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 \( \hat{A}_{QMC,l} \) 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

\[
\sum_{l=1}^{l_1l_2} C_l \exp(-C_l s \eta) m \log(m_+) \leq \max \{C_l\} \exp\left(\log(n_1n_2) - \min \{C_l\} m \log(m_+)\right).
\]

\( \square \)

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

\[
\left| \hat{f}(0) - f(0) \right| = O_P \left( \sqrt{\frac{\log(n_+)}{Nh}} + \frac{a_N}{\sqrt{n_1n_2}} \right).
\]

Proposition A.1. Let

\[
D_{N,h}(A) = \frac{1}{Nh} \sum_{i=1}^{N} K \left( \frac{Y_i - \text{tr}(X_i^T A)}{h} \right).
\]

To prove the proposition, without loss of generality, we can assume that \( \|A - A_*\|_F \leq a_N \). It follows that \( \hat{f}(0) = \frac{1}{Nh} \sum_{i=1}^{N} K \left( \frac{Y_i - \text{tr}(X_i^T A)}{h} \right) \).

We denote \( A_* = (A_{*,1}, \ldots, A_{*,n_1n_2}) \). For every \( s \) and \( t \), we divide the interval \( [A_{s,t} - a_N, A_{s,t} + a_N] \) 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 \( \mathbb{R}^{n_1\times n_2} \), \( \{A_{(k)}, 1 \leq k \leq s_N\} \) with \( s_N \leq (n_1n_2)^{M(n_1n_2)} \) and \( \|A_{(k)} - A_*\|_F \leq a_N \), such that for any \( A \) in the ball \( \{A : A \in \mathbb{R}^{n_1\times n_2}, \|A - A_*\|_F \leq a_N\} \), we have \( \|A - A_{(k)}\|_F \leq 2\sqrt{\sqrt{n_1n_2} a_N/(n_1n_2)^{M}} \) for some \( 1 \leq k \leq s_N \). Therefore

\[
\left| \frac{1}{Nh} K \left( \frac{Y_i - \text{tr}(X_i^T A)}{h} \right) - \frac{1}{Nh} K \left( \frac{Y_i - \text{tr}(X_i^T A_{(k)})}{h} \right) \right| \leq Ch^{-2} \left| \text{tr}(X_i^T (A - A_{(k)})) \right|.
\]

This yields that

\[
\sup_{1 \leq k \leq s_N} |D_{N,h}(A_{(k)}) - f(0)| \leq C N \sqrt{n_1n_2} a_N \leq \frac{CN \sqrt{n_1n_2} a_N}{(n_1n_2)^{M+1/2}}.
\]

By letting \( M \) large enough, we have

\[
\left| D_{N,h}(A_{(k)}) - f(0) \right| \leq O_P \left( n_+^{-\gamma} \right).
\]

It is enough to show that \( \sup_{k} |D_{N,h}(A_{(k)}) - \mathbb{E} D_{N,h}(A_{(k)})| \) and \( \sup_{k} \mathbb{E} |D_{N,h}(A_{(k)}) - f(0)| \) satisfy the bound in the lemma. Let \( E_{\ast}(\cdot) \) denote the conditional expectation given \( \{X_{k}\} \). We have

\[
E_{\ast}\left\{ \frac{1}{h} K \left( \frac{e_i - \text{tr}(X_i^T (A - A_*))}{h} \right) \right\} = \int_{-\infty}^{\infty} K(x) f \left\{ hx + \text{tr}(X^T (A - A_*)) \right\} dx
\]

\[
= f(0) + O \left( h + \left| \text{tr}(X^T (A - A_*)) \right| \right).
\]
Under Condition (C1), with the fact that $E|\text{tr}\{X_i^T(A - A_*)\}| \leq (n_1n_2)^{-1}a_N$ and $\text{Var}[\text{tr}\{X_i^T(A - A_*)\}] \leq (n_1n_2)^{-1}a_N^2$, we have
\[|E D_{N,h}(A) - f(0)| \leq C \left( h + (n_1n_2)^{-1/2} \|A_k - A_*\|_F \right) = O(h + (n_1n_2)^{-1/2}a_N). \]

It remains to bound $\sup_k |D_{N,h}(A_k) - E D_{N,h}(A_k)|$. Put
\[\xi_{i,k} = K \left( \epsilon_{i} - \frac{\text{tr}\{X_i^T(A_k - A_*)\}}{h} \right). \]

We have
\[E\xi_{i,k}^2 = h \int_{-\infty}^{\infty} \{K(x)\}^2 f \{h x + \text{tr}(X_i^T(A_k - A_*)\}) dx \leq Ch. \]

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

By letting $\gamma > M$, we can obtain that
\[\sup_k |D_{N,h}(A_k) - E D_{N,h}(A_k)| = O\left( \sqrt{\log(n_+)} / Nh \right). \]

This completes the proof.

**Lemma A.1.** We have for any $\gamma > 0$, $|u|_2 = 1$ and $|v|_2 = 1$, there exists a constant $C > 0$ such that
\[P \left( \frac{1}{N} \sum_{i=1}^{N} (|v^T X_i u| - E|v^T X_i u|) \geq C \frac{\log(n_+)}{n_{\min}^2 N} \right) = O(n_+^{-\gamma}). \]

**Proof of Lemma A.1.** On one hand, we have $E|v^T X_i u| = O(n_+^{-1})$. On the other hand, to apply Lemma 1 in Cai & Liu (2011), we only need to find $B_N$ so that $\sum_i^N E(|v^T X_i u|^2 \exp \eta|v^T X_i u|) \leq B_N^2$. For each $i = 1, \ldots, N$, we have
\[E(|v^T X_i u|^2 \exp \eta|v^T X_i u|) \leq \frac{7}{n_1n_2} \sum_{s=1}^{n_1} \sum_{t=1}^{n_2} u_s v_t^2 \exp \eta|u_s v_t| \leq \frac{7}{n_1n_2} \sum_{s=1}^{n_1} \sum_{t=1}^{n_2} u_s v_t^2 \exp \eta v_t^2 \leq \frac{C(n_1 + n_2)}{n_1n_2} = \frac{C}{n_{\min}}. \]

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

Denote $B_N(A) \in \mathbb{R}^{n_1 \times n_2}$ where
\[B_N(A) = \frac{1}{N} \sum_{i=1}^{N} [X_i I\{\epsilon_i \leq \text{tr}\{X_i^T(A - A_*)\}\}] - X_i f (\text{tr}\{X_i^T(A - A_*)\}) - \frac{1}{N} \sum_{i=1}^{N} [X_i I\{\epsilon_i \leq 0\} - X_i f(0)]. \]

Let $\Theta = \{A : \|A - A_*\|_F \leq c\}$ for some $c > 0$.

**Lemma A.2.** We have for any $\gamma > 0$, there exists a constant $C > 0$ such that
\[\sup_{|u|_2 = 1} \sup_{|v|_2 = 1} \text{Pr} \left( \sup_{A \in \Theta} \frac{\sqrt{n_1n_2} |v^T B_N(A) u|}{\sqrt{n_1n_2} \|v^T B_N(A) u\|} \right) \geq C \frac{\log(n_+)}{n_{\min}^2 N} = O(n_+^{-\gamma}). \]

**Proof of Lemma A.2.** We define $\mathbb{R}^{n_1 \times n_2}$, $\{A_k\}, 1 \leq k \leq s_N$ as in the proof of Proposition A.1 with by replacing $a_N$ with $c$. Then for any $A \in \Theta$, there exists $A_k$ with $\|A - A_k\|_F \leq 2c\sqrt{n_1n_2 / (n_1n_2)^2}$ and we have
\[\begin{aligned} &\left| \frac{\sqrt{n_1n_2} |v^T B_N(A) u|}{\sqrt{n_1n_2} \|v^T B_N(A) u\|} \right| - \left| \frac{\sqrt{n_1n_2} |v^T B_N(A) u|}{\sqrt{n_1n_2} \|v^T B_N(A) u\|} \right| \leq \left| \frac{\sqrt{n_1n_2} |v^T B_N(A) u|}{\sqrt{n_1n_2} \|v^T B_N(A) u\|} \right| + \frac{\sqrt{n_1n_2} |v^T B_N(A) u|}{\sqrt{n_1n_2} \|v^T B_N(A) u\|} =: f_1 + f_2. \end{aligned}\]
It is easy to see that
\[
|I_1| \leq C \sum_{i=1}^{N} |v^T X_i u| \left| \text{tr}(X_i^T (A_{(k)} - A_\ast)) \right| \times \frac{1}{\sqrt{n_1 n_2} \times c_{\sqrt{n_1 n_2}} (n_1 n_2)^{M} (c + n_{\max} \log(n_+) / N)^{3/2}} =: I_3.
\]
With Lemma A.1, we can show that
\[
\Pr \left( I_3 \geq C \sqrt{\frac{\log(n_+)}{n_{\min} N}} \right) = O(n_+^{-\gamma}),
\]
for any \( \gamma > 0 \) by letting \( M \) be sufficiently large. For \( I_2 \), noting that
\[
\left| f \left( \text{tr}(X_i^T (A - A_\ast)) \right) - f \left( \text{tr}(X_i^T (A_{(k)} - A_\ast)) \right) \right| \leq C \frac{\sqrt{n_1 n_2 / (n_1 n_2)^M}}{n_1 n_2}.
\]
we have
\[
|I_2| \leq \sqrt{n_1 n_2} \left( \frac{c n_{\max} \log(n_+)}{N} \right)^{-1/4} \frac{1}{N} \sum_{i=1}^{N} |v^T X_i u| \times
\]
\[
I \left[ |\epsilon_i - \text{tr}(X_i^T (A_{(k)} - A_\ast))| \leq 2C \sqrt{n_1 n_2 / (n_1 n_2)^M} \right] \\
+ C \left( \frac{n_1 n_2}{M} \right)^{1/4} \times \frac{1}{N} \sum_{i=1}^{N} |v^T X_i u|
\]
\[
=: I_4 + I_5.
\]
It is easy to show that \( \mathbb{E}(I_4) = o(\sqrt{\log(n_+) / n_{\min} N}) \) with \( M \) large enough and
\[
\Pr \left( I_5 \geq C \sqrt{\frac{\log(n_+)}{n_{\min} N}} \right) \leq \frac{1}{N} \sum_{i=1}^{N} \Pr \left( \left| v^T X_i u \right| \geq \frac{(n_1 n_2)^{M - 2} N^{1/4}}{n_{\max} \log(n_+)} \right)
\]
\[
= O(n_+^{-\gamma}),
\]
for any \( \gamma > 0 \) by letting \( M \) be sufficiently large. Also for some \( \eta > 0 \),
\[
\mathbb{E}(v^T X_i u)^2 \exp(\eta |v^T X_i u|) \times
\]
\[
I \left[ |\epsilon_i - \text{tr}(X_i^T (A_{(k)} - A_\ast))| \leq 2c \sqrt{n_1 n_2 / (n_1 n_2)^M} \right] \\
\leq C \left( \frac{n_1 n_2}{M} \right)^{1/2} \mathbb{E}(v^T X_i u)^2 \exp(\eta |v^T X_i u|)
\]
\[
= O(1 / ((n_1 n_2)^{M - 1/2} n_{\min})).
\]
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 \( 1 / ((n_1 n_2)^{M - 1/2} n_{\min}) = o(B_n^2) \)), we have for large \( C > 0 \),
\[
\Pr \left( |I_4 - \mathbb{E}(I_4)| \geq C \sqrt{\frac{\log(n_+)}{n_{\min} N}} \right)
\]
\[
= O(n_+^{-\gamma}).
\]
As \( s_N \leq (n_1 n_2)^M (n_1 n_2)^M \), by choosing \( C \) sufficiently large such that \( \gamma > M \), it is enough to show that for any \( \gamma > 0 \),
\[
\sup_{|v|_2 = 1} \sup_{|u|_2 = 1} \max_{k} \Pr \left( \frac{v^T B_N (A_{(k)}) u}{n_1 n_2 \log(n_+) / N} \right) \geq C \frac{1}{n_{\min} N} \log(n_+)
\]
\[
= O(n_+^{-\gamma}).
\]
Set
\[
Z_i(A) = \mathbb{I} \left[ \epsilon_i \leq \text{tr}(X_i^T (A - A_\ast)) \right] - f \left( \text{tr}(X_i^T (A - A_\ast)) \right).
\]
Then we have
\[
\mathbb{E}(v^T X_i u)^2 (Z_i(A) - Z_i(A_\ast)^2 \exp(\eta |v^T X_i u|)
\]
\[
\leq C \frac{(n_1 n_2)}{n_{\min} N} \exp(\eta |v^T X_i u|)
\]
\[
\leq C \frac{(n_1 n_2)}{n_{\min} N} \log(n_+) / n_{\min} \text{ and } z^2 = \gamma \log(n_+) \text{ in Lemma 1 in (Cai & Liu, 2011), we can show (A.2) holds.}
\]
Let
\[
U_N = \sup_{\|A - A_\ast\|_F \leq s_N} \|B_N(A)\|
\]
For a unit ball \( B \) in \( R^s \), we have the fact that there exist \( q_0 \) balls with centers \( x_1, \ldots, x_{q_0} \) and radius \( z \) (i.e., \( B_i = \{ x \in R^s : \| x - x_i \| \leq z \}, 1 \leq i \leq q_0 \}) such that \( B \subset \cup_{1 \leq i \leq q_0} B_i \) and \( q_0 \) satisfies \( q_0 \leq \left( 1 + \frac{2}{\gamma} \right)^N \). Then by a standard \( \varepsilon \)-net argument, for any matrix \( A \in R^{n_1 \times n_2} \), there exist \( v_1, \ldots, v_{q_0} \) and \( u_1, \ldots, u_{q_0} \) (which do not depend on \( A \)) with \( |v_i|_2 = 1 \) and \( |u_i|_2 = 1 \), \( b_i \leq 9n^s \) and \( b_0 \leq 9n^s \) such that
\[
\|A\| \leq 5 \max_{1 \leq i \leq q_0} \max_{1 \leq j \leq b_0} |v_i^T A u_j|.
\]
So we have
\[
U_N \leq 5 \max_{1 \leq i \leq q_0} \max_{1 \leq j \leq b_0} |v_i^T B_N(A_{(k)}) u_j| \
\]
Assume the initial value \((n_1 n_2)^{-1/2} \|A_\ast - \hat{A}_0\|_F = o_p(1)\). By Lemma A.2, we have
\[
U_N = O_p \left( \sqrt{\frac{\log(n_+)}{n_1 n_2 n_{\min} N}} \right) + \log(n_+),
\]
So we have the following lemma.

**Lemma A.3.** Assume that Conditions (C1)-(C6) hold. We have
\[
U_N = O_p \left( \sqrt{\frac{\log(n_+)}{n_1 n_2 n_{\min} N}} \right) + \log(n_+).
\]
Distributed Median Matrix Completion

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 \( \mathbf{A}_* \in \mathcal{B}(a, n_1, n_2) \). By choosing the bandwidth \( h \approx (n_1n_2)^{-1/2}a_N \) and taking

\[
\lambda_N = C \left( \sqrt{\frac{\log(n_+)}{n_{\min}N} + \frac{a_N^2}{n_{\min}(n_1n_2)}} \right),
\]

where \( C \) is a sufficient large constant, we have

\[
\frac{\| \hat{\mathbf{A}}(1) - \mathbf{A}_* \|_F}{n_1n_2} = O_P \left[ \max \left\{ \sqrt{\frac{\log(n_+)}{N}}, r_* \left( \frac{n_{\max} \log(n_+)}{N} + \frac{a_N^4}{n_{\min}^2(n_1n_2)} \right) \right\} \right]. \tag{A.4}
\]

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 \( \mathbf{A}_* \in \mathcal{B}(a, n_1, n_2) \). By choosing the bandwidth \( h \approx (n_1n_2)^{-1/2}a_N \), we have

\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \xi_i^{(1)} \mathbf{X}_i \right\| = O_P \left( \sqrt{\frac{\log(n_+)}{n_{\min}N} + \frac{a_N^2}{n_{\min}(n_1n_2)}} \right).
\]

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

\[
\mathbf{H}_N(\mathbf{A}) = \frac{\hat{\mathbf{f}}^{-1}(0)}{N} \sum_{i=1}^{N} \mathbf{X}_i \{ f \left[ \text{tr} \left\{ \mathbf{X}_i^T \left( \mathbf{A} - \hat{\mathbf{A}}_0 \right) \right\} \right] - f(0) \}
+ \frac{1}{N} \sum_{i=1}^{N} \mathbf{X}_i \text{tr} \left\{ \mathbf{X}_i^T \left( \mathbf{A} - \mathbf{A}_* \right) \right\}.
\]

We have

\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \xi_i^{(1)} \mathbf{X}_i \right\| \leq \left\| \frac{\hat{\mathbf{f}}^{-1}(0)}{N} \sum_{i=1}^{N} \mathbf{X}_i \{ f \left[ \text{tr} \left\{ \mathbf{X}_i^T \left( \mathbf{A} - \hat{\mathbf{A}}_0 \right) \right\} \right] - f(0) \}
+ \frac{1}{N} \sum_{i=1}^{N} \mathbf{X}_i \text{tr} \left\{ \mathbf{X}_i^T \left( \mathbf{A} - \mathbf{A}_* \right) \right\} \right\|
\]

By Proposition A.1 and \((n_1n_2)^{1/2} \log(n_+) = o(Na_N)\), we have \( \hat{f}(0) \geq c \) for some \( c > 0 \) with probability tending to one. Therefore, for the last term, by Lemma A.3, we have

\[
\left| \hat{f}^{-1}(0) \right| U_N = O_P \left( \sqrt{\frac{a_N \log(n_+)}{n_1n_2n_{\min}N} + \frac{\log(n_+)}{n_{\min}N}} \right).
\]

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

\[
\left| \hat{f}^{-1}(0) \right| \left\| \frac{1}{N} \sum_{i=1}^{N} \mathbf{X}_i \left[ \mathbb{E} \left[ e_i \leq 0 \right] - f(0) \right] \right\| = O_P \left( \sqrt{\frac{\log(n_+)}{n_{\min}N}} \right).
\]

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

\[
\frac{\hat{f}^{-1}(0)}{N} \sum_{i=1}^{N} \mathbf{v}_i^T \mathbf{X}_i \left[ f \left( \text{tr} \left\{ \mathbf{X}_i^T \left( \mathbf{A} - \hat{\mathbf{A}}_0 \right) \right\} \right) - f(0) \right]
= \frac{\hat{f}^{-1}(0)}{N} f(0) \sum_{i=1}^{N} \mathbf{v}_i^T \mathbf{X}_i \text{tr} \left\{ \mathbf{X}_i^T \left( \mathbf{A} - \hat{\mathbf{A}}_0 \right) \right\}
+ O(1) \left( \frac{\hat{f}^{-1}(0)}{N} \sum_{i=1}^{N} | \mathbf{v}_i^T \mathbf{X}_i| \left[ \text{tr} \left\{ \mathbf{X}_i^T \left( \mathbf{A} - \hat{\mathbf{A}}_0 \right) \right\} \right]^2 \right).
\]

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
\[
\left\| v_i^T H_N (\widehat{A}_0) u_j \right\| \leq \left| \widehat{f}^{-1} (0) f (0) - 1 \right| \times \\
\left\| \frac{1}{N} \sum_{i=1}^{N} v_i^T X_i u_i \right\| \text{tr}\left\{ \left( \begin{array}{c} \text{A}_* - \widehat{A}_0 \end{array} \right) \right\} \\
+ C \widehat{f}^{-1} (0) \left( \frac{1}{N} \sum_{i=1}^{N} \left| v_i^T X_i u_j \right| \right)^2 \leq \frac{C}{n_{\min} (n_1 n_2)} \left\| \text{A}_* - \widehat{A}_0 \right\| F.
\]

We can easily have
\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \xi_i^{(1)} X_i \right\| = O_p \left( \frac{\log(n_1)}{n_{\min} N} + \frac{\sqrt{\log(n_1)} n_2}{n_{\min} n_1 n_2} \right) \\
+ a_N \left( \frac{\log(n_1)}{n_{\min} N} + \frac{n_2}{n_{\min} n_1 n_2} \right).
\]

The lemma is proved.

### 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 (\(N(0,1)\)) to the ground truth matrix \(\text{A}_*\) 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.

<table>
<thead>
<tr>
<th>(T)</th>
<th>RMSE</th>
<th>MAE</th>
<th>rank</th>
<th>rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1(4)</td>
<td>0.6646 (0.0238)</td>
<td>0.4826 (0.0232)</td>
<td>63.74 (5.57)</td>
<td></td>
</tr>
<tr>
<td>S2(5)</td>
<td>0.6985 (0.0415)</td>
<td>0.6708 (0.0404)</td>
<td>67.59 (6.76)</td>
<td></td>
</tr>
<tr>
<td>S2(6)</td>
<td>0.4460 (0.0080)</td>
<td>0.3179 (0.0067)</td>
<td>43.07 (6.00)</td>
<td></td>
</tr>
<tr>
<td>S4(4)</td>
<td>0.8522 (0.0208)</td>
<td>0.6229 (0.0210)</td>
<td>45.21 (5.52)</td>
<td></td>
</tr>
</tbody>
</table>

#### 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 \(n_2 = 918\).

<table>
<thead>
<tr>
<th>(T)</th>
<th>RMSE</th>
<th>MAE</th>
<th>rank</th>
<th>rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>RawA</td>
<td>0.9253</td>
<td>0.7264</td>
<td>54</td>
<td>45</td>
</tr>
<tr>
<td>RawB</td>
<td>0.9268</td>
<td>0.7344</td>
<td>57</td>
<td>44</td>
</tr>
<tr>
<td>OutA</td>
<td>1.0580</td>
<td>1.0867</td>
<td>48</td>
<td>33</td>
</tr>
<tr>
<td>OutB</td>
<td>0.8707</td>
<td>0.8632</td>
<td>24</td>
<td>31</td>
</tr>
</tbody>
</table>

##### B.2.2. MovieLens-1M

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

---

Distributed Median Matrix Completion

Proof of Theorem A.1. Due to the basic inequality, we have
\[
\left\| \frac{1}{N} \sum_{k=1}^{N} \left( \xi_k^{(1)} - \text{tr}(X_k^T \widehat{A}) \right) \right\|^2 \leq \lambda_N \left\| \widehat{A} \right\|_*
\]
which implies
\[
\left\| \frac{1}{N} \left\| \Omega \left( \text{A}_* - \widehat{A} \right) \right\|_F \right\|^2 + \lambda_N \left\| \widehat{A} \right\|_*
\]
\[
\leq 2 \left\| \widehat{A} - \text{A}_*, \Sigma^{(1)} \right\| + \lambda_N \left\| \text{A}_* \right\|_*
\]
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 \(\lambda_N\) 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.
1M 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 $A_*$ is of dimension $4290 \times 2505$.

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.

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**

