Supplementary Material for "Median Matrix Completion: from Embarrassment to Optimality"

Weidong Liu [1] **Xiaojun Mao** [2] **Raymond K. W. Wong** [3]

A. Proofs

Proof of Theorem 1. As for the (i) in Theorem 1, we obtain the upper bound directly from Theorem 4.6 of Alquier et al. (2019).

As for (ii), by putting these $n_1n_2/(m_1m_2)$ estimators $\widehat{\mathbf{A}}_{\mathrm{QMC},l}$ together, we focus on both the first and second term of the right hand side of the upper bound (3.1) respectively. It is easy to verify that the upper bound in the right hand side hold.

In terms of the probability, we can conclude that

$$\sum_{l=1}^{l_1l_2} C_l \exp(-C_l s_l m_{\max}\log(m_+)) \leq \max\{C_l\}\exp(\log(n_1n_2) - \min\{C_l\} m_{\max}\log(m_+)).$$

□

Proposition A.1. *Suppose that Conditions (C1)-(C5) hold. Let $h \geq c\log(n_+)/N$ for some $c > 0$ and $h = O((n_1n_2)^{-1/2}a_N)$. We have*

$$\left|\widehat{f}(0) - f(0)\right| = O_{\mathsf{P}}\left(\sqrt{\frac{\log(n_+)}{Nh}} + \frac{a_N}{\sqrt{n_1n_2}}\right).$$

Proof of Proposition A.1. Let

$$D_{N,h}(\mathbf{A}) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{Y_i - \mathrm{tr}(\mathbf{X}_i^{\mathrm{T}}\mathbf{A})}{h}\right).$$

To prove the proposition, without loss of generality, we can assume that $\|\mathbf{A} - \mathbf{A}_\star\|_F \leq a_N$. It follows that $\widehat{f}(0) = D_{N,h}(\mathbf{A})$ and

$$\left|\widehat{f}(0) - f(0)\right| \leq \sup_{\|\mathbf{A}-\mathbf{A}_\star\|_F \leq a_N} \left|D_{N,h}(\mathbf{A}) - f(0)\right|.$$

We denote $\mathbf{A}_\star = (A_{\star,11}, \ldots, A_{\star,n_1n_2})$. For every s and t, we divide the interval $[A_{\star,st} - a_N, A_{\star,st} + a_N]$ into $(n_1n_2)^M$ small sub-intervals and each has length $2a_N/(n_1n_2)^M$, where M is a large positive number. Therefore, there exists a set of matrices in $\mathbb{R}^{n_1\times n_2}$, $\{\mathbf{A}_{(k)}, 1 \leq k \leq s_N\}$ with $s_N \leq (n_1n_2)^{M(n_1n_2)}$ and $\|\mathbf{A}_{(k)} - \mathbf{A}_\star\|_F \leq a_N$, such that for any $\mathbf{A}$ in the ball $\{\mathcal{A} : \mathbf{A} \in \mathbb{R}^{n_1\times n_2}, \|\mathbf{A} - \mathbf{A}_\star\|_F \leq a_N\}$, we have $\|\mathbf{A} - \mathbf{A}_{(k)}\|_F \leq 2\sqrt{n_1n_2}a_N/(n_1n_2)^M$ for some $1 \leq k \leq s_N$. Therefore

$$\left|\frac{1}{h}K\left(\frac{Y_i - \mathrm{tr}(\mathbf{X}_i^{\mathrm{T}}\mathbf{A})}{h}\right) - \frac{1}{h}K\left(\frac{Y_i - \mathrm{tr}(\mathbf{X}_i^{\mathrm{T}}\mathbf{A}_{(k)})}{h}\right)\right| \leq Ch^{-2}\left|\mathrm{tr}\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A} - \mathbf{A}_{(k)}\right)\}\right|.$$

This yields that

$$\sup_{\|\mathbf{A}-\mathbf{A}_\star\|_F \leq a_N} \left|D_{N,h}(\mathbf{A}) - f(0)\right| - \sup_{1\leq k\leq s_N}\left|D_{N,h}\left(\mathbf{A}_{(k)}\right) - f(0)\right| \leq \frac{CN\sqrt{n_1n_2}a_N}{(n_1n_2)^{M+1}h^2}.$$

By letting M large enough, we have

$$\sup_{|\mathbf{A}-\mathbf{A}_\star|_2 \leq a_N} \left|D_{N,h}(\mathbf{A}) - f(0)\right| - \sup_{1\leq k\leq s_N}\left|D_{N,h}\left(\mathbf{A}_{(k)}\right) - f(0)\right| = O_{\mathsf{P}}\left(n_+^{-\gamma}\right).$$

It is enough to show that $\sup_k |D_{N,h}(\mathbf{A}_{(k)}) - \mathbb{E}D_{N,h}(\mathbf{A}_{(k)})|$ and $\sup_k |\mathbb{E}D_{N,h}(\mathbf{A}_{(k)}) - f(0)|$ satisfy the bound in the lemma. Let $\mathbb{E}_*(\cdot)$ denote the conditional expectation given $\{\mathbf{X}_k\}$. We have

$$\mathbb{E}_*\left\{\frac{1}{h}K\left(\frac{\epsilon_i - \mathrm{tr}\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A} - \mathbf{A}_\star\right)\}}{h}\right)\right\} = \int_{-\infty}^{\infty} K(x)\, f\left\{hx + \mathrm{tr}\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A} - \mathbf{A}_\star\right)\}\right\}dx$$
$$= f(0) + O\left(h + \left|\mathrm{tr}\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A} - \mathbf{A}_\star\right)\}\right|\right).$$

*Equal contribution [1]School of Mathematical Sciences and MoE Key Lab of Artificial Intelligence Shanghai Jiao Tong University, Shanghai, 200240, China [2]School of Data Science, Fudan University, Shanghai, 200433, China [3]Department of Statistics, Texas A&M University, College Station, TX 77843, U.S.A.. Correspondence to: Xiaojun Mao <maoxj@fudan.edu.cn>.

Proceedings of the 37th International Conference on Machine Learning, Online, PMLR 119, 2020.

Under Condition (C1), with the fact that $\mathbb{E}|\mathrm{tr}\{\mathbf{X}_i^{\mathrm{T}}(\mathbf{A}-\mathbf{A}_\star)\}| \le (n_1n_2)^{-1}a_N$ and $\mathrm{Var}|\mathrm{tr}\{\mathbf{X}_i^{\mathrm{T}}(\mathbf{A}-\mathbf{A}_\star)\}| \le (n_1n_2)^{-1}a_N^2$, we have

$$\left|\mathbb{E}D_{N,h}\left(\mathbf{A}_{(k)}\right)-f\left(0\right)\right| \le C\left(h+(n_1n_2)^{-1/2}\left\|\mathbf{A}_{(k)}-\mathbf{A}_\star\right\|_F\right) = O(h+(n_1n_2)^{-1/2}a_N).$$

It remains to bound $\sup_k|D_{N,h}(\mathbf{A}_{(k)})-\mathbb{E}D_{N,h}(\mathbf{A}_{(k)})|$. Put

$$\xi_{i,k}=K\left(\frac{\epsilon_i-\mathrm{tr}\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A}_{(k)}-\mathbf{A}_\star\right)\}}{h}\right).$$

We have

$$\mathbb{E}_*\xi_{i,k}^2 = h\int_{-\infty}^{\infty}\{K\left(x\right)\}^2 f\left\{hx+\mathrm{tr}(\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A}_{(k)}-\mathbf{A}_\star\right))\right\}dx \le Ch.$$

Since $K(x)$ is bounded, we have by the exponential inequality (Lemma 1 in (Cai & Liu, 2011)) and the fact that $\log(n_+)=O(Nh)$, we have for any $\gamma>0$, there exists a constant $C>0$ such that

$$\sup_k\mathbb{P}\left(\left|\sum_{i=1}^N(\xi_{i,k}-\mathbb{E}\xi_{i,k})\right|\ge C\sqrt{Nh\log(n_+)}\right) = O\left(n_+^{-\gamma}\right).$$

By letting $\gamma>M$, we can obtain that

$$\sup_k\left|D_{N,h}(\mathbf{A}_{(k)})-\mathbb{E}D_{N,h}(\mathbf{A}_{(k)})\right| = O_{\mathsf{P}}\left(\sqrt{\frac{\log(n_+)}{Nh}}\right).$$

This completes the proof. □

Lemma A.1. *We have for any $\gamma>0$, $|\mathbf{u}|_2=1$ and $|\mathbf{v}|_2=1$, there exists a constant $C>0$ such that*

$$\Pr\left(\frac{1}{N}\sum_{i=1}^N\left(|\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}|-\mathbb{E}|\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}|\right)\ge C\sqrt{\frac{\log(n_+)}{n_{\min}N}}\right) = O(n_+^{-\gamma}).$$

Proof of Lemma A.1. On one hand, we have $\mathbb{E}|\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}|=O(n_{\min}^{-1})$. On the other hand, to apply Lemma 1 in Cai & Liu (2011), we only need to find B_N so that $\sum_i^N\mathbb{E}(|\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}|^2\exp\eta|\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}|)\le B_N^2$. For each $i=1,\dots,N$, we have

$$\begin{aligned}&\mathbb{E}(|\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}|^2\exp\left(\eta|\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}|\right))\\ &\le\frac{\bar{c}}{n_1n_2}\sum_{s=1}^{n_1}\sum_{t=1}^{n_2}u_s^2v_t^2\exp\left(\eta|u_sv_t|\right)\\ &\le\frac{\bar{c}}{n_1n_2}\sum_{s=1}^{n_1}\sum_{t=1}^{n_2}u_s^2v_t^2\exp\left(\eta u_s^2\right)\exp\left(\eta v_t^2\right)\\ &\le\frac{C(n_1+n_2)}{n_1n_2}=\frac{C}{n_{\min}}.\end{aligned}$$

Take $x^2=\gamma\log(n_+)$ and $B_N^2=C\gamma^{-1}Nn_{\min}^{-1}$ in Lemma 1 of Cai & Liu (2011), we can get the conclusion. □

Denote $\mathbf{B}_N(\mathbf{A})\in\mathbb{R}^{n_1\times n_2}$ where

$$\begin{aligned}B_N(\mathbf{A})=&\frac{1}{N}\sum_{i=1}^N\left[\mathbf{X}_i\mathbb{I}\left[\epsilon_i\le\mathrm{tr}\left\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A}-\mathbf{A}_\star\right)\right\}\right]\right.\\ &\left.-\mathbf{X}_if\left(\mathrm{tr}\left\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A}-\mathbf{A}_\star\right)\right\}\right)\right]\\ &-\frac{1}{N}\sum_{i=1}^N\left[\mathbf{X}_i\mathbb{I}\left[\epsilon_i\le0\right]-\mathbf{X}_if\left(0\right)\right].\end{aligned}\quad\text{(A.1)}$$

Let $\mathbf{\Theta}=\{\mathbf{A}:\|\mathbf{A}-\mathbf{A}_\star\|_F\le c\}$ for some $c>0$.

Lemma A.2. *We have for any $\gamma>0$, there exists a constant $C>0$ such that*

$$\sup_{|\mathbf{v}|_2=1}\sup_{|\mathbf{u}|_2=1}\Pr\left(\sup_{\mathbf{A}\in\mathbf{\Theta}}\frac{\sqrt{n_1n_2}|\mathbf{v}^{\mathrm{T}}B_N(\mathbf{A})\mathbf{u}|}{\sqrt{\|\mathbf{A}-\mathbf{A}_\star\|_F+n_{\max}\log(n_+)/N}}\ge C\sqrt{\frac{\log(n_+)}{n_{\min}N}}\right)=O(n_+^{-\gamma}).$$

Proof of Lemma A.2. We define $\mathbb{R}^{n_1\times n_2}$, $\{\mathbf{A}_{(k)},1\le k\le s_N\}$ as in the proof of Proposition A.1 with by replacing a_N with c. Then for any $\mathbf{A}\in\mathbf{\Theta}$, there exists $\mathbf{A}_{(k)}$ with $\|\mathbf{A}-\mathbf{A}_{(k)}\|_F\le2c\sqrt{n_1n_2}/(n_1n_2)^M$ and we have

$$\begin{aligned}&\left|\frac{\sqrt{n_1n_2}|\mathbf{v}^{\mathrm{T}}B_N(\mathbf{A})\mathbf{u}|}{\sqrt{\|\mathbf{A}-\mathbf{A}_\star\|_F+n_{\max}\log(n_+)/N}}-\frac{\sqrt{n_1n_2}|\mathbf{v}^{\mathrm{T}}B_N(\mathbf{A}_{(k)})\mathbf{u}|}{\sqrt{\|\mathbf{A}_{(k)}-\mathbf{A}_\star\|_F+n_{\max}\log(n_+)/N}}\right|\\ &\le\left|\frac{\sqrt{n_1n_2}|\mathbf{v}^{\mathrm{T}}B_N(\mathbf{A}_{(k)})\mathbf{u}|}{\sqrt{\|\mathbf{A}-\mathbf{A}_\star\|_F+n_{\max}\log(n_+)/N}}-\frac{\sqrt{n_1n_2}|\mathbf{v}^{\mathrm{T}}B_N(\mathbf{A}_{(k)})\mathbf{u}|}{\sqrt{\|\mathbf{A}_{(k)}-\mathbf{A}_\star\|_F+n_{\max}\log(n_+)/N}}\right|\\ &\quad+\frac{\sqrt{n_1n_2}|\mathbf{v}^{\mathrm{T}}B_N(\mathbf{A})\mathbf{u}-\mathbf{v}^{\mathrm{T}}B_N(\mathbf{A}_{(k)})\mathbf{u}|}{\sqrt{\|\mathbf{A}-\mathbf{A}_\star\|_F+n_{\max}\log(n_+)/N}}\\ &=:I_1+I_2.\end{aligned}$$

It is easy to see that

$$|I_1| \le C\frac{\sum_{i=1}^N |\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}|\mathrm{tr}\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A}_{(k)}-\mathbf{A}_\star\right)\}}{N} \times \frac{\sqrt{n_1n_2}\times c\sqrt{n_1n_2}}{(n_1n_2)^M(c+n_{\max}\log(n_+)/N)^{3/2}} =: I_3.$$

With Lemma A.1, we can show that

$$\Pr\left(I_3 \ge C\sqrt{\frac{\log(n_+)}{n_{\min}N}}\right) = O(n_+^{-\gamma}),$$

for any $\gamma > 0$ by letting M be sufficiently large. For I_2, noting that

$$\left|f\left(\mathrm{tr}\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A}-\mathbf{A}_\star\right)\}\right) - f\left(\mathrm{tr}\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A}_{(k)}-\mathbf{A}_\star\right)\}\right)\right| \le C\sqrt{n_1n_2}/(n_1n_2)^M,$$

we have

$$\begin{aligned}&|I_2| \le \sqrt{n_1n_2}\Big(\frac{cn_{\max}\log(n_+)}{N}\Big)^{-1/4}\frac{1}{N}\sum_{i=1}^N|\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}|\times\\ &\mathbb{I}\left[|\epsilon_i - \mathrm{tr}\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A}_{(k)}-\mathbf{A}_\star\right)\}| \le 2c\sqrt{n_1n_2}/(n_1n_2)^M\right]\\ &+C\frac{cn_1n_2}{(n_1n_2)^M}\Big(\frac{cn_{\max}\log(n_+)}{N}\Big)^{-1/4}\times\\ &\frac{1}{N}\sum_{i=1}^N|\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}|\\ &=: I_4 + I_5.\end{aligned}$$

It is easy to show that $\mathbb{E}(I_4) = o(\sqrt{\log(n_+)/(n_{\min}N)})$ with M large enough and

$$\begin{aligned}&\Pr\left(I_5 \ge C\sqrt{\frac{\log(n_+)}{n_{\min}N}}\right) \le\\ &\sum_{i=1}^N\Pr\left(|\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}| \ge \frac{(n_1n_2)^{M-2}N^{1/4}}{n_{\max}\log(n_+)}\right)\\ &= O(n_+^{-\gamma}),\end{aligned}$$

for any $\gamma > 0$ by letting M be sufficiently large. Also for some $\eta > 0$,

$$\begin{aligned}&\mathbb{E}(|\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}|^2\exp(\eta|\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}|)\times\\ &\mathbb{I}\left[|\epsilon_i - \mathrm{tr}\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A}_{(k)}-\mathbf{A}_\star\right)\}| \le 2c\sqrt{n_1n_2}/(n_1n_2)^M\right])\\ &\le C\sqrt{n_1n_2}(n_1n_2)^{-M}\mathbb{E}|\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}|^2\exp(\eta|\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}|)\\ &= O(1/((n_1n_2)^{M-1/2}n_{\min})).\end{aligned}$$

Now by the exponential inequality in (Cai & Liu, 2011) (taking $x = \sqrt{\gamma\log(n_+)}$, $B_n = \sqrt{\gamma^{-1}N\log(n_+)/n_{\min}}$ and noting that $1/((n_1n_2)^{M-1/2}n_{\min}) = o(B_N^2)$), we have for large $C > 0$,

$$\begin{aligned}&\Pr\left(|I_4 - \mathbb{E}(I_4)| \ge C\sqrt{\frac{\log(n_+)}{n_{\min}N}}\right)\\ &= O(n_+^{-\gamma}).\end{aligned}$$

As $s_N \le (n_1n_2)^{M(n_1n_2)}$, by choosing C sufficiently large such that $\gamma > M$, it is enough to show that for any $\gamma > 0$,

$$\begin{aligned}&\sup_{|\mathbf{v}|_2=1}\sup_{|\mathbf{u}|_2=1}\max_k \Pr\Big(\sqrt{n_1n_2}|\mathbf{v}^{\mathrm{T}}B_N(\mathbf{A}_{(k)})\mathbf{u}|\times\\ &\frac{1}{\sqrt{\|\mathbf{A}_{(k)}-\mathbf{A}_\star\|_F + n_{\max}\log(n_+)/N}} \ge C\sqrt{\frac{\log(n_+)}{n_{\min}N}}\Big)\\ &= O(n_+^{-\gamma}).\end{aligned} \tag{A.2}$$

Set

$$Z_i(\mathbf{A}) = \mathbb{I}\left[\epsilon_i \le \mathrm{tr}\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A}-\mathbf{A}_\star\right)\}\right] - f\left(\mathrm{tr}\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A}-\mathbf{A}_\star\right)\}\right).$$

Then we have

$$\begin{aligned}&\mathbb{E}(\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u})^2(Z_i(\mathbf{A}) - Z_i(\mathbf{A}_\star))^2\exp(\eta|\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}|)\\ &\le C(n_1n_2)^{-1}\|\mathbf{A}-\mathbf{A}_\star\|_F\times\\ &\sup_{|\mathbf{v}|_2=1,|\mathbf{u}|_2=1}\mathbb{E}(\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u})^2\exp(\eta|\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}|)\\ &\le C(n_1n_2)^{-1}\|\mathbf{A}-\mathbf{A}_\star\|_F n_{\min}^{-1}.\end{aligned}$$

Now letting $B_N^2 = C\gamma^{-1}(N\|\mathbf{A}_{(k)}-\mathbf{A}_\star\|_F/(n_1n_2) + N\log(n_+)/n_{\min})$ and $x^2 = \gamma\log(n_+)$ in Lemma 1 in (Cai & Liu, 2011), we can show (A.2) holds. □

Let

$$U_N = \sup_{\|\mathbf{A}-\mathbf{A}_\star\|_F \le a_N}\|\mathbf{B}_N(\mathbf{A})\|.$$

For a unit ball B in R^s, we have the fact that there exist q_s balls with centers $\mathbf{x}_1,\ldots,\mathbf{x}_{q_s}$ and radius z (i.e., $B_i = \{\mathbf{x} \in R^s : |\mathbf{x}-\mathbf{x}_i| \le z\}$, $1 \le i \le q_s$) such that $B \subseteq \cup_{i=1}^{q_s}B_i$ and q_s satisfies $q_s \le (1+2/z)^s$. Then by a standard $\mathcal{E}-$net argument, for any matrix $\mathbf{A} \in \mathbb{R}^{n_1\times n_2}$, there exist $\mathbf{v}_1,\ldots,\mathbf{v}_{b_1}$ and $\mathbf{u}_1,\ldots,\mathbf{u}_{b_2}$ (which do not depend on $\mathbf{A}$) with $|\mathbf{v}_i|_2 = 1$ and $|\mathbf{u}_i|_2 = 1$, $b_1 \le 9^{n_1}$ and $b_2 \le 9^{n_2}$ such that

$$\|\mathbf{A}\| \le 5\max_{1\le i\le b_1}\max_{1\le j\le b_2}|\mathbf{v}_i^{\mathrm{T}}\mathbf{A}\mathbf{u}_j|. \tag{A.3}$$

So we have $U_N \le 5\max_{1\le i\le b_1}\max_{1\le j\le b_2}|\mathbf{v}_i^{\mathrm{T}}B_N(\mathbf{A}_{(k)})\mathbf{u}_j|$. Assume the initial value $(n_1n_2)^{-1/2}\|\mathbf{A}_\star - \widehat{\mathbf{A}}_0\|_F = o_{\mathsf{P}}(1)$. By Lemma A.2, we have

$$U_N = O_{\mathsf{P}}\left(\sqrt{\frac{\left\|\widehat{\mathbf{A}}_0-\mathbf{A}_\star\right\|_F\log(n_+)}{n_1n_2n_{\min}N}} + \frac{\log(n_+)}{n_{\min}N}\right).$$

So we have the following lemma.

Lemma A.3. *Assume that Conditions (C1)-(C6) hold. We have*

$$U_N = O_P\Big(\sqrt{\frac{a_N\log(n_+)}{n_1n_2n_{\min}N}} + \frac{\log(n_+)}{n_{\min}N}\Big).$$

To obtain Theorem 2 which related to the repeated refinements, we consider the following one-step refinement result at first.

Theorem A.1 (One-step refinement)**.** *Suppose that Conditions (C1)–(C5) hold and* $\mathbf{A}_\star \in \mathcal{B}(a, n_1, n_2)$*. By choosing the bandwidth* $h \asymp (n_1 n_2)^{-1/2} a_N$ *and taking*

$$\lambda_N = C\left(\sqrt{\frac{\log(n_+)}{n_{\min} N}} + \frac{a_N^2}{n_{\min}(n_1 n_2)}\right),$$

where C *is a sufficient large constant, we have*

$$\frac{\left\|\widehat{\mathbf{A}}^{(1)} - \mathbf{A}_\star\right\|_F^2}{n_1 n_2} = O_P\left[\max\left\{\sqrt{\frac{\log(n_+)}{N}}, r_\star\left(\frac{n_{\max}\log(n_+)}{N} + \frac{a_N^4}{n_{\min}^2(n_1 n_2)}\right)\right\}\right]. \quad \text{(A.4)}$$

To obtain Theorems A.1 and 2, we require Lemmas A.4 and 1 respectively.

Lemma A.4. *Suppose that Conditions (C1)–(C5) hold and* $\mathbf{A}_\star \in \mathcal{B}(a, n_1, n_2)$*. By choosing the bandwidth* $h \asymp (n_1 n_2)^{-1/2} a_N$*, we have*

$$\left\|\frac{1}{N}\sum_{i=1}^N \xi_i^{(1)} \mathbf{X}_i\right\| = O_P\left(\sqrt{\frac{\log(n_+)}{n_{\min} N}} + \frac{a_N^2}{n_{\min}(n_1 n_2)}\right).$$

Lemma A.4 obtains the upper bound for the stochastic error term that appears in the first update iteration of the initial estimator $\widehat{\mathbf{A}}_0$ fulfill condition (C5). It is easy to verify that our initial estimator $\widehat{\mathbf{A}}_{\text{LADMC},0}$ proposed in section 2.2 satisfy condition (C5).

Proof of Lemma A.4. Denote $\mathbf{H}_N(\mathbf{A}) \in \mathbb{R}^{n_1 \times n_2}$ where

$$H_N(\mathbf{A}) = \frac{\widehat{f}^{-1}(0)}{N}\sum_{i=1}^N \mathbf{X}_i \left\{f\left[\text{tr}\left\{\mathbf{X}_i^{\mathrm{T}}(\mathbf{A} - \mathbf{A}_\star)\right\}\right] - f(0)\right\} + \frac{1}{N}\sum_{i=1}^N \mathbf{X}_i \text{tr}\left\{\mathbf{X}_i^{\mathrm{T}}(\mathbf{A} - \mathbf{A}_\star)\right\}.$$

We have

$$\begin{aligned} &\left\|\frac{1}{N}\sum_{i=1}^N \xi_i^{(1)} \mathbf{X}_i\right\| \le \\ &\left\|-\frac{\widehat{f}^{-1}(0)}{N}\sum_{i=1}^N \mathbf{X}_i\left(\mathbb{I}\left[Y_i \le \text{tr}\{\mathbf{X}_i^{\mathrm{T}}\widehat{\mathbf{A}}_0\}\right] - \tau\right) + \frac{1}{N}\sum_{i=1}^N \mathbf{X}_i \text{tr}\left\{\mathbf{X}_i^{\mathrm{T}}\left(\widehat{\mathbf{A}}_0 - \mathbf{A}_\star\right)\right\}\right\| \le \\ &\left\|\mathbf{H}_N(\widehat{\mathbf{A}}_0)\right\| + \left|\widehat{f}^{-1}(0)\right| \left\|\frac{1}{N}\sum_{i=1}^N \left[\mathbf{X}_i \mathbb{I}\left[\epsilon_i \le 0\right] - \mathbf{X}_i f(0)\right]\right\| + \left|\widehat{f}^{-1}(0)\right| U_N. \end{aligned}$$

By Proposition A.1 and $(n_1 n_2)^{1/2}\log(n_+) = o(N a_N)$, we have $\widehat{f}(0) \ge c$ for some $c > 0$ with probability tending to one. Therefore, for the last term, by Lemma A.3, we have

$$|\widehat{f}^{-1}(0)| U_N = O_{\mathsf{P}}\left(\sqrt{\frac{a_N \log(n_+)}{n_1 n_2 n_{\min} N}} + \frac{\log(n_+)}{n_{\min} N}\right).$$

For the second term of the right hand side, by (A.3) and the exponential inequality in (Cai & Liu, 2011), follow the same proof with Lemma A.1, we have

$$\left|\widehat{f}^{-1}(0)\right| \left\|\frac{1}{N}\sum_{i=1}^N \mathbf{X}_i\left[\mathbb{I}\left[\epsilon_i \le 0\right] - f(0)\right]\right\| = O_{\mathsf{P}}\left(\sqrt{\frac{\log(n_+)}{n_{\min} N}}\right).$$

By second order Taylor expansion, under condition (C1) we have,

$$\begin{aligned} &\frac{\widehat{f}^{-1}(0)}{N}\sum_{i=1}^N \mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}\left[f\left(\text{tr}\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A}_\star - \widehat{\mathbf{A}}_0\right)\}\right) - f(0)\right] \\ &= \frac{\widehat{f}^{-1}(0) f(0)}{N}\sum_{i=1}^N \mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}\,\text{tr}\left\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A}_\star - \widehat{\mathbf{A}}_0\right)\right\} \\ &+ O(1)\frac{\widehat{f}^{-1}(0)}{N}\sum_{i=1}^N |\mathbf{v}^{\mathrm{T}}\mathbf{X}_i\mathbf{u}|\left[\text{tr}\left\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A}_\star - \widehat{\mathbf{A}}_0\right)\right\}\right]^2. \end{aligned}$$

Let $\mathbf{v}_1, ..., \mathbf{v}_{b_1}$ and $\mathbf{u}_1, ..., \mathbf{u}_{b_2}$ be defined as in the argument above Lemma A.3. Together with Lemma A.1,we have

$$\begin{aligned}\left|\mathbf{v}_k^{\mathrm{T}}\mathbf{H}_N\left(\widehat{\mathbf{A}}_0\right)\mathbf{u}_j\right| &\le \left|\widehat{f}^{-1}(0)f(0)-1\right|\times \\ &\left|\frac{1}{N}\sum_{i=1}^{N}\mathbf{v}_k^{\mathrm{T}}\mathbf{X}_i\mathbf{u}_j\mathrm{tr}\left\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A}_\star-\widehat{\mathbf{A}}_0\right)\right\}\right| \\ +C\widehat{f}^{-1}(0)\frac{1}{N}\sum_{i=1}^{N}\left|\mathbf{v}_k^{\mathrm{T}}\mathbf{X}_i\mathbf{u}_j\right|&\left[\mathrm{tr}\left\{\mathbf{X}_i^{\mathrm{T}}\left(\mathbf{A}_\star-\widehat{\mathbf{A}}_0\right)\right\}\right]^2 \\ \le C\left(\sqrt{\frac{\log(n_+)}{Nh}}+\frac{a_N}{\sqrt{n_1n_2}}\right)&\frac{\left\|\mathbf{A}_\star-\widehat{\mathbf{A}}_0\right\|_F}{n_{\min}\sqrt{n_1n_2}} \\ &+C\frac{1}{n_{\min}(n_1n_2)}\left\|\mathbf{A}_\star-\widehat{\mathbf{A}}_0\right\|_F^2\end{aligned}$$

We can easily have

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}\xi_i^{(1)}\mathbf{X}_i\right\| = O_{\mathsf{P}}\Bigg(&\sqrt{\frac{\log(n_+)}{n_{\min}N}}+\sqrt{\frac{a_N\log(n_+)}{n_1n_2n_{\min}N}} \\ &+a_N\sqrt{\frac{\log(n_+)}{n_{\min}^2n_1n_2Nh}}+\frac{a_N^2}{n_{\min}(n_1n_2)}\Bigg).\end{aligned}$$

The lemma is proved. □

Define the observation operator $\Omega:\mathbb{R}^{n_1\times n_2}\to\mathbb{R}^N$ as $(\Omega(\mathbf{A}))_k=\langle\mathbf{X}_k,\mathbf{A}\rangle$.

Proof of Theorem A.1. Due to the basic inequality, we have

$$\frac{1}{N}\sum_{k=1}^{N}\left(\widetilde{Y}_k^{(1)}-\mathrm{tr}(\mathbf{X}_k^{\mathrm{T}}\widehat{\mathbf{A}})\right)^2+\lambda_N\left\|\widehat{\mathbf{A}}\right\|_*\le \frac{1}{N}\sum_{k=1}^{N}\left(\widetilde{Y}_k^{(1)}-\mathrm{tr}(\mathbf{X}_k^{\mathrm{T}}\mathbf{A}_\star)\right)^2+\lambda_N\left\|\mathbf{A}_\star\right\|_*,$$

which implies

$$\begin{aligned}&\frac{1}{N}\left\|\Omega\left(\mathbf{A}_\star-\widehat{\mathbf{A}}\right)\right\|_F^2+\lambda_N\left\|\widehat{\mathbf{A}}\right\|_* \\ &\le 2\left\langle\widehat{\mathbf{A}}-\mathbf{A}_\star,\mathbf{\Sigma}^{(1)}\right\rangle+\lambda_N\left\|\mathbf{A}_\star\right\|_* \\ &\le 2\left\|\mathbf{\Sigma}^{(1)}\right\|\left\|\widehat{\mathbf{A}}-\mathbf{A}_\star\right\|_*+\lambda_N\left\|\mathbf{A}_\star\right\|_*.\end{aligned}$$

Together with Lemma A.4 and follow the proof of Theorem 3 in Klopp (2014), it complete the proof. □

Proof of Lemma 1. Replacing the tuning parameter λ_N by $\lambda_{N,t}$, Lemma 1 follows directly from the proof of Lemma A.4. □

Proof of Theorem 2. Similar with the proof of Theorem A.1, together with the result in Lemma 1 we complete the proof. □

B. Experiments (Cont')

B.1. Synthetic Data (Cont')

In the following, we tested the proposed method DLADMC with the initial estimator synthetically generated by adding standard Gaussian noises ($\mathcal{N}$(0,1)) to the ground truth matrix $\mathbf{A}_\star$ and reported all the results in Table S1.

Table S1. The average RMSEs, MAEs, estimated ranks and their standard errors (in parentheses) of modified DLADMC over 500 simulations. The number in the first column within the parentheses represents T in Algorithm 1.

(T)	RMSE	MAE	rank
S1(4)	0.6364 (0.0238)	0.4826 (0.0232)	63.74 (5.37)
S2(5)	0.8985 (0.0407)	0.6738 (0.0404)	67.59 (6.76)
S3(5)	0.4460 (0.0080)	0.3179 (0.0067)	43.07 (6.00)
S4(4)	0.8522 (0.0203)	0.6229 (0.0210)	45.21 (5.52)

B.2. Real-World Data (Cont')

B.2.1. Effect of Iteration Number

To understand the effect of the iteration number, we ran 10 iterations and report all the details in Table S2. Briefly, the smallest and largest RMSEs among these iterations are (0.9226,0.9255), (0.9344,0.9381), (1.0486,1.0554) and (1.0512,1.0591) with respect to the 4 datasets in Section 4.2. Even with the worst RMSEs, we achieve a similar conclusion as shown in Section 4.2 of the paper.

Table S2. The RMSEs, MAEs and estimated ranks of DLADMC with different iteration number under dimensions $n_1 = 739$ and $n_2 = 918$.

	t	1	2	3	4	5
RawA	RMSE	0.9253	0.9253	0.9229	0.9252	0.9233
	MAE	0.7241	0.7267	0.7224	0.7264	0.7230
	rank	54	50	53	45	59
RawB	RMSE	0.9368	0.9381	0.9344	0.9373	0.9363
	MAE	0.7315	0.7344	0.7291	0.7340	0.7310
	rank	57	51	59	44	40
OutA	RMSE	1.0550	1.0543	1.0509	1.0549	1.0506
	MAE	0.8659	0.8648	0.8609	0.8673	0.8595
	rank	28	35	48	29	33
OutB	RMSE	1.0591	1.0569	1.0532	1.0583	1.0527
	MAE	0.8707	0.8679	0.8632	0.8713	0.8627
	rank	24	33	45	31	30
	t	6	7	8	9	10
RawA	RMSE	0.9253	0.9235	0.9250	0.9227	0.9255
	MAE	0.7265	0.7233	0.7264	0.7219	0.7268
	rank	41	41	45	55	44
RawB	RMSE	0.9362	0.9352	0.9369	0.9345	0.9370
	MAE	0.7328	0.7300	0.7333	0.7292	0.7339
	rank	49	51	46	58	44
OutA	RMSE	1.0544	1.0486	1.0553	1.0491	1.0554
	MAE	0.8671	0.8568	0.8695	0.8569	0.8697
	rank	31	38	35	40	33
OutB	RMSE	1.0572	1.0521	1.0577	1.0512	1.0582
	MAE	0.8699	0.8616	0.8706	0.8602	0.8716
	rank	30	28	31	30	33

B.2.2. MovieLens-1M

To further demonstrate the scalability of our proposed method, we tested various methods on a larger MovieLens-

1M[1] dataset. This data set consists of 1,000,209 movie ratings provided by 6040 viewers on approximate 3900 movies. The ratings also range from 1 to 5. To evaluate the performance of different methods, we keep one fifth of the data to be test set and remaining to be training set. We refer it to as Raw. Similar to Alquier et al. (2019), we added artificial outliers by randomly changing 20% of ratings that are equal to 5 in the train set to 1 and constructed Out. To avoid rows and columns that contain too few observations, we only keep the rows and columns with at least 40 ratings. The resulting target matrix $\mathbf{A}_\star$ is of dimension 4290×2505. For the proposed DLADMC, we fix the iteration number to 10. For the proposed BLADMC, to faster the speed, we split the data matrix so that the number of row subsets $l_1 = 4$ and number of column subsets $l_2 = 3$. To save times, the tunning parameters for all the methods were chosen by the one-fold validation. The RMSEs, MAEs, estimated ranks and the total computing time (in seconds) are reported in Table 2. For a fair comparison, we recorded the time of each method in the experiment with the selected tuning parameter.

Table S3. The RMSEs, MAEs and estimated ranks of DLADMC, BLADMC, ACL and MHT under dimensions $n_1 = 4290$ and $n_2 = 2505$.

		DLADMC	BLADMC	MHT
Raw	RMSE	0.8632	0.9733	0.8520
	MAE	0.6768	0.7865	0.6680
	rank	111	1911	156
	t	19593.58	1203.45	2113.55
Out	RMSE	0.9161	0.9733	0.9757
	MAE	0.7331	0.7865	0.8021
	rank	125	1913	45
	t	14290.16	1076.69	1053.58

As ACL is not scalable to large dimensions, we could not obtain the results of ACL within five times of the running time of the proposed DLADMC. It is noted that under the raw data Raw, the proposed DLADMC performed similarly as the least squares estimator MHT. BLADMC lost some efficiency due to the embarrassingly parallel computing. For the dataset with outliers, the proposed DLADMC performed better than MHT.

References

Alquier, P., Cottet, V., and Lecué, G. Estimation bounds and sharp oracle inequalities of regularized procedures with lipschitz loss functions. *The Annals of Statistics*, 47 (4):2117–2144, 2019.

Cai, T. T. and Liu, W. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106(494):672–684, 2011.

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

[1] https://grouplens.org/datasets/movielens/1m/