Assignment 4

Randomization in Numerical Linear Algebra (PCMI)

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

$$\left\|\mathbf{A} - (\mathbf{AZ})(\mathbf{AZ})^{\dagger}\mathbf{A}\right\|_{\xi} \leq \|\mathbf{A} - \mathbf{A}_k\|_{\xi} + \left\|(\mathbf{A} - \mathbf{A}_k)\,\mathbf{Z}(\mathbf{V}_k^T\mathbf{Z})^{\dagger}\right\|_{\xi}. \qquad (1)$$

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

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
$$(\mathbf{AZ})^{\dagger}\mathbf{A} = \arg\min_{\mathbf{X}\in\mathbb{R}^{r\times n}} \|\mathbf{A} - (\mathbf{AZ})\,\mathbf{X}\|_{\xi}.$$

2. Use the above result to prove that
$$\left\|\mathbf{A} - (\mathbf{AZ})(\mathbf{AZ})^{\dagger}\mathbf{A}\right\|_{\xi} \leq \left\|\mathbf{A} - \mathbf{AZ}\,(\mathbf{A}_k\mathbf{Z})^{\dagger}\,\mathbf{A}_k\right\|_{\xi}.$$

3. Use the triangle inequality to prove that
$$\begin{aligned}\left\|\mathbf{A} - (\mathbf{AZ})(\mathbf{AZ})^{\dagger}\mathbf{A}\right\|_{\xi} &\leq \|\mathbf{A} - \mathbf{A}_k\|_{\xi} + \left\|\mathbf{A}_k - \mathbf{A}_k\mathbf{Z}(\mathbf{A}_k\mathbf{Z})^{\dagger}\mathbf{A}_k\right\|_{\xi} \\ &+ \left\|(\mathbf{A} - \mathbf{A}_k)\,\mathbf{Z}(\mathbf{A}_k\mathbf{Z})^{\dagger}\mathbf{A}_k\right\|_{\xi}.\end{aligned}$$

4. Prove that $\mathbf{V}_k^T\mathbf{Z}\left(\mathbf{V}_k^T\mathbf{Z}\right)^{\dagger} = \mathbf{I}_k$.

5. Prove that $(\mathbf{A}_k\mathbf{Z})^{\dagger} = \left(\mathbf{V}_k^T\mathbf{Z}\right)^{\dagger}\mathbf{\Sigma}_k^{-1}\mathbf{U}_k^T$.

6. Prove that $\left\|\mathbf{A}_k - \mathbf{A}_k\mathbf{Z}(\mathbf{A}_k\mathbf{Z})^{\dagger}\mathbf{A}_k\right\|_{\xi} = 0$.

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 $\mathbf{Z}$ (the matrix of Lemma 1) be a sampling-and-rescaling matrix $\mathbf{S}$ indicating the columns of $\mathbf{A}$ 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 $\mathbf{V}_k^T\mathbf{S}$ has full rank, with constant probability.

2. Prove that for any two matrices $\mathbf{X}$ and $\mathbf{Y}$ of appropriate dimensions, $\|\mathbf{XY}\|_F \leq \|\mathbf{X}\|_F\,\|\mathbf{Y}\|_2$ (similarly, $\|\mathbf{XY}\|_F \leq \|\mathbf{X}\|_2\,\|\mathbf{Y}\|_F$). This property is known as strong submultiplicativity.

3. Apply the above property and eqn. (1) to get:

$$\left\|\mathbf{A}-(\mathbf{AS})(\mathbf{AS})^{\dagger}\mathbf{A}\right\|_F \leq \|\mathbf{A}-\mathbf{A}_k\|_F + \|(\mathbf{A}-\mathbf{A}_k)\,\mathbf{S}\|_F \left\|(\mathbf{V}_k^T\mathbf{S})^{\dagger}\right\|_2 .$$

4. Prove a lower bound for the smallest singular value of the matrix $\mathbf{V}_k^T\mathbf{S}$ (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 $(\mathbf{V}_k^T\mathbf{S})^{\dagger}$?

5. Prove that the expectation of $\|(\mathbf{A}-\mathbf{A}_k)\,\mathbf{S}\|_F^2$ is equal to $\|\mathbf{A}-\mathbf{A}_k\|_F^2$. Combine this with Markov's inequality to get a bound (with constant probability) for $\|(\mathbf{A}-\mathbf{A}_k)\,\mathbf{S}\|_F$.

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 $\mathbf{A}$ that are included in the matrix $\mathbf{C}$. However, it is easy to prove that

$$\left\|\mathbf{A}-(\mathbf{AS})(\mathbf{AS})^{\dagger}\mathbf{A}\right\|_F = \left\|\mathbf{A}-(\mathbf{AS}')(\mathbf{AS}')^{\dagger}\mathbf{A}\right\|_F ,$$

where $\mathbf{S}'$ is simply $\mathbf{S}$ without the rescaling (e.g., the entries of $\mathbf{S}'$ 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.