Solutions for Problem Set 2

Problem 1: Online Leverage Scores

Part 1
Let $A = U \Sigma V^T$ be the singular value decomposition with $U \in \mathbb{R}^{n \times d}$, $\Sigma \in \mathbb{R}^{d \times d}$ and $V \in \mathbb{R}^{d \times d}$. As the matrix $A$ has rank $d$, we have that all the diagonal entries of $\Sigma$ are nonzero and hence the square matrix $\Sigma V^T$ is invertible. Thus,

$$U = A(\Sigma V^T)^{-1}$$

is an orthonormal basis for the column space of $A$. By definition of the leverage scores,

$$\ell_i = \|U_i\|_2^2 = \|a_i^T(\Sigma V^T)^{-1}\|_2^2 = a_i^T(\Sigma V^T)^{-1}(a_i^T(\Sigma V^T)^{-1})^T.$$  

As $V^TV = I = VV^T$, we obtain that $(V^T)^{-1} = V$ and therefore the above expression simplifies to

$$\ell_i = a_i^T(V\Sigma^{-2}V^T)a_i.$$  

Finally, using the singular value decomposition, we also obtain that $A^TA = V\Sigma^2V^T$ and that $(A^TA)^{-1} = V\Sigma^{-2}V^T$ thereby proving that

$$\ell_i = a_i^T(A^TA)^{-1}a_i.$$  

Part 2
First we note that for any $\lambda > 0$, the matrix $\lambda I + A_iA_i^T$ is positive definite and hence invertible. Letting $A_0 = 0$, we have for any $i \geq 1$,

$$\det(\lambda I + A_iA_i^T) = \det(\lambda I + A_iA_i^T + a_i^Ta_i^T) = \det(\lambda I + A_{i-1}A_{i-1}^T(1 + a_i^T(\lambda I + A_{i-1}A_{i-1}^T)^{-1}a_i))$$

where in the last equality, we used the matrix determinant lemma and the fact that $\lambda I + A_{i-1}A_{i-1}^T$ is invertible. Now, as $\ell_i$ is defined as $\min(1, a_i^T(\lambda I + A_{i-1}A_{i-1}^T)^{-1}a_i)$, we obtain that

$$\det(\lambda I + A_iA_i^T) \geq \det(\lambda I + A_{i-1}A_{i-1}^T)(1 + \ell_i).$$

Now using the fact that $A_n = A$, we obtain using the above inequality repeatedly that

$$\det(\lambda I + AA^T) = \det(\lambda I + A_nA_n^T)$$

$$\geq (1 + \ell_n) \cdots (1 + \ell_1) \det(\lambda I).$$

By definition, $0 \leq \ell_i \leq 1$ for all $i$ and therefore for all $i$, $1 + \ell_i \geq \exp(\ell_i/2)$. Hence, the above inequality implies

$$\det(\lambda I + AA^T) \geq \exp\left(\sum_{i=1}^{n} \ell_i/2\right)\lambda^d.$$
where we also used the fact that $\det(\lambda I_d) = \lambda^d$. Thus,

$$\sum_{i=1}^{n} \ell_i \leq 2 \ln(\lambda^{-d} \det(\lambda I + AA^T)).$$

We now upper bound $\det(\lambda I + AA^T)$. Recall that

$$\det(\lambda I + AA^T) = \prod_{i=1}^{d} \lambda_i(\lambda I + AA^T)$$

where $\lambda_i(\cdot)$ denotes the $i$-th largest eigen value of the matrix. Finally as $\lambda I + AA^T$ is positive semi-definite,

$$\lambda_i(\lambda I + AA^T) \leq \lambda_1(\lambda I + AA^T) = \|\lambda I + AA^T\|_2 \leq \lambda + \|AA^T\|_2 = \lambda + \|A\|_2^2.$$

Thus, $\prod_{i=1}^{d} \lambda_i(\lambda I + AA^T) \leq (\lambda + \|A\|_2^2)^d$ which then implies

$$\sum_{i} \ell_i \leq 2d \ln \left(1 + \|A\|_2^2/\lambda\right).$$
Problem 2: Gap-Dependent Bounds for Low Rank Approximation

Part 1

Let $A = UΣV^T$ be the singular value decomposition of the matrix $A$ with $U, Σ, V ∈ \mathbb{R}^{n×n}$. We obtain that

$$(AA^T)^r A = UΣ^{2r+1}V^T.$$  

Let $y = (AA^T)^r A g = UΣ^{2r+1}V^T g$ where $g$ is the random Gaussian vector. Using the rotational invariance of the $n$ dimensional Gaussian distribution, the vector $g' = V^T g$ is also distributed as a Gaussian vector with coordinates being independent standard normal random variables.

To summarize, we have $y = (AA^T)^r A g = UΣ^{2r+1} g' = \sum_{i=1}^n u_i σ_i^{2r+1} g'_i$. We also have

$$\|u^T A\|^2 = \frac{\|y^T A\|^2}{\|y\|^2} = \frac{\|\sum_{i=1}^n σ_i^{2r+1} g'_i (σ_i v_i^T)\|^2}{\|\sum_{i=1}^n u_i σ_i^{2r+1} g'_i\|^2}.$$  

Here we used the fact that for all $i$, $u_i^T A = σ_i v_i^T$. Now using the fact that $v_i$s are orthonormal and $u_i$s are orthonormal, we get

$$\|u^T A\|^2 = \frac{\sum_{i=1}^n σ_i^{4r+4} (g'_i)^2}{\sum_{i=1}^n σ_i^{4r+2} (g'_i)^2}. $$

We now condition on the following events:

1. $|g'_i| ≥ 1/n^3$ which happens with probability $≥ 1 - 1/n^3$ and
2. $\max_i |g'_i| ≤ \sqrt{10 ln n}$ which happens with probability $≥ 1 - 1/n^3$.

By union bound, the events hold true simultaneously with probability $≥ 1 - 2/n^3$. We now have two cases: (i) $σ_1 ≥ 2σ_2$ and (ii)$σ_1 ≤ 2σ_2$. Consider the first case and assume that the above events hold:

$$\|u^T A\|^2 = \frac{σ_1^{4r+4} (g'_1)^2 [1 + (g'_2/g'_1)^2 (σ_2/σ_1)^{4r+4} + \cdots + (g'_n/g'_1)^2 (σ_n/σ_1)^{4r+4}]}{σ_1^{4r+2} (g'_1)^2 [1 + (g'_2/g'_1)^2 (σ_2/σ_1)^{4r+2} + \cdots + (g'_n/g'_1)^2 (σ_n/σ_1)^{4r+2}]}.$$ 

Using the fact that both the events hold, we have $\max_i (g'_i/g'_1)^2 ≤ 10n^6 ln n$ and therefore,

$$\|u^T A\|^2 ≥ \frac{σ_1^2}{1 + (σ_2/σ_1)^2 (10n^7 ln n)/2^{4r}}$$ 

where we used that $σ_i ≤ σ_1/2$ for all $i ≥ 2$. Now if $r = Ω(\log(n/γ))$,

$$\|u^T A\|^2 ≥ \frac{σ_1^2}{1 + γ(σ_2/σ_1)^2} ≥ σ_1^2 - γσ_2^2.$$ 

Now consider the case of $σ_2 ≥ σ_1/2$. Let $m$ be the largest index such that $σ_m ≥ (1 - γ)σ_1$. For $j > m$, $(σ_j/σ_1)^r ≤ (1 - γ)^r ≤ 1/poly_1(n)$ if $r = Ω(\log(n/γ))$ for a polynomial $poly_1$. Then,

$$\|u^T A\|^2 ≥ \frac{σ_1^2 [1 + (g'_2/g'_1)^2 (σ_2/σ_1)^{4r+4} + \cdots + (g'_m/g'_1)^2 (σ_m/σ_1)^{4r+4}]}{[1 + (g'_2/g'_1)^2 (σ_2/σ_1)^{4r+2} + \cdots + (g'_m/g'_1)^2 (σ_m/σ_1)^{4r+2} + 1/poly_2(n)]}$$

$$≥ \frac{σ_1^2 [1 + (σ_m/σ_1)^2 Θ]}{1 + Θ + 1/poly_2(n)}.$$
where

\[
\Theta := \left( \frac{g_2'/g_1'}{\sigma_2/\sigma_1} \right)^2 + \cdots + \left( \frac{g_m'/g_1'}{\sigma_m/\sigma_1} \right)^2 \Theta^2.
\]

We then have,

\[
\|u^T A\|_2^2 \geq \sigma_1^2 - \sigma_2^2 \left( 1 - \frac{\sigma_m/\sigma_1^2}{\sigma_1^2} + \frac{1}{\text{poly}_2(n)} \right)
\]

\[
\geq \sigma_1^2 - \sigma_2^2 \frac{1}{\text{poly}_2(n)} - (\sigma_1^2 - \sigma_m^2)
\]

\[
\geq \sigma_1^2 - \sigma_2^2 \frac{1}{\text{poly}_2(n)} - 2\gamma \sigma_1^2
\]

\[
\geq \sigma_1^2 - \sigma_2^2 (8\gamma + 4/\text{poly}_2(n)).
\]

Rescaling \(\gamma\) and using the fact that \(\gamma \geq 1/\text{poly}(n)\), we obtain the proof.

**Part 2**

Let \(u\) be the unit vector returned by \(\text{PowerMethod}(A, O((\log n)/\gamma))\). From above, we have that with probability \(\geq 1 - 1/\text{poly}(n)\),

\[
\|u^T A\|_2^2 \geq \sigma_1^2 - \gamma \sigma_2^2.
\]

Let \(v = A^T u\). Now,

\[
\|uv^T - A\|_F^2 = \|uu^T A - A\|_F^2 = \|A\|_F^2 - \|uu^T A\|_F^2.
\]

where the last equality follows from Pythagorean theorem using the fact that columns of \(uu^T A\) are orthogonal to columns of \(uu^T A - A = -(I - uu^T)A\). As \(u\) is a unit vector, we further obtain that \(\|uu^T A\|_F^2 = \|u^T A\|_F^2\). Hence with probability \(\geq 1 - 1/\text{poly}(n)\),

\[
\|uv^T - A\|_F^2 \leq \|A\|_F^2 - (\sigma_1^2 - \gamma \sigma_2^2) = \sum_{i=1}^{n} \sigma_i^2 - \sigma_1^2 + \gamma \sigma_2^2 = \|A - A_1\|_F^2 + \gamma \sigma_2^2.
\]

Now using the assumption that \(\sigma_2^2 \leq \alpha \|A - A_1\|_F^2\), we have

\[
\|uv^T - A\|_F^2 \leq (1 + \alpha \gamma) \|A - A_1\|_F^2.
\]

Setting \(\gamma = \min(1, \epsilon/\alpha)\), then gives that \(uv^T\) is a \(1 + \epsilon\) rank 1 approximation of the matrix \(A\). Note that overall, the algorithm needs \(O(\log n \max(1, \alpha/\epsilon))\) matrix vector products with \(A\) and hence can be performed in \(O(\text{nnz}(A)(1 + \alpha/\epsilon) \log n)\) time.
Problem 3: CUR Decompositions

Part 1

Let \( X^\ast \) be the optimal solution for:

\[
\min_X \|AX - B\|_F
\]

and \( \hat{X} \) be the optimal solution for

\[
\min_X \|SAX - SB\|_F.
\]

Using normal equations, we have the following equalities for any matrix \( X \):

\[
\|AX - B\|_F^2 = \|AX^\ast - B\|_F^2 + \|AX - AX^\ast\|_F^2.
\]

\[
\|SAX - SB\|_F^2 = \|SA\hat{X} - SB\|_F^2 + \|SAX - SA\hat{X}\|_F^2.
\]

Also note that \( AX^\ast - B \), the residual matrix, is a fixed matrix independent of \( S \). Hence, \( E[\|S(AX^\ast - B)\|_F^2] = \|AX^\ast - B\|_F^2 \) and using Markov inequality, with probability \( \geq 19/20 \),

\[
\|S(AX^\ast - B)\|_F^2 \leq 20\|AX^\ast - B\|_F^2.
\]

Hence, by union bound, with probability \( \geq 9/10 \), the following properties hold true simultaneously:

for all \( x \), \((1/2)\|Ax\|_2 \leq \|S Ax\|_2 \leq 2\|Ax\|_2 \)

and

\[
\|S(AX^\ast - B)\|_F^2 \leq 20\|AX^\ast - B\|_F^2.
\]

We now bound \( \|A\hat{X} - B\|_F^2 \) conditioning on the above two events. First, by using normal equations,

\[
\|A\hat{X} - B\|_F^2 \leq \|AX^\ast - A\hat{X}\|_F^2 + \|AX^\ast - B\|_F^2.
\]

Using the subspace embedding property, we have \( \|Ax\|_2^2 \leq 4\|S Ax\|_2^2 \) for all \( x \) which implies for all matrices \( X \), \( \|AX\|_2^2 \leq 4\|SAX\|_2^2 \). Therefore,

\[
\|A\hat{X} - B\|_F^2 \leq 4\|SAX^\ast - SA\hat{X}\|_F^2 + \|AX^\ast - B\|_F^2.
\]

Again, using the normal equations, we have \( \|SAX^\ast - SA\hat{X}\|_F^2 \leq \|SAX^\ast - SB\|_F^2 \) which implies,

\[
\|A\hat{X} - B\|_F^2 \leq 4\|SAX^\ast - SB\|_F^2 + \|AX^\ast - B\|_F^2.
\]

Finally, using the fact that \( \|SAX^\ast - SB\|_F^2 \leq 20\|AX^\ast - B\|_F^2 \), we obtain

\[
\|A\hat{X} - B\|_F^2 \leq 81\|AX^\ast - B\|_F^2.
\]
Part 2

Let $S$ be the leverage score sampling matrix for $\hat{U}$ that has $O(k \log k)$ rows. As $\hat{U}$ has only $k$ columns, from above, we have with probability $\geq 9/10$,

$$\|\hat{U} \hat{X} - A\|_F^2 \leq O(1) \min_X \|\hat{U} X - A\|_F^2$$

where $\hat{X} = \min_X \|S\hat{U} X - SA\|_F^2$. Further, $\min_X \|\hat{U} X - A\|_F^2 \leq \|\hat{U} \hat{V} - A\|_F^2 = O(1)\|A - A_k\|_F^2$. Thus,

$$\|\hat{U} \hat{X} - A\|_F^2 \leq O(1)\|A - A_k\|_F^2$$

Now, we note that $\hat{X} = (S\hat{U})^{-1} SA$. As $S$ only samples and rescales $O(k \log k)$ rows of $A$, we obtain a $O(k \log k) \times n$ submatrix $R$ of $A$ such that $SA = DR$ for some diagonal matrix $D$ and therefore,

$$\min_{U \in \mathbb{R}^{n \times O(k \log k)}} \|UR - A\|_F^2 \leq \|\hat{U} (S\hat{U})^{-1} DR - A\|_F^2 \leq O(1)\|A - A_k\|_F^2$$

Part 3

From the above part, we obtain a subset of rows $R$ satisfying that span an $O(1)$ bicriteria approximation to the rank $k$ Frobenius norm LRA. Now consider the following problem:

$$\min_{U \in \mathbb{R}^{O(k \log k \times n)}} \|R^T U - A^T\|_F^2.$$

If $S'$ is the leverage score sampling matrix with $O(k \log^2 k)$ rows for $R^T$, again using the part 1, we obtain with probability $\geq 9/10$,

$$\|R^T \hat{U} - A^T\|_F^2 \leq O(1) \min_{U} \|R^T U - A^T\|_F^2 \leq O(1)\|A - A_k\|_F^2$$

where $\hat{U}$ is the solution to

$$\min_{U} \|\hat{S} R^T U - \hat{S} A^T\|_F^2.$$

We again have $\hat{U} = (\hat{S} R^T)^{-1} (\hat{S} A^T)$. Thus,

$$\|R^T (\hat{S} R^T)^{-1} \hat{S} A^T - A^T\|_F^2 \leq O(1)\|A - A_k\|_F^2.$$

Note that $\hat{S} A^T = \hat{D} C^T$ where $C$ is a sub-matrix formed by $O(k \log^2 k)$ columns of $A$ and $\hat{D}$ is some diagonal matrix. Finally, we have

$$\|C (\hat{D})^T ((\hat{S} R^T)^{-1})^T R - A\|_F^2 \leq O(1)\|A - A_k\|_F^2.$$

Thus, with probability $\geq 8/10$, the submatrices $C$ and $R$ satisfy

$$\min_{U \in \mathbb{R}^{O(k \log k) \times O(k \log^2 k)}} \|CUR - A\|_F^2 \leq O(1)\|A - A_k\|_F^2.$$