Consistent Collective Matrix Completion under Joint Low Rank Structure: Supplementary Material

Suriya Gunasekar suriya@utexas.edu
Makoto Yamada makotoy@yahoo-inc.com
Dawei Yin daweiy@yahoo-inc.com
Yi Chang yichang@yahoo-inc.com

A Operator Bernstein Inequality

Theorem 1 (Operator Bernstein Inequality [4])**.** *Let S_i, $i = 1, 2, \ldots, m$ be i.i.d self–adjoint operators of dimension N. If there exists constants R and σ^2, such that $\forall i \ \|S_i\|_{op} \leq R$ a.s., and $\sum_i \|E[S_i^2]\|_{op} \leq \sigma^2$,*

$$\text{then} \quad \forall\, t > 0 \quad Pr\big(\|\textstyle\sum_i S_i\|_{op} > t\big) \leq N \exp\left(\frac{-t^2/2}{\sigma^2 + \frac{Rt}{3}}\right) \tag{25}$$

B Proof of Lemma 1

Recall that:

- $T(\mathcal{X}) = \text{aff}\{\mathcal{Y} \in \bar{\mathfrak{X}} : \forall\, v, \text{rowSpan}(\mathbb{Y}_{r_v}) \subseteq \text{rowSpan}(\mathbb{X}_{r_v}) \textit{ or } \text{rowSpan}(\mathbb{Y}_{c_v}) \subseteq \text{rowSpan}(\mathbb{X}_{c_v})\}$
- $T^{\perp}(\mathcal{X}) = \{\mathcal{Y} \in \bar{\mathfrak{X}} : \forall\, v, \text{rowSpan}(Y_v) \perp \text{rowSpan}(M_v) \textit{ and } \text{colSpan}(Y_v) \perp \text{colSpan}(M_v)\}$

We need to show that $\forall \mathcal{X} \in \bar{\mathfrak{X}}$, $\mathcal{X} \in T^{\perp}$ iff $\langle \mathcal{X}, \mathcal{Y} \rangle = 0$, $\forall \mathcal{Y} \in T$.

$\Longrightarrow$ Let $\mathcal{X} \in \{\mathcal{X} \in \bar{\mathfrak{X}} : \langle \mathcal{X}, \mathcal{Y} \rangle = 0, \forall \mathcal{Y} \in T\}$, if $\mathcal{X} \notin T^{\perp}$, then $\exists v$ such that atleast one of the statements below hold true:

(a) $\text{rowSpan}(X_v) \not\perp \text{rowSpan}(M_v)$, or
(b) $\text{colSpan}(X_v) \not\perp \text{colSpan}(M_v)$

WLOG let us assume that (a) is true, the proof for the other case is analogous. Consider the decomposition $X_v = X_v^{(1)} + X_v^{(2)}$ such that $\text{rowSpan}(X_v^{(1)}) \perp \text{rowSpan}(M_v)$ and $\text{rowSpan}(X_v^{(2)}) \subseteq \text{rowSpan}(M_v)$. Consider the collective matrix $\mathcal{Y}$ such that $Y_{v'} = X_v^{(2)}$ if $v' = v$, and $Y_{v'} = 0$ otherwise. Clearly, $\mathcal{Y} \in T$ as $\forall\, v, \text{rowSpan}(\mathbb{Y}_{r_v}) \subseteq \text{rowSpan}(\mathbb{X}_{r_v})$, but $\langle \mathcal{X}, \mathcal{Y} \rangle \neq 0$, a contradiction.

$\Longleftarrow$ If $\mathcal{X} \in T^{\perp}$, then by the definitions, $\forall \mathcal{Y} \in T$, $\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_v \langle X_v, Y_v \rangle = 0$.

C Proof of Lemma 3

Recall $\mathcal{R}_s$ and $\mathcal{R}_\Omega$ from (18) and (19). Also recall that $\forall \mathcal{X} \in \mathfrak{X}$, $\mathcal{X} = \sum_{v=1}^{V} \sum_{(i,j) \in \mathcal{I}(v)} \langle \mathcal{X}, \mathcal{E}^{(v,i,j)} \rangle \mathcal{E}^{(v,i,j)}$.

Thus, $P_T(\mathcal{X}) = \sum_{v=1}^{V} \sum_{(i,j) \in \mathcal{I}(v)} \langle P_T(\mathcal{X}), \mathcal{E}^{(v,i,j)} \rangle \mathcal{E}^{(v,i,j)} = \sum_{v=1}^{V} \sum_{(i,j) \in \mathcal{I}(v)} \langle \mathcal{X}, P_T(\mathcal{E}^{(v,i,j)}) \rangle \mathcal{E}^{(v,i,j)}$

Define $\mathcal{V}_s := P_T \mathcal{R}_s P_T : \mathcal{X} \to \frac{1}{p(v_s, i_s, j_s)} \langle \mathcal{X}, P_T(\mathcal{E}^{(s)}) \rangle P_T(\mathcal{E}^{(s)})$, where $p(v, i, j) = \frac{|\Omega_{r_v}|}{2 n_{r_v} m_{r_v}} + \frac{|\Omega_{c_v}|}{2 n_{c_v} m_{c_v}}$.

We then have $E[\mathcal{V}_s] = \frac{1}{|\Omega|} P_T$, and

$$\begin{aligned}\|\mathcal{V}_s\|_{\text{op}} &= \sup_{\|\mathcal{X}\|_F=1} \frac{1}{p(v_s,i_s,j_s)} \langle \mathcal{X}, P_T(\mathcal{E}^{(s)})\rangle \|P_T(\mathcal{E}^{(s)})\|_F = \frac{1}{p(v_s,i_s,j_s)} \|P_T(\mathcal{E}^{(s)})\|_F^2 \\ &\overset{(a)}{\leq} \frac{1}{p(v_s,i_s,j_s)} \left(\frac{\mu_0 R}{m_{r_{v_s}}} + \frac{\mu_0 R}{m_{c_{v_s}}}\right) \overset{(b)}{\leq} \frac{1}{c_0 \beta \log N},\end{aligned} \tag{26}$$

where (a) follows from the incoherence condition in Assumption 2, and (b) follows as $\forall k$, $|\Omega_k| > c_0 \mu_0 n_k R \beta \log N$.

(i) Bound on $\|\mathcal{V}_s - E[\mathcal{V}_s]\|_{\text{op}}$

$$\|\mathcal{V}_s - E[\mathcal{V}_s]\|_{\text{op}} \overset{(a)}{\leq} \max\left(\|\mathcal{V}_s\|_{\text{op}}, \|E[\mathcal{V}_s]\|_{\text{op}}\right) \leq \max\left(\frac{1}{c_0\beta \log N}, \frac{1}{\Omega}\right) = \frac{1}{c_0 \beta \log N} \tag{27}$$

where (a) follows as both $\mathcal{V}_s$ and $E[\mathcal{V}_s]$ are positive semidefinite.

(ii) Bound on $\sum_{s=1}^{|\Omega|} \|E[(\mathcal{V}_s - E[\mathcal{V}_s])^2]\|_{\text{op}}$.

$$\begin{aligned}E[(\mathcal{V}_s)^2(X)] &= E\left[\frac{1}{p(v_s,i_s,j_s)^2} \langle \mathcal{X}, P_T(\mathcal{E}^{(s)})\rangle \|P_T(\mathcal{E}^{(s)})\|_F^2 P_T(\mathcal{E}^{(s)})\right] \\ &\leq \frac{1}{c_0\beta\log N} E\left[\frac{1}{p(v_s,i_s,j_s)} \langle \mathcal{X}, P_T(\mathcal{E}^{(s)})\rangle P_T(\mathcal{E}^{(s)})\right] = \frac{1}{|\Omega| c_0 \beta \log N} P_T(\mathcal{X}).\end{aligned} \tag{28}$$

$$\|E[(\mathcal{V}_s - E[\mathcal{V}_s])^2]\|_{\text{op}} = \|E[\mathcal{V}_s^2] - (E[\mathcal{V}_s])^2]\|_{\text{op}} \leq \max\left(\|E[\mathcal{V}_s^2]\|_{\text{op}}, \|(E[\mathcal{V}_s])^2\|_{\text{op}}\right) \overset{(a)}{\leq} \frac{1}{|\Omega| c_0 \beta \log N}, \tag{29}$$

where (a) follows as $\|P_T\|_{\text{op}} \leq 1$.

Thus, $\sigma^2 := \sum_{s=1}^{|\Omega|} \|E[(\mathcal{V}_s - E[\mathcal{V}_s])^2]\|_{\text{op}} \leq \frac{1}{c_0 \beta \log N}$

(iii) The lemma follows by using (i) and (ii) above in the operator Bernstein inequality in (25).

D Proof of Lemma 4

Recall that under the assumptions made in the paper $\|\cdot\|_{\mathscr{A}}$ is norm, and by the sub differential characterization of norms we have the following:

$$\partial \|\mathcal{M}\|_{\mathscr{A}} = \{\mathcal{E} + \mathcal{W} : \mathcal{E} \in \mathscr{E}(\mathcal{M}) \cap T, \mathcal{W} \in T^\perp, \|\mathcal{W}\|_{\mathcal{A}}^* \leq 1\} \tag{30}$$

Recall $\mathscr{E}(\mathcal{M})$ from (4). In particular the set $\{\mathcal{E}_{\mathcal{M}} + \mathcal{W} : \mathcal{W} \in T^\perp, \|\mathcal{W}\|_{\mathcal{A}}^* \leq 1\} \subset \partial\|\mathcal{M}\|_{\mathscr{A}}$, where $\mathcal{E}_{\mathcal{M}}$ is the sign vector from Assumption 2.

Given any Δ, with $P_\Omega(\Delta) = 0$, consider any $\mathcal{W} \in T^\perp$, such that $\|P_{T^\perp}(\Delta)\|_{\mathscr{A}} = \langle \mathcal{W}, P_{T^\perp}(\Delta)\rangle$ and $\mathcal{E}_{\mathcal{M}} + \mathcal{W} \in \partial\|\mathcal{M}\|_{\mathscr{A}}$. Let $\mathcal{Y} = P_\Omega(\mathcal{Y})$ be a dual certificate satisfying the conditions stated in the Lemma.

$$\begin{aligned}\|\mathcal{M} + \Delta\|_{\mathscr{A}} &\overset{(a)}{\geq} \|\mathcal{M}\|_{\mathscr{A}} + \langle \mathcal{E}_{\mathcal{M}} + \mathcal{W} - \mathcal{Y}, \Delta\rangle = \|\mathcal{M}\|_{\mathscr{A}} + \langle \mathcal{E}_{\mathcal{M}} - P_T(\mathcal{Y}), P_T(\Delta)\rangle + \langle \mathcal{W} - P_{T^\perp}(\mathcal{Y}), P_{T^\perp}(\Delta)\rangle \\ &\overset{(b)}{\geq} \|\mathcal{M}\|_{\mathscr{A}} - \|\mathcal{E}_{\mathcal{M}} - P_T(\mathcal{Y})\|_F \|P_T(\Delta)\|_F + \|P_{T^\perp}(\Delta)\|_{\mathscr{A}}(1 - \|P_{T^\perp}(\mathcal{Y})\|_{\mathscr{A}}^*) \\ &\overset{(c)}{\geq} \|\mathcal{M}\|_{\mathscr{A}} - \frac{1}{2}\kappa_\Omega(N) \|\mathcal{E}_{\mathcal{M}} - P_T(\mathcal{Y})\|_F \|P_{T^\perp}(\Delta)\|_F + \frac{1}{2}\|P_{T^\perp}(\Delta)\|_{\mathscr{A}} \overset{(d)}{>} \|\mathcal{M}\|_{\mathscr{A}},\end{aligned} \tag{31}$$

where (a) follows as $\langle \Delta, \mathcal{Y}\rangle = 0$, (b) follows from Holder's inequality, (c) follows as $\|P_{T^\perp}(\mathcal{Y})\|_{\mathscr{A}}^* \leq \frac{1}{2}$ and $\frac{1}{2}\kappa_\Omega(N)\|P_{T^\perp}(\Delta)\|_F \geq \|P_T(\Delta)\|_F$ w.h.p. (from (22)), and (d) follows as $\|\mathcal{E}_{\mathcal{M}} - P_T(\mathcal{Y})\|_F < \frac{1}{\kappa_\Omega(N)}$ and using $\|\mathcal{X}\|_{\mathscr{A}} = \min_{Z \succcurlyeq 0} tr(Z) \;\; s.t. P_v[Z] = X_v \; \forall v \geq \min_{Z\succcurlyeq 0} \|Z\|_F \;\; s.t. P_v[Z] = X_v \; \forall v \geq \|\mathcal{X}\|_F$.

E Dual Certificate–Bound on $\|P_{T^\perp}\mathcal{Y}_p\|_{\mathscr{A}}^*$

Recall that $\mathcal{Y}_p$ was constructed through a iterative process described in Sec. 5.2 following a golfing scheme introduced by Gross et al. [1]. The proof for the second property of the dual certificate, extends directly from the analogous proof for matrix completion by Recht [2]. We note that:

$$\|P_{T^\perp}\mathcal{Y}_p\|_{\mathscr{A}}^* \leq \sum_{j=1}^{p}\|P_{T^\perp}\mathcal{R}_{\Omega^{(j)}}\mathcal{W}_{j-1}\|_{\mathscr{A}}^* = \sum_{j=1}^{p}\|P_{T^\perp}(\mathcal{R}_{\Omega^{(j)}}-\mathcal{I})\mathcal{W}_{j-1}\|_{\mathscr{A}}^* \leq \sum_{j=1}^{p}\|(\mathcal{R}_{\Omega^{(j)}}-\mathcal{I})\mathcal{W}_{j-1}\|_{\mathscr{A}}^* \tag{32}$$

Denote $\max_{(v,i,j)}|\langle \mathcal{X}, \mathcal{E}^{(v,i,j)}\rangle| = \|\mathcal{X}\|_{\max}$.

We state the following lemmas which are directly adapted from Theorem 3.5 and Lemma 3.6 in [2]:

Lemma 5. *Let Ω be any subset of entries of size $|\Omega|$ sampled independently according to Assumption 4, such that $E[\mathcal{R}_s(\mathcal{W})] = \frac{1}{|\Omega|}\mathcal{W}$, then for all $\beta > 1$ and $N \geq 2$, the following holds with probability greater than $1 - N^{1-\beta}$ provided $|\Omega| > 6N\beta\log N$, and $\frac{|\Omega_k|}{n_k m_k} \geq \frac{|\Omega|}{N^2}; \forall k$:*

$$\|(\mathcal{R}_\Omega - \mathcal{I})\mathcal{W}\|_{\mathscr{A}}^* \leq \|\mathcal{B}(\mathcal{R}_\Omega\mathcal{W} - \mathcal{W})\|_2 \leq \sqrt{\frac{8\beta N^3 \log N}{3|\Omega|}}\|\mathcal{W}\|_{max} \tag{33}$$

Proof. The proof is obtained by applying the steps described for the analogous proof in [2] on $\|\mathcal{B}(\mathcal{R}_\Omega\mathcal{W} - \mathcal{W})\|_2$. For $s = 1, 2, \ldots, |\Omega|$, let $\mathcal{V}_s = \mathcal{B}(\mathcal{R}_s(\mathcal{W}))$, then $\mathcal{B}(\mathcal{R}_\Omega\mathcal{W} - \mathcal{W}) = \sum_{s=1}^{|\Omega|}(\mathcal{V}_s - E[\mathcal{V}_s])$ is a sum of independent zero mean random variables. From the proof of Theorem 3.5 in the work by Recht [3], we have that for any $N \times N$ matrix Z, $\|Z\|_2 \leq N\|Z\|_{\max}$.

(i) $\|\mathcal{V}_s - E[\mathcal{V}_s]\|_2 \leq \|\mathcal{V}_s\|_2 + \|E[\mathcal{V}_s]\|_2 \overset{(a)}{\leq} \frac{N^2}{|\Omega|}\|\mathcal{W}\|_{\max} + \frac{N}{|\Omega|}\|\mathcal{W}\|_{\max} \leq \frac{3N^2}{2|\Omega|}\|\mathcal{W}\|_{\max}$ for $N \geq 2$, where (a) follows as $\frac{1}{p(v,i,j)} \leq \frac{1}{\min_k \frac{|\Omega_k|}{n_k m_k}} \leq \frac{N^2}{|\Omega|}$ if $\frac{|\Omega_k|}{n_k m_k} \geq \frac{|\Omega|}{N^2}, \forall k$; and $\|E[\mathcal{V}_s]\|_2 = \frac{1}{|\Omega|}\|\mathcal{B}(\mathcal{W})\|_2$.

(ii) $\|E[(\mathcal{V}_s - E[\mathcal{V}_s])^2]\|_2 = \|E[\mathcal{V}_s^2] - (E[\mathcal{V}_s])^2\|_2 \leq \max\{\|E[\mathcal{V}_s^2]\|_2, \|(E[\mathcal{V}_s])^2\|_2\}$.

Now, $\|(E[\mathcal{V}_s])^2\|_2 = \frac{1}{|\Omega|^2}\|\mathcal{B}(\mathcal{W}) * \mathcal{B}(\mathcal{W})\|_2 \leq \frac{N^2}{|\Omega|^2}\|\mathcal{W}\|_{\max}^2$.

Also, $\|E[\mathcal{V}_s^2]\|_2 = \frac{1}{|\Omega|}\Big\|\sum_{v=1}^{V}\sum_{(i,j)\in\mathcal{I}(v)}\frac{1}{p(v,i,j)}\langle\mathcal{W}, \mathcal{E}^{(v,i,j)}\rangle\mathcal{B}(\mathcal{E}^{(v,i,j)})\Big\|_2 \leq \frac{N^4}{|\Omega|^2}\|\mathcal{W}\|_{\max}^2$.

Thus $\sigma^2 := \|E[(\mathcal{V}_s - E[\mathcal{V}_s])^2]\|_2 \leq \frac{N^4}{|\Omega|^2}\|\mathcal{W}\|_{\max}^2$

The proof follows by using the above bounds in operator Bernstein's inequality with $t = \sqrt{\frac{8\beta N^3\log N}{3|\Omega|}}\|\mathcal{W}\|_{\max}$

Lemma 6. *If $\forall k$, $|\Omega_k| \geq c_0\beta n_k R\log N$, and the Assumptions in 3.1 are satisfied, then for sufficiently large c_0, the following holds with probability greater that $1 - N^{1-\beta}$:*

$$\forall\, \mathcal{W} \in T\ \|P_T\mathcal{R}_\Omega\mathcal{W} - \mathcal{W}\|_{max} \leq \frac{1}{2}\|\mathcal{W}\|_{max} \tag{34}$$

Using the above lemmas in (32), we have:

$$\begin{aligned}\|P_{T^\perp}\mathcal{Y}_p\|_{\mathscr{A}}^* &\leq \sum_{j=1}^{p}\|(\mathcal{R}_{\Omega^{(j)}}-\mathcal{I})\mathcal{W}_{j-1}\|_{\mathscr{A}}^* \overset{(a)}{\leq} \sum_{j=1}^{p}\sqrt{\frac{8\beta N^3\log N}{3|\Omega^{(j)}|}}\|\mathcal{W}_{j-1}\|_{\max} \\ &\overset{(b)}{\leq} 2\sum_{j=1}^{p}2^{-j}\sqrt{\frac{8\beta N^3\log N}{3|\Omega^{(j)}|}}\|\mathcal{E}_{\mathcal{M}}\|_{\max} \overset{(c)}{\leq} 2\sum_{j=1}^{p}2^{-j}\sqrt{\frac{8\beta\mu_1 RN\log N}{3|\Omega^{(j)}|}} \overset{(d)}{\leq} \frac{1}{2},\end{aligned} \tag{35}$$

where (a) follows from Lemma 5, (b) from Lemma 6 as $\mathcal{W}_j = \mathcal{W}_{j-1} - P_T\mathcal{R}_\Omega\mathcal{W}_{j-1}$, (c) from the second incoherence condition in Assumption 2, and finally (d) if for large enough c_1, $|\Omega^{(j)}| > c_1\mu_1\beta RN\log N$.

Finally, the probability that the proposed dual certificate $\mathcal{Y}_p$ fails the conditions of Lemma 4 is given by a union bound of the failure probabilities of (24), Lemma 5, and 6 for any partition $\Omega^{(j)}$: $3c_1\log(N\kappa_\Omega(N))N^{1-\beta}$; thus proving Theorem 1.

E.1 Proof of Lemma 6

Using union bound and noting that $\sum_v n_{r_v}n_{c_v} \leq N^2$, we have:

$$Pr(\|P_T\mathcal{R}_\Omega\mathcal{W} - \mathcal{W}\|_{\max} > \frac{1}{2}\|\mathcal{W}\|_{\max}) \leq Pr(\langle P_T\mathcal{R}_\Omega\mathcal{W} - \mathcal{W}, \mathcal{E}^{(v,i,j)}\rangle > \frac{1}{2}\|\mathcal{W}\|_{\max} \text{ for any (v,i,j)})N^2$$

For each (v,i,j), sample $s' = (v_{s'}, i_{s'}, j_{s'})$ according to the sampling distribution in Assumption 4. Define $\Psi_{(v,i,j)} = \langle\mathcal{E}^{(v,i,j)}, P_T\mathcal{R}_{s'}\mathcal{W} - \frac{1}{|\Omega|}\mathcal{W}\rangle$. Recall the definition of $\mathcal{R}_s$ from the paper. Now each entry of $P_T\mathcal{R}_\Omega\mathcal{W} - \mathcal{W}$ is distributed as $\sum_{s=1}^{|\Omega|}\Psi_{(v,i,j)}^{(s)}$, where $\Psi_{(v,i,j)}^{(s)}$ are iid samples of $\Psi_{(v,i,j)}$.

We have that : $|\Psi_{(v,i,j)}| \leq \frac{1}{p(v,i,j)}\|P_T(\mathcal{E}^{(v,i,j)})\|_F^2\langle\mathcal{E}^{(v,i,j)}, \mathcal{W}\rangle| \leq \frac{1}{c'\beta\log N}\|\mathcal{W}\|_{\max}$

Also, $E[\Psi_{(v,i,j)}^2] = E[\frac{1}{p(v,i,j)^2}\langle\mathcal{E}^{(v,i,j)}, \mathcal{W}\rangle^2\langle\mathcal{E}^{(v,i,j)}, \mathcal{E}^{(s')}\rangle^2] \leq \frac{1}{|\Omega|c'\beta\log N}$, where the expectation is over s'. Standard Bernstein inequality can be used with the above bounds to prove the lemma.

References

[1] D Gross. Recovering low-rank matrices from few coefficients in any basis. *Information Theory, IEEE Transactions on*, 2011.

[2] B. Recht. A simpler approach to matrix completion. *JMLR*, 2011.

[3] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 2010.

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