Supplementary Material for Submission 48

January 21, 2024

Lemma 2: Given two transition matrices T and $\tilde{T}$ of the same size $n \times n$,

$$\|T^k - \tilde{T}^k\|_F \leq k \cdot \sqrt{n} \cdot \|T - \tilde{T}\|_F.$$

Proof. First, we need to prove the claim that if T is a transition matrix, then $\forall k > 0$, T^k is also a transition matrix satisfying that each element is between 0 and 1 and the row sum is 1. A simple proof using Mathematical Induction.

The claim holds when $k' = 1$. Supposed that when $k' = k-1$ the claim holds, consider $T^k = T^{k-1} \cdot T$, $\forall 0 \leq i, j \leq n, T^k_{i,j} = \sum_{a=1}^n T^{k-1}_{i,a} \cdot T_{a,j}$, so $T^k_{i,j} > 0$. For row i, $\sum_{j=1}^n T^{k-1}_{i,j} = 1$ and $\sum_{j=1}^n T_{i,j} = 1$ .Next , calculate the row sum of T^k.

$$\begin{aligned}
&\sum_{j=1}^n T^k_{i,j} = \sum_{j=1}^n [\sum_{a=1}^n T^{k-1}_{i,a} \cdot T_{a,j}] \\
&= \sum_{j=1}^n [T^{k-1}_{i,1} \cdot T_{1,j} + T^{k-1}_{i,2} \cdot T_{2,j} \cdots + T^{k-1}_{i,n} \cdot T_{n,j}] \\
&= [T^{k-1}_{i,1} \cdot T_{1,1} + T^{k-1}_{i,2} \cdot T_{2,1} \cdots + T^{k-1}_{i,n} \cdot T_{n,1}] + [T^{k-1}_{i,1} \cdot T_{1,2} + T^{k-1}_{i,2} \cdot T_{2,2} \\
&+ \cdots + T^{k-1}_{i,n} \cdot T_{n,2}] \cdots + [T^{k-1}_{i,1} \cdot T_{1,n} + T^{k-1}_{i,2} \cdot T_{2,n} + \cdots + T^{k-1}_{i,n} \cdot T_{n,n}] \\
&= \sum_{a=1}^n T^{k-1}_{i,a} \cdot (T_{1,1} + T_{1,2} \cdots + T_{1,n}) \\
&= \sum_{a=1}^n T^{k-1}_{i,a} \cdot 1 = 1
\end{aligned}$$

Moreover, because each element of T is larger than or equal to 0 and row sum of T is 1, each element of T^k is less than or equal to 1.

Next, prove that Forbenius norm of T satisfies $\|T\|_F \leq \sqrt{n}$, so $\|T^k\|_F \leq \sqrt{n}$.

Because $\sum_{j=1} |T_{i,j}| = 1$, when $0 \leq T_{i,j} \leq 1$, $\sum_{j=1} |T_{i,j}|^2 \leq (\sum_{j=1} |T_{i,j}|)^2 = 1$. Therefore,

$$\|T\|_F = \sqrt{\sum_{i=1} \sum_{j=1} |T_{i,j}|^2} \leq \sqrt{\sum_{i=1} 1} = \sqrt{n}, \Longrightarrow \|T^k\|_F \leq \sqrt{n}$$

.

At last, prove the lemma. Because of $\|T^k\|_F \le \sqrt{n}$, we have

$$\begin{aligned}
&\|T^k - \tilde{T}^k\|_F \\
=& \|(\tilde{T} - T)(\tilde{T}^{k-1} + \tilde{T}^{k-2} \cdot T + \cdots + \tilde{T} \cdot T^{k-2} + T^{k-1})\|_F \\
\le& \|(\tilde{T} - T)\|_F \|(\tilde{T}^{k-1} + \tilde{T}^{k-2} \cdot T + \cdots + \tilde{T} \cdot T^{k-2} + T^{k-1})\|_F \\
\le& \|\tilde{T} - T\|_F \cdot k \cdot \max\{\|T^k\|_F, \|\tilde{T}^k\|_F\} \\
\le& \|\tilde{T} - T\|_F \cdot k \cdot \sqrt{n}
\end{aligned}$$

□

Lemma 3: Given a graph G, let T be its transition matrix. Given $\rho(\le rank(T))$, let T_ρ be the optimal ρ-rank approximation of T. If algorithm CUR-Trans sets the number of sampled columns and rows as $c = r = O(\rho \cdot \log(\rho))$, we have

$$\Pr\left(\|T - CUR\|_F \le (1+\phi)\|T - T_\rho\|_F\right) \ge 0.81, \tag{1}$$

where $\phi = \varphi^2 + 2\varphi$, $\varphi = \sqrt{\rho \cdot c} + \sqrt{c}$.

In order to proof Lemma 3, we need to proof some claims at first.

Claim 1: For $T \in \mathbf{R}^{n\times n}$, $C \in \mathbf{R}^{n\times c}$ is constructed by sampling c columns from T without replacement using the probability of equation 2. Then, sample rows from C without replacement using the probability of equation 3. Let S be a diagonal matrix denoted the sampled rows, if row i is sampled, the $S_{i,i} = 1$, else $S_{i,i} = 0$. Perform SVD decomposition on C and choose the top ρ singular vectors and values $C_\rho = U_{C,\rho}\Sigma_{C,\rho}V_{C,\rho}^T$.

For $c = r = O(100\rho\cdot\log(\rho))$, suppose $\tilde{\rho} = rank\left(SU_{C,\rho}\right)$, then with probability at least 0.9, we have $\rho = \tilde{\rho} = rank\left(U_{C,\rho}\right) = rank(C_\rho)$.

Moreover, $(SC_\rho)^+ = V_{C,\rho}\Sigma_{C,\rho}^{-1}(SU_{C,\rho})^+$, where A^+ is Moore–Penrose pseudoinverse of matrix A.

$$p_j^c = \frac{\sum_{i=1}^n |T_{i,j}|^2}{\sum_{j=1}^n \sum_{i=1}^n \left|T_{i,j}\right|^2} \tag{2}$$

$$p_i^r = \frac{\sum_{j=1}^n |T_{i,j}|^2}{\sum_{i=1}^n \sum_{j=1}^n \left|T_{i,j}\right|^2} \tag{3}$$

Proof. D is a diagonal matrix denoted the scaling of sampled rows, if row i is sampled, $D_{i,i} = \frac{1}{\sqrt{rp_i}}$, else $D_{i,i} = 0$. Let $\mathbb{S} = D \cdot S$. We have

$$\begin{aligned}
|1 - \sigma_i^2(\mathbb{S}U_{C,\rho})| &= |\sigma_i(U_{C,\rho}^T U_{C,\rho}) - \sigma_i(U_{C,\rho}^T \mathbb{S}^T \mathbb{S} U_{C,\rho})| && (4)\\
&\le \|U_{C,\rho}^T U_{C,\rho} - U_{C,\rho}^T \mathbb{S}^T \mathbb{S} U_{C,\rho}\|_2 && (5)\\
&\le 10\mathbf{E}[\|U_{C,\rho}^T U_{C,\rho} - U_{C,\rho}^T \mathbb{S}^T \mathbb{S} U_{C,\rho}\|_2] && (6)\\
&\le 10O(1)\sqrt{\frac{\log(r)}{r}}\|U_{C,\rho}^T U_{C,\rho}\|_F \|U_{C,\rho}^T U_{C,\rho}\|_2 && (7)\\
&= O(1)10\sqrt{\frac{\log(r)}{r}}\sqrt{\rho} && (8)
\end{aligned}$$

Equation (4) follows from properties of SVD $U_{C,\rho}^T U_{C,\rho} = I$, and $\sigma_i^2(SU_{C,\rho}) = \sigma_i(U_{C,\rho}^T S^T SU_{C,\rho})$. Equation (5) follows from $|\sigma(C+E) - \sigma(C)| \leq \|E\|_2$ [2]. Equation (6) follows from the Markov's inequity [3], which is that given $a > 0$, $P(X \geq a) \leq \frac{E(X)}{a}$. Then, we use it and give a probability $\frac{E(X)}{a} = 0.9$, so $a = 10\mathbf{E}[\|U_{C,\rho}^T U_{C,\rho} - U_{C,\rho}^T \mathbb{S}^T \mathbb{S} U_{C,\rho}\|_F]$. Equation (7) follows because of Theorem 7 of work [1]. Equation (8) follows because of $\|U_{C,\rho}^T U_{C,\rho}\|_F^2 = \rho$ and $\|U_{C,\rho}^T U_{C,\rho}\|_2 = 1$.

When use the $r = O(\rho \cdot \log(\rho))$ combined with section6.3.5 and Lemma 1 in previous work [1], we have $|1 - \sigma_i^2(\mathbb{S}U_{C,\rho})| < 1$ the same. D is a linear transformation matrix for $SU_{C,\rho}$, which will not change the rank. Therefore, $rank(\mathbb{S}U_{C,\rho}) = rank\left(SU_{C,\rho}\right)$, and with at least 0.9 probability, we have $rank(\mathbb{S}U_{C,\rho}) = rank\left(U_{C,\rho}\right) = rank(C_\rho) = \rho = \tilde{\rho} = rank\left(SU_{C,\rho}\right)$.

Then, prove $(SC_\rho)^+ = V_{C,\rho}\Sigma_{C,\rho}^{-1}(SU_{C,\rho})^+$.

$$
\begin{aligned}
(SC_\rho)^+ &= (SU_{C,\rho}\Sigma_{C,\rho}V_{C,\rho}^T)^+ \\
&= (U_{SU_{C,\rho}}\Sigma_{SU_{C,\rho}}V_{SU_{C,\rho}}^T\Sigma_{C,\rho}V_{C,\rho}^T)^+ \\
&= V_{C,\rho}(\Sigma_{SU_{C,\rho}}V_{SU_{C,\rho}}^T\Sigma_{C,\rho})^+U_{SU_{C,\rho}}^T \\
&= V_{C,\rho}(\Sigma_{SU_{C,\rho}}V_{SU_{C,\rho}}^T\Sigma_{C,\rho})^{-1}U_{SU_{C,\rho}}^T \qquad (9) \\
&= V_{C,\rho}\Sigma_{C,\rho}^{-1}V_{SU_{C,\rho}}\Sigma_{SU_{C,\rho}}^{-1}U_{SU_{C,\rho}}^T \qquad (10) \\
&= V_{C,\rho}\Sigma_{C,\rho}^{-1}(SU_{C,\rho})^+
\end{aligned}
$$

Equation (9) follows since $\rho = \tilde{\rho}$ with at least probability 0.9, all three matrices $\Sigma_{SU_{C,\rho}}$, $V_{SU_{C,\rho}}$, and $\Sigma_{C,\rho}$ are full rank square $\rho \times \rho$ matrices, they are invertible. When the matrix is invertible, the pseudo-inverse is equal to the inverse. If A,B,C are invertible matrices, then $(ABC)^{-1} = C^{-1}B^{-1}A{-}1$, so equation (10) follows. □

Claim 2: Define $\Omega = (SU_{C,\rho})^+ - (SU_{C,\rho})^T$, then

$$\|\Omega\|_2 = \|\Sigma_{SU_{C,\rho}}^{-1} - \Sigma_{SU_{C,\rho}}\|_2 \leq 1.$$

Proof.

$$
\begin{aligned}
\|\Omega\|_2 &= \|(SU_{C,\rho})^+ - (SU_{C,\rho})^T\|_2 \\
&= \|(U_{SU_{C,\rho}}\Sigma_{SU_{C,\rho}}V_{SU_{C,\rho}}^T)^+ - (U_{SU_{C,\rho}}\Sigma_{SU_{C,\rho}}V_{SU_{C,\rho}}^T)^T\|_2 \\
&= \|V_{SU_{C,\rho}}(\Sigma_{SU_{C,\rho}}^{-1} - \Sigma_{SU_{C,\rho}})U_{SU_{C,\rho}}^T\|_2 \\
&= \|\Sigma_{SU_{C,\rho}}^{-1} - \Sigma_{SU_{C,\rho}}\|_2 \\
&= \max_{i,j\in[1,\rho]} |\sigma_i(SU_{C,\rho}) - \frac{1}{\sigma_j(SU_{C,\rho})}|
\end{aligned}
$$

With $\sigma_i(SU_{C,\rho}) \geq 0$, therefore, $\sigma_i(SU_{C,\rho}) \in [0,1]$. Use the SVD decomposition of $SU_{C,\rho}$, and $V_{SU_{C,\rho}}$, $U_{SU_{C,\rho}}$ are matrices with orthonormal columns. We have

$$\|\Omega\|_2 = \max_{i,j\in[1,\rho]} |\sigma_i(SU_{C,\rho}) - \frac{1}{\sigma_j(SU_{C,\rho})} \leq |0 - 1| = 1.$$

□

Claim 3: $U_{C,\rho}$ is an $n\times\rho$ matrix contained the top ρ left singular vectors of C. $U^{\perp}_{C,\rho}$ is an $n\times(n-\rho)$ contained the bottom $r-\rho$ non-zero left singular vectors of C. $V^{\perp}_{C,\rho}$ and $\Sigma^{\perp}_{C,\rho}$ are defined by the same way. We have that

$$\|U^T_{C,\rho}S^TSU^{\perp}_{C,\rho}(U^{\perp}_{C,\rho})^TT\|_F \le \sqrt{r\rho}\|U^{\perp}_{C,\rho}(U^{\perp}_{C,\rho})^TT\|_F.$$

Proof.

$$\begin{aligned}
&\|U^T_{C,\rho}S^TSU^{\perp}_{C,\rho}(U^{\perp}_{C,\rho})^TT\|_F\\
&= \|U_{C,\rho}U^T_{C,\rho}S^TSU^{\perp}_{C,\rho}(U^{\perp}_{C,\rho})^TT\|_F \qquad (11)\\
&\le \|S^TS\|_F\|U_{C,\rho}U^T_{C,\rho}\|_F\|U^{\perp}_{C,\rho}(U^{\perp}_{C,\rho})^TT\|_F\\
&= \sqrt{r}\sqrt{\rho}\|U^{\perp}_{C,\rho}(U^{\perp}_{C,\rho})^TT\|_F
\end{aligned}$$

Equation (11) follows since $U_{C,\rho}$ is an orthogonal matrix $U_{C,\rho}U^T_{C,\rho} = U^T_{C,\rho}U_{C,\rho} = I_\rho$. □

Claim 4:

$$\begin{aligned}
\|SU^{\perp}_{C,\rho}(U^{\perp}_{C,\rho})^TT\|_F &\le \|S\|_F\|U^{\perp}_{C,\rho}(U^{\perp}_{C,\rho})^TT\|_F\\
&= \sqrt{r}\|U^{\perp}_{C,\rho}(U^{\perp}_{C,\rho})^TT\|_F.
\end{aligned}$$

Claim 5: For $T\in\mathbf{R}^{n\times n}$ and $C\in\mathbf{R}^{n\times c}$, there is the bound for minimize the $\underset{X}{min}\|T-CX\|$. $\|T-CX\|$ means T project on the subspace formed by C. For $r=O(100\rho\cdot\log(\rho))$, given a probability 0.9, we have

$$\Pr\left(\|T-C(SC)^+ST\| \le (1+\varphi)\|T-CC^+T\|\right) \ge 0.9,$$

where $\varphi=\sqrt{\rho r}+\sqrt{r}$. The same for $\underset{X}{min}\|T-XC\|$, the bound will be

$$\Pr\left(\|T-ST(SC)^+C\| \le (1+\varphi)\|T-TC^+C\|\right) \ge 0.9.$$

Proof. Suppose rank of C is μ , and the SVD of C is $C_{n\times c} = U_C\Sigma_CV^T_C = C_{n\times c} = U_{C,\mu}\Sigma_{C,\mu}V^T_{C,\mu}$. Given $\rho\le\mu$, $C\approx C_\rho = U_{C,\rho}\Sigma_{C,\rho}V^T_{C,\rho}$.

$X=\underset{Y}{min}\|T-CY\|_F = C^+T$, $\tilde{X}=(SC)^+(ST)$. Find the bound resulting from this approximation by calculating the distance between $\|T-CX\|$ and $\|T-C\tilde{X}\|$.

$$\begin{aligned}
T-C_\rho\tilde{X} &= T-C_\rho(SC_\rho)^+(ST)\\
&= T-U_{C,\rho}(SU_{C,\rho})^+(ST) \qquad (12)\\
&= T-U_{C,\rho}(SU_{C,\rho})^+SU_{C,\rho}U^T_{C,\rho}T - U_{C,\rho}(SU_{C,\rho})^+SU^{\perp}_{C,\rho}(U^{\perp}_{C,\rho})^TT \qquad (13)\\
&= U^{\perp}_{C,\rho}(U^{\perp}_{C,\rho})^TT - U_{C,\rho}(SU_{C,\rho})^+SU^{\perp}_{C,\rho}(U^{\perp}_{C,\rho})^TT \qquad (14)\\
&= U^{\perp}_{C,\rho}(U^{\perp}_{C,\rho})^TT + U_{C,\rho}(SU_{C,\rho})^TSU^{\perp}_{C,\rho}(U^{\perp}_{C,\rho})^TT + U_{C,\rho}\Omega SU^{\perp}_{C,\rho}U^{\perp}_{C,\rho}T.
\end{aligned}$$

Equation (12) follows from claim 1 with probability 0.9. Equation (13) follows by inserting $U_{C,\rho}U_{C,\rho}^T + U_{C,\rho}^{\perp}(U_{C,\rho}^{\perp})^T = I_n$. Equation (14) follows since $(SU_{C,\rho})^{+}SU_{C,\rho} = I_\rho$ and $\Omega = (SU_{C,\rho})^{+} - (SU_{C,\rho})^T$ in claim 2.

At last, combined claim 2-4, with probability 0.9, we have that

$$
\begin{aligned}
&\|T - C\tilde{X}\|_F \leq \|T - C_\rho \tilde{X}\|_F \\
&\leq \|U_{C,\rho}^{\perp}(U_{C,\rho}^{\perp})^T T\|_F + \|U_{C,\rho}(SU_{C,\rho})^T SU_{C,\rho}^{\perp}(U_{C,\rho}^{\perp})^T T\|_F + \|U_{C,\rho}\Omega SU_{C,\rho}^{\perp}U_{C,\rho}^{\perp}T\|_F \\
&\leq \|U_{C,\rho}^{\perp}(U_{C,\rho}^{\perp})^T T\|_F + \|U_{C,\rho}^T S^T SU_{C,\rho}^{\perp}(U_{C,\rho}^{\perp})^T T\|_F + \|\Omega\|_2 \|SU_{C,\rho}^{\perp}(U_{C,\rho}^{\perp})^T T\|_F, \\
&\leq (1 + \sqrt{\rho r} + \sqrt{r})\|U_{C,\rho}^{\perp}(U_{C,\rho}^{\perp})^T T\|_F \\
&= (1 + \sqrt{\rho r} + \sqrt{r})\|T - C_\rho C_\rho^{+} T\|_F \\
&\leq (1 + \sqrt{\rho r} + \sqrt{r})\|T - CC^{+}T\|_F
\end{aligned}
$$

At last, let $\varphi = \sqrt{\rho r} + \sqrt{r}$. □

At last, prove the lemma 3.

Proof. Let S_C be the diagonal matrix that records the sampled columns, where if column i is sampled, $(S_C)_{i,i} = 1$, otherwise, $(S_C)_{i,i} = 0$. Similarly, S_R is the diagonal matrix that records the sampled rows. According to claim 5, given $c = O(\rho \cdot \log(\rho))$, we have

$$
\begin{aligned}
\mathsf{Pr}(\|T - CUR\|_F &= \|T - C(S_R C)^{+} S_R T\|_F \\
&\leq (1 + \varphi_1)\|T - CC^{+}T\|_F) \geq 0.9,
\end{aligned} \tag{15}
$$

where $\varphi_1 = \sqrt{\rho r} + \sqrt{r}$, C^{+} is the Moore–Penrose pseudo-inverse of C and $\|T - CC^{+}T\|$ is the projection of T on the subspace spanned by the columns of C.

According to claim 5, given $r = O(\rho \cdot \log(\rho))$ and T_ρ is the optimal ρ-rank approximation of T, $\varphi = \sqrt{\rho c} + \sqrt{c}$, we have

$$
\begin{aligned}
\mathsf{Pr}(\|T - CC^{+}T\|_F &= \|T - TS_C(TS_C)^{+}T\|_F \\
&\leq \|T - TS_C(T_\rho S_C)^{+}T_\rho\|_F \\
&\leq (1 + \varphi_2)\|T - TT_\rho^{+}T_\rho\|_F \\
&= (1 + \varphi_2)\|T - T_\rho\|_F \geq 0.9
\end{aligned} \tag{16}
$$

Combining Equations 15 and 16, we have

$$
\mathsf{Pr}(\|T - CUR\|_F \leq (1 + \varphi_1)(1 + \varphi_2)\|T - T_\rho\|_F) \geq 0.81.
$$

Because $c = r = O(\rho \cdot \log(\rho))$, we can set $c = r$. Therefore, we can set $\phi = 2\varphi + \varphi^2$ and $\varphi = \sqrt{\rho r} + \sqrt{r}$, Equation 1 can be derived, and thus the proof is complete. □

Lemma 5: According to algorithm T^2-Approx, we have

$$
\|T^2 - X \cdot Y\|_F \leq \sqrt{(n - c)}\|T\|_F^2.
$$

Proof. Consider $T^2 = T^{(L)} \cdot T^{(R)}$. Let S be a diagonal matrix, where if column i of $T^{(L)}$ and row i of $T^{(R)}$ are sampled, $S_{i,i} = 1$, otherwise, $S_{i,i} = 0$. Then, we have $X = T \cdot S$ and $Y = S^T \cdot T$.

$$\begin{aligned}
\|T^{(L)}T^{(R)} - X \cdot Y\|_F &= \|T^{(L)}T^{(R)} - T^{(L)}SS^TT^{(R)}\|_F \\
&\le \|T^{(L)}\|_F\|T^{(R)}\|_F\|I - SS^T\|_F \\
&= \sqrt{n-c}\|T^{(L)}\|_F\|T^{(R)}\|_F \\
&= \sqrt{(n-c)}\|T\|_F^2
\end{aligned}$$

□

References

[1] Drineas, P., Mahoney, M. W., and Muthukrishnan, S. Relative-error CUR matrix decompositions. *SIAM J. Matrix Anal. Appl. 30* (2008), 844–881.

[2] Golub, G. H., and Van Loan, C. F. *Matrix computations.* JHU press, 2013.

[3] Wilhelmsen, D. R. A markov inequality in several dimensions. *J. Approx. Theory 11* (1974), 216–220.