Supplementary Material: Matrix Completion from Non-Uniformly Sampled Entries

Paper ID: 3555

A Supplementary Analysis

In this section, we first give the supporting theorems that we will use in this analysis. Then we provided the omitted proofs.

A.1 Supporting Theorems

The following results are used throughout the analysis.

Lemma 6. *(Theorem 1.1 of Tropp [2012]) Let $\mathcal{X}$ be a finite set of PSD matrices with dimension k (means the size of the square matrix is $k \times k$). $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ calculate the maximum and minimum eigenvalues respectively. Suppose that*

$$\max_{X \in \mathcal{X}} \lambda_{\max}(X) \leq B.$$

Sample $\{X_1, \cdots, X_l\}$ uniformly at random from $\mathcal{X}$ independently. Compute

$$\mu_{\max} = \lambda_{\max}\left(\sum_{i=1}^{l} \mathrm{E}[X_i]\right), \mu_{\min} = \lambda_{\min}\left(\sum_{i=1}^{l} \mathrm{E}[X_i]\right).$$

Then

$$\begin{aligned}
&\Pr\left\{\lambda_{\max}\left(\sum_{i=1}^{l} X_i\right) \geq (1+\rho)\mu_{\max}\right\} \\
\leq & k \exp\frac{-\mu_{\max}}{B}\left[(1+\rho)\ln(1+\rho) - \rho\right] \text{ for } \rho \geq 0, \\
&\Pr\left\{\lambda_{\min}\left(\sum_{i=1}^{l} X_i\right) \leq (1-\rho)\mu_{\min}\right\} \\
\leq & k \exp\frac{-\mu_{\min}}{B}\left[(1-\rho)\ln(1-\rho) + \rho\right] \text{ for } \rho \in [0, 1).
\end{aligned}$$

Lemma 7. *(Lemma 1 of Laurent and Massart [2000]) Let $x \sim \chi_d^2$. Then with probability at least $1 - 2\delta$ the following holds*

$$-2\sqrt{d\ln(1/\delta)} \leq x - d \leq 2\sqrt{d\ln(1/\delta)} + 2\ln(1/\delta).$$

Lemma 8. *Let $x_1, \cdots, x_n \sim \mathcal{N}(0, \sigma^2)$. Then with probability at least $1 - \delta$ the following holds*

$$\max_{i \in [n]} |x_i| \leq \sigma\sqrt{2\ln(2n/\delta)}.$$

Lemma 9. *(Corollary 5.35 of Vershynin [2010]) Let R be an $n \times t$ random matrix with independent and identically distributed standard Gaussian entries. Then for every $\epsilon \geq 0$ with probability at least $1 - 2\exp(-\epsilon^2/2)$ the following holds*

$$\sqrt{n} - \sqrt{t} - \epsilon \leq \sigma_{\min}(R) \leq \sigma_{\max}(R) \leq \sqrt{n} + \sqrt{t} + \epsilon.$$

Lemma 10. *(Theorem 2 in Drineas* et al. *[2006]) Suppose $M \in \mathbb{R}^{m \times n}$. Let A and $\hat{U}$ be constructed by Algorithm 1. Then we have*

$$\|M - \hat{U}\hat{U}^T M\|_F^2 \leq \|M - M_r\|_F^2 + 2\sqrt{r}\|MM^T - AA^T\|_F.$$

Lemma 11. *(Lemma 2 in Smale and Zhou [2007]) Let $\mathcal{H}$ be a Hilbert space and let ξ be a random variable with values in $\mathcal{H}$. Assume $\|\xi\| \leq M \leq \infty$ almost surely. Denote $\sigma^2(\xi) = \mathbb{E}\left[\|\xi\|^2\right]$. Let $\{\xi_i\}_{i=1}^d$ be $d\,(d < \infty)$ independent drawers of ξ. For any $0 < \delta < 1$, with confidence $1 - \delta$*

$$\left\|\frac{1}{d}\sum_{i=1}^{d}[\xi_i - \mathbb{E}[\xi_i]]\right\| \leq \frac{2M\ln(2/\delta)}{d} + \sqrt{\frac{2\sigma^2(\xi)\ln(2/\delta)}{d}}.$$

A.2 Proof of Lemma 1

Let $i_1, \cdots, i_d$ are the d selected columns. Define $S = (\mathbf{e}_{i_1}/\sqrt{dp_{i_1}}, \mathbf{e}_{i_2}/\sqrt{dp_{i_2}}, \cdots, \mathbf{e}_{i_d}/\sqrt{dp_{i_d}}) \in \mathbb{R}^{n \times d}$ where $\mathbf{e}_i$ is the i-th canonical basis. Such that we have $A = MS$, that is, A is composed of d selected and rescaled columns of M. Let the SVD of M be $M = \bar{U}\bar{\Sigma}\bar{V}^T$, where $\bar{U} \in \mathbb{R}^{m \times r}, \bar{\Sigma} \in \mathbb{R}^{r \times r}, \bar{V} \in \mathbb{R}^{n \times r}$. We have $A = \bar{U}\bar{\Sigma}\bar{V}^T S$. To prove $\mathrm{rank}(A) = r$, we need to bound the minimum eigenvalue of $\Psi\Psi^T$, where $\Psi = \bar{V}^T S \in \mathbb{R}^{r \times d}$. We have

$$\Psi\Psi^T = \bar{V}^T S S^T \bar{V} = \sum_{j=1}^{d} \frac{1}{dp_{i_j}} \bar{V}_{(i_j)}^T \bar{V}_{(i_j)}$$

where $\bar{V}_{(i)}, i \in [n]$ is the i-th row vector of $\bar{V}$.
It is straightforward to show that

$$\mathrm{E}\left[\bar{V}_{(i_j)}^T \bar{V}_{(i_j)}\right] = \sum_{i=1}^{n} p_i \frac{1}{dp_i} \bar{V}_{(i)}^T \bar{V}_{(i)} = \frac{1}{d} I_r$$

and

$$\mathrm{E}\left[\Psi\Psi^T\right] = I_r.$$

To bound the minimum eigenvalue of $\Psi\Psi^T$, we need Lemma 6, where we first need to bound the maximum eigenvalue of $\frac{1}{dp_i}\bar{V}_{(i)}^T\bar{V}_{(i)}$, which is a rank-1 matrix, whose eigenvalue

$$\max_{i\in[n]}\lambda_{\max}\left(\frac{1}{dp_i}\bar{V}_{(i)}^T\bar{V}_{(i)}\right) \leq \frac{1}{dp_{\min}}\max_{1\leq i\leq n}\|\bar{V}_{(i)}\|_2^2 \leq \frac{1}{dp_{\min}}\mu(r)\frac{r}{n}$$

and

$$\lambda_{\min}\left(\sum_{j=1}^{d}\mathrm{E}\left[\bar{V}_{(i_j)}^T\bar{V}_{(i_j)}\right]\right) = \lambda_{\min}\left(\mathrm{E}[\Psi\Psi^T]\right) = 1.$$

Thus, we have

$$\begin{aligned}&\Pr\left\{\lambda_{\min}\left(\Psi\Psi^T\right)\leq(1-\rho)\right\}\\ \leq& r\exp\frac{-1}{r\mu(r)/(ndp_{\min})}[(1-\rho)\ln(1-\rho)+\rho]\\ =& r\exp\frac{-ndp_{\min}}{r\mu(r)}[(1-\rho)\ln(1-\rho)+\rho].\end{aligned}$$

By setting $\rho=1/2$, we have,

$$\Pr\left\{\lambda_{\min}\left(\Psi\Psi^T\right)\leq\frac{1}{2}\right\} \leq r\exp\frac{-ndp_{\min}}{7r\mu(r)} = re^{-ndp_{\min}/(7r\mu(r))}.$$

Let $d \geq 7\mu(r)r(t+\ln r)/(np_{\min})$, we have $re^{-ndp_{\min}/(7r\mu(r))}\leq e^{-t}$. Then, we have

$$\Pr\left\{\sigma_{\min}\left(\Psi\right)\geq\sqrt{\frac{1}{2}}\right\} = \Pr\left\{\lambda_{\min}\left(\Psi\Psi^T\right)\geq\frac{1}{2}\right\} \geq 1-e^{-t}.$$

This means $\mathrm{rank}(\Psi)=r$, so $\mathrm{rank}(A)=\mathrm{rank}(\bar{U}\bar{\Sigma}\Psi)=r$.

A.3 Proof of Lemma 2

According the previous definition, $\hat{U}_{(j)}$, $j\in\mathcal{O}_i$ is the j-th row vector of $\hat{U}$. We have

$$\hat{U}_{\mathcal{O}_i}^T\hat{U}_{\mathcal{O}_i} = \sum_{j\in\mathcal{O}_i}\hat{U}_{(j)}^T\hat{U}_{(j)}.$$

It is straightforward to show that

$$\mathrm{E}\left[\hat{U}_{(j)}^T\hat{U}_{(j)}\right] = \frac{1}{m}I_{\hat{r}} \text{ and } \mathrm{E}\left[\hat{U}_{\mathcal{O}_i}^T\hat{U}_{\mathcal{O}_i}\right] = \frac{s}{m}I_{\hat{r}}.$$

To bound the minimum eigenvalue of $\hat{U}_{\mathcal{O}_i}^T\hat{U}_{\mathcal{O}_i}$, we need Lemma 6, where we first need to bound the maximum eigenvalue of $\hat{U}_{(j)}^T\hat{U}_{(j)}$, which is a rank-1 matrix, whose eigenvalue

$$\max_{j\in[m]}\lambda_{\max}\left(\hat{U}_{(j)}^T\hat{U}_{(j)}\right) = \max_{j\in[m]}\|\hat{U}_{(j)}\|_2^2 \leq \hat{\mu}\left(\hat{r}\right)\frac{\hat{r}}{m}$$

and

$$\lambda_{\min}\left(\sum_{j\in\mathcal{O}_i}\mathrm{E}\left[\hat{U}_{(j)}^T\hat{U}_{(j)}\right]\right) = \lambda_{\min}\left(\mathrm{E}\left[\hat{U}_{\mathcal{O}_i}^T\hat{U}_{\mathcal{O}_i}\right]\right) = \frac{|\mathcal{O}_i|}{m}.$$

Thus, we have

$$\begin{aligned}&\Pr\left\{\lambda_{\min}\left(\hat{U}_{\mathcal{O}_i}^T\hat{U}_{\mathcal{O}_i}\right)\leq(1-\rho)\frac{|\mathcal{O}_i|}{m}\right\}\\ \leq& \hat{r}\exp\frac{-|\mathcal{O}_i|/m}{\hat{r}\hat{\mu}(\hat{r})/m}[(1-\rho)\ln(1-\rho)+\rho]\\ =& \hat{r}\exp\frac{-|\mathcal{O}_i|}{\hat{r}\hat{\mu}(\hat{r})}[(1-\rho)\ln(1-\rho)+\rho].\end{aligned}$$

By setting $\rho=1/2$, we have

$$\Pr\left\{\lambda_{\min}\left(\hat{U}_{\mathcal{O}_i}^T\hat{U}_{\mathcal{O}_i}\right)\leq\frac{|\mathcal{O}_i|}{2m}\right\} \leq \hat{r}\exp\frac{-|\mathcal{O}_i|}{7\hat{r}\hat{\mu}(\hat{r})} = \hat{r}e^{-|\mathcal{O}_i|/7\hat{r}\hat{\mu}(\hat{r})}$$

where with $|\mathcal{O}_i|\geq 7\hat{\mu}(\hat{r})\hat{r}(t+\ln\hat{r}+\ln n)$, we have $\hat{r}e^{-|\mathcal{O}_i|/7\hat{r}\hat{\mu}(\hat{r})}\leq\frac{1}{n}e^{-t}$, that is

$$\Pr\left\{\lambda_{\min}\left(\hat{U}_{\mathcal{O}_i}^T\hat{U}_{\mathcal{O}_i}\right)\geq\frac{|\mathcal{O}_i|}{2m}\right\}\geq 1-\frac{1}{n}e^{-t}.$$

A.4 Proof of Lemma 4

According to our algorithm, we have selected $\hat{A} = [M^{(i_1)}, M^{(i_2)}, \cdots, M^{(i_d)}]$ from M and rescaled it to A, which means $\mathrm{span}(A) = \mathrm{span}(\hat{A})$. Let $\hat{C} = [C^{(i_1)}, C^{(i_2)}, \cdots, C^{(i_d)}]$, $\hat{R} = [R^{(i_1)}, R^{(i_2)}, \cdots, R^{(i_d)}]$ and the top-$\hat{r}$ SVD of $\hat{A} = \hat{C}+\hat{R}$ be $\hat{U}\hat{\Sigma}\hat{V}^T$, where $\hat{U}\in\mathbb{R}^{m\times\hat{r}}, \hat{\Sigma}\in\mathbb{R}^{\hat{r}\times\hat{r}}, \hat{V}\in\mathbb{R}^{d\times\hat{r}}$. First, according to the definition of SVD and span of columns, we have

$$\begin{aligned}\mathrm{span}(\hat{U}) =& \mathrm{span}(\hat{A}\hat{V}) = \mathrm{span}\left((\hat{C}+\hat{R})\hat{V}\right)\\ \subseteq& \mathrm{span}(\hat{C}\hat{V})\cup\mathrm{span}(\hat{R}\hat{V})\\ \subseteq& \mathrm{span}(\bar{U})\cup\mathrm{span}(\hat{R}\hat{V}).\end{aligned}$$

Then, let $\widetilde{R} = \hat{R}\hat{V}\in\mathbb{R}^{m\times\hat{r}}$ and B_{ij} be the entry of i-th row and the j-th column for any matrix B. We have $\widetilde{R}_{ij} = \sum_{k=1}^{d}\hat{R}_{ik}\hat{V}_{kj}$. Because $\hat{R}_{ij}\sim\mathcal{N}(0,\sigma^2)$, $\sum_{k=1}^{d}\hat{V}_{kj}^2 = \|\hat{V}_{(j)}\| = 1$ and the properties of Gaussian random variables, we have $\widetilde{R}_{ij}\sim\mathcal{N}(0,\sigma^2)$, which means that $\widetilde{R}$ is also a Gaussian random matrix. Let $\hat{\mathbf{r}}_i = [\hat{R}_{i1}, \hat{R}_{i2}, \cdots, \hat{R}_{id}]^T$, we have that $\mathrm{E}(\hat{\mathbf{r}}_i\hat{\mathbf{r}}_j^T) = \mathbf{0}_{d\times d}$ for $i\neq j$ and $\mathrm{E}(\hat{\mathbf{r}}_i\hat{\mathbf{r}}_j^T) = \sigma^2 I_d$ for $i=j$. Furthermore, for any two different entries $\widetilde{R}_{i_1j_1}$ and $\widetilde{R}_{i_2j_2}$ of $\widetilde{R}$, we have

$$\begin{aligned}&\mathrm{Cov}(\widetilde{R}_{ij_1}, \widetilde{R}_{ij_2})\\ =& \mathrm{E}((\widetilde{R}_{i_1j_1}-\mathrm{E}(\widetilde{R}_{i_1j_1}))(\widetilde{R}_{i_2j_2}-\mathrm{E}(\widetilde{R}_{i_2j_2})))\\ =& \mathrm{E}(\widetilde{R}_{i_1j_1}\widetilde{R}_{i_2j_2}) = \mathrm{E}(\hat{V}_{j_1}^T\hat{\mathbf{r}}_{i_1}\hat{\mathbf{r}}_{i_2}^T\hat{V}_{j_2}).\end{aligned}$$

When $i_1\neq i_2$, we have

$$\begin{aligned}\mathrm{Cov}(\widetilde{R}_{ij_1}, \widetilde{R}_{ij_2}) =& \mathrm{E}(\hat{V}_{j_1}^T\hat{\mathbf{r}}_{i_1}\hat{\mathbf{r}}_{i_2}^T\hat{V}_{j_2})\\ =& \hat{V}_{j_1}^T\mathrm{E}(\hat{\mathbf{r}}_{i_1}\hat{\mathbf{r}}_{i_2}^T)\hat{V}_{j_2} = 0\end{aligned}$$

and when $i_1 = i_2$ and $j_1 \neq j_2$, we have

$$\begin{aligned}\mathrm{Cov}(\widetilde{R}_{ij_1}, \widetilde{R}_{ij_2}) =& \mathrm{E}(\hat{V}_{j_1}^T \hat{\mathbf{r}}_{i_1} \hat{\mathbf{r}}_{i_2}^T \hat{V}_{j_2}) \\ =& \sigma^2 \hat{V}_{j_1}^T \hat{V}_{j_2} = 0\end{aligned}$$

which means any two different entries $\widetilde{R}_{i_1 j_1}$ and $\widetilde{R}_{i_2 j_2}$ of $\widetilde{R}$ are independent, where the last equality is due to the fact that $\hat{V}_{j_1}$ and $\hat{V}_{j_2}$ are the orthogonal columns of $\hat{V}$. Consequently, with probability at least $1 - 2\exp(-\epsilon^2/2) - \delta/2$, we have the following bound as

$$\begin{aligned}&\|P_{\hat{U}} \mathbf{e}_i\|_2^2 \\ \leq& \|P_{\widetilde{R}} \mathbf{e}_i\|_2^2 + \|P_{\bar{U}} \mathbf{e}_i\|_2^2 \\ \leq& \|\widetilde{R}\|_2^2 \|(\widetilde{R}^T \widetilde{R})^{-1}\|_2^2 \|\widetilde{R}^T \mathbf{e}_i\|_2^2 + \frac{r\mu(r)}{m} \\ \leq& \frac{(\sqrt{m} + \sqrt{\hat{r}} + \epsilon)^2 \sigma^2}{(\sqrt{m} - \sqrt{\hat{r}} - \epsilon)^4 \sigma^4} \sigma^2 \left(\hat{r} + 2\sqrt{\hat{r} \ln(2/\delta)} + 2\ln(2/\delta)\right) \\ &+ \frac{r\mu(r)}{m}\end{aligned}$$

where $\|\widetilde{R}\|_2^2$ and $\|(\widetilde{R}^T \widetilde{R})^{-1}\|_2^2$ are bounded by Lemma 9, and $\|\widetilde{R}^T \mathbf{e}_i\|_2^2$ is bound by Lemma 7. The last inequality holds with probability at least $1 - \delta$ by setting $\epsilon = \sqrt{2\ln(4/\delta)}$. Note that the fraction $(\sqrt{m}+\sqrt{\hat{r}}+\epsilon)^2/(\sqrt{m}-\sqrt{\hat{r}}-\epsilon)^4$ is approximately $O(1/m)$, when $\hat{r} \leq m/2$ and δ is not exponentially small (e.g., $\sqrt{2\ln(4/\delta)} \leq \frac{\sqrt{m}}{4}$). Then, with probability at least $1 - m\delta$, we have

$$\begin{aligned}\hat{\mu}(\hat{r}) =& \frac{m}{\hat{r}} \max_{i \in [m]} \|P_{\hat{U}} \mathbf{e}_i\|_2^2 \\ \leq& \frac{r\mu(r)}{\hat{r}} + O\left(\frac{\hat{r} + \sqrt{\hat{r}\ln(1/\delta)} + \ln(1/\delta)}{\hat{r}}\right) \\ =& O\left(\frac{r\mu(r) + \hat{r} + \sqrt{\hat{r}\ln(1/\delta)} + \ln(1/\delta)}{\hat{r}}\right) \\ =& O\left(\frac{r\mu(r)\ln(1/\delta)}{\hat{r}}\right).\end{aligned}$$

Setting $\delta = \delta'/m$, we prove (5).

The projected vector $P_{\hat{U}^\perp} \mathbf{m}_i$ can be wrote as $P_{\hat{U}^\perp} \mathbf{m}_i = \hat{\mathbf{c}} + \hat{\mathbf{r}}$, where $\hat{\mathbf{c}} = P_{\hat{U}^\perp} \mathbf{c}$ and $\hat{\mathbf{r}} = P_{\hat{U}^\perp} \mathbf{r}$. By definition, $\hat{\mathbf{c}}$ lies in $\mathrm{span}(\hat{U}^\perp) \cap \mathrm{span}(\bar{U})$, and $\hat{\mathbf{r}}$ lies in $\mathrm{span}(\hat{U}^\perp) \cap \mathrm{span}(\bar{U}^\perp)$ with rank at least $m - r - \hat{r}$. Note that $\hat{\mathbf{r}}$ is still a Gaussian random vector. As a result, with probability at least $1 - \delta$, we have

$$\begin{aligned}&\mu(P_{\hat{U}^\perp} \mathbf{m}_i) \\ =& m \frac{\|\hat{\mathbf{c}} + \hat{\mathbf{r}}\|_\infty^2}{\|\hat{\mathbf{c}} + \hat{\mathbf{r}}\|_2^2} \\ \leq& 3m \frac{\|\hat{\mathbf{c}}\|_\infty^2 + \|\hat{\mathbf{r}}\|_\infty^2}{\|\hat{\mathbf{c}}\|_2^2 + \|\hat{\mathbf{r}}\|_2^2} \leq 3m \frac{\|\hat{\mathbf{c}}\|_\infty^2}{\|\hat{\mathbf{c}}\|_2^2} + 3m \frac{\|\hat{\mathbf{r}}\|_\infty^2}{\|\hat{\mathbf{r}}\|_2^2} \\ \leq& 3r\mu(r) + \frac{6m\sigma^2 \ln(4mn/\delta)}{\sigma^2(m - r - \hat{r}) - 2\sigma^2 \sqrt{(m - r - \hat{r})\ln(2n/\delta)}}\end{aligned}$$

for all partially observed $\mathbf{m}_i$, where $\|\hat{\mathbf{r}}\|_\infty^2$ is bounded by Lemma 8 and $\|\hat{\mathbf{r}}\|_2^2$ is bounded by Lemma 7. Note that when $r \leq m/4$ and $\ln(2n/\delta) \leq m/64$, the denominator

$$\sigma^2(m - r - \hat{r}) - 2\sigma^2 \sqrt{(m - r - \hat{r})\ln(2n/\delta)} \geq \sigma^2 m/4.$$

Subsequently, we have

$$\mu(P_{\hat{U}^\perp} \mathbf{m}_i) \leq 3r\mu(r) + 24\ln(2mn/\delta)$$

for for all partially observed $\mathbf{m}_i$.

A.5 Proof of Lemma 5

Let $\xi_t = d\mathbf{a}_t \mathbf{a}_t^T$, where $\mathbf{a}_t$ is the t-th column of A constructed by Algorithm 1 with uniform sampling. We have

$$\begin{aligned}\|\xi_t\|_F =& \|d\mathbf{a}_t \mathbf{a}_t^T\|_F = n\|\mathbf{m}_{i_t}\|_2^2 \\ \leq& n \max_{i \in [n]} \|\mathbf{m}_i\|_2^2 = \mu(M)\|M\|_F^2\end{aligned}$$

and

$$\mathrm{E}\left[\|\xi_t\|_F^2\right] = n \sum_{i=1}^n \|\mathbf{m}_i\|_2^4 \leq \mu(M)\|M\|_F^4,$$

$$\mathrm{E}[\xi_t] = \sum_{i=1}^n \mathbf{m}_i \mathbf{m}_i^T = MM^T.$$

According to Lemma 11, with a probability $1 - \delta$, we have,

$$\begin{aligned}\|AA^T - MM^T\|_F =& \left\| \frac{1}{d} \sum_{t=1}^d [\xi_t - \mathrm{E}[\xi_t]] \right\|_F \\ \leq& \frac{2\ln(2/\delta)\mu(M)\|M\|_F^2}{d} + \sqrt{\frac{2\ln(2/\delta)\mu(M)\|M\|_F^4}{d}}.\end{aligned}$$

We complete the proof by substituting the above inequality into Lemma 10.