n_2)}{\sqrt{n_1n_2}}\right\}. \tag{S1.8}$$

S1.4 Proof of Theorems in Section S1.3

First, we review some basic facts about matrices which will be useful in the following development. For any $\boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{n_1 \times n_2}$, we have

- Trace Duality Property:

$$|\mathrm{tr}\,(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{C})| \leq \|\boldsymbol{C}\|\,\|\boldsymbol{B}\|_{*}\,. \tag{S1.9}$$

- Norm Inequalities:

$$\|\boldsymbol{B}\|_F \leq \|\boldsymbol{B}\|_{*} \leq r_{\boldsymbol{B}}^{1/2}\,\|\boldsymbol{B}\|_F \quad \text{and} \quad \|\boldsymbol{B}\| \leq \|\boldsymbol{B}\|_F \leq r_{\boldsymbol{B}}^{1/2}\,\|\boldsymbol{B}\|\,, \tag{S1.10}$$

  where $r_{\boldsymbol{B}}$ is the rank of matrix $\boldsymbol{B}$.

In this section, we provide the proofs of general n_1, n_2 and f cases of main theorems presented in Section S1.3. Their corresponding $n_1 = n_2 = n$ and the choice of inverse link function f to be logit cases can be directly derived from the general cases.

To prove Theorem S1 and Theorem S2, in additional to the Hellinger distance, we also adopt the Kullback-Leibler (KL) distance. For any $\boldsymbol{S}, \boldsymbol{T} \in [0,1]^{n_1 \times n_2}$, define $D(\boldsymbol{S}\|\boldsymbol{T})$ as $(n_1 n_2)^{-1}\sum_{i,j=1}^{n_1,n_2} D(s_{ij}\|t_{ij})$ where $D(s\|t) = s\log(s/t) + (1-s)\log((1-s)/(1-t))$ for $s,t \in [0,1]$ is the KL distance between two Bernoulli distributions. The KL distance is bounded below by the Hellinger distance, i.e, $d_H^2(s,t) \leq D(s\|t)$. Given any matrix $\boldsymbol{M} \in \mathbb{R}^{n_1 \times n_2}$, we work on the

function

$$\bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M}) = \ell_{\boldsymbol{W}}(\boldsymbol{M}) - \ell_{\boldsymbol{W}}(\boldsymbol{M}_\star),$$

rather than on $\ell_{\boldsymbol{W}}$ itself. Now we present several lemmas and proofs.

Lemma S1. *Under Condition C2, assume that matrix $\boldsymbol{M} = \mu\boldsymbol{J} + \boldsymbol{Z}$ such that $(\mu, \boldsymbol{Z}) \in \widetilde{\mathcal{C}}_{n_1,n_2}(\alpha_1, \alpha_2)$ as defined in* (5.1)*. Then, for any $s > 0$, we have*

$$\begin{aligned}
&\Pr\left[\sup_{(\mu,\boldsymbol{Z})\in\widetilde{\mathcal{C}}_{n_1,n_2}(\alpha_1,\alpha_2)} \left|\bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M}) - E_{\boldsymbol{W}}\left\{\bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M})\right\}\right| \geq C_0 s(\alpha_1 + \alpha_2 r_{\boldsymbol{Z}_\star}^{1/2})(n_1 n_2)^{1/2}\right] \\
\leq & C_1\left\{\frac{8L_{\alpha_0}(n_1 \vee n_2)^{1/2}}{s}\right\}^{\log(n_1+n_2)}, \qquad \text{(S1.11)}
\end{aligned}$$

where C_0 and C_1 are absolute constants and, the probability and the expectation are both taken over $\boldsymbol{W}$.

Proof of Lemma S1. Noting that for any $h > 0$, by Markov's inequality, we have that

$$\begin{aligned}
&\Pr\left[\sup_{(\mu,\boldsymbol{Z})\in\widetilde{\mathcal{C}}_{n_1,n_2}(\alpha_1,\alpha_2)} \left|\bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M}) - E_{\boldsymbol{W}}\left\{\bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M})\right\}\right| \geq C_0 s(\alpha_1 + \alpha_2 r_{\boldsymbol{Z}_\star}^{1/2})(n_1 n_2)^{1/2}\right] \\
= & \Pr\left[\sup_{(\mu,\boldsymbol{Z})\in\widetilde{\mathcal{C}}_{n_1,n_2}(\alpha_1,\alpha_2)} \left|\bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M}) - E_{\boldsymbol{W}}\left\{\bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M})\right\}\right|^h \geq \left(C_0 s(\alpha_1 + \alpha_2 r_{\boldsymbol{Z}_\star}^{1/2})(n_1 n_2)^{1/2}\right)^h\right] \\
\leq & \frac{E_{\boldsymbol{W}}\left[\sup_{(\mu,\boldsymbol{Z})\in\widetilde{\mathcal{C}}_{n_1,n_2}(\alpha_1,\alpha_2)} \left|\bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M}) - E_{\boldsymbol{W}}\left\{\bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M})\right\}\right|^h\right]}{\left(C_0 s(\alpha_1 + \alpha_2 r_{\boldsymbol{Z}_\star}^{1/2})(n_1 n_2)^{1/2}\right)^h}. \qquad \text{(S1.12)}
\end{aligned}$$

The bound in (S1.11) will follow by combining (S1.12) with an upper bound on $E_{\boldsymbol{W}}[\sup_{(\mu,\boldsymbol{Z})\in\widetilde{\mathcal{C}}_{n_1,n_2}(\alpha_1,\alpha_2)} |\bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M}) - E_{\boldsymbol{W}}\{\bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M})\}|^h]$ and setting $h = \log(n_1+n_2)$.

Note that we can write the $\bar{\ell}_{\boldsymbol{W}}$ as

$$\bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M}) = \sum_{i,j=1}^{n_1,n_2} \left[\mathbb{I}_{[w_{ij}=1]} \log \left\{ \frac{f(m_{\star,ij} + (m_{ij} - m_{\star,ij}))}{f(m_{\star,ij})} \right\} + \mathbb{I}_{[w_{ij}=0]} \log \left\{ \frac{1 - f(m_{\star,ij} + (m_{ij} - m_{\star,ij}))}{1 - f(m_{\star,ij})} \right\} \right].$$

By a symmetrization argument (Lemma 6.3 in Ledoux and Talagrand (2013)),

$$E_{\boldsymbol{W}} \left[\sup_{(\mu,\boldsymbol{Z}) \in \widetilde{\mathcal{C}}_{n_1,n_2}(\alpha_1,\alpha_2)} \left| \bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M}) - E_{\boldsymbol{W}} \left\{ \bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M}) \right\} \right|^h \right] \leq 2^h E_{\boldsymbol{W},\boldsymbol{\Xi}} \left[\sup_{(\mu,\boldsymbol{Z}) \in \widetilde{\mathcal{C}}_{n_1,n_2}(\alpha_1,\alpha_2)} \left| \sum_{i,j=1}^{n_1,n_2} \xi_{ij} \left[\mathbb{I}_{[w_{ij}=1]} \log \left\{ \frac{f(m_{\star,ij} + (m_{ij} - m_{\star,ij}))}{f(m_{\star,ij})} \right\} + \mathbb{I}_{[w_{ij}=0]} \log \left\{ \frac{1 - f(m_{\star,ij} + (m_{ij} - m_{\star,ij}))}{1 - f(m_{\star,ij})} \right\} \right] \right|^h \right], \tag{S1.13}$$

where ξ_{ij} are i.i.d. Rademacher random variables and the expectation in the upper bound is taken with respect to $\boldsymbol{W}$ as well as ξ_{ij}.

For any $|m_{\star,ij} + x| \leq \alpha_0$, define $\phi_{1,ij}(x) = L_{\alpha_0}^{-1} \log(f(m_{\star,ij} + x)/f(m_{\star,ij}))$ and $\phi_{2,ij}(x) = L_{\alpha_0}^{-1} \log\{[1 - f(m_{\star,ij} - x)]/[1 - f(m_{\star,ij})]\}$. By the definition of L_{α_0}, it is not hard to show that $|\phi_{1,ij}(x_1) - \phi_{1,ij}(x_2)| \leq |x_1 - x_2|$, $\phi_{1,ij}(0) = 0$ and $|\phi_{2,ij}(x_1) - \phi_{2,ij}(x_2)| \leq |x_1 - x_2|$, $\phi_2(0) = 0$. For $\boldsymbol{M} - \boldsymbol{M}_\star$ satisfies $|m_{ij} - m_{\star,ij}| \leq 2\alpha_0$, we can apply a contraction principle by Theorem 4.12 in Ledoux and Talagrand (2013). Thus, up to a factor of 2, the right hand side of (S1.13) can only decrease when $\phi_{1,ij}(m_{ij} - m_{\star,ij})$ and $\phi_{2,ij}(m_{\star,ij} - m_{ij})$ are

replaced by $m_{ij} - m_{\star,ij}$ and $m_{\star,ij} - m_{ij}$ respectively. We obtain

$$
\begin{aligned}
& E_{\boldsymbol{W}}\left[\sup_{(\mu,\boldsymbol{Z})\in\widetilde{\mathcal{C}}_{n_1,n_2}(\alpha_1,\alpha_2)}\left|\bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M}) - E_{\boldsymbol{W}}\left\{\bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M})\right\}\right|^h\right] \\
\leq & 2^h\left(2L_{\alpha_0}\right)^h E_{\boldsymbol{W},\boldsymbol{\Xi}}\left[\sup_{(\mu,\boldsymbol{Z})\in\widetilde{\mathcal{C}}_{n_1,n_2}(\alpha_1,\alpha_2)}\left|\sum_{i,j=1}^{n_1,n_2}\xi_{ij}\left\{\mathbb{I}_{\left[w_{ij}=1\right]}\left(m_{ij}-m_{\star,ij}\right)+\mathbb{I}_{\left[w_{ij}=0\right]}\left(m_{\star,ij}-m_{ij}\right)\right\}\right|^h\right] \\
= & \left(4L_{\alpha_0}\right)^h E_{\boldsymbol{W},\boldsymbol{\Xi}}\left[\sup_{(\mu,\boldsymbol{Z})\in\widetilde{\mathcal{C}}_{n_1,n_2}(\alpha_1,\alpha_2)}\left|\langle\boldsymbol{\Xi}\circ\left(2\boldsymbol{W}-\boldsymbol{J}\right),\boldsymbol{M}-\boldsymbol{M}_\star\rangle\right|^h\right], \qquad \text{(S1.14)}
\end{aligned}
$$

where $\boldsymbol{\Xi}$ denotes the matrix with entries ξ_{ij}. Using the facts that the distribution of $\boldsymbol{\Xi}\circ\left(2\boldsymbol{W}-\boldsymbol{J}\right)$ is the same as the distribution of $\boldsymbol{\Xi}$ and trace duality inequality given in (S1.9), we have that

$$
\begin{aligned}
& E_{\boldsymbol{W},\boldsymbol{\Xi}}\left[\sup_{(\mu,\boldsymbol{Z})\in\widetilde{\mathcal{C}}_{n_1,n_2}(\alpha_1,\alpha_2)}\left|\langle\boldsymbol{\Xi}\circ\left(2\boldsymbol{W}-\boldsymbol{J}\right),\boldsymbol{M}-\boldsymbol{M}_\star\rangle\right|^h\right] \\
= & E_{\boldsymbol{\Xi}}\left[\sup_{(\mu,\boldsymbol{Z})\in\widetilde{\mathcal{C}}_{n_1,n_2}(\alpha_1,\alpha_2)}\left|\langle\boldsymbol{\Xi},\boldsymbol{M}-\boldsymbol{M}_\star\rangle\right|^h\right] \\
\leq & E_{\boldsymbol{\Xi}}\left[\sup_{(\mu,\boldsymbol{Z})\in\widetilde{\mathcal{C}}_{n_1,n_2}(\alpha_1,\alpha_2)}\left\|\boldsymbol{\Xi}\right\|^h\left\|\boldsymbol{M}-\boldsymbol{M}_\star\right\|_*^h\right] \\
= & \left(2(\alpha_1+\alpha_2 r_{\boldsymbol{Z}_\star}^{1/2})(n_1n_2)^{1/2}\right)^h E_{\boldsymbol{\Xi}}\left\|\boldsymbol{\Xi}\right\|^h, \qquad \text{(S1.15)}
\end{aligned}
$$

To bound $E_{\boldsymbol{\Xi}}\|\boldsymbol{\Xi}\|^h$, observe that $\boldsymbol{\Xi}$ is a matrix with i.i.d. zero mean entries and thus by Theorem 1.1 of Seginer (2000),

$$
E_{\boldsymbol{\Xi}}\left\|\boldsymbol{\Xi}\right\|^h \leq C_1\left(n_1^{\frac{h}{2}}+n_2^{\frac{h}{2}}\right),
$$

for a positive constant C_1. This in turn implies that

$$
\left(E_{\boldsymbol{\Xi}}\left\|\boldsymbol{\Xi}\right\|^h\right)^{\frac{1}{h}} \leq C_1^{\frac{1}{h}}\left(n_1^{\frac{1}{2}}+n_2^{\frac{1}{2}}\right) \leq 2C_1^{\frac{1}{h}}\left(n_1\vee n_2\right)^{1/2}. \qquad \text{(S1.16)}
$$

Combining (S1.16) with (S1.14) and (S1.15), we obtain

$$\left\{E_{\boldsymbol{W}}\left|\bar{\ell}_{\boldsymbol{W}}\left(\boldsymbol{M}\right)-E\bar{\ell}_{\boldsymbol{W}}\left(\boldsymbol{M}\right)\right|^{h}\right\}^{\frac{1}{h}}\leq 16C_{1}^{\frac{1}{h}}L_{\alpha_0}(\alpha_1+\alpha_2 r_{\boldsymbol{Z}_\star}^{1/2})(n_1n_2)^{1/2}\left(n_1\vee n_2\right)^{1/2}.$$

Plugging this into (S1.12) and take $h=\log(n_1+n_2)$, the probability in (S1.12) is upper bounded by

$$C_1\left\{\frac{8L_{\alpha_0}\left(n_1\vee n_2\right)^{1/2}}{s}\right\}^{\log(n_1+n_2)},$$

which establishes the lemma. □

Lemma S1 presents the general version with complete observations of indicators $\boldsymbol{W}$. A comparable result is Lemma 1 of Davenport et al. (2014) which provide the incomplete indicators version. The following remark holds directly by setting $s=C_0L_{\alpha_0}(n_1\vee n_2)^{1/2}$ in our Lemma S1.

Remark 1. Under Condition C2, assume that matrix $\boldsymbol{M}=\mu\boldsymbol{J}+\boldsymbol{Z}$ such that $(\mu,\boldsymbol{Z})\in\widetilde{\mathcal{C}}_{n_1,n_2}(\alpha_1,\alpha_2)$ as defined in (5.1). Take $s=C_{L_{\alpha_0},n_1,n_2}=C_0L_{\alpha_0}(\alpha_1+\alpha_2 r_{\boldsymbol{Z}_\star}^{1/2})(n_1n_2)^{1/2}(n_1\vee n_2)^{1/2}$ and provided $C_0>16e$, we can simplify (S1.11) to be

$$\Pr\left(\left|\bar{\ell}_{\boldsymbol{W}}\left(\boldsymbol{M}\right)-E_{\boldsymbol{W}}\left\{\bar{\ell}_{\boldsymbol{W}}\left(\boldsymbol{M}\right)\right\}\right|\geq C_{L_{\alpha_0},n_1,n_2}\right)\leq\frac{C_1}{n_1+n_2}.$$

Next we will prove Theorem S1.

Proof of Theorem S1. Due to the fact that $|\mu_\star|\leq\alpha_1$ and $\|\boldsymbol{Z}_\star\|_\infty\leq\alpha_2$, we can easily have $\|\boldsymbol{M}_\star\|_\infty=\|\mu_\star\boldsymbol{J}+\boldsymbol{Z}_\star\|_\infty\leq\alpha_0$. Similarly, due to the feasible set

$\mathcal{C}_{n_1,n_2}(\alpha_1,\alpha_2)$, we have $|\widehat{\mu}| \leq \alpha_1$ and $\|\widehat{\boldsymbol{Z}}\|_\infty \leq \alpha_2$. These implies the bounds only related to α_1 and α_2 of all the three terms in Theorem S2.

Next we focus on the bounds related to $r_{\boldsymbol{Z}_\star}$. According to the definition of $\bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M})$ we have

$$E_{\boldsymbol{W}}\left[\bar{\ell}_{\boldsymbol{W}}\left(\widehat{\boldsymbol{M}}\right)\right] = E_{\boldsymbol{W}}\left[\ell_{\boldsymbol{W}}\left(\widehat{\boldsymbol{M}}\right) - \ell_{\boldsymbol{W}}\left(\boldsymbol{M}_\star\right)\right]$$
$$= \sum_{i,j=1}^{n_1,n_2}\left[f\left(m_{\star,ij}\right)\log\left\{\frac{f\left(\widehat{\mu}+\widehat{z}_{ij}\right)}{f\left(m_{\star,ij}\right)}\right\} + \left\{1-f\left(m_{\star,ij}\right)\right\}\log\left\{\frac{1-f\left(\widehat{\mu}+\widehat{z}_{ij}\right)}{1-f\left(m_{\star,ij}\right)}\right\}\right],$$

where the expectation is taken over the indicators $\boldsymbol{W}$. This term equals to $-n_1n_2D(\mathcal{F}(\boldsymbol{M}_\star)\|\mathcal{F}(\widehat{\boldsymbol{M}}))$. Then for $\widehat{\boldsymbol{M}} = \widehat{\mu}\boldsymbol{J} + \widehat{\boldsymbol{Z}}$, we have

$$\bar{\ell}_{\boldsymbol{W}}\left(\widehat{\boldsymbol{M}}\right) \leq -n_1n_2D\left\{\mathcal{F}\left(\boldsymbol{M}_\star\right)\|\mathcal{F}\left(\widehat{\boldsymbol{M}}\right)\right\} + \left|\bar{\ell}_{\boldsymbol{W}}\left(\widehat{\boldsymbol{M}}\right) - E_{\boldsymbol{W}}\left\{\bar{\ell}_{\boldsymbol{W}}\left(\widehat{\boldsymbol{M}}\right)\right\}\right|.$$

Moreover, from the definition of $(\widehat{\mu}, \widehat{\boldsymbol{Z}})$ given in (5.1), we also have that $\bar{\ell}_{\boldsymbol{W}}(\widehat{\boldsymbol{M}}) = \ell_{\boldsymbol{W}}(\widehat{\boldsymbol{M}}) - \ell_{\boldsymbol{W}}(\boldsymbol{M}_\star)$ and $\ell_{\boldsymbol{W}}(\widehat{\boldsymbol{M}}) = \ell_{\boldsymbol{W}}(\widehat{\mu}\boldsymbol{J} + \widehat{\boldsymbol{Z}}) \geq \ell_{\boldsymbol{W}}(\boldsymbol{M}_\star)$. Thus

$$0 \leq -n_1n_2D\left\{\mathcal{F}\left(\boldsymbol{M}_\star\right)\|\mathcal{F}\left(\widehat{\boldsymbol{M}}\right)\right\} + \left|\bar{\ell}_{\boldsymbol{W}}\left(\widehat{\boldsymbol{M}}\right) - E_{\boldsymbol{W}}\left\{\bar{\ell}_{\boldsymbol{W}}\left(\widehat{\boldsymbol{M}}\right)\right\}\right|.$$

Applying Remark 1 for matrix $\boldsymbol{M} = \widehat{\boldsymbol{M}} = \widehat{\mu}\boldsymbol{J} + \widehat{\boldsymbol{Z}}$, for any $C_0 > 16e$, we obtain that with probability at least $1 - C_1/(n_1+n_2)$,

$$0 \leq -n_1n_2D\left\{\mathcal{F}\left(\boldsymbol{M}_\star\right)\|\mathcal{F}\left(\widehat{\boldsymbol{M}}\right)\right\} + C_{L_{\alpha_0},n_1,n_2}.$$

Furthermore, we obtain that with probability at least $1 - C_1/(n_1+n_2)$,

$$d_H^2\left\{\mathcal{F}\left(\boldsymbol{M}_\star\right), \mathcal{F}\left(\widehat{\boldsymbol{M}}\right)\right\} \leq D\left\{\mathcal{F}\left(\boldsymbol{M}_\star\right)\|\mathcal{F}\left(\widehat{\boldsymbol{M}}\right)\right\} \leq \frac{C_{L_{\alpha_0},n_1,n_2}}{n_1n_2}.$$

By Lemma 2 of Davenport et al. (2014), we have

$$\frac{1}{n_1 n_2}\left\|\widehat{\boldsymbol{M}}-\boldsymbol{M}_\star\right\|_F^2 \leq \gamma_{\alpha_0} d_H^2\left\{\mathcal{F}\left(\widehat{\boldsymbol{M}}\right), \mathcal{F}\left(\boldsymbol{M}_\star\right)\right\} \leq \frac{\gamma_{\alpha_0} C_{L_{\alpha_0},n_1,n_2}}{n_1 n_2}.$$

Due to the decomposition that $\|\widehat{\boldsymbol{M}}-\boldsymbol{M}_\star\|_F^2 = n_1 n_2(\widehat{\mu}-\mu_\star)^2 + \|\widehat{\boldsymbol{Z}}-\boldsymbol{Z}_\star\|_F^2$, it implies the bounds related to $r_{\boldsymbol{Z}_\star}$ that

$$|\mu_\star-\widehat{\mu}| \leq \frac{\gamma_{\alpha_0}^{1/2} C_{L_{\alpha_0},n_1,n_2}^{1/2}}{(n_1 n_2)^{1/2}}, \quad \frac{1}{n_1 n_2}\left\|\widehat{\boldsymbol{Z}}-\boldsymbol{Z}_\star\right\|_F^2 \leq \frac{\gamma_{\alpha_0} C_{L_{\alpha_0},n_1,n_2}}{n_1 n_2}, \quad \text{and}$$
$$\frac{1}{n_1 n_2}\left\|\widehat{\boldsymbol{M}}-\boldsymbol{M}_\star\right\|_F^2 \leq \frac{\gamma_{\alpha_0} C_{L_{\alpha_0},n_1,n_2}}{n_1 n_2}.$$

□

To prove Theorem S2, we measure the similarity of estimated distribution $\widehat{\boldsymbol{\Theta}}_\beta = \mathcal{F}(\widehat{\boldsymbol{M}}_\beta)$ and the constrained distribution $\widehat{\boldsymbol{\Theta}}_{\star,\beta} = \mathcal{F}\{\widehat{\mu}\boldsymbol{J}+\mathcal{T}_\beta(\boldsymbol{Z}_\star)\}$ at first in next Theorem. Similar as the definition of $\bar{\ell}_{\boldsymbol{W}}(\boldsymbol{M})$, define

$$\widetilde{\ell}_{\boldsymbol{W}}(\boldsymbol{M}) = \ell_{\boldsymbol{W}}(\boldsymbol{M}) - \ell_{\boldsymbol{W}}\left\{\widehat{\mu}\boldsymbol{J}+\mathcal{T}_\beta(\boldsymbol{Z}_\star)\right\},$$

for any matrix $\boldsymbol{M} \in \mathbb{R}^{n_1\times n_2}$.

Remark 2. Under Condition C2, let $\mathcal{G}_{n_1,n_2}(\alpha_1,\beta) \subset \mathbb{R}\times\mathbb{R}^{n_1\times n_2}$ be

$$\mathcal{G}_{n_1,n_2}(\alpha_1,\beta) = \left\{(\mu,\boldsymbol{Z}) \in \mathbb{R}\times\mathbb{R}^{n_1\times n_2} : |\mu| \leq \alpha_1, \|\boldsymbol{Z}\|_\infty \leq \beta, \|\boldsymbol{Z}\|_* \leq \beta\sqrt{r_{\boldsymbol{Z}_\star} n_1 n_2}\right\}.$$

Assume that matrix $\boldsymbol{M} = \mu\boldsymbol{J}+\boldsymbol{Z}$ such that $(\mu,\boldsymbol{Z}) \in \mathcal{G}_{n_1,n_2}(\alpha_1,\beta)$ as defined in (5.3). Take $s = C_{L_{\alpha_1+\beta},n_1,n_2} = C_0 L_{\alpha_1+\beta}(\alpha_1+\beta r_{\mathcal{T}_\beta(\boldsymbol{Z}_\star)}^{1/2})(n_1 n_2)^{1/2}(n_1\vee n_2)^{1/2}$ and provided $C_0 > 16e$, we can simplify (S1.11) to be

$$\Pr\left(\left|\widetilde{\ell}_{\boldsymbol{W}}(\boldsymbol{M}) - E_{\boldsymbol{W}}\left\{\widetilde{\ell}_{\boldsymbol{W}}(\boldsymbol{M})\right\}\right| \geq C_{L_{\alpha_1+\beta},n_1,n_2}\right) \leq \frac{C_1}{n_1+n_2}.$$

Theorem S4. *Assume that Conditions C1-C3 hold, let L_{α_0} be the quantities as in* (S1.3)*, $\widehat{\mu}$ be the estimator defined in* (5.1)*. Consider $\widehat{\boldsymbol{M}}_\beta = \widehat{\mu}\boldsymbol{J} + \widehat{\boldsymbol{Z}}_\beta$ where $\widehat{\boldsymbol{Z}}_\beta$ is the solution to* (5.3) *and $\beta > 0$, for any $C_0 > 16e$, $C_{L_{\alpha_1+\beta},n_1,n_2} = C_0 L_{\alpha_1+\beta}(\alpha_1+\beta r^{1/2}_{\mathcal{T}_\beta(\boldsymbol{Z}_\star)})(n_1n_2)^{1/2}(n_1\vee n_2)^{1/2}$, there exist some positive constants C_1 and C_3' such that with probability at least $1-4C_1/(n_1+n_2)$, we have*

$$d^2\left\{\widehat{\boldsymbol{Z}}_\beta, \mathcal{T}_\beta\left(\boldsymbol{Z}_\star\right)\right\} \leq C_3' \min\left[\beta^2, \frac{\gamma_{\beta+\alpha_1}\left[C_{L_{\alpha_1+\beta},n_1,n_2} + L_{\alpha_1+\beta}\beta\left\{8N_\beta + L_{\alpha_1+\beta}\left(n_1n_2 - N_\beta\right)|\mu_\star - \widehat{\mu}|\right\}\right]}{n_1n_2}\right].$$

Proof of Theorem S4. For the proof of Theorem S4, according to the definition of $\widetilde{\ell}_{\boldsymbol{W}}(\boldsymbol{M})$ we have

$$\begin{aligned}
&E_{\boldsymbol{W}}\left\{\widetilde{\ell}_{\boldsymbol{W}}\left(\widehat{\boldsymbol{M}}_\beta\right)\right\} = E_{\boldsymbol{W}}\left[\ell_{\boldsymbol{W}}\left(\widehat{\boldsymbol{M}}_\beta\right) - \ell_{\boldsymbol{W}}\left\{\widehat{\mu}\boldsymbol{J} + \mathcal{T}_\beta\left(\boldsymbol{Z}_\star\right)\right\}\right] \\
&= \sum_{i,j=1}^{n_1,n_2}\left[f\left(m_{\star,ij}\right)\log\left\{\frac{f\left(\widehat{\mu}+\widehat{z}_{ij,\beta}\right)}{f\left(\widehat{\mu}+T_\beta\left(z_{\star,ij}\right)\right)}\right\} + \left\{1-f\left(m_{\star,ij}\right)\right\}\log\left\{\frac{1-f\left(\widehat{\mu}+\widehat{z}_{ij,\beta}\right)}{1-f\left(\widehat{\mu}+T_\beta\left(z_{\star,ij}\right)\right)}\right\}\right] \\
&= \sum_{i,j=1}^{n_1,n_2}\left[f\left\{\widehat{\mu}+T_\beta\left(z_{\star,ij}\right)\right\}\log\left\{\frac{f\left(\widehat{\mu}+\widehat{z}_{ij,\beta}\right)}{f\left(\widehat{\mu}+T_\beta\left(z_{\star,ij}\right)\right)}\right\} + \left\{1-f\left(\widehat{\mu}+T_\beta\left(z_{\star,ij}\right)\right)\right\}\log\left\{\frac{1-f\left(\widehat{\mu}+\widehat{z}_{ij,\beta}\right)}{1-f\left(\widehat{\mu}+T_\beta\left(z_{\star,ij}\right)\right)}\right\}\right] \\
&+ \sum_{z_{\star,ij}>\beta}\left[\left\{f\left(m_{\star,ij}\right)-f(\widehat{\mu}+\beta)\right\}\log\left\{\frac{f\left(\widehat{\mu}+\widehat{z}_{ij,\beta}\right)}{f\left(\widehat{\mu}+\beta\right)}\right\} + \left\{f\left(\widehat{\mu}+\beta\right)-f\left(m_{\star,ij}\right)\right\}\log\left\{\frac{1-f\left(\widehat{\mu}+\widehat{z}_{ij,\beta}\right)}{1-f\left(\widehat{\mu}+\beta\right)}\right\}\right] \\
&+ \sum_{z_{\star,ij}<-\beta}\left[\left(f\left(m_{\star,ij}\right)-f(\widehat{\mu}-\beta)\right)\log\left(\frac{f\left(\widehat{\mu}+\widehat{z}_{ij,\beta}\right)}{f\left(\widehat{\mu}-\beta\right)}\right) + \left(f\left(\widehat{\mu}-\beta\right)-f\left(m_{\star,ij}\right)\right)\log\left(\frac{1-f\left(\widehat{\mu}+\widehat{z}_{ij,\beta}\right)}{1-f\left(\widehat{\mu}-\beta\right)}\right)\right] \\
&+ \sum_{-\beta<z_{\star,ij}<\beta}\left[\left\{f\left(m_{\star,ij}\right)-f(\widehat{\mu}+z_{\star,ij})\right\}\log\left\{\frac{f\left(\widehat{\mu}+\widehat{z}_{ij,\beta}\right)}{f\left(\widehat{\mu}+z_{\star,ij}\right)}\right\}\right. \\
&\left.+\left\{f\left(\widehat{\mu}+z_{\star,ij}\right)-f\left(m_{\star,ij}\right)\right\}\log\left\{\frac{1-f\left(\widehat{\mu}+\widehat{z}_{ij,\beta}\right)}{1-f\left(\widehat{\mu}+z_{\star,ij}\right)}\right\}\right],
\end{aligned}$$

where the expectation is taken over the indicators $\boldsymbol{W}$. The first term equals to $-n_1n_2D(\mathcal{F}(\widehat{\mu}\boldsymbol{J}+\mathcal{T}_\beta(\boldsymbol{Z}_\star))\|\mathcal{F}(\widehat{\boldsymbol{M}}_\beta))$. For the second term, similar to $\phi_{1,ij}(x)$ and $\phi_{2,ij}(x)$, we can construct corresponding $\phi_3(x)$ and $\phi_4(x)$ for any $|t| \leq \alpha_1$ and $|x| \leq \beta$ where $\phi_3(x) = L^{-1}_{\alpha_1+\beta}\log(f(t+x)/f(t))$ and $\phi_4(x) = L^{-1}_{\alpha_1+\beta}\log\{[1-f(t-x)]/[1-f(t)]\}$. Due to the similar facts for $\phi_3(x)$ and $\phi_4(x)$ that $|\phi_3(x_1)-$

$\phi_3(x_2)| \leq |x_1 - x_2|$ and $|\phi_4(x_1) - \phi_4(x_2)| \leq |x_1 - x_2|$, we have

$$\sum_{z_{\star,ij}>\beta} \left[\{f(m_{\star,ij}) - f(\widehat{\mu}+\beta)\} \log \left\{ \frac{f(\widehat{\mu}+\widehat{z}_{ij,\beta})}{f(\widehat{\mu}+\beta)} \right\} + \{f(\widehat{\mu}+\beta) - f(m_{\star,ij})\} \log \left\{ \frac{1-f(\widehat{\mu}+\widehat{z}_{ij,\beta})}{1-f(\widehat{\mu}+\beta)} \right\} \right]$$

$$\leq 2L_{\alpha_1+\beta} \sum_{z_{\star,ij}>\beta} |f(m_{\star,ij}) - f(\widehat{\mu}+\beta)| \, |\beta + \widehat{\mu} - \widehat{z}_{ij,\beta} - \widehat{\mu}| \leq 8L_{\alpha_1+\beta}\beta \sum_{i,j=1}^{n_1,n_2} \mathbb{I}_{[z_{\star,ij}>\beta]}.$$

Similarly we can bound the third term in the same way. For the fourth term, we have

$$\begin{aligned}
&\sum_{-\beta<z_{\star,ij}<\beta} \left[\{f(m_{\star,ij}) - f(\widehat{\mu}+z_{\star,ij})\} \log \left\{ \frac{f(\widehat{\mu}+\widehat{z}_{ij,\beta})}{f(\widehat{\mu}+z_{\star,ij})} \right\} + \right. \\
&\left. \{f(\widehat{\mu}+z_{\star,ij}) - f(m_{\star,ij})\} \log \left\{ \frac{1-f(\widehat{\mu}+\widehat{z}_{ij,\beta})}{1-f(\widehat{\mu}+z_{\star,ij})} \right\} \right] \\
\leq & 2L_{\alpha_1+\beta} \sum_{-\beta<z_{\star,ij}<\beta} |f(m_{\star,ij}) - f(\widehat{\mu}+z_{\star,ij})| \, |z_{\star,ij} + \widehat{\mu} - \widehat{z}_{ij,\beta} - \widehat{\mu}| \\
\leq & L^2_{\alpha_1+\beta}\beta(n_1n_2 - N_\beta) |\mu_\star - \widehat{\mu}|.
\end{aligned}$$

Together we have

$$\begin{aligned}
E_{\boldsymbol{W}}\left\{\widetilde{\ell}_{\boldsymbol{W}}\left(\widehat{\boldsymbol{M}}_\beta\right)\right\} \leq & -n_1n_2D\left[\mathcal{F}\{\widehat{\mu}\boldsymbol{J} + \mathcal{T}_\beta(\boldsymbol{Z}_\star)\} \| \mathcal{F}\left(\widehat{\boldsymbol{M}}_\beta\right)\right] + 8L_{\alpha_1+\beta}N_\beta\beta \\
& + L^2_{\alpha_1+\beta}\beta(n_1n_2 - N_\beta)|\mu_\star - \widehat{\mu}|.
\end{aligned}$$

Then for $\widehat{\boldsymbol{M}}_\beta = \widehat{\mu}\boldsymbol{J} + \widehat{\boldsymbol{Z}}_\beta$, we have

$$\begin{aligned}
\ell_{\boldsymbol{W}}\left(\widehat{\boldsymbol{M}}_\beta\right) - \ell_{\boldsymbol{W}}\{\widehat{\mu}\boldsymbol{J} + \mathcal{T}_\beta(\boldsymbol{Z}_\star)\} = \widetilde{\ell}_{\boldsymbol{W}}\left(\widehat{\boldsymbol{M}}_\beta\right) \leq & -n_1n_2D\left[\mathcal{F}\{\widehat{\mu}\boldsymbol{J} + \mathcal{T}_\beta(\boldsymbol{Z}_\star)\} \| \mathcal{F}\left(\widehat{\boldsymbol{M}}_\beta\right)\right] \\
& + \left|\widetilde{\ell}_{\boldsymbol{W}}\left(\widehat{\boldsymbol{M}}_\beta\right) - E_{\boldsymbol{W}}\left\{\widetilde{\ell}_{\boldsymbol{W}}\left(\widehat{\boldsymbol{M}}_\beta\right)\right\}\right| + L_{\alpha_1+\beta}\beta\{8N_\beta + L_{\alpha_1+\beta}(n_1n_2 - N_\beta)|\mu_\star - \widehat{\mu}|\}.
\end{aligned}$$

Moreover, from the definition of $\widehat{\boldsymbol{Z}}_\beta$, we also have that $\ell_{\boldsymbol{W}}(\widehat{\mu}\boldsymbol{J} + \widehat{\boldsymbol{Z}}_\beta) \geq$

$\ell_{\boldsymbol{W}}\{\widehat{\mu}\boldsymbol{J}+\mathcal{T}_\beta(\boldsymbol{Z}_\star)\}$. Thus

$$\begin{aligned}0 \leq & -n_1n_2D\left[\mathcal{F}\left\{\widehat{\mu}\boldsymbol{J}+\mathcal{T}_\beta\left(\boldsymbol{Z}_\star\right)\right\}\|\mathcal{F}\left(\widehat{\boldsymbol{M}}_\beta\right)\right] \\ & +\left|\widetilde{\ell}_{\boldsymbol{W}}\left(\widehat{\boldsymbol{M}}_\beta\right)-E_{\boldsymbol{W}}\left\{\widetilde{\ell}_{\boldsymbol{W}}\left(\widehat{\boldsymbol{M}}_\beta\right)\right\}\right| \\ & +L_{\alpha_1+\beta}\beta\left\{8N_\beta+L_{\alpha_1+\beta}\left(n_1n_2-N_\beta\right)\left|\mu_\star-\widehat{\mu}\right|\right\}.\end{aligned}$$

Applying Remark 2 for matrix $\boldsymbol{M}=\widehat{\boldsymbol{M}}_\beta=\widehat{\mu}\boldsymbol{J}+\widehat{\boldsymbol{Z}}_\beta$, for any $C_0>16e$, we obtain that with probability at least $1-C_1/(n_1+n_2)$,

$$\begin{aligned}0 \leq & -n_1n_2D\left\{\mathcal{F}\left(\boldsymbol{M}_{\star,\beta}\right)\|\mathcal{F}\left(\widehat{\boldsymbol{M}}_\beta\right)\right\}+C_{L_{\alpha_1+\beta},n_1,n_2} \\ & +L_{\alpha_1+\beta}\beta\left\{8N_\beta+L_{\alpha_1+\beta}\left(n_1n_2-N_\beta\right)\left|\mu_\star-\widehat{\mu}\right|\right\}.\end{aligned}$$

Combining with Theorem S1, with probability at least $1-2C_1/(n_1+n_2)$, we obtain

$$\begin{aligned}d_H^2\left[\mathcal{F}\left\{\widehat{\mu}\boldsymbol{J}+\mathcal{T}_\beta\left(\boldsymbol{Z}_\star\right)\right\},\mathcal{F}\left(\widehat{\boldsymbol{M}}_\beta\right)\right] & \leq D\left[\mathcal{F}\left\{\widehat{\mu}\boldsymbol{J}+\mathcal{T}_\beta\left(\boldsymbol{Z}_\star\right)\right\}\|\mathcal{F}\left(\widehat{\boldsymbol{M}}_\beta\right)\right] \\ & \leq \frac{C_{L_{\alpha_1+\beta},n_1,n_2}+L_{\alpha_1+\beta}\beta\left\{8N_\beta+L_{\alpha_1+\beta}\left(n_1n_2-N_\beta\right)\left|\mu_\star-\widehat{\mu}\right|\right\}}{n_1n_2}.\end{aligned}$$

It implies that

$$d^2\left\{\mathcal{T}_\beta\left(\boldsymbol{Z}_\star\right),\widehat{\boldsymbol{Z}}_\beta\right\} \leq \frac{\gamma_{\beta+\alpha_1}\left[C_{L_{\alpha_1+\beta},n_1,n_2}+L_{\alpha_1+\beta}\beta\left\{8N_\beta+L_{\alpha_1+\beta}\left(n_1n_2-N_\beta\right)\left|\mu_\star-\widehat{\mu}\right|\right\}\right]}{n_1n_2}.$$

□

Next we will prove Theorem S2 in two parts.

Proof of $\widehat{\boldsymbol{M}}_\beta$ part in Theorem S2. Combining $\|\mathcal{T}_\beta(\boldsymbol{Z}_\star)\|_\infty \leq \beta$ and $\|\boldsymbol{Z}_\star\|_\infty \leq \alpha_2$, by Lemma 2 of Davenport et al. (2014), Theorems S1 and S4, we have with

probability at least $1 - 4C_1/(n_1 + n_2)$,

$$\begin{aligned} d^2\left(\widehat{\boldsymbol{M}}_\beta, \boldsymbol{M}_\star\right) \leq & 2\left[(\widehat{\mu} - \mu_\star)^2 + d^2\left\{\widehat{\boldsymbol{Z}}_\beta, \mathcal{T}_\beta\left(\boldsymbol{Z}_\star\right)\right\} + d^2\left\{\mathcal{T}_\beta\left(\boldsymbol{Z}_\star\right), \boldsymbol{Z}_\star\right\}\right] \\ \leq & C_2 \min\left\{\alpha_1^2, \frac{\gamma_{\alpha_0} C_{L_{\alpha_0}, n_1, n_2}}{n_1 n_2}\right\} + \frac{2(\alpha_2 - \beta)_+^2 N_\beta}{n_1 n_2} \\ & + C_3' \min\left[\beta^2, \frac{\gamma_{\beta+\alpha_1}\left\{C_{L_{\alpha_1+\beta}, n_1, n_2} + L_{\alpha_1+\beta}\beta\left(8N_\beta + L_{\alpha_1+\beta}\left(n_1 n_2 - N_\beta\right)|\mu_\star - \widehat{\mu}|\right)\right\}}{n_1 n_2}\right], \end{aligned}$$

where $x_+ = \max(x, 0)$. □

Proof of $\widehat{\boldsymbol{\Theta}}_\beta^\dagger$ part in Theorem S2. By Taylor's theorem, there is some η between $\widehat{\mu} + \widehat{z}_{ij,\beta}$ and $\mu_\star + T_\beta(z_{\star,ij})$ so that

$$\frac{1}{f\left(\widehat{z}_{ij,\beta}\right)} - \frac{1}{f\left(\mu_\star + T_\beta\left(z_{\star,ij}\right)\right)} = -\frac{f'\left(\eta\right)}{f^2\left(\eta\right)}\left(\widehat{\mu} + \widehat{z}_{ij,\beta} - \mu_\star - T_\beta\left(z_{\star,ij}\right)\right).$$

By $\|\widehat{\boldsymbol{M}}_\beta\|_\infty \leq \alpha_1 + \beta$, $\|\boldsymbol{M}_{\star,\beta}\|_\infty \leq \alpha_1 + \beta$, and the definition of $L_{\alpha_1+\beta}$ and $h_{\alpha_1,\beta}$, we have

$$\frac{f'\left(\eta\right)}{f^2\left(\eta\right)} \leq \frac{1}{h_{\alpha_1,\beta}} \sup_{|\eta| \leq \alpha_1+\beta} \frac{|f'\left(\eta\right)|}{f\left(\eta\right)\left(1 - f\left(\eta\right)\right)} = \frac{L_{\alpha_1+\beta}}{h_{\alpha_1,\beta}},$$

which further implies

$$d^2\left(\widehat{\boldsymbol{\Theta}}_\beta^\dagger, \boldsymbol{\Theta}_{\star,\beta}^\dagger\right) \leq \frac{L_{\alpha_1+\beta}^2}{h_{\alpha_1,\beta}^2} d^2\left(\widehat{\mu}\boldsymbol{J} + \widehat{\boldsymbol{Z}}_\beta, \boldsymbol{M}_{\star,\beta}\right) \leq \frac{2L_{\alpha_1+\beta}^2}{h_{\alpha_1,\beta}^2}\left[(\widehat{\mu} - \mu_\star)^2 + d^2\left\{\widehat{\boldsymbol{Z}}_\beta, \mathcal{T}_\beta\left(\boldsymbol{Z}_\star\right)\right\}\right]. \tag{S1.17}$$

As it is easy to bound the remaining part by $\|\boldsymbol{\Theta}_{\star,\beta}^\dagger - \boldsymbol{\Theta}_\star^\dagger\|_F^2 \leq 4N_\beta/\theta_L^2$, we

have

$$\begin{aligned} d^2\left(\widehat{\boldsymbol{\Theta}}_{\beta}^{\dagger}, \boldsymbol{\Theta}_{\star}^{\dagger}\right) \leq & 2\left\{d^2\left(\widehat{\boldsymbol{\Theta}}_{\beta}^{\dagger}, \boldsymbol{\Theta}_{\star,\beta}^{\dagger}\right)+d^2\left(\boldsymbol{\Theta}_{\star,\beta}^{\dagger}, \boldsymbol{\Theta}_{\star}^{\dagger}\right)\right\} \\ \leq & \frac{2 L_{\alpha_1+\beta}^2}{h_{\alpha_1,\beta}^2} d^2\left(\widehat{\mu} \boldsymbol{J}+\widehat{\boldsymbol{Z}}_{\beta}, \boldsymbol{M}_{\star,\beta}\right)+\frac{8 N_{\beta}}{n_1 n_2 \theta_L^2}. \end{aligned}$$

□

Write $\boldsymbol{J}_{ij} = e_i(n_1)e_j^{\mathrm{T}}(n_2)$, where $e_i(n) \in \mathbb{R}^n$ is the standard basis vector with the i-th element being 1 and the rest being 0. Define

$$\zeta_{\boldsymbol{W}} = \max\left\{\left\|\boldsymbol{W} \circ \left(\widehat{\boldsymbol{\Theta}}_{\beta}^{\dagger} - \widehat{\boldsymbol{\Theta}}_{\star,\beta}^{\dagger}\right)\right\|_{\infty,2}, \left\|\boldsymbol{W} \circ (\widehat{\boldsymbol{\Theta}}_{\beta}^{\dagger} - \widehat{\boldsymbol{\Theta}}_{\star,\beta}^{\dagger})^{\mathrm{T}}\right\|_{\infty,2}\right\}.$$

Lemma S2. *Assume Conditions C1-C4 hold, denote* $\Psi^{(1)} = \sum_{i,j=1}^{n_1,n_2} w_{ij}\epsilon_{ij}\boldsymbol{J}_{ij}/(n_1 n_2 \widehat{\theta}_{ij,\beta})$*, for some positive constants* C_1 *and* δ*, there exists* $\Delta^{(1)}$ *such that*

$$\left\|\Psi^{(1)}\right\| \leq \Delta^{(1)} \asymp \max\left[\frac{c_\sigma h_{(2),\beta}\left\{(n_1 \vee n_2)\log(n_1+n_2)\right\}^{1/2}}{n_1 n_2}, \frac{\eta \zeta_{\boldsymbol{W}} \log^{1+\delta}(n_1+n_2)}{n_1 n_2}\right],$$

holds with probability at least $1 - 1/(n_1+n_2) - 1/(n_1+n_2)^{\delta}$.

Proof of Lemma S2. Due to the triangle inequality, we have

$$\begin{aligned} \left\|\Psi^{(1)}\right\| = & \frac{1}{n_1 n_2}\left\|\sum_{i,j=1}^{n_1,n_2} \frac{\epsilon_{ij} w_{ij}}{\widehat{\theta}_{ij,\beta}} \boldsymbol{J}_{ij}\right\| \leq \frac{1}{n_1 n_2}\left\|\sum_{i,j=1}^{n_1,n_2} \frac{\epsilon_{ij} w_{ij}}{\theta_{\star,ij,\beta}} \boldsymbol{J}_{ij}\right\| \\ & + \frac{1}{n_1 n_2}\left\|\sum_{i,j=1}^{n_1,n_2} \epsilon_{ij} w_{ij}\left(\frac{1}{\widehat{\theta}_{ij,\beta}} - \frac{1}{\theta_{\star,ij,\beta}}\right) \boldsymbol{J}_{ij}\right\|. \end{aligned} \tag{S1.18}$$

Since ϵ_{ij} is independent of w_{ij}, we have

$$E\left(\frac{\epsilon_{ij} w_{ij}}{\theta_{\star,ij,\beta}} \boldsymbol{J}_{ij}\right) = E(\epsilon_{ij})\, E\left(\frac{w_{ij}\boldsymbol{J}_{ij}}{\theta_{\star,ij,\beta}}\right) = \mathbf{0}.$$

Write $\eta_H = \eta h_{(2),\beta}/\theta_L^{1/2}$, where η is the constant in Condition C1. Then we have

$$E\left(\frac{\epsilon_{ij}w_{ij}}{\theta_{\star,ij,\beta}}\right)^l = E\left\{E\left(\frac{\epsilon_{ij}w_{ij}}{\theta_{\star,ij,\beta}}\,|\{w_{ij}\}\right)^l\right\} \le E\left\{\frac{l!}{2}\cdot\left(\frac{c_\sigma h_{(2),\beta}w_{ij}}{\theta_{\star,ij}^{1/2}}\right)^2\left(\frac{\eta w_{ij}}{\theta_{\star,ij}^{1/2}}\right)^{l-2}\right\}$$
$$\le \frac{l!}{2}\cdot\left(c_\sigma h_{(2),\beta}\right)^2\eta_H^{l-2} \quad \text{for } l = 2,3,4,\ldots.$$

Similar as the proof of Lemma S1.1 in Mao et al. (2019), we can show that with probability at least $1 - 1/(n_1+n_2)$,

$$\frac{1}{n_1 n_2}\left\|\sum_{i,j=1}^{n_1,n_2}\frac{\epsilon_{ij}w_{ij}}{\theta_{\star,ij,\beta}}\boldsymbol{J}_{ij}\right\| \le \frac{2c_\sigma h_{(2),\beta}\left\{2\left(n_1\vee n_2\right)\log\left(n_1+n_2\right)\right\}^{1/2}}{n_1 n_2}.$$

Denote $\psi_{ij} = \epsilon_{ij}w_{ij}(\widehat{\theta}_{ij,\beta}^{-1} - \theta_{\star,ij,\beta}^{-1})$. By the matrix Bernstein inequality (Tropp, 2012, Theorem 6.2), we show that, for all $t > 0$,

$$\Pr\left(\left\|\sum_{i,j=1}^{n_1,n_2}\psi_{ij}\boldsymbol{J}_{ij}\right\| \ge t \mid \{w_{ij}\}\right) \le (n_1+n_2)\cdot\exp\left\{\frac{-t^2/2}{c_\sigma^2\zeta_{\boldsymbol{W}}^2 + \eta\left\|\boldsymbol{W}\circ\left(\widehat{\boldsymbol{\Theta}}_\beta^\dagger - \boldsymbol{\Theta}_{\star,\beta}^\dagger\right)\right\|_\infty t}\right\}.$$

We further have

$$\Pr\left(\left\|\sum_{i,j=1}^{n_1,n_2}\psi_{ij}\boldsymbol{J}_{ij}\right\| \ge t \mid \{w_{ij}\}\right) \le (n_1+n_2)\cdot\exp\left\{\frac{-t^2/2}{c_\sigma^2\zeta_{\boldsymbol{W}}^2 + \eta\zeta_{\boldsymbol{W}}t}\right\}. \tag{S1.19}$$

For any $\delta > 0$, take $t = \eta\zeta_{\boldsymbol{W}}\log^{1+\delta}(n_1+n_2)$ so that $c_\sigma^2\zeta_{\boldsymbol{W}}^2 \le \eta\zeta_{\boldsymbol{W}}t$, we have that

$$(n_1+n_2)\cdot\exp\left[\frac{-t^2/2}{c_\sigma^2\zeta_{\boldsymbol{W}}^2 + \eta\zeta_{\boldsymbol{W}}t}\right] \le (n_1+n_2)\cdot\exp\left[-\frac{t^2}{4\eta\zeta_{\boldsymbol{W}}t}\right] \le \frac{1}{(n_1+n_2)^{2\delta}}.$$

Combining all two terms in (S1.18) together, for any $\delta > 0$, with probability

at least $1-1/(n_1+n_2)-1/(n_1+n_2)^{2\delta}$, we have

$$\frac{1}{n_1n_2}\left\|\sum_{i,j=1}^{n_1,n_2}\frac{\epsilon_{ij}w_{ij}}{\widehat{\theta}_{ij,\beta}}\boldsymbol{J}_{ij}\right\| \leq \frac{1}{n_1n_2}\left\|\sum_{i,j=1}^{n_1,n_2}\frac{\epsilon_{ij}w_{ij}}{\theta_{\star,ij,\beta}}\boldsymbol{J}_{ij}\right\| + \frac{1}{n_1n_2}\left\|\sum_{i,j=1}^{n_1,n_2}\epsilon_{ij}w_{ij}\left(\frac{1}{\widehat{\theta}_{ij,\beta}}-\frac{1}{\theta_{\star,ij,\beta}}\right)\boldsymbol{J}_{ij}\right\|$$
$$= \Delta^{(1)} \asymp \max\left[\frac{c_\sigma h_{(2),\beta}\left\{(n_1\vee n_2)\log(n_1+n_2)\right\}^{1/2}}{n_1n_2}, \frac{\eta\zeta_{\boldsymbol{W}}\log^{1+\delta}(n_1+n_2)}{n_1n_2}\right].$$

□

Lemma S3. *Assume Conditions C1-C4 hold, $\widehat{\boldsymbol{A}}_\beta$ is the solution to* (4.3)*, denote* $\Psi^{(2)}=\sum_{i,j=1}^{n_1,n_2}(a_{\star,ij}-\hat{a}_{ij,\beta})(w_{ij}/\widehat{\theta}_{ij,\beta}-w_{ij}/\theta_{\star,ij,\beta})\boldsymbol{J}_{ij}/(n_1n_2)$*, there exists*

$$\Delta^{(2)} \asymp \frac{a_0\left\|\widehat{\boldsymbol{\Theta}}^\dagger_\beta-\boldsymbol{\Theta}^\dagger_{\star,\beta}\right\|_F}{n_1n_2},$$

such that $\|\Psi^{(2)}\|\leq\Delta^{(2)}$ *holds.*

Proof of Lemma S3. By the inequality (S1.10), we have

$$n_1n_2\left\|\Psi^{(2)}\right\| \leq \left\|\boldsymbol{W}\circ\left(\widehat{\boldsymbol{\Theta}}^\dagger_\beta-\boldsymbol{\Theta}^\dagger_{\star,\beta}\right)\circ\left(\boldsymbol{A}_\star-\widehat{\boldsymbol{A}}_\beta\right)\right\|_F \leq 2a_0\left\|\widehat{\boldsymbol{\Theta}}^\dagger_\beta-\boldsymbol{\Theta}^\dagger_{\star,\beta}\right\|_F.$$

□

For a $0\leq r\leq(n_1\wedge n_2)$, we consider the following constraint set

$$\mathcal{C}(r)=\left\{\boldsymbol{A}\in\mathbb{R}^{n_1\times n_2}:\|\boldsymbol{A}\|_\infty=1,\left\|\boldsymbol{\Theta}^\S_\star\circ\boldsymbol{\Theta}^\ddagger_{\star,\beta}\circ\boldsymbol{A}\right\|_F^2\geq\sqrt{\frac{64\log(n_1+n_2)}{\log(6/5)n_1n_2\pi_L}},\|\boldsymbol{A}\|_*\leq\sqrt{r}\|\boldsymbol{A}\|_F\right\},\tag{S1.20}$$

where $\boldsymbol{\Theta}^\S_\star=(\theta^{1/2}_{\star,ij})$. It is easy to see that once $\mathrm{rank}(\boldsymbol{A})\leq r$, $\|\boldsymbol{A}\|_*\leq\sqrt{r}\|\boldsymbol{A}\|_F$ holds.

Again, let ξ_{ij} are i.i.d. Rademacher random variables and define $\Psi^{(3)}=(n_1n_2)^{-1}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\xi_{ij}w_{ij}\theta^{-1}_{\star,ij,\beta}$.

Lemma S4. *Assume Conditions C1,C2 hold, let ξ_{ij} are i.i.d. Rademacher random variables. Then, for $n_1 n_2 \theta_L \geq c(n_1 \wedge n_2)\log^3(n_1+n_2)$ where c is some constant, there exist an absolute constant C_2 such that*

$$E\left\|\Psi^{(3)}\right\| \leq \frac{C_2 h_{(2),\beta}\left\{(n_1 \vee n_2)\log\left(n_1+n_2\right)\right\}^{1/2}}{n_1 n_2}.$$

Proof of Lemma S4. The proof can be completed by following the proof of Lemma 6 in Klopp (2014). □

Lemma S5. *Assume Conditions C1,C2 and C4 hold, then for all $\boldsymbol{A} \in \mathcal{C}(r)$ where $\mathcal{C}(r)$ is defined in* (S1.20)*, we have*

$$\frac{1}{n_1 n_2}\left\|\boldsymbol{W}\circ\boldsymbol{\Theta}^{\ddagger}_{\star,\beta}\circ\boldsymbol{A}\right\|_F^2 \geq \frac{1}{2n_1 n_2}\left\|\boldsymbol{\Theta}^{\S}_{\star}\circ\boldsymbol{\Theta}^{\ddagger}_{\star,\beta}\circ\boldsymbol{A}\right\|_F^2 - C_3 n_1 n_2 h_{(1),\beta}\left(E\left\|\Psi^{(3)}\right\|\right)^2,$$

where $\boldsymbol{\Theta}^{\S}_{\star} = (\theta^{1/2}_{\star,ij})$ with probability at least $1-1/(n_1+n_2)$ for some constant C_3.

Proof of Lemma S5. The proof can be completed by following the proof of Lemma 12 in Klopp (2014). □

Proof of Theorem S3. It follows from the definition of $\widehat{\boldsymbol{A}}_\beta$ that,

$$\frac{1}{n_1 n_2}\left\|\boldsymbol{W}\circ\widehat{\boldsymbol{\Theta}}^{\ddagger}_{\beta}\circ\left(\widehat{\boldsymbol{A}}_\beta - \boldsymbol{Y}\right)\right\|_F^2 + \tau\left\|\widehat{\boldsymbol{A}}_\beta\right\|_* \leq \frac{1}{n_1 n_2}\left\|\boldsymbol{W}\circ\widehat{\boldsymbol{\Theta}}^{\ddagger}_{\beta}\circ(\boldsymbol{A}_\star - \boldsymbol{Y})\right\|_F^2 + \tau\left\|\boldsymbol{A}_\star\right\|_*. \tag{S1.21}$$

Since we can rewrite the first term in both the left and right hand sides of

(S1.21) as

$$\frac{1}{n_1 n_2}\left\|\boldsymbol{W}\circ\widehat{\boldsymbol{\Theta}}^{\ddagger}_{\beta}\circ\left(\widehat{\boldsymbol{A}}_{\beta}-\boldsymbol{Y}\right)\right\|_F^2=\frac{1}{n_1 n_2}\left\|\boldsymbol{W}\circ\widehat{\boldsymbol{\Theta}}^{\ddagger}_{\beta}\circ\left(\widehat{\boldsymbol{A}}_{\beta}-\boldsymbol{A}_{\star}+\boldsymbol{A}_{\star}-\boldsymbol{Y}\right)\right\|_F^2,$$

the inequality (S1.21) is equivalent to

$$\begin{aligned}\frac{1}{n_1 n_2}\left\|\boldsymbol{W}\circ\widehat{\boldsymbol{\Theta}}^{\ddagger}_{\beta}\circ\left(\widehat{\boldsymbol{A}}_{\beta}-\boldsymbol{A}_{\star}\right)\right\|_F^2\leq&\frac{2}{n_1 n_2}\left\langle\boldsymbol{W}\circ\widehat{\boldsymbol{\Theta}}^{\ddagger}_{\beta}\circ\left(\widehat{\boldsymbol{A}}_{\beta}-\boldsymbol{A}_{\star}\right),\boldsymbol{W}\circ\widehat{\boldsymbol{\Theta}}^{\ddagger}_{\beta}\circ(\boldsymbol{Y}-\boldsymbol{A}_{\star})\right\rangle\\&+\tau\left(\|\boldsymbol{A}_{\star}\|_*-\left\|\widehat{\boldsymbol{A}}_{\beta}\right\|_*\right)\\=&\frac{2}{n_1 n_2}\left\langle\widehat{\boldsymbol{A}}_{\beta}-\boldsymbol{A}_{\star},\boldsymbol{W}\circ\widehat{\boldsymbol{\Theta}}^{\dagger}_{\beta}\circ\boldsymbol{\epsilon}\right\rangle+\tau\left(\|\boldsymbol{A}_{\star}\|_*-\left\|\widehat{\boldsymbol{A}}_{\beta}\right\|_*\right).\end{aligned}$$

For the left hand side, we have

$$\begin{aligned}\left\|\boldsymbol{W}\circ\widehat{\boldsymbol{\Theta}}^{\ddagger}_{\beta}\circ\left(\widehat{\boldsymbol{A}}_{\beta}-\boldsymbol{A}_{\star}\right)\right\|_F^2=&\left\langle\boldsymbol{W}\circ\widehat{\boldsymbol{\Theta}}^{\ddagger}_{\beta}\circ\left(\widehat{\boldsymbol{A}}_{\beta}-\boldsymbol{A}_{\star}\right),\boldsymbol{W}\circ\widehat{\boldsymbol{\Theta}}^{\ddagger}_{\beta}\circ\left(\widehat{\boldsymbol{A}}_{\beta}-\boldsymbol{A}_{\star}\right)\right\rangle\\=&\left\langle\widehat{\boldsymbol{A}}_{\beta}-\boldsymbol{A}_{\star},\boldsymbol{W}\circ\widehat{\boldsymbol{\Theta}}^{\dagger}_{\beta}\circ\left(\widehat{\boldsymbol{A}}_{\beta}-\boldsymbol{A}_{\star}\right)\right\rangle\\=&\left\langle\widehat{\boldsymbol{A}}_{\beta}-\boldsymbol{A}_{\star},\boldsymbol{W}\circ\left(\widehat{\boldsymbol{\Theta}}^{\dagger}_{\beta}-\boldsymbol{\Theta}^{\dagger}_{\star,\beta}\right)\circ\left(\widehat{\boldsymbol{A}}_{\beta}-\boldsymbol{A}_{\star}\right)\right\rangle\\&+\left\langle\widehat{\boldsymbol{A}}_{\beta}-\boldsymbol{A}_{\star},\boldsymbol{W}\circ\boldsymbol{\Theta}^{\dagger}_{\star,\beta}\circ\left(\widehat{\boldsymbol{A}}_{\beta}-\boldsymbol{A}_{\star}\right)\right\rangle.\end{aligned}$$

It implies that we can turns the above inequality to be

$$\begin{aligned}\frac{1}{n_1 n_2}\left\|\boldsymbol{W}\circ\boldsymbol{\Theta}^{\ddagger}_{\star,\beta}\circ\left(\widehat{\boldsymbol{A}}_{\beta}-\boldsymbol{A}_{\star}\right)\right\|_F^2\leq&\frac{1}{n_1 n_2}\left\langle\widehat{\boldsymbol{A}}_{\beta}-\boldsymbol{A}_{\star},\boldsymbol{W}\circ\left(\widehat{\boldsymbol{\Theta}}^{\dagger}_{\beta}-\boldsymbol{\Theta}^{\dagger}_{\star,\beta}\right)\circ\left(\boldsymbol{A}_{\star}-\widehat{\boldsymbol{A}}_{\beta}\right)\right\rangle+\\&\frac{2}{n_1 n_2}\left\langle\widehat{\boldsymbol{A}}_{\beta}-\boldsymbol{A}_{\star},\boldsymbol{W}\circ\widehat{\boldsymbol{\Theta}}^{\dagger}_{\beta}\circ\boldsymbol{\epsilon}\right\rangle+\tau\left(\|\boldsymbol{A}_{\star}\|_*-\left\|\widehat{\boldsymbol{A}}_{\beta}\right\|_*\right).\end{aligned}\tag{S1.22}$$

In the following, we use $\Psi^{(i)}$, for $i=1,2$, which are defined in Lemmas S2-S3. Under Conditions C1-C4, Lemmas S2-S3 show that there exist constants $\Delta^{(1)}$ and $\Delta^{(2)}$ such that with probability at least $1-1/(n_1+n_2)-1/(n_1+n_2)^{2\delta}$

and 1 respectively. As defined in (S1.7), we have for a positive constant C_4, $2\Delta^{(1)} + \Delta^{(2)} \leq C_4\widetilde{\Delta}$ with probability at least $1 - 2/(n_1 + n_2)$. Thus we can simplify (S1.22) due to the trace duality (S1.9) to be

$$\frac{1}{n_1 n_2}\left\|\boldsymbol{W} \circ \boldsymbol{\Theta}^{\ddagger}_{\star,\beta} \circ \left(\widehat{\boldsymbol{A}}_\beta - \boldsymbol{A}_\star\right)\right\|_F^2 \leq C_4\widetilde{\Delta}\left\|\widehat{\boldsymbol{A}}_\beta - \boldsymbol{A}_\star\right\|_* + \tau\left(\|\boldsymbol{A}_\star\|_* - \left\|\widehat{\boldsymbol{A}}_\beta\right\|_*\right).$$

Write the singular value decomposition of $\boldsymbol{A}_\star$ as $\sum_{i=1}^{r_{\boldsymbol{A}_\star}} \sigma_i(\boldsymbol{A}_\star) u^{(i)}_{\boldsymbol{A}_\star} v^{(i)T}_{\boldsymbol{A}_\star}$. Let $\mathcal{A}_{\star u}$ be the linear span of $u^{(1)}_{\boldsymbol{A}_\star}, \ldots, u^{(r_{\boldsymbol{A}_\star})}_{\boldsymbol{A}_\star}$ and $\mathcal{A}_{\star v}$ be the linear span of $v^{(1)}_{\boldsymbol{A}_\star}, \ldots, v^{(r_{\boldsymbol{A}_\star})}_{\boldsymbol{A}_\star}$. For any matrix $\boldsymbol{B}$, define the operators

$$\boldsymbol{P}_{\mathcal{A}_\star{}^\perp}(\boldsymbol{B}) = P_{\mathcal{A}_{\star u}{}^\perp}\boldsymbol{B}P_{\mathcal{A}_{\star v}{}^\perp} \quad \text{and} \quad \boldsymbol{P}_{\mathcal{A}_\star}(\boldsymbol{B}) = \boldsymbol{B} - \boldsymbol{P}_{\mathcal{A}_\star{}^\perp}(\boldsymbol{B}).$$

To prove the remaining bounds, note the fact that for any $\widehat{\boldsymbol{A}}_\beta$,

$$\|\boldsymbol{A}_\star\|_* - \left\|\widehat{\boldsymbol{A}}_\beta\right\|_* \leq \left\|\boldsymbol{P}_{\mathcal{A}_\star}\left(\boldsymbol{A}_\star - \widehat{\boldsymbol{A}}_\beta\right)\right\|_* - \left\|\boldsymbol{P}_{\mathcal{A}_\star{}^\perp}\left(\boldsymbol{A}_\star - \widehat{\boldsymbol{A}}_\beta\right)\right\|_*. \quad \text{(S1.23)}$$

This, the triangle inequality and $\tau \geq \frac{3}{2}C_4\widetilde{\Delta}$ lead to

$$\frac{1}{n_1 n_2}\left\|\boldsymbol{W} \circ \boldsymbol{\Theta}^{\ddagger}_{\star,\beta} \circ \left(\widehat{\boldsymbol{A}}_\beta - \boldsymbol{A}_\star\right)\right\|_F^2 \leq \frac{5}{3}\tau\left\|\boldsymbol{P}_{\mathcal{A}_\star}\left(\widehat{\boldsymbol{A}}_\beta - \boldsymbol{A}_\star\right)\right\|_*. \quad \text{(S1.24)}$$

Since $\boldsymbol{P}_{\mathcal{A}_\star}(\boldsymbol{B}) = P_{\mathcal{A}_{\star u}{}^\perp}\boldsymbol{B}P_{\mathcal{A}_{\star v}} + P_{\mathcal{A}_{\star u}{}^\perp}\boldsymbol{B}$, $\text{rank}(P_{\mathcal{A}_{\star u}}\boldsymbol{B}) \leq \text{rank}(\boldsymbol{B})$ and $\text{rank}(P_{\mathcal{A}_{\star v}}\boldsymbol{B}) \leq \text{rank}(\boldsymbol{B})$, we have that $\text{rank}(\boldsymbol{P}_{\mathcal{A}_\star}(\boldsymbol{B})) \leq 2\text{rank}(\boldsymbol{A}_\star)$. From (S1.24), we compute

$$\frac{1}{n_1 n_2}\left\|\boldsymbol{W} \circ \boldsymbol{\Theta}^{\ddagger}_{\star,\beta} \circ \left(\widehat{\boldsymbol{A}}_\beta - \boldsymbol{A}_\star\right)\right\|_F^2 \leq \frac{5}{3}\tau\sqrt{2\text{rank}(\boldsymbol{A}_\star)}\left\|\widehat{\boldsymbol{A}}_\beta - \boldsymbol{A}_\star\right\|_F.$$

Together with Lemma S6 and above, it implies that

$$\left\|\widehat{\boldsymbol{A}}_\beta - \boldsymbol{A}_\star\right\|_* \leq 6\left\|\boldsymbol{P}_{\mathcal{A}_\star}\left(\boldsymbol{A}_\star - \widehat{\boldsymbol{A}}_\beta\right)\right\|_* \leq \sqrt{72\text{rank}(\boldsymbol{A}_\star)}\left\|\widehat{\boldsymbol{A}}_\beta - \boldsymbol{A}_\star\right\|_F.$$

Set $a = \|\widehat{\boldsymbol{A}}_\beta - \boldsymbol{A}_\star\|_\infty$. By definition of $\widehat{\boldsymbol{A}}_\beta$, we have that $a \leq 2a_0$.

Suppose that $\|\boldsymbol{\Theta}^{\S}_\star \circ \boldsymbol{\Theta}^{\ddagger}_{\star,\beta} \circ (\widehat{\boldsymbol{A}}_\beta - \boldsymbol{A}_\star)\|_F^2 < a^2 h_{(3),\beta}\sqrt{\frac{64\log(n_1+n_2)}{\log(6/5)n_1n_2}}$, then it further implies

$$\frac{1}{n_1n_2}\left\|\widehat{\boldsymbol{A}}_\beta - \boldsymbol{A}_\star\right\|_F^2 \leq \frac{h_{(1),\beta}}{n_1n_2}\left\|\boldsymbol{\Theta}^{\S}_\star \circ \boldsymbol{\Theta}^{\ddagger}_{\star,\beta} \circ (\widehat{\boldsymbol{A}}_\beta - \boldsymbol{A}_\star)\right\|_F^2.$$

Thus we have

$$\frac{1}{n_1n_2}\left\|\widehat{\boldsymbol{A}}_\beta - \boldsymbol{A}_\star\right\|_F^2 \leq \frac{C_5 h_{(1),\beta}h_{(3),\beta}\log^{1/2}(n_1+n_2)}{\sqrt{n_1n_2}}.$$

Otherwise if $\|\boldsymbol{\Theta}^{\S}_\star \circ \boldsymbol{\Theta}^{\ddagger}_{\star,\beta} \circ (\widehat{\boldsymbol{A}}_\beta - \boldsymbol{A}_\star)\|_F^2 \geq a^2 h_{(3),\beta}\sqrt{\frac{64\log(n_1+n_2)}{\log(6/5)n_1n_2}}$, we have that $a^{-1}(\widehat{\boldsymbol{A}}_\beta - \boldsymbol{A}_\star) \in \mathcal{C}(72\text{rank}(\boldsymbol{A}_\star))$ due to the definition in (S1.20). By applying lemma S4 we can have with probability at least $1 - 3/(n_1+n_2)$, there exist some constants C_6 and C_7 such that

$$\begin{aligned}\frac{1}{n_1n_2}\left\|\widehat{\boldsymbol{A}}_\beta - \boldsymbol{A}_\star\right\|_F^2 \leq& C_6 n_1n_2 h_{(1),\beta}^2 r_{\boldsymbol{A}_\star}(\tau^2 + \left(E\left\|\Psi^{(3)}\right\|\right)^2)\\ \leq& C_6 n_1n_2 h_{(1),\beta}^2 r_{\boldsymbol{A}_\star}\tau^2 + \frac{C_7 h_{(1),\beta}^2 h_{(2),\beta}^2 r_{\boldsymbol{A}_\star}\log(n_1+n_2)}{(n_1 \wedge n_2)}.\end{aligned}$$

For the above result, let f to be logit inverse link function, we have $h_{(1),\beta}h_{(3),\beta} \leq e^{2\alpha_2 - 2\beta}$, $h_{(1),\beta}h_{(2),\beta} \leq e^{-\mu_\star/2+\alpha_2-\beta+|\alpha_2/2-\beta|}$ and $h_{(2),\beta} \leq e^{\mu_\star/2+\alpha_1+|\alpha_2/2-\beta|}$ which leads to the final version of Theorem (S3). □

Lemma S6. *If* $\tau \geq \frac{3}{2}(2\|\Psi^{(1)}\| + \|\Psi^{(2)}\|)$ *where* $\Psi^{(1)}$ *and* $\Psi^{(2)}$ *are defined in Lemmas S2-S3,*

$$\left\|\boldsymbol{P}_{\mathcal{A}_\star{}^\perp}\left(\boldsymbol{A}_\star - \widehat{\boldsymbol{A}}_\beta\right)\right\|_* \leq 5\left\|\boldsymbol{P}_{\mathcal{A}_\star}\left(\boldsymbol{A}_\star - \widehat{\boldsymbol{A}}_\beta\right)\right\|_*.$$

Proof of Lemma S6. Due to the inequality (S1.22) in the proof of Theorem S3, use the fact that the left hand side always nonnegative, we have that

$$\tau\left(\left\|\widehat{\boldsymbol{A}}_\beta\right\|_* - \|\boldsymbol{A}_\star\|_*\right) \leq \frac{2}{3}\tau\left\|\widehat{\boldsymbol{A}}_\beta - \boldsymbol{A}_\star\right\|_*.$$

This and (S1.23) implies that

$$\left\|\boldsymbol{P}_{\mathcal{A}_\star{}^\perp}\left(\boldsymbol{A}_\star - \widehat{\boldsymbol{A}}_\beta\right)\right\|_* \leq 5\left\|\boldsymbol{P}_{\mathcal{A}_\star}\left(\boldsymbol{A}_\star - \widehat{\boldsymbol{A}}_\beta\right)\right\|_*.$$

□

S1.5 Simulation Study (Cont')

S1.6 Real data application (Cont')

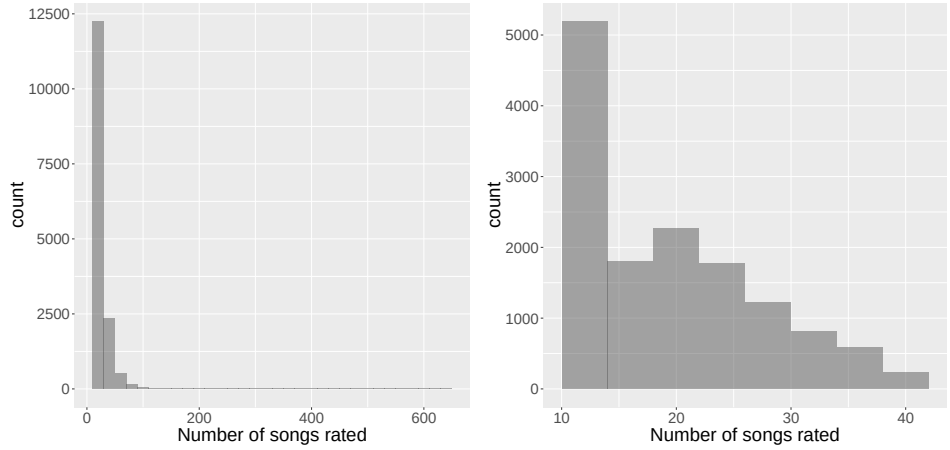

Figure S1: Left: The histogram of the number of songs rated per user in the Yahoo! Webscope dataset. Right: Similar to the left figure but restricted to no more than 40 songs rated per user.

We also demonstrate the proposed methodology by analyzing the Coat dataset

available at *https://www.cs.cornell.edu/ schnabts/mnar/*. It contains (incomplete) ratings from 290 Turkers on 300 items. The dataset consists of two subsets, a training set and a test set. The training set records approximately 7,000 ratings for 24 self-selected coats given by the aforementioned 290 Turkers. The test set was consisted of the ratings for 16 randomly picked coats that are not rated in the training set. The missing rates are 0.92 overall, 0.9172 for each Turker, and 0.7067 to 0.9833 across coats. In this experiment, we applied those methods as described in Section 6 to the training set and evaluated the test errors based on the corresponding test set.

Table S4 reports the root mean squared prediction errors. In additional to the same ten versions of proposed methods, we also report the result of $\mathsf{Prop\_}\widehat{\Theta}_{\mathsf{prop}}$ which missing probability is provided by the propensities in Schnabel et al. (2016). Note that $\mathsf{Prop\_}\widehat{\Theta}_{\beta}\mathsf{\_0.1}$ performs the best among all ten versions of proposed methods. With the propensities estimated by logistic regression in Schnabel et al. (2016), $\mathsf{Prop\_}\widehat{\Theta}_{\mathsf{prop}}$ outperforms the existing methods NW, KLT and MHT. However, even compared with $\mathsf{Prop\_}\widehat{\Theta}_{\mathsf{prop}}$ which is the best, our proposed method perform significantly better in terms of root mean squared prediction errors, and achieve as much as 10% improvement. This suggests that a more flexible modeling of missing structure improves the prediction power.

References

Beck, A. and M. Teboulle (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences* **2**(1), 183–202.

Boyd, S. and L. Vandenberghe (2004). *Convex optimization*. Cambridge university press.

Chen, C., B. He, Y. Ye, and X. Yuan (2016). The direct extension of admm for multi-block convex minimization problems is not necessarily convergent. *Mathematical Programming* **155**(1-2), 57–79.

Davenport, M. A., Y. Plan, E. van den Berg, and M. Wootters (2014). 1-bit matrix completion. *Information and Inference* **3**(3), 189–223.

Klopp, O. (2014). Noisy low-rank matrix completion with general sampling distribution. *Bernoulli* **20**(1), 282–303.

Koltchinskii, V., K. Lounici, and A. B. Tsybakov (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics* **39**(5), 2302–2329.

Ledoux, M. and M. Talagrand (2013). *Probability in Banach Spaces: Isoperimetry and Processes*. Springer Science & Business Media.

Mao, X., S. X. Chen, and R. K. Wong (2019). Matrix completion with covariate information. *Journal of the American Statistical Association* **114**(525), 198–210.

Mazumder, R., T. Hastie, and R. Tibshirani (2010). Spectral regularization algorithms for learning large incomplete matrices. *Journal of Machine Learning Research* **11**(80), 2287–2322.

Negahban, S. and M. J. Wainwright (2012). Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *Journal of Machine Learning Research* **13**(53), 1665–1697.

Schnabel, T., A. Swaminathan, A. Singh, N. Chandak, and T. Joachims (2016). Recommendations as

treatments: Debiasing learning and evaluation. Volume **48** of *Proceedings of Machine Learning Research*, New York, New York, USA, pp. 1670–1679. PMLR.

Seginer, Y. (2000). The expected norm of random matrices. *Combinatorics, Probability and Computing* **9**(2), 149–166.

Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics* **12**(4), 389–434.

Algorithm 1: The ADMM used to solve (3.3)

Input: Initialize $k = 0$, and select u, $\boldsymbol{H}^{(k)}$, $\boldsymbol{Z}^{(k)}$, $\boldsymbol{G}_1^{(k)}$, $\boldsymbol{G}_2^{(k)}$ such that $\boldsymbol{Z}^{(k)}$ is a solution of (3.3) without constraints, $\mathbf{1}_{n_1}^{\mathrm{T}} \boldsymbol{G}_1^{(k)} \mathbf{1}_{n_2} = 0$ and $\|\boldsymbol{G}_2^{(k)}\|_\infty \leq \alpha_2$.

1 Minimize $\mathcal{L}_u(\boldsymbol{Z}, \boldsymbol{G}_1^{(k)}, \boldsymbol{G}_2^{(k)}; \boldsymbol{H}^{(k)})$ with respect to $\boldsymbol{Z}$:

$$\boldsymbol{Z}^{(k+1)} = \mathcal{SVT}_{(uL)^{-1}\lambda}\{1/2(\boldsymbol{G}_1^{(k)} + \boldsymbol{G}_2^{(k)} + 1/u\boldsymbol{H}_1^{(k)} + 1/u\boldsymbol{H}_2^{(k)})\}.$$

Here $\mathcal{SVT}_c$ is the singular value soft-thresholding operator defined as

$$\mathcal{SVT}_c(\boldsymbol{D}) = U\mathrm{diag}(\{(\sigma_i - c)_+\})V^{\mathrm{T}} \quad \text{for any } c \geq 0,$$

where $x_+ = \max(x, 0)$, and $U\Sigma V^{\mathrm{T}}$, with $\Sigma = \mathrm{diag}(\{\sigma_i\})$, is the singular value decomposition of a matrix $\boldsymbol{D}$.

2 Minimize $\mathcal{L}_u(\boldsymbol{Z}^{(k+1)}, \boldsymbol{G}_1, \boldsymbol{G}_2^{(k)}; \boldsymbol{H}^{(k)})$ with respect to $\boldsymbol{G}_1$:

$$\boldsymbol{G}_1^{(k+1)} = \underset{\mathbf{1}_{n_1}^{\mathrm{T}} \boldsymbol{G}_1 \mathbf{1}_{n_2} = 0}{\arg\min} \quad \frac{1}{2}\left\|\boldsymbol{G}_1 - \left(\boldsymbol{Z}^{(k+1)} - 1/u\boldsymbol{H}_1^{(k)}\right)\right\|_F^2,$$

Let $\boldsymbol{B}_1 = \boldsymbol{Z}^{(k+1)} - 1/u\boldsymbol{H}_1^{(k)}$ and simplifies to

$\boldsymbol{G}_1^{(k+1)} = \boldsymbol{B}_1 - (n_1 n_2)^{-1}\mathbf{1}_{n_1}^{\mathrm{T}}\boldsymbol{B}_1\mathbf{1}_{n_2}\boldsymbol{J}$.

3 Minimize $\mathcal{L}_u(\boldsymbol{Z}^{(k+1)}, \boldsymbol{G}_1^{(k+1)}, \boldsymbol{G}_2; \boldsymbol{H}^{(k)})$ with respect to $\boldsymbol{G}_2$:

$$\boldsymbol{G}_2^{(k+1)} = \underset{\|\boldsymbol{G}_2\|_\infty \leq \alpha_2}{\arg\min} \quad \left\|\boldsymbol{G}_2 - \left\{\boldsymbol{Z}_{\text{old}} + \frac{1}{L}\nabla_{\boldsymbol{Z}}\ell_{\boldsymbol{W}}(\mu_{\text{old}}\boldsymbol{J} + \boldsymbol{Z}_{\text{old}}) - \boldsymbol{H}_2^{(k)} + u\boldsymbol{Z}^{(k+1)}\right\}/(1+u)\right\|_F^2.$$

Let $\boldsymbol{B}_2 = \{\boldsymbol{Z}_{\text{old}} + \frac{1}{L}\nabla_{\boldsymbol{Z}}\ell_{\boldsymbol{W}}(\mu_{\text{old}}\boldsymbol{J} + \boldsymbol{Z}_{\text{old}}) - \boldsymbol{H}_2^{(k)} + u\boldsymbol{Z}^{(k+1)}\}/(1+u)$ and simplifies to $\boldsymbol{G}_2^{(k+1)}(i,j) = \min\{\alpha_2, \max\{-\alpha_2, \boldsymbol{B}_2(i,j)\}\}$.

4 Update the dual variable $\boldsymbol{H}^{(k+1)} = (\boldsymbol{H}_1^{(k+1)\mathrm{T}}, \boldsymbol{H}_2^{(k+1)\mathrm{T}})^{\mathrm{T}}$ by

$$\boldsymbol{H}_1^{(k+1)} = \boldsymbol{H}_1^{(k)} - u(\boldsymbol{Z}^{(k+1)} - \boldsymbol{G}_1^{(k+1)}) \quad \text{and} \quad \boldsymbol{H}_2^{(k+1)} = \boldsymbol{H}_2^{(k)} - u(\boldsymbol{Z}^{(k+1)} - \boldsymbol{G}_2^{(k+1)}).$$

5 Return $\boldsymbol{Z} = \boldsymbol{Z}^{(k+1)}$ if converged. Otherwise, increment k and repeat Steps 1-5.

Algorithm 2: The ADMM used to solve (S1.2)

Input: Initialize $k = 0$, and select u, $\boldsymbol{H}^{(k)}$, $\boldsymbol{Z}^{(k)}$, $\boldsymbol{G}^{(k)}$ such that $\boldsymbol{Z}^{(k)}$ is a solution of (S1.2) without constraints, $\|\boldsymbol{G}^{(k)}\|_\infty \leq \beta$.

1 Minimize $\mathcal{L}_u(\boldsymbol{Z}, \boldsymbol{G}^{(k)}; \boldsymbol{H}^{(k)})$ with respect to $\boldsymbol{Z}$:

$$\boldsymbol{Z}^{(k+1)} = \mathcal{SVT}_{(uL)^{-1}\lambda'}\{\boldsymbol{G}^{(k)} + 1/u\boldsymbol{H}^{(k)}\}.$$

2 Minimize $\mathcal{L}_u(\boldsymbol{Z}^{(k+1)}, \boldsymbol{G}; \boldsymbol{H}^{(k)})$ with respect to $\boldsymbol{G}$:

$$\boldsymbol{G}^{(k+1)} = \underset{\|\boldsymbol{G}\|_\infty \leq \beta}{\arg\min} \quad \left\| \boldsymbol{G} - \left\{ \boldsymbol{Z}_{\text{old}} + \frac{1}{L}\nabla_{\boldsymbol{Z}}\ell_{\boldsymbol{W}}\left(\widehat{\mu}\boldsymbol{J} + \boldsymbol{Z}_{\text{old}}\right) - \boldsymbol{H}^{(k)} + u\boldsymbol{Z}^{(k+1)} \right\} /(1+u) \right\|_F^2.$$

Let $\boldsymbol{B} = \{\boldsymbol{Z}_{\text{old}} + \frac{1}{L}\nabla_{\boldsymbol{Z}}\ell_{\boldsymbol{W}}(\widehat{\mu}\boldsymbol{J} + \boldsymbol{Z}_{\text{old}}) - \boldsymbol{H}^{(k)} + u\boldsymbol{Z}^{(k+1)}\}/(1+u)$ and simplifies to $\boldsymbol{G}^{(k+1)}(i,j) = \min\{\beta, \max\{-\beta, \boldsymbol{B}(i,j)\}$.

3 Update the dual variable $\boldsymbol{H}^{(k+1)}$ by

$$\boldsymbol{H}^{(k+1)} = \boldsymbol{H}^{(k)} - u(\boldsymbol{Z}^{(k+1)} - \boldsymbol{G}^{(k+1)}).$$

4 Return $\boldsymbol{Z} = \boldsymbol{Z}^{(k+1)}$ if converged. Otherwise, increment k and repeat Steps 1-4.

Algorithm 3: Estimation of target matrix $\widehat{\boldsymbol{A}}_\beta$.

Input: Covariate matrix $\boldsymbol{A}$, incomplete data matrix $\boldsymbol{Y}$, estimated probability matrices $\widehat{\boldsymbol{\Theta}}_\beta$ (or $\widehat{\boldsymbol{\Theta}}$), tuning parameter candidates $\tau^{(1)}, \ldots, \tau^{(k)}$, where k is the grid length used for the search of parameter τ and a k evaluation matrix $\boldsymbol{Q} = (Q_{ij})$ to be $\boldsymbol{Q} = 0$.

1 Randomly partition the observed entries of $\boldsymbol{Y}$ into 5 equal sized subsamples. These subsamples are used in turn as a test set. When subsample l is used as test data, the remaining 4 subsamples are used as training data. Denote the corresponding indicator matrix of test data by $\boldsymbol{W}_*^{(l)}$ and that of training data by $\boldsymbol{W}^{(l)}$.

2 For each $i = 1, \ldots, k_1$ and $l = 1, \ldots, 5$, calculate $\widehat{\boldsymbol{A}}_\beta^{(l),\tau^{(i)}}$ by plugging $\boldsymbol{W}^{(l)}$ and $\tau^{(i)}$ in (4.3).

3 For $i = 1, \ldots, k$, $Q_i = \sum_{l=1}^{5} \|\boldsymbol{W}_*^{(l)} \circ \boldsymbol{\Theta}_\beta^\ddagger \circ (\widehat{\boldsymbol{A}}_\beta^{(l),\tau^{(i)}} - \boldsymbol{Y})\|_F^2$.

4 Output the best parameters $\tau^{(j)}$ that minimize Q_i among the entries of $\boldsymbol{Q}$.

5 Calculate $\widehat{\boldsymbol{A}}_\beta^{\tau^{(j)}}$ by plugging $\boldsymbol{W}$ and $\tau^{(j)}$ in (4.3).

Table S1: Root mean squared errors, test errors, estimated ranks $r_{\widehat{\boldsymbol{A}}_\beta}$ and their standard deviations (in parentheses) under the low-rank, missing-observation mechanism, for three existing methods and 10 versions of the proposed methods, where Prop indicates the estimators are obtained by solving problem (4.3), while $\widehat{\boldsymbol{\Theta}}_\beta$, $\widehat{\boldsymbol{\Theta}}_{\mathsf{Win},\beta}$, $\widehat{\boldsymbol{\Theta}}_\alpha$, $\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\beta}$, $\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\mathsf{Win},\beta}$, and $\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\alpha}$ represent the probability estimators used in (4.3), as described in Section 6.1, and $t = 0.05$ or 0.1 denotes the winsorized proportion for which β is chosen.

(n_1, n_2) = (1000, 1000)	RMSE($\widehat{\boldsymbol{A}}_\beta, \boldsymbol{A}_\star$)	Test Error	$r_{\widehat{\boldsymbol{A}}_\beta}$
Prop_$\widehat{\boldsymbol{\Theta}}_{\mathsf{Win},\beta}$_0.05	1.3975 (0.0142)	0.2375 (0.0035)	114.67 (19.73)
Prop_$\widehat{\boldsymbol{\Theta}}_{\beta}$_0.05	1.3909 (0.0064)	0.2391 (0.0023)	90.04 (6.51)
Prop_$\widehat{\boldsymbol{\Theta}}_{\mathsf{Win},\beta}$_0.1	1.3878 (0.0078)	0.2354 (0.0023)	100.69 (16.20)
Prop_$\widehat{\boldsymbol{\Theta}}_{\beta}$_0.1	1.3852 (0.0062)	0.2375 (0.0022)	81.79 (4.75)
Prop_$\widehat{\boldsymbol{\Theta}}_{\alpha}$	1.4024 (0.0242)	0.2389 (0.0062)	115.40 (22.21)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\mathsf{Win},\beta}$_0.05	1.4068 (0.0062)	0.2430 (0.0022)	98.97 (2.55)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\beta}$_0.05	1.3920 (0.0072)	0.2383 (0.0027)	97.88 (6.06)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\mathsf{Win},\beta}$_0.1	1.4121 (0.0062)	0.2449 (0.0022)	105.50 (1.16)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\beta}$_0.1	1.3913 (0.0064)	0.2383 (0.0023)	100.94 (7.12)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\alpha}$	1.3894 (0.0084)	0.2353 (0.0029)	113.92 (11.35)
NW	1.8519 (0.3534)	0.4081 (0.1405)	246.64 (82.34)
KLT	2.3207 (0.0053)	0.5964 (0.0016)	1.00 (0.00)
MHT	1.5083 (0.0084)	0.2857 (0.0033)	77.47 (5.31)
(n_1, n_2) = (1200, 1200)	RMSE($\widehat{\boldsymbol{A}}_\beta, \boldsymbol{A}_\star$)	Test Error	$r_{\widehat{\boldsymbol{A}}_\beta}$
Prop_$\widehat{\boldsymbol{\Theta}}_{\mathsf{Win},\beta}$_0.05	1.3389 (0.0168)	0.2171 (0.0040)	135.84 (25.41)
Prop_$\widehat{\boldsymbol{\Theta}}_{\beta}$_0.05	1.3226 (0.0057)	0.2157 (0.0020)	106.13 (5.81)
Prop_$\widehat{\boldsymbol{\Theta}}_{\mathsf{Win},\beta}$_0.1	1.3270 (0.0073)	0.2148 (0.0019)	112.28 (19.72)
Prop_$\widehat{\boldsymbol{\Theta}}_{\beta}$_0.1	1.3144 (0.0054)	0.2135 (0.0018)	97.71 (5.49)
Prop_$\widehat{\boldsymbol{\Theta}}_{\alpha}$	1.3453 (0.0287)	0.2187 (0.0071)	138.51 (29.08)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\mathsf{Win},\beta}$_0.05	1.3415 (0.0054)	0.2202 (0.0019)	115.63 (1.37)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\beta}$_0.05	1.3237 (0.0066)	0.2146 (0.0025)	115.07 (8.29)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\mathsf{Win},\beta}$_0.1	1.3489 (0.0054)	0.2226 (0.0019)	125.48 (1.04)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\beta}$_0.1	1.3259 (0.0058)	0.2157 (0.0019)	119.25 (10.60)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\alpha}$	1.3289 (0.0103)	0.2141 (0.0025)	137.05 (17.28)
NW	1.5528 (0.3693)	0.3016 (0.1382)	214.89 (100.18)
KLT	2.3494 (0.0044)	0.6041 (0.0013)	1.00 (0.00)
MHT	1.4649 (0.0062)	0.2706 (0.0024)	84.03 (4.49)

[2] With $r_{\boldsymbol{M}_\star} = 11$, $r_{\boldsymbol{A}_\star} = 11$, (n_1, n_2) = (1000, 1000), (1200, 1200), and SNR = 1. The three existing methods are proposed, respectively, in Negahban and Wainwright (2012)(NW), Koltchinskii et al. (2011)(KLT), and Mazumder et al. (2010)(MHT)

Table S2: Root mean squared errors, test errors, estimated ranks $r_{\widehat{\boldsymbol{A}}_\beta}$, and their standard deviations (in parentheses) under the low-rank, missing-observation mechanism, for three existing methods and 10 versions of the proposed methods, where Prop indicates the estimators are obtained by solving problem (4.3), while $\widehat{\boldsymbol{\Theta}}_\beta$, $\widehat{\boldsymbol{\Theta}}_{\mathsf{Win},\beta}$, $\widehat{\boldsymbol{\Theta}}_\alpha$, $\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\beta}$, $\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\mathsf{Win},\beta}$, and $\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\alpha}$ represent the probability estimators used in (4.3), as described in Section 6.1, and $t = 0.05$ or 0.1 denotes the winsorized proportion for which β is chosen.

(n_1, n_2) = (600, 600)	RMSE$(\widehat{\boldsymbol{A}}_\beta, \boldsymbol{A}_\star)$	Test Error	$r_{\widehat{\boldsymbol{A}}_\beta}$
Prop_$\widehat{\boldsymbol{\Theta}}_{\mathsf{Win},\beta}$_0.05	1.8460 (0.1167)	0.4080 (0.0470)	25.88 (23.71)
Prop_$\widehat{\boldsymbol{\Theta}}_{\beta}$_0.05	1.7310 (0.0321)	0.3625 (0.0135)	43.55 (5.29)
Prop_$\widehat{\boldsymbol{\Theta}}_{\mathsf{Win},\beta}$_0.1	1.7601 (0.0760)	0.3745 (0.0308)	44.63 (16.16)
Prop_$\widehat{\boldsymbol{\Theta}}_{\beta}$_0.1	1.7275 (0.0217)	0.3611 (0.0091)	42.85 (3.96)
Prop_$\widehat{\boldsymbol{\Theta}}_{\alpha}$	1.8992 (0.1061)	0.4289 (0.0429)	14.75 (21.17)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\mathsf{Win},\beta}$_0.05	1.8113 (0.1117)	0.3932 (0.0443)	31.20 (20.73)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\beta}$_0.05	1.7846 (0.1045)	0.3827 (0.0417)	34.66 (18.38)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\mathsf{Win},\beta}$_0.1	1.8245 (0.1129)	0.3984 (0.0444)	28.58 (22.20)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\beta}$_0.1	1.7888 (0.1118)	0.3845 (0.0449)	33.49 (18.47)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\alpha}$	1.8358 (0.1133)	0.4024 (0.0446)	25.45 (22.45)
NW	2.0240 (0.2553)	0.4941 (0.1225)	149.33 (53.68)
KLT	2.3087 (0.0079)	0.5997 (0.0025)	1.00 (0.00)
MHT	1.8147 (0.0087)	0.4033 (0.0038)	43.61 (2.60)
(n_1, n_2) = (800, 800)	RMSE$(\widehat{\boldsymbol{A}}_\beta, \boldsymbol{A}_\star)$	Test Error	$r_{\widehat{\boldsymbol{A}}_\beta}$
Prop_$\widehat{\boldsymbol{\Theta}}_{\mathsf{Win},\beta}$_0.05	1.6632 (0.0151)	0.3355 (0.0061)	76.87 (7.80)
Prop_$\widehat{\boldsymbol{\Theta}}_{\beta}$_0.05	1.6700 (0.0066)	0.3389 (0.0026)	63.18 (4.26)
Prop_$\widehat{\boldsymbol{\Theta}}_{\mathsf{Win},\beta}$_0.1	1.6647 (0.0128)	0.3362 (0.0054)	71.94 (3.7)
Prop_$\widehat{\boldsymbol{\Theta}}_{\beta}$_0.1	1.6673 (0.0067)	0.3380 (0.0026)	61.29 (3.79)
Prop_$\widehat{\boldsymbol{\Theta}}_{\alpha}$	1.6657 (0.0194)	0.3362 (0.0076)	77.60 (9.58)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\mathsf{Win},\beta}$_0.05	1.6816 (0.0080)	0.3427 (0.0028)	69.62 (9.42)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\beta}$_0.05	1.6740 (0.0070)	0.3396 (0.0027)	66.47 (6.59)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\mathsf{Win},\beta}$_0.1	1.6800 (0.0064)	0.3431 (0.0027)	74.12 (3.95)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\beta}$_0.1	1.6701 (0.0070)	0.3389 (0.0026)	70.28 (6.25)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\alpha}$	1.6722 (0.0111)	0.3386 (0.0044)	70.46 (5.63)
NW	2.1071 (0.3128)	0.5377 (0.1498)	222.81 (84.29)
KLT	2.3572 (0.0067)	0.6102 (0.0021)	1.00 (0.00)
MHT	1.7604 (0.0076)	0.3823 (0.0031)	61.07 (3.2)

[S1] With $r_{\boldsymbol{M}_\star} = 11$, $r_{\boldsymbol{A}_\star} = 31$, (n_1, n_2) = (600, 600), (800, 800), and SNR = 1. The three existing methods are proposed, respectively, in Negahban and Wainwright (2012)(NW), Koltchinskii et al. (2011)(KLT), and Mazumder et al. (2010)(MHT)

Table S3: Root mean squared errors, test errors, estimated ranks $r_{\widehat{\boldsymbol{A}}_\beta}$, and their standard deviations (in parentheses) under the low-rank, missing-observation mechanism, for three existing methods and 10 versions of the proposed methods, where Prop indicates the estimators are obtained by solving problem (4.3), while $\widehat{\boldsymbol{\Theta}}_\beta$, $\widehat{\boldsymbol{\Theta}}_{\mathsf{Win},\beta}$, $\widehat{\boldsymbol{\Theta}}_\alpha$, $\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\beta}$, $\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\mathsf{Win},\beta}$, and $\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\alpha}$ represent the probability estimators used in (4.3), as described in Section 6.1, and $t = 0.05$ or 0.1 denotes the winsorized proportion for which β is chosen.

(n_1, n_2) = (1000, 1000)	RMSE($\widehat{\boldsymbol{A}}_\beta, \boldsymbol{A}_\star$)	Test Error	$r_{\widehat{\boldsymbol{A}}_\beta}$
Prop_$\widehat{\boldsymbol{\Theta}}_{\mathsf{Win},\beta}$_0.05	1.6123 (0.0096)	0.3134 (0.0037)	108.31 (18.72)
Prop_$\widehat{\boldsymbol{\Theta}}_{\beta}$_0.05	1.6155 (0.0054)	0.3163 (0.0020)	86.32 (5.28)
Prop_$\widehat{\boldsymbol{\Theta}}_{\mathsf{Win},\beta}$_0.1	1.6045 (0.0069)	0.3104 (0.0026)	94.81 (7.74)
Prop_$\widehat{\boldsymbol{\Theta}}_{\beta}$_0.1	1.6140 (0.0056)	0.3154 (0.0020)	79.20 (5.26)
Prop_$\widehat{\boldsymbol{\Theta}}_{\alpha}$	1.6167 (0.0156)	0.3149 (0.0057)	109.57 (20.21)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\mathsf{Win},\beta}$_0.05	1.6285 (0.0050)	0.3205 (0.0020)	94.35 (2.47)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\beta}$_0.05	1.6173 (0.0063)	0.3167 (0.0024)	96.33 (6.07)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\mathsf{Win},\beta}$_0.1	1.6362 (0.0052)	0.3231 (0.0020)	87.25 (6.15)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\beta}$_0.1	1.6210 (0.0055)	0.3175 (0.0020)	84.35 (4.62)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\alpha}$	1.6166 (0.0111)	0.3150 (0.0038)	96.14 (14.22)
NW	1.9631 (0.2975)	0.4638 (0.1336)	236.97 (85.86)
KLT	2.3312 (0.0051)	0.5984 (0.0016)	1.00 (0.00)
MHT	1.6924 (0.0077)	0.3512 (0.0030)	77.64 (4.73)
(n_1, n_2) = (1200, 1200)	RMSE($\widehat{\boldsymbol{A}}_\beta, \boldsymbol{A}_\star$)	Test Error	$r_{\widehat{\boldsymbol{A}}_\beta}$
Prop_$\widehat{\boldsymbol{\Theta}}_{\mathsf{Win},\beta}$_0.05	1.5805 (0.0094)	0.3018 (0.0035)	132.90 (24.35)
Prop_$\widehat{\boldsymbol{\Theta}}_{\beta}$_0.05	1.5756 (0.0051)	0.3020 (0.0019)	106.59 (6.45)
Prop_$\widehat{\boldsymbol{\Theta}}_{\mathsf{Win},\beta}$_0.1	1.5746 (0.0051)	0.2991 (0.0018)	110.47 (21.53)
Prop_$\widehat{\boldsymbol{\Theta}}_{\beta}$_0.1	1.5722 (0.0047)	0.3007 (0.0017)	99.41 (5.04)
Prop_$\widehat{\boldsymbol{\Theta}}_{\alpha}$	1.5860 (0.0192)	0.3036 (0.0067)	134.72 (27.92)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\mathsf{Win},\beta}$_0.05	1.5937 (0.0041)	0.3073 (0.0017)	108.50 (2.12)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\beta}$_0.05	1.5785 (0.0070)	0.3020 (0.0026)	111.98 (7.35)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\mathsf{Win},\beta}$_0.1	1.5952 (0.0041)	0.3087 (0.0017)	116.63 (1.24)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\beta}$_0.1	1.5774 (0.0048)	0.3026 (0.0018)	120.85 (8.72)
Prop_$\widehat{\boldsymbol{\Theta}}_{\text{1-bit},\alpha}$	1.5727 (0.0060)	0.2993 (0.0022)	132.53 (16.82)
NW	1.6896 (0.2417)	0.3538 (0.1054)	193.73 (80.51)
KLT	2.3525 (0.0043)	0.6060 (0.0014)	1.00 (0.00)
MHT	1.6638 (0.0046)	0.3416 (0.0019)	93.87 (2.34)

[S2] With $r_{\boldsymbol{M}_\star} = 11$, $r_{\boldsymbol{A}_\star} = 31$, (n_1, n_2) = (1000, 1000), (1200, 1200), and SNR = 1. The three existing methods are proposed, respectively, in Negahban and Wainwright (2012)(NW), Koltchinskii et al. (2011)(KLT), and Mazumder et al. (2010)(MHT)

Table S4: Root mean squared prediction errors based on coat data set for the 10 versions of the proposed method, $\mathsf{Prop\_}\widehat{\boldsymbol{\Theta}}_{\mathsf{prop}}$ and the three existing methods proposed, respectively, in Negahban and Wainwright (2012)(NW), Koltchinskii et al. (2011)(KLT), and Mazumder et al. (2010)(MHT).

	$\mathsf{Prop\_}\widehat{\boldsymbol{\Theta}}_{\mathsf{Win},\beta}\mathsf{\_0.05}$	$\mathsf{Prop\_}\widehat{\boldsymbol{\Theta}}_{\beta}\mathsf{\_0.05}$	$\mathsf{Prop\_}\widehat{\boldsymbol{\Theta}}_{\mathsf{Win},\beta}\mathsf{\_0.1}$
RMSPE	1.0241	0.9592	1.0210
	$\mathsf{Prop\_}\widehat{\boldsymbol{\Theta}}_{\beta}\mathsf{\_0.1}$	$\mathsf{Prop\_}\widehat{\boldsymbol{\Theta}}_{\alpha}$	$\mathsf{Prop\_}\widehat{\boldsymbol{\Theta}}_{\mathsf{1\text{-}bit,Win},\beta}\mathsf{\_0.05}$
RMSPE	0.9376	1.0190	1.0172
	$\mathsf{Prop\_}\widehat{\boldsymbol{\Theta}}_{\mathsf{1\text{-}bit},\beta}\mathsf{\_0.05}$	$\mathsf{Prop\_}\widehat{\boldsymbol{\Theta}}_{\mathsf{1\text{-}bit,Win},\beta}\mathsf{\_0.1}$	$\mathsf{Prop\_}\widehat{\boldsymbol{\Theta}}_{\mathsf{1\text{-}bit},\beta}\mathsf{\_0.1}$
RMSPE	1.0987	1.0171	1.0878
	$\mathsf{Prop\_}\widehat{\boldsymbol{\Theta}}_{\mathsf{1\text{-}bit},\alpha}$	$\mathsf{Prop\_}\widehat{\boldsymbol{\Theta}}_{\mathsf{prop}}$	NW
RMSPE	1.0162	1.0206	1.0329
	KLT	MHT	
RMSPE	2.1758	1.5436	