Spectral Gap Error Bounds for Improving CUR Matrix Decomposition and the Nyström Method: Supplementary Material

1 Preliminaries

First we prove a useful theorem, which is similar to [3] theorem 3.5.

Theorem S.1. *Let Q_c be an $m \times c$ column-orthonormal matrix matrix. Let Q_r be a $n \times r$ column-orthonormal matrix. Let B_k be the rank-k truncated SVD of $Q_c^T A Q_r$. We have:*

$$\min_{rank(B) \leq k, B \in \mathbb{R}^{c \times r}} \left\| A - Q_c B Q_r^T \right\|_F^2 = \left\| A - Q_c B_k Q_r^T \right\|_F^2 \tag{S.1}$$

In addition:

$$\begin{aligned} \left\| A - Q_c B_k Q_r^T \right\|_F^2 \leq \left\| A - A_k \right\|_F^2 + \left\| \left(I - Q_c Q_c^T \right) A_k \right\|_F^2 + \left\| A_k \left(I - Q_r Q_r^T \right) \right\|_F^2 \\ - 2\mathbf{tr} \left(Q_c Q_c^T A_k Q_r Q_r^T \left(A - A_k \right)^T \right) \end{aligned} \tag{S.2}$$

Proof. We start by taking column-orthogonal matrices of dimensions $m \times (m-c)$ and $n \times (n-r)$ labeled $\hat{Q}_c$ and $\hat{Q}_r$, respectively, so that $\begin{pmatrix} Q_c & \hat{Q}_c \end{pmatrix}$ and $\begin{pmatrix} Q_r & \hat{Q}_r \end{pmatrix}$ are both orthogonal matrices. Then, the unitary invariance of the Frobenious norm and orthogonality give

$$\begin{aligned} \left\| A - Q_c B Q_r^T \right\|_F^2 &= \left\| \begin{pmatrix} Q_c^T A Q_r - B & Q_c^T A \hat{Q}_r \\ \hat{Q}_c^T A Q_r & \hat{Q}_c^T A \hat{Q}_r \end{pmatrix} \right\|_F^2 \\ &= \left\| A - Q_c \left(Q_c^T A Q_r \right) Q_r^T \right\|_F^2 + \left\| Q_c^T A Q_r - B \right\|_F^2 \end{aligned}$$

Thus, the last term in the expression above is minimized when $B = B_k$, which gives us (S.1). Since B_k is the minimizer, we can replace it with $Q_c^T A_k Q_r$ to get the inequality

$$
\begin{aligned}
\|A - Q_c B_k Q_r^T\|_F^2 \leq & \|A - Q_c \left(Q_c^T A_k Q_r\right) Q_r^T\|_F^2 \\
&= \|A - Q_c Q_c^T A_k + Q_c Q_c^T A_k - Q_c \left(Q_c^T A_k Q_r\right) Q_r^T\|_F^2 \\
&= \|A - A_k + A_k - Q_c Q_c^T A_k\|_F^2 + \|Q_c Q_c^T \left(A_k - A_k Q_r Q_r^T\right)\|_F^2 \\
&\quad + 2\mathbf{tr}\left(\left(A - A_k\right)^T Q_c Q_c^T A_k \left(I - Q_r Q_r^T\right)\right) \\
&= \|A - A_k\|_F^2 + 2\mathbf{tr}\left((A - A_k) A_k^T (I - Q_c Q_c^T)\right) \\
&\quad + \left\|\left(I - Q_c Q_c^T\right) A_k\right\|_F^2 + \|Q_c Q_c^T \left(A_k - A_k Q_r Q_r^T\right)\|_F^2 \\
&\quad + 2\mathbf{tr}\left(\left(A - A_k\right)^T Q_c Q_c^T A_k \left(I - Q_r Q_r^T\right)\right) \\
&\leq \|A - A_k\|_F^2 + \|\left(I - Q_c Q_c^T\right) A_k\|_F^2 + \|A_k \left(I - Q_r Q_r^T\right)\|_F^2 \\
&\quad - 2\mathbf{tr}\left(\left(A - A_k\right)^T Q_c Q_c^T A_k \left(Q_r Q_r^T\right)\right),
\end{aligned}
$$

which is (S.2). In the last inequality, we have used once the fact that $Q_c Q_c^T$ is an orthogonal projection and twice the fact that $(A - A_k)A_k^T = 0$ via the SVD. □

2 Deterministic Analysis

We begin with some notes about partitioning A by columns and rows. Let $\Pi_c \in \mathbb{R}^{n \times c}$ and $\Pi_r \in \mathbb{R}^{m \times r}$ be matrices that represent the column and row choices, respectively, of our algorithm such that $\begin{pmatrix}\Pi_c & \Pi_c^\perp\end{pmatrix} \in \mathbb{R}^{n \times n}$ and $\begin{pmatrix}\Pi_r & \Pi_r^\perp\end{pmatrix} \in \mathbb{R}^{m \times m}$ are a permutation matrices.

$$
\begin{aligned}
\begin{pmatrix}\Pi_r & \Pi_r^\perp\end{pmatrix}^T A \begin{pmatrix}\Pi_c & \Pi_c^\perp\end{pmatrix} &= U \Sigma V^T \begin{pmatrix}\Pi_c & \Pi_c^\perp\end{pmatrix} \\
&= \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} \begin{pmatrix} \Sigma_1^{(T)} & 0 \\ 0 & \Sigma_1^{(O)} \end{pmatrix} & 0 \\ 0 & \Sigma_2 \end{pmatrix} \begin{pmatrix} V_{11}^T & V_{21}^T \\ V_{12}^T & V_{22}^T \end{pmatrix}
\end{aligned}
$$

From this point on, we refer to

$$
\begin{aligned}
\Omega &\stackrel{def}{=} \begin{pmatrix} \Omega_1 \\ \Omega_2 \end{pmatrix} \stackrel{def}{=} \begin{pmatrix} V_{11}^T \\ V_{12}^T \end{pmatrix} \\
\Psi &\stackrel{def}{=} \begin{pmatrix} \Psi_1 \\ \Psi_2 \end{pmatrix} \stackrel{def}{=} \begin{pmatrix} U_{11}^T \\ U_{12}^T \end{pmatrix}
\end{aligned}
$$

We change notation at this point because these principles go far beyond column and row selection. For example, if either Π_c or Π_r were an iid Gaussian matrix, the following results will still hold.

By definition, the matrix $C \in \mathbb{R}^{m\times c}$ produced by our algorithm is $A\Pi_c$.

$$C = A\Pi_c = U \begin{pmatrix} \begin{pmatrix} \Sigma_1^{(T)} & 0 \\ 0 & \Sigma_1^{(O)} \end{pmatrix} & 0 \\ 0 & \Sigma_2 \end{pmatrix} \begin{pmatrix} \Omega_1 \\ \Omega_2 \end{pmatrix}$$

Now, we are interested in the matrix $CX = CC^\dagger A$ In order to get a grip on the orthogonal projector $CC^\dagger$, we will study the column space of C via post-multiplying by a judiciously chosen square invertible matrix $Y_c \in \mathbb{R}^{c\times c}$ (cf. [3].) This may change the matrix, but it preserves the column space.

$$\begin{aligned} CY_c &:= C \left[\begin{array}{c:c} \Omega_1^\dagger \begin{pmatrix} \Sigma_1^{(T)} & 0 \\ 0 & \Sigma_1^{(O)} \end{pmatrix}^{-1} & Z_c \end{array} \right] \\ &= U \begin{pmatrix} \begin{pmatrix} \Sigma_1^{(T)} & 0 \\ 0 & \Sigma_1^{(O)} \end{pmatrix} & 0 \\ 0 & \Sigma_2 \end{pmatrix} \begin{pmatrix} \Omega_1 \\ \Omega_2 \end{pmatrix} \left[\begin{array}{c:c} (\Omega_1)^\dagger \begin{pmatrix} \Sigma_1^{(T)} & 0 \\ 0 & \Sigma_1^{(O)} \end{pmatrix}^{-1} & Z_c \end{array} \right] \\ &= U \begin{pmatrix} I_k & 0 & 0 \\ 0 & I_{p-k} & 0 \\ H_1 & H_2 & H_3 \end{pmatrix} \end{aligned}$$

where we assume that $\Omega_1 \in \mathbb{R}^{c\times p}$ is full rank and $Z_c \in \mathbb{R}^{c\times(c-p)}$ is a matrix such that $\Omega_1 Z_c = 0$. This gives us that

$$H_1 = \Sigma_2\Omega_2\Omega_1^\dagger \begin{pmatrix} \left(\Sigma_1^{(T)}\right)^{-1} \\ 0 \end{pmatrix},\ H_2 = \Sigma_2\Omega_2\Omega_1^\dagger \begin{pmatrix} 0 \\ \left(\Sigma_1^{(O)}\right)^{-1} \end{pmatrix},\ H_3 = \Sigma_2\Omega_2 Z_c$$

By the same procedure we can select rows from A to form $R = \Pi_r^T A$. As before, there is an invertible matrix $Y_r \in \mathbb{R}^{r\times r}$ such that

$$Y_r R = \begin{pmatrix} I_k & 0 & 0 \\ 0 & I_{p-k} & 0 \\ G_1 & G_2 & G_3 \end{pmatrix}^T V^T$$

where

$$G_1 = \Sigma_2\Psi_2\Psi_1^\dagger \begin{pmatrix} \left(\Sigma_1^{(T)}\right)^{-1} \\ 0 \end{pmatrix},\ G_2 = \Sigma_2\Psi_2\Psi_1^\dagger \begin{pmatrix} 0 \\ \left(\Sigma_1^{(O)}\right)^{-1} \end{pmatrix},\ G_3 = \Sigma_2\Psi_2 Z_r$$

Following [3], we are interested in upper bounds on $||H_1||_2$ and $||G_1||_2$.

$$||H_1||_2 \le \frac{\sigma_{p+1}}{\sigma_k} \left|\left|\Omega_2\Omega_1^\dagger\right|\right|_2, \text{ and } \left|\left|\left(I + H_1^T H_1\right)^{-1/2}\right|\right|_2 \ge \frac{1}{\sqrt{1 + \left(\frac{\sigma_{p+1}}{\sigma_k}\right)^2 \left|\left|\Omega_2\Omega_1^\dagger\right|\right|_2^2}}$$

$$||G_1||_2 \le \frac{\sigma_{p+1}}{\sigma_k} \left|\left|\Psi_2\Psi_1^\dagger\right|\right|_2, \text{ and } \left|\left|\left(I + G_1^T G_1\right)^{-1/2}\right|\right|_2 \ge \frac{1}{\sqrt{1 + \left(\frac{\sigma_{p+1}}{\sigma_k}\right)^2 \left|\left|\Psi_2\Psi_1^\dagger\right|\right|_2^2}}.$$

To develop lower bounds on computed singular values, let

$$U\begin{pmatrix} I_k & 0 & 0 \\ 0 & I_{p-k} & 0 \\ H_1 & H_2 & H_3 \end{pmatrix} =: \widehat{Q}\widehat{R} =: \begin{pmatrix} \widehat{Q}_1 & \widehat{Q}_2 & \widehat{Q}_3 \end{pmatrix}\begin{pmatrix} \widehat{R}_{11} & \widehat{R}_{12} & \widehat{R}_{13} \\ 0 & \widehat{R}_{22} & \widehat{R}_{23} \\ 0 & 0 & \widehat{R}_{33} \end{pmatrix}, \quad \text{(S.3)}$$

$$V\begin{pmatrix} I_k & 0 & 0 \\ 0 & I_{p-k} & 0 \\ G_1 & G_2 & G_3 \end{pmatrix} =: \widetilde{Q}\widetilde{R} =: \begin{pmatrix} \widetilde{Q}_1 & \widetilde{Q}_2 & \widetilde{Q}_3 \end{pmatrix}\begin{pmatrix} \widetilde{R}_{11} & \widetilde{R}_{12} & \widetilde{R}_{13} \\ 0 & \widetilde{R}_{22} & \widetilde{R}_{23} \\ 0 & 0 & \widetilde{R}_{33} \end{pmatrix}. \quad \text{(S.4)}$$

It follows from (cf. [3]) that

$$Q_c Q_c^T = \widehat{Q}\widehat{Q}^T, \; Q_r Q_r^T = \widetilde{Q}\widetilde{Q}^T.$$

Consider the first k columns of the above expression, i.e.

$$U\begin{pmatrix} I \\ 0 \\ H_1 \end{pmatrix} = \widehat{Q}_1\widehat{R}_{11}$$

Since $R_{11}^T R_{11} = I + H_1^T H_1$, the polar decomposition tells us that R_{11} can be written in the form

$$\widehat{R}_{11} = W_c\left(I + H_1^T H_1\right)^{1/2}$$

for some orthogonal matrix $W_c \in \mathbb{R}^{k\times k}$. Thus, we can write

$$\widehat{Q}_1 = U\begin{pmatrix} I \\ 0 \\ H_1 \end{pmatrix}\left(I + H_1^T H_1\right)^{-1/2} W_c^T$$

By the same reasoning, we also have

$$\widetilde{Q}_1 = V\begin{pmatrix} I \\ 0 \\ G_1 \end{pmatrix}\left(I + G_1^T G_1\right)^{-1/2} W_r^T$$

for some orthogonal matrix $W_r \in \mathbb{R}^{k\times k}$. Next, by the interlacing theorem for singular values, we have

$$\begin{aligned} \sigma_k\left(CUR\right) &= \sigma_k\left(\widehat{Q}^T A \widetilde{Q}\right) \\ &\geq \sigma_k\left(\widehat{Q}_1^T A \widetilde{Q}_1\right) \\ &= \sigma_k\left(\left(I + H_1^T H_1\right)^{-1/2}\left(\Sigma_1 + H_1^T \Sigma_3 G_1\right)\left(I + G_1 G_1^T\right)^{-1/2}\right) \\ &\geq \frac{\sigma_k - \sigma_{p+1}\left(\frac{\sigma_{p+1}}{\sigma_k}\right)^2 \left\|\Omega_2\Omega_1^\dagger\right\|_2 \left\|\Psi_2\Psi_1^\dagger\right\|_2}{\sqrt{1 + \left(\frac{\sigma_{p+1}}{\sigma_k}\right)^2 \left\|\Omega_2\Omega_1^\dagger\right\|_2^2}\sqrt{1 + \left(\frac{\sigma_{p+1}}{\sigma_k}\right)^2 \left\|\Psi_2\Psi_1^\dagger\right\|_2^2}}. \end{aligned}$$

Now we bound the Frobenius-norm using Theorem S.1. By Theorem 4.4 of [3],

$$\left\|\left(I-\widehat{Q}\widehat{Q}^T\right)A_k\right\|_F^2 \le k\sigma_{p+1}^2\left\|\Omega_2\Omega_1^\dagger\right\|_2^2,\ \left\|A_k\left(I-\widetilde{Q}\widetilde{Q}^T\right)\right\|_F^2 \le k\sigma_{p+1}^2\left\|\Psi_2\Psi_1^\dagger\right\|_2^2. \tag{S.5}$$

To bound the last term in equation (S.2), we re-partition equations (S.3) and (S.4)

$$\begin{pmatrix} I_k & 0 & 0 \\ 0 & I_{p-k} & 0 \\ H_1 & H_2 & H_3 \end{pmatrix} \stackrel{def}{=} \begin{pmatrix} I_p & 0 \\ \widehat{H}_1 & \widehat{H}_2 \end{pmatrix},\ \begin{pmatrix} I_k & 0 & 0 \\ 0 & I_{p-k} & 0 \\ G_1 & G_2 & G_3 \end{pmatrix} \stackrel{def}{=} \begin{pmatrix} I_p & 0 \\ \widehat{G}_1 & \widehat{G}_2 \end{pmatrix}$$

and define accordingly

$$U^T\widehat{Q} = \begin{pmatrix} \widehat{Q}_{11} & \widehat{Q}_{12} \\ \widehat{Q}_{21} & \widehat{Q}_{22} \end{pmatrix},\ \widehat{\Sigma}_2 = \mathbf{diag}(\Sigma_1^{(O)}, \Sigma_2),\ V^T\widetilde{Q} = \begin{pmatrix} \widetilde{Q}_{11} & \widetilde{Q}_{12} \\ \widetilde{Q}_{21} & \widetilde{Q}_{22} \end{pmatrix}.$$

Now we rewrite

$$\begin{aligned}
&\mathbf{tr}\left(Q_cQ_c^TA_kQ_rQ_r^T\left(A-A_k\right)^T\right) = \mathbf{tr}\left(\widehat{Q}\widehat{Q}^TA_k\widetilde{Q}\widetilde{Q}^T\left(A-A_k\right)^T\right)\\
&= \mathbf{tr}\left(\begin{pmatrix} \widehat{Q}_{11} & \widehat{Q}_{12} \\ \widehat{Q}_{21} & \widehat{Q}_{22} \end{pmatrix}\begin{pmatrix} \widehat{Q}_{11} & \widehat{Q}_{12} \\ \widehat{Q}_{21} & \widehat{Q}_{22} \end{pmatrix}^T\begin{pmatrix} \Sigma_1^{(T)} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \widetilde{Q}_{11} & \widetilde{Q}_{12} \\ \widetilde{Q}_{21} & \widetilde{Q}_{22} \end{pmatrix}\begin{pmatrix} \widetilde{Q}_{11} & \widetilde{Q}_{12} \\ \widetilde{Q}_{21} & \widetilde{Q}_{22} \end{pmatrix}^T\begin{pmatrix} 0 & 0 \\ 0 & \widehat{\Sigma}_2 \end{pmatrix}\right)\\
&= \mathbf{tr}\left(\begin{pmatrix} 0 & 0 \\ \widehat{Q}_{21} & \widehat{Q}_{22} \end{pmatrix}\begin{pmatrix} \widehat{Q}_{11} & \widehat{Q}_{12} \\ 0 & 0 \end{pmatrix}^T\begin{pmatrix} \Sigma_1^{(T)} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \widetilde{Q}_{11} & \widetilde{Q}_{12} \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ \widetilde{Q}_{21} & \widetilde{Q}_{22} \end{pmatrix}^T\begin{pmatrix} 0 & 0 \\ 0 & \widehat{\Sigma}_2 \end{pmatrix}\right)\\
&= \mathbf{tr}\left(\left(\left(\widehat{Q}_{21}\ \widehat{Q}_{22}\right)\left(\widehat{Q}_{11}\ \widehat{Q}_{12}\right)^T\Sigma_1^{(T)}\right)\left(\left(\widetilde{Q}_{11}\ \widetilde{Q}_{12}\right)\left(\widetilde{Q}_{21}\ \widetilde{Q}_{22}\right)^T\widehat{\Sigma}_2\right)\right).
\end{aligned}$$

To continue, we notice that

$$\begin{aligned}
&\left\|\left(\widehat{Q}_{21}\ \widehat{Q}_{22}\right)\left(\widehat{Q}_{11}\ \widehat{Q}_{12}\right)^T\Sigma_1^{(T)}\right\|_F\\
&\le \left\|\begin{pmatrix} I-\left(\widehat{Q}_{11}\ \widehat{Q}_{12}\right)\left(\widehat{Q}_{11}\ \widehat{Q}_{12}\right)^T & -\left(\widehat{Q}_{11}\ \widehat{Q}_{12}\right)\left(\widehat{Q}_{21}\ \widehat{Q}_{22}\right)^T \\ -\left(\widehat{Q}_{21}\ \widehat{Q}_{22}\right)\left(\widehat{Q}_{11}\ \widehat{Q}_{12}\right)^T & I-\left(\widehat{Q}_{21}\ \widehat{Q}_{22}\right)\left(\widehat{Q}_{21}\ \widehat{Q}_{22}\right)^T \end{pmatrix}\begin{pmatrix} \Sigma_1^{(T)} \\ 0 \end{pmatrix}\right\|_F\\
&= \left\|\left(I-\begin{pmatrix} \widehat{Q}_{11} & \widehat{Q}_{12} \\ \widehat{Q}_{21} & \widehat{Q}_{22} \end{pmatrix}\begin{pmatrix} \widehat{Q}_{11} & \widehat{Q}_{12} \\ \widehat{Q}_{21} & \widehat{Q}_{22} \end{pmatrix}^T\right)\begin{pmatrix} \Sigma_1^{(T)} \\ 0 \end{pmatrix}\right\|_F\\
&= \left\|\left(I-\widehat{Q}\widehat{Q}^T\right)A_k\right\|_F \le \sqrt{k}\sigma_{p+1}\left\|\Omega_2\Omega_1^\dagger\right\|_2.
\end{aligned}$$

Similarly, we can derive

$$\left\|\left(\widetilde{Q}_{11}\ \widetilde{Q}_{12}\right)\left(\widetilde{Q}_{21}\ \widetilde{Q}_{22}\right)^T\widehat{\Sigma}_2\right\|_F \le \sqrt{k}\sigma_{p+1}\left\|\Psi_2\Psi_1^\dagger\right\|_2.$$

By the Cauchy-Schwartz inequality, we combine these two upper bounds to get

$$\left|\mathbf{tr}\left(Q_c Q_c^T A_k Q_r Q_r^T (A - A_k)^T\right)\right| \leq k\sigma_{p+1}^2 \left|\left|\Omega_2 \Omega_1^\dagger\right|\right|_2 \left|\left|\Psi_2 \Psi_1^\dagger\right|\right|_2.$$

Plugging this and upper bounds (S.5) into equation (S.2), we have

$$\begin{aligned} &||A - CUR||_F^2 \leq ||A - (CUR)_k||_F^2 \\ &\leq \sum_{j=k+1}^{\rho} \sigma_j^2 + k\sigma_{p+1}^2 \left(\left|\left|\Omega_2 \Omega_1^\dagger\right|\right|_2 + \left|\left|\Psi_2 \Psi_1^\dagger\right|\right|_2\right)^2 \end{aligned}$$

as desired. The spectral norm bound is simply the consequence of Theorem 3.5 of [3].

3 Proof of Deterministic Algorithm

Apply equation (7) to $\mathbf{U}_1$ and $\mathbf{V}_1^T$, we have

$$\begin{aligned} \left|\left|\mathbf{\Omega}_1^\dagger\right|\right|_2 &\leq \frac{\sqrt{c} - \sqrt{p}}{\sqrt{(\sqrt{n-p} + \sqrt{c})^2 + (\sqrt{c} - \sqrt{p})^2}}, \\ \left|\left|\mathbf{\Psi}_1^\dagger\right|\right|_2 &\leq \frac{\sqrt{r} - \sqrt{p}}{\sqrt{(\sqrt{m-p} + \sqrt{r})^2 + (\sqrt{r} - \sqrt{p})^2}}. \end{aligned}$$

Further notice that $\mathbf{\Omega}_2$ and $\mathbf{\Psi}_2$ are submatrices of $\mathbf{U}_2^T$ and $\mathbf{V}_2^T$, respectively, and so their singular values are at most 1.

4 Proof of Uniform Sampling Algorithm

By [2], it is shown that

$$\left|\left|\mathbf{\Omega}_1^\dagger\right|\right|_2^2 \leq \frac{n}{(1-\epsilon)\, c}$$

By symmetry, we have

$$\left|\left|\mathbf{\Psi}_1^\dagger\right|\right|_2^2 \leq \frac{m}{(1-\epsilon)\, r}.$$

5 Proof of Leverage Score Sampling Algorithm

By the proof of lemma 1 from [1], we have that with probability at least 0.9,

$$1 - \sigma_{\min}^2(\mathbf{\Omega}_1) \leq \epsilon$$

Therefore, we have

$$\left|\left|\mathbf{\Omega}_1^{\dagger}\right|\right|_2^2 \leq \frac{1}{1-\epsilon}$$

By symmetry, we have

$$\left|\left|\mathbf{\Psi}_1^{\dagger}\right|\right|_2^2 \leq \frac{1}{1-\epsilon}.$$

References

[1] P. Drineas, M. W. Mahoney, and S. Muthukrishnan. Relative-error CUR matrix decompositions. *SIAM Journal on Matrix Analysis and Applications*, 30(2):844–881, 2008.

[2] A. Gittens. The spectral norm error of the naive Nyström extension. *arXiv preprint arXiv:1110.5305*, 2011.

[3] M. Gu. Subspace iteration randomization and singular value problems. *arXiv preprint arXiv:1408.2208*, 2014.