1. We will prove a structural inequality that can be used as a starting point in many RandNLA algorithms. More specifically, we will prove the following lemma.

**Lemma 1** Given $A \in \mathbb{R}^{m \times n}$, let $V_k \in \mathbb{R}^{n \times k}$ the matrix of the top $k$ right singular vectors of $A$. Let $Z \in \mathbb{R}^{n \times r}$ $(r \geq k)$ be any matrix such that $V_k^T Z$ has full rank. Then, for any unitarily invariant norm $\xi$, we have that

$$\| A - (A Z) (A Z)^\dagger A \|_\xi \leq \| A - A_k \|_\xi + \left\| (A - A_k) Z (V_k^T Z)^\dagger A_k \right\|_\xi.$$

(1)

Recall that $A_k = U_k \Sigma_k V_k^T$ denotes the best rank-$k$ approximation to $A$, where $U_k \in \mathbb{R}^{m \times k}$ is the matrix of the top $k$ left singular vectors of $A$ and $\Sigma_k \in \mathbb{R}^{k \times k}$ is the diagonal matrix of the top $k$ singular values of $A$.

We will actually only prove the above lemma for the special case where $\xi = 2$ and $\xi = F$; all the following problems should be proven only for $\xi = 2$ and $\xi = F$.

1. Prove that $(A Z)^\dagger A = \arg \min_{X \in \mathbb{R}^{r \times n}} \| A - (A Z) X \|_\xi$.

2. Use the above result to prove that

$$\| A - (A Z)(A Z)^\dagger A \|_\xi \leq \| A - A Z (A_k Z)^\dagger A_k \|_\xi.$$

3. Use the triangle inequality to prove that

$$\| A - (A Z)(A Z)^\dagger A \|_\xi \leq \| A - A_k \|_\xi + \left\| (A - A_k) Z (A_k Z)^\dagger A_k \right\|_\xi + \left\| (A - A_k) Z (A_k Z)^\dagger A_k \right\|_\xi.$$

4. Prove that $V_k^T Z \left( V_k^T Z \right)^\dagger = I_k$.

5. Prove that $(A_k Z)^\dagger = \left( V_k^T Z \right)^\dagger \Sigma_k^{-1} U_k^T$.

6. Prove that $\| A_k - A_k Z (A_k Z)^\dagger A_k \|_\xi = 0$.

7. Combine all the above to prove eqn (1).

2. We will now leverage the structural inequality of Lemma 1 to prove that the algorithm of slides 104-106 achieves a constant factor approximation for the CX decomposition.

1. Let $Z$ (the matrix of Lemma 1) be a sampling-and-rescaling matrix $S$ indicating the columns of $A$ that were sampled using the algorithm of slides 104-105 with the sampling probabilities of slide 106 (basically, the column leverage scores). Prove that the matrix $V_k^T S$ has full rank, with constant probability.

2. Prove that for any two matrices $X$ and $Y$ of appropriate dimensions, $\| X Y \|_F \leq \| X \|_F \| Y \|_2$ (similarly, $\| X Y \|_F \leq \| X \|_2 \| Y \|_F$). This property is known as strong submultiplicativity.
3. Apply the above property and eqn. (1) to get:

\[
\|A - (AS)(AS)^\dagger A\|_F \leq \|A - A_k\|_F + \|(A - A_k)S\|_F \| (V_k^TS)^\dagger \|_2.
\]

4. Prove a lower bound for the smallest singular value of the matrix \(V_k^TS\) (this is similar to part 1 of this problem); the bound will hold with constant probability. What does that imply for the spectral norm of \((V_k^TS)^\dagger\)?

5. Prove that the expectation of \(\|(A - A_k)S\|_F^2\) is equal to \(\|A - A_k\|_F^2\). Combine this with Markov’s inequality to get a bound (with constant probability) for \(\|(A - A_k)S\|_F\).

6. Combine all the above to prove that the algorithm of slide 104-105 with the sampling probabilities of slide 106 returns a CX decomposition that achieves a constant factor approximation.

Notice that we constructed a CX decomposition that rescales the columns of \(A\) that are included in the matrix \(C\). However, it is easy to prove that

\[
\|A - (AS)(AS)^\dagger A\|_F = \|A - (AS')(AS')^\dagger A\|_F,
\]

where \(S'\) is simply \(S\) without the rescaling (e.g., the entries of \(S'\) are either zero or one, go to slide 16, Remark 2).

Finally, we note that we can also get a relative error approximation that holds with very high probability for the same algorithm with a more careful, albeit somewhat longer, analysis.