Supplementary Material of "CUR Algorithm for Partially Observed Matrices"

Miao Xu XUM@LAMDA.NJU.EDU.CN
National Key Laboratory for Novel Software Technology, Nanjing University
Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing, China

Rong Jin RONGJIN@CSE.MSU.EDU
Institute of Data Science and Technologies at Alibaba Group, Seattle, USA

Zhi-Hua Zhou ZHOUZH@LAMDA.NJU.EDU.CN
National Key Laboratory for Novel Software Technology, Nanjing University
Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing, China

In the supplementary material, we will prove the theorems in Section 1 and give some additional experiments in Section 2.

1. Analysis

We will first give the supporting theorems we will use in the analysis. Then we will give the detailed proof of the three theorems in the paper.

1.1. Supporting Theorems

The following results are used throughout the analysis.

Theorem 1. *(Theorem 9.1 in (Halko et al., 2011)) Let M be an $n \times m$ matrix with singular value decomposition $M = U\Sigma V^\top$. There is a fixed $r > 0$. Choose a test matrix $\Psi \in \mathbb{R}^{m\times d}$ and construct sample matrix $Y = M\Psi$. Partition M as in (1)*

$$M = U\Sigma V^\top = \begin{matrix} r & m-r \\ [U_1 & U_2] \end{matrix} \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^\top \\ V_2^\top \end{bmatrix} \quad (1)$$

and define $\Psi_1 = V_1^\top \Psi$ and $\Psi_2 = V_2^\top \Psi$. Assuming Ψ_1 has full row rank, the approximation error satisfies

$$\|M - P_Y(M)\|_2^2 \le \|\Sigma_2\|_2^2 + \|\Sigma_2\Psi_2\Psi_1^\dagger\|_2^2$$

where $P_Y(M)$ projects column vectors in M in the subspace spanned by the column vectors in Y and † denotes the pseudoinverse.

Theorem 2. *(Derived From Theorem 2.2 of (Tropp, 2011)) 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 eigen value respectively.*

Suppose that

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

Sample $\{X_1, \ldots, X_\ell\}$ uniformly at random from $\mathcal{X}$ without replacement. Compute

$$\mu_{\max} = \ell\lambda_{\max}(\mathrm{E}[X_1]), \quad \mu_{\min} = \ell\lambda_{\min}(\mathrm{E}[X_1])$$

Then

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

Theorem 3. *Let $A = S^\top HS$ and $\tilde{A} = S^\top \tilde{H} S$ be two symmetric matrices of size $n\times n$. Let $\lambda_i, i \in [n]$ and $\tilde{\lambda}_i, i \in [n]$ be the eigenvalues of A and $\tilde{A}$, respectively, ranked in descending order. Let $U_A, \tilde{U}_A \in \mathbb{R}^{n\times r}$ include the first r eigenvectors of A and $\tilde{A}$, respectively. Let $\|\cdot\|$ be any invariant norm. Define*

$$\begin{aligned} \Delta_\lambda &= \min\left(\sqrt{2}\left(1 - \frac{\lambda_{r+1}}{\lambda_r}\right), \frac{1}{\sqrt{2}}\right) \le \frac{1}{\sqrt{2}} \\ \Delta_H &= \frac{\|H^{-1}\|\|H - \tilde{H}\|}{\sqrt{1 - \|H^{-1}\|\|H-\tilde{H}\|}} \end{aligned}$$

If $\Delta_\lambda \ge \Delta_H/2$, we have

$$\|\sin\Theta(U_A, \tilde{U}_A)\| \le \frac{\Delta_H}{\Delta_\lambda - \Delta_H/2}\left(1 + \frac{\Delta_H \Delta_\lambda}{16}\right)$$

where

$$\begin{aligned} \Theta(X, \tilde{X}) &= \arccos((X^*X)^{-1/2}X^*\tilde{X}(\tilde{X}^*\tilde{X})^{-1} \\ & \quad l\tilde{X}^*X(X^*X)^{-1/2})^{1/2} \end{aligned}$$

defines the angle matrix between X and $\tilde{X}$.

Note that the above Theorem 3 follows directly from Theorem 4.4 and discussion in Section 5 from (Li, 1999).

1.2. Proof of Theorem 2

We will first provide the key result for our analysis, and then bound each component of the key result, that is, first, we will show that $\|M - P_{\widehat{U}} M P_{\widehat{V}}\|_2^2$ is small; then, we will bound the strong convexity of the objective function.

The following theorem shows that the difference between M and $\widehat{M}$ is well bounded if both $\|M - P_{\widehat{U}} M P_{\widehat{V}}\|_2^2$ and the strong convexity of Eq.2 are well bounded,

Theorem 4. *Assume (i) $\|M - P_{\widehat{U}} M P_{\widehat{V}}\|_2^2 \leq \Delta$, and (ii) the strong convexity of the objective function is no less than $|\Omega|\gamma$. Then*

$$\|M - \widehat{M}\|_2^2 \leq 2\left(\Delta + \frac{\Delta}{\gamma}\right).$$

where strongly convexity is defined as,

Definition 5. *A function $f : \mathcal{D} \to \mathbb{R}$ is ξ-strongly convex w.r.t. norm $\|\cdot\|$ if f is everywhere differentiable and*

$$f(\mathbf{w}) \geq f(\mathbf{w}') + \nabla f(\mathbf{w}')(w - w') + \frac{\xi}{2}\|w - w'\|^2.$$

Then ξ is the strongly convexity of f.

Proof. Set $Z = \widehat{U}^\top M \widehat{V}$. Since $\|M - P_{\widehat{U}} M P_{\widehat{V}}\|_2^2 \leq \Delta$, we have

$$\|M - \widehat{U} Z \widehat{V}^\top\|_2^2 \leq \Delta,$$

implying

$$\|\mathcal{R}_\Omega(M) - \mathcal{R}_\Omega(\widehat{U} Z \widehat{V}^\top)\|_F^2 \leq \Delta$$

Let Z_* be the optimal solution to Eq.2. Using the strongly convexity of Eq.2, we have

$$\frac{1}{2}\gamma|\Omega|\|Z - Z_*\|_F^2 \leq \frac{1}{2}|\Omega|\Delta,$$

i.e. $\|Z - Z_*\|_F^2 \leq \Delta/(\gamma|\Omega|)$.

This is because $f(Z) = \frac{1}{2}\|\mathcal{R}_\Omega(M) - \mathcal{R}_\Omega(\widehat{U} Z \widehat{V}^\top)\|_F^2$, such that $\nabla f(Z) = \widehat{U}^T[\mathcal{R}_\Omega(\widehat{U} Z \widehat{V}^T) - \mathcal{R}_\Omega(M)]\widehat{V}$, and $\nabla f(Z_*) = 0$

$$\begin{aligned}
& \frac{|\Omega|\gamma}{2}\|Z - Z_*\|_F^2 \\
\leq\ & \frac{1}{2}\|\mathcal{R}_\Omega(M) - \mathcal{R}_\Omega(\widehat{U} Z \widehat{V}^\top)\|_F^2 - \\
& \frac{1}{2}\|\mathcal{R}_\Omega(M) - \mathcal{R}_\Omega(\widehat{U} Z_* \widehat{V}^\top)\|_F^2 \\
\leq\ & \frac{1}{2}\|\mathcal{R}_\Omega(M) - \mathcal{R}_\Omega(\widehat{U} Z \widehat{V}^\top)\|_F^2 \leq \frac{|\Omega|\Delta}{2}
\end{aligned}$$

We thus have,

$$\begin{aligned}
& \|M - \widehat{M}\|_2^2 \\
\leq\ & l2\|M - P_{\widehat{U}} M P_{\widehat{V}}\|_2^2 + 2\|P_{\widehat{U}} M P_{\widehat{V}} - \widehat{U} Z_* \widehat{V}^\top\|_2^2 \\
\leq\ & 2\|M - P_{\widehat{U}} M P_{\widehat{V}}\|_2^2 + 2\|P_{\widehat{U}} M P_{\widehat{V}} - \widehat{U} Z_* \widehat{V}^\top\|_F^2 \\
\leq\ & 2\|M - P_{\widehat{U}} M P_{\widehat{V}}\|_2^2 + 2\|Z - Z_*\|_F^2 \\
\leq\ & 2\left(\Delta + \frac{\Delta}{\gamma}\right)
\end{aligned}$$

□

In order to bound Δ, we need the following theorem,

Theorem 6. *With a probability $1 - 2e^{-t}$, we have,*

$$\|M - M P_{\widehat{V}}\|_2^2 \leq \sigma_{r+1}^2\left(1 + 2\frac{m}{d}\right)$$

and

$$\|M - P_{\widehat{U}} M\|_2 \leq \sigma_{r+1}^2\left(1 + 2\frac{n}{d}\right)$$

provided that $d \geq 7\mu(r)r(t + \ln r)$.

Proof. Let $i_1, \ldots, i_d$ are the d selected columns. Define $\Psi = (\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_d}) \in R^{m\times d}$, where $\mathbf{e}_i$ is the ith canonical basis. Such that we have $A = M \times \Psi$, that is, A is composed of the d selected columns of M. To utilize Theorem 1, we need to bound the minimum eigenvalue of $\Psi_1\Psi_1^\top$, where $\Psi_1 = V_1^T\Psi \in R^{r\times d}$ is full rank. We have

$$\Psi_1\Psi_1^\top = V_1^\top \Psi\Psi^\top V_1$$

Let $\widetilde{\mathbf{v}}_i^\top, i \in [d]$ be the ith row vector of V_1. We have,

$$\Psi_1\Psi_1^\top = \sum_{j=1}^d \widetilde{\mathbf{v}}_{i_j}\widetilde{\mathbf{v}}_{i_j}^\top$$

It is straightforward to show that

$$\mathrm{E}\left[\Psi_1\Psi_1^\top\right] = \frac{d}{m}I_r$$

and

$$\mathrm{E}\left[\widetilde{\mathbf{v}}_{i_j}\widetilde{\mathbf{v}}_{i_j}^\top\right] = \frac{1}{m}I_r.$$

To bound the minimum eigenvalue of $\Psi_1\Psi_1^\top$, we need Theorem 2, where we first need to bound the maximum eigen value of $\widetilde{\mathbf{v}}_{i_j}\widetilde{\mathbf{v}}_{i_j}^\top$, which is a rank-1 matrix, whose eigen value

$$\begin{aligned}
& \max_{1\leq i\leq m} \lambda_{\max}(\widetilde{\mathbf{v}}_{i_j}\widetilde{\mathbf{v}}_{i_j}^\top) \\
=\ & \max_{1\leq i\leq m} |\widetilde{\mathbf{v}}_i|^2 \leq \mu(r)\frac{r}{m},
\end{aligned}$$

and

$$\lambda_{\max}(\mathrm{E}\left[\widetilde{\mathbf{v}}_{i_j}\widetilde{\mathbf{v}}_{i_j}^\top\right]) = \lambda_{\min}(\mathrm{E}\left[\widetilde{\mathbf{v}}_{i_j}\widetilde{\mathbf{v}}_{i_j}^\top\right]) = \frac{1}{m}$$

Thus, we have,

$$\begin{aligned} & \Pr\left\{\lambda_{\min}(\Psi_1\Psi_1^\top)(1-\delta)\frac{d}{m}\right\} \\ \leq \quad & r\exp\frac{-d/m}{r\mu(r)/m}\left[(1-\rho)\ln(1-\rho)+\rho\right] \\ = \quad & r\exp\frac{-d}{r\mu(r)}\left[(1-\rho)\ln(1-\rho)+\rho\right] \end{aligned}$$

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

$$\begin{aligned} \Pr\left\{\lambda_{\min}(\Psi_1\Psi_1^\top) \leq \frac{d}{2m}\right\} \quad & \leq \quad r\exp\frac{-d}{7r\mu(r)} \\ & = \quad re^{-d/[7\mu(r)r]} \end{aligned}$$

where with $d \geq 7\mu(r)r(t+\ln r)$, we have $r\exp^{-d/[7\mu(r)r]} \leq e^{-t}$, that is,

$$\Pr\left\{\lambda_{\min}(\Psi_1\Psi_1^\top) \geq \frac{d}{2m}\right\} \quad \geq \quad 1-e^{-t}$$

With

$$\lambda_{\min}(\Psi_1\Psi_1^\top) \geq \frac{d}{2m}$$

according to Theorem 1, we have

$$\begin{aligned} \|M - MP_{\widehat{V}}\|_2^2 \quad & \leq \quad \|\Sigma_2\|_2^2 + \|\Sigma_2\Psi_2\Psi_1^\dagger\|_2^2 \\ & \leq \quad \sigma_{r+1}^2 + \left\|\Sigma_2\Psi_2\Psi_1^\dagger\right\|_2^2 \\ & \leq \quad \sigma_{r+1}^2 + \|\Psi_1^\dagger\|_2^2\|\Sigma_2\Psi_2\|_2^2 \\ & \leq \quad \sigma_{r+1}^2 + \frac{2m}{d}\|\Sigma_2\Psi_2\|_2^2 \\ & \leq \quad \sigma_{r+1}^2 + \frac{2m}{d}\|\Sigma_2\|_2^2\|\Psi_2\|_2^2 \\ & \leq \quad \sigma_{r+1}^2 + \frac{2m}{d}\sigma_{r+1}^2 \\ & \leq \quad \sigma_{r+1}^2\left(1+\frac{2m}{d}\right) \end{aligned}$$

- The 1st inequality is according to Theorem 1.
- The 3rd inequality is because the two facts, $\|M_1M_2\|_2 \leq \|M_1\|_2 \times \|M_2\|_2$
- The 4th inequality is becuase $\|\Psi_1^\dagger\|_2 = 1/\sigma_{\min}(\Psi_1) = \sqrt{1/\lambda_{\min}(\Psi_1\Psi_1^\top)} \leq \sqrt{2m/d}$
- The 6th inequality is because $\|\Sigma_2\|_2 = \sigma_{r+1}$ and $\|\Psi_2\|_2 \leq \|V_2\|_2\|\Psi\|_2 = 1$

□

We then bound Δ,

Theorem 7. *With a probability $1-2e^{-t}$, we have,*

$$\Delta := \|M - P_{\widehat{U}}MP_{\widehat{V}}\|_2^2 \leq 4\sigma_{r+1}^2\left(1+\frac{m+n}{d}\right)$$

if $d \geq 7\mu(r)r(t+\ln r)$.

Proof. Using Theorem 6, we have, with a probability $1-2e^{-t}$

$$\begin{aligned} & \|M - P_{\widehat{U}}MP_{\widehat{V}}\|_2^2 \\ \leq \quad & 2\|M - MP_{\widehat{V}}\|_2^2 + 2\|(M-P_{\widehat{U}}M)P_{\widehat{V}}\|_2^2 \\ \leq \quad & 2\|M - MP_{\widehat{V}}\|_2^2 + 2\|M-P_{\widehat{U}}M\|_2^2 \\ \leq \quad & 4\sigma_{r+1}^2\left(1+\frac{n+m}{d}\right) \end{aligned}$$

□

We will then bound the strong convexity of the objective function,

Theorem 8. *With a probability $1-e^{-t}$, we have that $\gamma|\Omega|$, the strongly convexity for the objective function in (2), is bounded from below by $|\Omega|/[2mn]$ (that is, $\gamma \geq 1/(2mn)$), provided that*

$$|\Omega| \geq 7\widehat{\mu}^2(r)r^2(t+2\ln r)$$

Proof. To bound the strong convexity, we could instead bound the smallest eigen value of the Hessian matrix. The Hessian matrix is an $r^2 \times r^2$ matrix. Assuming the second-order derivative of the (i_1, j_1)th and (i_2, j_2)th entry of Z is the $(r(i_1-1)+j_1, r(i_2-1)+j_2)$th entry of the Hessian matrix, the Hessian matrix could be written as,

$$H = \sum_{(i,j)\in\Omega}[\text{vec}(\widetilde{\mathbf{u}}_i^\top\widetilde{\mathbf{v}}_j)][\text{vec}(\widetilde{\mathbf{u}}_i^\top\widetilde{\mathbf{v}}_j)]^T$$

To bound the minimum eigenvalue of H, we will use Lemma 2. Thus first we need to bound

$$\begin{aligned} & \max_{i,j}\lambda_{\max}([\text{vec}(\widetilde{\mathbf{u}}_i^\top\widetilde{\mathbf{v}}_j)][\text{vec}(\widetilde{\mathbf{u}}_i^\top\widetilde{\mathbf{v}}_j)]^T) \\ = \quad & \max_{i,j}|\text{vec}(\widetilde{\mathbf{u}}_i^\top\widetilde{\mathbf{v}}_j)|^2 \\ \leq \quad & \max\|\widetilde{\mathbf{u}}_i^\top\widetilde{\mathbf{v}}_j\|_F^2 \leq \frac{\widehat{\mu}^2(r)r^2}{mn} \end{aligned}$$

and

$$\begin{aligned} & \lambda_{\min}\left(\text{E}([\text{vec}(\widetilde{\mathbf{u}}_i^\top\widetilde{\mathbf{v}}_j)][\text{vec}(\widetilde{\mathbf{u}}_i^\top\widetilde{\mathbf{v}}_j)]^T)\right) \\ = \quad & \frac{1}{mn}\lambda_{\min}\left((U\otimes V)^T \times (U\otimes V)\right) \\ = \quad & \frac{1}{mn} \end{aligned}$$

where $\otimes$ is the Kronecker product.

Based on Theorem 2, we have

$$\Pr\left\{\lambda_{\min}(H) \leq \frac{|\Omega|}{2mn}\right\} \quad \leq \quad r^2 e^{\frac{-|\Omega|}{7\widehat{\mu}^2(r)r^2}}$$

Hence, with a probability $1 - e^{-t}$, we have

$$\lambda_{\min}(H) \geq \frac{|\Omega|}{2mn}$$

provided that

$$|\Omega| \geq 7\widehat{\mu}^2(r)r^2(t + 2\ln r)$$

□

Theorem 2 can be easily proved combining Theorems 4, 7 and 8.

1.3. Proof of Theorem 1

The following theorem allows us to replace $\widehat{\mu}(r)$ in Theorem 8 with $\mu(r)$ when the rank of M is less than or equal to r.

Theorem 9. *With a probability* $1 - 2e^{-t}$*, we have* $\widehat{\mu}(r) = \mu(r)$*, if* $d \geq 7\mu(r)r(t + \ln r)$.

Proof. According to Theorem 7, with a probability $1 - 2e^{-t}$, we have $M = P_{\widehat{U}} M P_{\widehat{V}}$, provided that $d \geq 7\mu(r)r(t + \ln r)$. Hence $P_{U_1} = P_{\widehat{U}}$ and $P_{V_1} = P_{\widehat{V}}$, which directly implies that $\mu(r) = \widehat{\mu}(r)$. □

Theorem 1 can be proved directly from Theorem 2 and Theorem 9.

1.4. Proof of Theorem 3

Define

$$H_A = \eta I + \frac{1}{mn} M M^\top, \quad \widehat{H}_A = \eta I + \frac{1}{dn} A A^\top$$

and

$$H_B = \eta I + \frac{1}{mn} M^\top M, \quad \widehat{H}_B = \eta I + \frac{1}{dm} B B^\top$$

We can have the first r eigen vector of would be H_A, because

$$\begin{aligned} H_A &= \eta I + \frac{1}{mn} M M^\top \\ &= \eta U U^T + \frac{1}{mn} U(\Sigma \Sigma^T) U^T \\ &= U(\eta I + \frac{1}{mn} \Sigma \Sigma^T) U^T \end{aligned}$$

and

$$\begin{aligned} H_A^{-1/2} &= lU diag(\sqrt{\frac{mn}{\sigma_1^2 + mn\eta}}, \ldots, \sqrt{\frac{mn}{\sigma_m^2 + mn\eta}}) \\ &= \sqrt{mn} U T U^T \end{aligned}$$

where

$$T = diag(\sqrt{\frac{1}{\sigma_1^2 + mn\eta}}, \ldots, \sqrt{\frac{1}{\sigma_m^2 + mn\eta}})$$

1.4.1. Proof of Lemma 1

Proof. Just consider the maximization of the norm of rows of U, then we will have

$$\begin{aligned} \mu(\eta) &= \max_{i=1,\ldots,n} \sum_{j=1}^m \frac{n}{r(M,\eta)} \frac{\sigma_j^2}{\sigma_j^2 + mn\eta} U_{i,j}^2 \\ &= \max_{i=1,\ldots,n} \frac{n}{r} \sum_{j=1}^m r \frac{\sigma_j^2}{r(M,\eta)(\sigma_j^2 + mn\eta)} U_{i,j}^2 \\ &\geq \max_{i=1,\ldots,n} \frac{n}{r} \sum_{j=1}^m r \frac{a}{r} U_{i,j}^2 \\ &= a \max_{i=1,\ldots,n} \frac{n}{r} \sum_{j=1}^m U_{i,j}^2 \\ &= a\mu(r) \end{aligned}$$

when $\eta = \sigma_r^2/mn$, then $a \leq r/2r(M,\eta)$, then

$$\mu(r) \leq \frac{1}{a}\mu(\delta) \leq \frac{2r(M,\eta)}{r}\mu(\eta)$$

completes our proof.

□

1.4.2. Proof of Lemma 2

To this end, we need the following theorem.

Theorem 10. *With a probability* $1 - 4e^{-t}$*, we have*

$$\begin{aligned} &1 - \delta \leq \lambda_k(H_A^{-1/2} \widehat{H}_A H_A^{-1/2}) \leq 1 + \delta, \\ &l1 - \delta \leq \lambda_k(H_B^{-1/2} \widehat{H}_B H_B^{-1/2}) \leq 1 + \delta, \ \forall k \in [n] \end{aligned}$$

if

$$d \geq \frac{4}{\delta^2}(\mu(\eta) r(M,\eta) + 1)(t + \ln n)$$

Proof. It is sufficient to show the result for $\widehat{H}_A$.

Define

$$\mathcal{X} = \left\{ M_i = (H_A^{-1/2})^T \left(\frac{1}{n} M_{*,i} M_{*,i}^\top + \eta I\right) H_A^{-1/2}, \quad i = 1, \ldots, m \right\}$$

Note that if $\mathbf{a}_i$ is the jth column of matrix M, then,

$$M_{*,i} = U\Sigma(V_{i,*})^\top$$

Thus we have

$$\begin{aligned} M_i &= mnUTU^\top(\frac{1}{n}U\Sigma V_{i,*}^\top V_{i,*}\Sigma U^\top + \eta I)UTU^\top \\ &= U\left(mT\Sigma V_{i,*}^\top V_{i,*}\Sigma T + mn\eta T^2\right)U^\top \end{aligned}$$

In this way

$$\begin{aligned} \lambda_{\max}(M_i) &\leq \lambda_{\max}(mUT\Sigma V_{i,*}^\top V_{i,*}\Sigma TU^\top) \\ &\quad l + \lambda_{\max}(mn\eta UT^2U^\top) \\ &= m|UT\Sigma V_{i,*}^\top|_2^2 + \frac{mn\eta}{\sigma_m^2 + mn\eta} \\ &\leq \mu(\eta)r(M,\eta) + 1 \end{aligned}$$

(this is because $|Ax|_2^2 \leq \|A\|_2^2|x|_2^2 \leq \|A\|_F^2|x|_2^2$) and

$$\begin{aligned} &\lambda_{\max}(\mathrm{E}[M_i]) \\ =\ &\lambda_{\max}(U\left(T\Sigma V^\top V\Sigma T + mn\eta T^2\right)U^\top) \\ =\ &\lambda_{\max}(U\left(T\Sigma\Sigma T + mn\eta T^2\right)U^\top) \\ =\ &\frac{\sigma_1^2}{mn\eta + \sigma_1^2} + \frac{mn\eta}{mn\eta + \sigma_1^2} \\ =\ &1 \end{aligned}$$

So

$$\mu_{\max} = d\lambda_1(\mathrm{E}[M_i]) = d$$

we have (using Lemma 2),

$$\begin{aligned} &\Pr\left\{\lambda_{\max}\left(H_A^{-1/2}\widehat{H}_A H_A^{-1/2}\right) \geq 1+\delta\right\} \\ \leq\ & n\exp\left(-\frac{d}{\mu(\eta)r(M,\eta)+1}\left[(1+\delta)\ln(1+\delta)-\delta\right]\right) \end{aligned}$$

Using the fact that (at 0 they are the same, but the left increase faster than the right)

$$(1+\delta)\ln(1+\delta) \geq \delta + \frac{1}{4}\delta^2, \forall\delta\in[0,1],$$

we have

$$\begin{aligned} &\Pr\left\{\lambda_{\max}\left(H_A^{-1/2}\widehat{H}_A H_A^{-1/2}\right) \geq 1+\delta\right\} \\ \leq\ & n\exp\left(-\frac{d\delta^2}{4(\mu r(M,\eta)+1)}\right) \end{aligned}$$

We have the result by setting $d \geq 4(\mu(\eta)r(M,\eta)+1)(\ln n + t)/\delta^2$. Similarly, for the lower bound, we have (using Lemma 2)

$$\begin{aligned} &\Pr\left\{\lambda_{\min}\left(H_A^{-1/2}\widehat{H}_A H_A^{-1/2}\right) \leq 1-\delta\right\} \\ \leq\ & n\exp\left(-\frac{d}{\mu(\eta)r(M,\eta)+1}\left[(1-\delta)\ln(1-\delta)+\delta\right]\right) \end{aligned}$$

Using the fact that (by Taylor Expansion of $\ln(1-\delta)$)

$$(1-\delta)\ln(1-\delta) \geq -\delta + \frac{\delta^2}{2}$$

We have the result by setting $d \geq 2(\mu(\eta)r(M,\eta)+1)(\ln n + t)/\delta^2$. □

Using Theorem 10, we will prove Lemma 2,

Proof. To utilize Theorem 3, we rewrite H_A and $\widehat{H}_A$, as

$$H_A = H_A^{1/2}IH_A, \quad \widehat{H}_A = H_A^{1/2}DH_A^{1/2}$$

where $D = H_A^{-1/2}\widehat{H}_A H_A^{-1/2}$. According to Theorem 10, with a probability $1-2e^{-t}$, we have $\|D-I\|_2 \leq \delta$, provided that

$$d = \frac{4}{\delta^2}(\mu(\eta)r(M,\eta)+1)(t+\ln n)$$

We then compute Δ_H defined in Theorem 3 as

$$\Delta_H \leq \frac{\delta}{\sqrt{1-\delta}}$$

Because $d \geq 16(\mu(\eta)r(M,\eta)+1)(t+\ln n)$, we have

$$\begin{aligned} \frac{4}{\delta^2}(\mu(\eta)r(M,\eta)+1)(t+\ln n) \geq 16(\mu(\eta)r(M,\eta)+1) \\ l(t+\ln n) \end{aligned}$$

that is $\delta \leq 1/2$.

Because $\sigma_r \geq \sqrt{2}\sigma_{r+1}$, we have $1/2 \leq 1-\sigma_{r+1}^2/\sigma_r^2$. Since $\delta \leq 1/2 \leq 1-\sigma_{r+1}^2/\sigma_r^2$, we have $\Delta_H \leq \sqrt{2}\delta$.

Then according to Theorem 3, we have,

$$\begin{aligned} \|\sin\Theta(U_1,\widehat{U})\|_2 &\leq \frac{\sqrt{2}\delta}{\Delta_\lambda - \sqrt{2}\delta/2}(1+\frac{\sqrt{2}\delta\Delta_\lambda}{16}) \\ &\leq \frac{\sqrt{2}\delta}{\Delta_\lambda - \sqrt{2}\delta/2}(1+\frac{1}{32}) < 3\sqrt{2}\delta \end{aligned}$$

Similarly, we have,

$$\|\sin\Theta(V_1,\widehat{V})\|_2 < 3\sqrt{2}\delta$$

Thus, with a probability $1-4e^{-t}$, we have

$$\begin{aligned} &\widehat{\mu}(r) \leq \frac{2r(M,\eta)}{r}\mu(\eta) + \frac{n}{r}\|\sin\Theta(V_1,\widehat{V})\|_2^2 \\ \leq\ & l\frac{2r(M,\eta)}{r}\mu(\eta) + \frac{18n\delta^2}{r} \end{aligned}$$

□

Theorem 3 can be proved by combining the results of Theorems 4, 8, Lemma 1 and Lemma 2.

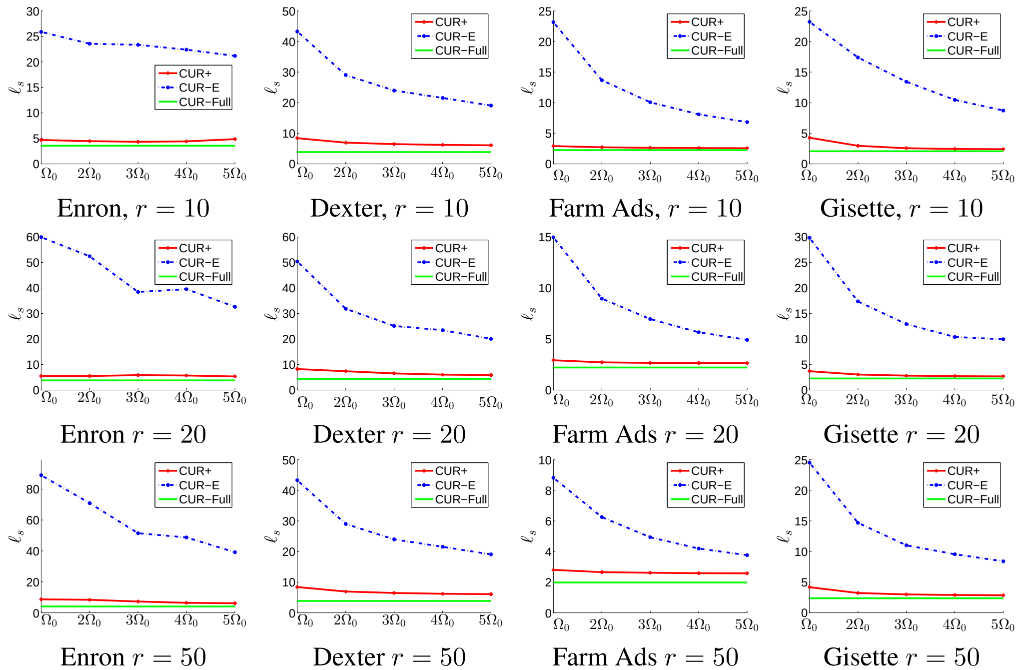

Figure 1. Comparison of CUR algorithms with the number of sampled columns (rows) fixed as $d_1 = 5r$ ($d_2 = 5d_1$), where $r = 10, 20, 50$. The number of observed entries $|\Omega|$ is varied from Ω_0 to $5\Omega_0$.

2. Additional Experiments

Comparison to the State-of-the-Art Low-Rank Approximation Algorithms The experimental settings are the same as that of Section 4.2, and we will compare here with CUR-E. We first construct an unbiased estimator M_e by using the randomly observed entries in Ω, and then estimate matrix Z by $Z = C^\dagger M_e R^\dagger$. Here, the unbiased estimation M_e is given by

$$[M_e]_{i,j} = \begin{cases} \frac{mn}{|\Omega|} M_{i,j} & (i,j) \in \Omega \\ 0 & (i,j) \notin \Omega \end{cases}$$

We call this algorithm **CUR-E**.

Results Figure 1 shows the results on spectral norm of low rank matrix approximation. We observe that the CUR+ works significantly better than the CUR-E method, and yields a similar performance as the CUR-F that has an access to the full target matrix M. We also observe that with larger α (i.e. increasing numbers of rows and columns), the approximation errors for CUR+ and CUR-E decrease. Figure 2 shows the results on Frobenius norm. We can see similar results as those on spectral norm.

References

Halko, N., Martinsson, P.-G., and Tropp, J. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.

Li, R.-C. Relative perturbation theory: (II) eigenspace and singular subspace variations. *SIAM J. Matrix Anal. Appl.*, 20:471–492, 1999.

Tropp, J. Improved analysis of the subsampled randomized hadamard transform. *Adv. Adapt. Data Anal*, 3:115–126, 2011.

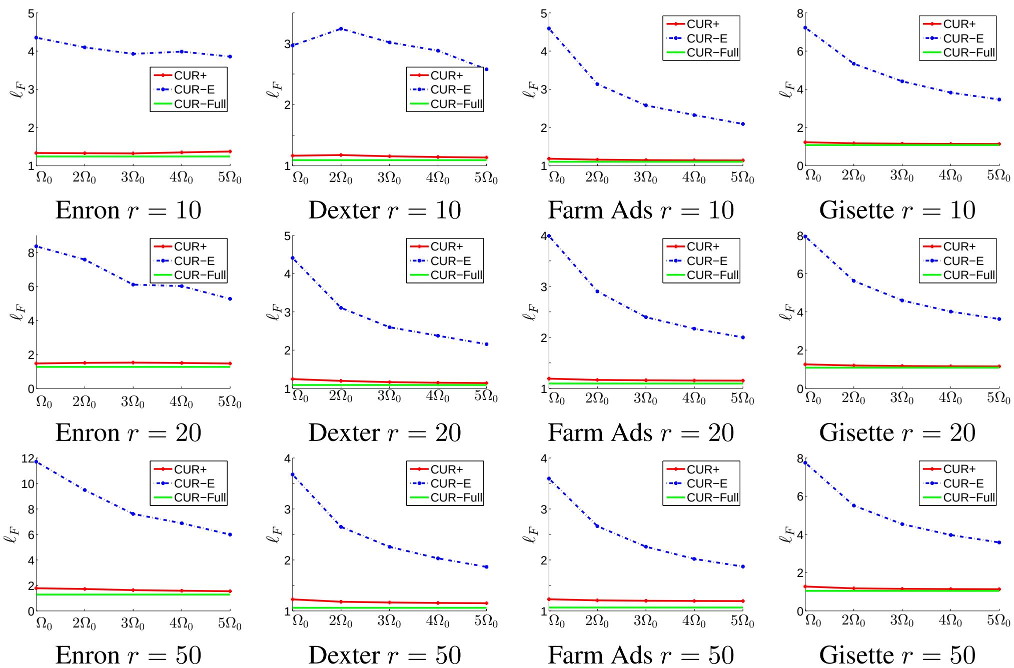

Figure 2. Comparison of CUR algorithms measured by Forbenius norm with the number of sampled columns and rows fixed as $d_1 = 5r$ and $d_2 = 5d_1$, respectively, where $r = 10, 20$ and 50. The number of observed entries $|\Omega|$ is varied from Ω_0 to $5\Omega_0$.